Natarajan Meghanathan
David C. Wyld (Eds)

Computer Science & Information Technology

8th International Conference on Computer Science, Engineering and
Applications (CCSEA 2018), February 17~18, 2018, Melbourne, Australia.

AIRCC Publishing Corporation

Volume Editors

Natarajan Meghanathan,
Jackson State University, USA
E-mail: nmeghanathan@jsums.edu

David C. Wyld,
Southeastern Louisiana University, USA
E-mail: David.Wyld@selu.edu

Preface

The 8th International Conference on Computer Science, Engineering and Applications (CCSEA 2018) was held in Melbourne, Australia, during February 17~18, 2018. The 4th International Conference on Networks & Communications (NCOM 2018), The 4th International Conference on Artificial Intelligence and Applications (AIFU 2018), The 6th International Conference on Data Mining & Knowledge Management Process (DKMP 2018), The 7th International Conference on Embedded Systems and Applications (EMSA 2018), The 4th International Conference on Signal and Image Processing (SIPRO 2018) and The 7th International Conference on Software Engineering and Applications (SEA 2018) was collocated with The 8th International Conference on Computer Science, Engineering and Applications (CCSEA 2018). The conferences attracted many local and international delegates, presenting a balanced mixture of intellect from the East and from the West.

The goal of this conference series is to bring together researchers and practitioners from academia and industry to focus on understanding computer science and information technology and to establish new collaborations in these areas. Authors are invited to contribute to the conference by submitting articles that illustrate research results, projects, survey work and industrial experiences describing significant advances in all areas of computer science and information technology.

The CCSEA-2018, NCOM-2018, AIFU-2018, DKMP-2018, EMSA-2018, SIPRO-2018, SEA-2018 Committees rigorously invited submissions for many months from researchers, scientists, engineers, students and practitioners related to the relevant themes and tracks of the workshop. This effort guaranteed submissions from an unparalleled number of internationally recognized top-level researchers. All the submissions underwent a strenuous peer review process which comprised expert reviewers. These reviewers were selected from a talented pool of Technical Committee members and external reviewers on the basis of their expertise. The papers were then reviewed based on their contributions, technical content, originality and clarity. The entire process, which includes the submission, review and acceptance processes, was done electronically. All these efforts undertaken by the Organizing and Technical Committees led to an exciting, rich and a high quality technical conference program, which featured high-impact presentations for all attendees to enjoy, appreciate and expand their expertise in the latest developments in computer network and communications research.

In closing, CCSEA-2018, NCOM-2018, AIFU-2018, DKMP-2018, EMSA-2018, SIPRO-2018, SEA-2018 brought together researchers, scientists, engineers, students and practitioners to exchange and share their experiences, new ideas and research results in all aspects of the main workshop themes and tracks, and to discuss the practical challenges encountered and the solutions adopted. The book is organized as a collection of papers from the CCSEA-2018, NCOM-2018, AIFU-2018, DKMP-2018, EMSA-2018, SIPRO-2018, SEA-2018.

We would like to thank the General and Program Chairs, organization staff, the members of the Technical Program Committees and external reviewers for their excellent and tireless work. We sincerely wish that all attendees benefited scientifically from the conference and wish them every success in their research. It is the humble wish of the conference organizers that the professional dialogue among the researchers, scientists, engineers, students and educators continues beyond the event and that the friendships and collaborations forged will linger and prosper for many years to come.

Natarajan Meghanathan
David C. Wyld

Organization

General Chair

David C. Wyld   Southeastern Louisisna University, USA
Jan Zizka   Mendel University in Brno, Czech Republic

Program Committee Members

Abe Zeid   Northeastern University, USA
Ahmad Khasawneh   Hashemite University, Jordan
Ahmed Yahya   Al-Azhar University, Egypt
Alaa Hamami   Princess Sumaya University for Technology, Jordan
Alfredo Cuzzocrea   University of Trieste, Italy
Ali AYDIN M.   Istanbul University, Turkey
Alimohammad Shahri   NESCOT, United Kingdom
Annamalai Annamalai   Prairie View A&M University, USA
Ayad Ghany Ismaeel   Al-Kitab University College, IRAQ.
Aysegul OZDEMIR TOPALOGLU   Trakya University, Turkey
B.N.Gupta   Institute of Management & Science, India
Bapu Bandgar   JSPM's Jayawant Technical Campus, India
Basar Oztaysi   Istanbul Technical University, Turkey
Bedir Tekinerdogan   Wageningen University, Netherlands
Bing Zhou   Sam Houston State University, USA
Chanda C K   IIEST, India
Chin-Chih Chang   Chung Hua University,Taiwan
Chin-Wang Tao   National ILan University,Taiwan
Christophe NICOLLE   University of Bourgogne Franche, France
Da Yan   The University of Alabama at Birmingham, USA
Dabin Ding   University of Central Missouri, United States
Davoud Mougouei   Monash University, Australia
Debabrata Datta   St. Xavier's College, India
Debi Prasanna Acharjya   VIT University, India
Di Pietro   Roberto-HBKU-CSE, Qatar
Diego Reforgiato Recupero   University of Cagliari,Italy
Donggil Song   Sam Houston State University, USA
Doo Heon Song   Yong-in SongDam College, Korea
Ekbal Rashid   Aurora's Technical and Research Institute, India
Fabio Silva   Federal University of Pernambuco, Brazil
Fadi Wedyan   Hashemite University, Jordan
Farhad pourfarzi   Ardabil University of Medical Sciences, Iran
Fatih Korkmaz   Çankiri Karatekin University, Turkiye
Figen Balo   Firat University, Turkey
Guoqing Xiao   Hunan University, China
Haeng-Kon Kim   Catholic University of Daegu, Korea
Haibat Jadhav   Flora Institute of Technology, India

HASSINI Noureddine	University of Mostaganem, Algeria
Hiromi Ban	Nagaoka University of Technology, Japan.
Houda KHROUF	Atos Innovation Lab, France
Irena Patasiene	Kaunas University of Technology, Lithuania
Ishfaq Ahmad	The University of Texas at Arlington, U.S.A
Israa Sh. Tawfic	University of Gaziantep, Turkey.
Ivo Pierozzi Junior	Embrapa Agricultural Informatics, Brazil
Iyad Alazzam	Yarmouk University, Jordan
Jafar Mansouri	Ferdowsi University of Mashhad, Iran
Jamaiah Yahaya	National University of Malaysia, Malaysia
Jamal El Abbadi	Mohammadia V University Rabat, Morocco
Jasmine Seng K. P.	Charles Sturt University, Australia
Jatindra Kumar Deka	Indian Institute of Technology Guwahati, India
Jeong-Hyun Park	ETRI, South Korea
Jia Zhu	South China Normal University, China
Jinsul Kim	Chonnam National University, Korea.
Lim Tien Sze	Multimedia University, Malaysia
Madjid Tavana	La Salle University, USA
Madya Dr. Yuhanis binti Yusof	Universiti Utara Malaysia, Malaysia
Mamoun Alazab	Macquarie University, Australia
Mehran Garmehi	University of Bojnord, Iran
Min Kyung An	Sam Houston State University, USA
Mohd Nordin Abdul Rahman	University of Sultan, Malaysia
Mokhairi Makhtar	Universiti Sultan Zainal Abidin, Malaysia.
Nasser Tairan	King Khalid University, Saudi Arabia
Nizar	University of Carthage, Tunisia
Oleksii Tyshchenko	University of Ostrava,Ukraine
Orhan Dagdeviren	Ege University, Turkey
Ovidiu Valentin Drugan	DNV GL, Norway
Philippe Thomas	Université de Lorraine, France
Rahil Hosseini	Islamic Azad University, Iran
Ramgopal Kashyap	Sagar Institute of Science and Technology, India
Rana Rahim	Lebanese University, Lebanon
Razieh malekhoseini	Islamic Azad University, Iran
Roberto Paiano	University of Salento, Italy
Samia Nefti-Meziani	University of Salford, UK
Saraee Mo	University of Salford-Manchester, UK
Sherif Rashad	Florida Polytechnic University, USA
Shifei Ding	China University of mining and technology, China
Stefania Tomasiello	University of Salerno, Italy
Tony Tsang	Honfg Kong College of Techology, China
TSE GUAN	Universiti Malaysia Kelantan, Malaysia
Veton Kepuska	Florida Institute of Technology, USA
Vittoriano Muttillo	University of L'Aquila, Italy
Xiaoxiong Zhong	Tsinghua University, China
Xin-Wei Yao	Zhejiang University of Technology, China
Yagoubi Belabbas	University of Oran, Algeria
Yao-Nan Lien	Asia University, Taiwan
Zihao Zhou	University of California, USA

Technically Sponsored by

Computer Science & Information Technology Community (CSITC)

Networks & Communications Community (NCC)

Information Technology Management Community (ITMC)

Organized By

Academy & Industry Research Collaboration Center (AIRCC)

TABLE OF CONTENTS

8th International Conference on Computer Science, Engineering and Applications (CCSEA 2018)

4th International Conference on Networks & Communications (NCOM 2018)

4th International Conference on Artificial Intelligence and Applications (AIFU 2018)

6th International Conference on Data Mining & Knowledge Management Process (DKMP 2018)

7th International Conference on Embedded Systems and Applications (EMSA 2018)

4th International Conference on Signal and Image Processing (SIPRO 2018)

7th International Conference on Software Engineering and Applications (SEA 2018)

LOAD BALANCING MANAGEMENT USING FUZZY LOGIC TO IMPROVE THE REPORT TRANSFER SUCCESS RATE

Sanghyeok Lim[1]and Taeho Cho[2]

[1]College of Information and Communication Engineering, Sungkyunkwan University, Republic of Korea
[2]College of Software, Sungkyunkwan University, Republic of Korea

ABSTRACT

A wireless sensor network (WSN) consists of multiple sensor nodes and base stations (BS) that collect information over widely deployed sensor nodes. Sensor nodes have limited energy source and low computing power. Due to those features, there is a disadvantage that user's individual node management is difficult and they are easily captured by attackers. Therefore, efficient energy allocation of nodes is important and network security protocol is needed. The Probabilistic Voting Filtering System (PVFS) is a system that prevents false vote injection attack and false report attack injected from attackers. The reason for the existence of this protocol is for energy management of nodes through defence against those attacks and in order to efficiently manage the network based on PVFS, load balancing of nodes should be performed. In the proposed scheme, fuzzy logic is applied to each cluster head node (CH) to perform load balancing by determine whether to perform a role as a verification node and an event forwarding node. The experiment shows that the event detection rate and the report delivery success rate are improved in proposed scheme compare with original PVFS.

KEYWORDS

Network Protocols, Wireless Sensor Network, Fuzzy logic, False Report Attack, False Vote Injection Attack

1. INTRODUCTION

WSNs are used for data collection and event detection in various fields such as home networks, military systems, and forest fire monitoring [1] and are composed of many sensor nodes and a base station (BS). When an event occurs, the sensor node detects the event and makes report of that event and send it over multiple hops of the sensor nodes to the BS[2]. However, sensor nodes are vulnerable to attack because of the disadvantages of limited computation, limited energy, random distribution in an open environment that operates independently, and difficulties in individual management [3], [4].Attackers exploit these vulnerabilities to attack WSNs by injecting reports containing false information or injecting false Message Authentication Codes (MACs). Figures 1 shows a schematic of these attacks. These attacks reduce the energy of the node, shorten the lifetime of the network, and prevent detection and reporting of normal events. Several protocols have been developed to prevent false report attacks[5],[6] andLi and Wu proposed a probabilistic voting-based filtering scheme (PVFS) [7]to prevent both false report and false vote injection attacks.
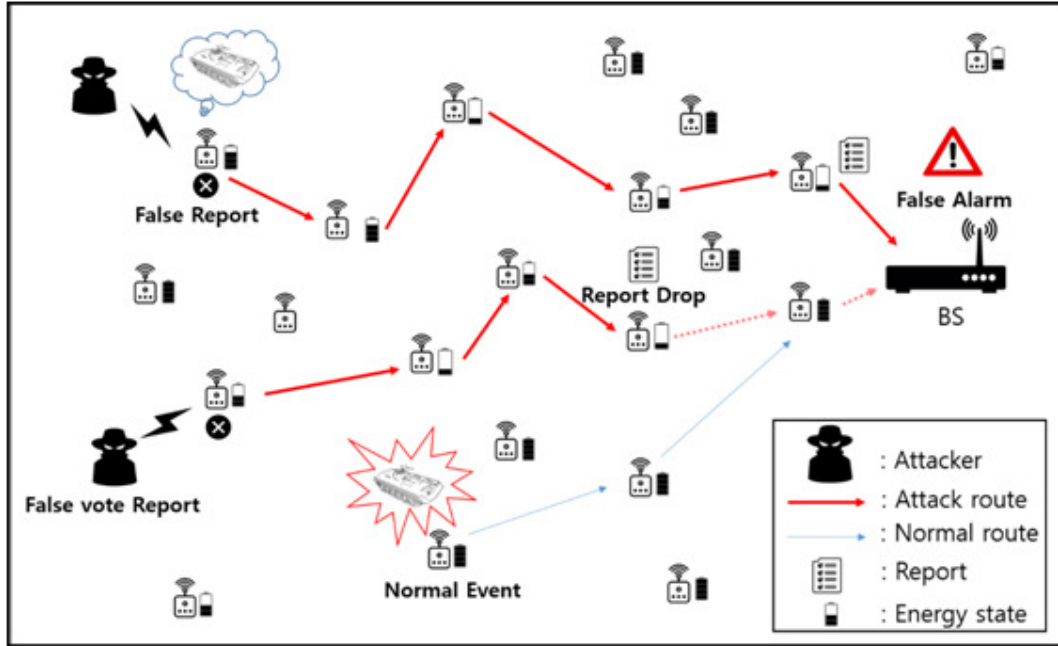
Figure 1. False report and false vote injection attacks

In PVFS, all nodes constitute a network that exploits cluster-based organization. When a cluster head (CH) recognizes an event, it generates a report of that event. It then sends this report to the member nodes. Next, the member nodes judge the authenticity of the report and generate their own MACs, alternatively referred to as votes in PVFS.
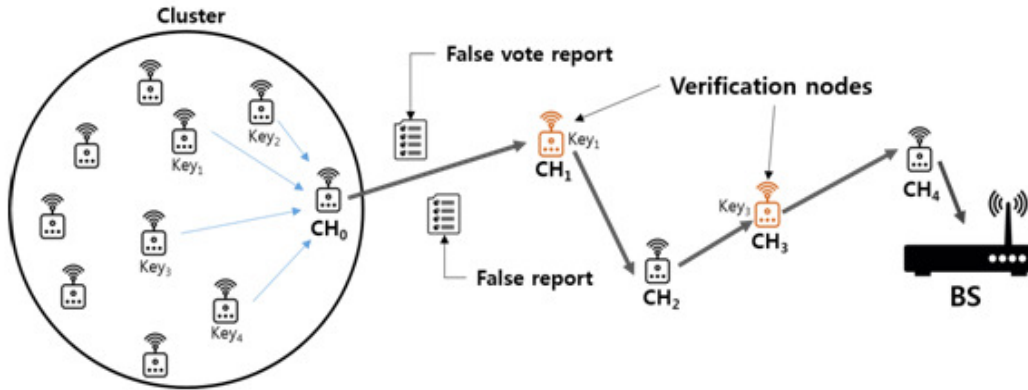
Figure 2. Report sending and verification process

Nodes that act as verification nodes in PVFS are more energy intensive than other member nodes, and if these nodes die, report generation and delivery of the region becomes impossible. In this paper, each CH's load balancing is performed by fuzzy logic based node role determination to increase event detection success rate and report transmission success rate with PVFS.Fuzzy logic is based on the fuzzy set thinking concept introduced by Professor L. A. Zadeh of the University of California at Berkeley in 1965 to quantitatively express the ambiguity of natural language and the like [8]. In the proposed method, fuzzy logic is used to help determine the role balancing of nodes by quantifying ambiguous properties such as how big the amount of nodes' energy is,how important nodes are.

2. RELATED WORK

2.1. PVFS

PVFS uses a true threshold value (Tt) and a false threshold value (Tf) to detect and filter false reports and false vote injected reports with validation nodes. PVFS has three phases; a key distribution phase, a report generation phase, and a verification phase.
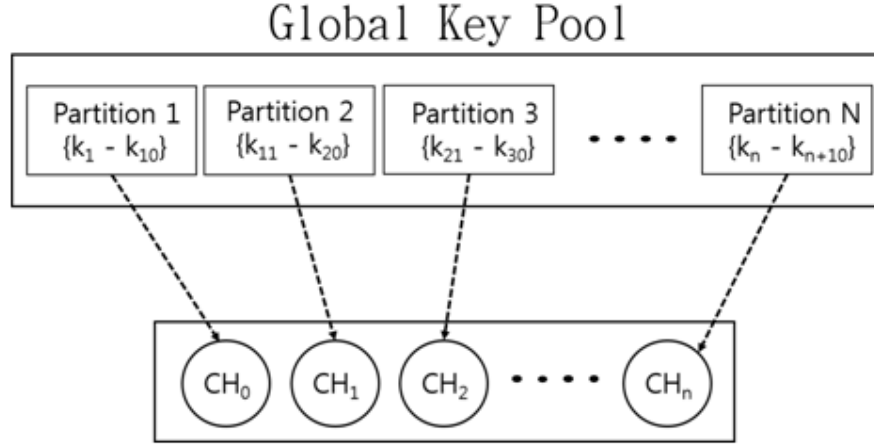
Figure 3. Key distribution process

Figure 3 shows the key allocation step in which the BS divides the key pool into N partitions and delivers them to each CH. Each partition contains L keys. The CH uses one of the keys in the partition as its own key and distributes the remaining L-1 keys to it's member nodes. A key is allocated to the member nodes according to the partition of the key pool. With this process, every node gets a single key from a global key pool.
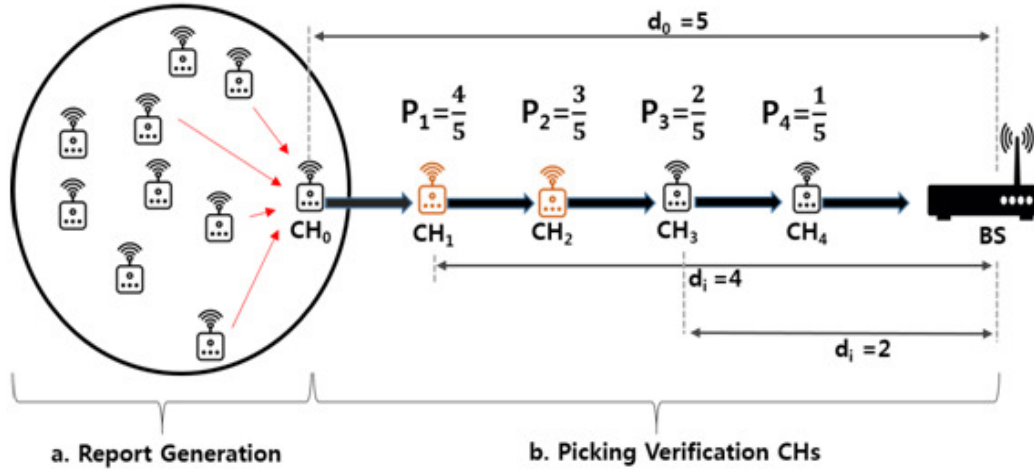
Figure 4. Report generation and verification node selection process

Figure 4 shows the report generation and validation node selection process. In the initial network configuration, nodes are divided into cluster units and the CHs responsible for report generation are selected ineach cluster. When an event occurs, the cluster closest to the event area wins the chance to generate a report of that event. CH generates a report and broadcasts it to member

nodes of its cluster. The member nodes confirm this, and if the report is judged to be a normal report, the MAC created by its own key is transmitted to the CH. CH extracts a predetermined number of MACs received from the member nodes and adds them to the report. The node selected as the verification node stores the keys of the member nodes of the event occurrence cluster one by one. In the report verification process, the verifying nodes compare their own key indexes with that of in the report. If they have same key index, they extract the MAC through their key.. If the MAC generated by the same key is different, the vote is regarded as fake and Tf is increased. In the filtering process, if the false count reaches the threshold value, the report is judged to be false and is immediately dropped. If the true count value reaches the threshold value, the report is considered legitimate and is sent to the BS without further validation. The verification node selection process is shown in Figure 4-b. The verification nodes among the CHson the path to the BS are probabilistically selected to verify the report. The probability 'p' uses the distance d_0 which is the hopcount from the BS to the event cluster and the distance d_i between the BS and CH_i. The closer the CHs are to the event cluster, the higher the probability of verification.

2.2. Fuzzy logic system

Fuzzy logic is based on the fuzzy set thinking concept introduced by Professor L. A. Zadeh of the University of California at Berkeley in 1965 to quantitatively express the ambiguity of natural language and the like. [9], [10], [11]. The concept of a fuzzy set is a set out of binary logic that each object belongs to or belongs to a certain group and represents the degree to which each object belongs to the group as a membership function. Also, the fuzzy measure indicates the ambiguity of the fact that the ambiguous element a in the general set A belongs to the subset P of A, so that the relation between P and A is a mathematically continuous attribute. The configuration of the Fuzzy Logic Controller is shown in figure 5.
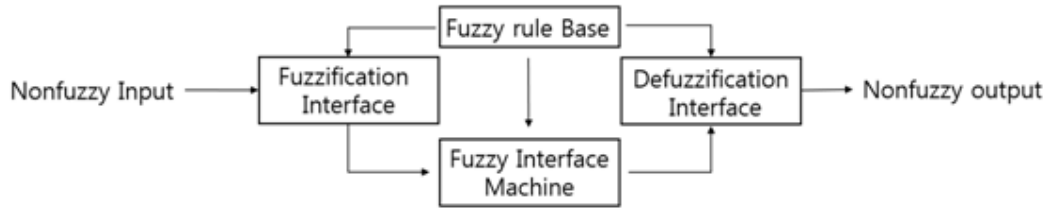
Figure 5. Fuzzy logic controller

The fuzzy controller consists of a fuzzy rule base, a fuzzy inference engine, and a defuzzifier. Fuzzification Interface is responsible for Scale Mapping task that changes the value range of the input variable to the universe of discourse. Fuzzy Rule Base is a set of linguistic control rules written in IF-THEN form. Fuzzy Inference Machine is decision-Making Logic that adopts rules from Fuzzy Input and Fuzzy Rule and in defuzzication Interface, scale mapping work is performed to convert into Universe of Discourse which is matched with the range of output variable value.

3. PROPOSED METHOD

3.1. Problem statement

In an environment where a WSN is installed, there are many areas where the user cannot distribute nodes directly and the nodes are scattered randomly. Therefore, there is a high possibility that a good routing path based on energy efficiency and characteristics of PVFS is not generated every time. One of the main purposes of the WSN is to gather information about events that occur in vast areas where it is difficult for a user to directly reach and respond appropriately. If the report of the detected event is not successfully transmitted to the BS, the main purpose of

the WSN is to lose, or if the energy of the node is depleted so that event detection in the corresponding region is not possible. Also, if the residual energy of the sensor nodes is high but the events to be delivered are not properly transmitted, the sensor nodes distributed in those areas become useless. Therefore, it is important to improve the event detection rate and report delivery success rate even if the total energy consumption of the WSN increases by modifying the node to be available for those multiple purposes. Since PVFS shows probabilistic security, there is a tendency to show extremely high and low performance in random routing environments. Therefore, it is important to use nodes efficiently in the WSN by adjusting the role of nodes appropriately and raising the event detection rate. Figure 6 shows the part of the field where the WSN is configured. As shown in the figure, CHs that detect the event have limited communication distances, so they need to cooperate with other nodes to deliver the report to the BS through hop movement.
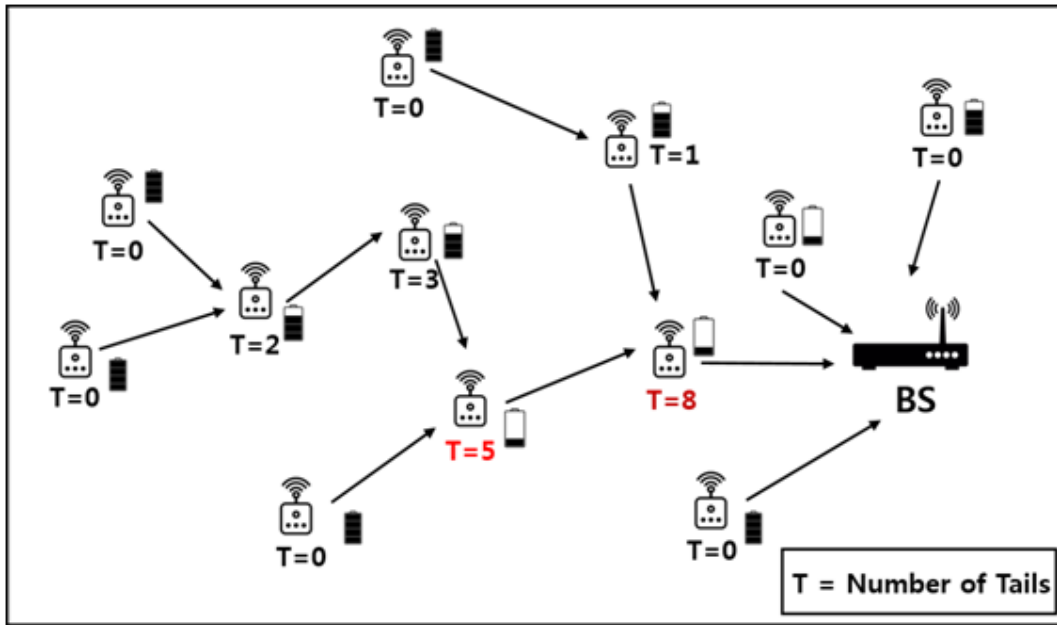
Figure 6. Calculating tail number

Therefore, each node has different number of downstream nodes depending on its location. Tail refers to the number of CH nodes that use the node to forward detected events. If the CH is positioned at the end of the field, the tail value of the corresponding node is zero. A node with many tails can consume a large amount of energy for sending, receiving, and verifying the report, which plays an important role in transferring report of the events occurring in other regions. If such node dies, the event detection node must forward the report through the new path except that node, and the event detected in the area where the dead node was receiving is no longer detectable because the CH is selected only once in the initial node dispatch phase.

3.2. System overview

Figure7 shows an example of the network on a field using the proposed method. Every CH has a list of the IDs of the nodes that are closer to the BS within its communication range. This list is used for the formation of new routing except the nodes selected by the tagging node among the list members. The sensor nodes have enough memory to store this brief information.
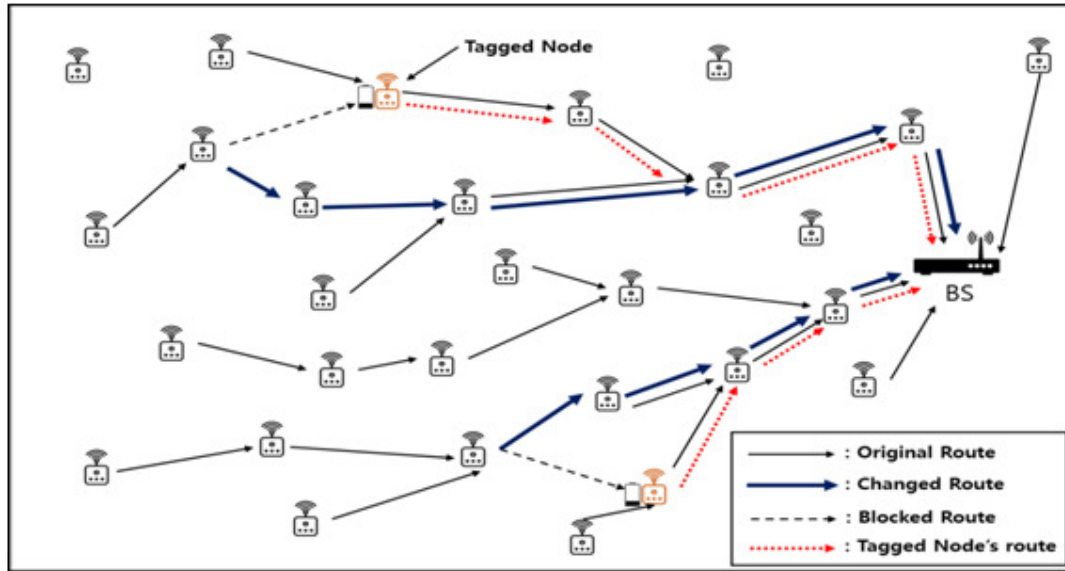
Figure 7. Dynamic routing in proposed scheme

Nodes with depleted energy are dropped from the list and the tagged nodes remain in the list but are excluded from the routing priority. If all the nodes in the list are tagged, the node closest to the BS is forcibly routed again. The node tagged by the fuzzy logic based decision system on the original route requests the previous route node to set up new routing, and the tagged node performs only event detection and report generation during the remaining period. The fuzzy function has three inputs: the energy of the node, the geographical importance of the node, and the frequency of occurrence of the event in the region. The calculation of the fuzzy function is performed with a constant period when it falls below a certain energy because frequent fuzzy calculations shorten the lifetime of the node.
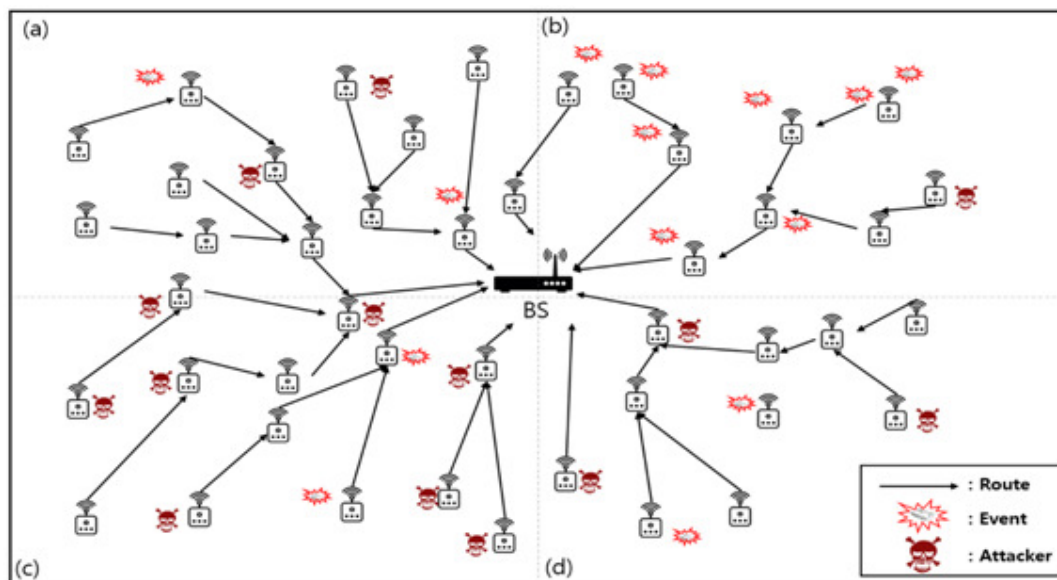
Figure 8. dynamic event and attacks in WSN

In the actual WSN field, there is very little chance that events and attacks will occur uniformly locally. In particular, in the case of an attack, it is difficult for the attacker to travel around the entire field, and attack occurs through a few nodes that have been hijacked. Therefore, load balancing considering the local attack rate is necessary. This is also why the frequency of occurrences of fuzzy input values in the proposed method is getting into.Figure 8 shows how these attacks and normal events can occur in an actual field.The proposed scheme extend the lifetime of the network through load balancing of nodes considering these characteristics, and improve report transmission success rate and event detection rate.

3.3. Fuzzy based fitness evaluation

The proposed method exploits a fuzzy rule based system for load balancing. Part of the appeal of fuzzy rule based systems is that they can be used for approximate reasoning. Which is particularly important when there is uncertainty in reasoning processes in addition to imprecision in data. The fuzzy input includes the amount of residual energy of the node, the importance of the position of the node, and finally the frequency of local event occurrence. The fuzzy output for these three inputs helps determine whether to use the CH as an event delivery and verification node.

3.4. Fuzzy membership function

The functions for the three inputs of the fuzzy logic are shown in the figure 9.The following is the reason for selecting the values used in the fuzzy input.The fuzzy functions contain the importance, energy amount, and frequency of occurrence of event and attacks respectively. Optimization of functions can be performed using genetic algorithms[12],[13],[14].In the proposed method, several functions are randomly simulated and a function that most satisfying the if-then rule is used.

Figure 9. Fuzzy membership functions

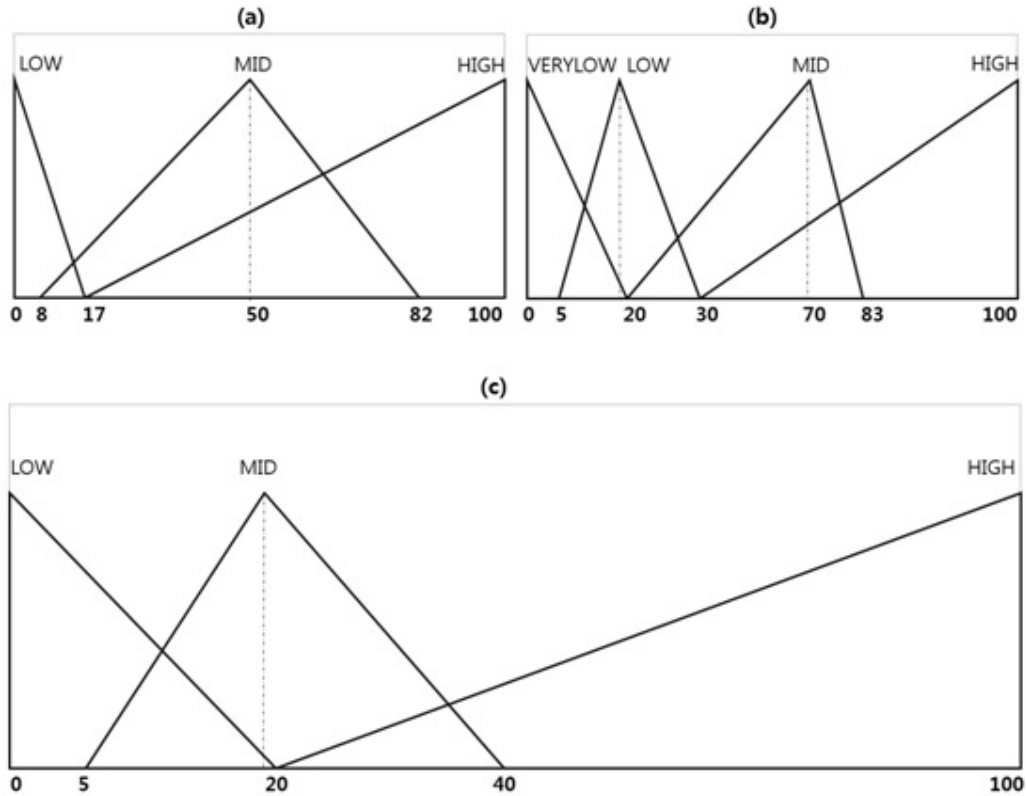1. Importance (a): Importance is a measure of whether a given CH plays an important role in the network. It is a value considering the probability that the CH becomes the verification node of the downstream nodes and the tail number of the corresponding CH.A node with a high probability of being a verification node is expected to have a high energy consumption and its importance is increased. In general, in PVFS, the closer to BS, the more tail increases and the probability of becoming a verification node becomes lower. Therefore, the distance with BS does not lower the importance of corresponding nodes every time and it's hard to measure it just considering distance. Therefore, the proposed method uses the measurement function to measure the importance. The evaluation function of i-th node E_{tag}^i is expressed as the sum of α, β and γ, which are the amount of energy used during the role of the forwarding node, the amount of energy consumed during the role of the event detection node and the energy used by the verification node to report transmission after event detection, respectively. α is given by $\alpha = e_t * \frac{N_{evt}}{N_n}$, where e_t is the transmitting cost, N_n is the number of nodes in the field and N_{evt} is the number of events. β is given by $\beta = T * (e_t + e_r) * \left(\frac{N_{evt}}{N_n}\right)$, where e_r denotes the receiving cost and T is the i-th CH's tail node. γ is more complicated, and can be expressed as the sum of probabilities that the i-th node is a verification node of each tail node, all times the probability of having the same key when it is a verification node. The sum of probabilities that a node is a verification node of the tail nodes is given by $e_{cal} * \sum_{j=1}^n P_i^j$, where $P_i^j = \frac{HopCount_{N_j}}{HopCount_{N_i}}$ and e_{cal} denotes calculating cost.Then, the sum of probabilities of each node is $P_i^1 + p_i^2 + \cdots + P_i^T = \frac{1}{HopCount_{N_i}} * \sum_{j=1}^T HopCount_{N_j}$. When the node performs verifying, the probability that the key of the corresponding node overlaps with the key of the report is $\frac{S}{L}$, where s is the number of votes in the report and L denotes the number of member nodes from each tail nodes. Therefore, the evaluation function is as follows:

$$E_{tag}^i = e_t * \frac{N_{evt}}{N_n} + T * (e_t + e_r) * \left(\frac{N_{evt}}{N_n}\right) + \frac{S}{L} e_{cal} * \frac{1}{HopCount_{N_i}} * \sum_{j=1}^T HopCount_{N_j}.$$

The output of the function is integerized into the fuzzy input.

2. Energy (b): The energy of the node is the most important element of the fuzzy inputs. When nodes with sufficient energy are tagged, the routing path is unnecessarily increased, which has the adverse effect of increasing transmission and reception costs. Also, the probability of encountering the hijacked node during the hop movement increases. On the contrary, if the CH is tagged in a situation where the amount of energy is too small, there is not enough residual energy to perform an event detection function, so that a node is killed during the generation of a report or waiting for a report, Report passing rates are also lower.

3. Event occurrence frequency (c): The occurrence frequency of events indicates the ratio of events occurring in the area where the corresponding node is located among all the fields. If the event frequency is high, the probability of event detection and report generation increases, and the probability of being tagged increases accordingly. On the other hand, if the frequency of occurrence of events in the area is low, it is desirable from the viewpoint of the whole network to delay the tagging timing and induce the node to perform the role of the verification and transmission node for a longer time. The frequency of occurrence is difficult to measure in real time, and the amount of energy consumed in real-time measurement increases, which is rather counterproductive.

The following is if-then rule of fuzzy system. It contains those 3 inputs that previously explained. Each input values have 3, 4 and 3 levels, respectively and it's the output of the rule is determined by the expert.

Table 1. Fuzzy If-Then rules

No.	INPUT			OUTPUT
	Importance	Energy	Event occurrence	ON/OFF
1	LOW	VERYLOW	LOW	ON
13	MID	LOW	LOW	OFF
15	MID	LOW	HIGH	ON
18	HIGH	LOW	HIGH	ON
31	MID	HIGH	LOW	OFF
36	HIGH	HIGH	HIGH	OFF

Table 1 shows some of the 36 if-then rules used in the proposed fuzzy system. The 3 input values are Importance, Energy, and event occurrence, and the output contains whether or not to tag the node as a report-generating-only node. For example, if the importance is LOW, the energy is VERYLOW, and the event occurrence is LOW, the node is selected as a node for event generation only. If the importance is MID, energy is LOW and event occurrence is LOW, It is decided not to tagged yet. The input can't be arbitrarily adjusted, and it must have a total of 36 values. The output of the rule reflects the case with the best results through experiments with several combinations. Rule with inconsistent output results in system performance degradation, so the exorbitant output was excluded from the experiment.

4. EXPERIMENTAL RESULT

4.1. Experimental environment

Table 2. Experiment parameters

Item	Value
Sensor field size(m)	1000× 1000
Number of sensor nodes	2000
Number of cluster head nodes	200
S	5
L	10
Packet size(byte)	24
Transmission range(m)	150

Experiments were conducted assuming that the attack rates were 0 to 90% The reason a 100% attack rate is excluded from experiments is that this experiment is an experiment that evaluates the failure rate of a normal report. The transmission and reception costs are set to 16.25μj and 12.5μj, respectively, and the calculated cost of voting was set to 15μj [15].The threshold values Tt and Tf used in the experiments were in accordance with the experimental environment of the original PVFS[7].The reason for this is to increase the reliability of the proposed method by making the environment of original version of PVFS equal to the proposed method.

4.2. Assumptions

The CH's location and key distribution phase does not change during the event. If the initial energy of the node is too high, it will be difficult to calculate the event detection failure and report transmission failure accurately because the node's energy depletion does not occur during the

experiment. Therefore, the initial energy of the node is set to an appropriate amount. The BS has all the keys distributed to the node and has the computing power to verify all false reports, false votes.

4.3. Experimental result
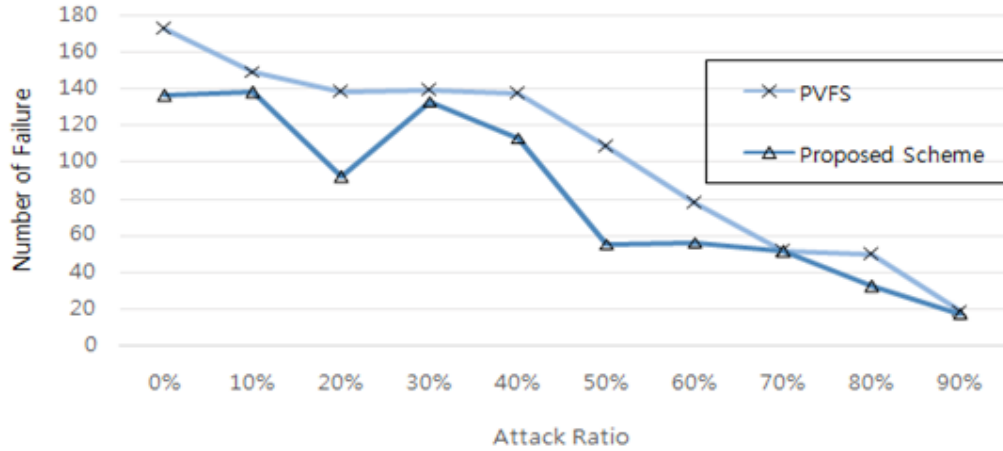
Figure 10. Number of transfer Failure

Figure 10 shows the number of report failures according to the attack rate at 600 events when the local occurrence frequency is 20, 40, 30, and 10%.The experiment was conducted with the exception of the transmission success rate for false report attacks because the network users are not mean to receive false reports derived from the attacker. It can be seen that the number of report transmission failures of the existing PVFS and the proposed technique decreases as the attack rate increases because when an attack ratio increases, the number of normal event's report sending itself is decreases. And as the attack rate increases, it can be seen that the difference between the number of report transmission failures of the proposed technique and PVFS is also reduced.
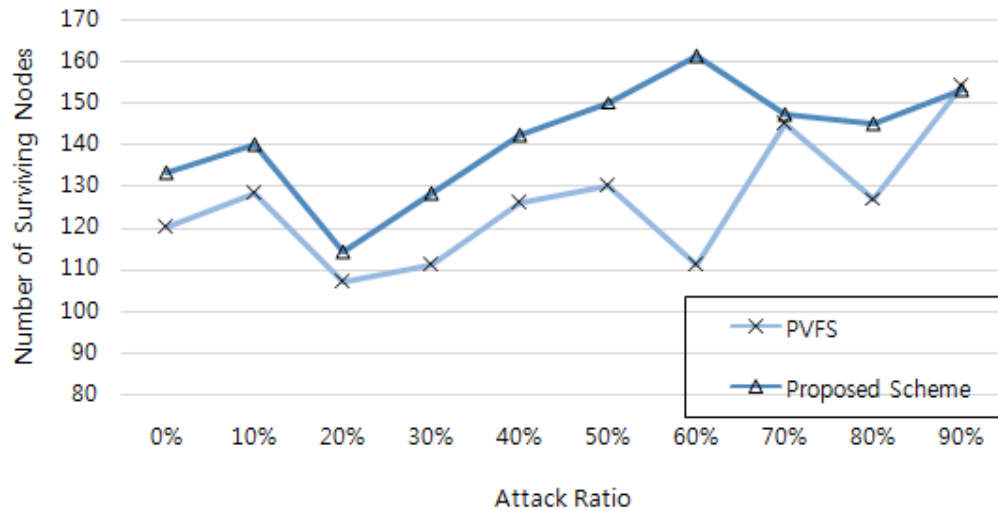
Figure 11. Number of surviving Nodes

Figure 11 shows the number of surviving nodes according to the attack rate of the existing PVFS and the proposed scheme. The number of surviving nodes was measured at every cycle after the experiment, and there is some part with little difference from original version of PVFS (70%, 90%).The reason for this result is that the proposed scheme has more number of surviving nodes than the existing PVFS during the experiment, resulting in additional energy consumption due to event reception, report generation, and report transmission. These results show that the number of failures of the proposed method and the report delivery failure of the existing PVFS are significantly different from those of the Figure 10, but the number of surviving nodes is not so different.
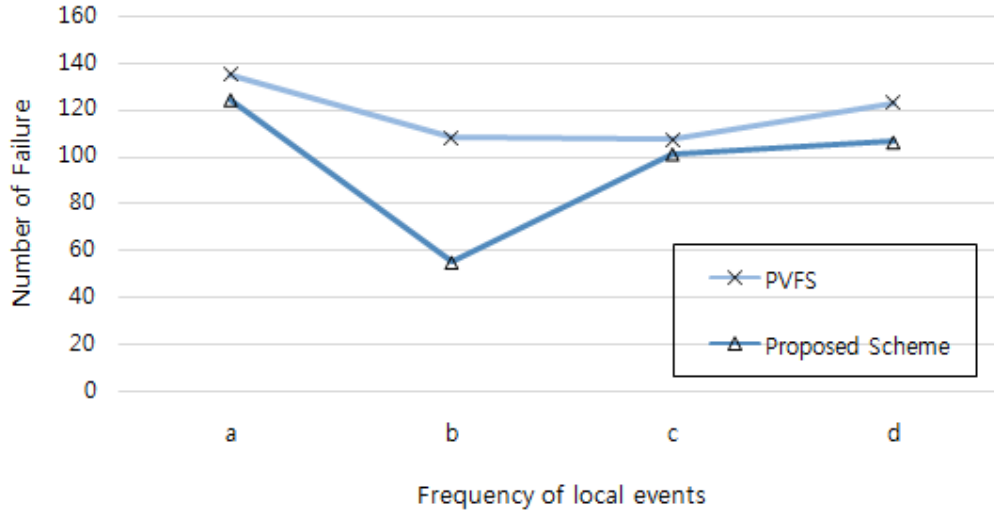
Figure 12. Number of Transfer Failure in Dynamic WSN Field Environment

Figure 12 shows the performance difference between the proposed method and the existing PVFS according to the change of local incidence a, b, c, and d of Figure 12 represent the frequency of events occurring in the regions of Figures 8-a, 8-b, 8-c, and 8-d, respectively and each variable means a = 25,25,25,25(%), b = 30,40,20,10(%), c = 20,20,50,10(%), d = 0,50,0, 50(%).As can be seen from the graph, load balancing with fuzzy logic increases the lifetime of the entire network and increases the success rate of the normal report on the event. A large number of surviving nodes means that event detection for the region can be longer and more. The residual energy of all the nodes can be measured somewhat lower in the proposed method because the report about the event that occurred in the dead node is generated and transmitted, the event detection is not performed, and report transmission more.

3. CONCLUSIONS

In the proposed method, we experimentally confirmed that a WSN's event detection success rate, report transmission success rate, and overall network lifetime are increased by organically changing the roles of the nodes through fuzzy logic based decision making system that evaluate the node status. What the user wants through the WSN is successful reporting of normal events. Therefore, the experiment of the proposed technique excluded how successful the transmission of false reports was. The WSN user periodically replaces the node battery placed in the entire field or picks up the nodes placed in the field when the local network is not needed. Therefore, the total sum of the residual energy amount of the entire nodes placed in the field does not influence the performance of the network. Although it is relatively low energy, it should be the criterion for determining the performance of the network whether or not the node can generate a report on the

live event and send it to BS. The proposed method is a method of raising the report success rate and the number of surviving nodes simultaneously. For further study, the load balancing using the membership function of the fuzzy logic which is optimized through the genetic algorithm will be implemented. The routing environment changes every moment and it is impossible for humans to optimize the membership function at every moment, so the genetic algorithm should be applied to automatically adapt the changing routing environment to obtain the optimal output.

REFERENCES

[1]   Zhang, Wensheng, and Guohong Cao. "Group rekeying for filtering false data in sensor networks: A predistribution and local collaboration-based approach." INFOCOM 2005. 24th Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings IEEE. Vol. 1. IEEE, 2005.

[2]   Perrig, Adrian, John Stankovic, and David Wagner. "Security in wireless sensor networks." Communications of the ACM 47.6 (2004): 53-57.

[3]   Al-Karaki, Jamal N., and Ahmed E. Kamal. "Routing techniques in wireless sensor networks: a survey." IEEE wireless communications 11.6 (2004): 6-28.

[4]   Perrig, Adrian, John Stankovic, and David Wagner. "Security in wireless sensor networks." Communications of the ACM 47.6.

[5]   Zhu, Sencun, et al. "An interleaved hop-by-hop authentication scheme for filtering of injected false data in sensor networks." Security and privacy, 2004. Proceedings. 2004 IEEE symposium on. IEEE, 2004.

[6]   Yang, Hao, and Songwu Lu. "Commutative cipher based en-route filtering in wireless sensor networks." Vehicular Technology Conference, 2004. VTC2004-Fall. 2004 IEEE 60th. Vol. 2. IEEE, 2004.

[7]   Li, Feng, and Jie Wu. "A probabilistic voting-based filtering scheme in wireless sensor networks." Proceedings of the 2006 international conference on Wireless communications and mobile computing. ACM, 2006.

[8]   Zadeh, Lotfi A. "Fuzzy sets." Information and control 8.3 (1965): 338-353.

[9]   J. Yen and R. Langari, Fuzzy Logic: Intelligence, Control, and Information. Prentice-Hall, Inc., 1998.

[10] G. Klir and B. Yuan, Fuzzy Sets and Fuzzy Logic. Prentice hall New Jersey, 1995.

[11] R. Babuška, "Fuzzy Systems, Modeling and Identification," Delft University of Technology, Department of Electrical Engineering Control Laboratory, Mekelweg, vol. 4, 1996.

[12] D. E. Golberg, "Genetic algorithms in search, optimization, and machine learning," Addion Wesley, vol. 1989, 1989.

[13] C. L. Karr, "Design of an adaptive fuzzy logic controller using a genetic algorithm." in Icga, 1991, pp. 450-457.

[14] A. Geyer-Schulz, Fuzzy Rule-Based Expert Systems and Genetic Machine Learning. Physica Verlag, 1997.

[15] Ye, Fan, et al. "Statistical en-route filtering of injected false data in sensor networks." IEEE Journal on Selected Areas in Communications 23.4 (2005): 839-850.

AUTHORS

Sanghyeok Lim

received a B.S. degree in Digital Information Engineering from Hanguk University of Foreign Studies in 2017, and is now working toward an M.S. degree in the Department of Electrical and Computer Engineering at Sungkyunkwan University.
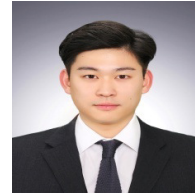
Taeho Cho

received a Ph.D. degree in Electrical and Computer Engineering from the University of Arizona, USA, in 1993, and B.S. and M.S.degrees in Electrical and Computer Engineering from Sungkyunkwan University, Republic of Korea, and the University of Alabama, USA, respectively. He is currently a Professor in the College of Information and Communication Engineering, Sungkyunkwan University, Korea.

INTENTIONAL BLANK

PERFORMANCE EVALUATION OF MOBILE IP ON MOBILE AD HOC NETWORKS USING NS2

Kingsley K. Ofosu[1], Jamal-Deen Abdulai (PhD)[2] and
Ferdinand A. Katsriku (PhD)[3]

Department of Computer Science, University of Ghana, Legon-Accra, Ghana

ABSTRACT

Mobile computing devices equipped with transceivers form Mobile Ad Hoc Networks (MANET) when two or more of these devices find themselves within transmission range. MANETs are stand-alone (no existing infrastructure needed), autonomous networks that utilise multi-hop communication to reach nodes out of transmitter range. Unlike infrastructure networks e.g. the Internet with fixed topology, MANETs are dynamic. Despite the heterogeneous nature of these two networks, integrating MANETs with the Internet extends the network coverage and adds to the application domain of MANETs. One of the many ways of combining MANETs with the Internet involves using Mobile Internet Protocol (Mobile IP) and a MANET protocol to route packets between the Internet and the MANET via Gateway agents. In this paper, we evaluate the performance of Mobile IP on MANET in Network Simulator 2 (NS2). We have implemented Mobile IP on Ad hoc On-demand Distance Vector (AODV), Ad hoc On-demand Multiple Distance Vector (AOMDV) and Destination-Sequenced Distance Vector (DSDV) routing protocols and compared performances based on Throughput, End-to-End Delay (E2ED), Packet Delivery Ratio (PDR) and Normalized Packet Ratio (NPR). The simulation results suggest that on-demand routing within the MANET better serves Mobile IP on MANETs.

KEYWORDS

Mobile Ad Hoc Network, Mobile IP, AODV, AOMDV &DSDV

1. INTRODUCTION

Communication has shifted from a predominantly wired setup towards an entirely wireless setup or a merger. Computer networks aid in faster and reliable communications over long distances. The Internet, a network of networks, has become a vital utility in our lives that enable us to communicate around the globe. Mobility as a feature in communication has gained the acceptance of end-users, and so, it is not surprising that Mobile Ad hoc Networks (MANETs) have attracted much attention from researchers. A MANET is an autonomous, infrastructure-less, self-forming and self-repairing data network of mobile devices that support multi-hop communication. MANET could be used to provide Internet connectivity beyond the reach of fixed or cellular infrastructure [1].

The Mobile Internet Protocol (MIP), which was designed to allow mobile nodes to move inside the fixed Internet without losing connectivity, has been experimented in MANET to provide Internet connectivity. We have implemented three variants of Mobile IP on MANET:

- ▪ Mobile IP on Ad Hoc On-demand Distance Vector (AODV)

- ▪ Mobile IP on Ad Hoc On-demand Multiple Distance Vector (AOMDV)

- ▪ Mobile IP on Destination-Sequenced Distance Vector (DSDV)

and evaluated their performances using Network Simulator 2 (NS2).

2. LITERATURE REVIEW

Not much work has been published concerning providing Internet connectivity for MANETs. A number of publications have suggested corporation between Mobile IP [2] and an ad hoc routing protocol to provide Internet access for MANETs. In "Ad Hoc Networking with Mobile IP" [3], a solution was presented whereby a proactive MANET routing protocol was used with Mobile IP. This solution was not compatible with on-demand routing in MANET. In "MIPMANET – Mobile IP for Mobile Ad Hoc Networks" [4], an on-demand MANET routing protocol (AODV) is used alongside Mobile IP with foreign agent care-of-addresses to connect a MANET to the Internet. Likewise MIPMANET, the Internet draft "Global Connectivity for IPv4 Mobile Ad hoc Networks" (Global4) [5] presented a solution whereby AODV interacts with Mobile IP however, the foreign agent discovery mechanism is incorporated into AODV.

Other publications have used different gateway discovery methods within the MANET routing protocol instead of Mobile IP. The publication titled "Wireless Multihop Internet Access: Gateway Discovery, Routing and Addressing" [6] discusses how MANET nodes can discover gateways to the Internet and the issue of routing and addressing in heterogenous environments. The master's thesis "A Study of Internet Connectivity for Mobile Ad Hoc Networks in NS2" [7] implemented the Internet draft "Global Connectivity for IPv6 Mobile Ad hoc Networks" [8] also known as Global6 in NS2. The AODV routing protocol was modified to include three different gateway detection mechanisms that were tested using NS2 simulations.

Regarding papers that focused more on performance analysis, [9] and [10] conducted separate performance evaluation of Mobile IP on proactive and reactive MANET routing respectively.We will return to these papers as we discuss related works to thus thesis in section 2.3, after we have discussed Mobile IP and Mobile Ad hoc Networks in Sections 2.1 and 2.2 respectively.

2.1. Mobile IP

Mobile IP works by allowing the Mobile Node (MN) to maintain two IP addresses; the home address and the Care-of-Address (CoA) [2]. The home address is static and binds MN to a known network called the Home Network (HN). This address is used by transport and application layer protocols to maintain connectivity with MN. On the other hand, the CoA is a dynamic address that MN obtains as it moves outside HN into an unknown network termed as a Foreign Network (FN). The CoA reflects the physical location of MN outside of HN.

When MN moves from the HN into FN, it sends the CoA obtained to the Home Agent (HA) on its HN via a similar agent on FN known as Foreign Agent (FA). This message prompts HA to receive packets destined for MN and arriving at HN on behalf of MN and then tunnels the packets to the CoA. MN repeats this procedure each time MN obtains a new CoA. There are three building blocks of Mobile IP: agent discovery, registration and datagram delivery.

2.1.1. Agent Discovery

This process is an extension of the router advertisement procedure, as specified in the Internet Control Message Protocol (ICMP) router discovery messages in RFC 1256 [11]. Specifically, the only difference is that an agent advertisement has, also, one or more CoAs made available by FA.

[2] explains that an agent advertisement does the following:

- Allows the detection of mobility agents;
- Lists the available CoA;
- Informs MN of the services offered by FA e.g. alternative encapsulation techniques;
- Aids MN to determine its network address and the status of the link to the Internet;
- Helps MN to determine whether the agent is a Home Agent, Foreign Agent or both, and therefore whether it is inside HN or FN.

Agents broadcast advertisements periodically, once a second or once every few seconds. However, an MN may solicit for an agent advertisement regardless of this schedule

2.1.2. Registration

When MN has a new CoA, it must inform HA to render its services to MN. This procedure is known as registration [1]. MN sends a registration request containing:

- Current CoA
- How long MN intends to use the CoA i.e. Time-to-Live (TTL)
- Parameters and flags that specify how HA should forward packets
- Special services that MN requests of HA.

When HA receives the registration request, it then authenticates MN and then decides whether to accept or reject depending on the outcome of the authentication process. HA sends a reply to MN via FA. If HA accepts the request, it maintains in its cache the home address, CoA and the TTL that MN specified in the request. This trio is known as a binding information, and so registration request is sometimes referred to as a binding update. Although the FA remains passive in this process, it maintains a visitors list of each MN that successfully registers with HA through it.
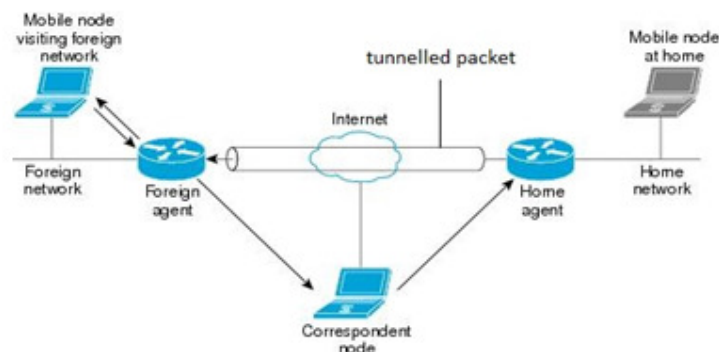
2.1.3. Datagram Delivery

Figure 1. Mobile IP architecture.

As illustrated in Figure 1 above, the Corresponding Node (CN) transmits packets destined for MN towards HN through standard IP routing. When the packets reach HN, the HA accepts these packets as though it is MN. HA then encapsulates the packet with the CoA as the new destination and forwards the packets to MN. This process is known as tunnelling. Two tunnelling protocols often used are IP within IP [12] and minimal IP within IP [13]. HA continues to receive packets on behalf of MN until TTL runs out. If MN wishes to continue this service, it must register again.

2.2. Mobile Ad Hoc Networks
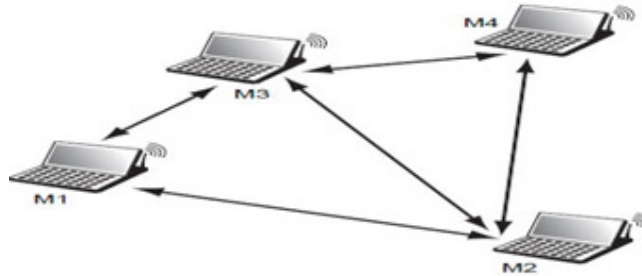
Figure 2. Mobile Ad Hoc Network.

A MANET is an autonomous, infrastructure-less, self-forming and self-repairing network of mobile devices that support multi-hop communication. The mode of transmission is wireless and has a very dynamic topology due to node movements. MANET nodes are usually small, and battery powered. MANETs are deployed in disaster recovery [14], military applications [15], sensor networks [16], etc.

2.2.1. Manet Routing Protocols

Further Initially, MANET routing protocols were proactive. They store and update periodically, information on existing paths to all possible destinations for data delivery. Proactive protocols are quicker in choosing routes for data delivery, however, they do not easily converge [17]. E.g., Destination-Sequenced Distance Vector (DSDV) [18].

Reactive or On-demand protocols, unlike the proactive protocols do not anticipate routes between nodes. Nodes discover routes only when there is data to transmit. These protocols only maintain a node neighbour list necessary for route solicitation. E.g. Ad hoc On-demand Distance Vector (AODV) [19].

Destination-Sequenced Distance Vector (DSDV)

DSDV is based on a distance-vector algorithm and guarantees a loop-free route unlike earlier distance vector based routing protocols by introducing sequence numbers [18]. Nodes maintain a routing table of all destinations with information on the next-hop to the destination, the number of hops and destination sequence number.

Periodically, a node shares its route table information with neighbouring nodes known as route updates. Attached to each update is a sequence number. An odd sequence number indicates an unreachable destination whereas an even sequence number indicates a reachable destination. The greater the sequence number, the fresher the update.

It is possible to have different updates to the same destination with the same sequence number. In such scenarios, nodes prefer the route with a lesser number of hops to the destination.

Ad Hoc On-Demand Distance Vector (AODV)

AODV [19] is a reactive protocol also based on the distance-vector algorithm. In AODV, nodes only seek for route information only when they have data to transmit. Loop-free routes are guaranteed with the use of sequence numbers. A node maintains route information only when transmitting data or it finds itself along an active path.

Each time a node has data to share, it solicits for a route to the destination by broadcasting a route request (RREQ) message. A node which is the destination itself or another node that happens to know a route to the destination may unicast a route reply (RREP) to the source or to the intermediate node from which it received the forwarded RREQ and maintains a pointer to that node. This process is known as a forward route setup.

Otherwise, the intermediate node forwards the RREQ and maintains a pointer to the node from which it received the RREQ. This process is known as a reverse route setup. Data transmission may begin as soon as the source receives the RREP via the forward route setup to the destination. Nodes maintain the forward route until there is no more data to transmit or the path becomes broken due to an unreachable intermediate node or destination in which case route solicitation restarts.

Ad Hoc On-Demand Multiple Distance Vector (AOMDV)

AOMDV [20] is a reactive protocol that improves on AODV. In AOMDV, nodes maintain multiple forward route setups during a single route solicitation procedure. These multiple routes are link disjoint which means that no two routes share a common intermediate node. This procedure reduces the probability that all routes fail at once. A primary route for data transmission is then selected, and nodes maintain the rest as alternate routes. When the primary route fails, nodes select an alternate route for data delivery, unlike AODV that repeats route solicitation whenever there is route failure.

Nodes repeat route solicitation when there is a new transmission session or when all alternate routes and the primary path fails.

2.3. Related Works

In this section, we discuss some closely related works to this thesis as we build up the motivation for this study.

In [3], a proposal for connecting MANET to the Internet using Mobile IP is presented in which a modified Routing Information Protocol (RIP) like DSDV was used in routing packets within the MANET. In this proposal, a single routing table is used and shared by Mobile IP and the MANET routing protocol to reduce the management tasks involved in maintaining separate routing tables for Mobile IP and the MANET protocol. Therefore, a route manager is introduced to coordinate route table management between Mobile IP and the modified RIP on the shared routing table. With this, neither Mobile IP or the modified RIP could modify the routing table directly. Route manipulation requests are sent to the route manager which then acts on behalf of the protocols. Although this proposal was successful in providing Internet access to the MANET, it was more proactive and did not support reactive MANET protocols since it relied on the periodic control messages of the MANET routing protocol to propagate agent advertisements.

To fill the gap, [4] presented a master thesis that detailed the use of Mobile IP on a reactive MANET routing protocol. The authors implemented Mobile IP on AODV in NS2 and then

performed some simulations to validate their study. In this proposal, the authors used Mobile IP with Foreign Agent care-of-addresses to reach Mobile Nodes from the Internet. Packets were exchanged between Home Agent and the Mobile Node via reverse tunnelling to minimize the requirements on AODV. A novel Internetworking unit was introduced between Mobile IP and AODV to ensure that no modifications be made to Mobile IP outside the MANET. The emphasis of the thesis was not on performance evaluation and so the simulations were not quite extensive. Also, the Internet draft "Global Connectivity for IPv4 Mobile Ad Hoc Networks" [5] sometimes referred to as Global4, presented a solution which involves the use of Mobile IP with foreign agent care-of-addresses and AODV just as was done in MIPMANET. However, foreign agent discovery was made a part of AODV while mobile node registrations with the foreign agent via Mobile IP is maintained.

In the paper titled "Wireless Multihop Internet Access: Gateway Discovery, Routing and Addressing" [6], an alternative approach to providing Internet Access for MANETs is describes. In this approach, Mobile IP is eliminated. The use of specific routers that serve as gateways resolve the heterogeneity between the fixed Internet and the MANET. Although the solution looks promising, it was based on IPv6 networks which has not gained popularity over IPv4 networks.

Again, the master's thesis "A Study of Internet Connectivity for Mobile Ad Hoc Networks in NS2" [7] implemented the Internet draft "Global Connectivity for IPv6 Mobile Ad hoc Networks" [8] also known as Global6 in NS2. Global6 is an enhancement on Global4 however, Mobile IP services are not used. The AODV routing protocol was modified to include three different gateway detection mechanisms that were tested using NS2 simulations. The gateway detection mechanisms were classified as proactive, reactive and hybrid. The results from simulations showed no significant difference in packet delivery ratio between the three gateway discovery mechanisms. However, the proactive and hybrid mechanisms performed slightly better than the reactive mechanism regarding average end-to-end delay.

With regards to measuring performance of Mobile IP on MANET, [9] evaluated the performance of Mobile IP on a proactive MANET protocol, DSDV. They created simulation scenarios involving a Home Agent, three Routers, four Foreign Agents, a Corresponding Node and a varying number of Mobile Nodes at varying speeds. Number of nodes were varied to test the robustness of the solution as traffic increases within the simulation space. This was done to rightly model the real Internet world, as we are witnesses to the increasing number of mobile devices on the Internet. They evaluated performance of Mobile IP based on received packets, lost packets, throughput and End-to-End Delay (E2ED).

Also, [10] performed simulations in NS2 to evaluate the performance of Mobile IP on AODV. They used a single Home Agent and one Foreign Agent on a 670 ×670 rectangular field with a few Mobile Nodes. Performance metrics used were throughput, delay and packet overhead. Although the methodology was convincing, the simulation setup was poorly described.

In our thesis, we take these works further as we present on the same simulation platform, scenarios to evaluate the performance of Mobile IP on both reactive and proactive MANET routing on IPv4 networks. Our choice of solution involving corporation between Mobile IP and a MANET routing protocol over methods involving IPv6 is mainly because IPv4 as of now is still widely used over IPv6.

We evaluate the performance of Mobile IP on MANET by performing simulations in Network Simulator 2 (NS2). Likewise in [9], we vary number of nodes in our simulation scenarios to account for increasing traffic within the network.

3. METHODOLOGY

As mentioned earlier, we modelled a Mobile Ad Hoc Network using Network Simulator 2 software. Due to the overwhelming dynamism and cost involved in running a live experiment, we chose a computer simulation model specifically using NS2 which has a widely accepted error margin within the networking research community.

3.1. Simulation Scenario

We used a square flat surface of dimension 670m×670m with a simulation time of 200sec as seen in Figure 3. We modelled the fixed Internet using a Home Agent and four Foreign Agents all with a transmission range of 100m. We placed a gap of 5m between any two adjacent agents. The gap was to ensure that Mobile Nodes decide between Foreign Agents quickly. We simulate up to 175 Mobile Nodes. For each number of Mobile Nodes selected, we performed at least 10 different simulation runs and averaged our results.

Table 1. Simulation Parameters.

Parameter	Value
Movement model	Random waypoint
Traffic type	CBR
MAC layer	802.11
Transmitter range	~100m
Bandwidth	5Mbps
Simulation time	200sec
Simulation field	670m×670m
Number of foreign agents	4
Packet rate	1Mbps
Pause time	0
Maximum speed	5m/s
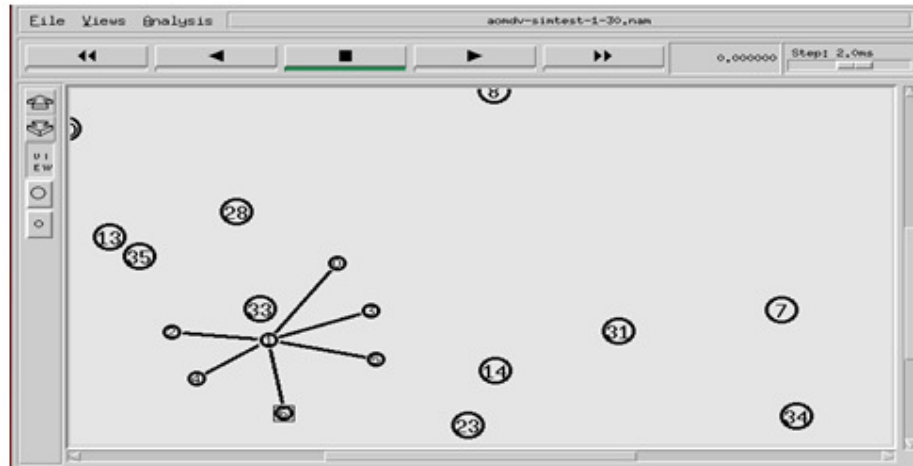
Figure 3. Wired cum wireless scenario.

4. RESULTS AND DISCUSSION
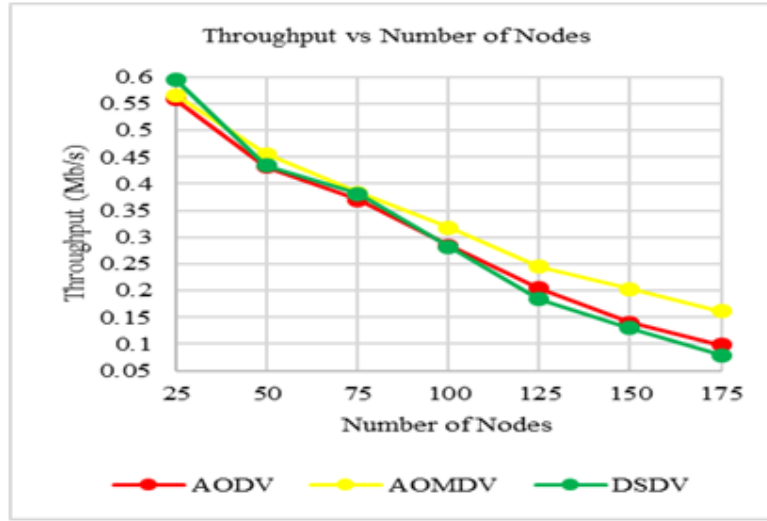
Figure 4. Throughput in Mb/s.

Figure 4 above shows the network throughput as recorded in our simulations. Throughput is a measure of the actual data that can be sent across a channel per unit time. Throughput is usually lesser than channel bandwidth because, channel bandwidth is just a theoretical estimation of how much data could be sent across a channel per unit time. In measuring the network throughput, we excluded control packets and focused on the CBR data packets only. This is sometimes referred to as Goodput. Our results show that, throughput decreases as number of nodes increase. AOMDV had the best throughput performance whereas DSDV had the worst, as number of nodes increased. The observed difference in performance is because, AOMDV sends fewer control messages to nodes than AODV and DSDV and so the channel is less occupied and data packets can be transmitted with ease.
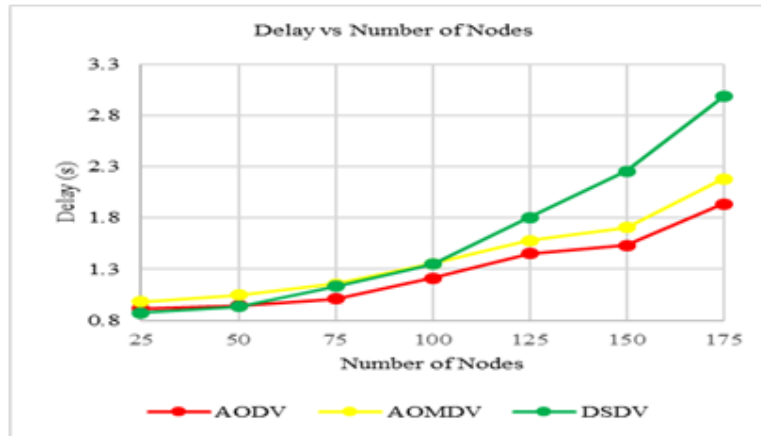
Figure 5. Average end-to-end delay.

Average End-to-End Delay (E2ED) is a measure of how fast a unit size of data can be transmitted across a network. E2ED is a sum of propagation delay, queueing delay, transmission delay and processing delay at intermediate nodes. E2ED is influenced by node mobility and transmission distance. Frequent mobility and longer distances result in longer E2ED. In our simulations, the

same mobility files were used in all three implementations to nullify the effects of mobility and distances when comparing E2ED performances. In all 3 implementations, E2ED increased as nodes increased. AODV and AOMDV showed similar trends as number of nodes increased. Since they are both On-demand routing protocols, path discovery affects E2ED. AOMDV and AODV use the same route discovery mechanism and deliver packets in a similar fashion therefore, within the same mobility scenario they are expected to have the same E2ED. The extra delay recorded by AOMDV is because it spends additional time during each route discovery process to discover multiple routes. DSDV had the worst E2ED performance with a sharp rise in E2ED trend as number of nodes increased. The observed difference is because, intermediate nodes spend longer processing delays as they check their route tables for next hops to forward data, whereas AOMDV and AODV use forward route setups resulting in less processing delays at intermediate nodes.
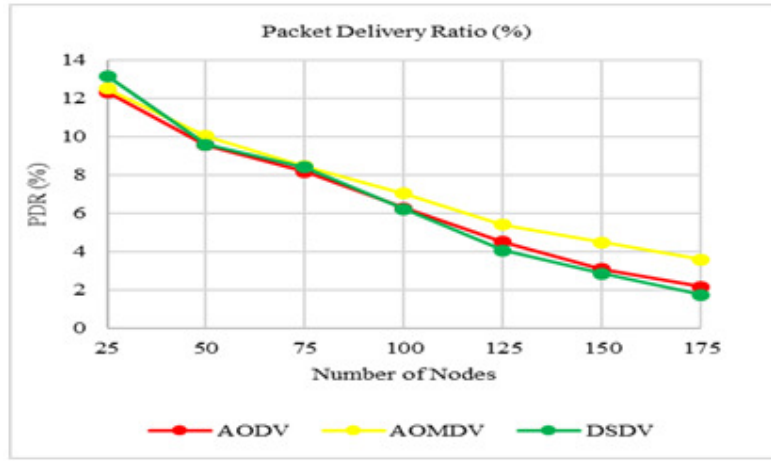
Figure 6. Packet delivery ratio.

Packet Delivery Ratio (PDR) is the ratio of delivered packets to the total number of packets generated. In figure 6 above, we expressed this ratio as a percentage. PDR is inversely correlated to Packet Loss Ratio. A higher packet delivery ratio means fewer packet loss. AOMDV proved to be the most reliable in terms of packet delivery. As mentioned earlier, AOMDV discovers multiple disjoint paths between source and destination. These disjoint paths are less probable to fail all at once hence, a higher reliability in delivering packets from source to destination.
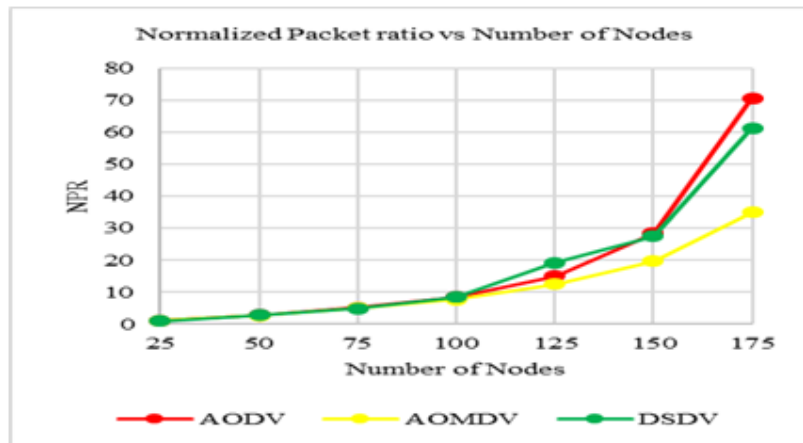
Figure 7. Normalized packet ratio.

Normalized Packet Ratio (NPR) is the ratio of control packets to actual data packets. NPR can be interpreted as the number of control messages needed by the MANET routing protocol to send a single data packet from source to destination. A lesser NPR is most preferred in networks as control messages are not useful to end users. In figure 7, it is observed that, all three protocols have similar trend and values from 25 to 100 nodes. The significant difference in trend begins after 100 nodes. AOMDV had the best NPR performance whereas AODV recorded the worst performance. The observed pattern is attributed to the fact that AODV nodes repeat the route discovery procedure each time a link is broken unlike AOMDV that resorts to a secondary path earlier discovered during the immediate past route discovery phase. It is only when all the multiple paths fail that AOMDV nodes repeat the route discovery procedure.

In figure 8 and figure 9 below, we zoom in on figure 7 to show a better picture of the trends between 25 and 100 nodes.
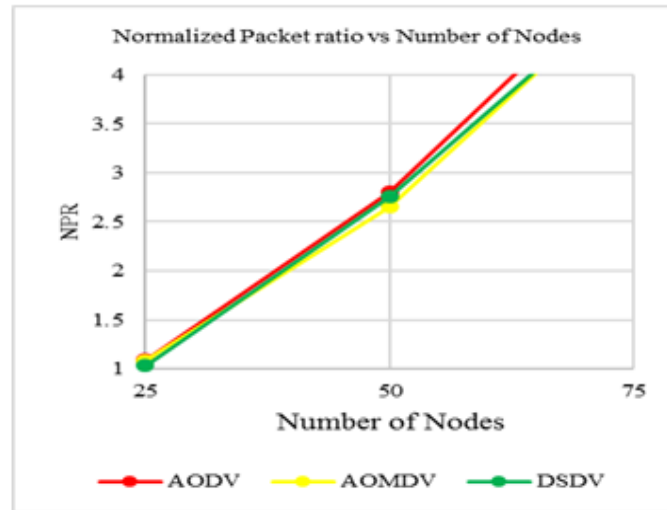
Figure 8. Normalized packet ratio (A).
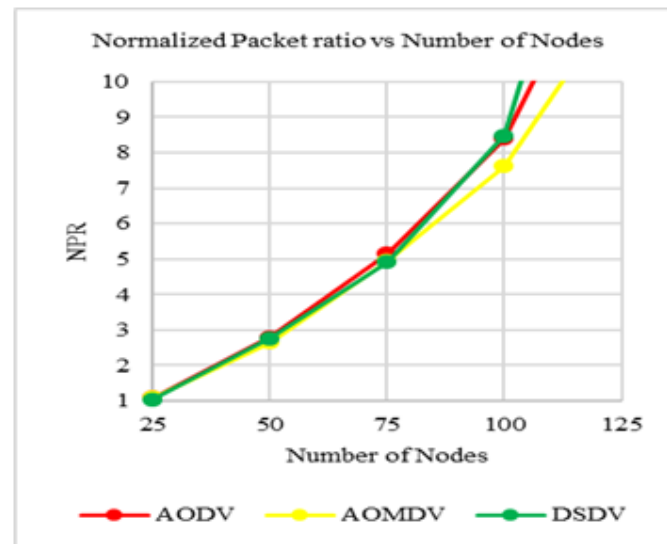
Figure 9. Normalized packet ratio (B).

5. CONCLUSIONS

We have presented a performance evaluation of Mobile IP on MANET. We have studied for the first time the performance of Mobile IP on MANET using Ad Hoc On-demand Multiple Distance Vector (AOMDV) for routing packets inside the MANET. We have also studied the performance of Mobile IP using Destination-Sequenced Distance Vector (DSDV) and Ad Hoc On-demand Distance Vector (AODV) for routing packets inside the MANET.

We modelled Mobile IP on MANET using a wired cum wireless scenario in NS2. Our scenario included one Home Agent, a Corresponding Node and four Foreign Agents. Mobile nodes moved randomly between these sub networks and registered with an agent whenever they entered a new sub network. Communication between Mobile Nodes and Corresponding Node continued even as Mobile Nodes moved across sub networks.

Results obtained from simulation suggested that On-demand routing improved the performance of Mobile IP on MANET regarding the average end-to-end delay, throughput, packet delivery ratio and normalized packet ratio. Specifically, Mobile IP on AOMDV outperformed AODV and DSDV.

ACKNOWLEDGEMENTS

REFERENCES

[1] U. Jönsson, F. Alriksson, T. Lasson, P. Johansson, and G. Maguire, "MIPMANET - Mobile I P for Mobile Ad Hoc Networks," MobiHoc '00 Proc. 1st ACM Int. Symp. Mob. ad hoc Netw. Comput., pp. 75–85, 2000.

[2] C. E. Perkins, "Mobile networking through mobile IP," IEEE Internet Comput., vol. 2, no. 1, pp. 58–69, 1998.

[3] H. Lei and C. Perkins., "Ad Hoc Networking with Mobile IP," in 2nd European Personal Mobile Communications Conf. (EPMCC 97), IEE., 1997, p. 197–202.

[4] U. Jönsson, F. Alriksson, T. Larsson, P. Johansson, G. Maguire, and Jr, "MIPMANET: mobile IP for mobile ad hoc networks," MobiHoc '00 Proc. 1st ACM Int. Symp. Mob. ad hoc Netw. Comput., pp. 75–85, 2000.

[5] E. M. Belding-Royer, Y. Sun, and C. E. Perkins, "Global Connectivity for IPv4 Mobile Ad hoc Networks, IETF Internet Draft." Work in progress, 2001.

[6] J. Xi and C. Bettstetter, "WIRELESS MULTIHOP INTERNET ACCESS : GATEWAY DISCOVERY , ROUTING , AND ADDRESSING," Proc. Int. Conf. Third Gener. Beyond, 2002.

[7] A. A. Hamidian, "A Study of Internet Connectivity for Mobile Ad Hoc Networks in NS 2," no. January, 2003.

[8] R. Wakikawa, J. T. Malinen, C. E. Perkins, A. Nilsson, and A. J. Tuominen, "Global Connectivity for IPv6 Mobile Ad hoc Networks, IETF Internet Draft." Work in progress, 2001.

[9] E. Hassan, A. Alsaied, S. M. Alshareefmodatheir, and I. Hal, "Performance Evaluation of Mobile IP with DSD V Routing Protocol using NS2," 2015.

[10] S. N. Mane, N. V. Mane, and D. G. Khairnar, "Performance of mobile node between different MANET with Mobile IP," 2015 Int. Conf. Ind. Instrum. Control. ICIC 2015, no. Icic, pp. 1662–1664, 2015.

[11] S. Deering and Ed., "ICMP Router Discovery Messages," no. RFC 1256. pp. 1–19, 1991.

[12] C. E. Perkins, "IP Encapsulation within IP," no. RFC 2003. pp. 1–14, 1996.

[13] C. E. Perkins, "Minimal Encapsulation within IP," no. RFC 2004. pp. 1–6, 1996.

[14] D. G. Reina et al., "A survey on ad hoc networks for disaster scenarios," in Proceedings - 2014 International Conference on Intelligent Networking and Collaborative Systems, IEEE INCoS 2014, 2014, pp. 433–438.

[15] R. Ruppe, S. Griswald, P. Walsh, and R. Martin, "Near Term Digital Radio (NTDR) System." IEEE, pp. 1282–1287, 1997.

[16] M. M. N. Aldeer, "A summary survey on recent applications of wireless sensor networks," Res. Dev. (SCOReD), 2013 IEEE Student Conf., no. December, pp. 485–490, 2013.

[17] C. E. Perkins, "Ad Hoc Networking," p. 370, 2000.

[18] C. E. Perkins and P. Bhagwat, "Highly Dynamic (DSDV) for Mobile Computers Routing," Proc. ACM SIGCOMM94, London, UK, pp. 234–244, 1994.

[19] C. E. Perkins and E. M. Royer, "Ad-hoc on-demand distance vector routing," Proc. - WMCSA'99 2nd IEEE Work. Mob. Comput. Syst. Appl., pp. 90–100, 1999.

[20] M. K. Marina and S. R. Das, "Ad hoc On-demand Multipath Distance Vector Routing," ACM SIGMOBILE Mob. Comput. Commun. Rev., vol. 6, no. 3, pp. 92–93, 2002.

AUTHORS

Kingsley K. Ofosu is currently an MPhil student at the Department of Computer Science, University of Ghana. He received his BSc. degree in Psychology and Computer Science (Combined) from University of Ghana in 2013. His research interests include mobile networks, cloud computing and intelligent systems.
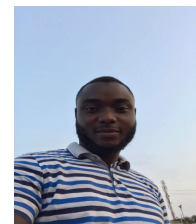
Jamal-Deen Abdulai received his BSc. degree in Computer Science in 2002 from the Kwame Nkrumah University of Science and Technology (KNUST). In 2006, he was awarded an MPhil and later in 2009 a PhD in Computer Science by the Department of Computing Science at the University of Glasgow, UK.

Dr. Abdulai's is currently a lecturer at the Department of Computer Science, University of Ghana, investigating how probabilistic methods can be used to optimize the performance of both wired and wireless networks. His research interests include Performance modelling and evaluation of Mobile Wireless Ad hoc and Sensor Networks, Network Security and Management, Embedded Systems, Parallel and Distributed Systems, Artificial Intelligence and its application in information security.

Ferdinand A. Katsriku is the current head of the department of computer science, University of Ghana. In 1989, he received an M. Eng. in computer systems engineering with distinction from Kharkov Polytechnic Institute. He was awarded an MSc in Laser Engineering and Pulsed Power Technology in August 1992 and later awarded postgraduate certificate in education from King's College London in 1996. In 2000, he received a PhD in Information Engineering from City University, London. His current research interests are wireless sensor networks, cognitive radio and water quality monitoring. He is also a technical Reviewer of EPSRC grants and for major publishers including John Wiley and Elsevier.

INTENTIONAL BLANK

ON THE DUALITY FEATURE OF P-CLASS PROBLEMS AND NP COMPLETE PROBLEMS

WenhongTian[1, 2]

[1]University of Electronic Science and Technology of China,
Chengdu, China.
[2]Chongqing Institute of Green and Intelligent Technology,
Chinese Academy of Science.

ABSTRACT

In term of computational complexity, P-class (abbreviated as P) problems are polynomial-time solvable by deterministic Turing machine while NP complete (abbreviated as NPC) problems are polynomial-time solvable by nondeterministic Turing machine. P and NPC problems are regularly treated in different classes. Determining whether or not it is possible to solve NPC problems quickly, is one of the principal unsolved problems in computer science today. In this paper, a new perspective is provided: both P problems and NPC problems have the duality feature in terms of computational complexity of asymptotic efficiency of algorithms.

KEYWORDS

P Problems; NP problems; NP Complete Problems; the P versus NP Problem; The Duality Feature.

1. INTRODUCTION

In 1971, Cook [1] firstly established a theorem that a class of problems can be P-reducible (polynomial time reducible) to each other and each of them can be P-reducible to Boolean Satisfiability (SAT) problem, this class of problems is called NP (nondeterministic polynomial time) problems. Karp [2] applied Cook's 1971 theorem that the SAT problem is NP Complete (also called the Cook-Levin theorem) to show that there is a polynomial time reduction from the SAT problem to each of 21 combinatorial problems, thereby showing that they are all NP complete (NPC). This was one of the first demonstrations that many natural computational problems occurring throughout computer science are computationally intractable, and it drove interest in the study of NP-completeness and the P versus NP problem [3]. The P versus NP problem, determining whether or not it is possible to solve NP problems quickly, is one of the principal unsolved problems in computer science today and listed as one of seven millennium problems [3,4], challenging tens of thousands of researchers.

Simply speaking, P problems mean that the class of problems can be solved exactly in polynomial time while NPC problems stands for a class of problems which might not be solvable in polynomial time. NPC problems has far-reaching consequences to other problems in mathematics, biology, philosophy and cryptography. More specifically, in Big O-notation (refer

to Definition 1 and 2) of computational complexity in asymptotic efficiency of algorithms, P problems can be solved in polynomial time of $O(n^k)$ for some constant k where n is the size of input to the problem, while NPC problems may have computational complexity of $O(2^{cn})$ including both exponential time and sub-exponential time, where c is a positive constant larger than zero.

Karp [2] ever claimed that if any of NPC problems have efficient polynomial time algorithms, then they all do. It is for this reason that research into the P versus NP problem centers on NPC problems, i.e., looking for efficient polynomial time algorithms for NPC problems. Through decades' efforts by many researchers, it is still an open question. There are quite many results but none of them is commonly accepted yet by the research community.

The current author also classified NPC problems based on their natures and other methods [5] and found that NPC problems are not equivalent in computational complexity.

In the following, we show a new perspective: both P problems and NPC problems have duality features of each other in terms of computational complexity of asymptotic efficiency of algorithms, especially by considering the representation of the input.

2. PROBLEM FORMULATIONS

The following definitions are based on the computational complexity of different problems in the worst case.

Definition 1: The asymptotic efficiency of algorithms [6,7]: concerns with how the running time of an algorithm increases with the size of the input *in the limit*, as the size of the input increases without bound.

Definition 2: The O-notation of computational complexity of an algorithm: asymptotically bounds a function from above within a constant factor [6,7]. For a given function *g(n)*, we denote by O(*g(n)*) :

O(*g(n)*)={*f(n)*: there exist positive constants c and n_0 such that $0 \leq f(n) \leq cg(n)$ for $n \geq n_0$}.

Fig.1 shows the intuition behind O-notation. For all values n to the right of n_0, the value of function *f(n)* is on or below *g(n)*.
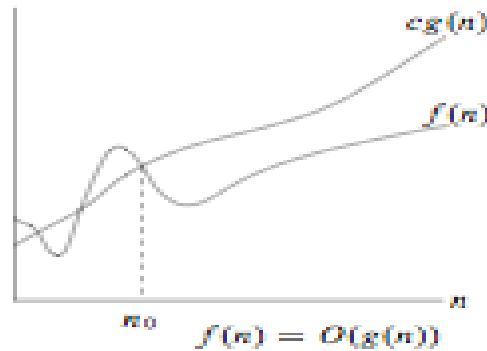
Figure 1.f(n)=O(g(n)) [6]

Since O-notation describes an upper bound, when we use it to bound the worst-case running time of an algorithm, we have a bound on the running time of the algorithm on every input. For example, the doubly nested loop structure of the insertion sort algorithm has an $O(n^2)$ upper bound; equivalently, we mean that the worst-case running time is $O(n^2)$.

Definition 3: P-class problems in term of O-notation of computational complexity: in the worst case, their exact algorithms have computational complexity of $O(n^k)$ for some constant k where n is the size of input to the problem.

Definition 4: NPC problems: in the worst case, their exact solutions may have computational complexity of $O(2^{cn})$ **in O-notation**, where n is the size of input to the problem and c is a positive constant.

Table 1 provides a summary of variables used in this paper. Table 2 shows computational complexities of exact solutions to some NPC problems (abstracted from [8]).

Table 1. The summary of variables

Variables	Meaning
k, c, n_0	a positive constant
n,e	The size of inputs to a problem
s, m	The number of bits (binary)
B_0	**min**(2^{cn}, n^k, 2^m, 2^s)
W, C, b	a positive number

Table 2. The complexities of exact solutions to some NPC problems from [8]

Problem	Complexity	Genre
MIS (maximum independent set)	1.1996^n	Packing
Set covering problem (SCP) *	1.1996^n	Covering
3CP (3-coloring of planar graph)	1.3446^n	Partitioning
SSP (subset sum problem) [7]	nW	Numerical
3SAT	1.4802^n	Satisfaction
TSP	$n^2 2^n$	Sequencing

*: derived from MIS since SCP and MIS are complementary [7].

We also consider that the representation of the input to the problem by *m* bits (in binary) and the space complexity of a problem by *s* bits (in binary) in memory**.**

Fact 1: A regular system (the computer hardware, called the system in the following) has capability of handling B_0 numbers in term of computational and space complexity, i.e., handling the input efficiently within reasonable time (may be within a few hours or minutes depending on the applications) without overflowing the system, where B_0 may be related to the memory size or whichever (CPU, memory, disk et.) is the bottleneck of the system.

Based on the definitions above, the capabilityB_0can be represented by

$$B_0 = \boldsymbol{min}(2^{cn}, n^k, 2^m, 2^s) \tag{1}$$

3. RESULTS

Lemma 1. Dynamic programming algorithm for Subset Sum problem (SSP) is pseudo-polynomial time.

Proof. removed∎

Lemma 2. Any NPC problem can be reduced to SSP.

Theorem 1:The instances of SSP can have computational complexity of NPC problems and P-class problems, this is called the duality feature in this paper.

Proof: removed∎

Fact 2: The instances of other NPC problems also have the duality feature.

Proof: removed.∎

In Table 3, records of optimum solutions to TSP problem are provided. One can see that the size of the problem is increasing as year goes. Notice that the instance with 1904711 nodes is still not yet solved exactly but just has a good lower bound [14]. Similar results are also observed for other NPC problems. For real-life problems with small or moderate number of variables, they can be solved to exactly very easily (LKH [15] and Concorde [16] are considered as two of the best such solvers which can now find exact solutions for medium size TSP Problems). However, if the problem size increases to very large, finding efficient solutions to NPC problems just become intractable.

Theorem 2: The P-class Problems can have the duality feature in terms of computational complexity.

Proof: removed. ∎

Remarks: Actually SSP can be treated as polynomial time solvable problem (P-class problem) when both n and W is not very large while it is of exponential complexity when both n and W are very large.

Observation 1: Even one can find efficient solution (polynomial time algorithm) to one or more NPC problems, some P problems may become exponential time solvable in computational complexity when the representation of the input to them become very large.

4. DISCUSSIONS AND CONCLUSION

NPC problems have different natures, they can be classified into six basic genres [2, 7], i.e., Satisfaction, Packing, Covering, Partitioning, Sequencing, Numerical computing.

Originally, PRIMES and Graph Isomorphism were hard to determine to be in NPC or not in Cook's paper [1]. In 2004, PRIMES is found in P-class and accepted by the research community [10]. And in 2015, Graph isomorphism is reported to have Quasipolynomial time solution [11,

12], though the results are still under verification. These show some new perspectives and trends on NPC problems.

As another perspective, it is recently proved mathematically that memcomputing machines (a novel non-Turing paradigm) have the same computational power of nondeterministic Turing machines [13]. Therefore, they can solve NPC problems in polynomial time with resources that only grow polynomially with the input size.

NPC problems and the P versus NP problem challenge many researchers to tackle them through decades of efforts. In this paper, a new perspective is provided: both P problems and NPC problems have the duality feature in terms of computational complexity of asymptotic efficiency of algorithms.

Table 3.   Records of optimum solutions to TSP problems [14] where *n* is the number of nodes in TSP. All TSP problems in the table are solved to optimum except for the last one

n	Year (solved)	Node type
48	1954	USA cities
64	1971	random nodes
80	1975	random nodes
120	1977	Germany cities
318	1987	cities
532	1987	USA cities
666	1987	World cities
1002	1987	cities
2392	1987	cities
3038	1992	cities
13509	1998	USA cities
15112	2001	cities
24978	2004	Sweden cities
85900	2006	cities
100000	2009	Japan
1904711	2010*	World TSP Challenge

ACKNOWLEDGMENTS

REFERENCES

[1]   S.A. Cook (1971). The Complexity of Theorem Proving Procedures, Proceedings of the third annual ACM symposium on Theory of computing. pp. 151158, March of 1971.

[2]   R. M. Karp (1972), Reducibility Among Combinatorial Problems, In R. E. Miller and J. W. Thatcher (editors). Complexity of Computer Computations. New York: Plenum. pp. 85-103.

[3]   Clay   Mathematics   Institute,   http://www.claymath.org/millennium-problems/millennium-prize-problems.

[4]   Wikipedia, http://en.wikipedia.org/wiki/NP-complete.

[5]   WenhongTian, NP Complete Problems Are not All Equivalent, under review, submitted 2017.

[6]   T. Cormen,C. Leiserson, R.L. Rivest,C. Stein, Introduction to Algorithm, 2nd edition, MIT Press,
       2004.

[7]   J. Kleinberg, E. Tardos, Algorithm Design, 2006, Pearson Education Asia Limited, pp. 475-479.

[8]   G. J. Woeginger, Exact algorithms for NP-hard Problems: A Survey, Combinatorial optimization -
       Eureka, you shrink! Pages 185 - 207 , Springer-Verlag New York, Inc. New York, NY, USA ©2003

[9]   WenhongTian, C. Huang, X. Wang, A Near Optimal Approach for Symmetric Traveling Salesman
       Problem in Eclidean Space, In Proceedings of ICORES 2017, Portugal, also available at arxiv
       https://arxiv.org/pdf/1502.00447.pdf

[10]   M. Agrawal, N. Kayal, N. Saxena, PRIMES is in P, The Annals of Mathematics, Pages 781-793 from
       Volume 160, Issue 2, 2004.

[11]   L. Babai, Graph Isomorphism in Quasipolynomial Time, https://arxiv.org/abs/1512.03547, Dec 11,
       2015.

[12]   A. Cho , Science news, mathematician-claims-breakthrough-complexity-theory,
       http://www.sciencemag.org/news/2015/11/mathematician-claims-breakthrough-complexity-theory,
       Nov. 11, 2015.

[13]   F. L. Traversa, C. Ramella, F. Bonani and M.D. Ventra, memcomputing NP-complete problems in
       polynomial time using polynomial resources and collective states, Science, Vol. 1, no. 6, e1500031,
       Nov. 2015.

[14]   W. Cook, In Pursuit of the Traveling Salesman, Princeton University Press, 2012.

[15]   LKH codes, http://www.akira.ruc.dk/~keld/research/LKH/

[16]   Concorde codes, http://www.math.uwaterloo.ca/tsp/concorde.html

PERCEPTION OF HOME-SHARING INDUSTRY: A STUDY OF COMPARING AIRBNB AND COUCHSURFING

JinTao Huang, KangNyeon Yi and JiHye Han

Department of Interaction Science, SungKyunKwan University,
Seoul, South Korea

ABSTRACT

The aim of this paper is to examine people's perceptions of home sharing industry through comparing Airbnb and Couchsurfing, using python to crawling twitter's data. We conducted semantic network analysis by using UCINET, which is embedded with NETDRAW, for calculating betweenness centrality and visualizing semantic network based on multidimensional scaling(MDS). And we also used LIWC to analysis public sentimental perceptions. The present study's results show that the characteristics discovered about Airbnb and Couchsurfing in cyberspace have possible future directions in view of word usage frequency, centrality and semantic networks. Moreover, the results show that in sentimental aspects, there are different public trends of emotions on Airbnb and Couchsurfing. Through those results, researchers provide information to understand which sectors should entrepreneurs put more efforts and money.

KEYWORDS

Sharing Economy, Home-sharing, Airbnb, Couchsurfing, Semantic Network Analysis

1. INTRODUCTION

Sharing is a phenomenon as old as humankind, while collaborative consumption and the "sharing economy" are phenomena born of the Internet age (Russell, 2014). The sharing economy is a socio-economic system that coordinates "the peer-to-peer-based activity of obtaining, giving, or sharing the access to goods and services" through "community-based online services" (Hamari et al., 2015). This economy system is a new cooperative production and consumption mode that enables the sharing of various goods and services by virtue of internet. These types of economies have recently become one of the important forms of achieving sustainable development by use of scientific and technological innovations. And especially, we focus on space sharing, which is one of the sharing economy. Space sharing includes releasing the information and prices of idle properties through various platforms on the Internet for tourists, people who are taking business trips, and other people with a need for, or those who are looking for housing, to select (Xinyuea et al., 2016). The home sharing economy already has a foothold in the vacation rental market, which accounted for 9% of the traditional hotel market in 2014. This market is now becoming more organized and gaining more widespread use as trust in these services increases (Saussier, 2015).

However, not many sharing economy platforms survive and only a few are economically successful. The success of these platforms in developing from start-ups to sustainable businesses depends on their ability first to mobilize initial user networks and then to maintain users' active participation and ensure growth of the user base sufficient for sustainable revenue generation (Constantiou et al., 2016). We investigate the case of Airbnb and Couchsurfing, which is the mainstream of home sharing. Airbnb and Couchsurfing are sharing platforms network in the home sharing industry. Our study is about examine public sentiments on these home-sharing companies. We could find public positive and negative emotion about Airbnb and Couchsurfing in Twitter from 2010 to 2017. We discover the different characteristics and identify the historical Changes of tweets. Using UCINET, the authors analysis the betweenness centrality and visualizing semantic network based on multidimensional scaling(MDS). Our study provides public perceptions of home sharing industry through comparing Airbnb and Couchsurfing.

2. LITERATURE REVIEW

2.1. Airbnb

Airbnb is an online community for vacation rentals that allows travelers and accommodation providers to list, find and rent lodging. Airbnb was founded in 2008 in San Francisco. With the increasing demands in travel accommodation, the company discovered the great potential of this market; thus, Airbnb was then created to provide low-cost and unique travel experience to capture the market demand. Airbnb has focused on the economic impact they can bring internationally and endeavors to align it with the goal of maintaining the current development in the neighborhoods in which they operate. On average, Airbnb generated more than $300 million in economic activity in several major cities, such as New York, Paris, Berlin, and London (Chan et al., 2016).

2.2. Couchsurfing

Couchsurfing is a non-profit organization that 'seeks to internationally network people and places, create educational exchanges, raise collective consciousness, spread tolerance and facilitate cultural understanding' (Couchsurfing, 2006). The project was originally launched in 2003 (Molz & Jennie Germann, 2013). This community works on the premise that a user has potential friends around the world. These friends are hosts who are willing to let a stranger sleep over at their home rather than staying at a hotel. People can find hosts and share their experiences via the website, which is made up of a community of 10 million people in 200,000 cities (Global Investor, 2015).

2.3. Applicable theory

The authors found applicable theories for applying the results.

Cognitive dissonance theory

"Dissonance is aroused whenever a person engages in an unpleasant activity to obtain some desirable outcome. From the cognition that the activity is unpleasant, it follows that one would not engage in the activity; the cognition that the activity is unpleasant is dissonant with engaging

in the activity. Dissonance should be greater, the greater the unpleasant effort required to obtain the outcome." (Harmon-Jones & Mills, 1999)

Consider someone who stays an expensive Airbnb house but discovers that it is not comfortable to stay long. Dissonance exists between their beliefs that they have paid a good Airbnb house and that a good house should be comfortable. Dissonance could be eliminated by deciding that it does not matter since the house is mainly used for short trips (reducing the importance of the dissonant belief) or focusing on the house's strengths such as safety, appearance (thereby adding more reasonable beliefs). Thus people can rationalize their conduct by thinking so.

2.4. Mere exposure effect

The mere exposure effect is a psychological phenomenon by which people tend to develop a preference for things or people that are more familiar to them than others. Repeated exposure increases familiarity. This effect is therefore also known as the familiarity effect. Therefore misattribution of positive feelings (due to fluency) to the ad or brand comes into effect. This means that with mere exposure the more often you see an information the better you can process it and the more fluent it is. Human beings like easy processing rather than difficult cognitive processes so they stick with familiar things. This may be the explanation for their positive feelings towards an ad or brand (Falkenbach et al., 2013). When people evaluate Airbnb and Couchsurfing, there are many factors why people evaluate that way. Among them, mere exposure effect theory can be a factor. We can see Airbnb's commodity advertisement a lot more than Couchsurfing. Through the advertisement, people are exposed several times with Airbnb, and it affects raising people's positive evaluation about Airbnb because of the familiarity that improved by exposure.

2.5. Research questions

RQ 1. *Semantic Networks*. In the datasets from tweets "Airbnb" and "Couchsurfing", what were the different characteristics discovered?
RQ 1-1. In view of word usage frequency, centrality and semantic networks, what were the characteristics found in tweets about "Airbnb" and "Couchsurfing" respectively?
RQ 1-2.As time went by, what were the remarkable differences discovered from tweets about "Airbnb" and "Couchsurfing" respectively?
RQ 2. What are the theoretical and practical implications of the study's findings?

3. METHODS

3.1. Data gathering

The datasets were retrieved from Twitter by using Python. We gathered past tweets from Twitter through a python library. Twitter is an online news and social networking service which users post and interact with messages, and it is frequently used by the general public. In addition, to investigate how the home-sharing industry has developed since related enterprises showed up, we used two keywords, "Airbnb" and "Couchsurfing", which are two representative companies in the home-sharing industry. The time periods of datasets were divided into four groups respectively: before 2015(2010-2014), 2015, 2016 and 2017.

3.2. Semantic network analysis

This study employed semantic network analysis (SNA), which is a numerical and graphical measurement method for a variety of text data. Semantic network analysis is a semi-automated text analysis for determining the most frequently used words and the differences among texts in terms of word frequency and clusters (Lee, Kim, & Rosen, 2009; Doerfel & Barnett, 1999). Python was employed for collecting data from Twitter. 'WORDij' (Danowski, 2010) was also employed to convert text data from Twitter, into co-occurrence matrices. 'WORDij' is word-pair based software for semantic structure retrieval without any prior categories (Danowski, 2010). Moreover, according to Danowski(2010), word pairs are made based on word proximity which is calculated by a "window that slides through the text, counting all word pairs inside as it moves from word in the full text". For calculating betweenness centrality and visualizing semantic network based on multidimensional scaling (MDS), we used UCINET6 for calculating centrality from co-occurrence matrix. Besides, we also used NETDRAW, which is embedded from UCINET6, for drawing semantic network based on co-occurrence matrix and centrality.

4. RESULTS

Table 1 indicates top 20 words based on frequency and betweenness centrality from the data sets (Tweets gathered from Twitter based on keyword 'airbnb' from 2010to 2017). In the first period (before 2016) 'airbnb', 'uber', 'travel', 'sharing',' business', 'economy' locate at top 6 words in order of frequency. Because 'airbnb' is the search keyword of data collecting process, it always locates at the first order on tables of frequency and centrality. 'uber' has the second highest order at both frequency and centrality. In the second period (2016) 'airbnb', 'uber', 'travel', 'sharing',' business', 'economy' locate at top 6 words in order of frequency, and 'airbnb', 'app', 'travel', 'ceo', 'twitter', 'uber' locate at top 6 words based on betweeness centrality. In the third period (2017) 'airbnb', 'travel', 'hosts', 'first', 'hi', 'dm'(direct message) locate at top 6 words by frequency, and 'airbnb', 'first', 'dm' (direct message), 'help', 'users', 'man' locate at the upper 6 words based on centrality. Table 2 indicates the upper 20 words based on frequency and betweeness centrality from the data sets (Tweets gathered from Twitter based on keyword 'couchsurfing' from 2010 to 2017). In the first period (before 2016) 'couchsurfing', 'travel', 'site', 'people', 'airbnb', 'moscow' list on the table from the highest frequency to 6th highest frequency, and 'couchsurfing', 'data', 'tour', 'business', 'check', 'app' locate at top 6 highest centrality. In the second period (2016) 'couchsurfing', 'travel', 'moscow', 'host', 'airbnb', 'world' locate at top six based on frequency, and 'couchsurfing', 'people', 'lot', 'better', 'Italian', 'trvel', list on table 3 based on centrality order.

Figure 1 is about semantic network of Airbnb from Twitter. 'uber', 'sharing', and, 'economy' have high frequency and betweenness centrality. There are many names of cities such as 'Amsterdam', 'San Francisco', 'New York' in the network. There are words related to a case of hidden camera issue such as 'german', 'hidden', 'camera', 'suing'. In 2015, a german woman argued that she was unaware of the camera, which allegedly captured both naked footage and private conversations. In addition, there are four clusters: the first cluster (A) is about 'Airbnb'-relevant issues; the second cluster (B) stands for user experience on mobile app and web site; the third cluster (C) is about trip-relevant startups; the fourth cluster (D) describes car-sharing industry (Uber).

Table 1. Frequency and centrality of 'Airbnb' data from 2010 to 2017.

Airbnb	2010-2015		2016		2017	
	Frequency	Centrality	Frequency	Centrality	Frequency	Centrality
1	airbnb	airbnb	airbnb	airbnb	airbnb	airbnb
2	uber	Uber	travel	app	travel	first
3	travel	Mashable	ceo	travel	hosts	dm
4	sharing	Birdbnb	twitter	ceo	first	help
5	business	Money	uber	twitter	hi	users
6	economy	Sharing	bitcoin	uber	dm	man
7	stay	Economy	needs	bitcoin	study	Managed
8	hosts	tech	bookings	most	reject	hear
9	love	Order	grow	users	guests	Code
10	company	trip	full-fledged	chesky	email	Study
11	amsterdam	business	hub	gift	help	Case
12	hotel	amsterdam	hosts	business	thanks	Interiordesign
13	startup	hosts	app	housing	stay	Baboons
14	apartment	share	rent	customers	business	Disabilities
15	place	site	stay	start	rent	Reject
16	san	mobile	hotel	hosts	finds	People
17	klm	hotel	business	tenants	disabilities	travel
18	tourist	host	first	hub	trip	Hosts
19	nyc	app	updates	grow	people	Business
20	rent	people	front	Full-fledged	love	Email

Table 2. Frequency and centrality of 'Couchsurfing' data from 2010 to 2017.

| Couchsurfing	2010-2015		2016		2017	
	Frequency	Centrality	Frequency	Centrality	Frequency	Centrality
1	couchsurfing	couchsurfing	couchsurfing	couchsurfing	couchsurfing	couchsurfing
2	travel	Data	travel	people	travel	Travel
3	site	Tour	moscow	lot	host	Airbnb
4	people	Business	host	better	traveling	City
5	airbnb	Check	airbnb	italian	experience	Host
6	moscow	app	world	Travel	airbnb	Experience
7	check	Her	people	Italy	first	Traveler
8	great	Shes	italian	Couches	people	First
9	host	Building	site	Policeman	ttot	ttot
10	website	Ever	first	Rape	friends	Added
11	love	Hook-up	couch	Site	moscow	Ways
12	friends	travel	photo	Host	great	People
13	time	use	great	World	check	Best
14	thanks	sharing	friends	Moscow	best	Why
15	chrome	management	journey	Airbnb	free	Youtube
16	app	brand	time	First	world	Casey
17	couch	fail	round	Website	time	free
18	good	poor	experience	Women	meet	Fenton
19	android	kill	photographers	Accused	couch	Friends
20	cs	airbnb	sleep	News	love	Fiona
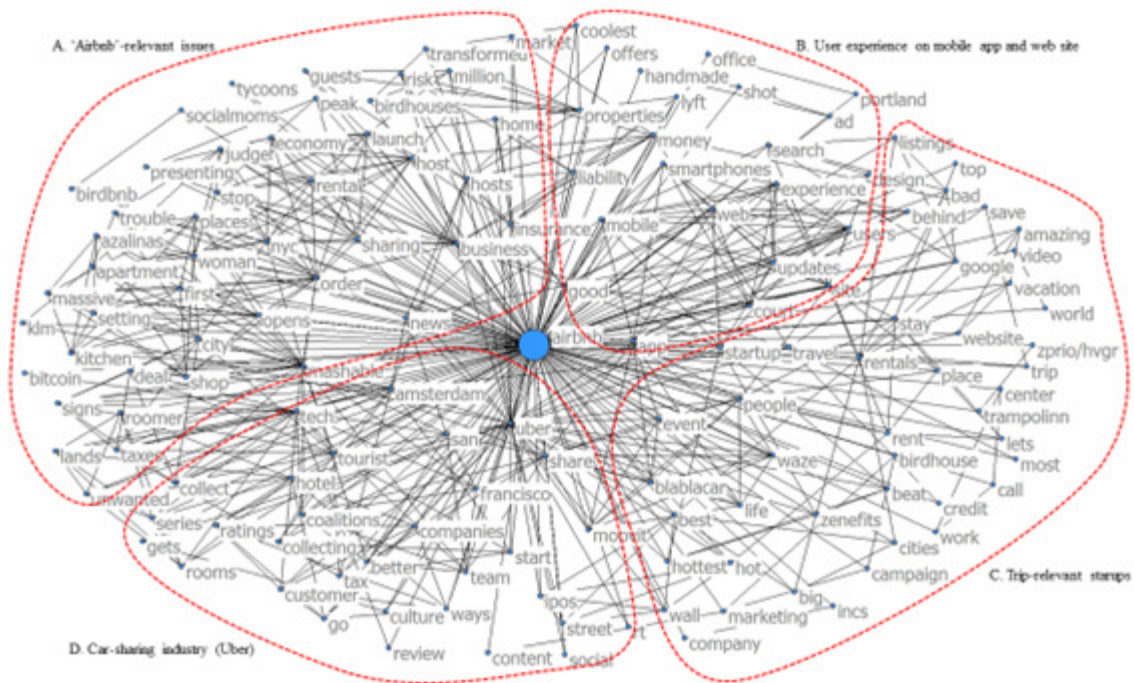
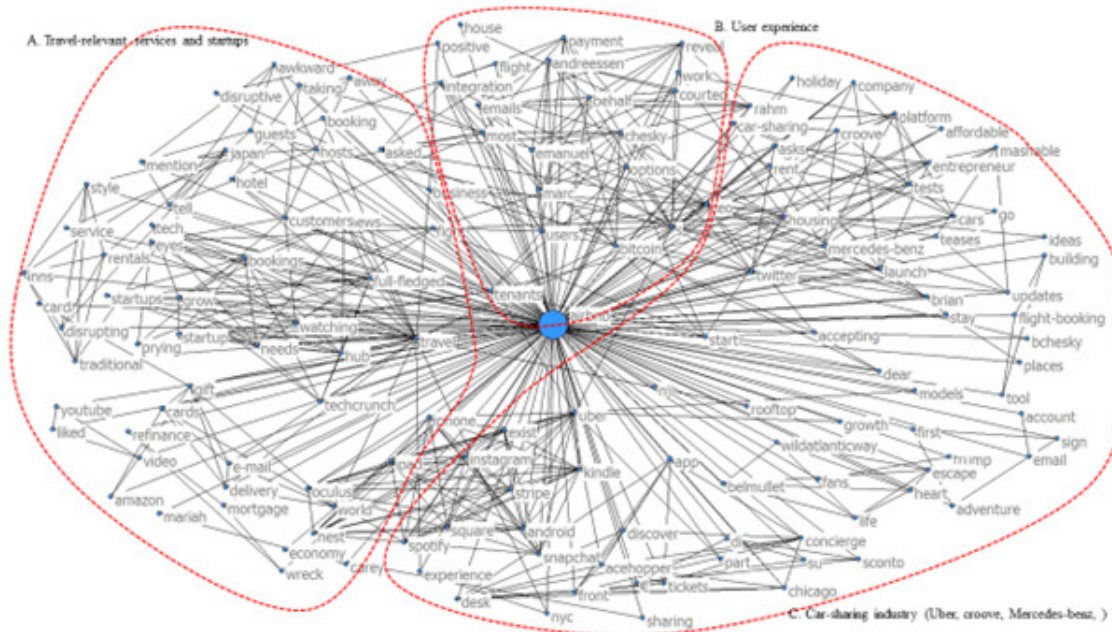Figure 1. Semantic Network from Twitter on 'Airbnb' from 2010 to 2015

Figure 2. Semantic Network from Twitter on 'Airbnb' in 2016

There are words relevant to a suit case about contract on lease of apartment such as 'tenants', 'fights' in Figure 2. There are words about blockchain technology such as 'bitcoin'. There are three clusters: the first cluster (A) depicts travel-relevant services and starups; the second cluster (B) stands for user experience; the third cluster (C) is about car-sharing industry (Uber, Crrove, Mercedes-benz).

There are words about discrimination on disabilities from academia such as 'disabilities', 'researchers', 'study' in Figure 3. There are three clusters: the first cluster (A) describes house-relevant comments (interior design, architecture, decoration); the second cluster (B) stands for'Airbnb'-relevant issues including enter china market; the third cluster (C) is about travel-relevant reports and perception.
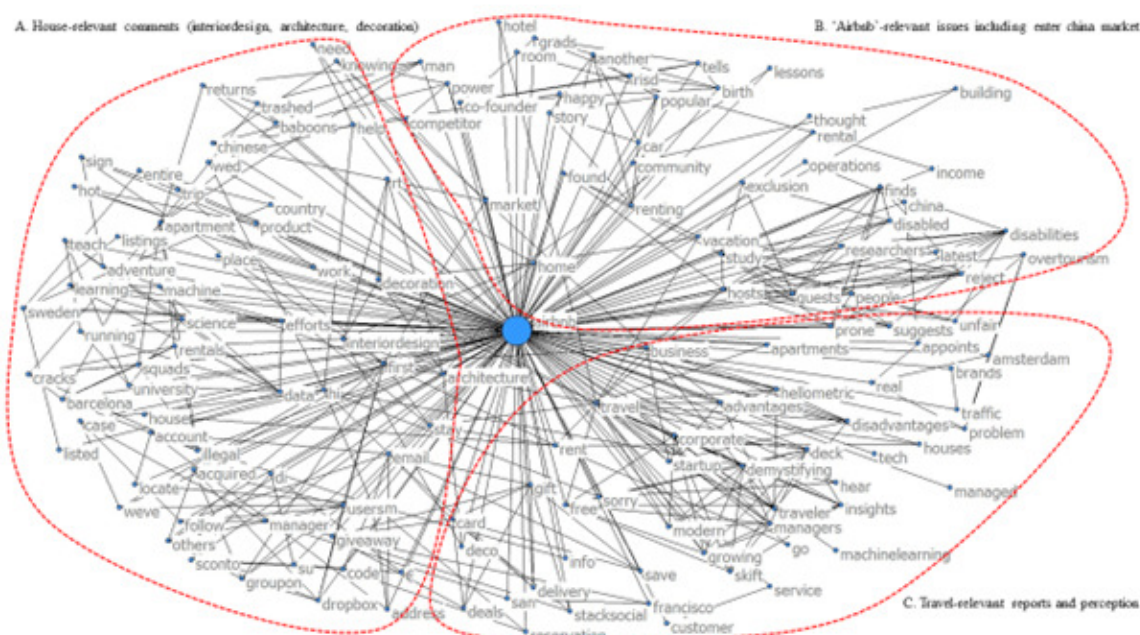
Figure 3. Semantic Network from Twitter on 'Airbnb' in 2017

There are words related to sex culture such as 'hook-up', 'secret', 'sex' in Figure 4. There are names of cities such as 'Moscow', 'San Francisco', 'Amsterdam'. There are three clusters: the first cluster (A) is about violence-relevant news; the second cluster (B) describes user experience (os, android, site, apps, web, etc.); the third cluster (C) stands for Competitors and other starups.

The word 'ttot' is abbreviation of 'Travel Talk on Twitter'. There are words related to a murder case in Nepal such as 'Nepal', 'American', 'murdered' in Figure 5. There are three clusters: the first cluster (A) depicts travel-relevant issues; the second cluster (B) is about violence-relevant news; the third cluster (C) stands for competitors and other startups.

There are many words relevant to travel experience such as 'experience', 'thailand', 'sudan', 'cambodia', 'iran', 'toronto', 'quebec', 'san francisco', 'europe', 'russia' in Figure 6. There are three clusters: the first cluster (A) describes travel experiences; the second cluster (B) is about violence-relevant news; the third cluster (C) stands for competitors-relevant issues.

Figure 4. Semantic Network from Twitter on 'Couchsurfing' from 2010 to 2015

Figure 5. Semantic Network from Twitter on 'Couchsurfing' in 2016
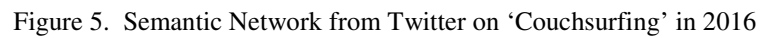
Figure 6. Semantic Network from Twitter on 'Couchsurfing' in 2017

5. CONCLUSIONS

By analysing the semantic networks of text data from Twitter and the words which have high frequency and centrality in the networks from four different time periods, the current study seeks to figure out how home sharing economy change over time from early stage to 2017. In most semantic networks in this study, some kind of flash and horrible event including murder or sexual assault cases has been men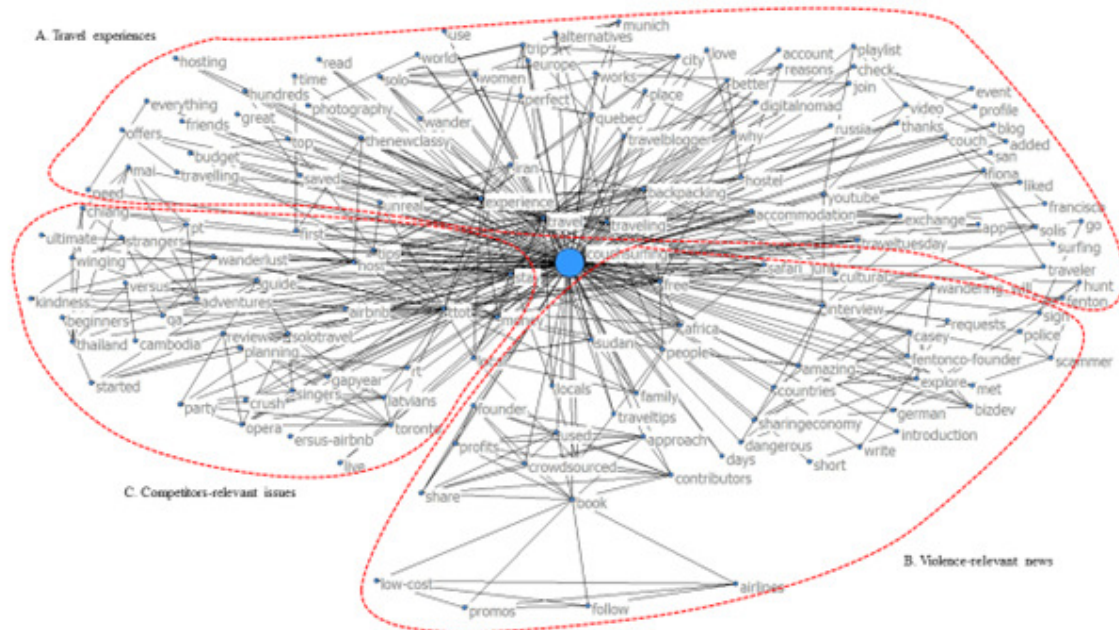tioned frequently. As a result, their frequency and centrality are high. In the semantic networks from keyword 'airbnb', 'uber' appears frequently, because 'uber' is another popular sharing economic service for travelling. In the networks from search keyword 'couchsurfing', 'airbnb' has often been mentioned, because 'airbnb' service hosts can provide their houses to travelers for 'couchsurfing'. Since 2016 the word 'Youtube' has been mentioned frequently than ever in the networks on 'couchsurfing'. Perhaps a lot of couchsurfing hosts make Youtube main advertisement tools. Additionally, there are more violence-relevant issues about 'couchsurfing' than 'airbnb'.

Time span of data from twitter in this study is not that large enough to cover the development periods of Airbnb and couchsurfing. In addition, Twitter is not only one representative social media website so that the dataset of universality had limitations. For further development, time span of data should expand to around 10 years ago. The variety of data sources should be used by not only twitter but also facebook, reddit and other popular social media websites. Deep Learning method based on Natural Language could develop Semantic Network Analysis including clustering and categorizing data set. (e.g., Deep Learning Open Source including Google Tensor Flow, etc). Innovation theories including 'Diffusion of innovation theory' could be employed in the further study.

REFERENCES

[1]   Belk, Russell. "You are what you can access: Sharing and collaborative consumption online." Journal of Business Research 67.8 (2014): 1595-1600.

[2]   Borgatti, S. P., Everett, M. G., & Freeman, L. C. (2002). UCINET 6 for Windows. Harvard: Analytic Technologies, 185.

[3]   Borgatti, S. P. (2005). Centrality and network flow. Social networks, 27(1), 55-71.

[4]   Chan, Cheuk, et al. "BUSINESS 478 SECTION E200." (2016).

[5]   Constantiou, Ioanna, Ben Eaton, and Virpi Kristiina Tuunainen. "The Evolution of a Sharing Platform into a Sustainable Business." System Sciences (HICSS), 2016 49th Hawaii International Conference on. IEEE, 2016.

[6]   CouchSurfing Project Web 2.0 (2006) Our Mission. Available at: http://www.couchsurfing.com/mission.html (consulted November 2008).

[7]   Danowski, J. A. (2010). WORDij 3.0 [computer program]. Chicago: University of Illinois at Chicago.

[8]   Doerfel, M. L., & Barnett, G. A. (1999). A semantic network analysis of the International Communication Association. Human communication research, 25(4), 589-603.

[9]   Falkenbach, Katja, et al. "Mere Exposure Effect." (2013).

[10]   Freeman, L. C. (1977). A set of measures of centrality based on betweenness. Sociometry, 35-41.

[11]   Hamari, Juho, Mimmi Sjöklint, and Antti Ukkonen. "The sharing economy: Why people participate in collaborative consumption." Journal of the Association for Information Science and Technology 67.9 (2016): 2047-2059.

[12]   Harmon-Jones, Eddie Ed, and Judson Ed Mills. "Cognitive dissonance: Progress on a pivotal theory in social psychology." Scientific Conferences Program, 1997, U Texas, Arlington, TX, US; This volume is based on papers presented at a 2-day conference at the University of Texas at Arlington, winter 1997.. American Psychological Association, 1999.

[13]   Lee, S., Hyun Kim, J., & Rosen, D. (2009). A semantic network and categorical content analysis of Internet and online media research. The Open Communication Journal, 3(1).

[14]   Molz, Jennie Germann. "Social networking technologies and the moral economy of alternative tourism: The case of couchsurfing. org." Annals of tourism research 43 (2013): 210-230.

[15]   Saussier, Julie. "Sharing as a disruptive force" Global Investor Nov. 2015: 18. Print.

[16]   Xinyuea, Zhu, and Guo Rongb. "SHARING ECONOMY." (2016)

[17]   "PEER NETWORKS ARE CONNECTING THE DOTS" Global Investor Nov. 2015: 2. Print

[18]   "THE SHARING MARKETPLACE: BRIDGING INTERESTS, COUNTRIES AND AGES" Global Investor Nov. 2015: 32. Print.

AUTHORS

JinTao Huang
Master of interaction science, SungKyunKwan University

KangNyeon Yi
Ph.D of interaction science, SungKyunKwan University

JiHye Han
Master of interaction science, SungKyunKwan University

INTENTIONAL BLANK

SOCIAL LIVE-STREAMING: TWITCH.TV AND USES AND GRATIFICATION THEORY SOCIAL NETWORK ANALYSIS

James Dux

Department of Interaction Science, Sungkyunkwan University,
Seoul, South Korea

ABSTRACT

The recent explosive growth of live-streaming, particularly video game streaming on Twitch.tv, has led to the development of young communities in a diverse and frequently evolving virtual social space. This study is a Social Network Analysis that aims to uncover how communities are interlocked based on attributes such as Community membership and stream style/type. Additionally, the network geography is related to the Uses and Gratification Theory (UG) that Twitch has been proven to fulfill according to genre and stream type (Sjoblom, Torhonen, Hamari, & Macey, 2017). By collecting data in a user-based fashion and using tools like Gephi, this project presents the Twitch community in a way that reveals users' interest seeking behavior trends. Implications from this study are a resource for: understanding Twitch and live-streaming communities at large, streamers looking for an insight into their audience's interests, and developers seeking the best way to provide and display beneficial content.

KEYWORDS

Social Network Analysis, Uses and Gratification, Twitch.tv, Social Live Streaming, Social Communities

1. INTRODUCTION

Today's internet speeds allow for a new type of media to flourish: live streaming video. Twitch.tv (colloquially known as just 'Twitch') is a prime example of this medium that creates a virtual 'Third Place' where users watch live-streamed videos while interacting with each other simultaneously (Brundl, Matt, & Hess, 2017). 'Streamers' are the performers who broadcast the video while engaging with their live audience 'followers' to varying degrees. Twitch focuses on the streaming of video games such as 'PLAYERUNKNOWN'S BATTLEGROUNDS' (PUBG) and 'League of Legends' (LoL). Why is live-streaming an important subject? Twitch.tv was only founded in 2011, but was bought by Amazon in 2014 for nearly 1 billion US dollars. Twitch.tv boasts 100+ million unique monthly visitors, and about 10 million daily active users (Smith, 2017). Clearly this new type of virtual 'living room experience' has struck a chord with audiences, and deserves further and deeper research.

After thorough literature review of related topics and papers, it is clear that Twitch presents a unique opportunity to learn more about this new digital form of interaction. Previous research has

proven that live-stream users enjoy and benefit from the social interaction provided by Twitch and similar sites. In particular, the "Co-Experience" provided by this medium has been shown to have a strong influence on the Active and Passive Enjoyment of live-streaming, marking the social aspect as a significant use motivation (Brundl et al, 2017).

Twitch is a social live-streaming site heavily driven by its community (and sub-communities) of streamers and the users who watch, follow, and subscribe to them. The availability of this connection information and the nature of this kind of organization lends itself very well to social network analysis. Social network analysis is "a visual and a mathematical analysis of human relationships", which gives researchers the opportunity to dissect how users in a network are connected by peers (Krebs, n.d.). Not only do social networks reveal connections, but can also be used to discover groupings based on node (individuals') attributes like common interests, communities, and other values. Analysis of these social networks provides benefits to the overall organization and members themselves. For Twitch streamers this might mean insight in where to "strengthen relationships within and across groups"; for Twitch.tv it could mean the ability to provide better 'related content' or to identify key nodes worthy of increased sponsorship (Serrat, 2009). It's clear that social network analysis of Twitch is a worthwhile endeavor that can provide real value for the community and site.

2. LITERATURE REVIEW

2.1. Twitch

Live streaming platforms such as Twitch allow streamers to project live video from their home to viewers over the internet. This video can be sourced directly from their PC for the intent of displaying video game game-play, program use, or any other entertaining/instructional visuals. Twitch in particular is primarily a video game streaming site. Streaming video directly from a webcam is also available. It's very common for streamers to overlay their webcam video inside of their game-play video, so audiences can see their face as the play and talk (see Figure 1). In fact, this method has become expected; viewers enjoy having the opportunity to share in the emotions of the streamer through their facial expressions (Hamilton, Garretson, & Kerne, 2014).

The stream is only half of the Twitch/live streaming equation; the user chat is the other half. The user chat is a simple chat room in which viewers of the stream can communicate with each other (and the streamer) with text and emoticons. The chat is a deeply significant part of the Twitch experience, providing users a way to chat with each other about what's happening on the stream as events unfold. In a 2015 survey of Taiwanese live stream users, 58% of users said they used the chat, and chat room design influenced the intent to use the live stream service (Ho & Yang, 2015). The chat is an integral part of the Twitch experience, morphing what could be a passive video watching experience into a modern participatory, social occasion.

Twitch exists only because of its community. Streamers form the base; they are the ones who produce the content (live streams), displayed on their 'channel'. Viewers support them by tuning in and watching their streams. Revenue for the site comes from ad space and 'subscriptions'. Subscribing to a streamer allows a user to voluntarily pay (certain amount on a monthly basis) a streamer they enjoy. In return, the subscribing user can benefit from things like 'subscriber only mode', when only subscribers are allowed to use the chat, and having unique emoticons available to them (Hamilton et al., 2014). The streamer backbone of Twitch becomes apparent in a 2016

survey which showed that 45.8% of respondents chose 'Specific twitchers' as a reason to watch Twitch, followed by 'Entertainment based on owned game' at 29.1% (Gandolfi, 2016).
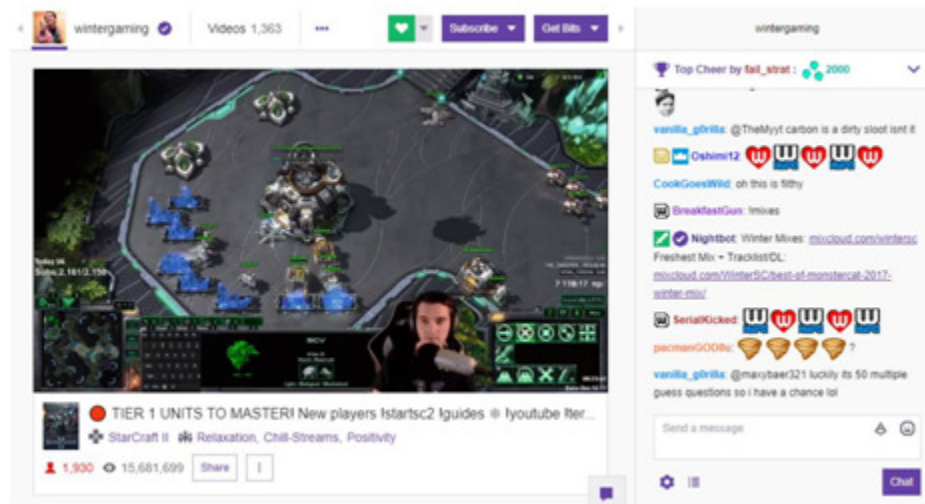

Figure 1. Twitch streamer 'wintergaming' plays Starcraft 2 with a webcam overlay of his face. Users participate in the chat on the right of the screen.

2.2. UGT and Twitch

Knowing what functions Twitch offers is the foundation to understanding why people use it. The interest of games appears to be the initial draw, as respondents to a 2014 survey most frequently cited their desire "to learn something about a particular game" as the reason they started watching streams (Hamilton et al., 2014). Thanks to this shared interest, Twitch forms what it known as a 'Third Place' for its users, an informal digital meeting place where people happily congregate to enjoy a common interest and socialize with each other (Hamilton et al., 2014). Holbert, Kwak, & Shah (2001) also state "chatrooms and multiplayer games increasingly provide the illusion of face-to-face social interaction and belonging", further reinforcing Twitch as 'Third Place'.

To delve deeper into the reasons for using Twitch, researchers have commonly turned toward the 'Uses and Gratification" (UG) theory. UG is a prominent theory when dealing with personal motivations behind using media ranging from passive television watching to participatory social media use, making UG an appropriate measurement tool for Twitch as well (Sjoblom, Torhonen, Hamari, & Macey, 2017). In UG theory, users are considered to be a conscious 'fulfillment seeker'. That is, users knowingly and self-reflectively seek out media that will satisfy their personal motivations according to certain guidelines as demonstrated in Table 1 (Sjoblom & Hamari, 2016).

Sjoblom and Hamari (2016) compared the UG need types with Twitch users' Hours Watched, Streamers Watched, Streamers Followed, and Subscription; their results showed a positive correlation for all need types excepting Personal Integration with 'Hours Watched' and 'Streamers Watched'. Personal Integrative was only positively correlated to 'Streamers Followed', and the only need that was negatively correlated to 'Hours Watched' and 'Streamers Watched', demonstrating a unique characteristic (Sjoblom & Hamari, 2016).

Table 1. The five aspects of UG that were tested by Sjoblom & Hamari, 2016

Table 1
UG need types (West & Turner, 2010, p. 398).

Need type	Description	Media examples
Cognitive	Acquiring information, knowledge, comprehension	Television (news), video ("How to Install Ceramic Tile"), movies (documentaries or films based on history e.g., The Other Boleyn Girl)
Affective	Emotional, pleasant, or aesthetic experience	Movies, television (sitcoms, soap operas)
Personal integrative	Enhancing credibility, confidence, and status	Video ("Speaking With Conviction")
Social integrative	Enhancing connections with family, friends, and so forth	Internet (e-mail, chat rooms, Listservs, IM)
Tension release	Escape and diversion	Television, movies, video, radio, Internet

Overall, Twitch displays the ability to satisfy UG motivations with "social interaction, learning and entertainment" as "important aspects of spectatorship" (Sjoblom & Hamari, 2016). Twitch UG satisfaction was further investigated by Sjoblom et al. (2017) where the unique characteristics of game genre and stream type were related to UG needs, revealing that the unique characteristics of game genres and stream types watched were correlated with different motivations. Even more significant was the fact that stream type had more impact on affecting user gratifications than did game genre (Sjoblom et al., 2017). This reinforces the notion of Twitch being a 'Third Place', allowing users to informally seek use gratification by dividing themselves into groups/streams based on interest in not just game genre, but also in the stream type or 'atmosphere' that has been developed by the streamer (Hamilton et al., 2014).

It is clear that all UG needs contribute in some way to the way users watch streams, and that these UG needs also affect users' choice of stream (Sjoblom & Hamari, 2016; Sjoblom et al., 2017). Beyond how stream type and content (such as genre) affect the UG needs for Twitch, there is the effect of stream size and possibly other factors playing into users' stream choices.

In addition to the above-mentioned stream attributes, the size of the stream also plays a role in the interactivity and social usability of Twitch. Hamilton et al. (2014) make specific note of the effect of stream size: "as streams scale up, information overload renders chat unreadable…… participants become frustrated with the difficulty of interacting in these streams", possibly leading to users seeking smaller, less populated streams. This is directly backed by Sjoblom et al.'s (2017) finding that Social Integrative and Personal Integrative UG needs are a part of Twitch users' motivation. The interaction between streamers and their audience is affected by the size of the stream.Streamers generally reported that they could interact effectively with 100-150 viewers by reading comments on stream, taking game play suggestions, and so on; beyond that number interaction was often unmanageable and the social aspects, even between users, was diminished (Hamilton et al. 2014).

A curious finding about the relation between stream type and game genre emerges when looking at the size of communities as a whole. Churchill and Xu (2016) found in a social network analysis of streamers that the Casual Streaming culture is the largest, followed by the Speed-Running culture, and finally the Competitive culture. The claim is that this ranking comes from the technical difficulty of games, Casual being the easiest to understand and Competitive being the most difficult and technically intensive (Churchill & Xu, 2016). However, Sjoblom et al. (2017) show in their UG study that when factoring game genre and stream/culture type together, Competitive stream types offer more, and more significant UG need satisfaction than Speed-running stream types, indicating that perhaps the Competitive community should be larger

(Sjoblom et al., 2017). These two types provide similar levels of motivation when only considering stream type. Taking this anomaly into account, extra consideration should be given to what streams users choose and the multitude of factors at play.

During their social network analysis of Twitch, Churchill and Xu (2016) visualized communities by specific games and community/culture types (Competitive, Casual, and Speedrun). Strong connections between game franchises were found, along with fairly well integrated cultures. During this study, however, only large, popular streamers who followed other streamers were considered. The inclusion of smaller streams of varying types might have an impact in subsequent studies. This is because, as mentioned by Hamilton et al. (2014), the size of streams/streamers impact the quality of socialization and therefore possibly users' habits.

Twitch itself provides a platform for every game (e.g. League of Legends, Warframe), every genre (e.g. Action, Real-time Strategy), and every stream/culture type (e.g. Speed Running, Casual). Users then filter this multitude of streams based on their personal interest, style, and motivations such as entertainment based on owned games, evaluation of non-owned games, or the ability to comment on the game (Gandolfi, 2016).

3. METHOD

This study aims to construct an exploratory social network analysis of Twitch, expanding on previous efforts with a focus on how communities are possibly affected by UG afforded by different types/cultures of streams, as well as game genre and stream size. Recently introduced by Twitch is browsing by 'Communities' that includes sections like 'VarietyStreaming', 'Positivity', and 'eSports', allowing another level of social network comparison. There is also the availability of coding streams based on their culture/type. This represents a move away from the traditional classification of streams by game or genre, and presents the opportunity for a fresh look at community structure.

3.1. Research Questions

RQ1) How does Twitch's social network look considering stream size, but before partitioning based on stream attributes?

RQ2) How does Twitch's social network look when based on streams' Community Membership Type?

RQ3) How does Twitch's social network look when based on stream Type (eg Competitive, Casual, Speedrunning,Social, or Learn)?

RQ4) How does Twitch's social network look when based on game genre?

RQ5) How does Twitch's social network look when based on a modularity algorithm?

RQ6) What can detailed views of single nodes reveal about the network?

RQ7)Which streams are key in terms of degree centrality, and how does this compare to their follower size and other traits?

3.2. Collection/Sampling

A total of 175 streams were selected. First, 75 streams were selected from Twitch's Communities. A specific community was selected and automatically sorted by live-viewers. A stream must have had more than one live-viewer and more than 100 followers to have been selected. Streams were selected systematically by amount of live viewers; approximately 10% of the Community's streams were selected each time. Additionally, 100 streams were selected from 'TwitchSwitch', a website that generates a link to a random Twitch stream, with at least one live-viewer.

Streams were required to have at least 100 followers to be more likely to guarantee links between nodes. Streams of all sizes were desired for diversity. Smaller streams are valued for their representation of increased streamer-follower interaction, and larger streams for generating edges as popular media (Hamilton et al. 2014).

3.3. Nodes

Each node represents a stream. Each node has certain attributes to help group them. Existing attributes are:

-Number of Followers (stream size)

-Game Played

 Other attributes had to be assigned by the researcher. These are:

-Type of Stream (casual – streamers play a game, competitive – streamers play competitively against other players [in leagues or for ranks], speed running – players race to complete a game as fast as possible, social – streamers are doing some other activity, learn – streamers note on their stream that they try to teach a game)

-Genre (the genre of the game being played, ie First Person Shooter (FPS), Multiplayer Online Battle Area (MOBA), Massively Multiplayer Online (MMO), In-Real-Life (irl), Sports, Horror, Strategy, Card, Sandbox, Adventure, Fighting, Role Playing Game (RPG), etc)

-Community Membership Type (General – belongs only to social communities, Competitive – belongs only   to competitive/game communities, Mixed – belongs to at least one social and at least one competitive/game community, None – belongs to no communities. Streamers choose which communities they want to join. Streams can belong to up to three communities at a time. Due to the infinite amount of communities available, streamers were coded into one of the four classifications.)

3.4. Edges

Edges are determined by actual users. That is, followers that streams share in common. The weight of the edge is determined by the amount of followers in common.

3.5. Programs/Resources

Twitch.tv, TwitchTools.com, TwitchSwitch.tv, Gephi, Python, Microsoft Excel

3.6. Process

Once node information from Twitch was recorded into an Excel sheet, the stream's ID was given to TwitchTools.com to retrieve a list of its followers in CSV format. A Python script was used to compare stream follower files. The followers were sorted by 'Notification' and then 'Follow Date, only the first 5000 followers were compared. 'Notification' means the user receives a notification when that stream comes online, 'Follow Date' is the date when the user started to follow the streamer. These were chosen to best represent habitual stream watchers. The amount of followers in common between streamers was used to determine the weight of edges between streams. This node and edge information was fed into and analyzed with Gephi.

When analyzing, edges with a weight of less than two were filtered out. This is to slightly increase the meaningfulness of the graph, as there were quite a few edges with a weight of only one. This caused eight of the nodes to be excluded from the final graph, leaving 167 nodes. The layout algorithm used was Force Atlas 2, and edges are undirected. Edge weight plays a role in determining the shape of the network.

4. RESULTS

RQ1) How does Twitch's social network look considering stream size, but before partitioning based on stream attributes?
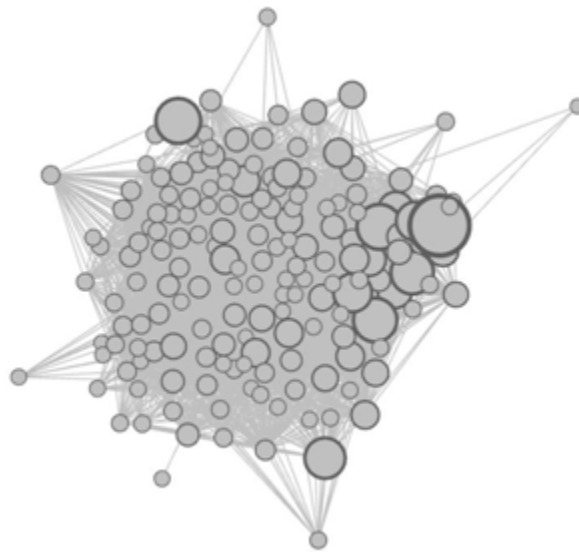
Figure 2. Nodes are presented simply, without any groupings. The size of the nodes is relative to the amount of followers they have.

Overall, the social graph is quite compact with few outliers. Nearly every node has a connection to more than one other node. Analysis provides an Average Degree of 38.299, an Average Path Length of 1.87, Diameter of 4, and Density of .221. This data and Figure 2 show a quite tightly knit community despite the number of nodes. Notable is that some of the largest nodes are actually near the outside of the graph, when generally the largest nodes would be expected to be more central. In fact, many of the central nodes small and medium sized. This shows the significant factor of interactivity that smaller streams can deliver above larger streams (Hamilton

et al. 2014). Due to the data collection method, it can be hypothesized that many people visit and follow these larger streams, but they do not visit them as habitually (no 'Notification') as smaller streams.

RQ2) How does Twitch's social network look when based on streams' Community Membership Type?

Introduced as a beta feature in February 2017, Twitch's 'Communities' are "self-organizing, public groups centered around specific activities"; these communities act as a way for users to more easily discover streams that are more pertinent to their interests (Karsevar, 2017). Figure 3 reveals that 'General' community streams are strewn throughout the graph, likewise with 'None'. General (more social focused) streams form the 'core' of the network. This comes as no surprise seeing as relaxed and socially open streams such as these represent several of the Uses and Gratification traits like Social and Personal Integration. 'Competitive' community streams are a bit scattered, but show a couple of clear groupings, indicating users' interest in certain gaming mechanics, and then branching off to social streams as well. Interestingly, 'Mixed' community streams are a bit scattered as well, but are often found clumped next to the two Competitive groupings. It's possible these mixed streams form a type of bridge as followers express their intents of satisfying different UG needs while moving between more social and competitive communities.
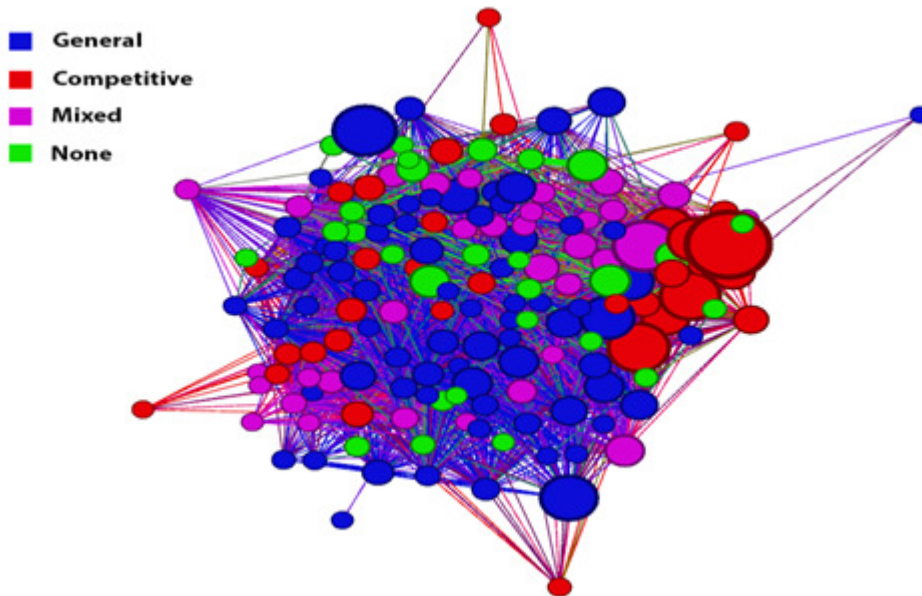

Figure 3. Nodes (streams) are partitioned by Community Membership Type, and sized depending on the number of overall followers.

RQ3) How does Twitch's social network look when based on stream Type (eg Competitive, Casual, Speedrunning, Social, or Learn)?

This classification for streams is based on previous research by Sjoblom et al. (2017). Streams were coded by the researcher based on how the streamer described their stream, or lacking a self-description, a stream type was assigned depending on the style of game-play or culture. For instance, streams which did not display a particular game were assigned 'Social', and streams that mentioned competitive play were assigned 'Competitive'.

Immediately obvious in Figure 4 is that groupings are much clearer by stream Type than they are by Community Membership (Figure 3). This may suggest a couple of things. First, Community Membership is more complicated than the coding allowed for. Second, Community Membership (which is user-decided) is a more esoteric [social] marker than an appropriate stream identifier. More research is needed.
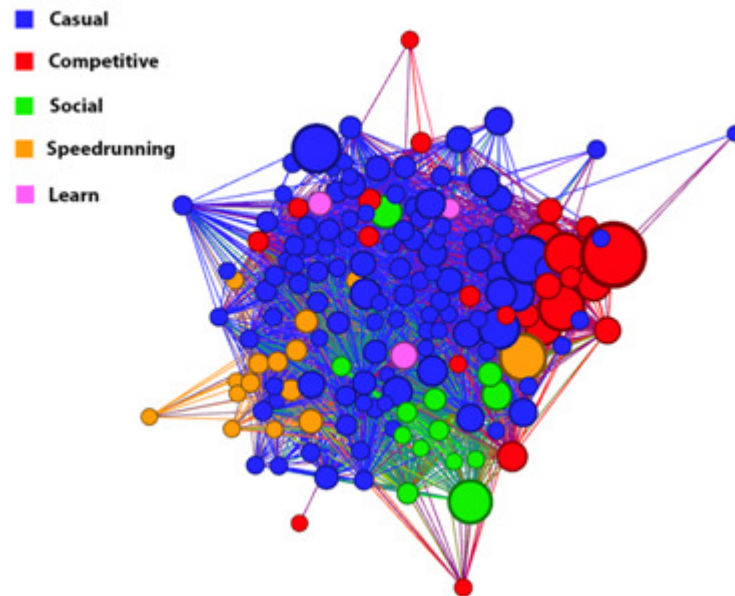
Figure 4. Nodes (streams) are partitioned by stream Type and sized depending on the number of overall followers.

As compared to Figure 3 (General community being moderately related to 'Casual' type streams), Figure 4 again shows that Casual streams to make up the bulk and core of the network. Casual type streams had the highest degrees of centrality, acting as a bond between more focused cultures. 'Speedrunning', 'Competitive', and 'Social Streams' were quite tight-knit. This suggests that users congregate to a specific culture/interest, and then branch out to satisfy other needs like Tension use gratifications (or vice versa). No community really has a monopoly on large or small streams, signifying that users still can attain Social or Personal Integrative gratification across cultures and stream styles. Speedrunning does however have the largest concentration of small streams, possibly adding credence to Churchill and Xu's (2016) study that stated the small attraction of highly technical streams compared to less technical general streams.

RQ4) How does Twitch's social network look when based on game genre?

Figure 5 shows the social network based on game genres, listing the top ten. Heavily prevalent is the First Person Shooter (FPS) genre, likely due in part to the current popularity of PUBG. FPS games like PUBG and Counter-Strike are scattered across the Competitive and Casual regions

(Figure 4) likely because of their ability to satisfy multiple UG stream needs; they can be highly technical (Cognitive UG), provide in-game socialization (Social/Affective UG), and can be cinematic (Tension UG). Adventure games make up almost the entire Speedrunning type streams due to their single-player nature, but are still found elsewhere throughout the network. Multiplayer Online Battle Arena (MOBA) games lend themselves well to the Competitive streams thanks to their top-down views of small scale battles, but are still found throughout the network. In-Real-Life (IRL) is relegated to Social streams, as Social streams are defined as doing something other than a game. The IRL designation is given unless the streamer lists a specific activity (like music or art), so general IRL activities (chatting) greatly trump specific non-game activities at this time.
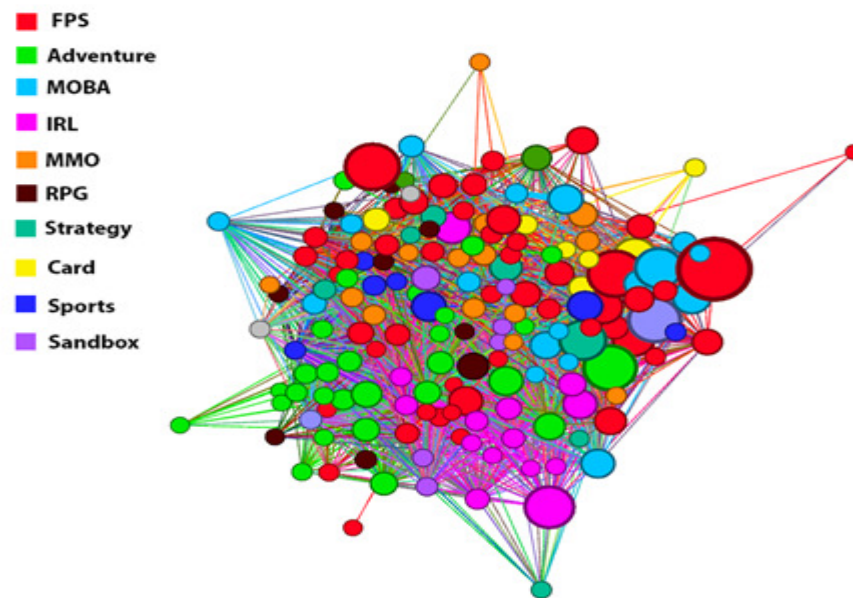
Figure 5. Nodes (streams) are partitioned by game genre and sized depending on the number of overall followers.

In general, game genres don't have too many boundaries when it comes to Community Membership or even stream Type. Streamers are simply using the games as a tool to interact with their community, a common bonding type. Sometimes a specific tool (Adventure or MOBA) helps reach a certain type of follower, but the social heart of Twitch allows genres to span the entire network.

RQ5) How does Twitch's social network look when based on a modularity algorithm?

Running the social network through a modularity algorithm produces results close to consistent with assigned variables like stream type. In this algorithm, the network was divided into 6 sections with no knowledge about the nodes' attributes stream communities, types, genres, etc. Figure 6 is the result of automatically generated 'communities' or sections of the network. Without assigning meaning to the nodes, parts of the network still match together. These 'organic' communities provide the insight of the mixing and dividing of users' interests. Again, competitive and speedrunning cultures stand out clearly. As communities from Figure 3 were mixed at the top, they are here as well. At the bottom, there is the collusion of General/Mixed communities and casual/social type streams.
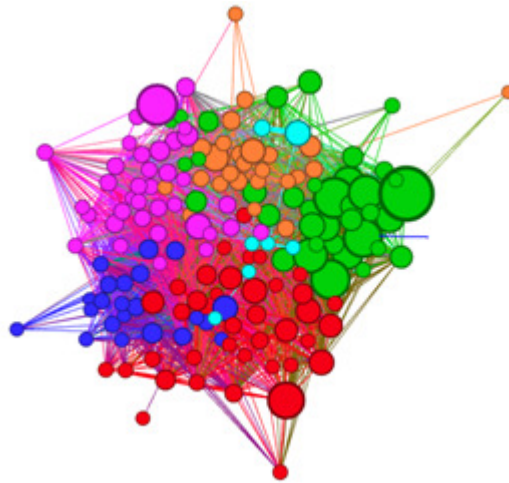
Figure 6. The network when run through a modularity algorithm to produce 6 sections.

RQ6) What can detailed views of single nodes reveal about the network?

By focusing on specific nodes, 'case studies' about the network and its nodes are possible. The stream with the most followers, 'dreamhackcs', is a 'Competitive' type stream, and belongs to the 'Competitive'community. It's connected to nearly all of the other Competitive streams as kind of a key figure for that stream type (Figure 6). Despite its size and importance to Competitive streams, it's not nearly as central as some of the other Casual type streams, or even compared to the other Competitive streams. This is likely due to the fact that 'dreamhackcs' is a special use stream that live-broadcasts large competitive events (Dreamhack Tournament) and reruns, but is not a typical stream run by a single streamer.
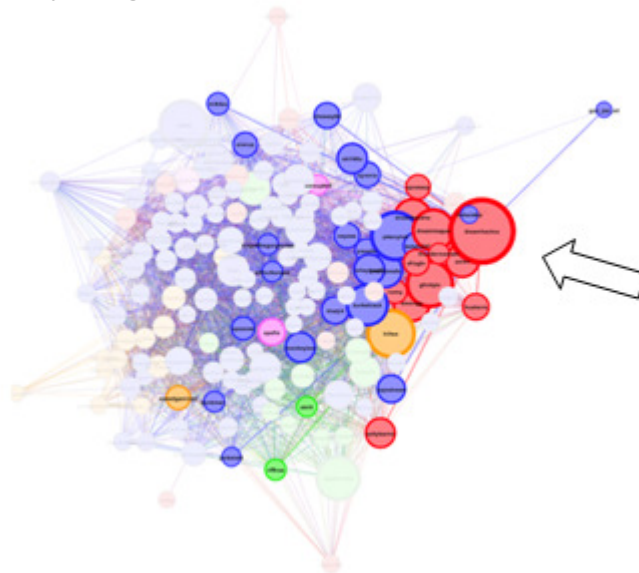
Figure 7. A focused view of 'dreamhackcs'. See Figure 4 for network legend.
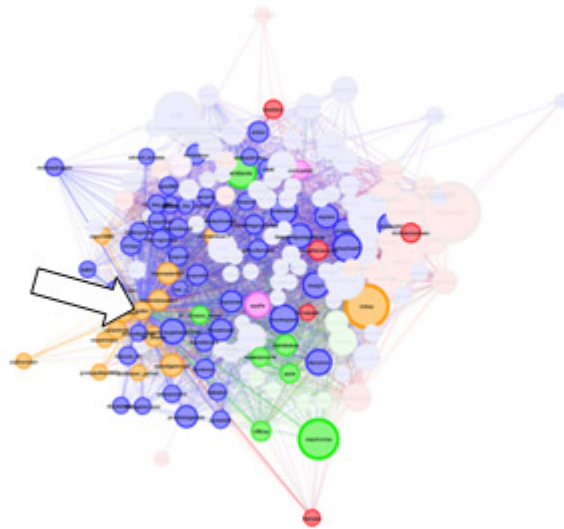
Figure 8. A focused view of 'reabs'. See Figure 4 for network legend.

'reabs' is a Speedrun type node (Figure 7). Despite being close to the outskirts of the network, it is still connected to almost every other Speedrun node, showing a tight-knit community. 'reabs' is also connected to the majority of Casual nodes, despite being a purely Competitive Community Member. Like Figure 3, this might suggest that Community Membership might not have too big of an impact in what users stay within. Instead, users may allow their UG needs to lead the way and search out a stream like 'reabs' for certain gratifications, and the Casual streams for other gratifications.

RQ7) Which streams are key in terms of degree centrality, and how does this compare to their follower size and other traits?

Initially, the most noticeable trait of the centrality heat map is that the largest nodes/streams do not represent the most connective nodes (Figure 9). When compared to Figure 4, the most central nodes are Casual. As Twitch satisfies UG needs, these nodes seems to provide a great even ground for viewers, a kind of comfortable starting point that attends to many needs like Cognitive gameplay and Affective streamer interaction. Also notable is the fact that highly central nodes are often medium sized. Too small, and a stream doesn't provide enough interaction. Too large, and a stream overwhelms the user with interaction.

Competitive nodes are often also quite central. The base reason for Twitch is to act as a platform for streaming game content, and it's clear users still move between social/Casual streams and more serious Competitive streams. Speedrunning is quite mixed in terms of centrality. Speedrunning can be quite a specific interest, and this is represented in the centrality map. It's likely the furthest streams have a specific focus, drawing a dedicated crowd. And yet, as in Figure 7, this crowd still mingles with very central/Casual streams.
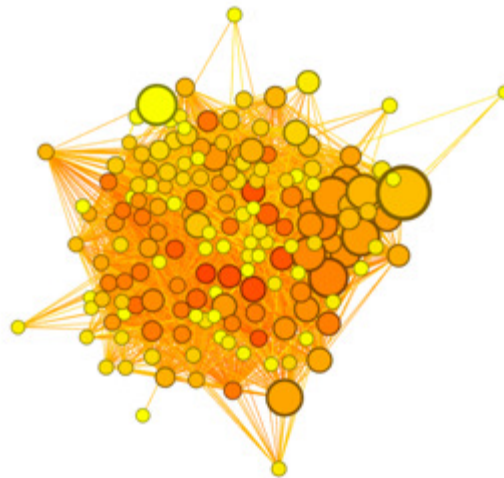
Figure 9. A network 'heat map' of centrality. Darker means a higher degree of centrality (more connections).

5. CONCLUSION / DISCUSSION

Compared to traditional media, social live-streaming sites like Twitch are relatively new. Twitch itself is often in a state of change as trends and streamers come and go, updates are applied, technologies become more accessible, and the internet culturefluctuates. Even Twitch's Communities section had been added within one year of this study. This study was performed to impart a beginning understanding of the nature of Twitch's social network onto the audience.

This study found that Community Membership didn't segregate nodes into sections as well as stream types. This suggests that Communities may be kind of vague, being that they are streamer-defined (ie streamers can create/join any community), or, the method used to define them wasn't selective enough. There are an infinite amount of communities, which may make it difficult for streams to define themselves, or for users to find an appropriate stream that includes all/most of the streams that would satisfy their UG needs. Well defined communities, however, could be extremely beneficial.

Therefore, this researcher would like to suggest an improved way of classifyingstreams. Communities (especially large ones) as they are, often don't correlate quite well with stream type or the game played. A way of breaking down and defining streams may help users find the streams they desire. Since Communities are already streamer-defined, it's not impossible that Twitch could provide a streamer or user-driven filter system. Streamers could add certain filters to their stream regarding the usual peak viewer count, usual game and game genre, type of stream, and so on. Likewise, users could also be given the power to help define streams/filters.

The method of connecting streams, by followers-in-common and by Notification, was successful. It helped avoid large streams dominating the network by size alone. Streams of all sizes were allowed shine in the network based on habitual users as they do on Twitch itself. Smaller and mid-sized streams often had high centrality (especially Casual types) due to their increased interactivity (smaller streams better suited for socializing). Larger streams like 'Dreamhackcs' (Competitive type) helped form the hubs of their respective stream types.Frequent connections between all social circles were found, even from very focused cultures such as Speedrunning.

Twitch is indeed an open community, offering a diverse range of topic and interests, yet linked together by the social experience provided. The existence of specific communities lets users swim around within the overall network to find a stream to satisfy whatever their needs might be.

Due to the ever-evolving nature of Twitch and social sites, this research has some limitations. For instance, streams can have membership in multiple communities. For this study, this node attribute had to be codified based on their Community Memberships, not on their exact communities due to the infinite possibilities. Future research may want to consider a way in which to include this variable. Additionally, stream type and game played can change. Again, these variables were assigned based on the stream at the time of data collections. Classification is unending for continually changing social sites and evolving media. It is also very difficult to dictate the perfect genre for a game, as a game may have both RPG and Adventure elements, so the researcher must code the genre as what appears to be the primary genre. Following studies may benefit from incorporating a system where the genre of games in a stream can be better defined and compared. As this was exploratory research, future research on a specific network trend, or expansion of a specific community, genre, or stream type might be interesting.

REFERENCES

[1] Brundl, S., Matt, C.,& Hess, T. (2017, June). Consumer Use of Social Live Streaming Services: The Influence of Co-Experience and Effectance on Enjoyment. In Proceedings of the 25th European Conference on Information Systems (ECIS),   1775-1791. ISBN 978-989-20-7655-3. Retrieved from http://aisel.aisnet.org/ecis2017_rp/114

[2] Smith, C. (2017, July). By the Numbers: 46 Amazing Twitch Stats and Facts. DMR. Retrieved from https://expandedramblings.com/index.php/twitch-stats/

[3] Churchill, B. C.B.,& Xu, Wen. (2016). The Modern Nation: A First Study on Twitch.TV Social Structure and Player/Game Relationships. 2016 IEEE International Conferences on Big Data and Cloud Computing, Social Computing and Networking, Sustainable Computing and Communications, 223-228. DOI 10.1109/BDCloud-SocialCom-SustainCom.2016.43

[4] Sjoblom, M., Torhonen, M., Hamari, J., & Macey, J. (2017, March). Content Structure is King: An Empiriical Study on Gratifications, Game Genres and Content Type on Twitch. Computers in Human Behavior 73 (2017), 161-171. DOI: 10.1016/j.chb.2017.03.036

[5] Ho, C., & Yang, C. (2015). A Study on Behavior Intention to Use Live Streaming Video Platform Based on TAM Model.

[6] Hamilton, W., Garretson, O., & Kerne, A. (2014). Streaming on Twitch: Fostering Participatory Communities of Play within Live Mixed Media. DOI: 10.1145/2556288.2557048

[7] Sjoblom, M., Hamari, J. (2016, August). Why Do People Watch Others Play Video Games? An Empirical Study on the Motivations of Twitch Users. Computers in Human Behavior, 2016. DOI: 10.1016/j.chb.2016.10.019

[8]   Gandolfi, E. (2016). To watch or to play, it is in the game: The game culture on Twitch.tv among performers, plays and audiences. Journal of Gaming &Virtual Worlds, 8: 1, pp. 63–82, doi: 10.1386/jgvw.8.1.63_1

[9]   Holbert, R.L., Kwak, N., Shah, D.V. (2001). "Connecting" and "Disconnecting" With Civic Life: Patterns of Internet Use and the Production of Social Capital. Political Communication, 18:2, 141-162, DOI: 10.1080/105846001750322952

[10]   Krebs, V. (n.d.). Social Network Analysis: An Introduction. Orgnet. Retrieved from http://www.orgnet.com/sna.html

[11]   Serrat, O. (2009, February). Social Network Analysis. Knowledge Solutions.

[12]   Karsevar, A. (2017, February). Communities Are Here: Discover Broadcasts Faster Than Ever. Twitch. Retrieved from https://blog.twitch.tv/communities-93de2c1a2e4d

INTENTIONAL BLANK

EXAMINING RISK COMMUNICATION OF SAMSUNG BY ANALYSING CUSTOMER REACTION ON SOCIAL MEDIA: FROM GALAXY NOTE 7 EXPLOSION TO GALAXY S8 RELEASE

Jaewon Lim and Jang Hyun Kim

Department of Interaction Science, Sungkyunkwan University, Seoul, Korea

ABSTRACT

Crisis of product defect causes recalls, which not only causes financial loss but also devastates brand reputation. This study explores customers' response to the twice recall of the Galaxy Note 7 in Korea. The authors examined the semantic networks of consumers comments from the Samsung official social media 'Samsung Newsroom (blog and Facebook)' in five time periods(i.e., release, first recall, recommendation to discontinue use, second recall, and compensation program). Customers in each period showed different emotional states which were positive in the first recall but negative in the second recall. Therefore, timely recall and factual communication is needed to the product-harm crisis management, but root cause investigation should be taken to avoid further recalls.

KEYWORDS

Galaxy Note7, Product harm crisis, Recall, Semantic network analysis

1. INTRODUCTION

The crisis of the Galaxy note 7 recall occurred within a month of its release. Galaxy note 7 was unveiled on August 19, 2016. However, several Galaxy note 7 explosion cases were reported and the epidemic was spreading. Samsung diagnosed the cause of the explosion is a failure of the batteries and recalled the Note 7. Although they provided replacement phones, those began to explode as well. In the end, Samsung issued the second recall and recalled the entire line of Note 7 phones.

[1]Recalling a product can make the involved firms' brand image worse.[2]Therefore, they should try to reduce negative impacts, one of which is to convey their official position to the public through media.

[3]Since the mid-2000s, Social media platforms increasingly tend to act as both a place for the firms and customers. Focal firms' risk communication message can be delivered instantly and broadly. With social media, customers can voice their opinion. Samsung posted a statement in

their official social media channel, 'Samsung Newsroom(blog and Facebook)' during a series of explosion event. At the same time, the customers commented on each statement to voice their opinion.

[4]As far as the authors investigated, Samsung is perhaps the only company to recall twice and has the largest recall in terms of volume. However, as customers become more active and social network service develops, product recalls may happen more frequently in the future. By assessing Samsung's crisis management to the Galaxy note 7, this study aims to understand the successful product crisis management. Using the semantic network analysis, the authors analyzed the customers' comments written under the Samsung official announcements in their social channel. Therefore, we proposed the following research questions :

RQ1: How customers react to Samsung's risk communication on social media(blog and Facebook)?

RQ1-1. What are semantic traits of the blog and Facebook comments?

RQ1-2. What are common and unique words for each period?

RQ1-3. How did emotional words change over time?

2. METHOD

Figure 1 shows the overall research process.
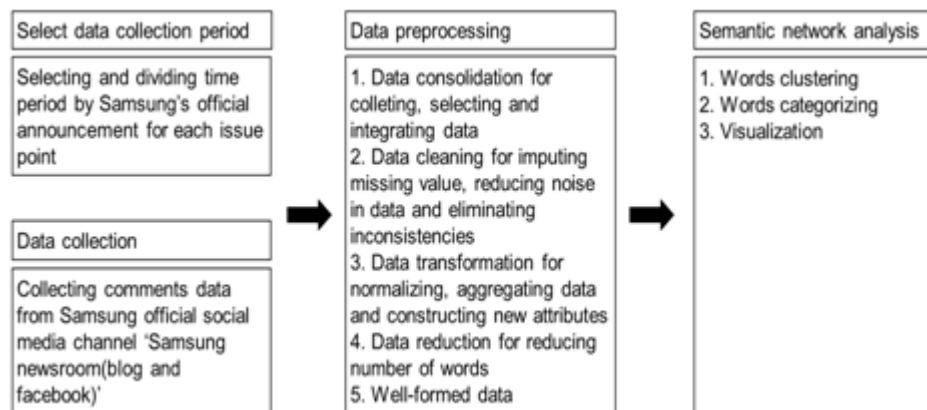
Figure 1. Overall research process

2.1 Select data collection period

The authors estimated the change of the word by period and analyze the change of customers' awareness about the issue. The data collection period was divided into six periods. The first data collection period is from August 11, 2016 to August 17. It covers six days from the product launch. The second data collection period is September 2, 2016, when Samsung voluntarily announced the recall. The third data collection period is from September 10, 2016 to September 20, 2016. During this period, Samsung advised their users to discontinue the use of the Galaxy Note7 and announced instructions for commencing the exchange of products. The fourth data

collection period is October 11, 2016, when Samsung officially announced the discontinuation of Galaxy Note7. The fifth data collection period is October 13, 2016 to October 27. It was time for Samsung to announce compensation program, called 'Galaxy upgrade program', including exchange and refund. Last period is April 19, 2017, when Galaxy s8 released. Galaxy s8 is the first and new Samsung galaxy series product after the Galaxy Note 7 explosion.

2.2 Data Collection

Customer comments were gathered from the Samsung's official social media channel 'Samsung Newsroom (blog & Facebook)'. We scraped all the comments manually. A total of 4,016 comments were collected during the data collection period. All of the comments are in Korean as the authors examined only the newsroom for Koreans.

Only Galaxy s8 customer comments were gathered from six news articles posted in the Naver.

2.3 Data preprocessing

Text preprocessing was carried out through Textom, an online text analysis service which is a Korea-based big data processing solution site.

The authors uploaded the collected text data to Textom and preprocessed. In this process, the main words were identified through the frequency of extracted words. Basically, it provides listing of nouns, adjectives, and verb phrases that remove special letters and postposition through morphological analysis. Words with similar meanings are grouped together into represented word. In this way, only the meaningful words were selected from the word list formed by refining and the co-occurrence frequency matrix data was produced.

2.4 Semantic Network Analysis

[5]Semantic network analysis (SNA) was conducted, which examines the relationships among words such as co-occurrence, frequency, and distance. As SNA adequately shows the significant words regarding the customers opinion and emotions about Samsung's product crisis management, it is opted to employ for the present study.

[6]To visualize the relationship between words, we used computational text analysis tools NETDRAW from UCINET6.

Among a variety of centrality measures, betweenness centrality was selected for this study.

3. RESULTS

To visualize semantic traits of words, the authors categorized words into four groups. The words which are related to facts of issue were categorized into group 'Fact' with black nodes. Also, emotional words were considered for grouping. Positive words were specified 'Positive feeling' with blue nodes and negative words were grouped into 'Negative feeling' with red nodes. Finally, words about cost were categorized into 'Cost' with yellow nodes.

3.1 Six days from the product launch- August 11, 2016 to August 17, 2016

For the first period, 73 comments were collected. The words with the highest frequency next to '*Samsung*' are *article, opening, Galaxy Note7, good, consultation, head office, change* and others.

Figure 2. Semantic network of comments on Samsung Newsroom from August 11, 2016 to 17

In Figure2, words are related to 'Current Galaxy users (S7, Note5)' and 'General expectation for Note7'. For example, words explaining current Galaxy users are *product, service center, repair,* and *Galaxy S7*. Such words as *design, software, hardware, pre-order, lens* and *better* were mentioned for general expectation for Note7. In this period, most words are categorized into group 'Fact' and 'Positive feelings'.

3.2. Recall announcement - September 2, 2016

For the second period, 462 comments were collected. The words with the highest frequency next to 'Samsung' are *product, exchange, decision, battery, refund, pre-order, recall, new product, problem, use, customer, corporation,* and others.

In Figure3, words are related to two topics, which are 'Explosion issue and disappointed feeling' and 'Positive opinion for Samsung decision.' For example, words about the first topic are *explosion, tiresome, disappointment, problem, defect, bomb, return, pity,* and *error*. For the second topic, *good decision, trust, thanks,* and *cheer up* were mentioned.

At this time, lots of words are categorized into group 'Fact' and 'Positive feelings' like the first period semantic network. Also, words with 'Negative feelings' were salient from this point.
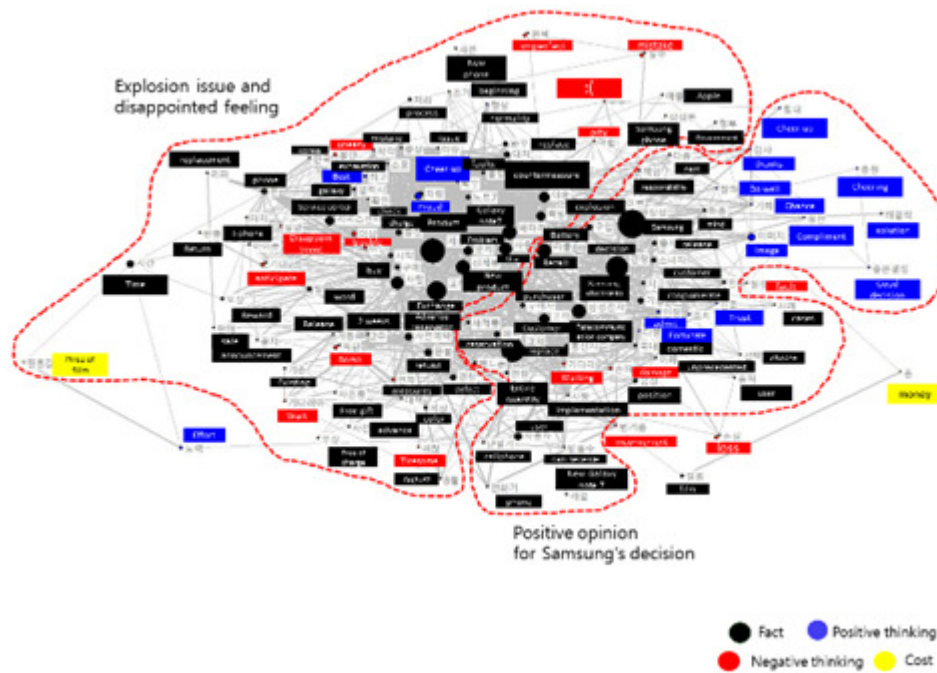
Figure 3. Semantic network of comments on Samsung Newsroom on September 2, 2016

3.3. From recommendation for stop using to exchanging replacement - September 10, 2016 to September 20, 2016

For the third period, 515 comments were collected. The most frequent words next to 'Samsung' are exchange, customer, battery, product, problem, wait, Galaxy Note7,authorized dealer, explosion, reservation, recall, purchase and other words.

In Figure4, words described 'Explosion issue and product problems' and 'Complaint for exchanging and disappointment.' For example, the words related to explosion issue and product problems are *explosion, fault, dead battery, defect,* and *safety*. For the complaint for exchanging and disappointment, the words such as *it's annoying, inconvenience, refusal* and *apology* were mentioned.

During this period, while blue nodes with 'Positive feeling' decreased sharply compare to previous two periods, red nodes with 'Negative feeling' increased.

3.4. Second recall - October 11, 2016

For the fourth period, 1,002 comments were collected. Word '*Exchange*' is the most frequently mentioned at this time. *Refund, Samsung, Galaxy Note7, customer, purchase, use, think, product, compensation, change, wait,* and *cancellationcharge* are the highest frequency words with 'Exchange'

Figure 4. Semantic network of comments on Samsung Newsroom from September 10, 2016 to September 20, 2016

Figure 5. Semantic network of comments on Samsung Newsroom on October 11, 2016
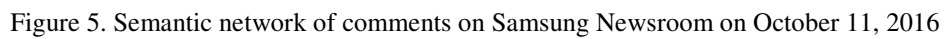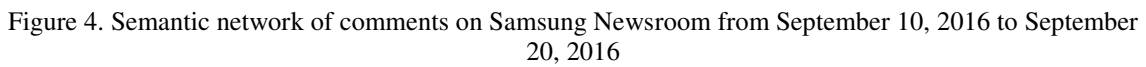
In Figure5, words are demonstrating two topics, which are 'Loss, compensation and disappointment' and 'Countermeasures for Galaxy Note7.' For example, the words regarding the first topic are *cancellation, charge, damage, halt production, absurd,* and *compensation*. For the second topic, *Galaxy Note8, Galaxy S8, iPhone, new product,* and *rental phone* were mentioned.

At this time, lots of words are categorized into group 'Cost' and 'Negative feeling' unlike other three periods.

3.5. Compensation program notice - October 13, 2016 to October 27, 2016

For the last period, 2,022 comments were collected. The most frequent words next to 'Samsung' are exchange, customer, refund, compensation, opinion, Galaxy Note7, use, change, product, program, purchase, return, Galaxy Note8, and others.

Figure 6. Semantic network of comments on Samsung Newsroom from October 13, 2016 to October 27, 2016

In Figure6, most words in categories 'Fact', 'Negative feeling', and 'Cost' are mentioned similarly. Black, red and yellow nodes are spotted sporadically. For example, words like money, installment, (smartphone) case, 30 dollars, and coupon are showing yellow nodes 'Cost'. Dissatisfaction, trash, fool, wait, fail, discomfort, phone replacement, and pushover present red nodes 'Negative feeling'. This network has only one blue node for 'Positive feeling' with word 'good'.

3.6. Galaxy s8 release

We can see the unique word 'Sakura Edition'. Despite having to be cautious after the battery explosion of the Galaxy Note 7, the phenomenon that the display screen is reddish immediately after the launch of the Galaxy S8 was controversial. It got a humiliating nickname of'Sakura Edition', and consumers blamed it as 'Trash Galaxy' and kicked the tongue. In addition, it is the

first new product after the explosion, so consumers' worries about 'Safety' have increased. The words 'Note 7', 'Battery' and 'Explosion' related to Note 7 were also mentioned.

Although there are few numbers, positive words also appeared. A few consumers were satisfied and admitted with the design of the Galaxy S8 and cheered on Samsung('Fighting'
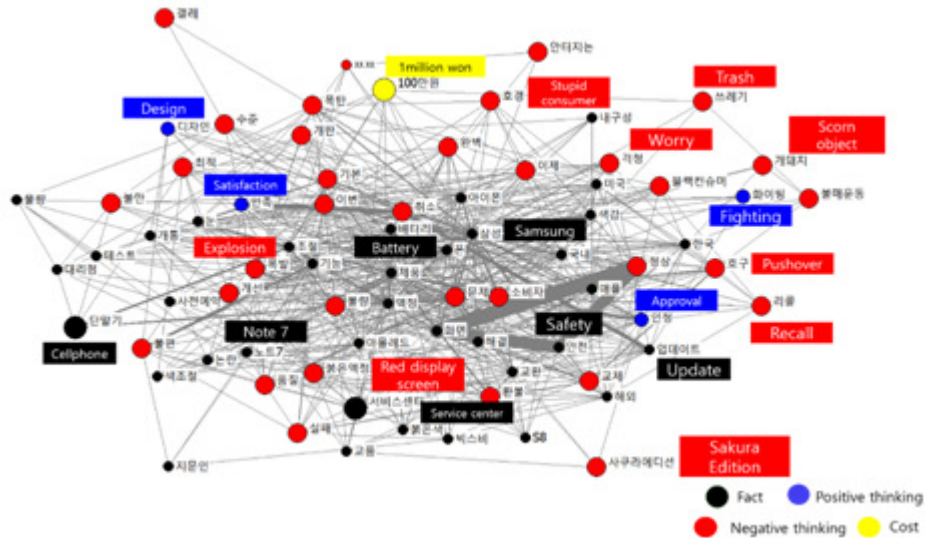
Figure 7.Semantic network of comments on Galaxy s8

4. DISCUSSION

As a result of examining the crisis response of Samsung in five periods as above, it can be seen that Samsung's product crisis management on the Galaxy note 7 was not so effective as desired. It was a good decision to admit the product defect and to make a swift recall. As we can see the words in the second period, people felt "trust" in Samsung's recall. However, due to the rapid exchange of products without properly identifying the cause, replacements also exploded. This led to second recalls and customers had to do "tiresome" exchanges even twice. Customers even need to seek a smartphone to replace the Galaxy Note 7 and to use 'iPhone' or Galaxy s7 edge which is older than Galaxy Note 7.

When Samsung announced the first recall, it would have been possible to reduce the damage if they had enough time to properly identify the cause. It is a smart move to admit product defects in detail and voluntarily recall products as early as possible if the cause of the defect is the problem of the product itself. Without uncovering the root cause, rashness can result in the two times of recalls like Galaxy note 7. Not only timely but accurate and factual communication is needed.

Second, results show that the importance of an appropriate compensation program. After Samsung announced that it has halted sales and production of the Galaxy Note 7 worldwide, Samsung is offering a new compensation program. However, it is not much different from the existing 'Galaxy program', which was packaged as if it were a special reward program for Galaxy note 7 customers. So when the authors analyzed the comments, the word 'trickery' appeared frequently. It should be noted that anger that is worse than a product defect came from the users

feeling that the company may have deceived its customers.

Third, the performance of the device in launching a new product after the crisis must be emphasized. For the new products released after the crisis, consumers expressed their negative opinions due to the previous explosion (the 'red display screen' phenomenon of the Galaxy S8 also contributed to the negative comments), but at the same time it showed a positive response to the 'performance' and 'design' of the s8.Therefore, when it comes to product launches in the future, it will be more important to improve the 'performance' and 'safety' of products.

5. CONCLUSION AND LIMITATION

[7]Samsung sits at 49th place in 2017 Reputation Quotient Ratings, which ranks one of American's 100 most visible companies. This is comparable to the 7th rank in 2016 and all three consecutive years when Samsung secured a spot in the top 10 from 2014 to 2016.

The limitation of this study is in that it explored comments from Korean Facebook and blog, although Galaxy note7 recall occurred worldwide.

REFERENCES

[1]   LAUFER, Daniel; COOMBS, W. Timothy. How should a company respond to a product harm crisis? The role of corporate reputation and consumer-based cues. Business Horizons, 2006, 49.5:379-38

[2]   J. HOLMES, Bev, et al. Communicating with the public during health crises: experts' experiences and opinions. Journal of Risk Research, 2009, 12.6: 793-807.

[3]   RUTSAERT, Pieter, et al. The use of social media in food risk and benefit communication. Trends in Food Science & Technology, 2013, 30.1: 84-91.

[4]   Aditya Madanapalle. (2016) Samsung Galaxy Note 7 recall: Here's a look at a few device recalls from the past. http://www.firstpost.com/tech/news-analysis/samsung-galaxy-note-7-recall-her es-a-look-at-a-few-device-recalls-from-the-past-3688793.html (accessed Oct.18, 2017)

[5]   LEE, Sungjoon; HYUN KIM, Jang; ROSEN, Devan. A semantic network and categorical content analysis of Internet and online media research. The Open Communication Journal, 2009, 3.1.

[6]   BORGATTI, Stephen P.; EVERETT, Martin G.; FREEMAN, Linton C. Ucinet for Windows: Software for social network analysis. 2002.

[7]   The Harris Poll. (2017) 2017 Reputation Quotient Ratings. http://www.theharrispoll.com/reputation-quotient (accessed Oct.18, 2017)

AUTHORS

Jaewon Lim is studying Interaction and Science in Sungkyunkwan University. Main research interests lie in social and semantic network analyses and media and communication.

INTENTIONAL BLANK

CONCEPTUALIZING AI RISK

Martin Ciupa[1] and Keith Abney[2]

[1]CTO calvIO Inc., Webster, New York, USA
[2]Senior Lecturer, CalPoly, California, USA

ABSTRACT

AI advances represent a great technological opportunity, but also possible perils. This paper undertakes an ethical and systematic evaluation of those risks in a pragmatic analytical form of questions, which we term 'Conceptual AI Risk analysis'. We then look at a topical case example in an actual industrial setting and apply that methodology in outline. The case involves Deep Learning Black-Boxes and their risk issues in an environment that requires compliance with legal rules and industry best practices. We examine a technological means to attempt to solve the Black-box problem for this case, referred to as "Really Useful Machine Learning" (RUMLSM). DARPA has identified such cases as being the "Third Wave of AI." Conclusions to its efficacy are drawn.

KEYWORDS

AI Risk, Deep Neural Network, Black-box Problem, Really Useful Machine Learning, RUMLSM, DARPA Third Wave of AI.

1. INTRODUCTION

A common worry about AI is that it poses an unacceptable risk to humanity (or individual humans) in some way. An extensive literature has begun to emerge about various aspects of Artificial Intelligence (AI) risk, much of it focused on existential risk from Artificial Generic Intelligence (AGI). But AI poses other risks, from how driverless cars solve the 'trolley problem', to whether autonomous military robots attack only legitimate targets, to trust in the safety of AI/Robotics in industrial and commercial settings. More generally, the discussion of risks from AI has paid insufficient attention to the nature of risk itself, as well as how decisions about the acceptability of the risks of AI compare to worries about convergent technologies. For example, global debates about autonomous weapons have focused predominantly on robotics, but AI also can be weaponized. For instance, in robotics serious concern exists over a possible lack of "meaningful human control" [1]. Missing is a similar concern for autonomous AI-controlled cyber attacks that would lack the very same control [2]. The Vice Chairman of the Joint Chiefs of Staff understands, saying, "In the [Defense] Department, we build machines and we test them until they break. You can't do that with an artificial intelligence, deep learning piece of software. We're going to have to figure out how to get the software to tell us what it's learned" [3]. Significant aspects of AI risk can thus be characterized as a "black-box" problem.

So, how best to understand the risks of AI, judge them (un)acceptable, and then apply our insights on risk to determine what policies to pursue?

2. DEFINING RISK, AND HOW TO THINK ABOUT IT

So, AI poses many different types of risk – but what exactly is risk? Andrew Maynard [4] suggests that we start with the idea of "value." If innovation is defined as creating value that someone is willing to pay for, then he suggests risk as a *threat to value*, and not just in the ways value is usually thought of when assessing risk, such as health, the environment or financial gain/loss, but also well-being, environmental sustainability, deeply held beliefs, even a sense of cultural or personal identity. Risk is thus a potential, but not certain, harm.

Extending a schema based on previous work [5], the major factors in determining 'acceptable risk' in AI will include (but are not limited to):

2.1 Acceptable-Risk Factor: Consent

Consent: Is the risk voluntarily endured, or not? For instance, secondhand smoke is generally more objectionable than firsthand, because the passive smoker did not consent to the risk, even if the objective risk is smaller. Will those who are at risk from AI reasonably give consent? When would it be appropriate to deploy or use AI without the meaningful consent of those affected? Would non-voluntariness (in which the affected party is unaware of the risk/ cannot consent) be morally different from involuntariness (in which the affected party is aware of the risk and does not consent)? [6]

2.2 Acceptable-Risk Factor: Informed Consent

Even if AIs only have a 'slave morality' in which they always follow orders [7], and citizens consent to their use (through, say, political means), that still leaves unanswered whether the risk (of malfunction, unintended consequences, or other error) to *unintended* parties is morally permissible. After all, even if widespread consent is in some sense possible, it is completely unrealistic to believe that all humans affected by AI could give *informed* consent to their use. So, does the morality of consent require adequate knowledge of what is being consented to?

Informed consent: Are those who undergo the risk voluntarily fully aware of the true nature of the risk? Or would such knowledge undermine their efficacy in fulfilling their (risky) roles? Or are there other reasons for preferring ignorance? Thus, will all those at risk from AI know that they are at risk? If not, do those who know have an obligation to inform others of the risks? What about foreseeable but unknown risks—how should they (the 'known unknowns') be handled? Could informing people that they are at risk ever be unethical, even akin to terrorism?

2.3 Acceptable-Risk Factor: The Affected Population

Even if consent or informed consent do not appear to be morally required with respect to some AI, we may continue to focus on the affected population as another factor in determining acceptable risk:

Affected population: Who is at risk—is it merely groups that are particularly susceptible or innocent, or those who broadly understand that their role is risky, even if they do not know the particulars of the risk? For example, in military operations civilians and other noncombatants are usually seen as not morally required to endure the same sorts of risks as military personnel, even (or especially) when the risk is involuntary or non-voluntary.

2.4 Acceptable-Risk Factor: Step risk versus State risk

A state risk is the risk of being in a certain state, and the total amount of risk to the system is a direct function of the time spent in the state. Thus, state risk is time-dependent; total risk depends (usually linearly) on the time spent in the state. So, for us living on the surface of the Earth, the risk of death by asteroid strike is a state risk (it increases the longer we're here).

Step risk, on the other hand, is a discrete risk of taking the next step in some series or undergoing some transition; once the transition is complete, the risk vanishes. In general, step risk is not time-dependent, so the amount of time spent on step matters little (or not at all). [8] Crossing a minefield is usually a step risk – the risk is the same whether you cross it in 1 minute or 10 minutes. The development of AGI poses a clear step risk; but, if there is a 'fast takeoff,' the state risk may be negligible.

2.5 Acceptable-Risk Factors: Seriousness and Probability

We thereby come to the two most basic facets of risk assessment, seriousness, and probability: how bad would the harm be, and how likely is it to happen?

Seriousness: A risk of death or serious physical (or psychological) harm is understandably seen differently than the risk of a scratch or a temporary power failure or slight monetary costs. But the attempt to make serious risks nonexistent may turn out to be prohibitively expensive. What (if any) serious risks from AIs are acceptable—and to whom: users, nonusers, the environment, or the AI itself?

Probability: This is sometimes conflated with seriousness but is intellectually quite distinct. The seriousness of the risk of a 10-km asteroid hitting Earth is quite high (possible human extinction), but the probability is reassuringly low (though not zero, as perhaps the dinosaurs discovered). What is the probability of harm from AIs? How much certainty can we have in estimating this probability? What probability of serious harm is acceptable? What probability of moderate harm is acceptable? What probability of mild harm is acceptable?

2.6 Acceptable-Risk Factors: Who Determines Acceptable Risk?

In various other social contexts, all of the following have been defended as proper methods for determining that a risk is unacceptable [9]:

Good faith subjective standard: It is up to each individual as to whether an unacceptable risk exists. That would involve questions such as the following: Can the designers or users of AI be trusted to make wise choices about (un)acceptable risk? The idiosyncrasies of human risk aversion may make this standard impossible to defend, as well as the problem of involuntary/ non-voluntary risk borne by nonusers.

The reasonable-person standard: An unacceptable risk is simply what a fair, informed member of a relevant community believes to be an unacceptable risk. Can we substitute a professional code or some other basis for what a 'reasonable person' would think for the difficult-to-foresee vagaries of conditions in the rapidly emerging AI field, and the subjective judgment of its practitioners and users? Or what kind of judgment would we expect an autonomous AI to have— would we trust it to accurately determine and act upon the assessed risk? If not, then can AI never be deployed without teleoperators—like military robots, should we always demand a human in the loop? But even a 'kill switch' that enabled autonomous operation until a human doing remote surveillance determined something had gone wrong would still leave unsolved the first-generation problem.

Objective standard: An unacceptable risk requires evidence and/or expert testimony as to the reality of (and unacceptability of) the risk. But there remains the first-generation problem: how do we understand that something is an unacceptable risk unless some first generation has already endured and suffered from it? How else could we obtain convincing objective evidence?

2.7 Acceptable-Risk Factors: The Wild Card: Existential Risk?

Plausibly, a requirement for extensive, variegated, realistic, and exhaustive pre-deployment testing of AIs in virtual environments before they are used in actual human interactions could render many AI risks acceptable under the previous criteria. But one AI risk may remain unacceptable even with the most rigorous pre-deployment testing. An existential risk refers to a risk that, should it come to pass, would either annihilate Earth-originating intelligent life or permanently and drastically curtail its potential. Existential disasters would end human civilization for all time to come. For utilitarians, existential risks are terribly important: doing what we can to mitigate even a small chance that humanity comes to an end may well be worth almost any cost. And for deontologists, the idea that 'one always has a moral obligation never to allow the extinction of all creatures capable of moral obligation' is at least a plausible *prima facie* (and perhaps absolute) duty; such a survival principle appears required for any viable ethics [10]. If there is even a tiny risk that developing AGI would pose an existential risk, this 'Extinction Principle' may well imply that we have a duty to stop it.

3. SPECIFIC CASE STUDY, POSSIBLE SOLUTION, AND RISK ANALYSIS

3.1 Specific Case Study

As noted above, a well-known problem of AI risk is applying Deep Learning Neural Networks, specifically the black-box problem. This occurs when the Neural Network has a large set of training data; the consequential connectivity arrangements are such that for a given input, an output is provided to a reinforced goal of the system – typically prediction, decision or identification. These bottom-up systems can learn well, especially in tight domains, and in them can surpass human performance – but, without any means of validation, or clear compliance to regulations. This renders the AI potentially untrustworthy, a serious deployment risk that may be deemed unacceptable. A human can usually provide a top-down rationalization of their behavior, responding to "why did you do that" questions. But a Deep Learning system cannot easily answer such queries. It is not rule-based and cannot easily track its "reasoning." [11]

We propose a possible solution. Our case study is a Deep Learning system applied to the Path Planning of a Robot Arm, in which vials of biohazardous materials are to be moved from point A to point B in an optimum path. This path is constrained by parameters such as speed, power-usage, minimization of actuator acceleration and deceleration (that causes wear of the actuators) and collision avoidance. See Fig 1. Keep in mind that cost of production, as well as quality/safety, are value factors to be balanced in this manufacturing example. And the use of AI Robots in this case example is a very real-world example of potential benefit and ethical concerns.

A key part of the solution proposed is an "Extractor" which will build a rationalization of the Deep Learning System into a Rule-Based Decision tree that can be validated against risk analysis/compliance needs, i.e., answering questions related to risks. This is depicted in Fig 2.

3.2 Possible Solution: Extraction of Heuristics from Deep Learning Neural Net

The means by which Expert Heuristics are extracted from the Deep Learning Neural Networks has been studied by other teams [12], [13] and [14]. The specific means by which we propose to do so in RUML^{SM} is an innovative patent pending process [15], [16] and [17]. Expert Heuristic/Rule extraction can be defined as "…given a trained neural network and the data on which it was trained, produce a description of the network's hypothesis that is comprehensible yet closely approximates the network's predictive behavior." Such extraction algorithms are useful for experts to verify and cross-check Neural Network systems. Earlier this year, John Launchbury, director of DARPA's Information Innovation Office said, "There's been a lot of hype and bluster about AI." They published their view of AI into "Three Waves," to explain what AI can do, what AI can't do, and where AI is headed. See Fig 3 [18]. We consider the example we have outlined above falls into this "Third Wave of AI."
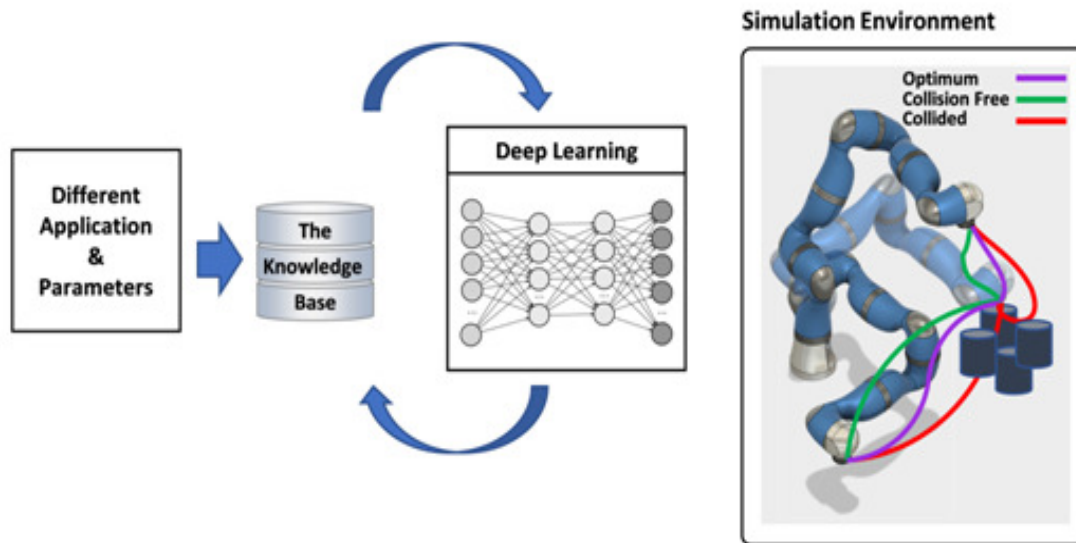
Figure 1: Deep Learning System Applied to Robot Arm Path Planning (Source: CalvIO Inc)
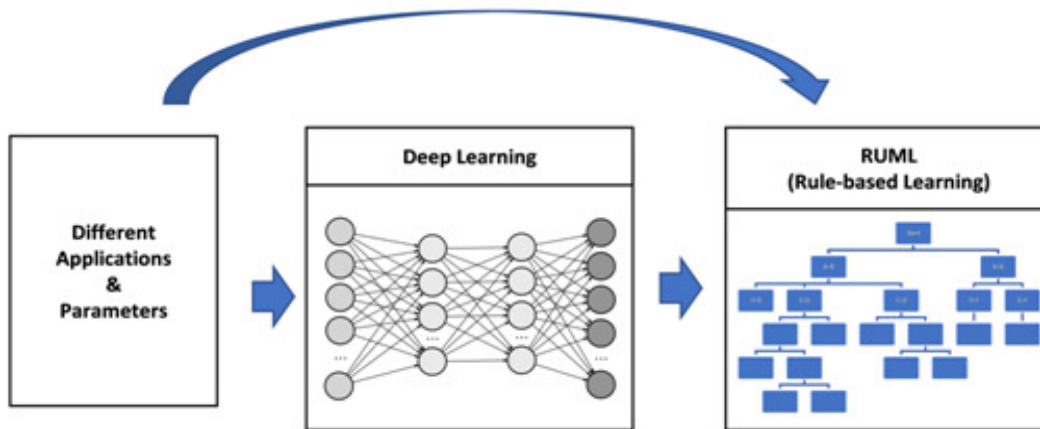
Figure 2: Deep Learning System Rule Extractor (RUML^{SM}) (Source: CalvIO Inc)
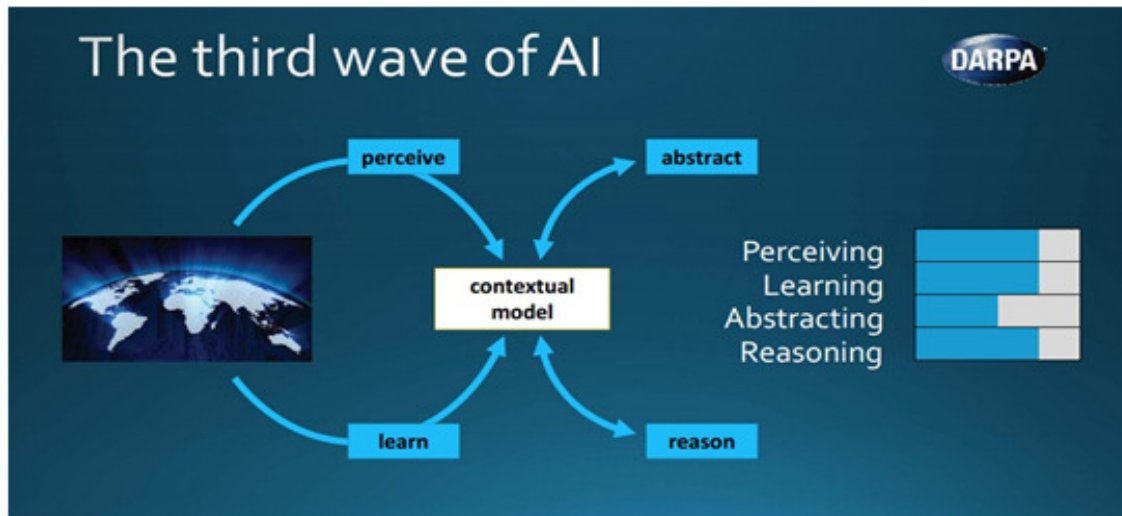
Figure 3: DARPA's Third Wave of AI (Source: DARPA)

3.3 AI Risk Conceptual Analysis applied to the Case Example

1/ Acceptable-Risk Factor: Consent
The use of a robot (in a protective clean room cell) reduces the need for human operator exposure to Biohazards.
2/ Acceptable-Risk Factor: Informed Consent
However, if the AI directing the robot causes breaches of the clean/safe room (e.g., collisions with the cell walls), then what was thought safe might not be. It is necessary to test any black-box defined behavior will not violate these rules.

3/ Acceptable-Risk Factor: The Affected Population	
The affected population might not be limited to the factory; conceivably, an extended exposure could cause health and safety threats to those outside, or violations of FDA regulations, etc. Again, the black-box must be validated against the regulations/laws applicable to the domain.	
4/ Acceptable-Risk Factor: Step risk versus State risk	
Both state and step risks need to be exposed through the black-box rule extraction process.	
5/ Acceptable-Risk Factors: Seriousness and Probability	
The seriousness of a biohazard breach can be evaluated in principle, but the probability may need testing. This should be done in simulation (to avoid physical exposure) as well as an assessment of the black-box extracted rule-set.	
6/ Acceptable-Risk Factors: Who Determines Acceptable Risk?	
There are industry bodies that set standards (e.g., GAMP5) as well as government entities that set regulations in this case example (e.g., US FDA).	
7/ Acceptable-Risk Factors: The Wild Card: Existential Risk?	
If the biohazard agent was severe enough, as might be possible with nuclear materials and/or live chemical/biological agents, then the impacts could be existential, if the AI goes "rogue." The severity relates properly to steps 3 and 5 above. The risk of 'going rogue' is conceivable, but presumably with complete AI automation of the industrial facility, and absent proper safeguards against hacking or rampancy. A solution may involve a software system over-riding ethical kernel that ensures "no harm."	

4. CONCLUSIONS

We reviewed the concept of AI Risk and picked a well-known one, i.e., the AI Neural Network Black-box problem. We proceeded to outline a means of structuring a dialog of these "Conceptual AI Risks."

We provided a near-term AI/Robotics case example (Smart Robotics Path Planning for a Pick and Place application for Biohazardous material) and applied the dialog to it; we think the result is an actual beneficial one for highlighting the AI risk concerns and start the process of handling them objectively.

The case example method applies a meta-level/hybrid AI system to "extract" heuristics from the neural network black-boxes (applying a top-down AI system on a bottom-up AI system). The system is based on technology called Really Useful Machine Learning (RUML[SM]). The approach taken is an example of DARPA's Third Wave of AI. In a sense this process can be considered as a Cybernetic Self-Regulation (extraction of a rationalized model for control).

As such, we believe the resulting techno-philosophy methodology to be a potentially useful early step in the building of tools for conceptualizing and assessing acceptable AI Risk. Further work is needed to develop these concepts, and trial them in real-world applications.

REFERENCES

[1]   UNIDIR: The weaponization of increasingly autonomous technologies: considering how Meaningful Human Control might move the discussion forward. UNIDIR Resources, no. 2, 2014. http://unidir.org/files/publications/pdfs/considering-how-meaningful-human-control-might-move-the-discussion-forward-en-615.pdf. Last Referenced 22nd October 2017.

[2]   Roff, H.: Monstermind or the doomsday machine? Autonomous cyberwarfare. Duck of Minerva, 13 August 2014. http://duckofminerva.com/2014/08/monstermind-or-the-doomsday-machine-autonomous-cyberwarfare.html. Last Referenced 22nd October 2017.

[3]   Clevenger, A.: 'The Terminator conundrum': Pentagon weighs ethics of pairing deadly force, AI. Army Times, 23 January 2016. http://www.armytimes.com/story/defense/policy-budget/budget/2016 /01/23/terminator-conundrum-pentagon-weighs-ethics-pairing-deadly-force-ai/79205722/. Last Referenced 22nd October 2017.

[4]   Andrew Maynard, "Thinking innovatively about the risks of tech innovation". The Conversation, January 12, 2016. https://theconversation.com/thinking-innovatively-about-the-risks-of-tech -innovation-52934. Last Referenced 22nd October 2017.

[5]   Lin, P., Mehlman, M., and Abney, K.: Enhanced warfighters: risk, ethics, and policy. Report funded by the Greenwall Foundation. California Polytechnic State University, San Luis Obispo. 1 January 2013. http://ethics.calpoly.edu/Greenwall_report.pdf. Last Referenced 22nd October 2017.

[6]   Abney, K., Lin, P., and Mehlman, M. "Military Neuroenhancement and Risk Assessment" in James Giordano (ed.), Neuroscience and Neurotechnology in National Security and Defense: Practical Considerations, Ethical Concerns (Taylor & Francis Group, 2014)

[7]   Lin, P., Mehlman, M., and Abney, K.: Enhanced warfighters: risk, ethics, and policy. Report funded by the Greenwall Foundation. California Polytechnic State University, San Luis Obispo. 1 January 2013. http://ethics.calpoly.edu/Greenwall_report.pdf . Last Referenced 22nd October 2017.

[8]   Nick Bostrom, Superintelligence. (Oxford University Press, 2014)

[9]   Abney, K., Lin, P., and Mehlman, M. "Military Neuroenhancement and Risk Assessment" in James Giordano (ed.), Neuroscience and Neurotechnology in National Security and Defense: Practical Considerations, Ethical Concerns (Taylor & Francis Group, 2014)

[10] Keith Abney, "Robots and Space Ethics," ch 23 in Robot Ethics 2.0, eds. Lin, P., Jenkins, R., and Abney, K. (Oxford University Press, 2017)

[11] Will Knight, (MIT Press), The Dark Secret at the Heart of AI - No one really knows how the most advanced algorithms do what they do. That could be a problem. April 11, 2017, https://www.technologyreview.com/s/604087/the-dark-secret-at-the-heart-of-ai/   Last Referenced 22nd October 2017.

[12] Tameru Hailesilassie, (IJCSIS) International Journal of Computer Science and Information Security,Vol. 14, No. 7, July 2016 "Rule Extraction Algorithm for Deep Neural Networks: A Review" https://arxiv.org/ftp/arxiv/papers/1610/1610.05267.pdf   Last Referenced 22nd October 2017.

[13] Jan Ruben Zilke, Master Thesis, TUD, "Extracting Rules from Deep Neural Networks" http://www.ke.tu-darmstadt.de/lehre/arbeiten/master/2015/Zilke_Jan.pdf   Last Referenced 22nd October 2017.

[14] Zhiting Hu, Xuezhe Ma, Zhengzhong Liu, Eduard Hovy, Eric P. Xing, School of Computer Science Carnegie Mellon University, 2016 "Harnessing Deep Neural Networks with Logic Rules" http://www.cs.cmu.edu/~epxing/papers/2016/Hu_etal_ACL16.pdf   Last Referenced 22nd October 2017.

[15] M Ciupa "Hybrid Machine Learning Design Based on a Bottom-up/Top-Down Methodology," US 62/476,068 United States Patent and Trademark Office, 2017.

[16] M. Ciupa, "Is AI in Jeopardy? The Need to Under Promise and Over Deliver - The Case for Really Useful Machine Learning," in Computer Science {\&} Information Technology (CS {\&} IT), 2017.

[17] Martin Ciupa, Nicole Tedesco, and Mostafa Ghobadi, "Automating Automation: Master Mentoring Process" 5th International Conference on Artificial Intelligence and Applications (AIAP-2018), Jan 2018, Zurich, Switzerland

[18] Steve Crowe, 21 February 21, 2017, Robotics Trends, "What AI Can and Can't Do: DARPA's Realistic View, http://www.roboticstrends.com/article/what_ai_can_and_cant_do_darpas_realistic_view/Artificial_Intelligence

AUTHORS

Martin Ciupa

Martin Ciupa is the CTO of calvIO Inc., a company (associated with the Calvary Robotics group of companies) focused on simplifying the cybernetic interaction between man and machine in the industrial setting. Martin has had a career in both technology, general management and commercial roles at senior levels in North America, Europe and Asia. He has an academic background in Physics and Cybernetics. He has applied AI and Machine learning systems to applications in decision support for Telco, Manufacturing and Financial services sectors and published technical articles in Software, Robotics, AI and related disciplines.

Keith Abney

Keith Abney is senior fellow in the Ethics + Emerging Sciences Group and senior lecturer in the Philosophy Department of California Polytechnic State University - San Luis Obispo. He is co-editor of Robot Ethics (MIT Press) and the newly published Robot Ethics 2.0 (Oxford UP).

INTENTIONAL BLANK

CNN Features Are Also Great at Unsupervised Classification

Joris Guérin, Olivier Gibaru, Stéphane Thiery and Eric Nyiri

Laboratoire des Sciences de l'Information et des Systèmes (CNRS UMR 7296)
Arts et Métiers ParisTech, Lille, France

ABSTRACT

This paper aims at providing insight on the transferability of deep CNN features to unsupervised problems. We study the impact of different pretrained CNN feature extractors on the problem of image set clustering for object classification as well as fine-grained classification. We propose a rather straightforward pipeline combining deep-feature extraction using a CNN pretrained on ImageNet and a classic clustering algorithm to classify sets of images. This approach is compared to state-of-the-art algorithms in image-clustering and provides better results. These results strengthen the belief that supervised training of deep CNN on large datasets, with a large variability of classes, extracts better features than most carefully designed engineering approaches, even for unsupervised tasks. We also validate our approach on a robotic application, consisting in sorting and storing objects smartly based on clustering.

KEYWORDS

Transfer learning, Image clustering, Robotics application

1. INTRODUCTION

In a close future, it is likely to see industrial robots performing tasks requiring to make complex decisions. In this perspective, we have developed an automatic sorting and storing application (see section 1.1.2) consisting in clustering images based on semantic content and storing the objects in boxes accordingly using a serial robot (`https://youtu.be/NpZIwY3H-gE`). This application can have various uses in shopfloors (workspaces can be organized before the workday, unsorted goods can be sorted before packaging, ...), which motivated this study of image-set clustering.

As of today, deep convolutional neural networks (CNN) [1] are the method of choice for supervised image classification. Since [2] demonstrated astounding results on ImageNet, all other methods have rapidly been abandoned for ILSVRC [3]. As suggested by [4], performances of CNN are highly correlated to the amount of labeled training data available. Nevertheless, even when few labels are available, several recent papers [5, 6, 7] have shown that CNN can still outperform any other approach by transferring knowledge learned on large datasets. In particular, [5] has shown that extracting features from OverFeat [8] pretrained on ImageNet, and training a simple Support Vector Machine (SVM) [9] classifier on these features to fit the new dataset

provides better results than many more complex approaches for supervised classification. These results demonstrate that a CNN, trained on a large and versatile dataset, learns information about object characteristics that is generic enough to transfer to objects that are not in the original dataset.

While developing the automatic robotic sorting and storing application, we needed to classify sets of images based on their content, in an unsupervised way. Multiple papers introduced methods to solve unsupervised object classification from sets of images (see section 1.1.1), producing relatively good results. However, we wanted to know if the information from a large and versatile dataset, stored in the weights of a CNN, could be used straightforwardly to outperform state-of-the-art algorithms at unsupervised imageset classification. The goal of this paper is to answer the following question: *How good are features extracted with a CNN pretrained on a large dataset, for unsupervised image classification tasks?* To do so, we use a similar approach to [5], consisting in applying classic clustering methods to features extracted from the final layers of a CNN (see section 2 for more details) and comparing it with state-of-the-art image set clustering algorithms [10, 11] on several public datasets.

The intuition behind such approach for unsupervised object classification is that, as it works with SVM [5], the CNN must project data in a feature space where they are somehow linearly separable. Thus, simple clustering algorithms such as K-means might be working well. However, this study is interesting as the performance of such simple clustering algorithms often depends on the notion of distance between points, on which we remain uncertain.

1.1. Previous work

1.1.1. Image-set clustering

Given a set of unlabeled images, the image-set clustering problem consists in finding subsets of images based on their content: two images representing the same object should be clustered together and separated from images representing other objects. Figure 1 illustrates the expected output from an image-set clustering algorithm in the case of our robotics application. This problem should not be confused with image segmentation [12], which is also sometimes called image clustering.

Image-set clustering has been widely studied for two decades. It has applications for searching large image database [13, 14, 15], concept discovery in images [16], storyline reconstruction [17], medical images classification [18], ... The first successful methods focused on feature selection and used sophisticated algorithms to deal with complex features. For instance, [19] represents images by Gaussian Mixture Models fitted to the pixels and clusters the set using the Information Bottleneck (IB) method [20]. [21] uses features resulting from image joint segmentation and sequential IB for clustering. [11] uses Bags of Features with local representations (SIFT, SURF, ...) and defines commonality measures used for agglomerative clustering. Recently, image-set clustering algorithms have shifted towards using deep features. [10] uses deep auto-encoders combined with ensemble clustering to generate feature representations suitable for clustering. [22, 18] learns jointly the clusters and the representation using alternating optimization [23].

1.1.2. Robotic automatic sorting application

Together with showing that deep *features + simple clustering* outperforms other approaches on unsupervised object classification, we apply this pipeline to solve an automatic smart robot sorting application first introduced in [24]. The idea is that the robot, equipped with a camera, visualizes a set of objects placed on a table. Given the number of storage boxes available, it needs to figure out the best way of sorting and storing the objects before physically doing it. The approach proposed in this paper exploits semantic information (deep features) while [24] uses a computer vision algorithm to extract shapes and colors. A video of the application can be seen at (`https://youtu.be/NpZIwY3H-gE`). An example of inputs/output of the application is shown in Figure 1. The robustness of this application is also investigated in section 4 by changing the lighting conditions, the position and orientation of the objects as well as the background. For this robustness validation, we built a dataset that appears to be a challenging one for image-set clustering (`http://www.usine-agile.fr/datas/22-cnn-datas-1.html`).
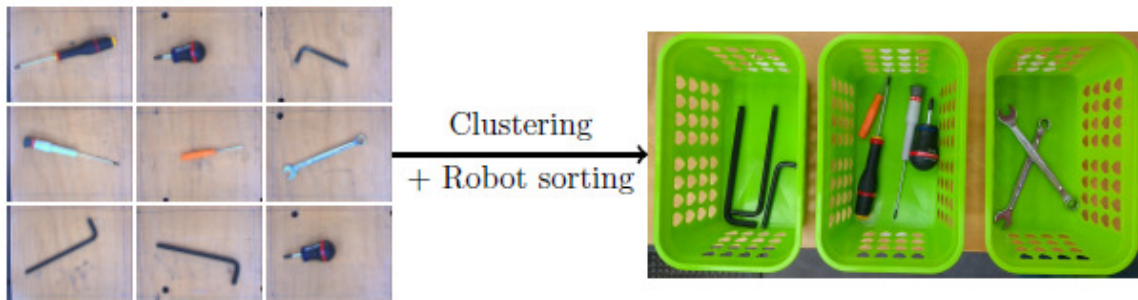
Figure 1: Robotic application description

1.2. Contributions

The main contribution of this paper is to convey further insight into deep CNN features by showing their scalability to unsupervised classification problems. We also propose a new baseline on several image clustering tasks.

Other contributions include the implementation of an application combining unsupervised image classification with robotic sorting. The method proposed to solve this problem, constitutes a new step towards autonomous decision-making robots. The dataset introduced in this paper, which is relatively challenging for image-set clustering, is also a contribution that can be used in further research to investigate robustness to background and lighting conditions for image clustering algorithms.

2. CLUSTERING IMAGES WITH DEEP FEATURE EXTRACTION

2.1. Pipeline description

The pipeline we propose for image set clustering is fairly straightforward. It consists in extracting deep features from all the images in the set, by using a deep convolutional neural network pretrained on a large dataset for image classification and then apply a "standard" clustering algorithm to these features. We initially tried this approach as a first step towards developing a

better clustering algorithm, however, it appears that this simple approach outperforms state-of-the-art algorithm at image-set clustering.

To implement this unsupervised image classification pipeline, we first need to answer four questions:

- What dataset should be used for pretraining?
- What CNN architecture should we use?
- Which layer output should be chosen for feature extraction?
- What clustering algorithm should be used?

As of today, ImageNet is the only very large labelled public dataset which has enough variety in its classes to be a good feature extractor for a variety of tasks. Moreover, there are plenty of CNN pretrained on ImageNet already available online. Hence, we will use a CNN pretrained on ImageNet. The three other questions are answered experimentally. We use the VOC2007 [25] test set without labels, which is new for the pretrained net, to compare performances of the different options.

To ease development and accelerate implementation, we compare the Keras [26] implementations of ResNet50 [27], InceptionV3 [28], VGG16, VGG19 [29] and Xception [30] with the pretrained weights provided by Keras. For the clustering algorithms, we use the scikit-learn [31] implementations of K-means (KM) [32], Minibatch K-means (MBKM) [33], Affinity Propagation (AP) [34], Mean Shift (MS) [35], Agglomerative Hierarchical Clustering (AC) [36], DBScan (DBS) [37] and Birch (Bi) [38]. For each CNN, the layers after which the features are extracted can be found in Table 1 (Layers names are the same as in the Keras implementations).

In the image-set clustering problem, the expected classes are represented by objects present on the picture and for this reason we need semantic information, which is present in the final layers of the network. Thus, we only choose layers among the last layers of the networks. On the one hand, the last one or two layers might provide better results as their goal is to separate the data (at least for the fully-connected layers). On the other hand, the opposite intuition is also relevant as we can imagine that these layers are too specialized to be transferable. These two contradictory arguments motivated the following experiment.

We also note that the test set of VOC2007 has been modified for this validation. We removed all the images presenting two or more labels in order to have ground truth to compute Normalized Mutual Information (NMI) scores. Indeed, if an image possesses several labels we cannot judge if the clustering pipeline classified it properly or not. We note VOC2007-SL (single label) the modified VOC2007 test set.

2.2. Hyperparameters Choice

To answer the questions stated above, we try to cluster the VOC2007-SL set using all combinations of CNN architectures, layer choices and clustering algorithms. To compare performances, we use NMI scores. We also report clustering time for completeness. Only scikit-learn default hyperparameters of the different clustering algorithms are used, which illustrate the simplicity of this approach. For KM and MBKM, as the results depend on random initialization, experiments are run ten times and reported results are averaged over the different runs.

Looking at the results, we choose Agglomerative Clustering on features extracted from the final layer of an Xception CNN pretrained on ImageNet for image-set clustering. This pipeline is then compared to state-of-the-art methods in the next section in order to see how transferable CNN ImageNet features are for unsupervised categorization.

3. VALIDATION ON SEVERAL PUBLIC DATASETS

3.1. Datasets Description

The efficiency of the proposed method is demonstrated by comparing it to other recent successful methods (see section 3.2) on several public datasets which characteristics are described in Table 2.

Table 1: NMI scores (in black) and time in seconds (in blue, italics) on Pascal VOC2007-SL test set using different CNN, different output layers and different clustering algorithms. (Layers names are the same as in the Keras implementations).

		KM	MBKM	AP	MS	AC	DBS[1]	Bi
Inception V3	mixed9	0.108	0.105	0.219	0.153	0.110	0	0.110
		374	*7.3*	*12.7*	*16281*	*525*	*138*	*577*
	mixed10	0.468	0.401	0.442	0.039	0.595	0	0.595
		609	*5.1*	*8.5*	*12126*	*525*	*119*	*567*
	avg pool	0.674	0.661	0.621	0.024	0.686	0	0.686
		6.3	*0.2*	*7.7*	*230*	*8.5*	*1.8*	*8.9*
Resnet 50	avg pool	0.6748	0.641	0.587	0.043	0.640	0	0.640
		7.0	*0.1*	*4.6*	*197*	*8.0*	*1.9*	*8.9*
VGG 16	block4 pool	0.218	0.085	0.133	0.124	0.277	0	0.277
		278	*3.6*	*6.0*	*10010*	*391*	*82.8*	*436*
	block5 pool	0.488	0.048	0.262	0.194	0.530	0	0.530
		78	*1.1*	*9.3*	*2325*	*99*	*21*	*107*
	fc1	0.606	0.458	0.421	0.187	0.617	0	0.617
		17	*0.2*	*4.8*	*365*	*17*	*3.8*	*19*
	fc2	0.661	0.611	0.551	0.085	0.673	0	0.673
		16	*0.2*	*4.3*	*373*	*15.9*	*3.8*	*19.7*
VGG 19	block4 pool	0.203	0.139	0.124	0.135	0.234	0	0.234
		220	*3.7*	*6.3*	*10298*	*388*	*83*	*435*
	block5 pool	0.522	0.321	0.250	0.198	0.540	0	0.540
		74	*0.9*	*9.3*	*2353*	*97*	*20*	*106*
	fc1	0.607	0.471	0.449	0.188	0.628	0	0.628
		17	*0.2*	*9.3*	*365*	*17*	*3.9*	*18*
	fc2	0.672	0.615	0.557	0.083	0.671	0	0.671
		15	*0.2*	*5.6*	*391*	*17*	*3.9*	*18*
Xception	block13 pool	0.376	0.264	0.351	0.044	0.473	0	0.473
		410	*4.8*	*10*	*9677*	*403*	*87*	*444*
	block14 act	0.574	0.428	0.584	0.071	0.634	0	0.634
		935	*10*	*10*	*24809*	*820*	*180*	*901*
	avg pool	0.692	0.636	0.636	0.052	**0.726**	0	0.726
		7.1	*0.1*	*4.9*	*201*	***8.5***	*5.5*	*9.1*

Table 2: Several key features about the datasets used for method validation.

	Problem type	Image Size	# Classes	# Instances
COIL100 [39]	Object recognition	128×128	100	7201
Nisters [40]	Object recognition	640×480	2550	10200
ORL [41]	Face recognition	92×112	40	400
VOC2007-5-ML [2]	Object recognition	variable	5	3376

The clustering tasks involved by these datasets are different from each others (Face recognition, grouping different objects, recognizing different pictures of the same object). In addition, the content of the classes differs from the ones in ImageNet. For these reasons, the four datasets constitute a good benchmark to quantify the robustness of transfer learning for unsupervised object categorization.

3.2. Results Comparison

We propose a comparison with the results reported in the following papers dealing with image set clustering:

[11] proposes different clustering algorithms applied on bags of features. In Table 3, we note "BoF" the best results obtained by such pipeline on the different datasets.

[10] proposes a method called infinite ensemble clustering (IEC). In the paper, IEC algorithm is compared to several other deep clustering algorithms and ensemble clustering algorithms. In Table 3, we report the best results obtained using Deep Clustering (DC) and Ensemble Clustering (EC) for each datasets. We note that for VOC2007-5-ML, [10] also uses deep features as clustering inputs (the CNN used is not reported).

[22] proposes a method called Joint Unsupervised Learning (JULE) of Deep Representations and Image Clusters, based on Alternating optimization between clustering and weight optimization of the CNN feature extractor. Results from this work are reported in Table 3.

For each dataset groundtruth is known as they are intended for supervised classification. We compute both NMI scores and purity for each dataset/method pair.

Table 3 shows that features extracted by the final layer of Xception combined with Agglomerative Clustering performs better than or close to state-of-the-art methods at unsupervised object classification as well as fine-grained image clustering (ORL). Results on the ORL dataset are interesting as they show that pretrained Xception is able to classify different faces without supervision, although ImageNet does not deal with human faces at all.

---

[1]The poor results with DB-Scan might come from the default parameters. We might get better results using different configurations, but this is out of the scope of this paper.

[2]The data used for VOC2007 in [10] are irrelevant for clustering with hard assignment. The VOC2007 subset used in [10] contains images with several objects but only one label. However, we still ran our clustering method on this data to be able to compare results. This second modified VOC2007 set is denoted VOC2007-5-ML (5 classes - multiple labels)

Table 3: NMI scores and purity comparison on various public datasets. (A result that is not reported in the papers cited above is denoted N.R.)

NMI scores					
	DC	EC	BoF	JULE	Ours (Xception + AC)
COIL100	0.779	0.787	N.R.	0.985	0.951
Nisters	N.R.	N.R.	0.918	N.R.	0.991
ORL	0.777	0.805	0.878	N.R.	0.93
VOC2007-5-ML	0.265	0.272	N.R.	N.R.	0.367

Purity			
	DC	EC	Ours (Xception + AC)
COIL100	0.535	0.546	0.882
Nisters	N.R.	N.R.	0.988
ORL	0.578	0.630	0.875
VOC2007-5-ML	0.513	0.536	0.622

This is an important result as it shows that, *with today's methods, given an unlabeled image-set, we can extract more information from a large labeled dataset, with a large variety of classes, than from the set itself*. It is better than hand-engineered approaches (BoF) as well as unsupervised trained deep networks (Deep clustering and Ensemble clustering). It also raises the question of how the representation learning problem should be handled. Indeed, although less satisfactory from a research perspective, it might be more appropriate to work on the creation of a larger database and train networks on it so that it can be used as a knowledge base for many other problems.

We underline the very good results of JULE ([22]) at clustering COIL100. Regarding the results of this methodology on Scene clustering [18], it appears that fine tuning feature extraction using alternating optimization is a good way of improving clustering results. However, the simple approach proposed here still keeps the advantage of being very fast (as it only requires to evaluate the network once for each sample and apply AC), which is useful for our application for instance.

It is also interesting to notice that [10] is also using CNN features for clustering VOC2007-5-ML. The fact that our pipeline outperform largely DC and EC for this dataset illustrate that when using transfer learning for a new problem, careful selection of the CNN used is important, which is often neglected.

3.3. A Word on Scene Clustering

The problem studied in this paper is the one raised by the robotic application, unsupervised objects sorting. For this reason, we compared our result at different object categorization tasks as well as fine-grained classification for clustering of similar objects. Another interesting image classification problem is the one of scene clustering, studied in [18] on two datasets ([42, 43]). For this task, the pipeline proposed in this paper cannot perform as well as for object classification. Indeed, ImageNet does not contain any class which requires to group objects together. Thus, although the features are good at supervised scene classification, without further training they are not able to group objects together as such behaviour is not encoded in the final layers of the CNN. However, this issue is not inherent to the method defined and we believe that with bigger and more versatile datasets, results would be as good as any other method.

4. APPLICATION VALIDATION

In the setting tested initially (Figure 1), where the set of objects to cluster is composed of screw drivers, at keys, allen keys and clamps, and where the background is a table, the success rate of the application described in section 1.1.2 is 100%. Although for certain classes (flat and allen keys) the intra-cluster similarity is high, it is not the case for the two others. This task is also difficult to solve because we carried out the experiments in a shopfloor under unmastered lighting.

For further testing of the application robustness, we have built a dataset for pixel-based object clustering. The full dataset, together with its description, can be found at (http//www.usine-agile.fr/datas/22-cnn-datas-1.html) and example images can be seen on Figure 2. The dataset statistics are reported in table 4. This dataset is difficult because some classes have low intra-cluster similarity (usb) and extra-cluster similarity between some classes is relatively high (pens/screws). The lighting conditions as well as the background also change within the dataset, which makes the task even harder.

Table 4: Several key features about the constructed dataset.

Problem type	Image Size	# Classes	# Instances
Object recognition	640×480	7	560

For each conditions, we randomly pick one out of the several pictures (different position /orientation) of each object. The results reported in the top subtable of Table 5 are averaged over 100 random combinations. The clustering results are not perfect, looking at the misclassifications, the main source of error comes from classes with low intra-cluster similarity (pens, usb) and from background containing sharp edges (conditions 4 and 5).

Figure 2: Example images from the robustness validation dataset. Objects in the same row are expected to be clustered in the same group. The five different background/lighting conditions are represented in this figure.

We also carry out an experiment of fine grained classification with this dataset. Within each class, we try to group together the pictures representing the exact same object. Purity results can be found in the bottom subtable of Table 5. An interesting fact can be noticed about object

recognition within a category. Classes responsible for decreasing the clustering quality in the first subtable are objects with the highest purity in the second table. Such remark makes sens as low intra-cluster similarity is good for the second task but harmful for the first one.

Looking at the results, this robustness validation dataset appears to be a challenging one for image-set clustering and could be used to validate further research in the field.

Table 5: Clustering accuracies for different background, lighting and orientation conditions on the tool clustering dataset.

Grouping by category

Metric \ Condition	1	2	3	4	5	Mixed
Purity	0.85	0.94	0.84	0.69	0.79	0.58
NMI score	0.87	0.94	0.83	0.71	0.82	0.54

Recognizing object inside a category

Object \ Condition	1	2	3	4	5
Allen	0.58	0.58	0.67	0.67	0.67
Clamp	0.83	0.92	0.67	0.58	0.67
Driver	0.75	0.75	0.75	0.63	0.75
Flat	0.58	0.83	0.67	0.58	0.58
Pen	1.0	1.0	0.75	0.69	1.0
Screws	0.5	0.56	0.75	0.50	0.56
USB	0.80	0.65	0.85	0.55	0.65

5. CONCLUSION AND PERSPECTIVES

5.1. Conclusive Remarks

This paper extends the interesting work of [5, 6, 7] about the transferability of CNN features. It shows that, even for unsupervised classification tasks, features extracted from deep CNN trained on large and diverse datasets, combined with classic clustering algorithms, can compete with more sophisticated and tuned image-set clustering methods. The fairly simple and naive pipeline proposed outperforms the best results reported in recent work, which raises the question of which research direction should be chosen to reach generic knowledge. Are efforts spent in developing image representation extractors more useful than simply building larger and more diverse datasets?

This approach is used to implement a robotic application using unsupervised image classification to store objects smartly. To validate this application, we also built a challenging dataset for image clustering that is made available to the research community.

5.2. Future Work

The proposed improvements mainly go in the direction of the robotics application, which is still not robust enough to adapt perfectly to very different looking objects within a cluster and to difficult backgrounds and lighting conditions. If we want to make it work in difficult

environments, the clustering pipeline needs to be improved. One possible direction is to tune the final clustering algorithm, indeed, the scikit clustering algorithms are used without any parameter tuning, setting hyperparameters to their default values.

The sorting application can also be improved by introducing automatic image segmentation, which would make it more suitable for practical uses. To do this, we could use a pretrained region proposal network [44] and cluster objects in the proposed regions.

ACKNOWLEDGEMENTS

REFERENCES

[1]   Y. LeCun, Y. Bengio, and G. Hinton, \Deep learning," Nature, vol. 521, no. 7553,pp. 436{444, 2015.

[2]   A. Krizhevsky, I. Sutskever, and G. E. Hinton, \Imagenet classification with deep convolutional neural networks," in Advances in neural information processing systems, 2012, pp. 1097{1105.

[3]   O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein et al., \Imagenet large scale visual recognition challenge," International Journal of Computer Vision, vol. 115, no. 3, pp. 211{252, 2015.

[4]   S. Uchida, S. Ide, B. K. Iwana, and A. Zhu, \A further step to perfect accuracy by training cnn with larger data," in Frontiers in Handwriting Recognition (ICFHR), 2016 15th International Conference on. IEEE, 2016, pp. 405{410.

[5]   A. Sharif Razavian, H. Azizpour, J. Sullivan, and S. Carlsson, \Cnn features off-the-shelf: an astounding baseline for recognition," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2014, pp. 806{813.

[6]   J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, \How transferable are features in deep neural networks?" in Advances in neural information processing systems, 2014, pp. 3320{3328.

[7]   J. Donahue, Y. Jia, O. Vinyals, J. Hoffman, N. Zhang, E. Tzeng, and T. Darrell, \Decaf: A deep convolutional activation feature for generic visual recognition." in Icml, vol. 32, 2014, pp. 647{655.

[8]   P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun, \Over-feat: Integrated recognition, localization and detection using convolutional networks," arXiv preprint arXiv:1312.6229, 2013.

[9]   S. Theodoridis and K. Koutroumbas, \Pattern recognition," 2003.

[10] H. Liu, M. Shao, S. Li, and Y. Fu, \Infinite ensemble for image clustering," in Proceedings of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016.

[11] T. Fukui and T. Wada, \Commonality preserving image-set clustering based on diverse density," in International Symposium on Visual Computing. Springer, 2014, pp. 258{269.

[12] N. Dhanachandra and Y. J. Chanu, \A survey on image segmentation methods using clustering techniques," European Journal of Engineering Research and Science, vol. 2, no. 1, 2017.

[13] M. Flickner, H. Sawhney, W. Niblack, J. Ashley, Q. Huang, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovic et al., \Query by image and video content: The qbic system," computer, vol. 28, no. 9, pp. 23{32, 1995.

[14] Y. Gong, M. Pawlowski, F. Yang, L. Brandy, L. Bourdev, and R. Fergus, \Web scale photo hash clustering on a single machine," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 19{27.

[15] Y. Avrithis, Y. Kalantidis, E. Anagnostopoulos, and I. Z. Emiris, \Web-scale image clustering revisited," in Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 1502{1510.

[16] Y. J. Lee and K. Grauman, \Shape discovery from unlabeled image collections," in Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on. IEEE, 2009, pp. 2254{2261.

[17] G. Kim, L. Sigal, and E. P. Xing, \Joint summarization of large-scale collections of web images and videos for storyline reconstruction," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 4225{4232.

[18] X. Wang, L. Lu, H.-c. Shin, L. Kim, M. Bagheri, I. Nogues, J. Yao, and R. M. Summers, \Unsupervised joint mining of deep features and image labels for large-scale radiology image categorization and scene recognition," arXiv preprint arXiv:1701.06599, 2017.

[19] J. Goldberger, S. Gordon, and H. Greenspan, \Unsupervised image-set clustering using an information theoretic framework," IEEE transactions on image processing, vol. 15, no. 2, pp. 449{458, 2006.

[20] N. Tishby, F. C. Pereira, and W. Bialek, \The information bottleneck method," arXiv preprint physics/0004057, 2000.

[21] Y. Seldin, S. Starik, and M. Werman, \Unsupervised clustering of images using their joint segmentation," in Proceedings of the 3rd International Workshop on Statistical and Computational Theories of Vision (SCTV 2003), 2003.

[22] J. Yang, D. Parikh, and D. Batra, \Joint unsupervised learning of deep representations and image clusters," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 5147{5156.

[23] J. C. Bezdek and R. J. Hathaway, \Some notes on alternating optimization," in AFSS International Conference on Fuzzy Systems. Springer, 2002, pp. 288{300.

[24] J. Guerin, O. Gibaru, S. Thiery, and E. Nyiri, \Clustering for different scales of measurement-the gap-ratio weighted k-means algorithm," arXiv preprint arXiv:1703.07625, 2017.

[25] M. Everingham, A. Zisserman, C. K. Williams, L. Van Gool, M. Allan, C. M. Bishop, O. Chapelle, N. Dalal, T. Deselaers, G. Dorko et al., \The pascal visual object classes challenge 2007 (voc2007) results," 2007.

[26] F. Chollet, \Keras," 2015.

[27] K. He, X. Zhang, S. Ren, and J. Sun, \Deep residual learning for image recognition," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770{778.

[28] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, \ Rethinking the inception architecture for computer vision," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 2818{2826.

[29] K. Simonyan and A. Zisserman, \Very deep convolutional networks for large-scale image recognition," arXiv preprint arXiv:1409.1556, 2014.

[30] F. Chollet, \Xception: Deep learning with depthwise separable convolutions," arXiv preprint arXiv:1610.02357, 2016.

[31] L. Buitinck, G. Louppe, M. Blondel, F. Pedregosa, A. Mueller, O. Grisel, V. Niculae, P. Prettenhofer, A. Gramfort, J. Grobler, R. Layton, J. VanderPlas, A. Joly, B. Holt, and G. Varoquaux, \API design for machine learning software: experiences from the scikit-learn project," in ECML PKDD Workshop: Languages for Data Mining and Machine Learning, 2013, pp. 108{122.

[32] D. Arthur and S. Vassilvitskii, \k-means++: The advantages of careful seeding," in Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms. Society for Industrial and Applied Mathematics, 2007, pp. 1027{1035.

[33] D. Sculley, \Web-scale k-means clustering," in Proceedings of the 19th international conference on World wide web. ACM, 2010, pp. 1177{1178.

[34] D. Dueck and B. J. Frey, \Non-metric affinity propagation for unsupervised image categorization," in Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on. IEEE, 2007, pp. 1{8.

[35] D. Comaniciu and P. Meer, \Mean shift: A robust approach toward feature space analysis," IEEE Transactions on pattern analysis and machine intelligence, vol. 24, no. 5, pp. 603{619, 2002.

[36] F. Murtagh, \A survey of recent advances in hierarchical clustering algorithms," The Computer Journal, vol. 26, no. 4, pp. 354{359, 1983.

[37] M. Ester, H.-P. Kriegel, J. Sander, X. Xu et al., \A density-based algorithm for discovering clusters in large spatial databases with noise." in Kdd, vol. 96, no. 34, 1996, pp. 226{231.

[38] T. Zhang, R. Ramakrishnan, and M. Livny, \Birch: an efficient data clustering method for very large databases," in ACM Sigmod Record, vol. 25, no. 2. ACM, 1996, pp. 103{114.

[39] S. Nayar, S. Nene, and H. Murase, \Columbia object image library (coil 100)," Department of Comp. Science, Columbia University, Tech. Rep. CUCS-006-96, 1996.

[40] D. Nister and H. Stewenius, \Scalable recognition with a vocabulary tree," in Computer vision and pattern recognition, 2006 IEEE computer society conference on, vol. 2. Ieee, 2006, pp. 2161{2168.

[41] F. S. Samaria and A. C. Harter, \Parameterisation of a stochastic model for human face identification," in Applications of Computer Vision, 1994., Proceedings of the Second IEEE Workshop on. IEEE, 1994, pp. 138{142.

[42] A. Quattoni and A. Torralba, \Recognizing indoor scenes," in Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on. IEEE, 2009, pp. 413{420.

[43] Z. Xu, D. Tao, Y. Zhang, J. Wu, and A. C. Tsoi, \Architectural style classification using multinomial latent logistic regression," in European Conference on Computer Vision. Springer, 2014, pp. 600{615.

[44] R. B. Girshick, \Fast R-CNN," CoRR, vol. abs/1504.08083, 2015. [Online]. Available: http://arxiv.org/abs/1504.08083

AUTHORS

Joris Guérin received the diplome d'ingénieur (equivalent to M.Sc. degree) from Arts et Metiers ParisTech and the M.Sc. in Industrial Engineering from Texas Tech University, both in 2015. He is currently a Ph.D student at Laboratoire des Sciences de l'Information et des Systémes (LSIS), at Arts et Métiers ParisTech, Lille, France. His current research focuses on Clustering and Reinforcement Learning for Robotics manipulation.

Olivier Gibaru is currently full professor at the Department of Mathematics and Computer Science at ENSAM, Lille campus. He obtained his PhD in applied mathematics in 1997. His main research interests includes: applied mathematics, estimation for robotic applications, geometry, control engineering and high precision mechanical systems. He is the coordinator of the EU Horizon 2020 ColRobot project www.colrobot.eu. He is an active member of the SMAI-SIGMA group which is a national learned society dedicated to Applied Mathematics for the Industrial Applications.

Stéphane Thiery received the Ph.D degree in Automatics from University of Nice-Sophia Antipolis, France, in 2008. He was a post-doctoral fellow in the NON-A team in INRIA-Lille, for eight months in 2009-2010, and joined Arts et Métiers ParisTech Engineering School, as assistant professor in Applied Mathematics and Automatics, in 2010. His current research includes machine learning, real-time parameters estimation, and control of mechanical systems.
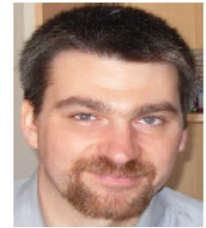
Eric Nyiri received the Ph.D degree in Computer Science from University of Lille I, France, in 1994. He joined Arts et Métiers ParisTech Engineering School, as an assistant professor in Applied Mathematics and Computer Science, in 1995. His initial research domain was L1 interpolation and approximation. In 2010, he joined the LSIS Lab and his current research includes machine learning and path planning for robots. Since 2016, he is a member of the COLROBOT European project.

INTENTIONAL BLANK

COMPARISON OF BANKRUPTCY PREDICTION MODELS WITH PUBLIC RECORDS AND FIRMOGRAPHICS

Lili Zhang[1], Jennifer Priestley[2] and Xuelei Ni[3]

[1]Program in Analytics and Data Science,
Kennesaw State University, Georgia, USA
[2]Analytics and Data Science Institute,
Kennesaw State University, Georgia, USA
[3]Department of Statistics, Kennesaw State University, Georgia, USA

ABSTRACT

Many business operations and strategies rely on bankruptcy prediction. In this paper, we aim to study the impacts of public records and firmographics and predict the bankruptcy in a 12-month-ahead period with using different classification models and adding values to traditionally used financial ratios. Univariate analysis shows the statistical association and significance of public records and firmographics indicators with the bankruptcy. Further, seven statistical models and machine learning methods were developed, including Logistic Regression, Decision Tree, Random Forest, Gradient Boosting, Support Vector Machine, Bayesian Network, and Neural Network. The performance of models were evaluated and compared based on classification accuracy, Type I error, Type II error, and ROC curves on the hold-out dataset. Moreover, an experiment was set up to show the importance of oversampling for rare event prediction. The result also shows that Bayesian Network is comparatively more robust than other models without oversampling.

KEYWORDS

Bankruptcy Prediction, Public Records, Firmographics, Classification, Oversampling

1. INTRODUCTION

Bankruptcy prediction has been studied since the 1960s,to improve decision making related to business operations conducted with reliable counterparties [3]. For example, investors want to make investments to organizations that have high potential to succeed. Banks want to lend to the organizations that are less likely to default. Business entities want to do business and build relationships with the ones that can prosper and survive in a long term. Hence, it is valuable to foresee the possibility of the bankruptcy of a business customer or partner.

To improve the accuracy of bankruptcy prediction, researchers and practitioners have pursued two primary paths of study. First, explore important variables for bankruptcy prediction. For example, the predictive ability of financial ratio variables has been thoroughly studied. Second, improve the methodologies used for the bankruptcy prediction, benefiting from the development of both the algorithm theories and computation infrastructure. Besides significant variables and high-performance methods, we observe that appropriate data sampling before modeling is also

important for improving bankruptcy prediction, considering that frequently the proportion of bankruptcy cases is substantively lower than the proportion of non-bankruptcies.

In this paper, we aim to make contributions from all above perspectives. First, we explore the impacts of public records and firmographics on bankruptcy prediction to add values to widely used financial ratio variables. Both univariate analysis and multiple variable analysis were conducted to measure statistical association and significance. With significant variables selected, we comprehensively compare seven classification models from the statistics and machine learning domains, including Logistic Regression, Decision Tree, Random Forest, Gradient Boosting, Support Vector Machine, Neural Network, and Bayesian Network. The performance of the models are evaluated on the hold-out dataset. The overall classification accuracy, Type I error, Type II error, and ROC curves are evaluated. Finally, we demonstrate the importance of oversampling for the rare event prediction like bankruptcy prediction, and demonstrate the robustness of the Bayesian Network for rare event modeling.

The paper is structured as follows. In Section 2, related work is reviewed. In Section 3, the data processes are described. In Section 4,the univariate analysis between the dependent variable and each individual input variable is performed. In Section 5, the models are developed, diagnosed, evaluated, and compared. In Section 6 and 7, conclusions and future work are discussed.

2. RELATED WORK

Because of its importance in business decisions like investment and loan lending, the bankruptcy prediction problem has been studied through deriving significant predictors and developing novel prediction models. Altman proposed a set of traditional financial ratios, including Working Capital/Total Assets, Retained Earnings/Total Assets, Earnings before Interest and Taxes/Total Assets, Market Value Equity/Book Value of Total Debt, and Sales/Total Assets, and used them in the multiple discriminant analysis for the corporate bankruptcy prediction [2].Those financial ratios were widely adopted and extended later [13] [4]. Amircame up with some novel financial ratio indicators, including BookValue/Total Assets, Cashflow/Total Assets, Price/Cashflow, Rate of Change of Stock Price, and Rate of Change of Cashflow perShare, in addition to Altman's ones, for a neural network model, and increased the prediction accuracy by 4.04% for a three-year-ahead forecast [4]. Everett et al. studied the impact of external risk factors(i.e. macro-economic factors) on small business bankruptcy prediction and proposed a logistic regression model [7]. Chavaet al. demonstrated the statistical significance of industry effects by grouping firms into finance/insurance/real estate, transportation/communications/utilities, manufacturing /mineral, and miscellaneous industries [6].

From the methodology perspective, various statistical methods, machine learning algorithms, and hybrid models have been applied and compared for the bankruptcy prediction problem. Odom et al. proposed the first neural network model for bankruptcy prediction [13]. Zhang et al. showed that the neural network performed better than logistic regression and were robust to sampling variations [17].Shin et al. found that the support vector machine outperformed the neural network on small training datasets [14]. Min et al. applied support vector machine with optimal kernel function hyper parameters[12]. Zibanez had showed the acceptable prediction ability of decision tree on the bankruptcy prediction problem and determined the most important financial ratios [8]. Zikeba et al. proposed and evaluated a novel gradient boosting method for learning an ensemble of trees [18]. Sun et al. studied the application of Bayesian network on the bankruptcy prediction problem in respects of the influence of variable selection and variable discretization on the model performance [15]. Ahn et al. presented a hybrid methodology by combining rough set theory and neural network [1]. Huanget al. proposed a hybrid model by incorporating static and trend analysis in the neural network training [9]. Kumar et al. provided a comprehensive review on both

the financial ratio variables and methods used for the bankruptcy prediction from 1968 to 2005, discussed merits and demerits of each method, and listed some important directions for future research [11]. Bellovary et al.reviewed 165 existing studies for the bankruptcy prediction and made some suggestions, where one suggestion was that the model accuracy was not guaranteed with the number of factors [5].

Most models proposed for bankruptcy prediction in the literature were directly developed on the dataset with a balanced proportion of bankruptcy and non-bankruptcy observations. However, data imbalance is a common issue in practice. Kim et al. proposed a geometric mean based boosting algorithm to address the data imbalance problem in the bankruptcy prediction, but only compared it with other boosting algorithms to show its advantage [19]. Zhou studied the effect of sampling methods for five bankruptcy prediction models, but the models were not tuned to their optimal hyperparameters [20].

The models applied to the bankruptcy prediction utilize a variety of algorithms. Logistic Regression formulates a function between the probability of the event ($\hat{p}$) and input variables $(x_1, x_2, \dots, x_n)$ defined as:

$$\hat{p} = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \dots + \beta_n x_n)}}$$

The coefficients $(\beta_1, \beta_2, \dots, \beta_n)$ in the functionare estimated by optimizing the maximum likelihood functiondefined as below, where y is the actual value with the event denoted as 1 and the nonevent denoted as 0.

$$\max y log\hat{p} + (1 - y) \log(1 - \hat{p})$$

Decision Tree defines hierarchical rules by searching for optimal splits on input variables based on the Entropy or Gini index. The Entropy and Gini index of an input variable are defined below, where x is a given input variable,$1, \dots, k$ are levels in the dependent variable, and $p(i|k)$ is the conditionalprobability for the dependent variable taking value i given$x$ [16].

$$Entropy(x) = -\sum_{i=1}^{k} p(i|k) log_2\big(p(i|k)\big)$$

$$Gini(x) = 1 - \sum_{i=1}^{k} [p(i|k)]^2$$

Random Forest and Gradient Boosting are an ensemble of multiple decision tree models through bagging and boosting, respectively. In Random Forest, each tree is trained independently on a bootstrap dataset created from the original training dataset and then combined to a single prediction model by taking the average of all trees. In Gradient Boosting, each tree is trained sequentially based on a modified version of the original training dataset by utilizing the information of previously trained trees [10]. In tree-based models, a summary of variable importance can be obtained. The importance of each input variable is measured based on the Entropyor Gini reduction by splitting a given input variable. The larger thevalue is, the more important an input variable is.

Support Vector Machine defines a hyperplane for two-class classification by maximizing the marginal distance. To handle the nonlinear relationship, a kernel function can be first applied to project the input variables to a higher feature space. Neural Network learns the relationship between the dependent variable and input variables by first transforming input variables with an activation function(Tanh, Sigmoid, etc.) through each hidden unit in one or more hidden layers and then adjusting the weights through backpropagation iteratively to minimize a loss function. Bayesian Network represents the probability relationship and conditional dependencies between the dependent variable and input variables via a directed a cyclic graph.

3. DATA

The bankruptcy indicator, public records and firmographics information of 11,787,287 U.S. companies in the 4th Quarter of 2012and 2013 was collected by a national credit reporting agency, and were approved for use in this study. From the data, a bankruptcy flag indicates whether a corporate is in bankruptcy or in business at the capture time point. Firmographics in the data include industry, location, size, and status and structure. Each corporate is identified by its unique Market Participant Identifier (MPID).Public Records include judgements and liens reported.

From the dataset provided, we aim to answer the following question explicitly, which can provide decision makers with insights into improved bankruptcy prediction.

Given the public records and firmographics indicators of an organization in one quarter, can we predict its operation status one year in the future?

To answer the question above, the dependent variable Bankruptcy Indicator Change (i.e. BrtIndChg) was created and is provided in Table 1. Originally, Bankruptcy Indicator (i.e.BrtInd) has two levels, 0 and 1, where 0 indicates that the organization is operating and 1 indicates a bankruptcy. If an organization in business in 2012 went to bankruptcy in 2013, then BrtIndChg was assigned to 1. If the organization was still in business in 2013, then BrtIndChg was assigned to 0.

The raw data had to be cleaned and transformed prior to modeling, to address missing values, abnormal/incorrect values, and correlated variables. The following steps were applied to the data.

(1) Only keep observations with the level value 0 in the original 2012 BrtInd.

(2) Create the dependent variable BrtIndChg by comparingBrtInd in the dataset of 2012 and 2013 as shown in Table 1.

(3) Drop interval variables if the percentage of coded values or missing values is greater than 30%. A value of 30% was selected to optimize the percent of variance explained in the dataset.

(4) Drop observations in an interval variable or a categorical variable if the percentage of the abnormal/incorrect values in that variable is less than 5%.

(5) Continuous variables were binned into nominal variables. For example, the variable Number of Current Liens or Judgment was binned into Current Liens or Judgment Indicator (i.e. curLiensJudInd) with two levels, 0and 1, where 0 means an organization does not have a lien or judgment currently and 1 means an organization has one or more liens or judgments currently.

(6) Retain the variable with the best predictive ability among several correlated variables. For example, based on both the variable definition and the Chi-Square value, the following variables are correlated: Current Liens/Judgment Indicator, Number of Current Liens/Judgment and Total Current Dollar Amounts on All Liens/Judgments. After comparing their performance, only the variable Current Liens/Judgment Indicator was kept.

Table 1. Creation of Dependent Variables

BrtInd 2012	BrtInd 2013	BrtIndChg
0	1	1
0	0	0

After the data was cleaned, the variables in Table 2 were prepared for further analysis and modeling. As described above, the bankruptcy is a rare event, which can be further confirmed by the distribution of the dependent variable BrtIndChg, as shown in Table 3. In our dataset, there are 0.12% of observations going into bankruptcy from 2012 to 2013 and 99.88% of observations staying in business from 2012 to 2013. Because the proportion of event cases is much less than the proportion of nonevent cases, we need to consider oversampling to have sufficient event cases to train the model and achieve better performance, which will be discussed in detail in Section 5.

Table 2. Variables for Analysis and Modelling.

Variable	Type	Description
MPID	Nominal	Market Participant Identifier
BrtIndChg	Binary	Bankruptcy Indicator Change
curLiensJudInd	Nominal	Current Liens/Judgment Indicator
histLiensJudInd	Nominal	Historical Liens/Judgment Indicator
Industry	Nominal	Industry
LargeBusinessInd	Nominal	Large Business Indicator
Region	Nominal	Geographical Region
PublicCompanyFlag	Nominal	Public Company Flag
SubsidiaryInd	Nominal	Subsidiary Indicator
MonLstRptDatePlcRec	Interval	Number of Months Since Last Report Date on Public Records

Table 3. Frequency of Dependent Variable.

BrtIndChg	Frequency	Percent(%)
1	1031	0.12
0	843330	99.88

4. EXPLORATORY ANALYSIS

To examine the statistical association and significance between each individual input variable and the dependent variable, bivariate analysis was performed. The results of odds ratio and Chi-square test can be found in Table 4. Based on the Chi-Square results, all the variables are significantly associated with the dependent variable except the variable Public Company Flag. Based on the odds ratio, we have the following observations regarding their relationship:

- Current Lien/Judgment Indicator: The organizations which currently do not have any lien/judgment is about 47.1% less likely to go into bankruptcy in the following year than those which currently have liens or judgments.
- Historical Lien/Judgment Indicator: The organizations which did not have any lien/judgment is about 32% less likely to go into bankruptcy in the following year than the ones which historically had liens or judgments.
- Large Business Indicator: The organizations which are not large are about 45.8% less likely to go into bankruptcy in the following year than the ones which are large.
- Subsidiary Indicator: The organizations which are not subsidiaries are 74.5% more likely to go into bankruptcy in the following year than those organizations which are subsidiaries.

- Industry: By using the industry group 8 as the reference level, the organizations in the industry group 3 is about2 times more likely going to the bankruptcy in the following year than the ones in the industry group 8.
- Region: By using the region group 9 as the reference level, the organizations in the region group 2 are about 55.7% less likely to go into bankruptcy in the following yearthan the ones in the region group 9.
- Number of Months Since Last Report Date on Public Records(i.e. MonLstDatePlcRec): Figure 1 shows that the distribution of MonLstDatePlcRec is very different in different levels of BrtIndChg, indicating their strong relationship.

Table 4. Univariate Odds Ratio and Chi-Square p-value.

Effect	Odds Ratio	95% Confidence Interval	Chi-Squarep-value
curLiensJudInd0 vs 1	0.529	[0.447, 0.627]	<.0001
histLiensJudInd 0 vs 1	0.680	[0.601, 0.768]	<.0001
LargeBusinessIndN vs Y	0.542	[0.474, 0.620]	<.0001
LargeBusinessIndU vs Y	0.202	[0.165, 0.249]	
PublicCompanyFlagN vs Y	0.295	[0.104, 0.838]	0.065
PublicCompanyFlagU vs Y	0.370	[0.138, 0.989]	
SubsidiaryIndN vs Y	1.745	[0.997, 3.053]	<.0001
SubsidiaryIndU vs Y	0.411	[0.261, 0.648]	
Industry1 vs 8	1.538	[0.947, 2.496]	<.0001
Industry2 vs 8	3.085	[1.118, 8.514]	
Industry3 vs 8	2.079	[1.545, 2.797]	
Industry4 vs 8	1.971	[1.365, 2.847]	
Industry5 vs 8	1.648	[1.136, 2.392]	
Industry6 vs 8	2.421	[1.704, 3.439]	
Industry7 vs 8	1.386	[1.033, 1.859]	
Industry9 vs 8	1.348	[1.012, 1.795]	
Industry10 vs 8	0.885	[0.216, 3.629]	
IndustryU vs 8	0.473	[0.343, 0.651]	
Region1 vs 9	0.699	[0.479, 1.019]	<.0001
Region2 vs 9	0.443	[0.358, 0.549]	
Region3 vs 9	0.627	[0.505, 0.779]	
Region4 vs 9	0.913	[0.686, 1.215]	
Region5 vs 9	0.636	[0.525, 0.772]	
Region6 vs 9	1.203	[0.928, 1.558]	
Region 7 vs 9	1.084	[0.875, 1.343]	
Region 8 vs 9	1.194	[0.920, 1.549]	
MonLstRptDatePlcRec	0.971	[0.969, 0.973]	<.0001

5. METHODOLOGY

To better train and evaluate the models, the dataset was first oversampled and then split into training dataset and validation dataset, where the training dataset was used for training the models and the validation dataset was used as the hold-out dataset for evaluating the performance of models. Seven different models were developed, including Logistic Regression, Decision Tree, Random Forest, Gradient Boosting, Support Vector Machine, Bayesian Network, and Neural Network. Their respective performances were then evaluated by overall accuracy, Type I error, Type II error, and ROC curve.
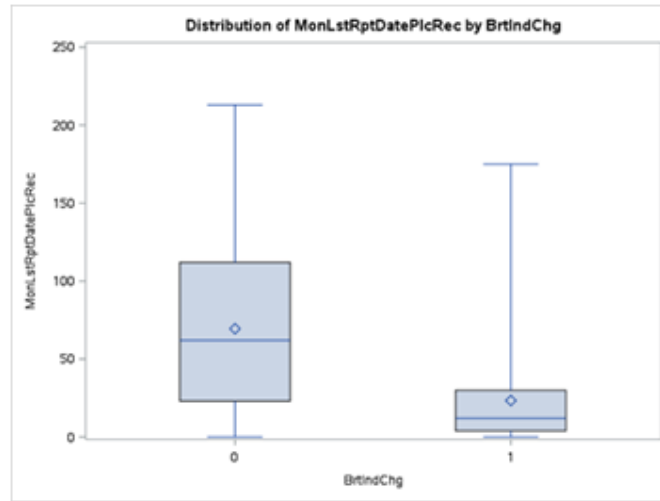

Figure 1. Boxplot of MonLstRptDatePlcRec by BrtIndChg

5.1. Sampling

Sampling was done in three steps.

(1) Oversampling: The proportion of the events is 0.12%, as indicated in Table 3. To have sufficient event cases to train the model and achieve better performance, the oversampling technique is used to adjust the proportion of event observations and non-event observations to 50% versus 50%, which include all the event observations and an equal number of randomly selected non-event observations. That ends up with1031 bankruptcy observations and 1031 non-bankruptcy observations.

(2) Training Dataset and Validation Dataset Split: The out-of-sample test is used by evaluating the models on the hold-out data set. Hence the dataset is split into training and validation by70% versus 30%, respectively.

(3) Oversampling Adjustment: Prior probability and inverse prior weights are applied to the results to adjust oversampling.

5.2. Model Development and Evaluation

The models were developed using SAS Enterprise Miner. All variables in Table 4 are specified as initial inputs for all models. Every model is tuned to their best performance by trying different hyper parameter values.

In Logistic Regression, backwards selection is used to select significant variables with the significance level set to 0.05. The multi variate odds ratio and Chi-Square p-value of the resulting model can be found in Table 5. The significant variables include curLiensJudInd,histLiensJudInd,

Large Business Ind, Region, and MonLstDatePlcRec.Their multivariate odds ratio is consistent with their univariate oddsratio. For example, univariate odds ratio shows that curLiensJudIndis negatively associated with the dependent variable, which is the same as indicated by the multivariate odds ratio of curLiensJudInd.

Decision Tree, Gradient Boosting, and Random Forest are all tree-based models. Entropy is used as the criteria of searching and evaluating candidate splitting rules for Decision Tree, while Gini index is used for Gradient Boosting and Random Forest. The important variables selected by these models include MonLstDatePlcRec,Region, Industry, curLiensJudInd, histLiensJudInd, and LargeBusinessInd. Their importance measure can be found in Table 6. Note that for Decision Tree and Gradient Boosting, the importance measure presented here is the total Entropy or Gini reduction, while for Random Forest, the importance measure is the marginal Gini reduction.

Table 5. Multivariate Odds Ratio and Chi-Square p-value.

Effect	Odds Ratio	Chi-Squarep-value
curLiensJudInd0 vs 1	0.573	0.0046
histLiensJudInd0 vs 1	0.508	<.0001
LargeBusinessIndN vs Y	0.796	<.0001
LargeBusinessIndU vs Y	0.332	
Region1 vs 9	1.067	0.0002
Region2 vs 9	0.411	
Region3 vs 9	0.583	
Region4 vs 9	0.839	
Region5 vs 9	0.558	
Region6 vs 9	0.858	
Region 7 vs 9	0.881	
Region 8 vs 9	1.261	
MonLstRptDatePlcRec	0.976	<.0001

Table 6. Variable Importance.

Variable	Decision Tree	Gradient Boosting	Random Forest
MonLstRptDatePlcRec	1.0000	1.0000	0.0911
Region	0.2423	0.2880	0.0048
Industry	0.1663	0.3516	0.0110
curLiensJudInd	0.1550	0.0820	0.0024
histLiensJudInd	0.1192	0.1205	0.0038
LargeBusinessInd	0.0308	0.2752	0.0100

In Support Vector Machine, linear kernel function performs better than polynomial kernel function. In Neural Network, Tanh is used as the activation function in the hidden layer while Sigmoid is used in the output layer. Its architecture can be found in Figure 2.In Bayesian Network, the significant variables selected by G-Square with the significance level 0.2 include MonLstDatePlcRec, Region, Industry, curLiensJudInd, histLiensJudInd, LargeBusinessInd, and SubsidiaryInd. The resulting Bayesian Network can be found in Figure 3.
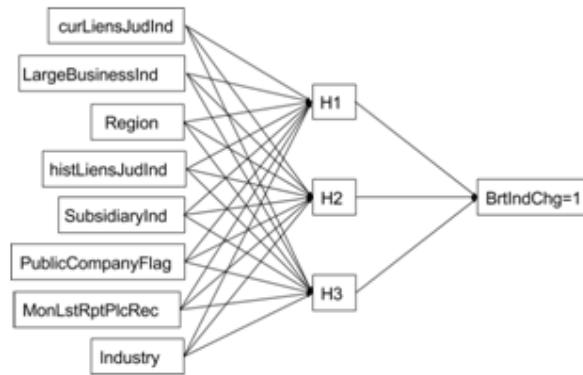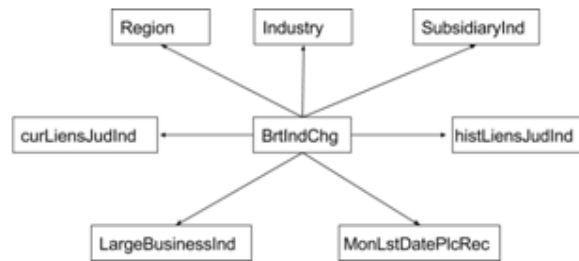
Figure 2. Neural Network Architecture

Figure 3. Bayesian Network Diagram

The accuracy, Type I error, and Type II error of all models are summarized in Table 7. Overall speaking, there is always a tradeoff between Type I error and Type II error, where a large Type I error causes less profits by classifying organizations with low bankruptcy risk into high risky ones and a large Type II error brings more losses by classifying organizations with high bankruptcy risk into low risky ones. For example, according to Type II error, Support Vector Machine performs the best, but it gives the worst overall accuracy and Type I error in the meantime. Practitioners are suggested to select the model by making a balance among model characteristics (accuracy, Type I error, Type II error, interpretability,etc.) based on their expectations. For example, Neural Network, Bayesian Network, and Logistic Regression give the same Type IIError in this case, Bayesian Network and Logistic Regression may be favored than Neural Network because of their high interpretability.

Table 7. Performance of Models.

Model	Accuracy	Type I Error	Type II Error
Support Vector Machine	73.39%	40.32%	12.90%
Decision Tree	75.16%	36.45%	13.87%
Gradient Boosting	74.51%	31.29%	17.42%
Random Forest	75.80%	29.35%	19.03%
Neural Network	74.52%	29.35%	19.35%
Bayesian Network	74.35%	31.93%	19.35%
Logistic Regression	74.35%	31.61%	19.35%

To more comprehensively compare these models, ROC curves onboth the training and validation dataset are provided in Figure 4and Figure 5, respectively. For all models, there is no large

difference between training ROC and validation ROC, so there is no overfitting. Moreover, all models overall perform similar, because there is no large gap among their ROCs.

5.3. Experiment without Oversampling

To show the influence of oversampling in the rare event prediction, the dataset and models were fitted without oversampling. Theresulting performance measures on the validation dataset can be found in Table 8. All models, except Bayesian Network, classify all bankruptcy observations to non-bankruptcy, as indicated by Type II error, although they have high overall accuracy 99.88% which is exactly the proportion of non-bankruptcy observations in the original data. For Bayesian Network without, its overall accuracy and Type II error decreased by 9.92% and 2.03%, respectively, while its Type I error increased by 3.66%. The ROC curve in Figure 6 further shows that all other models perform no better than random selection, except Bayesian Network.
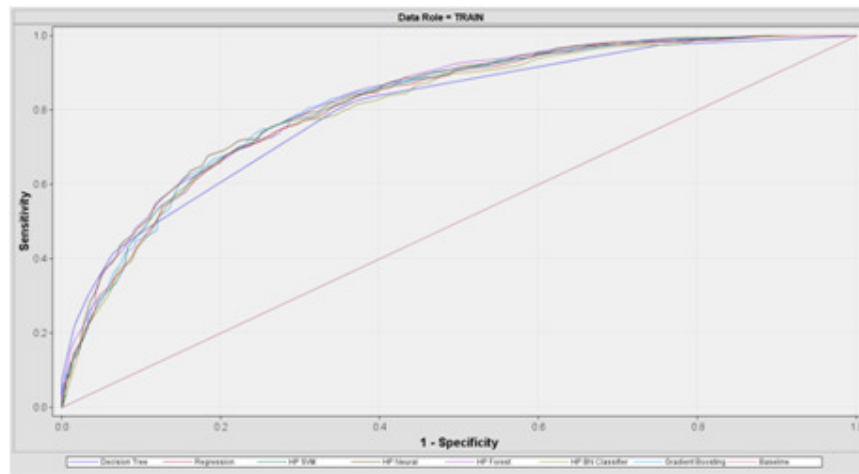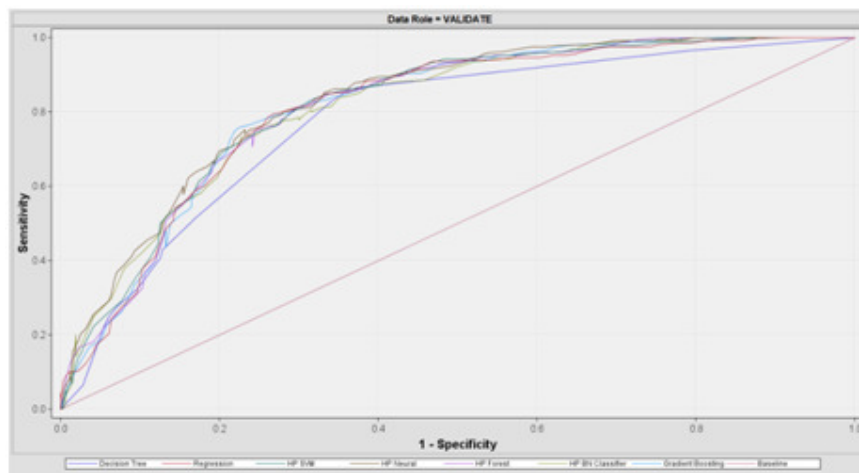

Figure 4. ROC Curve on Training Dataset

Figure 5. ROC Curve on Validation Dataset

Table 8. Performance of Models without Oversampling.

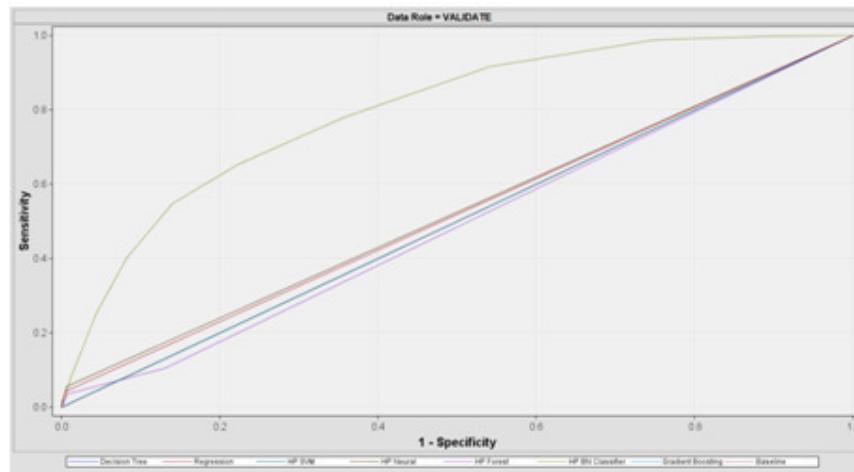| Model	Accuracy	Type I Error	Type II Error
Support Vector Machine	99.88%	0%	100%
Decision Tree	99.88%	0%	100%
Gradient Boosting	99.88%	0%	100%
Random Forest	99.88%	0%	100%
Neural Network	99.88%	0%	100%
Bayesian Network	64.43%	35.59%	17.32%
Logistic Regression	99.88%	0%	100%

Figure 6. ROC Curve on Validation Dataset without Oversampling

6. DISCUSSIONS AND CONCLUSIONS

Based on the univariate analysis and multivariate analysis, the impacts of public records and firmographics indicators were comprehensively studied. With them as input variables of different classification models, the model results show that public records and firmographics indicators play an important role in the bankruptcy prediction. This may serve as a reference for practitioners and researchers to include these information in the bankruptcy prediction model.

Different classification models generate quite different Type I/II error, although their overall accuracy is similar. Support Vector Machine gives the lowest Type II error, and Logistic Regression gives the lowest Type I error. Regarding the interpretability, Logistic Regression, Decision Tree and Bayesian Network might be favorable choices. We also find that on the dataset with small/medium size, simple models may outperform complicated models like ensemble models and Neural Network. Practitioners may handle the tradeoff between Type I error and Type II error as well as the model interpretability and accuracy based on their expectations.

For rare event prediction, the oversampling is necessary before modeling to achieve better performance. Bayesian Network is quite robust for rare event prediction without oversampling, compared to other classification models.

7. FUTURE WORK

In this study, we only focused on the public records and firmographics indicators. In the future, we may collect other information like financial ratios to further reduce Type I/II error, and test the model performance in a wider time spanning.

REFERENCES

[1]   Ahn, B. S., S. S. Cho, and C. Y. Kim. "The integrated methodology of rough set theory and artificial neural network for business failure prediction." Expert systems with applications 18.2 (2000): 65-74.

[2]   Altman, Edward I. "Financial ratios, discriminant analysis and the prediction of corporate bankruptcy." The journal of finance23.4 (1968): 589-609.

[3]   Altman, Edward I., and Paul Narayanan. "An international survey of business failure classification models." Financial Markets, Institutions & Instruments 6.2 (1997): 1-57.

[4]   Atiya, Amir F. "Bankruptcy prediction for credit risk using neural networks: A survey and new results." IEEE Transactions on neural networks 12.4 (2001): 929-935.

[5]   Bellovary, Jodi L., Don E. Giacomino, and Michael D. Akers. "A review of bankruptcy prediction studies: 1930 to present." Journal of Financial education (2007): 1-42.

[6]   Chava, Sudheer, and Robert A. Jarrow. "Bankruptcy prediction with industry effects." Financial Derivatives Pricing: Selected Works of Robert Jarrow. 2008. 517-549.

[7]   Everett, Jim, and John Watson. "Small business failure and external risk factors." Small Business Economics 11.4 (1998): 371-390.

[8]   Foroghi, Daryush, and Amirhassan Monadjemi. "Applying decision tree to predict bankruptcy." Computer Science and Automation Engineering (CSAE), 2011 IEEE International Conference on. Vol. 4. IEEE, 2011.

[9]   Huang, Shi-Ming, et al. "A hybrid financial analysis model for business failure prediction." Expert Systems with Applications35.3 (2008): 1034-1040.

[10]   James, Gareth, et al. An introduction to statistical learning. Vol. 112. New York: springer, 2013.

[11]   Kumar, P. Ravi, and Vadlamani Ravi. "Bankruptcy prediction in banks and firms via statistical and intelligent techniques–A review." European journal of operational research 180.1 (2007): 1-28.

[12]   Min, Jae H., and Young-Chan Lee. "Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters." Expert systems with applications 28.4 (2005): 603-614.

[13]   Odom, Marcus D., and Ramesh Sharda. "A neural network model for bankruptcy prediction." Neural Networks, 1990., 1990 IJCNN International Joint Conference on. IEEE, 1990.

[14]   Shin, Kyung-Shik, Taik Soo Lee, and Hyun-jung Kim. "An application of support vector machines in bankruptcy prediction model." Expert Systems with Applications 28.1 (2005): 127-135.

[15]   Sun, Lili, and Prakash P. Shenoy. "Using Bayesian networks for bankruptcy prediction: Some methodological issues." European Journal of Operational Research 180.2 (2007): 738-753.

[16]   Tan, Pang-Ning. Introduction to data mining. Pearson Education India, 2006.

[17] Zhang, Guoqiang, et al. "Artificial neural networks in bankruptcy prediction: General framework and cross-validation analysis." European journal of operational research 116.1 (1999): 16-32.

[18] Zięba, Maciej, Sebastian K. Tomczak, and Jakub M. Tomczak. "Ensemble boosted trees with synthetic features generation in application to bankruptcy prediction." Expert Systems with Applications 58 (2016): 93-101.

[19] Kim, Myoung-Jong, Dae-Ki Kang, and Hong Bae Kim. "Geometric mean based boosting algorithm with over-sampling to resolve data imbalance problem for bankruptcy prediction." Expert Systems with Applications 42.3 (2015): 1074-1082.

[20] Zhou, Ligang. "Performance of corporate bankruptcy prediction models on imbalanced dataset: The effect of sampling methods." Knowledge-Based Systems 41 (2013): 16-25.

AUTHORS

Lili Zhang is currently a Ph.D. candidate in Analytics and Data Science at Kennesaw State University, working on the data mining, machine learning, and data management.

INTENTIONAL BLANK

Predicting Players' Performance in One Day International Cricket Matches Using Machine Learning

Kalpdrum Passi and Niravkumar Pandey

Department of Mathematics and Computer Science,
Laurentian University,
Sudbury, Canada

ABSTRACT

Player selection is one the most important tasks for any sport and cricket is no exception. The performance of the players depends on various factors such as the opposition team, the venue, his current form etc. The team management, the coach and the captain select 11 players for each match from a squad of 15 to 20 players. They analyze different characteristics and the statistics of the players to select the best playing 11 for each match. Each batsman contributes by scoring maximum runs possible and each bowler contributes by taking maximum wickets and conceding minimum runs. This paper attempts to predict the performance of players as how many runs will each batsman score and how many wickets will each bowler take for both the teams. Both the problems are targeted as classification problems where number of runs and number of wickets are classified in different ranges. We used naïve bayes, random forest, multiclass SVM and decision tree classifiers to generate the prediction models for both the problems. Random Forest classifier was found to be the most accurate for both the problems.

KEYWORDS

Naïve Bayes, Random Forest, Multiclass SVM, Decision Trees, Cricket

1. INTRODUCTION

Cricket is a sport played by two teams with each side having eleven players. Each team is a right blend of batsmen, bowlers and all rounders. The batsmen's role is to score maximum runs possible and the bowlers have to take maximum wickets and restrict the other team from scoring runs at the same time. All rounders are the players who can both bat and bowl and they contribute by scoring runs and taking wickets. Each player contributes towards the overall performance of the team by giving his best performance in each match. Each player's performance varies with factors like the team he is playing against and the ground at which the match is being played. It is important to select the right players that can perform the best in each match. The performance of a player also depends on several factors like his current form, his performance against a particular team, his performance at a particular venue etc. The team management, the coach and the captain analyze each player's characteristics, abilities and past stats to select the best playing XI for a given match. In other words, they try to predict the players' performance for each match.

In this paper, we predict the players' performance in One Day International (ODI) matches by analyzing their characteristics and stats using supervised machine learning techniques. For this, we predict batsmen's and bowlers' performance separately as how many runs will a batsman score and how many wickets will a bowler take in a particular match.

2. LITERATURE REVIEW

An extensive online search produced very few articles related to players' performance prediction in the game of cricket. A very small number of researchers have studied the performance of cricket players. Muthuswamy and Lam [1] predicted the performance of Indian bowlers against seven international teams against which the Indian cricket team plays most frequently. They used back propagation network and radial basis network function to predict how many runs a bowler is likely to concede and how many wickets a bowler is likely to take in a given ODI match. Wikramasinghe [2] predicted the performance of batsmen in a test series using a hierarchical linear model. Barr and Kantor [3] defined a criterion for comparing and selecting batsmen in limited overs cricket. They defined a new measure P(out) i.e. probability of getting out and used a two-dimensional graphical representation with Strike Rate on one axis and P(out) on another. Then they define a selection criterion based on P(out), strike rate and batting average of the batsmen. Iyer and Sharda [4] used neural networks to predict the performance of players where they classify batsmen and bowlers separately in three categories – performer, moderate and failure. Based on the number of times a player has received different ratings, they recommend if the player should be included in the team to play World Cup 2007. Jhanwar and Paudi [5] predict the outcome of a cricket match by comparing the strengths of the two teams. For this, they measured the performances of individual players of each team. They developed algorithms to model the performances of batsmen and bowlers where they determine the potential of a player by examining his career performance and then his recent performances. Lemmer [6] defined a new measure called Combined Bowling Rate to measure the performance of bowlers. The combined bowling rate is a combination of three traditional bowling measures: bowling average, strike rate and economy. Bhattacharjee and Pahinkar. [7] used this combined bowling rate to analyze the performance of bowlers in Indian Premier League (IPL). They also determined other factors that affect the performance of bowlers and applied multiple regression model to identify the factors that are empirically responsible for the performance of bowlers. Mukharjee. [8] applied Social Network Analysis to rate batsmen and bowlers in a team performance. He generated a directed and weighted network of batsmen-bowlers using player-vs-player information available for test and ODI cricket. He also generated a network of batsmen and bowlers using the dismissal record of batsmen in the history of cricket. Shah [9] also defined new measures to measure players' performance. The new measure for batsmen takes into account the quality of each bowler he is facing and the new measure for bowlers considers the quality of each batsman he is bowling to. The aggregate of individual performance of a batsman against each bowler is the total performance index of the batsman. Similarly, the aggregate of individual performance of a bowler against each batsman is the total performance index of the bowler. Parker, Burns and Natarajan. [10] defined a model for valuation of players for IPL auction. Their model considered factors like previous bidding price of the player, experience of the player, strike rate etc. Prakash, Patvardhan. and Lakshmi [11] defined batting index and bowling index to rank players' performance for their models to predict outcomes of IPL matches. Ovens and Bukiet [12] applied a mathematical approach to suggest optimal batting orders for ODI matches.

Our work is probably the first generalized approach to predict how many runs will a batsman score and how many wickets will a player take on a particular match day. Muthuswamyand Lam [1] carried out a similar study predicting how many wickets will a bowler take using neural networks but their work was limited to eight Indian bowlers and is difficult to generalize for all the bowlers in the world. We used some supervised machine learning algorithms to build prediction models that can be used to predict the performance of any player in a given match.

3. DATA AND TOOLS

We obtained all our data from www.cricinfo.com using scraping tools, parsehub [13] and import.io [14]. For batting, we considered matches played from January 14, 2005 to July 10, 2017. The senior most player during this span was SR Tendulkar, so we collected innings by innings list of the performance of all the batsmen from December 18, 1989 when he played his first ODI match. For bowling, we considered matches played from January 2, 2000 to July 10, 2017. The senior most player during this span was PA de Silva, so we collected innings by innings list of the performance of all the batsmen from March 31, 1984 when he played his first ODI match. Since the past stats of the players such as average, strike rate etc. are not available directly online for each match they played, we calculated from the innings by innings list for each match. We imported all the data in MySQL tables and used php to manipulate them.

For predictive analytics, we used Weka and Dataiku. Both these tools are a collection of machine learning algorithms for data mining and also provide some preprocessing functionalities. All the results in this study have been obtained from Weka 3-9-1-oracle-jvm and Dataiku Data Science Studio on Mac OS 10.11.6 and Windows 10.

4. DATA PREPROCESSING

4.1 Calculating the Traditional Attributes

As mentioned in the previous section, the stats of the players such as average, strike rate etc. are not available directly for each game, we calculated these attributes from the innings by innings list using aggregate functions and mathematical formulae. These attributes are generally used to measure a player's performance. These attributes are as follows:

4.1.1 Batting Attributes

No. of Innings: The number of innings in which the batsman has batted till the day of the match. This attribute signifies the experience of the batsman. The more innings the batsman has played, the more experienced the player is.

Batting Average: Batting average commonly referred to as average is the average number of runs scored per innings. This attribute indicates the run scoring capability of the player.

$$Average = Runs\ Scored\ /\ Number\ of\ times\ dismissed$$

Strike Rate (SR): Strike rate is the average number of runs scored per 100 balls faced. This attribute indicates how quickly the batsman can score runs.

$$Strike\ Rate:\ (Runs\ Scored\ /\ Balls\ Faced)\ *\ 100$$

Centuries: Number of innings in which the batsman scored more than 100 runs.

Fifties: Number of innings in which the batsman scored more than 50 (and less than 100) runs.

Zeros: Number of innings in which the batsman was dismissed without scoring a single run.

Highest Score: The highest runs scored by a batsman in any (single) innings throughout his career.

4.1.2 Bowling Attributes

No. of Innings: The number of innings in which the bowler has bowled till the day of the match.

Overs: The number of overs bowled by a bowler.

Bowling Average: Bowling average is the number of runs conceded by a bowler per wicket taken.

Average: Number of runs conceded / Number of wickets taken

Bowling Strike Rate: Bowling strike rate is the number of balls bowled per wicket taken.

Strike Rate: Number of balls bowled / Number of wickets taken

Four/Five Wicket Haul: Number of innings in which the bowler has taken more than four wickets.

4.2 Rating The Attributes

Each traditional attribute is given a rating from 1 to 5 based on its value with 1 being the minimum and 5 being the maximum.

These ratings are as follows:

No. of Innings:

For Consistency:
   1 – 49: 1
   50 – 99: 2
   100 – 124: 3
   125 – 149: 4
   >=150: 5

For Form:
   1 – 4: 1
   5 – 9: 2
   10 – 11: 3
   12 – 14: 4
   >=15: 5

For Opposition:
   1 – 2: 1
   3 – 4: 2
   5 – 6: 3
   7 – 9: 4
   >=10: 5

For Venue:
   1: 1
   2: 2
   3: 3
   4: 4
   >=5: 5

Batting Average (for all derived attributes):

    0.0   - 9.99: 1
    10.00 - 19.99: 2
    20.00 - 29.99: 3
    30.00 - 39.99: 4
    >=40: 5

Batting Strike Rate (for all derived attributes):

    0.0   - 49.99: 1
    50.00 - 59.99: 2
    60.00 - 79.99: 3
    80.00 - 100.00: 4
    >=100.00: 5

Centuries:

    For Consistency:
        1 – 4: 1
        5 – 9: 2
        10 – 14: 3
        15 – 19: 4
        >=20: 5

    For Form:
        1: 1
        2: 2
        3: 3
        4: 4
        >=5: 5

    For Opposition:
        1: 3
        2: 4
        >=3: 5

    For Venue:
        1: 4
        >=2: 5

Fifties:

    For Consistency:
        1 – 9: 1
        10 – 19: 2
        20 – 29: 3
        30 – 39: 4
        >=40: 5

    For Form & Opposition:
        1 – 2: 1
        3 – 4: 2
        5 – 6: 3
        7 – 9: 4

>=10: 5

For Venue:
1: 4
>=2:– 5

Zeros:

For Consistency:
1 – 4: 1
5 – 9: 2
10 – 14: 3
15 – 19: 4
>=20: 5

For Form & Opposition:
1: 1
2: 2
3: 3
4: 4
>=5: 5

Highest Score (For Venue Only):
1 – 24: – 1
25 – 49: 2
50 – 99: 3
100 – 150: 4
>=150: 5

Overs:

For Consistency:
1 – 99: 1
100 – 249: 2
250 – 499: 3
500 – 1000: 4
>=1000: 5

For Form & Opposition:
1 – 9: 1
10 – 24: 2
25 – 49: 3
50 – 100: 4
>=100: 5

For Venue:
1 – 9: 1
10 – 19: 2
20 – 29: 3
30 – 39: 4
>=40: 5

Bowling Average (for all derived attributes):
    0.00 - 24.99: 5
    25.00 - 29.99: 4
    30.00 - 34.99: 3
    35.00 - 49.99: 2
    >=50.00: 1

Bowling Strike Rate (for all derived attributes):
    0.00 - 29.99: 5
    30.00 -39.99: 4
    40.00 -49.99: 3
    50.00 -59.99: 2
    >=60.00: 1

Four/Five Wicket Haul:

For Consistency:
    1 – 2: 3
    3 – 4: 4
    >=5: 5

For Form, Opposition & Venue:
    1 – 2: 4
    >=3: 5

4.3 Calculating The Derived Attributes

We have derived four attributes for batsmen and bowlers to reflect their abilities and characteristics. These four attributes have been derived by taking a weighted average of the traditional attributes described above. The weights were determined using analytic hierarchy process developed by Thomas L. Saaty [15] [16]. These attributes are described below:

4.3.1 Consistency

This attribute describes how experienced the player is and how consistent he has been throughout his career. All the traditional attributes used in this formula are calculated over the entire career of the player.

Formula for batting:

$$Consistency = 0.4262*average + 0.2566*no. \ of \ innings + 0.1510*SR + 0.0787*Centuries + 0.0556*Fifties - 0.0328*Zeros$$

Formula for bowling:

$$Consistency = 0.4174*no. \ of \ overs + 0.2634*no. \ of \ innings + 0.1602*SR + 0.0975*average + 0.0615*FF$$

4.3.2 Form

Form of a player describes his performance over last one year. All the traditional attributes used in this formula are calculated over the matches played by the player in last 12 months from the day of the match.

Formula for batting:

$$Form = 0.4262*average + 0.2566*no. \ of \ innings + 0.1510*SR + 0.0787*Centuries + 0.0556*Fifties – 0.0328*Zeros$$

Formula for bowling:

$$Form = 0.3269*no. \ of \ overs + 0.2846*no. \ of \ innings + 0.1877*SR + 0.1210*average + 0.0798*FF$$

4.3.3 Opposition

Opposition describes a player's performance against a particular team. All the traditional attributes used in this formula are calculated over all the matches played by the player against the opposition team in his entire career till the day of the match.

Formula for batting:

$$Opposition = 0.4262*average + 0.2566*no. \ of \ innings + 0.1510*SR + 0.0787*Centuries + 0.0556*Fifties – 0.0328*Zeros$$

Formula for bowling:

$$Opposition = 0.3177*no. \ of \ overs + 0.3177*no. \ of \ innings + 0.1933*SR + 0.1465*average + 0.0943*FF$$

4.3.4 Venue

Venue describes a player's performance at a particular venue. All the traditional attributes used in this formula are calculated over all the matches played by the player at the venue in his entire career till the day of the match.

Formula for batting:

$$Venue = 0.4262*average + 0.2566*no. \ of \ innings + 0.1510*SR + 0.0787*Centuries + 0.0556*Fifties + 0.0328*HS$$

Formula for bowling:

$$Venue = 0.3018*no. \ of \ overs + 0.2783*no. \ of \ innings + 0.1836*SR + 0.1391*average + 0.0972*FF$$

4.3.5 Other Input Attributes

Batting Hand: The dominant hand of the batsman while batting. It has two possible values: Left or Right.

Bowling Hand: The dominant hand of the batsman while bowling. It has two possible values: Left or Right.

Batting Position: The number at which the batsman bats in the batting order.

Match Type: The type of the match. This attribute has four possible values: Normal, quarter-final, semi-final or final.

Match Time: The time at which the match is played. There are two possible values: Day or Day-night.

Strength of opposition: This is the batting/bowling strength of the opposition team. It is the average of the consistency measure of the batsmen/bowlers of the opposition team.

Ven: The relative venue for the teams. It has three possible values: Home, Away or Neutra l.
Oppo: The opposition team.

The following attributes are used only for predicting runs.

Role: The playing role of the player. It can take following values:

Opening Batsman (OBT) – The two batsmen who usually bat at position one or two are called opening batsmen.

Top Order Batsman (TOB) – The batsmen who usually bat at position three or five are called top order batsmen.

Middle Order Batsman (MOB) - The batsmen who ususally bat at position five to eight are called middle order batsmen.

Batsman – The batsmen who usually bat at different positions are categorised simply as batsmen here.

Allrounder – The players who are equally skilled at both batting and bowling are called allrounders.

Batting Allrounder – The players who can both bat and bowl but are more skilled at batting than bowling, are called batting allrounders.

Bowling Allrounder – The players who can both bat and bowl but are more skilled at bowling than batting, are called bowling allrounders.

Bowler – The players who are expert bowlers but not so skilled at batting, are categorised as bowlers.

Captain: This is a binary attribute indicating whether a player is captain of the team.

WK: This is a binary attribute indicating whether a player is a wicketkeeper.

Innings: This attributes indicates if it is the first or the second innings of the match.

Tournament: The type of tournament in which the match is being played. It can take the following values:

Two Team Tournament (TT)

Three-Four Team Tournament (TFT)

Five Team Tournament (FT)

Toss: Indicates whether the player's team won or lost the toss.

Pressure: Indicates mental and psychological pressure on the player. It takes values from 1 to 5. Its value depends on the type of match being played and the teams that are playing the match. The values are defined as follows:

Normal matches: 1

Quarter Finals: 3

Semi Finals: 4

Finals: 5

Above values are incremented by 1 if the match is between India and Pakistan or Australia and England as these countries are strong rivals of each other.

Host: The country in which the match is being played.

Ground: The ground on which the match is being played.

4.3.6 Outputs

Both the problems are treated as classification problems.

Runs are predicted in five classes:

1 – 24: 1
25 – 49: 2
50 – 74: 3
75 – 99: 4
>=100: 5

Wickets are predicted in three classes:

0 - 1: 1
2 – 3: 2
>=4: 3

4.3.7 Data Cleaning

A large number of values of Opposition and Venue were zero. This is because a player has not played any match against a particular team or at a venue before the day of play. We treated such values as missing values and replaced them with the class average of corresponding attributes.

4.3.8 Class Imbalance

We observed that majority of the records fall within class 1 in both batting and bowling. This created a major imbalance in the distribution of values and affected the performance of the learning algorithms. To solve this problem, we applied an oversampling technique Supervised Minority Oversampling Technique (SMOTE) [17] on minority classes to make all the classes equally distributed. SMOTE over-samples minority classes by creating synthetic example tuples [17].

5. LEARNING ALGORITHMS

For generating the prediction models, we used supervised machine learning algorithms. In supervised learning algorithms, each training tuple is labeled with the class to which it belongs [18]. We used naïve bayes, decision trees, random forest and multiclass support vector machines for our experiments. These algorithms are explained in brief.

5.1 Naïve Bayes

Bayesian classifiers are statistical classifiers that predict the probability with which a given tuple belongs to a particular class [18]. Naïve Bayes classifier assumes that each attribute has its own individual effect on the class label, independent of the values of other attributes. This is called class-conditional independence. Bayesian classifiers are based on Bayes' theorem.

Bayes Theorem: Let X be a data tuple and C be a class label. Let X belongs to class C, then

$$P(C|X) = \frac{P(X|C)P(C)}{P(X)}$$

where;

- P(C|X) is the posterior probability of class C given predictor X.
- P(C) is the prior probability of class.
- P(X|C) is the posterior probability of X given the class C.
- P(X) is the prior probability of predictor.

The classifier calculates P(C|X) for every class Ci for a given tuple X. It will then predict that X belongs to the class having the highest posterior probability, conditioned on X. That is X belongs to class Ci if and only if

$$P(Ci|X) > P(Cj|X) \quad \text{for } 1 \le j \le m, j \ne i.$$

5.2 Decision Trees

Decision tree induction is the process of creating decision trees for class-labeled training tuples [18]. A decision tree is basically a tree structure like a flowchart [18]. Each internal node of the tree represents a test on an attribute and each branch is the outcome of the test. Each leaf node is a class label. The first node at the top of the tree is the root node. To classify a given tuple X, the attributes of the tuple are tested against the decision tree starting from the root node to the leaf node which holds the class prediction of the tuple. Ross Quinlan introduced a decision tree algorithm called ID3 in his paper [19]. Later he introduced a successor of ID3 called C4.5 in [20] to overcome some shortcomings such as over-fitting. Unlike ID3, C4.5 can handle both continuous and discrete attributes, training data with missing values and attributes with differing costs. In a basic decision tree induction algorithm, all the training tuples are at the root node at start. The tuples are then partitioned recursively based on selected attributes. The attributes are selected based on an attribute selection method which specifies a heuristic procedure to determine the splitting criterion. The algorithm terminates if all the training tuples belong to the same class or there are no remaining attributes for further partitioning or all training tuples are used. ID3 uses the attribute selection measure called information gain, which is simply the difference of the information needed to classify a tuple and the information needed after the split. These two can be formularized as follows:

Expected information needed to classify a tuple in the training set D

$$Info(D) = -\sum_{i=1}^{m} p_i \, log_2(p_i)$$

where; pi is the nonzero probability that a tuple in D belongs to class Ci.

Information needed after the splitting (to arrive at the exact classification)

$$Info_A(D) = \sum_{j=1}^{v} \frac{|D_j|}{|D|} \times Info(D_j)$$

where A is the attribute on which the tuples are to be partitioned.

Then, information gain

$$Gain(A) = Info(D) - Info_A(D)$$

The attribute with highest information gain is selected as the splitting attribute.

C4.5 uses gain ratio as the attribute selection measure. Gain ratio is an extension to information gain in a sense because it normalizes information gain by using a split information value;

$$SplitInfo_A(D) = -\sum_{j=1}^{v} \frac{|D_j|}{|D|} \times log_2\left(\frac{|D_j|}{|D|}\right)$$

Then,

$$GainRatio(A) = \frac{Gain(A)}{SplitInfo_A(D)}$$

The attribute with the highest gain ratio is selected as the splitting attribute.

5.3 Random Forest

Random Forests is an ensemble method for classification and regression [18]. Random forests are a set of decision trees where each tree is dependent on a random vector sampled independently and with the same distribution of all the trees in the forest [21]. The algorithm generates a number of decision trees creating a forest. Each decision tree is generated by selecting random attributes at each node to determine the split [18]. Tim Kam Ho introduced the first method for random forests using random subspace method in his paper [22]. Later, Breiman Leo extended the algorithm in his paper [21] and this method was official known as Random Forests. The general procedure to generate decision trees for random forests starts with a dataset D of d tuples. To generate k decision trees from the dataset, for each iteration k, a training set Di of d tuples is sampled with replacement from the dataset D. To construct a decision tree classifier, at each node, a small number of attributes from the available attributes are selected randomly as candidates for the split at the node. Then Classification And Regression Trees (CART) [23] method is used to grow the trees. The trees are then grown to maximum size and are not pruned. CART is a non-parametric decision tree induction technique that can generate classification and regression trees. CART recursively selects rules based on variables' values to get the best split. It stops splitting when it detects that no further gain can be made or some pre-determined stopping conditions are met.

5.4 Support Vector Machine

Vladimir Vapnik, Bernhard Boser and Isabell Guyon introduced the concept of support vector machine in their paper [24]. SVMs are highly accurate and less prone to overfitting. SVMs can be used for both numeric prediction and classification. SVM transforms the original data into a higher dimension using a nonlinear mapping. It then searches for a linear optimal hyperplane in this new dimension separating the tuples of one class from another. With an appropriate mapping to a sufficiently high dimension, tuples from two classes can always be separated by a hyperplane. The algorithm finds this hyperplane using support vectors and margins defined by the support vectors. The support vectors found by the algorithm provide a compact description of the learned prediction model. A separating hyperplane can be written as:

$$W \cdot X + b = 0$$

where W is a weight vector, W = {w1, w2, w3,..., wn}, n is the number of attributes and b is a scalar often referred to as a bias. If we input two attributes A1 and A2, training tuples are 2-D, (e.g., X = (x1, x2)), where x1 and x2 are the values of attributes A1 and A2, respectively. Thus, any points above the separating hyperplane belong to Class A1:

$$W \cdot X + b > 0$$

and any points below the separating hyperplane belong to Class A2:

$$W \cdot X + b < 0$$

SVM was originally used for binary classification. However, several multiclass SVM algorithms have also been developed. In Weka, we used LIBSVM package developed by Chih-Chung Chang and Chih-Jen Lin [25]. The package can be downloaded from http://www.csie.ntu.edu.tw/cjlin/libsvm. LIBSVM is an easy to use package to apply multiclass SVM and has gained a wide popularity in machine learning.

6. RESULTS AND DISCUSSIONS

We used different sizes of training and test sets to find the best combination that gives the most accuracy. We used four machine learning algorithms: Naïve Bayes, Decision Trees, Random Forest and Support Vector Machine in our experiments. The results are tabulated below. Table 1 shows the accuracies of the algorithms for predicting runs and table 2 shows the accuracies of the algorithms for predicting wickets.

As we can see, Random Forest builds the most accurate prediction models for both batting and bowling in all the cases. Also, the accuracy of the models increases as we increase the size of the training dataset for all algorithms except in case of Naïve Bayes for batting where the accuracy decreases as we increase the size of the training set. Random Forest predicts runs with the highest accuracy of 90.74% when we use 90% of the dataset for training. Similarly, Random Forest predicts wickets with highest accuracy of 92.25% when we use 90% of the dataset for training. On the other hand, Naïve Bayes predicts runs with the least accuracy of 42.5% when we use 90% of the dataset for training. Naïve Bayes predicts wickets too with the least accuracy of 57.05% when we use 60% of the dataset for training. Decision Trees performs reasonably well with the maximum accuracy of 80.46% and the minimum accuracy of 77.93% for predicting runs. It predicts wickets with the maximum accuracy of 86.5% and the minimum accuracy of 84.4%, which is again reasonably well against the performance of Random Forest. The prediction models of SVM for predicting runs showed the maximum accuracy of 51.45% with 90% training data and the minimum accuracy of 50.54% with 60% training data. Also for wickets, SVM had the maximum accuracy of 68.78% with 90% training data and the minimum accuracy of 67.45% with

60% training data. Thus, surprisingly SVM was beaten by Random Forest and Decision Trees on both the datasets.

Table 1 Predicting runs

Classifier	Accuracy (%)			
	60% train 40% test	70% train 30% test	80% train 20% test	90% train 10% test
Naïve Bayes	43.08	42.95	42.47	42.50
Decision Trees	77.93	79.02	79.38	80.46
Random Forest	89.92	90.27	90.67	90.74
SVM	50.54	50.85	50.88	51.45

Table 2 Predicting wickets

Classifier	Accuracy (%)			
	60% train 40% test	70% train 30% test	80% train 20% test	90% train 10% test
Naïve Bayes	57.05	57.18	57.48	58.12
Decision Trees	84.40	85.12	85.99	86.50
Random Forest	90.68	91.26	91.80	92.25
SVM	67.45	67.53	68.35	68.78

Muthuswamy and Lam [1] achieved an accuracy of 87.10% with backpropagation network (BPN) and 91.43% with radial basis function network (RBFN) table 3. They predicted wickets in two classes: 0 or 1 and 2 or more. Their study was limited only to eight Indian bowlers who played ODI matches for India since year 2000 against seven countries. Our approach works for any player in the world and is also applicable to new players who will be playing for their country in future. Moreover, they considered only three input parameters: An ID assigned manually to a bowler, an ID assigned manually to the opposition team and the number of overs bowled by the bowler. We carried out a detailed study of the stats and characteristics of the players by considering a number of attributes that can potentially impact the performance of a player in any match.

Table 3 Predicting wickets using BPN and RBFN [1]

Model	Accuracy (%)
BPN	87.10
RBFN	91.43

7. CONCLUSION AND FUTURE WORK

Selection of the right players for each match plays a significant role in a team's victory. An accurate prediction of how many runs a batsman is likely to score and how many wickets a bowler is likely to take in a match will help the team management select best players for each match. In this paper, we modeled batting and bowling datasets based on players' stats and characteristics.

Some other features that affect players' performance such as weather or the nature of the wicket could not be included in this study due to unavailability of data. Four multiclass classification algorithms were used and compared. Random Forest turned out to be the most accurate classifier for both the datasets with an accuracy of 90.74% for predicting runs scored by a batsman and 92.25% for predicting wickets taken by a bowler. Results of SVM were surprising as it achieved an accuracy of just 51.45% for predicting runs and 68.78% for predicting wickets.

Similar studies can be carried out for other formats of the game i.e. test cricket and T20 matches. The models for these formats can be shaped to reflect required characteristics of the players; e.g. batsmen need to have patience and ability to play longer innings in test matches whereas score more runs in less overs in T20 matches. Similarly, bowlers need to have stronger wicket taking abilities in test matches and better economy rate i.e. conceding less runs in T20 matches. Moreover, attempts can be made to improve accuracies of the classifiers for ODI matches.

REFERENCES

[1]   S. Muthuswamy and S. S. Lam, "Bowler Performance Prediction for One-day International Cricket Using Neural Networks," in Industrial Engineering Research Conference, 2008.

[2]   I. P. Wickramasinghe, "Predicting the performance of batsmen in test cricket," Journal of Human Sport & Excercise, vol. 9, no. 4, pp. 744-751, May 2014.

[3]   G. D. I. Barr and B. S. Kantor, "A Criterion for Comparing and Selecting Batsmen in Limited Overs Cricket," Operational Research Society, vol. 55, no. 12, pp. 1266-1274, December 2004.

[4]   S. R. Iyer and R. Sharda, "Prediction of athletes performance using neural networks: An application in cricket team selection," Expert Systems with Applications, vol. 36, pp. 5510-5522, April 2009.

[5]   M. G. Jhanwar and V. Pudi, "Predicting the Outcome of ODI Cricket Matches: A Team Composition Based Approach," in European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECMLPKDD 2016 2016), 2016.

[6]   H. H. Lemmer, "The combined bowling rate as a measure of bowling performance in cricket," South African Journal for Research in Sport, Physical Education and Recreation, vol. 24, no. 2, pp. 37-44, January 2002.

[7]   D. Bhattacharjee and D. G. Pahinkar, "Analysis of Performance of Bowlers using Combined Bowling Rate," International Journal of Sports Science and Engineering, vol. 6, no. 3, pp. 1750-9823, 2012.

[8]   S. Mukherjee, "Quantifying individual performance in Cricket - A network analysis of batsmen and bowlers," Physica A: Statistical Mechanics and its Applications, vol. 393, pp. 624-637, 2014.

[9]   P. Shah, "New performance measure in Cricket," ISOR Journal of Sports and Physical Education, vol. 4, no. 3, pp. 28-30, 2017.

[10]   D. Parker, P. Burns and H. Natarajan, "Player valuations in the Indian Premier League," Frontier Economics, vol. 116, October 2008.

[11]   C. D. Prakash, C. Patvardhan and C. V. Lakshmi, "Data Analytics based Deep Mayo Predictor for IPL-9," International Journal of Computer Applications, vol. 152, no. 6, pp. 6-10, October 2016.

[12]   M. Ovens and B. Bukiet, "A Mathematical Modelling Approach to One-Day Cricket Batting Orders," Journal of Sports Science and Medicine, vol. 5, pp. 495-502, 15 December 2006.

[13]   "Free web scraping - Download the most powerful web scraper | ParseHub," parsehub, [Online]. Available: https://www.parsehub.com.

[14] "Import.io | Extract data from the web," Import.io, [Online]. Available: https://www.import.io.

[15] T. L. Saaty, "A scaling method for priorities in a hierarchichal structure," Mathematical Psychology, vol. 15, pp. 234-281, 1977.

[16] T. L. Saaty, The Analytic Hierarchy Process, New York: McGraw-Hill, 1980.

[17] N. V. Chavla, K. W. Bowyer, L. O. Hall and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," Artificial Intelligence Research, vol. 16, p. 321–357, June 2002.

[18] J. Han, M. Kamber and J. Pei, Data Mining: Concepts and Techniques, 3rd Edition ed., Waltham: Elsevier, 2012.

[19] J. R. Quinlan, "Induction of Decision Trees," Machine learning, vol. 1, no. 1, pp. 81-106, 1986.

[20] J. R. Quinlan, C4.5: Programs for Machine Learning, Elsevier, 2015.

[21] L. Breiman, "Random Forests," Machine Learning, vol. 45, no. 1, pp. 5-32, 2001.

[22] T. K. Ho, "The Random Subspace Method for Constructing Decision Forests," IEEE transactions on pattern analysis and machine intelligence, vol. 20, no. 8, pp. 832-844, August 1998.

[23] L. Breiman, J. Friedman, C. J. Stone and R. A. Olshen, Classification and regression trees, CRC Press, 1984.

[24] B. E. Boser, I. M. Guyon and V. N. Vapnik, "A Training Algorithm for Optimal Margin Classifiers," in Fifth Annual Workshop on Computational Learning Theory, Pittsburgh, 1992.

[25] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," ACM Transactions on Intelligent Systems and Technology, vol. 2, no. 3, April 2011.

AUTHORS

Kalpdrum Passi received his Ph.D. in Parallel Numerical Algorithms from Indian Institute of Technology, Delhi, India in 1993. He is an Associate Professor, Department of Mathematics & Computer Science, at Laurentian University, Ontario, Canada. He has published many papers on Parallel Numerical Algorithms in international journals and conferences. He has collaborative work with faculty in Canada and US and the work was tested on the CRAY XMP's and CRAY YMP's. He transitioned his research to web technology, and more recently has been involved in machine learning and data mining applications in bioinformatics, social media and other data science areas. He obtained funding from NSERC and Laurentian University for his research. He is a member of the ACM and IEEE Computer Society.

IMPLEMENTING UHF RFID READER ON SMARTPHONE PLATFORM FOR IOT SENSING

Douglas Lautner, Xiayu Hua, Scott DeBates and Shangping Ren

Department of Computer Science, Illinois Institute of Technology,
Chicago, IL 60616, USA
Device+ Research and Technology, Lenovo Motorola,
Chicago, IL 60654, USA

ABSTRACT

As a core component of the Internet of Things technology (IoT), Radio Frequency Identification (RFID) tagged items will add billions, perhaps trillions, of objects to the Internet. As a result, uses of Ultra High Frequency (UHF) RFID sensing become massive ranging from logistics, retail and healthcare to homes and even entire smart cities. Under this trend, mobile UHF RFID scanners also need to evolve. Consumers will interact with their surroundings via tagged RFID items taking full advantage of the advancing IoT. For mainstream consumer smartphones, unfortunately, UHF RFID connectivity has yet to be fully integrated. The major challenges are: 1) the compatibility of an RFID reader module to the host platform, 2) Radio Frequency (RF) signal coexistence interference between the RFID reader and other sensor/RF technologies, and 3) the unacceptable high current drain caused by RFID active scanning. In this paper, we present a design and implementation of a novel modular UHF RFID scanning subsystem, the UHF RFID reader module, on a Motorola Moto-Z smartphone. This module is fully integrated with an Android 7.0 Operating System (OS) and directly interconnects with the low-level smartphone hardware and software framework. With the new antenna design and the signal spectrum analysis, we guarantee the RF isolation of the Mod with the smartphone's other native wireless components and sensors. Our design and implementation also address the current drain issue and extends the battery life of Moto-Z smartphone up to 30.4 hours with IoT RFID scanning.

KEYWORDS

UHF RFID, Smartphone design, mobile system architecture, mobile sensing

1. INTRODUCTION

The advance of the Internet of Things (IoT) technology introduces new application scenarios to smartphone designers and manufactures. For example, using smartphones to gather information from nearby IoT objects through RFID and providing surrounding status to the user is one of the potentials. Thanks to the massive production of UHF RFID tags (from 8.9 billion tags sold in

2015 to 10.4 billion in 2016 [1]), its price becomes affordable [2], hence transforming common everyday items of a wide variety, such as jeans or brushes, into IoT objects [3]. Likewise, smartphone shipments have increased every year since 2006 with an estimated global market of 2.1 billion units shipped in 2021 [4]. With high-performance computing power and constant Internet connectivity, nowadays smartphones are already capable to associate mobile computing with the IoT [5]. Once equipped with a UHF RFID reader, a smartphone is able to interact with the surrounding objects tagged by UHF RFID, gather and process information from those tags locally or on the internet, then represent feedback to the user via its multimedia user interfaces.

However, to-date there has been no successful widespread deployment of a fully integrated UHF RFID reader on mainstream smartphones. Common UHF RFID reader solutions are either stand-alone devices or are deployed as add-on implementation on smartphones that are not fully integrated with the hardware and software of a smartphone. For example, a UHF RFID reader on microSD which can be inserted into a smartphone through external memory slot was developed in [6]. Another solution utilizes the proprietary audio port to connect a UHF RFID reader with a smartphone [7]. Current integrated solutions, such as Invengo XC-1003 mobile IoT device and Chafon CF-H802 run on older Android platforms, are more industrial and are not widely distributed [8], [9]. Moreover, since hardware and software of these devices are specially designed, they are expensive and their functionality are monotonous. Through the research and prototyping experience in our preliminary works, we have identified three major challenges that causes this absence of RFID reader on a smartphone:

1) No existing interface that enables the UHF RFID reader to establish a direct high-speed connection to the lower level smartphone hardware without significant modifications on the phones hardware framework.

2) Severe signal interferences amongst different RF sensors and communication modules while the RFID reader antenna is active.

3) Unbearable current drain to a consumer smartphone when the UHF RFID reader is up and running.

In this work, we address the three challenges and empower a commercial smartphone with UHF RFID reading capability. Specifically, we design and implement a fully functional consumer UHF RFID Reader Mod (URRM), as illustrated in Fig. 1, based on a mainstream Android-driven smartphone, Motorola Moto-Z, which has been "sold around the world" [10]. By leveraging the advantage of the unique modularization features of Moto-Z, i.e. the support of third-party Mods [11], we are able to utilize the GreyBus protocol to connect our RFID reader module, i.e. the Impinj RS500 RFID module, to the phones lower-level hardware framework without changing the phones native hardware. Comparing to add-on implementations that connect through USB, audio jack or memory card ports [9,10], direct connection to lower-level hardware framework provides with high energy-efficiency, low communication overhead as well as wider applicable system support such as power management or chip-level control and scheduling. To guarantee URRMs RF performance and avoid interference with phones native wireless functionalities, we design an antenna for the URRM which isolates to the phones native wireless signal system. We further utilize the power-efficient sensor fusion core on the URRMs hardware infrastructure to manage the scanning and data processing tasks. We minimize the impact of the high current drain of RFID scanning to the phones battery life by optimizing the control and management system and coupling to an external power source. Our experiments prove that our prototype overcomes

these three challenges and hence becomes practical and ready-to-manufacture. In addition, our design and implementation of the key hardware and software architecture is extensible and scalable for continued research. Other IoT sensors on different mobile platforms, such as WiFi or Bluetooth modules on NuttX-driven Rasperry Pi platforms, can also apply our solutions.

The rest of the paper is organized as follows. We discuss related work in Section II. Section III presents the design and implementation of the portable UHF RFID reader architecture. In Section IV we study the radio performance of the UHF RFID reader Mod with respect to signal isolation, sensing range and sensitivity. We conclude in Section V.

2. RELATED WORK

The fast growing IoT technology and its expanding scope of applications demand for a wide spectrum of both the sensor type and the sensing content [12]. In [13], researchers developed CARISMA (Context-Aware Reflective Middleware System for Mobile Applications), a software framework for IoT contextual sensing and awareness tasks on mobile systems. This framework is generic with respect to mobile operation systems and sensor types, hence it can be implemented on different smartphone platforms and work with different types of sensors. Similar middleware system, the SOCAM (Service Oriented Context-Aware Middleware), was also developed by researchers in [14]. To utilize the advantages of heterogeneous sensing sources in IoT scheme, the e-SENSE [15] mobile sensing system enables ambient intelligence using wireless multi-sensor networks for making IoT oriented context-rich information available to applications and services. e-SENSE combines body sensor networks (BSN), object sensor networks (OSN), and environment sensor networks (ESN) to capture context in the IoT paradigm. Later, Hydra3 [16] was developed as an IoT middleware that aims to integrate wireless devices and sensors into ambient intelligence systems. Hydra3 identifies context reasoning rule engine, context storage, context querying, and event/action management as the key components of a context-aware framework. In the IoT sensing scheme, RFID reader plays a key role. In [17], researchers illustrate eleven different categories of use cases that need to utilize RFID technology to interact with the IoT surroundings. In [18], researchers introduce a semi-passive, reconfigurable UHF RF identification (RFID) sensing tag operating as the generic sensing platform (GSP). The tag is highly configurable and can be dynamically switched between a Continuous data transmitting platform (Online mode) or a Data logging platform (Offline mode). Researchers in [19] develop a suite of web-based, user-level tools and applications designed to empower users by facilitating their understanding, management, and control of personal RFID data and privacy settings. Besides researches on RFID technology itself, research topics of implementing RFID based IoT systems on other application fields, such as massive manufacture workflow optimization [20], Geofencing item tracking [21] and industrial-level quality control [22], have drawn increasing attention not just from computer developers and researchers, but from all the industrial and science community. IoT is advancing fast, RFID and their combined technologies are also becoming ubiquitous and playing a critical role of identifying objects. However, smartphones, one of the most convenient and widely used portable computing devices, is still out of the picture of research on RFID based IoT systems. Although it matches all the requirements from both the hardware and the software perspective, severe RFID integration challenges remain.

3. SYSTEM DESIGN AND IMPLEMENTATION

The core concept of our system design is the URRM and its control and support components. In particular, our system architecture design contains five subsystems: the UHF RFID reader Mod (URRM) subsystem, the Mod control and management subsystem, the Mod support subsystem, the Battery subsystem and the Antenna subsystem. The following subsections discuss the five subsystems in details, followed by a discussion of the system workflow.

A. UHF RFID Reader Mod Subsystem

To embed a RFID reader module into the low-level of a native smartphone system, the major challenge is to establish a physical connection from the RFID module directly to the low-level smartphone system without significantly modifying the native system. We achieve this goal by designing the URRM for a Moto Z smartphone, both of which are showed in Fig. 1, respectively. In particular, we embedded the RFID reader module into the MotoMod platform [23] with a HAT (Hardware Attached on Top) adapter board [24]. Due to the modularization design of the Moto-Z mobile system, the URRM utilizes GreyBus [25] protocol to communicate the Moto Zs hardware layer through Serial Peripheral Interface (SPI) once the URRM is attached to a Moto Zs backboard. The picture and the brief architecture of the Moto-Z system and the URRM are depicted in Fig.1(a) and Fig.1(b), respectively.

(a) MotoMod and Moto Z products   (b) MotoMod and Moto Z structure
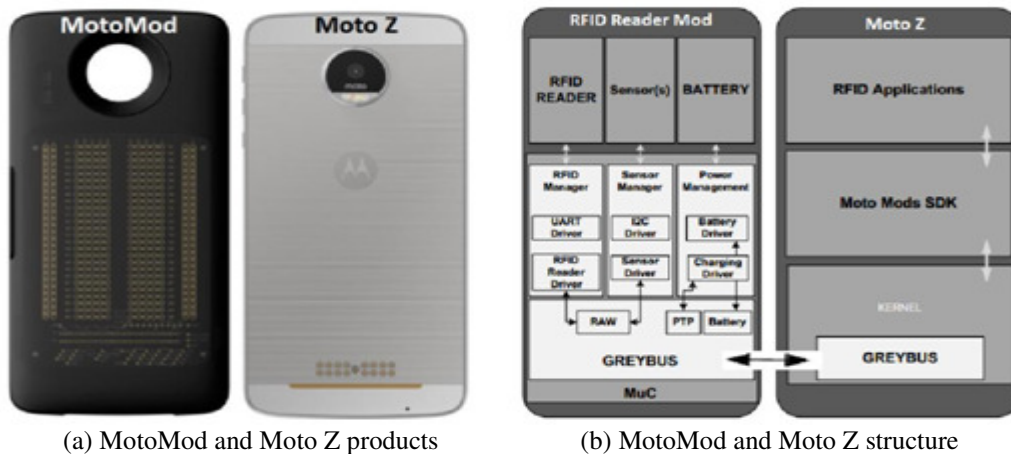
Fig. 1. Moto Mods and Moto Z

GreyBus is developed by Google and Motorola in 2014, it supports hot plug and play communications between the GreyBus Master and GreyBus Slave. The Moto Zs Mod physical hardware interface contains 16 metal touch points which supports USB2.0, USB3.1, SPI, Mobility Display Port (MyDP), Inter-IC Sound (I2S) and other mainstream communication protocols. Up to 15 Mps of high speed data exchange with a dedicated transmit and receive Direct Memory Accesses (DMA) between Master and Slave devices. To drive up the reader module, we use Universal Synchronous/Asynchronous Receiver/Transmitter (UsART) to port the reader Mod to the HAT board. To improve the energy efficiency, we further design a two-layer architecture which utilizes both the MuC on the MotoMod, i.e., the Mod Control and Management Subsystem, and the Application Processor (AP) on the Moto-Z, i.e., the Mod Support subsystem, to control and manage the RFID reader module jointly. Moto-Z smartphone system is equipped with a

SnapDragon 820 CPU as its AP which has 2.2GHz main frequency and 512 KB + 1 MB L2 cache. It is designed to support the latest version of Android system, i.e., Android 8.0, and smartphone applications. However, the current drain is high (28.8mA in full performance and 4.8mA in idle state). For scan-related tasks such as conducting IoT sensing or processing the raw data, the AP is over-powerful and hence becomes very energy-inefficient. The MuC on the MotoMod, on the other hand, has higher energy-efficiency (4.3mA current drain in full performance and 1.1uA in idle state). This system is powered by a STM-32L4 based core that has 80MHz main frequency and 1MB flash memory. It's computing capacity is less powerful comparing to the AP, but it is sufficient to handle the scanning routings such as sending commands or parsing events to or from the RFID module. In our stress test, the MuC can process no less than 1200 RFID tags per second without depleting its computing power and buffer. In our design, we keep computing and control oriented tasks of UHF RFID sensing, such as the scanning routings or raw data process procedures, on the MCU and create the Mod Control and Management subsystem. For tasks that can only run on Moto Z, such as the scan application with Graphic User Interface (GUI), we shift them to the AP and create Mod support subsystem. Both subsystems are discussed in the following two subsections.

B. Mod Control and Management Subsystem

The MuC and its baseline software framework [26] are the core components of the MotoMod system and provide Mod developers a POSIX and ANSI standards develop environment. The software framework is built upon a real-time operating system, the NuttX Real-time operation system (NuttX) [27], to provide basic I/O control, task scheduler, power management and drivers for specific circuits running on the MotoMod. In particular, the slave-end GreyBus protocol is implemented in the NuttX driver layer and is abstracted as a virtual file. All the data transmissions between Moto Z and MotoMod are implemented through reading or writing this virtual file. The system architecture is illustrated in Fig.1(b).
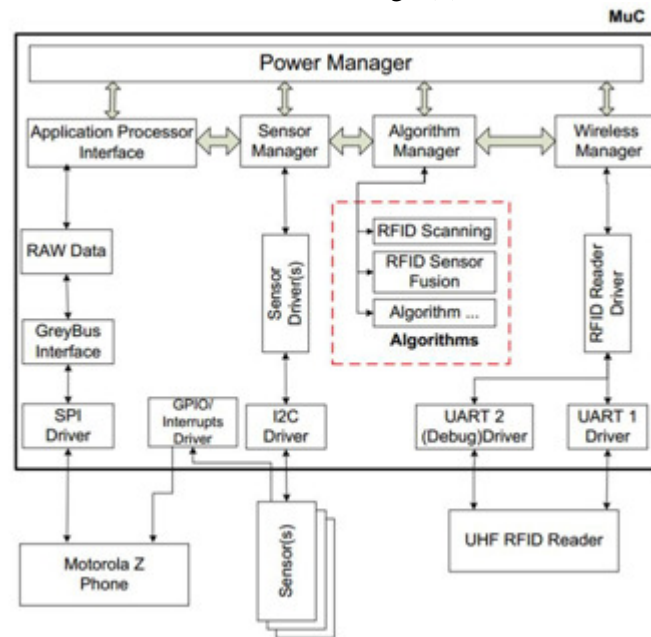

Fig. 2. Sensor Fusion Architecture

To control the RFID reader module, we deploy the sensor fusion architecture modified from in our previous works [28], [29], which is illustrated in Fig.2, on the baseline software framework. The core components of the sensor fusion architecture are: Power Manager, Sensor Manager, Algorithm Manager, Application Processor Interface and Wireless Manager. We specify their functions separately.

Power Manager: responsible for managing the tasks power mode votes and configuring the MuC's power mode based on the votes. The power manager also controls and monitors the status of the Mods battery.

Sensor Manager: responsible for configuring the sensors and managing the raw sensor data obtained from the sensors.

Algorithm Manager: responsible for configuring, scheduling, and providing the raw sensor data for all data processing algorithms.

Application Processor Interface: responsible for all communications between the sensor fusion core and the application processor. Its tasks include managing features turned on/off by the application processor, and interrupting applications when sensor data is available.

Wireless Manager: responsible for configuring the UHF RFID reader IC, and write/reading data of surrounding UHF RFID tags.

The design of the Mod Control and Management subsystem has two advantages. First, the architecture is extensible which means not only the RFID reader, but other similar wireless technologies, such as Bluetooth, WiFi, and NFC, can be deployed on the Mod and can work jointly following the same approach. Second, all the drivers and implementation details are encapsulated on the MuC and are transparent to the smartphone, which means even a URRM with different configurations, like with a different RFID reader module, is attached on, systems on the smartphone are not affected.

C. Mod Support Subsystem

On the smartphone side, we design a vertical system framework running on Android system which communicates through from the top (Application) layer of the Android system to the bottom (GreyBus protocol over the SPI transport) layer of the smartphone system. All functionalities of the whole framework are encapsulated and embedded with an API library, the Motlib, and are exposed to the Androids application layer. Our application, as well as any other third-party applications, utilizes the framework by simply including the Motlib library.

The architecture of the vertical system, including the Mod support subsystem and Mod control and management subsystem, are illustrated in Fig.3. For the Mod support subsystem,each component in this framework from top layer to bottom layer is discussed as follows:

RFID Application an Android application with a GUI that allows users to interact with the URRM by sending commands and displaying the reader Mod status and scan results. This application utilizes functions in the Motlib library to interact with the Mod Manager layer.

Mod Manager: The Mod Manager is a component of the Hardware Abstract Layer (HAL) in the NuttX system. It encapsulates the implementation details of hardware under NuttX and maps system calls from the Application Layer to the device driver functionalities. In our case, the URRM is registered to the Mod Manager with its all functionalities. Plus, Mod Manager can directly update the firmware of the MuC in URRM.
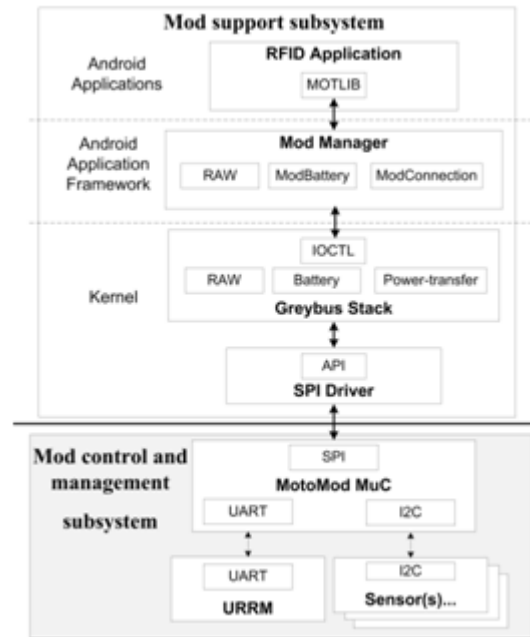

Fig. 3. Architecture of URRM and Mod Support Subsystems

GreyBus Stack (master end): GreyBus is running on the Kernel layer of the Android system. GreyBus defines multiple class protocols that are used to communicate and provide function to specific hardware. In our design, we enable three protocols: RAW, Battery and Power Transfer Protocol (PTP) which enable us to use our own designed communication package format (RAW), enable URRM to manage its own battery (Battery) and enable Moto Z can provide power to URRM (PTP).

SPI Driver: SPI Driver is running in the Android kernel. It controls the SPI Bus which is the physical communication interface between the MotoMod and Moto-Z mobile device for GreyBus communications to the MuC.

This design has two advantages. First, it allows Android applications to flash the MuC's firmware, either for driver updates or for upgrades of functionality, algorithms or even for the baseline system. This feature brings more flexibility and convenience to the application developer since without this function, such update or upgrade procedures can only be done manually by opening up the Mods enclosure and flash new firmware through J-tag or CC-debugger that is attached to the MuC. Second, the power management function in the Mod Support subsystem allows us to embed the external power source in the URRM and enable URRM to gain power supply from either the external power source or the Moto Z. The detailed implementation of the external power source is discussed in the next subsection.

D. Battery Subsystem

In our design, the RFID Mods MuC is responsible for controlling the power management. We use a MAX17050X chipset for monitoring the battery when in-use and charging. To maximize device power when RFID scanning is active, we couple a regenerative 3.7 volt battery with a capacity of 1500 mAH in the RFID Mod with the Moto Z through the MuC controller. Total energy capacity available is 5000 mAH when combined with Moto Z's native battery capacity of 3500mAH. Utilizing the MuC's power management component, our configuration uses the Mods battery until the battery level reaches 5%, at that point the MuC will switch in the Moto Zs battery for continuous use ensuring no degradation in performance.

UR Scanning mode	Measured Current Drain	Time (minutes)
Background Scanning (Sensor Fusion Core only)	**291mA**	**1824**
Foreground scanning (application running and display on)	388mA	1422

Fig. 4. Battery Life Evaluation

The method we use to measure the battery consumption is a NI USB-6251 DAQ to monitor current consumption at the battery contacts of the Moto Z. Once the DAQ is connected, we place the phone into an airplane mode to disable all wireless processing, and measure the current consumption with the UHF RFID application launched but not on. This provides us the idle screen current drain baseline measurement. We then enable the UHF RFID reader Mod. The configuration of the scanning is 10sec on and 10sec idle. The results in Fig.4 show that our design is capable of scanning for an entire day in the background. As we continue to develop and refine existing algorithms we anticipate further reductions in battery consumption.

E. Antenna Subsystem

When active, the reader Mod actively project energy to the RFID receiver, such as RFID tags, and capture their echoes as responses. Therefore, the performance of the reader as well as the interference amongst other wireless communication modules could be severe and becomes one of the main blockage for the deployment of RFID readers on smartphone platform. We provide our antenna design here and will evaluate the antennas performance in the next section.

We design the antenna to operate at 902MHz-928MHz, and an ambient distance of one meter for small passive tags, and three meters for active, or larger UHF RFID tags. We use Rogers 3010 material [30] to reduce the effects of increasing the Z height in our design. The antenna dimensions are 80x52x1.2mm. The feed impedance of the antenna was 50 Ohms. The direction of the antenna's radiation is perpendicular to the Moto Z's plane, and the polarization is across the Moto Z's antennas. The isolation between the UHF RFID antenna and the Moto Z's communication frequencies is larger than 30dB. The radiation efficiency of our design from 902MHz-928MHz was 5 - 10%.

F. System Workflow

Once the UHF RFID reader Mod is attached to the back of the Moto Z, the Mod transmits its manifest file from the MuC to the Motorola Mod Manager. The Motorola Mod Manager then parses the manifest file and determines what type of Mod is connected and its communication capabilities to expose for applications use. After the reader Mod is registered, applications can use Motorola Mod Manager to send RFID commands over the RAW protocol to the Mod. Once the reader Mod receives the RAW packet it will parse the data and determine which Manager (Sensor, Algorithm, Wireless) gets the payload. The Manager will then determine which driver or algorithm to send the payload. The driver or algorithm then parses the payload to perform the requested action. Upon a completion of the requested action a returned status (success or failure) is returned to the application.

4. URRM SIGNAL FREQUENCY ISOLATION ANALYSIS

As a part of the smartphone wireless system, the reader Mod works in a very close range with other native sensors and wireless modules. Therefore, the radio signal interference becomes a key feature which determines if the reader Mod is practical for commercial use or is just a prototype. Moreover, the radio performance of the reader Mod is another key feature as a sensor that needs to be evaluated, too short sensor range or too low sensitivity may significantly reduce the applicable range of the reader Mod in real-world scenarios. In the following subsections, we experimentally evaluate these two key features.

A. RF Signal Interference Validation

In Section I, we discuss about the antenna design in the hardware level that isolates the RFID signal from other sensors. In specific, as antennas of different wireless components are tightly coupled in a small enclosure, we need to confirm that our new RFID reader has no 1) Radiated spurious emissions [31] and 2) Inter-modulation distortion [32], [33]) against the components in the original system. In this section, we experimentally proof the signal isolation effectiveness of our design from the radio spectrum perspective.

Our experiment testbed is setup as follow. We use a RF shielded box (Will Technology model SH-120) to provide a signal-free environment and use two developer version Moto Z smartphones to perform data transmission via different wireless modules. One smartphone equips with the RFID reader Mod. A spectrum analyzer (Rohde & Schwarz model FSP13) is connected to the RF shield box to monitor and record the RF signals in the box from frequency 13Mhz to 6Ghz. Another antenna is wired through the shield and is implanted into the enclosure to provide cellular signal. The experiment testbed design is shown in Fig.5(b) and its setup illustrated in Fig.5(b).

To perform the interference test on a wireless module or sensor, we take the following steps. First, we place RFID tags into the RF shielded enclosure. Then, we engage the RF modules on both smartphones and start transferring data from one smartphone to another or to the implanted antenna. The data volume is 10G, hence its transmission provides us enough time to place both phones into the RF shield box and complete the setup. In the third step, we engage the reader Mod and keep it scanning the RFID tags. We then close the shield box and record the spectrums with and without the reader Mod. Following sections illustrates the test results for each wireless

module/sensor, respectively. We label the start, peak and end of the spectrum with green dot and specific frequency and decibel values.

(a) Tested design   (b) Tested setup
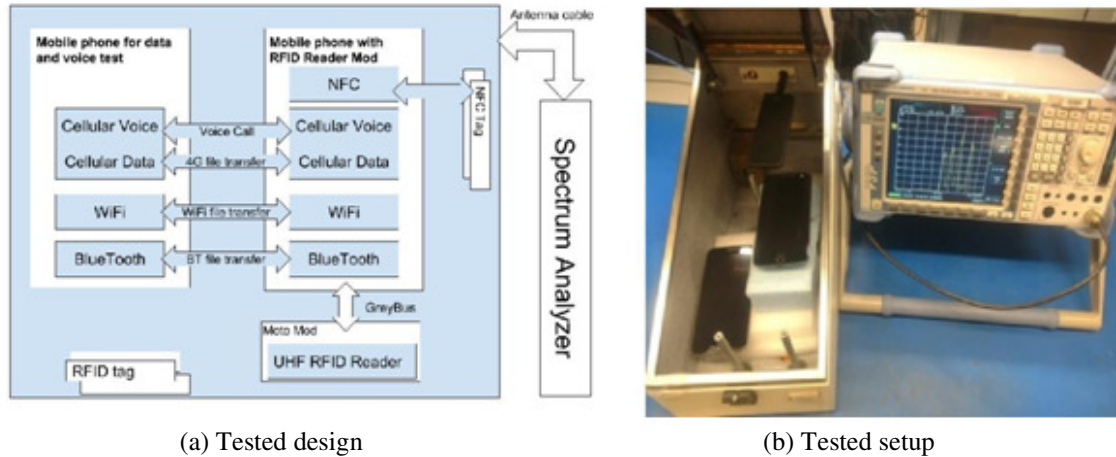
Fig. 5. Testbed for Frequency Isolation Analysis

B. Radio Spectrum of RFID Scanning Signal

In Fig.6 we show our design that is active at frequency band ranging from 864 MHz to 944MHz with energy peak at 912MHz. The strength at the peak is 15 dB at distance zero meters.
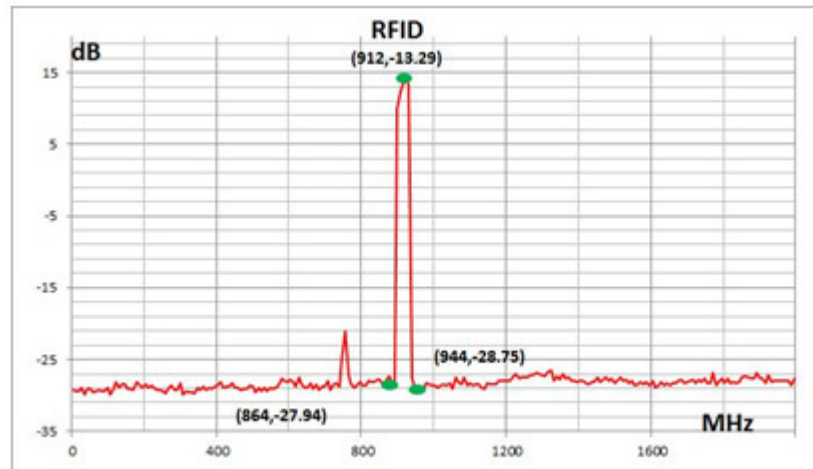

Fig. 6. UHF RFID Spectrum

1) Cellular Signal Interference: Testing for Cellular Interference is performed by connecting the Moto-Z with our UHF RFID reader Mod, and a live Verizon SIM. In both cases, with or without engaging the RFID reader, the radio spectrum of the Cellular voice call signal remains the same, as illustrates in Fig.7(a), which indicates that there is no interference between RFID reader signal and the Cellular signal. For the Cellular data streaming, there is also no interference. The signal spectrum of data streaming with and without RFID reader

scanning both follow the pattern depicted in Fig.7(b), all datasets streamed where received without retransmission.
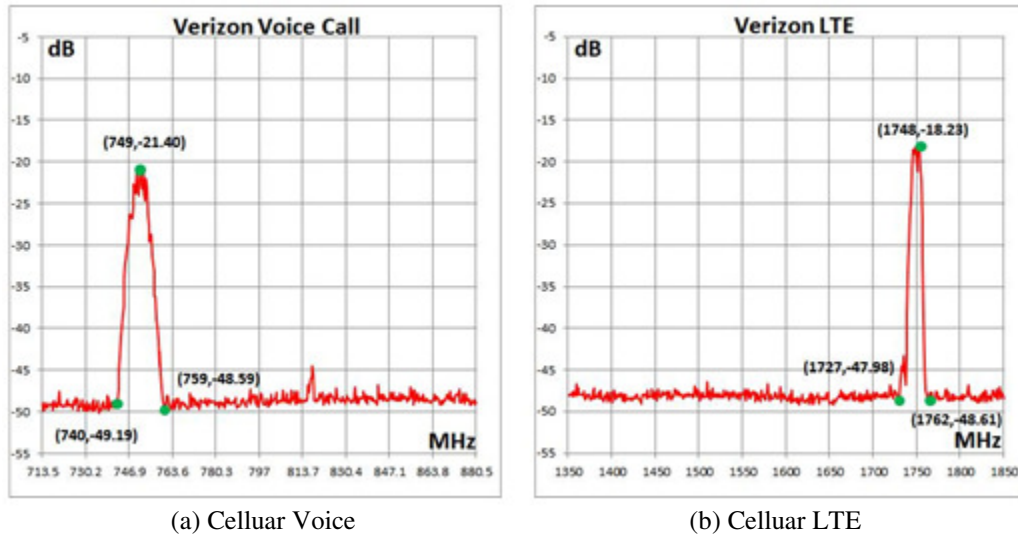

(a) Celluar Voice   (b) Celluar LTE

Fig. 7. Frequency analysis I

2) Bluetooth Signal Interference: To test Bluetooth interference, we capture the signal spectrums of all four stage of BT communications, i.e., BT scanning, advertising, paring and data transmission. The accumulated peak spectrums of all stages of BT communication, both with and without RFID scanning activity, are followed the same pattern depicted in Fig.8(a). The test results in Fig. 8 show no signal interference with Bluetooth, and we do not find transmission errors with the dataset received on the PC side, which indicate there is no interference between BT and RFID reader Mod.

3) Near Field Communication (NFC) Signal Interference : Testing for NFC interference is performed by connecting the Moto Z with Mod, and using a NXP NFC tag. While reading the tags memory, RFID background scanning is running. The test results in Fig.8(b) show no signal interference with NFC, and neither errors when reading or writing to the NFC tag.

4) WiFi Signal Interference: For WiFi test, we measure the two modes of WiFi communication: in 2.4G channel and in 5G channel. Both test are performed by set one smartphone as a hotspot and use another one to connect to the hotspot. Experimental data shows that for both modes, WiFi communication has no interference to the RFID reader Mod. The specific spectrum data are illustrated in Fig.9(a) and Fig.9(b) for 2.4G mode and 5G mode, respectively.

As illustrated in the five experiments above, our UHF RFID reader Mod has no interference on mainstream smartphone connectivity systems, which means our design is able to coexist with the native wireless sensors and communication modules.
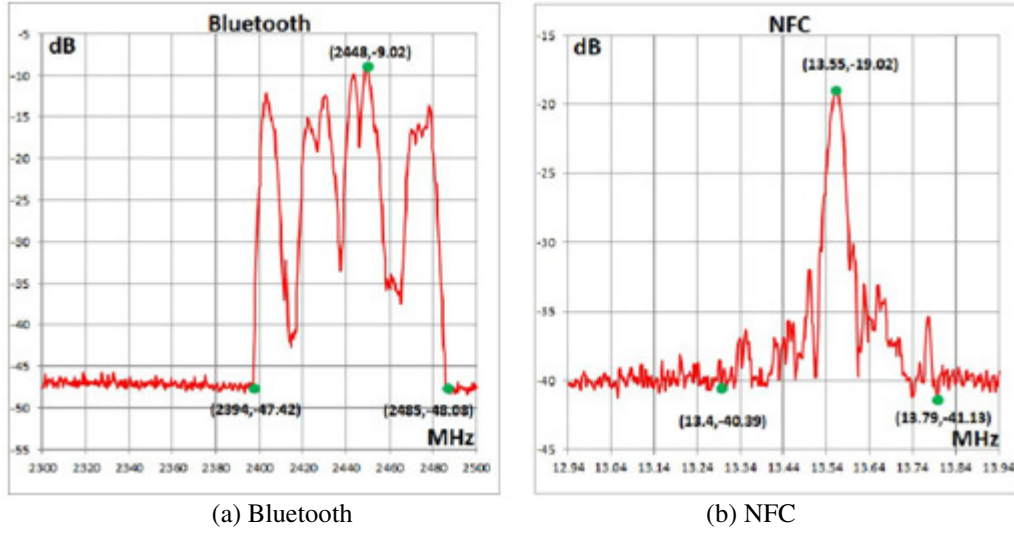
(a) Bluetooth   (b) NFC

Fig. 8. Frequency analysis II

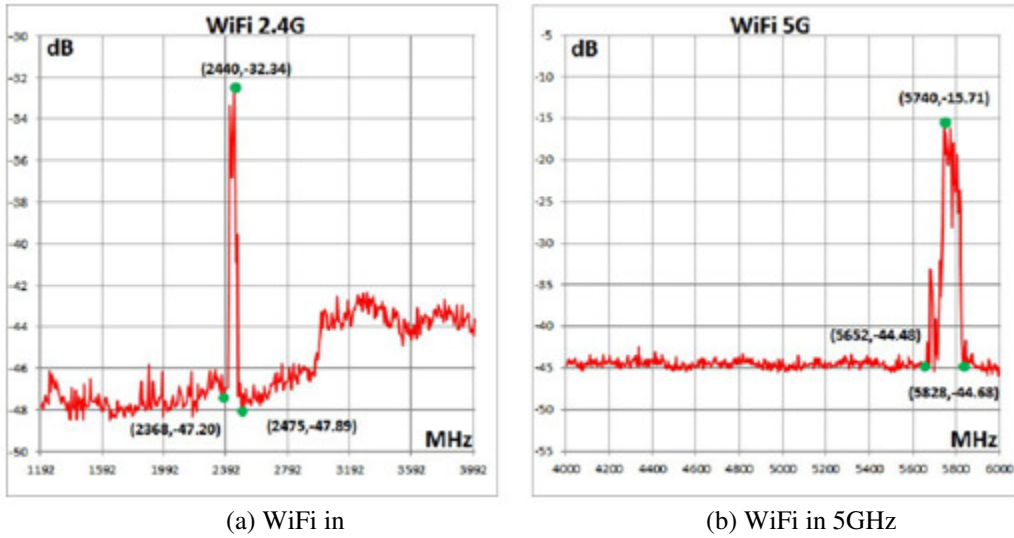(a) WiFi in   (b) WiFi in 5GHz

Fig 9. 2.4GHz Frequency analysis III

C. Sensor Performance Evaluation

As a IoT sensor, the sensing range and the signal attenuation are two key parameters which determine the quality of service of IoT sensing. In the following, we design two experiments to evaluate these two features, respectively. For these two experiments, the environmental temperature is 18.3 Celsius, the target RFID tag is SMARTRAC belt RFID paper tag [34] and the experiment lab is covered with radio absorb material to eradicate the radio reflection from wall faces. Battery is charged to 70%. Each value is the average value of 50 readings.

Our first experiment determines the sensing range of the UHF RFID reader Mod. From range 0 meter to 5 meters, we use our Mod to send 150 scanning signals at each distance and count the

successful Electronic Product Code (EPC) reads from RFID tag. Fig.10(a) illustrates the Successful EPC read count/rate of a RFID tag at different distances. As the picture shows, the signal picking up ability reduces when the distance increases but still very reliable before 3.5 meters. From 3.5 meters to 4.5 meters, it degrades dramatically and tags with range longer than 4.5 meter cannot be recognized. We determine the sensing range as 4 meters where the expected retry time is four, which is acceptable to most IoT sensing scenarios.

(a) EPC Read Evaluation Result   (b) Attenuation Evaluation
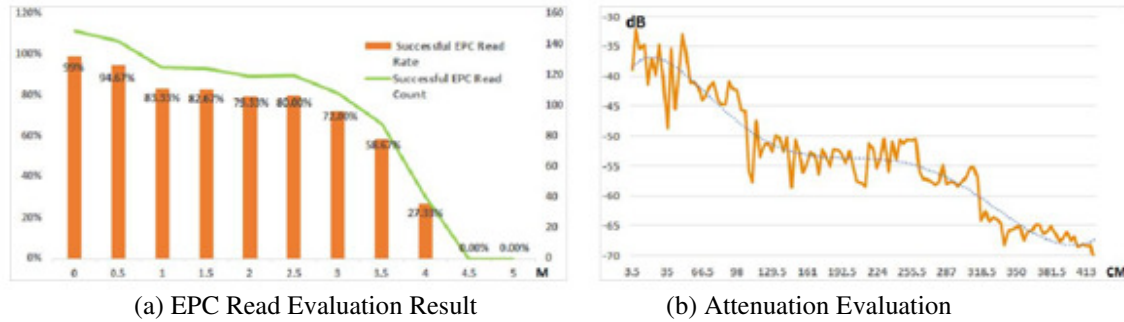
Fig. 10. Evaluation Result

Our second experiment evaluates the signal sensitivity and attenuation of our reader Mod. In this experiment, we measure the RSSI (Received Signal Strength Indicator) provide from the reader IC from range 0 Meter to 4.2 Meter. The result is depicted in Fig.10(b) where the solid line is the actual value and the dot line is the trendline. With the result from the previous experiment, the attenuation distribution shows that the reader Mod has stable radio performance when signal strength is higher than -65 dB (3.5 meters) and signals that are weaker than -70dB cannot be detected. One interesting finding is that, in average, the attenuation increases with the grows of the distance, but it also shows dramatic decrease in some distance pints. In our case, there are three: 0.97 Meter, 3.18 Meter and 4.21 Meter. We are not able to explain this phenomenon for now, but we will further study it in our future work.

5. CONCLUSION

In this paper, we design and implement a UHF RFID reader Mod with direct connection to a mainstream smartphone, Moto-Z, without changing the hardware of the phone. The experimental data proves that our design of the UHF RFID reader Mod can coexist with other smartphone sensors and communication components with no conflict. Also, the battery life with active RFID scanning is extended significantly due to the control system and the external power source design.

Our research and architecture has been tested for commercial deployment and we have high-lighted the key hardware and software architecture components which are extensible and scalable for continued research.

Through our work, mobile devices will take advantage of UHF RFID capabilities by evolving into sensory data accumulators passing raw and fused sensor states to cloud infrastructure for further compute-intensive processing. In our future work, we will optimize the algorithms running on the sensor fusion core to further reduce the power consumption to the acceptable range enabling us to remove the external power source. Next we will introduce RFID as a Sensor (RaaS), providing even wider sensing spectrum and more comprehensive environmental

information. Coupled with a highly optimized Sensor Fusion Core (SFC) and Bluetooth Low Energy (BLE) these sensors jointly provide a real-time, always-on and comprehensive sensory system optimized for a mobile devices limited battery life. Optimized wireless monitoring algorithms will be developed minimizing power consumption while guaranteeing an applications Quality of Service (QoS) requirements. We will also further study the signal sensativitiy and attenuation behaviors when the reader Mod is activated.

REFERENCES

[1] IDTechEx, R. Das, and P. Harrop, RFID forecasts, players and opportunities 2011-2021. IDTechEx, 2011.

[2] L. Yang, A. Rida, R. Vyas, and M. M. Tentzeris, "Rfid tag and rf structures on a paper substrate using inkjet-printing technology," IEEE Transactions on Microwave Theory and Techniques, vol. 55, no. 12, pp. 2894–2901, 2007.

[3] L. Atzori, A. Iera, and G. Morabito, "The internet of things: A survey," Computer networks, vol. 54, no. 15, pp. 2787–2805, 2010.

[4] B. Intelligenceo, "The global smartphone report: The forces behind the global deceleration in smartphone sales," http://www.businessinsider.com/, 2016.

[5] A. Kamilaris and A. Pitsillides, "Mobile phone computing and the internet of things: A survey," IEEE Internet of Things Journal, vol. 3, no. 6, pp. 885–898, 2016.

[6] S. Park, "A development of uhf rfid device for mobile iot service," Ubiquitous Computing, Electronics and Mobile Communication Conference (UEMCON), 2016.

[7] A. M. Requist and C. A. Requist, "Method for connecting a portable rfid reader to a handheld computing device via an audio connection," Mar. 14 2017, uS Patent 9,594,932.

[8] http://www.invengo.com/, "Invengo users manual," 2014.

[9] http://chafon.com/, "Chafon users manual," 2014.

[10] R. Triggs, "Lenovos sold 1 million moto z handsets globally," http://www.androidauthority.com/1-million-moto-z-sales-727170/, 2016.

[11] M. Mobility, https://developer.motorola.com/, 2016.

[12] C. Perera, A. Zaslavsky, P. Christen, and D. Georgakopoulos, "Context aware computing for the internet of things: A survey," IEEE Communications Surveys & Tutorials, vol. 16, no. 1, pp. 414–454, 2014.

[13] L. Capra, W. Emmerich, and C. Mascolo, "Carisma: Context-aware reflective middleware system for mobile applications," IEEE Transactions on software engineering, vol. 29, no. 10, pp. 929–945, 2003.

[14] T. Gu, H. K. Pung, and D. Q. Zhang, "A service-oriented middleware for building context-aware services," Journal of Network and computer applications, vol. 28, no. 1, pp. 1–18, 2005.

[15] A. Gluhak and W. Schott, "A wsn system architecture to capture context information for beyond 3g communication systems," in Intelligent Sensors, Sensor Networks and Information, 2007. ISSNIP 2007. 3rd International Conference on. IEEE, 2007, pp. 49–54.

[16] A. Badii, M. Crouch, and C. Lallah, "A context-awareness framework for intelligent networked embedded systems," in Advances in Human-Oriented and Personalized Mechanisms, Technologies and Services (CENTRIC), 2010 Third International Conference on. IEEE, 2010, pp. 105–110.

[17] X. Jia, Q. Feng, T. Fan, and Q. Lei, "Rfid technology and its applications in internet of things (iot)," in Consumer Electronics, Communications and Networks (CECNet), 2012 2nd International Conference on. IEEE, 2012, pp. 1282–1285.

[18] M. S. Khan, M. S. Islam, and H. Deng, "Design of a reconfigurable rfid sensing tag as a generic sensing platform toward the future internet of things," IEEE Internet of things journal, vol. 1, no. 4, pp. 300–310, 2014.

[19] E. Welbourne, L. Battle, G. Cole, K. Gould, K. Rector, S. Raymer, M. Balazinska, and G. Borriello, "Building the internet of things using rfid: the rfid ecosystem experience," IEEE Internet computing, vol. 13, no. 3, 2009.

[20] M. Liukkonen and T.-N. Tsai, "Toward decentralized intelligence in manufacturing: recent trends in automatic identification of things," The International Journal of Advanced Manufacturing Technology, vol. 87, no. 9-12, pp. 2509–2531, 2016.

[21] K. Ding, P. Jiang, P. Sun, and C. Wang, "Rfid-enabled physical object tracking in process flow based on an enhanced graphical deduction modeling method," IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2017.

[22] F. Bibi, C. Guillaume, N. Gontard, and B. Sorli, "A review: Rfid technology having sensing aptitudes for food industry and their contribution to tracking and monitoring of food products," Trends in Food Science & Technology, 2017.

[23] M. M. LLC, "Bring your moto mod from concept to market," in https://enterprise.indiegogo.com/motomods. Motorola Mobility, 2016.

[24] M. Mobility, "Mdk user guide: Hat adapter board," https://developer.motorola.com/build/mdk-user-guide/hat-adapter-board, 2016.

[25] A. ray, "Greybus protocol an application layer for unipro," Project Ara Developers Conference, http://www.modularphonesforum.com/news/greybus-protocol-an-application-layer-for-unipro-491/, 2015.

[26] M. Mobility, "Moto mods system architecture," https://developer.motorola.com/explore/system-architecture, 2016.

[27] G. Nutt, "Nuttx operating system user s manual," 2014.

[28] D. Lautner, X. Hua, S. Debates, M. Song, J. Shah, and S. Ren, "Baas (bluetooth-as-a-sensor): conception, design and implementation on mobile platforms," in Proceedings of the Symposium on Applied Computing. ACM, 2017, pp. 550– 556.

[29] D. Lautner, S. DeBates, J. Shah, M. Song, and S. Ren, "Sensor-based low power management for mobile platforms," in High Performance Computing and Communications (HPCC), 2015 IEEE 7th International Symposium on Cyberspace Safety and Security (CSS), 2015 IEEE 12th International Conferen on Embedded Software and Systems (ICESS), 2015 IEEE 17th International Conference on. IEEE, 2015, pp. 934–940.

[30] J. C. Rautio, "Measurement of uniaxial anisotropy in rogers ro3010 substrate material," in Microwaves, Communications, Antennas and Electronics Systems, 2009. COMCAS 2009. IEEE International Conference on. IEEE, 2009, pp. 1–4.

[31] Q. Yu, "Radiated spurious emissions measurement by substitution method," in Electromagnetic Compatibility (EMC), 2010 IEEE International Symposium on. IEEE, 2010, pp. 159–164.

[32] C. H. Z. X. S. Linfang, "Inter-modulation distortion analysis for single-drive mach-zehnder modulator [j]," Acta Optica Sinica, vol. 11, p. 004, 2006.

[33] J. McRory and R. Johnston, "An rf amplifier for low inter-modulation distortion," in Microwave Symposium Digest, 1994., IEEE MTT-S International. IEEE, 1994, pp. 1741–1744.

[34] K. Ding, P. Jiang, P. Sun, and C. Wang, "Smartrac belt rfid paper tag," https://www.smartrac-group.com/belt.html, 2017.

IMAGE QUALITY ASSESSMENT- A SURVEY OF RECENT APPROACHES

Noor Al Madeed, Zainab Awan and Somaya Al Madeed

College of Computer Science and Engineering, Qatar University, Doha, Qatar

ABSTRACT

Image Quality Assessment (IQA) is the process of quantifying degradation in image quality. With the increased image-based applications IQA deserves extensive research. In this paper we have presented popular IQA methods for the three types namely, Full Reference (FR), No Reference (NR) and Reduced Reference (RR). The paper gives comparison of the approaches in terms of the database used, the performance metric and the methods used.

KEYWORDS

Full reference, image quality assessment, no reference, reduced reference

1. INTRODUCTION

Image quality assessment refers to the evaluation of the quality of a distorted image with respect to the original image. There are two main categories of IQA; namely, subjective methods and objective methods. The former is difficult and time-consuming for real-time assessment needs. Subjective methods require human intervention and are very expensive. The later refers to the algorithmic models to estimate the image quality. For the real-time assessment we require automated or objective methods. There are three dimensions of objective IQA, namely, Full Reference FR, Reduced Reference RR and No Reference NR. FR IQA methods need the whole information of the reference image, while NR IQA methods stipulate that distorted image be assessed without any information of the reference image. RR methods are a compromise between FR and NR methods that require partial information of the reference image.

The rest of the paper has been organised into following sections. Section 2 defines image quality, section 3 explains image quality assessment, section 4 outlines types of IQA, and section 5 highlights few benchmark databases, followed by section 6 that discusses IQA approaches, followed by the discussion and conclusion.

2. IMAGE QUALITY

Image quality has been defined [1] as a composite of three factors, fidelity, perception and aesthetics. Fidelity is the exactness of a distorted image relative to its original image. Perception is inspired from characteristics of human visual systems HVS. This type of metric considers for example visual attention [2], contrast masking [3] etc. Aesthetics is subjective and may contain [1] visual constancy, visual attention and visual fatigue etc.

3. IMAGE QUALITY ASSESSMENT

Due the continuous development in the multimedia technologies, images have become a great source of information. While transmitting an image over wired or wireless network channel its quality is degraded due to the noise/distortions introduced. The image on the receiver side is no longer the same as the one sent at sender side. IQA is hard as the algorithm needs to be reliable, fast, robust and should require less information from the reference image.

4. THREE TYPES OF IQA

IQA methods are categorized into the following three categories. Methods are categorized based on the amount of information needed from a reference/original image.

4.1. Full-Reference IQA

Full Reference QA algorithms have the original/reference image available at hand for the comparison with a distorted version of the same image. These kinds of methods are best suited for in-lab testing as in the real use cases the reference image is not available. Figure 1 is reproduced from [1].

Figure 1: FR IQA Framework [1]

4.2. No-Reference IQA

In NR QA methods the reference image is not available at all. QA algorithm must be able to evaluate the quality of a distorted image as that is the only information available at hand. Figure 2 is reproduced from [1].
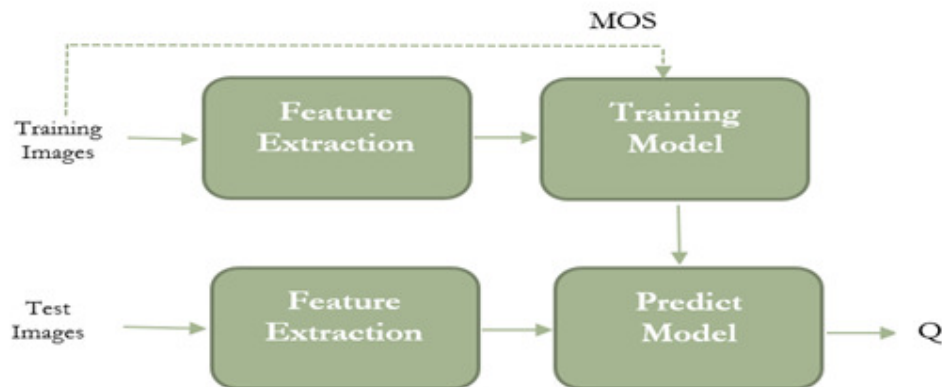

Figure 2: NR IQA Framework [1]

4.3. Reduced-Reference IQA

RR QA methods serve as a compromise between NR and FR methods. These methods require partial information from the reference image in the form of features. That partial information is sent to the receiver side through an ancillary channel. The goal here is to reduce the features required. Figure 3 taken from [4] indicates the framework for reduced reference IQA.
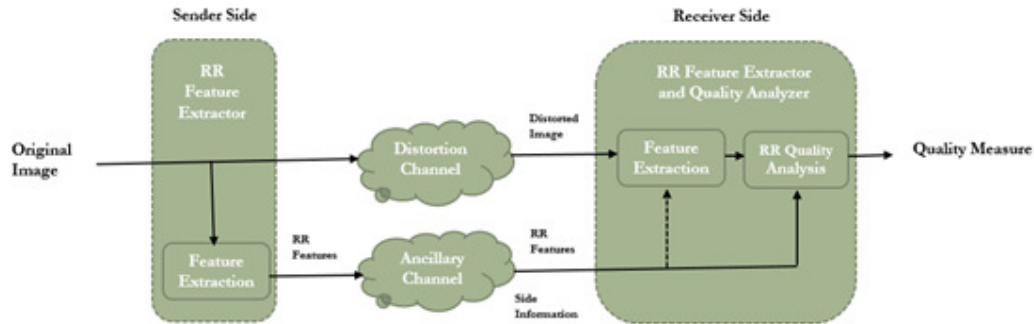

Figure 3: RR IQA Framework [4]

5. BENCHMARK DATABASES FOR IQA

In this section we will discuss some standard databases available for the evaluation of IQA methods. In order to evaluate a newly proposed IQA metric with the ones that are currently existing, one needs to perform experiments on the publicly available databases for the purpose. The databases contain reference images for which various distorted images are produced by using different distortion types. Each distorted image has as associated subjective score. To evaluate a new metric you need to compute the objective score of each distorted image and compute correlations between the objective and subjective scores.

5.1. CSIQ Database

This database [5] contains 30 reference images where each image is introduced with 6 types of distortions at 4 to 5 different levels of distortions. The distortions types include JPEG compression, JPEG-2000 compression, global contrast decrements, additive pink Gaussian noise, and Gaussian blurring. The subjective ratings are given in Differential Mean Opinion Scores (DMOS). The database is freely downloadable for the experimentation.

5.2. TID2013 Database

This database [6] contains larger test images set than its earlier release TID2008. It contains 25 reference images, 24 distortion types and 5 levels for each distortion type, which makes 3000 test images in total. The images and associated MOS values are freely available for the download and research purposes.

5.3. Tampere Image Database 2008 (TID2008) Database

This database [7] contains 25 reference images and 17 types with 4 levels of distortions for each image, in total 1700 images and subjective ratings are reported in the form of Mean Opinion Scores (MOS). The database is available freely for investigation of the existing and new metrics.

5.4 Laboratory for Image and Video Engineering (LIVE) Database

This database [8], [9] contains 29 reference images, each image distorted with 5 distortion types, total of 799 distorted images. Subjective ratings are provided in the form of DMOS.

5.5 IVC Database

This database [10] contains 8 original images distorted at 3 different processing and 5 different compression rates. In total there are 120 distorted test images available.

5.6 Cornell A-57 Database

This database [11] is relatively a small database. It contains 3 reference images, 6 distortion types at 3 different distortion levels. This database is publicly available for research.

5.7 Wireless Imaging Quality (WIQ) Database

This database [12] contains 7 reference images and 80 distorted images.

6. STATE-OF-THE-ART APPROACHES

This section discusses some recent approaches that have been proposed for the three types of assessment methods.

Table 1: Full reference approaches

Serial No	Year	Method	Database	Performance Metric
1	2013	Discrete Wavelet Transform [13]	IVC, LIVE II and TID 2008	SSIM and PSNR
2	2012	Machine Learning [14]	LIVE and TID2008	LCC, KROCC and SROCC
3	2011	In spatial domain [15]	Standard MATLAB Vegetable Image	SCC
4	2015	Binocular Visual Characteristic [16]	NBU, LIVE I, LIVE II, MICT and CML	PLCC, SRCC and RMSE
5	2010	Gabor Filters [17]	Their own database explained in the paper	PCC, RMSE MOS, SCC and Outlier Ratio

6.1. Full Reference Approaches

Table 1 contains the studies that have been reviewed for FR methods. In this paper [13], a novel framework for calculating Image Quality Metric (IQM) is given in discrete wavelet domain using Haar wavelet. The framework is applicable to map based such as SSIM that generate quality maps and nonmap based metrics for example PSNR as a final score, with reduced complexity and improved accuracy. The framework has been applied to various well-known assessment methods SSIM, VIF, PSNR and AD. The framework provides trade-off between accuracy and complexity, hence, could be efficiently applied in wavelet based image/video processing applications.

This paper [14] presents a machine learning expert based on SVM classification and regression, known as Machine Learning based IQM (MLIQM), shown in Figure 4. The feature is composed

of local quality features on which multi class SVM classifier is applied to classify test image into one of the five quality classes as recommended by ITU. After this SVM regression is applied to score the quality of the test image. The proposed technique is compared with four FR IQA techniques i.e. VSNR, PSNR, VIF and MS-SSIM, have been reported to perform better than them.
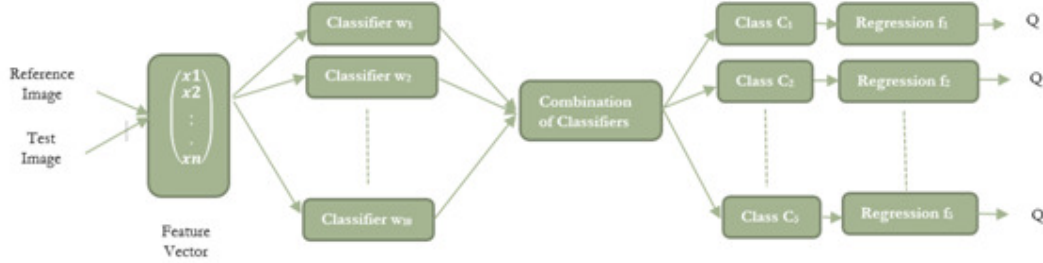

Figure 4: Framework for quality score calculation [14]

This paper [15] proposes new FR IQA metric for objective assessment in spatial domain. This metric is a combination of four factors, namely, sensitivity, edge performance index, contrast improvement and structural correlation. The proposed framework is shown in Figure 5. The proposed metric is reported to be more reliable than SSIM across different distortion types.
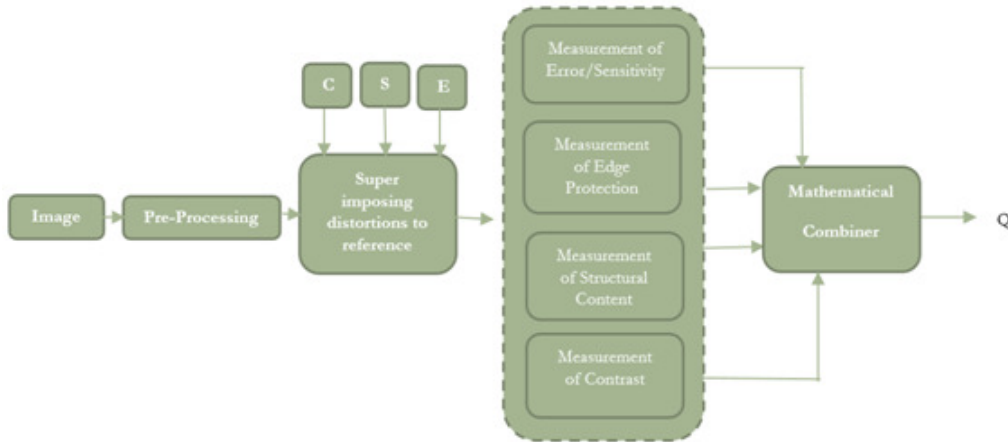
Figure 5: Spatial domain quality metric [15]

This paper [16] proposes a framework that gives IQA method for stereoscopic 3D images. The proposed methodology consists of the two phases; training and quality estimation. Binocular receptive field properties are learnt to simulate in simple and complex cells in cortex and to project their impact in quality estimation.

This paper [17] proposed FRIQA for the images distorted with local geometric distortions. The framework is shown in Figure 6. At first, the displacement field and a set of features by using Gabor filters from the image structure are extracted. Local quality scores are then calculated by evaluating the impact of displacement field on the features. Error pooling is performed to calculate the overall quality metric.
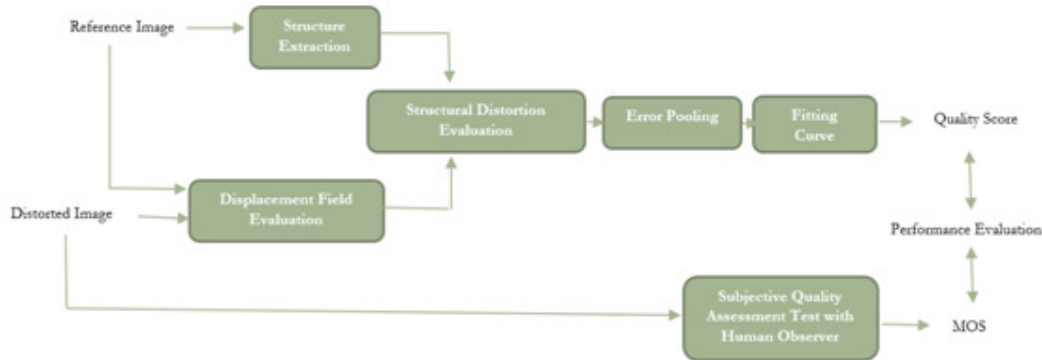
Figure 6: FR IQA framework for stereoscopic images [17]

Table 2: Reduced reference approaches

Serial No	Year	Method	Database	Performance Metric
1	2015	Deep Learning, Restricted Boltzmann Machine Similarity Measure RBMSim [18]	LIVE Multiply Distorted Image Quality and CSIQ	SROCC and Pearson Correlation Coefficient
2	2012	Entropic Differencing [19]	Tampere Image and LIVE	SROCC
3	2011	2011 Reorganized DCT Based Image Representation [20]	LIVE Image	CC, SROCC, RMSE
4	2012	2012 Image Statistics in Pixel Domain[21]	LIVE	CC, SROCC, RMSE, MAE
5	2012	2012 Structural Similarity Estimation [22]	] LIVE, Cornell, IVC, Toyama-MICT, TID 2008	CSIQ PLCC, MAE, RMS, SRCC, KRCC
6	2014	Phase Information in Complex Wavelet Domain [23]	LIVE	Correlation Coefficient (Prediction Accuracy) and Rank-Order Correlation(Prediction Monotonicity)
7	2015	Entropy Differences in DCT Domain [24]	CSIQ , Toyama and LIVE	Pearson Linear and Spearman Rank-Order Correlation Coefficient (SROCC)

6.2. Reduced Reference Approaches

Table 2 contains the studies that have been reviewed for RR methods. Novel stochastic RR-IQA metric [18], automatic stochastic procedure capable of assessing the quality of distorted images, independent of the type of image of distortions. Compared with subjective IQA algorithms and two benchmark databases. Restricted Boltzmann Machines are stochastic neural networks and contain two layers of neurons, the hidden layer h and the visible layer v, used for unsupervised

learning. The learning algorithm used for BMs is called Contrastive Divergence CD. Framework is called Gaussian Bernoulli Restricted Boltzmann Machine (GRBM). Each pixel from the RI is modelled on the visible layer v, with individual RGB value encoded as one neuron. For easy weight manipulation the image was split into sub-images. The distorted image DI quality can be assessed by initializing the v of GRBM learned on the RI by the sub images from DI, this is called vDI. After this step, Gibbs sampling is performed to infer the values of hidden neuron and then values of visible neurons are reconstructed based on the hidden neuron values. The reconstructed values are called vDI*. The Root Mean Squared Error RMSE distance (or any other metric) between vDI and vDI* reflects the quality of distorted image. GRBM trained on the RI has to be sent to the client side and very less information is required from the reference image. The performance is comparable with widely known FR- IQA metrics and its computation time is less. So, it can be applied to video/multimedia applications to be able to do real-time performance evaluation of network services.

In [19] the quality score is estimated by measuring the changes in weighted entropies of reference and distorted images in the wavelet domain. Wavelet coefficients of natural image and distorted are modelled as GSM distributions. The quality metric is obtained as a distance between reference and natural image approximation of a distorted image. The proposed algorithm perform much better than FR MSE algorithm.

In this paper [20] the IQA metric is based on Discrete Cosine Transform DCT coefficients, the method follows three steps, shown in Figure 7. In the first step, block-based DCT is applied, as a result of which we have DC coefficients where the most of the energy of the image resides and AC coefficients which can be easily ignored because energy is very low within these coefficients. The DCT coefficients are reorganized by decomposing them into 10 sub bands. Then the coefficients of the same sub bands from different blocks are grouped together such that there is a structural similarity between all the sub bands. The reorganized DCT coefficients look like wavelet image representation. In the second step, generalized Gaussian density GGD curve fitting is done to model DCT coefficient distribution GGD model efficiently represent the coefficient histogram for each DCT subband. In addition to two standard parameters of GGD another parameter, prediction error is introduced. In the last step, city block distance is computed between actual distribution and the fitted distribution. The proposed method has been tested on LIVE database and compared with the well-known RR WNISM and FR PSNR and SSIM. The proposed method outperforms the RR WNISM and FR PSNR but slightly underperforms the FR SSIM.

Figure 7: RR IQA DCT based metric [20]

In [21] a novel RR IQA algorithm based on image statistics is proposed. Image statistics are modelled in pixel domain which is based on gradient distribution. RR quality features are extracted based on statistical model. The quality is measured by determining the similarity between reference and distorted image. The proposed method is reported to perform as good as widely known FR PSNR.

In this paper [22], an approximation attempt for RR SSIM metric from FR SSIM has been made by divisible normalization transform DNT of the natural image. In addition to quality assessment, image repair concept has been introduced by matching the sub band statistical properties of the distorted image with those of the reference image. Features are extracted from the reference image by applying multi-scale linear transform (wavelet transform) and then DNT representation is calculated by dividing each coefficient of wavelet by a local energy measure based on neighbouring coefficients. For the efficient summarization of the statistical properties of the reference image zero mean GGD is fitted over DNT coefficients. For the estimation of RR SSIM, the effect of the distortions on the statistical properties need to be coherent with corresponding FR SSIM. The proposed method is compared with FR (PSNR and SSIM) and RR (Wavelet Marginal and DNT marginal) and the proposed metric is highly competitive in most of the cases.

In this paper, [23] IQA is based on the Complex Wavelet Transform (CWT) as it carries not only the magnitude information but also the phase information. Choosing the right RR features for IQA is critical, they have to be representative and yet small. Here different features are used namely, KLD, the standard deviation, the kurtosis, the skewness in the distribution of the relative phase and directional information from CWT coefficients. Generalized complex wavelet coefficient known as dual-tree complex wavelet transform (DT-CWT) has been used rather than simplistic CWT. The central idea of the algorithm is that the different distortion types affect the distribution in different ways. The distribution of phases of CWT coefficients is uniform and therefore contain no information. But the distribution of relative phases of different images follows a similar pattern when affected by a single distortion type. In case of different distortion type the pattern would be different. General Regression Neural Network GRNN has been used to predict the quality of distorted image by regression. The performance has been compared with 2 FR-IQA and 5

RR-IQA metrics and the proposed method is reported to be as good as popular FR-IQA metric MS-SSIM while using smaller features. To assess the quality of the distorted image we need the histograms of the relative phase of the distorted and the original image and then compute the Kullback-Liebler divergence.

In this paper [24], Figure 8, the differing importance of spectral sub-bands in DCT domain of images have been exploited by considering the fact that human eye corresponds differently to different bands. Quality of image should be assessed separately in each sub-band. Quality of the distorted image is computed via Shannon entropy differencing of DCT co-efficients between the reference and the distorted image. The overall quality of the distorted image is calculated by the weighted summation over entropy difference of all the sub bands. On the sender side the feature extraction is performed in three steps i.e. firstly, block-based DCT of the RI, secondly, the coefficient rearrangement and lastly the entropy calculation of each band. The same process is performed on the receiver side with the DI. The entropy features of RI at the sender side are sent to the receiver side over the ancillary error free channel which is required at the receiver side for the differencing. Higher prediction accuracy was attained using less reference data, similar frequency bands were merged into one term. The proposed algorithm was compared with already existing RRED, WNISM and RRVIF and overall outperforms over the three methods.
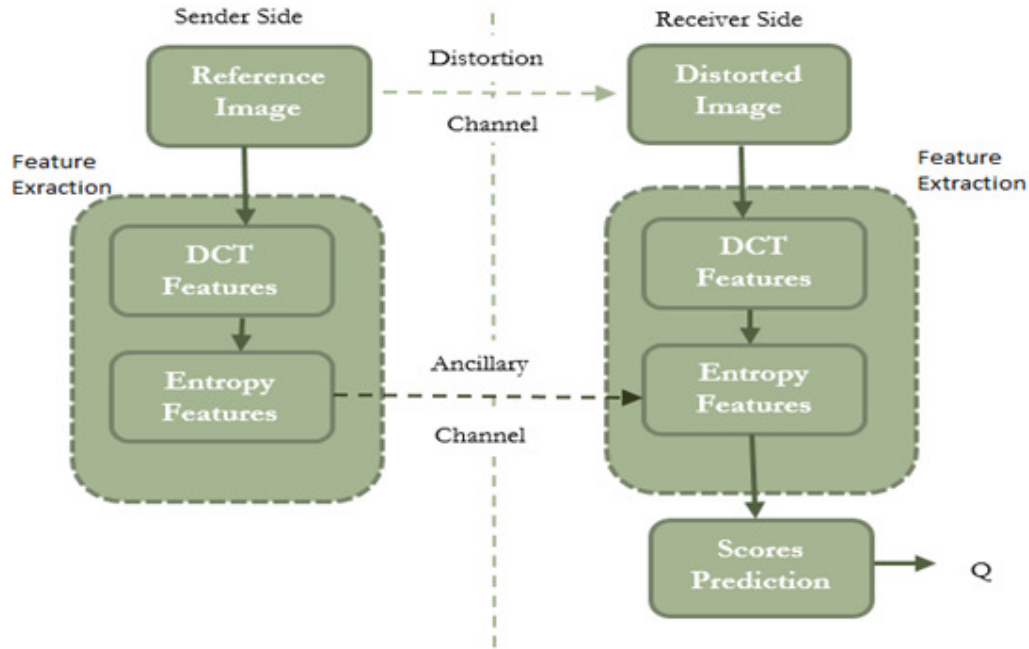
Figure 8: Entropy Difference in DCT domain base RR-IQA algorithm [24]

Table 3: No reference approaches

Serial No	Year	Method	Database	Performance Metric
1	2013	Combination of Free energy theory and Structural Degradation [25] model	LIVE	PLCC, SROCC and RMSE
2	2014	Statistical Characterization in Shearlet Domain [28]	LIVE, Multiply Distorted LIVE and TID2008	LCC and SROCC
3	2012	Visual Codebooks [30]	CSIQ and LIVE	SROCC and LCC
4	2013	Filter Learning [31]	LIVE, TID2008, Sharpness OCR Correlation SOC dataset and News dataset	LCC and SROCC
5	2014	Spatial and Spectral Entropies [32]	LIVE and TID 2008	SROCC, LCC and RMSE

6.3. No Reference Approaches

Table 3 contains the studies that have been reviewed for NR methods. The proposed framework [25] is a combination of two existing NR-IQA approaches, shown in Figure 9. The two approaches are called free energy based distortion metric FEDM and structural degradation model SDM. The newly proposed metric no reference free energy and structural degradation model NFSDM is reported to outperform two FR metrics, SSIM and PSNR and NR metrics BLIINDS-II [26] and DIIVINE [27]. This metric makes use of a relatively novel feature which is a non-linear

combination of free energy and structural degradation information. This metric has been tested on LIVE database.
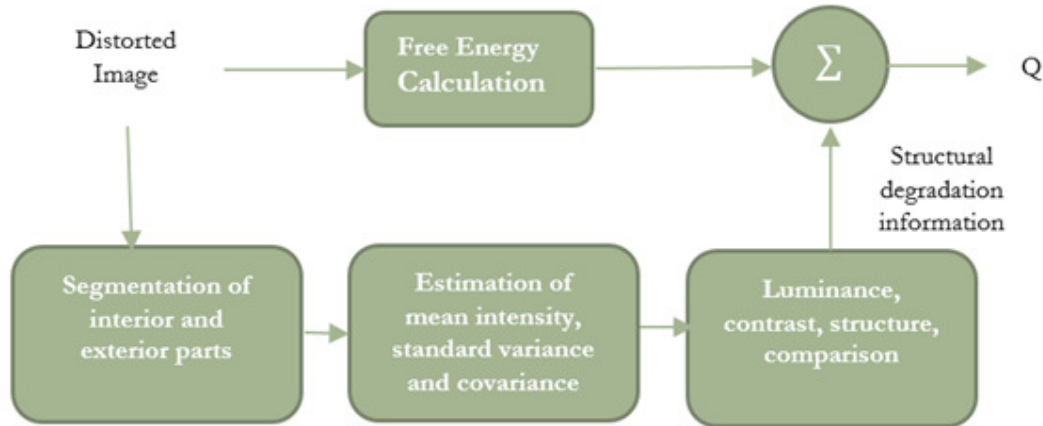

Figure 9: NFSDM Framework [25]

This paper [28] proposes shearlet based no reference image quality assessment SHANIA framework, Figure 10. The method makes use of natural scene statistics NSS model in shearlet domain [29]. The central idea behind approach is that the statistical properties of natural images remain constant in shearlet domain and change for the distorted images. The natural parts from the distorted image are used as reference and quality score is computed between reference and the distorted parts. The problem is mapped into a classification problem. The features are obtained from the distorted and reference parts and softmax classifier is used to obtain the quality score. The proposed framework is comparable in performance to three FR metrics including PSNR, SSIM and Multi-Scale SSIM and four NR methods BIQI, DIIVINE, BLINDS II and BRISQUE.
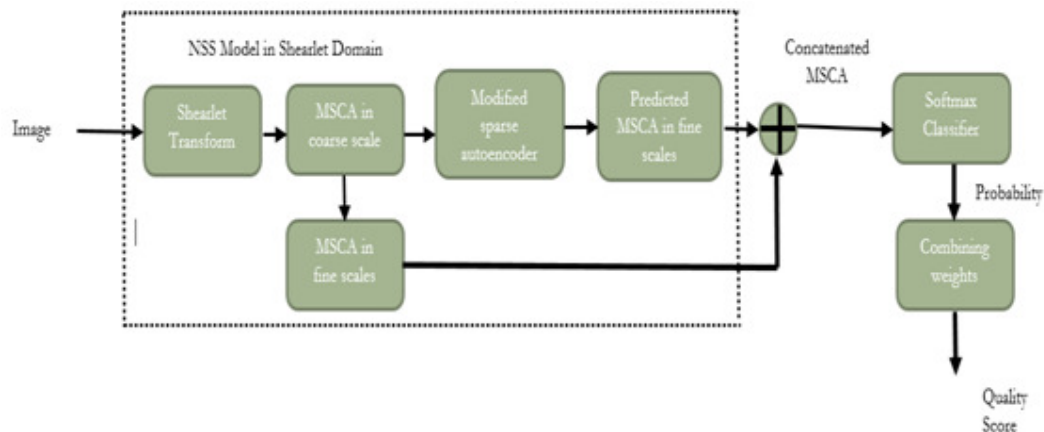

Figure 10: SHANIA Framework [28]

The paper [30] proposes NR-IQA framework based on visual codebooks which is a generalized framework not particular for some types of distortions, as shown in Figure 11. The proposed framework is reported to outperform FR-IQA SSIM and PSNR on LIVE database. The proposed method makes use of local feature extracted by using Gabor-filters. After this stage, visual codebook is constructed by forming a large set of block based Gabor feature vectors for the training images. Codebook C is constructed from that set by using clustering algorithm. Images

are represented in the form of codewords taken from codebook. Afterwards, regression is applied to calculate the quality score.
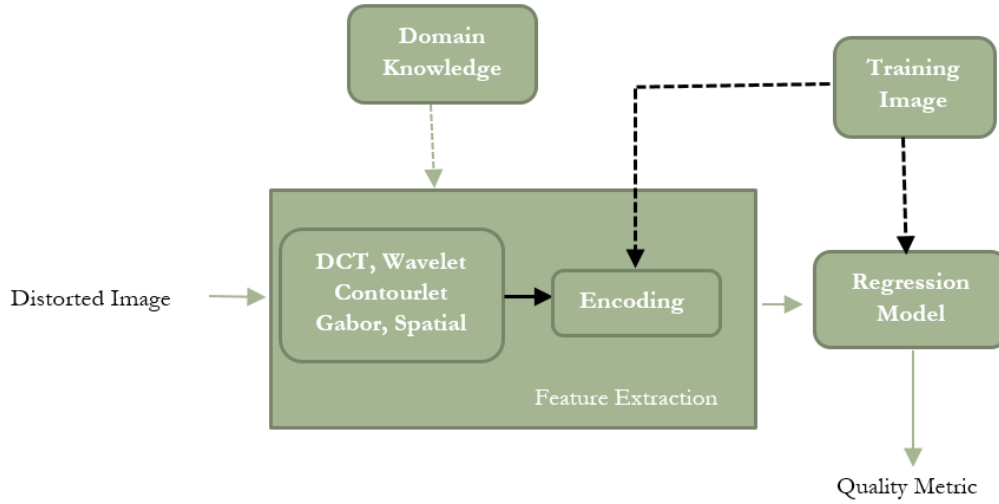

Figure 11: Visual Codebooks based NR IQA Framework [30]

The proposed framework [31] is for multiple domain images in real time applications, shown in Figure 12. There are three steps involved in this method, (1) local feature extraction, (2) global feature extraction and (3) regression model. In this approach step 1 and 3 unlike previous approaches are not treated independently rather they are linked together by back projection. Back projection allows to choose less number of discriminative features. The method utilizes relatively small set of features obtained from raw image patches directly which makes it a fast method. It has been applied to the document image datasets to prove its cross-domain applicability.
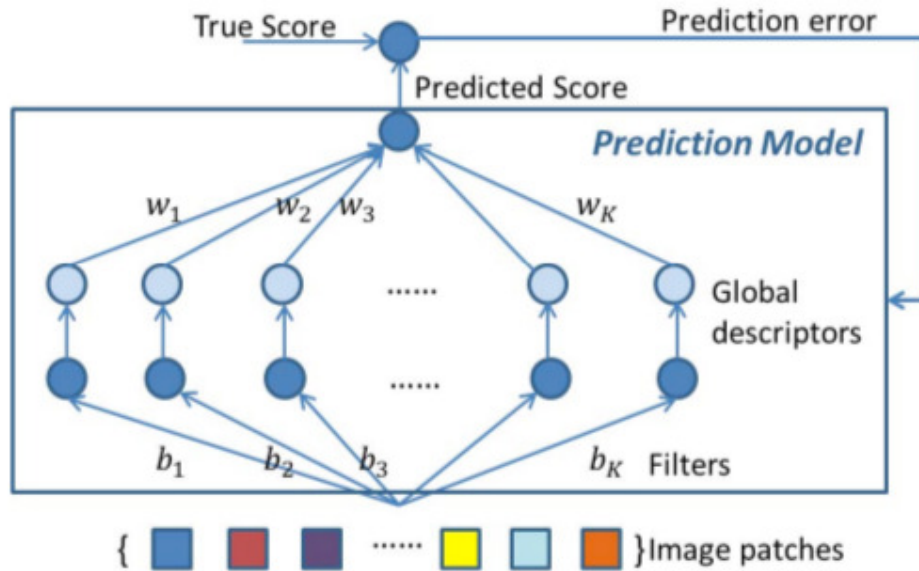

Figure 12: Framework for NR-IQA based on filter learning [31]

The proposed framework [32] utilizes local spatial and spectral features of distorted images, overview in Figure 13. Two types of features have been used in the method, namely, spatial and spectral entropies. The method involves four stages, i.e., pre-processing(down-sampling) of the distorted image, block-based partition and computation of spectral and spatial entropies for each block, percentile pooling from the two feature sets is done and lastly the quality score is computed. The metric is called Spatial-Spectral Entropy based Quality index SSEQ and have been reported to outperform FR SSIM and NR DIIVINE, BLINDS II and BIQI on LIVE and TID2008 database.
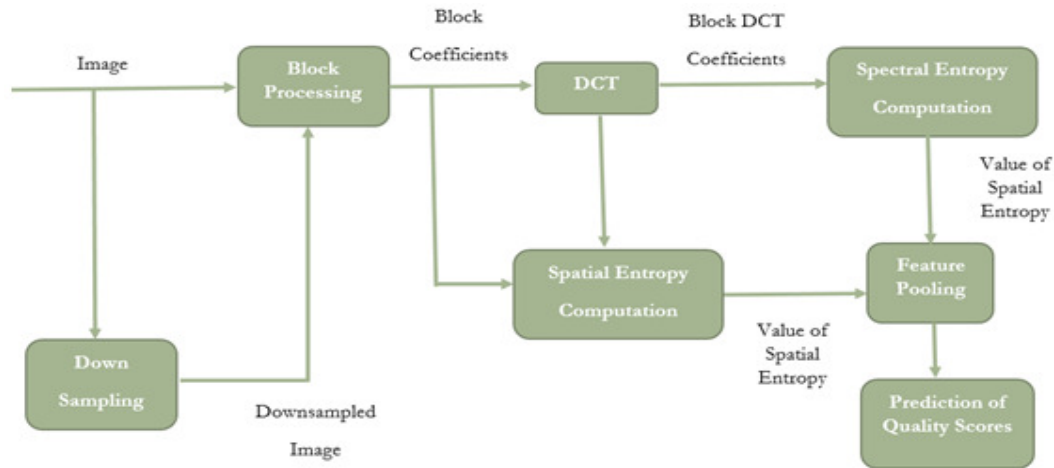

Figure 13: Framework for NR-IQA based on spatial and spectral entropies [32]

7. DISCUSSION

The problem of IQA is inherently hard. It has been dealt by using different tools. In this paper we have discussed entropy based methods, gradient based methods and DCT based methods. We have also shown a framework where deep learning has been used to deal with it. The fact that images from different domains have different properties. We represent those properties as features to a machine. Therefore, we cannot have universal set of features that could be applicable to all types of applications. State-of-the-art methods for the three types of IQA have been discussed previously. It shows that we need new IQA metrics that possess greater power than widely used PSNR and MSE.

8. CONCLUSIONS

We have reviewed some state-of-the-art approaches for the three types of IQA. Of the three types, NR IQA poses most challenges. As it requires no information from the reference image at the same time without strict assumptions regarding distortion level and types. Till date no system have been developed that meets the requirements of real-time quality prediction, data-base independence and generalized for all the distortion types and levels.

REFERENCES

[1] Lihuo He, Fei Gao, Weilong Hou, and Lei Hao. Objective image quality assessment: a survey. International Journal of Computer Mathematics, 91(11):2374–2388, 2014

[2] Ali Borji and Laurent Itti. State-of-the-art in visual attention modelling. IEEE transactions on pattern analysis and machine intelligence, 35(1):185–207, 2013

[3] Stanley A Klein, Thom Carney, Lauren Barghout-Stein, and Christopher W Tyler. Seven models of masking. In Electronic Imaging'97, pages 13–24. International Society for Optics and Photonics, 1997

[4] Qiang Li and Zhou Wang. Reduced-reference image quality assessment using divisive normalization-based image representation. Selected Topics in Signal Processing, IEEE Journal of, 3(2):202–211, 2009.

[5] Eric C Larson and Damon M Chandler. Most apparent distortion: full reference image quality assessment and the role of strategy. Journal of Electronic Imaging, 19(1):011006–011006, 2010

[6] Nikolay Ponomarenko, Lina Jin, Oleg Ieremeiev, Vladimir Lukin, Karen Egiazarian, Jaakko Astola, Benoit Vozel, Kacem Chehdi, Marco Carli, Federica Battisti, et al. Image database tid2013: Peculiarities, results and perspectives. Signal Processing: Image Communication, 30:57–77, 2015.

[7] Nikolay Ponomarenko, Vladimir Lukin, Alexander Zelensky, Karen Egiazarian, M Carli, and F Battisti. Tid2008-a database for evaluation of full-reference visual quality assessment metrics. Advances of Modern Radioelectronics, 10(4):30–45, 2009.

[8] Hamid Rahim Sheikh, Muhammad Farooq Sabir, and Alan Conrad Bovik. A statistical evaluation of recent full reference image quality assessment algorithms. Image Processing, IEEE Transactions on, 15(11):3440–3451, 2006

[9] Zhou Wang, Alan Conrad Bovik, Hamid Rahim Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. Image Processing, IEEE Transactions on, 13(4):600–612, 2004.

[10] Patrick Le Callet and Florent Autrusseau. Subjective quality assessment irccyn/ivc database, 2005. http://www.irccyn.ec-nantes.fr/ivcdb/

[11] Damon M Chandler and Sheila S Hemami. Vsnr: A wavelet-based visual signal-to-noise ratio for natural images. Image Processing, IEEE Transactions on, 16(9):2284–2298, 2007

[12] Ulrich Engelke, Maulana Kusuma, Hans-J¨urgen Zepernick, and Manora Caldera. Reduced-reference metric design for objective perceptual quality assessment in wireless imaging. Signal Processing: Image Communication, 24(7):525–547, 2009

[13] Soroosh Rezazadeh and St´ephane Coulombe. A novel discrete wavelet transform framework for full reference image quality assessment. Signal, Image and Video Processing, 7(3):559–573, 2013

[14] Christophe Charrier, Olivier L´ezoray, and Gilles Lebrun. Machine learning to design full-reference image quality assessment algorithm. Signal Processing: Image Communication, 27(3):209–219, 2012.

[15] Akshat Jain and Vikrant Bhateja. A full-reference image quality metric for objective evaluation in spatial domain. In Communication and Industrial Application (ICCIA), 2011 International Conference on, pages 1–5. IEEE, 2011.

[16] Feng Shao, Weisi Lin, Shanbo Gu, Gangyi Jiang, and Thambipillai Srikanthan. Perceptual full-reference quality assessment of stereoscopic images by considering binocular visual characteristics. Image Processing, IEEE Transactions on, 22(5):1940–1953, 2013

[17] Angela D Angelo, Li Zhaoping, and Mauro Barni. A full-reference quality metric for geometrically distorted images. Image Processing, IEEE Transactions on, 19(4):867–881, 2010.

[18] Decebal Constantin Mocanu, Georgios Exarchakos, Haitham Bou Ammar, and Antonio Liotta. Reduced reference image quality assessment via boltzmann machines. In Integrated Network Management (IM), 2015 IFIP/IEEE International Symposium on, pages 1278–1281. IEEE, 2015

[19] Rajiv Soundararajan and Alan C Bovik. Rred indices: Reduced reference entropic differencing for image quality assessment. Image Processing, IEEE Transactions on, 21(2):517–526, 2012

[20] Lin Ma, Songnan Li, Fan Zhang, and King Ngi Ngan. Reducedreference image quality assessment using reorganized dct-based image representation. Multimedia, IEEETransactionson,13(4):824–829,2011

[21] Xiaolin Chen, Shibao Zheng, and Rui Zhang. Reduced reference image quality assessment based on image statistics in pixel domain. In Advances on Digital Television and Wireless Multimedia Communications, pages 148–155. Springer, 2012.

[22] Abdul Rehman and Zhou Wang. Reduced-reference image quality assessment by structural similarity estimation. Image Processing, IEEE Transactions on, 21(8):3378–3389, 2012

[23] Zhichao Lin, Zhufeng Zheng, Ronghua Guo, and Liangfeng Pei. Reduced-reference image quality assessment based on phase information in complex wavelet domain. In Signal Processing (ICSP), 2014 12th International Conference on, pages 966–971. IEEE, 2014

[24] YazhongZhang, JinjianWu, GuangmingShi ,and XuemeiXie. Reduced reference image quality assessment based on entropy differences in dct domain. In Circuits and Systems (ISCAS), 2015 IEEE International Symposium on, pages 2796–2799. IEEE, 2015

[25] Ke Gu, Guangtao Zhai, Xiaokang Yang, Wenjun Zhang, and Longfei Liang. No-reference image quality assessment metric by combining free energy theory and structural degradation model. In 2013 IEEE International Conference on Multimedia and Expo (ICME), pages 1–6. IEEE, 2013

[26] Michele A Saad, Alan C Bovik, and Christophe Charrier. Dct statistics model-based blind image quality assessment. In 2011 18th IEEE International Conference on Image Processing, pages 3093–3096. IEEE, 2011

[27] Anush Krishna Moorthy and Alan Conrad Bovik. Blind image quality assessment: From natural scene statistics to perceptual quality. IEEE Transactions on Image Processing, 20(12):3350–3364, 2011.

[28] Yuming Li, Lai-Man Po, Xuyuan Xu, and Litong Feng. No-reference image quality assessment using statistical characterization in the shearlet domain. Signal Processing: Image Communication, 29(7):748–759, 2014

[29] Gitta Kutyniok, Wang-Q Lim, and Xiaosheng Zhuang. Digital shearlet transforms. In Shearlets, pages 239–282. Springer, 2012

[30] Peng Ye and David Doermann. No-reference image quality assessment using visual codebooks. Image Processing, IEEE Transactions on, 21(7):3129–3138, 2012

[31] Peng Ye, Jayant Kumar, Le Kang, and David Doermann. Real-time no-reference image quality assessment based on filter learning. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 987–994, 2013

[32] Lixiong Liu, Bao Liu, Hua Huang, and Alan Conrad Bovik. Noreference image quality assessment based on spatial and spectral entropies. Signal Processing: Image Communication, 29(8):856–863, 2014.

THE WEB-BASED EDUCATION JOURNEY: A CONSTANT LIFELINE

Vidhu Mitha

Department of Information Technology, SRM University, Chennai, India

ABSTRACT

E-learning has revolutionized our realm in more than just a listable number of ways. But it took a paradigm shift when it entered the threshold of the varsity system. With the prevailing spoon-feeding era, are the students really ¬industry ready? We answer that by confirming a fact: web-based learning has become the oxygen of freshers in the IT Industry instead of the traditional learning done through graduation. Furthermore, are university enforced e-learning assessment systems a true representation of a student's proficiency? This paper is a peep into what web-based e-learning systems are to a student of today's world, by giving an overview of university-level e-learning in India deploying an example from SRM University's organizational framework. It assesses a key e-learning trend, the implementation of which bridges the gap between universities and the industry. It is proposed to provide constructive feedback to the e-learning community and shine some light on areas of scope for future developments.

KEYWORDS

web-based learning, e-learning, university

1. INTRODUCTION

We see trends move toward a NO-PAPER system in education today. But what is the reality?

Ranging from high schools' e-labs, to finding it's traces in fully developed experts' CVs as nano degrees, e-learning has found its way in, everywhere including in India with an estimated $3 billion market. With the rapidly changing definition of e-learning today, it has swept countless fields and has a universal set of users starting from middle schoolers who've started submitting their assignments online, to corporate biggies that cannot function without it.

Gone are the days that a student accredited an institution's tutoring for his/her learning. Each student now has a pet go-to e-learning facility to thank for saving their tests and scores! This what tempts us to ask if this was used as a last-minute trick and if it helped? If yes, we will discuss if it is really a true representation of a student's skill as compared to a hardworking offline-study student.

But on the other hand, is this system being utilized to its maxed potential?

In universities, fully equipped hi-tech labs are available for training graduates. During the college years, we are taught the practice of multiple technologies, coding languages and its practical implementations. But the fact remains that over 90% of it, is learned online. Throughout this learning process, students are being asked to line up to a system bound by the university, to complete the courses they have defined for the them. This leads to the said underutilization of e-learning. While sticking to pre-defined university courseware, we lose track of the fast-paced industry's growing needs. This would lead us to the discussion on "bridging the gap".

2. THROUGH THE STUDENT LENS

On an average, a graduate level student uses books less than once a week. But online content is accessed at least 4-5 days a week and multiple times during a study day. The virtual notebook has taken over the blackboard. A few scenarios can help explain how:

When a student does not know a common definition or is unaware of the meaning of certain phrases or terminology used somewhere, he no longer looks for an encyclopedia, but looks up his phone's google dictionary instead. It is a lot faster than manual searching since it's just a few clicks away.

Another example could be when an assignment or project must be submitted. The student has many options open to him on the internet with an ocean of similar projects that can easily help set precedence to his project. There are websites which provide free expert reviews of student's projects and help them create it from scratch, from building blueprints onward and even for free. This was obviously much more economic compared to when we they had to purchase a book for the same purpose.

Yet another instance could be the paramount help e-learning is during competitive exams and university tests. Instead of spending many hours at tedious and monotonous lectures, there are interesting and fun ways of learning concepts via puzzles, questionnaires, online group activities etc. on the internet which have been proven to be more effective due to shortcut methods available, even if it is a last-minute preparation. They also have access to large databases of question papers to refer and study from. And based on each student's caliber, there are specialized test questions available. The mock tests online also help give them an approximate idea of their preparation level and an expected score range.

So, we can see that a student's learning curve is vague but still dependent on the method of e-learning that he chooses. To make a wise choice he must pick the method that is most comfortable to him and suited to his understanding and grasping power. The different choices he has include:

2.1. STATIC LEARNING

This is a one-way street where the student's only interaction with the web is to access the disseminated and stored information or content. This includes but is not limited to PowerPoint slides, E-books, question papers, videos, recorded lectures and related material. On picking up the relevant content from sources such as WhatsApp class groups, University-Notes portals etc., the student conducts domestic self-study with the help of these collectives.

2.2. DYNAMIC LEARNING

This usually consists of a regular online community that is balanced by two poles: The 'teacher' and the 'pupil'.

This is further divided into:

- Private tutoring: The teacher imparts knowledge and ideas to be learned and explains them to the pupil in a one-on-one session.

- Group E-Learn Meets: Two or more people who are well versed in a particular common area of interest get together and exchange ideas to help each other out. Both are equal participants in this exchange.

- The Online Classroom: A qualified individual schedules a class and delivers a lecture to an entire class of interested individuals live via Skype, Blackboard, FaceTime, Google Plus Hangouts, Udacity etc., and also conducts a questionnaire session for doubt discussions.

Most of the higher education institutes and universities in India are adopting the online classroom method. Students are getting on to the E-learning journey either by choice to skill up in new areas or as part of their university curriculum. Once he selects what is personally best for him, he is all set to go on his E-learning journey.

3. ALONG THE WAY

The student starts walking down his selected road, but he encounters a lot of deviations and obstacles.

With universities enforcing online tests and online submissions as a focal part of education, there is the rising threat to authenticity. While e-learning enabled students to work time-effectively, it still had its shortcomings with respect to college level testing. Students could always duplicate content from online easily using plugins for every type of testing software. This is majorly due to the general technical drawback of the online world: there was always a way to break in, hack and alter the functioning of most systems easily. This has made it increasingly easy for students to indulge in malpractice since every step to performing such unethical hacks are present openly online paradoxically.

So, in an era where cybercrimes are a matter of fact, online testing might not be a completely apt representation of a student's true potential and proficiency.

But since it's our best bet for development and still more effective than traditional learning, we move on to analyzing it further. At university level, to be successful at web learning, there are multiple factors that contribute apart from the students learning capacity and tutor/instructor:

- The student's attitude.

- The content organization or knowledge base.

- An interactive tool or the web-based platform used as media for communication, which is more important than the content itself.

Many universities are launching distance education programs using web-based learning and online libraries. With India being the current third largest online market for education in the world, it has launched NPTEL-a Govt. initiative and website where regular courses from reputed Indian institutions hosted for free. One such similar but private e-learning instance in India is eLab.

3.1. THE REVIEW OF eLAB

SRM university's eLab is an auto-evaluation tool for learning programming. eLab helps learners to practice and acquire programming skills. Sophomore students are warmed up to the new coding languages they would be learning through a semester using this tool. With a step-by-step approach, it levels up in competitiveness with every question. There are several sections present in the tool, to cater to all the subjects the student will be completing in the current semester. Each section within a subject would pertain to a chapter of the subject. Thus, all areas of the student's study, would be thoroughly tested. Each student is required to complete a set of programs whose problem statements will be defined to them along with example test cases (as shown in the screenshot below). The student is required to analyze the question and code to satisfy the test cases. This system executes the written code and returns a result or highlights the issues with the code if insufficient. The system is also designed to award a completion score in percentage. If the code was ideal and satisfied all conditions, it recorded a 100%. If the code worked in favor of the question but would not satisfy all conditions, it recorded a 33,50 or 75 percent based on how much improvement the code needed. Finally, an average of all these scores makes the final cut to be added to the student's grade point average(GPA).

Initially the access to this tool was limited to devices connected via the university's private network. Now it has been made available for use from any network. So now students can use it from anywhere including home, to complete their graded college work and submit it as well. But the list of pros doesn't stop there. Teachers and respective faculty gain well from this tool too. They can check the student's progress at any given time, review & reframe the question set, and download a progress report for all students that registered under them.

The tool's access specifications are sealed in the backend, due to which the students cannot change of edit their grades or scores. Additionally, since the same software is used for multiple subjects, it works well to satisfy the university's needs while constantly acting as an effective learning tool for the student learners. Though the tool has a very good infrastructure for web learning, the content must match real-world problems, else it would defeat the purpose of learning programming. With students dealing with university enforced tools with a lot of loopholes, they lag behind when they are sent into the industry
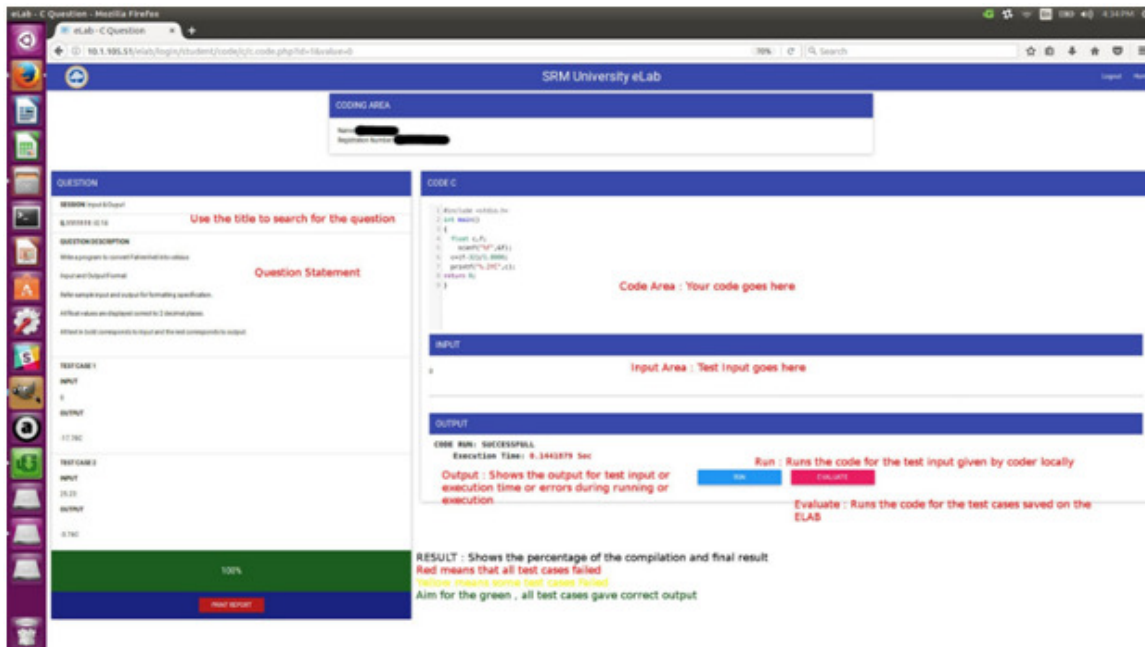
Figure 1.The eLab user interface

4. ZOOMING IN ON THE GAP

With the advent of e-learning in colleges, the universities curriculum started to get greased and easy. But this was only one side of the learning curve because on the other end of the curve, were graduates passing out of the universities with no knowledge of the industry whatsoever. They were incomplete products whom the companies seem to devour and shape to suit their specific needs.

According to HR reports, around 1.5 million engineers are released into the job/working market every year. But only about 25% of them are directly employable from the college campus. So, this goes to say that employers today look for more than just the basic skill set in their employees, they look for a long-term relationship with the organization. Which is why they focus on reskilling. Graduate students are required to unlearn and learn to fit into the IT world. So, the educators need to start thinking ahead of the curve.

This is where companies like Embibe and Simplilearn enter the picture with their courses for mid-level professionals. They help create a smooth segway from college to industry. Many businesses empires have been built around helping this university-to-company transition: Coursera, Udacity, edX etc. These e-learning websites primarily focus on delivering certified industry level courses which are now even being prescribed by companies itself! MNCs like Google, Microsoft and IBM now ask their employees to refer to such common e-learning portals instead of conducting intense and tedious training programs. A Brandon-Hall Study proved that learning online required almost 40% to 60% less employee time than learning the same topic in a traditional classroom setup. For instance, when IBM implemented an e-learning program in their company, its employees learned five times more without increasing their time spent in training.

When the freshers felt unprepared for the market, they ran to these web-based systems for a revival. On completion of these online courses from Coursera etc., they would've added jewels of value to their CVs which immediately helped them get hired and have also eased their transition into the walls of the industry since companies preferred such candidates over newer unprepared college minds.

4.1. HOW THE GAP COULD BE HANDLED

- Course content for universities should start to be: prescribed by or delivered directly from industry partners or companies, as the constant updates that are done by college seem insufficient. Therefore, the responsibility comes upon the system to update the syllabus and the pedagogy based on the society's demands.

- Universities should use professional content developers who have the industry experience.

- Web-based learning systems should consider using Artificial Intelligence. The tools should be smart and more interactive in nature. It should learn by itself and adjust the complexity of the delivered course based on the interaction and feedback of the students.

- The evaluation tools should predict the student understanding level and adjust the complexity level and provide the set of questions dynamically based on their response and award accordingly.

- Colleges should create space for students with a more direct inclination to the industry and inculcate that as part of their graded courseware, as outcome-driven academic systems flourish better.

- More web-based courses should be introduced into university curriculum to help industry freshers since they come with no experience of the workplace pressure, workload volumes or co-ordination required at companies at a global level. They need to be thought how to unlearn what they have collected in domestic college environments and learn what real-time systems require.

5. IN CONCLUSION

With the rigorously increasing E-learning appetite, the education system has improved by leaps and bounds as compared to what it was in the past decade. But ever since the comfort of studying on your own couch was introduced, the lazy students sense the main turbulence of studying at their own pace - the procrastination. Students need to have a highly motivated self-learning drive since web-based learning eliminates travel, timings and sometimes deadlines too and makes it all too flexible. But we cannot deny the fact that both the economy and literacy levels have gone up since the advent of web-based e-learning. Furthermore, tailored courses with flexible completion timing do enhance majority of students' inclination toward learning. The stepping stool into the IT industry has become much larger with e-learning. More and more companies are starting to recognize this trend. It is due to the abstraction feature of web-learning that, from a sea of unrelated information we get straightforward and tailored content delivered to us for learning. Thus, employees are now able to deliver better and effective results at within shorter time spans.

So, web-learning has contributed to nothing but good to the technical revolution. But with just a few more changes and guideline enforcements, in the foreseeable future, e-learning would have revamped the book era entirely and become the new worldwide education standard. Universities would be technology enabled with a much leaner infrastructure. And in this changed paradigm, the learner is the new center of the knowledge-universe with his personalized learning path.

REFERENCES

[1]   Noman Hasany, (2017) "E-Learning Student Assistance Model For The First Computer Programming Course", International Journal on Integrating Technology in Education (IJITE) Vol.6, No.1, March 2017.

[2]   PengiranShaiffadzillahPengiranOmarali,(2017) "Designing A Survey Study To Measure The Diversity Of Digital Learners", International Journal on Integrating Technology in Education (IJITE) Vol.6, No.1, March 2017.

[3]   Prof (Dr) Rajan Saxena, (2016) "Future of jobs and its implications on Indian higher education".

[4]   Othman, A. A., Pislaru, C., &Impes, A. M. (2014). Improving the quality of technology-enhanced learning for computer programming courses. International Journal of Information and Education Technology, 4(1), 83.

[5]   Shipra Sharma and Shalini Garg, (2016) "Web based learning: A research on assessing its effectiveness in contrast to traditional classroom learning environment", Web based learning- IEEE 2016:International Conference on Computational Techniques in Information and Communication Technologies (ICCTICT)

[6]   Joel A. Mejia, Wade Goodridge and Christopher Green, (2015) "Using web-based learning logs to analyse students' conceptual understanding of truss analysis in an engineering statics course", IEEE 2014: Frontiers in Education Conference (FIE).

[7]   Anil Agarwal, Web-Based Learning, and Teaching Technologies: Opportunities and Challenges, Idea Group Publishing.

[8]   Ulrik Schroeder, (2008) "Web-Based Learning-Yes We Can", Advances in Web-Based Learning - ICWL 2009: 8th International Conference, ppl 25-33.

AUTHOR

Vidhu Mitha Goutham is a second-year student pursuing Engineering in Information Technology at SRM University in Chennai, India with a keen interest in voicing to the online community.

INTENTIONAL BLANK

IMPROVED HANDOVER ALGORITHM FOR PROXY MIPv6 BASED ON AAA SERVER

Hewei Yu[1] and Meiling Zhou[2]

[1]School of Computer Science & Engineering, South China University of Technology, Guang Zhou, P.R.China
[2]GF Fund Management Co. Ltd., Guang Zhou, P.R.China

ABSTRACT

This paper proposes an improved handover algorithm which does not need authenticating again if Mobile Node moves within the same Proxy MIPv6 domain. When MN enters PMIPv6 domain at the first time, it needs to make an AAA authentication. But when MN moves between MAGs in the same domain, it can perform handover procedure without the second times of AAA authentication, and speed the handover process. We built a structure of PMIPv6 including AAA server on NS-2, and proposed an improved handover algorithm for PMIPv6 based on AAA authentication. The simulation results show that the new scheme can effectively reduce the handover latency and ratio of packet loss, and improve network performance.

KEYWORDS

PMIPv6; AAA; handover; fast handover

1. INTRODUCTION

Nowadays, the huge commercial demand for mobile Internet has brought new opportunity and challenge of technology. In 2008, the Internet Engineering Task Force (IETF) proposed a network-based mobility management protocol called Proxy MIPv6[1]. Because of the bright prospect of PMIPv6, researchers all over the world have carried out extensive and in-depth study on PMIPv6, and provides various effective handoff scheme. As a whole, the research on PMIPv6 has many aspects, including security, handover performance etc. On the security mechanism of PMIPv6, RFC5779[2] defined AAA for authentication ,authorization and accounting. Document[3] designed and implemented the authentication and authorization of PMIPv6 network system, the testing results proved the validity and feasibility of the certification system. On the handover performance of PMIPv6, document[4] proposed a seamless handover scheme using the neighbor discovery message of IPv6 to reduce the handover latency and packet buffering at the Mobile Access Gateway (MAG) to avoid the on-the-fly packet loss during a handover for PMIPv6. However, it is unable to satisfy the real-time demand of higher business communication. Document[5] proposes a novel low latency handover scheme for PMIPv6 using Media Independent Handover (MIH) services. Document[6] tested the handover performance of PMIPv6 with and without the IEEE 802.21 MIH services in a heterogeneous wireless networks' environment. The results showed that the handover performance had improved if the IEEE 802.21 MIH services is used, while there have some problem such as Signal overload. Of course, A scheme to Reduce Packet Loss during PMIPv6 Handover is called PL-PMIPv6, which is similar to the fast handover for MIPv6, and FH-PMIPv6 was improved based on it .

In the PMIPv6 system include AAA severs, when MN enters PMIPv6 domain at the first time, it is required to be authenticated. This paper proposes a scheme, MN only need to be authenticated once in the same domain. In another word, when MN moves into a new MAG in the same domain, it does not need to be authenticated again, so this schema can minimize the authentication latency and packet loss.

The rest of this paper is arranged as follows: the implementation of AAA server for PMIPv6 is introduced in section 2. Section 3 introduces the improved handover schema to minimize typical authentication procedures. Simulation and results analysis are provided in section 4. This paper is concluded in section 5.

2. IMPLEMENTATION OF AAA SERVER FOR PMIPV6 [7]

PMIPv6 introduces two network entities: MAG (Mobile Access Gateway) and LMA(Local Mobility Anchor). LMA is the home agent for the MN in a PMIPv6 domain. It is the topological anchor point for the MN's home network prefix(es) and is the entity that manages the MN's binding state. MAG is a function on an access router that manages the mobility-related signalling. It is responsible for tracking the MN's movements to and from the access link and for signalling the MN's local mobility anchor.

In this paper, the functions of AAA server are implemented in NS2 using C++ language and TCL scripts. We add two structures that is AAA server and list of policy profile. Figure1. (a) and Figure1. (b) show the working flow chart of AAA server.

After AAA server is implemented, Figure.2 shows the signal flow when MN moves from P-MAG(pre-MAG) to N-MAG (new-MAG) in the same PMIPv6 domain.
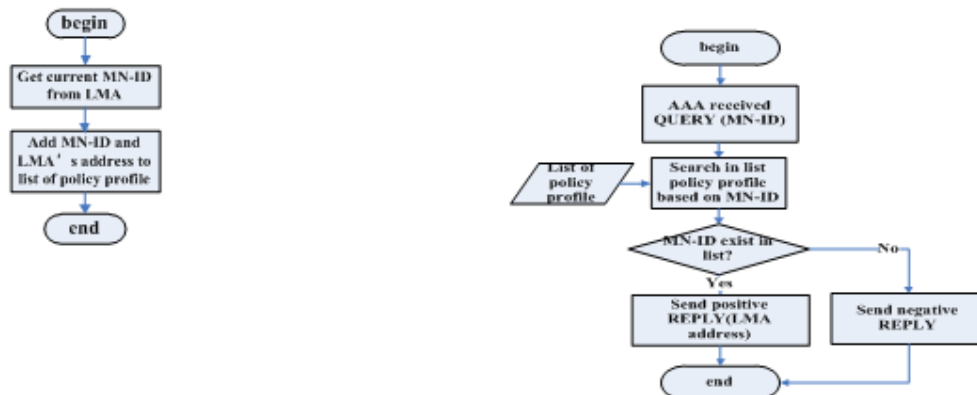

Figure 1. (a) Create list of policy profile   (b) Judging the legality of MN

The whole handover procedure goes as follows:

1) P-MAG sends DeReg PBU message (including MN-HNP and Proxy-CoA) to LMA when it detects that MN is going to leave. Then LMA can release the binding between MN-HNP and Proxy-CoA of MN.

2) LMA starts the delay timer immediately when it receivs DeReg PBU message. So LMA allows N-MAG to wait for a period of time before updating the binding cache entry, and replies DeReg

PBA message to P-MAG simultaneously.

3) P-MAG receives DeReg PBA message.

4) When N-MAG detects the attachment of MN, N-MAG obtains information about MN. Then MN sends RS message to N-MAG.

5) N-MAG sends QUERY message to AAA server after received RS message, then gets policy profile of MN.

6) After MN's successful access authentication, N-MAG sends PBU message to LMA to update the LMA about the current location of MN.

7) Upon receiving the PBU message, LMA updates binding cache entry. LMA sends PBA message including MN-HNP.

8) After receiving the PBA message, N-MAG sets up a tunnel to LMA and adds a default route over tunnel to LMA. N-MAG then sends RA message to MN on the access link to advertise the MN-HNP.

9) When MN receives RA and know that there are no change on L3 interface after checking, it keeps the same IP address and uses it for packet delivery.
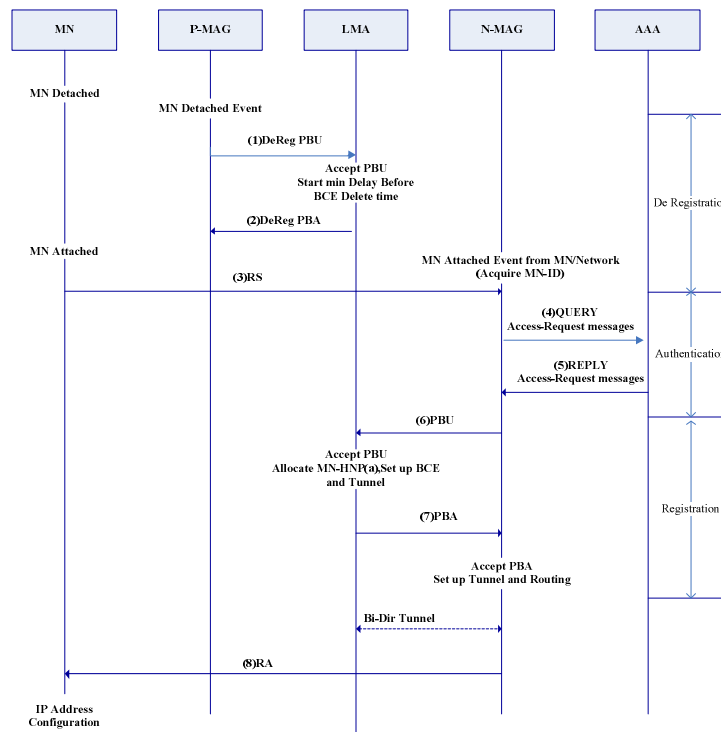
Figure 2. Signal flow of Proxy MIPv6 based on Inter-Domain Handover

As signal flow chart shows, the handover schemes for PMIPv6 networks includes deregistration (1)-3)),authentication(4)-5)) and registration(6)-9)) phases.

3. IMPROVED HANDOVER SCHEME

In order to reduce the longer latency and extra cost caused by combining the AAA function while guaranteeing the handover performance, some handover procedure have been proposed as follows: in document[8], the AAA server selects the most suitable LMA based on the load information received from the LMAs. Moreover, the LMA performs admission control based on the number of current mobile nodes registered to the LMA. Document[9] proposed adaptive context transfer schemes for a fast handoff in Proxy MIPv6 which reduces the latency in AAA authentication and specifies context transfer scenarios taking into account proactive and reactive handover. When MNs attempt to handover inter/intra domain, the AAA context information stored in LMAs and MAGs will be used to support the handover without visiting the AAA server. Other documents put forward an configuration file based intra-domain handover method in PMIPv6. Optimize the intra-domain handover performance by expanding the contents of the configuration file and adding the proxy care-of address option. Of course, there are many improved schemes based on AAA server. In this section we discuss our proposed scheme to reduce the handover latency of Proxy MIPv6 which is based on intra domain.

As we know, handover latency in PMIPv6 is caused due to following four latencies[10]:

- ■ Link switching latency

- ■ AAA authentication latency

- ■ MN registration latency

- ■ Latency due to Router Solicitation(RS) message and Router Advertisement(RA) message

In the proposed scheme, the handover latency is reduced due to AAA authentication again in the same domain.

As mentioned previously, the MAG detects the movements of MN and performs mobility-related signalling with the LMA in place of the MN. As Fig3 shows, the serving MAG(i.e.,P-MAG) will send handover initial (HI) message to the target MAG(i.e.,N-MAG) when MN is going to leave. Then N-MAG will be told the MN has been authentication. First ,N-MAG stores information (such as the AAA initiation information of MN(i.e.,MN-ID) and the address of LMA) from HI message. Then N-MAG sends back a handover acknowledge(HACK) message to P-MAG. During this time, MN has not disconnected from P-MAG, so all the data packets are forwarded through P-MAG as before.

P-MAG sends DeReg PBU message to LMA while detects MN is disconnect. The deregistration period is the same as mentioned previously.

When N-MAG detects that MN enters or gets the Route Solicitation(RS) message from MN, it will check whether the MN's information (including MN-ID and LMAA) have already existed. If N-MAG has MN's information, it means that MN has been authenticated. So it is not need to registration again. Here we give some analysis about handover latency. Usually handover latency is marked by the maximum time interval between MN receives the last packet from the P-MAG and the first packet from N-MAG. Using some marks such as T_{DeReg} (deregistration) ,T_{Auth} (authentication), T_{Reg}(registration), $T_{MN\text{-}MAG}$ (MN and MAG), $T_{MAG\text{-}LMA}$ (MAG and LMA), $T_{MAG\text{-}}$

$_{AAA}$(MAG and AAA) and T_{L2} (layer2 handover), we can get handover latency (HL) of traditional algorithm according to Fig4.1.
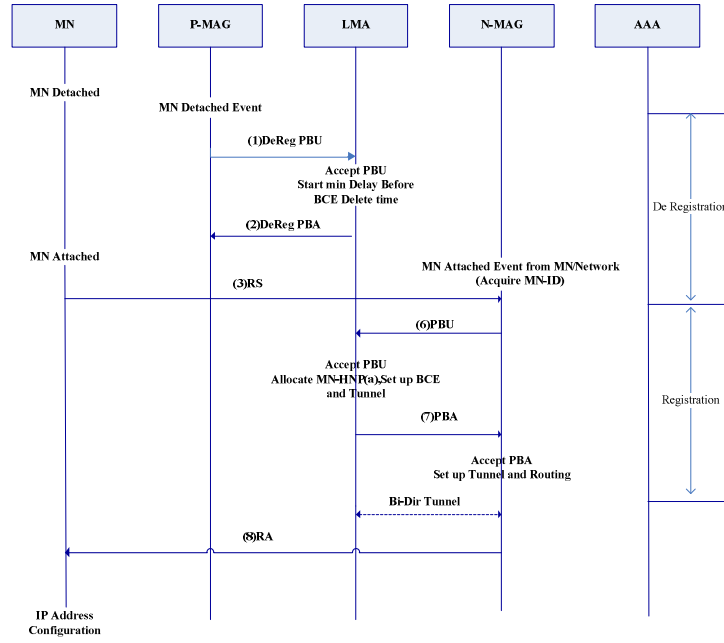

Figure 3. Signal flow of proposed Proxy MIPv6 based on intra-domain.

$$HL= T_{DeReg} +T_{L2} +T_{RS} + T_{Auth} + T_{Reg} + T_{RA}$$

$$= 4T_{MAG,LMA}+T_{L2} +2T_{MN, MAG} +2T_{MAG,AAA} \qquad (1)$$

The handover latency (HL) of proposed algorithm is as the following according to Figure 4 (b).

$$HL= T_{DeReg} +T_{L2} +T_{RS} + T_{Reg} + T_{RA}$$

$$= 4T_{MAG,LMA}+T_{L2}+2T_{MN,MAG} \qquad (2)$$

Compare equation (1) and (2), you will find that the delay of AAA authentication is eliminated in our algorithm.
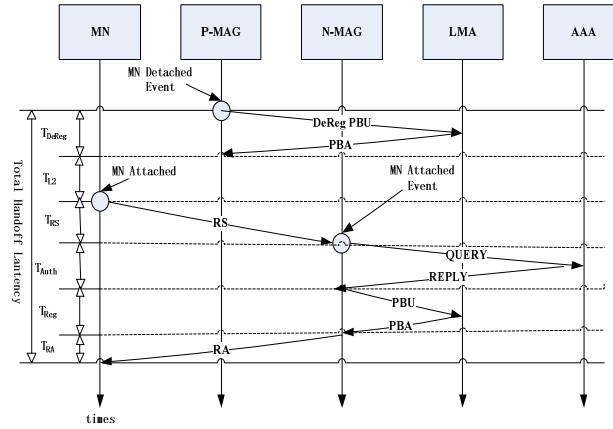
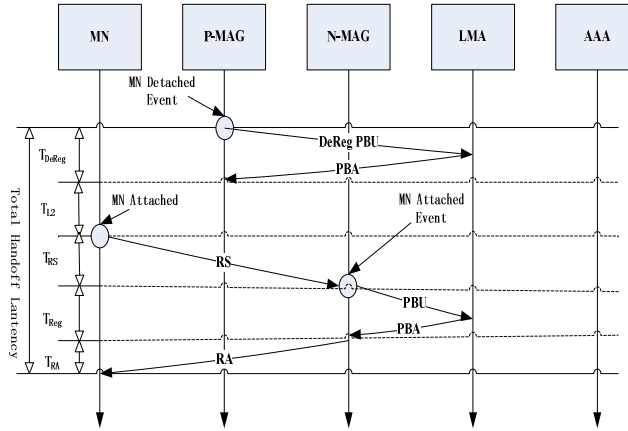Figure 4. (a) Handover latency of PMIPv6

Figure 4. (b) Handover latency of proposed PMIPv6

4. SIMULATION AND ANALYSIS

We have taken simulation using NS-2 and compared the performance of the proposed scheme with the previous and the ordinary one when MN moves between MAGs in the same domain. In the simulation, we assumed correspondent node(CN) sent packets to MN and MN moved form P-MAG to N-MAG , the packets from CN get to MAG through LMA and router. The simulation model is shown in Figure 5. The time in the simulation is controlled in 20 seconds. CN sent data packets at intervals of 0.05s after 1.0 second. At 14.4 second, MN moved away from P-MAG and went toward N-MAG.
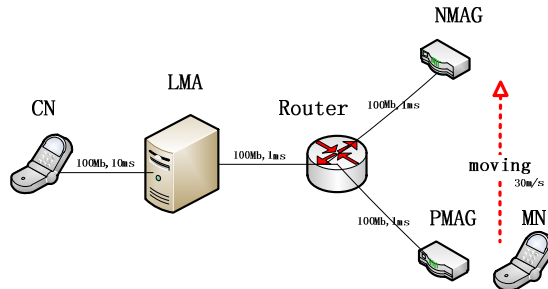

Figure 5. Practical simulation model

2.1. Handover latency comparison

The handover latency comparison between ordinary (marked by black) and proposed (marked by red) scheme is show in Fig6. The abscissa represents the simulation time, the ordinate represents handover latency. Because of MN has not entered PMIPv6 network at the starting time, there are not handover latency and the delay time is 0s. The packets transfer delay is 0.05s. There are two high vertical line, the first line represents the latency that MN enters a new Proxy MIPv6   domain and attaches to P-MAG. The graph shows, handover latency is sharp increased up to 1.02s at 1s. It is the same as the ordinary scheme. However, there are visible difference in second line. The second line represents the latency that MN enters N-MAG. The handover latency of ordinary scheme is sharp increased up to 0.5s while it is 0.45s in proposed scheme, which shows that the proposed scheme can reduce the handover latency.
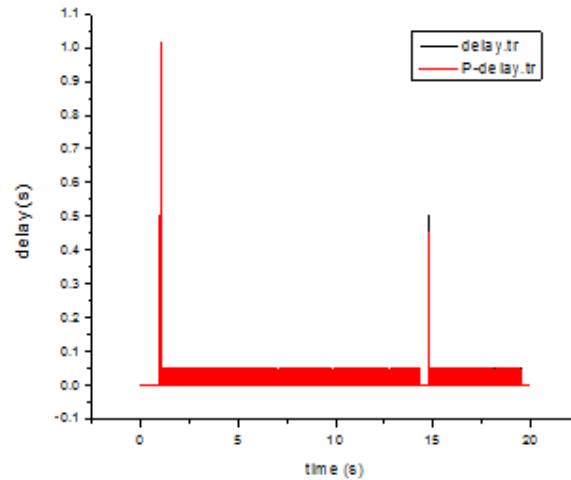

Figure 6. Handover latency comparison between ordinary and proposed scheme

Table 1 shows some parameters about handover latency in the simulation.

Table 1. Some parameters in the simulation

Scheme	Start time of handover(s)	Finish time of handover(s)	Handover latency(s)
Previous scheme	14.4	14.9	0.5
Proposed scheme	14.4	14.85	0.45

Both of the start time of handover of the two scheme are 14.4s. The proposed scheme has finished handover at 14.85s, and the ordinary one has finished at 14.9s.

Handover latency of the ordinary scheme =0.5s

Handover latency of the proposed scheme =0.45s

Obviously, the handover latency of the proposed scheme has been reduced 0.05s (0.5-0.45=0.05s). It is decreased by 10% ((0.05/0.5)*100%=10%) according to our proposal.

2.2. Packet loss rate

Figure 7 shows the number of packet loss comparison of ordinary and proposed scheme. In this simulation , ordinary scheme lost 9 packets. As while as the proposed scheme lost 8 which is less than the ordinary one.

In the simulation, the largest number of data packet is 383 in two scheme by checking the simulation output trace file. So we can get table2.
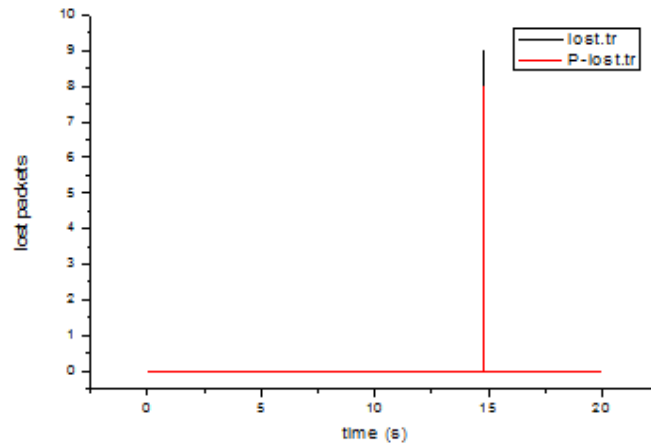

Figure 7. Number of packet loss of ordinary and proposed scheme

All of the forward data packets are 383.

The number of packet loss are 9 and 8 in ordinary and proposed scheme respectively.

Obviously, the ratio of packet loss of our proposed scheme has been reduced 1(9-8=1). So the ratio of packet loss is decreased by 11.11% ((1/9)*100%≈11.11%) according to our proposal.

To sum up, the proposed scheme is characterized by completed N-MAG authentication in advance of handover, so there is no need of authentication again. It is clear from the result of simulation, the proposed scheme reduces the handover latency and the ratio of packet loss without increased throughput.

Table 2. packet loss comparison of ordinary and proposed scheme

Scheme	The number of loss	The number of forward	Packet loss ratio
Previous scheme	9	383	≈2.35%
Proposed scheme	8	383	≈2.09%

5. CONCLUSIONS

In this paper, we introduce the present development of PMIPv6 and the way to add AAA server to PMIPv6 in NS-2. What's more, we have proposed an improved handover scheme which can minimize handover latency and ratio of packet loss by omitting operation of AAA authentication again in the same domain. We evaluate the performance of both the ordinary and the proposed scheme in NS-2 using a PMIPv6 system model. The simulation results show that our proposed scheme can effectively reduce the handover latency and ratio of packet loss. However, there are some drawbacks, for example, it is only satiable for intra-domain handover, the security issues is not perfect etc. Further work should include the improvement and simulation jobs aimed at the problems above.

REFERENCES

[1]   S. Gundavelli (Ed.), K. Leung, V. Devarapalli, K. Chowdhury and B. Patil, 'Proxy Mobile IPv6', Internet Engineering Task Force RFC 5213, August 2008.

[2]   J. Korhonen, Ed.Diameter Support for Proxy Mobile IPv6[S].IETF RFC 5779,2010.

[3]   LV Jianhua, ZHOU Huachun, QIU Luwei,et al. Implementation of authentication and authorization of proxy mobile IPv6. Computer Engineering and Application, 2012,48(22):109-116.

[4]   Ju Eun Kang, Dong Won Kun, Yang Li, You Ze Cho. Seamless Handover Scheme for Proxy Mobile IPv6[J]. IEEE Computer Society, 2008,10.1109/WiMob:410-414.

[5]   Igor Kim, Young Chul Jung, Yong Tak Kim. Low Latency Proactive Handover Scheme for Proxy MIPv6 with MIN[J],SpringerLink, 2008, Volume 5297:344-353.

[6]   Linoh A.Magagula, Olabisi E.Falowo, H.Anthony Chan.Handover Optimization in Heterogeneous Wireless Networks: PMIPv6 vs. PMIPv6 with MIH [C] .Southern Africa Telecommunication Networks and Applications Conference, 2009.

[7]   Nitesh M. Tarbani, B.R. Chandavarkar. Implementation of AAA Server for PMIPv6 in NS-2[J]. PDCTA,2011. IEEE, 2011.

[8]   Viswanathan M, Yi M K, Yun S Y, et al. A novel local mobility anchor selection scheme for proxy mobile IPv6 networks[C], Proceedings of the 6th International Conference on Ubiquitous Information Management and Communication. ACM, 2012: 49.

[9]   Baek J J, Song J S. Adaptive context transfer scheme for fast handoff in Proxy Mobile IPv6[C], Next Generation Mobile Applications, Services and Technologies, 2008. NGMAST'08. The Second International Conference on. IEEE, 2008: 127-130.

[10]   Banerjee K, Tahasin Z I, Uddin R. An Efficient Handover Scheme for PMIPv6 in IEEE 802.16/WiMAX Network[J]. International Journal of Electrical & Computer Sciences, 2011, 11(5)

AUTHORS

Hewei Yu

Professor of South China University of Technology. His research interests focus on integration technology of heterogeneous wireless networks. He directed several projects supported by the National Natural Science Foundation of China, Science & Technology projects of Guangdong Province, and Guangzhou City.

Meiling Zhou

Got her Master degree from South China University of Technology. Her research direction is handoff technology of Proxy Mobile IPv6. Now she is working in the GF Fund Management Co. Ltd.

INTENTIONAL BLANK

ANALYSIS OF MOBILE PAYMENT INFLUENCING FACTORS

Sitalakshmi Venkatraman

School of Engineering, Construction and Design (IT),
Melbourne Polytechnic, Australia

ABSTRACT

With the exponential proliferation of mobile devices in the consumer market, wireless e-business is emerging as a key area to revolutionise industries. In the past few years, industry has witnessed an increase in the adoption of mobile payment and billing methods that leverage on wireless technologies. Yet, the success of mobile payments in businesses much depends on many factors such as, type of wireless technologies used, security options available, the players involved and their influencing m-business models. This paper examines mobile payments in both technical as well as business perspectives. It identifies and analyses the influencing factors from multi-dimensions that would be useful for adopting mobile payments.

KEYWORDS

Mobile payments, Wireless technologies, Business models. M-commerce

1. INTRODUCTION

Wireless e-business termed as m-commerce has the main advantage of portability with ubiquitous business transactions that can be performed anywhere and anytime [1]. In addition to the key mobile features of location-based and personalised information services, payments embedded in mobile apps or websites provide value-added facility that aid in the business agility of organisations in this competitive business environment [2][3]. With recent developments in wireless communication technologies using 3G evolved to 4G networks, mobile Internet and multimedia streaming have gained popularity and provide increased channels of interaction with customers [4][5]. Hence, organisations expect that m-commerce provides a cost-effective improvement in their operational efficiency, resulting in an increase in profitability. Businesses have started to consider mobile payments as value-added services providing opportunities for their marketing, promotion and branding [6][7].

Apart from business perspectives, looking at consumer trend in the last decade, most consumers have been using mobile phones beyond the traditional services of voice calls and SMS. By the end of 2016, two thirds of the world's population acquired a mobile subscription with a surge in mobile broadband adoption via smart phones [8]. Mobile devices now account for 60% of the connections with new services and dedicated apps leveraging on mobile internet that could make common man's daily activities quite convenient. Hence, consumers' expectations from these smart devices have risen dramatically and mobile payment services are now being recognised to be important. Many consumers from countries such as, Finland, UK, North America, Japan, Korea and Singapore started to adopt mobile payments more than a decade back [9][10] as their preferred mode of cashless transactions in their everyday lives - while transacting with vending

machines, online ticketing agencies, public transport systems and merchant stores. A recent survey in UK indicates that 56% of smart phone owners purchased a product using a mobile app transaction in 2015 and mobile banking transactions via apps have now been envisaged by consumers [11]. However, global reports indicate that mobile revenue growth is slow [8][12]. There is lack of research in analysing this situation from current technological and business perspectives and such investigations could have a positive influence on mobile payment adoption [7][13]. This provides the key motivation for the study reported here.

This paper is aimed at providing an analysis of the trends in mobile payments, from technology, business, as well as consumer perspectives for analysing the influencing factors. The remainder of the paper is organised as follows. Section 2 describes the need for the study. The influencing factors from various facets of mobile payments are identified and analysed in section 3. Finally, concluding remarks are provided in section 4.

2. THE NEED FOR THE STUDY

Despite potential advantages of m-commerce], mobile payments face major challenges as financial transactions via mobile devices demands a multi-faceted approach involving technology innovations, apps and service offerings as well as global business trends [14][15]. In addition, with the emergence of player-centric mobile payment services, new mobile payment business models are being introduced [16]. Both remote and proximity type of mobile payment systems have witnessed evolution of technologies such as Infrared (IR) and Bluetooth combined with services such as Interactive Voice Response (IVR), Short Message Service (SMS), Radio Frequency Identification (RFID)/2D bar codes, Wireless application Protocol (WAP), contactless interface to chip-enabled payment technology, Near-Field Communication (NFC) enabled mobile phones, contactless stickers, tags, mobile browsers, mobile apps and m-wallets [5][17].

With the existing plethora of wireless technologies, both merchants and consumers are clouded with a range of options available to them for adopting mobile payments and are not sure if they are taking the road to success or failure [4][7]. In the past decade, mobile payments were used for predominantly non-banking transactions with several failed initiatives such as Dencoin and Google Wallet even in developed countries [18][19].

On one hand, the banking sector is gradually embracing digital payments capitalising on the IT players such as Apple and Google to compete in the market that may shape its future [12]. On the other hand, the success or failure of mobile payments is based on the consumer acceptance [13]. Recent studies conducted worldwide indicate that mobile payment growth has been slower than anticipated [8][12]. This is because, mobile payments are not just a new technology added on to existing consumer networks, but actually have their own set of historical and normative roots in development practices with an emphasis to be given on market subjects and consumer subjects [7]. Previous studies have predominantly focussed on the technology and payment options available to a consumer based on factors such as location (remote/proximity), payment value (micro/macro amount), and charging method (post-paid/pre-paid), and the associated security protocols [20][21][22]. There is lack of literature that analyses the current multi-faceted dimensions of mobile payments [23][24][25]. There is little guidance available for merchants and consumers to review the various mobile payment options and to adopt those mobile payment methods that suit their requirements.

Due to the paucity of research in this area, this paper analyses various influencing factors for adopting mobile payments. This would help businesses and consumers to make a more systematic review before they embark on the right kind of mobile payments that would add value to their everyday activities. This paper takes a modest step in this direction.

3. INFLUENCING FACTORS FOR MOBILE PAYMENTS

Although some industries such as, retail, entertainment and banking have been successfully adopting mobile options for billing and payment in certain countries, many industries are still cautious in offering mobile payment methods for their customers. This is partly because there are several issues related to mobile payments in an enterprise. On analysing the influencing factors, these could be broadly classified under three main dimensions:

- mobile payment players,
- payment methods and
- enabling wireless technologies

Mobile Payment Players

There are many parties or players contributing to the mobile payment value chain.   In order to access a payment service, a user or subscriber initially has to come into an agreement with a network provider for accessing telecommunication and the payment services provided by the specific operator. If mobile internet is used, an additional contract between a subscriber and an ISP is required. However, with the introduction of 3G/4G technologies, mobile internet could be accessed directly through their mobile operator. Recent evolutions have enabled users to adopt new mobile payment and billing models for accessing various networks and services independent of the network provider [5][26].   In Table 1, we have summarised a list of players involved in influencing mobile payments in businesses.

Table 1. Mobile Payment Players and Influencing Factors

Mobile Payment Players	Influencing Factors
Equipment / device manufacturers	Device usability, features & constraints, Wireless technology, Native software, use of third party H/W and S/W
Network service providers	Telecommunication infrastructure, Billing models and contracts, Integration with payment service providers, QoS, Network interoperability and Security
Financial service providers	Types of e-cash systems, Credit systems, Security and Payment settlement
Payment service providers	Interfacing (S/W, H/W and other players), Security and Cost
Market regulators and consortiums	Governing laws (bank, government and other agencies), Rights protection, Price regulations and Tax policies

The abovementioned players need not be exactly mapped onto various business parties but could be an aggregation of business entities.   For instance, the network service provider could be aggregating the services of mobile network operators, content providers and Internet service providers. Alternatively, there are cases where the network operator covers the functionality of an Internet service provider, while the content provider caters to payment services as well. Proprietary mobile payment solutions are now being introduced as issuer wallets to compete with device manufacturers such as Apple, Samsung, etc. Hence, an evolution of such industry players have a major influence on the consumer adoption of mobile payments.

Payment Methods

Though mobile payment services were traditionally managed by mobile network operators for many years using player-centric models, recently we find the emergence of new payment and billing models that allow consumers to access the payment services independent of who the

owner of the service is [13][27].   The issue of mobile payment adoption is further compounded by the types of payment methods supported here. Commonly used payment methods include pre-paid cash card, micro-billing, person-to-person (P2P) and credit card payments. Some of the main factors that influence these payment methods in m-commerce are identified in Table 2.

Table 2. Payment Methods and Influencing Factors

Payment Methods	Influencing Factors
Pre-paid cash card	Small transaction amount, Top-up methods and Availability to anyone
Micro-billing	Billing options, Service charge and Transaction maintenance
Person-to-person (P2P)	Direct credit or debit options, Service charge, Account maintenance and Security
Credit card	Service provider trust, M-wallets (H/W and S/W integration) and Security

Similar to Internet-based online payment methods that support pre-paid (debit) and post-paid (credit) services,   mobile payments require an integrated flexible model that support quality of service (QoS) with one-stop billing schemes for the end-users [21].

Enabling Wireless Technologies

The aforesaid payment methods could be adopted to make mobile payment transactions using different wireless technologies. Such mobile payments go through a set of wireless security levels which is a key factor in providing consumer trust and confidence [28]. The enabling wireless technologies could support mobile security elements that are either removable or non-removable from the mobile device [24]. Removable security elements could be a single chip, a dual chip or a dual slot, while non-removable security elements could support embedded hardware and software modules, including personalised biometric technologies [13]. In table 3, we identify a list of wireless technologies commonly adopted in businesses and their influencing factors for mobile payment adoption.

Table 3. Enabling Wireless Technologies and Influencing Factors

Enabling Wireless Technologies	Influencing Factors
Radio Frequency Identification (RFID)	Transponder & reader frequencies / range, Encryption methods and Payment methods
Short Message Service (SMS)	SMS gateway QoS, Payment methods, Merchant port ID and Push technology
Wireless Access Protocol (WAP) using WIM chip	Dual chip phones, WIM certificate, Payment methods, Security and Account Maintenance
Infrared (IR)	Point-to-point range, Fault-tolerance and Payment methods
Bluetooth	Security, Transmission time and Data volume
Near Field Communication (NFC)	Tokenisation, Mobile phones, NFC reader, Payment methods
Internet	Mobile Apps for Payments, M-wallets, Security, Payment Methods, Account Maintenance & Fees

A mobile payment system typically espouses several wireless technologies with different possible payment methods and with different player-centric influencing factors. With security being a major concern in mobile payments, security modules with digital signatures are being implemented in the form of WAP Identity Module (WIM) cards. From traditional, simple Interactive Voice Response (IVR) systems adopted in phone banking to enhanced WIM chips in mobile devices and now with sophisticated mobile apps, the use of different wireless technologies

involves different levels of security as the device moves in and out of multiple wireless networks leading to many security issues [17].

4. CONCLUSIONS

The adoption of mobile payment is constricted by multi-faceted issues that can be approached from several angles. This paper identified three main dimensions such as mobile payment players, payment methods and enabled wireless technologies. An analysis resulted in their influencing factors for a successful mobile payment adoption. These influencing factors would facilitate consumers, businesses and researchers to consider the different mobile payment solutions, from technology as well as business perspectives for decision making.

REFERENCES

[1] Koss, V., Miles, J. and Boyes, E. (2013). "M-commerce Comes of Age: Collaborate to Succeed," Journal of Payments Strategy & Systems, vol. 7, no. 1., pp. 50-60

[2] Buhan, D., Cheong, Y.C. and Tan, C.L. (2002) 'Mobile payments in m-commerce', Wireless Internet Centre Report, Gemini Ernst & Young.

[3] Wyse, J.E. (2003) 'Supporting m-commerce transactions incorporating locational attributes: an evaluation of the comparative performance of a location-aware method of locations repository management', International Journal of Mobile Communications, Vol. 1, Nos. 1–2, pp.119–147.

[4] Mathew M, Balakrishnan N, Pratheeba S (2010) A study on the success potential of multiple mobile payment technologies. In: Technology management for global economic growth (PICMET), Proceedings of PICMET '10

[5] Pukkasenung P. and Chokngamwong R. 2016 Review and Comparison of Mobile Payment Protocol, J.J.(Jong Hyuk) Park et al. (eds.), Advances in Parallel and Distributed Computing and Ubiquitous Services, Lecture Notes in Electrical Engineering, Springer, pp 11-20.

[6] Lee, C.P., Warkentin, M. and Choi, H. (2004), The Role of Technological and Social Factors on the Adoption of Mobile Payment Technologies, proceedings of the 10th Americas Conference on Information Systems (AMCIS), New York, August, pp. 2781-2786.

[7] Kremers, R. and Brassett J. (2017) Mobile Payments, Social Money: Everyday Politics of the Consumer Subject, New Political Economy, DOI: 0.1080/13563467.2017.1306503.

[8] GSMA White Paper (2017), The Mobile Economy, GSMA.

[9] Anil, S., Ting, L.T., Moe, L.H. and Jonathan, G.P. (2003) 'Overcoming barriers to the successful adoption of mobile commerce in Singapore', International Journal of Mobile Communications, Vol. 1, Nos. 1–2, pp.194–231.

[10] Elliott, G. and Tang, H. (2004) 'The wireless mobile internet: an international and historical comparison of the European and American wireless application protocol (WAP) and the Japanese iMode service', International Journal of Information Technology and Management, Vol. 3, No. 2/3/4.

[11] EMV White Paper, (2017), Payment Tokenisation Specification EMVCo.

[12] EY White paper (2017), Mobile money — the next wave of growth, Ernst and Young.

[13] BCG White Paper (2017), Digital Payments 2020, BCG Boston Consulting Group.

[14] Ondrus, J. and Pigneur, Y. (2006),Towards a holistic analysis of mobile payments: a multiple perspectives approach, Electronic Commerce Research and Applications, 5(3), pp. 246-257. Ondrus, J. and Pigneur, Y. (2007), An assessment of NFC for future mobile payment systems, proceedings of the International Conference on Mobile Business (ICMB'07),Toronto, July, pp. 43-53.

[15] Mobile Wallet Report (2017) Study of Mobile Banking & Payments First Annapolis Consulting, Inc. |

[16] Yuan, Y. and Zhang, J.J. (2003) 'Towards an appropriate business model for m-commerce', International Journal of Mobile Communications, Vol. 1, Nos. 1–2, pp.35–56.

[17] Hayashi, F. (2012). "Mobile Payments: What's in It for Consumers?" Fed¬eral Reserve Bank of Kansas City, Economic Review, vol. 97, no. 1, pp. 35-66.

[18] Mallat, N. (2007), Exploring consumer adoption of mobile payments - A qualitative study, The Journal of Strategic Information Systems, 16(4), pp. 413-432.

[19] Carton, F, Hedman, J, Dennehy, D, Damsgaard, J, Tan, K and McCarthy, J, B. (2012), "Framework for Mobile Payments Integration" The Electronic Journal Information Systems Evaluation Vol. 15 No.1, pp.14 -25.

[20] Ghosh, A. and Swaminatha, T. (2001) 'Software security and privacy risks in mobile e-commerce', Communications of the ACM, Vol. 44, No. 2, pp.51–57.

[21] Koutsopoulou, M. et al. (2004) 'An integrated charging, accounting and billing management platform for the support of innovative business models in mobile networks', International Journal of Mobile Communications, Vol. 2, No. 4, pp. 418-434.

[22] Fun TS, Beng LY, Razali MN (2013) Review of mobile macro-payments schemes. Journal of Advances in Computer Networks, Vol. 1, No. 4, pp. 323-328

[23] Singh A, Shahazad KS (2012) A review: secure payment system for electronic transaction. International Journal of Advanced Research in Computer Science and Software Engineering, Vol. 2, No. 3, pp. 236-243

[24] Ahamad SS, Udgata SK, Nair M (2014) A secure lightweight and scalable mobile payment framework. In: FICTA 2013. Advances in intelligent system and computing, vol 247. Springer International Publishing, Switzerland, pp. 545-553.

[25] Venkatraman, S. (2008), 'Mobile Payment Implementation – A Reference Framework', International Journal of Business Information Systems, 3, 3, pp. 252-271.

[26] Xu, W. (2017), The Study of WeChat Payment Users Willingness Factor, Journal of Service Science and Management, 10, 251-259

[27] Au, Y.A. and Kauffman, R.J. (2008), The economics of mobile payments: Understanding stakeholder issues for an emerging financial technology application, Electronic Commerce Research and Applications, (7), pp. 141–164.

[28] PwC Mobile Payment Report 2017, Price Waterhouse Coopers.

AUTHORS

Dr. Sita Venkatraman earned her PhD in Computer Science, with a doctoral thesis titled "Efficient Parallel Algorithms for Pattern Recognition", from National Institute of Industrial Engineering in 1993.   In the past 30 years, Sita's work experience involves both industry and academics - developing turnkey projects for IT industry and teaching a variety of IT courses for tertiary institutions, in India, Singapore, New Zealand, and more recently in Australia since 2007. She currently works as Lecturer (Information Technology) at the School of Engineering, Construction & Design, Melbourne Polytechnic, Australia. She also serves as Member of Register of Experts at Australia's Tertiary Education Quality and Standards Agency (TEQSA).

Dr. Venkatraman's. recent research areas are predominantly in Business Intelligence, Data Mining, Cloud Computing, Information Security and Health Informatics.   She has published eight book chapters and more than 120 research papers in internationally well-known refereed journals and conferences..

AUTHOR INDEX