CONFIGURING ASSOCIATIONS TO INCREASE TRUST IN PRODUCT PURCHASE

Pegah Moslemipoor [1] and Ali Haroon Abadi [2]

[1] Department of Computer Engineering, Kish International Branch, Islamic Azad University, Kish, Iran
[2] Department of Computer Engineering, Islamic Azad University, Science and Research Branch, Tehran, Iran

ABSTRACT

Clustering is categorizing data into groups with similar objects. Data mining adds to complexities of clustering a large dataset with various features. Among these datasets, there are electronic business stores which offer their products through web. These stores require recommendation systems which can offer products to the user which the user might require them with higher probability. In this study, previous purchases of users are used to present a sorted list of products to the user. Identifying associations related to users and finding centers increases precision of the recommended list. Configuration of associations and creating a profile for users is important in current studies. In the proposed method, association rules are presented to model user interactions in the web which use time that a page is visited and frequency of visiting a page to weight pages and describes users' interest to page groups. Therefore, weight of each transaction item describes user's interest in that item. Analyzing results show that the proposed method presents a more complete model of users' behavior because it combines weight and membership degree of pages simultaneously for ranking candidate pages. This method has obtained higher accuracy compared to other methods even in higher number of pages.

KEYWORDS

Data mining, association rule, clustering, previous behavior of user, recommender system

1. INTRODUCTION

Data mining is analyzing data to explore unknown relationships which offer useful information [1]. Data mining includes using complex analytical tools to explore valid and unknown patterns and existing dependencies in large data sets. Therefore, data mining technology is more than collecting and managing data; it includes decomposition, analysis and prediction [2]. In another definition, data mining is referred to extracting hidden information or patterns and specific relations in a large volume of data in one or more large information banks [3]. Based on the pattern which should be extracted, data mining tasks are categorized as below [4]:

- Summarization
- Classification
- Association
- Clustering
- Trend

Main duty of data mining is to classify and assign each record of database to one of pre-defined classes. Next task is clustering in which groups of records close to a sample record are found. Association is next duty which can be used to define implicit rules based on record features [5]. Data mining has wide applications which are classified as descriptive and predictive [6]. In descriptive data mining, purpose is to increase data recognition and its contents, while in predictive data mining, purpose is to predict to orient decision procedure [7].

2. ASSOCIATION

Association rules square measure supported the notion of group action that is an observation of the co-occurrence of a collection of things [8].Association means exploring union or relationship between objects. Association is based on certain rules. These rules are known as association rules. These rules show that there is a cooperative relationship among objects. In fact, association finds this relationship in a set of objects [2]. Association rule mining is an unsupervised learning method in data mining [9]. Association rule mining is used to find hidden correlations among different transactions in databases. An association rule might be any regulation which describes association relationship among different objects or items. In general, these association rules are If-Then rules which operate on conditional probabilities [10]. Association rules are the most important factor in recommender motors. In other words, association rules are the basis of each recommender motor. Association rules refer to exploring dependency among items which occur simultaneously. These rules are represented as argumentative proposition A->B. important scales in association rules which are used to evaluate explored rules are support and confidence which results in profitability and certainty of explored rules, respectively. If $A \Rightarrow B$, values of support and confidence are calculated using Equations 1 and 2.

$$Support \ (A \Rightarrow B) = P(A \cup B) \qquad (1)$$
$$Confidence \ (A \Rightarrow B) = P(B \mid A) \qquad (2)$$

Most common method used to find association rules is to break the problem into two parts [11]:

1. Finding all repetitive item sets
2. Generating strong association rules from repetitive items

Finding all repetitive items set is difficult while generating strong rules costs lower. Association rule mining is a main research field in data mining. Apriori and FT-Growth algorithms are basis algorithms for many association rule mining algorithms [12].

2.1. APRIORI ALGORITHM

Apriori is an association rule mining algorithm which finds repetitive algorithms. This algorithm is a classic algorithm for generating repetitive patterns [6]. This algorithm is constructed on databases including transactions. Purpose of this algorithm is to find correlation between different data sets. Sometimes, it is called shopping bag. Algorithm input is a set of items and its output is a set of rules. In this algorithm, association rules are usually generated in two steps. First, minimum Support is used to find all repetitive item sets in the database, then this set along with minimum Confidence is used to constitute rules [12].

Algorithm performs recursively and generates a lot of rules. Constraint of this algorithm is generating candidate items at each processing level [13]. Scalability is the most important problem in Apriori algorithm and computational complexity increases significantly [9].

3. CLUSTERING

Clustering is identification of clusters or groups for a set of objects with unknown classes. Clustering should be performed such that similarities between objects of similar clusters is maximum and similarity between clusters of different clusters is minimum. Purpose of clustering is to divide existing data to multiple groups such that data of different groups have maximum possible difference and data of a group are similar as much as possible. Clustering procedure includes different steps. 4 main steps are as follows [5]:

- Feature selection or Extraction
- Designing clustering algorithm
- Validation
- Result analysis

Clustering is similar to classification. Clustering is an unsupervised learning method or unsupervised clustering of data items (including feature vectors (FV), patterns or observations). Clustering is finding clusters or classifying data in a set of unlabeled data. In Artificial intelligence, clustering technique is one of the main tasks [14]. Different clustering methods have been proposed as various algorithms. Figure 1 shows clustering algorithms employed in big data mining. Most important algorithm in clustering context is K-mean algorithm, in which each cluster is represented through its average objects (cluster center) [3].
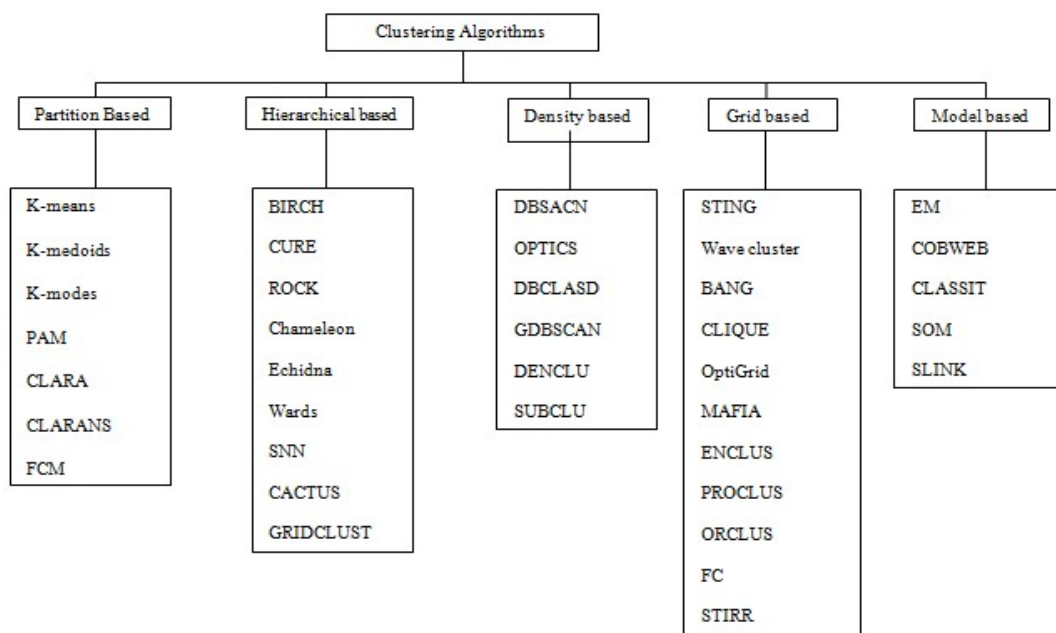

Figure 1. Clustering algorithms in data mining [15]

3.1. K-MEANS ALGORITHM

K-means algorithm which is also called Lloyd algorithm [16] is one of the partition-based k series algorithms, one of the simplest unsupervised learning algorithms and one of the most important clustering algorithms. In this algorithm, clusters are separated through dividing a data set of N objects to k separate sets of points. Advantage of K-means is that when number of variables is high, computations might be performed faster than hierarchical clustering. Disadvantage of this algorithm are as follows [14]:

- Comparing number of generated clusters is difficult.
- Constant number of clusters might make predicting value of k difficult.
- It does not work well with non-spherical clusters

Main idea of K-means algorithm is defining k centers for each cluster. These centers should be selected with high precision, because different centers result in different conclusions. Therefore, best option is putting centers as far as possible from each other. In the next step, each pattern is assigned to the closest center. When all points are assigned to existing centers, an initial grouping

is performed and first step is complete. This way, new value of K is calculated and points are assigned to new clusters; this is continued until other k centers are not displaces. In fact, this procedure looks to minimize cost function or objective function. This function exists in Equation 3:

$$Cost = \sum_{i=1}^{N} argmin_j \left(\left\| X_i - C_j \right\| \right)^2 \tag{3}$$

4. CONCEPTS

4.1. USER PROFILE

Number of users who search Internet and purchase through Web is increasing every day. Owners of web sites provide the possibility for users to create profile. Internet users can have specific profiles at each web site. For instance, if a user intends to purchase a product from a commercial site can sign up in that web site and set a user profile by filling fields like name, age, job, education and etc along with username and password. As users can see a history of their activities in that website, website owners can also use this profile to provide better services for the users.

4.2. WEB RECOMMENDER SYSTEMS

Recommendation system is employed in various applications and different aspects. Recommender systems have emerged since beginning of Internet but not in personal form. Many methods are used in recommendation system which have some advantages and disadvantages. Currently, these methods include [17]:

- Collaborative filtering
- Content-based filtering
- Demographic information-based filtering
- Knowledge-based filtering
- Hybrid filtering methods

Recommender systems are categorized as personal and impersonal classes. Personal recommender systems include all of the above including, collaboration, content, demographic, knowledge and hybrid. Collaborative filtering is also divided into user-based and item-based classes [18].

4.3. TRUST IN INTERNET SHOPPING

Internet shopping requires trust [19]. In recent years, the term "Trust" has become significant. For a virtual organization, trust is the most important problem and one of the management priorities [7]. Trust is a Feeling accompanied by acceptance of behavior [6]. Trusting user purchase depends on different factors. These factors include personal trust, background factors, knowledge and etc. Customer loyalty is the result of customer trust. Compared to conventional market, competition challenge is Internet market is higher and customer loyalty is less. Loyalty is purchasing again from the same brand or market. Customer loyalty is customer tendency to refer to a specific website again or his tendency to purchase a specific product. Different loyalty degrees form a pyramid, where the most important one is brand loyalty and customer tendency from a specific brand .

5. THE PROPOSED APPROACH

Using data mining technique, website managers can analyze preferences and purchase patterns of online customers for product recommendations. Data mining tool helps offering services proportional to user requirements. Purpose of this study is to identify and classify customers of internet stores based on sexual similarity, geographical location and age. Therefore, the proposed method is presented in three steps including:

- Feature extraction from user profile
- User clustering
- Extracting association rules

First step is to extract features from user profiles which is clustering algorithm perquisite and includes features used in clustering. Features used for clustering are extracted from user profiles which include: geographical location, age and sexual similarity.

In clustering users, following items are investigated considering a minimum threshold value and maximum threshold value:

- If distance of user from all clusters is greater than maximum threshold, a new cluster is created and user record is located in that cluster.
- If distance of user from cluster is smaller than threshold, he is located in its current cluster; otherwise, maximum threshold is checked and if user distance is greater than threshold, a new cluster is created.
- This way, all clusters are checked for the users and in user distance from clusters is not smaller than threshold, and user distance from all clusters is greater than threshold, a new cluster is created and current record is located in the new cluster.

After performing the steps, there are still people which are not located in any of the clusters. Therefore, mentioned steps are repeated with this difference that maximum and minimum threshold values are changed, that is maximum value is decreased and minimum value is increased so that remained people are also clustered. In order to extract association rules, most applicable association rules are used in shopping bag analysis. Using these association rules, one can detect which products are bought together and describe confidence percentage. Following sentence is an example of output interpretation of association rules in analysis of shopping bag: "customers whom have soap in their shopping bag have bought paper tissue also in 74% of cases".

Using these rules, sellers can modify arrangement of their products to sell as much as possible, for instance, considering the mentioned rule, seller puts paper tissues beside soaps to increase selling paper tissues.

5.1. IMPLEMENTATION

Simulation is performed in MATLAB and employed dataset is related to real NASA data. The proposed algorithm evaluates overlap of two sets, that is ratio of suitable recommendations to total number of recommendations through precision measure presented in Equation 4.

$$precision(rs, rp) = \frac{|rs \cap rp|}{|rs|} \tag{4}$$

rw is rize of the recommended window, rt is the recommended threshold and rt is $rp=\{x_{rw+1},x_{rw+2},\ldots,x_{rw+|rs|}\}$ set of visited pages by the user followed by real subsidence.

5.2. RESULTS

Method used to generate recommendations based on association rules is compared with 4 common methods based on conventional association rules and learning automata. A brief review of performance of these methods based on precision of recommendations vs. number of recommended pages can be seen in Diagram 1. According to this figure, association rule based method, obtains better results compared to other methods.
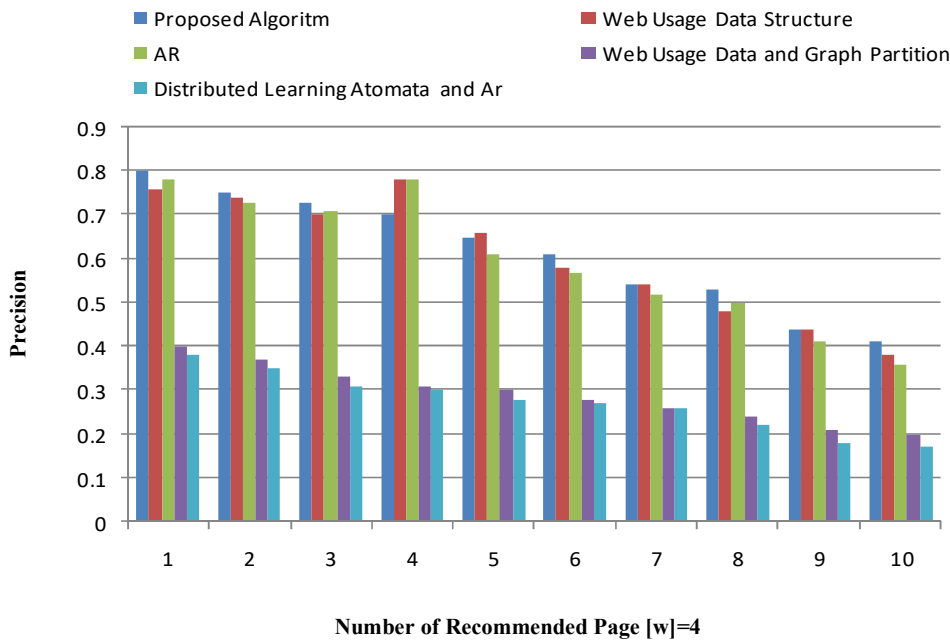

Figure 2: Comparing the Proposed algorithm and some other methods in terms of Precision

5.3. RESULTS ANALYSIS

Investigating recommendations of these methods, it can be seen that in low number of recommended pages, recommendations are generally associated to pages which have been visited by the user immediately after visiting the current page. In fact, each method offers most reliable recommendations therefore, this set of recommendations have relatively high precision. On the contrary, by increasing number of recommended pages, each method offers its subsequent recommendations with lower score. In this case, our proposed method recommends pages which will be visited by the user in next session with higher precision. In general, performance of these methods degrades as number of pages increases. But association rule based method has proposed a more complete model of user behavior due to combining weight and membership degree of pages simultaneously for ranking candidate pages. This method outperforms other methods even in high number of pages.

6. CONCLUSION

User model is important since it is one of the main components of web personalization system. The better is this model, recommendations offered based on the model would be more precise and

deeper. In this paper, association rules are offered to model user interactions in web which use duration of visiting a page and frequency of visiting a page to weight pages and describe user interest in pages. Analysis shows that proposed method based on association rules has proposed a more complete model of user behavior due to combining weight and membership degree of pages simultaneously for ranking candidate pages. This method outperforms other methods in terms of precision even in high number of pages.

REFERENCES

[1] Yeh, I., & Lien, C. (2008). The comparisons of data mining techniques for the predictive accuracy of probability of default of credit card clients, Expert Systems with Applications 36 (2) (2008) 2473–2480.

[2] Castellón González, Pamela, and Juan D. Velásquez., (2013), "Characterization and detection of taxpayers with false invoices using data mining techniques."Expert Systems with Applications 40.5 (2013): 1427-1436.

[3] Richa Gupta ,(2014),"Journey from Data Mining to Web Mining to Big Data", International Journal of Computer Trends and Technology (IJCTT) ,page 18-20, volume 10 number 1 , Apr 2014.

[4] http://academic.csuohio.edu/fuy/Pub/pot97.pdf

[5] Mansi Gera, Shivani Goel, (2015), Data Mining - Techniques, Methods and Algorithms: A Review on Tools and their Validity, International Journal of Computer Applications (0975 – 8887) Volume 113 – No. 18, March 2015, pages 22-29.

[6] M. Sinthuja, N. Puviarasan and P. Aruna, (2017), Evaluating the Performance of Association Rule Mining Algorithms, World Applied Sciences Journal 35 (1): 43-53, 2017.

[7] E. W. T. Ngai, "Customer relationship management research (1992-2002): An academicliterature review and classification," Mark. Intell. Plan., vol. 23, no. 6, pp. 582–605, Jan.2005.

[8] K.Karthikeyan and Dr.V.Karthikeyani, (2014), Association Rule Mining Based Extraction of Semantic Relations Using Markov Logic Network, International Journal of Web & Semantic Technology (IJWesT) Vol.5, No.4, pages: 33-51.

[9] Hoda Khanali, Babak Vaziri, (2017), A Survey on Improved Algorithms for Mining Association Rules, International Journal of Computer Applications (0975 – 8887) Volume 165 – No.9, May 2017, pages:6-11.

[10] Amuit Kumar Chandan, Kavita & M K Shukla, ,(2017), ASSOCIATION RULE MINING USING MODIFIED BPSO, International Journal of Computer Science Engineering and Information Technology Research (IJCSEITR) ISSN(P): 2249-6831; ISSN(E): 2249-7943 Vol. 7, Issue 2, Apr 2017, 29-36.

[11] Rekha Jain , Dr. G. N. Purohit, (2011), "Page Ranking Algorithms for Web Mining" , International Journal of Computer Applications (0975 – 8887) Volume 13– No.5, January 2011.

[12] K. S. Ranjith, Yang Zhenning, Ronnie D. Caytiles and N. Ch. S. N. Iyengar,(2017) ,Comparative Analysis of Association Rule Mining Algorithms for the Distributed Data, International Journal of Advanced Science and Technology Vol.102 (2017), pp.49-60.

[13] MS.J.OMANA, MS.S.MONIKA, MS.B.DEEPIKA, (2017), SURVEY ON EFFICIENCY OF ASSOCIATION RULE MINING TECHNIQUES, J.OMANA et al, International Journal of Computer Science and Mobile Computing, Vol.6 Issue.4, April- 2017, pg. 5-8.

[14] Prachi Surwade1, Prof. Satish S. Banait, (2016), A Survey On Clustering Techniques For Mining Big Data, International Journal of Advanced Research in Science Management and Technology, Volume 2, Issue 2, February 2016.

[15] T. Sajana, C. M. Sheela Rani and K. V. Narayana , (2016), A Survey on Clustering Techniques for Big Data Mining, Indian Journal of Science and Technology, Vol 9(3), January 2016.

[16] Barkha Narang, Poonam Verma, Priya Kochar, (2016), Application based, advantageous K-means Clustering Algorithm in Data Mining – A Review, International Journal of Latest Trends in Engineering and Technology (IJLTET), ISSN: 2278-621X Vol 7 issue 2 July 2016.

[17] Rahul Singh, Kanika chuchra and Akshama Rani, (2017), A Survey on the Generation o Recommender Systems, I.J. Information Engineering and Electronic Business, 2017, 3, 26-35 Published Online May 2017.

[18] Debashis Das, Laxman Sahoo, Sujoy Datta ,(2017) , A Survey on Recommendation System, International Journal of Computer Applications (0975 – 8887) Volume 160 – No 7, February 2017.

[19] Kwek Choon Ling (Corresponding author) , The Effects of Shopping Orientations, Online Trust and Prior Online Purchase Experience toward Customers' Online Purchase Intention , International Business Research , ISSN 1913-9004 ,pp.63-76,Vol. 3, No. 3; July 2010.