SARCASM AS A CONTRADICTION BETWEEN A TWEET AND ITS TEMPORAL FACTS: A PATTERN-BASED APPROACH

Santosh Kumar Bharti and Korra Sathya Babu

Department of Computer Science and Engineering, National Institute of Technology, Rourkela, Odisha, India

ABSTRACT

In the context of Indian languages, sarcasm detection in Hindi is a tedious job as it is rich in morphology and complex in structure. The annotated resources for sarcastic Hindi sentences are almost negligible for machine learning analysis. Here, we propose a pattern-based framework for sarcasm detection in Hindi tweets. It has been observed that a tweet is sarcastic if it contradicts its temporal facts intentionally. The temporal fact is a collection of time-dependent facts which may change over the period. We used Hindi news with timestamp as a corpus of temporal facts. The timestamp describes the fact period of any entity. In this research, a temporal fact is represented as a <key, value> pair. To form a <key, value> pair, one need to extract triplets i.e. subject, verb and object for every sentence. Next, a key is formed using subject and verb together. The value is formed using object and timestamp together. To predict the sarcastic tweet; one needs to extract the triplets from input tweet and form a <key, value> pair. Now, the <key, value> pair of the input tweet is mapped with related <key, value> pair in the corpus of temporal facts and are checked if they coincide. If they contradict, the input tweet is considered as sarcastic. The achieved accuracy of the proposed approach outperforms the state-of-the-arts techniques for Hindi sarcasm detection as it attains an accuracy of 82.8%.

KEYWORDS

Natural Language Processing, Sentiment, Sarcasm, Social Network, Tweets

1. INTRODUCTION

Sarcasm is an indirect way to negate a sentence to feel someone insult or stupid. While writing or speaking sarcastic sentences, one often use positive words or intensified positive words. Therefore, detection of sarcasm is one of the interesting problems in the text, speech, video, etc. In textual data, physical clues are missing such as hands movement, rolling eyes, intentional tonal stress, *etc*., unlike speech and video data. This makes the analysis of sarcasm in text ever more difficult.

Twitter is one of the popular social media sites that allow users to post a message within 140 characters. The data generated by Twitter is much higher compared to other social media [1]. For instance, Twitter claims to have more than 500 million users, out of which more than 332 million are active [2]. They post more than 340 million tweets and 1.6 billion search queries every day [1]. Due to this, Twitter generates such a large volume of data every day that poses many challenges such as accessing, storing, processing, analyzing etc. [3]. Among these challenges, analysis of sarcasm in tweets is the most difficult task especially in the context of Indian language such as Hindi, Telugu, Bengali, *etc.*

DOI: 10.5121/ijnlc.2018.7507

Hindi is one of the popular Indian languages and is highly used by Indians while communicating on social media such as Facebook, Twitter, WhatsApp, *etc*. Hindi stands fourth in popularity after Mandarin, Spanish, and English [4]. It is widely used for speaking in countries like India, Mauritius, Fiji, Suriname, Guyana, Trinidad & Tobago and Nepal [5]. It has 490 million speakers across the world, and majority of them are from India [6]. In the past, several authors have proposed sarcasm detection system for tweets scripted in English [7-15]. These methods require an extensive set of annotated training data to detect sarcasm. In the domain of sarcasm detection, availability of Hindi annotated dataset is almost negligible. Due to this, the research related to the Hindi language is not much explored so far. A sarcasm detection system was developed [16] for Hindi tweets using a similar set of features used for English tweets namely, #tag, emoticons, punctuation marks etc. However, in real scenario natural Hindi tweets 1 are different in structure unlike English scripted tweets or Hindi tweets translated from English tweets. A sample list of Hindi sarcastic tweets is shown in Figure 1.

1. काले धन पे पेनल्टी 200% से घटा के 10% कर दी? काला धन वालों के सामने मोदी जी ने घुटने टेक दिए?- @ArvindKejriwal
2. दो दिन बाद शाहरुख खान अपना 51वां जन्मदिन मनाने वाले हैं, लेकिन उनकी हीरोइन की उम्र लगातार कम होती जा रही है
3. @Rajrrsingh #सुना_है! #iphone7 टिम कुक के टकले पे रख के चार्ज किया जायेगा!
4. आज सुबह मुझे सवच्छता भारत अभियान सड़क पर बिखरा हुआ मिला! #swachbharat #Hindi #clean #mock #sarcasm
5. #JioOffer का आधा से ज्यादा डेटा तो लोग सिर्फ ट्विटर पे अरविन्द केजरीवाल को ट्रोल करने में इस्तेमाल करते है.

Figure 1: Sample Hindi sarcastic tweets.

To identify sarcasm in such natural Hindi tweets, the same feature set used for English scripted tweets might not be applicable efficiently. Therefore, one needs to rely on other parameters such as news context, specific patterns, rules, *etc*., to identify sarcastic Hindi tweets. In this research, we propose a pattern-based framework for sarcasm detection in Hindi tweets by observing manually annotated set of natural Hindi sarcastic tweets. After analyzing these tweets, it is observed that a tweet is sarcastic if it contradicts to its temporal fact. A temporal fact can be defined as a time-dependent fact. Unlike a universal fact, the temporal fact may change over the period. This approach is wholly based on timestamp that covers the facts period. The news is one of the primary sources of temporal facts over the period. Hence, this research uses a corpus of Hindi news from top Hindi news sources as a corpus of temporal facts. While prediction, the framework verifies the occurrences of any contradiction between an input tweet and its temporal fact. If they contradict, the input tweet is considered as sarcastic; otherwise not sarcastic.

The rest of the paper is organised as follows: Section 2 describes related work. All the preliminaries are explained in Section 3. The proposed scheme is discussed in Section 4. Analysis of the results are given in Section 5 and conclusion of the article is drawn in Section 6.

2. RELATED WORK

In recent past, several authors have worked on sarcasm detection in Twitter data. They proposed various sarcasm detection systems for tweets scripted in English [7-15]. According to the literature, Twitter sarcasm detection techniques can be classified into four categories, namely lexicon-based approach, machine learning-based approach, context-based approach and pattern-based approach.

2.1 LEXICON-BASED APPROACH

This approach utilizes a bags-of-lexicons which comprise of unigram, bigram, trigram, *etc*. phrases to identify sarcasm in tweets. Riloff *et al*. [9] developed two bags-of-lexicons, namely positive sentiment and negative situation using bootstrapping technique. These lexicon file consists of unigram, bigram, and trigram phrases. Further, they utilized these phrases to identify sarcasm in tweets for the structure, positive sentiment in a negative situation. Similarly, Bharti *et al.* [12] developed four bag-of-lexicons, namely positive sentiment, negative situation, negative sentiment and positive situation using parsing technique. Further, they utilized these phrases to identify sarcasm in tweets for the both structures, namely positive sentiment in a negative situation and negative sentiment in a positive situation.

2.2 MACHINE LEARNING-BASED APPROACH

In this approach, one needs to get labelled dataset to extract features. Further, these features are used to train the classifiers such as SVM, DT, NB, *etc*. Liebrecht *et al.* [8] designed a machine learning model to detect sarcasm in Dutch tweets. Peng *et al.* [17] used a supervised approach to classify movie reviews into two classes after performing subjective feature extractions. Tayal *et al.* [18] utilized machine learning classifier to detect the sarcastic political tweets. Tungthamthiti *et al.* [19], explored concept level knowledge using the hyperbolic words in sentences and gave an indirect contradiction between sentiment and situation such as raining, bad weather that are conceptually the same. Therefore, if 'raining' is present in any sentence, then one can assume 'bad weather'. They build a system for sarcasm detection using SVM classifier and used these indirect contradiction feature for training. Bharti *et al.* [20] developed a machine learning framework to detect sarcastic tweets. They deployed several classical classifiers such as SVM, DT, NB, Maximum Entropy (ME), *etc*.

2.3 CONTEXT-BASED APPROACH

The relationship between an author and audience followed by the immediate communicative context can be helpful to improve the sarcasm prediction accuracy [21]. Message-level sarcasm detection on Twitter using a context-based model were used for sarcasm detection [22]. Chains of tweets that work in a context were considered. They introduce a complex classification model that works over an entire tweet sequence and not on one tweet at a time. Integration between linguistic and contextual features extracted from the analysis of visuals embedded in multimodal posts was deployed for sarcasm detection [23]. A framework based on the linguistic theory of context incongruity and an introduction of inter-sentential incongruity by considering the history of the posts in the discussion thread was considered for sarcasm detection [11]. A quantitative evidence of historical tweets of an author can provide additional context for sarcasm detection [24]. The author`s past sentiment on the entities in a tweet was exploited to detect the sarcastic intent.

2.4 PATTERN-BASED APPROACH

In this approach, one can observe the sarcastic dataset and identify unique patterns for sarcastic text. Riloff *et al*. identified a unique pattern for sarcasm detection in the tweet *i.e.* "sarcasm is a contrast between positive sentiment and negative situation" [9]. Similarly, another pattern-based sarcasm detection system using parsing technique was proposed [12]. They generalised the concept of [9] approach and suggested that sarcasm is a contradiction between sentiment and situation. It may follow both the patterns namely, sarcasm is a contrast between positive sentiment and negative situation, sarcasm is a contrast between negative sentiment and positive situation. Bouazizi and Ohtsuki identified three patterns for sarcasm in tweets; namely, sarcasm is

a wit, sarcasm is an evasion and sarcasm is a whimper [14]. Based on these three patterns, they proposed an efficient sarcasm detection system that enhance the accuracy of sentiment analysis.
In the context of Indian languages, sarcasm detection is less explored due to the unavailability of benchmark resources for training and testing. Desai *et al*. proposed a Support Vector Machine (SVM) based sarcasm detector for Hindi sentences [16]. They used Hindi tweets as the dataset for training and testing using SVM classifier. In the absence of annotated datasets for training and testing, they converted English tweets into Hindi. Therefore, they focused on a similar set of features like emoticons and punctuation marks for sarcasm detection in English text. These methods when applied directly, are not suitable for the Hindi sarcastic tweets. Similarly, a context-based approach for sarcasm detection in Hindi tweets is proposed [25]. They exploit online Hindi news as a context to determine the sarcasm in tweets. They experimented using a small set of manually collected and annotated Hindi tweets. Dalal *et al.* proposed an insult detection in Hindi text using logistic regression and SVM classifiers [26]. They have used n-grams, negation and second person as features to build the feature vector to train these classifiers.

3. PRELIMINARIES

This section describes all the tools used in this article namely, Hindi POS tagging, Hindi parsing and triplet extraction of a sentence. The details are explained below:

3.1 POS TAGGING

To identify the POS tag information in Hindi sentences, we have developed a Hidden Markov Model (HMM) based POS tagger [30]. It uses Indian Language (IL) standard tagset which consists of 24 tags [27]. The POS tagger is available on URL:http//www.taghindi.herokuapp.com. The implementation details are released on URL: https://github.com/rkp768/hindi-pos-tagger. An example Hindi POS tagging is shown in Figure 2.

मुझे सोना पसंद हैं| (Mujhe Sona Pasand Hain.)

मुझे- PRP | सोना--NN | पसंद--NN | हैं-VAUX

Figure 2: An example POS tagging in Hindi text.

3.2 PARSING

Parsing is the way toward breaking down the linguistic structure of a language. Given a succession of words, a parser shapes units like subject, verb, object and identify the relations between these units, as per the tenets of a formal sentence structure and produce a parse tree (Bharti, 2015). This article deploys Hindi shallow parser to get the parsing tree information for Hindi text [28]. The Hindi shallow parser is available on URL:http://ltrc.iiit.ac.in/analyzer/hindi/run.cgi. Figure 3 depict the examples of the parsing information for Hindi sentence राम के ला खाता ह| (Ram kela khata hain) (Ram eats banana).

राम केला खाता हैं|

```
<Sentence id="1">
1     ((     NP   <fs af='राम,n,m,sg,3,d,0,0' head="राम">
1.1   राम   NNP  <fs af='राम,n,m,sg,3,d,0,0' name="राम">
      ))
2     ((     NP   <fs af='केला,n,m,sg,3,d,0,0' head="केला">
2.1   केला  NN   <fs af='केला,n,m,sg,3,d,0,0' name="केला">
      ))
3     ((     VGF  <fs af='खा,v,m,pl,1,,ता_है,wA' vpos="tam1_2" head="खाता">
3.1   खाता  VM   <fs af='खा,v,m,sg,any,,ता,wA' name="खाता">
3.2   |      SY   <fs af='.,punc,,,,,,'>
      ))
</Sentence>
```

Figure 3: Parsing information for Hindi text "राम के ला खाता ह"| (Ram kela khata hain) (Ram eats banana)

3.3 TRIPLET EXTRACTION

This article used Rusu_Triplets extraction algorithm [29] to extract the triplets i.e. subject, object and verb of any sentence using parse tree information. For example, triplets information of the Hindi sentence "राम के ला खाता ह"| (Ram kela khata hain) (Ram eats banana) are राम (Ram) as subject, के ला (Kela) as object and खाता (Khata) as verb.

4. PROPOSED SCHEME

This section describes the proposed framework for sarcasm detection in Hindi tweets using its temporal fact. In this research, a corpus of Hindi news is considered as a set of temporal facts and prediction of sarcasm in Hindi tweet is made through a pattern "sarcasm as a contradiction between a tweet and its temporal facts". While detecting a sarcastic tweet, it checks the contradiction between a tweet and its temporal fact. If contradiction occurs, then the tweet is classified as sarcastic; otherwise non-sarcastic. The pipeline procedure of proposed framework is shown in Figure 4.

The framework starts with a Hindi news corpus and identifies the parsing (syntax tree) information of every news in the corpus. Next, the parse information is used to extract triplets *i.e.* subject, object and verb of every news sentence. Based on the extracted triplets, possible tweets are collected from Twitter. For testing a tweet as sarcastic or not, it forms (key, value) pairs for both tweet and the related news. While forming (key, value) pair, subject and verb together from triplets are treated as key. Object and its timestamp together are considered as value. Finally, it checks the contradiction between a tweet and its temporal fact. If they contradict each other, the input tweet is considered as sarcastic in the context of time-dependent fact; otherwise, given tweet is not sarcastic.

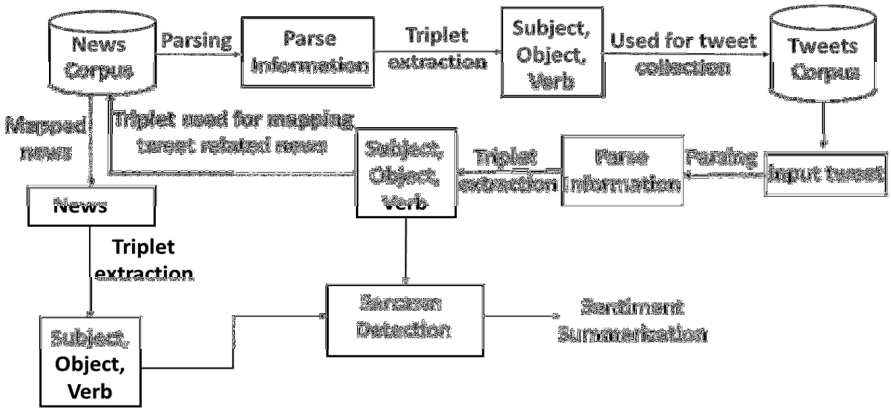

Figure 4: Proposed framework for sarcasm detection in Hindi tweets using temporal facts.

4.1 News And Tweets Collection

This After browsing several online news sources such as ABP News Hindi, AajTak, Zee News Hindi, *etc*., on Twitter, we have collected a total of around 500 processed Hindi news with timestamp manually using the trending topics such as #EVMs, #EVMtampering, #Bahubali2, #khelratna, etc. A sample Hindi news on EVMs is given in Figure 5. While pre-processing, the news related to murder, rape, bomb blast, etc., were eliminated. We believe that sarcastic tweets will not be floated on such grave topics. Using the similar trending topics, approximately 2500 Hindi tweets were collected manually in the same timestamp. A sample of tweets on #EVMs is shown in Figure 6. The datasets of both the news and tweets are released on URL: https://github.com/sbharti1984/Hindi-Tweets.

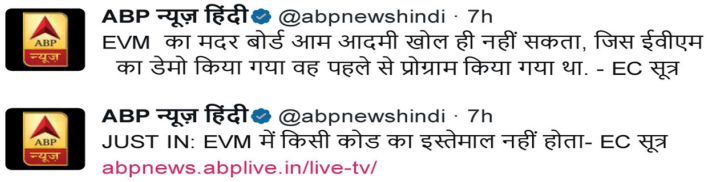

Figure 5: A sample Hindi news on EVM.

Further, among 2500 Hindi tweets, 2000 Hindi tweets were used as observation set and remaining 500 Hindi tweets were used as testing set. Next, both sets of Hindi tweets were distributed for annotation among three professionals in the Hindi language who are teachers and practitioners. They annotated these 2500 tweets manually to identify the tweet as sarcastic or not. After collecting the annotation results from all the individuals and we calculate the Inter-annotator agreement (IAA). To compute the IAA, Fleiss' Kappa coefficient [31] is used as it is more suitable when the number of annotators is more than two. In this experiment, the attained IAA value is 0.89.

Figure 6: A sample Hindi tweets related to EVM.

The annotation result of testing set is used as ground truth to check the performance of the proposed system and error analysis. After verifying the annotation results of all the three annotators, the result is shown in Table 1.

Table 1: Annotation result of 2500 Hindi tweets as sarcastic or not.

	Total tweets	sarcastic	not sarcastic	ambiguous
Observation set	2000	724	1140	136
Testing set	500	184	290	26

After annotation, we observed that most of the sarcastic tweets contradict the temporal facts. Therefore, we proposed a pattern-based sarcasm detection system using pattern "sarcasm as a contradiction between a tweet and its temporal facts".

4.2 Generation Of (Key, Value) Pairs From Temporal Facts (News)

This section generates the <key, value> pairs of news corpus as a set of temporal facts. The procedure of generating <key, value> pairs is given in Algorithm 1.

Algorithm 1 takes a corpus of news as temporal facts and finds parse tree information of every news using Hindi shallow parser [28]. Next, the parse tree information is used to extract triplets (subject, verb and object) using Rusu_Triplets extraction algorithm [29]. Finally, it generates the (key, value) pair for every news sentence with the help of the extracted triplets. In the formation of (key, value) pair of news, a key is made up using subject and verb from its triplets. Value is consists of an object and its timestamp. The times- tamp is the date of which the news was released.

Algorithm 1: *Generating_Key_Value_Pairs*

Data: *dataset* := Corpus of temporal facts (news).
Result: *classification*:= A $\langle key, value \rangle$ pairs
Notation: N: News, SoN: Subject of News, VoN: Verb of News, OoN: Object of News, $\mathbb{C}$: Corpus, PIF: Parsing information file, $NWPP$: News-wise parsing phrases, TFF: Temporal fact file, TS: Timestamp .
Initialization : $PIF = \{ \phi \}, TFF = \{ \phi \}$
while N *in* $\mathbb{C}$ **do**
    $NWPP$ = find_parsing_information (N)
    $PIF \leftarrow PIF \cup NWPP$
end
while $NWPP$ *in* PIF **do**
    SoN = find_subject_information ($NWPP$)
    VoN = find_verb_information ($NWPP$)
    OoN = find_object_information ($NWPP$)
    Key $\leftarrow$ (SoN + VoN)
    Value $\leftarrow$ (OoN + TS)
    $TFF \leftarrow \langle key, value \rangle$
end

4.3 SARCASM DETECTION ALGORITHM

This section identifies a sarcastic tweet based on the contradiction between a tweet and its temporal facts (TF). In this approach, a news corpus is used as the set of temporal facts. The procedure of identifying sarcastic tweets is given in Algorithm 2.

Algorithm 2: *Sarcasm_Detection_using_TF*

Data: *dataset* := Corpus of tweets ($\mathbb{C}$), TFF.
Result: *classification*:= Sarcastic or not sarcastic
Notation: T: Tweet, SoT: Subject of Tweet, VoT: Verb of Tweet, OoT: Object of Tweet, $\mathbb{C}$: Corpus, PIF: Parsing information File, $TWPPI$: Tweet-wise_parse_phrase_information, TS: Timestamp
Initialization : $PIF = \{ \phi \}$
while T *in* $\mathbb{C}$ **do**
    $TWPPI$ = find_parsing_information (T)
    $PIF \leftarrow PIF \cup TWPPI$
end
while $TWPPI$ *in* PIF **do**
    SarcasticFlag=1;
    SoT = find_subject_information ($TWPPI$)
    VoT = find_verb_information ($TWPPI$)
    OoT = find_object_information ($TWPPI$)
    $Key \leftarrow (SoT + VoT)$
    $Value \leftarrow (OoT + TS)$
    forms $\langle key, value \rangle$ pair for all tweets in corpus
    if *(key in* TFF*) && (val==TFF[key].value)* **then**
        SarcasticFlag=0;
    end
    return SarcasticFlag;
end

Algorithm 2 takes tweets corpus and temporal fact file (TFF) as the input and finds the parse tree information of every testing tweet. Next, it ex- tracts triplets of every tweet using parse information and generates the <key, value> pair. Further, it compares the <key, value> pair of a testing tweet with one of the <key, value> pair in temporal fact file to identify sarcastic tweet. If both the key and value of both tweet and temporal facts are matching, then the tweet is not sarcastic. Otherwise, tweets are sarcastic.

Let us consider an example shown in Figure 7 to test if a tweet is sarcastic or not. The example consists of a tweet and corresponding related news. The <key, value> pair of given tweet in the example is <(Sourav Bhardwaj, dabaya), (EVM secret code, 10 May)>. The temporal fact from the related news is "there is no secret code in EVMs". The <key, value> pair of related news is <(EVMs, istemal nahi), (secret code, 10 May)>. Hence, given tweet in the example is sarcastic as it contradicts its temporal facts.

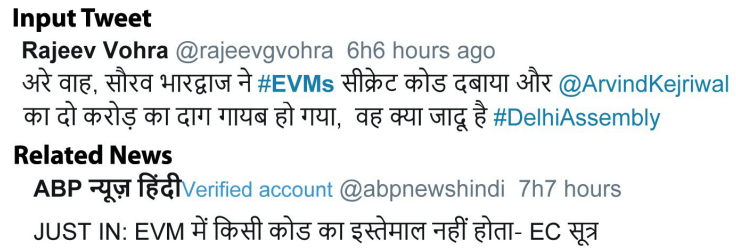

Figure 7: An example of input tweet and related news based on #EVM.

Similarly, let us consider another example of input tweet and related news based on #khelratna as shown in Figure 8. According to proposed algorithm, the <key, value> pair of the given tweet in the example is <(Devendra Jhajharia, Hasil Karne), (KhelRatna, 03 Aug)>. The temporal fact from the related news is "Devendra Jhajharia and hockey player Sardar Singh will get KhelRatna". The <key, value> pair of related news is <(Devendra Jhajharia and Hockey Khiladi Sardar Singh, Ko Milega), (KhelRatna, 03 Aug)>. Hence, given tweet in the example is not sarcastic as tweet infer similar meaning as temporal facts. Here, the meaning of "Hasil Karna" and "Milega" is semantically same in this context of tweet and news.

Figure 8: Example of input tweet and related news based on #khelratna.

5. EXPERIMENTAL RESULTS AND DISCUSSION

This section describes the experimental results of the proposed approach to identify sarcasm in Hindi tweets. There are four statistical parameters considered namely, accuracy, precision, recall and F1-measure to evaluate the performance of proposed approach. The accuracy determines the performance of a system. A formula to calculate accuracy is given in Equation 1.

$$Accuracy = (Tp + Tn) / (Tp + Tn + Fp + Fn) \quad (1)$$

Precision and recall are the other parameters which determine the performance of a system. Here, precision determines how much relevant information system identified. Whereas, recall determines how much identified information is relevant. The formula to ascertain precision and recall appeared in equations 2and 3.

$$Precision = Tp / (Tp + Fp) \quad (2)$$

$$Recall = Tp / (Tp + Fn) \quad (3)$$

The value of precision and recall may vary application to application. For example, an application attains high precision but low recall. Similarly, another application attains low precision but high recall. To deal with this situation, one can rely on another statistical parameter i.e. F1- measure. It is a harmonic mean of precision and recall. To obtain the F1-measure, a formula is given in Equation 4.

$$F1-measure = \frac{2 * Precision * Recall}{Precision + Recall} \quad (4)$$

Where,

T_p = true positive, F_p = false positive, F_n = false negative.

5.1 EXPERIMENTAL RESULTS

A set of 500 random Hindi tweets is used as a testing set to experiment. After the experiment, a confusion matrix of 500 Hindi tweets is given in Table 2 for error analysis. To identify sarcasm in tweets, the proposed system identified 161 tweets correctly out of 184 sarcastic tweets. The ground truth of sarcastic and non-sarcastic tweets are given in Table 1. Using the confusion matrix given in Table 2, the values of precision, recall, F1-measure and accuracy, attained by the proposed approach for identifying sarcasm in tweets are given in Table 3.

Table 2: Confusion matrix for sarcasm detection in Hindi tweets.

Proposed approach	No. of tweets	T_p	T_n	F_p	F_n
Identifying sarcasm	500	161	253	39	47

Table 3: Accuracy, Precision, Recall and F1-measure.

Proposed approach	*Precision*	*Recall*	*F1-measure*	*Accuracy* (%)
Identifying sarcasm	0.805	0.774	0.786	82.8

We made a comparison of the proposed approach with state-of-the-art approaches and is shown in Table 4.

Table 4: Comparison of proposed approach with some of the state-of-the-art techniques.

Study	*Precision*	*Recall*	*F*1-measure	*Accuracy* (%)
Desai *et al.*	0.732	0.674	0.705	71.4
Bharti *et al.*	0.736	0.717	0.726	79.4
Proposed Approach	**0.834**	**0.831**	**0.832**	**82.4**

5.2 DISCUSSION

While identifying sarcasm of Hindi tweets in the context of temporal facts, the algorithm checks for the contradiction between input tweet and the temporal fact obtained from related news in the same timestamp. If the user posts any tweet which negates the temporal fact intentionally on the same timestamp, then the input tweet is sarcastic. Here, another example is discussed and shown in Figure 9 in support of proposed approach.

Figure 9: Another example of input tweet and related news based on #EVM.

The example consists of a tweet and corresponding related news. The <key, value> pair of the given tweet in the example is <(Kalyugi Yudhishthira, hack), (EVM, 10 May)>. The temporal fact from the related news is "EVMs cannot be hacked". The <key, value> pair of related news is <(AAP ka Dava, kharij karna), (EVM hack, 10 May)>. Hence, given tweet in the example is sarcastic as it contradicts its temporal facts.

6. CONCLUSIONS AND FUTURE DIRECTION

In the absence of sufficient annotated sarcastic Hindi tweets for training and testing, a set of sarcastic Hindi tweets were collected and annotated manually. After observing the sarcastic Hindi tweets, we found that most of the sarcastic tweets were written intentionally and it contradicts the time-dependent fact of the same timestamp. Therefore, this article proposed a pattern-based approach to identify sarcastic Hindi tweets using temporal facts and timestamp as the context. Here, a set of online news is treated as temporal facts. The proposed approach attains an accuracy of 82.8%. The performance of the proposed approach is compared with state-of-the-arts techniques for Hindi sarcasm detection and observe that proposed approach outperforms existing approaches. In future, we will use some new #tag trending to collect a large set of tweets as observation set and formulate more pattern for sarcastic tweets in Hindi.

REFERENCES

[1] Chaffey, D. 2016. Global social media research summary 2016. URL http://www.smartinsights.com/social-media-marketing/social-media-strategy/new-globalsocial-media-research.

[2] Tan, W., Blake, M. B., Saleh, I. and Dustdar, S. 2013. Social-network sourced big data analytics. Internet Computing, 17(5), pp. 62-69.

[3] Gastelum, Z. N., Whattam, K. M. 2013. State-of-the- Art of Social Media Analytics Research. Pacific Northwest National Laboratory (PNNL), pp. 1-9.

[4] Lee, Parkvall, M. 2007. Varldens 100 storsta sprak 2007. The Worlds 100, 2007.
[5] Mesthrie, R. 1992. Language in indenture: a sociolinguistic history of Bhojpuri-Hindi. South Africa, Routledge, pp. 30-32. ISBN 978-0415064040.
[6] Language and Culture. 2015. Top 30 Languages by Number of Native Speakers, 2005. http://www.vistawide.com/languages/top 30 languages.htm.
[7] Gonzalez-Ibanez, R., Muresan, S. and Wacholder, N. 2011. Identifying sarcasm in twitter: a closer look. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, vol. 2, pp. 581-586.
[8] Liebrecht, C. Kunneman, F. and van den Bosch, A. 2013. The perfect solution for detecting sarcasm in tweets #not. In proceedings of the Association for Computational Linguistics, pp. 29-37.
[9] Riloff, E., Qadir, A., Surve, P., De Silva, L., Gilbert, N. and Huang, R. 2013. Sarcasm as contrast between a positive sentiment and negative situation. In proceedings of the Empirical methods in natural language processing, pp. 704-714.
[10] Rajadesingan, A., Zafarani, R. and Liu, H. 2015. Sarcasm detection on Twitter: A behavioural modelling approach. In Proceedings of the Eighth ACM Inter- national Conference on Web Search and Data Mining, pp. 97-106.
[11] Joshi, A., Sharma, V. and Bhattacharyya, P. 2015. Harnessing Context Incongruity for Sarcasm Detection. In proceedings of the Association for Computational Linguistics vol. 2, pp. 757-762.
[12] Bharti, S., Sathya Babu, K. and Jena, S. 2015. Parsing- based sarcasm sentiment recognition in Twitter data. In International Conference on Advances in Social Networks Analysis and Mining (ASONAM), pp. 1373-1380, IEEE/ACM.
[13] Joshi, A., Tripathi, V., Patel, K., Bhattacharyya, P. and Carman, M. 2016. Are Word Embedding-based Features Useful for Sarcasm Detection?. ArXiv preprint arXiv: 1610.00883.
[14] Bouazizi, M. and Ohtsuki, T. O. 2016. A pattern-based approach for sarcasm detection on twitter. IEEE Access, vol. 4, pp. 5477-5488.
[15] Bharti, S.K., Vachha, B. Pradhan, R.K., Babu, K.S. and Jena, S.K. 2016. Sarcastic sentiment detection in tweets streamed in real time: a big data approach. Digital Communications and Networks 2(3), pp. 108-121.
[16] Desai, N. and Dave, A.D. 2016. Sarcasm Detection in Hindi sentences using Support Vector machine. International Journal of Advance Research in Computer Science and Management Studies, 4(7), pp. 8- 15.
[17] P. Liu, W. Chen, G. Ou, T. Wang, D. Yang, and K. Lei. 2014. Sarcasm detection in social media based on imbalanced classification. In Web Age Information Management, pp. 459–471.
[18] D. Tayal, S. Yadav, K. Gupta, B. Rajput, and K. Kumari. 2014. Polarity detection of sarcastic political tweets. In proceedings of International Conference on Computing for Sustainable Global Development (INDIACom), pp. 625–628, IEEE.
[19] P. Tungthamthiti, K. Shirai, and M. Mohd. 2014. Recognition of sarcasm in tweets based on concept level sentiment analysis and supervised learning approaches. ACL, pp. 404–413.
[20] S. K. Bharti, R. Pradhan, K. S. Babu, and S. K. Jena. 2017. Sarcasm analysis on twitter data using machine learning approaches. In Trends in Social Network Analysis, pp. 51–76, Springer.
[21] D. Bamman and N. A. Smith. 2015. Contextualized sarcasm detection on Twitter. In ICWSM, pp. 574–577.
[22] Z. Wang, Z. Wu, R. Wang, and Y. Ren. 2015. Twitter sarcasm detection exploiting a context-based model. In International Conference on Web Information Systems Engineering, pp. 77–91, Springer.
[23] R. Schifanella, P. de Juan, J. Tetreault, and L. Cao. 2016. Detecting sarcasm in multimodal social platforms. In Proceedings of the 2016 ACM on Multimedia Conference, pp. 1136–1145, ACM.
[24] A. Khattri, A. Joshi, P. Bhattacharyya, and M. J. Carman. 2015. Your sentiment precedes you: Using an author's historical tweets to predict sarcasm. In Proceedings of the WASSA@ EMNLP, pp. 25–30.
[25] Bharti, S., Sathya Babu, K. and Jena, S. 2017. Harnessing Online News for Sarcasm Detection in Hindi Tweets. Proceeding of PReMI 2017, LNCS, Springer.
[26] Dalal, C., Tandon, S. and Mukerjee, A. 2014. Insult Detection in Hindi. Technical report on Artificial Intelligence, pp. 1-8.
[27] Bharati, A., Sangal, R., Sharma, D. and Bai, L. 2006. Anncorra: Annotating corpora guidelines for POS and chunk annotation for Indian languages. LTRC- TR31, pp. 1-38.
[28] Bharati, A. and Mannem, P.R. 2007. Introduction to shallow parsing contest on south Asian languages. In Proceedings of the IJCAI and the Workshop On Shallow Parsing for South Asian Languages (SPSAL), pp. 1-8.

[29] Rusu, D., Dali, L., Fortuna, B., Grobelnik, M. and D. Mladenic. 2007. Triplet extraction from sentences. In Proceedings of the 10th International Multiconference on Information Society-IS, pp. 8-12.

[30] Doug, et al. 1992. A practical part-of-speech tagger. In Proceedings of the third conference on applied natural language processing. Association for Computational Linguistics, pp. 133-140.

[31] Fleiss, J.L., Cohen, J. and Everitt, B.S. 1969. Large sample standard errors of kappa and weighted kappa. In Psychological Bulletin. American Psychological Association, 72(5), pp. 323-

AUTHORS

Santosh Kumar Bharti is currently pursuing his Ph.D. in CSE from National Institute of Technology Rourkela, India. His research interest includes opinion mining and sarcasm sentiment detection resume.

Korra Sathya Babu is working as an Assistant Professor in the Dept. of CSE, National Institute of Technology Rourkela India. His research interest includes Data engineering, Data privacy, Opinion mining and Sarcasm sentiment detection.