Research Article

Information verification improvement by textual entailment methods

Arefeh Yavary[1] · Hedieh Sajedi[1] · Mohammad Saniee Abadeh[2,3]

Abstract

Information verification is significant because the rate of information generation is high and growing every day, generally in social networks. This also causes social networks to be invoked as a news agency for most of the people. Accordingly, information verification in social networks becomes more significant. Therefore, in this paper, a method for information verification on Twitter is proposed. The proposed method employs textual entailment methods for enhancement of verification methods on Twitter. Aggregating the results of entailment methods in addition to the state-of-the-art methods can enhance the outcomes of tweet verification. In addition, as writing style of tweets is not perfect and formal enough for textual entailment, we used the language model to supplement tweets with more formal and proper texts for textual entailment. Although singly utilizing entailment methods for information verification may result in acceptable results, it is not possible to provide relevant and valid sources for all of the tweets, especially in early times by posting tweets. Therefore, we utilized other sources as a user conversational tree (UCT) besides utilizing entailment methods for tweet information verification. The analysis of UCT is based on the pattern extraction from the UCT. Experimental results indicate that using entailment methods enhances tweet verification.

Keywords Information verification · Textual entailment · Rumor detection

1 Introduction

These days, reffering to social networks, we face to several messages which we not sure, do we trust or believe in them or not. This distrust makes the social networks the unpleasant environment, especially in a crisis, which results in concern between people. Also, by the high rate of data generation in social networks like Twitter, this social media is generally used for getting news. Hence, it is vital and important to check the validity of information, which is spread over the Twitter. Despite the reasons mentioned, in this paper, we are going to check the validity of tweets. By now, some approaches were suggested for rumor detection of tweets. In addition, rumor diffusion is studied in some cases. The main challenge of rumor detection is that there is not available some

reliable and credible source for determining the validation of a tweet, in all cases. Therefore, in our proposed method, we consider two different sources for checking the validity of a tweet. In our proposed method of this paper, we are going to get better income in information validity of tweets. Therefore, we aggregate the textual entailment methods with the results of analysis in a UCT of intended tweet for checking validity in order to enhance the results of information validation. This aggregation is done using a weighted voting classifier on the result of entailment on the tweet and some references and the analysis of the belonging UCT. Furthermore, in our suggested method, we are faced with several challenges. The most important one is that, as the context and writing style of tweets are tidy and also the length of tweets is short, it is hard to get worthy outcomes in using

✉ Hedieh Sajedi, hhsajedi@ut.ac.ir; Arefeh Yavary, yavary_rf@ut.ac.ir; Mohammad Saniee Abadeh, saniee@modares.ac.ir | [1]Department of Computer Science, School of Mathematics, Statistics and Computer Science, College of Science, University of Tehran, Tehran, Iran. [2]Faculty of Electrical and Computer Engineering, Tarbiat Modares University, Tehran, Iran. [3]School of Computer Science, Institute for Research in Fundamental Science (IPM), Tehran, Iran.

tweet in textual entailment methods. Hence, we used a language model in order to make tweets language style more acceptable. Overall, our contribution in this paper for information validation in Twitter is:

- Using textual entailment to enhance rumor detection on Twitter
- Using a language model for making tweets more acceptable in writing style
- Considering subtree in analyzing UCT
- Proposing a weighted voting classifier in order to aggregate the result of entailment method and UCT analysis

In our experiments, we used the just available public benchmark dataset for rumor detection on Twitter. The experimental result shows that our proposed method improved the result of information validation in Twitter with respect to other proposed method, which is tested on the benchmark. In addition, the results show that entailment methods boost the results of information validation. Furthermore, results of information validation using textual entailment are very astonishing. But, as maybe it is not possible to collect valid information sources for all of the tweets, textual entailments must be used in combination with other methods of information verifications like aggregating the results of UCT analysis and texual entailment by a weighted voting classifier.

In the subsequent paper, first we review the related works of rumor detection, textual entailment and voting classifiers. Then, some preliminary knowledge has been stated before expressing suggested approach. Afterward, the results and discussion come to account. At the end, we conclude in conclusion.

2 Literature review

In this part, recent studies in information verification, textual entailments and voting classifier are reviewed.

2.1 Information verification

Ma et al. [1] used multitask learning based on the neural framework for rumor detection. Thakur et al. [2] studied rumor detection on Twitter using supervised learning method. Rumor diffusion was investigated by Li et al. [3]. Majumdar et al. [4] proposed a method for rumor detection in financial scope using a high volume of data. Mondal et al. [5] focused on the fast and timely detection of rumor, which is so important in disaster and crisis.
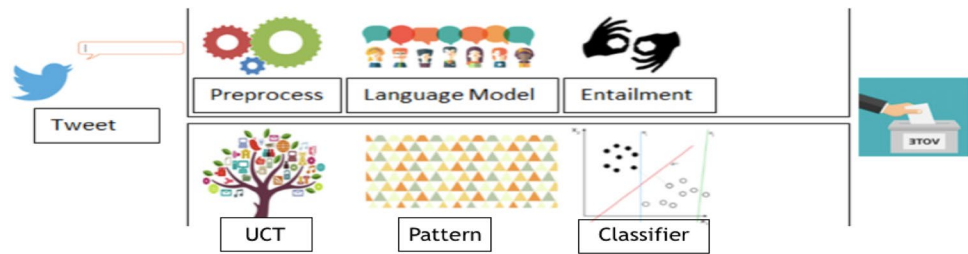
2.2 Textual entailment

In textual entailment methods, we have a hypothesis H and theory T. Our task is to decide whether we can entail H from the T or not. This task can have three or two labels: entailment/nonentailment/contradiction or positive/negative, where positive class equals entailment and negative class equals non-entailment and contradiction. Entailment means that when we can entail T from H, contradiction means H contradicts T and non-entailment means when H gets any conclusion about T [6]. Silva et al. [6] proposed a textual entailment method using the definition of graph. Rocha et al. and Almarwani et al. [7, 8] studied textual entailments in Portuguese and Arabic languages, respectively. Balazs et al. [9] suggested a representation method for sentences to be used for textual entailment in attention-based neural networks. Burchardt et al. [10] annotated a textual entailment corpus using FrameNet.

2.3 Ensemble classifier

Ensemble classifiers are a family of classifiers which ensemble a number of classifiers with an aggregating method to combine the results of the classifier to get a better classifier. The difference between different ensemble classifiers is the difference between their manners of integration [11]. Onan et al. [12] used a weighted voting classifier combined with differential evolution method for sentiment analysis. An online-weighted ensemble classifier for the geometrical data stream is suggested by Bonab et al. [11]. Gu et al. [13] proposed a rule-based ensemble classifier for remote sensing scenes.

2.4 Extreme learning machine

Extreme learning machine (ELM) is a single-hidden-layer feed-forward neural network includes three layers: input layer, hidden layer and output layer. The special property of ELM is its interesting learning method as follows: First, the weights between input and hidden layers are set randomly. Then, using some kind of matrix transpose, the weights of edges between hidden and output layers are computed. ELM can use different activation functions and have several extensions like a kernel extension [14] or multilayer ELM, which has more than one hidden layer [15].

3 Proposed method

Considering different challenges in rumor detection, we consider results in rumor detection for analysis of two sources: (1) textual entailment method on source news, (2) analysis of UCT. Then, for each of these two sources, we train two classifiers separately, and after that, by using weighted ensemble voting classifier, the results of these two separated classifiers are combined to create a new classifier. The process of the proposed method is illustrated in Fig. 1. In the following, each subprocess of the method is described.

3.1 Entailment-based classifier

In this part, modeling the entailment classifier is described. First, the tweets are corrected by language modeling, and then, the textual entailment is utilized on the tweets:

1. *Formal modeling of tweet* As the tweets are short and concise, they are not following the formal English writing style. Because of that, we used a language model [16] to correct the tweet writing style.
2. *Textual entailment* The used entailment methods are as follows [17]:

   - Edit distance involves using algorithms casting textual entailment as the problem of mapping the whole content of *T* into the content of *H*. Mappings are performed as sequences of editing operations (i.e., insertion, deletion and substitution) on text portions needed to transform *T* into *H*, where each edit operation has a cost associated with it. The underlying intuition is that the probability of an entailment relation between *T* and *H* is related to the distance between them [18]. In Ed-RW, the knowledge resource is WordNet.
   - Maximum Entropy classifier combines the outcomes of several scoring functions. The scoring functions extract a number of features at various linguistic levels (bag-of-words, syntactic dependencies, semantic dependencies, named entities). The approach was thoroughly described in [19].

In M-TVT version, the knowledge resource is VerbOcean. In M-TWT, the resource is WordNet, and in M-TWVT, resources are WordNet and VerbOcean.
   - Entailment decision algorithms are at the top level in the entailment core. They compute an entailment decision for a given text/hypothesis (*T/H*) pair and use components that provide standardized algorithms or knowledge resources. The EOP ships with several EDAs. In PRPT, the resource is paraphrase table [17].

The mentioned methods have been implemented in Excitement Open Platform.[1]

3.2 UCT-based classifier

UCT has several uses in different cases. According to that we are working on the context of Twitter, we define UCT in use of Twitter. Our intended UCT [20] is structured as follows: First, the tweet which we want to analyze for veracity is put to the root of the tree. Then, any reply to this tweet is considered as the children of the root. Then, any undirected reply is considered as the child of the tweet, which we reply to. In this way, the UCT is created. After that, each reply in the UCT is labeled corresponding to its opinion with respect to the main tweet and its parents. These labels are Support, Deny, Query and Comment. Support and Deny mean that the reply agrees and disagrees with the corresponding tweet, respectively. Query means that the reply asks for some reference about the main tweet. A Comment means that the reply tweet just gives some comment, without any indication to deny or support of the tweet. For analyzing UCT, two groups of patterns: branched and unbranched patterns, are proposed as the following: (1) unbranched subtree: As the name of the pattern shows, these groups of patterns are subtrees of patterns, which do not have any branches like N-gram. (2) Branched subtree: These patterns are the one which already have at least one branch.

---

[1] http://hlt-services4.fbk.eu/eop/index.php.

Table 1 Results of using entailment methods

Method	Evaluation measures		
	Score	Confidence RMSE	Final score
Ed-RW	0.778	0.947	0.041
M-TVT	0.445	0.929	0.032
M-TWT	0.445	0.925	0.033
M-TWVT	0.445	0.934	0.030
PRPT	**0.778**	**0.629**	**0.289**

3.3 Weighted voting ensemble classifier

In this phase, the results of entailment-based classifier and UCT-based classifier are aggregated using a weighted voting classifier. We used grid search for tuning weights of voting classifier [21].

4 Experiments and discussion

In this section, first the experimental environment is explained. Then, the proposed method is compared with other methods. Then, the experimental results are discussed.

4.1 Experimental Environment

The dataset is defined in [20], and also this dataset is a just publicly available one for rumor detection. The train set contains 137, 62 and 98 tweets for true, false and unverified labels, respectively. The test set contains eight, 12 and eight tweets for true, false and unverified labels, respectively. During preprocessing, UCT analysis and ensemble voting classifier are implemented in Python. Textual entailment method is implemented in Java. ELM method is implemented in MATLAB 2016a.

4.2 Comparing proposed methods with other methods

Methods of comparison are the systems introduced in [20]. Evaluation measures are the same as used in [20]. These measures are score, confidence RMSE and final score. The score is the same as accuracy, confidence RMSE is measured to compute value of confidence error and final score is computed by multiplying score and (1-confidence RMSE).

4.3 Results and discussion

In the following tables, results of just using entailment methods, best results of using UCT patterns in train and

Table 2 Best results of using patterns of UCT in train set

Method	Evaluation measures		
	Score	Confidence RMSE	Score %
Elm-kernel (RBF kernel)	0.960	0.063	0.900
Elm-kernel (linear kernel)	0.467	0.846	0.072
Elm-sine	0.971	0.037	0.935
Elm-RBFs	0.960	0.063	0.900
Elm-sine	0.971	0.037	0.935
Elm- RBFs	0.960	0.063	0.900
Elm-tribas	0.971	0.037	0.935
Multinomial Naive Bayes	0.684	0.478	0.357
Support vector machine	0.820	0.180	0.672
Multilayer perceptron	0.184	0.816	0.034

Table 3 Best results of using patterns of UCT in test set

Method	Evaluation measures		
	Score	Confidence RMSE	Final score
Elm-kernel (RBF kernel)	0.536	0.607	0.210
Elm-kernel (Linear kernel)	0.321	0.679	0.103
Elm-sig	**0.642**	**0.301**	**0.424**
Elm-hardlim	0.607	0.536	0.282
Elm-sine	**0.642**	**0.301**	**0.424**
Elm–RBFs	0.607	0.536	0.282
Elm-tribas	0.643	0.301	0.424
Multinomial Naive Bayes	0.500	0.679	0.161
Support vector machine	0.429	0.571	0.184
Multilayer perceptron	0.500	0.679	0.161
DFKI DKT	0.393	0.845	0.061
ECNU	0.464	0.736	0.122
IITP	0.286	0.807	0.055
IKM	0.536	0.736	0.142
NileTMRG	0.536	0.672	0.176
Baseline	0.571	–	–

Table 4 Results in combination using entailment and patterns

Method	Evaluation measures		
	Score	Confidence RMSE	Final score
Elm + entailment	**0.714**	**0.401**	**0.428**
DFKI DKT	0.393	0.845	0.061
ECNU	0.464	0.736	0.122
IITP	0.286	0.807	0.055
IKM	0.536	0.736	0.142
NileTMRG	0.536	0.672	0.176
Baseline	0.571	–	–

Fig. 2 Diagram for comparison of different entailment methods
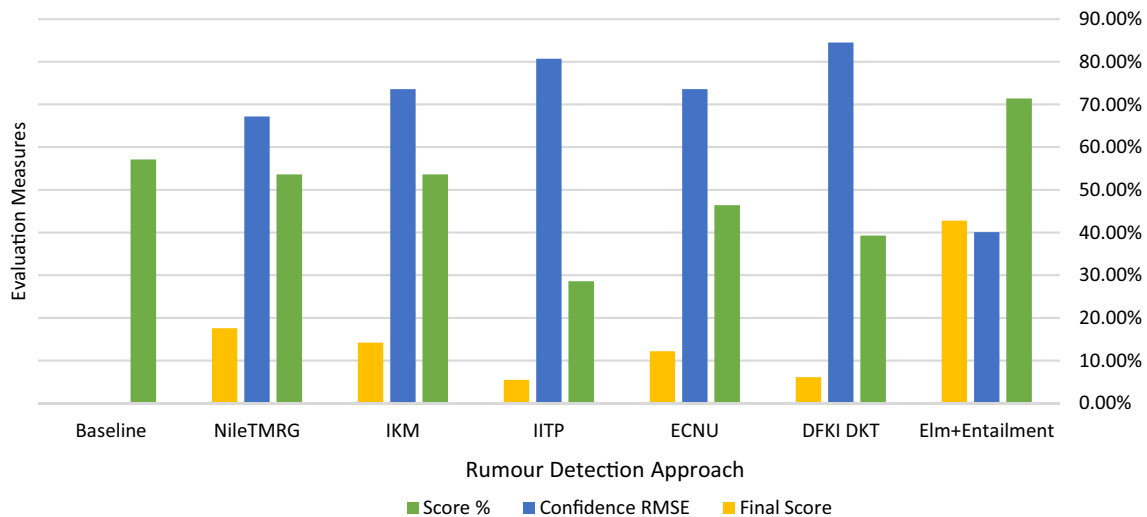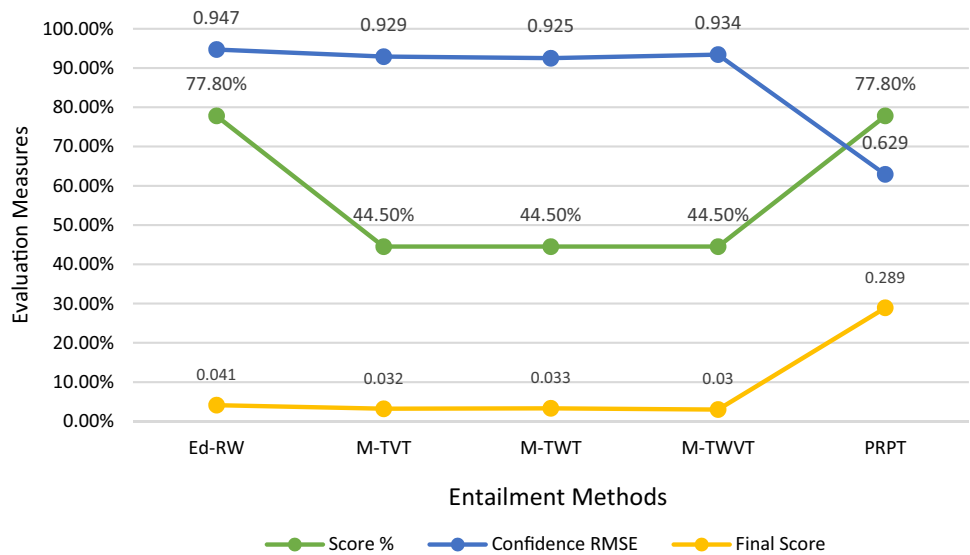
Fig. 3 Comparison of different rumor detection methods in different evaluation measures

test sets, results of combination of entailment methods and patterns are presented in Tables 1, 2, 3 and 4, respectively. In each of the mentioned tables, best results are shown in bold case. In Tables 3 and 4, the results of systems for comparison are shown in gray cell. As Table 1 shows, the best result for textual entailment is P1EDA method. A recent Twitter increases the length of tweets because of the fact that some languages have larger encoding of characters; this results in longer tweets which make an easier analysis of tweets in textual entailment. In addition, in analyzing UCT, as tweets are short, we can suppose that each tweet just gets one of the labels of Support, Deny, Query and Comment. Also, normalizing the patterns with the maximum length of the pattern category, could be useful for affecting the long patterns, too. Parameters

like the time interval between posting replies could be an important feature, too. In Fig. 2, the diagram for comparison of different entailment methods is illustrated. As we can see, P1EDA results are better compared to other methods. In Fig. 3, different rumor detection methods are comprised. The results show that ELM combined with entailment methods achieved the best results.

5 Conclusion and future works

Rumor detection is a hot and open research area. This research topic is very challenging, especially because there is no reliable source for determining the validity of all of the tweets. In addition, these days rumors are

mainly spreading through social networks. Between different social networks, Twitter is more disposed for rumor spreading, because of the high rate of information generation rate and the length of the tweet. Therefore, we selected Twitter as the social media for rumor detection study. By the challenge of rumor detection, we consider two kinds of resources for rumor detection, which are user feedbacks and news resources. Our method is analyzing UCT and entailment method to consider the sources for rumor detection, respectively. Also, as tweets are somehow untidy, we used the language model to clean the tweets in entailment methods. Then, the results of them are aggregated using an ensemble classifier. Experimental results of our method on the benchmarks in rumor detection show that our method has overpassed the state-of-the-art methods. To continue our method in the future, we propose to extend our method by studying more special patterns in UCTs and special entailment methods.

Compliance with ethical standards

Conflict of interest The author declares that they have no conflict of interest.

References

1. Ma J, Gao W, Wong K-F (2018) Detect rumor and stance jointly by neural multi-task learning. In: Companion of the the web conference 2018—WWW 18
2. Thakur HK, Gupta A, Bhardwaj A, Verma D (2018) Rumor detection on twitter using a supervised machine learning framework. Int J Inf Retr Res 8(3):1–13
3. Li D, Gao J, Zhao J, Zhao Z, Orr L, Havlin S (2018) Repetitive users network emerges from multiple rumor cascades. arXiv preprint arXiv:1804.05711
4. Majumdar A, Bose I (2018) Detection of financial rumors using big data analytics: the case of the Bombay Stock Exchange. J Organizational Comput Electron Commer 28(2):79–97
5. Mondal T, Pramanik P, Bhattacharya I, Boral N, Ghosh S (2018) Analysis and early detection of rumors in a post disaster scenario. Inf Syst Front 20:961–979
6. Silva VS, Freitas A, Handschuh S (2017) Recognizing and justifying text entailment through distributional navigation on definition graphs. In: Thirty-second AAAI conference on artificial intelligence AAAI
7. Rocha G, Cardoso HL (2018) Recognizing textual entailment: challenges in the Portuguese language. Information 9(4):76
8. Almarwani N, Diab M (2017) Arabic textual entailment with word embeddings. In: Proceedings of the third arabic natural language processing workshop
9. Balazs J, Marrese-Taylor E, Loyola P, Matsuo Y (2017) Refining raw sentence representations for textual entailment recognition via attention. In: Proceedings of the 2nd workshop on evaluating vector space representations for NLP
10. Burchardt A, Pennacchiotti M (2017) FATE: annotating a textual entailment corpus with FrameNet. Handbook of linguistic annotation, pp 1101–1118
11. Bonab HR, Can F (2018) GOOWE: geometrically optimum and online-weighted ensemble classifier for evolving data streams. ACM Trans Knowl Discov Data 12(2):1–33
12. Onan A, Korukoğlu S, Bulut H (2016) A multiobjective weighted voting ensemble classifier based on differential evolution algorithm for text sentiment classification. Expert Syst Appl 62:1–16
13. Gu X, Angelov PP, Zhang C, Atkinson PM (2018) A massively parallel deep rule-based ensemble classifier for remote sensing scenes. IEEE Geosci Remote Sens Lett 15(3):345–349
14. Huang G-B (2014) An insight into extreme learning machines: random neurons, random features and kernels. Cogn Comput 6(3):376–390
15. Yang Y, Wu QMJ (2016) Multilayer extreme learning machine with subnetwork nodes for representation learning. IEEE Trans Cybern 46(11):2570–2583
16. Ng AH, Gorman K, Sproat R (2017) Minimally supervised written-to-spoken text normalization. In: 2017 IEEE automatic speech recognition and understanding workshop (ASRU)
17. Magnini B, Zanoli R, Dagan I, Eichler K, Neumann G, Noh T-G, Padó S, Stern A, Levy O (2014) The excitement open platform for textual inferences. In: Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations
18. Kouylekov M, Magnini B (2005) Recognizing textual entailment with tree edit distance algorithms. In: Proceedings of the first PASCAL challenges workshop on recognising textual entailment, Southampton, UK, pp 17–20
19. Wang R, Neumann G (2007) Recognizing textual entailment using a subsequence kernel method. In: Proceedings of AAAI, Vancouver, BC, pp 937–945
20. Derczynski L, Bontcheva K, Liakata M, Procter R, Hoi GWS, Zubiaga A (2017) SemEval-2017 Task 8: RumourEval: determining rumor veracity and support for rumours. In: Proceedings of the 11th international workshop on semantic evaluation (SemEval-2017)
21. Lavalle SM, Branicky MS (2004) On the relationship between classical grid search and probabilistic roadmaps. In: Springer tracts in advanced robotics algorithmic foundations of robotics V, pp 59–75