ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

IJARCCE

International Journal of Advanced Research in Computer and Communication Engineering

PCON-2019

National Conference and Seminar on Innovations in Engineering & Technology

Govt Polytechnic College, Palakkad

Vol. 8, Special Issue 1, January 2019

Detection and Verification of Rumour in Social Media: A Survey

Pavithra C P[1], Shibily Joseph[2]

PG Student, Dept of Computer Science and Engineering, Government Engineering College, Palakkad, Kerala, India[1]

Associate Professor, Department of Computer Science and Engineering, Government Engineering College, Palakkad, Kerala, India[2]

Abstract: Social media have gained wide popularity as a medium that enable the users to extract information and breaking news. But all the information that spreading across this social media is accurate. One of the features that characterizes social media is the rapid emergence and spread of new information. This leads to the circulation of rumours. A rumour is defined as an unverified or unconfirmed statement or report circulating in a community. The rumour resolution process has been defined as a pipeline involving four sub-tasks: (1) rumour detection, determining whether a claim is worth verifying rather than the expression of an opinion; (2) rumour tracking, collecting sources and opinions on a rumour because it unfolds; (3) stance classification, determining the attitude of the sources or users towards the truthfulness of the rumour, and (4) rumour verification, as the ultimate step where the veracity value of the rumour is predicted. Here express the rumour resolution process as a multitask problem that needs to address a number of challenges, where the veracity classification task is the main task and therefore the remainder of the parts area unit auxiliary tasks that can be leveraged to boost the performance of the veracity classifier. Multitask learning refers to the joint training of multiple tasks, which has gained popularity recently for a range of tasks in Machine Learning and Natural Language Processing and has been connected in various diverse errands and machine learning structures. Its adequacy is essentially credited to learning shared portrayals of firmly related assignments. Here rumour verification is the main task and others are considered to be auxiliary tasks.

Keywords: Rumour Detection, Rumour Verification, Stance Classification, Multi-task Learning

I. INTRODUCTION

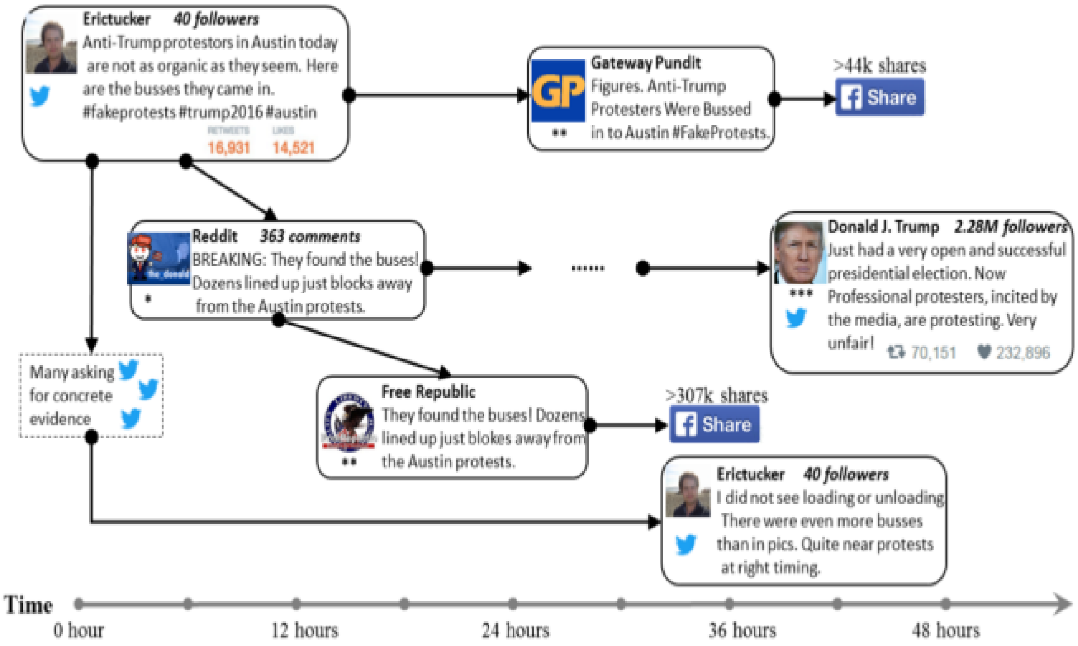
Social media platforms are increasingly being used as a tool for gathering information about, for example, societal issues and to find out about the latest developments during breaking news stories. This is possible because these platforms enable anyone with an internet connected device to share in real-time about their thoughts and/or to post an update about an unfolding event that they may be witnessing. Hence, social media has become a powerful tool for journalists but also for ordinary citizens [6].

Rumour Verification generally means to verify the authenticity of a rumour. A rumour is a piece of information which may or might not be true, so it is very important to determine whether it is fake or genuine before believing in it. Fig. 1. shows that how the rumour spread across the social media.

Therefore, a rumour resolution system has to undergo a set of steps, from detecting that a new circulating claim is a rumour, to the ultimate step of determining its veracity value.

The rumour resolution method has been outlined as a pipeline involving four sub-tasks :

- Rumour detection: decisive whether or not a claim is value verifying rather than the expression of an opinion.
- Rumour tracking: assembling sources and opinions on a rumour as it unfolds.
- Stance classification: determining the attitude of the sources or users towards the truthfulness of the rumour.
- Rumour verification: as the ultimate step where the veracity value of the rumour is predicted.


Erictucker 40 followers
Anti-Trump protestors in Austin today are not as organic as they seem. Here are the busses they came in. #fakeprotests #trump2016 #austin
RETWEETS 16,931 LIKES 14,521

Gateway Pundit
Figures. Anti-Trump Protesters Were Bussed in to Austin #FakeProtests.

Reddit 363 comments
BREAKING: They found the buses! Dozens lined up just blocks away from the Austin protests.

Donald J. Trump 2.28M followers
Just had a very open and successful presidential election. Now Professional protesters, incited by the media, are protesting. Very unfair!
70,151 232,896

Many asking for concrete evidence

Free Republic
They found the buses! Dozens lined up just blokes away from the Austin protests.

Erictucker 40 followers
I did not see loading or unloading There were even more busses than in pics. Quite near protests at right timing.

Fig. 1. Example of how the rumour spread in social media

The applications of this include PHEME, Emergent, Rumor Lens, Rumour Flow, Hoaxy. etc. The main objective of this work is a rumor is an unverified claim about any event, transmitting from person to person. It may refer to an incident, object or problem of public concern. Numerous recent studies have looked at characterizing the emergence and spread of rumours in social media. Insights from these studies can, in turn, be helpful to tell the event of rumour classification systems Rumour classification is a complex task that can be represented as a pipeline of sub-tasks. While there are different possibilities for structuring this pipeline a rumour classification system is expressed as a sequence of subtasks, namely rumour detection, rumour tracking, rumour stance classification leading to rumour verification. Rumour detection, stance and verification, are the main tasks and rumour tracking does not involve classification but rather consists of collecting tweets following up on a rumour in the form of replies [12].

The major challenges of this work is :

- Non-structural, incomplete and noisy nature of online data.
- Semantics understanding: Most rumours are deliberately fabricated to misleading the public.
- Huge variations: Rumours can cover all kinds of topics and take various language styles.
- Heterogeneous propagation structure: During the diffusion of rumours on the social network, users can discuss and make comments.

II. RUMOUR DETECTION AND VERIFICATION

Automatic resolution of rumours may be a difficult task that can be attenuated into smaller parts that create up a pipeline, including rumour detection, rumour tracking and stance classification, resulting in the ultimate outcome of determining the veracity of a rumour. The problem of rumour detection and verification can be formally defined as: Given a twitter information N, the goal of rumour detection and verification is to model and discover the twitter content in N and to classify it is either rumour or not.

A. Different Approaches for rumour detection and Verification

There are different approaches for the task of rumour detection. The work by Zubiaga et.al. [2], deals with stance classification. Rumour stance classification, defined as classifying the stance of specific social media posts into one of supporting, denying, querying or commenting on an earlier post, is becoming of increasing interest to researchers [2]. While most work has focused on using individual tweets as classifier inputs, this work report on the performance of sequential classifiers that exploit the discourse features inherent in social media interactions or conversational threads.

Testing the effectiveness of four sequential classifiers Hawkes Processes, Linear-Chain Conditional Random Fields (Linear CRF), Tree Structured Conditional Random Fields (Tree CRF) and Long Short Term Memory networks (LSTM) on eight datasets associated with breaking news stories, and looking at different types of local and contextual features, this work sheds new light on the development of accurate stance classifiers. Fig. 2. is an example of a tree-structured conversation, with two overlapping branches highlighted.

Main Contribution includes:

- Perform an analysis of whether or not and also the extent to that use of the sequential structure of conversational threads can improve stance classification in comparison to a classifier that determines a tweets stance from the tweet in isolation.
- Perform a close analysis of the results broken down by dataset and by depth of tweet within the thread, as well as an error analysis to further understand the performance of the different classifiers.

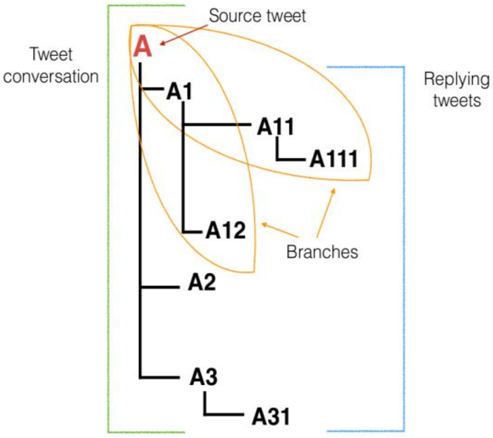
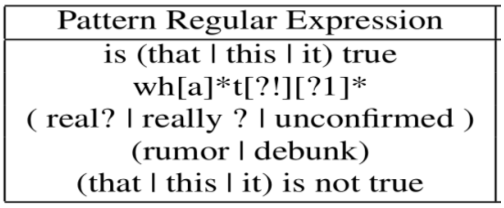


Fig. 2. Example of a tree-structured conversation, with two overlapping branches highlighted

The work by Zhao et.al, present a technique to identify trending rumours that we tend to outline as topics that embrace disputed factual claims [3]. Putting aside any attempt to assess whether the rumours are true or false, it is valuable to identify trending rumours as early as possible. This is a new way to detect rumours as early as possible in their life cycle. The new method utilizes the enquiry behaviour of social media users as sensors. The key insight is that some people who are exposed to a rumour, before deciding whether or not to believe it will take a step of information enquiry to seek more information or to express skepticism without asserting specifically that it is false. Some of them can build their enquiries by tweeting. Fig. 3. represents some regular expression used to find enquiry phrases.

This technique is based on searching for the enquiry phrases, clustering similar posts together, and then collecting related posts that don't contain these simple phrases.



Pattern Regular Expression
is (that | this | it) true
wh[a]*t[?!][?1]*
(real? | really ? | unconfirmed)
(rumor | debunk)
(that | this | it) is not true

Fig. 3. Patterns used to filter Enquiries

For rumour detection another work by [4] compare a novel approach using Conditional Random Fields that learns from the sequential dynamics of social media posts with the present state-of-the-art rumour detection system, as well as other baselines. In contrast to existing work, this classifier does not need to observe tweets querying the stance of a post to deem it a rumour but, instead, exploits context learned throughout the event. For which two types of features are considered: content-based features and social features, testing them individually as well as combined. These two types of features are intended to capture the role that both textual content and user behaviour play in the detection of rumours.

Content-based Features: Consider seven different features extracted from the content of the tweets:

- Word Vectors: to create vectors representing the words in each tweet, to built word vector representations using Word2Vec.
- Capital Ratio: the ratio of capital letters among all alphabetic characters in the tweet.
- Use of Period: Punctuation may be indicative of good writing and hence careful reporting.
- Word Count: the number of words in the tweet, counted as the number of space-separated tokens.

Social Features: Consider five social features, all of which can be inferred from the data related to author of the tweet.

- Tweet Count: Inferred this feature from the number of tweets a user had posted on Twitter.
- Follow Ratio: Looked at the reputation of a user as reflected by their number of followers.
- Verified: a binary feature representing if the user had been verified by Twitter or not.

The work done by Omar et.al[5] is for determining rumour veracity and support for rumours. They have participated in 2 subtasks: SDQC (Subtask A) which deals with tracking how tweets orient to the accuracy of a rumourous story, and Veracity Prediction (Subtask B) which deals with the goal of predicting the veracity of a given rumour [5]. This is a closed task variant, in which the prediction is made solely from the tweet itself. For subtask A, linear support vector classification was applied to a model of bag of words, and the help of a naive Bayes classifier was used for semantic feature extraction. For subtask B, a similar approach was used. Many features were used during the experimentation process but only a few proved to be useful with the data set provided. Below is the complete list of features which apply for both subtask A and B.

- Question Existence
- Denial term detection
- Support words detection
- Hashtag Existence
- Source tweet user is verified

The Elena Kochkina et.al [1] while there are different possibilities for structuring this pipeline, here adopt the architecture of a rumour classification system shown in Fig. 4. The roles of important components are:

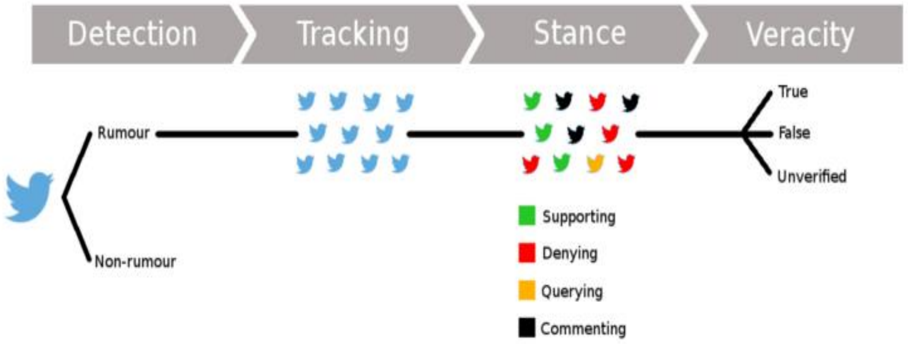


Fig. 4. Rumour Resolution Pipeline

- Rumour detection: In the first instance, a rumour classification system has to identify whether a piece of information constitutes a rumour. A typical input to a rumour detection part will be a stream of social media posts, whereupon a binary classifier must verify if each post is deemed a rumour or a non-rumour.
- Rumour tracking: Once a rumour is known, either because it is known a priori or because it is identified by the rumour detection component, the rumour tracking component collects and filters posts discussing the rumour.
- Stance classification: While the rumour tracking component retrieves posts related to a rumour, the stance classification component determines how each post is orienting to the rumours veracity.
- Veracity classification: The final veracity classification component attempts to determine the actual truth value of the rumour. It can use as input the set of posts collected in the rumour tracking component, as well as the stance labels produced in the stance classification component. In this approach 2 models are compared.

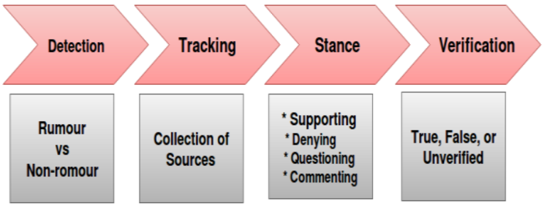


Fig. 5. Rumour Resolution Pipeline

- Sequential approach: The benefits of using a sequential approach were suggested by previous studies of rumour stance classification and rumour detection tasks following the branchLSTM approach. First split the conversations into linear branches and use them as training instances that become an input to a model consisting of an LSTM layer followed by several dense ReLU layers and a softmax layer that predicts class probabilities.

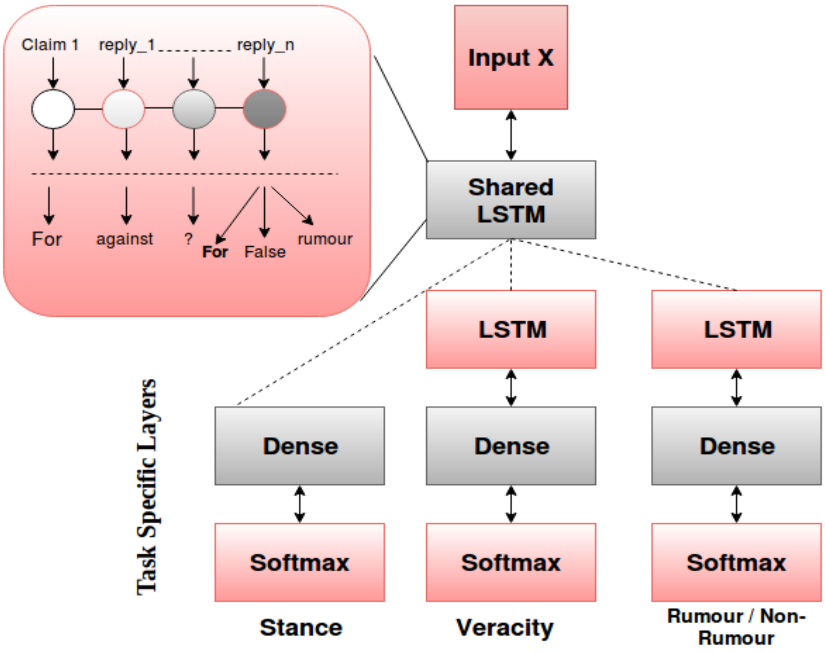


Fig. 6. Multi-task learning models

- Multi-task learning approach: Leverage the relationship between the tasks from the rumour classification pipeline in a joint multi-task learning setup. Figure [6] gives a detailed step-by-step procedure of Multi-task learning. At the base of it is a sequential approach, represented by a shared LSTM layer (hard parameter sharing), which is followed by a number of task-specific layers. The experiment is performed in three set ups: joint training of Stance with veracity classification, Rumour detection with veracity classification, Learning all three tasks together.

Conclusion and Future Work

Research on the development of rumour detection and verification tools has be- come increasingly popular as social media penetration has increased, enabling both ordinary users and professional practitioners to gather news and facts in a real-time fashion, but with the problematic side effect of the diffusion of information of un- verified nature. This survey article has summarised studies reported in the scientific literature toward the development of rumour classification systems, defining and characterising social media rumours, and has described the different approaches to the development of their four main components: (1) rumour detection, (2) rumour tracking, (3) rumour stance classification, and (4) rumour veracity classification. In so doing, the survey provides a guide to the state of the art in the development of these components. The survey has focused particularly on the classification of rumours circulating in social media. Most of the general aspects, such as rumour definition and the classification architecture, etc. However, the specific approaches described for each of the four components are usually designed for social media and are not necessarily directly applicable to other genres.

Further improvements in these work includes:

- In recent years, research in rumour classification has largely focused on the later stages of the pipeline, namely rumour stance classification and veracity classification. Future research should focus on rumour detection and tracking. Further research in this direction would then enable development of entirely automated rumour classification systems.
- An important limitation toward the development of rumour classification systems has been the lack of publicly available datasets. So encourage researchers to release their own datasets so as to enable further research over different datasets and so enable the scientific community to compare their approaches with one another.
- When it comes to stance classification, recent work has shown the effectiveness of leveraging context in social media streams and conversations to develop state-of-art classifiers for the stance of individual posts. Research in this direction is, however, still in its infancy and more research is still needed to best exploit this context for maximising the performance of stance classifiers

Despite substantial progress in the research field, as shown in this survey, also show that this is still an open research problem that needs further study.

REFERENCES

[1]. Elena Kochkina, Maria Liakata, Arkaitz Zubiaga, All-in-one: Multitask Learn- ing for Rumour Verification , Proceedings of the 27th International Conference on Computational Linguistics, August 20-26, 2018.
[2]. Zubiaga, Elena Kochkina, Maria Liakata, Discourse-Aware Rumour Stance Clas- sification in Social Media Using Sequential Classifiers , Information Processing Management, 2018.
[3]. Zhe Zhao, Paul Resnick, Qiaozhu Mei, Enquiring Minds: Early Detection of Ru- mors in Social Media from Enquiry Posts , In Proceedings of the 24th International Conference on World Wide Web, 2016.
[4]. Arkaitz Zubiaga, Maria Liakata, and Rob Procter, Exploiting Context for Rumour Detection in Social Media, In International Conference on Social Informatics, 2017.
[5]. Omar Enayet and Samhaa R El-Beltagy, Niletmrg at semeval-2017 task 8: De- termining rumour and veracity support for rumours on twitter , In Proceedings of the 11th International Workshop on Semantic Evaluation, 2017.
[6]. Gustavo Aguilar, Suraj Maharjan, Adrian Pastor Lpez Monroy, and Thamar Solorio, A multi-task approach for named entity recognition in social media data, In Proceedings of the 3rd Workshop on Noisy User-generated Text, 2017.
[7]. Tong Chen, Lin Wu, Xue Li, Jun Zhang, Hongzhi Yin, and Yang Wang. Call at- tention to rumors: Deep attention based recurrent neural networks for early rumor detection. arXiv:1704.05973, 2017.
[8]. Michal Lukasik, PK Srijith, Duy Vu, Kalina Bontcheva, Arkaitz Zubiaga, and Trevor Cohn. Hawkes processes for continuous time sequence classification: an application to rumour stance classification in twitter. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), volume 2, pages 393398, 2016.
[9]. Jing Ma, Wei Gao, Prasenjit Mitra, Sejeong Kwon, Bernard J Jansen, Kam-Fai Wong, and Meeyoung Cha. Detecting rumors from microblogs with recurrent neural networks. In IJCAI, pages 38183824, 2016.
[10]. Arkaitz Zubiaga, Maria Liakata, Rob Procter, Geraldine Wong Sak Hoi, and Peter Tolmie. Analysing how people orient to and spread rumours in social media by looking at conversational threads . PloS one, 11(3):e0150989, 2016.
[11]. Arkaitz Zubiaga, Ahmet Aker, Kalina Bontcheva, Maria Liakata, and Rob Proc- ter. Detection and resolution of rumours in social media: A survey ACM Comput. Surv., 51(2):32:132:36, February,2018a.
[12]. Gustavo Aguilar, Suraj Maharjan, Adrian Pastor Lpez Monroy, and Thamar Solorio. A multi-task approach for named entity recognition in social media data. In Proceedings of the 3rd Workshop on Noisy User-generated Text, pages 148153, 2017.
[13]. Arkaitz Zubiaga, Elena Kochkina, Maria Liakata, Rob Procter, Michal Lukasik, Kalina Bontcheva, Trevor Cohn, and Isabelle Augenstein. 2018b. Discourse-aware rumour stance classification in social media using sequential classifiers. Information Processing Management, 54(2):273290.
[14]. Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu. 2017. Fake news detection on social media: A data mining perspective. ACM SIGKDD Explorations Newsletter, 19(1):2236.
[15]. Marcelo Mendoza, Barbara Poblete, and Carlos Castillo. 2010. Twitter under crisis: Can we trust what we RT? In 1st Workshop on Social Media Analytics, SOMA10, pages 7179.
[16]. Man Lan, Jianxiang Wang, Yuanbin Wu, Zheng-Yu Niu, and Haifeng Wang. 2017. Multi-task attention-based neural networks for implicit discourse relationship representation and identification. In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, pages 12991308.