Stance Classification for Rumour Verification in Social Media Conversations

Elena Kochkina[1,2], Maria Liakata[1,2,3], Arkaitz Zubiaga[3]
[1] University of Warwick, Coventry, United Kingdom
[2] Alan Turing Institute, London, United Kingdom
[3] Queen Mary University of London, London, United Kingdom

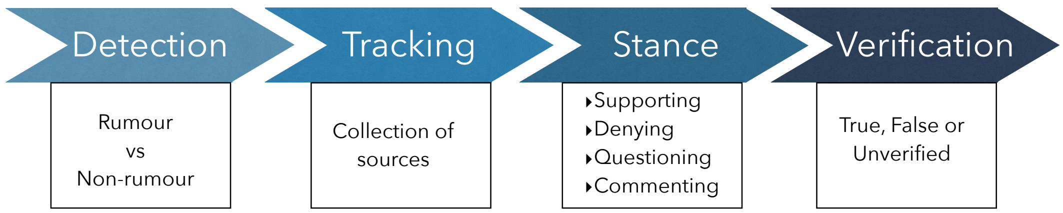

Figure 1: Rumour Verification Pipeline [15]

KEYWORDS

rumours, rumour verification, stance classification, neural networks

Social media platforms have gained popularity as news sources, often delivering updates faster than traditional media. However, unfiltered malicious posts can have a significant negative impact, that has highlighted the importance of fact-checking and information verification on social media. At the time of writing this, the public is facing an 'infodemic', wide spread of rumours and conspiracy theories regarding COVID-19 and the associated pandemic. It can create panic, affect rates of transmission; encourage trade in untested treatments, effectively putting people's lives in danger[1]. Thus, the WHO, governments and platforms have to assign significant resources to combat the infodemics. While various initiatives have been launched by journalists in recent years to address this problem (e.g. www.emergent.info, fullfact.org), manual fact-checking and verification cannot scale to address the amount of unverified information (rumours) circulating and cannot be easily performed in real-time.

Due to the risks posed by the proliferation of unverified content online, there is a need to develop Machine Learning (ML) methods to assist with the verification of circulating rumours, statements unverified at the time of posting. Rumour verification can be formulated as a classification problem, where a model is trained to predict if a rumour is true, false or unverified, given posts discussing a rumour as the input. Rumour verification is a time-sensitive problem that involves a set of subtasks, as proposed in Zubiaga et al. [15]. These could be seen as a pipeline or can be learnt jointly and in different combinations.

Figure 1 illustrates the possible sequence of tasks in this pipeline : (a) Rumour Detection: identifying check-worthy stories of unverified veracity status that are spreading widely; (b) Rumour tracking: collecting all relevant sources and responses to a particular rumour; (c) Rumour stance classification: identifying the attitude of users towards the truthfulness of the rumour as either Supporting, Denying, Querying or Commenting (i.e. not addressing rumour veracity); (d) Rumour verification: determining if a rumour is true, false or remains unverified.

Previous works have explored a variety of approaches to automated rumour verification using linguistic features [1]; user information and their social network connections [8]; incorporating temporal and structural propagation features [12]; media and images [5]; as well as external information [13] in order to find and exploit features that are indicative of the truthfulness of a rumour. In particular, a discussion around a rumour, in which users share their opinion, links to extra sources and evidence, can be proven useful. An example of such discussion around a false rumour is shown in Figure 2. Previous research [14] has shown that rumours attracting a lot of sceptical and denying reactions are more likely to be proven false later. Thus classifying the stance of posts towards rumours automatically is an important task that aids rumour verification. We propose a talk discussing the relation between the tasks of rumour stance and veracity classification in social media conversations, giving the overview of recent advances leveraging that relation based on our work in this domain and experience from organising a shared task.

The RumourEval shared task [2, 4] was proposed to test the hypothesis regarding the synergy between stance and rumour veracity. RumourEval consists of 2 sub-tasks: (A) rumour stance classification and (B) rumour veracity classification, where the input is a

[1]https://www.un.org/en/un-coronavirus-communications-team/un-tackling-%E2%80%98infodemic%E2%80%99-misinformation-and-cybercrime-covid-19

MEDIATE '20, June 08–11, 2020,

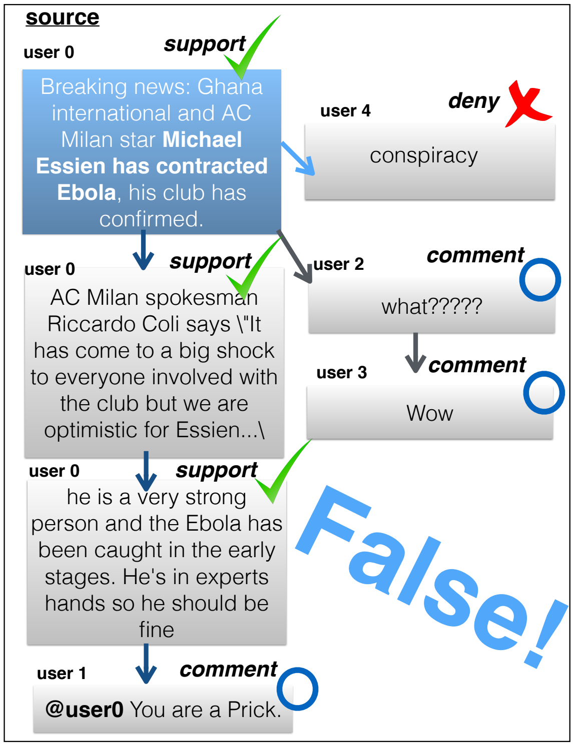

Figure 2: Example of a conversation discussing a rumour

collection of Twitter conversations discussing rumours related to news breaking events. In its first edition in 2017, the winning system of subtask B was the only system that used the predicted stance labels as features for their classifier. We also noticed that improving stance classification has a positive effect on a rumour verification system that utilises stance as a feature [7]. In the second edition of RumourEval in 2019 more systems utilised stance as a feature to help determine the veracity of a rumour. The winning system of subtask B, which outperformed strong baselines (winners of the previous edition) and other competitors, used an ensemble of classifiers and stance extracted from subtask A.

Furthermore, we have explored the incorporation of stance classification into rumour verification as an auxiliary task in a multitask learning set up, when a deep learning model was trained to perform several tasks simultaneously [7]. The results show that the joint learning of two tasks from the rumour verification pipeline outperforms a single-learning approach to rumour verification for RumourEval and larger PHEME dataset. The combination of three tasks (stance classification, detection and verification) leads to further improvements. Independently, Ma et al. [11] came to similar conclusions using datasets from RumourEval, Liu et al. [10] and Fake News Challenge[2]. Dungs et al. [3], proposed a competitive approach using Hidden Markov Models using stance and response posting times as features for rumour verification. Recent work by Lillie et al [9] suggested that stance-based veracity works across languages and platforms. Rumour verification is a complex task as rumours can concern a wide variety of topics; discussions around these rumours use different vocabulary and attract the attention of a variety of audiences. Identifying and measuring the degree of support and denial of users and various sources towards a rumour is a cross-domain feature that may lead to improved generalisability of a rumour verification system.

To conclude we will outline open challenges that rumour verification models are facing, and share our view on how to tackle them. The need to create a reliable system for automated rumour verification poses high requirements to the researchers. The system should be accurate; generalisable to unseen rumours; time-sensitive to update predictions over time; real-time to aim for early predictions; provide justification or explanations for its predictions; inform humans of its uncertainty [6], and be impartial towards biases such as source bias. Absence of these qualities in an automated model leads to a lack of trust from end-users, journalists or the wider public. Automating rumour verification is, therefore, an extremely challenging goal, which calls for collaboration between companies, platforms, journalists, government and researchers.

[2]http://www.fakenewschallenge.org/

REFERENCES

[1] Alton Yeow Kuan Chua and Snehasish Banerjee. 2016. Linguistic predictors of rumor veracity on the Internet. (2016).
[2] Leon Derczynski, Kalina Bontcheva, Maria Liakata, Rob Procter, Geraldine Wong Sak Hoi, and Arkaitz Zubiaga. 2017. SemEval-2017 Task 8: RumourEval: Determining rumour veracity and support for rumours. In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*. 69–76.
[3] Sebastian Dungs, Ahmet Aker, Norbert Fuhr, and Kalina Bontcheva. 2018. Can Rumour Stance Alone Predict Veracity?. In *Proceedings of the 27th International Conference on Computational Linguistics*. 3360–3370.
[4] Genevieve Gorrell, Elena Kochkina, Maria Liakata, Ahmet Aker, Arkaitz Zubiaga, Kalina Bontcheva, and Leon Derczynski. 2019. SemEval-2019 task 7: RumourEval, determining rumour veracity and support for rumours. In *Proceedings of the 13th International Workshop on Semantic Evaluation*. 845–854.
[5] Aditi Gupta, Hemank Lamba, Ponnurangam Kumaraguru, and Anupam Joshi. 2013. Faking sandy: characterizing and identifying fake images on Twitter during hurricane sandy. In *Proceedings of the 22nd international conference on World Wide Web*. ACM, 729–736.
[6] Elena Kochkina and Maria Liakata. 2020. Estimating predictive uncertainty for rumour verification models. In *Proceedings of the 58th annual meeting of the Association for Computational Linguistics (ACL)*.
[7] Elena Kochkina, Maria Liakata, and Arkaitz Zubiaga. 2018. All-in-one: Multi-task Learning for Rumour Verification. In *Proceedings of the 27th International Conference on Computational Linguistics*. 3402–3413.
[8] Quanzhi Li, Qiong Zhang, and Luo Si. 2019. Rumor Detection by Exploiting User Credibility Information, Attention and Multi-task Learning. In *Proceedings of the 57th Conference of the Association for Computational Linguistics*. 1173–1179.
[9] Anders Edelbo Lillie, Emil Refsgaard Middelboe, and Leon Derczynski. 2019. Joint Rumour Stance and Veracity. (2019).
[10] Xiaomo Liu, Armineh Nourbakhsh, Quanzhi Li, Rui Fang, and Sameena Shah. 2015. Real-time rumor debunking on Twitter. In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*. ACM, 1867–1870.
[11] Jing Ma, Wei Gao, and Kam-Fai Wong. 2018. Detect rumor and stance jointly by neural multi-task learning. In *Companion Proceedings of the The Web Conference 2018*. International World Wide Web Conferences Steering Committee, 585–593.
[12] Jing Ma, Wei Gao, and Kam-Fai Wong. 2018. Rumor detection on Twitter with tree-structured recursive neural networks. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Vol. 1. 1980–1989.
[13] Yumeng Qin, Dominik Wurzer, Victor Lavrenko, and Cunchen Tang. 2016. Spotting rumors via novelty detection. *arXiv preprint arXiv:1611.06322* (2016).
[14] Zhe Zhao, Paul Resnick, and Qiaozhu Mei. 2015. Enquiring minds: Early detection of rumors in social media from enquiry posts. In *Proceedings of the 24th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 1395–1405.
[15] Arkaitz Zubiaga, Ahmet Aker, Kalina Bontcheva, Maria Liakata, and Rob Procter. 2018. Detection and resolution of rumours in social media: A survey. *ACM Computing Surveys (CSUR)* 51, 2 (2018), 1–36.