Stance Classification for Rumour Verification in Social Media Conversations

Elena Kochkina¹², Maria Liakata¹²³, Arkaitz Zubiaga³
¹ University of Warwick, Coventry, United Kingdom
² Alan Turing Institute, London, United Kingdom
³ Queen Mary University of London, London, United Kingdom

ABSTRACT
Due to the risks posed by the proliferation of unverified content online, there is a need to develop automated methods to assist with the verification. Previous works have explored a variety of approaches to automated rumour verification. In particular, a discussion around a rumour, in which users may share their opinion and link extra relevant information, can be useful as rumours attracting a lot of sceptical and denying reactions are more likely to be proven false later. Here we discuss 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.

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. 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±

Breaking news: Ghana international and AC Milan star Michael Essien has contracted Ebola, his club has confirmed.

Furthermore, we have explored the incorporation of stance classification into rumour verification as an auxiliary task in a multitask learning setup, 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 Challenge2. 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.

REFERENCES


http://www.fakenewschallenge.org/