

Big Data and quality data for fake news and misinformation detection

Big Data & Society
January–June 2019: 1–14
© The Author(s) 2019
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/2053951719843310
journals.sagepub.com/home/bds



Fatemeh Torabi Asr and Maite Taboada 

Abstract

Fake news has become an important topic of research in a variety of disciplines including linguistics and computer science. In this paper, we explain how the problem is approached from the perspective of natural language processing, with the goal of building a system to automatically detect misinformation in news. The main challenge in this line of research is collecting quality data, i.e., instances of fake and real news articles on a balanced distribution of topics. We review available datasets and introduce the MisInfoText repository as a contribution of our lab to the community. We make available the full text of the news articles, together with veracity labels previously assigned based on manual assessment of the articles' truth content. We also perform a topic modelling experiment to elaborate on the gaps and sources of imbalance in currently available datasets to guide future efforts. We appeal to the community to collect more data and to make it available for research purposes.

Keywords

Fake news, misinformation, labelled datasets, text classification, machine learning, topic modelling

Introduction

Fake news is a problem. It is a Big Data problem. We are trying to solve it with small amounts of data.

Those are, in a nutshell, the three main points of our paper. We will not retread the familiar territory covered by many recent papers, reports and news media articles regarding how fast fake news spreads; why the presence of fake news is a problem because so many people get their news through online sources; and how the inability to trust news is a problem for democracy. We will only provide brief literature reviews of how each of those aspects has been addressed in recent literature. Our focus is on providing technological solutions to a problem that has been not necessarily created, but certainly exacerbated, by technology. We provide a comprehensive account of fake news detection as a text classification problem, to be solved using natural language processing (NLP) tools, and show that, in our experiments with two general classes of algorithms, fake news articles are detectable, especially given enough training data. And this need for data leads to our call to arms to the research community, to news

media and social media companies: We want your fake news data.

In this introduction, we first define and delimit the problem and its historical roots. Then in the section on Approaches to the fake news problem, we discuss general approaches, from multiple points of view (educating the public, stopping the spread, human and automatic identification). The approach we take concentrates on automatic identification by using the *text* of the fake news article (rather than metadata of information about spread). This is why, in text classification for fake news, we introduce text classification methods, including both classic and more recent algorithms used in research on fake news. The following section (Data: where and how to find fake news) discusses the problem of lack of quality data in this case. Although media would have us believe that instances of fake news are

Simon Fraser University, Canada

Corresponding author:

Maite Taboada, Department of Linguistics, Discourse Processing Lab, Simon Fraser University, Burnaby, British Columbia, Canada.
Email: mtaboada@sfu.ca



everywhere, we have found it challenging to compile a large enough dataset of *reliably labelled* fake news articles. We discuss how we have compiled MisInfoText, our relatively large, but still insufficient dataset, and what steps can be taken to add to this data. This repository is built with a focus on quality data collection and based on the continuous effort of fact-checking websites in finding and labelling instances of fake news. We make available the full text of the news articles, together with veracity labels previously assigned based on manual assessment of the articles' truth content. We have conducted experiments on the data that we have so far collected to show the gaps and sources of data imbalance in the topics covered by fact-checkers section. For building text classification systems to help distinguish misinformation from real news, we need big and reliably labelled training data, and that is why in A call to arms we propose ways to add to the data repository that we have built (<https://github.com/sfu-discourse-lab/MisInfoText>).

Let us start, then, with defining the 'fake news problem'. In the most recent incarnation of this problem, and especially during the 2016 US presidential election, the problem refers to the creation and spread of news articles that favoured or attacked one of the two main candidates, Hillary Clinton and Donald Trump. In a more general sense, the issue is one of disinformation (false information that is purposely spread to deceive people) and misinformation (false or misleading information; Lazer et al., 2018), but it also includes the bias that is inherent in news produced by humans with human biases. Lazer et al. (1094) define this most recent phenomenon as 'fabricated information that mimics news media content in form but not in organizational process or intent.' More comprehensive definitions and classifications of different types of falsehoods and disinformation can be found in Jack (2017), Shu et al. (2017), or Wardle (2016),.

Some researchers and media analysts object to the use of the term *fake news*, because of its recent use as a political weapon, when politicians label as fake news a story, or even an entire news media organization, because they dislike what is being said about them (Nielsen and Graves, 2017). We use it because it helps draw attention to the problem (Lazer et al., 2018), and because it is a convenient shorthand. It should be clear, however, that present-day fake news are not only about politics, but about health, celebrities, or aspects of the economy.

Historically, misinformation has been seen as the normal state of affairs, and news sources were routinely considered untrustworthy. Virginia Woolf, in *Three Guineas*, stated that 'if you want to know any fact about politics, you must read at least three different papers, compare at least three different versions of the

same fact, and come in the end to your own conclusion.'¹ The expectation that news articles and news organizations exhibit impartiality is a development of the 20th century (Darnton, 2017; Marwick and Lewis, 2017; Wardle, 2016). We have, perhaps, come to take it for granted, but it has not traditionally been the norm (Lazer et al., 2017).

What makes present-day fake news most alarming is the speed with which it spreads. It is worth noting that concern about the role of technologies in the spread of false, inaccurate or misleading information has quickly followed the invention of such technologies. Darnton (2017) discusses how, in 18th-century London, invented stories or gossip made it into newspapers which had just began to circulate among a broad public. A *Harper's* magazine story in 1925 warned of the vulnerability of news wire services like the Associated Press: 'Once the news faker obtains access to the press wires all the honest editors alive will not be able to repair the mischief he can do.'² Similar concerns were raised after the invention of the printing press. See also Marlin (2002) for a thorough analysis of the history of propaganda.

It is, however, undeniable that social media have enabled the speediest form of spread in this long history. Vosoughi et al. (2018) studied the spread of stories through Twitter, and found that false stories diffused significantly farther, faster, deeper and more broadly than true stories, and, within the false stories, political stories had the fastest and broadest spread. Most of the spread was viral, i.e., it was distributed not centrally, but through peer-to-peer diffusion. There are potential psychological explanations for this, as fake news articles tend to be more novel, or more shocking, and we as humans are attracted to such stimuli, perhaps the result of negativity bias, our tendency to pay more attention to negative events (Rozin and Royzman, 2001). That is why sensationalist stories sell newspapers (Glogger et al., 2016; Sachsman, 2017).

The speed and extent of the spread is probably directly tied to political and financial incentives, both online and offline. The political incentives are clear: Voters (and even non-voters) wish for their preferred candidate to win an election, be successful in political initiatives, or pursue and defend their political agenda. More complex is the role of ideologues, conspiracy theorists and hate groups (Marwick and Lewis, 2017). The financial incentives have complicated matters online, as many fake news producers do so simply for financial gain. The widely discussed case of the 'Macedonian teenagers' illustrates this. Teenagers in a small Macedonian town created website content in many areas, including health, sports, and politics. They found that US politics provided the most ad revenue and, within US politics, pro-Trump stories were the

most profitable, so they diligently set out to post such stories. They plagiarized the content of those stories from American fake news sites (Silverman and Alexander, 2016; Subramanian, 2017).

The effects of fake news in specific situations are being documented. For instance, Allcott and Gentzkow (2017) analysed browsing data, archives of fact-checking websites and online surveys to conclude that social media was an important source of news for many Americans during the 2016 US presidential election. They attempted to quantify the amount of exposure to and engagement with fake news stories, and concluded that those stories were widely shared, and tended to feed confirmation bias: People were more likely to believe stories that favoured their preferred candidate. This is certainly a problem in the specific situation of the election, because the results of a particular election may have been affected by stories that were proven to be false. It is, however, a much more general problem as it promotes the impression that we cannot believe anything we find online, in printed media, or in radio and television broadcasts. That is a problem for democracy, because it erodes trust in public institutions.

Now that we have briefly discussed the problem, its spread and effects, we will move, in the next section, to what approaches have been taken so far in tackling it.

Approaches to the fake news problem

The root causes, the spread and the consequences of fake news are all complex issues. One can take multiple approaches and, indeed, individuals, researchers and organizations have undertaken efforts to address the issue. Lazer et al. (2018) propose interventions along two lines: empowering individuals to evaluate potential fake news they encounter; and structural changes to stop or minimize exposure to such 'news.' We would like to break those down a bit further, into: (1) educating the public; (2) analysing and curtailing the spread; (3) performing manual checking; or (4) performing automatic fact-checking and classification. We agree with Lazer et al. (2018) that this is a problem that requires an interdisciplinary approach. In this section, we provide brief descriptions of the possible interventions, but, in the rest of the paper, will focus on the last one: performing automatic text classification to determine whether a news article seems to contain fake or false information.

Educating the public

Education efforts can be enhanced, starting at the school level, with media literacy, and a general education towards empowering a responsible citizenship,

raised in civil and democratic values, who is also able to understand the competing pressures of capitalist societies, including the influence of lobby groups, political parties, and the simple financial gain of creating online content that generates advertising revenue for the creator (and of course for the hosting site). We should pause and think for a moment what the internet would have been like had it not taken the route of using advertising as a form of revenue.

More focused forms of education concentrate on news and sources of news specifically, such as an infographic prepared by the International Federation of Library Associations and Institutions, which encourages readers to examine the source, read beyond a headline, or ensure that the content is not meant to be humorous or satirical.³ Another excellent initiative is the course designed by Harvard Kennedy School's Shorenstein Centre, aimed at both journalists and the general public, and providing tools to verify information.⁴

Calling for better education, whatever form it takes, is, nonetheless, an easy way out, and one that places undue burden on the individual to acquire such education.

It is not our place to advise governments on how to create and administer education policy. We would like, however, to offer a caveat, that evidence supporting the belief that a higher level of education inures news consumers to outrageous claims is not conclusive. For instance, Allcott and Gentzkow (2017) found that level of education was not statistically significantly associated with how likely readers were to believe an ideologically aligned story (but people with higher education tended to have more accurate beliefs about news). Furthermore, Greenhill and Oppenheim (2017) found that education, income, age and gender, what they describe as commonly cited factors in receptivity to rumours, did not seem to have a correlation with how likely people were to believe a rumour.

Analysing and curtailing the spread

Fake news spread fast. It spreads faster and penetrates social networks to a larger extent than credible news (Mustafaraj and Metaxas, 2017; Vosoughi et al., 2018). This may be due to its novelty, its capacity to generate outrage (which generates attention), or its role in confirming the preexisting biases of the reader. The novelty and outrage may explain why Facebook's effort of flagging debunked fake stories backfired (Constine, 2018). Users actually shared flagged stories more.

Part of the problem resides in echo chambers or filter bubbles, which means that some people will be exposed to only one point of view, and will find it easier to believe stories that reaffirm that point of view

(Bechmann and Nielbo, 2018; Del Vicario et al., 2016; Greenhill and Oppenheim, 2017). This is why, in their agenda for research and action, Lazer et al. (2017) encourage communication online across partisan or ideological lines.

We also know that people tend to remember facts and events that have been repeatedly mentioned, even when the repeated mention is in the context of a retraction or myth debunking (Ecker et al., 2017; Swire et al., 2017). It makes sense, then, to stop false information on its tracks, before it reaches too many people and becomes entrenched in their minds. Research in this area includes linguistic signals of a rumour (Zhao et al., 2015) and models of spread, which help in determining how to contain it, and how many fact-checkers are needed to contain a hoax (Tambuscio et al., 2015). Hoaxy, an open platform to study misinformation and fact-checking on Twitter, is useful in modelling how to disrupt the spread of a rumour (Shao et al., 2018).

Despite popular belief that bots play a crucial role in spreading misinformation, Vosoughi et al. (2018) found that rumours spread with the same speed, depth and reach, whether they originated or were retweeted by either bots or humans. Therefore, while identifying bots may be useful, humans are still a major source of misinformation spread.

Manual checking

Manual checking of false statements, rumours and fake news articles online plays a vital role in containing the spread. Two broad classes of efforts can be identified: using fact-checking websites, and performing manual checking on specific social media sites.

Fact-checking websites (e.g., Snopes, Politifact, Emergent) provide verification of claims that they find, or that users submit. They have the advantage of using qualified journalists and other professionals, who are able to research and verify claims and news stories. They do have some downsides, however. The first one is, as with education, the process makes the responsibility rest with the individual. Lazer et al. (2018) also point out that people may not be likely to fact-check a story that aligns with their pre-existing beliefs. Fact checking could even be counterproductive, as fact-checking a story or a rumour leads to familiarity with it, and familiarity breeds not contempt but acceptance (Berinsky, 2017; Ecker et al., 2017; Pennycook et al., 2018; Swire et al., 2017). Lewandowsky et al. (2012) recommend that, if a myth or rumour is to be debunked, that it not be repeated. The correct facts should rather be reported, without mentioning the false information.

Large technology companies and social media sites have responded to social pressure and the common

belief that they played a role in abetting, or at least not curtailing, the spread of fake news by announcing that they will hire (more) content moderators. Human monitoring is desirable, because it ensures that claims are accurately verified. It has many potential pitfalls, however, ranging from the possibility that moderators' bias will be propagated to the mental toll placed on individuals performing the checking (Chen, 2017). Facebook partnered with fact-checking organizations to reduce and contain the impact of fake news. A recent report on the partnership (Ananny, 2018) documents mixed success, stemming from a lack of common goals. Partners also worried that the effort was not transparent, to them or to the wider public. Pavleska et al. (2018) have also documented the problems with fact-checking organizations, including: lack of coordination among each other; excessive reliance on human expertise without, in some cases, a plan for long-term sustainability; or an absence of measures of impact.

The Credibility Coalition⁵ is developing a framework for *credibility indicators*, signals that help human and automated systems determine whether content is credible (Zhang et al., 2018). The indicators may be within the text (clickbait headline, relationship between headline and text, logical fallacies, emotional tone), or in the publisher's metadata (presence of ads, indicators of sources of revenue, type of outbound links).

Automatic checking

There are clear benefits to performing verification automatically: It can be done at scale and it saves moderators from having to sort through at best unpleasant content. This form of automatic checking is about the content and claims in the story itself, not about metadata such as source or rate of spread.

Computational fact-checking attempts to find unverified claims in a story or rumour, and check them against reliable sources. Ciampaglia et al. (2015) find factual information by transforming Wikipedia into a network of knowledge graphs. Unverified statements can be checked against this network. A statement known to be true in Wikipedia will be present as an edge of the knowledge graph, or will have its subject and object linked via a short path in the graph. Presumably, untrue statements should not be found as connected in the graph.

Jaradat et al. (2018) have created ClaimRank, a computational system that detects claims that may need verifying (available for both Arabic and English). The claims can then be sent to fact-checking websites (which typically employ humans to do the verification), or to automatic systems. One such

system by Mohtarami et al. (2018) finds documents that may be relevant to a given claim, and snippets of evidence. While the system may not be used in a completely automatic way, it can assist human verification experts.

Another form of automatic checking involves assessing the language of the story itself, i.e., finding cues in the language of the story that point to exaggerated claims, overly emotional language or a style that is uncommon in mainstream news sources. This is, in essence, a text classification problem, one commonly addressed by computational linguists using NLP tools. Potthast et al. (2018) describe this type of classification as *style-based* fake news detection, as opposed to *context-based* (exploring the social network of the posts and the posters) or *knowledge-based* detection (fact-checking).

We argue that computational linguists are uniquely positioned to determine whether there is a ‘language of fake news.’ We discuss the potential of text classification in the next section.

Text classification for fake news

An intuitive framing of the fake news problem in NLP would be to ask how we can classify news text into fake and legitimate instances. This applies especially to the case of full text – as opposed to tweets or headlines distributed on social media – because text classification relies mainly on the linguistic characteristics of longer text. Deception detection in text has a broad literature in NLP, and fake news articles can be considered a category of deceptive text (Chen et al., 2015; Feng et al., 2012; Pérez-Rosas and Mihalcea, 2015). Methods used for text classification vary from classic machine learning algorithms using a set of pre-defined linguistic features to modern neural network models which mainly rely on pre-trained word vectors and embedded representations resulting from processing large amounts of textual data. In this section, we briefly introduce text classification methods used in the domain of deception detection and, in particular, in fake news detection.

Feature-based approaches

In NLP, the feature-based approach, which involves the extraction and analysis of linguistic cues for identification of specific target phenomena (e.g., fake product reviews from real ones) has been a very powerful model with relatively interpretable results. Features such as n-grams, subjectivity and polarity markers, lexical semantic classes, syntactic or discourse-level features have been explored in previous work on deception detection in general and on news classification in

particular (Afroz et al., 2012; Conroy et al., 2015; Horne and Adali, 2017; Pérez-Rosas and Mihalcea, 2015; Rashkin et al., 2017; Rubin et al., 2015; Ruchansky et al., 2017; Volkova et al., 2017). These features can be used with a variety of traditional supervised algorithms. Feature-based modelling usually involves feature engineering and a feature selection phase. Based on comparative experiments in different machine learning applications, it has also been shown that the performance of these classic models plateaus at some point as the training data size increases (Ng, 2011). Thus, in problems where Big Data is available, deep neural network models are being preferred, as they usually achieve impressively better results (for a recent overview of the NLP trends see Young et al., 2018).

Deep learning models

Deep learning has taken over most NLP tasks but usually in domains where large-scale training data is available. In text classification, recurrent neural networks (RNNs), convolutional neural networks (CNNs) and Attention models have been competing with feature-based models (Conneau et al., 2017; Le and Mikolov, 2014; Medvedeva et al., 2017; Yang et al., 2016; Zhang et al., 2015). RNNs are capable of encoding sequential information and are most suitable for modelling short text semantics. CNNs are composed of convolution and pooling layers, which provide an abstraction of the input. These models are employed in specific NLP tasks where the presence or absence of features is a more distinguishing factor than their location or order. For example, presence of specific words and phrases in a product review is usually indicative of it being a positive or negative review. Therefore, CNNs are well suited for the purpose of longer text classification. Neural network models have also been applied in previous work within the domain of misinformation and fake news (Rashkin et al., 2017; Wang, 2017; Yang et al., 2017).

All leading machine learning techniques for text classification, including feature-based and neural network models, are heavily data-driven. Therefore, training data is the first requirement to build these models. Quality training data for misinformation detection should consist of a balanced, sufficiently diverse and carefully labelled set of legitimate and fake news articles. While building such a dataset may sound trivial, the following section explains the challenges in gathering a dataset of this kind by referring to the datasets we have found through a review of previous work. Initial experiments suggest that existing data is still insufficient for building a robust misinformation detection system.

Data: Where and how to find fake news

The first question we need to answer in addressing fake news detection through text classification is what we consider as a representative instance of fake news. In other domains related to deceptive text, such as fake product review detection, objective criteria can be designed when labelling the fake instances: a review written by someone who has not bought or used the product, or someone recruited by the product seller for the specific duty of writing a review would be considered fake. Fake news can also be defined as news articles written by amateurs (rather than journalists) recruited with the express purpose of generating content in favour or against an entity or policy, to promote a specific idea, or for financial gain such as attracting clicks for ads. Professional journalists can also fabricate stories, for various reasons. One recent case is Claas Relotius, journalist for *Der Spiegel* in Germany, who was found to have made up stories, details and quotes from multiple sources over a long period of time.⁶ This definition considers authors and their intention as the key factor to determine whether a news article is an instance of fake. In this study, our focus is on misinformation, which entails a definition of fake news with respect to the validity of its content. So a news article that simply contains wrong information (contrary to fact) is considered as an instance of the fake class (*false*), and a news article containing verified information is a sample of real news (*true*).

The data collection strategy for building a fake news detection system depends on the definition one adopts for the task. In the majority of previous work, instances of fake news were collected from a list of suspicious websites. A relatively large collection of this type is a dataset of roughly 20,000 news articles collected by Rashkin et al. (2017). This data contains texts harvested from eight news publishers categorized into four classes: *propaganda* (The Natural News and Activist Report), *satire* (The Onion, The Borowitz Report, and Clickhole), *hoax* (American News and DC Gazette) and *trusted* (Gigaword News). This dataset is balanced across classes, and split into training, validation and test sets. However, the noisy strategy to label all articles of a publisher based on its reputation would bias a classifier trained on this data, limiting its power to distinguish individual truthful news articles from misinformation instances. In other words, data collected in this fashion would not be suitable for learning linguistic patterns of deception; it would rather help distinguish general writing style of a group of news websites (the rumour or clickbait style). We would also like to point out that newswire (what Gigaword contains) is not exactly the same as news articles. Newswire or press releases have a slightly different

audience (mostly journalists) and structure (collections of facts) than articles published by mainstream media.

In order to build a text classification system to detect false from true content based on linguistic cues, we need news articles assessed individually and labelled with respect to their level of veracity. This type of data collection is labour-intensive, as it involves fact-checking for each and every news article. A variety of fact-checking websites perform this analysis on real news. Therefore, one way to collect data on rumours and false news is to take advantage of these websites and to try to automatically scrape information such as the true vs. false headlines and hopefully their sources. Previous attempts to collect large data in this manner did not focus on the text of the news articles where the rumour was originally distributed; they rather cared more about the headlines and the annotations of the fact-checking websites.⁷

A few relatively small datasets have been collected and used in previous work that indeed contain news article texts and veracity labels assigned to them in a one-by-one fashion (see Table 1). For example, Allcott and Gentzkow (2017) collected 156 news articles by manually checking three fact-checking websites (Snopes, Politifact and BuzzFeed) and downloading the source page of the debunked rumours. The Emergent dataset (Ferreira and Vlachos, 2016) is a similar collection obtained from another fact-checker, Emergent. This collection contains 2595 news articles, but only 1238 can be considered the source of the rumours (taking a ‘for’ position towards the discussed claims).⁸ Rubin et al. (2016) took a different strategy by building a dataset of satirical news articles across nine pre-selected topics and from different publishers. This dataset has a distinguishing property: Each satirical article is matched with a legitimate article on the same topic, making the dataset very well balanced. They also checked each news article closely for a set of satirical cues to make sure the data would be representative of the register. A similar effort has shaped the Credibility Coalition project (Zhang et al., 2018), where annotators manually check the text of a news article for a set of credibility indicators. These indicators include both content-related and context-related features such as *Logical Fallacies* and *Number of Ads* on the news page, respectively. The currently published dataset, however, contains annotations for only 40 news articles. Finally, Pérez-Rosas et al. (2017) collected legitimate data on diverse topics from credible websites and matched them with fake versions by asking Mechanical Turkers to modify the content while imitating the language of journalists. This effort resulted in a dataset of 240 legitimate news articles and 240 fake news articles. In addition, they manually collected 100 celebrity-focused fake and 100 similar topic

Table 1. Currently available datasets with texts individually labelled for veracity.

Dataset	Size and type	Labelling system	Notes
Allcott and Gentzkow (2017)	156 news articles	5-Way (false to true)	Collected from Snopes, Politifact and BuzzFeed fact-checking pages, focused on 2016 US election
Ferreira and Vlachos (2016)	1612 news articles	2-Way (false/true)	Unbalanced, originally developed for stance-detection
Rubin et al. (2016)	360 news articles	2-Way (satirical/legitimate)	Balanced by topic and label. A variety of topics.
Zhang et al. (2018)	40 news articles	Multiple (credibility indicators)	Continuous effort with the future goal of annotating 10,000 articles
Pérez-Rosas et al. (2017)	480 news articles	2-Way (fake/legitimate)	Balanced by topic and label. Fake items were artificially generated by Turkers.
Pérez-Rosas et al. (2017)	200 news articles	2-Way (fake/legitimate)	Balanced by topic and label. Focused on celebrity stories.
Wang (2017)	12.8K short statements	6-Way (false to true)	Collected using the Politifact API
Thorne et al. (2018)	185K short statements and supporting/refuting Wikipedia documents	2-Way (original/mutated)	Originally developed for stance-detection. Mutated claims were artificially generated.

legitimate articles to build a balanced dataset in a specific domain from real web data.

The Liar dataset (Wang, 2017) is the first large dataset collected through reliable annotation, but it contains only short statements, not full news article texts. Another recently published large dataset is FEVER (Thorne et al., 2018), which contains both claims and texts from Wikipedia pages that support or refute them, together with veracity labels for the claims. This dataset, however, has been built to serve the slightly different purpose of stance detection (Hanselowski et al., 2018; Mohtarami et al., 2018); the false claims have been artificially generated; and the documents are not news articles, but Wikipedia pages (as true text instances) and their modified version obtained from crowd-sourcing (as fake or false instances). We provide a summary of these datasets in Table 1.

MisInfoText: A repository of assessed news texts

In order to address the lack of data with reliable labels, we have built a repository of news article texts that have been labelled by fact-checking websites. The entire dataset, plus links to datasets listed in Table 1, is available from our lab GitHub space⁹ and from our demo page.¹⁰ This repository contains three data categories:

- Links to all publicly available datasets of news that contain (1) the text of news articles, and (2) veracity labels assigned to them, are collected and maintained

in our repository. This will facilitate both theoretical and application-based studies on fake news and automatic misinformation detection.

- In addition to datasets originally published in previous work, we perform scraping on top of datasets that contain veracity labelled claims and URLs of their sources, but not necessarily the text of news articles. For example, we have found two datasets of links with veracity labels on the BuzzFeed News repository. These links become useful in finding news article instances that have already been assessed for their factual content. We make available both the original data containing links and veracity labels, as well as augmented data that we scrape from the associated news web pages including body text, title, author and date of publication. The BuzzFeed dataset will be introduced in the rest of the current section.
- Finally, we maintain and use a list of potential fact-checking websites to harvest larger amounts of data and provide a scrape-and-clean service on top of them. In collecting data directly from fact-checking websites such as Snopes, we apply a mixture of automatic and manual procedures. We have so far scraped the entire archive of the Snopes, Politifact and Emergent websites, and then automatically followed the links mentioned by each fact-checking article on these websites to the sources of discussed rumours. Figure 1 shows a screenshot of the web-scraping service that we have built and made available online for public use. Manual checking is necessary to verify that the text is valid and it in

Figure 1. Screenshot of our web service to scrape data from fact-checking websites and links to the original news articles, available at <https://fakenews.ngrok.io>.

fact supports the discussed claim. We have done the manual assessment on a subset of the automatically collected Snopes articles. This manual effort and the resulting dataset are introduced in the following.

Buzzfeed dataset

The first source of information that we used to harvest full news articles with veracity labels is from the BuzzFeed media company. BuzzFeed has published a collection of links to Facebook posts, originally compiled for a study around the 2016 US election (Silverman et al., 2016). Each URL in this dataset was given to human experts so they could rate the amount of false information contained in the linked article. The links were collected from nine Facebook pages (three right-wing, three left-wing and three mainstream publishers).¹¹ We had to follow the Facebook URLs and then the links to the original news articles to obtain the news texts. We scraped the full text of each news article from its original source. The resulting dataset includes a total of 1380 news articles on a focused topic (US election and candidates). Veracity labels come in a four-way classification scheme including 1090 *mostly true*, 170 *mixture of true and false*, 64 *mostly false* and 56 articles containing *no factual content*. Another interesting collection of URLs published by BuzzFeed News points to the top 50 fake news stories in 2017.¹² The available dataset contains only links, not the full text of the articles. Same as for the above collected data, we scraped news articles from their source of publication by following each URL, cleaned the text and augmented the original datafile by adding new columns for article title, body text, author and date of

publication. The resulting datafile contains these new pieces of information for 33 news articles that were still available online. Contrary to the US election dataset, this data contains only false news stories, with the articles covering a variety of topics.

Snopes dataset

The second source of information that we used to harvest full news articles with veracity labels is Snopes, a well-known rumour debunking website run by a team of expert editors. In addition to finding rumours and mentioning distributing websites, Snopes provides elaborate explanations of the rumour and its effects. We scraped the entire archive of fact-checking pages. On each page, Snopes discusses a claim, cites the sources (news articles, forums or social networks where the claim was distributed) and provides a veracity label for the claim. We automatically extracted all links mentioned on a Snopes page, followed the link to each original news article, and extracted the text. The resulting datafile includes roughly 4000 rows, each containing a claim discussed by Snopes annotators, the veracity label assigned to it, and the text of a news article related to the claim. The main challenge in using this data for training/testing a fake news detector is that some of the links on a Snopes page that we collect automatically do not actually point to the discussed news article, i.e., the source of the claim. Many links are to pages that provide contextual information for the fact-checking of the claim. Therefore, not all the texts in our automatically extracted dataset are reliable or simply the ‘supporting’ source of the claim. To come up with a reliable set of veracity-labelled news articles, we randomly selected 312 items and assessed them manually. Two annotators performed independent assessments on the 312 items.

A third annotator went through the entire list of items for a final check and to resolve disagreements. Snopes has a fine-grained veracity labelling system. We selected *[fully] true*, *mostly true*, *mixture of true and false*, *mostly false*, and *[fully] false* stories. Among the 312 assessed items, 145 came out as the supporting source of the claim, thus reliable news text articles with veracity labels suitable for training/testing of automatic misinformation detection. The next section will provide more details on the content of news in this collection.

Topics covered by fact-checkers

In this section, we perform a topic modelling experiment to explore the data we have collected from fact-checking websites and to get a sense of what type of news articles are covered in the available datasets. The issue of topics is important because training datasets that are skewed in terms of topic will result in classifiers that are unable to generalize to different topic distributions. More generally, research to date has not explored what topics are more or less likely to be featured in fake news stories, although it seems that news about politics, the environment and health are prevalent. Vargo et al. (2018) investigate the media landscape, and the interaction between mainstream media, fake news publishers and fact checkers, testing the hypothesis that fake news media and fact checkers have the power to set the agenda of news media, by the types of stories that they respectively cover or fact-check. While their study found that fake news does not set the agenda for mainstream media, it is intricately connected to partisan news, taking cues from partisan sites with regard to what types of topics and stories are covered in fake news. Even more worrisome is the connection between fake news and emerging media, perhaps, Vargo et al. hypothesize, because emerging media is, like fake news sites, predominantly online. These are all issues that need further exploring, and will affect how fake news datasets ought to be built in terms of topic distributions.

Since the Emergent dataset is the largest and most similar dataset to ours (because it was also collected from a fact-checking website), we also include this data in our experiment. The objectives of this experiment are two-fold:

- Discover what topics in news are covered by fact-checking websites, and how the distribution varies between true vs. false news stories.
- Find the gaps and sources of imbalance in currently available data to provide useful directions for future data collection efforts.

In order to address these points, we need a sufficiently large reference corpus of news text – as training

data to the topic model – that is representative of news stories regardless of their content being misinformation or not. For this purpose, we employ a collection of 16,000 texts from the training portion of Rashkin et al.'s (2017) data, which we briefly introduced in the previous section. Projecting our labelled news articles into the topic space constructed based on this diverse data will then reveal the topic distribution in fake news as well as preferences of the fact-checking websites in picking and debunking rumours.

To build the topic model, we preprocessed documents in Rashkin's training set (by tokenizing, normalizing and removing punctuations and stopwords) and fed the resulting word-document vectors into an Latent Dirichlet Allocation model in the Gensim python library. We tuned the number of topics so each topic represents a clear category of news that is not too fine or coarse-grained for visual investigation. The final number of topics that gave clearest results was 10. Figure 2 (bottom section) shows the word clouds we obtained from the 10 most important words in each topic, with their weight represented by the font size. In a similar fashion, we preprocessed documents from our two datasets extracted from Buzzfeed and Snopes, as well as the Emergent dataset. We then projected each subset of these datasets (split based on veracity labels) into the pre-trained topic space. Doing so provides us with some interesting observation regarding the distribution of important news topics in the labelled collections (see the top section of Figure 2).

The Buzzfeed dataset (1380 articles), which is mostly focused on news related to the US election in 2016, comes out as the least diverse dataset. This was to be expected, as this dataset covers the topics of election, personal stories (of the presidential candidates) and other political topics such as stories related to police and the legislation system. The Snopes dataset (145 articles) is relatively more diverse: In addition to political topics, it includes some news on sports, environment and health. Notice that the Buzzfeed top fake news collection (33 articles) has a more similar distribution to that of the Snopes collection, and this is because Buzzfeed in fact collected that dataset by looking at Snopes and Politifact websites. Finally, the Emergent dataset (1612 articles) stands out as the most varied collection. This dataset is also relatively larger, which might indirectly contribute to topic diversity. While the three datasets put together cover a variety of news stories, it does seem that stories on certain topics such as the market (economy) and technology are less represented in this collection.

By looking closely at each row of the heat map, we also find that some topics are more frequent in false than true news. For example, in the Snopes dataset, the topic of police is found more in false news articles.

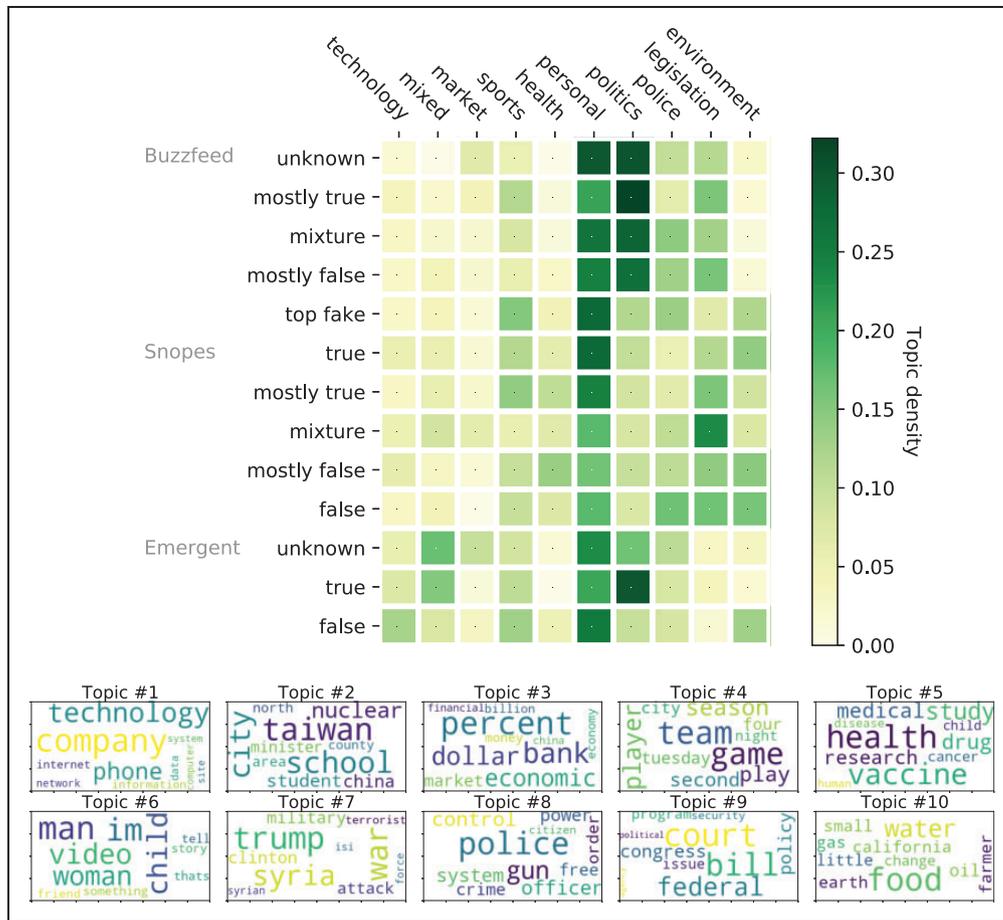


Figure 2. Topic distribution across news text corpora obtained from fact-checking websites: BuzzFeed, Snopes and Emergent.

In the Emergent dataset, the technology and environment topics are more frequent in false news, whereas the opposite pattern is observed for the politics topic. These differences can be indicative of an inherent difference between misinformation and real news, or they might just mean that the studied fact-checking websites are biased towards certain types of stories. Personal stories, in particular, appear frequently across all datasets and all veracity labels. This type of pattern is particularly interesting, as it can indeed be a consistent feature of the rumour type of news, but not necessarily a sign of misinformation.

We used the datasets introduced here for text classification experiments which we do not include in this paper, but which are interesting to briefly mention. Using unbalanced data in terms of topics leads to high accuracy classification, even using very simple features (such as tokens or word n-grams) when train and test data are sampled from a similar distribution of news topics. However, reporting such high accuracies is misleading because what we are looking for is in fact a fake news detection system that can generalize to new topics, i.e., a classifier that detects high-level features

that can be considered as signs of deception, regardless of a news article's specific topic. Small data collections would not offer cross-topic generalization because what the models learn in this situation is the vocabulary differences between fake and real news and the vocabulary depends strongly on the topics. For example, if we train a classifier on the data depicted in Figure 2, any test instance that comes from the technology topic would likely be classified as 'false' (because we have few training instances of this topic in the collection and most of them are from the false class in the Emergent data). Therefore, it is important to collect both fake and real news instances on a variety of topics to make sure that what our systems learn about deception can be generalized to unseen instances of news across topics.¹³

A call to arms

Our efforts at collecting data to build a robust fake news classifier have taught us a valuable lesson: Reliably labelled fake news articles are actually hard to come by. Although many fake news publishers

exist, we have no assurance that every story on those sites constitutes misinformation. Thus, we need instances of individually labelled stories, labelled by humans with some expertise on the topic of the stories, or at least with some general training in journalism.

We have modestly contributed to this effort, with two datasets from BuzzFeed and Snopes, amounting to a total of 1558 individually labelled articles with veracity scores in a five-way spectrum. While this dataset has allowed us to investigate certain aspects of fake news, such as the types of topics covered, it is certainly not sufficient for modern text classification methods, especially for deep learning models. We need Big Data to solve this problem.

Our call to arms encourages researchers in this field to share datasets, and to work towards a standard for labelling and organizing the data. This is not about who gets a paper published first; it is about addressing an important problem, and finding solutions by working together. Lazer et al. (2017) call for developing datasets that are useful for studying the spread of misinformation, and suggest pressuring social media companies to share important data. We join this chorus, and would like to have access to datasets to study not only spread, but also the fake news articles themselves.

Conclusions

We have discussed the different approaches to the problem of fake news and misinformation, some of them relating to how to educate the public or to how to stop the spread of such pernicious news. We focus on tackling the problem as a text classification problem, i.e., attempting to automatically detect whether a particular news article is fake or not. By ‘fake’ we mean an article that contains unverified or untrue claims, or attempts to disseminate information that is not accurate.

In order to perform automatic classification of news texts, modern NLP and machine learning methods require large amounts of training data. As computational linguistics researchers, we feel, however, that we cannot decide by ourselves which articles are instances of fake or real news. This is why we propose relying on datasets containing articles that have been individually labelled for veracity by experts. We have found, unfortunately, that there are very few such datasets, because individual labelling is a time-consuming task. Nevertheless, one source of such labels are fact-checking websites, which perform this task for the public good. We have scraped, cleaned up and organized individual articles harvested from these sites, together with their labels (true, false, or similar labels). We introduce this dataset, MisInfoText, as a resource for text classification efforts. We also carried

out analyses based on topics, and discovered that the datasets are unbalanced with respect to topics, an issue that needs to be addressed for text classification.

More work in this regard is certainly needed, and we encourage the community to organize and contribute their own datasets, so that we can address this problem in a collaborative fashion.

Our future work involves using this dataset and any other that we can find to build robust classifiers. We are experimenting with both ‘classic’ feature-based approaches and deep learning methods.

Acknowledgements

We thank members of the Discourse Processing Lab at Simon Fraser University, especially Yajie Zhou and Jerry Sun, for their help checking individual stories in the datasets and building the website.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by a Discovery Grant from the Natural Sciences and Engineering Research Council of Canada, and by NVIDIA Corporation, with the donation of a Titan Xp GPU.

Notes

1. Virginia Woolf, *Three Guineas*, originally published in 1938, reprinted in Lapham’s Quarterly, *A History of Fake News* (2018).
2. Cited in Lazer et al. (2017).
3. <https://www.ifla.org/publications/node/11174>
4. <https://shorensteincenter.org/free-course-identifying-misinformation/>
5. <https://credibilitycoalition.org/>
6. <http://www.spiegel.de/international/zeitgeist/claas-reliotius-reporter-forgery-scandal-a-1244755.html>
7. Examples of such datasets include a collection shared on Kaggle at <https://www.kaggle.com/arminehn/rumor-citation> and a dataset mentioned, but not published, in the supplementary material of Vosoughi et al. (2018) at http://science.sciencemag.org/content/sci/suppl/2018/03/07/359.6380.1146.DC1/aap9559_Vosoughi_SM.pdf
8. These numbers are extracted from the Emergent final cleaned train and test data files, which did not include veracity labels (because this dataset was originally published for a stance detection task): <https://github.com/willferreira/mscproject>. However, we explored their raw data file, which contained veracity labels, and we discovered 1612 unique articles categorized as taking the ‘for’ stance towards the claims. This subset of the data included 359 true, 742 false and 511 unknown items. We

use this dataset along with our collected datasets in the topic modelling experiments presented later on in this paper.

9. <https://github.com/sfu-discourse-lab/MisInfoText>
10. <http://fakenews.ngrok.io/>
11. <https://github.com/BuzzFeedNews/2016-10-facebook-fact-check/blob/master/data/facebook-fact-check.csv>
12. https://github.com/BuzzFeedNews/2017-12-fake-news-top-50/blob/master/data/fact_check.csv
13. Besides topic variety in data, considering more sophisticated features such as semantic categories (e.g., emotion words), phrase and sentence-level features may help the classifiers learn about cross-domain characteristics of deceptive and truthful text. This falls within the scope of a future study with an emphasis on suitable text classification methods and feature analysis for general deception detection in text. Our data collection can provide a good sample of texts in the domain of news for such studies.

ORCID iD

Maite Taboada  <http://orcid.org/0000-0002-6750-8891>

References

- Afroz S, Brennan M and Greenstadt R (2012) Detecting hoaxes, frauds, and deception in writing style online. In: *IEEE symposium on security and privacy (SP) 2012*, San Francisco, CA, IEEE, pp. 461–475.
- Allcott H and Gentzkow M (2017) Social media and fake news in the 2016 election. *Journal of Economic Perspectives* 31: 211–236.
- Ananny M (2018) *The partnership press: Lessons for platform-publisher collaborations as Facebook and news outlets team to fight misinformation*. Technical report, Tow Center for Digital Journalism, Columbia University, New York.
- Bechmann A and Nielbo KL (2018) Are we exposed to the same “news” in the News Feed? An empirical analysis of filter bubbles as information similarity for Danish Facebook users. *Digital Journalism* 6(8): 990–1002.
- Berinsky AJ (2017) Rumors and health care reform: Experiments in political misinformation. *British Journal of Political Science* 47(2): 241–262.
- Chen A (2017) The human toll of protecting the Internet from the worst of humanity. *New Yorker*, 28 January.
- Chen Y, Conroy NJ and Rubin VL (2015) Misleading online content: Recognizing clickbait as false news. In: *Proceedings of the 2015 ACM on workshop on multimodal deception detection*, ACM, pp. 15–19.
- Ciampaglia GL, Shiralkar P, Rocha LM, et al. (2015) Computational fact checking from knowledge networks. *PLoS One* 10(6): e0128193.
- Conneau A, Schwenk H, Barrault L, et al. (2017) Very deep convolutional networks for text classification. In: *Proceedings of the 15th conference of the European chapter of the association for computational linguistics*, Valencia, Spain, pp. 1107–1116.
- Conroy NJ, Rubin VL and Chen Y (2015) Automatic deception detection: Methods for finding fake news. *Proceedings of the Association for Information Science and Technology* 52(1): 1–4.
- Constine J (2018) Facebook shrinks fake news after warnings backfire. *Tech Crunch*, 28 April. Available at: <https://tcrn.ch/2jb7gcp> (accessed April 24, 2019)
- Darnton R (2017) The true history of fake news. *The New York Review of Books*, 13 February.
- Del Vicario M, Bessi A, Zollo F, et al. (2016) The spreading of misinformation online. *Proceedings of the National Academy of Sciences* 113(3): 554–559.
- Ecker UK, Hogan JL and Lewandowsky S (2017) Reminders and repetition of misinformation: Helping or hindering its retraction? *Journal of Applied Research in Memory and Cognition* 6(2): 185–192.
- Feng S, Banerjee R and Choi Y (2012) Syntactic stylometry for deception detection. In: *Proceedings of the 50th annual meeting of the Association for Computational Linguistics*, pp. 171–175.
- Ferreira W and Vlachos A (2016) Emergent: a novel data-set for stance classification. In: *Proceedings of the 2016 conference of the North American chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 1163–1168.
- Gloger I, Otto L and Boukes M (2016)) The softening of journalistic political communication: A comprehensive framework model of sensationalism, soft news, infotainment, and tabloidization. *Communication Theory* 27(2): 136–155.
- Greenhill KM and Oppenheim B (2017) Rumor has it: The adoption of unverified information in conflict zones. *International Studies Quarterly* 61(3): 660–676.
- Hanselowski A, Avinesh PVS, Schiller B, et al. (2018) A retrospective analysis of the fake news challenge stance detection task. *arXiv preprint arXiv:1806.05180*.
- Horne BD and Adali S (2017) This just in: Fake news packs a lot in title, uses simpler, repetitive content in text body, more similar to satire than real news. *arXiv preprint arXiv:1703.09398*.
- Jack C (2017) *Lexicon of lies: Terms for problematic information*. Technical report, Data & Society Research Institute, New York, NY.
- Jaradat I, Gencheva P, Barrón-Cedeño A, et al. (2018) ClaimRank: Detecting check-worthy claims in Arabic and English. In: *Proceedings of the 2018 conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations*, New Orleans, LA, pp. 26–30.
- Lazer D, Baum M, Grinberg N, et al. (2017) *Combating fake news: An agenda for research and action*. Harvard Kennedy School, Shorenstein Center on Media, Politics and Public Policy, 2 May.
- Lazer D, Baum MA, Benkler Y, et al. (2018) The science of fake news. *Science* 359(6380): 1094–1096.
- Le Q and Mikolov T (2014) Distributed representations of sentences and documents. In: *Proceedings of the 31st international conference on machine learning*, Beijing, China, pp. II-1188–II-1196.
- Lewandowsky S, Ecker UK, Seifert CM, et al. (2012) Misinformation and its correction: Continued influence

- and successful debiasing. *Psychological Science in the Public Interest* 13(3): 106–131.
- Marlin R (2002) *Propaganda and the Ethics of Persuasion*. Toronto: Broadview Press.
- Marwick A and Lewis R (2017) *Media manipulation and disinformation online*. Technical report, Data & Society Research Institute, New York, USA.
- Medvedeva M, Kroon M and Plank B (2017) When sparse traditional models outperform dense neural networks: The curious case of discriminating between similar languages. In: *Proceedings of the 4th workshop on NLP for Similar Languages, Varieties and Dialects (VarDial)*, Valencia, Spain, pp. 156–163.
- Mohtarami M, Baly R, Glass J, et al. (2018) Automatic stance detection using end-to-end memory networks. In: *Proceedings of the 2018 conference of the North American chapter of the Association for Computational Linguistics: Human Language Technologies, volume 1 (long papers)*, New Orleans, LA, pp. 767–776.
- Mustafaraj E and Metaxas PT (2017) The fake news spreading plague: Was it preventable? In: *Proceedings of the 2017 ACM on web science conference*, ACM, pp. 235–239.
- Ng A (2011) Why is deep learning taking off?. Available at: <https://www.coursera.org/lecture/neural-networks-deep-learning/why-is-deep-learning-taking-off-praGm>
- Nielsen RK and Graves L (2017) ‘News you don’t believe’: Audience perspectives on fake news. *Reuters Institute for the Study of Journalism Report*. Available at: https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2017-10/Nielsen%26Graves_factsheet_1710v3_FINAL_download.pdf
- Pavleska T, Školokay A, Zankova B, et al. (2018) Performance analysis of fact-checking organizations and initiatives in Europe: A critical overview of online platforms fighting fake news. In: *EIDS6 (European Integration and Democracy Series)*, pp. 1–29.
- Pennycook G, Cannon T and Rand DG (2018) Prior exposure increases perceived accuracy of fake news. *Journal of Experimental Psychology: General* 147(12): 1865–1880.
- Pérez-Rosas V, Kleinberg B, Lefevre A, et al. (2017) Automatic detection of fake news. *arXiv preprint arXiv:1708.07104*.
- Pérez-Rosas V and Mihalcea R (2015) Experiments in open domain deception detection. In: *Proceedings of the conference on empirical methods in natural language processing*, pp. 1120–1125.
- Potthast M, Kiesel J, Reinartz K, et al. (2018) A stylometric inquiry into hyperpartisan and fake news. In: *Proceedings of the 56th annual meeting of the Association for Computational Linguistics (volume 1: long papers)*, Melbourne, Australia, pp. 231–240.
- Rashkin H, Choi E, Jang JY, et al. (2017) Truth of varying shades: Analyzing language in fake news and political fact-checking. In: *Proceedings of the 2017 conference on empirical methods in natural language processing*, Copenhagen, Denmark, pp. 2921–2927.
- Rozin P and Royzman EB (2001) Negativity bias, negativity dominance, and contagion. *Personality and Social Psychology Review* 5(4): 296–320.
- Rubin VL, Chen Y and Conroy NJ (2015) Deception detection for news: Three types of fakes. *Proceedings of the Association for Information Science and Technology* 52(1): 1–4.
- Rubin VL, Conroy NJ, Chen Y, et al. (2016) Fake news or truth? Using satirical cues to detect potentially misleading news. In: *Proceedings of NAACL-HLT*, San Diego, CA, pp. 7–17.
- Ruchansky N, Seo S and Liu Y (2017) CSI: A hybrid deep model for fake news detection. In: *Proceedings of the 2017 ACM on conference on information and knowledge management*, Singapore, pp. 797–806.
- Sachsman DB (2017) *Sensationalism: Murder, Mayhem, Mudslinging, Scandals, and Disasters in 19th-Century Reporting*. New York, NY: Routledge.
- Shao C, Hui PM, Wang L, et al. (2018) Anatomy of an online misinformation network. *PLoS ONE* 13(4): e0196087.
- Shu K, Sliva A, Wang S, et al. (2017) Fake news detection on social media: A data mining perspective. *ACM SIGKDD Explorations Newsletter* 19(1): 22–36.
- Silverman C and Alexander L (2016) How teens in the Balkans are duping Trump supporters with fake news. *Buzzfeed News*. Available at: <https://www.buzzfeed.com/craigsilverman/how-macedonia-became-a-global-hub-for-pro-trump-misinfo>
- Silverman C, Strapagiel L, Shaban H, et al. (2016) Hyperpartisan Facebook pages are publishing false and misleading information at an alarming rate. *BuzzFeed News*. Available at: <https://www.buzzfeed.com/craigsilverman/partisan-fb-pages-analysis>
- Subramanian S (2017) Inside the Macedonian fake-news complex. *Wired Magazine*, 15 February.
- Swire B, Ecker UK and Lewandowsky S (2017) The role of familiarity in correcting inaccurate information. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 43(12): 1948.
- Tambuscio M, Ruffo G, Flammini A, et al. (2015) Fact-checking effect on viral hoaxes: A model of misinformation spread in social networks. In: *Proceedings of the 24th international conference on world wide web*, ACM, pp. 977–982.
- Thorne J, Vlachos A, Christodoulopoulos C, et al. (2018) FEVER: A large-scale dataset for Fact Extraction and VERification. In: *Proceedings of the 2018 conference of the North American chapter of the Association for Computational Linguistics: Human Language Technologies, volume 1 (long papers)*, New Orleans, LA, pp. 809–819.
- Vargo CJ, Guo L and Amazeen MA (2018) The agenda-setting power of fake news: A big data analysis of the online media landscape from 2014 to 2016. *New Media & Society* 20(5): 2028–2049.
- Volkova S, Shaffer K, Jang JY, et al. (2017) Separating facts from fiction: Linguistic models to classify suspicious and trusted news posts on Twitter. In: *Proceedings of the 55th annual meeting of the Association for Computational Linguistics*, Vol. 2, Vancouver, Canada, pp. 647–653.
- Vosoughi S, Roy D and Aral S (2018) The spread of true and false news online. *Science* 359(6380): 1146–1151.

- Wang WY (2017) ‘Liar, liar pants on fire’: A new benchmark dataset for fake news detection. In: *Proceedings of the 55th annual meeting of the Association for Computational Linguistics*, Vol. 2, Vancouver, Canada, pp. 422–426.
- Wardle C (2016) Six types of misinformation circulated this election season. *Columbia Journalism Review*, 18 November.
- Yang F, Mukherjee A and Dragut E (2017) Satirical news detection and analysis using attention mechanism and linguistic features. In: *Proceedings of the 2017 conference on empirical methods in natural language processing*, pp. 1979–1989.
- Yang Z, Yang D, Dyer C, et al. (2016) Hierarchical attention networks for document classification. In: *Proceedings of the 2016 conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 1480–1489.
- Young T, Hazarika D, Poria S, et al. (2018) Recent trends in deep learning based natural language processing. *IEEE Computational Intelligence Magazine* 13(3): 55–75.
- Zhang AX, Ranganathan A, Metz SE, et al. (2018) A structured response to misinformation: Defining and annotating credibility indicators in news articles. In: *Proceedings of the web conference 2018*, pp. 603–612. Lyon, France: International World Wide Web Conferences Steering Committee.
- Zhang X, Zhao J and LeCun Y (2015) Character-level convolutional networks for text classification. In: *Proceedings of the 28th international conference on neural information processing systems*, Montréal, Canada, pp. 649–657.
- Zhao Z, Resnick P and Mei Q (2015) Enquiring minds: Early detection of rumors in social media from enquiry posts. In: *Proceedings of the 24th international conference on world wide web*, pp. 1395–1405.