Information Processing and Management 57 (2020) 102025

Contents lists available at ScienceDirect

Information Processing and Management

journal homepage: www.elsevier.com/locate/infoproman

An overview of online fake news: Characterization, detection, and discussion

Xichen Zhang*, Ali A. Ghorbani

Canadian Institute for Cybersecurity (CIC), Faculty of Computer Science, University of New Brunswick (UNB), Fredericton, NB E3B 5A3, Canada

ARTICLE INFO

Keywords:
Social media
Online fake news
Fake news detection

ABSTRACT

Over the recent years, the growth of online social media has greatly facilitated the way people communicate with each other. Users of online social media share information, connect with other people and stay informed about trending events. However, much recent information appearing on social media is dubious and, in some cases, intended to mislead. Such content is often called fake news. Large amounts of online fake news has the potential to cause serious problems in society. Many point to the 2016 U.S. presidential election campaign as having been influenced by fake news. Subsequent to this election, the term has entered the mainstream vernacular. Moreover it has drawn the attention of industry and academia, seeking to understand its origins, distribution and effects.

Of critical interest is the ability to detect when online content is untrue and intended to mislead. This is technically challenging for several reasons. Using social media tools, content is easily generated and quickly spread, leading to a large volume of content to analyse. Online information is very diverse, covering a large number of subjects, which contributes complexity to this task. The truth and intent of any statement often cannot be assessed by computers alone, so efforts must depend on collaboration between humans and technology. For instance, some content that is deemed by experts of being false and intended to mislead are available. While these sources are in limited supply, they can form a basis for such a shared effort.

In this survey, we present a comprehensive overview of the finding to date relating to fake news. We characterize the negative impact of online fake news, and the state-of-the-art in detection methods. Many of these rely on identifying features of the users, content, and context that indicate misinformation. We also study existing datasets that have been used for classifying fake news. Finally, we propose promising research directions for online fake news analysis.

1. Introduction

1.1. Background

The exploding development of World Wide Web after the mid-1990s has significantly advanced the way that people communicate with each other. Online social media, like *Twitter* and *Facebook*, can facilitate the distribution of real-time information among users from all over the world. With the characteristics of ease-of-use, low cost, and rapid rate, social media has become the major platform for online social interaction and information transmission (Shu, Sliva, Wang, Tang, & Liu, 2017). Nowadays, nearly two-thirds of

* Corresponding author.
E-mail addresses: xichen.zhang@unb.ca (X. Zhang), ghorbani@unb.ca (A.A. Ghorbani).

https://doi.org/10.1016/j.ipm.2019.03.004
Received 10 August 2018; Received in revised form 12 February 2019; Accepted 8 March 2019
Available online 20 March 2019

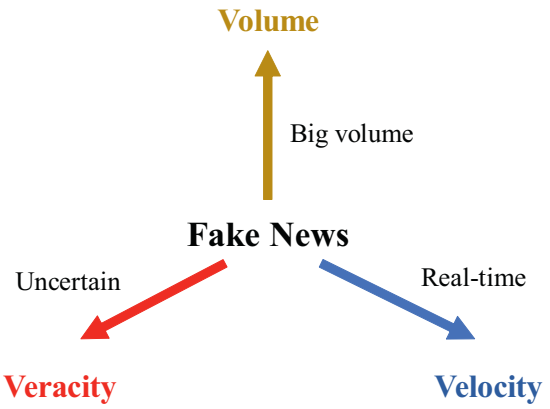

Fig. 1. The volume, velocity and veracity of fake news.

American adults get access to news via online channels (News use across social media platforms 2016), and this number is still growing exponentially (Dale, 2017; News use across social media platforms 2016).

However, owing to the increasing popularity of online social media, the Internet becomes an ideal breeding ground for spreading fake news, such as misleading information, fake reviews, fake advertisements, rumors, fake political statements, satires, and so on. Now fake news is more popular and widely spread through social media than mainstream media (Balmas, 2014). Being extensively used for confusing and persuading online users with biased facts, fake news has become the major concern for both industry and academia. Furthermore, a massive amount of incredible and misleading information is created and displayed through the Internet, which has arisen as a potential threat to online social communities, and had a deep negative impact on the Internet activities, such as online shopping, and social networking.(Fig. 1)

The issue of online fake news has gained more attention by both researchers and practitioners, especially after 2016 U.S. presidential election (Horne & Adali, 2017). Fake news has been accused of increasing political polarization and partisan conflict during the election campaign (Riedel, Augenstein, Spithourakis, & Riedel, 2017), and the voters can also be easily influenced by the misleading political statements and claims. Many latest online fact-checking systems, such as FactCheck.org and PolitiFact.com are based on manual detection approaches by professionals, where time latency is the main issue. Also, most of the existing online fact-checking resources are mainly focusing on the verification of political news, so the practical applicability of those systems is limited, due to the high variety of news types and formats, and the widely and quickly propagation of fake information in the social network. In addition, a large amount of real-time information is created, commented, and shared via online social media everyday, which makes online real-time fake news detection even more difficult.

In recent years, to help online users identify useful and valuable information, there has been extensive research on establishing an effective and automatic framework for online fake news detection. Identifying credible social information from millions of messages, however, is challenging, due to the heterogeneous and dynamic nature of online social communication. More specifically, it is difficult to distinguish online truthful signals from the fake and anomalous information, since the fake news is intentionally written to mislead readers (Shu et al., 2017). Meanwhile, the linguistic-based features extracted from the news content are not sufficient for revealing the in-depth underlying distribution patterns of fake news (Shu et al., 2017; Zhao et al., 2014). Auxiliary features such as the credibility of the news author and the spreading patterns of the news, play more important roles for online fake news prediction. Furthermore, online social data is time-sensitive, which means that they occur in a real-time pattern, and represent the trending topics and events. As a result, an online real-time detection system should be designed for detecting, exploring and interpreting fake information in online social media.

1.2. Significances of fake news detection

Over the recently years, the fast and explosive development of social media have witnessed the extensive growth in the number of fake news. Nowadays, fake news is annoying, obtrusive, distracting and all over the places. It has profound impacts on both individuals and the society. So it is significant for building an effective detection system for fake news identification. The basic characteristics of fake news can be summarized as follows:

- **The volume of fake news**: Without any verification procedure, everyone can easily write fake news on the Internet (Ahmed, 2017). There are lots of webpages which are established purposely to publish fake news and stories, such as denverguardian.com, wtoe5news.com, ABCnews.com.co, and so on. Those websites often resemble legitimate news organizations (Allcott & Gentzkow, 2017), and are deliberately created to distribute hoaxes, propaganda, and disinformation, often for financial or political gain. Therefore, a massive amount of fake contents are distributed through the Internet, even without users' awareness.
- **The variety of fake news**: There are several close definitions of fake news, such as rumors, satire news, fake reviews, misinformation, fake advertisements, conspiracy theories, false statement by politicians etc., which affect every aspect of people'

lives. With the increasing popularity of social media, fake news can dominate public's opinions, interests and decisions. In addition, fake news change the way that people interact with real news. Some fake news are created intentionally to mislead and confuse social media users, especially young students and old people who are empty of self-protection consciousness (Forbes.com). For example, some rumors were propagated on Twitter immediately after the 2010 earthquake in Chile, which increased the public panic and chaos in the local population (Castillo, Mendoza, & Poblete, 2011). More recently, a story shared on Facebook used selective TV ratings data to make the misleading claim that Cable News Network (CNN) was not one of the 10 most watched cable networks in 2018 (Fichera). Another fake science news reported that Physicist Stephen Hawking warned "aliens existed on the far side of the Moon" (Daily). We can see that online fake news is profound and far-reaching into every aspect of our daily life.

- **The velocity of fake news**: Fake news creators tend to be short-lived (Allcott & Gentzkow, 2017). For example, many active fake news webpages during 2016 U.S. election no longer exist after the campaign. As more attention is paid to fake news in recent years, more fake news generators are nothing but a transient flash in order to avoid detection by the detection systems. Furthermore, most of the fake news on social media are focusing on the current events and hot affairs to bring more attention to the online users. The real-time nature of fake news on social media makes identifying online fake news even more difficult. It is complicated to evaluate how many online users are involved with a certain piece of instant message, and it is hard to tell when and how the far-reaching consequences of fake news stop.

By conveying biased and false information, fake news can destroy folk's faith and beliefs in authorities, experts and the government. For instance, 88% of customers rely on online reviews, and 72% of them firmly believe a business with positive reviews (Ahmed, 2017). Another example is 2016 U.S. presidential election. During this campaign, hundreds or thousands of Russian fake accounts posted anti-Clinton messages, such as "*Hillary was sick*", "*Hillary was a criminal*", "*Obama had a secret army*", and so on, to influence soft Hillary Clinton supporters (huffingtonpost.com; nytimes.com). The voters can easily be affected by the false information, and even work as fake news spreaders by sharing the fake content and commenting on the fake news. There is a view that Donald Trump's victory in 2016 U.S. president election is somehow regarded as the outcome of fake news (straitstimes.com; Allcott & Gentzkow, 2017). The fake news continues to dominate the Internet these days, which brings fateful consequences to the society, to the politics, to IT and financial matters, and to everyone who may live in a cyber environment with the crisis of trust. There is an immediate necessity for generating a well-established, accurate-oriented real-time system for online fake news detection and identification.

1.3. Motivations and contributions of this paper

As fake news detection has become an emerging topic, more and more technical giant companies are seeking future solutions for recognizing online fake information. With the help from fact-checking professionals, *Facebook* allows users to flag and report satires or news that are potentially suspicious and anomalous (News feed fyi; Mark zuckerberg). Most recently, a new online service called "*Google News Initiative*" is announced by *Google*, in order to fight fake news, misinformation, and contentious breaking stories (Google news initiative). This project will spend *$392* millions over the next several years, which could make it easier for readers to subscribe to quality publication. Also, it can help readers on how to spot misleading news and reports (Google announcement).

However, accurate fake news detection, is still challenging, due to the dynamic nature of the social media, and the complexity and diversity of online communication data. In addition, the limited availability of high-quality training data is a big issue for training supervised learning models. It is necessary to design a framework which is able to identify anomalous or suspicious online information even without the knowledge of anomalous samples (Zhao et al., 2014). Under the circumstances, both industry and academia are actively involved in the trend of combating online fake news. It is significant to design effective, automatic and applicable approaches for online fake news detection.

The motivations of this paper can be summarized as follows. (1) The analysis of fake news content is not sufficient to establish an effective and reliable detection system. So other important relevent aspects, such as author and user analysis, news social context are also described in this paper, in order to generate an overall understanding of online social information. (2) The studies on online fake news detection are diverse in terms of objectives, methodologies and domains. It is a necessity to summarize different types of techniques and methods in this area, compare representative hand-crafted features, and evaluate the existing detection systems. By presenting a comprehensive view of online fake news detection, our survey can provide practical conveniences for both researchers and participators. (3) Potential promising data mining algorithms and methods are introduced in this paper, which are valuable for addressing the aforementioned challenging and improving the existing detection frameworks.

Recently, some survery papers also cover the topic of online fake news and false information detection. In Shu et al. (2017), the authors discuss online fake news detection on social media. Especially, they focus on characterizing fake news in the perspectives of psychology and social theories. Also, the existing data mining algorithms and the evaluation metrics are demonstrated in their paper. In Kumar and Shah (2018), the authors present a comprehensive study on the distribution of online false information. They mainly discuss how false information proliferates on the Internet and why it succeeds in deceiving online readers. Also, they quantify the impact of false information and summarize some useful algorithms for detecting false information. Different from their studies, in our work, the fake news is characterized by four major aspects: news creator, news content, social context, and the targets. We believe that in this way the readers can have a better understanding of the nature of online fake news, like who is the news sources, what is their purpose for creating online false information, what writing skills are more likely to be used in fake news, how fake news is distributed via the Internet, and how it can effect online readers. The major contributions of this paper can be summarized as follows.

(1) We summarize both practical-based approaches and research-based approaches for online fake news detection. And for the readers, no matter they are researchers, industrial participates, or random Internet users, can find helpful and useful knowledge from our work. (2) We propose an up-to-date and comprehensive set of features which can be used for online fake news identification. This feature set contains three different subcategories: creator and user-based features, news content-based features, and social context-based features. With our proposed features, researchers can not only conduct a task of online fake news detection, but also work on other similar domains, like botnet detection, malicious or fake account detection, unknown news creator detection, sentiment analysis, stance detection, news similarity analysis, and so on. This is practically significant for researchers whose research interests are data mining in social media, natural language processing, and false information detection. (3) At the end of this survey, some potential technologies like unsupervised learning algorithms, one-class classification algorithms, and real-time detection are proposed as future research directions. Also a comprehensive fake news detection ecosystem is designed with three layers (alert layer, detection layer, and intervention layer). We provide a well-structured work on the topic of fake news detection, from characterization, detection to final discussion.

The rest of this paper is organized as follows. Section 2 demonstrates the definition and other important aspects of online fake news, such as the author and target users of fake news, the news content body and the social context of online fake news. Section 3 summarizes practical-based approaches for fake news detection, includes online fact-checking resources and some useful social guides. Section 4 presents the latest research based studies for online fake news detection and analysis, lists the influential features for fake news representation, and evaluates the available fake news datasets. Section 5 discusses the open issues and some promising research directions in online fake news analysis. And finally Section 6 recaps the conclusions and the contributions of this paper.

2. Fake news characterization

Nowadays online fake news tend to be intrusive and diverse in terms of topics, styles and platforms (Shu et al., 2017). And it is not easy to construct a generally accepted definition for "fake news". *Stanford University* provides the definition of fake news as: "*the news articles that are intentionally and verifiably false, and could mislead readers* (Detecting fake news with nlp)". According to *Wikipedia* (Fake news), fake news is: "*a type of yellow journalism or propaganda that consists of deliberate misinformation or hoaxes spread via traditional print and broadcast news media or online social media.*"

As the tremendous development of the Internet, more and more fake news are distributed via social networks and word-of-mouth. In this paper, we proposed our definition as: "*fake news refers to all kinds of false stories or news that are mainly published and distributed on the Internet, in order to purposely mislead, befool or lure readers for financial, political or other gains.*" By emphasizing the various types of online fake information and without loosing of generality, fake reviews, fake advertisements are also discussed and covered in our study.

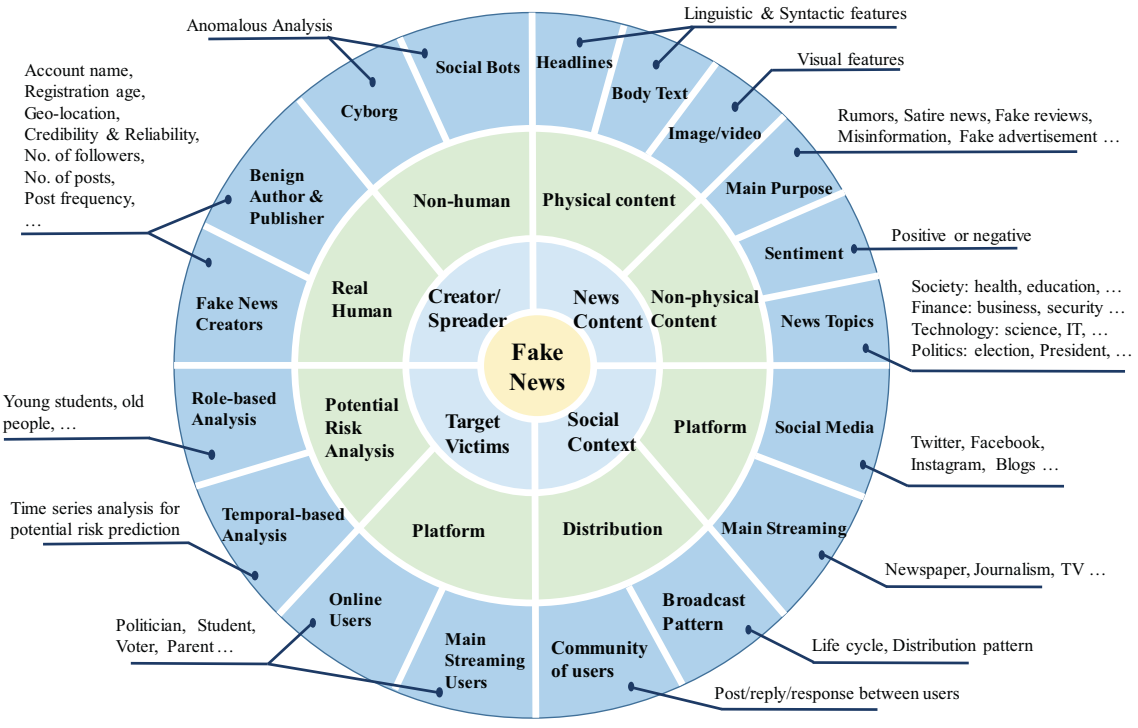

Fig. 2. Fake news and everything related to it.

Fig. 3. An example of fake news shared by a Facebook user.

To clearly understand the scope and variety of online fake information, some important aspects for defining fake news are shown in Fig. 2. In Fig. 2, the term "Fake News" is in the core of the onion-shaped graph, and it contains four major component: Creator/Spreader, Target Victims, News Content, and Social Context. All the components are in the first inner layer around "Fake News".

- Creator/Spreader: The creators of online fake news can be either real human or non-human. Real human fake news creators include both benign authors and users who publish fake news unintentionally, and malicious users who create false information on purpose. More details can be seen in Section 2.1.
- Target Victims: Victims are the main targets of the online fake news. They can be users from online social media or other online news platforms. Based on the purposes of the news, the targets can be students, voters, parents, senior people, and so on. More details can also be seen in Section 2.1.
- News Content: News content refers to the body of the news. It contains both physical content (e.g., title, body text, multimedia) and non-physical content (e.g., purpose, sentiment, topics). More information can be found in Section 2.2.
- Social Context: Social content indicates how the news is distributed through the Internet. Social context analysis includes user network analysis (how online users are involved in the news) and broadcast pattern analysis (temporal pattern of the dissemination). This aspect is mainly discussed in Section 2.3.

Fig. 3illustrates an example of fake news shared by a Facebook user. In this Figure, A, B, C, and D represents creator/spreader, news content, social context, and target respectively. We can see that, www.dailypresser.com is the sources of the news, and the Facebook user *Bob* is the news spreader. Both of them can be considered as A. B includes the title of the news, the body of the news, the multimedia of the news if available, and the comment from *Bob*. C includes all the interactions between other users and this news (e.g., comments, likes or dislikes, the timestamp, and so on). And any users who involve with this news by the above mentioned ways can be considered as potential targets, which is D.

2.1. Fake news creators and target users

It is significant to demonstrate who is behind the fake news and why the fake news is written and shared throughout the social media. The creator/spreader of the fake news can be either real human beings or non-humans.

- **Non-humans**: Social bots and cyborgs are the most common non-human fake news creators. Social bots are computer algorithms that are designed to exhibit human-like behaviors, and automatically produce content and interact with humans on social media (Ferrara, Varol, Davis, Menczer, & Flammini, 2016). Although some social bots perform important roles in the spread of legitimate information (Ratkiewicz et al., 2011), many bots are designed specifically to distribute rumors, spam, malware, misinformation, slander, or even just noise (Ferrara et al., 2016). For example, millions of social bots are created to support either Trump or Clinton in 2016 U.S. election, injecting thousands of tweets pointing to websites with fake news (Shu et al., 2017). Cyborgs refers

to either bot-assisted humans and human-assisted bots (Chu, Gianvecchio, Wang, & Jajodia, 2012). After being registered by a human, the cyborg account can post tweets and participate with the social community. Similar to social bots, malicious cyborg accounts mislead and exploit online social users by disseminating fake information and messages, which may result in damaging the social belief and trust. With the nature of bot, cyborg becomes an essential platform for spreading fake news fast and easily. In this paper, the cyborg is considered as one type of non-human fake news creators.

- **Real humans**: Real humans are crucial sources for fake news diffusion. Actually, social bots and cyborgs are only the carriers of fake news on social media, those automate accounts are being programmed to spread false messages by humans. No matter the fake news is spread manually or automatically, real humans, who aim to disturb the credibility of online social community, are the ultimate creators for the untreatable information. The fake contents are generated intentionally by the malicious online users, so it is really difficult to distinguish between fake information and truth information only by content and linguistic analysis (Shu et al., 2017). Furthermore, some benign online users can also contribute to the distribution of fake news. For example, the following news: "*FBI agent suspected in Hillary E-mail leaks found dead in apparent murder-suicide*" is completely false, but this news is shared on *Facebook* over half a million times (Denver guardian). It is obvious that many legitimate users become the spreaders of the news. This message can be posted and shared in the certain community groups, where the friends and followers of the legitimate users may behave as the next-generation spreaders as well. Therefore, an echo chamber is formed which makes the propagation of the false news widespread. Due to the anonymous nature of the Internet, online users do not need to take responsibilities for what they post, share and comment. This is problematic since the unidentified messages may undergo far-reaching dissemination, and may have material impacts on the Internet.

The target users of the fake news may be varied, depending on different purposes of the dishonest information. Voters and citizens can be the target users of fake political claims; online customers can be the target users of online fake reviews and fake advertisements; parents can be the target users of fake educational news; senior people can be the target user of fake health news. For instance, a *11-year-old* girl reported that she was followed by a man who tried to cut off her hijab with scissors in Canada (Cbc news). Although it had been identified as false by Toronto policemen later, this news still underwent widespread dissemination with the powerful capability of the Internet, and it is appears that everyone can fall victim to fake news.

2.2. News content

Each piece of news is consist of physical news content and non-physical news content.

- **Physical News Content**: As shown in Fig. 2, the physical content of fake news contains the title of the news, the main body of the news, and the other elements such as images or videos of the news. As the extensive growth of social network, online social data like tweets and *Facebook* posts become powerful medium of information sharing. The contents of the online social data are valuable and meaningful sources for news content. As they occur in a real-time manner, the online social messages are good representations of the hot trending events (Kwak, Lee, Park, & Moon, 2010; Russell, 2013). In this case, every component contained in the online social data, such as an *Uniform Resource Locator* (URL) of a webpage, a hashtag, a mention signal, an emoji, an image, a video, are all considered as the physical content of the news. Due to the certain meanings and functionalities, those components are important features for fake news detection.
- **Non-physical News Content**: Physical contents are the carriers and formats of the news, and non-physical contents are the opinions, emotions, attitudes and sentiments that the news creators want to express.
  As proposed in our fake news definition, fake news may have different categories, like fake reviews, fake advertisements, fake political news, and so on. Everyday, millions of reviews are produced via online shopping platforms like *Amazon* and *eBay*. Low quality and biased reviews are big issues for both online customers and brands. The fake reviews can not only affect the decision-making process, they can easily destroy a brand's reputation as well. Similar to fake reviews, fake advertisements are written specifically to mislead customers by advertising products with false and unproven information. Both fake reviews and fake advertisements are dangerous for damaging the credibility of online e-commerce. As aforementioned, fake political news play a pivotal role in the 2016 U.S. presidential election (Uscinski, Klofstad, & Atkinson, 2016). With the dramatic circulation of the false information, fake political news has become a major concern in the society. And it is important to track and detect the misleading political claims in order to build a credible online social environment.
  Non-physical content is the main kernel of the fake news, since it contains all the important ideas, feelings and views that the authors want to pass to the readers. Sentiment polarity is another important feature of non-physical content for fake news. In order to make their news persuasive, authors often express strong positive or negative feeling in the text body (Devitt & Ahmad, 2007). Apart from different categories and sentiment polarities, the fake news may target certain domains and themes, like fake social news, fake financial news, fake IT news, and so on.

2.3. Social context

Social context refers to the entire activity system and social environment in which the dissemination of the news operates, includes how the social data is distributed and how online users interact with each other. Today, the ways of sharing and spreading information are increasingly dominated by the interactive technologies on social media (Olteanu, Kıcıman, & Castillo, 2018). And the

social context of an online social news can provide important information on differentiating valuable social news from the huge amounts of messages.

Depend on different media platforms, nowadays news can be shared and transferred through mainstream media like TV, rideo and print, or online social medial like *Twitter, Facebook*, and *Instagram*. Due to the low-cost and easy-access of the Internet, an increasing amount of traditional media, such as *NBC News, New York Times* and *Washington Post*, undergoes dramatic transformations from mainstream platforms to digital platforms. Online social media is changing the way we consume news and information, online users can not only learn about the trending events, they can share their stories and advocate for problems and issues as well. From friends to followers, online social users share experiences and interactions within certain social groups. If a group of like-minded individuals post, share, forward a certain piece of information, the influence of this message may be amplified, this is called "echo chamber effect". This effect can facilitate the diffusion of fake news since online users may be exposed to the social community in an exaggerated distorted form Shu et al. (2017).

As a another key aspect for fake news, social context is an influential indicator of the distribution patterns of both the false news and the real news. The social communities of online users and the social context of fake news dominate how widely and quickly the news propagate in the social network, and they are essential attributes for fake news detection (Castillo et al., 2011; Yang, Liu, Yu, & Yang, 2012).

3. Fake news detection – practical-based approaches

Simply speaking, fake news detection is the task of assessing the truthfulness of a certain piece of news (Vlachos & Riedel, 2014). And the current fake news detection resources can be summarized into two categories: (1) the practical-based detection approaches, from the perspective of Internet users, and (2) the existing research-based detection methods, from the perspective of academia and research. And in this section we mainly discuss the practical-based approach for online fake news identification.

3.1. Online fact-checking resources

Fact-checking resources are commonly performed by mainstream media organizations. Because real-time news is always the mixture of information, sometimes a binary classification result can not fully explain the overall problem. So many evaluation criteria or visual metrics are used to determine the truthful level of the news in current fact-checking resources.

By telling users what is true, false, or in-between, fact-checking is a good way for fake news identification (Factcheck). In the following, we evaluate and compare some popular fact-checking online resources.

- Classify.news is an online platform for fake news article identification. Their goal is to build a model to discern the credibility of an article based solely on its textual content using machine learning algorithms (Classify.news). More specifically, they collect labeled news articles from *OpenSources* (Opensource), and implement a daily learning on those credible and non-credible website samples. There are two types of prediction models: "Content-only" model with multinomial Nave Bayes classifier and "Context-only" model with adaptive Boosting classifier.
- FackCheck.org is a nonprofit "consumer advocate" webpage for voters that aims to reduce the level of deception and confusion in U.S. politics (Factcheck). They evaluate the factual truthfulness of the claims or statements by major U.S. political players. Those claims and statements are originated from various platforms, include TV advertisements, debates, speeches, interviews, new releases and social media. They mainly focus on presidential candidates in presidential election years, and evaluate the factual accuracy of their statements systematically. With the help of trustworthy inside and outside experts, they finally analyse and report the reliability of each piece of information. FactCheck.org is also consist of several other components, such as *SciCheck* for science-based claims fact-checking, *Health Watch* for health care debate fact-checking, *Facebook Initiative* for debunking fake news on *Facebook*.
- Factmata.com is a *Google* fully-funded project for statistical fact-checking and claim detection (Factmata). The most significant feature of this project is that the claim checking task is depend entirely on artificial intelligence and machine learning algorithms. Based on the advanced Natural Language Processing (NLP) techniques, Factmata.com can identify and check statistical claims by numerical relation extraction (Dale, 2017). The objective of this platform is to detect and verify misinformation and provide a better informed opinion on the world. Also, they can help advertisers avoid placing advertising on fake news, hate speech and extremist content.
- Hoaxy.iuni.iu.edu is a framework for collection, detection, and analysis of online misinformation and its related fack-checking efforts (Shao, Ciampaglia, Flammini, & Menczer, 2016). In this platform, a preliminary analysis of a sample from public tweets containing both fake news and fact-checking information are illustrated via interactive visualization. In their online visualization system, users can search for topics that they are interested in, and visualize the distribution of fake claims and the corresponding fact-checking information. They can also visualize the networks and activities of fake news spreaders and fact-checkers.
- Hoax-Slayer.com focuses on debunking email hoaxes, thwarting Internet scammers, combating spam, and educating web users about email and Internet security issues (Hoaxslayer.com). They also counteract criminal activities by publishing information about Internet scams, sharing anti-spam tips, publishing computer and email security information. They thoroughly research all the potentials hoaxes based on information available via a variety of credible sources, including reputable websites, news articles, press releases, government or company publications and consumer alerts. Also they may contact companies, government departments, or other relevant entities directly to enquire about the veracity of particular messages.

- PolitiFact.com is an U.S. website that rates the accuracy of claims or statements by elected officials, pundits, columnists, bloggers, political analysts and other members of the media. PolitiFact.com is an independent, no-partisan source of online fact-checking system for political news and information (Politifact). They use the following judgments to rate the truthfulness of a certain claim: True, Mostly True, Half True, Mostly False, False and Pants on Fire. The editors examine the specific word and the full context of a claim carefully, and then verify the reliability of the claims and statements. The PolitiFact website also provides API for users to get access to the full text of statements, stories, promises and updates that have been checked.
- Snopes.com is widely known as one of the first online fact-checking websites for validating and debunking urban legends and similar stories in American popular culture (Snopes.com). This webpage covers a wide ranges of disciplines including automobiles, business, computers, crime, fraud and scams, history, and so on. With the knowledge from professional individuals and organizations, Snopes.com can provide comprehensive evaluations for various types of printed resources, and assign truth ratings to them. And this site has been profiled by many major news organizations, including *CNN, MSNBC, Fortune, The Washington Post*, and *New York Times*.
- TruthOrFiction.com is a non-partisan online website where Internet users can quickly and easily get information about e-rumors, warnings, hoaxes, virus warnings, and humorous or inspirational stories that are distributed through emails (Truthorfiction). This website mainly focuses on misleading information that are popular via forwarded emails. And they rate stories or information by the following categories: Truth, Fiction, Reported to be Truth, Unproven, Truth and Fiction, Previously Truth, Disputed and Pending Investigation.
- Other related online resources such as OpenSecrets.org, OpenSources, FakeNewsWatch.com, fakespot.com, reviewmeta.com and so on.
  OpenSecrets.org is a non-partisan guide to track money matters in U.S. politics and their effects on elections and public policy (Opensecrets). OpenSources is a professional online sources which can provide a continuously updated database of fake, false, conspiratorial and misleading news (Opensource). FakeNewsWatch.com is an online blacklist-based platform that tracks hoax, fake news, satire and clickbait websites. fakespot.com and reviewmeta.com are popular online review checking websites. With the implementation of machine learning techniques, fakespot.com helps consumers to filter out suspicious online reviews on Amazon.com (Fakespot). reviewmeta.com uses statistical modeling on suspicious online reviews and helps users to navigate millions of feedbacks on the products that they are interested in Reviewmeta.

Table 1shows detailed comparison of the existing fact-checking resources, in terms of “Topic coverage”, “Source of the fake news”, “Rating levels”, “Dashboard & visualization”, “API”, “Detection technology” and “Others”. Overall speaking, 100% automate fake news detection is still difficult and has a long way to go. Popular fact-checking websites like Snopes.com, PolitiFact.com and FactCheck.org are solely depend on manual detection by professional experts and organizations. However, this process may be time-consuming and expensive, and a large human involvement is necessary for maintaining such detection systems (Dale, 2017). It is essential to develop automatic detection approaches, in order to improve the generality and performance of the existing systems.

There are some newly emerging online platforms such as Classify.news and Factmata.com, which leverage the most advanced artificial intelligence and machine learning algorithms in the task of fake news identification. However, Factmata.com can only check claims or statements which contain statistical information. And Classify.news is mainly on supervised machine learning techniques which need high-quality training dataset and labels. The open issues and difficulties in fake news detection keep motivating researchers and practitioners to build more systematical models for adaptive, automate and comprehensive prediction approaches.

3.2. Social practical guide for fake news detection

As aforementioned, there are some issues and limitations of the current fact-checking resources, such as the detection process is time-consuming, the results are always delayed, and a large amount of manual labour should be involved. So it is essential for online Internet users to improve their own distinguishing capacities for online fake information. And in this section, we present some useful practical-based social theories for fake news identification. Same as the characterization of fake news discussed in Section 2, the practical social guidance can be summarized as creator-based approach, new content based approach and social context based approach.

- Creator-based approach: More and more researchers believe that the best chance for detecting fake news is not focusing on the claims themselves, but on the news sources. There are many useful social hints that can help Internet users to detect suspicious fake sources and the potential false information. For example, if the news source is from a popular web domain or an unknown domain? It is even possible to identify a malicious website only by checking the lexical property of the URL, such as if there is any abnormal domain name (e.g., “.com.co”) or suspicious tokens in the URL (Zhang, Lashkari, & Ghorbani, 2017). Sometimes the “About Us” or “Disclaimer” section offers useful information about the webpage, and can be used as a credibility indicator.
- News content based approach: As thousands of news are spread online everyday, it is perhaps easier for Internet users to share the news with eye-popping headlines than even read them. In addition, in many studies, Internet users have shown their weak abilities for discerning false from true information (Kumar & Shah, 2018). Here are some practical social theories which can help users to detect suspicious news content.
  - Do not stop at headline: In fake news, the headlines are always outrageous and eye-popping in order to receive more clicks and attention. The news content could be nothing related to the headlines or even conflict with the facts that expressed in the headlines. Reading the whole story rather than stopping at the headline is a good strategy for defining any skeptical online

Table 1
The comparison of some existing online fact-checking systems.

Fact-checking Websites	Topics Covered					Source of the news	Rating levels	Dashboard or Visualization	API	Detection Technology	Related Information Provided
	Civic	Politics	Technology	Business	Others						
Classify.news	No	Yes	No	No	No	Online websites	Credible, Non-credible	Yes	No	Automate detection by machine learning algorithms	Simple interactive visualization with users
FactCheck.org	Yes	Yes	Yes	Yes	Yes	Claims or statements from political players	N/A	No	No	Manual detection by fact-checkers or related expertise	Users can ask questions on certain topics
Factmata.com	No	Yes	No	Yes	No	N/A	N/A	No	No	Automate detection by artificial intelligence	N/A
Hoaxy.iuni.iu.edu	Yes	Yes	Yes	Yes	Yes	Social media and fact-checking websites	N/A	Yes	Yes	Detection based on other online fact-checking resources	Interactive visualization of fake claims and fact-checking information
Hoax-Slayer.com	Yes	Yes	Yes	Yes	Yes	Emails	N/A	No	No	Manual detection by fact-checkers	Information about computer security, malware threats, email security and spam control
PolitiFact.com	Yes	Yes	Yes	Yes	Yes	Claims or statements from political players	True, Mostly True, Half True, Mostly False, False, Pants on Fire	No	Yes	Manual detection by fact-checkers or related expertise	It also rates an official's consistency on an issue
Snopes.com	Yes	Yes	Yes	Yes	Yes	Various platforms	True, Mostly True, Mostly False, False Mixture				
	No	No	Manual detection by journalists, reporters or researchers, writers, etc.	The original full text of the news, Major claim of the news, True part of the news, False part of the news, Undetermined part of the news							
TruthOrFiction.com	Yes	Yes	Yes	Yes	Yes	Emails	Truth, Unproven, Truth and Fiction, Previously Truth, Disputed, Pending	No	No	Manual detection by fact-checkers	N/A

information.

- Check the supporting resources: In order to convince readers, news authors always include plenty of facts, such as the knowledge from experts, some statistical or survey data, supporting documents, references and even external links in the news content. Take a time to read those supporting resources since it could also help Internet users to understand the truthfulness of the news content.
- Check the sentiment and sensitive topics of the news: Most of the fake news purposely play on readers' fears, anxieties, sympathy, curiosity, trepidation, and so on. So before following the emotions and opinions in the news, Internet users should be responsible for detecting the sensitive sentiment level of the news, such as if the news make you angry or sad. Also, online readers need to figure out some sensitive topics and always ask themselves, if the news is too funny or interesting to be true; if the news talks about some miserable stories; if the news predict a future disaster like earthquake or epidemic disease; if a major illness is being cured, and so on.

- Social context based approach: Rather than focusing on the news content, another practical way is to capture skeptical social context of the online news. Some useful examples include: check if the news from the same source are incredulous or not; check the date of news or the supporting resources; check if there is any other online news platforms report the same or similar stories, and what is their credibility.

4. Fake news detection – research-based approaches

4.1. The existing research-based approaches

In this section, we review and discuss the stat-of-the-art studies on fake news detection. Table 2 illustrates the overall categorizations of the current research on online fake news detection, from which we can discover the differences of detecting different types of false information in terms of features, data mining algorithms, and platforms.

Recently, the development of online social media rise the widespread dissemination of online fake news. The information distributed via social networks is massive, fast, wide-ranging, diverse, and heterogeneous, thus online false information can cause severe impact to the whole society. As a result, more and more researchers are working on detecting false information and fake news on online social media, and this trend can also be discovered in Table 2. From Table 2, the number of studies focusing on fake news detection is more significant than other topics (e.g., rumor or satire detection).

Because online fake reviews and rumors are always compacted and information-intensive, their content lengths are often shorter than online fake news. As a result, traditional linguistic processing and embedding techniques such as *bag-of-words* or *n-gram* are good for processing reviews or rumors, but they are not powerful enough for extracting the underlying relationship for fake news. So for online fake news detection, sophisticated embedding approaches are necessary in order to capture the key opinion and semantic sequential order in news content. As aforementioned, with the development of deep learning techniques in recent years, algorithms like *Recurrent Neural Network* or *Auto-Encoder* are powerful tools for embedding natural language, remembering the important semantic sequential orders, and capturing the underlying semantic relationships. So deep learning algorithms are used more frequently in online fake news detection. In addition, in the area of "rumor or satire detection", most of the studies only focus on content based analysis. However, most of the online fake news studies combine information from creator, news content and social context.

There are mainly two types of systems: *practical-based approaches* and *research-based approaches*. Practical-based approaches are discussed in Section 3. And in this section, we present the detailed descriptions of research-based approaches for fake news detection. As aforementioned, the verification of a piece of news can not only depend on the news content, creator of the news and the social context of the news are also influential factors. With the additional information such as the credibility of the news creator, and the underlying distribution pattern of the news, we can have better understanding of the news, and make more accurate prediction. There are three types of models for research-based approaches: *component-based category, data mining based category*, and *implement-based category*. And each model and its subcategories are discussed as follows.

4.1.1. Component-based category

As mentioned in Section 2, fake news contains four major components: fake news creator, target victim, fake news content, and fake news social context. Based on the analysis of different components, fake news detection approaches can be divided as: *creator and user analysis, news content analysis* and *social context analysis*.

- **Creator and user analysis**: There are extensive attempt and efforts on the analysis of malicious accounts on social media. In this section, we review and discuss the major methodologies used for creator and user analysis.
  Being exposed to a large amount of unproven messages, online users lack the clues to evaluate the credibility of the social information (Castillo et al., 2011). Malicious social media accounts intent to manipulate people's decision and pollute the truth news content by purposely spreading misinformation (Davis, Varol, Ferrara, Flammini, & Menczer, 2016). So creator and user analysis is a critical aspect for fake news detection. With unique characteristics, malicious social media accounts behave different from legitimate users. And the creator and user analysis can be categorized across the following differences: *user profiling analysis, temporal and posting behavior analysis, credibility-related analysis*, and *sentiment-related analysis*.
  (1) User profiling analysis: The basic user profiling information includes the language used by the account, the geographic locations of the account, the account creation time, if the account is verified or not, how many posts/tweets does the account have, and so on (Ferrara et al., 2016; Zhao et al., 2014). The user profiling analysis describes how active and suspicious a social

Table 2
The categorization of the existing research-based online fake news detection approaches.

	Component-based category			Data mining based category	
	Creator analysis	Content analysis	Context analysis	Supervised learning	
				Deep learning	Machine learning
Fake News	Yang et al. (2019) Del Vicario et al. (2018) Kim, Tabibian, Oh, Schölkopf, and Gomez-Rodriguez (2018) Farajtabar et al. (2017) Shao, Ciampaglia, Varol, Flammini, and Menczer (2017) Ruchansky et al. (2017) Tschiatschek, Singla, Rodriguez, Merchant, and Krause (2017) Cao et al. (2016) Mitra and Gilbert (2015) Chen, Conroy, and Rubin (2015) Dickerson et al. (2014) Zhao et al. (2014) Afroz et al. (2012) Chu et al. (2012) Castillo et al. (2011)	Dong et al. (2018b) Della Vedova et al. (2018) Shu, Bernard, and Liu (2019a) Wang et al. (2018) Del Vicario et al. (2018) Horne and Adali (2017) Singhania, Fernandez, and Rao (2017) Ruchansky et al. (2017) Tacchini et al. (2017) Tschiatschek et al. (2017) Shao et al. (2017) Pérez-Rosas, Kleinberg, Lefevre, and Mihalcea (2017) Kumar et al. (2016) Mitra and Gilbert (2015) Conroy et al. (2015) Chen et al. (2015) Dickerson et al. (2014) Afroz et al. (2012) Castillo et al. (2011)	Yang et al. (2019) Dong et al. (2018b) Shu et al. (2019b) Shu et al. (2019a) Della Vedova et al. (2018) Del Vicario et al. (2018) Farajtabar et al. (2017) Tacchini et al. (2017) Ruchansky et al. (2017) Kumar et al. (2016) Cao et al. (2016) Conroy et al. (2015) Chen et al. (2015) Zhao et al. (2014) Chu et al. (2012) Castillo et al. (2011)	Dong et al. (2018b) Shu et al. (2019a) Wang et al. (2018) Farajtabar et al. (2017) Singhania et al. (2017) Ruchansky et al. (2017)	Shu et al. (2019b) Della Vedova et al. (2018) Del Vicario et al. (2018) Kim et al. (2018) Tacchini et al. (2017) Shao et al. (2017) Tschiatschek et al. (2017) Pérez-Rosas et al. (2017) Kumar et al. (2016) Conroy et al. (2015) Chen et al. (2015) Dickerson et al. (2014) Afroz et al. (2012) Chu et al. (2012) Castillo et al. (2011)
Fake Review or Fake Ad	Karumanchi, Fu, and Deng (2018) Youli, Hong, Ruitong, Lutong, and Li (2018) Di, Wang, Fang, and Zhou (2018) Dematis, Karapistoli, and Vakali (2018) Dong et al. (2018a)	Karumanchi et al. (2018) Narayan, Rout, and Jena (2018) Youli et al. (2018) Di et al. (2018) Dematis et al. (2018) Dong et al. (2018a)	Karumanchi et al. (2018) Youli et al. (2018) Di et al. (2018) Dong et al. (2018a)	Dong et al. (2018a)	Karumanchi et al. (2018) Narayan et al. (2018) Youli et al. (2018) Di et al. (2018) Dematis et al. (2018)
Rumor or Satire	Yang et al. (2012) Afroz et al. (2012)	Wu et al. (2017) Rubin et al. (2016) Rubin et al. (2016) Yang et al. (2012) Afroz et al. (2012)	Ma et al. (2016) Ma et al. (2015) Zhao et al. (2015) Yang et al. (2012)	Ma et al. (2016)	Wu et al. (2017) Rubin et al. (2016) Ma et al. (2016) Rubin et al. (2016) Zhao et al. (2015) Yang et al. (2012) Afroz et al. (2012)

	Data mining based category	Implementation- based category		Platform-based category	
	Unsupervised learning	Online	Offline	Social media	Other online news platform
Fake News	Yang et al. (2019) Shao et al. (2017) Cao et al. (2016) Chen et al. (2015) Zhao et al. (2014)	Farajtabar et al. (2017) Cao et al. (2016) Mitra and Gilbert (2015) Shao et al. (2016) Zhao et al. (2014)	Yang et al. (2019) Dong et al. (2018b) Shu et al. (2019b) Shu et al. (2019a) Della Vedova et al. (2018) Wang et al. (2018) Del Vicario et al. (2018) Kim et al. (2018) Shao et al. (2017) Ruchansky et al. (2017) Singhania et al. (2017) Tacchini et al. (2017)	Yang et al. (2019) Della Vedova et al. (2018) Shu et al. (2019a) Wang et al. (2018) Del Vicario et al. (2018) Kim et al. (2018) Shao et al. (2017) Singhania et al. (2017) Tacchini et al. (2017) Farajtabar et al. (2017)	Dong et al. (2018b) Shu et al. (2019b) Pérez-Rosas et al. (2017) Tschiatschek et al. (2017) Ruchansky et al. (2017) Kumar et al. (2016) Conroy et al. (2015) Shao et al. (2016) Chen et al. (2015) Afroz et al. (2012)

(*continued on next page*)

Table 2 (*continued*)

	Data mining based category	Implementation- based category		Platform-based category	
	Unsupervised learning	Online	Offline	Social media	Other online news platform
Fake Review or Fake Ad			Tschiatschek et al. (2017) Pérez-Rosas et al. (2017) Kumar et al. (2016) Chen et al. (2015) Afroz et al. (2012) Chu et al. (2012) Castillo et al. (2011) Karumanchi et al. (2018) Narayan et al. (2018) Youli et al. (2018) Di et al. (2018) Dematis et al. (2018) Dong et al. (2018a)	Cao et al. (2016) Mitra and Gilbert (2015) Shao et al. (2016) Zhao et al. (2014) Chu et al. (2012) Castillo et al. (2011)	Karumanchi et al. (2018) Narayan et al. (2018) Youli et al. (2018) Di et al. (2018) Dematis et al. (2018) Dong et al. (2018a)
Rumor or Satire	Ma et al. (2015) Zhao et al. (2015)		Wu et al. (2017) Ma et al. (2016) Zhao et al. (2015) Yang et al. (2012) Afroz et al. (2012)	Wu et al. (2017) Ma et al. (2016) Zhao et al. (2015) Ma et al. (2015) Yang et al. (2012)	Rubin et al. (2016) Ma et al. (2016) Rubin et al. (2016) Afroz et al. (2012)

account is, and has been shown useful for suspicious social account detection (Zhao et al., 2014). (2) Temporal and posting behavior analysis: Temporal behavior reveal the temporal patterns of the online social account, such as the signal similarity to a Poisson process (Ghosh, Surachawala, & Lerman, 2011), the average time between two consecutive posts, the frequency of replying, sharing, mentioning, and so on. Driven by timers or automatic programs, suspicious accounts like social bots and cyborgs are more active in a certain time period (Gianvecchio, Wu, Xie, & Wang, 2009; Gianvecchio, Xie, Wu, & Wang, 2008). In contrast, legitimate human users have complex timing behaviors (Chu et al., 2012). In 2016 U.S. presidential campaign, by disguising their geographic locations and being active in replying and mentioning activities, social bots play a disproportionate role in spreading and repeating misinformation (Murthy et al., 2016). The high intensity of reply and mention behaviors may indicate the high suspicious level of a social account. (3) Credibility-related information: The numbers of friends and followers are also good features for differentiating malicious accounts and legitimate users. The number of followers of a legitimate social user is often close to its friends. However, social bots usually have much more friends than followers (Mislove, Marcon, Gummadi, Druschel, & Bhattacharjee, 2007). Chu et al. (2012) proposes an equation for calculating account reputation with the number of followers and friends. The equation is defined as: $Account\_Reputation = \frac{follower}{follower + friend}$, and they observe that a famous celebrity always has a reputation score that close to 1, whereas the score for a suspicious social bot is close to 0. (4) Sentiment-related analysis: Sentiment-related factors are also key attributes for suspicious account identification (Dickerson, Kagan, & Subrahmanian, 2014). By triggering anomalous emotional response, malicious accounts can exaggerate the facts and mislead legitimate users. Sentiment analysis is a useful way for illustrating the emotion, attitude and opinion that are conveyed by online social media. Psychological keywords analysis is a common way for indicating the original author's emotion and sentiment (Zhao et al., 2014). Various approaches have been proposed for extracting sentiment-related analysis, like arousal-valence-dominance score (Warriner, Kuperman, & Brysbaert, 2013), happiness score (Dodds et al., 2015), emotion score (Agarwal, Xie, Vovsha, Rambow, & Passonneau, 2011) and polarization and strength (Wilson, Wiebe, & Hoffmann, 2005). By combining various sentiment variables, Dickerson et al. (2014) shows that sentiment-related behavior is sufficient for distinguishing human account and social bot account.

- **News Content analysis**

  As shown in Fig. 2, a piece of fake news contains the physical content (like title, body text, image or video), and the non-physical content (like purpose, sentiment, and news topics). By exploiting in-depth news content analysis, we can analyze linguistic patterns and writing styles for both truth news and fake news, and then capture the most discriminative features for online fake news detection. The current studies on news content analysis can be categorized as: *linguistic and semantic-based analysis, knowledge-based analysis*, and *style-based analysis*.

  - **Linguistic and semantic-based analysis**: Both linguistic and semantic-based analysis are classic and scientific studies of natural language. By extracting useful information from the news content, linguistic and semantic analysis can analyze the associate language patterns, structures and meanings of the news.

    (1) Linguistic-based analysis: Online fake news is generated intentionally by the fake news creators for financial, political, or other gains (Zhao et al., 2014). Most of the fake news creators use specific writing strategies to avoid being detected (Conroy, Rubin, & Chen, 2015). The primary goal of linguistic analysis is to match the news creator's language competence by observing the language formats and discovering the writing patterns (Raskin, 1987).

    "*Bag-of-words*" and "*n-grams*" are the most common methods for representing raw news texts (Ahmed, 2017; Conroy et al., 2015). In "*bag-of-words*", by regarding each word as a single and equal unit, the raw news text can be represented as the set of its words, disregarding language grammar and the word order. In "*n-gram*", the raw news text is represented by a contiguous sequence of *n* items, the items can be phonemes, syllables, letters, or words. However, the simplicity of these two approaches also leads to some obvious shortcomings for raw text processing, for instance, "*n-gram*" model is extreme sparsity, and it can not interpret news samples that contain unknown tokens; "*bag-of-words*" may lose significant information by ignoring the context and semantic of the words. In recent years, other techniques have been proposed and used for natural language representation and document classification, such as deep syntax analysis (Popel & Žabokrtskỳ, 2010), *word2vec* (Goldberg & Levy, 2014), *long short-term memory* (LSTM) neural network (Sundermeyer, Schlüter, & Ney, 2010), *sequence-2-sequence* based deep neural network (Sutskever, Vinyals, & Le, 2014), and so on.

    (2) Semantic-based analysis: Semantic-based analysis refers to the process of characterizing the syntactic structures of the news from phrases levels to semantics level. By uniting *"n-gram"* model with deep syntax model, semantic-based analysis can discover the degree of compatibility and consistency between the news creator's personal experience and the news content (Conroy et al., 2015). For example, the fake news creators often use exaggerated title to attract readers' attention, so the title of the fake news is usually unrelated or in conflict with the news content. However, the title of a true news should be consistent with the content of the news body. A fake online review may contain contradictions or mistakes in the comments, since the deceptive reviewers have no experience with the functionalities and services about the products. So semantic-based analysis can provide important clues for assessing the suspicious level of online news. With the combined information from news creator analysis and semantic-based analysis, researchers can verify the compatibility between the user's background and the news content, which has shown good improvement for false information classification and detection (Feng & Hirst, 2013).

  - **Knowledge-based analysis**: Knowledge-based analysis refers to the attempts to directly check the truthfulness of the major claims in a news (Shu et al., 2017). The aforementioned fact-checking websites like Snopes.com, PolitiFact.com and FactCheck.org are typical examples of knowledge-based fake news detection websites. In these websites, external and professional resources, like the knowledge from an expert or an organization, are necessary for assigning truthful value for a piece of news.

Knowledge-based analysis is a fundamental component of online fake news detection in terms of the following two perspectives. First, artificial intelligence (AI)-based learning models are feasible solutions for online news evaluation. However, misleading online messages with different writing patterns and purposes are emerging everyday, which make it difficult for the AI-based models to maintain a high detection performance. Although many novel techniques are proposed recently for automatic fake news detection, nowadays fact-checking tasks are still mainly depend on human's knowledge. Second, by considering fake news detection as a binary classification task, we can build supervised machine learning models for classifying fake news from true news. And the very first task for establishing an automatic machine learning model is to collect a high quality dataset with labels. As mentioned in Section 2, online fake news is diverse in terms of topics, purposes, domains, styles, and platforms. So it is difficult to generate a complete fake news dataset for training such models. In addition, the real-world datasets of fake news are always incomplete, unstructured, unlabeled and noisy (Shu et al., 2017), which make automatic detection even more problematic. In this case, a large amount of human efforts are essential for collecting and labeling fake news datasets, and knowledge-based analysis is a critical aspect for generating effective machine learning models, and for identifying online fake news.

- **Style-based analysis**: Legitimate online users express their opinions, emotions, and feelings toward certain products, events, and services via social media (Ahmed, 2017). Whereas, malicious online accounts express deceptive information by intentionally obfuscating their writing style or attempting to imitate other users (Afroz, Brennan, & Greenstadt, 2012). By trying to capture the distinguishing characteristics of writing styles between legitimate users and anomalous accounts, style-based analysis plays an important role in online fake news identification. From the perspective of analytic target, style-based analysis can be divided as *physical style analysis* and *non-physical style analysis*. (1) Physical style analysis is the process of extracting influential physical features to distinguish fake news from honest news. These features can illustrate the writing style, the text syntax and the personal attitude of the news, such as the number of verbs and nouns, the number of emotion words and casual words (Horne & Adali, 2017). The presence of suspicious tokens, such as the number of URLs, hashtags, mentions, and the uppercase words in social communication data are also good features for authorship identification and writing style analysis (Castillo et al., 2011; Horne & Adali, 2017; Jin, Cao, Zhang, & Luo, 2016). (2) Non-physical style analysis analyzes the non-physical aspects of the news, such as the complexity and readability of the news text (Horne & Adali, 2017). Based on Ahmed (2017) and Banerjee, Feng, Kang, and Choi (2014), fake news creators usually take longer time and make more mistakes during their writing. So some specific keystroke patterns can be traced for writing style analysis. For instance, the key "*backspace*" and "*delete*" are used more often when a fake news creator want to write some false messages (Ahmed, 2017). The authenticity of a news or a document heavily depends on the authenticity of its author (Afroz et al., 2012). Style-based analysis can provide important information on representing an author's writing style, therefore, it should obtain more attention in online fake news detection.
- **Social context analysis**: As mentioned in Section 2.3, social context is the social environment in which the news disseminates. Social context analysis is the study of how quickly and widely the social data is distributed, and how online users interact with each other. However, most of the recent approaches for online fake news detection are related to direct news content analysis, few studies use social context analysis for predicting anomalous online information (Ma, Gao, Wei, Lu, & Wong, 2015; Shu et al., 2017). In this survey, we propose two types of social context analysis, which can be potential candidates for enhancing the performance of the existing fake news detection methods. They are *user network analysis* and *distribution pattern analysis*. (1) User network analysis: It is believed that the truthfulness of a piece of online news can be identified on the network of the news creator (Markines, Cattuto, & Menczer, 2009; Ruchansky, Seo, & Liu, 2017; Tacchini, Ballarin, Della Vedova, Moret, & de Alfaro, 2017). Over the recent years, social media like *Twitter* and *Facebook* are fast growing social network services, and they can provide strong and interactive communication platforms for online users. Different online users have different education backgrounds, working experiences, and interests, so users on social media tend to form groups containing like-minded users that are similar to them (Shu et al., 2017). The distribution of fake news is typically considered as an epidemic via certain social networks (Castillo, Mendoza, & Poblete, 2013; Jin, Dougherty, Saraf, Cao, & Ramakrishnan, 2013; Kumar, West, & Leskovec, 2016; Ma et al., 2015). In this case, it is reasonable to make the hypothesis: the online users that are highly interact with the news creator can be used to predict the truthful level of the news. For example, if many anomalous or unreliable accounts "*like*" or "*comment*" on a piece of news, then this news is more likely to contain false and misleading information (Castillo et al., 2011; Tacchini et al., 2017). In addition, the credibility of the news creator's social network can also be a good indicator for the credibility of the news (Wu, Yang, & Zhu, 2015; Yang et al., 2012). (2) Distribution pattern analysis: User network analysis is the study to reveal the interaction between online users, whereas distribution pattern analysis is the study to analyze the characteristics of information spreading. Everyday, millions of online news and messages are published, shared, and forwarded by over one billion active online social accounts (Zhao et al., 2014). And it is important to distill trustable news from the abundant information on social media. The diffusion patterns of online news can provide valuable and informative information for suggesting anomalous events and messages (Diakopoulos, Naaman, & Kivran-Swaine, 2010; Song, Wen, Lin, & Davis, 2013). In recent years, many researchers are working on anomalous pattern detection for social information, which can help online users discover data that is different, unusual or unexpected (Chu et al., 2012; Ho, Li, & Lin, 2011; Romero, Meeder, & Kleinberg, 2011; Zhao et al., 2014). So in this survey, we propose that distribution pattern analysis can be considered as one future solution for online fake news detection. However, detect suspicious distribution pattern for fake news is complicate and challenging, due to the heterogeneous and dynamic nature of online social behaviors (Zhao et al., 2014). As a result, advanced visualization systems are always incorporated with classic machine learning algorithms in order to address the above challenges (Adams, Phung, & Venkatesh, 2011; Cui et al., 2011; Wang Baldonado, Woodruff, & Kuchinsky, 2000; Zhao et al., 2014). As the future work for fake news detection, the establishment

of visualization system for online social news is discussed in Chapter 5. In summary, overwhelming information is diffused via social media everyday. And most of the time, news content analysis is not sufficient for building an effective system for false information detection. By quantitatively model the distribution pattern for social communication data, social context analysis is capable of making early predictions of the propagation speed, scale, and impact of information diffusion (Yang & Counts, 2010), and providing diverse and in-depth perspectives for representation online fake news.

4.1.2. Data mining based category

Based on different categories of component analysis, different types of features can be extracted from online communication data, and be leveraged for learning model construction. Following traditional ways of classifying data mining techniques, the existing machine learning models can be summarized as: supervised, semi-supervised and unsupervised models (Han, Pei, & Kamber, 2011; Shu, Wang, Tang, Zafarani, & Liu, 2017). And in this paper, *supervised learning models* and*unsupervised learning models* are mainly discussed for online fake news detection.

- **Supervised learning**: Supervised machine learning algorithms like *Decision Tree, Random Forest, Support Vector Machine* (SVM), *Logistic Regression, K-nearest Neighbour* are extensively used in previous literatures for online hoaxes, frauds, and deceptive information classification (Afroz et al., 2012; Castillo et al., 2011; Davis et al., 2016; Horne & Adali, 2017; Kwon, Cha, Jung, Chen, & Wang, 2013; Tacchini et al., 2017; Yang & Counts, 2010). And many evaluation criterias are used for assessing the performance of different machine learning techniques. The most common metrics are *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), *False Negative* (FN), *Precision* (Pr), *Recall* (Re), *False Positive Rate* (FPR), *False Negative Rate* (FNR), *F-score* and *Accuracy* (Acc). TP, TN, FP, FN are detected fake news, detected true news, misclassified true news, and undetected fake news respectively. And the formulas for calculating other evaluation metrics are: $Pr = \frac{TP}{TP+FP}$, $Re = \frac{TP}{TP+FN}$, $FPR = \frac{FP}{FP+TN}$, $FNR = \frac{FN}{TP+FN}$, $F_{-score} = \frac{2\times Pr\times Re}{Pr+Re}$, $Acc = \frac{TP+TN}{TP+TN+FP+FN}$. Each criteria evaluate the classification performance from a different perspective (Shu et al., 2017). And more detailed explanations for the aforementioned metrics can be found in Han et al. (2011).
  Over the recent years, deep learning algorithms have gained great successes in the domains of speech recognition and visual object recognition (Goodfellow, Bengio, Courville, & Bengio, 2016; LeCun, Bengio, & Hinton, 2015; Schmidhuber, 2015). Different from conventional machine learning techniques which require hand-craft feature extraction, deep learning algorithms can be fed with raw data and discover the representations automatically (LeCun et al., 2015). More specifically, deep learning algorithms like *Recurrent Neural Network* (RNN) is good at revealing sequence structures in high-dimensional data, and has shown dramatic potential in natural language processing, such as topic classification (Collobert et al., 2011), sentiment analysis (Glorot, Bordes, & Bengio, 2011), question answering (Bordes, Chopra, & Weston, 2014), and language translation (Jean, Cho, Memisevic, & Bengio, 2014; Sutskever et al., 2014). Therefore, deep learning based methods are good solutions for online fake news representation and detection, and have been introduced in Ma et al. (2016) and Ruchansky et al. (2017). With deep learning algorithms like LSTM, bidirectional LSTM, *Gated Recurrent Unit* (GRU), fake news detection methods are not rely on hand-crafted textual features, and can capture the hidden implications of news contextual information and author information over time. In addition, the classification performance of deep learning can be further improved via sophisticated computation units and extra hidden layers (Ma et al., 2016).
- **Unsupervised learning**: The performance of a supervised learning model heavily depend on the quality of a labeled dataset. However, it is difficult to generate a wide-covered, good-quality dataset for fake news detection, for the following reasons: (1) the real-world online dataset is usually big, incomplete, unstructured, unlabeled, and noisy (Shu et al., 2017); (2) everyday a large amount of false information with diverse intentions and different linguistic characteristics is created via social media (Ruchansky et al., 2017; Zhao et al., 2014). It is difficult to obtain the ground truth label for the data. Thus, an unsupervised learning model is more practical and feasible for solving real-world problem. However, there are only few studies directly work on detecting online fake news in unsupervised manners. Most of them focus on semantic similarity analysis or sentiment analysis. Ahmed (2017) proposes an unsupervised similarity measurement for online fake reviews. By combining word similarity and word-order similarity, their proposed approach is able to identify near-duplicate online reviews with high accuracy. Wang, Wang, Tang, Liu, and Li (2015) presents an unsupervised sentiment analysis framework for social media images. By exploiting the relations between visual information and the relevant contextual information, their method can predict the sentiment of social images from two large-scale datasets. Mukherjee, Liu, and Glance (2012) leverages an unsupervised generative Bayesian model for online fake review analysis. Based on cosine similarity measurement, they use rating and temporal features to automatically discern separating features of truthful and fraudulent reviewers. We believe that unsupervised learning algorithms are practical and essential directions for online fake news detection, and should be given top priority in the future research. In Chapter 5, some promising unsupervised learning approaches are introduced, including clustering analysis, outlier detection analysis, semantic similarity analysis, and unsupervised news embedding.

4.1.3. Implement-based category

In terms of how the system is executed, fake news detection can be categorized as *real-time detection* and *offline detection*.

In the offline detection system, batch-sized machine learning models are usually applied for fake news identification. Offline detection system is important for online fake news classification, since they can analyze anomalous information in a descriptive manner, such as select the most influential features for discriminate false information among large amounts of social messages. Based on the types of online information, offline classification can be divided as *fake review detection,satire news detection, hoaxes detection*,

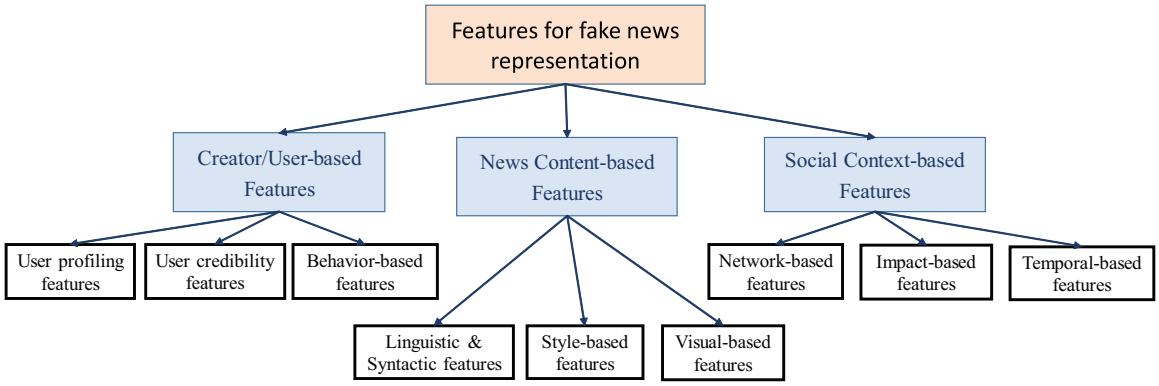

Fig. 4. Different types of features for fake news representation and identification.

and *political news detection*. Shu et al. (2017) also gives a brief summary of some other related areas, such as *rumor classification, truth discovery, clickbait detection*, and *spammer and bot detection*.

However, offline system is limited. The datasets used may not represent the underlying characteristics of online fake news, and the learning models trained in one offline system may not be applied to other circumstances. In the real-time detection system, different real-time analysis techniques are used to determine if the ongoing social information is fake or not. By using predictive analytics methods on the real-time information, real-time analysis can improve the applicability of offline methods, and bring practical significance for online fake news prediction. There are only few studies work on real-time fake news detection, Zhao et al. (2014) builds an real-time visualization system for analyzing anomalous information spreading on *Twitter*, Shao et al. (2016) generates an online platform for tracing real-time misinformation and the related fact-checking records. Establishing an effective online system is challenging, since online communication data is time-sensitive, continuous, and heterogeneous. Despite this, online detection system is a powerful tool for capturing the dynamic nature of online information and fighting with online fake news, and it should obtain more attention in the future.

4.2. Features for fake news detection

In this section, we discuss the important and commonly used features for online fake news detection. As discussed in Section 4.1, auxiliary information such as the news creator analysis and the social context analysis are important resources for identifying the truth level of online news, so the features for suspicious user profiling and anomalous distribution patterns are also introduced in this paper. Based on the components of fake news that discussed in Fig. 2, there are three main types of feature sets: *creator/user-based feature set, news content-based feature set*, and *social context-based feature set*. In each type of feature set, there are different categories of features, as shown in Fig. 4.

4.2.1. Creator/user-based features

Creator/User-based features have been widely used for suspicious online account detection, these features aim to capture the unique characteristics of suspicious user accounts or non-human accounts, and can be categorized as *user profiling features, user credibility features* and *user behavior features*.

- **User profiling features**: User profiling features include the basic user information such as account name, geolocation information, the data of registration of the user, verified or not, has description or not, and so on.
- **User credibility features**: User credibility features record the impact and the credibility of the online account, include the credibility score of the user (Chu et al., 2012), the number of friends and followers of the user, the ratio between the user's friends and followers, the total number of tweets/posts of the user (Castillo et al., 2011).
- **User behavior features**: User behavior features can be considered as part of social context feature set, which we may discuss in the following. User behavior features aim to obtain user behavior pattern for both deceptive users and legitimate users. A typical user behavior feature is the user anomaly score, which computed by the number of the user's interaction in a time window divided by the user's monthly average in Zhao et al. (2014) for online anomalous information detection.

4.2.2. News content-based features

News content-based features are explicit clues for fake news detection, and they are most commonly used attributes for fake news representation and detection analysis. They can be categorized as *linguistic and syntactic-based features, style-based features* and *visual-based features*.

- **Linguistic and Syntactic-based features**
  Linguistic and syntactic-based features refer to the fundamental component, structure and semantics for natural language. Although fake news content are always generated intentionally for misleading online users, linguistic and syntactic-based features

are still valuable sources for suspicious news analysis. And they can be categorized as*word-level features, sentence-level features* and *content-level features*.

- **Word-level features**: bag-of-words, n-gram, term frequency (TF), term frequency-inverted document frequency (TF-IDF) are the most commonly used linguistic features for natural language processing. Also, the presence of special and suspicious tokens in the news content is used in Castillo et al. (2011) for fake information identification. The special and suspicious tokens include exclamation, question mark, multiple exclamation, stoke symbol, user mention, hashtag, emoticon smile, emoticon frown, uppercase token, bold word, the fraction of tweet/post containing the suspicious tokens and so on. Similar to suspicious tokens, the present of stylistic words can also be used for online fake news detection (Horne & Adali, 2017). The stylistic words include the stop-words, punctuations, quotes, negations (such as no, never, not, nope, despite, doubt, bogus, debunk, pranks, retract, deny, fake, fraud, false etc.), informal/swear words, interrogative (how, when, what, why), nouns, personal pronouns, possessive pronouns, determinants, cardinal numbers, adverbs, interjections, verbs, quantifying words, comparison words, exclamation marks, online slang terms (such as *lol, brb*, etc.) and so on. Linguistic Inquiry and Word Count (LIWC) is a transparent text analysis program that counts words in psychologically meaningful categories (Pennebaker, Mehl, & Niederhoffer, 2003). LIWC contains five major categories and various subcategories, like social, affective, cognitive, perceptual. Based on LIWC, researchers can count the number of sentiment words in the news content, which may help to determine the overall sentimental level of the news. The sentiment words include analytic words, insightful words, causal words, discrepancy words, tentative words, certainty words, differentiation words, affiliation words, power words, reward words, risk words, personal concern words (work, leisure, religion, money), emotional tone words, emotion words (anger, sad) and so on Horne and Adali (2017) and Tausczik and Pennebaker (2010). Other word-level linguistic features, such as word readability (the grade level reading score based on the number of syllables in words) and type-token ratio (the number of unique words divided by the total number of words in the documents/posts/tweets) can indicate the lexical simplicity and diversity of the vocabulary in the news, and can also be used for fake news content analysis.
- **Sentence-level features**: Sentence-level features refer to all the important attributes that based on sentence scale, they include parts of speech tagging (POS), the average sentence length (Horne & Adali, 2017), the average length of a tweet/post (Castillo et al., 2011), the frequency of punctuations, function words, and phrase in a sentence (Castillo et al., 2011), the average polarity of the sentence (positive, neutral or negative), the sentence complexity (De Marneffe, MacCartney, Manning et al., 2006), and so on. More specifically, sentence complexity compute each sentence's syntax tree depth, noun phrase syntax tree depth and verb phrase syntax tree depth using the *Stanford Parser*, and is used in Horne and Adali (2017) for online fake news detection.
- **Content-level features**: Content-level features refer to the raw information of the meta news content (Shu et al., 2017). The overall sentiment score of a news content has been verified as a powerful tool for fake news and suspicious author analysis (Dickerson et al., 2014). In the previous studies, *SentiStrength* has been used to identify the intensity of positive and negative emotion in raw document (Horne & Adali, 2017; Thelwall, Buckley, Paltoglou, Cai, & Kappas, 2010). Similarly, stance-based features indicate the supporting or denying attitude of a tweet/post, and can be used to evaluate the emotional posture of the news. Other content-level features, such as the present of bold news title, the news topics (social lives, politics, technology and cybersecurity, business and financial, etc.), the certainty of the news (Castillo et al., 2011), the number of special tags or symbols in the whole news, the present of external links or URLs (Castillo et al., 2011) are also important clues for online fake news identification.

- **Style-based features**: Style-based features aim to reveal the different characteristics of writing styles for fake news authors. Although most of the time, fake news authors try to mimic the writing style of a normal news author to deceive the online readers, there are still some differences which can used to discriminate fake news creators and true news creators. In order to detect the deceptive opinions and reviews, Ahmed (2017) studies the keystroke features for both fake content creators and real online reviewers. With the focus on editing pattern features (such as the number of deletion, "MouseUp", and arrow keystrokes) and timespan features (such as the timespan of the entire document, the average time span of word, the average interval between words), Ahmed (2017) finds that fake content creators need longer time to finish writing, and they tend to make more mistakes. Besides keystrokes features, Castillo et al. (2011) and Zhao et al. (2014) leverage many pattern-based features, which can be used to define the unique content pattern of the fake news. Those pattern-based features include the fraction of tweets/posts that contain external links, user mentions, hashtags during a time window, if the external link uses a popular domain name, and so on.
- **Visual-based features** The images or videos contained in a new content are critical cues for detecting suspicious or deceptive information. Recent studies (Gupta, Lamba, Kumaraguru, & Joshi, 2013; Jin, Cao, Zhang, Zhou, & Tian, 2017) explore visual-based features for online misinformation identification. Those visual-based features include but not limit to the number of image or videos, clarity score, coherence score, similarity distribution histogram, diversity score, clustering scores, image ratio, multi-image ratio, hot image ratio, long image ratio, and so on.

4.2.3. Social context-based features

Social context-based features are designed to reflect the distribution pattern of the online news, and the interaction between online users. And they can be summarized into the following three types: *network-based features, distribution-based features*, and *temporal-based features*.

- **Network-based features**: Network-based analysis intent to focus on a group of similar online users, in term of different perspectives, like location, education background, and habits. And network-based features are selected and extracted based on

specific networks and can be used to study the unique characteristics of certain networks, and the similarity and dissimilarity of different online accounts. Shu et al. (2017) gives a concise summarization of different network-based analysis, such as stance network (Jin et al., 2016; Tacchini et al., 2017), cooccurrence network (Ruchansky et al., 2017), friendship network (Kwon et al., 2013), and diffusion network (Kwon et al., 2013).

- **Distribution-based features**: Distribution-based features can help to capture the distinct diffusion pattern of online news. Usually a propagation tree can be built to facilitate characterizing the distribution nature of a piece of news (Castillo et al., 2011; Watts, Peretti, & Frumin, 2007). And the features related to the propagation tree include the degree of the root in a propagation tree, the maximum number of subtrees, the maximum/average degree and depth of the tree, and so on. In addition, some other features like the number of retweets/reposts for the original tweet/post, the fraction of tweets/posts that are retweeted for an online account, the in-degree/out-degree of an online user's ego net can also be used to assess the impact, popularity and suspicious level of online fake news.
- **Temporal-based features**: Temporal-based features can be used to describe the posting behavior of online news creator in a time series manner. They are good attribute to detect suspicious posting activities, and can be used to indicate the false level of online news. The commonly used temporal-based features include the interval between two post, the frequency of posting, replying and commenting for a certain account, the time of the day when the original information is posted/shared/commented, and the day of the week in which the post is published.

4.3. Evaluation analysis of the current works

In this section, the evaluation results of fake news detection in the existing studies are compared and discussed in terms of both qualitative and quantitative aspects.

Horne and Adali (2017) studies the distinguishing characters between fake news and real news article. They find that *fake news is more closely related to satire news in terms of the complexity and style of the content*. With machine learning algorithm, their model could achieve 91% accuracy for fake news identification. Wu, Li, Hu, and Liu (2017) proposes a framework which uses patterns from prior labeled data to help reveal emergent rumors. They find that *similar rumors usually trigger similar reactions, such as curiosity, inquiry, and anxiety*. And their proposed model can achieve *F*-score as high as 90%. Mitra and Gilbert (2015) introduces a large-scale social media corpus for credibility annotations. Their work comprises more than 60 million tweets grouped into 1049 real-world events. After analyzing by human annotators, *roughly 24% of the events in the global tweet stream are not perceived as credible*. Rubin, Conroy, Chen, and Cornwell (2016) builds a predictive model for online satire detection. Their results support a fact that *the rhetorical component of satire provides reliable cues to its identification*. Their model can reach high accuracy rates with 90% precision and 87% *F*-score. Ma et al. (2016) introduces a deep learning based approach for online rumor detection. They construct two microblog datasets using Twitter and Sina Weibo. They find their *RNN based model can capture the hidden implications and dependencies of online rumor over time*. Kumar et al. (2016) studies the false information in Wikipedia by focusing on hoax articles. Their major findings are: 1) *A small number of false information survive long and are well cited across the Web*; 2) *article authors, structure and content are the influential features for hoax detection*; and 3) *human is not doing a good job for online false information detection*. Conroy et al. (2015) drafts an approach for online fake news detection combining both linguistic analysis and network-based analysis. With a final 91% accuracy, they discovers that *the contextual information describes how online fake news is distributed, and is good indicators for identifying credible sources*. By collecting online misinformation cross online social media and its related fact-checking efforts (e.g., detection results from Snopes), Shao et al. (2016) builds an online platform for online fake news analysis. They note that *the sharing of fact-checking content typically lags that of misinformation by 1–20 h*. Shu, Wang, and Liu (2019b) studies the tri-relationship during news dissemination process on social media, which is the relationship among publishers, news pieces and the users. With publisher-news relations and user-news interactions, their proposed model shows effectiveness for fake news detection with accuracy around 90%. According to their results, since fake news is often intentionally written to mislead users, *other auxiliary information like social context is essential and necessary for online fake news identification*.

4.4. Fake news dataset

The quality of the dataset is one of the fundamental factors for building an effective supervised learning model. In this Section, we propose some important evaluation metrics for assessing the quality of fake news datasets, and compare the available datasets for fake news detection.

- **Benjamin Political News Dataset**[1]: This dataset is created by Horne and Adali (2017) for online political and satire stories detection. The dataset contains 75 stories from the following categories of news: real, fake and satire. The fake news sources are collected from Zimdar's list of fake and misleading news websites (False mislaeding clickabit or satirical news sources), and the real sources are collected from Business Insider's "Most Trusted" list (Business insider most trust news list).
- **Burfoot Satire News Dataset**[2]: This dataset is manually collected by Burfoot and Baldwin (2009), which contains 4000 real news samples and 233 satire news samples. The real news stories are collected using newswire documents samples from the English

[1] https://github.com/rpitrust/fakenewsdata1
[2] http://www.csse.unimelb.edu.au/research/lt/resources/satire/

Gigaword Corpus, and the satire news stories are selected that are closely related in topic to the real ones (Horne & Adali, 2017).

- **BuzzFeed News**[3]: This dataset contains more than 2000 news samples that are published in September 2016 from *Facebook* (Horne & Adali, 2017; Buzzfeednews). All the news sample are verified by professional journalists from BuzzFeed, and has been categorized as mostly true, not factual content, mixture of true and false, and mostly false. Also for each news sample, this dataset provides other relative information such as URL of the news post, published data, number of shares, reactions and comments.
- **Credbank Dataset**[4]: This corpus is a collection of streaming tweets that are tracked between October 2014 and February 2015. The overall dataset comprises more than 60 million tweets with the coverage of 1049 real-word events, and the truthfulness of the tweets are evaluated by 30 annotators (Mitra & Gilbert, 2015).
- **Fake News Challenge dataset**[5]: This dataset provides 50,000 "stance" tuples, each tuple contains the headline of the news, the body of the news and the stance which identifies if the news headline agrees, disagrees, discusses, or unrelated to the news article (Riedel et al., 2017). By checking the consistency between the title and the article, stance-based analysis can also help to detect suspicious false information.
- **FakeNewsNet**[6]: This dataset is provided in Shu et al. (2017), which contains 211 fake news and 211 true news that are collected from both BuzzFeed.com and PolitiFact.com. For each news sample, it also contains the important features such as publisher information, news content, and social engagements information. A more detail analysis for the dataset can be found in Shu, Wang, and Liu (2017).
- **LIAR**[7]: This data is proposed and published in Wang (2017) for online fake news detection. It contains 12,800 manually labeled short statements in various contexts from PolitiFact.com. Each data sample is marked as the following 6 ratings: true, mostly true, half true, barely true, false and pants-fire. Also, detailed analysis report and links to source documents are provided for each case.

In addition to the above datasets, there are still other fake news datasets which are available online and can be used to train the prediction models. For example, Kaggle.com provides a wide range of datasets for fake news prediction and identification. In Ahmed (2017), two fake review datasets are used for deceptive review and opinion analysis, more details can be seen in Ott, Choi, Cardie, and Hancock (2011) and Banerjee et al. (2014). In order to provide an overview of available fake news datasets and compare the important attributes for different datasets, a comparison table is created for further illustration. In Table 3, the aforementioned datasets are compared based on *Scope coverage, Label, Size, Source & Platform, Date of generation*, and *Feature coverage*. We can see that the available datasets are different in terms of different evaluation metrics. But for most datasets, they only contain the news content information. This is challenging for building an effective detection model, due to the existing linguistic features and patterns may not reflect the underlying characteristics of the data from real world. Also, other important attributes such as the size and the positive/negative distribution should also be considered when selecting a suitable fake news dataset for machine learning based assessments. A high-quality dataset plays an extremely important role in the task of supervised based learning. However, the lack of labeled fake news dataset is the bottleneck for building an effective detection system for online misleading information (Wang, 2017).

5. Open issues and future work

In this section, some challenges and open issues for automatic online fake news detection are discussed, along with some promising research directions in this area. Finally, we present how to build an effective online fake news detection ecosystem.

5.1. Unsupervised learning for fake news analysis

As we mentioned before, the limited accessibility of high-quality labeled dataset is one of the major challenges for online fake news detection. And unsupervised learning methods can be applied for practical analysis of the real world dataset. In this paper, we propose three types of unsupervised learning model for fake news detection, they are: *cluster analysis, outlier analysis, semantic similarity analysis*, and *unsupervised news embedding*.

- **Cluster analysis**: Cluster analysis studies data objects without consulting labels. In cluster analysis, the data can be grouped based on the principle of maximizing the intraclass similarity and minimizing the interclass similarity, and it can generate class labels for a group of data (Han et al., 2011). In fake news detection, cluster analysis can be used to identify the homogeneous of individual groups of news and authors.
- **Outlier analysis**: Outlier analysis is the study of detecting the abnormal behavior of objects (Rousseeuw & Leroy, 2005). By learning the statistical distribution of the unlabeled online social data, outlier analysis can uncover fraudulent information and suspicious authors based on statistical measures, distance measures and density-based methods (Hodge & Austin, 2004).
- **Semantic similarity analysis**: Semantic similarity analysis is used to detect near-duplicated news content (Ahmed, 2017). Due to the lack of relative knowledge and imagination, online fake news creators usually reuse the existing news content (Li, McLean,

[3] https://github.com/BuzzFeedNews/2016-10-facebook-fact-check/tree/master/data
[4] http://compsocial.github.io/CREDBANK-data/
[5] https://github.com/FakeNewsChallenge/fnc-1
[6] https://github.com/KaiDMML/FakeNewsNet
[7] https://www.cs.ucsb.edu/~william/data/liar_dataset.zip

Table 3
The comparison of different fake news datasets.

Dataset	Scope coverage	Label (Count)	Size	Sources & platform	Date of generation	Feature Coverage		
						User Features	Content Features	Context Features
Benjamin Political News Dataset	Political news	Real (75) Fake (75) Satire (75)	225	Wall Street Journal The Economist Washington Post Ending the Fed True Pundit Infowars The Onion Borowitz Report Satire Wire and so on	2017	No	Yes	No
Burfoot Satire News Dataset	Satirical news	Real (4000) Satire (233)	4233	English Gigaword Corpus	2009	No	Yes	No
BuzzFeed News	Social posts and linked articles	Most true (1669) No factual content (264) Mixture of true and false (245) Mostly false (104)	2282	Facebook	2016	Yes	Yes	Yes
Credbank Dataset	Social posts	Human credibility judgements	> 60 million	Twitter	2015	Yes	Yes	Yes
Fake News Challenge Dataset	Fake news	Unrelated (36545) Discuss (8910) Agree (3678) Disagree (839)	49,972	Fake News Challenge	2017	No	Yes	No
FakeNewsNet	Fake news	Fake (211) True (211)	422	BuzzFeed.com PolitiFact.com	2017	Yes	Yes	Yes
LIAR	Fake news	True (2009) Mostly true (2457) Half true (2729) Barely true (2099) False (2502) Pants fire (1040)	12,836	PolitiFact.com	2017	No	Yes	No

Bandar, O'shea, & Crockett, 2006). For example, an online fake reviewer can change only a few words of an online review to mislead customers. So semantic similarity analysis provide a good way for detecting the copied or semi-copied news content manipulated by unfriendly authors, and can be used for potential fake news detection.

- **Unsupervised news embedding**: As aforementioned, due to the textual nature of the online news, semantic similarity analysis, sentiment analysis and other related tasks are important components for online fake news detection. Embedding is an essential step in natural language processing, it refers to a process of extracting distributed representations of raw textual data. In fake news detection, the numeric representations can then be used as input for further analysis. Different embedding technologies can capture the characteristics of data in different perspectives. How to choose a good embedding method plays a significant role in obtaining the underlying nature of the news, and thus in the successful detection of online false information. Some popular unsupervised-based embedding techniques include Word2vec (Mikolov, Sutskever, Chen, Corrado, & Dean, 2013), FastText (Bojanowski, Grave, Joulin, & Mikolov, 2017), Sent2vec (Pagliardini, Gupta, & Jaggi, 2017), and Doc2vec (Le & Mikolov, 2014).

5.2. One-class classification for fake news analysis

In addition to unsupervised learning, one-class classification algorithms, which can generate abstract scores and latent variables, are also potential solutions for dealing with the unlabeled real-world data. In view of the fact that there are far more true news stories than false or satirical news on the Internet (Burfoot & Baldwin, 2009; Chandola, Banerjee, & Kumar, 2009), the real-world social media dataset is a good resource for training one-class classification algorithms. For example, by measuring how the information spreading pattern is different from a set of unlabeled training examples with an anomaly score, one-class conditional random fields (OCCRF) model is applied to *Twitter* data in Zhao et al. (2014) for anomalous information analysis. Other classic one-class classification algorithms include one-class support vector machines (OSVMs) (Khan & Madden, 2009) and Non-OSVMs methods (such as one-class neural network (Chalapathy, Krishna Menon, & Chawla, 2018), one-class k-nearest neighbor (Khan & Madden, 2014), one-class random forests (Désir, Bernard, Petitjean, & Heutte, 2013)).

5.3. Real-time visualization for fake news detection

Based on the difficulty and complexity of online fake news detection, a binary classification model is far from enough to discern the characteristics of online anomalous information. Due to the real-time and heterogeneous nature of social communication data, data visualization is a powerful tool for illustrating different aspects and distribution patterns of online social information. An interactive visualization system can provide diverse dimensions and views of the data, facilitate human supervision and understanding, reveal temporal-based patterns and behaviors of the data, and summarize important features in a more clear way. Social media distribution and visualization has been extensively studied in Nishi et al. (2016), Kwak et al. (2010), Bakshy, Hofman, Mason, and Watts (2011), Cogan et al. (2012), Cha, Mislove, and Gummadi (2009), Gómez, Kappen, Litvak, and Kaltenbrunner (2013) and Wang et al. (2011). However, apart from Zhao et al. (2014), few studies focus on the interactive exploration of visualizing false information on social media. Such visualization platforms may indicate the nature of the information diffusion, and can be considered as informative resources about the relationships between online users (Nishi et al., 2016). With an online real-time detection system, false information or user abnormal behaviors can be detected the moment it happens. Then proper actions can be taken to limit the negative impact of online fake news. The online system is a real-time safeguard that attempts to protect online readers instantly. Also, with real-time system, Internet users or cyber experts can stay one step ahead of the fully distribution of online false information, which can mitigate the effects of such information attacks. Often combined with unsupervised learning algorithms and visualization techniques, a real-time system can keep up with the new trends of fake news, and can be applied in more scenarios. As a result, real-time visualization systems are important components for online fake news detection and monitoring, and are worth exploring areas.

However, the massive data size, the high dimensionality of real-time data, and the heterogeneous nature of online data streaming pose unique challenges for building a real-time detection system, in terms of data storage and data computation. And all the challenges should be addressed to establish an effective and efficient online detection system.

5.4. Early prediction and intervention for online fake news

Online fact-checking resources can only detect fake news after the misleading information is created and spread through the Internet. Fact-checking websites can warn online users against similar claims or topics, but they can not fully stop the propagation of misinformation in online social networks. Beside the detection system, there are two promising research aspects that are also very important for fighting with fake news, they are (1) *fake news early prediction* and (2) *fake news intervention*.

Currently most of the studies attempt to assess whether the online communication information is true or false, however it is extremely significant to identify any trending or potential false news as early as possible. By learning from historical data, fake news early prediction aims to detect the newly emerging fake news or rumors even before they occur. With the information of user polarization and confirmation bias, Del Vicario, Quattrociocchi, Scala, and Zollo (2018) is able to identify topics that are susceptible to misinformation. Zhao, Resnick, and Mei (2015) and Wu et al. (2017) study the problem of rumor early detection in social media. And more efforts can be made in fake news early detection, in terms of suspicious news author/platform/data source analysis, potential fake news topic analysis, potential time peak analysis, and so on.

Fake news early prediction can remind online users of any potential false information before the fake news exists. Fake news intervention can help online users erase the negative impacts of fake news after the fake news happens. By combining reinforcement

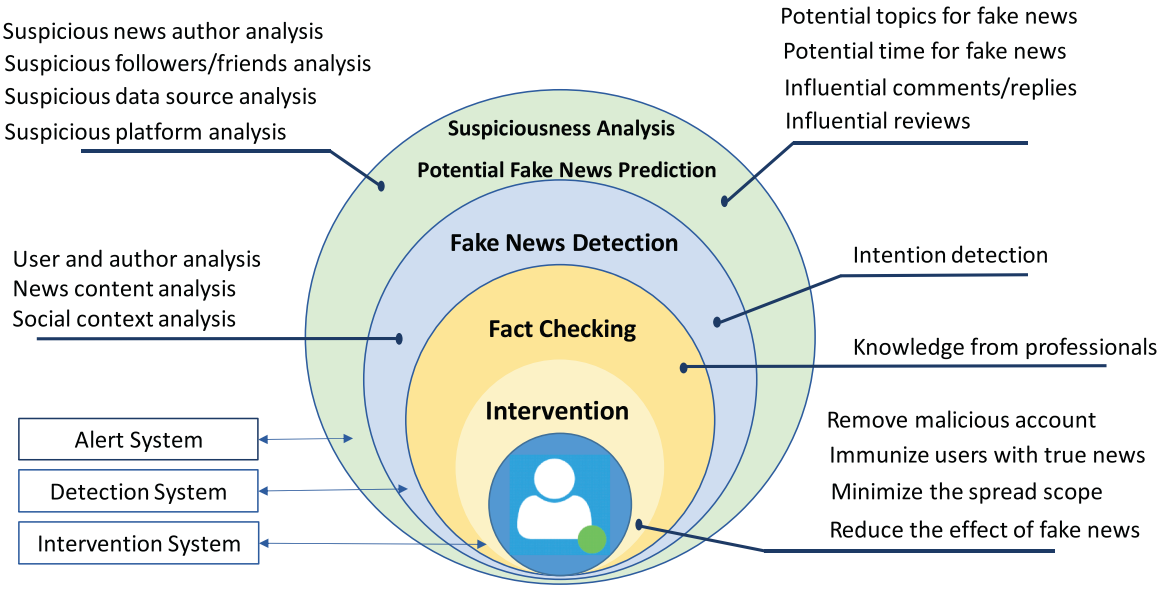

Fig. 5. A comprehensive fake news detection ecosystem.

learning with a point process network model, Farajtabar et al. (2017) mitigates the effect of fake news in social media. Shu et al. (2017) mentions some other ways for minimizing the spread of online fake news, such as removing suspicious online accounts, and immunizing users with true news. In conclusion, fake news intervention is a potential research direction for mitigating the physical spreading and the psychological impact of online fake news.

5.5. Evaluation for fake news detection system

In this section, based on previous knowledge, we mainly discuss the evaluation metrics for building an effective detection system for online fake news, and propose a comprehensive fake news detection ecosystem for further study. We present the following attributes for evaluating the performance of an online fake news detection.

- **Accurate Detection**: The bottom line of any detection system is always the accurate detection output. Due to the difficulty of online fake news identification, various types of features and methods should be used in order to improve the effectiveness of the system.
- **Interactive Visualization**: Visualization is the fundamental component of an online fake news monitoring system. It can facilitate human understanding and bring more perspectives for describing the time-sensitive data.
- **Early Warning and Post Intervention**: As we discussed in Section 5.4, early prediction and post intervention are promising research directions for online fake news identification, and are good factors by assessing the integrity of an fake news detection system.
- **Third Party Verification**: Third parties may also be involved in outcome verification, which can make a fake news detection system reliable and credible.

Finally we propose a comprehensive fake news detection ecosystem, see in Fig. 5. We believe that an effective fake news detection ecosystem should contain *Alert system, Detection system*, and *Intervention system*. The combination of these three systems can provide various types of analysis, alerts and detection, which are strong protection for social accounts against the in-depth impact of online fake news.

6. Conclusion

Recently, fake news is emerging as one of the most threatening harms on social media. Fake news can be used by malicious entities to manipulate people's options and decisions on important daily activities, like stock markets, health-care options, online shopping, education, and even presidential election. Automatic detection of online fake news is an extremely significant but challenging task for both industry and academia (Ruchansky et al., 2017). In this survey, we present a comprehensive overview of online fake news detection. And the key contributions of this paper can be summarized as follows. (1) We discuss the in-depth understanding of the important aspects of online fake news, such as the news creator/spreader, news targets, news content and social context. The clear characterization of online fake news can play a significant role in social communication data analysis and anomalous information detection. (2) By comparing the existing detection approaches, listing an exhaustive set of hand-crafted features, and

evaluating the existing datasets for training supervised models, we provide a fundamental review for fake news detection. As a detailed guideline, our survey can bring valuable knowledge and practical convenient for both researchers and participators. (3) Some potential research focuses are proposed in order to address the open issues, improve the existing detection frameworks, and establish an effective online fake news monitoring and detection system.

Acknowledgement

The authors generously acknowledge the funding from the Atlantic Canada Opportunity Agency (ACOA) through the Atlantic Innovation Fund (AIF) Project #201212 and through Discover Grant and Tier 1 Canada Research Chair Funding grant from the National Science and Engineering Research Council of Canada (NSERC 232074) to Dr. Ghorbani.

References

Adams, B., Phung, D., & Venkatesh, S. (2011). *Eventscapes: visualizing events over time with emotive facets. Proceedings of the 19th ACM international conference on multimedia*. ACM1477–1480.
Afroz, S., Brennan, M., & Greenstadt, R. (2012). *Detecting hoaxes, frauds, and deception in writing style online. Security and privacy (sp), 2012 IEEE symposium on*. IEEE461–475.
Agarwal, A., Xie, B., Vovsha, I., Rambow, O., & Passonneau, R. (2011). *Sentiment analysis of twitter data. Proceedings of the workshop on languages in social media*. Association for Computational Linguistics30–38.
Ahmed, H. (2017). *Detecting opinion spam and fake news using n-gram analysis and semantic similarity* Ph.D. thesis.
Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of Economic Perspectives, 31*(2), 211–236.
Bakshy, E., Hofman, J. M., Mason, W. A., & Watts, D. J. (2011). *Everyone's an influencer: Quantifying influence on twitter. Proceedings of the fourth ACM international conference on web search and data mining*. ACM65–74.
Balmas, M. (2014). When fake news becomes real: Combined exposure to multiple news sources and political attitudes of inefficacy, alienation, and cynicism. *Communication Research, 41*(3), 430–454.
Banerjee, R., Feng, S., Kang, J. S., & Choi, Y. (2014). *Keystroke patterns as prosody in digital writings: A case study with deceptive reviews and essays. Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*1469–1473.
Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2017). Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics, 5*, 135–146.
Bordes, A., Chopra, S., Weston, J. (2014). Question answering with subgraph embeddings. ArXiv e-prints.
Burfoot, C., & Baldwin, T. (2009). *Automatic satire detection: Are you having a laugh? Proceedings of the ACL-IJCNLP 2009 conference short papers*. Association for Computational Linguistics161–164.
Business insider most trust news list. http://www.businessinsider.com/here-are-the%20-most-and-least-trusted-news-outlets-in-america-2014-10. Accessed: 2018-04-19.
Buzzfeednews. https://github.com/BuzzFeedNews/2016-10-facebook-fact-check/blob/master/data/facebook-fact-check.csv. Accessed: 2018-04-19.
Cao, N., Shi, C., Lin, S., Lu, J., Lin, Y.-R., & Lin, C.-Y. (2016). Targetvue: Visual analysis of anomalous user behaviors in online communication systems. *IEEE Transactions on Visualization and Computer Graphics, 22*(1), 280–289.
Castillo, C., Mendoza, M., & Poblete, B. (2011). *Information credibility on twitter. Proceedings of the 20th international conference on world wide web*. ACM675–684.
Castillo, C., Mendoza, M., & Poblete, B. (2013). Predicting information credibility in time-sensitive social media. *Internet Research, 23*(5), 560–588.
Cbc news. http://www.cbc.ca/news/canada/toronto/scarborough-hijab-attack-1.4487716. Accessed: 2018-03-21.
Cha, M., Mislove, A., & Gummadi, K. P. (2009). *A measurement-driven analysis of information propagation in the flickr social network. Proceedings of the 18th international conference on world wide web*. ACM721–730.
Chalapathy, R., Krishna Menon, A., & Chawla, S. (2018). Anomaly detection using one-class neural networks. ArXiv e-prints,.
Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. *ACM Computing Surveys (CSUR), 41*(3), 15.
Chen, Y., Conroy, N. J., & Rubin, V. L. (2015). *Misleading online content: Recognizing clickbait as false news. Proceedings of the 2015 ACM on workshop on multimodal deception detection*. ACM15–19.
Chu, Z., Gianvecchio, S., Wang, H., & Jajodia, S. (2012). Detecting automation of twitter accounts: Are you a human, bot, or cyborg? *IEEE Transactions on Dependable and Secure Computing, 9*(6), 811–824.
Classify.news. https://makenewscredibleagain.github.io/. Accessed: 2018-01-25.
Cogan, P., Andrews, M., Bradonjic, M., Kennedy, W. S., Sala, A., & Tucci, G. (2012). *Reconstruction and analysis of twitter conversation graphs. Proceedings of the first ACM international workshop on hot topics on interdisciplinary social networks research*. ACM25–31.
Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K., & Kuksa, P. (2011). Natural language processing (almost) from scratch. *Journal of Machine Learning Research, 12*(Aug), 2493–2537.
Conroy, N. J., Rubin, V. L., & Chen, Y. (2015). Automatic deception detection: Methods for finding fake news. *Proceedings of the Association for Information Science and Technology, 52*(1), 1–4.
Cui, W., Liu, S., Tan, L., Shi, C., Song, Y., Gao, Z., et al. (2011). Textflow: Towards better understanding of evolving topics in text. *IEEE Transactions on Visualization and Computer Graphics, 17*(12), 2412–2421.
Daily, C.. http://global.chinadaily.com.cn/a/201801/10/WS5a55a685a3102e5b17371dd1_2.html. Accessed: 2019-02-09.
Dale, R. (2017). Nlp in a post-truth world. *Natural Language Engineering, 23*(2), 319–324.
Davis, C. A., Varol, O., Ferrara, E., Flammini, A., & Menczer, F. (2016). *Botornot: A system to evaluate social bots. Proceedings of the 25th international conference companion on world wide web*. International World Wide Web Conferences Steering Committee273–274.
De Marneffe, M.-C., MacCartney, B., Manning, C. D., et al. (2006). *Generating typed dependency parses from phrase structure parses. Proceedings of lrec6. Proceedings of lrec* Genoa Italy449–454.
Del Vicario, M., Quattrociocchi, W., Scala, A., & Zollo, F. (2018). Polarization and fake news: Early warning of potential misinformation targets. ArXiv e-prints,.
Della Vedova, M. L., Tacchini, E., Moret, S., Ballarin, G., DiPierro, M., & de Alfaro, L. (2018). *Automatic online fake news detection combining content and social signals. 2018 22nd conference of open innovations association (fruct)*. IEEE272–279.
Dematis, I., Karapistoli, E., & Vakali, A. (2018). *Fake review detection via exploitation of spam indicators and reviewer behavior characteristics. International conference on current trends in theory and practice of informatics*. Springer581–595.
Denver guardian. https://en.wikipedia.org/wiki/Denver_Guardian. Accessed: 2018-01-25.
Désir, C., Bernard, S., Petitjean, C., & Heutte, L. (2013). One class random forests. *Pattern Recognition, 46*(12), 3490–3506.
Detecting fake news with nlp. https://medium.com/@Genyunus/detecting-fake-news-with-nlp-c893ec31dee8. Accessed: 2018-03-15.
Devitt, A., & Ahmad, K. (2007). *Sentiment polarity identification in financial news: A cohesion-based approach. Proceedings of the 45th annual meeting of the association of computational linguistics*984–991.
Di, R., Wang, H., Fang, Y., & Zhou, Y. (2018). *Fake comment detection based on time series and density peaks clustering. International conference on algorithms and architectures for parallel processing*. Springer124–130.

Diakopoulos, N., Naaman, M., & Kivran-Swaine, F. (2010). *Diamonds in the rough: Social media visual analytics for journalistic inquiry. Visual analytics science and technology (vast), 2010 IEEE symposium on*. IEEE115–122.
Dickerson, J. P., Kagan, V., & Subrahmanian, V. (2014). *Using sentiment to detect bots on twitter: Are humans more opinionated than bots? Advances in social networks analysis and mining (asonam), 2014 IEEE/ACM international conference on*. IEEE620–627.
Dodds, P. S., Clark, E. M., Desu, S., Frank, M. R., Reagan, A. J., Williams, J. R., et al. (2015). Human language reveals a universal positivity bias. *Proceedings of the National Academy of Sciences, 112*(8), 2389–2394.
Dong, M., Yao, L., Wang, X., Benatallah, B., Huang, C., & Ning, X. (2018a). Opinion fraud detection via neural autoencoder decision forest. arXiv:1805.03379.
Dong, M., Yao, L., Wang, X., Benatallah, B., Sheng, Q. Z., & Huang, H. (Yao, Wang, Benatallah, Sheng, Huang, 2018b). *Dual: A deep unified attention model with latent relation representations for fake news detection. International conference on web information systems engineering*. Springer199–209.
Factcheck. https://www.factcheck.org/about/our-mission/. Accessed: 2018-01-24.
Factmata. https://medium.com/factmata/introducing-factmata-artificial-intelligence-for-political-fact-checking-db8acdbf4cf1. Accessed: 2018-01-25.
Fake news. https://en.wikipedia.org/wiki/Fake_news. Accessed: 2018-03-16.
Fakespot. https://www.fakespot.com. Accessed: 2018-02-09.
False mislaeding clickabit or satirical news sources. https://docs.google.com/document/d/10eA5-mCZLSS4MQY5QGb5ewC3VAL6pLkT53V_81ZyitM/preview. Accessed: 2018-04-19.
Farajtabar, M., Yang, J., Ye, X., Xu, H., Trivedi, R., Khalil, E., et al. (2017). Fake news mitigation via point process based intervention. ArXiv e-prints,.
Feng, V. W., & Hirst, G. (2013). *Detecting deceptive opinions with profile compatibility. Proceedings of the sixth international joint conference on natural language processing*338–346.
Ferrara, E., Varol, O., Davis, C., Menczer, F., & Flammini, A. (2016). The rise of social bots. *Communications of the ACM, 59*(7), 96–104.
Fichera, A.. https://www.factcheck.org/2019/01/story-cherry-picks-in-assessing-cnn-ratings/. Accessed: 2019-02-09.
Forbes.com Americans believe they can detect fake news. Studies show they can't.https://www.forbes.com/sites/brettedkins/2016/12/20/americans-believe-they-can-detect-fake-news-studies-show-they-cant/#1f4c85cf4022. Accessed: 2018-04-01.
Ghosh, R., Surachawala, T., & Lerman, K. (2011). Entropy-based classification of 'retweeting' activity on twitter. ArXiv e-prints,.
Gianvecchio, S., Wu, Z., Xie, M., & Wang, H. (2009). *Battle of botcraft: Fighting bots in online games with human observational proofs. Proceedings of the 16th acm conference on computer and communications security*. ACM256–268.
Gianvecchio, S., Xie, M., Wu, Z., & Wang, H. (2008). *Measurement and classification of humans and bots in internet chat. Usenix security symposium*155–170.
Glorot, X., Bordes, A., & Bengio, Y. (2011). *Domain adaptation for large-scale sentiment classification: A deep learning approach. Proceedings of the 28th international conference on machine learning (icml-11)*513–520.
Goldberg, Y., & Levy, O. (2014). word2vec explained: Deriving Mikolov et al.'s negative-sampling word-embedding method. ArXiv e-prints,.
Gómez, V., Kappen, H. J., Litvak, N., & Kaltenbrunner, A. (2013). A likelihood-based framework for the analysis of discussion threads. *World Wide Web, 16*(5–6), 645–675.
Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y. (2016). *Deep learning. 1*. MIT press Cambridge.
Google Announcement. Google anounces google news initiative to help quality journalism in digital age.http://canoe.com/technology/google-announces-google-news-initiative-to-help-quality-journalism-in-digital-age. Accessed: 2018-03-21.
Google News Initiative. https://newsinitiative.withgoogle.com/. Accessed: 2018-03-21.
Gupta, A., Lamba, H., Kumaraguru, P., & Joshi, A. (2013). *Faking sandy: Characterizing and identifying fake images on twitter during hurricane sandy. Proceedings of the 22nd international conference on world wide web*. ACM729–736.
Han, J., Pei, J., & Kamber, M. (2011). *Data mining: Concepts and techniques.* Elsevier.
Ho, C.-T., Li, C.-T., & Lin, S.-D. (2011). *Modeling and visualizing information propagation in a micro-blogging platform. Advances in social networks analysis and mining (asonam), 2011 international conference on*. IEEE328–335.
Hoaxslayer.com. https://www.technorms.com/454/get-your-facts-right-6-fact-checking-websites-that-help-you-know-the-truth. Accessed: 2018-01-25.
Hodge, V., & Austin, J. (2004). A survey of outlier detection methodologies. *Artificial Intelligence Review, 22*(2), 85–126.
Horne, B. D., & 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 e-prints,.
Jean, S., Cho, K., Memisevic, R., & Bengio, Y. (2014). On using very large target vocabulary for neural machine translation. arXiv:1412.2007.
Jin, F., Dougherty, E., Saraf, P., Cao, Y., & Ramakrishnan, N. (2013). *Epidemiological modeling of news and rumors on twitter. Proceedings of the 7th workshop on social network mining and analysis*. ACM8.
Jin, Z., Cao, J., Zhang, Y., & Luo, J. (2016). *News verification by exploiting conflicting social viewpoints in microblogs. AAAI*2972–2978.
Jin, Z., Cao, J., Zhang, Y., Zhou, J., & Tian, Q. (2017). Novel visual and statistical image features for microblogs news verification. *IEEE Transactions on Multimedia, 19*(3), 598–608.
Karumanchi, A., Fu, L., & Deng, J. (2018). *Prediction of review sentiment and detection of fake reviews in social media. Proceedings of the international conference on information and knowledge engineering (ike)*. The Steering Committee of The World Congress in Computer Science, Computer181–186.
Khan, S. S., & Madden, M. G. (2009). *A survey of recent trends in one class classification. Irish conference on artificial intelligence and cognitive science*. Springer188–197.
Khan, S. S., & Madden, M. G. (2014). One-class classification: Taxonomy of study and review of techniques. *The Knowledge Engineering Review, 29*(3), 345–374.
Kim, J., Tabibian, B., Oh, A., Schölkopf, B., & Gomez-Rodriguez, M. (2018). *Leveraging the crowd to detect and reduce the spread of fake news and misinformation. Proceedings of the eleventh acm international conference on web search and data mining*. ACM324–332.
Kumar, S., & Shah, N. (2018). False information on web and social media: A survey. arXiv:1804.08559.
Kumar, S., West, R., & Leskovec, J. (2016). *Disinformation on the web: Impact, characteristics, and detection of wikipedia hoaxes. Proceedings of the 25th international conference on world wide web*. International World Wide Web Conferences Steering Committee591–602.
Kwak, H., Lee, C., Park, H., & Moon, S. (2010). *What is twitter, a social network or a news media? Proceedings of the 19th international conference on world wide web*. ACM591–600.
Kwon, S., Cha, M., Jung, K., Chen, W., & Wang, Y. (2013). *Prominent features of rumor propagation in online social media. Data mining (icdm), 2013 ieee 13th international conference on*. IEEE1103–1108.
Le, Q., & Mikolov, T. (2014). *Distributed representations of sentences and documents. International conference on machine learning*1188–1196.
LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*(7553), 436.
Li, Y., McLean, D., Bandar, Z. A., O'shea, J. D., & Crockett, K. (2006). Sentence similarity based on semantic nets and corpus statistics. *IEEE Transactions on Knowledge and Data Engineering, 18*(8), 1138–1150.
Ma, J., Gao, W., Mitra, P., Kwon, S., Jansen, B. J., Wong, K.-F., et al. (2016). *Detecting rumors from microblogs with recurrent neural networks. Ijcai*3818–3824.
Ma, J., Gao, W., Wei, Z., Lu, Y., & Wong, K.-F. (2015). *Detect rumors using time series of social context information on microblogging websites. Proceedings of the 24th ACM international on conference on information and knowledge management*. ACM1751–1754.
Mark zuckerberg. https://www.facebook.com/zuck/posts/10103269806149061. Accessed: 2018-03-15.
Markines, B., Cattuto, C., & Menczer, F. (2009). *Social spam detection. Proceedings of the 5th international workshop on adversarial information retrieval on the web*. ACM41–48.
Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). *Distributed representations of words and phrases and their compositionality. Advances in neural information processing systems*3111–3119.
Mislove, A., Marcon, M., Gummadi, K. P., Druschel, P., & Bhattacharjee, B. (2007). *Measurement and analysis of online social networks. Proceedings of the 7th acm sigcomm conference on internet measurement*. ACM29–42.
Mitra, T., & Gilbert, E. (2015). *Credbank: A large-scale social media corpus with associated credibility annotations. ICWSM*258–267.
Mukherjee, A., Liu, B., & Glance, N. (2012). *Spotting fake reviewer groups in consumer reviews. Proceedings of the 21st international conference on world wide web*.

ACM191–200.
Murthy, D., Powell, A. B., Tinati, R., Anstead, N., Carr, L., Halford, S. J., et al. (2016). Automation, algorithms, and politics| bots and political influence: A socio-technical investigation of social network capital. *International Journal of Communication, 10*, 20.
Narayan, R., Rout, J. K., & Jena, S. K. (2018). *Review spam detection using opinion mining. Progress in intelligent computing techniques: Theory, practice, and applications*. Springer273–279.
News Feed fyi. Addressing hoaxes and fake newshttps://newsroom.fb.com/news/2016/12/news-feed-fyi-addressing-hoaxes-and-fake-news/. Accessed: 2018-03-15.
News use across social media platforms 2016. http://www.journalism.org/2016/05/26/news-use-across-social-media-platforms-2016/. Accessed: 2018-03-14.
News use across social media platforms 2016. http://www.journalism.org/2016/05/26/news-use-across-social-media-platforms-2016/. Accessed: 2018-02-09.
Nishi, R., Takaguchi, T., Oka, K., Maehara, T., Toyoda, M., Kawarabayashi, K.-i., et al. (2016). Reply trees in twitter: Data analysis and branching process models. *Social Network Analysis and Mining, 6*(1), 26.
Olteanu, A., Kıcıman, E., & Castillo, C. (2018). *A critical review of online social data: Biases, methodological pitfalls, and ethical boundaries. Proceedings of the eleventh ACM international conference on web search and data mining*. ACM785–786.
Opensecrets. https://www.opensecrets.org/about/. Accessed: 2018-01-25.
Opensource. http://www.opensources.co/. Accessed: 2018-01-25.
Ott, M., Choi, Y., Cardie, C., & Hancock, J. T. (2011). *Finding deceptive opinion spam by any stretch of the imagination. Proceedings of the 49th annual meeting of the association for computational linguistics: Human language technologies-volume 1*. Association for Computational Linguistics309–319.
Pagliardini, M., Gupta, P., & Jaggi, M. (2017). Unsupervised learning of sentence embeddings using compositional n-gram features. arXiv:1703.02507.
Pennebaker, J. W., Mehl, M. R., & Niederhoffer, K. G. (2003). Psychological aspects of natural language use: Our words, our selves. *Annual Review of Psychology, 54*(1), 547–577.
Pérez-Rosas, V., Kleinberg, B., Lefevre, A., & Mihalcea, R. (2017). Automatic detection of fake news. arXiv:1708.07104.
Politifact. https://politifactmediakit.hotims.com/. Accessed: 2018-01-25.
Popel, M., & Žabokrtskỳ, Z. (2010). *Tectomt: Modular nlp framework. International conference on natural language processing*. Springer293–304.
Raskin, V. (1987). *Linguistics and natural language processing. Machine translation: Theoretical and methodological issues.*. Cambridge University Press, Cambridge42–58.
Ratkiewicz, J., Conover, M., Meiss, M. R., Gonçalves, B., Flammini, A., & Menczer, F. (2011). Detecting and tracking political abuse in social media. *ICWSM, 11*, 297–304.
Reviewmeta. https://reviewmeta.com/blog/how-it-works/. Accessed: 2018-02-09.
Riedel, B., Augenstein, I., Spithourakis, G. P., & Riedel, S. (2017). A simple but tough-to-beat baseline for the fake news challenge stance detection task. ArXiv e-prints,.
Romero, D. M., Meeder, B., & Kleinberg, J. (2011). *Differences in the mechanics of information diffusion across topics: Idioms, political hashtags, and complex contagion on twitter. Proceedings of the 20th international conference on world wide web*. ACM695–704.
Rousseeuw, P. J., & Leroy, A. M. (2005). *Robust regression and outlier detection. 589*. John wiley & sons.
Rubin, V., Conroy, N., Chen, Y., & Cornwell, S. (2016). *Fake news or truth? Using satirical cues to detect potentially misleading news. Proceedings of the second workshop on computational approaches to deception detection*7–17.
Ruchansky, N., Seo, S., & Liu, Y. (2017). *Csi: A hybrid deep model for fake news detection. Proceedings of the 2017 ACM on conference on information and knowledge management*. ACM797–806.
Russell, M. A. (2013). *Mining the social web: Data mining facebook, twitter, linkedin, google+, github, and more*. "O'Reilly Media, Inc.".
Russia used fake news. This is how Russia used fake news on facebook to help elect donald trumphttps://www.huffingtonpost.com/entry/this-is-how-russia-used-fake-news-on-facebook-to-help_us_59b60b64e4b0bef3378ce1be. Accessed: 2018-02-09.
Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural networks, 61*, 85–117.
Shao, C., Ciampaglia, G. L., Flammini, A., & Menczer, F. (2016). *Hoaxy: A platform for tracking online misinformation. Proceedings of the 25th international conference companion on world wide web*. International World Wide Web Conferences Steering Committee745–750.
Shao, C., Ciampaglia, G. L., Varol, O., Flammini, A., & Menczer, F. (2017). The spread of fake news by social bots. arXiv:1707.07592 (pp. 96–104).
Shu, K., Bernard, H. R., & Liu, H. (Bernard, Liu, 2019a). *Studying fake news via network analysis: Detection and mitigation. Emerging research challenges and opportunities in computational social network analysis and mining*. Springer43–65.
Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake news detection on social media: A data mining perspective. *ACM SIGKDD Explorations Newsletter, 19*(1), 22–36.
Shu, K., Wang, S., & Liu, H. (2017). Exploiting tri-relationship for fake news detection. ArXiv e-prints,.
Shu, K., Wang, S., & Liu, H. (Wang, Liu, 2019b). *Beyond news contents: The role of social context for fake news detection. Proceedings of the twelfth ACM international conference on web search and data mining*. ACM312–320.
Shu, K., Wang, S., Tang, J., Zafarani, R., & Liu, H. (2017). User identity linkage across online social networks: A review. *ACM SIGKDD Explorations Newsletter, 18*(2), 5–17.
Singhania, S., Fernandez, N., & Rao, S. (2017). *3han: A deep neural network for fake news detection. International conference on neural information processing*. Springer572–581.
Snopes.com. https://en.wikipedia.org/wiki/Snopes.com. Accessed: 2018-01-24.
Song, Y., Wen, Z., Lin, C.-Y., & Davis, R. (2013). *One-class conditional random fields for sequential anomaly detection. IJCAI*1685–1691.
Sundermeyer, M., Schlüter, R., & Ney, H. (2010). Recurrent neural network based language model. *10*, 1045–1048.
Sutskever, I., Vinyals, O., & Le, Q. V. (2014). *Sequence to sequence learning with neural networks. Advances in neural information processing systems*3104–3112.
Tacchini, E., Ballarin, G., Della Vedova, M. L., Moret, S., & de Alfaro, L. (2017). Some like it hoax: Automated fake news detection in social networks. arXiv:1704.07506.
Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: Liwc and computerized text analysis methods. *Journal of Language and Social Psychology, 29*(1), 24–54.
The fake americans russia created to influence the election. https://www.nytimes.com/2017/09/07/us/politics/russia-facebook-twitter-election.html. Accessed: 2018-02-09.
The negative impact of fake news. http://www.straitstimes.com/opinion/the-negative-impact-of-fake-news. Accessed: 2018-02-10.
Thelwall, M., Buckley, K., Paltoglou, G., Cai, D., & Kappas, A. (2010). Sentiment strength detection in short informal text. *Journal of the Association for Information Science and Technology, 61*(12), 2544–2558.
Truthorfiction. https://www.truthorfiction.com/about-us/. Accessed: 2018-01-25.
Tschiatschek, S., Singla, A., Rodriguez, M. G., Merchant, A., & Krause, A. (2017). Detecting fake news in social networks via crowdsourcing. arXiv:1711.09025.
Uscinski, J. E., Klofstad, C., & Atkinson, M. D. (2016). What drives conspiratorial beliefs? The role of informational cues and predispositions. *Political Research Quarterly, 69*(1), 57–71.
Vlachos, A., & Riedel, S. (2014). *Fact checking: Task definition and dataset construction. Proceedings of the ACL 2014 workshop on language technologies and computational social science*18–22.
Wang, D., Wen, Z., Tong, H., Lin, C.-Y., Song, C., & Barabási, A.-L. (2011). *Information spreading in context. Proceedings of the 20th international conference on world wide web*. ACM735–744.
Wang, W. Y. (2017). "Liar, liar pants on fire": A new benchmark dataset for fake news detection. ArXiv e-prints,.
Wang, Y., Ma, F., Jin, Z., Yuan, Y., Xun, G., Jha, K., et al. (2018). *Eann: Event adversarial neural networks for multi-modal fake news detection. Proceedings of the 24th acm sigkdd international conference on knowledge discovery & data mining*. ACM849–857.
Wang, Y., Wang, S., Tang, J., Liu, H., & Li, B. (2015). *Unsupervised sentiment analysis for social media images. IJCAI*2378–2379.
Wang Baldonado, M. Q., Woodruff, A., & Kuchinsky, A. (2000). *Guidelines for using multiple views in information visualization. Proceedings of the working conference on advanced visual interfaces*. ACM110–119.

Warriner, A. B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 english lemmas. *Behavior Research Methods, 45*(4), 1191–1207.
Watts, D. J., Peretti, J., & Frumin, M. (2007). *Viral marketing for the real world.* Harvard Business School Pub.
Wilson, T., Wiebe, J., & Hoffmann, P. (2005). *Recognizing contextual polarity in phrase-level sentiment analysis. Proceedings of the conference on human language technology and empirical methods in natural language processing.* Association for Computational Linguistics347–354.
Wu, K., Yang, S., & Zhu, K. Q. (2015). *False rumors detection on sina weibo by propagation structures. Data engineering (ICDE), 2015 IEEE 31st international conference on.* IEEE651–662.
Wu, L., Li, J., Hu, X., & Liu, H. (2017). *Gleaning wisdom from the past: Early detection of emerging rumors in social media. Proceedings of the 2017 siam international conference on data mining.* SIAM99–107.
Yang, F., Liu, Y., Yu, X., & Yang, M. (2012). *Automatic detection of rumor on sina weibo. Proceedings of the ACM SIGKDD workshop on mining data semantics*. ACM13.
Yang, J., & Counts, S. (2010). Predicting the speed, scale, and range of information diffusion in twitter. *ICWSM, 10*(2010), 355–358.
Yang, S., Shu, K., Wang, S., Gu, R., Wu, F., & Liu, H. (2019). Unsupervised fake news detection on social media: A generative approach. *AAAI19*.
Youli, F., Hong, W., Ruitong, D., Lutong, W., & Li, J. (2018). *Detecting fake reviews based on review-rating consistency and multi-dimensional time series. International conference on algorithms and architectures for parallel processing.* Springer117–123.
Zhang, X., Lashkari, A. H., & Ghorbani, A. A. (2017). A lightweight online advertising classification system using lexical-based features. *International Conference on Security and Cryptography, 4*, 486–494. https://doi.org/10.5220/0006459804860494.
Zhao, J., Cao, N., Wen, Z., Song, Y., Lin, Y.-R., & Collins, C. (2014). # fluxflow: Visual analysis of anomalous information spreading on social media. *IEEE Transactions on Visualization and Computer Graphics, 20*(12), 1773–1782.
Zhao, Z., Resnick, P., & Mei, Q. (2015). *Enquiring minds: Early detection of rumors in social media from enquiry posts. Proceedings of the 24th international conference on world wide web*. International World Wide Web Conferences Steering Committee1395–1405.