Social Network Analysis and Mining (2020) 10:82
https://doi.org/10.1007/s13278-020-00696-x

REVIEW PAPER

Deep learning for misinformation detection on online social networks: a survey and new perspectives

Md Rafiqul Islam[1] · Shaowu Liu[1] · Xianzhi Wang[2] · Guandong Xu[1]

Received: 28 March 2020 / Revised: 11 September 2020 / Accepted: 12 September 2020 / Published online: 29 September 2020

Abstract

Recently, the use of social networks such as Facebook, Twitter, and Sina Weibo has become an inseparable part of our daily lives. It is considered as a convenient platform for users to share personal messages, pictures, and videos. However, while people enjoy social networks, many deceptive activities such as fake news or rumors can mislead users into believing misinformation. Besides, spreading the massive amount of misinformation in social networks has become a global risk. Therefore, misinformation detection (MID) in social networks has gained a great deal of attention and is considered an emerging area of research interest. We find that several studies related to MID have been studied to new research problems and techniques. While important, however, the automated detection of misinformation is difficult to accomplish as it requires the advanced model to understand how related or unrelated the reported information is when compared to real information. The existing studies have mainly focused on three broad categories of misinformation: false information, fake news, and rumor detection. Therefore, related to the previous issues, we present a comprehensive survey of automated misinformation detection on (i) false information, (ii) rumors, (iii) spam, (iv) fake news, and (v) disinformation. We provide a state-of-the-art review on MID where deep learning (DL) is used to automatically process data and create patterns to make decisions not only to extract global features but also to achieve better results. We further show that DL is an effective and scalable technique for the state-of-the-art MID. Finally, we suggest several open issues that currently limit real-world implementation and point to future directions along this dimension.

Keywords Deep learning · Neural network · Misinformation detection · Decision making · Online social networks

1 Introduction

On online social networks such as Facebook[1], Twitter[2], and Sina Weibo[3], people share their opinions, videos, and news on their various activities (Gao and Liu 2014; Islam et al. 2018a). While people enjoy social networks, many deceptive activities such as fake news, or rumors can mislead users into believing misinformation (Kumar et al. 2016). Therefore, MID in social networks has gained a great deal of attention and is considered an emerging area of research interest recently (Wu et al. 2019; Goswami and Kumar 2016). However, the automated detection of misinformation is difficult to accomplish as it requires the advanced model to understand how related or unrelated the reported information is when compared to real information (Wu et al. 2019). Also, to solve many complex MID problems, academia and industry researchers have applied DL to a large number of applications to make decisions (Xu et al. 2019; Yenala et al. 2018; Yin et al. 2020). Therefore, this survey seeks to provide such a systematic review of current research on MID based on DL techniques.

✉ Guandong Xu
Guandong.Xu@uts.edu.au

Md Rafiqul Islam
MdRafiqul.Islam-1@student.uts.edu.au

Shaowu Liu
Shaowu.Liu@uts.edu.au

Xianzhi Wang
XIANZHI.WANG@uts.edu.au

[1] Advanced Analytics Institute (AAi), University of Technology Sydney (UTS), Sydney, Australia

[2] School of Computer Science, University of Technology Sydney (UTS), Sydney, Australia

[1] https://www.facebook.com/

[2] https://twitter.com/

[3] https://weibo.com/

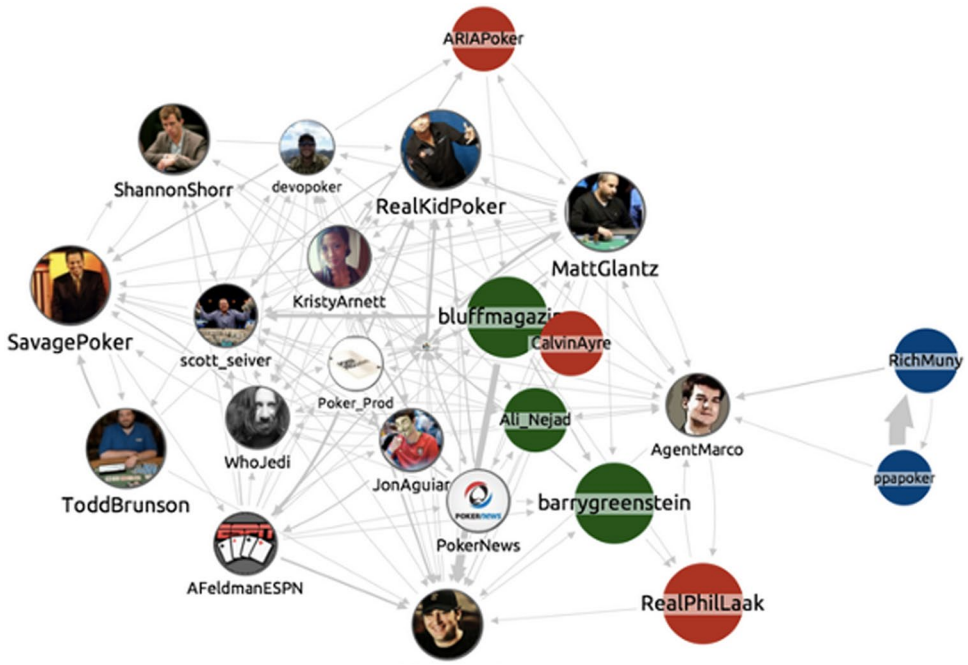

Fig. 1 Social relationships between different users

Social network (SN) sites are a dynamic platform that is now being utilized for different purposes such as education, business, medical purposes, telemarketing, but also, unfortunately, unlawful activities (Vartapetiance and Gillam 2014; Wu et al. 2019; Naseem et al. 2020). Generally, people use SN to socialize with their interested friends and colleagues. Additionally, it is utilized as a channel to speak with clients, and its information can be valuable for identifying new patterns in business insights (Bindu et al. 2017). Figure 1 illustrates the rapid information exchange between users regardless of their location. For example, on online social media, businesses share their product marketing, organizations share their daily activities, celebrities share news on their various activities, and government bodies share information on their various responsibilities. As a result, businesses can offer discounts on their products based on observations of current market demand, customer feedback, etc. Besides, they have realized that online marketing is spreading faster than manual marketing (Acquisti and Gross 2009; Tsui 2017; Nguyen et al. 2017a; Quah and Sriganesh 2008). Similarly, celebrities use SN to increase their public exposure, and the government uses it to collect public opinions. However, it also opens the door for unlawful activities which harm society, business markets, healthcare systems, etc. where incorrect or misleading information is intentionally or unintentionally spread (Bharti et al. 2017; Sun et al. 2018; Gao and Liu 2014). Therefore, SN has attracted a lot of attention and is considered to be a developing interdisciplinary research area that aims to analyze, combine, explore, and adjust techniques to investigate SN data globally. Although existing studies might consider the concept of misinformation in a different view, we consider MID in social media is a timely matter of concern (Sharma et al. 2019; Shu et al. 2020).

Misinformation is inaccurate information which is created to misguide the readers (Fernandez and Alani 2018; Zhang et al. 2018a). There are numerous terms related to misinformation including fake news, rumors, spam, and disinformation, which usually contain numerical, categorical, textual, image, etc., data and used to initiate terrible outcomes (Ma et al. 2016; Bharti et al. 2017; Helmstetter and Paulheim 2018). Due to the high dependency on social media, many dishonest people get a chance to spread misinformation via a false account (Kumar and Shah 2018; Shu et al. 2019c). Additionally, the information they provide is well written, long, and well referenced. So, the readers trust their activities. However, the spread of misinformation will be ineffective if people can identify the different types of misbehavior including fake reviews, false information, and rumors. And, identifying misinformation using SN data may provide early feedback on emerging issues, such as stock movement, political gossip, social issues, and business performance (Habib et al. 2019). In this regard, various techniques have been applied to differentiate genuine and fraudulent information or users over the past years (Islam et al. 2018a; De Choudhury et al. 2013a; De Choudhury et al. 2013b). However, it is very difficult for traditional methods to analyze all these types of misinformation. Therefore, deep learning-based detection approaches can be designed to fit various types of features for MID.

The development of machine learning (ML) and DL techniques have attracted significant attention for different

purposes both from industry and research communities (LeCun et al. 2015; Islam et al. 2019). In particular, DL-based detection approaches have become a major source of MID. For example, a large volume of research works have explored on automatic MID (Jain et al. 2016; Qazvinian et al. 2011; Zhang et al. 2016), as well as related terms, e.g., rumor (Sampson et al. 2016; Wu et al. 2017), fake information (Kwon et al. 2017; Ma et al. 2016; Ruchansky et al. 2017; Shu et al. 2017; Wu et al. 2017), and spam detection (Hu et al. 2013; Yin et al. 2018; Li and Liu 2018; Markines et al. 2009; Wang et al. 2011). Therefore, the success of DL for MID both in academia and industry requires a systematic review to better understand the scenarios of the existing problem and current research issues. Although there have been attempts to review and summarize the literature on MID in a very nice way, there are still enough spaces to review the literature on misinformation in a broader way. For example, in an existing survey, Shu et al. (2017) shows a fascinating association between psychological concept, fake news, and social network with data mining techniques. The literature surveyed by Zubiaga et al. (2018), Yu et al. (2017b), and Zhang et al. (2015) illustrates a related problem of rumor detection, where they differentiate between unverified and verified information, wherein the unverified information may remain unresolved or may turn out to be true or false. Additionally, Kumar and Shah (2018) addressed a broader scope of false information on the web which presents the existing work, current progress, and future directions together. However, from the existing studies, we find that (1) there is no clear boundary definition between misinformation, disinformation, and false information, and (2) there are no DL method-based systematic reviews on MID whether different types of misinformation problems have been summarized under each DL technique. We wish to emphasize that misinformation is getting more and more complex, making the conventional machine learning techniques incapable of detecting them. For example, due to the recent advances in large-scale pre-trained models (e.g., BERT, GPT-3) and adversarial learning, programs can generate misinformation in an automated and difficult to detect manner. This calls for the need of using high capacity models, such as DL. Therefore, although existing reviews are important in their own right, much like the detection of scam email (Saberi et al. 2007), fake followers (Cresci et al. 2015), or false web links (Lake 2014), we decide to focus on MID to provide detailed discussions of DL techniques and their limitations.

The existing surveys covered a broad range of techniques used for MID. However, given the increasing popularity of using DL methods to detect misinformation, we believe our survey provides a timely review of the use of DL techniques. For example, we reviewed how different MID problems are covered under various DL techniques, which were not covered in existing surveys. We hope this survey can benefit researchers to deep insight between related techniques and these issues. Moreover, despite the promising outcomes of DL techniques, we focus on some open issues such as data volume, data quality, explainability, domain complexity, interpretability, feature enrichment, model privacy, incorporating expert knowledge, and temporal modeling, which are necessary to understand the advances in this domain. In summary, the main contributions of our survey are as follows:

- We present a state-of-the-art systematic review of the existing problems, solutions, and validation of MID in online social networks based on various DL techniques.
- To identify the recent and future trends of MID research, we analyze the key strengths and limitations of the existing various techniques and describe the state-of-the-art DL as an emerging technique on massive social network data.
- We provide some open issues that contribute to this new exciting field based on DL techniques.

In the rest of the paper, first, we present the MID with the formal problem definition, types, impacts, and DL with the associated challenges. We then present the state-of-the-art DL techniques for MID. Further, to encourage researchers with rigorous evaluation and comparison, we include a list of open issues that outline promising directions for future research. Finally, we present the conclusion.

2 Background

2.1 Misinformation detection

In this section, we discuss the formal definition of misinformation, types, and its impact on SN. But we do not discuss in this section how to undertake MID, what techniques are used, and what techniques are effective. The introduction of this paper has shown that DL is now an emerging technique that plays a critical role in MID. As our main task is to review the learning process of DL for MID, in the following sections, we discuss the importance of DL for MID, overview its existing performance, and provide some open issues to work in the future.

2.1.1 What is misinformation?

Misinformation is a false statement to lead people astray by hiding the correct facts. It is also referred to as deception, ambiguity, falsehoods, etc. (Zhang et al. 2016). It generates feelings of mistrust that subsequently weaken relationships, which is a negative violation of expectations (Wu et al.

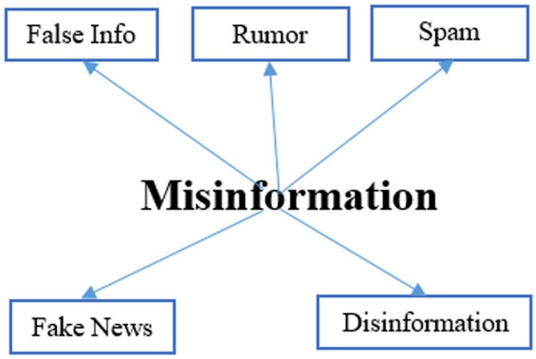

Fig. 2 Key types related to misinformation

2019; Ma et al. 2018). Additionally, people do not expect to receive misinformation from their close friends, relatives, or strangers. Instead, they expect truthful communication. For example, some people were involved in a Facebook discussion on a recently published product where there are both fake users and real users. The real users discuss the product's features honestly. However, fake users praise the product regardless of their true opinion.

Problem definition Suppose you have been given a restaurant review of E among N users to analyze user feedback a, which contains both genuine feedback and false reviews created by restaurant owners. It is very difficult to distinguish the false review from the true ones. Therefore, the researcher's role is to identify the real and false reviews.

$$F(a) = \begin{cases} 1, & \text{if a is false} \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

According to Wu et al. (2019), in the following section, we provide five major terms related to MID, namely rumor, fake news, false information, spam, and disinformation as shown in Fig. 2. We then describe a brief literature review of existing DL techniques for MID in Sect.3.

2.1.2 Types of misinformation

There are many terms related to misinformation such as rumor, fake news, false information, spam, and disinformation. Rumor is a story of circulating information from person to person whose veracity status is doubtful (Lin et al. 2019). Fake news is a news article that intentionally misleads the readers, and it is verifiably false (Shu et al. 2018). Misinformation can be broadly used to treat information as False information (Kumar and Shah 2018). Spam can be referred as an unsolicited message which sends over the internet for spreading malware, advertising, etc. (Rayana and Akoglu 2015; Hu et al. 2013). Disinformation is a piece of inaccurate information that is spread intentionally to mislead people (Galitsky 2015). Although misinformation and disinformation both refer to incorrect/false information, a big difference between them lies in the intention - without the intent misinformation is spread to deceive while disinformation is spread with the intent (Kumar et al. 2016; Hernon 1995; Fallis 2014). Several studies have been forwarded for misinformation identification on social media (Kumar and Shah 2018; Wu et al. 2019). Some works treat a microblog post an object, obtain the credibility of the post, and aggregate to the event level (Jin et al. 2017a), (Jindal et al.), (Qazvinian et al. 2011; Gupta et al. 2013). Additionally, some work extracts various features from the event level and identifies whether an event belongs to misinformation (Kwon et al. 2017; Ma et al. 2016; Zhang et al. 2016). Moreover, some other works extract more effective hand-crafted features, including conflict viewpoints (Jin et al. 2017c), temporal properties (Kwon et al. 2017; Ma et al. 2019), users' feedback (Shu et al. 2020), and signals tweets containing skepticism (Zhao et al. 2015). Therefore, to better understand misinformation in social media, Fig. 2 illustrates the five related terms of misinformation such as false information, rumors, fake news, spam, and disinformation. In this section, we define and describe each type of misinformation, respectively.

False information False information is a broader concept of misinformation. Intentionally, it is interchangeably used to define as a correct information. In social networks, some unscrupulous people exploit this for their interests by ensuring that the basic patterns of the original information are correct (Kumar and Shah 2018; Habib et al. 2019). For example, we generally expect honest users to provide positive reviews to good products and negative reviews to bad products, whereas dishonest users may not follow this behavior. Existing studies estimated that 20% of the reviews on Yelp are fake (Donfro 2013). This large number of fake reviews, which are getting more and more difficult to detect, call for the use of advanced DL techniques to extract meaningful features and to identify the review as fake or real accurately.

Rumor A rumor is a story of doubtful truth that is easy to spread widely online (Zubiaga et al. 2017). The rumor is spread by dishonest business people for their benefit (Zubiaga et al. 2018, 2016b). For example, a rumor spreads on social media that recently the price of salt and onion had increased in Bangladesh and some shops quickly increased their prices[4], [5]. In this regard, some people purchased more of these products than they needed. Additionally, Fig. 3 depicts an interesting rumor campaign about "Saudi Arabia's first female robot citizen beheaded," which shows how

[4] https://www.thedailystar.net/country/rumour-salt-price-raising-spreads-among-consumers-1829260

[5] https://www.dhakatribune.com/business/2019/11/19/rumour-drives-groceries-out-of-salt-stock-across-country

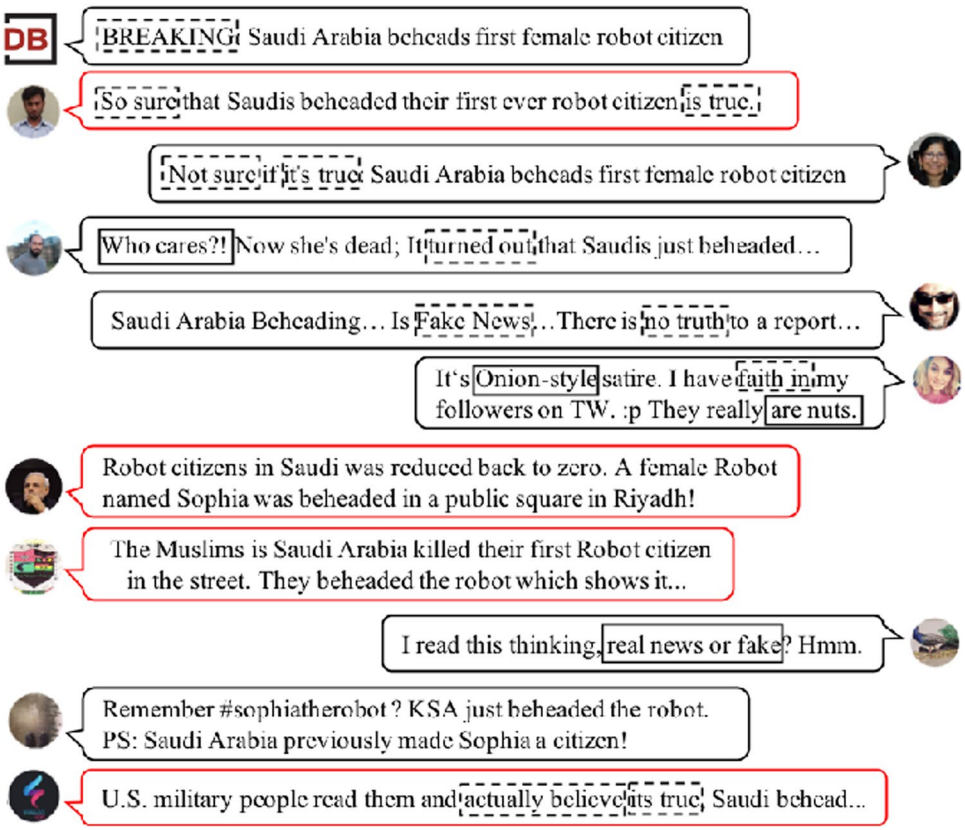

Fig. 3 Sample responses to a rumor claim. Figure courtesy by Ma et al. (2019)

popular index patterns are overwhelmed by propaganda terms expressing doubts and disagreements like "fake news."

Fake news Fake news is a modified version of an original news story which is spread intentionally and very difficult to identify (Cui et al. 2019). It mimics traditional news and spreads easily on social media, reaches a large number of people quickly, and deceives many (Kumar et al. 2019; Shu et al. 2017).

Spam Spam is an unwanted message that generally contains irrelevant or inappropriate information to mislead users (Yilmaz and Durahim 2018). It is difficult to distinguish spam from real messages, as spammers hack users' information (Çıtlak et al. 2019).

Disinformation Disinformation is a subset of misinformation, which is false or misleading information. It is intentionally spread online to deceive others, and its impact has continued to grow (Hernon 1995; Galitsky 2015). Misinformation is conveyed in the honest but mistaken belief that the relayed incorrect facts are true. However, disinformation defines false facts that are conceived to deliberately deceive an audience. One recent disinformation example is pure alcohol that can cure the coronavirus infections during the COVID-19 pandemic situation. However, pure alcohol can be very harmful to the human body[6], [7].

In summary, we compare the five related terms of misinformation as shown in Table 1. For example, on characteristics, disinformation provides misleading features that also have a specific objective. Additionally, different types of misinformation have different categories of side effects. On integrity, we use sure and not sure levels to evaluate five different types of misinformation.

2.1.3 Impact of misinformation

Misinformation can affect every aspect of life such as the social, political, economic, stock market, emergency response during natural disasters, and crisis events. It aims to intentionally or unintentionally mislead public opinions, influence political elections, and threaten public security and social stability (Wu et al. 2019). Most of the time, it reveals fabricated information related to fictional issues rather than

[6] https://ec.europa.eu/info/live-work-travel-eu/health/coronavirus-response/fighting-disinformation/tackling-coronavirus-disinformation-en

[7] https://www.lowyinstitute.org/the-interpreter/disinformation-and-coronavirus

Table 1 Comparison of different types of misinformation

Type	Characteristics	Objectiveness	Severity	Integrity	References
Rumors	Ambiguous	Not sure	Low	Not sure	Shu et al. (2020)
False information	Deception	Yes	High	False	Kumar and Shah (2018)
Fake news	Misguided	Yes	Medium	False	Sharma et al. (2019)
Spam	Confused	Yes	Low	Not sure	Çıtlak et al. (2019)
Disinformation	Mislead/deceive	Yes	Medium	False	Guo et al. (2019)

relevant information (Fernandez and Alani 2018). Nowadays, it has become easier to spread misinformation quickly due to social network platforms such as Facebook, Twitter, and Sina Weibo. In particular, when people engage in conversation, one can share information that is stated to be factual but that may not always be true. Additionally, fraudulent users share misleading information to look for personal gain in some way. For example, concerning political issues, some view being a misled resident as more regrettable than being an uninformed resident. Misguided residents express their opinions with certainty and thus influence in elections. This kind of deception originates from speakers not continually being forthright and clear.

With the advent of SN and technological advances around the world, there has been a great explosion of misinformation (Kumar et al. 2016; Sharma et al. 2019; Goswami and Kumar 2016). In the last few decades, several studies have been conducted to measure the impact of misinformation such as rumors, fake reviews, and fake news. For example, Friggeri et al. (2014) studied the spread of rumors on Facebook, and (Willmore) analyzed the use of fake election news articles on Facebook. Another work by Zubiaga et al. (2018) discussed how rumors spread quickly on social media (Twitter) and how this is becoming a threat to many people. They stated that misinformation has a significant negative impact on the workplace and daily life. For example, an organization can undermine reliable evidence through a purposeful deception campaign. In detail, tobacco companies utilized falsehood in half of the twentieth century to diminish the reliability of studies that showed the connection between smoking and lung disease (Brandt 2012). In the clinical field, misinformation can quickly prompt life endangerment as found in the case of the public's negative observation toward vaccines to treat diseases.

Overall, in the context of misinformation, existing studies focus on the text content mostly, whereas a few of them investigated image/video content (Jin et al. 2016; Gupta et al. 2014). Although many techniques used for MID, all the approaches have not been proved effective yet (Zhang et al. 2016; Shu et al. 2017). Additionally, existing approaches have some challenges for MID, e.g. data volume, data quality, domain complexity, interpretability, feature enrichment, model privacy, incorporating expert knowledge, temporal modeling, dynamic, etc. (Liu and Xu 2016; Ma et al. 2015). Therefore, we attempt to introduce DL as an emerging state-of-the-art technique for MID.

2.2 Deep learning

The term deep learning (DL) was first introduced to the machine learning community by Dechter (1986) and to artificial neural networks based on a Boolean threshold by Aizenberg (1999). In the field of ML in artificial intelligence, DL is an emerging technique which is used in various applications including computer vision (Wang and Yeung 2013), speech recognition (Hinton et al. 2012), natural language processing (Young et al. 2018), anomaly detection (Du et al. 2017), portfolio optimization (Vo et al. 2019), healthcare monitoring (Islam et al. 2018b), personality mining (Vo et al. 2018), novelty detection in robot behavior, traffic monitoring, visual data processing, social network analysis, etc. Nowadays, it is becoming increasingly used for processing data and creating patterns to assist the decision-making process. Furthermore, this state-of-the-art method helps to improve learning execution, expand the scope of the research area, and simplify the measuring procedure.

Over the few decades, various techniques have been proposed to solve many problems (fake news, misinformation, anomaly detection, etc.) in the online social network. Researchers are constantly finding and investigating research gaps in various domains and attempting to solve these problems using various techniques. Deep learning is one such technique and has become increasingly popular, being explored in a large number of domains with various neural networks such as convolutional neural networks (CNN) (Abdel-Hamid et al. 2014; Kim 2014), recurrent neural networks (RNN) (Cho et al. 2014b; Li and Wu 2015), and long short-term memory (LSTM) (Sun et al. 2018), which are introduced to help other researchers explore their knowledge in different applications.

Deep learning serves as the key to performing complex tasks of higher levels of sophistication. However, to successfully build and deploy them proves to be a challenge for data scientists and engineers all over the world (Liu and Wu 2020; Hardy et al. 2016). Although data training takes a little longer, testing can be done in a very short time. To accelerate DL processing, DL frameworks combine the implementation of

Table 2 List of popular deep learning framework

Framework	Key Point	Interface Support	CNN/RNN Support	References
Caffe	Caffe is one of the most popular deep learning network	C, C++, Python, MATLAB	Yes	Jia et al. (2014)
Torch	Because of using the fast scripting language LuaJIT, torch provides faster performance than other frameworks	C/C++	Yes	Collobert et al. (2002)
PyTorch	PyTorch is a port to torch deep learning framework	Python	Yes	Ketkar (2017)
DL4j	DL4j uses for text-mining, NLP, and image recognition	Java, Scala, and JVM	Yes	Parvat et al. (2017)
Neon	Neon is designed to ease of use and for extensibility	Python	Yes	Pouyanfar et al. (2018)
TensorFlow	TensorFlow is one of the best deep learning frameworks for natural language processing, speech recognition, image processing	Python, C++ and R	Yes	Abadi et al. (2015)
Keras	Keras is a part of the TensorFlow core API and uses for text generation, summarization, and classification	Python	Yes	Chollet (2018)
CNTK	To train deep learning models, CNTK is an open-source deep learning framework for for image, speech, and text-based data	Python, C++	Yes	Shi et al. (2018)
Theano	Theano allows users to define, optimize, and evaluate mathematical expressions on arrays and tensors	C, Python	Yes	Van Merriënboer et al. (2015)
Dlib	Dlib is an independent cross-platform open source software	C++	CNN-Yes RNN-No	King (2009)
Torch	Torch is a machine learning open source software library which provides a large number of algorithms for deep learning	C	Yes	Collobert et al. (2011)
BigDL	It is a distributed deep learning based framework	Python	Yes	Dai et al. (2019)

modularized DL algorithms, optimization techniques, distribution techniques, and support to infrastructures (Chalapathy and Chawla 2019; David and Netanyahu 2015). They are developed to simplify the implementation process and boost system-level development and research. Table 2 shows some popular DL frameworks such as Caffe, Torch, TensorFlow, MXNet, and CNTK, which allow researchers to develop tools and can offer a better level of abstraction and simplify difficult programming challenges. Each framework is built differently for different purposes. It can be observed from Table 2 that most of the frameworks are implemented in Python, which is the most common language for DL architecture design. It can make programming more efficient and easier by simplifying the programming process.

3 Deep learning for misinformation detection: state-of-the-art

Misinformation detection is defined as an observation that deviates greatly from other observations and thereby arouses suspicion that it was generated by a different mechanism. In Sect. 2.1.2, we have discussed different terms related to misinformation with examples. It is observed that the same type of problem has been solved by many techniques. Although many techniques are being used to detect misinformation in social network data, DL is one of the better approaches to use. However, the same type of misinformation problems has been solved with various DL techniques (Table 3). Additionally, these types of DL techniques are dependent on different data characteristics and used to automatically identify misinformation. Therefore, we have divided the DL techniques into three main categories based on the model as follows: (1) discriminative models, (2) generative models, and (3) hybrid models. All three categories have a large number of architectural models that are commonly used for MID. However, due to differences in performance, we only discussed 12 models namely convolutional neural networks (CNN), recurrent neural networks (RNN), recursive neural networks (RvNN), restricted Boltzmann machines (RBM), deep Boltzmann machines (DBM), deep belief networks (DBN), variational autoencoders (VAE), convolutional restricted Boltzmann machines (CRBM), convolutional

Table 3 Deep learning methods showing the performances within the applications of social network research

Models	Input Data		User Response	Problem Tackled	References
	Text	Visual			
CNN+LSTM	✓		✓	Disinformation detection	Dhamani et al. (2019)
LSTM+BiLSTM	✓			False claim detection	Popat et al. (2018)
RCNN	✓			False information detection	Wu et al. (2018)
BiLSTM	✓		✓	Misinformation detection	Zhang et al. (2019)
RNN+ GRU	✓		✓	Fake news detection	Shu et al. (2019a)
CNN+Attention	✓		✓	Review spam detection	Gong et al. (2020)
CNN+LSTM	✓			Spam detection	Shahariar et al. (2019)
LSTM+Attention	✓		✓	Early rumor detection	Chen et al. (2018)
Attention	✓		✓	Misinformation identification	Liu et al. (2018)
LSTM+Attention	✓	✓	✓	Fake news detection	Popat et al. (2018)
RNN	✓		✓	Fake news detection	Ruchansky et al. (2017)
CNN	✓		✓	Misinformation identification	Jia et al. (2016), Yu et al. (2017a)
LSTM+Attention	✓		✓	Rumor detection	Guo et al. (2018)
RNN	✓	✓	✓	Rumor detection	Jin et al. (2017b)
GRU	✓		✓	Rumor detection	Li et al. (2018a)
CNN+GRU			✓	Early detection of fake news	Liu and Wu (2018)
RNN	✓		✓	Rumor detection	Ma et al. (2016)
CNN+LSTM	✓		✓	Rumor detection	Nguyen et al. (2017b)

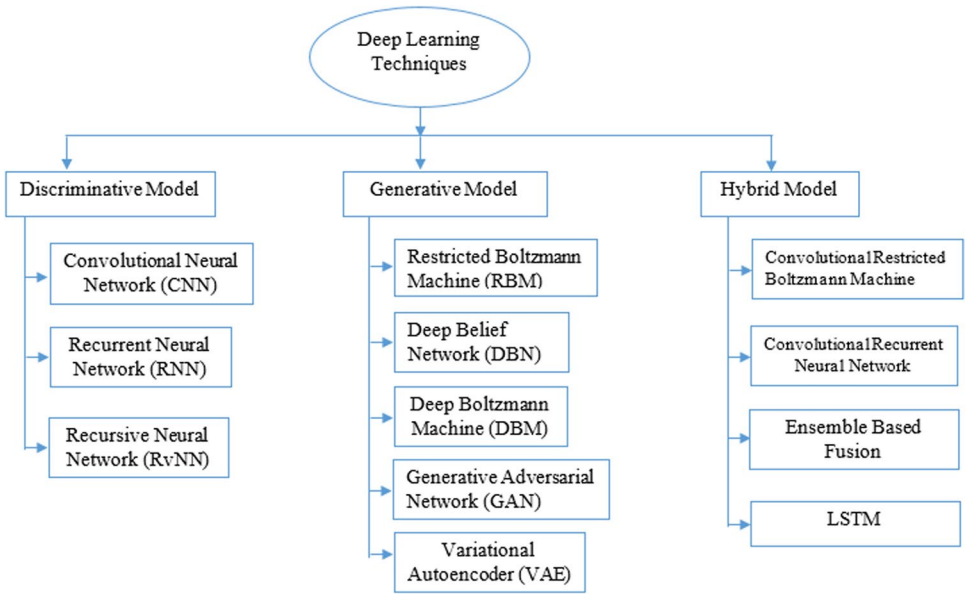

Fig. 4 Classification of deep learning models

recurrent neural networks (CRNN), ensemble-based fusion (EBF), and long short-term memory (LSTM), as shown in Fig. 4. We discuss each model that uses for MID, respectively.

3.1 Discriminative model for detecting misinformation

A variety of discriminative models used social content and context-based features for MID. In recent years, to tackle the problem of misinformation, several studies have been conducted and revealed some promising preliminary results. Therefore, we briefly review the three discriminative models

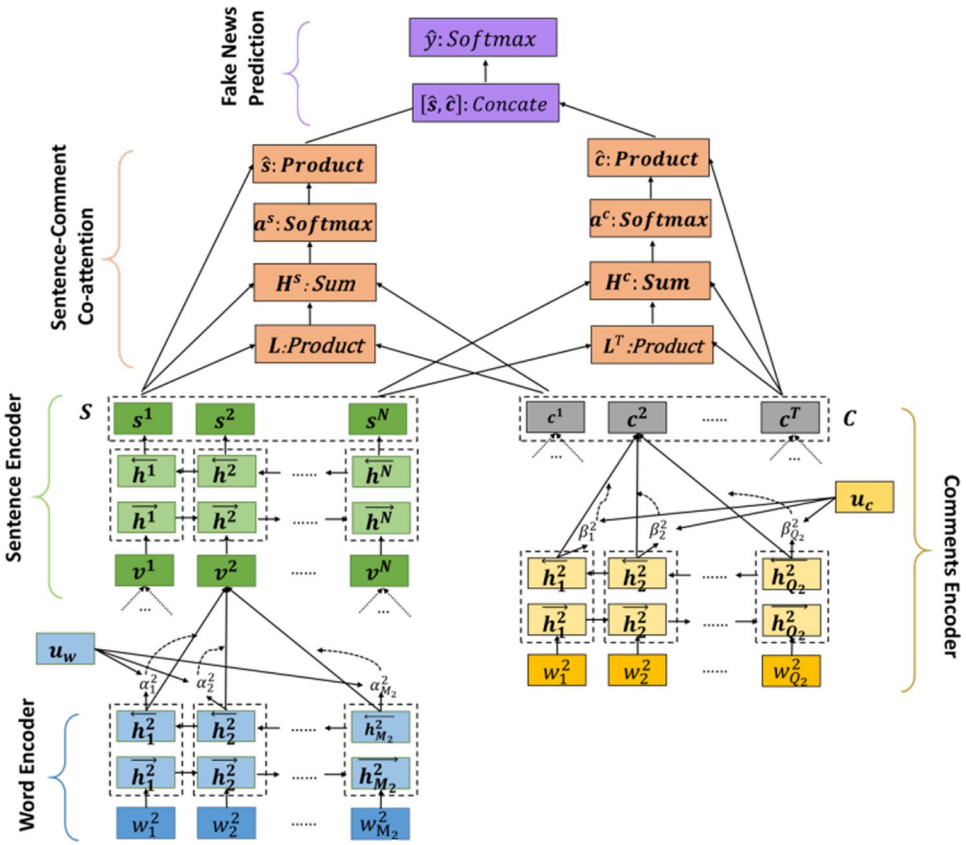

Fig. 5 RNN architecture used for fake news detection. Figure courtesy by Shu et al. (2019a)

namely CNN, RNN, and RvNN, respectively. It is noted that the discriminative-based models have demonstrated significant advances in text classification and analysis.

Convolutional Neural Network (CNN) CNN is one of the most popular and widely used models for the state-of-the-art of many computer vision tasks (LeCun et al. 2010). However, recently, it has been extensively applied in the NLP community as well (Jacovi et al. 2018). For example, Chen et al. (2017) introduced a convolutional neural network-based classification method with single and multi-word embedding for identifying both rumor and stance tweets. Kumar et al. (2019) introduced both a CNN and a bidirectional LSTM ensembled network with an attention mechanism to solve MID. Additionally, Yang et al. (2018) stated that online social media is continually growing in popularity and genuine users are being attacked by many fraudulent users. They informed that fake news is written to intentionally mislead users. In their paper, they applied the TI-CNN model to identify the explicit and latent features from the text and image information. They demonstrated that their model solves the fake news detection problem effectively.

Recurrent Neural Network (RNN) RNN utilizes the sequential information in the network which is essential in many applications where the embedded structure in the data sequence conveys useful knowledge (Alkhodair et al. 2020). The advantage of RNN is its ability to better capture contextual information. To detect rumors, existing methods rely on handcrafted features to employ machine learning algorithms that require a huge manual effort. To guard against this issue, the earliest adoption of RNNs for rumor detection is reported in Ma et al. (2016) and recurrent neural networks with attention mechanism in Chen et al. (2018) and Jin et al. (2017b). Figure 5 shows the RNN architecture used for the fake news detection proposed by (Shu et al. 2019a). Authors have proposed different RNN architectures, namely tanh-RNN, LSTM and Gated Recurrent Unit (GRU) (Cho et al. 2014a). Among the proposed architectures, GRU has obtained the best results in both the datasets considered, with 0.88 and 0.91 accuracy, respectively. Ma et al. (2016) proposed a RNN model to learn and that captures variations in relevant information in posts over time. Additionally, they described that RNN utilizes the sequential information in the network where the embedded structure in the data sequence conveys useful knowledge. They demonstrated that their proposed model can capture more data from hidden layers which give better results than the other models.

Recursive Neural Network (RvNN) Researchers are more concerned to identify unscrupulous users in SN and want to protect genuine users from fraudulent behavior (Guo et al. 2019). Therefore, RvNN is one of the most widely used and successful networks for many natural language processing (NLP) tasks (Socher et al. 2013; Zubiaga et al. 2016a). This

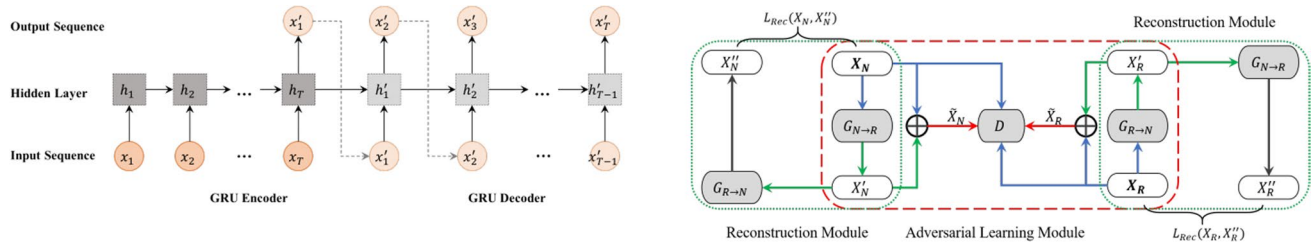

Fig. 6 The architecture of the generative adversarial learning model. Figure courtesy by Ma et al. (2019)

architecture processes objects that can make predictions in a hierarchical structure and classifies the outputs using compositional vectors. To reproduce the patterns of the input layer to the output layer, this network is trained by auto-association. Also, this model analyzes a text word by word and stores the semantics of all the previous texts in a fixed-sized hidden layer (Cho et al. 2014b). For instance, Zubiaga et al. (2016b) proposed a RvNN architecture for handling the input of different modalities. Ma et al. (2018) proposed a model that collects tweets from Twitter and extracts features from discriminating information. It follows a non-sequential pattern to present a more robust identification of the various types of rumour-related content structures.

3.2 Generative model for detecting misinformation

Over the last few decades, online social media platforms have become the main target space of deceptive opinions where deceptive opinions (such as rumor, spam, troll, fake news) are deliberately written to sound authentic. Several existing works for MID are based on syntactic and lexical patterns or features of opinion. Therefore, in this section, the successful use of five generative models on various classification applications, namely RBM, DBN, DBM, GAN, and VAE are discussed.

Restricted Boltzmann Machine (RBM) RBM is a generative stochastic artificial neural network. It can learn a probability distribution over its set of inputs (Liao et al. 2016). Although learning is impractical in general Boltzmann machines, it can be quite efficient in an architecture called the restricted Boltzmann machine. However, it does not allow intra-layer connections between hidden units (Papa et al. 2015). Therefore, this method of stacking RBMs makes it possible to train many layers of hidden units efficiently. RBMs have been applied in various applications, but very few works have been addressed in the context of MID. However, in the last few decades, researchers are attempting to fit this method to identify fake, rumour, spam, etc. on social media platforms. For instance, Da Silva et al. (2018), da Silva et al. (2016), and Silva et al. (2015) applied RBMs to automatically extract the features related to spam detection.

Deep Belief Network (DBN) DBN is a generative graphical model composed of multiple layers of latent variables (hidden units). It connects between the layers but not between units within each layer. DBNs can be viewed as a composition of simple, unsupervised networks such as restricted Boltzmann machines (RBMs) or autoencoders, where each subnetwork's hidden layer serves as the visible layer for the next. There are already many works that have used this network (Li et al. 2018b; Yepes et al. 2014; Alom et al. 2015; Selvaganapathy et al. 2018). For example, Tzortzis and Likas (2007) stated that spam is an unexpected message which contains inappropriate information and first applied to fit DBNs for spam detection. In another paper, Wei et al. (2018) proposed a DBN-based method to identify false data injection attacks in the smart grid. They demonstrated that the DBN-based method achieves a better result than the traditional SVM-based approach.

Deep Boltzmann Machine (DBM) DBM is a type of binary pair-wise markov random field with multiple layers of hidden random variables. This is a network of symmetrically coupled stochastic binary units which have been used to detect malicious activities (Zhang et al. 2012; Dandekar et al. 2017). For example, Jindal et al. used a multimodal benchmark dataset for fake news detection. They presented results from a Deep Boltzmann Machine-based multimodal DL model (Srivastava and Salakhutdinov 2012). Zhang et al. (2012) generated a model based on DBMs to detect spoken queries. They presented that their proposed method achieved 10.3% improvement compared to that with the previous Gaussian model.

Generative Adverserial Network (GAN) GAN is a class of ML systems (Goodfellow et al. 2014). Given a training set, this technique learns to generate new data with the same statistics as the training set. When considering earlier studies, we see that the widespread rumors usually result from the deliberate dissemination of information which is generally aimed at forming a consensus on rumor news events. Ma et al. (2019) proposed a generative adversarial network model to make automated rumor detection more robust and efficient and is designed to identify powerful features related to uncertain or conflicting voice production and rumors. Figure 6 illustrates the structure of a deep generative adversarial

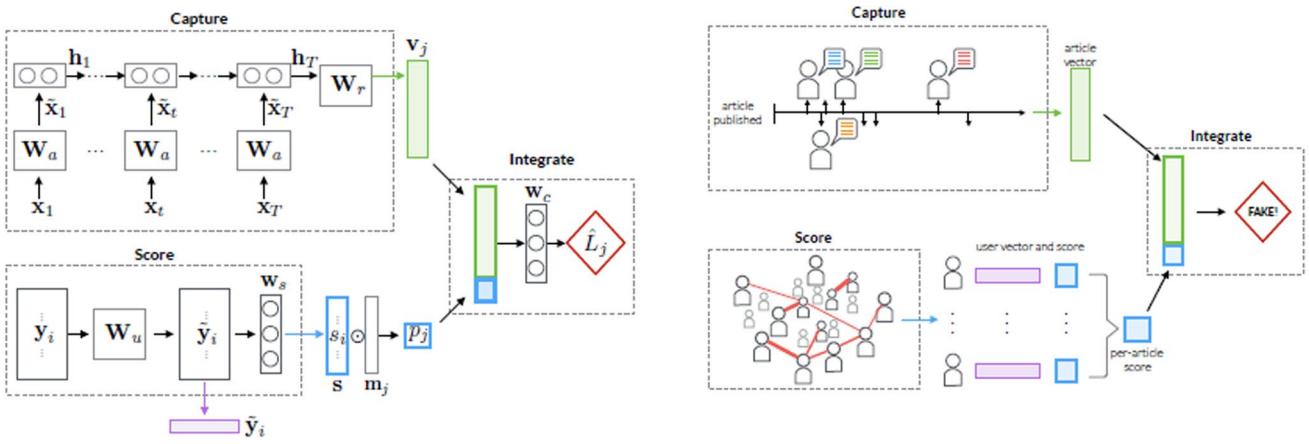

Fig. 7 An illustration of the hybrid model. Figure courtesy by Ruchansky et al. (2017)

learning model for rumors detection proposed by Ma et al. (2019).

Variational Autoencoder (VAE) VAE models make strong assumptions concerning the distribution of latent variables. The use of a variational approach for latent representation learning results in an additional loss component and a specific estimator for the training algorithm called the stochastic gradient variational bayes (SGVB) estimator. Qian et al. (2018) proposed a generative conditional VAE model to extract new patterns by analyzing a user's past meaningful responses on true and false news articles and played a vital role in detecting misinformation on social media. Wu et al. (2017) explored whether the knowledge from the historical data analysis can benefit rumor detection. The result of their study was that similar rumors always produce the same behaviors.

3.3 Hybrid model for detecting misinformation

The tasks of detecting misinformation (such as fake news, rumor, spam, troll, false information, and disinformation) have been made in a variety of ways. A lot of research works have been done using various DL models separately. However, to increase the performance of individual models, the need for hybrid models are immense. Therefore, over the last few decades, hybrid DL has been considered an emerging technique for various purposes. In this section, we review some related works on MID based on the deep hybrid model. The hybrid model consists of CRBM, CRNN, EBF, and LSTM.

Convolutional Recurrent Neural Network (CRNN) Currently, researchers are increasingly focusing on applying CNN and RNN models in a hybrid way to achieve better performance in various applications. They argue that real-world data are structured sequences, with spatio-temporal sequences. For example, several works utilized a blend of CNN and RNN such as spatial and temporal regularities (Lin et al. 2019; Xu et al. 2019; Wang et al. 2019). Their models can process time-shifting visual contributions for variable length expectations. These neural network architectures combine a CNN for visual element extraction with an RNN for grouping learning. Besides, such models have been effectively utilized for fake news, rumor, false information, and spammer detection. For example, to identify rumor for events on social media platform, Lin et al. (2019) proposed a novel rumor detection method based on a hierarchical recurrent convolutional neural network. They use the RCNN model to learn contextual information and utilize the bidirectional GRU network to learn time period information. Figure 7 shows the structure of a deep hybrid model for fake news detection proposed by Ruchansky et al. (2017). Xu et al. (2019) proposed a CRNN model to extract data from textual overlays, for example, captions, key ideas, or scene level summaries for rumor detection on Sina Weibo. They proposed this CRNN model to create training data intended for textual overlays regularly occurring in the online sina weibo platform. Zhang et al. (2018c) proposed an approach called deceptive review identification by recurrent convolutional neural network (DRI-RCNN) to identify the deceptive review of the content. They compared the neural network approaches (RvNN, LB-SVM, CNN, and RCNN, GRNN) to the widely used conventional strategies. Their experimental results demonstrated that the neural network approaches outperform the conventional techniques for all datasets.

Convolutional Restricted Boltzman Machine (CRBM) An extension of the RBM model, called the convolutional RBM (CRBM), was developed by Norouzi (2009). He informed that CRBM, like the RBM, is a two-layer model in which visible and hidden arbitrary factors are organized as matrices. He proposed a way to make a Boltzmann machine and convolutional limited Boltzmann machine, forming a deep

network to improve its presentation for both image processing and feature extraction. He also provided a simple and intuitive training method that jointly optimizes all RBMs in the network, which works well in practice. For instance, Norouzi et al. (2009) proposed a CRBM model for learning features specific to an object class. In which associations are nearby and loads are shared with the spatial structure of pictures and stack them over one another to construct a multilayer progressive system of exchanging, separating, and pooling.

Ensemble-Based Fusion To study profile information, Wang (2017) proposed a hybrid model where he used speaker profiles as a part of the input data. He also made the first large-scale fake news detection benchmark dataset with speaker details information such as location of speech, party affiliation, job title, credit history, as well as topic. Tschiatschek et al. (2018) investigated the vital problem of leveraging crowd signals to detect fake news. They analyzed user's flagging behaviors and applied novel algorithms detective to perform Bayesian inference to detect fake news. Their experiments performed well in identifying a genuine user's flagging behavior. Zhang et al. (2018b) explored a new idea to detect fake news on social media. They identified some deceptive words which can be used by these online fake users and harm offline society. Shu et al. (2018) discussed that social media has become a popular network for sharing misinformation and presented FakeNewsNet as fake news data respiratory for further analysis. Roy et al. (2018) presented misinformation and applied many various DL techniques such as CNN, Bi-LSTM, and MLP to detect fake news. They claimed that the rate of misinformation is increasing rapidly.

LSTM Density Mixture Model Although traditional methods have used lexical features to detect fake news automatically, the hybrid deep neural network has received a lot of attention globally. For example, Ruchansky et al. (2017) stated that fake news detection has gained a great deal of attention both from the research and academic communities. In their work, they identified three types of fake news: (1) the text of an article, (2) the user response it receives, and (3) the source on which users promote it. They analyzed that fake news has the importance to affect public opinion. Existing studies have mostly focused on tailoring solutions to one particular problem with their limited success. However, Ruchansky et al. (2017) proposed a hybrid model combining all their characteristics to predict the more accurate and automated result. Similarly, Long et al. (2017) proposed a novel method to incorporate speaker profiles into an attention-based LSTM model for fake news detection. Additionally, several studies described that LSTM-based hybrid model is proven to work better for long sentences and attention models are also proposed to weigh the importance of different words in their context (Tang et al. 2015; Prova et al. 2019).

In another study, Kudugunta and Ferrara (2018) stated that bots have been used to sway political elections by distorting online discourse, to manipulate the stock market that may have caused health epidemics. They applied LSTM-based architecture that exploits both content and metadata to detect bots. They claimed that their model can achieve an extremely high accuracy exceeding 0.96 AUC. Yenala et al. (2018) identified that the automatic detection and filtering of inappropriate messages or comments have become an important problem for improving the quality of conversations with users as well as virtual agents. They proposed a novel hybrid DL model to automatically identify the inappropriate language. Zhao et al. (2015) created a hybrid model namely C-LSTM where they combined CNN and LSTM for the sentiment analysis of movie reviews and question-type classification.

4 Discussion with open issues and future research

Over the few years, several researchers have applied the recent developments and easy access of DL techniques to fake news, rumor, spam, etc. in the online social networks. It enables frameworks to process, handle, and adapt a lot of information. It is now being utilized the most in business insight frameworks and predictive analytics, as well as in increasingly advanced learning management systems (LMS). Therefore, we followed various DL methods inspired by the guidelines of Pouyanfar et al. (2018), Pouyanfar and Chen (2016), Perozzi et al. (2014), and Savage et al. (2014). The development of DL can potentially benefit to MID research. However, existing studies are not directly comparable to each other due to the lack of large-scale publicly available datasets. There are still numerous improvements that can be made to the models. Furthermore, DL is one of the most effective methods for the present development of innovation in the world. This method has given computers remarkable power, for example, the capacity to perceive discourse similar to a human, to prepare a model with no requirement for feature extraction or data labeling. It is currently being utilized to guide and improve a wide range of key procedures. As previously stated there are many reasons to apply DL to MID, we summarize some of the strengths of the deep learning-based model in the following.

- First, deep learning techniques are more robust and effective than state-of-the-art baseline approaches and have shown their strength in various applications, in particular, misinformation (fake, spam, rumor, false informa-

Table 4 Summary of datasets used by existing efforts

Dataset	Problem Tackled	Text Content	Number of Instances	Number of Classes	Ground Truth	References
BuzzfeedPolitical	Fake news detection	✓	120	2	✓	Silverman et al. (2016)
LIAR	Fake news detection	✓	12.8K	6	✓	Wang (2017)
CREDBANK	Fact extraction	✓	4856	2	✓	Mitra and Gilbert (2015)
FakeNewsNet	Rumor detection	✓	CNN		✓	Shu et al. (2019b, (2017)
Twitter	Rumor detection	✓	1111	2	✓	Ma et al. (2018)
PHEME	Rumor detection	✓	6425	2	✓	Aiello et al. (2013), (Kochkina et al. 2018)
NewsFN-2014	Fake news detection	✓	221	5	✓	Nan et al. (2015), (Vlachos and Riedel 2014)
PolitiFact	Fake news detection	✓	488	2	✓	Bathla et al. (2018), Horne and Adali (2017)
Weibo	Rumor detection	✓	816	2	✓	Ma et al. (2016)
YelpChi	Fake review detection	✓	67K	2	✓	Mukherjee et al. (2013)
YelpNYC	Spam detection	✓	359K	2	✓	Rayana and Akoglu (2015)
YelpZip	Spam detection	✓	608K	2	✓	Rayana and Akoglu (2015)
Twitter dataset	Spam detection	✓	5.5M	2	✓	Concone et al. (2019)
KaggleEmergent[a]	Rumor detection	✓	2145	3	✓	
KaggleFN[b]	Fake news detection	✓	13K	1	✓	
FacebookHoax[c]	Hoax detection	✓	15.5K	2	✓	Tacchini et al. (2017)
BuzzfeedNews	Misleading detection	✓	2282	4	✓	Silverman et al. (2016)
Enron emai[d]	Disinformation detection	✓	.5M	2	✓	Dhamani et al. (2019)
Fraudulent emai[e]	Disinformation detection	✓	2500	2	✓	Dhamani et al. (2019)
Italian dataset	Disinformation detection	✓	160K	2	✓	Pierri et al. (2020)

[a] https://www.kaggle.com/arminehn/rumor-citation

[b] https://www.kaggle.com/mrisdal/fake-news

[c] https://github.com/gabll/some-like-it-hoax/tree/master/dataset

[d] https://www.kaggle.com/wcukierski/enron-email-dataset

[e] https://www.kaggle.com/rtatman/fraudulent-email-corpus

tion, disinformation, troll, etc.) and detection (Wu et al. 2019; Zhang et al. 2016).

- Second, deep learning architecture can be easily adapted to a new problem, e.g., using CNNs, RNNs, or LSTM, GAN, DBN, etc., which is valuable for MID.
- Third, deep learning techniques are highly flexibile especially with the advent of much popular deep learning frameworks such as Tensorflow, Keras, Caffe, PyTorch, and Theano.
- Fourth, deep learning techniques can deal with complex interaction patterns and precisely reflect users' preferences.

4.1 Existing dataset

The establishment of unique solutions for MID has often been dependent on limited and quality datasets. Therefore, to encourage future research work, we highlighted some recent and quality datasets related to the misinformation task in Table 4. Such datasets are needed to understand the reasons for applying DL techniques to MID and collectively improve the state-of-the-art. Although several established techniques have been used for MID in different domains, they are not similar to each other. Due to various research directions, the data collections could vary significantly. Moreover, the available data resource from existing research work is also hard for the collection. For instance, some datasets mainly focus on personal issues while others consist of political, business, and socially relevant issues. Additionally, datasets may vary depending on what sorts of text contents are incorporated, what labels are given, how labels are gathered, whether fraudulent information is recorded, etc.

4.2 Open issues

In this section, we summarise some limitations which we identified and proposed some ideas to address these limitations:

Semantics Understanding Misinformation which is fabricated or manipulated to mislead users. It is very difficult for

a machine to completely understand such semantics. Existing studies (Shu et al. 2019a; Braşoveanu and Andonie 2019) for MID covers various kinds of language styles. However, the understanding of semantic features is necessary to distinguish between different weapons and to improve the performance of MID.

Multimodal Data for Misinformation In the existing literature, there are several studies related to MID such as rumors detection, fake news detection, and spam detection based on multi-modal features are exist. According to the previous studies (Wang et al. 2018; Farajtabar et al. 2017; Jin et al. 2017b), we specify that misinformation on social media takes the form of text, images, or videos and the information in different modalities can provide clues for MID. However, how to extract these prominent features from each modality is challenging. Also, comprehensive and large-scale datasets are needed for MID.

Content Validation Due to misinformation, users are often confronted with misleading, confusing, controversial issues that need to be addressed very well. However, it is also true that too beautifully identifiable misinformation is difficult. Therefore, to easily identify incorrect information on online social media, we need a very good quality fact-checker and a special tool for crowdsourcing content validation can be developed

Spreader Identification Identifying the influential spreaders in social networks is a very important topic, which is conducive to deeply understand the role of nodes in information diffusion and epidemic spreading among a population. However, existing techniques are not able to quantify the nodal spreading capability correctly nor can they differentiate the influence of various nodes.

Misinformation Identification At present, many types of misidentification methods have been introduced in the existing research. However, most of the research works (a) tend to focus on alerting users but give no explanation as to why this is misinformation; (b) focus more on directly engaged users for the detection of misinformation. But if the users are not directly related, some users play an effective role in spreading misinformation on online social media. As they are not directly related, identifying them is a very difficult job.

Anomalous and Normal User Identification As the number of people who depend on online social media are growing, dishonest users try to exploit this opportunity. In most cases, dishonest people do this for their benefit (Zhao et al. 2014; Feng and Hirst 2013). Although researchers have used many methods to identify dishonest users, many more approaches can be investigated, for example, perhaps a new technique or modified version of an existing technique could be developed.

Bridging Echo Chamber Social media echo chambers play an important rule in spreading the presence of misinformation. One of the strategies for MID is to bridge conflicting echo chambers so that opposing opponents can be exchanged and considered. Therefore, data-driven models are an effective means which are needed to bridge these echo chambers. Also, researchers need to research to reduce the polarization effectiveness.

Mining Disinformation The widespread of disinformation can cause detrimental societal effects. Therefore, mining disinformation is desired to prevent a large number of people to be affected. From the discussion of existing studies, we find a recent improvement for disinformation in SN. However, due to its diversity, complexity, multi-modality, and costs of fact-checking, it is still non-trivial. Additionally, it is often unrealistic to obtain abundant labeled data. Existing studies argued that due to overfitting on small labeled datasets, the performance is largely limited (Wei and Wan 2017; Kim et al. 2018). In addition, models learned on one domain may be biased and might not perform well on a different target domain. Therefore, advanced DL strategies such as reinforcement learning can be utilized to tackle this problem, explore more information and better detect disinformation.

Misinformation Dynamic The spread of misinformation on social networking sites mainly depends on the content of the information, the impact of the users' behavior, and the network structure. Most studies on misinformation analyzed the various effects of static data but have not analyzed the effects of topology on real-time data (Wei and Wan 2017; Kim et al. 2018). Therefore, we need to consider dynamic model to capture the uncertainty of user behavior to reduce the spread of fake news and misinformation.

4.3 Future direction

As with anything, there are both good and bad aspects of technology dependence. Spreading misinformation in SN is one such example. There have been a lot of research works on MID, and good results have been achieved by various effective techniques. However, we have to keep in mind that the current age is knowledge-based and technology-dependent. Therefore, researchers have to think deeply about how their research can transform people's wellbeing in a technology-dependent era. Thus, in this paper, we have discussed some of the effective roles that elimination of misinformation can have on online SN with DL techniques. Moreover, we focused on the impact, characteristics, and detection of misinformation using DL techniques. In summary, the following are several findings of this article and possible future works:

- One of the important tasks of DL is that it can work with large-sized data which the other techniques cannot. However, DL also has difficulty to find and process massive datasets, and generally to train the model, DL networks require a lot of time. In today's competitive age, it is

worthwhile to research how to train large data in a short time with DL. We believe DL should be investigated in the future.

- Most current studies show that researchers can analyze static data on a given topic and predict the positive or negative aspects of that topic. However, it is high time to analyze dynamic data.
- The practice of detecting false facts from SN data is very popular and is benefiting people greatly. However, this involves descriptive research which is explainable MID, not just predictive research. With MID, if a new part can be added such as the description of why it is false, then maybe that research will be even more effective and acceptable to people.
- Deep reinforcement learning is a new area of machine learning that enables an agent to learn in a good interactive environment by experimenting with feedback from it own experience. So, if we combine reinforcement learning with DL to detect false facts, then better results can be obtained.
- Deep learning faces the over-fitting problem which impacts the execution of the model in real-life situations.

5 Conclusion

In this survey, we reviewed various research works on MID in social networks. In particular, we took a comprehensive view of five related terms of misinformation: false information, rumor, spam, fake news, and disinformation and discussed how misinformation misleads people on social networks. We also discussed the importance of earlier works as this may be helpful to other researchers who wish to investigate this area. Comparing with the most existing detection approaches, we considered DL is an efficient and effective technique to measure the misinformation problem on online social networks. We emphasized that DL is now the leading technique to solve MID problems because it helps in identifying false facts perfectly. The result and performance are excellent and it is similar to human performance. In all respects, DL is one of the best techniques to analyze social network data. We also demonstrated that DL can be utilized to improve MID given unlabeled and imbalanced data. However, there are several challenges (data volume, data quality, domain complexity, interpretability, explainability, feature enrichment, federated inference, model privacy, incorporating expert knowledge, temporal modeling, etc.) which need to be improved in further research work. Therefore, deep learning-based MID is still an active research topic and needs to be extended in future research. This paper can benefit others who will choose to investigate DL models.

Existing research works used DL and have done extensive work for MID. Most of the existing research works have discussed the connection of one user to another in SN, their historical activities, etc. in spreading false facts. However, there are very few studies that have incorporated user mental health conditions with the user's historical activity. Although in the above section, we introduced some future directions, one of the main future directions of our research is to expand modeling. Therefore, First, we can analyze user connections and their historical activity on SN where the user can reflect how they relate to the spread of fake news. Second, we can incorporate the human mental condition with the user's historical data, which can better analyze the user's activity, since the tendency to spread false information is related to the user's human mental condition.

References

Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, Corrado GS, Davis A, Dean J, Devin M, et al. (2015) Tensorflow: large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow. org 1

Abdel-Hamid O, Mohamed, A.r., Jiang, H., Deng, L., Penn, G., Yu, D., (2014) Convolutional neural networks for speech recognition. IEEE/ACM Trans Audio Speech Lang Process 22:1533–1545

Acquisti A, Gross R (2009) Predicting social security numbers from public data. Proc Nat Acad Sci 106:10975–10980

Aiello LM, Petkos G, Martin C, Corney D, Papadopoulos S, Skraba R, Göker A, Kompatsiaris I, Jaimes A (2013) Sensing trending topics in twitter. IEEE Trans Multimedia 15:1268–1282

Aizenberg IN (1999) Neural networks based on multi-valued and universal binary neurons: theory, application to image processing and recognition. In: International conference on computational intelligence, Springer, pp 306–316

Alkhodair SA, Ding SH, Fung BC, Liu J (2020) Detecting breaking news rumors of emerging topics in social media. Inf Process Manag 57:102018

Alom MZ, Bontupalli V, Taha TM (2015) Intrusion detection using deep belief networks. In: 2015 national aerospace and electronics conference (NAECON), IEEE, pp 339–344

Bathla G, Aggarwal H, Rani R (2018) Improving recommendation techniques by deep learning and large scale graph partitioning. Int J Adv Comput Sci Appl 9:403–409

Bharti SK, Pradhan R, Babu KS, Jena SK (2017) Sarcasm analysis on twitter data using machine learning approaches. In: Trends in social network analysis. Springer, pp 51–76

Bindu P, Thilagam PS, Ahuja D (2017) Discovering suspicious behavior in multilayer social networks. Comput Hum Behav 73:568–582

Brandt AM (2012) Inventing conflicts of interest: a history of tobacco industry tactics. Am J Public Health 102:63–71

Braşoveanu AM, Andonie R (2019) Semantic fake news detection: a machine learning perspective. In: International work-conference on artificial neural networks, Springer, pp 656–667

Chalapathy R, Chawla S (2019) Deep learning for anomaly detection: a survey. arXiv preprint arXiv:1901.03407

Chen T, Li X, Yin H, Zhang J (2018) Call attention to rumors: Deep attention based recurrent neural networks for early rumor detection. In: Pacific-Asia conference on knowledge discovery and data mining, Springer, pp 40–52

Chen YC, Liu ZY, Kao HY (2017) Ikm at semeval-2017 task 8: vonvolutional neural networks for stance detection and rumor

verification. In: Proceedings of the 11th international workshop on semantic evaluation (SemEval-2017), pp 465–469
Cho K, van Merriënboer B, Gulcehre C, Bahdanau D, Bougares F, Schwenk H, Bengio Y, (2014a) Learning phrase representations using RNN encoder–decoder for statistical machine translation. In: Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), Association for Computational Linguistics, Doha, Qatar. pp 1724–1734. https://www.aclweb.org/anthology/D14-1179, 10.3115/v1/D14-1179
Cho K, Van Merriënboer B, Gulcehre C, Bahdanau D, Bougares F, Schwenk H, Bengio Y (2014b) Learning phrase representations using rnn encoder-decoder for statistical machine translation. arXiv preprint arXiv:1406.1078
Chollet F (2018) Deep learning mit Python und Keras: Das Praxis-Handbuch vom Entwickler der Keras-Bibliothek. MITP-Verlags GmbH & Co, KG
Çıtlak O, Dörterler M, Doğru İA (2019) A survey on detecting spam accounts on twitter network. Soc Netw Anal Min 9:35
Collobert R, Bengio S, Mariéthoz J (2002) Torch: a modular machine learning software library. Technical Report, Idiap
Collobert R, Kavukcuoglu K, Farabet C (2011) Torch7: a matlab-like environment for machine learning. In: BigLearn, NIPS workshop
Concone F, Re GL, Morana M, Ruocco C (2019) Twitter spam account detection by effective labeling. In: ITASEC
Cresci S, Di Pietro R, Petrocchi M, Spognardi A, Tesconi M (2015) Fame for sale: efficient detection of fake twitter followers. Decis Support Syst 80:56–71
Cui L, Wang S, Lee D (2019) Same: sentiment-aware multi-modal embedding for detecting fake news. In: Proceedings of the 2019 IEEE/ACM international conference on advances in social networks analysis and mining, pp 41–48
Da Silva LA, Da Costa KA, Papa JP, Rosa G, De Albuquerque VHC (2018) Fine-tuning restricted boltzmann machines using quaternions and its application for spam detection. IET Netw 8:101–105
Dai JJ, Wang Y, Qiu X, Ding D, Zhang Y, Wang Y, Jia X, Zhang LC, Wan Y, Li Z, Wang J, Huang S, Wu Z, Wang Y, Yang Y, She B, Shi D, Lu Q, Huang K, Song G (2019) Bigdl: A distributed deep learning framework for big data. In: Proceedings of the ACM symposium on cloud computing, Association for Computing Machinery. pp 50–60. https://arxiv.org/pdf/1804.05839.pdf, 10.1145/3357223.3362707
Dandekar A, Zen RA, Bressan S (2017) Generating fake but realistic headlines using deep neural networks. In: International conference on database and expert systems applications, Springer, pp 427–440
David OE, Netanyahu NS (2015) Deepsign: Deep learning for automatic malware signature generation and classification. In: 2015 international joint conference on neural networks (IJCNN), IEEE, pp 1–8
De Choudhury M, Counts S, Horvitz E (2013a) Predicting postpartum changes in emotion and behavior via social media. In: Proceedings of the SIGCHI conference on human factors in computing systems, ACM, pp 3267–3276
De Choudhury M, Gamon M, Hoff A, Roseway A (2013b) âĂIJmoon phrasesâĂİ: a social media faciliated tool for emotional reflection and wellness. In: 2013 7th international conference on pervasive computing technologies for healthcare and workshops, IEEE, pp 41–44
Dechter R (1986) Learning while searching in constraint-satisfaction problems. University of California, Computer Science Department, Cognitive Systems âĂę
Dhamani N, Azunre P, Gleason JL, Corcoran C, Honke G, Kramer S, Morgan J (2019) Using deep networks and transfer learning to address disinformation. arXiv preprint arXiv:1905.10412
Donfro J (2013) A whopping 20% of yelp reviews are fake
Du M, Li F, Zheng G, Srikumar V (2017) Deeplog: Anomaly detection and diagnosis from system logs through deep learning. In: Proceedings of the 2017 ACM SIGSAC conference on computer and communications security, ACM, pp 1285–1298
Fallis D (2014) A functional analysis of disinformation. In: Conference 2014 proceedings
Farajtabar M, Yang J, Ye X, Xu H, Trivedi R, Khalil E, Li S, Song L, Zha H (2017) Fake news mitigation via point process based intervention. In: Proceedings of the 34th international conference on machine learning, vol 70, pp 1097–1106
Feng VW, Hirst G (2013) Detecting deceptive opinions with profile compatibility. In: Proceedings of the sixth international joint conference on natural language processing, pp 338–346
Fernandez M, Alani H (2018) Online misinformation: challenges and future directions. Companion Proc Web Conf 2018:595–602
Friggeri A, Adamic L, Eckles D, Cheng J (2014) Rumor cascades. In: Eighth international AAAI conference on weblogs and social media
Galitsky B (2015) Detecting rumor and disinformation by web mining. In: 2015 AAAI Spring symposium series
Gao H, Liu H (2014) Data analysis on location-based social networks. Mobile Soc Netw 11:165–194
Gong M, Gao Y, Xie Y, Qin A (2020) An attention-based unsupervised adversarial model for movie review spam detection. In: IEEE transactions on multimedia
Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y (2014) Generative adversarial nets. In: Advances in neural information processing systems, pp 2672–2680
Goswami A, Kumar A (2016) A survey of event detection techniques in online social networks. Soc Netw Anal Min 6:107
Guo B, Ding Y, Yao L, Liang Y, Yu Z (2019) The future of misinformation detection: new perspectives and trends. arXiv preprint arXiv:1909.03654
Guo H, Cao J, Zhang Y, Guo J, Li J (2018) Rumor detection with hierarchical social attention network. In: Proceedings of the 27th ACM international conference on information and knowledge management, ACM, pp 943–951
Gupta A, Kumaraguru P, Castillo C, Meier P (2014) Tweetcred: Real-time credibility assessment of content on twitter. In: International Conference on Social Informatics, Springer, pp 228–243
Gupta A, Lamba H, Kumaraguru P, Joshi A (2013) Faking sandy: characterizing and identifying fake images on twitter during hurricane sandy. In: Proceedings of the 22nd international conference on World Wide Web, ACM, pp 729–736
Habib A, Asghar MZ, Khan A, Habib A, Khan A (2019) False information detection in online content and its role in decision making: a systematic literature review. Soc Netw Anal Min 9:50
Hardy W, Chen L, Hou S, Ye Y, Li X (2016) Dl4md: a deep learning framework for intelligent malware detection. In: Proceedings of the international conference on data mining (DMIN), the steering committee of the world congress in computer science, computer âĂę, p 61
Helmstetter S, Paulheim H (2018) Weakly supervised learning for fake news detection on twitter. In: 2018 IEEE/ACM international conference on advances in social networks analysis and mining (ASONAM), IEEE, pp 274–277
Hernon P (1995) Disinformation and misinformation through the internet: findings of an exploratory study. Gov Inf Q 12:133–139
Hinton G, Deng L, Yu D, Dahl G, Mohamed Ar, Jaitly N, Senior A, Vanhoucke V, Nguyen P, Kingsbury B, et al. (2012) Deep neural networks for acoustic modeling in speech recognition. In: IEEE Signal processing magazine, p 29
Horne BD, 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. In: Eleventh international AAAI conference on web and social media
Hu X, Tang J, Zhang Y, Liu H (2013) Social spammer detection in microblogging. In: Twenty-third international joint conference on artificial intelligence
Islam MR, Kabir MA, Ahmed A, Kamal ARM, Wang H, Ulhaq A (2018a) Depression detection from social network data using machine learning techniques. Health Inf Sci Syst 6:8
Islam MR, Kamal ARM, Sultana N, Islam R, Moni MA, et al. (2018b) Detecting depression using k-nearest neighbors (knn) classification technique. In: 2018 international conference on computer, communication, chemical, material and electronic engineering (IC4ME2), IEEE, pp 1–4
Islam MR, Miah SJ, Kamal ARM, Burmeister O (2019) A design construct of developing approaches to measure mental health conditions. In: Australasian journal of information systems, p 23
Jacovi A, Shalom OS, Goldberg Y (2018) Understanding convolutional neural networks for text classification. arXiv preprint arXiv :1809.08037
Jain S, Sharma V, Kaushal R (2016) Towards automated real-time detection of misinformation on twitter. In: 2016 international conference on advances in computing. Communications and informatics (ICACCI), IEEE, pp 2015–2020
Jia Y, Shelhamer E, Donahue J, Karayev S, Long J, Girshick R, Guadarrama S, Darrell T (2014) Caffe: convolutional architecture for fast feature embedding. In: Proceedings of the 22nd ACM international conference on Multimedia, ACM, pp 675–678
Jia Y, Song X, Zhou J, Liu L, Nie L, Rosenblum DS (2016) Fusing social networks with deep learning for volunteerism tendency prediction. In: Thirtieth AAAI conference on artificial intelligence
Jin D, Ge M, Li Z, Lu W, He D, Fogelman-Soulie F (2017a) Using deep learning for community discovery in social networks. In: 2017 IEEE 29th international conference on tools with artificial intelligence (ICTAI), IEEE, pp 160–167
Jin Z, Cao J, Guo H, Zhang Y, Luo J (2017b) Multimodal fusion with recurrent neural networks for rumor detection on microblogs. In: Proceedings of the 25th ACM international conference on multimedia, pp 795–816
Jin Z, Cao J, Guo H, Zhang Y, Wang Y, Luo J (2017c) Detection and analysis of 2016 us presidential election related rumors on twitter. In: International conference on social computing, behavioral-cultural modeling and prediction and behavior representation in modeling and simulation, Springer, pp 14–24
Jin Z, Cao J, Zhang Y, Zhou J, Tian Q (2016) Novel visual and statistical image features for microblogs news verification. IEEE Trans Multimedia 19:598–608
Jindal S, Sood R, Singh R, Vatsa M, Chakraborty T (xxxx) Newsbag: a multimodal benchmark dataset for fake news detection
Ketkar N (2017) Introduction to pytorch. In: Deep learning with python. Springer, pp 195–208
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. In: Proceedings of the eleventh ACM international conference on web search and data mining, pp 324–332
Kim Y (2014) Convolutional neural networks for sentence classification. arXiv preprint arXiv:1408.5882
King DE (2009) Dlib-ml: a machine learning toolkit. J Mach Learn Res 10:1755–1758
Kochkina E, Liakata M, Zubiaga A (2018) All-in-one: Multi-task learning for rumour verification. arXiv preprint arXiv:1806.03713
Kudugunta S, Ferrara E (2018) Deep neural networks for bot detection. Inf Sci 467:312–322
Kumar S, Asthana R, Upadhyay S, Upreti N, Akbar M (2019) Fake news detection using deep learning models: a novel approach. Trans Emerg Telecommun Technol 5:e3767
Kumar S, Shah N (2018) False information on web and social media: a survey. arXiv preprint arXiv:1804.08559
Kumar S, West R, Leskovec J (2016) Disinformation on the web: Impact, characteristics, and detection of wikipedia hoaxes. In: Proceedings of the 25th international conference on World Wide Web, pp 591–602
Kwon S, Cha M, Jung K (2017) Rumor detection over varying time windows. PloS ONE 12:69
Lake JM (2014) Fake web addresses and hyperlinks. US Patent 8,799,465
LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521:436
LeCun Y, Kavukcuoglu K, Farabet C (2010) Convolutional networks and applications in vision. In: Proceedings of 2010 IEEE international symposium on circuits and systems, IEEE, pp 253–256
Li C, Liu S (2018) A comparative study of the class imbalance problem in twitter spam detection. Concurr Comput Pract Exp 30:e4281
Li L, Cai G, Chen N (2018a) A rumor events detection method based on deep bidirectional gru neural network. In: 2018 IEEE 3rd international conference on image. Vision and computing (ICIVC), IEEE, pp 755–759
Li X, Wu X (2015) Constructing long short-term memory based deep recurrent neural networks for large vocabulary speech recognition. In: 2015 IEEE international conference on acoustics. Speech and signal processing (ICASSP), IEEE, pp 4520–4524
Li Y, Nie X, Huang R (2018b) Web spam classification method based on deep belief networks. Expert Syst Appl 96:261–270
Liao L, Jin W, Pavel R (2016) Enhanced restricted boltzmann machine with prognosability regularization for prognostics and health assessment. IEEE Trans Industr Electron 63:7076–7083
Lin X, Liao X, Xu T, Pian W, Wong KF (2019) Rumor detection with hierarchical recurrent convolutional neural network. In: CCF international conference on natural language processing and chinese computing, Springer, pp 338–348
Liu Q, Yu F, Wu S, Wang L (2018) Mining significant microblogs for misinformation identification: an attention-based approach. ACM Trans Intell Syst Technol 9:1–20
Liu Y, Wu YFB (2018) Early detection of fake news on social media through propagation path classification with recurrent and convolutional networks. In: Thirty-second AAAI conference on artificial intelligence
Liu Y, Wu YFB (2020) Fned: a deep network for fake news early detection on social media. ACM Trans Inf Syst 38:1–33
Liu Y, Xu S (2016) Detecting rumors through modeling information propagation networks in a social media environment. IEEE Trans Comput Soc Syst 3:46–62
Long Y, Lu Q, Xiang R, Li M, Huang CR (2017) Fake news detection through multi-perspective speaker profiles. In: Proceedings of the eighth international joint conference on natural language processing, vol 2, pp 252–256
Ma J, Gao W, Mitra P, Kwon S, Jansen BJ, Wong KF, Cha M (2016) Detecting rumors from microblogs with recurrent neural networks. In: Ijcai, pp 3818–3824
Ma J, Gao W, Wei Z, Lu Y, Wong KF (2015) Detect rumors using time series of social context information on microblogging websites. In: Proceedings of the 24th ACM international on conference on information and knowledge management, pp 1751–1754
Ma J, Gao W, Wong KF (2018) Rumor detection on twitter with tree-structured recursive neural networks. In: Proceedings of the 56th annual meeting of the association for computational linguistics, vol 1, pp 1980–1989

Ma J, Gao W, Wong KF (2019) Detect rumors on twitter by promoting information campaigns with generative adversarial learning. In: The World Wide Web Conference, ACM, pp 3049–3055

Markines B, Cattuto C, Menczer F (2009) Social spam detection. In: Proceedings of the 5th international workshop on adversarial information retrieval on the web, pp 41–48

Mitra T, Gilbert E (2015) Credbank: A large-scale social media corpus with associated credibility annotations. In: ICWSM, pp 258–267

Mukherjee A, Venkataraman V, Liu B, Glance NS (2013) What yelp fake review filter might be doing? In: Icwsm, pp 409–418

Nan CJ, Kim KM, Zhang BT (2015) Social network analysis of tv drama characters via deep concept hierarchies. In: 2015 IEEE/ACM international conference on advances in social networks analysis and mining (ASONAM), IEEE, pp 831–836

Naseem U, Razzak I, Musial K, Imran M (2020) Transformer based deep intelligent contextual embedding for twitter sentiment analysis. Future Gener Comput Syst 6:91

Nguyen DT, Al Mannai KA, Joty S, Sajjad H, Imran M, Mitra P (2017a) Robust classification of crisis-related data on social networks using convolutional neural networks. In: Eleventh international AAAI conference on web and social media

Nguyen TN, Li C, Niederée C (2017b) On early-stage debunking rumors on twitter: Leveraging the wisdom of weak learners. In: International conference on social informatics, Springer, pp 141–158

Norouzi M (2009) Convolutional restricted Boltzmann machines for feature learning. Ph.D. thesis. School of Computing Science-Simon Fraser University

Norouzi M, Ranjbar M, Mori G (2009) Stacks of convolutional restricted boltzmann machines for shift-invariant feature learning. In: 2009 IEEE conference on computer vision and pattern recognition, IEEE, pp 2735–2742

Papa JP, Rosa GH, Marana AN, Scheirer W, Cox DD (2015) Model selection for discriminative restricted boltzmann machines through meta-heuristic techniques. J Comput Sci 9:14–18

Parvat A, Chavan J, Kadam S, Dev S, Pathak V (2017) A survey of deep-learning frameworks. In: 2017 international conference on inventive systems and control (ICISC), IEEE, pp 1–7

Perozzi B, Al-Rfou R, Skiena S (2014) Deepwalk: Online learning of social representations. In: Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, pp 701–710

Pierri F, Piccardi C, Ceri S (2020) A multi-layer approach to disinformation detection on twitter. arXiv preprint arXiv:2002.12612

Popat K, Mukherjee S, Yates A, Weikum G (2018) Declare: debunking fake news and false claims using evidence-aware deep learning. arXiv preprint arXiv:1809.06416

Pouyanfar S, Chen SC (2016) Semantic event detection using ensemble deep learning. In: 2016 IEEE international symposium on multimedia (ISM), IEEE, pp 203–208

Pouyanfar S, Sadiq S, Yan Y, Tian H, Tao Y, Reyes MP, Shyu ML, Chen SC, Iyengar S (2018) A survey on deep learning: algorithms, techniques, and applications. ACM Comput Surv 51:92

Prova AA, Akter T, Islam MR, Uddin MR, Hossain T, Hannan M, Hossain MS (2019) Analysis of online marketplace data on social networks using lstm. In: 2019 5th international conference on advances in electrical engineering (ICAEE), IEEE, pp 381–385

Qazvinian V, Rosengren E, Radev DR, Mei Q (2011) Rumor has it: Identifying misinformation in microblogs. In: Proceedings of the conference on empirical methods in natural language processing, Association for. Computational Linguistics, pp 1589–1599

Qian F, Gong C, Sharma K, Liu Y (2018) Neural user response generator: Fake news detection with collective user intelligence. In: IJCAI, pp 3834–3840

Quah JT, Sriganesh M (2008) Real-time credit card fraud detection using computational intelligence. Expert Syst Appl 35:1721–1732

Rayana S, Akoglu L (2015). Collective opinion spam detection: Bridging review networks and metadata. In: Proceedings of the 21th acm sigkdd international conference on knowledge discovery and data mining, pp 985–994

Roy A, Basak K, Ekbal A, Bhattacharyya P (2018) A deep ensemble framework for fake news detection and classification. arXiv preprint arXiv:1811.04670

Ruchansky N, Seo S, Liu Y (2017) Csi: A hybrid deep model for fake news detection. In: Proceedings of the 2017 ACM on conference on information and knowledge management, ACM, pp 797–806

Saberi A, Vahidi M, Bidgoli BM (2007) Learn to detect phishing scams using learning and ensemble? methods. In: 2007 IEEE/WIC/ACM international conferences on web intelligence and intelligent agent technology-workshops, IEEE, pp 311–314

Sampson J, Morstatter F, Wu L, Liu H (2016) Leveraging the implicit structure within social media for emergent rumor detection. In: Proceedings of the 25th ACM international on conference on information and knowledge management, pp 2377–2382

Savage D, Zhang X, Yu X, Chou P, Wang Q (2014) Anomaly detection in online social networks. Soc Netw 39:62–70

Selvaganapathy S, Nivaashini M, Natarajan H (2018) Deep belief network based detection and categorization of malicious urls. Inf Secur J Global Perspect 27:145–161

Shahariar G, Biswas S, Omar F, Shah FM, Hassan SB (2019) Spam review detection using deep learning. In: 2019 IEEE 10th annual information technology, electronics and mobile communication conference (IEMCON), IEEE, pp 0027–0033

Sharma K, Qian F, Jiang H, Ruchansky N, Zhang M, Liu Y (2019) Combating fake news: a survey on identification and mitigation techniques. arXiv preprint arXiv:1901.06437

Shi S, Wang Q, Chu X (2018) Performance modeling and evaluation of distributed deep learning frameworks on gpus. In: 2018 IEEE 16th international conference on dependable, autonomic and secure computing, 16th international conference on pervasive intelligence and computing, 4th international conference on big data intelligence and computing and cyber science and technology congress (DASC/PiCom/DataCom/CyberSciTech), IEEE, pp 949–957

Shu K, Cui L, Wang S, Lee D, Liu H (2019a) defend: Explainable fake news detection. In: Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery and data mining, pp 395–405

Shu K, Mahudeswaran D, Wang S, Lee D, Liu H (2018) Fakenewsnet: a data repository with news content, social context and dynamic information for studying fake news on social media. arXiv preprint arXiv:1809.01286

Shu K, Sliva A, Wang S, Tang J, Liu H (2017) Fake news detection on social media: a data mining perspective. ACM SIGKDD Explor Newslett 19:22–36

Shu K, Wang S, Lee D, Liu H (2020) Mining disinformation and fake news: concepts, methods, and recent advancements. arXiv preprint arXiv:2001.00623

Shu K, Wang S, Liu H (2019b) Beyond news contents: The role of social context for fake news detection. In: Proceedings of the twelfth ACM international conference on web search and data mining, pp 312–320

Shu K, Zhou X, Wang S, Zafarani R, Liu H (2019c) The role of user profiles for fake news detection. In: Proceedings of the 2019 IEEE/ACM international conference on advances in social networks analysis and mining, pp 436–439

Silva L, Ribeiro P, Rosa G, Costa K, Papa JP (2015) Parameter setting-free harmony search optimization of restricted boltzmann

machines and its applications to spam detection. In: 12th international conference applied computing, pp 142–150
da Silva LA, da Costa KAP, Ribeiro PB, de Rosa GH, Papa JP (2016) Learning spam features using restricted boltzmann machines. IADIS International Journal on Computer Science & Information Systems 11
Silverman C, Strapagiel L, Shaban H, Hall E, Singer-Vine J (2016) Hyperpartisan facebook pages are publishing false and misleading information at an alarming rate. Buzzfeed News 20:68
Socher R, Perelygin A, Wu J, Chuang J, Manning CD, Ng A, Potts C (2013) Recursive deep models for semantic compositionality over a sentiment treebank. In: Proceedings of the 2013 conference on empirical methods in natural language processing, pp 1631–1642
Srivastava N, Salakhutdinov RR (2012) Multimodal learning with deep boltzmann machines. In: Advances in neural information processing systems, pp 2222–2230
Sun X, Zhang C, Ding S, Quan C (2018) Detecting anomalous emotion through big data from social networks based on a deep learning method. Multimedia Tools Appl 5:1–22
Tacchini E, Ballarin G, Della Vedova ML, Moret S, de Alfaro L (2017) Some like it hoax: automated fake news detection in social networks. arXiv preprint arXiv:1704.07506
Tang D, Qin B, Liu T (2015) Document modeling with gated recurrent neural network for sentiment classification. In: Proceedings of the 2015 conference on empirical methods in natural language processing, pp 1422–1432
Tschiatschek S, Singla A, Gomez Rodriguez M, Merchant A, Krause A (2018) Fake news detection in social networks via crowd signals. In: Companion of the the web conference 2018 on the web conference 2018, international world wide web conferences steering committee, pp 517–524
Tsui D (2017) Predicting stock price movement using social media analysis. Technical Report. Stanford University, Technical Report
Tzortzis G, Likas A (2007) Deep belief networks for spam filtering. In: 19th IEEE international conference on tools with artificial intelligence (ICTAI 2007), IEEE, pp 306–309
Van Merriënboer B, Bahdanau D, Dumoulin V, Serdyuk D, Warde-Farley D, Chorowski J, Bengio Y (2015) Blocks and fuel: frameworks for deep learning. arXiv preprint arXiv:1506.00619
Vartapetiance A, Gillam L (2014) Deception detection: dependable or defective? Soc Netw Anal Min 4:166
Vlachos A, Riedel S (2014) Fact checking: Task definition and dataset construction. In: Proceedings of the ACL 2014 workshop on language technologies and computational social science, pp 18–22
Vo NN, He X, Liu S, Xu G (2019) Deep learning for decision making and the optimization of socially responsible investments and portfolio. Decis Support Syst 124:113097
Vo NN, Liu S, He X, Xu G (2018) Multimodal mixture density boosting network for personality mining. In: Pacific-Asia conference on knowledge discovery and data mining, Springer. pp 644–655
Wang D, Irani D, Pu C (2011) A social-spam detection framework. In: Proceedings of the 8th annual collaboration, electronic messaging, anti-abuse and Spam conference, pp 46–54
Wang N, Yeung DY (2013) Learning a deep compact image representation for visual tracking. In: Advances in neural information processing systems, pp 809–817
Wang W, Zhang F, Luo X, Zhang S (2019) Pdrcnn: precise phishing detection with recurrent convolutional neural networks. Secur Commun Netw 9:72
Wang WY (2017) " liar, liar pants on fire": a new benchmark dataset for fake news detection. arXiv preprint arXiv:1705.00648
Wang Y, Ma F, Jin Z, Yuan Y, Xun G, Jha K, Su L, Gao J (2018) Eann: Event adversarial neural networks for multi-modal fake news detection. In: Proceedings of the 24th ACM sigkdd international conference on knowledge discovery and data mining, pp 849–857
Wei L, Gao D, Luo C (2018) False data injection attacks detection with deep belief networks in smart grid. In: 2018 Chinese automation congress (CAC), IEEE, pp 2621–2625
Wei W, Wan X (2017) Learning to identify ambiguous and misleading news headlines. arXiv preprint arXiv:1705.06031
Willmore A (xxxx) This analysis shows how viral fake election news stories outperformed real news on facebook
Wu L, Li J, Hu X, Liu H (2017) Gleaning wisdom from the past: Early detection of emerging rumors in social media. In: Proceedings of the 2017 SIAM international conference on data mining, SIAM, pp 99–107
Wu L, Morstatter F, Carley KM, Liu H (2019) Misinformation in social media: definition, manipulation, and detection. ACM SIGKDD Explor Newslett 21:80–90
Wu L, Rao Y, Yu H, Wang Y, Nazir A (2018) False information detection on social media via a hybrid deep model. In: International conference on social informatics, Springer, pp 323–333
Xu Y, Wang C, Dan Z, Sun S, Dong F (2019) Deep recurrent neural network and data filtering for rumor detection on sina weibo. Symmetry 11:1408
Yang Y, Zheng L, Zhang J, Cui Q, Li Z, Yu PS (2018) Ti-cnn: Convolutional neural networks for fake news detection. arXiv preprint arXiv:1806.00749
Yenala H, Jhanwar A, Chinnakotla MK, Goyal J (2018) Deep learning for detecting inappropriate content in text. Int J Data Sci Anal 6:273–286
Yepes AJ, MacKinlay A, Bedo J, Garvani R, Chen Q (2014) Deep belief networks and biomedical text categorisation. In: Proceedings of the Australasian language technology association. Workshop, pp 123–127
Yilmaz CM, Durahim AO (2018) Spr2ep: a semi-supervised spam review detection framework. In: 2018 IEEE/ACM international conference on advances in social networks analysis and mining (ASONAM), IEEE, pp 306–313
Yin J, Zhou Z, Liu S, Wu Z, Xu G (2018) Social spammer detection: a multi-relational embedding approach. In: Pacific-Asia conference on knowledge discovery and data mining, Springer, pp 615–627
Yin J, Li Q, Liu S, Wu Z, Xu G (2020) Leveraging Multi-level Dependency of Relational Sequences for Social Spammer Detection. arXiv preprint arXiv:2009.06231
Young T, Hazarika D, Poria S, Cambria E (2018) Recent trends in deep learning based natural language processing. In: IEEE computational intelligence magazine, vol 13, pp 55–75
Yu F, Liu Q, Wu S, Wang L, Tan T et al. (2017a) A convolutional approach for misinformation identification
Yu S, Li M, Liu F (2017b) Rumor identification with maximum entropy in micronet. Complexity 2017
Zhang H, Alim MA, Li X, Thai MT, Nguyen HT (2016) Misinformation in online social networks: detect them all with a limited budget. ACM Trans Inf Syst 34:1–24
Zhang H, Kuhnle A, Smith JD, Thai MT (2018a) Fight under uncertainty: Restraining misinformation and pushing out the truth. In: 2018 IEEE/ACM international conference on advances in social networks analysis and mining (ASONAM), IEEE, pp 266–273
Zhang J, Cui L, Fu Y, Gouza FB (2018b) Fake news detection with deep diffusive network model. arXiv preprint arXiv:1805.08751
Zhang Q, Lipani A, Liang S, Yilmaz E (2019) Reply-aided detection of misinformation via bayesian deep learning. In: The world wide web conference, pp 2333–2343
Zhang Q, Zhang S, Dong J, Xiong J, Cheng X (2015) Automatic detection of rumor on social network. In: Natural language processing and Chinese computing. Springer, pp 113–122
Zhang W, Du Y, Yoshida T, Wang Q (2018c) Dri-rcnn: an approach to deceptive review identification using recurrent convolutional neural network. Inf Process Manag 54:576–592

Zhang Y, Salakhutdinov R, Chang HA, Glass J (2012) Resource configurable spoken query detection using deep boltzmann machines. In: 2012 IEEE international conference on acoustics. Speech and signal processing (ICASSP), IEEE, pp 5161–5164

Zhao J, Cao N, Wen Z, Song Y, Lin YR, Collins C (2014) Fluxflow: visual analysis of anomalous information spreading on social media. IEEE Trans Visual Comput Graphics 20:1773–1782

Zhao Z, Resnick P, Mei Q (2015) Enquiring minds: Early detection of rumors in social media from enquiry posts. In: Proceedings of the 24th international conference on world wide web, pp 1395–1405

Zhou C, Sun C, Liu Z, Lau F (2015) A c-lstm neural network for text classification. arXiv preprint arXiv:1511.08630

Zubiaga A, Aker A, Bontcheva K, Liakata M, Procter R (2018) Detection and resolution of rumours in social media: a survey. ACM Comput Surv 51:1–36

Zubiaga A, Kochkina E, Liakata M, Procter R, Lukasik M (2016a) Stance classification in rumours as a sequential task exploiting the tree structure of social media conversations. arXiv preprint arXiv:1609.09028

Zubiaga A, Liakata M, Procter R (2017) Exploiting context for rumour detection in social media. In: International conference on social informatics, Springer, pp 109–123

Zubiaga A, Liakata M, Procter R, Hoi GWS, Tolmie P (2016b) Analysing how people orient to and spread rumours in social media by looking at conversational threads. PLoS ONE 11:e0150989