Detecting Fake News on Social Media: A Multi-Source Scoring Framework

Huxiao Liu, Lianhai Wang*, Xiaohui Han*, Weinan Zhang, Xun He
Shandong Computer Science Center (National Supercomputer Center in Jinan),
Shandong Provincial Key Laboratory of Computer Networks,
Qilu University of Technology (Shandong Academy of Sciences)
Jinan, China
e-mail: liuhux@163.com, wanglh@sdas.org*, hanxh@sdas.org*, zhangwn330@gmail.com, 1091419861@qq.com

Abstract—**Social media has dramatically promoted the efficiency of news diffusion. However, as information is no longer verified by journalists or experts, it has also become a fruitful environment for fake news. Since fake news has long been a critical threat to our society, it has always been an important work for both social media sites and government agencies to combat fake news. Although a large body of research work and efforts have been focused on fake news detection in social media, most of the existing methods are single-source based, which can easily lead to subjective detection results. In this paper, we propose FNDMS, a framework that integrates the credibility scores of multiple news sources to detect fake news. FNDMS uses two sets of features, i.e., author-based features and content-based features, to measure the credibility of a single news source. Then a DST model is employed to integrate credibilities of multiple sources and produce a judgment on the truth of an event. To collect event-related reports, we also propose a three-step method to retrieve and filter news articles from social media sites. Experimental results on real social media data demonstrate the feasibility and advance of FNDMS.**

Keywords-Fake news; multi-source scoring; DST; credibil-ity; social media

I. Introduction

Social media have radically changed the way of both news consumption and distribution. In such a disintermediation environment, users can actively participate in news production and thus are both producer and consumer of news. This makes the diffusion of news more efficient via social media than through conventional news websites [1]. Consequently, social media have become an important source of news for many people. According to a recent report [2], in the year of 2018, roughly 68% of U.S. adults get news on social media sites (e.g., Facebook and Twitter). However, as information is no longer verified by journalists or experts, social media has also become a fruitful environment for fake news (or rumors). What's even worse, the ubiquity and easy access characteristics of social media can dramatically accelerate the speed of fake news spreading [1], [3]–[5].

Fake news has long been a critical threat to our society [6]–[9]. It has always been an important work for both social media companies and government agencies to combat fake news, especially to develop ingenious and automatic techniques that can help users differentiate between the truth and rumors. Although a large body of research work and efforts have been focused on fake news detection (FND) in social media, the common strategy is to evaluate the credibility of a news report based on various features extracted from its text content or the profile of its author [10]–[14]. We call these methods single-source-based methods as they focus on the properties of a single article or user. Single-source-based methods can easily lead to detection results that are heavily dependent on individual users and are not objective enough since it is typically difficult to judge whether a news article is telling the truth merely based on text words or user characteristics.

Typically, a news event can be reported by many social media users, some of whom tell people the truth, while some of the others spread rumors. Intuitively, if we collect a set of distinct articles reporting the same news event and analyze the credibility of them, we may reveal whether the event has really happened or is just a rumor by leveraging "collective intelligence". However, few previous studies have attempted to detect fake news in this way. Motivated by this, the aim of this paper is to develop an approach that can identify fake news by integrating the credibility of multiple news sources. To achieve this goal, we must find solutions to two questions: (1) How to collect articles reporting the same news event effectively? (2) How to appropriately fuse the credibilities of multiple new sources to get a decision on whether the news event has truly happened or not? For the first question, we develop a three-step method to acquire event-related reports from social media platforms. For the latter, we employ a Dempster-Shafer Theory (DST) model to integrate conflicting information from different news sources and produce a judgment on the truth of an event. We call the entire framework as **F**ake **N**ews **D**etector based on **M**ulti-source **S**coring (FNDMS). Experimental results on real social media data show that FNDMS outperforms several single-source-based approaches.

To be specific, our main contributions include the following:

- We propose identifying fake news by integrating the credibilities of multiple news sources, which is more objective than depending on clues from a single source.
- We propose a three-step method for retrieving and filtering news articles reporting the same event from social media platforms.
- We propose two sets of features, i.e., author-based fea-

978-1-7281-6024-5/20/$31.00 ©2020 IEEE

Authorized licensed use limited to: Hong Kong University of Science and Technology. Downloaded on July 19,2023 at 10:59:18 UTC from IEEE Xplore. Restrictions apply.

tures and content-based features, for evaluating the credibility of a single news source. Experimental results show that both of the two sets of features are effective in fake news detection.

The rest of this paper is arranged as follows. In Section 2, we briefly review related work on fake news detection. In Section 3, we present the details of FNDMS. In Section 4, we give the experiments on real social media data. Finally, in Section 5, we conclude this paper.

II. Related Work

Since fake news poses a serious threat to both social media sites and governments, sufficient research has been conducted in this area in recent years. Most of the existing methods evaluate the credibility of a news story based on the related user-generated content (UGC) in social media. Natural Language Processing (NLP) algorithms have been widely applied to extracting various features from UGC. For instance, Horne *et al.* [13] built an SVM classifier with four sets of features, i.e., readability features, emotional tendency features, writing style features, and syntactic features, to detect fake news. They found that the titles of fake news reports have fewer stop words and nouns but more verbs. The accuracy of their detection model can achieve 71%-91% in different scenarios. Similar to [13], Hadeer Ahmed *et al.* [14] extracted Term Frequency-Inverted Document Frequency (TF-IDF) from textual UGC, then learned a Linear Support Vector Machine (LSVM) to judge the credibility of news. Some recent methods also employ deep learning frameworks for fake news detection in social media. For example, Wang *et al.* [15] introduced a hybrid-CNN based surface language mode to detect fake news. Some other deep-learning-based methods can be found in [16]–[18]. Wang *et al.* [19] found that compared with traditional learning methods, deep-learning-based techniques can achieve better performance due to their superior feature extraction ability.

Besides UGC features, there are also FND methods using features extracted from other types of social media content. C. Castillo *et al.* [12] proposed a supervises learning method to evaluate the credibility of hot topics. They extracted features from tweets, user profiles, and tweet-propagation paths respectively, and then learned a decision tree to predict whether the topic is credible. Yang *et al.* [20] further improved C. Castillo *et al.*'s method by adding two additional features, i.e., client-based feature and location-based feature. Their experimental results indicated that the adding of new features led to a better FND performance in terms of accuracy.

A comprehensive review of existing fake news detection techniques can be found in [21]. In summary, nearly all the FND techniques mentioned above can be categorized into single-source-based methods. That is, they analyze the credibility of news only based on characteristics of a single news source (either the news article's content or the author's profile). By comparison, our method aims to identify fake news by integrating the credibility of multiple news sources reporting the same news event. Experimental results show that the accuracy of FND can be improved by leveraging "collective intelligence" (see Section IV).

III. Fake News Detector Based On Multi-Source Scoring

In this section, we give the details of the proposed FNDMS. We first present a method for news article collecting and filtering, and then propose two sets of features to evaluate the credibility of a single news source (report), as listed in Fig. 1. Finally, we introduce a DST-based model for integrating the credibility scores of multiple sources and making a decision on the truth of a news event. Fig. 2 shows an illustration of the entire FNDMS framework.

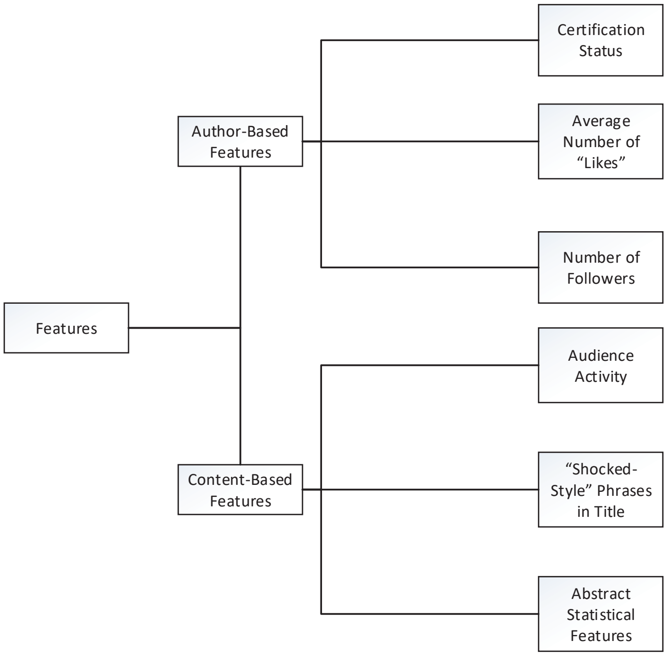

Figure 1. Feature Structure

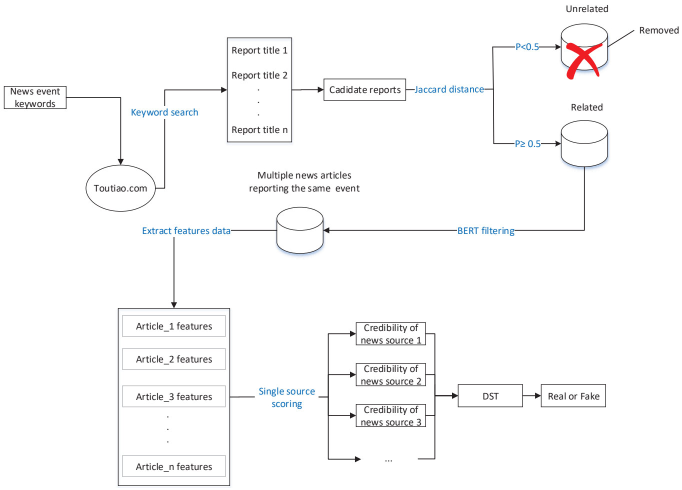

Figure 2. An illustration of the proposed FNDMS framework.

Authorized licensed use limited to: Hong Kong University of Science and Technology. Downloaded on July 19,2023 at 10:59:18 UTC from IEEE Xplore. Restrictions apply.

A. News Article Collecting and Filtering

We use a three-step method to collect articles reporting the same event from social media. For a news event, we select an article reporting it as a seed. In the first step, we perform a keyword searching using terms in the title of the seed as keywords to retrieve a set of candidate news articles. In the second step, we measure the similarities between the title of the seed and those of articles in the candidate set with Jaccard distance, then filter out unrelated articles having small similarities. Considering the situation that two titles reporting different events have a large number of identical keywords, in the third step, we employ the Bidirectional Encoder Representation from Transformers (BERT) [22] to further filter the candidate set. In the following, we give the details of the three steps.

1) Step 1: We first remove stop words from the title of the seed article. Then we take the remaining words as input and perform a keyword searching using the interface provided by social media platforms. For the result list returned by the searching interface, we keep the top 20 articles as candidates.

2) Step 2: For each article R^i_{cand} in the candidate set, we compare the similarity between its title and that of the seed R_{seed} using Jaccard distance, which is computed as follows:

$$Jaccard(R_{seed}, R^i_{cand}) = \frac{W_{\mathrm{seed}} \cap W^i_{\mathrm{cand}}}{W_{\mathrm{seed}} \cup W^i_{\mathrm{cand}}}, \tag{1}$$

where W_{seed} is the set of terms in the title of the seed, and W^i_{cand} is the set of terms in the title of the ith candidate article. Stop words are also removed from W^i_{cand} before computing the Jaccard distance. We remove unrelated candidate articles with distances larger than a threshold. In experiments, we set the value of the threshold to 0.5. The first row in Table I gives an example of unrelated articles that can be removed by Jaccard distance.

3) Step 3: To further filter out candidate articles that have a very similar title-term-set with the seed but report a different event (see the second row in Table I), we use BERT to classify the relationship between a candidate article and the seed into three categories, i.e., "*Same Event*", "*Refutation*", and "*Irrelevant*". We use the pre-training model provide by Google[1] and fine-tune the model with our training data to obtain the classification model we need. Each sample in our training dataset is a triple $< S_1, S_2, L >$, where S_1 and S_2 are two sentences, and L is their relationship category label. After classification, we select the articles with the category label "*Same Event*" and "*refutation*" as the final set. All of the articles in the final set are considered reporting the same events with the seed. We only use the titles of two articles to filter out irrelevant articles since this can reduce the computation cost than using the entire content of the article.

B. Credibility Scoring for a Single News Article

When evaluating the credibility of a single news article (i.e., a single news source), we consider both the impacts of its text content and the reputation of its author. We extract two sets of features, i.e., author-based features and content-based features, from the text content and the user profile, respectively. In the following, we give the detail of each feature.

1) Author-Based Features: As shown in Fig. 1, we use three author-based features to measure the credibility of the author of a news article. The features are defined and calculated as follows:

- **Certification Status**: Users with certification are those whose personal information has been verified by social media platforms. Typically, these users are more willing to take responsibility for the content they published on social media platforms. Therefore, it is reasonable to assume that users with certification are more likely to report the truth of a news event. We use a binary feature U_{certif} to represent the certification status of a user, where

$$U^i_{certif} = \begin{cases} 1, & \text{if user } i \text{ has been certified} \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

- **Average Number of "Likes"**: On social media platforms, such as Facebook and Twitter, users can light a "like" label for an article if they like or agree to the content of it. Typically, news reports that convey the truth of an event are easier to get praise from readers. Therefore, it is reasonable to use the number of "likes" obtained by an article to measure its credibility. Intuitively, if all or most of the articles published by a user get a large number of "likes", the user is more likely to publish articles with high credibility in the future. Based on this assumption, we use the average number of "likes" obtained by User i, which is represented by U^i_{like}, to rate the credibility level of all the articles she/he has published. U^i_{like} is computed as follows:

$$U^i_{like} = \lg(\frac{\sum_{i=1}^{N^i_{art}} N^{i,j}_{like}}{N^i_{art}}), \tag{3}$$

where N^i_{art} is the number of articles published by user i, and $N^{i,j}_{like}$ is the number of "likes" obtained by the jth article published by i. Because the difference between the of different users may be very large, Since there may be a very big difference between the U_{like} values of two distinct users, we use $\lg(\cdot)$ to limit the range of U^i_{like}. The value of U^i_{like} is further rescaled into [0,1] by the min-max normalization as follows:

$$G^i_{like} = \frac{U^i_{like} - \min(U_{like})}{\max(U_{like}) - \min(U_{like})}, \tag{4}$$

where $\max(U_{like})$ and $\min(U_{like})$ are the maximum value and the minimum value of U_{like} respectively for all users.

- **Number of Followers**: Social media users can receive constant updates from any other user they interested in by following that user. Intuitively, users with a large number of followers should be more cautious when they post

[1] https://github.com/google-research/bert

Authorized licensed use limited to: Hong Kong University of Science and Technology. Downloaded on July 19,2023 at 10:59:18 UTC from IEEE Xplore. Restrictions apply.

TABLE I
EXAMPLES OF SEARCHING SEEDS AND RETRIEVED ARTICLES.

Index	Title of the Seed Article	Title of a Retrieved Article	Relation
1	More than 10 people will be automatically monitored by the network police, and the possibility of being seized.	Emergency rumors! i 've heard that xiamen is going to be upgraded to a municipality directly under the jurisdiction of the municipality... the police have spoken	Unrelated (with dissimilar words)
2	The 003 nuclear-powered aircraft carrier has been built, with a displacement of 110,000 tons of carrier aircraft, 70 of which will be compared with the Ford class.	China's nuclear-powered aircraft carrier is about to appear, the displacement of 110000 tons of aircraft carrier 90, technology leading the world!	Unrelated (with similar words)
3	A billion years ago, there might have been a superior civilization on Mars. The Mayans?	Science says: 1 billion years ago, there could have been a superior civilization on mars!	Related

content, as the influence of the content will cover all their followers. Accordingly, these users are more likely to publish content stating the truth of a news event. On the other hand, the number of followers is also an indicator of reputation. A user with a large number of followers typically has a high reputation in her/his social media community. Therefore, we use the number of followers, which is represented by U_{follow}, as a measurement to evaluate the credibility of a user. As what we do to U_{like}, we limit the range of $U_{follower}$ and rescale its value into [0,1] using $\lg(\cdot)$ and min-max normalization, which are achieved with Equation (5) and (6), respectively.

$$U^i_{follower} = \lg(N^i_{follower}), \tag{5}$$

$$G^i_{follower} = \frac{U^i_{follower} - \min(U_{follower})}{\max(U_{follower}) - \min(U_{follower})}. \tag{6}$$

In (5), $N^i_{follower}$ denotes User i's raw number of followers. In (6), $\max(U_{like})$ and $\min(U_{like})$ are the maximum value and the minimum value of $U_{follower}$ respectively for all users.

Finally, we compute a total credibility score C_{user} for the author of a news article as follows:

$$C_{author} = \tau G_{follower} + \varepsilon G_{like} + \upsilon U_{certif}, \tag{7}$$

where τ, ε, and υ are the weights of $G_{follower}$, G_{like}, and U_{certif}, respectively, satisfying $\tau + \varepsilon + \upsilon = 1$. It is worth noticing that, based on our observation, we found that G_{like} plays a more important role in measuring the credibility of users. Therefore, we give this feature a larger weight in the experiments (see Section IV), i.e., we set τ, $\upsilon < \varepsilon$.

2) Content-Based Features: We also use three textual features to measure the credibility of a news article's content, which are defined and calculated as follows.

- **Audience Activity**: Typically, the content of an article reporting the truth is more reasonable than that of fake news and thus may attract more responses from the others. That means users may prefer to comment a credible article with their own opinions rather than just read it. Based on this intuition, we introduce "audience activity" as a feature to measure the credibility of an article. Let $S_{comment}$ denote the number of comments obtained by Article i, and S^i_{reader} denote the number of readers of i. We define the audience activity of a news article as the ratio of $S_{comment}$ and S^i_{reader}, which is computed using Equation (8):

$$N^i_{activity} = \frac{\left|S^i_{comment}\right|}{\left|S^i_{reader}\right|}. \tag{8}$$

- **"Shocked-Style" Phrases in Title**: After examining a great number of news articles (both stating the truth and rumors), we found that the authors of fake news are more likely to use "shocked-style" phrases, such as "really (adj.)", "I never knew this", "Can you believe that?", "you surely can't guess", and "Nobody can believe." Based on this observation, we construct a dictionary of "shocked-style" phrases and then check whether there are these phrases in the title of a news article. We use the proportion of "shocked-style" phrases in the title (denoted by $N_{shocked}$) as a feature to evaluate the credibility of Article i, which is computed as follows:

$$N^i_{shocked} = 1 - \frac{\left|S^i_{shocked}\right|}{\left|S^i_{all}\right|}, \tag{9}$$

where $S^i_{shocked}$ and S^i_{all} are the number of "shocked-style" phrases and the length of the title, respectively.
- **Abstract Statistical Features**: An abstract of a news article summarizes the full text with concise sentences. We consider a news article's abstract as an important part in evaluating the credibility of the article. Generally, the completeness of the abstract reflects the author's attitude towards writing. On the other hand, an article reporting the truth typically describes the facts of an event objectively, using as few emotional words and emotional punctuations as possible. Therefore, we uses the length of the abstract and the number of emotional words contained in the abstract to measure the completeness and objectivity of the abstract, Among them, the length of the abstract is represented by N_{length}, and $V_{emotional}$ represents the

Authorized licensed use limited to: Hong Kong University of Science and Technology. Downloaded on July 19,2023 at 10:59:18 UTC from IEEE Xplore. Restrictions apply.

number of emotional words contained in the abstract. N_{length} reflects the completeness of an abstract, the ratio of $V_{emotional}$ to N_{length} reflects the emotion of the author. The more emotional words there are in an abstract, the more subjectively the author describes the facts. We used $N^i_{aobject}$ to measure the objectivity of the abstract of the i-th news, and we compute $N^i_{aobject}$ of the ith news article using the following equation:

$$N^i_{aobject} = 1 - \frac{|V^i_{emotional}|}{N^i_{length}}, \tag{10}$$

where $V^i_{emotional}$ is the set of emotional words in the abstract of Article i.

We compute a final content-based credibility score for Article i by summing all the features above as follows:

$$C^i_{content} = \widetilde{N}^i_{activity} + \widetilde{N}^i_{shocked} + \widetilde{N}^i_{aobject} + \widetilde{N}^i_{length}, \tag{11}$$

where $\widetilde{N}^i_{activity}$, $\widetilde{N}^i_{shocked}$, $\widetilde{N}^i_{aobject}$, and $\widetilde{N}^i_{length}$ are the min-max normalizations of $N^i_{activity}$, $N^i_{shocked}$, $N^i_{aobject}$, and N^i_{length}, respectively.

C. Multi-Source Credibility Integration

Thus far, we have extracted sets of features to evaluate the credibility of a news source from both its author's profile and its textual content. The next question is how to integrate the credibility scores of multiple news sources to judge whether a news event has really happened or just a rumor. The credibility that different sources can provide is different, we can't judge whether news events really happen through a single source. DST as a method of uncertainty reasoning provides a powerful tool for the representation and fusion of decision-level uncertain information due to the fact that it is able to provide a fusion framework, and changes previous opinions based on new evidence combined with cumulative evidence [23]. It can better integrate the credibility representation of different sources of news events to decide whether the news is a rumor. So in the credibility integration work, we adopt the DST, which is designed to effectively fuse evidence from multiple sources, to address this question. Let Ω be a complete set composed of mutually exclusive basic propositions, and let $\mathbb{P}(\Omega) = 2^{\Omega}$ be the set of all possible subsets of Ω, including the empty set. The elements in $\mathbb{P}(\Omega)$ can be taken to represent propositions that we are interested in, by containing and only the states in which this proposition is true. The DST assigns a belief mass to each subset of $\mathbb{P}(\Omega)$, which can be formally represented by a function $m : \mathbb{P}(\Omega) \rightarrow [0, 1]$, satisfying $m(\varnothing) = 0$ and $\sum_{A \in \mathbb{P}(\Omega)} m(A) = 1$. Such an assignment is called a Basic Probability Assignment (BPA).

For our problem, the proposition set $\Omega = \{g, \neg g\}$ is composed of two elements, where g denotes that a news event has really happened, and $\neg g$ means that the event did not happen. The power set of Ω is $\mathbb{P}(\Omega) = \varnothing, g, \neg g, \Omega$, where Ω is an impossible proposition. Then we have:

$$\begin{cases} m(\varnothing) = 0 \\ m(g) = C_i \\ m(\neg g) = 1 - m(g) \\ m(\Omega) = 0 \end{cases}. \tag{12}$$

Here C_i is the credibility score of the ith news source, which is computed as follows:

$$C_i = \lambda C^i_{author} + (1 - \lambda) C^i_{content}, \tag{13}$$

where C^i_{author} and $C^i_{content}$ are author-based and content-based credibility scores of the ith news source, respectively, and λ is a weight parameter to balance the contributions of C^i_{author} and $C^i_{content}$.

As the original DST may have evidence conflicts in the data processing process, it can produce results inconsistent with common sense. Therefore, we employ the improved-DST [24] to fuse the credibility scores of multiple news sources, which is modeled mathematically as follows:

$$\begin{cases} m(\varnothing) = 0 \\ m(g) = \prod_{i=1}^{n} m_i(g) + kq(g) \\ m(\neg g) = 1 - m(g) \\ m(\Omega) = 0 \end{cases}, \tag{14}$$

where $k = 1 - \prod_{i=1}^{n} m_i(g) - \prod_{i=1}^{n} m_i(\neg g)$, $q(g) = \frac{1}{n}\sum_{i=1}^{n} m_i(g)$. We consider $m(g)$ as the probability that a news event has really happened, and $m(\neg g) = 1 - m(g)$ as the probability that it is a rumor. Thus, the event is real if $m(g) > m(\neg g)$ and is fake otherwise.

IV. Experiments And Analysis

In this section, we conduct comprehensive experiments to evaluate the performance of the proposed framework. All the experiments were implemented with python and run on a desktop with an Intel i7 CPU and 16GB memory.

A. Data

Our experimental dataset was collected from toutiao.com, which is one of the most popular news-releasing social media sites in China. We took the public dataset released by WS-DM Fake News Classification Challenge (FNCC)[2] as a data basis. FNCC provides a training dataset and a testing dataset, which are used for training fake news classifiers and testing, respectively. We only used the training set in data acquiring, which contains 320,767 samples. Each sample consists of a pair of news (titles only) and a label indicating the relation (agreed/disagreed/unrelated) between the pair. The first piece of news in each pair reports a fake event. We used its title as a seed and collect articles reporting the same event from toutiao.com with the three-step method proposed in Section III-A. The training dataset was also used in the fine-tuning of BERT. We finally collected 1,737 news articles reporting 392 fake events. In addition, we also manually selected seed

[2] https://www.kaggle.com/c/fake-news-pair-classification-challenge/data

Authorized licensed use limited to: Hong Kong University of Science and Technology. Downloaded on July 19,2023 at 10:59:18 UTC from IEEE Xplore. Restrictions apply.

articles for 178 real events and obtained another 889 related news reports using the same data collecting process.

B. Evaluation Metrics

To evaluate the performance of classification models, we use *accuracy*, *precision*, *recall*, and F_1 as performance metrics, which are expressed as follows:

$$accuracy = \frac{TP + TN}{Total}, \tag{15}$$

$$precision = \frac{TP}{TP + FP}, \tag{16}$$

$$recall = \frac{TP}{TP + FN}, \tag{17}$$

$$F_1 = \frac{2 * recall * precision}{recall + precision}, \tag{18}$$

where *TP* is the number of positive samples classified correctly, *FN* is the number of positive samples classified as negative, *FP* is the number of negative samples classified as positive, *TN* is the number of negative samples classified correctly, and *Total* is the total number of samples.

C. Effectiveness of Features

In the first experiment, we tested the effectiveness of the proposed features. We built FNDMS with different feature configurations separately, each of which has one feature removed, as listed in Table II. Fig. 3 gives the prediction results of the six configurations on our experimental data. From the results, we can see that removing any one of the six features can lead to a drop in performance in terms of all measurements. Among all features, “Average Number of Likes” has the greatest impact on the performance of FNDMS. This feature can reflect the extent that the audiences agree with the content of a news article. The removing of it causes the accuracy, the precision, the recall, and the F_1 to drop from 0.996, 0.997, 0.994, and 0.995 to 0.875, 0.85, 0.875, and 0.862, respectively. A possible explanation is that news articles reporting the truth could obtain more support from their audiences. By comparison, the feature “Audience Activity” has the smallest impact on the performance of FNDMS. Based on our analysis of the experimental data, we found that both articles reporting the truth and fake news could attract a great number of responses from audiences.

Another interesting observation is that author-based features have higher impacts on the performance of FNDMS than content-based features. When removing an author-based feature, the decrease in performance in terms of accuracy and F_1 is larger than that of removing a content-based feature. This reveals that the credibility of the author plays an important role in judging whether a news article tells the truth. We found that authors with high credibility were more likely to publish credible news all the time.

In summary, the results indicate that all the proposed features have contributions to fake news detection. A combination of author-based features and content-based features can improve the performance of the proposed method.

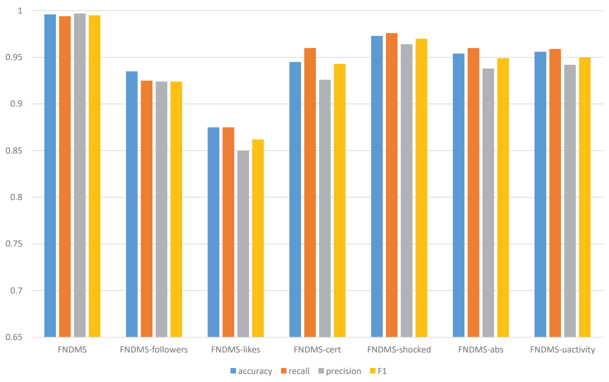

Figure 3. Validation of Feature Validity

D. Effectiveness of Multi-Source Scoring

We compared the proposed FNDMS with four common-used classifiers, including Support Vector Machine (SVM), Random Forest (RF), Logistic Regression (LR), and Adaboost, to validate its performance on fake news detection. We selected 2,040 news articles reporting 444 events (including 119 real events and 325 fake events) as the training set to learn the four baseline classifiers. The same set of features with our method was used in the training process of each classifier. We also chose 486 news articles related to 100 events (including 50 real events and 50 fake events) as the testing set. Note that for the baseline classifiers each news article is seen as a single example, which means these classifiers evaluate the credibility of a news event based on information from a single news source.

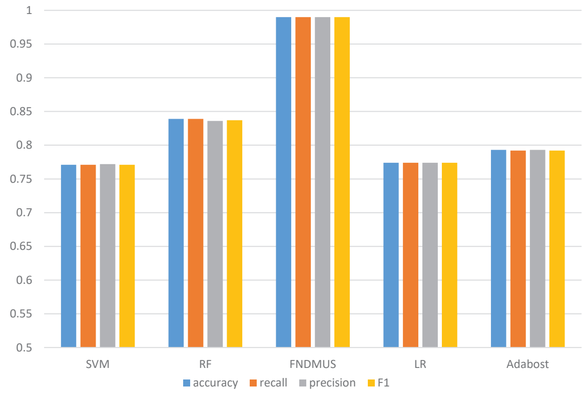

Figure 4. Comparison of Classification Results of Various Classifiers

The experimental results are shown in Fig. 4. As we can see from the results, our method outperformed the other four methods significantly in terms of all evaluation measurements. The accuracy of our method achieved 99%, which is over 15%

Authorized licensed use limited to: Hong Kong University of Science and Technology. Downloaded on July 19,2023 at 10:59:18 UTC from IEEE Xplore. Restrictions apply.

TABLE II
FEATURE CONFIGURATIONS USED TO BUILD FNDMS MODELS

Notations	Feature Configurations
$FNDMS$	A FNDMS model built with all features
$FNDMS_{-followers}$	A FNDMS model built with the “number of followers” feature removed.
$FNDMS_{-likes}$	A FNDMS model built with the “number of likes” feature removed.
$FNDMS_{-cert}$	A FNDMS model built with the “Certification Status” feature removed.
$FNDMS_{-abs}$	A FNDMS model built with the “Abstract Statistical” features removed.
$FNDMS_{-uactivity}$	A FNDMS model built with the “Audience Activity” feature removed.
$FNDMS_{-shocked}$	A FNDMS model built with the “Shocked-Style Phrases in Title” feature removed.

TABLE III
EXAMPLES OF FAKE NEWS DETECTION RESULTS

Method	New ID	News title	Publisher	User status	category
RF	16016	The first day of examination of the new regulations in Nanning, the rate of clearance is such	China Economic Net	Official account	Real
RF	16016	The first day of examination of the new regulations in Nanning, the rate of clearance is such	Kapok	Unauthorized users	Fake
FNDMU	16016	The first day of examination of the new regulations in Nanning, the rate of clearance is such	All authors reporting 16016	-	Fake

higher than that of the best baseline method (i.e., RF). This reveals that the effectiveness of our multi-source scoring strategy for fake news detection. For the four baseline classifiers, their performances are affected by the sample distribution in the training data to an extent. For example, Table III gives three prediction results of the same news event, including two outputs from the RF classifier and one output from our method. We can see that RF gave two different predictions for the same news event. This may because that in the training set most articles published by authors similar to “China Economic Net”, who is certified and has a large number of followers, report real events, while most articles published by authors similar to “Kapok” report rumors. The RF thus classified the article authored by “China Economic Net” as real and that authored by “Kapok” as fake. This problem is also with the other three baseline methods. However, our method can significantly reduce the negative impact of this problem by considering the credibility of multiple news sources. Moreover, the DST model is superior to traditional supervised learning algorithms in integrating multiple conflicting information.

In summary, the results of this experiment demonstrate the feasibility and advance of our framework. It can achieve better performance than baseline methods by integrating the credibilities of multiple news sources, which is more objective than depending on clues from a single source.

E. Impact of the Number of Event-related Articles

As the efficiency of our multi-source scoring strategy can be impacted by the number of collected articles reporting the same event, we also investigate the sensitivity of the proposed method to the number of event-related articles. We selected articles reporting 54 news events (including 30 fake events and 24 real events) as the testing set for this experiment. Each event has six articles reporting it. We ran our framework four times on the testing data, and in each time we randomly deleted 0∼3 articles respectively from the related articles of an event. The experimental results are shown in Table IV.

TABLE IV
IMPACT OF THE NUMBER OF ARTICLES ON CLASSIFICATION RESULTS

Runs	# of Removed Articles	$accuracy$	$recall$	$precision$	F_1
1	0	**1.0**	**1.0**	**1.0**	**1.0**
2	1	0.981	0.979	0.984	0.982
3	2	0.981	0.983	0.98	0.982
4	3	0.907	0.908	0.906	0.907

From Table IV we can see that the removing of articles can result in a performance decline of our method. A possible reason is that if the number of articles reporting an event is small, the performance of our framework will heavily be affected by a single article. The more related articles there are,

Authorized licensed use limited to: Hong Kong University of Science and Technology. Downloaded on July 19,2023 at 10:59:18 UTC from IEEE Xplore. Restrictions apply.

the more accurate our framework will be. If there is only one related article, our framework will degenerate into a single-source method. This experiment also reveals the importance of integrating the credibilities of multiple sources.

V. Conclusion

In this paper, we have proposed FNDMS, a framework designed for detecting fake news spreading on social media platforms. FNDMS uses a multi-source scoring strategy, which integrating the credibility scores of multiple news sources, to evaluate the truth of a news event. We propose two sets of features, i.e., author-based features and content-based features, to measure the credibility of a single news source. A DST model is employed for credibility fusion and making a final decision on whether the event has really happened. We also propose a three-step method to retrieve and filter articles reporting a news event. Experimental results on real social media data show that FNDMS is superior to several common-used machine learning algorithms in fake news detection. This demonstrates the feasibility and advance of our multi-source scoring strategy, which is more objective than depending on clues from a single source. Furthermore, the results also reveal the effectiveness of the proposed features. In the future work, we will work to improve the accuracy of news source collection, and at the same time, we will mine deeper author features and text features to participate in FNDMS.

Acknowledgment

This work was supported in part by National Natural Science Foundation of China (61602281, 61702309), Shandong Provincial Key Research and Development Program (2018CXGC0701, 2018GGX106005, 2017CXGC0701, and 2017CXGC0706), Natural Science Foundation of Shandong Province of China (ZR2015YL018, ZR2016YL011, and ZR2016YL014), and Cooperation Fund for the Youth Doctor of Qilu University of Technology (Shandong Academy of Sciences) (2017BSH2012).

References

[1] Matheus P Viana, Diego R Amancio, and Luciano da F Costa. On time-varying collaboration networks. *Journal of Informetrics*, 7(2):371–378, 2013.

[2] Newman Nic, Richard Fletcher, Antonis Kalogeropoulos, David AL Levy, and Rasmus Kleis Nielsen. Reuters institute digital news report 2018, [online]. available: http://media.digitalnewsreport.org/wp-content/uploads/2018/06/digital-news-report-2018.pdf, accessed on: 06, 2019.

[3] Jiaojiao Jiang, Sheng Wen, Shui Yu, Yang Xiang, and Wanlei Zhou. Rumor source identification in social networks with time-varying topology. *IEEE Transactions on Dependable and Secure Computing*, 15(1):166–179, 2016.

[4] Zhenhua Tan, Jingyu Ning, Yuan Liu, Xingwei Wang, Guangming Yang, and Wei Yang. Ecrmodel: An elastic collision-based rumor-propagation model in online social networks. *IEEE Access*, 4:6105–6120, 2016.

[5] Dazhen Lin, Ben Ma, Min Jiang, Naixue Xiong, Kai Lin, and Donglin Cao. Social network rumor diffusion predication based on equal responsibility game model. *IEEE Access*, 7:4478–4486, 2019.

[6] Aditi Gupta, Ponnurangam Kumaraguru, Carlos Castillo, and Patrick Meier. Tweetcred: Real-time credibility assessment of content on twitter. In *International Conference on Social Informatics*, pages 228–243. Springer, 2014.

[7] Nguyen Vo and Kyumin Lee. The rise of guardians: Fact-checking url recommendation to combat fake news. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, pages 275–284. ACM, 2018.

[8] Lee Howell et al. Digital wildfires in a hyperconnected world. *WEF Report*, 3:15–94, 2013.

[9] Chengcheng Shao, Pik-Mai Hui, Pengshuai Cui, Xinwen Jiang, and Yuxing Peng. Tracking and characterizing the competition of fact checking and misinformation: Case studies. *IEEE Access*, 6:75327–75341, 2018.

[10] Ruohan Li and Ayoung Suh. Factors influencing information credibility on social media platforms: Evidence from facebook pages. *Procedia computer science*, 72:314–328, 2015.

[11] Carlos Castillo, Marcelo Mendoza, and Barbara Poblete. Predicting information credibility in time-sensitive social media. *Internet Research*, 23(5):560–588, 2013.

[12] Carlos Castillo, Marcelo Mendoza, and Barbara Poblete. Information credibility on twitter. In *Proceedings of the 20th international conference on World wide web*, pages 675–684. ACM, 2011.

[13] Benjamin D Horne and Sibel Adali. 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*, 2017.

[14] Hadeer Ahmed, Issa Traore, and Sherif Saad. Detection of online fake news using n-gram analysis and machine learning techniques. In *International Conference on Intelligent, Secure, and Dependable Systems in Distributed and Cloud Environments*, pages 127–138. Springer, 2017.

[15] William Yang Wang. " liar, liar pants on fire": A new benchmark dataset for fake news detection. *arXiv preprint arXiv:1705.00648*, 2017.

[16] Jing Ma, Wei Gao, and Kam-Fai Wong. Rumor detection on twitter with tree-structured recursive neural networks. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, volume 1, pages 1980–1989, 2018.

[17] Feng Qian, Chengyue Gong, Karishma Sharma, and Yan Liu. Neural user response generator: Fake news detection with collective user intelligence. In *IJCAI*, pages 3834–3840, 2018.

[18] Natali Ruchansky, Sungyong Seo, and Yan Liu. Csi: A hybrid deep model for fake news detection. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, pages 797–806. ACM, 2017.

[19] Yaqing Wang, Fenglong Ma, Zhiwei Jin, Ye Yuan, Guangxu Xun, Kishlay Jha, Lu Su, and Jing Gao. Eann: Event adversarial neural networks for multi-modal fake news detection. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 849–857. ACM, 2018.

[20] Fan Yang, Yang Liu, Xiaohui Yu, and Min Yang. Automatic detection of rumor on sina weibo. In *Proceedings of the ACM SIGKDD Workshop on Mining Data Semantics*, pages 13:1–13:7. ACM, 2012.

[21] Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu. Fake news detection on social media: A data mining perspective. *ACM SIGKDD Explorations Newsletter*, 19(1):22–36, 2017.

[22] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

[23] Dehuai Zeng, Jianmin Xu, and Gang Xu. Data fusion for traffic incident detection using ds evidence theory with probabilistic svms. *Journal of computers*, 3(10):36–43, 2008.

[24] Li Bicheng and Qian Cengbo. An efficient combination rule of evidence theory. *Journal of Data Acquisition & Processing*, 17(1):33–36, 2002.

Authorized licensed use limited to: Hong Kong University of Science and Technology. Downloaded on July 19,2023 at 10:59:18 UTC from IEEE Xplore. Restrictions apply.