IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

# Fraud news identification: A Survey

**Ms. Indrani Sonkar, Mr. Neetish Chandrakar, Mr. L P Bhaiya, Mr Virendra Swarnkar**

Department of Computer Science

Bharti college of engineering and technology, Durg

## Abstract

Fraud news identification might be a basic yet testing issue. The quick ascent of person to person communication stages has not just yielded a colossal increment in data openness however has additionally quickened the spread of fake news. Given the enormous measure of site page, programmed Fraud news identification might be a reasonable NLP issue required by all online substance suppliers. This paper presents a study on Fraud news identification. Our review presents the difficulties of programmed Fraud news location. We deliberately audit the datasets and NLP arrangements that are produced for this errand. We likewise talk about the limits of those datasets and issue details, our bits of knowledge, and suggested arrangements.

Keyword : Fake news detection, fake-news images, multi domain, social media

## 1 Introduction

Automatic Fraud news identification is that the task of assessing the truthfulness of claims in news. this is often a replacement , but critical NLP problem because both traditional journalism and social media have huge social-political impacts on every individual within the society. for instance , exposure to Fraud news can cause attitudes of inefficacy, alienation, and cynicism toward certain political candidates (Balmas, 2014). The worst a part of the spread of faux news is that sometimes it does link to offline violent events that threaten public safety (e.g., the Pizza Gate (Kang and Goldman, 2016)). Detecting Fraud news is of crucial importance to the NLP community, because it also creates broader impacts on how technologies can facilitate the verification of the veracity of claims while educating the overall public.

The conventional solution to the present task is to ask professionals like journalists to see claims against evidence-based on previously spoken or written facts. However, it's time-consuming and costs tons of human resources, for instance, Poli-tiFact1 takes three editors to guage whether a bit of stories is real or not.

As the Internet community and therefore the speed of the spread of data are growing rapidly, auto-mated fact-checking on internet content has gained many interests within the AI research community. The goal of automatic Fraud news identification is to scale back the human time and energy to detect Fraud news and help us to prevent spreading them. The task of faux news identification has been studied from various perspectives with the event in subareas of computing, like Machine Learning (ML), data processing (DM), and NLP.

In this paper, we survey automated Fraud news identification from the attitude of NLP. Broadly, we introduce the technical challenges in Fraud news identification and the way researchers define different tasks and formulate machine learning solutions to tackle this problem. We discuss the pros and cons, also because the potential pitfalls and disadvantages of every task. More specifically, we offer a summary of research efforts for Fraud news identification and a scientific comparison of their task definitions, datasets, model construction, and performances. We also discuss a suggestion for future research during this direction. This paper also includes another aspect like social engagement analysis. Our contributions are three folds:

- We furnish the first complete evaluation on Natural Language Processing options for computerized Fraud information identification;

- We systematically analyze how Fraud information identification is aligned with present NLP tasks, and talk about the assumptions and tremendous troubles for one-of-a-kind formulations of the problem;
- We categorize and summarize handy datasets, NLP approaches, and results, pro-viding first-hand experiences and available introductions for new researchers involved in this problem.

| Name | Main Input | Data Size | Label | Annotation |
|---|---|---|---|---|
| LIAR | short claim | 12,836 | six-grade | editors, journalists |
| FEVER | short claim | 185,445 | three-grade | trained annotators |
| BUZZFEEDNEWS | FB post | 2282 | four-grade | journalists |
| BUZZFACE | FB post | 2263 | four-grade | journalists |
| some-like-it-hoax | FB post | 15,500 | hoaxes or non-hoaxes | none |
| PHEME | Tweet | 330 | true or false | journalists |
| CREDBANK | Tweet | 60 million | 30-element vector | workers |
| FRAUDNEWSNET | article | 23,921 | Fraud or real | editors |
| BS DETECTOR | article | - | 10 different types | none |

Table 1: A Summary of Various Fraud News Identification Related Datasets.

## 2 Datasets

A major challenge for automated Fraud news identification is the availability and quality of the datasets. There exists a variety of datasets for Fraud news identification. We categorize them and discuss their characteristics.

### 2.1 One-or-Few-Sentences Datasets

#### 2.1.1 Short Claims

A latest benchmark dataset for Fraud information identification is a LIAR (Wang, 2017). This dataset consists of 12,836 real-world brief statements gathered from PolitiFact, the place editors handpicked the claims from a range of events such as debate, campaign, Facebook, Twitter, interviews, ads, etc. Each assertion is labelled with six-grade truthfulness. The facts about the subjects, party, context, and audio system are additionally protected in this dataset. Vlachos and Riedel (2014) and Ferreira and Vlachos (2016) are the first to find out about PolitiFact data, however LIAR is orders of magnitude large and extra comprehensive. However, observe that the unique LIAR paper does no longer encompass the editor's justification or proof due to copyright concerns, and customers will want to retrieve the justification/evidence one after the other the usage of an API. Also, even even though both the claims and the proof are from real-world occasions, they are especially unstructured. Fact-checking stays pretty difficult for this dataset.

Fever (Thorne et al., 2018) is a dataset offering associated proof for Fraud information identification. Fever includes 185,445 claims generated from Wikipedia data. Each declaration is labelled as Sup-ported, Refuted, or Not Enough Info. They additionally marked which sentences from Wikipedia they use as evidence. Fever makes it viable to strengthen a machine which can predict the truthfulness of a declare collectively with the evidence, even although the kind of data and proof from Wikipedia may also nonetheless show off some foremost stylistic variations from these in real-world political campaigns.

#### 2.1.2 Posts On Social Networking Services

In addition to the web sites noted above, posts on Social Networking Services (SNS), such as Twitter and Facebook, can additionally be a supply of brief information statements. There are some datasets for Fraud information identification focusing on SNS, however they have a tendency to have a restrained set of subjects and can be much less associated to the news.
BUZZFEEDNEWS4 collects 2,282 posts from 9 information companies on Facebook. Each publish is fact-checked through 5 BuzzFeed journalists. The blessings of this dataset are that the articles are accumulated from each facets of left-leaning and right-leaning organizations, and they are enriched in

| Attributes | Value |
|---|---|
| ID of the statement | 11972 |
| Label | True |
| Statement | Building a wall on the U.S.-Mexico border will take literally years. |
| Subject(s) | Immigration |
| Speaker | Rick Perry |
| Speaker's job title | Governor of Texas |
| Party affiliation | Republican |
| Total Credit History Counts | 30,30,42,23,18 |
| Context | Radio Interview |

Table 2: An Example Entry from L IAR. The ordered total credit history counts are {barely true, false, half true, mostly true, pants on fire}.

Potthast et al. (2017) via including statistics such as the linked articles. BUZZFACE (Santia and Williams, 2018) extends the BuzzFeed dataset with the feedback associated to information articles on Facebook. It incorporates 2,263 information articles and 1.6 million comments. SOME-LIKE-IT-HOAX5 (Tacchini et al., 2017) consists of 15,500 posts from 32 Facebook pages, that is, the public profile of agencies (14 conspiracy and 18 scientific organizations). This dataset is labelled primarily based on the identification of the writer alternatively of post-level annotations so that it might also have imposed a sturdy assumption. A possible foremost pitfall for such dataset is that such sort of labelling method can end result in desktop gaining knowledge of fashions getting to know traits of every publisher, as an alternative than that of the Fraud news.

PHEME (Zubiaga et al., 2016) and CRED-BANK (Mitra and Gilbert, 2015) are two Twitter datasets. PHEME incorporates 330 twitter threads (a collection of related Tweets from one person) of 9 newsworthy events, labelled as genuine or false in accordance to thread constructions and follow-follower relationships. CREDBANK carries 60 mil-lion tweets masking ninety six days, grouped into 1,049 occasions with a 30-dimensional vector of truthful-ness labels. Each tournament used to be rated on a 5-point Likert scale of truthfulness via 30 human annotators. They in reality concatenate 30 rankings as a vector due to the fact they locate it hard to decrease it to a one-dimensional score.

As noted above, these datasets had been created for verifying the truthfulness of tweets. Thus they are restricted to a few numbers of matters and can encompass tweets with no relationship to the news. Hence each datasets are now not so plenty best for Fraud information identification so that they are greater often used for hearsay identification.

### 2.2 Entire-Article Datasets

There are quite a few datasets for Fraud information identification focusing on Fraud information identification primarily based on the complete article. For example, FRAUDNEWSNET (Shu et al., 2017a,b, 2018) is an ongoing statistics series venture for Fraud information research. It consists of headlines and physique texts of Fraud information articles from BuzzFeed and PolitiFact. It additionally collects facts about the social engagements of these articles from Twitter.

BS DETECTOR6 is gathered from a browser extension named BS Detector, which shows its la-bels are the outputs of BS Detector, now not human an-notators. BS Detector searches all hyperlinks on a net web page at trouble for references to unreliable sources via checking towards a manually compiled listing of unreliable domains. Note that the primary problem with the usage of this dataset is that the computer gaining knowledge of fashions educated on this dataset are studying the parameters of the BS Detector.

Websites such as BLUFF THE LISTENER and THE ONION create sarcastic and humorous (Rubin et al., 2015a) Fraud information intentionally. Note that the sorts of Fraud information from these sources are limited. Moreover, it is pretty handy to classify them in opposition to regular new media articles. A dataset consists of articles from a variety of publishers can be higher (Rashkin et al., 2017), although person claims ought to be checked. We need to additionally notice that one need to keep away from the usage of combination labels honestly based totally on internet site source, as it provides extra confounding variables and it is greater of a internet site classification task..

## 3 Tasks

The general goal of Fraud news identification is to identify Fraud news. However, this task can be formulated in various ways.

### 3.1 Input

In this paper, we focus on Fraud news identification of text content. The input can be text ranging from short statements to entire articles. Additional information such as speakers' identity can be ap-pended. Inputs are related to which dataset is used (see Section 2).

### 3.2 Output

In most studies, Fraud news identification is formulated as a classification or regression problem, but the classification is more frequently used.

#### 3.2.1 Classification

The most common way is to formulate the Fraud news identification as a binary classification problem. However, categorize all the news into two classes (Fraud or real) is difficult because there are cases where the news is partly real and partly Fraud. To address this problem, add additional classes is a common practice. There are mainly two ways of adding additional classes. One is to set a category for the news which is neither completely real nor completely Fraud. The other one is to set more than two degrees of truthfulness, like LIAR and CREDBANK. The latter method reflects hu-man judgments more delicately. When using these datasets, the expected outputs are multi-class labels, and those labels are learned as independent labels with i.i.d assumptions (Rashkin et al., 2017; Wang, 2017).

#### 3.2.2 Regression

Fraud news identification can also be formulated as a regression task, where the output is a numeric score of truthfulness. This approach is used by Nakashole and Mitchell (2014). Formulating the task in this way can make it less straightforward to do the evaluation. Usually, evaluation is done by calculating the difference between the predicted scores and the ground truth scores or using Pearson/Spearman Correlations. However, since the available datasets have discrete ground truth scores, the challenge here is how to convert the discrete labels to numeric scores.

#### 3.2.3 Clustering

One of the conditions for Fraud news classifiers to achieve good performances is to have sufficient labelled data. However, to obtain reliable labels requires a lot of time and labour. Therefore, semi-supervised and unsupervised methods are proposed. The task is then formulated as a clustering problem instead of a classification one Rubin and Vashchilko (2012).

## 4 Methods

### 4.1 Preprocessing

Preprocessing usually includes tokenization, stemming, and generalization or weighting words.

To convert tokenized texts into features, Term Frequency-Inverse Document Frequency (TF-IDF) and Linguistic Inquiry and Word Count

(LIWC) are frequently used. For word sequences, pre-learned word embedding vectors such as word2vec (Mikolov et al., 2013) and GloVe (Pennington et al., 2014) are commonly

used. Appropriate preprocessing is necessary for a better understanding of Fraud news. Mihalcea and Strapparava (2009) use LIWC and find there is a difference in word usage between deceptive language and non-deceptive ones, so using word classification may have significant meaning on identification.

When using entire articles as inputs, an additional preprocessing step is to identify the central claims from raw texts. Thorne et al. (2018) rank the sentences using TF-IDF and DrQA system (Chen et al., 2017). Solutions to the text summarization task can also be applied.

### 4.2 Collecting Evidences

The RTE-based (Recognizing Textual Entailment) method is frequently used to gather and utilize evidence. RTE is the task of recognizing relationships between sentences, which can be applied to Fraud news identification. By gathering sentences which is for or against input from data sources such as news articles using RTE method, we can predict whether the input is correct or not. RTE-based models need textual evidence for fact check; thus this approach can be used only when the dataset includes evidence, such as FEVER and Emergence. Besides, RTE models cannot learn correctly when a claim in a dataset does not have enough information as evidence. To address this problem Thorne et al. (2018) develop a new approach which simulates training instances labelled as Not Enough Info by sampling evident sentences. Thus RTE models can use data without evidence.

### 4.3 Rhetorical Approach

Rhetorical Structure Theory (RST), sometimes combined with the Vector Space Model (VSM), is often used for Fraud news identification (Rubin et al., 2015b). RST is an analytic framework for the coherence of a story. Through defining functional relations (e.g., Circumstance, Evidence, and Pur-pose) of text units, this framework can systematically identify the essential idea and analyze the characteristics of the input text. Fraud news is then identified according to its coherence and structure.

To explain the results by RST, VSM is used to convert news texts into vectors, which are compared to the centre of true news and Fraud news in high-dimensional RST space. Each dimension of the vector space indicates the number of rhetorical relations in the news text.

### 4.4 Machine Learning Models

As mentioned in section 3, the majority of existing research use supervised method while semi-supervised or unsupervised methods are commonly used. In this section, we mainly describe classification models with several actual examples.

#### 4.4.1 Non-Neural Network Models

The most frequently used classification models in Fraud news identification are Support Vector Machine (SVM) and Naive Bayes Classifier (NBC). These two models differ a lot in structure thus comparing among them is meaningful. Logistic regression (LR) and decision tree such as Random Forest Classifier (RFC) are also used.

### 4.4.2 Neural Network Models

Many types of neural network models such as multi-layer perceptrons work for Fraud news identification, and many combinations of models are shown.

Recurrent Neural Network (RNN) is very popular in Natural Language Processing, especially Long Short-Term Memory(LSTM), which solves the vanishing gradient problem. LSTMs can capture longer-term dependencies. For example, Rashkin et al. (2017) set up two types of LSTM model, one put simple word embeddings initialized with GloVe into LSTM, and the other concatenate LSTM output with LIWC feature vectors before undergoing the activation layer. In both cases, they were more accurate than NBC and Maximum Entropy(MaxEnt) models, even though slightly.

Ruchansky et al. (2017) extract representations of both users and articles as low-dimensional vectors, and for the representation of articles, they use LSTM for each article. Textual information of each social engagement for an article is processed by doc2vec and put in LSTM, and are integrated with the score of the user in the last layer to classify.

Convolutional neural networks (CNN) are also widely used since they succeed in many text classification tasks. Wang (2017) uses a model based on Kim's CNN (Kim, 2014). They concatenate the max-pooled text representations with the meta-data representation from the bi-directional LSTM. CNN also used for analyzation using a variety of meta-data. For example, Deligiannis et al. give graph-like data of relationships between news and publishers to CNN and assess news from them.

Karimi et al. (2018) proposed Multi-source Multi-class Fraud news Identification framework (MMFD), in which CNN analyzes local patterns of each text in a claim and LSTM analyze temporal dependencies in the entire text. This model takes advantage of the characteristics of both models because LSTM works better for long sentences.

Attention mechanisms are often incorporated into neural networks to achieve better performance. Long et al. (2017) use attention model that incorporates the speaker's name and the statements topic to attend to features first, then weighted vectors are fed into an LSTM. Doing this increases accuracy by about 3 % (shown in Table 3, id 3,4). Kirilin and Strube (2018) use a very similar attention mechanism.

Memory networks, which is a kind of attention-based neural network, also shares the idea of attention mechanism. Pham (2018) uses Single Layer Memory network to learn a different representation of words by memorizing the set of words in the memory. When judging, input sentences weight the words in memory by attention mechanism. Thus the model can extract related words from its memory.

## 5 Experimental Results

We compare empirical results on classification datasets via various machine learning models in this section. Table 3 summaries the results on four datasets: LIAR, FRAUDNEWSNET, FEVER, and PHEME.

In Table 3, we collect and compare the exist-ing results of Fraud news classification research. For comparison, we use accuracy, which is a commonly used metric. Other evaluation metrics (Shu et al., 2017a) such as Precision, Recall, F-scores and ROC-AUC are also discussed.

## 6 Observations, Discussions, & Recommendations

### 6.1 Datasets and Inputs

There is nine requirements for Fraud news identification corpus, and we agree:

1. Availability of both truthful and deceptive in-stances;
2. Digital textual format accessibility;
3. Verifiability of "ground truth";
4. Homogeneity in lengths;
5. Homogeneity in writing matte;
6. Predefined timeframe;
7. The manner of news delivery;
8. Pragmatic concerns;
9. Language and culture.

### 6.2 Models

First, we compare how each model process textual content based on NLP. Most models we shared in Table 3 used word embeddings, especially word2vec, for taking the meanings of each text. The key to applying machine learning to Fraud news identification is choosing efficient features from just text with redundant information because features differ among Fraud news and real news, not among news topics or publishers of the news, should be extracted.

| Author | Input | Base Model | Acc. |
|---|---|---|---|
| Wang (2017) | Text+Meta | CNNs | **0.274** |
| Karimi et al. (2018) | Text+Meta | MMFD | **0.348** |
| Long et al. (2017) | Text+Meta | LSTM+Att | **0.415** |
| Pham (2018) | Text+Credit | MM(4.4.2) | **0.442** |
| Kirilin and Strube (2018) | Text+Meta+Sp2C | LSTM | **0.457** |
| Bhattacharjee et al. (2017 | | Deep (CNN) | 0.962 |
| Shu et al. (2017b | BuzzFeed | TriFN | **0.864** |
| Della Vedova et al. (2018) | | HC-CB-3(6.2) | **0.856** |
| Deligiannis et al. | | GCN(4.4.2) | **0.944** |
| Shu et al. (2017b) | PolitiFact | TriFN | **0.878** |
| Deligiannis et al. | | GCN | **0.895** |
| Della Vedova et al. (2018) | | HC-CB-3(6.2) | **0.938** |
| Thorne et al. (2018) | claim & evidences | Decomposable Att | **0.319** |
| Yin and Roth (2018) | | TWOWINGOS | **0.543** |
| Hanselowski et al. (2018) | | LSTM (ESIM-Att) | **0.647** |
| UNC-NLP | | (not yet announced) | **0.640** |
| UCL Machine Reading | | | **0.623** |
| Athene UKP TU Darmstadt | | | **0.613** |
| Kochkina et al. (2018) | 9 events | NileTMRG | **0.492** |

Table 3: The Current Results for Fraud News Identification. *Papers are sorted by the accuracy of the most accurate model. The highest result in each paper is in bold. "Att" is short for "Attention". Acc.: Accuracy. In* FEVER, N = 3* *means that the task is combined with evidence collection, and strictly speaking, it is not a classification task but claims verification, but we put it for your information.*

There are some imperative facets to extract in particular in Fraud information identification. First, the psycholinguistic classes of phrases used in the Fraud information have been established to be distinctive in some researches on the grounds that Mihalcea and Strapparava (2009) locate traits of the phrase used in misleading languages. Shu et al. (2017b) obtain 64% accuracy on FRAUDNEWSNET by means of only examining phrase utilization in LIWC. Thus it is clear analyzation on phrase utilization contributes tons to detecting Fraud news. Second, the rhetorical points can also vary in Fraud news. Rubin and Vashchilko (2012) exhibit that there must be some variations in the shape of sentences in misleading languages. In Table 3, RST (4.3) is the solely framework to study such features, and acquire 61% accuracy on FAK-ENEWSNET.

However, these home made points extraction may additionally be changed by using neural networks. Rashkin et al. (2017) indicates that including LIWC did now not enhance the overall performance of the LSTM mannequin however even damage it whilst Naive Bayes and MaxEnt fashions are improved. It may additionally be due to the fact some neural community fashions like LSTM can research lexical data in LIWC by using themselves. There is no such a find out about on rhetorical elements so we can't conclude, however neural community fashions may additionally additionally lean them, thinking about the RST model(id 17,23) gain solely low accuracies in contrast to different methods.

Hence it might also be higher to use computerized getting to know methods. For Natural Language Processing, LSTM and attention-based technique such as interest attachments or reminiscence community are frequently used. It is due to the fact they can analyze long-term and content-transitional statistics so that they can use the considerable phrase facts of sentences and observe context. Actually, many lookup in Table three use interest strategies (id 7,9-14,19,20,25,26,29,30,33,34) or LSTM (id 5-9,13,14,33,34,42,46) to analyze textual models. A famous software of interest mechanism is to generate interest weights for hidden layers primarily based on meta-data.

Second, thinking about extra records different than textual content in claims or articles, such as speaker credibility or social engagements records is the different environment friendly and realistic method; consequently most latest research in most cases center of attention on this method. Most research on LIAR enhance accuracy by means of altering the way to introduce no longer texts however speakers' records due to the fact it is challenging to realize a lie from quick sentences. Kirilin and Strube (2018) enhance accuracy with the aid of 21% thru changing the credibility records in LIAR's with a large credibility supply they launched named speak2credit7 (id 13-14) They exhibit that their interest mannequin depends on speaker's credibility via 43%, tons greater than 17% on a declaration of claim, with the aid of case study.

However, the tendency to depend their judgments on audio system or publishers can also reason some problem. Vlachos stated that the most risky misinformation comes from the sources we trust, and upgrading or downgrading unique sources purpose silencing minorities' voice(Graves, 2018).

Thus he developed new datasets FEVER which include proof so that it can be used for declare verification no longer solely for classification. Such content-based tactics have to be developed extra in the future. For declare verification on FEVER, Yin and Roth (2018) enhance the precision price to 45% from 10% (the benchmark score). The factor is that thinking about the recall fee does now not exchange that dramatically (from 46% to 50%), this mannequin has much less threat of verifying Fraud declare incorrectly. Research on FEVER is fewer than that on others due to the fact this dataset was once posted very these days and the accuracy, recall and precision fee are surprisingly low in most studies. There are very modern day consequences in Table three (id 35-40), however their performances do now not make plenty difference.

Social engagements statistics additionally suggests to be effective. For example, in Shu et al. (2017b) the mannequin the usage of solely social engagements facts (id 19,25) defeated the mannequin the usage of solely textual records (id 17,18,23,24). The identical as the usage of audio system credibility, we need to assume about the suited use of extra statistics as Della Vedova et al. (2018)(id 21,28) developed mannequin which makes use of the content-based technique when there are no longer sufficient social-engagements-based facts and in any other case use on the whole social-based one.

# 7 Related Problems

## 7.1 Fact-Checking

Fact-checking is the task of assessing the truthfulness of claims made by public figures such as politicians, pundits, etc (Vlachos and Riedel, 2014). Many researchers do not distinguish Fraud news identification and fact-checking since both of them are to assess the truthfulness of claims. How- to very high rate so that we should extend the way ever, Fraud news identification usually only focuses on verification with evidence as to the content-based on news events while fact-checking is broader.

## 7.2 Rumor Identification

There is not a consistent definition of rumor identification. A recent survey (Zubiaga et al., 2018) de-fines rumour identification as separating personal statements into rumor or non rumor, where rumor is defined as a statement consisting of unverified pieces of information at the time of posting. In other words, rumor must contain information that is worth verification rather than subjective opinions or feelings.

## 7.3 Stance Identification

Stance identification is the task of assessing whether the document supports a specific claim or not. It is different from Fraud news identification in that it is not for veracity but for consistency. Stance identification can be a subtask of Fraud news identification since it can be applied to searching documents for evidence (Ferreira and Vlachos, 2016).

## 7.4 Sentiment Analysis

Sentiment analysis is the task of extracting emotions, such as customer's favorable or unfavorable impression of a restaurant. Different from rumor identification and Fraud news identification, sentiment analysis is not to do an objective verification of claim but to analyze personal emotions.

# 8 Conclusion

We described and compared previous datasets and proposed new requirements for future datasets; 1. Easy to make from raw data in internets, 2.Have enough classes of truthfulness, 3.Quote claims or articles from different speakers or publishers. Be-sides, We compared the accuracy of many previous experiments and made some challenging task to our future Fraud news identification model; 1. More textual content-based method on multi-class Fraud news identification based on Natural Language Pro-cessing should be developed for realizing reliable identification. 2. We need a more logical explanation for Fraud news characteristics 3. There should be a limitation in language-based Fraud news identification in the case that there are not enough linguistic differences to improve identification accuracy

# References

[1] Meital Balmas. 2014. When fake news becomes real: Combined exposure to multiple news sources and political attitudes of inefficacy, alienation, and cyni-cism. *Communication Research*, 41(3):430–454.

[2] Sreyasee Das Bhattacharjee, Ashit Talukder, and Bala Venkatram Balantrapu. 2017. Active learning based news veracity identification with feature weight-ing and deep-shallow fusion. In *Big Data (Big Data), 2017 IEEE International Conference on*, pages 556–565. IEEE.

[3] Cody Buntain and Jennifer Golbeck. 2017. Auto-matically identifying fake news in popular twitter threads. In *Smart Cloud (SmartCloud), 2017 IEEE International Conference on*, pages 208–215. IEEE.

[4] Danqi Chen, Adam Fisch, Jason Weston, and An-toine Bordes. 2017. Reading wikipedia to answer open-domain questions. *arXiv preprint arXiv:1704.00051*.

[5] Nikos Deligiannis, Tien Huu Do, Duc Minh Nguyen, and Xiao Luo. Deep learning for geolocating social media users and detecting fake news.

[6] Marco L Della Vedova, Eugenio Tacchini, Stefano Moret, Gabriele Ballarin, Massimo DiPierro, and Luca de Alfaro. 2018. Automatic online fake news detection combining content and social signals. In *2018 22nd Conference of Open Innovations Association (FRUCT)*, pages 272–279. IEEE.

[7] William Ferreira and Andreas Vlachos. 2016. Emer-gent: a novel data-set for stance classification. In *Proceedings of the 2016 conference of the North American chapter of the association for computa-tional linguistics: Human language technologies*, pages 1163–1168.

[8] Lucas Graves. 2018. Understanding the promise and limits of automated fact-checking. *Factsheet*, 2:2018–02.

[9] Andreas Hanselowski, Hao Zhang, Zile Li, Daniil Sorokin, Benjamin Schiller, Claudia Schulz, and Iryna Gurevych. 2018. Ukp-athene: Multi-sentence textual entailment for claim verification. *arXiv preprint arXiv:1809.01479*.

[10] Cecilia Kang and Adam Goldman. 2016. In washing-ton pizzeria attack, fake news brought real guns. *the New York Times*.

[11] Hamid Karimi, Proteek Roy, Sari Saba-Sadiya, and Jiliang Tang. 2018. Multi-source multi-class fake news detection. In *Proceedings of the 27th Inter-national Conference on Computational Linguistics*, pages 1546–1557.

[12] Yoon Kim. 2014. Convolutional neural net-works for sentence classification. *arXiv preprint arXiv:1408.5882*.

[13] Angelika Kirilin and Micheal Strube. 2018. Exploit-ing a speakers credibility to detect fake news. In *Proceedings of Data Science, Journalism & Media workshop at KDD (DSJM18)*.

[14] Elena Kochkina, Maria Liakata, and Arkaitz Zubiaga. 2018. All-in-one: Multi-task learning for rumour verification. *arXiv preprint arXiv:1806.03713*.

[15] Yunfei Long, Qin Lu, Rong Xiang, Minglei Li, and Chu-Ren Huang. 2017. Fake news detection through multi-perspective speaker profiles. In *Pro-ceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, volume 2, pages 252–256.

[16] Rada Mihalcea and Carlo Strapparava. 2009. The lie detector: Explorations in the automatic recognition of deceptive language. In *Proceedings of the ACL-IJCNLP 2009 Conference Short Papers*, pages 309– 312. Association for Computational Linguistics.

[17] Tomas Mikolov, Kai Chen, Greg Corrado, and Jef-frey Dean. 2013. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

[18] Tanushree Mitra and Eric Gilbert. 2015. Credbank: large-scale social media corpus with associated credibility annotations. In *ICWSM*, pages 258–267.

*[19]* Ndapandula Nakashole and Tom M Mitchell. 2014. Language-aware truth assessment of fact candidates. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume Long Papers)*, volume 1, pages 1009–1019.

[20] Jeffrey Pennington, Richard Socher, and Christopher Manning. 2014. Glove: Global vectors for word representation. In *Proceedings of the 2014 confer-ence on empirical methods in natural language processing (EMNLP)*, pages 1532–1543.

[21] Trung Tin Pham. 2018. A study on deep learning for fake news detection.

[22] Martin Potthast, Johannes Kiesel, Kevin Reinartz, Janek Bevendorff, and Benno Stein. 2017. A stylometric inquiry into hyperpartisan and fake news. *arXiv preprint arXiv:1702.05638*.

[23] Hannah Rashkin, Eunsol Choi, Jin Yea Jang, Svitlana Volkova, and Yejin Choi. 2017. Truth of varying shades: Analyzing language in fake news and polit-ical fact-checking. In *Proceedings of the 2017 Con-ference on Empirical Methods in Natural Language Processing*, pages 2931–2937.

[24] Victoria L Rubin, Yimin Chen, and Niall J Conroy. 2015a. Deception detection for news: three types of fakes. In *Proceedings of the 78th ASIS&T An-nual Meeting: Information Science with Impact: Research in and for the Community*, page 83. American Society for Information Science.

[25] Victoria L Rubin, Niall J Conroy, and Yimin Chen. 2015b. Towards news verification: Deception de-tection methods for news discourse. In *Proceedings of the Hawaii International Conference on System Sciences (HICSS48) Symposium on Rapid Screening Technologies, Deception Detection and Credibility Assessment Symposium, January*, pages 5–8.

[26] Victoria L Rubin and Tatiana Vashchilko. 2012. Iden-tification of truth and deception in text: Application of vector space model to rhetorical structure theory. In *Proceedings of the Workshop on Computational Approaches to Deception Detection*, pages 97–106. Association for Computational Linguistics.

[27] Natali Ruchansky, Sungyong Seo, and Yan Liu. 2017. 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.

[28] Giovanni C Santia and Jake Ryland Williams. 2018. Buzzface: A news veracity dataset with facebook user commentary and egos. In *ICWSM*, pages 531– 540.

[29] Kai Shu, Deepak Mahudeswaran, Suhang Wang, Dongwon Lee, and Huan Liu. 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*.

[30] Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu. 2017a. Fake news detection on social me-dia: A data mining perspective. *ACM SIGKDD Ex-plorations Newsletter*, 19(1):22–36.

[31] Kai Shu, Suhang Wang, and Huan Liu. 2017b. Exploit-ing tri-relationship for fake news detection. *arXiv preprint arXiv:1712.07709*.

[32] Eugenio Tacchini, Gabriele Ballarin, Marco L Della Vedova, Stefano Moret, and Luca de Al-faro. 2017. Some like it hoax: Automated fake news detection in social networks. *arXiv preprint arXiv:1704.07506*.

[33] James Thorne and Andreas Vlachos. 2018. Automated fact checking: Task formulations, methods and fu-ture directions. *COLING*.

[34] James Thorne, Andreas Vlachos, Christos Christodoulopoulos, and Arpit Mittal. 2018. Fever: a large-scale dataset for fact extraction and verification. *arXiv preprint arXiv:1803.05355*.

[35] Andreas Vlachos and Sebastian Riedel. 2014. Fact checking: Task definition and dataset construction. In *Proceedings of the ACL 2014 Workshop on Lan-guage Technologies and Computational Social Science*, pages 18–22.

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

[37] Wenpeng Yin and Dan Roth. 2018. Twowingos: Atwo-wing optimization strategy for evidential claim verification. *arXiv preprint arXiv:1808.03465*.

[38] Arkaitz Zubiaga, Ahmet Aker, Kalina Bontcheva,Maria Liakata, and Rob Procter. 2018. Detection and resolution of rumours in social media: A survey.*ACM Computing Surveys (CSUR)*, 51(2):32.

[39] Arkaitz Zubiaga, Maria Liakata, Rob Procter, Geral-dine Wong Sak Hoi, and Peter Tolmie. 2016. Analysing how people orient to and spread rumours in social media by looking at conversational threads. *PloS one*, 11(3):e0150989.