International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 01, January-2020

Fake News Identification on Social Media

Moin Khan
Department of Computer Engineering,
Sinhgad College of Engineering,
Savitribai Phule Pune University,
Pune, India.

Amisha Jain
Department of Computer Engineering,
Sinhgad College of Engineering,
Savitribai Phule Pune University,
Pune, India.

Rishi Chouhan
Department of Computer Engineering,
Sinhgad College of Engineering,
Savitribai Phule Pune University,
Pune, India.

Sakeeb. H. Sheikh
Department of Computer Engineering,
Sinhgad College of Engineering,
Savitribai Phule Pune University,
Pune, India.

***Abstract*— Fake news proves to have a great impact on society as well as the public. It not only affects people's perception but also fails to preserve the traditional news ecosystem based on the pillars of truth and reality. Considering this situation that affects the public worldwide, here we propose an application that can identify any false information that gets circulated through social media. Our system is proposed with a goal to identify the fake news by making comparisons with the existing facts and data which are available in our datasets. The text information given by the user as an input to the system can be easily distinguished either as fake or real with respective tags attached in the output. Our proposed model enables the ability to identify fake and misleading information and thus retain the trust of the public, leading to the protection of society from the negative impacts of fake news.**

Keywords*— Word embeddings, sentence semantics, fake message, word2vec, doc2vec.

I. INTRODUCTION

In the past decade, social media has become more and more trendy for news utilization due to its easy access, fast propagation, and low cost. However, social media also enables the ample proliferation of "fake news," i.e., news with purposely false information. Fake news on social media can have noteworthy negative societal effects. Therefore, fake news discovery on social media in recent times has become a promising research area that is attracting great attention.

The spread of fake news through various message sharing applications and social media has been increasing on a tremendous scale. During natural calamities, news spread's without any authentication thus affecting the trust of common people as they tend to believe all the contents forwarded to them. Social media has become one of the best media for the widespread of fake news. This extensive propagation of information may have a negative impact on individuals as well the society breaking the authenticity balance of the news ecosystem. It not only changes the way of interpretation but also intentionally persuades consumers to accept biased or false beliefs. Some news is intentionally created to trigger people's distrust and make them confused, impeding their abilities to differentiate what is wrong and what is true.

The propagation of misleading data in everyday access media outlets such as social media feeds, news blogs, and online newspapers have made it exigent to identify truthful news sources, thus increasing the need for computational tools able to endow with insights into the reliability of online content. In this paper, the spotlight is on the automatic identification of fake content in online news.

First, we introduce data in the database via various online sources eg. timesofindia.com, news.google.co.in for the task of fake news detection, covering the domain of the politics. If in case the relevant information is not found, then the content is checked on the Search Engine and topmost five to seven results are taken into consideration.

Secondly, we perform a set of learning experiments to build accurate fake news detectors where different NLP operations are performed to standardize the message and to transform the message into the vector form. By comparing the given transformed message vector with the stored vectors in the database, a score is generated. Based on this score, 'fake' or 'real' tag is attached to the message and shown along with the message.

II. LITERATURE SURVEY

In this section ,we are going to discussed some past research that have been done in fake message identification ,their benefits. limitations and technologies used.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 01, January-2020

Table 1:Summary of Literature Survey

SR. No.	TITLE	PUBLISHER	DESCRIPTION	BENEFITS	LIMITATIONS
01	Design Exploration of Fake News: A Transdisciplinary Methodological Approach to Understanding Content Sharing and Trust on Social Media[1]	Jaigris Hodson, *Royal Roads University;* Brian Traynor, *Mount Royal University* (2018 IEEE International Professional Communication Conference)	In this paper, Publishers recommended a transdisciplinary transfer in the direction of researching fake news that takes into account algorithmic tactics, psychometric data, and qualitative explorations of user actions.	1) Provided fresh tactics to identifying fake news, via both algorithmically and user experience research. 2) Planned how diverse tools and methods may need to be employed collectively, in a transdisciplinary tactic to indulgent user habits, in order to fully address the stuff of trust, news sharing, and the stretch of misrepresentation.	The shortcoming of the experiment is that it requires both labelled datasets and communities of experts to help train applications to recognize and categorize contents. So, work is still essential to recognize how human judgement of fake news takes place.
02	Fake and Spam Messages: Detecting Misinformation during Natural Disasters on Social Media[2]	Meet Rajdev and Kyumin Lee, *Department of Computer Science, Utah State University (*2015 IEEE/WIC/ACM International Conferenceon Web Intelligence and Intelligent Agent Technology)	1) In this paper, Publishers conducted a case study of 2013 Moore Tornado and Hurricane Sandy. 2) Pilot grades showed that the projected tactics classify spam and fake messages with 96.43% accuracy and 0.961 F-measure	1) In this paper, Publishers conducted a case study of 2013 Moore Tornado and Hurricane Sandy. 2) Pilot grades showed that the projected tactics classify spam and fake messages with 96.43% accuracy and 0.961 F-measure	One mislaid study is when would be a right prompt to develop fake and spam tweet forecaster while streaming data is looming.
03	Classifying Fake News Articles Using Natural Language Processing to Identify In-Article Attribution as a Supervised Learning Estimator.[3]	Terry Traylor, U.S. Marine Corps, Fargo, ND Jeremy Straub, Gurmeet, Nicholas Snell (Department of Computer Science North Dakota State University), Fargo, ND (2019 IEEE 13th International Conference on Semantic Computing (ICSC))	1) In this paper, the research process, methodical analysis, technical semantics work, and classifier performance and results are offered. 2) The paper concludes with a discourse of how the current system will advance into an impact mining system.	1) Influence mining technique is offered as a mode that can be used to enable fake news and even advertising detection 2) This paper presented the consequences of a study that produced a restricted fake news detection system. 3) The initial performance of system is usually encouraging, because fake news is intended to deceive human targets, so a initial classification tool with only one removal feature seems to do well.	1) The attribution-based fake news discovery tool that uses the quote ascription classifier, however, like the attribution classifier, it did not perform well enough for production use. 2) Upon review, some of the missed labels were attributable to fake news forms with no quotations, fake news documents with credited quotes of imprecise statements, and fake news documents that quoted or cited other fake news documents. 3) The overall act results for this system are not as robust as desired.
04	Credibility Assessment of Textual Claims on the Web[4]	Kashyap Popat, Subhabrata Mukherjee, Jannik Strötgen, Gerhard Weikum (Max Planck Institute for Informatics, Saarbrücken, Germany)	1) This paper recommends the use of a method lever-ages the joint communication between the language of articles about the claim and the reliability of the basic web sources. 2) Experiments with claims from the popular website snopes.com and from reported cases of Wikipedia hoaxes prove the viability of the projected methods and their superior accuracy over various baselines.	1)This paper addresses the assessing the integrity of arbitrary claims made in natural-language text - in an open-domain setting deprived of any assumptions about the structure of the claim, or the community where it is made. 2) Solution is based on automatically finding sources in news and social media, and feeding these into a far supervised classifier for measuring the credibility of a claim (i.e., true or fake). 3) The work in this paper aims to replace this blue-collar confirmation with an automated system.	Can't examine the role of attribution or speaker information, refined linguistic aspects like denial, and understanding the article's view about the claim.
05	Message Authentication System for Mobile Messaging Applications[5]	Ankur Gupta, Purnendu Prabhat, Rishi Gupta, Sumant Pangotra, Suave Bajaj, (Department of Computer Science and	This paper proposes a system enabling users to confirm the authenticity of messages recognised through message sharing applications.	1) The Message Authentication System (MAS) builds a hierarchical database of faithful information through mining a multitude of trusted sources on the internet and social media feeds.	The proposed system works as a third-party data authentication service capable of working with a wide-variety of message sharing requests or social network platforms which

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 01, January-2020

		Engineering, Model Institute of Engineering and Technology, Jammu, India) 2017 International Conference on Next Generation Computing and Information Systems (ICNGCIS)		2) The user can choose to forward the authenticated message to her contacts with the appended legitimacy index thereby helping prevent the spread of spin.	involve large-scale data sharing and forwarding.

III. GAP ANALYSIS

In [2]. User experience, more specifically how user gets engaged and how they tend to recognize fake news is one concept which are focused only by a few. This paper thus proposes a comparatively different approach for analyzing user behaviour. It uses factors such as trust, loyalty, appearance as well as the usability inorder to identify fake news. It also provides results of their analysis performed on new sites.

In [3], Natural Language Processing is a novel method used for detection of fake news. In 2019 Terry Taylor et.al proposed methods for detection of fake news using natural language processing. The paper benefits us the future researchers with detailed technical analysis. Also it uses textblob, sciPy Toolkit to develop a unique classifier that could later detect fake and misleading information. It also provides results based on the performance of classifier as well as the precision value of the proposed system.
In [4],Credibility of messages has been a crucial problem .The paper published in 2016 by Kashyap Popat et.al concentrates on the credibility of messages that are being spread. It proposes methods in which the sources of news are feeded to a supervised classifier and thus checking the credibility. The paper succeeds by providing various results extracted by performing practical implications on sites such as snopes.com and Wikipedia.

In [5], Rajdev et. al (2015) proposed a system which basically focuses on detection of spam messages during natural calamities using classification as well as feature detection of information that gets tweeted during natural calamities. The classification approach consists of flat as well as hierarchical classification of information as either legitimate or fake but the feature detection focuses on identifying the piece of information based on specific features. The proposed system is claimed to be successful by giving an accuracy of 96.43percent.

IV. PROPOSED SYSTEM

In this we have proposed a system where the user will be provided with the messaging application which have an extra feature of identifying the legitimacy of the message it has received by verifying it over our server. Our server contains a machine learning model which is based on word embedding based machine learning algorithm trained on data gathered from various sources like news sites, fact checking sites and search engine in case if the data relevant to users' data is not available in our database. Our model transforms the message sent by the user into numerical vector using word2vec and doc2vec word embedding techniques .Next, it identifies the legitimacy of the message using score generator module which takes help of distance calculator techniques like cosine similarity. Our proposed system then attaches the tag of fake or real to the user message and sends back to the user. Hence Our proposed mechanism doesn't intend to stop the extensive spread of false news but instead, it makes the user aware of the authenticity of the information, thus developing an application that can identify the fake news.

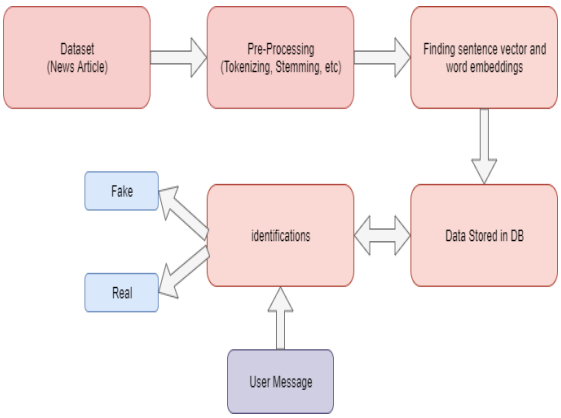

Fig. 1. Workflow Model

V. ARCHITECTURE

In our proposed system,we have three main modules which are shown in the fig . namely user application side ,web server side and database server.The modules interacts with each other ,the flow is initiated by actor(user) which interacts with the user application(an android application) ,it sends a message for verification over the HTTP network to the web server where different NLP operations are performed to standardize the message and to transform the message into the vector form which is required to generate the score by comparing given transformed message vector with the stored vectors in the database.Based on the score 'fake' or 'real' tag is attached to the message and sent back to the actor.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 01, January-2020

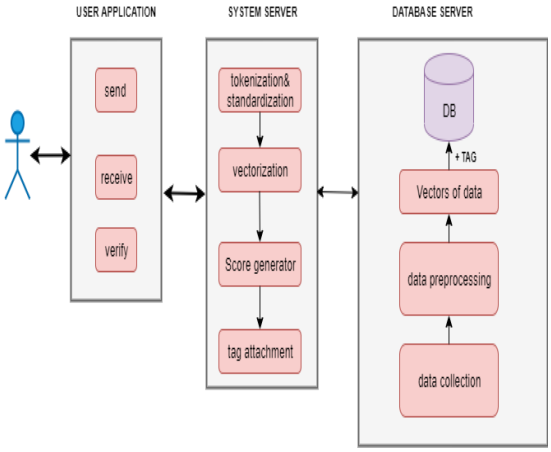

Fig 2. System Architecture

A. *User application*

This module is nothing but a messaging application developed in android. It provides basic chatting functionality to the user like sending ,receiving with extra functionality of verification of the message.

B. *System Server*

This module is responsible for identification of received message either 'real' or 'fake' before reverting back the message to the user. Before sending back ,the message goes through number of stages as follows:

C. *Tokenization and Standardization*

In this module, the received message which is in text form ,which is first broken in tokens and then it is converted into standardized form by performing different NLP techniques.

D. *Vectorization*

The converted message into standardized form is transformed into numerical vector using pre-trained machine learning model based on word embedding technique namely word2vec and USE(Universally Sentence Encoder).

E. *Score generator And Tag Attacher*

Message vector of given message is compared with the vectors of data stored in the database. It makes use of vector similarity algorithms like cosine similarity, which gives results in 0 to 1.The average of given scores are find out and if the average is greater than 0.5 then the message is identified as real otherwise fake. The respective tag of real or fake is attached to the message and it is sent back to the actor(user).

F. *Database Server*

This module is responsible for storing the data in vector form. The data gathered from different sources are converted into vectors before storing into the database.

G. *Data Collection*

The data is collected from different sources like fact checking sites like snopes.com ,news sites by web scraping and search engine.

H. *Data Pre-processing*

This module deals with pre-processing of the data .Different operations like tokenization and standardization are performed to convert all the data into standard form.

I. *Data Vectors*

The web embedding techniques like word2vec and USE are used to convert the data in text form into numerical vector form. This transformed data is then stored into the database.

VI. METHODOLOGY

There are different approaches proposed by the experts to identify the authenticity of the message. But we found out that semantics of the message plays an important role in finding the context of the messages. Context of the message gives the intent of the overall messages ,which is essential for identification of the given message as real or fake. It is difficult to compare two sentence in their original form to find the semantics difference in them, so we make use of a method which is based on finding the relation between neighbouring words called as word embedding, which transforms given text into numerical vectors .It is easy to compare two or more data in vector form than text form. So we used a methodology where we used word embedding based techniques to find the semantics of the given data.

A. *Word Embeddings*

Word embedding[6] is a form of word representation that is used to bridge the gap between human language and machine. It represents the words, phrases or vectors in n-dimensional space called as vectors or encoded vectors. Word embedding is an approach which learns from corpus of data and finds the encoded vectors for each word in such a manner that words with similar meaning or context have similar vector representation. This is contrasted to the traditional bag of words where each unique word has different distributed representation , regardless of how they are used. Here it is based on deeper linguistic theory , called as "distributional hypothesis" by Zellig Harris[10] that could be summarized as: words that have similar context will have similar meanings.

B. *Word2Vec*

Word2Vec is a method to construct word-embedding between the words. It was developed by Tomas Mikolov, et al. [7]at Google in 2013 as a response to make the neural-network-based training of the embedding more efficient. It intends to give that: a numeric representation for each word, that will be able to capture such relations . There can be different relation for each distributed representation between such word like synonyms, antonyms, or analogies, such as this one:

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 01, January-2020

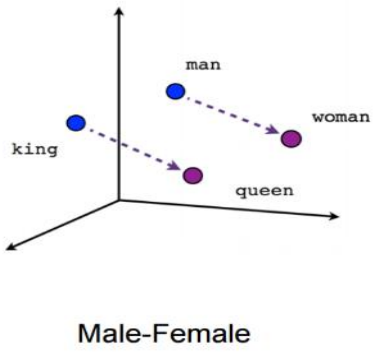

Fig 3. Word Embedding

Here, woman = man – king + queen.

Word2Vec forms embedding uses Skip-gram[8] and Common Bag of Words (CBOW).

- *Common Bag of Words (CBOW)*

CBOW learns embedding for current words by creating a sliding a window around the current word to predict it context from surrounding words. Consider this example: Word2Vec is a word embedding technique.

Let the input to the Neural Network be the word, WORD2VEC. Notice that here we are trying to predict a target word (Word2vec) using a single context input words{word, embedding, technique }. More specifically, we use the one hot encoding of the input word and measure the output error compared to one hot encoding of the target word (Word2Vec). In the process of predicting the target word, we learn the vector representation of the target word.

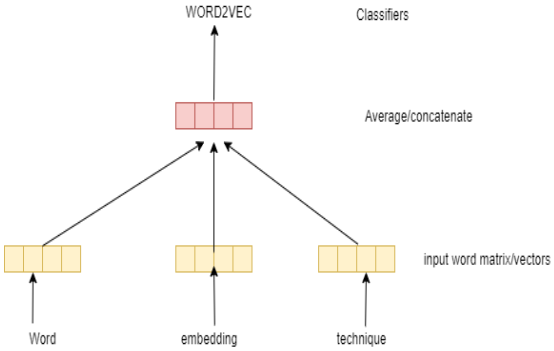

Fig 4. CBOW Algorithm Sketch

- *Skip Gram*

This approach is the opposite of CBOW ,instead of predicting current word from surrounding words, it predicts surrounding words vectors from current word. Like here we are predicting vectors for {word, embedding ,technique} from input word {word2vec}.

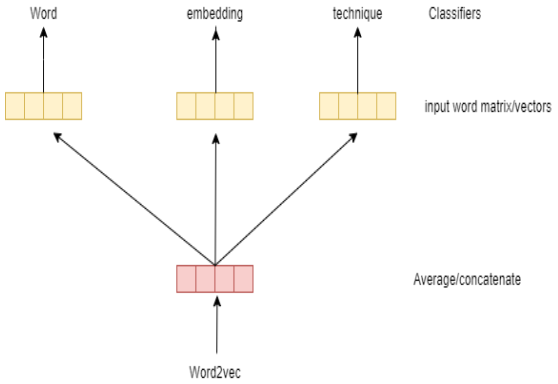

Fig 5.Skip Gram

Both have their own advantages and disadvantages. According to Mikolov, Skip Gram works well with small amount of data and is found to represent rare words well. On the other hand, CBOW is faster and has better representations for more frequent words.

C. Doc2Vec

Doc2vec [9] can be considered as an extension to word2vec ,here we are finding the hot encoded vector for each document ,paragraph or sentence. But compared to words, representing documents in numerical form is difficult as they don't always come in structural manner, so there was a need to handle such unstructured documents. The concept that Mikilov and Le have used was simple: they have used the word2vec model, and added another vector (Paragraph ID below). Similar to word2vec ,Doc2vec also has two methods.

- *Distributed Memory version of Paragraph Vector (PV-DM)*

It's similar to CBOW but instead of using just words to predict the next word, we also added another feature vector, which is document-unique. So, when training the word vectors W, the document vector D is trained as well, and in the end of training, it holds a numeric representation of the document.

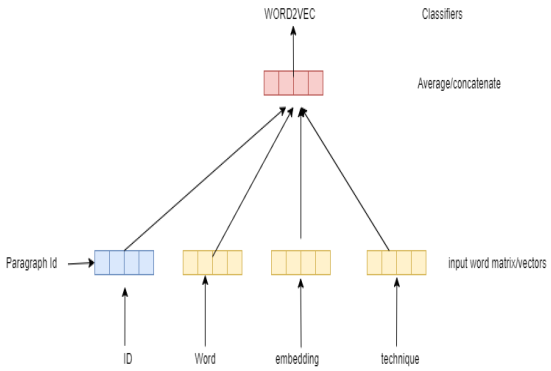

Fig 6. .PV-DM

- Distributed Bag of Words version of Paragraph Vector (PV-DBOW)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 01, January-2020

It's similar to skip gram model of word2vec Here, this algorithm is actually faster as compared to word2vec and consumes less memory, since there is no need to save the word vectors. The doc2vec models may be used in the following way: for training, a set of documents is required. A word vector W is generated for each word, and a document vector D is generated for each document.

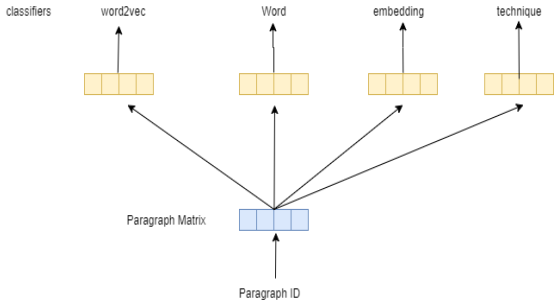

Fig 7. PVDBOW

VII. CONCLUSION

In this paper, we conducted a literature survey of existing fake detection systems which were proposed earlier. Observation of gaps from existing studies in the domain led to the revised system architecture which is explained in detail. A conceptual framework is proposed based on the gaps in fake news detection for public. The framework compares the already existing methods for detecting political news or any other fake news on social media. Focusing on these results, fake news identification application is proposed and practical implications are also discussed.

REFERENCES

[1] Jaigris Hodson,Brian Traynor, "Design Exploration of Fake News: A Transdisciplinary Methodological Approach to Understanding Content Sharing and Trust on Social Media" 2018 IEEE International Professional Communication Conference

[2] Terry Traylor, Jeremy Straub,Gurmeet and Nicholas Snell, "Classifying Fake News Articles Using Natural Language Processing to Identify In-Article Attribution as aSupervised Learning Estimator," in IEEE (ICSC),2019.

[3] Terry Traylor, Jeremy Straub,Gurmeet and Nicholas Snell, "Classifying Fake News Articles Using Natural Language Processing to Identify In-Article Attribution as aSupervised Learning Estimator," in IEEE (ICSC),2019.

[4] KashyapPopat, Subhabrata Mukherjee, JannikStrotgen and Gerhard Weikum, "Credibility Assessment of Textual Claims on the Web, " in ACM,2016.

[5] Meet Rajdev and KyuminLee , "Fake and Spam Messages: Detecting Misinformation during Natural Disasters on Social Media ," in IEEE/WIC/ACM International Conference , 2015.

[6] Zhang, Ye & Rahman, Md Mustafizur & Braylan, Alex & Dang, Brandon & Chang, Heng-Lu & Kim, Henna & McNamara, Quinten & Angert, Aaron & Banner, Edward & Khetan, Vivek & McDonnell, Tyler & Nguyen, An & Xu, Dan & Wallace, Byron & Lease, Matthew. (2016). Neural Information Retrieval: A Literature Review.

[7] T. Mikolov, K. Chen, G. Corrado, J. Dean, "Efficient Estimation of Word Representations in Vector Space," Proc. Workshop at ICLR, 2013.

[8] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, J. Dean, "Distributed Representations of Words and Phrases and their Compositionality", *Proc. NIPS*, 2013.

[9] Quoc V. Le and Tomas Mikolov "Distributed Representations of Sentences and Documents" Proc NIPS, 2014.

[10] Harris, Zellig. Distributional structure. Word, 1954.

[11] Mikolov, Tomas. Statistical Language Models based on Neural Networks. PhD thesis, Brno University of Technology, 2012.

[12] Mikolov, Tomas, Le, Quoc V., and Sutskever, Ilya. Exploiting similarities among languages for machine translation. CoRR, abs/1309.4168, 2013b.

[13] Mikolov, Tomas, Sutskever, Ilya, Chen, Kai, Corrado, Greg, and Dean, Jeffrey. Distributed representations of phrases and their compositionality. In Advances on Neural Information Processing Systems, 2013c.