INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

Fake Detector Effective Fake News Detection with Deep Diffusion Neural Network

C.Sukumar, K. Preethi

Assistant Professor, Department of CSE, Annapoorana Engineering College (Autonomous), Sankari Main Road,

Periya Seeragapadi, Tamil Nadu, India

PG Student, Department of CSE, Annapoorana Engineering College (Autonomous), Sankari Main Road,

Periya Seeragapadi, Tamil Nadu, India

ABSTRACT: In recent years, due to the booming development of online social networks, fake news for various commercial and political purposes has been appearing in large numbers and widespread in the online world. With deceptive words, online social network users can get infected by these online fake news easily, which has brought about tremendous effects on the offline society already. An important goal in improving the trustworthiness of information in online social networks is to identify the fake news timely. This paper aims at investigating the principles, methodologies and algorithms for detecting fake news articles, creators and subjects from online social networks and evaluating the corresponding performance. This paper addresses the challenges introduced by the unknown characteristics of fake news and diverse connections among news articles, creators and subjects. This paper introduces a novel automatic fake news credibility inference model, namely FAKEDETECTOR. Based on a set of explicit and latent features extracted from the textual information, FAKEDETECTOR builds a deep diffusive network model to learn the representations of news articles, creators and subjects simultaneously. Extensive experiments have been done on a real-world fake news dataset to compare FAKEDETECTOR with several state-of-the-art models, and the experimental results have demonstrated the effectiveness of the proposed model

KEY WORDS: Fake News Detection; Diffusive Network; Text Mining; Data Mining

I. INTRODUCTION

The spams usually exist in personal emails or specific review websites and merely have a local impact on a small number of audiences, while the impact fake news in online social networks can be tremendous due to the massive user numbers globally, which is further boosted by the extensive information sharing and propagation among these users audiences' initiative: instead of receiving spam emails passively, users in online social networks may seek for, receive and share news information actively with no sense about its correctness; and (3) identification difficulty: via comparisons with abundant regular messages (in emails or review websites), spams are usually easier to be distinguished; meanwhile, identifying fake news with erroneous information is incredibly challenging, since it requires both tedious evidence-collecting and careful factchecking due to the lack of other comparative news articles available. These characteristics aforementioned of fake news pose new challenges on the detection task. Besides detecting fake news articles, identifying the fake news creators and subjects will actually be more important, which will help completely eradicate a large number of fake news from the origins in online social networks.

These characteristics aforementioned of fake news pose new challenges on the detection task. Besides detecting fake news articles, identifying the fake news creators and subjects will actually be more important, which will help completely eradicate a large number of fake news from the origins in online social networks. Generally, for the news creators, besides the articles written by them, we are also able to retrieve his/her profile information from either the social network website or external knowledge libraries, e.g., Wikipedia or government-internal database, which will provide fundamental complementary information for his/her background check. Meanwhile, for the news subjects, we can also obtain its textual descriptions or other related information, which can be used as the foundations for news subject credibility inference. From a higher-level perspective, the tasks of fake news article, creator and subject detection are highly correlated, since the articles written from a trustworthy person should have a higher credibility, while the person who frequently posting unauthentic information will have a lower credibility on the other hand. Similar correlations can also be observed between news articles and news subjects. In the following part of this paper, without clear specifications, we will use the general fake news term to denote the fake news articles, creators and subjects by default.

1.1 Background

Due to the increasing number of users in social media, news articles can be quickly published or share among users without knowing its credibility and authenticity. Fast spreading of fake news articles using different social media platforms can create inestimable harm to society. These actions could seriously jeopardize the reliability of news media platforms. So it is imperative to prevent such fraudulent activities to foster the credibility of such social media platforms. An efficient automated tool is a primary necessity to detect such misleading articles. Considering the issues mentioned earlier, in this paper, we propose a hybrid model using multiple branches of the convolutional neural network (CNN) with Long Short Term Memory (LSTM) layers with different kernel sizes and filters.

II. LITERATURE SURVEY

Title -Fake News Detection Using Machine Learning Approaches

Author-Z Khanam[1], B N Alwasel[1], H Sirafi[1] and M Rashid[2]

The fake news on social media and various other media is wide spreading and is a matter of serious concern due to its ability to cause a lot of social and national damage with destructive impacts. A lot of research is already focused on detecting it. This paper makes an analysis of the research related to fake news detection and explores the traditional machine learning models to choose the best, in order to create a model of a product with supervised machine learning algorithm, that can classify fake news as true or false, by using tools like python scikit-learn, NLP for textual analysis. This process will result in feature extraction and vectorization; we propose using Python scikit-learn library to perform tokenization and feature extraction of text data, because this library contains useful tools like Count Vectorizer and Tiff Vectorizer. Then, we will perform feature selection methods, to experiment and choose the best fit features to obtain the highest precision, according to confusion matrix results.

Title- A Hybrid Model for Effective Fake News Detection with a Novel COVID-19 Dataset

Author- Rohit Kumar Kaliyar1 ,Anurag Goswami1 and Pratik Narang2

Due to the increasing number of users in social media, news articles can be quickly published or share among users without knowing its credibility and authenticity. Fast spreading of fake news articles using different social media platforms can create inestimable harm to society. These actions could seriously jeopardize the reliability of news media platforms. So it is imperative to prevent such fraudulent activities to foster the credibility of such social media platforms. An efficient automated tool is a primary necessity to detect such misleading articles. Considering the issues mentioned earlier, in this paper, we propose a hybrid model using multiple branches of the convolutional neural network (CNN) with Long Short Term Memory (LSTM) layers with different kernel sizes and filters. To make our model deep, which consists of three dense layers to extract more powerful features automatically. In this research, we have created a dataset (FN-COV) collecting 69976 fake and real news articles during the pandemic of COVID-19 with tags like social-distancing, covid19, and quarantine. We have validated the performance of our proposed model with one more real-time fake news dataset: PHEME. The capability of combined kernels and layers of our C-LSTM network is lucrative towards both the datasets. With our proposed model, we achieved an accuracy of 91.88% with PHEME, which is higher as compared to existing models and 98.62% with FN-COV dataset.

Title - A Study of Fake News Detection using Machie Learning Algorithms

K.ArunKumar, G.Preethi, K.Vasanth

In this paper, we explore the application of Natural Language Processing techniques to identify when a news source may be producing fake news. We use a corpus of labeled real and fake new articles to build a classififier that can make decisions about information based on the content from the corpus. We use a text classification approach, using four different classification models, and analyze the results. The best performing model was the LSTM implementation. The model focuses on identifying fake news sources, based on multiple articles originating from a source. Once a source is labeled as a producer of fake news, we can predict with high confidence that any future articles from that source will also be fake news. Focusing on sources widens our article misclassification tolerance, because we then have multiple data points coming from each source.

Title -FakeDetector: Effective Fake News Detection with Deep Diffusive Neural Network

In recent years, due to the booming development of online social networks, fake news for various commercial and political purposes has been appearing in large numbers and widespread in the online world. With deceptive words, online social network users can get infected by these online fake news easily, which has brought about tremendous effects on the offline society already. An important goal in improving the trustworthiness of information in online social networks is to identify the fake news timely. This paper aims at investigating the principles, methodologies and algorithms for detecting fake news articles, creators and subjects from online social networks and evaluating the

corresponding performance. This paper addresses the challenges introduced by the unknown characteristics of fake news and diverse connections among news articles, creators and subjects. This paper introduces a novel gated graph neural network, namely FAKEDETECTOR. Based on a set of explicit and latent features extracted from the textual information

III. EXISTING SYSTEM

In this paper, we propose to study the fake news detection (including the articles, creators and subjects) problem in online social networks. Based on various types of heterogeneous information sources, including both textual contents/profile/descriptions and the authorship and article subject relationships among them, we aim at identifying fake news from the online social networks simultaneously. We formulate the fake news detection problem as a credibility inference problem, where the real ones will have a higher credibility while unauthentic ones will have a lower one instead.

DISADVANTAGES
The fake news detection problem is not easy to address due to the following reasons:
- Problem Formulation: The fake news detection problem studied in this paper is a new research problem, and a formal definition and formulation of the problem is required and necessary before studying the problem.
- Textual Information Usage: For the news articles, creators and subjects, a set of their textual information about their contents, profiles and descriptions can be collected from the online social media. To capture signals revealing their credibility, an effective feature extraction and learning model will be needed.
- Heterogeneous Information Fusion: In addition, as mentioned before, the credibility labels of news articles, creators and subjects have very strong correlations, which can be indicated by the authorship and article-subject relationships between them. An effective incorporation of such correlations in the framework learning will be helpful for more precise credibility inference results of fake news.

IV. PROPOSED SYSTEM

The fake news on social media and various other media is wide spreading and is a matter of serious concern due to its ability to cause a lot of social and national damage with destructive impacts. A lot of research is already focused on detecting it. This paper makes an analysis of the research related to fake news detection and explores the traditional machine learning models to choose the best, in order to create a model of a product with supervised machine learning algorithm, that can classify fake news as true or false, by using tools like python scikit-learn, NLP for textual analysis. This process will result in feature extraction and vectorization; we propose using Python scikit-learn library to perform tokenization and feature extraction of text data, because this library contains useful tools like Count Vectorizer and Tiff Vectorizer. Then, we will perform feature selection methods, to experiment and choose the best fit features to obtain the highest precision, according to confusion matrix results.

ADVANTAGES
- Fake News contains misleading information that could be checked. This maintains lie about a certain statistic in a country or exaggerated cost of certain services for a country, which may arise unrest for some countries like in Arabic spring.
- There are organizations, like the House of Commons and the Crosscheck project, trying to deal with issues as confirming authors are accountable.
- A solution could be, by the development of a system to provide a credible automated index scoring, or rating for credibility of different publishers, and news context.

V. MODULE DESCRIPTION

FAKE NEWS CONSEQUENCES
 There has always been fake news since the beginning of human civilization. However, the spread of fake news is increased by modern technologies and the conversion of the global media landscape. The major consequences on social, political, and economic environments may be caused by fake news. Fake information and fake news have various faces. As information molds our view toward the world, fake news has a huge impact. We make critical decisions based on the information

DATA PRE-PROCESSING

Data pre-processing is utilized to represent complex structures with attributes, binaries attributes, change discrete attributes, persist, and manage lost and obscure attributes. Containing faces wearing masks or faces not wearing masks. During data pre-processing, different visualization procedures are helpful. A cautious pre-processing strategy is required to ingest the data in a neural network for fake news detection because social media data sources are fragmented, unstructured, and noisy. It is a popular fact that amid the learning stage, data pre-processing saves computational time and space. In addition, limiting the impact of artifacts during the learning process, text pre-processing avoids every ingests of noisy data. The data becomes a logical representation after proper text pre-processing. It also included the most representative descriptive words. Umer et al. [42] experimented on a fake news detection model in which the accuracy was only 78% when they used the features excluding data cleaning or pre-processing, which is surprisingly poor. After performing the pre-processing steps and removing unnecessary data, the accuracy increases dramatically to 93.0%. Data quality assessment, dimensionality reduction, and splitting of the dataset are the data pre-processing steps used in various studies

DATA QUALITY ASSESSMENT

Data are frequently taken from numerous sources that are ordinarily reliable and are in completely different formats. When working on a machine learning problem, more time is invested in managing data quality issues. It is unreasonable to anticipate that the data would be perfect. There may be some issues due to a human blunder, defects within the data collection process, or restrictions on measuring gadgets. The quality of a dataset is often responsible for the poor performance of fake news detection models.

TRAIN/VALIDATION/TEST SPLIT BASED

The dataset may be divided into train, test, and validation sets. The sample of data that is utilized to adjust the parameters is called the training set. The validation set is a series of examples used to fine-tune the parameters of a model. A set of examples applied only for assessing a fully-specified model's performance is regarded as the test set. Although many studies on fake news detection have divided theirdataset into training, validation, and test sets, few studies have used only the training, and test sets.

VI. CONCLUSION

Fake news is escalating as social media is growing. Researchers are also trying their best to find solutions to keep society safe from fake news. This survey covers the overall analysis of fake news classification by discussing major studies. A thorough understanding of recent approaches in fake news detection is essential because advanced frameworks are the front-runners in this domain. Thus, we analyzed fake news identification methods based on NLP and advanced DL strategies. We presented a taxonomy of fake news detection approaches. We explored different NLP techniques and DL architectures and provided their strength and shortcomings. We have explored diverse assessment measurements. We have given a short description of the experimental findings of previous studies. In this field, we briefly outlined possible directions for future research. Fake news identification will remain an active research field for some time with the emergence of novel deep learning network architectures. There are fewer chances of inaccurate results using deep learning-based models. We strongly believe that this review will assist researchers in fake news detection to gain a better, concise perspective of existing problems, solutions, and future directions.

VII. FUTURE WORK

In modern years, due to the active development of online social networks, fake news for various marketable and political purposes has been appearing in large numbers and pervasive in the online world. With ambiguous words, online social network users can get contaminated by this online fake news easily, which has brought about incredible effects on the offline society. An important goal in improving the honesty of information in online social networks is to recognize the fake news sensible. This paper aims at investigate the values, methodologies and algorithms for detect fake news article, creators and subjects from online social networks and evaluating the equivalent performance. This paper address the challenges introduce by the unknown distinctiveness of fake news and various connections among news articles, creators and subjects. This paper introduces a new regular fake news credibility inference model, namely FAKEDETECTOR. Based on a set of unambiguous and hidden features extracted from the textual information, FAKEDETECTOR builds a deep diffusive network model to learn the representation of news article, creators and subjects concurrently. Broad experiments have been done on a real-world fake news dataset to evaluate FAKEDETECTOR with quite a few state-of-the-art models, and the new results have demonstrated the efficiency of the proposed model.

REFERENCES

[1] H. Allcott and M. Gentzkow, ''Social media and fake news in the 2016 election,'' J. Econ. Perspect., vol. 31, no. 2, pp. 36–211, 2017.

[2] T. Rasool, W. H. Butt, A. Shaukat, and M. U. Akram, ''Multi-label fake news detection using multi-layered supervised learning,'' in Proc. 11th Int. Conf. Comput. Autom. Eng., 2019, pp. 73–77.

[3] X. Zhang and A. A. Ghorbani, ''An overview of online fake news: Characterization, detection, and discussion,'' Inf. Process. Manage., vol. 57, no. 2, Mar. 2020, Art. no. 102025. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0306457318306794

[4] Abdullah-All-Tanvir, E. M. Mahir, S. Akhter, and M. R. Huq, ''Detecting fake news using machine learning and deep learning algorithms,'' in Proc. 7th Int. Conf. Smart Comput. Commun. (ICSCC), Jun. 2019, pp. 1–5.

[5] K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, ''Fake news detection on social media: A data mining perspective,'' ACM SIGKDD Explorations Newslett., vol. 19, no. 1, pp. 22–36, 2017.

[6] R. Oshikawa, J. Qian, and W. Y. Wang, ''A survey on natural language processing for fake news detection,'' 2018, arXiv:1811.00770.

[7] S. B. Parikh and P. K. Atrey, ''Media-rich fake news detection: A survey,'' in Proc. IEEE Conf. Multimedia Inf. Process. Retr. (MIPR), Apr. 2018, pp. 436–441.

[8] A. Habib, M. Z. Asghar, A. Khan, A. Habib, and A. Khan, ''False information detection in online content and its role in decision making: A systematic literature review,'' Social Netw. Anal. Mining, vol. 9, no. 1, pp. 1–20, Dec. 2019.

[9] M. K. Elhadad, K. F. Li, and F. Gebali, ''Fake news detection on social media: A systematic survey,'' in Proc. IEEE Pacific Rim Conf. Commun., Comput. Signal Process. (PACRIM), Aug. 2019, pp. 1–8.

[10] A. Bondielli and F. Marcelloni, ''A survey on fake news and rumour detection techniques,'' Inf. Sci., vol. 497, pp. 38–55, Sep. 2019. [Online]. Available: http://www.sciencedirect. com/science/article/pii/S0020025519304372

[11] P. Meel and D. K. Vishwakarma, ''Fake news, rumor, information pollution in social media and web: A contemporary survey of state-of-the-arts, challenges and opportunities,'' Expert Syst. Appl., vol. 153, Sep. 2020, Art. no. 112986.

[12] K. Sharma, F. Qian, H. Jiang, N. Ruchansky, M. Zhang, and Y. Liu, ''Combating fake news: A survey on identification and mitigation techniques,'' ACM Trans. Intell. Syst. Technol., vol. 10, no. 3, pp. 1–42, May 2019.

[13] X. Zhou and R. Zafarani, ''A survey of fake news: Fundamental theories, detection methods, and opportunities,'' ACM Comput. Surv., vol. 53, no. 5, pp. 1–40, 2020.

[14] B. Collins, D. T. Hoang, N. T. Nguyen, and D. Hwang, ''Trends in combating fake news on social media—A survey,'' J. Inf. Telecommun., vol. 5, no. 2, pp. 247–266, 2021.

[15] A. Zubiaga, A. Aker, K. Bontcheva, M. Liakata, and R. Procter, ''Detection and resolution of rumours in social media: A survey,'' ACM Comput. Surveys, vol. 51, no. 2, pp. 1–36, Jun. 2018

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

📱 9940 572 462 🟢 6381 907 438 ✉ ijircce@gmail.com