IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 9, Issue 6, June 2021

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.542

9940 572 462 | 6381 907 438 | ijircce@gmail.com | www.ijircce.com

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542 |

|| Volume 9, Issue 6, June 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0906185 |

Fake Detector: Effective Fake News Detection with Deep Diffusive Neural Network

Nishigandha Bhor[1], Komal Nathe[2], Sakshi Ugale[3], Gayatri Raut[4], Ms. Shraddha Banne[5]

UG Student, Dept. of Computer., BVCOERI, Nashik, Maharashtra, India[1,2,3,4]

Assistant Professor, Dept. of Computer., BVCOERI, Nashik, Maharashtra, India[5]

ABSTRACT: Fake news has recently leveraged the power and scale of online social media to effectively spread misinformation which not only erodes the trust of people on traditional presses and journalisms, but also manipulates the opinions and sentiments of the public. Detecting fake news is a daunting challenge due to subtle difference between real and fake news. As a first step of fighting with fake news, this paper characterizes hundreds of popular fake and real news measured by shares, reactions, and comments on Facebook from two perspectives: domain reputations and content understanding. Our domain reputation analysis reveals that the Web sites of the fake and real news publishers exhibit diverse registration behaviors, registration timing, domain rankings, and domain popularity. In addition, fake news tends to disappear from the Web after a certain amount of time. The content characterizations on the fake and real news corpus suggest that simply applying term frequency-inverse document frequency (tf-idf)and Latent Dirichlet Allocation(LDA)topic modeling is in efficient in detecting fake news, while exploring document similarity with the term and word vectors is a very promising direction for predicting fake and real news. To the best of our knowledge, this is the first effort to systematically study domain reputations and content characteristics of fake and real news, which will provide key insights for effectively detecting fake news on social media.

KEYWORDS: Fake News Detection; Diffusive Network; Text Mining; Data Mining

I. INTRODUCTION

The last decade has witnessed the rapid growth and success of online social networks, which has disrupted traditional media by fundamentally changing how, who, when, and where on the distribution of the latest news stories. Unlike traditional newspapers or magazines, anyone can spread any information at any time on many open and always-on social mediaplat forms with out real-world authentications and circulation and spreading's of fake news, social spams, and misinformation Driven by the political or financial incentives, the creators of fake news generate and submit these wellcrafted news stories on online social media, and subsequently recruit social bots or paid spammers to push the news to a certain popularity. The recommendation and ranking algorithms on social media, if failed to immediately detect such fake news, likely surface such news to many other innocent users who are interested in the similar topics and content of the news, thus leading to a viral spreading process on social media. These rising social spams, click baits, and fake news, mixed with real news and credible content, create challenges and difficulties for regular Internet users to distinguish credible and fake content. Towards effectively detecting, characterizing, and modeling Internet fake news on online social media, this paper proposes a new framework which systematically characterizes the Web sites and reputations of the publishers of the fake and real news articles, analyzes the similarity and dissimilarity of the fake and real news on the most important terms of the news articles via term frequency-inverse document frequency (tf-idf) and Latent Dirichlet Allocation (LDA) topic modeling, as well as explores document similarity analysis via Jaccard similarity measures between fake, real, and hybrid news articles. The contributions of this paper can be summarized as Follows:

We systematically characterize the Web sites and reputations of the publishers of the fake and real news articles on their registration patterns, Web site ages, domain rankings, domain popularity, and the probabilities of news disappearance from the Internet. We analyze the similarity and dissimilarity of the fake and real news on the most important terms of the news articles via tf-idf and LDA topic modeling.We explore document similarity between fake, realor hybrid news articles via Jaccard similarity to distinguish, classify, and predict fake and real news.

II. LITERATURE SURVEY

A fake news detection system aims to assist users in detecting and filtering out varieties of potentially deceptive news. The prediction of the chances that a particular news item is intentionally deceptive is based on the analysis of

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542 |

|| Volume 9, Issue 6, June 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0906185 |

previously seen truthful and deceptive news. A scarcity of deceptive news, available as corpora for predictive modeling, is a major stumbling block in this field of natural language processing (NLP) and deception detection. This paper discusses three types of fake news, each in contrast to genuine serious reporting, and weighs their pros and cons as a corpus for text analytics and predictive modeling. Filtering, vetting, and verifying online information continues to be essential in library and information science (LIS), as the lines between traditional news and online information are blurring.

The problem of fake news has gained a lot of attention as it is claimed to have had a significant impact on 2016 US Presidential Elections. Fake news is not a new problem and its spread in social networks is well-studied. Often an underlying assumption in fake news discussion is that it is written to look like real news, fooling the reader who does not check for reliability of the sources or the arguments in its content. Through a unique study of three data sets and features that capture the style and the language of articles, we show that this assumption is not true. Fake news in most cases is more similar to satire than to real news, leading us to conclude that persuasion in fake news is achieved through heuristics rather than the strength of arguments. We show overall title structure and the use of proper nouns in titles are very significant in differentiating fake from real. This leads us to conclude that fake news is targeted for audiences who are not likely to read beyond titles and is aimed at creating mental associations between entities and claims.
We introduce the novel task of determining whether a newswire article is “true” or satirical. We experiment with SVMs, feature scaling, and a number of lexical and semantic feature types, and achieve promising results over the task.
Automatic fake news detection is a challenging problem in deception detection, and it has tremendous real-world political and social impacts. However, statistical approaches to combating fake news has been dramatically limited by the lack of labeled benchmark datasets. In this paper, we present liar: a new, publicly available dataset for fake news detection. We collected a decade-long, 12.8K manually labeled short statements in various contexts from PolitiFact.com, which provides detailed analysis report and links to source documents for each case. This dataset can be used for fact-checking research as well. Notably, this new dataset is an order of magnitude larger than previously largest public fake news datasets of similar type. Empirically, we investigate automatic fake news detection based on surface-level linguistic patterns. We have designed a novel, hybrid convolutional neural network to integrate meta-data with text. We show that this hybrid approach can improve a text-only deep learning model.

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542 |

|| Volume 9, Issue 6, June 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0906185 |

III. PROPOSED ALGORITHM

TFID ALGOEITHM:

TFID ALGORITHM

Input: Chat History ch, URL

Output: result r. (Ranked Statement)

Let,

D is data extract from URL,

words=getWords(chat); //Using Split("\\s+")

Term Frequency TF:

for each statement s ?D

for each word, w ? ch

$$tf = \sum_{k=0}^{n} \alpha/\beta$$

α= Number of times w appears in a s;

β= Total number of terms in s;

end for

end for

Inverse Document Frequency IDF:

for each word, w ? ch

$$idf = \sum_{k=0}^{n} \log\left(\frac{\text{Total No. of Statements}}{\text{No. of Statements w presnet}}\right)$$

end for

return tf*idf;

End;

Fig.1. TFID Algorithm

IV. SYSTEM ANALYSIS

aterfall approach was first SDLC Model to be used widely in Software Engineering to ensure success of the project. In "The Waterfall" approach, the whole process of software development is divided into separate phases. In this Waterfall model, typically, the outcome of one phase acts as the input for the next phase sequentially.

The following illustration is a representation of the different phases of the Waterfall Model

• Requirement Gathering and analysis − All possible requirements of the system to be developed are captured in this phase and documented in a requirement specification document.

• System Design − The requirement specifications from first phase are studied in this phase and the system design is prepared. This system design helps in specifying hardware and system requirements and helps in defining the overall system architecture.

• Implementation − With inputs from the system design, the system is first developed in small programs called units, which are integrated in the next phase. Each unit is developed and tested for its functionality, which is referred to as Unit Testing.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542 |

|| Volume 9, Issue 6, June 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0906185 |

• Integration and Testing − All the units developed in the implementation phase are integrated into a system after testing of each unit. Post integration the entire system is tested for any faults and failures.

• Deployment of system − Once the functional and non-functional testing is done; the product is deployed in the customer environment or released into the market.

• Maintenance − There are some issues which come up in the client environment. To fix those issues, patches are released. Also to enhance the product some better versions are released. Maintenance is done to deliver these changes in the customer environment.

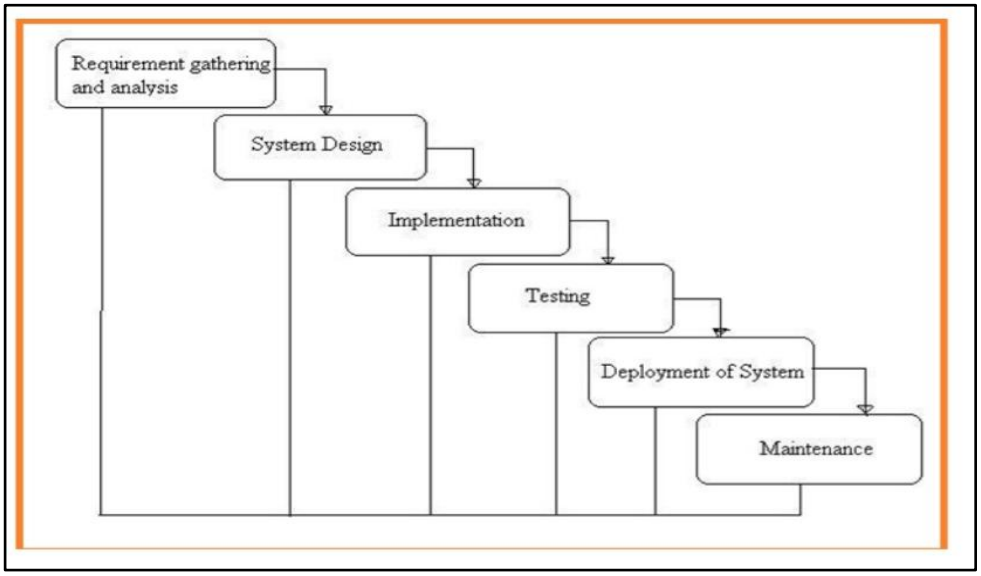

Fig.2. System Analysis

V. CONCLUSION AND FUTURE WORK

As fake news and disinformation continue to grow in online social media, it becomes imperative to gain in depth understanding on the characteristics of fake and real news articles for better detecting and filtering fake news. Towards effectively combating fake news, this paper characterizes hundreds of very popular fake and real news from a variety of perspectives including the domains and reputations of the news publishers, as well as the important terms of each news and their word embeddings. Our analysis shows that the fake and real news exhibit substantial differences on the reputations and domain characteristics of the news publishers. On the other hands, the difference on the topics and word embeddings shows little or subtle difference between fake and real news.

REFERENCES

1. K. Shu, A. Sliva, S. H. Wang, J. L. Tang, and H. Liu, Fake newsdetectiononsocialmedia: Adataminingperspective, ACM SIGKDD Explorati., vol. 19, no. 1, pp. 22–36, 2017.
2. E. Tacchini, G. Ballarin, M. L. D. Vedova, S. Moret, and L. de Alfaro, Some like it hoax: Automated fake news detection in social networks, Tech. Rep. UCSC-SOE-1705, School of Engineering, University of California, Santa Cruz, CA, USA, 2017.
3. M. M. Waldrop, News feature: The genuine problem of fakenews,Proc.Natl.Acad.Sci.USA,vol.114,no.48,pp. 12631–12634, 2017.
4. Z.B.He,Z.P.Cai,andX.M.Wang,Modelingpropagation dynamics and developing optimized countermeasures for rumor spreading in online social networks, in Proc. IEEE 35th Int. Conf. on Distributed Computing Systems, Columbus, OH, USA, 2015.
5. Z. B. He, Z. P. Cai, J. G. Yu, X. M. Wang, Y. C. Sun, and Y. S. Li, Cost-efficient strategies for restraining rumor spreading in mobile social networks, IEEE Trans. Veh. Technol., vol. 66, no. 3, pp. 2789–2800, 2017.
6. C. C. Shao, G. L. Ciampaglia, O. Varol, A. Flammini, and F. Menczer, The spread of fake news by social bots, arXiv preprint arXiv: 1707.07592v1, 2017.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542 |

|| Volume 9, Issue 6, June 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0906185 |

7. N. J. Conroy, V. L. Rubin, and Y. M. Chen, Automatic deception detection: Methods for finding fake news, Proc. Assoc. Inf. Sci. Technol., vol. 52, no. 1, pp. 1–4, 2015.
8. B. Markines, C. Cattuto, and F. Menczer, Social spam detection, in Proc. 5th Int. Workshop on Adversarial Information Retrieval on the Web, Madrid, Spain, 2009.
9. Y. M. Chen, N. J. Conroy, and V. L. Rubin, Misleading online content: Recognizing clickbait as “false news”, in Proc. 2015 ACM on Workshop on Multimodal Deception Detection, Seattle, WA, USA, 2015.

INNO SPACE
SJIF Scientific Journal Impact Factor
Impact Factor: 7.542

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

Scan to save the contact details