Assignment 4: Domain related paper search

Group id: 2 **Submitted to:** Rajeshwari Goudar Ma'am
Group Members: Sanskar Sharma 90 **Date:** 24/02/2020
Pratiksha Sabale 87
Nakul Aggarwal 74

Research Paper 1:

FAKEDETECTOR: Effective Fake News Detection with Deep Diffusive Neural Network

Published on: 10th August, 2019

Authors:

- Jiawei Zhang (IFM Lab, Department of Computer Science, Florida State University, FL, USA)
- Bowen Dong, Philip S. Yu (BDSC Lab, Department of Computer Science, University of Illinois at Chicago, IL, USA)

Summary:

Fake news denotes intentionally presents misinformation or hoaxes spreading through both traditional print news media and recent online social media. Fake news has been existing for a long time, since the "**Great moon hoax**" published in 1835.

In recent years, due to the booming developments 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

This paper includes the study of fake news detection (like 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, it aims at identifying fake news from the online social networks simultaneously. This paper formulates the fake news detection problem as a credibility inference problem, where the real ones will have a higher credibility whereas the unauthenticated ones will have a lower one instead. This work is also supported in part by NSF through grants IIS-1763365 and IIS-1763325.

Based on the news from heterogeneous social networking sites, a set of explicit and latent features capable enough to classify news as fake or real can be extracted from the textual information of news articles, creators and subjects respectively. Furthermore, based on the connections among news articles, creators and news subjects, a deep diffusive network model has also been proposed to incorporate the network structure information into model learning. This paper also introduces a new diffusive unit model, namely **GDU**. Model GDU accepts

multiple inputs from different sources simultaneously, and can effectively fuse these input for output generation with content **"forget"** and **"adjust" gates**.

Research Paper 2:

Machine Learning for Detection of Fake News

Published on: 17th May, 2018

Authors:

- Nicole O'Brien (Master of Engineering in Electrical Engineering and Computer Science at the Massachusetts Institute of Technology)

Summary:

Recent political events has led to an increase in the popularity and spread of fake news. As demonstrated by the widespread effects of the large onset of fake news, humans are inconsistent if not outright poor detectors of fake news. With this, efforts have been made to automate the process of fake news detection. The most popular of such attempts include "**blacklists**" of sources and authors that are unreliable.

The goal of this research paper was to create a tool for detecting the language patterns that characterize fake and real news through the use of machine learning and natural language processing techniques.

The main contribution of this paper is support for the idea that **machine learning** could be useful in a novel way for the task of classifying fake news. Its findings show that after much pre-processing of relatively small dataset, a simple **CNN** is able to pick up on a diverse set of potentially subtle language patterns that a human may (or may not) be able to detect. Many of these language patterns are intuitively useful in a humans manner of classifying fake news. Some such intuitive patterns that our model has found to indicate fake news include generalizations, colloquialisms and exaggerations.

Other contributions of this paper includes the creation of a dataset for the task and the creation of an application that aids in the visualization and understanding of the neural nets classification of a given body text. It could also be useful in researchers trying to develop improved models through the use of improved and enlarged datasets, different parameters, etc.

Research Paper 3:

Fake News: A Survey of Research, Detection Methods, and Opportunities

Published on: 2nd December, 2018

Authors:

- XINYI ZHOU (Syracuse University, USA)
- REZA ZAFARANI (Syracuse University, USA)

Summary:

The rise of fake news during the 2016 U.S. Presidential Election highlighted not only the dangers of the effects of fake news but also the challenges presented when attempting to separate fake news from real news. Recent political events have led to an increase in the popularity and spread of fake news. As demonstrated by the widespread effects of the large onset of fake news, humans are inconsistent if not outright poor detectors of fake news. These days' fake news is creating different issues from sarcastic articles to a fabricated news and plan government propaganda in some outlets. Fake news and lack of trust in the media are growing problems with huge ramifications in our society. Obviously, a purposely misleading story is "fake news " but lately blathering social media's discourse is changing its definition. Some of them now use the term to dismiss the facts counter to their preferred viewpoints. Fake news is one of the biggest scourges in our digitally connected world. That is no exaggeration. It is no longer limited to little squabbles – fake news spreads like wildfire and is impacting millions of people every day.

"Fake news (also known as junk news, pseudo-news, or hoax news) is a form of news consisting of deliberate **disinformation** or hoaxes spread via traditional **news media** (print and broadcast) or online **social media**."

An Overview of this Survey:

This survey aims to present a comprehensive framework to study fake news. Fake news can be studied with respect to four perspectives:

- knowledge-based
- style-based
- credibility-based

This survey compare several fake-news related terms and concepts. Besides this it also provides a definition for fake news. This survey provides the most comprehensive list of fundamental theories that can be utilized when studying fake news.

The goal of this survey has been to comprehensively and extensively review , summarize ,compare and evaluate the current research on fake news. The open issues and challenges are also presented in this survey with potential research tasks that can facilitate further development in fake news research.

Research Paper 4:

Fake News Detection Using Machine Learning

Published on: 05 April 19

Authors:

- Lilapati Waikhoma (Department of Computer Science & Engineering, NIT, Arunachal Pradesh, India)
- Rajat Subhra Goswami (Department of Computer Science & Engineering, NIT, Arunachal Pradesh, India)

Summary:

The Internet has become compulsory in our life. It is now very easy to access the Internet than it was before. There is no doubt that many young people prefer the internet to get their news rather than the newspaper, radio, etc. The Internet provides many opportunities for us, we can search for anything on the internet to clear our doubt and for research purpose also. Simply saying, we can't even imagine our life without the internet.

In a diverse country like India where Internet access has become cheap as compared to the past decade , a lot of people have a convenient access of news through their digital devices relevant to the field of interest If it is about the news, the internet plays a very important role because through the internet, the news widespread very fast. There are so many consequences of fake news, it can cause harm to innocent people. Fake news may be made intentionally or accidentally to give harms to an individual or a group for any purposes, such as for political issues, for religious purposes and so on.

Automatic fake news detection has already been studied for some years. Rubin, et.al in his book along with N.J Conroy and Y. Chen titled **"Automatic deception detection: Methods for finding fake news"** gave a hybrid approach which combines the linguistic features of a language with the network analysis approach which may not be always suitable as the network information may be restricted or not available.

Majority of the datasets available contain short statements as the language used for political information broadcasting on TV interviews, social media posts and tweets which are mostly of short length statements, that's why the detection of fake news is more challenging. Following are the important methodologies that play a crucial role in the making of fake news detection algorithm :

- Textual features extraction(includes bag of words concept , N-grams, TFIDF[term frequency inverse document frequency])
- Categorial features (include hot encoding and label encoding)
- Numerical features (including the calling and normalization)
- Ensemble techniques
- Evaluation

In this paper, we present the task of automatic detection of fake news. We have used a new publicly available fake news dataset the LIAR-dataset. The classification of fake news from

the real news is very crucial task nowadays. Our best performing models achieved accuracies that are comparable to the human ability to spot fake content.

Research Paper 5:

Fake News Detection using Machine Learning and Natural Language Processing

Published on: March 25, 2019

Authors:

- Kushal Agarwalla (Department of Computer Science and Engineering, SRM Institute of Science and Technology, Chennai, India.)
- Shubham Nandan (Department of Computer Science and Engineering, SRM Institute of Science and Technology, Chennai, India.)
- Shubham Nandan (Department of Computer Science and Engineering, SRM Institute of Science and Technology, Chennai, India.)
- D. Deva Hema (Department of Computer Science and Engineering, SRM Institute of Science and Technology, Chennai, India.)

(Summary:

Modern life has become quite suitable and the people of the world have to thank the vast contribution of the internet technology for transmission and information sharing. There is no doubt that internet has made our lives easier and access to surplus information viable.

But at the same time it unfocussed the line between true media and maliciously forged media. Today anyone can publish content – credible or not – that can be consumed by the world wide web. Sadly, fake news accumulates a great deal of attention over the internet, especially on social media. People get deceived and don't think twice before circulating such mis-informative pieces to the far end of the world. This kind of news vanishes but not without doing the harm it intended to cause.

Various models are used to provide an accuracy range of 60-75%. Which comprises of Naïve Bayes classifier,
Linguistic features based, Bounded decision tree model, SVM etc. The parameters that are taken in consideration do not yield high accuracy. The motive of the following paper tends to increase the accuracy of detecting fake news more than the present results that are available.

The following were the relational models that are found useful for making of the algorithm :

- Logistic regression: The LR model uses gradient descent to converge onto the optimal set of weights (θ) for the training set.
- Support vector machine : A Support Vector Machine (SVM) is a supervised machine learning algorithm that can be used for both classification and regression purposes. SVMs are mostly used in classification problems. SVMs are founded on the idea of finding a hyperplane that best divides a dataset into two classes. Support vectors are the data points nearest to the hyperplane, the points of a data set that, if

deleted, would alter the position of the dividing hyperplane. Because of this, they can be considered the critical elements of a data

- **Naïve Bayes Classification with Lid stone smoothing :**In machine learning, Naive Bayes classifiers are a family of simple "probabilistic classifiers" based on applying Bayes' theorem with powerful (naive) independent assumptions between the features. Naive Bayes classifiers are highly scalable, requiring a number of parameters linear in the number of variables (features/predictors) in a learning problem.

A lot of our results circle back to the need for acquiring more accuracy. Generally speaking, simple algorithms perform better on less (less variant) data. Since we had a huge set of data, SVM, Naive Bayes and Logistic Regression underperformed.

Plagiarism Checker