Proposing a Novel Method for Fake News Classification

Sepehr Janghorbani
sj620@rutgers.edu

Kshitij Shah
ks1223@rutgers.edu

Souvick Gosh
sg1223@rutgers.edu

Abstract

Fake news Detection has gained widespread attention especially after the events of the 2016 US presidential election. As a result, many studies have been conducted to tackle the spreading of Fake news. The first step of such tasks would be to classify claims associated based on their credibility. In this study, we try to provide a comprehensive overview of what has already been done in this domain and other similar fields, and then come up with a generalized method based on Deep Neural Nets to classify fake news data based content, style and other features of the given claim. Our experiments conducted on benchmark datasets show that for the given classification task we can obtain up to 72% accuracy by comparing a claim against a knowledge base using information retrieval techniques and 80% accuracy using the hidden style features in the text.

1 Introduction

Many intellectuals have named the year 2016 as the beginning of a new era in modern politics, the "Post-truth" era. Having this in mind, choosing "Post-truth" as Oxford Dictionary's Word of the year (oxf) seems to be less and less surprising, especially after the trends observed during the 2016 US presidential elections and the rise of the far-right movement both in US politics as well as in other parts of the world. For this reason, there is an increasing need to put a restraint on the exploding trends of fake news, thus it seems vital to come up with algorithms and methods to combat this ever-increasing problem. To be more precise, it might seem helpful to have a clearer definition of what we call fake news. Cambridge dictionary defines the above term as "False stories that appear to be news, spread on the Internet or using other media, usually created to influence political views or as a joke".(cam, 2017)

Having this definition in mind, some other more specific terms such as "satire", "propaganda news", and even "rumor" can also be considered subcategories of the wider and broader definition of "fake news". In this study, we try to address the problem of coming up with a classifier for tackling the mentioned problem, on several benchmark datasets which comprise short sentences containing fake and real news obtained from several credible and fake sources (The details are provided in the dataset section)

It is also necessary to mention that although the problem of fake news detection has been addressed in several previous studies along with a challenge called "fake news challenge", no universal model has been proposed yet that is able to give acceptable results. That is, classifying the claim correctly by comparing the claim against the factually supported material for every type of claim regardless of the source, whether it is some blog post, twitter data, mainstream media or just an oral speech.

In the next section we provide a thorough overview of what has already been done, thus giving us some insight on what directions should be taken in the future.

2 Related Works

The massive popularity of social media has led to the availability of large amount of user-generated, unregulated information which lacks in quality and are often unverifiable. Also, the content is generated in real-time in huge volumes (big data) and cannot be filtered or checked manually for veracity. This has resulted in the inundation of the Web with wrong or fake information - some of which are generated with malicious intent, and some for the purposes of humor. Linguistically speaking, a wrong information may be a result of inefficient reporting and may not be intended for the purpose of misleading the audience or readers. However, the word fake involves intended actions for the purposes of presenting a false information as true.

The rise of fake news in social media in recent years and the significant effects of it on the 2016 US elections, several studies have been conducted which relates to fake news, its influence and automatic detection. In this section, we try to mention and analyze the important researches which we find relatable to our work. We categorize them into two subsections: Traditional NLP approaches, and Deep Learning Approaches.

2.1 Traditional Natural Language Processing Approaches

Rubin et al. (2015) identified three types of fake news in their work. They categorized fake news into three distinct categories - serious fabrications, large-scale hoaxes, and humorous fake news. The ability of the social media like Facebook and Twitter to influence the opinions of audiences has led to an increased use of fake information. Not only could such fake information be shared extensively to a large number of readers (or followers) but also at a great speed which makes dissemination of such fake information to have significant impact on politics (voter opinions during election) and online retailers (fake product reviews). It could influence the decision making process of the audience by deceiving them into believing fabricated facts. Consequently, the area of automatically detecting fake news has attracted a number of researches in recent years

Fake news have often been compared to satires where implicit humor is used for criticisms which can be easily detected. Therefore, researchers have often resorted to using linguistic approaches for detecting fake news. Both supervised and unsupervised approaches were used in these researches. Papadopoulou et al. (2017) used a two-level text-based classifier to detect clickbaits. They used a wide variety of morphological, grammatical, stylistic, and word-based features. The authors also used sentiment present in the text to make the classifications. Rubin et al. (2016) used satirical cues to differentiate between fake and true news. Their approach depended on absurdity of the text, punctuations, and grammatical features, and achieved a precision and recall of 90% and 87% respectively. The length and complexity of the sentence, the number of clauses present, and the presence of slangs and swear words were some of the more informative features. Ahmed et al. (2017) used Support Vector Machines with n-gram features in their work. They used tf-idf for feature extraction and linear SVM for the classification, achieving 92% accuracy on 50000 features.

Some researches used hybrid approaches which used network analysis, sentiments, and behavioral information in addition to linguistic features. Conroy et al. (2015)) was one of the first researchers to use network analysis in fake news detection while Mukherjee et al. (2013) used words and the respective part-of-speech tags, together with bigrams to achieve a 68.1% accuracy on Yelp data. It must be noted that the accuracy of the different approaches depend largely on the dataset which is being used. For example, the research on Yelp dataset worked on a unbalanced dataset with only 14% fake reviews.

The authors also used a number of behavioral features like maximum number of reviews, the percentage of positive reviews, review length, reviewer deviation, and maximum content similarity. Similarly, Bhelande et al. (2017) used sentiment analysis using bag of positive and negative words for his Naive Bayesian classifier.

Researchers have also utilized discourse analysis with linguistics to identify instances of deception. Using language markers and rhetorical relations, Pisarevskaya (2017) achieved a f-score of 0.65 using SVM and Random Forest classifiers.

2.2 Deep Learning Approaches

One of the more famous problems of this kind was proposed in the "Fake News Challenge 2017"(FNC) where the participators tried to tackle "stance detection" which can be thought of as a simpler subtask of the original problem. Stance Detection refers to classifying the stance of a claim towards an article as one of the following: Agree, Disagree, or Irrelevant. A number of approaches have been investigated for solving this problem, which includes deep learning and traditional NLP techniques. Studying these approaches can be quite useful as they provide valuable insights for the problem at hand.

Surprisingly, the top teams in the competition use simple but highly optimized methods to tackle the problem. For example, the second and third teams used only simple multilayer Deep Neural Networks with highly optimized hyper-parameters and achieve accuracies of 85-88% (Riedel et al., 2017a; Hanselowski, 2017). The first team introduced a slightly more complicated approach by combining two classifiers, a deep learning model (made up of CNN layers + DL layers) and a gradient boosted tree classifier. In addition, they use hand-made optimized features. (Largent, 1970)

Other more complicated approaches have also been investigated: bidirectional LSTM/GRU architectures some with modifications (Chopra and Jain, 2017; Qi Zeng), ensemble of classifiers (Thorne et al., 2017), vanilla CNNs, Independent Encoders, Conditional Encoder(Neel Rakholia), Multipass conditional encoders, Attentive Readers(in one case with weighted cross entropy function) (Kurt Miller) and bidirectional LSTMs. One team also treated the problem as a regression problem and introduced a new model called Siamese Regression model.(Akshay Agrawal)

Although fake news challenge might seem to be a very similar problem to the problem at hand, the general problem formulation is not ideal. The reason is that the data is very much imbalanced, around 75% of which is unrelated articles (although this is somewhat natural); so, by classifying all the data into the majority class we can achieve 75% accuracy. Also, the fake news training data is much less compared to the real news (details in the data analysis section)

Besides the fake news challenge other studies have also been conducted:

In (Aymanns et al., 2017) the problem of fake news detection in social media is treated as finding spread patterns in a graph of the social media also taking into account whether people support or reject a claim. Their solution uses reinforcement learning. Also in (Kumar) a similar problem formulation has been proposed.

In (Avrahamov, 2017) some kind of a knowledge-base graph is constructed by annotating each article with the information about its authors, topics and main keywords. So the problem is again formulated as finding patterns in a hyper graph. (Chen et al., 2017) have used a dataset of articles obtained from fake and genuine labeled source of news. (having equal numbers of fake/real articles) Their design consists of a 3 layer hierarchical deep attentive reader + pooling to classify test articles. However, one thing to mention in their study is that the dataset can somewhat be distinguished only based on the language since even the simple count-based methods can achieve significant accuracies.

Another somewhat similar problem is detecting rumors on tweets. In (Jin et al., 2017) the related task of detecting rumors based on twitter posts is addressed.The dataset for which is obtained by matching tweets with verified rumor articles.In (Derczynski et al., 2017), Semeval 2017 Rask 8 : RumorEval the task of 4 way classification of rumor tweets have been addressed in which the proposed methods with the highest accuracy used ensemble methods, LSTMs and CNN.

In (Ma et al., 2017) again the problem of classifying rumor/non rumor tweets is modeled as a graph classification, thus mostly finding the pattern of spread of the tweets instead of checking the text itself.In (Ma et al., 2016) RNN is used for classifying rumor/non-rumor tweets.

Another similar problem is the identification of clickbaits. This can be useful and relevant to the problem at hand because many of the fake news posts are also clickbaits,i.e. They entice the user to click on a link. In (Biyani et al., 2016) different types of clickbait posts are mentioned and the gradient boosted decision tree classifier relies heavily on feature engineering (such as similarity of the headline and the article and informality of these posts).

Cao et al. (2017) investigates linear regression, logistic regression and random forests on the clickbait dataset, again using similar feature engineering for tweet's text, title and also keywords. In (Zhou, 2017a) a self attentive network with GRU cells have been proposed to tackle this problem. Similarly in (Zhou, 2017b) attentive based LSTMs have been used for event based twitter/weibo posts.

In (Yang et al., 2017) the problem of detecting satirical news have been addressed using BiRNN architectures with GRU cells and 4 levels of hierarchy augmented with attention Mechanism. Their dataset is articles labeled as satirical/real news labeled only based on the source of the article.

Dataset	Mean Number of Instances	Mean Number of Words	Mean # Characters
Uni. of Washington Fake news data	**60481**	**530**	**NA**
Kaggle Fake News (text)	12999	637	NA
Kaggle Fake News (title)	12138	10.55	65
Fake News Challenge	**49974**	**11**	**69**
LIAR	12791	18	107

Figure 1: Comparison of Candidate Datasets

In (Karadzhov et al., 2017) the problem is formulated as detecting fake news using external online sources. This study combines the BiLSTMs of data from different websites and the given claim to classify it.

In (Wang, 2017), the new LIAR benchmark is introduced for the fake news detection problem (6-way classification), annotated with metadata as well. This dataset is described in details in the dataset section.

In [38] a similar 6-way classification problem is solved using the data obtained from politifact website using Naive Bayes, LSTM and MAXENT algorithms, and the accuracy obtained is around 20%, however when converted to a 2-way classification problem it increases 56%.

(Ruchansky et al., 2017) takes a more universal approach of classifying fake news tweets by not only taking the text itself but also the images accompanying it and the number of likes, reactions, shares and tags for each post. However, for the machine learning part they only use simple models such as SVM, random forest and Logistic Regression.

Finally, in (Shu et al., 2017) a complete overview of the recent approaches towards this and other similar problems has been provided.

3 Dataset

In this section, the detailed analysis of the datasets used for this study are provided.

There is still a lack of standard benchmark datasets for the fake news detection problem; partly because the term fake news contains a wide variety of subcategories and partly because the topic has gained major attention very recently. In addition, these datasets are vastly different from one another some designed for totally different tasks, which makes the evaluation even more challenging. Over a dozen dataset in total (mentioned in the in the related work section) have been investigated and only four datasets, which we found pertinent to the task at hand were chosen. The datasets are further categorized into two types based on the length and structure of the sentences. Class I datasets are for relatively short, tweet like news or statements, typically 70 to 150 characters in length. Class II datasets, however are generally made up of longer texts, like proper news articles, typically 400 to 700 words. Fig. 1 compares details of these datasets. The summary of datasets in consideration is provided in the following subsections.

3.1 Class I

3.1.1 LIAR

The dataset known as LIAR is first introduced in (Wang, 2017) as a benchmark to fake news problem, containing around 12,000, statements from various sources each accompanying an associated number which represents the truthfulness and credibility of the claim on a scale of 0-5.(0 being completely false or as the website calls it "Pants on Fire!" and 5 being completely accurate) The statements and labels are from Politifact website, which specializes in verifying the veracity of political statements by expert journalists. In addition, the dataset includes metadata information containing the speaker of each claim, position of the speaker, his/her home state if the speaker is a political representative, the history of his/her statements and other similar information. The metadata associated with the claims can be of use since there are often observed patterns in one's way of speaking. The mentioned dataset has a large amount of news claims related to United States politics and generally is hard to classify due to the of lack of reference sources to verify.

3.1.2 Kaggle's Fake News Dataset

The dataset provided in the kaggle.com (Risdal, 2016) has about 12,500 instances of fake claims, each claim containing a header alongside a body article. The headlines of such articles can be categorized into Class I, while the text is categorized as Class II explained earlier. This dataset also contains some metadata such as crawl time and news id for each of the instances, however for the task at hand these information turn out to be partly irrelevant to the task at hand.

3.1.3 Fake News Challenge 2017 Dataset

This dataset contains around 13,000 short headlines and 2587 full articles(Rubin et al., 2015).

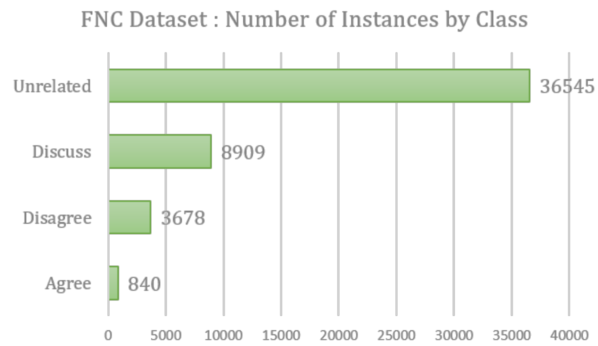

Figure 2: Label frequencies for Fake News Challenge Dataset

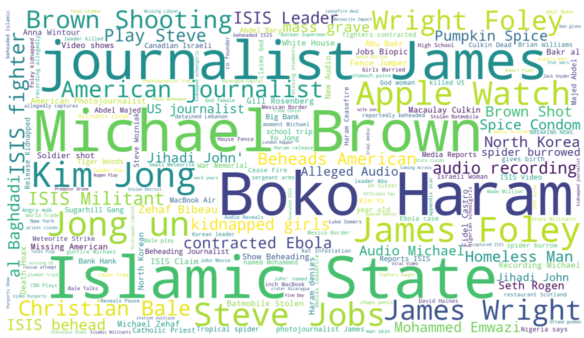

Figure 3: Word cloud for Fake News Dataset

Each sentence refers to one of the articles and the label is the stance of the article towards the claim, showing whether it agrees, disagrees or is irrelevant with the claim. The fake news challenge took a unique and interesting approach for formulating the problem, but the downside is that for every claim a reference article is required to do the classification. On the other hand, the proposed approach in this report combines information retrieval and deep learning to address the shortcomings. Fig. 2 shows the data inn FNC dataset by frequency. In Fig. 3 the word cloud for this dataset is shown which gives some interesting insights into the dataset showing the mostly dominant topics.

3.2 Class II

3.2.1 University of Washington Fake news data

This data boasts about 49,000 article paragraphs collected from websites known to only contain fake and even satirical news such as onion alongside credible news sources. As such, each claim has either of the four labels: hoax, propaganda or satire, true news. The length of each sentence is quite long, averaging between 500 to 600 words. Although the dataset was proposed for a similar problem, still some slight modifications need to be added to the dataset to suit it better to the proposed problem. More specifically, all the sentences labeled as satire are removed from the dataset, as satire is not exactly fake news and it is mostly intended for humor rather than actual misleading information.(Rashkin et al., 2017)

4 Methods

As mentioned earlier, our proposed model consists of several smaller submodules each responsible for classifying the instances based on a set of features and then combining the results through a voting process, which can be either majority voting or some wighted average, the weights of which are also learned by the model. For this study we mostly focus on two main submodules: Veracity Detection submodule (based on IR and knowledge base), and style-based submodule. This model can be extended further by adding other submodules to it such as author's background information, history and so on each having its own formulation. In the following subsections we will discuss the details of the two main modules that are implemented.

4.1 Veracity Detection Submodule

The first submodule is responsible for checking the veracity of each claim given that we have already constructed a knowledge base. In order to do so, two steps are taken: In the first step, the most relevant documents are retrieved from the knowledge base. In the second step, given those documents, the stance of the claim towards the documents is inferred. The overall flow of the process is depicted in Fig. 4. This can be interpreted as checking the validity of claim given that a knowledge base of credible news sources are provided (For example it is crawled from online news sources in advance or it is obtained from some universal knowledge base). The number of the retrieved documents is also controlled by a hand-picked hyper-parameter of the model noted by k. It is evident that as we increase the hyper parameter k, the precision of the retrieved documents would suffer.

For retrieval we used TF-IDF method as baseline and more advanced algorithms for comparison and improved performance. The following three algorithms have been implemented and tested:

- BM25: BM25 algorithm (BM standing for Best Matching) is a ranking function scoring based on probabilistic retrieval frameworks. It uses bag-of-words representation of documents to rank each document with respect to the different query words occurring in it. However, BM25 ignores the relative ordering of query terms as well as their proximity within the documents.

- Vector Space Model: The Vector Space Model is another retrieval algorithm which is implemented alongside the Boolean model of Information Retrieval in the Lucene framework. All the documents initially returned by the Boolean model are scored by the Vector Space Model and returned in ranked order. The ranking score is the cosine similarity between the query and the document vectors in a multidimensional word vector space. The advantages of this scoring method are partial matching and a continuous ranking scale.

- Language Model: This is another probabilistic model where conditional probability $P(d|q)$ is calculated for the given query q and document d vectors.It assumes Dirichlet priors for the probability to smooth the function with a document normalization component.

After the k related articles are retrieved, the second step of the algorithm starts in which each article is classified into three labels 'Fake', 'Suspicious' or 'Legit'. For the classification, any deep learning architecture can be used. In our case a simple Feed Forward Neural Net is used as shown in the Figure 5. This specific architecture is inspired by one of three winning entries in Fake news challenge (Riedel et al., 2017b). However, modification are made to transform it to the reformulated problem. The input features of the classifier are two bag of word vector of size 5000, one corresponding to the news statement and the other to the article. Both of these vectors are fitted on the vocabulary of 5000 most frequent used words in knowledge base. Additionally, it takes the cosine similarity between these two vectors as an additional input, hence, extending the final size of input vector to 10001. The hidden layer of the model has 100 Rectified Linear Units(ReLU) and the final layer is a softmax layer with three output classes as mentioned before.

4.2 Style Detection Submodule

The second main submodule of the model is responsible for extracting valuable information using the differences between the styles of fake/real news, since generally the style of the text itself can say a lot about the intentions of the authors. In most cases, the fake articles have a more aggressive and biased tone as well as stronger choice of words.(Rashkin et al., 2017). For this purpose, our model uses a Deep Bidirectional LSTM architecture, since Bi-LSTMs have proven very useful in storing and making efficient use of the information present in long sentences. The power of LSTMs come from their more complicated cell structures compared to standard RNNS. The equations governing an LSTM are as follows:(Hochreiter and Schmidhuber, 1997)

$$\begin{aligned} i(t) &= \sigma(W^{(i)}x^{(t)} + U^{(i)}h^{(t-1)}) \\ f(t) &= \sigma(W^{(f)}x^{(t)} + U^{(f)}h^{(t-1)}) \\ o(t) &= \sigma(W^{(o)}x(^{t)}+U^{(o)}h^{(t-1)}) \\ \bar{c}^{(t)} &= tanh(W^{(c)}x^{(t)} + U^{(c)}h^{(t-1)}) \\ c^{(t)} &= f^{(t)} \circ \bar{c}^{(t-1)} + i^{(t)} \circ \bar{c}^{(t)} \\ h^{(t)} &= o^{(t)} \circ tanh(c^{(t)}) \end{aligned} \tag{1}$$

In addition, using bidirectional neural networks instead of one-directional neural nets further improves the accuracy.

5 Experimental Results

For the experiments, the FNC dataset is used for training the veracity-based (IR-DL) submodule, while UW dataset is used for training the style-based submodule. The reason is that UW dataset is rich in terms of style content whereas the other datasets mostly focus on the actual fact-based differences.

Another thing to mention is that for training the veracity based module, the claims in the FNC dataset labeled as "unrelated" are discarded, since there is no article which can help in verifying their authenticity, thus it injects noise into training.

Since the average number of relevant documents in this dataset turned out to be 10, we chose hyperparameter k to be 10. In Fig. 6, Fig. 7 and Fig.8 recall and precision for different values of k can be seen. As expected, by increasing the number k, precision suffers but recall improves. Fig. 9

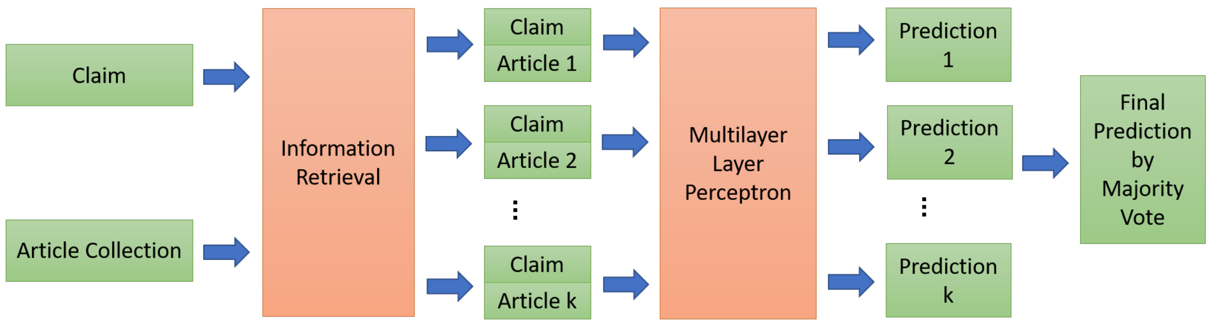

Figure 4: Overall Pipeline of the Veracity-based classifier

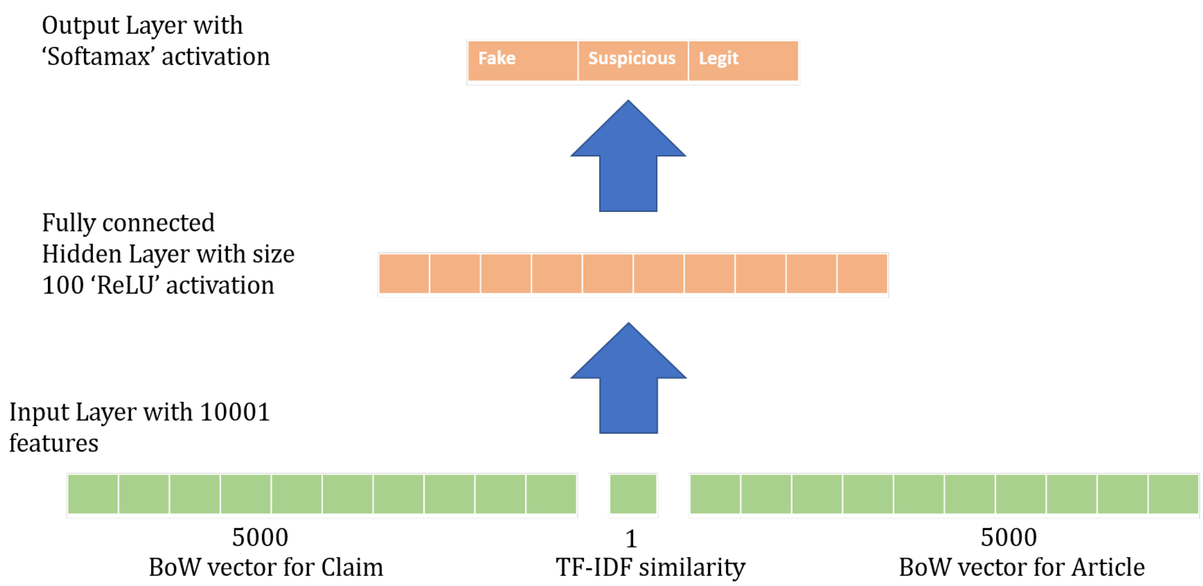

Figure 5: Architecture of the FFNN used

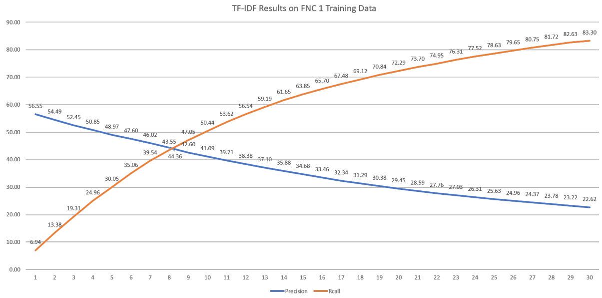

Figure 6: Precision and Recall for TF-IDF Method

also represents the confusion matrix after the classification is done on FNC dataset.

The accuracy of the veracity-based submodule after retrieving the documents is 67.09% for three way classification (Fake, Suspicious, Real) and 72.12% for binary classification. The accuracy of the style-based submodule is also evaluated separately on the UW test dataset and with the given architecture is 81.83%.

One issue faced while training and evaluating the veracity-based module was that due to the class imbalance (less number of instances in the class "Fake") the classifier might not improve as good. In order to tackle this problem several approaches, such as merging datasets, oversampling or undersampling can be taken. Another approach is to force the classifier to add extra penalty by modifying the cost function. This latter approach results in a higher precision and recall but lower accuracy.() Below 60%) Another thing to note is the accuracy performance of the IR-DL submodule should not be naively compared to the FNC challenge since the FNC challenge contains too many unrelated articles, making the task much easier and the accuracy metric somewhat misleading, because even assigning each data into the unrelated class in that case would give accuracy of 75% due to high class imbalance.

6 Conclusions and Future Work

There are certain areas which can be improved to make the model more robust. One of such improvements can be modifying the retrieval algorithm so that the retrieval and learning are jointly performed, thus improving the accuracy. Also other more advanced architectures can be investigated for the veracity-based part of the model. Other extensions can also be made to the model such as taking author's information and history or even the source of the news into account. For in-

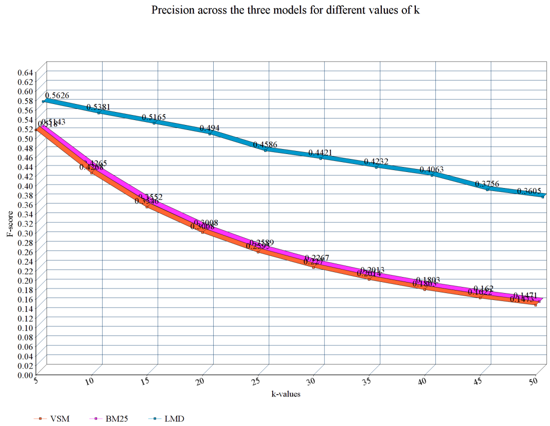

Figure 7: Precision for Advanced Algorithms

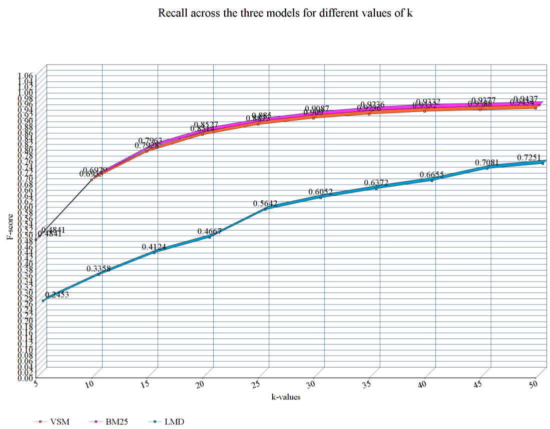

Figure 8: Recall for Advanced Algorithms

		Predicted		
		Suspicious	Fake	Legit
Actual	Suspicious	672	11	454
	Fake	138	12	173
	Legit	472	14	1889

Figure 9: Confusion Matrix for FNC Dataset

stance, one approach may be to construct a hypergraph of the authors and their articles and learn a Deep neural network on the graph. One important issue we also faced was the lack of a generalized and standard dataset for the task of fake news detection. So one approach could be providing a universal benchmark for these kinds of tasks.

References

????

???? Fake news challenge stage 1 (fnc-i): Stance detection. http://www.fakenewschallenge.org/.

2017. English dictionary, translations and thesaurus. https://dictionary.cambridge.org/.

Hadeer Ahmed, Issa Traore, and Sherif Saad. 2017. Detection of online fake news using n-gram analysis and machine learning techniques. In *International Conference on Intelligent, Secure, and Dependable Systems in Distributed and Cloud Environments*. Springer, pages 127–138.

Delenn Chin Kevin Chen Akshay Agrawal. ???? Cosine siamese models for stance detection. Technical report, Stanford University, year = 2017.

Nir Avrahamov. 2017. Machine learning, graphs, and the fake news epidemic (part 2) - dzone ai. https://dzone.com/articles/machine-learning-graphs-and-the-fake-news-epidemic-1.

Christoph Aymanns, Jakob Foerster, and Co-Pierre Georg. 2017. Fake news in social networks. *CoRR* abs/1708.06233. http://arxiv.org/abs/1708.06233.

Manish Bhelande, Aniket Sanadhya, Mrugang Purao, Abhay Waldia, and Vikas Yadav. 2017. Identifying controversial news using sentiment analysis. *Imperial Journal of Interdisciplinary Research* 3(2).

Prakhar Biyani, Kostas Tsioutsiouliklis, and John Blackmer. 2016. " 8 amazing secrets for getting more clicks": Detecting clickbaits in news streams using article informality.

Xinyue Cao, Thai Le, and Jason Zhang. 2017. Machine learning based detection of clickbait posts in social media. *CoRR* abs/1710.01977. http://arxiv.org/abs/1710.01977.

Tong Chen, Lin Wu, Xue Li, Jun Zhang, Hongzhi Yin, and Yang Wang. 2017. Call attention to rumors: Deep attention based recurrent neural networks for early rumor detection. *arXiv preprint arXiv:1704.05973* .

Sahil Chopra and Saachi Jain. 2017. Towards automatic identification of fake news: Headline-article stance detection with lstm attention models.

Niall J Conroy, Victoria L Rubin, and Yimin Chen. 2015. Automatic deception detection: Methods for finding fake news. *Proceedings of the Association for Information Science and Technology* 52(1):1–4.

Leon Derczynski, Kalina Bontcheva, Maria Liakata, Rob Procter, Geraldine Wong Sak Hoi, and Arkaitz Zubiaga. 2017. Semeval-2017 task 8: Rumoureval: Determining rumour veracity and support for rumours. *arXiv preprint arXiv:1704.05972* .

Andreas Hanselowski. 2017. Team athene on the fake news challenge âĂŞ andreas hanselowski âĂŞ medium. https://medium.com/@andre134679/team-athene-on-the-fake-news-challenge-28a5cf5e017b.

Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural Comput.* 9(8):1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735.

Zhiwei Jin, Juan Cao, Han Guo, Yongdong Zhang, Yu Wang, and Jiebo Luo. 2017. Rumor detection on twitter pertaining to the 2016 U.S. presidential election. *CoRR* abs/1701.06250. http://arxiv.org/abs/1701.06250.

Georgi Karadzhov, Preslav Nakov, Lluís Màrquez, Alberto Barrón-Cedeño, and Ivan Koychev. 2017. Fully automated fact checking using external sources. *CoRR* abs/1710.00341. http://arxiv.org/abs/1710.00341.

Sumeet Kumar. ???? Investigating the use of bots to spread fake news in social media .

Alex Oswalt Kurt Miller. ???? Fake news headline classification using neural networks with attention. Technical report, California State University, year = 2017.

William Largent. 1970. Talos targets disinformation with fake news challenge victory. https://blog.talosintelligence.com/2017/06/talos-fake-news-challenge.html.

Jing Ma, Wei Gao, Prasenjit Mitra, Sejeong Kwon, Bernard J Jansen, Kam-Fai Wong, and Meeyoung Cha. 2016. Detecting rumors from microblogs with recurrent neural networks. In *IJCAI*. pages 3818–3824.

Jing Ma, Wei Gao, and Kam-Fai Wong. 2017. Detect rumors in microblog posts using propagation structure via kernel learning. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. volume 1, pages 708–717.

Arjun Mukherjee, Vivek Venkataraman, Bing Liu, and Natalie S Glance. 2013. What yelp fake review filter might be doing? In *ICWSM*.

Shruti Bhargava Neel Rakholia. ???? Is it true?âĂİ âĂŞ deep learning for stance detection in news. Technical report, Stanford Uiversity, year = 2017.

Olga Papadopoulou, Markos Zampoglou, Symeon Papadopoulos, and Ioannis Kompatsiaris. 2017. A two-level classification approach for detecting clickbait posts using text-based features. *arXiv preprint arXiv:1710.08528* .

Dina Pisarevskaya. 2017. Deception detection in news reports in the russian language: Lexics and discourse. In *Proceedings of the 2017 EMNLP Workshop: Natural Language Processing meets Journalism*. pages 74–79.

Quan Zhou Shanshan Xu Qi Zeng. ???? Neural stance detectors for fake news challenge. Technical report, Stanford Uiversity, year = 2017.

Hannah Rashkin, Eunsol Choi, Jin Yea Jang, Svitlana Volkova, and Yejin Choi. 2017. Truth of varying shades: Analyzing language in fake news and political fact-checking. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*. pages 2921–2927.

Benjamin Riedel, Isabelle Augenstein, Georgios P. Spithourakis, and Sebastian Riedel. 2017a. A simple but tough-to-beat baseline for the fake news challenge stance detection task. *CoRR* abs/1707.03264. http://arxiv.org/abs/1707.03264.

Benjamin Riedel, Isabelle Augenstein, Georgios P Spithourakis, and Sebastian Riedel. 2017b. A simple but tough-to-beat baseline for the fake news challenge stance detection task. *arXiv preprint arXiv:1707.03264* .

Megan Risdal. 2016. Getting real about fake news. https://www.kaggle.com/mrisdal/fake-news.

Victoria L Rubin, Yimin Chen, and Niall J Conroy. 2015. Deception detection for news: three types of fakes. *Proceedings of the Association for Information Science and Technology* 52(1):1–4.

Victoria L Rubin, Niall J Conroy, Yimin Chen, and Sarah Cornwell. 2016. Fake news or truth? using satirical cues to detect potentially misleading news. In *Proceedings of NAACL-HLT*. pages 7–17.

Natali Ruchansky, Sungyong Seo, and Yan Liu. 2017. CSI: A hybrid deep model for fake news. *CoRR* abs/1703.06959. http://arxiv.org/abs/1703.06959.

Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu. 2017. Fake news detection on social media: A data mining perspective. *ACM SIGKDD Explorations Newsletter* 19(1):22–36.

James Thorne, Mingjie Chen, Giorgos Myrianthous, Jiashu Pu, Xiaoxuan Wang, and Andreas Vlachos. 2017. Fake news stance detection using stacked ensemble of classifiers. In *Proceedings of the 2017 EMNLP Workshop: Natural Language Processing meets Journalism*. pages 80–83.

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

Fan Yang, Arjun Mukherjee, and Eduard Constantin Dragut. 2017. Satirical news detection and analysis using attention mechanism and linguistic features. *CoRR* abs/1709.01189. http://arxiv.org/abs/1709.01189.

Yiwei Zhou. 2017a. Clickbait detection in tweets using self-attentive network. *CoRR* abs/1710.05364.

Yiwei Zhou. 2017b. Clickbait detection in tweets using self-attentive network. *CoRR* abs/1710.05364. http://arxiv.org/abs/1710.05364.