Universität Koblenz - Landau   18. Juni 2018
Department für Informatik,
Institute for Web Science and Technologies
Prof. Dr. Steffen Staab
Supervisor????
Lukas Schmelzeisen

Master Thesis Kandhasamy Rajasekaran

Fake News Detection using Neural Network models of Wikipedia

Abstract

The unprecedented growth of production and dissemination of information leads to an unprecedented growth of production and dissemination of Fake news. Fake news hinders the society from progressing by delaying the pursuit of right information. It is very essential to have a mechanism to detect and control fake news. Several organizations uses collaborative efforts of domain experts, a manual process which cannot withstand the proliferation of news production and dissemination. This research work will use Wikipedia as a ground reality and cross check claims automatically. Deep Neural Networks will be used to understand Wikipedia and the performance of different configurations of Neural Networks will be benchmarked against each other and the already available automated fake news detectors.

1   Introduction

Humans evolved into a superior race only because of transfering the knowledge to its descendants. 'Knowledge is Power', whether it is about events or about the working of anything, knowing has always put anyone into a better position than being ignorant. Only by sharing information the society improves its collective intelligence. With the rise in web and social networks, the production and dissemination of information happens at a rapid pace. As much as it became easier to disseminate information it has also become easier to spread fake news.

According to Lazer et al., Fake news resembles very similar to a news media information in form but not in organizational process or intent. Fake news overlaps with misinformation (misleading information shared unintentionally) and disinformation (false information spread purposefully). In essence the fake news publishers does not have the rigorous news media's editorial norms for making sure of accuracy and credibility [LBB+18].

Fake news is quite prevalent and studies refer that an average American exposed to atleast a mimimum of 3 fake news stories per month from well known publishers during 2016 election. Social Networking platforms such as for e.g. Twitter and Facebook have been used heavily by Fake news publishers during 2016 US and 2017 French elections [LBB+18].A large scale empirical study with twitter dataset by Vosoughi et al. reveals that fake news spread farther, faster, deeper and broader than the legitimate news. The effects of fake news are more prominent in political context than news about terrorism, natural disaster, science and other domains. Novelty is the most important factor for information to spread. People share information which are interesting and new information are considered as interesting. Fake news are received as much more novel than the True news. As against the understanding, it is the humans, not robots, who are more likely responsible for spread of fake news [VRA18].

Most often Fake news is detected by people and organizations through their common sense. There are many fact checking websites such as for e.g. snopes.com, factcheck.org, politfact which uses collaborative effort of domain experts and tag news with a fact meter to refer the authenticity. Although this is fairly good, since it requires manual effort, it is not available for all domains and not scalable. At the same time, it is unmatchable to the rate at which the information is produced. With many social networks, blogging sites we have seen a unparallel rise in the volume of content being generated and a manual intervention to cross check their authenticity is clearly no match. But overall the main idea to handle false news is to check against reliable source of information and claim its integrity.

Wikipedia is a free online encyclopedia available in more than 300 languages with a principle that anyone can edit. It has got a wide range of domains covered with many articles written under each domain [Wal05]. According to Alexa and SimilarWeb, Wikipedia is considered to be fifth most popular website.According to Wikipedia, the English Wikipedia consists highest number of articles at present amounting to approximately 500 million. The range of subject covered is wide such as for e.g. Art, Culture, Science, Mathematics, Religion etc. and it is maintained by many authors. The frequency of the udpate in English Wikipedia is very high and it is approximately equal to 10 updates per second and 600 articles per day. Although it can be edited by anyone, investigation carried out by Nature, reveals that the quality of content is similar to other encyclopedia such as Britannica [Wal05]. Although there can be malicious users in Wikipedia, the culture and the community ensures most of the high impactful errors are rectified very quickly[PCL+07]. Thus English Wikipedia is a reliable source of information.

Understanding Wikipedia is a complex task and it requires understanding the subtelities of English natural language and the context. Machine Learnining is the capability of systems learning patterns from raw data. Different features of raw data need to be extracted separately and fed to Machine learning algorithms to get good performance. This drawbeck is solved by using Deep Learning, which uses neural networks, a multi layer network of simple representations to learn complex data representations and then extracts patterns out of the data. Deep Learning provides state-of-the-art results in the field such as image recognition, speech recognition and natural language processing[GBC16].

The focus of the master thesis is to use wikipedia as a ground reality or as source of experts opinion and use this knowledge to cross check claims automatically. Whether the information is present in Wikipedia or not will be used as a proxy for information being considered as Truthy or Fake. Neural Networks with different configuration will be used to understand Wikipedia and the performance of each one will be benchmarked against each other.

2   Related work

Many researches have been conducted to detect Fake news in microblogging platforms such as Twitter. Most of these works classify the news as Truth or Fake by using the platform/user specific information such as how popular the post is, credibility of the user who shared it, diffusion patterns etc [LNL+15] [MGW+15]. Zhao, et al. have used cue terms such as 'not true', 'unconfirmed' etc in retweets or the comments to detect fake news. The assumption is that when people exposed to fake news they will comment or retweet with such words in their post[ZRM15]. Other studies focused on using the temporal characteristics of fake news during the spread. Kwon et al. used tweet volume in time series and Ma, et al. measured variations of social context features over time[KCJ+13] [MGW+15]. All the attempts made in the above researches involve handcrafted feature engineering which is critical, biased and very time consuming.

Jing Ma et al. efforts were focused on building a recurrent neural network (RNN) to detect rumors from Microblogs such as Twitter and Weibo effectively. The training dataset is obtained by using constructed fake and truthy news keywords from debunking services such as Snopes and Sina community management center. The keywords are used in Search API's of Microblog and labelled respectively. The social context information of a post and all its relevant posts such as comments or retweets is modeled as variable-length time series. RNNs with different configurations such as using one or two layers of GRU and LSTM are very good in capturing long distance dependencies of temporal and textual representations of posts under supervision. This method completely avoids all the handcrafted feature engineering efforts which are biased and time consuming. It produces better results with datasets from Twitter and Sina Weibo than all of the traditional Machine Learning methods. RNNs with two layers of GRU gave the best results and it was also very quick in predicting the rumor than the average time from debunking services[MGM+].

The method uses platform specific features such as the content of tweets, retweets, comments in a tweet and the temporal correlation between them to figure out whether the news is truthy or fake. All the features specified will not be available in many platforms and this methodolgy will not suit in such conditions. RNNs give best results for sequence data like text and validating a news as true or false completely based on its semantic is good in such cases.

Giovanni Luca Ciampaglia et al. used DBPedia for checking computationally whether a given information is factual or not. The work uses the knowledge graph built from DBPedia which represents infobox section in Wikipedia. This represents only non-controversial and factual information which is analagous to human collected information. The methodology formulates the problem of checking facts into a network analysis problem which is finding the shortest path between nodes (subject and object of a sentence) in a graph. The aggregated generalities of nodes along a path in a weighted undirected graph is used as a metric for measuring the authenticity of information. The more the elements are generic the weaker the truthfulness is. The genericness of a node is obtained by the degree of that node - no. of nodes connected to that node. The truthfulness of the information is improved if there exists at least one path from subject to an object with minimal non-generic nodes. This approach exploits the indirect connections to a great extent with distance constraints in a knowledge graph. The approach gave promising results when tested with datasets containing simple factual information about history, geography, entertainment and biography[CSR+15].

The above research is a good initial step towards an automated fact checker system using only semantics of data. The problem of fake news attempted is very primitive and uses only 'is' or 'type of' relation. Current fake news are very complex and subtle when it comes to ambiquities. In this approach, DBPedia is used and according to their sources the update/synch frequency is slower than wikipedia by 6 to 18 months.

3   Approach

Deep learning systems are giving better results in building intellectual systems nowadays.[Gol16] The results achieved in applications such as Autonomous car driving, playing chess are almost equivalent to the skillsets of a human. There are different neural network models exists such as Feed Forward, Convolutional, Recurrent Neural Network models. Wikipedia contains a lot of articles and each articles contains text. Text is a sequence data where as position of words in it is dependent on the other. Recurrent Neural Networks bring out best results in many applications involving text data such as machine translation, super tagging.

Every article or sentence in wikipedia can be fetched and inputted to the neural network as truthy value. But we lack false values. We need to use a semi supervised technique where in extraction of falsy values need to be carried out automatically. This is one of the challenges of this master thesis.

Some of the ideas in place are 1) Extract the articles and sentences from many of these fact checkers websites and use the ones which are labelled as false values. We need to make sure that they are opposites of sentences in wikipedia. 2) Build sentences which are opposite of sentences in wikipedia by using GLOVE technique or construct negative sentences. 3) Build sentences which are opposite of sentences in wikipedia by looking a semantic web representations. Word embeddings can be used to replace the verb opposites

The size of the article in wikipedia are long and arbitrary and RNN will face gradient diminishing problems. The long distance dependencies will be missed out and hence different configurations of the neural network should be used and compared 1) Different layers 2) Single/Multiple LSTM or GRU units 3) Usage of one hot vector vs word embeddings. The creation of word embeddings need to be thought through. It would be a good idea to do it from either wikipedia itself or from pre-trained word embeddiings

Recently convolutional neural network which character level input is giving out better results for some applications and this configuraiton should also be tried out.

4 Evaluation

The training, evaluation and test dataset needs to be curated.

More or less equal data should be present for each bin such as facts and lies.

Extract sentence from wikipedia and give it as it is.

Distort the sentence by swapping and give it.

Construct good facts and lies outside - a proper labelled dataset and see how it works.

Use the dataset provided by researches indicated in related work section.

Look for already curated community wide popular datasets for fake news.

Have a baseline.

Compare each configuration against baseline and measure the accuracy, time taken. If the curated dataset has many groups of varying complexity then state the results groupwise

5 Organizational matters

Duration of work: 01-July-2018 – 31-Dec-2018
Candidate: Kandhasamy Rajasekaran
E-Mail: kandhasamy@uni-koblenz.de
Student number: 216100855
Primary supervisor: Prof. Dr. Steffen Staab
Supervisor: Supervisor????
Secondary supervisor: Lukas Schmelzeisen

6 Time schedule

- Introduction and Literature: 01-May-2018 – 30-June-2018

- Initial Phase: 01-July-2018 – 15-Sep-2018

    - Prototyping: 01-July-2018 – 30-July-2018

    - Implementing ML Pipeline: 01-Aug-2018 – 15-Aug-2018

    - Baseline Implementation: 16-Aug-2018 – 30-Aug-2018

    - Testing and refining: 01-Sep-2018 – 15-Sep-2018

- Development Phase: 16-Sep-2018 – 30-Nov-2018

    - Prototyping: 01-July-2018 – 30-July-2018

    - Implementing ML Pipeline: 01-Aug-2018 – 15-Aug-2018

    - Baseline Implementation: 16-Aug-2018 – 30-Aug-2018

    - Testing and refining: 01-Sep-2018 – 15-Sep-2018

- Final Phase: 01-Dec-2018 – 30-Dec-2018

    - Comprehend Benchamark results: 01-Dec-2018 – 15-Dec-2018

    - Revision: 08-Dec-2018 – 22-Dec-2018

    - Thesis report: 01-Dec-2018 – 30-Dec-2018

References

[CSR⁺15] Giovanni Luca Ciampaglia, Prashant Shiralkar, Luis M. Rocha, Johan Bollen, Filippo Menczer, and Alessandro Flammini. Computational fact checking from knowledge networks. *PLoS ONE*, 2015.

[GBC16] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. www.deeplearningbook.org.

[Gol16] Yoav Goldberg. A Primer on Neural Network Models for Natural Language Processing. *Journal of Artificial Intelligence Research*, 57:345–420, 2016.

[KCJ⁺13] Sejeong Kwon, Meeyoung Cha, Kyomin Jung, Wei Chen, and Yajun Wang. Prominent features of rumor propagation in online social media. In *Proceedings - IEEE International Conference on Data Mining, ICDM*, 2013.

[LBB⁺18] David M. J. Lazer, Matthew A. Baum, Yochai Benkler, Adam J. Berinsky, Kelly M. Greenhill, Filippo Menczer, Miriam J. Metzger, Brendan Nyhan, Gordon Pennycook, David Rothschild, Michael Schudson, Steven A. Sloman, Cass R. Sunstein, Emily A. Thorson, Duncan J. Watts, and Jonathan L. Zittrain. The science of fake news. *Science*, 359(6380):1094–1096, 2018.

[LNL⁺15] Xiaomo Liu, Armineh Nourbakhsh, Quanzhi Li, Rui Fang, and Sameena Shah. Real-time Rumor Debunking on Twitter. In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management - CIKM '15*, 2015.

[MGM⁺] Jing Ma, Wei Gao, Prasenjit Mitra, Sejeong Kwon, Bernard J Jansen, Kam-Fai Wong, and Meeyoung Cha. Detecting Rumors from Microblogs with Recurrent Neural Networks.

[MGW⁺15] Jing Ma, Wei Gao, Zhongyu Wei, Yueming Lu, and Kam-Fai Wong. Detect Rumors Using Time Series of Social Context Information on Microblogging Websites. In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management - CIKM '15*, 2015.

[PCL⁺07] Reid Priedhorsky, Jilin Chen, Shyong (Tony) K. Lam, Katherine Panciera, Loren Terveen, and John Riedl. Creating, destroying, and restoring value in wikipedia. In *Proceedings of the 2007 international ACM conference on Conference on supporting group work - GROUP '07*, 2007.

[VRA18] Soroush Vosoughi, Deb Roy, and Sinan Aral. The spread of true and false news online. *Science*, 359(6380):1146–1151, 2018.

[Wal05] Jimmy Wales. Internet encyclopaedias go head to head, 2005.

[ZRM15] Zhe Zhao, Paul Resnick, and Qiaozhu Mei. Enquiring minds: Early detection of rumors in social media from enquiry posts. In *Proceedings of the 24th International Conference on World Wide Web*, WWW '15, pages 1395–1405, Republic and Canton of Geneva, Switzerland, 2015. International World Wide Web Conferences Steering Committee.

7   Signatures

_____   _____
        Kandhasamy Rajasekaran   Prof. Dr. Steffen Staab

_____   _____
            Supervisor????   Lukas Schmelzeisen

8 Declaration of Authorship

I hereby declare that the thesis submitted is my own unaided work. All direct or indirect sources used are acknowledged as references.

I am aware that the thesis in digital form can be examined for the use of unauthorized aid and in order to determine whether the thesis as a whole or parts incorporated in it may be deemed as plagiarism. For the comparison of my work with existing sources I agree that it shall be entered in a database where it shall also remain after examination, to enable comparison with future theses submitted. Further rights of reproduction and usage, however, are not granted here.

This paper was not previously presented to another examination board and has not been published.

Koblenz, on June 18, 2018

_____

Kandhasamy Rajasekaran