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

12. Juni 2018

Master Thesis Kandhasamy Rajasekaran

Fake News Detection using Neural Network models of Wikipedia

Abstract

With the unprecedented growth of production and dissemination of information, there exists an unprecedented growth in production and dissemination of fake news. There have been several adverse incidents that happened in the past due to rise of fake news. It is very essential to have a mechanism to detect and control fake news automatically. Several methods have been proposed which are automatic and semi automatic than the traditional manual checking systems. In this master thesis project, an attempt is to be made to understand online encyclopedias using deep learning technique. The performance of the system will be benchmarked against the results that have been published for those systems.

1 Introduction

Humans evolved and became a superior race than other species only by being able to pass the information to its descendants. The society advances as a whole by improving its collective intelligence and that happens only by sharing. With the advent of technology systems such as web, it is possible for anyone to become a publisher and publish their thoughts, ideas to the entire world. As much as it became easier to disseminate information it also brought in a lot of negative effects. One of it is the spread of false news or fake news. The dissemination of false news can harm the society very bad and it can reduce the progress or worsen the state of the society. There are several examples such as a misleading news refering to a fire in a US school made a lot of chaos in traffic and accidents because parents were rushing to save their kids. Humans have been relying on their collective intelligence and advice from experts on a particular domain to handle the fake news. There are several organizations such as snopes, factcheck.org, politfact which tags many information with a fact meter to label the authenticity of information. Although this is fairly good, it is not scalable and unmatchable to the rate at which the information is created. 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 crowdsourced online encylopedia which has versatile topics and a whole range of articles for each topic. It has been listed as one of the famous 10 websites or the one of the 10 most used website. The range of subject covered is very wide and it comes in different languages. The frequency of the udpate is really high and every content is peer reviewed by other as against the process in newspaper. Although the checking need not be done by experts opinion the reliability of common public on wikipeida is highly visible.

The focus of the master thesis is to use wikipedia a ground reality or as source of experts opinion and use this knowledge to cross check the claims which in the form of tweets, blogs, sentences through automated means. Recurrent Neural Network are a special kind of neural network models for sequence data. They are used to understand wikipedia and acts a bot to check the news as fake or not. In this we assume whatever information present in wikipedia is truthy and we rely on the collective and collaborative effort of humans to build this enormous information resource.

2 Related work

There are many attempts being made to counter attack fake news. Most of them are imbibed into one particular platform and uses the news, user characteristics, who shared it and what is their credibility and how the diffusion have happened. They used many supervised machine learning algorithms which uses handcrafted features and obtain substantial results.

Giovanni Luca Ciampaglia et al. used DBPedia for checking computationally whether a given information is factual or not. The work involves using 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.

Jing Ma et al. efforts were focused on building a recurrent neural network (RNN) to detect rumors from Microblogs effectively. 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.

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 Enough data should be present for each bin such as facts, slightly distorted lie, blatant 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

5 Organizational matters

Duration of work:	01.07.2018 – 31.12.2018
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

This needs to be changed. But for now just a filler.

- Introduction and Literature: 03.05.17 – 15.06.17
- Methodology: 09.06.17 – 30.07.17
  - Approach concept: 09.06.17 – 22.06.17
  - Implementation Plan: 18.06.17 – 29.06.17
  - Implementation: 29.06.17 – 30.07.17
  - Testing and refining: 01.07.17 – 30.07.17
- Approach results: 31.07.17 – 20.08.17
  - Sampling: 31.07.17 – 12.08.17
  - Interpretation: 13.08.17 – 20.08.17
- Evaluation: 21.08.17 – 25.09.17
  - Preparing evaluation: 21.08.17 – 01.09.17
  - Conducting evaluation: 02.09.17 – 11.09.17
  - Analyzing results: 12.09.17 – 25.09.17
- Revision: 26.09.17 – 31.10.17

References

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

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 12, 2018

_____

Kandhasamy Rajasekaran