

## **Master Thesis Kandhasamy Rajasekaran**

### **Fake News Detection using Neural Network models of Wikipedia**

#### **Abstract**

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Morbi nulla ligula, blandit quis mi ut, pellentesque mattis nisi. Praesent elementum diam sed nibh lobortis dictum. Sed eleifend posuere magna, ac rhoncus nisi feugiat in. Donec accumsan ac leo in cursus. Vestibulum id elit lacus. Ut nec rhoncus dolor, quis laoreet dui. Donec luctus fringilla elit.

Sed suscipit dolor dolor, sed consequat neque ornare non. Nunc laoreet, metus ac congue efficitur, neque felis varius metus, in pretium dui nibh eget nisi. Integer maximus porta turpis, et dignissim dolor tempus ac. In hac habitasse platea dictumst. Vivamus pretium dui massa, in volutpat orci semper in. Aliquam lorem elit, gravida pretium aliquet id, laoreet eget sem. Mauris mauris lacus, vehicula at felis sit amet, rutrum efficitur est. Pellentesque auctor vestibulum risus sit amet consectetur. Nunc lobortis velit sed nunc aliquam, non egestas nisl vulputate. Nam posuere mauris lectus, nec ullamcorper ipsum placerat eget. Duis quis justo sit amet nunc facilisis ullamcorper. Aliquam posuere suscipit urna. Donec mattis vestibulum odio. Sed ultricies, lorem ut tempor eleifend, elit lacus feugiat orci, sit amet varius lorem enim ac ante. Aliquam ac lacus vel nunc mollis scelerisque ut at erat. Sed pulvinar metus tellus, ut mollis lorem mattis at.

## **1 Introduction**

### **Fake news**

Fake news is a type of yellow journalism or propaganda that consists of deliberate misinformation or hoaxes spread via traditional print and broadcast news media or online social media.[1][2] This false information is mainly distributed by social media, but is periodically circulated through mainstream media.[3] Fake news is written and published with the intent to mislead in order to damage an agency, entity, or person, and/or gain financially or politically,[4][5][6] often using sensationalist, dishonest, or outright fabricated headlines to increase readership, online sharing, and Internet click revenue. In the latter case, it is similar to sensational online "clickbait" headlines and relies on advertising revenue generated from this activity, regardless of the veracity of the published stories.[4] Intentionally misleading and deceptive fake news differs from obvious satire or parody, which is intended to amuse rather than mislead its audience. The relevance of fake news has increased in post-truth politics. For media outlets, the ability to attract viewers to their websites is necessary to generate online advertising revenue. If publishing a story with false content attracts users, it may be worth producing it to benefit advertisers and improve ratings. Easy access to online advertisement revenue, increased political polarization, and the popularity of social media,[1] primarily the Facebook News Feed,[1] have all been implicated in the spread of fake news,[4][7] which has come to provide competition for legitimate news stories. Hostile government actors have also been implicated in generating and propagating fake news, particularly during elections.[8]

## **Wikipedia**

Wikipedia (About this sound listen) is a multilingual, web-based, free-content encyclopedia that is based on a model of openly editable content. It is the largest and most-popular general reference work on the Internet,[3][4][5] and is named as one of the most popular websites.[6] It is owned and supported by the Wikimedia Foundation, a non-profit organization which operates on whatever money it receives from its annual fund drives.[7][8][9] Wikipedia was launched on January 15, 2001 by Jimmy Wales and Larry Sanger.[10] Sanger coined its name,[11][12] a portmanteau of wiki[notes 3] and encyclopedia. There was only the English-language version initially, but similar versions in other languages quickly developed which differ in content and in editing practices. With 5,664,463 articles,[notes 4] the English Wikipedia is the largest of the more than 290 Wikipedia encyclopedias. Overall, Wikipedia comprises more than 40 million articles in 301 different languages[14] and had 18 billion page views and nearly 500 million unique visitors each month as of February 2014.[15]

## **Neural Networks**

The term neural network was traditionally used to refer to a network or circuit of neurons.[1] The modern usage of the term often refers to artificial neural networks, which are composed of artificial neurons or nodes. Thus the term may refer to either biological neural networks, made up of real biological neurons, or artificial neural networks, for solving artificial intelligence problems. The connections of the biological neuron are modeled as weights. A positive weight reflects an excitatory connection, while negative values mean inhibitory connections. All inputs are modified by a weight and summed. This activity is referred as a linear combination. Finally, an activation function controls the amplitude of the output. For example, an acceptable range of output is usually between 0 and 1, or it could be -1 and 1. Unlike von Neumann model computations, artificial neural networks do not separate memory and processing and operate via the flow of signals through the net connections, somewhat akin to biological networks. These artificial networks may be used for predictive modeling, adaptive control and applications where they can be trained via a dataset.

## **Master Thesis Topic**

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Morbi nulla ligula, blandit quis mi ut, pellentesque mattis nisi. Praesent elementum diam sed nibh lobortis dictum. Sed eleifend posuere magna, ac rhoncus nisi feugiat in. Donec accumsan ac leo in cursus. Vestibulum id elit lacus. Ut nec rhoncus dolor, quis laoreet dui. Donec luctus fringilla elit.

Sed suscipit dolor dolor, sed consequat neque ornare non. Nunc laoreet, metus ac congue efficitur, neque felis varius metus, in pretium dui nibh eget nisi. Integer maximus porta turpis, et dignissim dolor tempus ac. In hac habitasse platea dictumst. Vivamus pretium dui massa, in volutpat orci semper in. Aliquam lorem elit, gravida pretium aliquet id, laoreet eget sem. Mauris mauris lacus, vehicula at felis sit amet, rutrum efficitur est. Pellentesque auctor vestibulum risus sit amet consectetur. Nunc lobortis velit sed nunc aliquam, non egestas nisl vulputate. Nam posuere mauris lectus, nec ullamcorper ipsum placerat eget. Duis quis justo sit amet nunc facilisis ullamcorper. Aliquam posuere suscipit urna. Donec mattis vestibulum odio. Sed ultricies, lorem ut tempor eleifend, elit lacus feugiat orci, sit amet varius lorem enim ac ante. Aliquam ac lacus vel nunc mollis scelerisque ut at erat. Sed pulvinar metus tellus, ut mollis lorem mattis at.

## 2 Related work

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Morbi nulla ligula, blandit quis mi ut, pellentesque mattis nisi. Praesent elementum diam sed nibh lobortis dictum. Sed eleifend posuere magna, ac rhoncus nisi feugiat in. Donec accumsan ac leo in cursus. Vestibulum id elit lacus. Ut nec rhoncus dolor, quis laoreet dui. Donec luctus fringilla elit.

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.

limitations ——— 1) primitive - the complexity of news dealt with is very primitive. Current fake news are very complex and subtle when it comes to the ambiguities 2) Semantic proximity methodology is simple and primitive

But it is a good initial step towards an automated fact checker system.

What improvements are we going to bring? ————— 1) Ours would also be primitive :-) May be slightly better :-) 2) Frequency of conversion of wikipedia content into DBPedia is lesser ?? 3) DBPedia does not cover the whole of wikipedia ????

Things to do: ——— 1) Find the answers for above questions 2) Get the dataset that we can use? 3) Try to look for a demo of this system

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 Research Problem

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Morbi nulla ligula, blandit quis mi ut, pellentesque mattis nisi. Praesent elementum diam sed nibh lobortis dictum. Sed eleifend posuere magna, ac rhoncus nisi feugiat in. Donec accumsan ac leo in cursus. Vestibulum id elit lacus. Ut nec rhoncus dolor, quis laoreet dui. Donec luctus fringilla elit.

Sed suscipit dolor dolor, sed consequat neque ornare non. Nunc laoreet, metus ac congue efficitur, neque felis varius metus, in pretium dui nibh eget nisi. Integer maximus porta turpis, et dignissim dolor tempus ac. In hac habitasse platea dictumst. Vivamus pretium dui massa, in volutpat orci semper in. Aliquam lorem elit, gravida pretium aliquet id, laoreet eget sem. Mauris mauris lacus, vehicula at felis sit amet, rutrum efficitur est. Pellentesque auctor vestibulum risus sit amet consectetur. Nunc lobortis velit sed nunc aliquam, non egestas nisl vulputate. Nam posuere mauris lectus, nec ullamcorper ipsum placerat eget. Duis quis justo sit amet nunc facilisis ullamcorper. Aliquam posuere suscipit urna. Donec mattis vestibulum odio. Sed ultricies, lorem ut tempor eleifend, elit lacus feugiat orci, sit amet varius lorem enim ac ante. Aliquam ac lacus vel nunc mollis scelerisque ut at erat. Sed pulvinar metus tellus, ut mollis lorem mattis at.

## 4 Approach

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Morbi nulla ligula, blandit quis mi ut, pellentesque mattis nisi. Praesent elementum diam sed nibh lobortis dictum. Sed eleifend posuere magna, ac rhoncus nisi feugiat in. Donec accumsan ac leo in cursus. Vestibulum id elit lacus. Ut nec rhoncus dolor, quis laoreet dui. Donec luctus fringilla elit.

Sed suscipit dolor dolor, sed consequat neque ornare non. Nunc laoreet, metus ac congue efficitur, neque felis varius metus, in pretium dui nibh eget nisi. Integer maximus porta turpis, et dignissim dolor tempus ac. In hac habitasse platea dictumst. Vivamus pretium dui massa, in volutpat orci semper in. Aliquam lorem elit, gravida pretium aliquet id, laoreet eget sem. Mauris mauris lacus, vehicula at felis sit amet, rutrum efficitur est. Pellentesque auctor vestibulum risus sit amet consectetur. Nunc lobortis velit sed nunc aliquam, non egestas nisl vulputate. Nam posuere mauris lectus, nec ullamcorper ipsum placerat eget. Duis quis justo sit amet nunc facilisis ullamcorper. Aliquam posuere suscipit urna. Donec mattis vestibulum odio. Sed ultricies, lorem ut tempor eleifend, elit lacus feugiat orci, sit amet varius lorem enim ac ante. Aliquam ac lacus vel nunc mollis scelerisque ut at erat. Sed pulvinar metus tellus, ut mollis lorem mattis at.

Morbi in enim mollis, mollis orci congue, dapibus sem. Vivamus consectetur tortor a diam fringilla fringilla. Duis quis cursus ipsum. Vestibulum tincidunt semper enim, in aliquam elit hendrerit vitae. Vestibulum tincidunt libero non dui gravida feugiat. Quisque interdum sapien vel dui euismod, quis luctus arcu euismod. Vivamus consectetur, purus eu maximus fringilla, magna quam dapibus arcu, at consequat augue mauris vitae turpis. Proin non ipsum sodales, venenatis nisi at, sagittis urna. Vestibulum aliquam nunc quis risus porttitor, vitae egestas dui molestie. Pellentesque ultricies elementum sodales. Nam non molestie tellus. Aliquam erat volutpat. Etiam varius ex dui, non interdum massa pellentesque ac. Pellentesque eu risus ultrices, vulputate leo in, lacinia leo. Vivamus consectetur nunc erat, ut consequat ligula vulputate sed.

## 5 Evaluation

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Morbi nulla ligula, blandit quis mi ut, pellentesque mattis nisi. Praesent elementum diam sed nibh lobortis dictum. Sed eleifend posuere magna, ac rhoncus nisi feugiat in. Donec accumsan ac leo in cursus. Vestibulum id elit lacus. Ut nec rhoncus dolor, quis laoreet dui. Donec luctus fringilla elit.

Sed suscipit dolor dolor, sed consequat neque ornare non. Nunc laoreet, metus ac congue efficitur, neque felis varius metus, in pretium dui nibh eget nisi. Integer maximus porta turpis, et dignissim dolor tempus ac. In hac habitasse platea dictumst. Vivamus pretium dui massa, in volutpat orci semper in. Aliquam lorem elit, gravida pretium aliquet id, laoreet eget sem. Mauris mauris lacus, vehicula at felis sit amet, rutrum efficitur est. Pellentesque auctor vestibulum risus sit amet consectetur. Nunc lobortis velit sed nunc aliquam, non egestas nisl vulputate. Nam posuere mauris lectus, nec ullamcorper ipsum placerat eget. Duis quis justo sit amet nunc facilisis ullamcorper. Aliquam posuere suscipit urna. Donec mattis vestibulum odio. Sed ultricies, lorem ut tempor eleifend, elit lacus feugiat orci, sit amet varius lorem enim ac ante. Aliquam ac lacus vel nunc mollis scelerisque ut at erat. Sed pulvinar metus tellus, ut mollis lorem mattis at.

Morbi in enim mollis, mollis orci congue, dapibus sem. Vivamus consectetur tortor a diam fringilla fringilla. Duis quis cursus ipsum. Vestibulum tincidunt semper enim, in aliquam elit hendrerit vitae. Vestibulum tincidunt libero non dui gravida feugiat. Quisque interdum sapien vel dui euismod, quis luctus arcu euismod. Vivamus consectetur, purus eu maximus fringilla, magna quam dapibus arcu, at consequat augue mauris vitae turpis. Proin non ipsum sodales, venenatis nisi at, sagittis urna. Vestibulum aliquam nunc quis risus porttitor, vitae egestas dui molestie. Pellentesque ultricies elementum sodales. Nam non molestie tellus. Aliquam erat volutpat. Etiam varius ex dui, non interdum massa pellentesque ac. Pellentesque eu risus ultrices, vulputate leo in, lacinia leo. Vivamus consectetur nunc erat, ut consequat ligula vulputate sed.

## 6 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

## 7 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

## 8 Signatures

---

Kandhasamy Rajasekaran

---

Prof. Dr. Steffen Staab

---

Supervisor????

---

Lukas Schmelzeisen

## **9 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 10, 2018

---

Kandhasamy Rajasekaran