DETECTING HEALTH MISINFORMATION IN WEB PAGE TEXT USING DEEP LEARNING METHODS

Dione Morales

Bachelor of Engineering
Computer Engineering Stream

School of Engineering
Macquarie University

November XX, 2018

Supervisor: Associate Professor Adam Dunn

ACKNOWLEDGMENTS

I would like to acknowledge ...

STATEMENT OF CANDIDATE

I, (insert name here), declare that this report, submitted as part of the requirement for the award of Bachelor of Engineering in the School of Engineering, Macquarie University, is entirely my own work unless otherwise referenced or acknowledged. This document has not been submitted for qualification or assessment an any academic institution.

Student's Name:

Student's Signature:

Date:

ABSTRACT

This is where you write your abstract ...

Contents

List of Figures

List of Tables

Chapter 1

Introduction

One of the key components required to minimize the propagation of misinformation online is to have the ability of automatically evaluating and quantifying the credibility of articles. However, traditional automated methods - such as shallow learning-based techniques, still require the domain knowledge of experts to be able to develop the features required by the model. Thus, this project aims to investigate the performance of Deep Learning-based (DL) techniques in evaluating the credibility of information within domain-specific articles via the classification of set criteria that have deemed to be highly correlated with articles that have low credibility. Specifically, this project will focus on evaluating the credibility of online health articles related to vaccination due to the commonly misinformed and controversial views associated with its effects [3].

1.1 Project Overview

This section details the scope of the project and its associated outcomes outlining the various tasks that must be accomplished to successfully complete the project.

1.1.1 Project Scope

With the primary objective of this project being the evaluation on the effectiveness of deep learning models in determining the credibility of online health-related articles. Due to the complexity of this project, a set of activities - divided into main goals and stretch goals, have been defined to ensure that the completion of this project remains feasible in the given time frame. The completion of all activities categorized as main goals will signal the realization of the primary objective and the completion of the project. Stretch goals are activities of interest that have been identified as non-essential to the completion of the primary objective but (talk about the overarching goal that all stretch goals have in common e.g. understand the model, utilize the model etc.) and will be worked on after the completion of the project.

Main Goals

- Develop a set of criteria that will be used to determine the credibility of online articles.
- Evaluate the performance of common ML-based methods on the classification of online vaccine-related articles, based on the criteria developed.
- Evaluate the performance of the proposed DL model on the classification of online vaccine-related articles, based on the criteria developed.
- Evaluate the effect of transfer learning methods in the performance of the proposed DL method (assuming that the chosen method doesn't rely on transfer learning)
- Evaluate the effectiveness of various transfer learning methods for the classification task

Stretch Goals

- Utilize attention mechanisms to understand how the aforementioned DL model classifies the criteria for credibility.

Chapter 2

Background and Related Work

A literature review has been conducted to develop an understanding on the research that has been done in the assessment of the credibility of information, specifically in the context of information related to health and the limitations and capabilities of shallow learning techniques and how it differs from deep learning-based methods for the task of text classification.

2.1 Assessing Credibility of Information

To have the capability of automating the process of evaluating the quality or credibility of online information, a definition that outlines of what is required by an article to be considered as a credible source of information must be developed. While there has been a significant amount of research that has been done on the development of tools and frameworks that aim to assess the credibility of online health information, there is currently no standardized method or benchmark that is universally used. Commonly used tools and frameworks are (*Note: add citation for each one*): DISCERN [1], HealthNewsReview [2], QIMR [5] and (*Wiley 2017 - update this when you find the source*)

Talk about the work done in establishing the measurement of quality in online health information e.g. DISCERN, QIMR and the sources from that document in the slack channel

Try to understand the following for each source:

- How the scoring system works
- How credibility is defined

2.1.1 DISCERN

2.1.2 HealthNewsReview

2.1.3 QIMR

2.1.4 Wiley 2017

2.2 Prior Approaches

Discuss the prior work that has been done in terms of text classification e.g. spam, sentiment, topic

2.2.1 Shallow Learning Models

For each model, talk about the following:

- *How it works and the mechanisms involved*
- *Advantages*
- *Limitations*

Naive Bayes

Support Vector Machines

Artificial Neural Networks

2.2.2 Feature Selection

Talk about word embeddings e.g. GloVe, word2vec, fastText, ngrams and its variants (skip-grams, sn-grams), BoW etc. and justify which features I will be using for this project.

Bag of Words

N-Grams

GloVe

Word2Vec

Language Models

2.3 Deep Learning

Introduce the state-of-the-art DL based approaches for text classification and try to compare it performance with state-of-the-art ML approaches

For each model, talk about the following:
NOTE: REMEMBER WHEN WRITING THIS SECTION TO ALWAYS CONSIDER HOW IT DIFFERS TO ML TECHNIQUES

- *How it works and the mechanisms involved*
- *Advantages*
- *Limitations*

Deep learning models are a class of machine learning models that have the capability of automatically learning a hierarchical representation of data. These hierarchical representations are constructed through the use of artificial neural networks, the main underlying mechanism of deep learning models. Typically, large amounts of training data is required to train a model in learning the language model required to attain state of the art results, in the task of text classification for instance, the size of commonly used non-domain specific datasets range from hundreds of thousands of training examples to millions [4] [6] *(note: look into the datasets used by state of the art approaches)*. Due to these constraints, it is not feasible to procure a dataset for the domain specific task of this project due to the aforementioned knowledge expertise and time requirements to manually label the articles required. Hence, *(Talk about transfer learning/N-shot learning/domain adaptation here)* will be used to overcome this issue.

Introduce the typical architectures used for text classification e.g. RNNs, LSTMs, CNNs, GRUs?

2.3.1 Deep Learning Models

Recurrent Neural Networks

Gated Recurrent Unit Networks

Long Short-Term Memory Networks

Convolutional Neural Networks

2.3.2 GET PROPER NAME FOR THIS SECTION

Transfer Learning

Talk about transfer learning and how it works and how it is applicable to this project.

N-Shot Learning

Talk about zero/few/etc-shot learning and how it works and how it is applicable to this project.

2.4 Conclusion

Summarize lit review and describe why DL-based approaches should be preferred over ML-based for this type of problem. Also talk about Transfer/N-Shot learning and describe which one will be feasible given the project's time constraints

Chapter 3

Proposed Approach

3.1 Rationale

Introduce and discuss the factors that led to me choosing the proposed approach

3.2 Credibility Criteria

Introduce and discuss the 7 criteria that will be classified and describe how the criteria was determined

3.3 Study Data

Talk about the data I'll be using, how we got it, its characteristics etc.

3.4 System Model

Describe the architecture of the model

3.5 Experiments

Describe the experiments that I'm planning to do (in such a way that they are easily reproducible)

3.6 Outcome Measures

Talk about the type of analyses that I'll be doing to determine the performance of my proposed model

Chapter 4

Conclusions and Future Work

4.1 Conclusions

The end

4.2 Future Work

Chapter 5

Abbreviations

AWGN	Additive White Gaussian Noise
BC	Broadcast Channel
BS	Base Station
CSI	Channel State Information
CSIR	Channel State Information at Receiver
CSIT	Channel State Information at Transmitter
dB	Decibels
DPC	Dirty Paper Coding
GS	Gram-Schmidt
RVQ	Random Vector Quantisation
SISO	Single Input Single Output
SNR	Signal to Noise Ratio
SINR	Signal to Interference plus Noise Ratio
MISO	Multiple Input Single Output
SIMO	Single Input Multiple Output
MIMO	Multiple Input Multiple Output
MMSE	Minimum Mean Square Error
MRC	Maximum Ratio Combining
QoS	Quality of Service
TDD	Time Division Duplex
FDD	Frequency Division Duplex
ZF	Zero-Forcing
ZFBF	Zero-Forcing Beamforming
ZMCSCG	Zero Mean Circularly Symmetric Complex Gaussian

Appendix A

name of appendix A

A.1 Overview

here is the Overview of appendix A ...

A.2 Name of this section

here is the content of this section ...

Appendix B

name of appendix B

B.1 Overview

here is the Overview of appendix B ...

B.2 Name of this section

here is the content of this section ...

Bibliography

[1] "DISCERN - The DISCERN Instrument." [Online]. Available: http://www.discern.org.uk/discern{_}instrument.php

[2] "Our Review Criteria - HealthNewsReview.org." [Online]. Available: https://www.healthnewsreview.org/about-us/review-criteria/

[3] D. C. Burgess, M. A. Burgess, and J. Leask, "The MMR vaccination and autism controversy in United Kingdom 1998-2005: Inevitable community outrage or a failure of risk communication?" *Vaccine*, vol. 24, pp. 3921–3928, 2006. [Online]. Available: https://ac.els-cdn.com/S0264410X06002076/1-s2.0-S0264410X06002076-main.pdf?{_}tid=46d1dda6-f576-4f5e-ad53-d550f1cd9990{&}acdnat=1534726962{_}1b237371d8bb916694f34f0f951c84bc

[4] A. Conneau, H. Schwenk, Y. Le Cun, and L. Loıc Barrault, "Very Deep Convolutional Networks for Text Classification," 2017. [Online]. Available: https://arxiv.org/pdf/1606.01781.pdf

[5] D. Zeraatkar, M. Obeda, J. S. Ginsberg, and J. Hirsh, "The development and validation of an instrument to measure the quality of health research reports in the lay media," *BMC Public Health*, vol. 17, no. 1, p. 343, dec 2017. [Online]. Available: http://bmcpublichealth.biomedcentral.com/articles/10.1186/s12889-017-4259-y

[6] X. Zhang, J. Zhao, and Y. Lecun, "Character-level Convolutional Networks for Text Classification," 2015. [Online]. Available: https://arxiv.org/pdf/1509.01626.pdf