Identification of Vaccine Misinformation Online

Jonathan Waring and Dr. Shannon Quinn
Department of Computer Science, Franklin College of Arts and Sciences, University of Georgia, Athens, 30602 GA

Introduction

Vaccination provides the most effective method of preventing infectious diseases. While the effectiveness and safety of vaccines has been widely studied and verified, there is still opposition from the anti-vaccine movement. It has led to vaccine hesitancy, which is defined as a delay in acceptance or a refusal of vaccine services. It is an ever-growing and constantly changing problem that needs constant surveillance. The Internet plays a large role in disseminating vaccine misinformation to a large number of people, which contributes to the vaccine hesitancy problem. In order to combat the spread of misinformation online, it is important to first recognize true facts from false ones. In order to help solve this problem, we attempt to develop a machine learning strategy using natural language processing (NLP) that allows one to identify misinformation in vaccine-related webpages.

Objective

The results of this study could enable both public health practitioners and the general public to monitor vaccine misinformation online in order to reduce vaccine hesitancy and identify strategies to improve vaccine education.

Methods

Semi-Supervised Classification

Semi-Supervised Classification uses both labeled and unlabeled data in order to predict the classification of unseen data. For our project, the labeled data consists of 20 manually labeled vaccine webpage documents as either TRUE or MISINFORMED. The unlabeled data consists of 1095 vaccine webpage documents collected through the use of Google's Custom Search API [1] and Python's Goose-Extractor Library [2]. We then attempt to infer the label of the unlabeled examples in advance before building the classifier (transductive learning). This is accomplished through the use of the Doc2Vec algorithm [3].

Doc2Vec

Doc2Vec is a low-dimensional document embedding algorithm that represents a document in vector space. We build document embedding for all the collected documents on vaccines, and then use pairwise cosine similarities (formula shown below) between document vectors to infer labels. After inferring the labels of our unknown documents, our collection contained 1,021 TRUE documents (91.6%) and 94 MISINFORMED documents (8.4%).

$$\text{similarity} = \cos(\theta) = \frac{A \cdot B}{\|A\|\|B\|} = \frac{\sum_{i=1}^{n} A_i \times B_i}{\sqrt{\sum_{i=1}^{n} (A_i)^2} \times \sqrt{\sum_{i=1}^{n} (B_i)^2}}$$

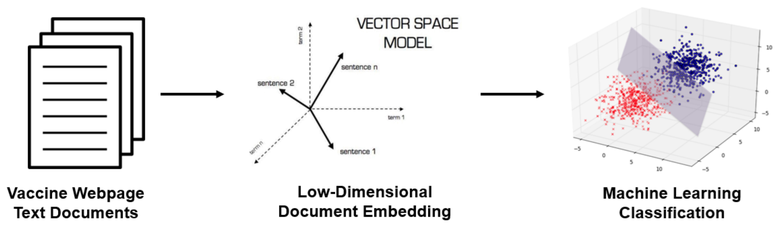

Figure 1: Visual representation of the methodology used to produce results.

Machine Learning Classifiers

We had four different classification tasks to address, as follows:

1. Build a supervised classification model using the inferred documents as the training data. Test the model on the known labeled documents.
2. Build 10-fold cross-validated classification models where we split into test and training data using a combination of the inferred and known labeled documents.
3. Build classification models using a combination of the inferred and known labeled documents. Predict to see what the model thinks the label for the unknown documents would be. Compute the proportion of what cosine distance inference and the model prediction results in.
4. Repeat Task 3 with known documents only as the training data.

Results

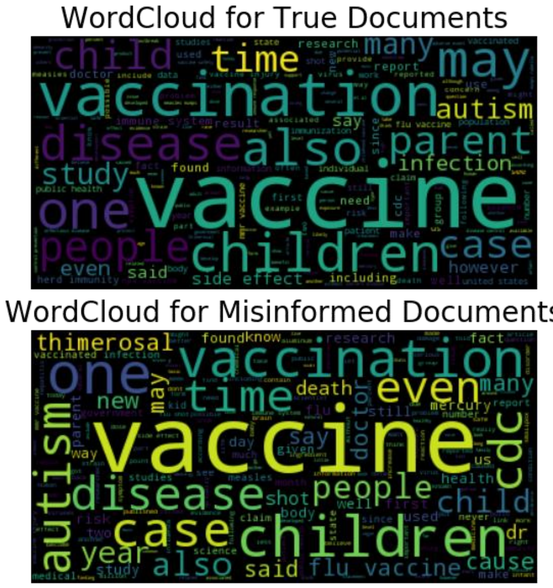

Figure 2: Word Cloud representations of TRUE and MISINFORMED documents.

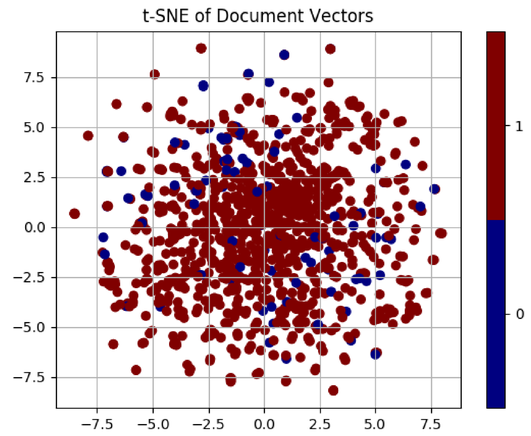

Figure 3: t-SNE scatterplot visualization of TRUE and MISINFORMED documents.

MACHINE LEARNING CLASSIFIERS

TASK	Logistic Regression	Naïve Bayes	SVM	Random Forest	KNN
1	80%	95%	80%	55%	50%
2	93.7% (± 1.56)	83.1% (±2.67)	93.6% (±1.51)	91.0% (±0.79)	91.6% (±1.05)
3	100%	84.0%	100%	99.9%	92.3%
4	89.3%	84.0%	91.1%	80.4%	92.8%

Table 1: Summary of the accuracies of the classification tasks

The above table shows the achieved accuracy of the machine learning classification tasks. All algorithms perform well, except for Random Forest and KNN in Task 1.

Discussion

First and foremost, Doc2Vec serves as a reasonably good starting point for document classification and, ultimately, a semi-supervised framework for identifying false or misleading documents. Given the results of the different classification tasks all falling within the 80%-95% range for most algorithms, it seems that this is a fairly reasonable assumption.

The more important question remains of what truly makes a document reliable or not. We gain a few insights from our Word Clouds in Fig. 2 as it relates to vaccination. It is clear that there are several terms that are mentioned more heavily in MISINFORMED versus TRUE documents, such as autism, flu vaccine, CDC, thimerosal, mercury, cause, death. On the other hand, terms such as, study, parent, case, may, side effect, and time appear more in TRUE documents. To gain further insight, we attempt to interpret the results of the t-SNE visualization to see if TRUE and MISINFORMED documents look different in 2-dimensional space, but there are no apparent patterns. Future works should look to expand upon this issue.

Acknowledgements

This work was supported by the University of Georgia's Center for Undergraduate Research Opportunities Office through the CURO Research Assistantship Award in the Spring 2017 term.

References

1. https://developers.google.com/custom-search/
2. Xavier Grangier, python-goose, Github Repository, https://github.com/grangier/python-goose.
3. Q. Le, T. Mikolov. 2014. Distributed Represenations of Sentences and Documents. *In Proceedings of ICML 2014*.

Contact Information

If you have any questions, feel free to contact Jonathan at jwaring8@uga.edu. The following QR code links to the GitHub Repository hosting the source code for Jonathan Waring's CURO 2017 Project.