Unfit to Print: Detecting Sarcasm in News Headlines

Ben Collins and Drishaan Jain (DJ)
University of Michigan
`{collb, drishaan}@umich.edu`

1 Project Description

In recent years, social media sites have become one of the, if not the predominate medium of communication. Social media serves as a platform for sharing information ranging from news to personal details about one's own life. However, due to its limited, digital form, online communication has historically lacked the sophistication required to express more nuanced aspects of in-person communication, such as sarcasm.

It is this ambiguity that makes text analysis for a computer or a human difficult with the limited range of expression. A technology that could automatically detect sarcasm would prove useful in a variety of contexts including, but not limited, to social media teams moderating conversations or user analytics.

Our goal is to recreate more primitive models that attempt to predict sarcasm in news headlines. These news headlines make a great source for our data due to the standardized structure and format of a headline. Our choice of model is intentionally unconstrained as we intend to compare the results from a variety of them.

2 Related Work

Over the years there have been many theories crafted about how to best detect the presence of sarcasm. Looking at more recent attempts we can see works such as that of Davidov (2010) who devised a very general KNN method for sarcasm detection and others like Riloff (2013) that sought to do so through a hand-crafted lexicon. At this moment, the leading edge technology comes from the work of Rajadesingan (2015) who improved upon previous methods by developing a feature that estimates a user's likelihood of being a sarcastic person that, provided enough data, achieves a rough accuracy of 83%.

Our goal will be to explore how the different framings of this task affect the accuracy score. Utilizing different methods and different feature representations we seek to analyze how each of this different models help or hinder its ability to discern sarcasm in news headlines.

3 Dataset and Usage

3.1 Summary of Data

This dataset is the News Headlines Dataset For Sarcasm Detection from Kaggle. It contains 26,709 rows of news headlines with 14,985 sourced from *The Huffington Post* and the other 11,724 from *The Onion*. This a reasonably well-balanced dataset. Each row contains the headline text, a 1 or 0 indicating sarcasm being present or not, and 'huffingtonpost' or 'theonion' indicating its source. It is worth noting that the last two columns are just different encodings of the same information as all the headlines from sarcastic headlines were written by *The Onion* and vice versa.

3.2 Data Usage

We will use the text to build our list of features and the 0 or 1 indicating sarcasm as our labels for training and testing purposes.

4 Methodology

4.1 Data Preparation

To pre-process the news headlines, we will convert all text to lowercase but otherwise leave the headlines intact. Headlines may use inconsistent upper- and lower-casing which we aim to normalize. However, stopwords and punctuation could be helpful indicators of sarcasm, so we will not remove these valuable portions of the headline. We will use 80% of the dataset for training, 10% for validation, and 10% for testing.

4.2 Classification

We will use three different classification models and compare their performance to determine the most effective type. The expected behavior of the classifier is to receive a headline as an input and accurately output whether it contains sarcasm (1) or not (0). The most simple model will be Naive Bayes utilizing a bag-of-words approach.

The second model is logistic regression. This model will be tested with both unigrams and bi-grams to determine how accuracy is impacted by the change.

The final model is a support vector machine (SVM). We would like to observe how decision boundaries are determined and if there are certain words or formats that are weighed heavily in predicting the presence of sarcasm.

4.3 Prediction

In addition to the classification models, we will implement a Hidden Markov Model to develop new headlines that are either sarcastic or not.

5 Evaluation

To evaluate each classification model, we will compute its F1 score, AUROC score, and accuracy. These metrics will provide insight regarding the precision, recall, and efficiency. The AUROC score will also serve as a standalone metric for each classifier, to determine how much better or worse it performed compared to a random classification (see Figure 1).

For our Hidden Markov Model, we will use human evaluation to determine its success. For example, we can generate a set of headlines that contain sarcasm and a set of non-sarcastic headlines using
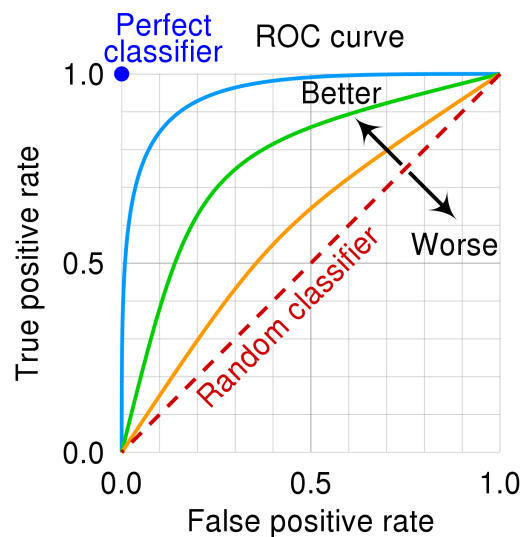

Figure 1: Interpretation of AUROC curve

the HMM. Then, we can survey other University of Michigan students to determine whether the headline is understandable and whether they would consider it sarcastic or not upon reading it.

References

O. Tsur D. Davidov and A. Rappoport. 2010. *In Proceedings of the Fourteenth Conference on Computational Natural Language Learning*. Association for Computational Linguistics, Cambridge, UK.

P. Surve E. Riloff, A. Qadir and Silva. 2013. *EMNLP*. ACL, Cambridge, UK.

Ashwin Rajadesingan, Reza Zafarani, and Huan Liu. 2015. Sarcasm detection on twitter. *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining*.