Sentiment Analysis vs. Human Inference for Detecting Sarcasm - Proposal

Chris McVeigh

1 Project Outline

It is often recognised that sarcasm is very difficult to parse in writing. Effective use of sarcasm in a conversation requires some degree of familiarity with both the topic in question and the speaker themselves. However, online discussions frequently replicate real-life conversations, which means that sarcastic statements can appear just as often as sincere ones. Although this familiarity is not present in an anonymous online discussion, humans are able to use contextual information to determine whether a given statement is meant ironically or sincerely (e.g. prevailing opinion of the group where the comment was posted). Using a dataset of Reddit comments, I will develop a sentiment analysis tool that aims to identify potentially sarcastic statements. My goal is to establish how the addition of context may improve the tool's results, and I will also use human responses to the data in a web-based solution to find out how the same applies to human understanding of natural language, and whether the effect is the same for humans and computers.

2 Method

The dataset I intend to use is SARC [Self-Annotated Reddit Corpus] 2.0 (https://nlp.cs.princeton.edu/SARC/2.0/) by Mikhail Khodak, Nikunj Saunshi, and Kiran Vodrahalli. This corpus contains Reddit comments from a variety of subreddits, posted between 2009 and 2017. The training data provided marks comments as being sarcastic or not, with options available that provide either a 50/50 split of sarcastic and sincere comments, or a more representative figure of less than 1% sarcastic.

The development of this project is divided into two parts - the sentiment analysis tool and the web-based tool for human responses.

2.1 Sentiment Analysis Tool

I will use the NLTK library for Python to develop a tool that can recognise sarcasm in Reddit comments. By analysing the features that distinguish sarcasm in text (word choices, punctuation, formatting, etc.), I will first attempt to classify sarcastic comments based purely on their content. I will then introduce context for the same dataset and run again. Context here means the comments featured in the rest of the thread, the subreddit to which the post belongs, the score the comment received, and perhaps information about the comment author, which SARC also provides. My hypothesis is that the precision/recall/F_1 scores for the data with context will be significantly higher, as the contextual information we use when parsing sarcasm in real life can be approximated by the comment's metadata.

2.2 Human Response Survey

In order to contextualise the results of the sentiment analysis tool, I intend to run a similar test using human participants. I will use the same dataset, and present the user with a random selection of comments. For each comment, the user will indicate whether they believe the statement is sarcastic or not. The same set of comments will then be given proper context, with users shown the thread each comment originally belonged to, along with the subreddit it was posted to, and its original score. The user will then answer again. I will be able to take the precision/recall/F_1 scores of each respondent and find the average human result, for which I expect the outcome to be similar to the sentiment analysis tool - that people are better at recognising sarcasm when they are presented with context.

3 Report Plan

My final report will focus on the performance of the sentiment analysis tool before and after being provided context, with the human data intended to legitimise my expected findings. If there are large disparities between the human and computer scores, I will examine which kinds of comments the victor was better at recognising, and if they score more closely I will see where the strongest agreements were. I will use my data to summarise how sentiment analysis responds to sarcasm compared to how humans infer the information ourselves.