Peter Riser

Article Sentiment Predictor

Finding a reasonable way to model the sentiment of a company or firm is in high necessity. To address this, I am going to try to quantify the sentiment of an article using a machine learning algorithm; The algorithm will use new articles of the past few days to calculate the public sentiment of the company. To do this I am first going to create a learning database of several hundred articles to create a keyword table that will learn to correlate the frequency of words with the sentiment of the articles. The table will have binary good/bad words which can then be weighted based on frequency and used to measure not only whether an article is positive or negative, but how positive it is.

This is interesting because a large part of Economics (my other major) is quantifying and modeling the sentiment and feelings of people to try to make predictions. With this, you can make decisions related to marketing, stock purchases or even customer reviews.

The input for the system is going to be a company name. The company will then be checked against Google and Yahoo databases where the most recent financial/economic articles will be and they will use those articles to calculate an average sentiment which will then be returned to the user. The sentiment is expressed as a numerical value with 1 being neutral, greater than 1 being positive and less than 1 being negative. The internet is full of articles but I am going to use data from newspapers to get the initial learning set as there is already api’s and data that you can pull.

The algorithm will use supervised learning. I will give the algorithm its initial data with the knowledge whether it is positive or negative.