NATURAL LANGUAGE CLASSIFICATION OBJECTIVES

EXECUTIVE SUMMARY

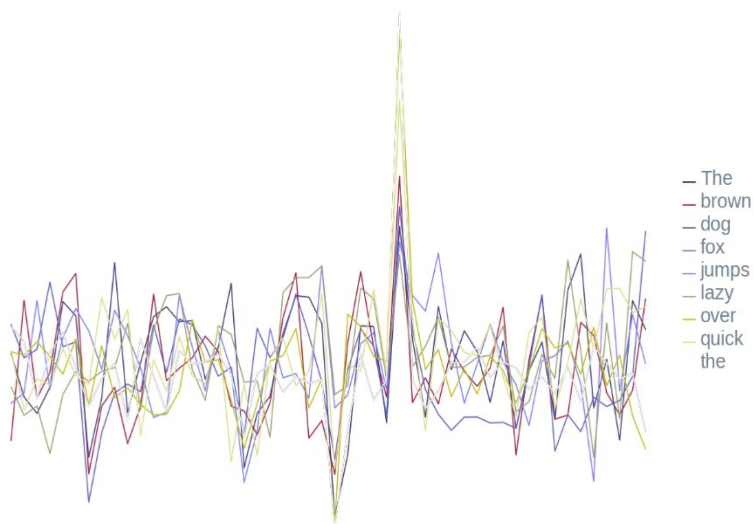
Establishing an effective and efficient customer management system is one of the most critical elements for success in any customer-facing organization, according to an article published in the Harvard Business Review, titled "The Four Things a Service Business Must Get Right." Any company, no matter how long established, can boost both top-line and bottom-line revenue by automating the process of categorizing customer feedback into homogenous, manageable and prioritized segments.

In the information age, where we are all perpetually connected by various social media platforms and smartphones, companies are especially being overwhelmed with real-time customer feedback posted publicly on this medium, which can have dramatic effects on even the largest companies. These situations can arise from public figures the likes of Carl Icon or Elon Musk, posting single tweets on social media which both caused drastic moves in the stock price of their respective mentioned companies, to a relatively unknown local reporter tweeting from a local communities courtroom.

These are extreme examples of public sentiment dictating a company's stock price. However, the point remains that a successful company must have a way to not only consume large volumes of user feedback on a variety of mediums, they must also be able to classify this written feedback into overall sentiment analysis of the content therein, and quickly dispatch the feedback to the appropriate operational departments as quickly as possible.

RESEARCH DESIGN

For this problem we used a set of two pre-trained word embeddings from GloVe, short for Global Vectors for Word Representation, developed by researchers at Stanford University, to transform the written language content into its numeric representation. These encodings allow us to transform a sequence of words, or a sentence, into a sequence of numeric vectors of which we can derive mathematical models.



As depicted on the right, we can see the numerical representation of the famous sentence, "The quick brown fox jumps over the lazy dog", which contains all the letters in the English alphabet. This sentence makes a perfect sample for a visual illustration of the data we are working within this model.

For the actual training and test data to conduct our research, we will use a pre-selected sample of 1,000 textual movie reviews. Each of the reviews has either a positive or negative sentiment, a thumbs up or a thumbs down, associated with the content. We will concatenate these reviews together as a single series with the pre-defined sentiment in an associative data structure. This design will allow us to shuffle the data and execute a split, train and test cross-validation approach that is industry standard for the validation phase of our research.

TECHNICAL OVERVIEW

The model construction and cross-validation methods were conducted purely in the cloud, leveraging a pre-canned environment from a world-class provider of machine learning solutions, Google, called Colabratory. This cloud environment enables us to execute research more efficiently with their hardware targeted machine learning accelerated platform, while also allowing maximum reproducibility of our results by taking specific environment setups out of the equation. We can also publish our research globally and allow any astute reader to reproduce our results for themselves in a sandboxed environment.

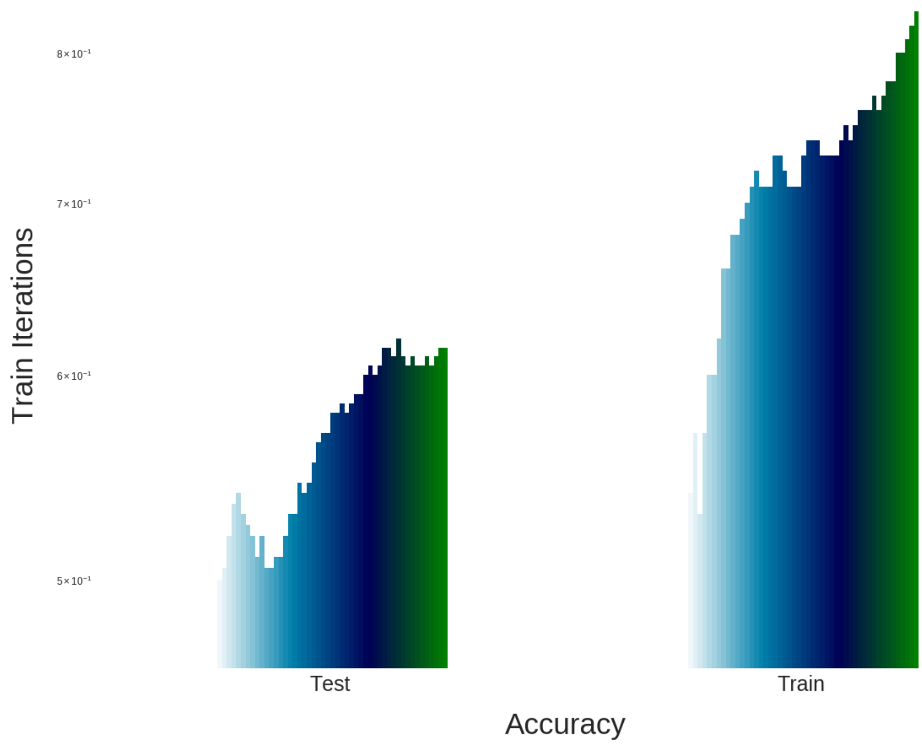
Additionally, for this research, we also leveraged an industry-leading machine learning framework, TensorFlow, which Google also produces. TensorFlow gives us access to the same underlying technology that powers several of the most advanced analytical systems in production today.

Using this framework, we set up an artificial neural network algorithm for training our classification system. For this problem, we chose a Recurrent Neural Network (RNN) algorithm that is configured to use the pre-trained word embeddings from the GloVe database. The RNN model is given a randomized mixture of both positive and negative reviews from the training dataset, which is 80% of the overall reviews, in batch sizes of one-hundred reviews each iteration, for a total of fifty iterations.

CONCLUSION

The results from the sentiment analysis conducted with the Recurrent Neural Network performed well with our cross-validation metrics. With the training dataset, our model was able to accurately predict the sentiment in 83% of our training data.

However, when we look at the results of our model running against the 20% of the dataset we set aside before we constructed our model, we can see that the accuracy lowers to 61.5% when predicting against data it hasn't seen before.



We also note that attempts to increase the training epochs, as it is evident that the number of training iterations leads to higher prediction accuracy during validation, lead to higher test data prediction accuracy, up to 96%. However, this over-fit the training data and the model performed even lower against the unseen test data, by approximately three-hundred basis points.

Overall, the analysis performed well enough that we would recommend it for a high-level classification system for assessing the customer's sentiment of written product reviews or customer experience surveys. We recommend further research to not only analysis the overall sentiment of the text, but also the degree of strength in which these feelings expressed. Such quantifiables would enable us to dispatch a specialist to the customers with the worst experiences to customer experience specialists for promotions on repeat transactions or to survey the warmest reviews for what gave them the great experience so that we can increase these facilities.

For additional information on how this research is conducted, please visit the [Colabratory Notebook](#).