

Final Project: Sentiment Analysis of Amazon Product Reviews

I will be using an Amazon product reviews dataset to conduct sentiment analysis based on the star rating of the review. I plan on testing TF-IDF, Word2Vec, and GloVe vectorizers to create the word vectors and embeddings. Reviews that are later considered too short will be removed from the dataset. The resulting vectors will then be run through an ensemble classifier such as Random Forests, a simple 1-2 hidden layer dense neural network, a 1D convolution neural network, and recurrent neural networks to compare accuracy of classification, F1 score, and training times with each pipeline. One study suggested SVMs are the best method among traditional machine learning algorithms for classifying Amazon reviews (Haque, Saber, & Shah, 2018). A separate pipeline can be trained on each category of products. For example, video games are likely rated different than beauty products. The vocabulary will differ and so will the resulting embeddings. Similar analysis has been done on a much smaller dataset with less category abstraction (Prakoso, 2018).

By developing on a few categories, I plan on fitting the best pipeline to as many categories as time allows. This will build on the results on previously cited studies. The resulting model can then be used to predict and flag future text reviews of a product in the modeled category. We can also go into the embeddings layers and analyze what vocabulary words are kept, how many words were needed to best classify the reviews, and compare them between categories. By analyzing the word clusters and embeddings, we may find that some products are reviewed with similar sentiment and lexicon.

References

- Haque, T. U., Saber, N. N., & Shah, F. M. (2018). Sentiment analysis on large scale Amazon product reviews. *2018 IEEE International Conference on Innovative Research and Development (ICIRD)*. doi: 10.1109/icird.2018.8376299

Prakoso, A. A., Yananta, B. W., Setyawan, A. F., & Muljono. (2018). A Lexicon-Based Sentiment Analysis for Amazon Web Review. *2018 International Seminar on Application for Technology of Information and Communication*. doi: 10.1109/isemantic.2018.8549812