Multi-Class Sentiment Analysis on Amazon Fine Food Reviews

Nikolaos Gounakis, AM: 1254

December 2021

1 Intro

In Web applications it is useful to understand various concepts of text. One of them is the ability to understand if a piece of text provides a negative, a positive or a neutral meaning. That can result to a new way of producing analytics for users when analyzing reviews or social media posts. The problem described is a multi-class classification problem and we propose a solution of constructing a ML model trained in Amazon Fine Food Reviews Dataset [2].

2 Dataset

This dataset consists of reviews of fine foods from amazon. The data span a period of more than 10 years, including all ≈500,000 reviews up to October 2012. Reviews include product and user information, ratings, and a plain text review. It also includes reviews from all other Amazon categories. `https://www.kaggle.com/snap/amazon-fine-food-reviews`

3 Methodology

There are several techniques to extract features from text. We will use and compare two existing methods **Doc2vec** (`https://radimrehurek.com/gensim/models/doc2vec.html`) and **TF-IDF** (`https://scikit-learn.org/stable/modules/feature_extraction.html`), used and tested on the following paper [1] and evaluate them on almost the same algorithms. It is

important to mention that in our case a feature selection must be performed because of the large corpus. We will do that by tuning the hyperparameters of the methods we mention like the authors did in [1].

After the feature extraction we will split the data (train,test) using **10-Fold Cross Validation** and then train and evaluate the resulting models of the following algorithms for multi-class classification: **KNN**, **SVM**, **Logistic Regression**, **Naive Bayes** and **Random Forest**.

For each algorithm and feature extraction combination we will compute the following metrics: **Accuracy**, **AUC** and **F1**.

Finally we will analyze the results and propose the final model.

References

[1] Avinash Madasu and Sivasankar E. A study of feature extraction techniques for sentiment analysis, 2019.

[2] Julian John McAuley and Jure Leskovec. From amateurs to connoisseurs: Modeling the evolution of user expertise through online reviews. In *Proceedings of the 22nd International Conference on World Wide Web*, WWW '13, page 897–908, New York, NY, USA, 2013. Association for Computing Machinery.