Queen Mary University of London
School of Electronic Engineering and Computer Science

Investigation in to Sentiment Analysis and Machine Learning to Develop a Food Recommendation Mobile Application

MIHAIL BUTNARU
150186618

ABSTRACT

The growth and popularity of social media has caused the creation of vast amounts of personal data. Large volumes of text and metadata in the form of feedback, comments, reviews and suggestions are generated every day, creating an opportunity for companies to provide truly bespoke services. Social media is regarded as an important source of knowledge, because of the breadth of topics discussed.

This thesis, follows the research and development of sentiment analysis and machine learning approaches. In order to address the problem of deciding where this cuisine can be procured. The sentiment analysis and machine learning will be centred around data collected from a user's Facebook account. This will then be extracted and used to identify and evaluate the sentiments of posts.

Sentiment analysis will classify sentiments from every post and assign it into the correct class: positive, negative and neutral. From each class, every word score will be computed by a supervised learning method and the analysed data split in seventy percent train data and thirty percent test data. The analysis and performance of the Naïve Bayes Classifier, Support Vector Machines Classifier and Maximum Entropy has a significant impact on bigram features rather than on unigram features. Bigrams features achieved a higher performance in all instances, giving an accuracy of eight-five percent. Unigrams achieved the least of eighty percent. Naïve Bayes classifier and Support Vectors Machines achieved a maximum of eighty-five percent and eighty-three percent respectively, indicating a significant impact on the performance of the algorithm.

The selected sentiment analysis and machine learning methods, were identified with respect to project requirements their high accuracy based on unigram and bigram features. Given these achievements, prediction of the user's desired cuisine can be calculated. Addressing the problem choice overload experienced by users when faces with many restaurant options.

CONCLUSION

In conclusion, the study showed positive results that the mobile application regarding the topic – 'deciding what and where to eat out in London'. The evaluation session and user participation during the tests, succeeded in achieving the project objective. The computing performance of the mobile application exceeds the traditional methods in finding a place to eat by achieving results in seconds. The mixture of the machine learning approaches allows for variations of different data to be observed and analysed to improve the classifier and to predict the most accurate results.

This project demonstrates the viability of using machine learning approaches in order to solve numerous problems in everyday life. The classification model used for this research, was trained on two different datasets in order to study the difference of sentiments based on the specific domain. Furthermore, two feature models were investigated, unigrams and bigrams, for training the Support Vector Machines classifier, the Naïve Bayes classifier and the Maximum Entropy.

Additionally, this research has shown that in order to achieve a high accuracy of the classifier, the classifier has to be trained and tested on the same type of the dataset because the correlation exists between the performance and efficiency. Moreover, the investigation on sentiment analysis and machine learning improved the cuisine experience on different users.