

SENTIMENT ANALYSIS OF CUSTOMER SUPPORT MESSAGES

Bogdan A.O.

***National Technical University «Kharkiv polytechnic institute»,
Kharkiv***

Daily, companies generate enormous amounts of customer support tickets. Based on some characteristics, these tickets are grouped and placed in specialized queues from where the customer support personnel (CSP) resolve them on a first-come-first-out basis. This is however not optimal because all support tickets are created equal, but they require different levels of urgency. Based on business policy, several heuristics can be applied in prioritizing the support tickets. One of such heuristics is sentiments.

This study investigates the automation of sentiment polarity predictions on customer support tickets using machine learning methods and natural language techniques.

In total, four experiments are conducted in this study. All the experiments, except experiment 3, had three factors with each factor having three labels. The factors include feature selection methods, vector representation and ML algorithms. The labels were combined to create 27 treatments. Each treatment is applied to each of the corpora, which represents the subjects of the study.

Data preprocessing is identified as the most important stage in sentiment analysis. It helps to eliminate noise while reducing the dimension of the corpus. During the study, it was observed that removing stop words from the corpus did not yield any significant impact on the performance of the models hence ignored.

Stemming is identified to contribute to reducing the dimension of the corpus but leads to a loss in sentiment information. POS tagging and negation tagging presents a solution to the problem of sentiment information loss by preserving the context but contributes to increasing the dimension.

Compared to unigrams, bigrams and trigrams contribute more to context retention but can quickly grow the dimension of the corpus. This leads to the curse of dimensionality problem. Though a challenge, this is a good cost to bear in building sentiment analysis models as it leads to sentiment information gain which increases the performance.

In this study, LR and SVM have proved to perform better than NB. However, it would be relevant to investigate how deep learning models can be used to make sentiment predictions. Based on the natural selection method used by genetic algorithms in developing models, it would be interesting to study how they can be used to build a colony based on the underlying sentiment information in the corpus. This colony could be used to supplement the cluster ensemble method proposed.

Sentiment analysis models built from the movie and product review domain cannot be used to make sentiment predictions on support tickets. Also, when POS Tagging and negation tagging is applied to a given corpus, the sentiment information would be an amplification for existent data.