Introduction

1. As there are millions of applications available today that serves various purposes, therefore its components like sentiment analysis of the user comments have become an important study of research.

2. Sentiment Analysis helps in evaluating whether the content is negative, neutral, or positive.

3. The library used to perform the sentiment analysis is TextBlob python library and Watson Natural Language Understanding API. They help us in extracting the meta-data such as entities, contents, sentiments, emotions, part-of-speech tagging, classification, NP(noun phrase) extraction and so on.

4. The results of Watson NLU API consist of keywords, semantic roles, entities, emotions, relations, and categories.

5. Problems Faced

   1. Data collection from various websites as it had limited API calls, data collaboration into a single source(each website had a different format) and data cleaning.

   2. Learning the API and training the dataset. Identification of data idioms/sarcasm.

   3. Generating the rules and implementing the machine learning algorithms.

Methodology

Design Considerations

1. The preprocessing techniques: data collection, data cleaning, and document conversion. The corpus was created by collecting data from disparate sources such as helpshift.com, appfollow.io, and searchman.com. Corpus was created by training 20% of data and creating a dataset of negative and positive words.

2. For the scope of this project, I had conducted my experiments on five application i.e. Snapchat, Instagram, Whatsapp, IMO, Skype

Architecture

1. This Python library i.e. TextBlob is an API used to perform the natural language activities such as extracting the noun phrase, translation, POS tagging, sentiment analysis, tokenization, n-grams etc.

2. Polarity is in the range [-1.0 to 1.0] where scores declined towards -1.0 denotes negative comments and scores declined towards 1.0 denotes positive comments. The Subjectivity is in the range [0.0 to 1.1] where scores declined towards 1.1 denotes subjective sentence.

3. These results act as an input to the machine learning algorithms written in Java. These algorithms are set of rules then help the expert system evaluate the overall results.

4. It combines the average ranking and number of downloads to provide an overall summary of the application.

5. The Watson NLU API performs analysis on the data and the results are in the form of keywords, semantic roles, entities, emotions, relations, and categories.

6. These results not only help the users but it also helps the developers in various aspects.

Conclusion

1. Sentiment analysis/opinion mining is an active area of research that studies the sentiments, emotions, or attitude of people towards a particular entity.

2. The textBlob library and Watson NLU API does a great job of providing the sentiment results and the implemented model helps in agitating the results.

3. The final results provide a summary about the application which consists of the sentiments of the users, averaging rating, and a number of downloads.

4. The results of NLU helps in understanding the frequently used keywords, demonstrates the part of speech tagging, identifies entities, and emotions in form of joy, anger, sadness, fear, and disgust.

Future Work

1. Real time data collection from the mobile application instead of depending on the earlier collected data.

2. Creating an expert system that can consider all the features of an application such as security, audio, video and perform the entire analysis.

3. Implementing the sentiment model using other libraries such as CoreNLP, WordSegmentor, Tokenizer, SentiStrength and compare the efficiency of each library.

Contact details

Forum kapadia : fjk9481@rit.edu