Team Members:

Amr El-Gendy – 2013 0 48 26  
Mohamed Maher – 2013 0 14 63

Sentiment Analysis

CIE 553 – Natural Language Processing Course Project

Table of Contents:

|  |  |
| --- | --- |
| Application Abstract | P.2 |
| Group Members | P.2 |
| Techniques | P.3 |
| References | P.5 |

**Application Abstract:**

Sentiment Analysis is the process of identifying, and categorizing opinions regarding a specific topic in order to determine the people attitude towards this topic whether positive, negative or neutral. This process is used widely to help in determining the marketing strategy, improving customer service.

In this course project, we are going to use the techniques learnt during the NLP course in addition what we have found in literature review to implement a sentiment analysis tool to overview people’s opinions regarding any uploaded video on YouTube. This tool will be able to classify each comment in the video as a positive, negative or neutral opinion. Our application will target comment in English, Arabic, and Arabizi (Franko Arabic) languages. Moreover, the application will be implemented using python language and deployed on a cloud computing platform to be able to use it as a web application. So, the user will be able to just enter a YouTube video link then a summary of classified comments will be generated in a web page in text and graph format.

**Team Members:**

1. Amr El-Gendy: 2013 0 48 26
2. Mohamed Maher: 2013 0 14 63

**Application Techniques:**

In this application, we are going to use various techniques used in the NLP course in addition to ideas we have found in literature review.

Firstly, the gathering of data process will include connection with YouTube API in order to retrieve comments available in a certain video. Then, pre-processing steps will take place that will include various techniques like language detection, word stemming and lemmatization. After that, according to literature review found, two possible ways can be done in order to classify a comment:

1. Unsupervised Classification: Using a lexicon of positive and negative words will help us to determine a certain weight of the opinion.
2. Supervised Classification: which will use a previous labeled dataset of opinions and using different machine learning algorithms like Bayesian Classifier or Neural Networks, we will be able to classify opinions.

However, in some literature, we have that some trials have merged both techniques in their classification by many different ways and called there techniques as semi-supervised approach. We are going to implement one of those semi-supervised techniques that will ensure a high classification accuracy of the comments.

Finally, we are going to use python flask and Amazon web services to deploy the python code as a web application that can be used by different users.

Note: A further investigation still needed to see how we can deal with the Arabizi comments as the literature review showed us that this language is still a crucial problem that hasn’t achieved much progress in previous research. However, one of the solutions that we can implement is to use an API of any (Arabizi - Arabic) online translators like google translate to convert Arabizi comments to Arabic Language and deal with it in the same procedure as classification of Arabic comments.

The main reasonable and insightful ideas in this course project can be summarized in the following points:

* Using python to connect to a well-known API like YouTube in order to gather data needed for a NLP application
* Performing pre-processing steps on data obtained to prepare data for further processing
* Dealing with different languages and finding a solution to Arabizi Language
* Performing both supervised and unsupervised text classification in many NLP application process
* Learning about deploying of python scripts into a web application on cloud computing services like Amazon web services.

# References:

# [1] - Sentiment analysis in Arabic: A review of the literature, 2017, retrieved from: https://www.sciencedirect.com/science/article/pii/S2090447917300862

# [2] – Google Developers (YouTube API) documentation <https://developers.google.com/youtube/>

# [3] Nile University Arabic Lexicon Words for Sentiment Analysis. Retrieved From: https://github.com/NileTMRG/NileULex

# [4] Hu & Lie English Lexicon Words for Sentiment Analysis Retrieved From: https://github.com/woodrad/Twitter-SentimentMining/tree/master/Hu%20and%20Liu%20Sentiment%20Lexicon

# [5] Semi-supervised Multi-view Sentiment Analysis Retrieved From: <https://www.researchgate.net/publication/300348684_Semi-supervised_Multi-view_Sentiment_Analysis>

# [6] Sentiment analysis in Arabic tweets Retrieved From: <https://www.researchgate.net/publication/271550479_Sentiment_Analysis_in_Arabic_tweets>

# [7] Labeled dataset for tweets

# <https://raw.githubusercontent.com/AmrMehasseb/Emotional-Tone/master/Emotional-Tone-Dataset.csv>