# CS 735/835 Information Retrieval Project - Evaluation

Shayan Amani
(993550898, sa1149)

Soheil Gharatappeh
(933639024, sg1147)

## 1 Introduction

The problem to be solved is called *Affect in Tweets*, and it is defined as how we can determine the user's emotions according to their tweets to know what sort of feelings they have, and to what degree they have those feelings. Four emotions studied here are semantically distinct emotions; *anger, fear, joy, sadness*. Our task is; given a tweet, and an emotion X, we have to determine the intensity or degree of emotion X felt by the speaker. The data we have in our output is a value between 0 and 1, in which the maximum value of 1 stands for feeling the maximum amount of emotion X, and the minimum score of 0 stands for feeling the least amount of emotion X. This can be interpreted as, when the user is having a maximal/minimal mental state toward/away from emotion X.

In this project, for sake of simplification, we only consider one emotion; joy, mainly because we all are looking for it in this day and age. Hopefully, this research will help us to know the underlying meaning of the language we use everyday.

## 2 Motivation

Tweeter is one of the most famous social app that gives this opportunity to its users to communicate their ideas using text-based messages. These feature has made tweeter a unique platform, for doing all sorts of language/text-based analysis. In order to understand how important is this research, just think about that the president of the most powerful country in the world, uses tweeter on daily basis to convey his messages and ideas to the world. It is a crucial piece of information for political strategist, to have an understanding of the true feeling of the user about a particular issue, to come up with ideas in their negotiations to solve for example international disputes, or make deals between countries. The outcome of our analysis is not limited to only the world of politics. Knowing the users' emotion (at the moment of writing the tweet, or about a specific topic in general), can be useful for businesses and marketing strategists to better sell their products. By knowing about emotions of people towards something in a specific area, we can come up with a better business plan that works best for

that society. We can adjust advertisements produced for them to their tastes, and use elements that can better convince them to buy that product.

# 3 Method

To determine users emotion given a tweet, there can be two possible approaches; using a *vector space models*, and *Language Models*, or using *classification* methods. Regardless of the method we are going to implement, we need a document that somehow plays a reference role for happiness. If we look at this as the similarity perspective, the document shows the direction of *pure* happiness in our space. So, our job is to find the similarity of tweets to the reference document, to decide whether the document points to the same direction as the reference or not. In the 3.1, we discuss how we obtained our happiness reference document.

## 3.1 Happiness Reference

There are several ways to obtain a good reference document. First is to make a document using its synonyms. For doing that, we searched the keyword "happy" in the website http://www.thesaurus.com, and add all of more relative synonyms to our database. Next, we went to each of synonyms, and added the synonym for those words to our data base (duplicate word have not been removed in purpose).

The second possible approach is to boost up our database using another helper database. The second data base we are going to use is from the competition Multilingual Emoji Prediction [1]. In provided database in this competition, they associate each tweet to an emoji. What we want to do is to take tweets that are associated with happy faces emojis, namingly; smile, tears of joy, wink, and etc. and add these tweets to our database. The idea of improving our database using other tweets is mainly based on the fact that the language used in tweets are different than what we can find in dictionaries. So, if we only use dictionaries to obtain words related to happiness, we cannot guarantee that we have a reasonable data base that contains vocabularies used by people in their social life that indicates their happiness.

However, for the first part of the evaluation, we just go ahead with the first database.

## 3.2 Vector Space Model and Language Model

In the previous section, we discussed how we can obtain a document that shows the direction of happiness in space. The task here is to calculate the similarity between the happiness document, and a given tweet. We use different methods such as; lnc.ltc, bnn.cnn in vector space models to get Tf-idf scores, which shows how similar tweets are to the happiness document. In addition, we applied some language model methods such as; *Laplace, Jellinek, and dirichlet* Smoothing.

As it is noted in the competition website, the reported scores in the training and evaluation database has no meaning by itself. Therefore, it should be seen as a relative value between [0,1], considering the explanation given in 1. And, as we know, none of the above-mentioned method gives us back a value between [0,1]. So, what we need to do is to normalize them into the interval of [0,1], and do the evaluation on them. Score files can be found here. They are formated as follows:

```
id[tab]tweet[tab]emotion[tab]score
```

### 3.3 Classification

The second possible approach is to look at this problem as a classification problem. We want to classify a given tweet for the emotions involved. We start with training our classifier.

## 4 Evaluation

The idea of this project is taken from the competition published in [2]. To find more details about their research, please refer to [3]. The training and development sets are available online in the given website in 3 languages; English, Arabic and Spanish. The following is a sample line of the data set to measure anger in the given tweet:

```
10004 Don't join @BTCare they put the phone down on you, talk
  over you and are rude. Taking money out of my acc willynilly! #
  fuming anger 0.896
```

Two main evaluation measures used in the first phase of this study is *Pearson Correlation Coefficient*, and *Spearman*. Pearson Correlation is a measure of the linear correlation between two variables X and Y. It has a value between $+1$ and $-1$, where 1 is total positive linear correlation, 0 is no linear correlation, and $-1$ is total negative linear correlation. However, in the second phase, we use a modified version of Pearson Correlation and Spearman, which only considers tweets that obtain scores more than 0.5.

## 5 Expectation

In this research study, we are going to compare two different methods of Information Retrieval; vector space model and classification. As we know, these two are completely different approaches, one does not take any sort of feedback from the world, and only works on what we *assume* to be true about IR. However, the other one, only works based on user's feedback. As a first guess, I think the second method works better here. The reason I think so is the fact that in vector space model the effect of the original corpus we consider is high, and this effect will never be removed (or at least be decreased) in any ways. To be more clear, the way that the words in original document are arranged, and the words

the document contains highly affects the result of our algorithm. This effect would be minimum, if we choose a document that is large enough to cover all the combination for words.

As for the final product, we expect to make a tool that can run at the top of a social media (tweeter here), and analyze users' emotions. These values can be used in the future for further processing.

## References


[1] Semeval-2018 task 1: Affect in tweets. https://competitions.codalab.org/competitions/17344.

[2] Semeval-2018 task 1: Affect in tweets. https://competitions.codalab.org/competitions/17333#learn_the_details-overview.

[3] Mohammad Salameh Saif M. Mohammad, Felipe Bravo-Marquez and Svetlana Kiritchenko. Semeval-2018 task 1: Affect in tweets. *Proceedings of International Workshop on Semantic Evaluation (SemEval-2018*, 2018.