A Protocol for Temporal Analysis of Emotions in Tweets During AI Debate Between Yoshua Bengio and Gary Marcus

Peyman Ghasemi
Tehran University of Medical Sciences
Department of Health Management and Economics, Tehran, Iran
p.ghasemi94@gmail.com

Samira Abbasgholizadeh Rahimi
McGill University
Department of Family Medicine, Montreal, Canada
samira.rahimi@mcgill.ca

Emotion and sentiment analysis of the language is one of the leading methods in the evaluation of social interest over controversial topics. Twitter is a modern and widespread way for people to demonstrate their feelings and opinions about ongoing events around their world in short phrases. Therefore, it is possible to use the data sources of twitter for the detection of people's feelings about specific subjects along the time. In this study, we are going to analyze the tweets during an AI-related debate between Dr. Yoshua Bengio and Dr. Gary Marcus. The debate's time is December 23rd, 2019 6:30 pm - 8:00 pm.

The time-frame of the analysis will be one week before and one week after the debate (December 16-30). This time-frame is selected based on the feedback from the twitter activities around the subject and also with the follow-ups on the speakers' accounts.

All tweets with the hashtag *#AIDebate* will be collected via twitter developer API. The data for every tweet is including the sender's username, tweet's text, tweet's time and tweet's location. Other possible hashtags will be investigated after the debate manually

Every tweet will be checked to determine whether its language is English or not. This is possible with Google's pre-trained models for language detection. After this, the English tweet will be fed to a trained recurrent neural network model [1] to predict the emotion. This model works based on three emotion classification systems:

- Ekman's six basic emotions,
- Plutchik's eight basic emotions,
- Profile of Mood States (POMS) six mood states.

In this study, we will use Ekman's classification system [2] for the emotions, because its accuracy is better than other in the pre-trained RNN models (figure 1). Paul Ekman has defined a set of six universally recognizable basic emotions: **anger**, **disgust**, **fear**, **joy**, **sadness,** and **surprise**. The pre-trained RNN model will provide a probability value for every emotion described by Ekman.

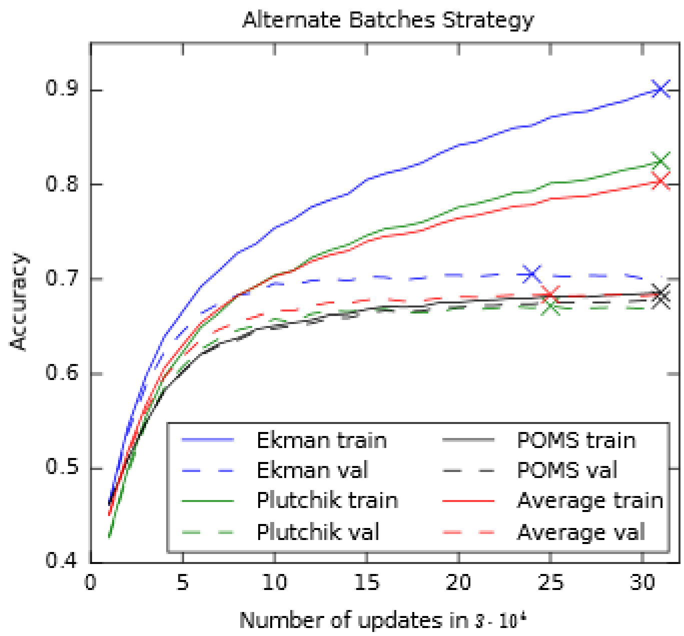

Figure 1. Accuracy of different emotion classification systems in the pre-trained RNN model

Furthermore, every tweet will be checked in terms of being positive or negative. Firstly, we will preprocess the texts in order to clean the URLs, emojis, stop words, etc. and keep the important parts, using the NLTK library. After this, we will use sentiment analysis methods provided by Pattern Library to measure the level of polarity in the text. The sentiment analysis method provided in this library is based on a lexicon dataset for English adjectives which has been labeled with a number between -1.0 to +1.0 in the aspect of polarity. -1.0 means the adjective has a very negative feeling (like *awful*), 0 is neutral (like *physical*), and the +1.0 means it is very positive (like *awesome*).

The algorithm for the sentiment analysis is essentially working based on a simple Naïve Bayes classifier exploiting a large dataset of positive/negative feeling probability related to each adjective in English (which is explained above). A naïve Bayes model for this problem can be described as [3]:

$$P(Text \ \&\& \ Classification\ n) = P(Text \mid Classification\ n)\ P(Classification\ n)$$

Assuming that our classes' number is 2 (positive vs. negative) and the possibility of their occurrence *P(Classification n)* is the same, we can model the problem only with knowing the *P(Text | Classification n)* which is achievable with a proper lookup table. Regarding these assumptions and methodology, It is possible to assign the most valuable probability of a word to the whole sentence, in order to have a simple sentiment analyzer.

Another analysis will be the number of repetition of the speaker's names in every tweet. This may measure the level of people's attention and attraction to every speaker. We will be able to assign the corresponding value for each tweet contained any of the speakers' names to evaluate the amount of that specific emotion for the speaker.

After applying the mentioned methods on every tweet, the summation value for the emotions, sentiments, and every speaker's name repetition will be plotted along the time and the location, and the dynamics of the emotions and feelings during the debate can be observable.

We can visualize the temporal values of the emotions and also sentiments in two ways:

- Diagram for every emotion and sentiment along the time
- Diagram of the emotions and sentiments in the tweets containing every speaker's name

We can also plot the frequency of tweets and emotions associated with them based on location to analyze the effect of location on people's feelings about the debate.

We will also provide a words cloud diagram for the whole collected tweets to show the most interesting words that audiences have used during the debate. The most frequent words may be an interesting keyword for further analysis (like temporal analysis of the word's repetition). The words cloud will be generated before and after the debate separately, so that we can investigate the effect of the debate in possible changes in audiences' minds and beliefs about the concepts like deep learning or symbolic AI.

After the creation of the words cloud, we will be able to select some most important words among them, in order to analyze their usage trends along the time (temporal analysis of the important words).

References

*1- Colneriĉ, Niko, and Janez Demsar. "**Emotion recognition on twitter: Comparative study and training a unison model**." IEEE Transactions on Affective Computing (2018).*

*2- Ekman, Paul. "**An argument for basic emotions**." Cognition & emotion 6.3-4 (1992): 169-200.*

*3- Falco, Xavier, et al. "**Sentiment and Objectivity Classification**." Stanford University, 2009, Natural Language Processing Group, available at: https://nlp.stanford.edu/courses/cs224n/2009/fp/24.pdf*