What do I mean with my emoji?

Emoji perception on Twitter

Computational Semantics, 2020/2021
Lucía Pitarch, Victor Šimić, David Moreno

1. Introduction

In recent years, different ways of communication have emerged. In fact, inside social networks such as Twitter, Facebook or Instagram, it is currently possible and really common to say a lot without saying a single word. This is achieved through stickers, gifs and the most common one, emojis. Our study is based on this last one, emojis, inside a really well known platform; Twitter. We aimed to analyze the usage and perception of emojis on Twitter, as it is obvious that emojis are not always used the way they were created for. For humans, most of the time it is easy to understand the meaning and the polarity of the emojis taking into account the context, or even knowing the country or the topic which is being discussed. We suspect it is not that easy for machines to understand the specific meaning of each single emoji, as its understanding and usage involves many extralinguistic and pragmatic factors. That is the reason why we found this event an interesting topic to research: the real perception of the emojis (by our computational models), instead of their original sentiment or their original meaning.

Getting a little bit deeper on the goal of our project, we wanted to analyze the usage and the perception of the emojis, as we believe that the perception is correlated with the usage of them. We approached the analysis through four different tasks: computation of the emoji embedding, prediction of an emoji given a particular text, emoji clustering and sentiment prediction comparing texts with and without emojis.

We used data that was already compiled for the sake of time. The first asset we used is a text file named "Emojitweets"[1]. Each line of the file contains a clean tweet (without hashtags, mentions or urls) and with, at least, one emoji. The other document we used (just for the sentiment analysis) is a file with the sentiment analysis already computed for 163K Tweets. The sentiment prediction on "TwitterSentiment"[2] file was done with TextBlob and the ratings given range from -1(negative sentiment), to 1(positive sentiment), being 0 neutral tweets.

2. Implementation

In order to go through our goals along the project, we set up different tasks to be able to properly analyze the emojis perception: The first one was to obtain the emojis' embeddings in order to work properly with them. Additionally, we implemented an emoji prediction method to verify that the embeddings were correctly created. Once the embeddings were a stable development inside the project, we implemented an Emoji Clustering taking into account the embeddings. Parallely, we developed a Sentiment Prediction system taking into account sentences with and without emojis.

2.1. Emoji embedding

As mentioned in the previous paragraph, the first implementation on our way to analyze the emojis was the emoji embedding. To make the embeddings of the emojis, we followed the next process:

1. We started by tokenizing the sentences from the 6 million tweets from our corpus with TweetTokens python package, due to the fact it also takes into account the emojis when making the tokenization, and eased our data treatment work.

2. Afterwards, we modelized the words and emojis that appeared in our corpus using Word2Vec, from gensim.models. That gave us the embeddings for each word and emoji that we had in the corpus. The way to access those embeddings, is by using the function model.wv.__getitem__(word). That would give us the embedding of that particular word in the model created with Word2Vec. However, if instead of a word we need to get the embedding of an emoji, if the emoji is already in the shape 'emoji' we just need to use the same function as before but using the emoji instead of the word. However, if the emoji is in the format of its name (for example 'thumbs_up') we first need to make the emoji of that, with the function emoji.emojize(':name_emoji:'). This change is needed because the corpus from twitter uses the emoji already built, not the names of the emojis.

2.2. Emoji prediction

After obtaining the emoji and word embeddings, we considered useful making a predictor of emojis given a sentence. The theory behind that was the same as making the similarity between two sentences, or more precisely, between a sentence and a word.

With that theory in mind, the implementation needed first a dictionary that held all the emojis with their embeddings to be able to compare them with the sentence given by the user. In order to build that dictionary, we went through all the tokenized tweets, and taking each emoji and their embedding (to identify an emoji, the function emoji.UNICODE_EMOJI resulted useful, as it has access to all the emojis). Afterwards, we made the cosine similarity between the emoji and the sentence given by the user, being the embedding of the sentence, the mean of the embeddings of its words.

However, we realized that, due to the fact that twitter often uses irony and sarcasm even with the emojis, sometimes the emoji that had the highest cosine similarity, was not the expected one from a logic and human point of view. Because of that, we decided to add a tolerance of 0.02 in the cosine, so if more than one emoji had the highest cosine (this time, taking into account the tolerance), all of the emojis were outputted as a prediction for the sentence.

On the figure below we can observe some examples of simple and multiple emoji prediction:

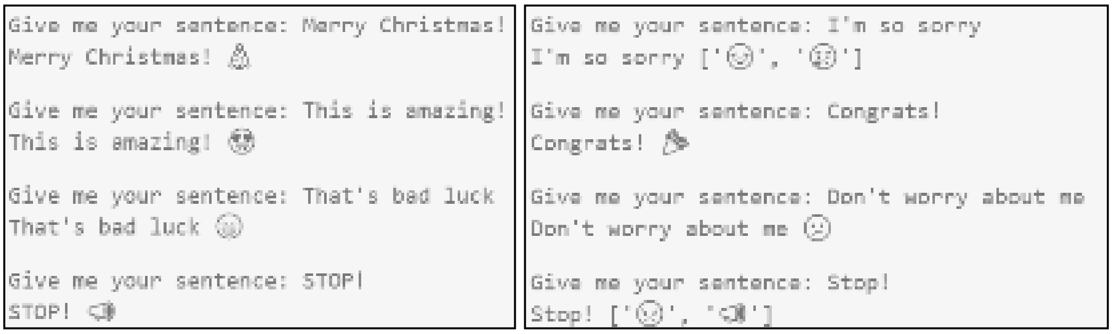

Figure 1: Outputs of emoji prediction

In the picture above is possible to see the example of irony when using emojis in the sentence 'That's bad luck'. Also, it is clear that the use of capital letters changes which emoji will be used. For example, the difference between using "stop" with all capital letters and the same word without that is that in the first case the predictor understands it as shouting, so the emoji goes according to that. However, in the second case, it also appears the emoji of the angry face.

2.3. Emoji clustering

Using the previously explained achievements, mainly the embeddings of the emojis in a twitter corpus, we found interesting and relevant to understand better how emojis are grouped by their use inside the 6 million twitter corpus we have used. This clustering could give some intuition on which emojis are related despite their original meaning without context may be far away. In this research, we have tried to approach the emoji clustering in two different ways. First we implemented it with 5 random centroids, which will result in 5 different clusters. The second one, we implemented it with 6 different clusters, and we tried to identify the 6 different Ekman's categories of facial emotion expressions.

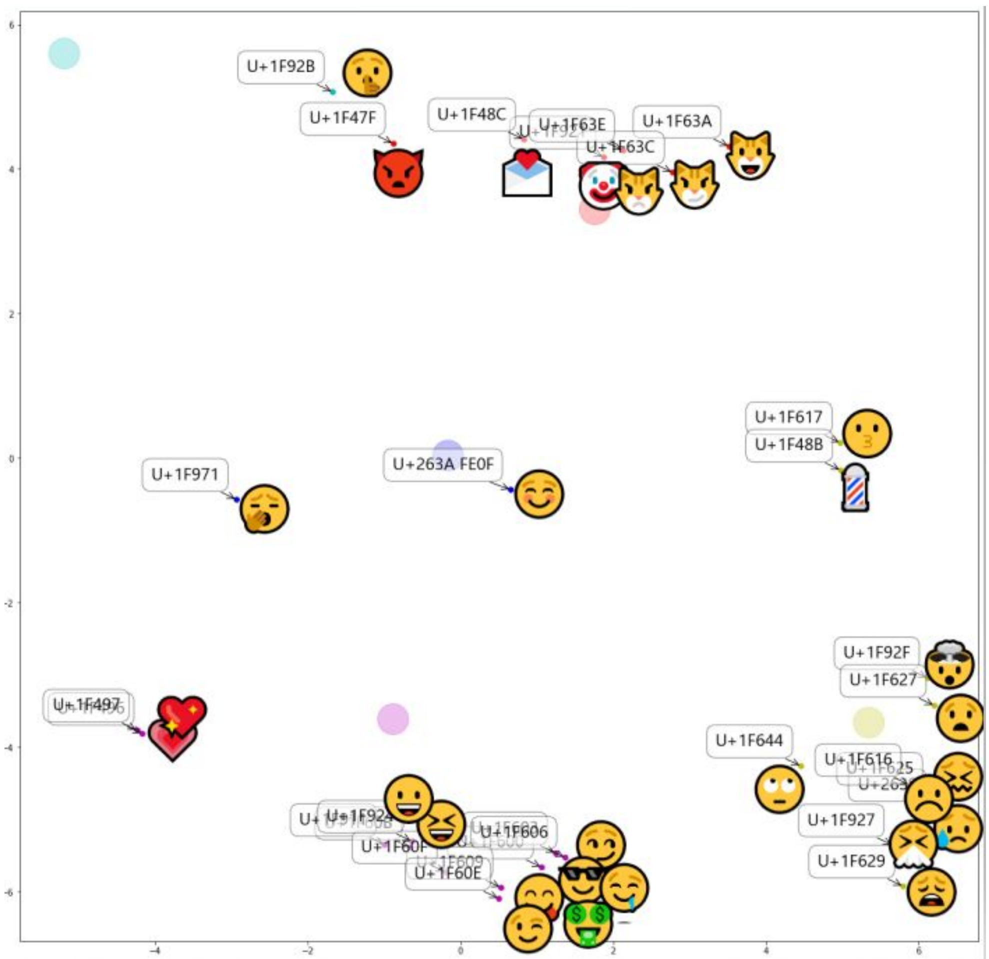

Figure 2: Emoji clustering with 5 random centroids

In the figure above we can observe how emojis are grouped according to the k-mean clustering algorithm. We randomly choose the emojis in order to avoid biases, as well as random centroids. The centroids are represented as bigger semi transparent dots, with the color of the corresponding cluster. A remarkable insight to make is that, as the centroids are randomly initialized, and they converge iteratively, the clustering algorithm provides a different configuration of the plots, even though the groups remain stable due to their embeddings

After watching the results of a 5 clustering model, we thought that it would be interesting to analyze it using a more formalized categorization. We decided to try to distinguish between the different categories for emotion expressions provided by Ekman. This intuition was based on the Guibon et al.[3], 2016, where Guibon et al. tried to classify all facial emojis on the different emotions. We wanted to tweak this approach a little bit, taking into account also the non-facial emojis, due to the belief that they can also represent emotions in nowadays society. The results can be observed in the figure below:

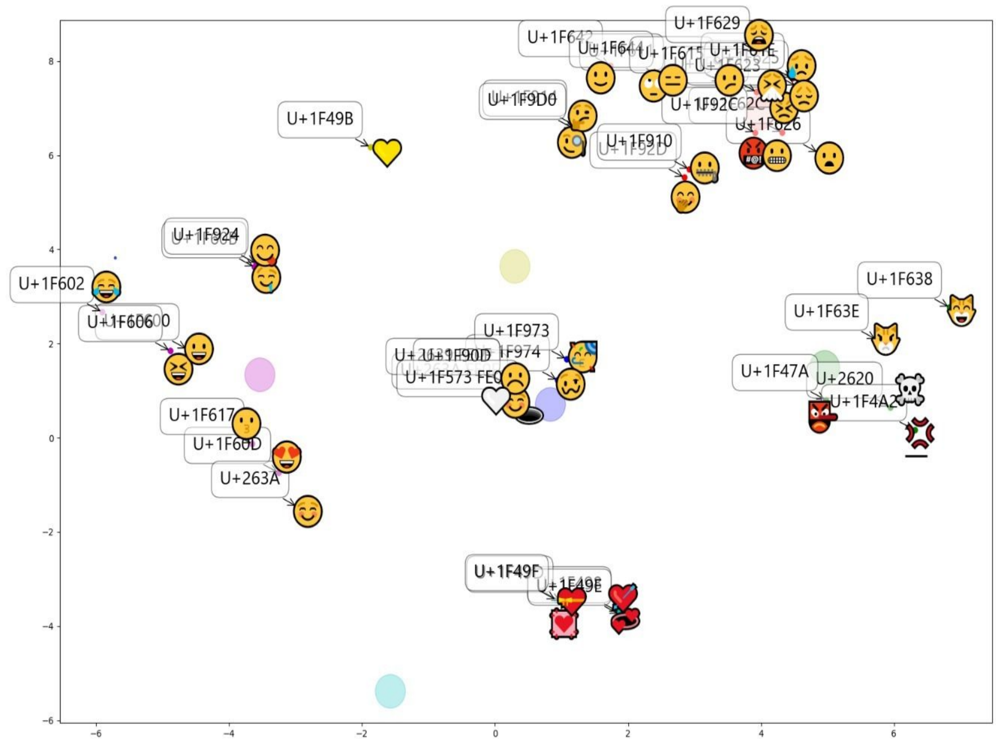

Figure 3: Emoji clustering with 6 random centroids

In this case, we can observe that we randomly sampled 40 emojis instead of 30. That was due to the fact that if we only sampled 30 random emojis, there was a cluster which was always empty. So, despite the difference, we believed that made more sense to have all clusters filled at least with one emoji.

It is relatively easy to see the 6 different groups in the cluster. On the left, with the magenta big spot as a centroid, we could consider the *happiness* emotions. In this case, the random sampling retrieved only facial emojis, despite we didn't set it as a requirement. On the bottom of the figure, and despite the centroid displacement, we have the hearts cluster. At this point we find a problem, as the remaining Ekman's categories are anger, disgust, fear, sadness, and surprise, and, perceptually, the heart cluster does not fit into any of these. In Guibon et al.[3], 2016, they used a more specific classification of such emotions, decoupling them a bit more, having 18 different labels. For instance, they used a specific label for excitement (love). Using this more complex but accurate labelling, our clustering figure makes much more sense, as we could express the heart cluster as excitement(love), the upper cluster as *sadness* and the right cluster as *angry*. The remaining two clusters, the yellow containing only a yellow heart, and the purple centroid which, unfortunately, varies quite a lot, as we can observe a sad emoji face, a happy/smiling face and also a partyish emoji.

These conclusions may lead us to believe that a classical emotion classification is not suitable for the use of emojis inside Twitter, as their use might be really ambiguous. Also, this clustering helps us to better understand the possible use of the emojis inside Twitter, understanding that, in some cases, the meaning or the polarity of such emojis represent the opposite in the majority of the cases, not just as an isolated case.

2.4. Sentiment prediction

The sentiment of a text has been already proven to vary through the use of emojis (Guibon et al.[3], 2016; Shiha and Ayvaz, 2017[4]). While Guibon et al. present more a state of the art study, Shiha and Ayvaz do, in fact, an emoji based sentiment analysis. Both of them agree, however, that there is still a lot to be studied.

Shiha and Ayvaz approach appears to be a good starting point, some parts seem problematic to us. In particular the way they focus their study on seeing whether emojis appear most in positive or negative events and how they assign emojis a sentiment depending on the sentence (with extracted emojis) they appeared in. We consider their approach might be problematic as the same emoji can be used with both positive and negative sentiments, and as Guibon et. Al. state emojis can be used for very heterogeneous purposes (sentiment enhancement, sentiment expression, sentiment modification...) this can also be seen in our cluster study where for example the smiling face could appear in two different clusters.

Other studies have made use of emoji analysis for sentiment prediction(Pak et al., Becker et al., Thelwall et al., Selmer et al. than are found in Guibon et al.[3]), however, they focus more on their number of appearances rather than on the meaning of emojis themselves or as we intend to do in our project how and in which contexts we use them. Thus we want to do a further analysis on it.

For the last part of our project by doing a sentiment prediction task we wanted to cover not only how emojis are selected and acquire meaning through the context they appear in but also the other way around: how emojis affect the context they appear in. To do so we first trained a Bag of Words and a TF-IDF model with 50000 Tweets with already predicted sentiments (from the TwittSentiment file). It is important to mention that for both models, as they take each word as a feature, the vectorizer function must be applied to both the train and the test (in our case, predicted) data.

Both models seemed to work fine and doing a manual revision of the predictions we realized they only differed on the Tweets that presented an ambiguous sentiment or meaning. Even us differed on how we would rate them.

Then we used both models to predict the sentiment of 5000 Tweets with emojis. At this point we needed to prove emojis were being recognized by the model and they were being analyzed as individual words, to do so we had to specify it in the Token Pattern (under the vectorizer function).

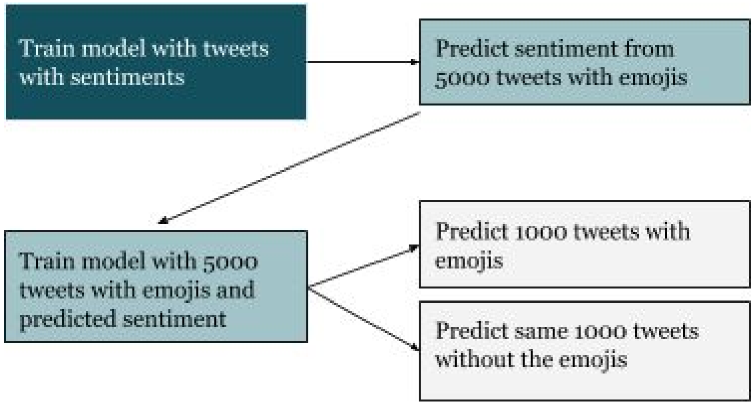

Figure 4: Pipeline of sentiment analysis and prediction

The next step was to train a new Bag of Words Model with the tweets with sentiment and emojis and then use it to predict two sets of tweets. Both tweet sets had the same words but in the first one we kept the emojis and in the second set we removed them. That way we could ensure that the only difference was the appearance of the emoji.

We can summarize sentiment prediction results in the following table:

Table 1: Sentiment prediction results. *Bag of Words, TF-IDF

	Positive	**Negative**	**Neutral**
Tweets with emojis	221 289	38 36	741 675
Tweets without emojis	272 309	36 32	692 659

Looking at these results it seems like emojis tend to neutralize sentiment prediction. An explanation for this might be that users in Twitter usually produce ironic and sharp messages that could easily be misinterpreted, therefore emojis could be used to soften their words or at least, warn irony is being used. Very similar results are found when using both the Bag of Words model and the TF-IDF model. Emojis seem to 'neutralize' the sentiment of the sentence.

3. Conclusion & Future work

Throughout our project two things were proven: emoji perception is affected by the context they appear in and also emojis affect the context surrounding them.
Besides, as emojis are overall described as ambiguous and as we have already seen in this short research project, an interesting further work would be to test the same kind of research on smaller areas and for different generations in order to observe usage differences across people from different social, cultural and age groups. Although we supposed that the emoji changes frequently in shape and meaning over time, we consider that a diachronic study would be interesting. In fact, for further work we recommend taking datasets for all of the tests we realized that cover the exact same time range. Finally, the results of these work would be more solid when contrasted with human judged results.

4. References

[1] Larionov, D. (2019, April 26). *EmojifyData-EN: English tweets, with emojis*. Retrieved December 17, 2020 from https://www.kaggle.com/rexhaif/emojifydata-en

[2] A, C. (2019, November 28). *Twitter and Reddit Sentimental analysis Dataset*. Retrieved December 17, 2020, from https://www.kaggle.com/cosmos98/twitter-and-reddit-sentimental-analysis-dataset?select=Twitter_Data.csv

[3] Guibon, G., Ochs, M., Bellot, P. (2016, June)"From Emojis to Sentiment Analysis". *WACAI 2016, Lab-STICC; ENIB; LITIS*, Brest, France.

[4] Shiha, M. O., and Ayvaz, S. (2017, May)"The Effects of Emoji in Sentiment Analysis," *International Journal of Computer and Electrical Engineering* vol. 9, no. 1, pp. 360-369.

[5] Miller, H., Thebault-Spieker, J., Chang, S., Johnson, I., Terveen, L., & Hecht, B. (2016). Blissfully happy" or "ready to fight": Varying Interpretations of Emoji. *Proceedings of ICWSM, 2016*.