William Zhang

7th April, 2020

The Data Story Contextual Progress Report

Emoji, developed from the 1990s, has become a complete and well-supported system today. We use emoji, sometimes to support our texts and express emotions, and sometimes, we spam emoji just to show the strong feeling within us. As emoji are so widely used, especially on social media, my data story project – “Emojified Social Media – how are emoji used on Twitter?” would focus on the use of emoji on Twitter, analyze specific texts with respect to its sentiment and emoji use.

So, firstly, how is emoji developed? People have long been expecting a set of typographical signs for emotions since the 1960s, and it is until the 1980s when the first :-) and :-( is invented. Emoji, however, didn’t appear until the late the 1990s in Japan. That year, a designer called Shigetaka Kurita (who is deemed as the inventor of Emoji) created a set of emoji for one of Japan’s telecom operators and later emoji went viral in Japan, and that explains why the emoji we have today includes so many Japanese culture, such as sushi (🍣), ramen (🍜) and Japanese Dolls (🎎). After that, people saw the potential in emoji and its use spread to the world.

In October 2010, The Universal Coded Character Set (Unicode), released Unicode 6.0.1 It is the first version of the Unicode Standard to support emoji. And since this release, the use of emoji has been unified and every year, there would be proposals to add new emoji to the whole set and Unicode would make decisions on which of these proposals go to the final yearly release. Still, companies would design their own emoji, for example, the mobile phone emoji on OS X and iOS is represented with an iPhone icon; and emoji on Twitter, Google, Apple or Microsoft are all slightly different in style.

Today, emoji are more than icons, they could represent the global trend of ideas and thoughts. We could notice that currently, emoji include mechanical limbs (🦾 & 🦿), all occupations of both genders, all choices of sexual orientation or gender cognition (💑 & 👨‍❤️‍👨 & 👩‍❤️‍👩, or 👯‍♀️ & 👯‍♂️ (Women or Men with Bunny Ears). And mosquitos (🦟) are added to the emoji to raise awareness for diseases spread by the insect, such as dengue and malaria.2

Furthermore, there are controversial emoji. If we notice, back in 2015, if we were to type the ‘gun’ emoji, it will be a revolver, but right now, ‘gun’ of all platforms has changed into a water pistol (🔫). Multiple arrests and imprisonments have followed usage of pistol (🔫), knife (🗡), and bomb (💣) emoji in ways that were deemed by authorities to constitute credible threats.3

Emoji are often misunderstood. This misunderstanding relates to the cultural or contextual interpretation of the emoji – when the author picks an emoji, they think about it in a certain way, but the same character may not trigger the same thoughts in the mind of the receiver. And in China, we have developed a new way of using emoji – a way of sarcasm: a smiley face (🙂) could be sent to convey a despising, mocking, and even obnoxious attitude.

As statistics show, half a billion tweets are sent out each day4, and the emoji – Face with Tears of Joy (😂), which is chosen as the Word of the Year 2015 by the Oxford Dictionaries – is used over 2.7 billion times.5 So, I deem that the data from Twitter could represent, at least some proportion of social media users, and I want to see how emoji are used on Twitter.

So, the data I found is a raw data set of 18 million unprocessed English tweets with at least one emojis, as well as a file about the how these emojis are used in Twitter: how many are used in positive, neutral and negative tweets, and base on that, I could calculate the sentiment score for each emoji – if it is used more in negative tweets, then it is more negative than positive.

So, what I planned to do with the 18 million raw tweets is to process it and separate the word content with emojis, I would use the Natural Language Token and VADER (Valence Aware Dictionary and sEntiment Reasoner) library of Python to give the text and emoji sentiment scores separately and see how emoji is contributing to the emotion of the tweets, whether they conform with each other or are emoji expressing more than the text, since some may imagine we are becoming more emotional with the aid of emoji online.

Also, I’d like to show how one emoji is used with other, for example, an eagle and a US national flag (🦅 & 🇺🇸) will be an expected combination, and what could be the unexpected ones? Moreover, I would process the data to show which emoji is used several times in one text, like 😂😂😂😂 should be a typical use. The above features are commonly used in our social media life, but we seldom notice them.

Aside from that, I would like to show the emotions and sentiment behind the emoji, and create an interactive interface to get the user see the difference of how they perceive emoji are and how they are really used in reality as well as how they are originally defined.

In general, I would like this project to be a window to how we use emoji, not necessary when or where we are using them, but the emotional context behind the use of emoji. And after viewing this project, we could reflect on our use of the emoji and maybe, change some of our ways of using them.

Data Sources:

1. EmojifyData-EN: English tweets, with emojis – 18 million English tweets, all with emoticons included, <https://www.kaggle.com/rexhaif/emojifydata-en>.
2. Emoji Sentiment Ranking v1.0, <http://kt.ijs.si/data/Emoji_sentiment_ranking/index.html>.
3. Full Emoji Database – Emoji names, groups, sub-groups, and codepoints, <https://www.kaggle.com/eliasdabbas/emoji-data-descriptions-codepoints>.
4. How To Perform Sentiment Analysis in Python 3 Using the Natural Language Toolkit (NLTK), <https://www.digitalocean.com/community/tutorials/how-to-perform-sentiment-analysis-in-python-3-using-the-natural-language-toolkit-nltk>.
5. VADER-Sentiment-Analysis, <https://github.com/cjhutto/vaderSentiment>.

Sources:

1. “Unicode Version 6.0”, Retrieved from <https://emojipedia.org/unicode-6.0/>.
2. Desmon, Stephanie. "Creating Buzz: Proposing a Mosquito Emoji for Public Health". Johns Hopkins Center for Communication Programs. Published September 18, 2017.
3. Kelly, Heather. "Apple replaces the pistol emoji with a water gun". CNN Tech. Published August 2, 2017.
4. “10 Twitter Statistics Every Marketer Should Know in 2020 [Infographic]”, Oberlo, <https://www.oberlo.com/blog/twitter-statistics>.
5. “Emojitracker: realtime emoji use on twitter”, <http://emojitracker.com/>.