#mood: The Emoji Context Project

Caroline Locke, Ryan Flint, Chris Muncey, Bryceton Bible, and Manny Abhidya

Abstract— This data analysis project uses Twitter posts to extract connotative meanings of emojis based on the context in which they are used.

I. OBJECTIVE

#Mood: The Emoji Project seeks to find the underlying meanings of emojis based on the words that typically accompany them. Most emojis have a face value that can be interpreted by just viewing the image — for example, the smiley face represents a smile or happiness; however, language mutates rapidly on social media and so the underlying meanings and usage of these emojis changes with current trends and slang. Sometimes, one must be engrossed in an internet subculture or social platform to be able to fully understand the true meaning of a message and the emoji used because of the layers of subtext that have been created in these bubbles. This project seeks to learn if the underlying meanings of emojis can be extracted purely from analyzing the surrounding text without being in the social media bubble in which the meaning is being created. By collecting words that commonly appear with an emoji and assigning a mood word or connotation to it, it may be possible to interpret these underlying lying meanings just from the textual evidence. This project will determine if this is a viable way to understand the ways in which users enhance their communications with emojis.

II. MOTIVATION

In today's world, a large proportion of communication takes place electronically — via social media, text messages, email, forums, websites, etc. Although emojis aren't typically used in professional or academic electronic communications, they are nearly ubiquitous in casual settings. From the dawn of the internet, users have created ASCII-based faces and caricatures meant to communicate a feeling or idea i.e. the smiley [:)], the surprise face [:O], etc. Overtime, these ASCII figures have evolved into "emojis" — small pictures representing everything from face emotions, to animals, to blood types. They have become integrated into the way we communicate with our peers and are available for use in most texting and communication apps. Text based communication is rife with misinterpretations and miscommunications due to lack of tone or context, but emojis help add back the emotion that is lost in translation. Because of the vast amount of emojis (with more being added all the time), users are finding new and interesting ways to use these pictures to emphasize their words, clarify meanings and enhance their words. Because of social media, emojis can take on brand new meanings with elaborate layers of context that requires the reader to be "in the know" in specific ways to fully understand the meaning of both the emoji and the contextual words. Many emojis often have multiple contextual meanings alongside their face value meaning. This project attempts to unearth these hidden meanings through textual analysis of the words accompanying these emojis to try to understand the mood or connotation being imbued by the users.

III. METHODS

This section discusses the methods in which data will be obtained, how it will be analyzed, responsibilities of each member, and a list of milestones and dates for the project.

A. Data Collection

The data to be analyzed in this project will be Tweets scraped from the social media site Twitter. Twitter is one of the major social media sites and a hub of emoji usage. Many viral memes, videos, quotes, and jokes come from this site. With over 300 million active users, this source provides a large sampling population. Twitter also has an easy to use API for scraping the data. To collect the data, we will create a Python script that searches Twitter using the API connection and collects any English language tweet that includes target emojis. These will be stored in a file that will later be parsed by emoji. Because of limitations placed on scraping to prevent overwhelming the twitter servers, the scraping scripts will have to run for a fair amount of time to collect meaningful data. We plan to have multiple scrapers running simultaneously, both looking for different emojis. Each scraping session will run for at least 8 hours. This may change depending on the volume of data we are able to scrape during this time. Because there are currently over 3000 emojis, we will use only the top 20 most frequently used emojis in our data analysis. There are several websites such as "EmojiFrequency" that track the most used emojis in real time. We will use this filtered list to choose the emojis we analyze to get the most meaningful data.

B. Data Analysis

Once all the data is obtained, we will separate the tweets and organize them by emoji. If multiple emojis appear in the tweet, the tweet will be included in the lists for all emojis indicated. We will extract all the words that appear alongside each emoji, remove any stop words (like 'of', 'but', 'and', and any slurs) and store them in a list along with the number of times that word has appeared. If any common misspellings emerge, they will be counted as the correct spelling of the word. We will then analyze the lists to determine the most common words associated with each

emoji. After we collect, parse and display the data, we will take the top 5 unique words associated with each emoji (once the by words are removed and accounted for). The words will likely be displayed in a word cloud or graph to denote the frequency. Once this data is extracted, we will compare the word lists to common connotative and mood word lists. We will use the mood words to assign meaning or feeling to each emoji. This will allow us to assign moods or meanings to each emoji. If multiple moods/meanings emerge, these will be noted. We will then compile a document detailing the intermediate results and the final results of the analysis along with the assigned moods for our chose emojis. This will include frequency charts for each emoji and the final assigned mood/meaning.

C. Responsibilities

All members will be responsible for project design including selecting emojis to scrape for, final assignment of moods and meanings, and evaluation of resource as well as some coding and parsing of the script. Each member will also be tasked with specific responsibilities as described below:

- Caroline Locke: project management (including documentation, timeline management, and planning), final data analysis
- Bryceton Bible: obtaining resources for emojis and mood/connotative word lists, parsing resultant data
- Ryan Flint: twitter scraping, API maintenance, scripting
- Chris Muncey: twitter scraping, API maintenance, stop word list creation
- Manny Bhidya: parsing resultant data, displaying the associated words data meaningfully via graphs/visuals

D. Timeline

Below is a timeline for the completion of this project:
10/10: Scraping bot completion
10/13: Emojis chosen for bot
10/14- 10/31: Collect tweets via scraping for all of the emojis.
10/20: Mood/connotation word sources selected
11/1: Begin scripting to parse out words for each emoji
11/11: Finish parsing and displaying data via Python
11/15: Begin assigning moods/connotations from collected data
11/30: Document findings and complete final report

IV. EXPECTED OUTCOME

For this project, we hope to find that we are able to interpret the meaning of the emoji in context based on the words that appear alongside them. We expect to find specific words that are more commonly associated with specific emojis. Although there will likely be some overlap, we expect some common emojis to have very different connotations emerge. We expect to see some emojis to have different moods or connotations than the "face value" interpretation. We also expect some emojis to have two distinct connotations or moods.