Varying linguistic purposes of emoji in (Twitter) context

Noa Na'aman, Hannah Provenza, Orion Montoya
Brandeis University
{nnaaman,hprovenza,obm}@brandeis.edu

Abstract

Research into emoji in textual communication has, thus far, focused on high-frequency usages and the ambiguity of interpretations. Investigation of emoji uses across a wide range of uses can divide them into different linguistic functions: function and content words, or multimodal affective markers. We report on an annotation task on English Twitter data with the goal of classifying emoji usage by these categories, and on the effectiveness of a classifier trained on these annotations. We find that PART OF IT IS STRAIGHTFORWARD, but ANOTHER PART OF IT IS COMPLICATED.

1 Background

Early work on Twitter emoticons (Schnoebelen, 2012) pre-dated the wide spread of Unicode emoji on mobile and desktop devices. Schnoebelen studied the Recent work (Miller et al., 2016) has explored the

2 Annotation task

2.1 Data collection and filtering

Tweets were pulled from the public Twitter streaming API using the `tweepy` library. The collected tweets were automatically filtered to include only tweets with characters from the Emoji Unicode ranges (i.e. generally U+1FXXX, U+26XX–U+27BF); only tweets labeled as being in English; to exclude tweets with embedded images or links (more below).

Tweets with links and images were excluded from consideration to reduce time investment and cognitive load for annotators. Our early explorations found frequent cases where emoji were tweeted to show a reaction to an attached image or linked page (especially a blog post or news story) and that these tended toward ambiguous interpretations akin to those found by Miller et al. A given tweet's U+1F62D 'loudly crying face' might be showing true sympathy, or sarcastically saying "cry my a river." The amount of annotator effort necessary for an annotator to determine this in context would require an understanding of the tweeter's past opinions and their stance on the parties involved in the story they linked. These are very interesting questions for future research, but we determined them to be out of scope for the present research, which focuses on high-level, coarse-grained distinctions.

Redundant/duplicate tweets were filtered by comparing tweet texts after removal of hashtags and @mentions; this left only a small number of cloned duplicates.

References

Hannah Miller, Jacob Thebault-Spieker, Shuo Chang, Isaac Johnson, Loren Terveen, and Brent Hecht. 2016. "Blissfully happy" or "ready to fight": Varying interpretations of emoji. In *Proceedings of the Tenth International Conference on Web and Social Media, ICWSM 2016, Cologne, Germany, May 17–20, 2016*. Association for the Advancement of Artificial Intelligence, May.

Tyler Schnoebelen. 2012. Do you smile with your nose? Stylistic variation in Twitter emoticons. In *University of Pennsylvania Working Papers in Linguistics*, volume 18, pages 117–125. University of Pennsylvania.