Emoji as Emotion Tags for Tweets

Ian D. Wood, Sebastian Ruder

Insight Centre for Data Analytics; Aylien Ltd.

National University of Ireland, Galway; Dublin, Ireland

`firstname.lastname@insight-centre.org`

Abstract

In many natural language processing tasks, supervised machine learning approaches have proved most effective, and substantial effort has been made into collecting and annotating corpora for building such models. Emotion detection from text is no exception; however, research in this area is in its relative infancy, and few emotion annotated corpora exist to date. A further issue regarding the development of emotion annotated corpora is the difficulty of the annotation task and resulting inconsistencies in human annotations. One approach to address these problems is to use self-annotated data, using explicit indications of emotions included by the author of the data in question. We present a study of the use of unicode emoji as self-annotation of a Twitter user's emotional state. Emoji are found to be used far more extensively than hash tags and we argue that they present a more faithful representation of a user's emotional state. A substantial set of tweets containing emotion indicative emoji are collected and annotated for emotions. The accuracy and utility of emoji as emotion labels are evaluated directly (with respect to annotations) and through trained statistical models. Results are cautiously optimistic and suggest further study of emotji usage.

Keywords: Twitter, hash tags, emotion annotation, emotion detection, emoji, emoticons

1. Previous Work

Purver and Battersby (2012a) also use distant supervision labels for detecting Ekman's six emotions in Twitter, in their case hashtags and emoticons. They conduct several experiments to assess the quality of classifiers to identify and discriminate between different emotions. A survey reveals that emoticons associated with anger, surprise, and disgust are ambiguous. Generally, they find that emoticons are unreliable labels for most emotions besides happiness and sadness. In another study, Suttles and Ide (2013) examine hashtags, emoticons, as well as emoji as distantly supervised labels to detect Plutchik's eight emotions, constructing a binary classifier for each pair of polar opposites. In order to create a multi-way classifier, they require four additional neutral binary classifiers. This approach has had success with using text emoticons and selected hash tags for sentiment annotation (Davidov et al., 2010) and emotion-specific hash tags for emotion annotation (Mohammad, 2012; Mohammad and Kiritchenko, 2015).

2. Emotion Expression in Text-only Communication

Facial expressions, voice inflection and body stance are all significant communicators of emotion (Johnston et al., 2015). Indeed, research into emotion detection from video and voice has found that arousal (the level of excitement or activation associated with an emotional experience) is difficult to detect in text transcripts, implying that those aspects are not strongly expressed in text. One might think, therefore, that text-only communication would be emotion-poor, containing less expression of emotion than face-to-face or vocal communication.

Research into text-only communication has, however, found that people find ways to communicate emotion, despite the lack of face, voice and body stance, and that text-only communication is no less rich in emotional content than face-to-face communication (Derks et al., 2008). Other research has found that text emoticons (text sequences that indicate facial expressions, such as *(-:*) produce similar brain responses to faces (Churches et al., 2014), and it is not unreasonable to expect that facial expression emoji (unicode characters whose glyphs are small images, such as 😄) function similarly.

In recent years, marketing researchers claim to have observed significant and continuing increases in the use of emoji in online media (emogi.com, 2015). This increase was not constrained to young internet users, but across all ages. Facial expression emoji have become a common method for emotion communication in online social media that appears to have wide usage across many social contexts, and are thus excellent candidates for the detection of emotions and author-specified labelling of text data.

3. Collecting Emoji Tweets

We selected a number of commonly used emoji[1] with clear emotional content as emotion indicators and collected tweets that contained at least one of the selected emoji. We used Ekman's emotion classification of six basic emotions (Ekman, 1992) for our experiments. Another common scheme for categorical emotion classification was presented by Plutchik (1980) and includes two extra basic emotions, trust and anticipation. However, there were no emoji we considered clearly indicative of these emotions, which is in line with previous research found (Suttles and Ide, 2013), which found that emoji thought to be associated with anticipation are few and unreliable emotion indicators. The selected emoji and their corresponding Unicode code points are displayed in Tables 1 and 2 respectively.

3.1. Challenges

There are a few choices and difficulties in selecting these emoji that should be noted. First, it was difficult to identify emoji that clearly indicated disgust. An emoji image with green vomit has been used in some places, including Facebook; however this is not part of the Unicode official emoji set (though is slated for release in 2016) and does not currently appear in Twitter.

---

[1] as indicated by `http://emojitracker.com/`

joy	😀😁😃😄😆😇😉😊😌😍😎😏🌞☺️😘😚😛 😜😝😺😸😻❤️💖💕😁❤️
anger	😬😠😐😑😡😤😣😫👊
disgust	💩
fear	😅😦😧😱😨😰😮
sad	😌😕☹️😫😩😢😥😪😭😿💔
surprise	😳😯😵😲

Table 1: Selected emoji for Ekman's six emotions

joy	U+1F600, U+1F602, U+1F603, U+1F604, U+1F606, U+1F607, U+1F609, U+1F60A
	U+1F60B, U+1F60C, U+1F60D, U+1F60E, U+1F60F, U+1F31E, U+263A, U+1F618
	U+1F61C, U+1F61D, U+1F61B, U+1F63A, U+1F638, U+1F639, U+1F63B, U+1F63C
	U+2764, U+1F496, U+1F495, U+1F601, U+2665
anger	U+1F62C, U+1F620, U+1F610, U+1F611, U+1F620, U+1F621, U+1F616, U+1F624
	U+1F63E
disgust	U+1F4A9
fear	U+1F605, U+1F626, U+1F627, U+1F631, U+1F628, U+1F630, U+1F640
sad	U+1F614, U+1F615, U+2639, U+1F62B, U+1F629, U+1F622, U+1F625, U+1F62A
	U+1F613, U+1F62D, U+1F63F, U+1F494
surprise	U+1F633, U+1F62F, U+1F635, U+1F632

Table 2: Selected emoji Unicode code points

The second difficulty concerns the interpretation and popular usage of emoji: All emoji have an intended interpretation (indicated by their description in the official unicode list). However it is not guaranteed that their popular usage aligns with this prescription. The choices made in this study were intended as a proof of concept, drawing on the personal experiences of a small group of people. Though these choices are likely to be, on the whole, reasonably accurate, a more thorough analysis of emoji usage through the analysis of associated words and contexts is in order.

3.2. Data collection

The "sample" endpoint of the Twitter public streaming API was used to collect tweets. This endpoint provides a random sample of 1-2% of tweets produced in Twitter. Tweets containing at least one of the selected emoji were retained. Twitter has a streaming endpoint, which can be queried with phrases (the "filter" endpoint) and provides only only tweets that mach the search criteria. This endpoint has two disadvantages that are relevant to the gathering of emotionally-indicative tweets: a) You can search only on entire (whitespace-delimited) words, and not on individual unicode glyphs; thus only emoticons surrounded by whitespace are retrieved; b) the volume of data for an emoji search phrase build from the above list is very large.

A substantial data set could be collected in a single day; however this would be subject to biases resulting from the particular trending topics during that day. Collecting a smaller proportion of tweets over a longer period mitigates this problem. To illustrate the magnitude of the trending topic problem in our initial experiment using the "filter" endpoint, we note that the most common hash tag (in 35,000 tweets) was "*#mrandmrssotto*", which relates to a prominent wedding in the US Filipino community.

We also considered a set of emotion-related hash tags (similar to (Mohammad, 2012)). However, we found that the number of such tweets was orders of magnitude lower than tweets with our emotion emoji. This fact combined with evidence from psycholinguistic research connecting emoji to emotion expression (see Section 2.) forms our primary motivation to focus on emoji in the context of this study.

3.3. Data Summary

We collected a total of over half a million tweets over a period of two weeks, of which 190,591 were tagged by Twitter as English. We show the tweet counts for the top 15 languages in Table 3. Note that all the counts provided do not include retweets as these are considered to bias the natural distribution of word frequencies due to the power-law distribution of retweet frequencies and the fact that a retweet contains verbatim text from the original tweet.

4. Evaluation

We carry out two forms of evaluation: a) In Section 4.1., we evaluate the quality of the chosen emoji as emotion indicators; b) in Section 4.2., we evaluate the quality of classifiers trained using emoji-labeled data.

4.1. Evaluation of emojis

For the first evaluation, we selected a random subset of 360 of the collected English tweets. For these, we removed emotion-indicative emoji and created an annotation task. The guidelines of the task ask the the annotator to annotate all emotions expressed in the text.

In past research using crowd-sourcing, a tweet is usually annotated by three annotators. As emotion annotation is notoriously ambiguous, we increased the number of annotators. In total, 17 annotators annotated between 60 and 360 of the provided tweets, providing us with a large sample of different annotations.

For calculating inter-annotator agreement, we use Fleiss' kappa. We weight each annotation with $6/n_{ij}$ where n_{ij} is the number of emotions annotated by annotator i for tweet j in order to prevent a bias towards annotators that favor multiple emotions. This yields κ of 0.51, which signifies moderate agreement, a value in line with previous reported research.

4.1.1. PMI

To gain an understanding of the correlation, between emotions and emoji, we calculate PMI scores between emoji and emotions. We first calculate PMI scores between emoji and the emotion chosen by most annotators per tweet (scores are similar for all emotions on which a majority agreed), which we show in Table 4.

Note that among all emoji, emotions are correlated most strongly with their corresponding emoji. Anger and – to a lesser degree – surprise emoji are also correlated with disgust, while we observe a high correlation between sadness emoji and fear. Additionally, some emoji that we have associated with sadness and fear seem to be somewhat ambiguous, showcasing a slight correlation with joy. This can be due to two reasons: a) Some fear and sad emoji can be equally used to express joy; b) some tweets containing these emoji are ambiguous without context and can be attributed to both joy and fear or sadness.

Language	Total	Joy	Sadness	Anger	Fear	Surprise	Disgust
en	190,591	136,623	36,797	7,658	6,060	2,943	510
ja	99,032	68,215	17,397	4,595	4,585	3,631	609
es	65,281	45,809	11,773	3,877	2,532	1,176	114
UNK	56,597	42,535	9,217	1,959	1,624	1,033	229
ar	44,026	29,976	11,216	1,114	1,084	5,72	64
pt	29,259	21,987	4,894	1,208	8,89	233	48
tl	20,438	14,721	4,096	752	656	176	37
in	18,910	13,578	3,175	1,018	738	323	78
fr	13,848	10,567	1,821	651	572	213	24
tr	8,644	6,935	773	419	305	201	11
ko	7,242	5,980	916	142	113	87	4
ru	5,484	4,024	646	411	317	74	12
it	4,086	3,391	376	156	119	34	10

Table 3: Number of collected tweets per emoji for the top 15 languages (displayed with their ISO 639-1 codes). UNK: unknown language.

	Joy	Dis.	Sur.	Fear	Sad.	Ang.	Ø
Joy	**.40**	-.53	.08	-.59	-.59	-.62	-.12
Dis.	.01	**.33**	-.11	-.02	-.24	-.27	.17
Sur.	-.49	.31	**.64**	-1.00	-.03	-.29	.15
Fear	.12	-.16	-.12	**.66**	-.14	-.07	-.03
Sad.	.11	-.68	-.58	**.76**	.66	-.37	-.69
Ang.	-.58	.71	-.22	-.13	-.35	**.87**	.06

Table 4: PMI scores between emojis and emotions chosen by most annotators per tweet. Emoji ↓, emotion →. Ø: No emotion.

Calculating PMI scores not only between emojis and those emotions, which have been selected by the most annotators for each tweet, but all selected emotions produces a slightly different picture, which we show in Table 5.

	Joy	Dis.	Sur.	Fear	Sad.	Ang.	Ø
Joy	**.32**	-.35	.04	-.24	-.56	-.46	-.27
Dis.	-.17	**.27**	-.36	-.14	.09	.11	.17
Sur.	-.23	.20	.35	**.63**	-.27	-.13	-.03
Fear	.23	-.31	.29	**.31**	.16	-.20	.22
Sad.	.16	-.33	-.08	-.13	**.26**	-.16	-.57
Ang.	-.50	.48	-.15	.09	.21	**.61**	.06

Table 5: PMI scores between emojis and all annotated emotions. Emoji ↓, emotion →. Ø: No emotion.

The overall correlations still persist; an investigation of scores where the sign has changed reveals new insights: Surprise and fear are closely correlated now, with surprise emojis showing a strong correlation with fear, while fear emojis are correlated with surprise. This interaction was not evident before, having been eclipsed by the prevalence of fear and sadness. Additionally, disgust emojis now show a slight correlation with sadness and anger, fear emojis with sadness, and anger emojis with fear and sadness.

4.1.2. Precision, recall, F1

Finally, we calculate precision, recall, and F1 using the emojis contained in each tweet as predicted labels. We calculate scores both using the emotion chosen by most anno-

tators per tweet (as in Table 4) and all emotions (as in Table 5) as gold label and show results in Table 6.

Emotion	P_{top}	R_{top}	$F1_{top}$	P_{all}	R_{all}	$F1_{all}$
Joy	0.51	0.45	0.48	0.67	0.41	0.51
Disgust	0.13	0.24	0.17	0.33	0.21	0.26
Surprise	0.24	0.33	0.28	0.57	0.29	0.38
Fear	0.03	0.33	0.06	0.13	0.24	0.17
Sadness	0.32	0.45	0.38	0.33	0.17	0.22
Anger	0.21	0.45	0.28	0.39	0.19	0.25

Table 6: Precision, recall, and F1 scores for emojis predicting annotated emotions. $_{top}$: emotion selected by most annotators used as gold label. $_{all}$: all annotations used as gold labels.

As we can see, joy emojis are the best at predicting their corresponding emotion, while fear is generally the most ambiguous. Fear emojis are present in many more tweets that are predominantly associated in fear and even when taking into account weak associations, only about every eighth tweet containing a fear emoji is also associated with fear. Disgust, anger, and sadness are similarly present in only about every third tweet containing a corresponding emoji, although sadness usually dominates when it is present. While surprise is less often the dominating emotion, its emoji are the second-best emotion indicators in tweets.

4.2. Evaluation of classifiers

We trained six support vector machine (SVM) binary classifiers with n-gram features (up to 5-grams) on the collected data (excluding annotated tweets — see Section 4.1.), one for each basic emotion, using a linear kernel and squared hinge loss. N-grams containing any of the selected emoji (for any emotion) were excluded from the feature set. Parameter selection was carried out via grid search. We show results of 3-fold cross-validation in Table 7.

Note that previous similar work reporting impressive accuracies (Purver and Battersby, 2012b) used artificially balanced train and sets. In contrast, the performance measures we report reflect the difficulty of classification with highly

Emotion	Precision	Recall	F1
Joy	0.80	0.97	0.87
Disgust	0.06	0.08	0.07
Surprise	0.07	0.12	0.09
Fear	0.07	0.36	0.11
Sadness	0.39	0.63	0.48
Anger	0.19	0.21	0.20

Table 7: Results of 3-fold cross-validation

imbalanced data and provide a more realistic estimate of performance in real-world application settings.

Final models were trained with parameters selected during optimization and applied to the classification of the annotated tweets, for which we show results in Table 8.

Emotion	P_{top}	R_{top}	$F1_{top}$	P_{all}	R_{all}	$F1_{all}$
Joy	0.08	0.81	0.14	0.51	0.87	0.64
Disgust	0.14	0.09	0.11	0.21	0.06	0.10
Surprise	0.01	0.08	0.02	0.50	0.19	0.28
Fear	0.20	0.38	0.26	0.13	0.50	0.20
Sadness	0.11	0.49	0.18	0.51	0.70	0.59
Anger	0.20	0.14	0.17	0.50	0.27	0.35

Table 8: Precision, recall, and F1 for SVM classifiers for predicting annotated emotions. $_{top}$: emotion selected by most annotators used as gold label. $_{all}$: all emotions chosen by annotators used as gold labels.

Results are comparable – in some cases even superior – to results in Table 6 using solely emoji as emotion predictors. This is encouraging, as it indicates the existence of lexical features associated with emoji and emotion usage, which can be leveraged by classifiers trained using distant supervision to capture some of the underlying emotional content. As the ability of emoji to predict emotions can be seen as a ceiling to classifier performance, classifiers will benefit from refining emoji labels. Finally, investigating emoji usage and potential differences across language will allow us to train language-specific emotional classifiers.

5. Conclusion

We have collected a substantial and multilingual data set of tweets containing emotion-specific emoji in a short time and assessed selected emoji as emotion labels, utilising human annotations as the ground truth. We found moderate correspondence between emoji and emotion annotations, indicating the presence of emotion indicators in tweet texts alongside the emoji and suggesting that emoji may be useful as distant emotion labels for statistical models of emotion in text. There was evidence of ambiguous emoji usage and interpretation. An investigation of these in future research, particularly in the multilingual setting, will help to produce more adequate emotion indicators that can be used for emotion detection in different languages. While our statistical models performed well on common emotions (joy and sadness), performance was poor on minority emotions due to class imbalance.

6. Bibliographical References

Churches, O., Nicholls, M., Thiessen, M., Kohler, M., and Keage, H. (2014). Emoticons in mind: An event-related potential study. *Social Neuroscience*, 9(2):196–202.

Davidov, D., Tsur, O., and Rappoport, A. (2010). Enhanced sentiment learning using twitter hashtags and smileys. In *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*, COLING '10, pages 241–249, Stroudsburg, PA, USA. Association for Computational Linguistics.

Derks, D., Fischer, A. H., and Bos, A. E. R. (2008). The role of emotion in computer-mediated communication: A review. *Computers in Human Behavior*, 24(3):766–785.

Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion*, 6(3-4):169–200.

emogi.com. (2015). 2015 emoji report. http://emogi.com/report.php.

Johnston, E., Norton, L. O. W., Jeste, M. D., Palmer, B. W., Ketter, M. D., Phillips, K. A., Stein, D. J., Blazer, D. G., Thakur, M. E., and Lubin, M. D. (2015). *APA Dictionary of Psychology*. Number 4311022 in APA Reference Books. American Psychological Association, Washington, DC.

Mohammad, S. M. and Kiritchenko, S. (2015). Using hashtags to capture fine emotion categories from tweets. *Computational Intelligence*, 31(2):301–326.

Mohammad, S. M. (2012). #emotional tweets. In *Proceedings of the First Joint Conference on Lexical and Computational Semantics*, SemEval '12, pages 246–255, Stroudsburg, PA, USA. Association for Computational Linguistics.

Plutchik, R. (1980). A general psychoevolutionary theory of emotion. *Theories of emotion*, 1:3–31.

Purver, M. and Battersby, S. (2012a). Experimenting with Distant Supervision for Emotion Classification. *Proceedings of the 13th Conference of the European Chapter of the Association for computational Linguistics (EACL 2012)*, pages 482–491.

Purver, M. and Battersby, S. (2012b). Experimenting with distant supervision for emotion classification. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*, EACL '12, pages 482–491, Stroudsburg, PA, USA. Association for Computational Linguistics.

Suttles, J. and Ide, N. (2013). Distant supervision for emotion classification with discrete binary values. In Alexander Gelbukh, editor, *Computational Linguistics and Intelligent Text Processing*, number 7817 in Lecture Notes in Computer Science, pages 121–136. Springer Berlin Heidelberg.