Emoji Recommendation in Private Instant Messages

Gaël Guibon, Magalie Ochs, Patrice Bellot

Emoji Recommendation in Private Instant Messages

Gaël Guibon
Caléa Solutions, LIS-CNRS
Marseille, France
gael.guibon@lsis.org

Magalie Ochs
LIS-CNRS
Marseille, France
magalie.ochs@lsis.org

Patrice Bellot
LIS-CNRS
Marseille, France
patrice.bellot@lsis.org

ABSTRACT

Emojis are some of the most common ways to convey emotions and sentiments in social messaging applications. In order to help the user choose emojis among a vast range of possibilities, we aim at developing an automatic recommendation system based on user message analysis and real emoji usage, which goes beyond the simple dictionnary lookup that is done in the industry (mainly Android and iOS). For this purpose, we present a novel automatic emoji prediction model trained and tested on real data and based on sentiment-related features. Such a model differ from the ones learnt from tweets and can predict emojis with a 84.48% f1-score and a 95.49% high precision, using MultiLabel-RandomForest algorithm on real private instant message corpus. We want to determine the best discriminative features for this task.

CCS CONCEPTS

• **Computing methodologies → Information extraction**;

KEYWORDS

emoji, messaging application, multi-label classification, natural language processing, recommendation

ACM Reference Format:
Gaël Guibon, Magalie Ochs, and Patrice Bellot. 2018. Emoji Recommendation in Private Instant Messages. In *SAC 2018: SAC 2018: Symposium on Applied Computing , April 9–13, 2018, Pau, France.* ACM, New York, NY, USA, Article 4, 3 pages. https://doi.org/10.1145/3167132.3167430

1 INTRODUCTION

Messenging applications are one of the most used means of communication. Nowadays 55% of teenagers send at least one instant message per day on their mobile phone [11], and 92% of online users send emojis [20]. Emojis are small pictures representing facial cues 😘, objects 🍓, or ideas 💤. In our research work, we focus on mobile instant text messaging applications with emojis. In these applications users have to scroll through thousands of different emojis to select one. Among the 2,389 emojis of the Unicode Consortium, 797 were added in 2015 and 233 in 2016[1].

We aim at developing a novel emoji recommendation system to help users, based on their emoji usage in context. This approach is different from emoji recommendation based on emojis metadata [4]. Different strategies are considered. One strategy, called Trust-based recommendation, consists in taking into account users' indications (ratings for instance) [14]. Another strategy, called content-based recommendation, consists in recommending items based on their description (such as metadata) [16]. In this paper, we use a content-based recommendation approach, the contents being the context in which emojis are used. To do so, we first need to focus on predicting emojis in context. It can be done for a sentence, or a whole conversation. We have decided to focus on sentences in the emoji prediction task, trying to know what features are the most discriminative for this task. It is worth noting that several emojis can be associated to one sentence, as users often combine emojis. This is why we use a multi-label classification approach, each emoji being a possible label. Moreover, we focus on a particular type of emojis: *sentiment-related emojis*. This type of emojis convey emotions, sentiments, and opinions, mainly through facial cues. For this reason, object-related emojis are excluded. Sentiment-related emojis are indeed the most widely used categories of emojis[2] - 71.63% of the emojis used being happy faces, sad faces, and hearts. The same goes on Twitter[3]. Thus, our classification models mainly exploits sentiment-based features for each sentence.

Hence, our contribution is twofold. First, we offer a novel model to automatically predict one or several sentiment-related emojis that could be recommended to a user's written sentences. As far as we know, we are the first to use a large private instant message corpus for this task, as opposed to the few other models made from public tweets [2] or public weibo messages [24]. Moreover, unlike the 20 emojis used for prediction until now, we predict up to 169 emojis. In addition, by creating an emoji prediction system from private text messages, we are able to predict emojis in a new context, different from tweets.

The paper is organized as follows. We summarize the related work (Section 2) before analysing the data (Section 3.1). Then we detail our methodology and analyse the results of our emoji prediction system (Section 3.1).

2 RELATED WORK

Emoticons (:-) , :P) and emojis (😘) are 2 different ways to represent facial cues. While the former are characters, the latter are pictures and tend to replace emoticons in social conversations [15]. According to Kelly *et al.* [8], emojis are used to improve the understanding of the message in 70% of cases.

To do so, several emoji prediction models have been proposed. Eisner [4] used word embeddings based on the Unicode[4] emoji descriptions to create an emoji vector space without emoji usage

[1]http://unicode.org/emoji/charts/full-emoji-list.html

[2]According to the SwiftKey Report [17]
[3]According to http://emojitracker.com/
[4]http://unicode.org/emoji/charts/full-emoji-list.html

ACM ISBN 978-1-4503-5191-1/18/04.
https://doi.org/10.1145/3167132.3167430

contexts. They evaluate their model through a multi-label classification of emoji descriptions and obtained 85% accuracy while predicting one emoji for several keywords. Also, Xie [24] used neural networks trained on Weibo[5] and predicted 10 possible emojis for conversations with 65% accuracy for the 3 mostly used emojis. Barbieri [2] predicted the 20 most used emojis in tweets using LSTM [6]. The first two considered keywords and conversation, and the other papers did not directly considered emoji prediction in sentences, which is our main contribution in this paper. We predict up to 169 possible emojis, going beyond the 20 most used emojis.

Text classification by emojis, emotions [3], or moods [13] can be close to each other. For instance, Li [12] did an emotion classification of blog messages using emotion-cause extraction (65% f1-score).

Our work differs by focusing on predicting emojis in sentences and by using several emojis, because users often choose several emojis for one sentence. Thus, we used a multi-label classification approach. Multi-label classification whose first aimed at associating domains to documents [18], or classifying musics by emotions [22], is a generalization of the classification task. Recent state-of-the-art reviews [23, 25] define two main approaches for multi-label classification: one by transformation and one by adaptation, respectively creating a binary classifier per label, each one being independent of the others [10], or adapting existing classification algorithms, resulting into one classifier [7].

In our work, we have a total of 169 possible emojis, so we prefered the adaptation approach for performance and computing time. The Multi-Label RandomForest algorithm was chosen, as it makes it possible to retrieve feature importance scores, and have a strong generalization capacity [19] suited to small datasets.

3 EMOJI PREDICTION

3.1 Data Analysis

To train and validate the models, we used a new text message corpus retrieved within a messaging application[6] upon acceptation from the users. The corpus is made of 9,700 sentences from 1,272 users, each sentence containing one or more emojis from 164 different emojis. The corpus is not topic filtered. The main caracteristics are described in Table 1. We automatically split the message into sentences using OpenNLP[7] [1], and then filter them in order to only keep the sentences with emojis. A sentence is represented by its text along with the list of labels (*i.e. emojis*). For instance: ("I heard about the news, it is quite depressing.", 😢😟).

Users	1,272	**Words**	69,930
Sentences	9,700	**Emojis**	18,384
Different Emojis	164	**Emojis/Sentence**	1.9
Average words/sentence	7	**Possible moods**	38
ssth* positive sentences	1,014	**ssth* negative sentences**	0
Echo positive sentences	1,532	**Echo negative sentences**	7,040
Echo negative sentences	1,128		

Table 1: Resource characteristics. *ssth = SentiStrength

A corpus of sentiment-related emojis. A key characteristic of our approach is that we only used a corpus made of messages with sentiment-related emojis. This means that emojis representing sentiments such as joy, fear or sadness, will be in our corpus, whereas the ones representing objects ☕ 🚗 will not.
We identified 169 sentiment-related emojis based on their representation in the EmojiSentimentRanking (ESR) [9], a polarity lexicon for emojis. The ESR gives 3 polarity scores for 751 emojis based on manually annotated tweets in context. From these scores we selected emojis that are sentiment-related. For instance, 😡 has a triplet negative;neutral;positive of {0.532 ; 0.108 ; 0.360}. This triplet includes this emoji as a sentiment-related one, because the neutral score is not the highest one. On the contrary, the emoji ✨ ({0.052 ; 0.545 ; 0.403}) was not selected as a sentiment-related emoji because of its neutral polarity score in the ESR.
Features used. In order to construct the prediction model, we have defined a set of features considering both textual elements and sentiment-related features. All the available features are the following. As textual features, we used bags of words or bags of characters, total word count, exclamation and interrogation marks and n-grams (up to 5-grams). As sentiment-related features, we used positive, negative, and neutral polarity scores from SentiStrength[8] [21], using the available model trained on MySpace comments and tweets, and from Echo[9] [5] trained on 9684 tweets. Another sentiment-related feature is the current mood selected by the user: users can choose between 38 moods and change it whenever they want. Hence, the current mood is attached to each message.
Token representation. Tokens can either be count vectors of words or characters gathered, then transformed using TF-IDF weighting scheme. Bags of characters can be really useful to deal with spelling variations and slang words without the need of a knowledge based or external lexicon.

In short, we used 9,700 sentences vectorized using TF-IDF on a bag of words/characters representation with computed features.

3.2 Methodology

Protocol. To predict emojis we used the *ML-RandomForest* algorithm[10]. Based on the empirical tests we conducted, we chose to use 20 trees with no depth limitation. Each model has been trained with the following methodology: 1) Preprocessing (tf-idf vectorization without stop words, and feature computation) 2) Cross validation (10 folds) 3) Classifier overall and per label evaluation.
Evaluation method. The evaluation was made from the average scores of each emoji. This means that we did not evaluate our classification as some powersets. Thus, in a sentence tagged with 😍 and 😘, each emoji will be considered separately. For instance, in our results, the accuracy, precision, and recall scores are the average score of each emoji with their frequency weight.

3.3 Results Analysis

By applying the methodology (Section 3.2) we obtained the results shown in Table 2. Higher scores (**bold**) are the ones resulting from sentiment-related features: the mood and the polarities from SentiStrength [21]. The average recall is even higher adding Echo [5]

[5]http://www.weibo.com/
[6]Mood Messenger. No SMS were retrieved, only instant messages
[7]We used the following model : http://opennlp.sourceforge.net/models-1.5/
[8]http://sentistrength.wlv.ac.uk/
[9]https://github.com/OpenEdition/echo
[10]http://scikit-learn.org/

	Accuracy	Precision	Recall	F1-score
Features	**Lemmas*, 1-5-grams, word count, tf-idf, ! mark, ? mark**			
BoW	62.01	93.57	64.43	75.22
BoC	67.96	94.17	70.56	79.72
Adding mood and SentiStrength polarity scores				
BoW	68.30	92.80	71.67	80.11
BoC	74.23	95.35	76.74	84.37
Adding mood and sentiment analysis (SentiStrength, Echo)				
BoW	68.35	92.99	71.62	80.15
BoC	**74.39**	**95.49**	**76.83**	**84.48**

Table 2: Emoji prediction cross validation scores in sentences. BoW/C = bags of words/characters. *Ignored for BoC

polarity labels. These scores show that using sentiment-related features yields better results for sentiment-related emojis.

RandomForest importance scores ranked the most discriminative features: mood came first in every run, followed by Echo Neutral Label. However, these scores do not take into account the combination of multiple features (*i.e.* sentiment-related features). Given these feature rankings, we wanted to quantify the impact of the mood as it is the first one in every run we made. To do so, we compared the scores using a baseline with bags of characters only, and then using bags of characters with an additional mood feature. The mood feature alone added 2.79% to precision, 5.04% to recall, and thus 4.74% to the f1-score measure. This is the main improvement factor in our emoji prediction models: the mood feature improves the recall, which was the weak point of our prediction model.

Nonetheless, by coupling these with the average accuracy and f1-score, our emoji prediction model performs well on private instant messages if we compare it to existing results on tweets [2], maximazing precision over recall. In Table 3, we compared the scores of the 3 most used emojis in [2] with our results.

	B-LSTM (Barbieri *et al.*)			**ML-RF**		
	P	**R**	**F1**	**P**	**R**	**F1**
😂	0.7	0.84	0.77	0.98	0.80	0.88
❤️	0.61	0.78	0.69	0.94	87.50	0.88
😍	0.52	0.30	0.38	0.98	0.71	0.82

Table 3: Sample emoji scores (Precision, Recall, F1-score)

Reproductibility. Because of privacy needs, we cannot release our corpus. And, as far as we know, there is no available private instant message corpus with emojis. However, this approach can be reproduced[11] on other public data such as tweets.

4 PERSPECTIVES

We contribute by proposing an emoji-prediction system using supervised multi-label classification through RandomForest and sentiment-related features to automatically predict up to 169 sentiment-related emojis, which is higher than the 20 emojis used so far. Our models obtained good prediction scores. For example, a 94.3% precision score was obtained using bags of characters with mood and polarity scores as features. From our results we can draw three conclusions. First, mood feature is important to improve emoji prediction, thus messaging applications should use it. Secondly, using bags of characters for private instant messages, instead of a common bag of words, drastically increases emoji recommendation. Finally, results show that using polarity scores does not help to give better sentiment-related-emoji recommendation.

In future work, we will use these results as a baseline to determine whether or not deep learning yields better results, considering our corpus is not excessively large. Finally, we will compare the results from private and public instant messages, taking different contexts into account: sentence, conversation, and user profile.

[11]Predicted emojis, scores, features, and example available here : https://gguibon.github.io/sac2018/index.html

REFERENCES

[1] Jason Baldridge. 2005. The opennlp project. (2005).
[2] Francesco Barbieri, Miguel Ballesteros, and Horacio Saggion. 2017. Are Emojis Predictable? *arXiv preprint arXiv:1702.07285* (2017).
[3] Soumaya Chaffar and Diana Inkpen. 2011. Using a heterogeneous dataset for emotion analysis in text. In *Canadian Conference on Artificial Intelligence*. Springer, 62–67.
[4] Ben Eisner, Tim Rocktäschel, Isabelle Augenstein, Matko Bošnjak, and Sebastian Riedel. 2016. emoji2vec: Learning Emoji Representations from their Description. *arXiv preprint arXiv:1609.08359* (2016).
[5] Hussam Hamdan, Patrice Bellot, and Frederic Bechet. 2015. Sentiment lexicon-based features for sentiment analysis in short text. In *In Proceeding of the 16th International Conference on Intelligent Text Processing and Computational Linguistics*.
[6] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long Short-Term Memory. *Neural Comput.* 9, 8 (Nov. 1997), 1735–1780.
[7] Shu Huang, Wei Peng, Jingxuan Li, and Dongwon Lee. 2013. Sentiment and topic analysis on social media: a multi-task multi-label classification approach. In *Proceedings of the 5th annual acm web science conference*. ACM, 172–181.
[8] Caroline Kelly. 2015. *Do you know what I mean >:(: A linguistic study of the understanding of emoticons and emojis in text messages*. Master's thesis. Halmstad University.
[9] Petra Kralj Novak, Jasmina Smailovic, Borut Sluban, and Igor Mozetic. 2015. Sentiment of Emojis. *PLOS ONE* 10, 12 (2015).
[10] Boris Lauser and Andreas Hotho. 2003. Automatic multi-label subject indexing in a multilingual environment. In *International Conference on Theory and Practice of Digital Libraries*. Springer, 140–151.
[11] Amanda Lenhart, Aaron Smith, Monica Anderson, Maeve Duggan, and Andrew Perrin. 2015. Teens, technology and friendships. (2015).
[12] Weiyuan Li and Hua Xu. 2014. Text-based emotion classification using emotion cause extraction. *Expert Systems with Applications* 41, 4 (2014), 1742–1749.
[13] Gilad Mishne. 2005. Experiments with mood classification in blog posts. In *Proceedings of ACM SIGIR 2005 workshop on stylistic analysis of text for information access*, Vol. 19. Citeseer, 321–327.
[14] John O'Donovan and Barry Smyth. 2005. Trust in recommender systems. In *Proceedings of the 10th international conference on Intelligent user interfaces*. ACM, 167–174.
[15] Umashanthi Pavalanathan and Jacob Eisenstein. 2015. Emoticons vs. emojis on Twitter: A causal inference approach. *arXiv preprint arXiv:1510.08480* (2015).
[16] Michael J. Pazzani and Daniel Billsus. 2007. Content-based recommendation systems. In *The adaptive web*. Springer, 325–341.
[17] SwiftKey PT. 2015. Most-used emoji revealed: Americans love skulls, Brazilians love cats, the French love hearts. (2015), 18 pages. http://swiftkey.com/en/blog/americans-love-skulls-brazilians-love-cats-swiftkey-emoji-meanings-report/
[18] Timothy N. Rubin, America Chambers, Padhraic Smyth, and Mark Steyvers. 2012. Statistical topic models for multi-label document classification. *Machine Learning* 88, 1 (2012), 157–208.
[19] Carolin Strobl, Anne-Laure Boulesteix, Thomas Kneib, Thomas Augustin, and Achim Zeileis. 2008. Conditional Variable Importance for Random Forests. *BMC Bioinformatics* 9, 1 (2008), 307.
[20] Emoji Research Team. 2015. Emoji Report. (2015), 40 pages. https://emogi.com/documents/Emoji_Report_2015.pdf
[21] Mike Thelwall, Kevan Buckley, Georgios Paltoglou, Di Cai, and Arvid Kappas. 2010. Sentiment strength detection in short informal text. *Journal of the American Society for Information Science and Technology* 61, 12 (2010), 2544–2558.
[22] Konstantinos Trohidis, Grigorios Tsoumakas, George Kalliris, and Ioannis P. Vlahavas. 2008. Multi-Label Classification of Music into Emotions.. In *ISMIR*, Vol. 8. 325–330.
[23] Grigorios Tsoumakas and Ioannis Katakis. 2006. Multi-label classification: An overview. *International Journal of Data Warehousing and Mining* 3, 3 (2006).
[24] Ruobing Xie, Zhiyuan Liu, Rui Yan, and Maosong Sun. 2016. Neural Emoji Recommendation in Dialogue Systems. *arXiv preprint arXiv:1612.04609* (2016).
[25] Min-Ling Zhang and Zhi-Hua Zhou. 2014. A Review on Multi-Label Learning Algorithms. *IEEE Transactions on Knowledge and Data Engineering* 26, 8 (2014), 1819–1837.