

Mining the Relationship between Emoji Usage Patterns and Personality

Weijian Li*, Yuxiao Chen*, Tianran Hu, Jiebo Luo

University of Rochester
Rochester, NY 14623

{wli69, ychen211, thu, jluo@cs.rochester.edu}

Abstract

Emojis have been widely used in textual communications as a new way to convey nonverbal cues. An interesting observation is the various emoji usage patterns among different users. In this paper, we investigate the correlation between user personality traits and their emoji usage patterns, particularly on overall amounts and specific preferences. To achieve this goal, we build a large Twitter dataset which includes 352,245 users and over 1.13 billion tweets associated with calculated personality traits and emoji usage patterns. Our correlation and emoji prediction results provide insights into the power of diverse personalities that lead to varies emoji usage patterns as well as its potential in emoji recommendation tasks.

Introduction

Emojis are a large set of pictorial representations (Miller et al. 2016) and are widely used for textual Computer-Mediated Communications (CMC). Researchers find that the same Emoji is used and interpreted differently by different users (Miller et al. 2017). Several studies (Hu et al. 2017; Ai et al. 2017) have been done to find the reason behind this observation. Other studies (Felbo et al. 2017; Barbieri, Ballesteros, and Saggion 2017) focus on real applications, shows the potential of predicting emojis given user text information. Their methods mainly based on emoji context meanings, sentiments, and linguistic effects and less focus on the differences of human attributes. Inspired by studies that extract divergent human personalities from social media data (Liu et al. 2016; Wei et al. 2017) and assessing participants' personalities from emojis (Marengo, Giannotta, and Settanni 2017) in a relatively small scale, we consider user personality may be one of the factors that lead to different emoji usage patterns and may also contribute to emoji recommendation tasks.

To this end, our paper dives deeper to explore the hidden relationship between user personality and emoji usage patterns from a data-driven and fully computational perspective. By doing this, we are able to extract personality traits and emoji usage patterns from collected volumetric tweets

data to avoid influences from subjective surveys or questionnaire answers.

In summary, our work makes two main contributions:

- We show personality as an influential factor in different emoji usage patterns by presenting a fully computational analysis using large-scale social media data.
- We present personality's discriminant power in the emoji prediction task with a superior prediction result when personality is fused as additional input features into the machine learning classifier.

Data Collection and Preprocessing

Our dataset contains 352,245 twitter users collected from March 2016 to June 2016. For each user, we collect his/her most recent 3,200 tweets until June 2016 which gives us 1.13 billion tweets in total. Since we are more interested in individual users instead of public accounts whose tweet context may not relate to the account owner, we keep users whose number of followers less than 1649 (90th percentile of follower distribution) and the number of followees less than 1180 (90th percentile of followee distribution). Secondly, for each remained user, we remove the tweets that are: not written in English, contain less than ten words after removing hashtags, URLs, mentions and email address. Users who have less than 500 tweets left are removed to keep enough context information for our later analysis. Lastly, to conduct our research with a focus on the majority of users, we filter users who frequently or seldom use emojis. To be more specific, for each user, his/her emoji usage frequency is defined as the number of tweets containing emojis divided by the total number of tweets. Users whose emoji usage frequency is more than 0.95 or less than 0.05 are filtered out. Eventually, we end up with 71,823 users, 75 million tweets.

Extraction of User Personality Traits

In this paper, we use Big Five personality traits (Big Five) (John and Srivastava 1999) to describe user personality. Big Five is one of the popular and widely adopted definitions of personality. It models human personality as five different dimensions: Openness (degree of appreciation to new, unusual ideas and experiences), Conscientiousness (the tendency to self-discipline), Extraversion (the extent of preference to being in social situations), Agreeableness (willing-

| Dimension | P-mean | N-mean. | t-value | p-value. |
|-----------|--------------|--------------|---------|--------------|
| Open. | 0.237 | 0.233 | 2.147 | 0.032 |
| Cons. | 0.229 | 0.238 | -5.001 | 5.7e-7 |
| Extra. | 0.225 | 0.273 | -25.58 | 4.14e-43 |
| Agree. | 0.261 | 0.237 | 12.69 | 7.67e-37 |
| Neuro. | 0.258 | 0.223 | 19.34 | 6.76e-83 |

Table 1: The result of t-test on mean emoji usage frequency between positive (P) and negative (N) sets in each five personality dimensions. The last column represents two-tailed p-values. All results are significant (two-tailed p-value < 0.005) except openness.

ness to cooperate or compassionate) and Neuroticism (level of emotional stability). We use abbreviations Open., Cons., Extra., Agree., Neuro., or O, C, E, A, N for each dimension if necessary. Each user’s Big Five is computed based on their tweet text data leveraging a LIWC (Pennebaker et al. 2015) based personality calculation model Receptiviti¹ which is suggested in the LIWC paper. We take each user’s whole processed tweets as input and obtain each user’s psychological attributes from Receptiviti. We keep the Big Five personality scores from the attributes as personality traits for the users.

Emoji Usage Analysis

In this section, we reveal the relationship between users’ personalities and their mean emoji usage frequencies, as well as the relationship with emoji usage preferences.

Usage Frequency Analysis

Here we are interested in how each dimension of the five personalities affects mean emoji usage frequency. For each personality dimension, we collect representative users who have the top and bottom quartile personality scores, mark them as the positive set and the negative set respectively. By doing this, we obtain two sets of users on each personality dimension which gives us ten sets of users in all. For each set in the pair sets (positive and negative), we compute mean emoji usage frequencies, as defined in the previous section, for the users within the set, and conduct two-tailed t-test over two sets to verify the significance of differences for mean usage frequencies. Our results are reported in Table 1.

- **Openness** It is obvious to see four out of five results are significant at two-tailed p-value < 0.005, except for openness. For openness, since its p-value is greater than the significant level, we cannot reject the null hypothesis on the equality of two mean values. We consider the reason is that emojis are predefined pictorial representations and have been used for online communications for a long time. Whether to use emojis or not does not depend on the willingness of accepting new things. Thus, we infer that users with different openness scores use a similar amount of emojis in their tweets.

- **Conscientiousness** Users with a higher score in conscientiousness use less emojis than users with lower conscientiousness scores. This is because of their vigilance and self-discipline which make them stick to traditional and plainer text contexts to conserve their emotions.
- **Extraversion** Users with a low score in extraversion use much more emojis than users with higher scores. What’s more, they are the users who use emojis the most. We consider the reason is that introversion users prefer implicit visual contexts than explicit text messages on which sender has to express themselves directly.
- **Agreeableness** Users with high agreeableness score use more emojis than users with lower scores. By definition, high agreeableness score indicates distinct attributes like empathetic and compliance. Messages convey these ideas are not easy to describe directly using text words. Instead, they are more expressible via vivid facial or object representations like emojis.
- **Neuroticism** Users with a high score in neuroticism have unstable emotions and have difficulty to control their emotions well. Given their unpredictable emotions, we consider neuroticism users use more emojis than plain text words to better describe their emotional feelings.

Usage Preference Analysis

Our previous analysis measures mean emoji usage frequencies and shows how emoji usage frequency varies along each of five personality dimensions. To extend our study quantitatively and more specifically, we investigate which emojis are more preferred on five personality dimensions. By the time we conducted our experiment, there were in total 2685 emojis² in the dataset. Here we focus on the most frequently used emojis based on the overall usage frequency. To obtain these emojis, we first collect overall emoji frequencies over the entire dataset and sort the popularity of emojis in descending order. Emojis whose frequency is larger than 0.1% of total emoji frequency are selected as our study objects. Following these steps, we obtain 67 most frequently used emojis. We compute Spearman’s Rank Correlation Coefficient (Kendall 1938) between five personality traits and their emoji usage frequencies. Our results are reported in Table 2. For each personality dimension, we list the top and bottom ten correlated emojis on the left column and corresponding correlation coefficient scores on the right column.

Openness In general, correlation coefficient scores in openness are relatively small in scale indicating similar usage preferences on each emoji. This result coincides with our previous findings on emoji usage frequency. Besides, there are no clear emoji preference patterns based on the definition of openness. Thus, we infer that openness has little correlation with emoji usage patterns.

Conscientiousness People with higher conscientiousness scores are more vigilant and self-controlled. They use less negative emotion emojis such as 😞 (-0.224), 😓 (Wijeratne et al. 2016) (-0.201), 😫 (-0.200) and 😩 (-0.184). Instead,

¹<https://www.receptiviti.ai/>, retrieved Jan. 2018.

²The emojis used in this study are extracted from Twemoji v2.4. <https://emojiemoji.com/twitter/>, retrieved Jan. 2018.

| Emoji | Open. | Emoji | Cons. | Emoji | Extra. | Emoji | Agree. | Emoji | Neuro. |
|--------------------|--------|----------------------|--------|----------------------|--------|----------------------|--------|------------------------|--------|
| Skull | 0.125 | Sun | 0.216 | TU | 0.167 | Two Hearts | 0.206 | Loudly Crying Face | 0.245 |
| Sparkling Heart | 0.102 | TU | 0.182 | WF | 0.136 | Red Heart | 0.186 | Weary Face | 0.234 |
| Weary Face | 0.085 | WF | 0.179 | TU: Light Skin | 0.102 | SF with Smiling Eyes | 0.169 | Unamused Face | 0.203 |
| Loudly Crying Face | 0.081 | Party Popper | 0.160 | Clinking Beer Mugs | 0.084 | Face Blowing a Kiss | 0.167 | Upside-Down Face | 0.202 |
| Sparkles | 0.078 | GF | 0.141 | GF | 0.083 | Sparkling Heart | 0.165 | Skull | 0.196 |
| Eyes | 0.067 | GF w/ Big Eyes | 0.129 | GF w/ Big Eyes | 0.074 | Loudly Crying Face | 0.147 | Face w/ Rolling Eyes | 0.194 |
| Pistol | 0.060 | See-No-Evil Monkey | 0.117 | WF w/ Tongue | 0.062 | Sparkles | 0.137 | Face w/ Tears of Joy | 0.177 |
| Two Hearts | 0.058 | TU: Light Skin Tone | 0.108 | See-No-Evil Monkey | 0.061 | Relieved Face | 0.135 | Pistol | 0.175 |
| Hundred Points | 0.053 | Yellow Heart | 0.095 | Sun | 0.060 | SF w/ Heart Eyes | 0.128 | SF w/ Heart Eyes | 0.129 |
| Red Heart | 0.050 | GF with Smiling Eyes | 0.093 | Eyes | 0.045 | Growing Heart | 0.127 | Thinking Face | 0.126 |
| See-No-Evil Monkey | -0.094 | Loudly Crying Face | -0.224 | Loudly Crying Face | -0.369 | Face w/ Medical Mask | -0.174 | WF | -0.139 |
| Flushed Face | -0.083 | Skull | -0.202 | Upside-Down Face | -0.263 | Clinking Beer Mugs | -0.165 | GF | -0.129 |
| WF | -0.075 | Face w/ Tears of Joy | -0.201 | Unamused Face | -0.252 | TU | -0.151 | TU | -0.124 |
| Sun | -0.070 | Weary Face | -0.200 | Two Hearts | -0.229 | Sleeping Face | -0.150 | GF w/ Big Eyes | -0.108 |
| TU: Light Skin | -0.059 | Unamused Face | -0.184 | GF with Sweat | -0.222 | See-No-Evil Monkey | -0.148 | Sun | -0.101 |
| TU | -0.056 | Thinking Face | -0.170 | Broken Heart | -0.219 | Pistol | -0.128 | GF w/ Smiling Eyes | -0.082 |
| Sleeping Face | -0.046 | Face w/ Rolling Eyes | -0.160 | Face w/ Rolling Eyes | -0.188 | Pouting Face | -0.121 | Face Screaming in Fear | -0.081 |
| Gramacing Face | -0.042 | Eyes | -0.137 | Expressionless Face | -0.186 | Sun | -0.116 | WF w/ Tongue | -0.074 |
| Party Popper | -0.037 | Relieved Face | -0.133 | Weary Face | -0.185 | TU: Light Skin Tone | -0.112 | GS Face | -0.073 |
| GF | -0.036 | Upside-Down Face | -0.127 | Relieved Face | -0.179 | Eyes | -0.102 | TU: Light Skin Tone | -0.071 |

Table 2: Top 10 positive and negative Spearman Correlation Coefficient Results on each Big Five Personality dimension. All coefficient scores significant at Bonferroni-corrected $p < 0.005$. Correlation results on each dimension are listed on the right columns. GF: Grinning Face, WF: Winking Face, GS: Grinning Squinting, TU: Thumbs Up.

they prefer expressing positive emotions through emojis like 🌞 (0.216), 👍 (0.182) 😊 (0.179) and 😄 (0.141). 🎉 (0.160) is also used frequently. Though this emoji is explained as “Party Popper”, we find people use it regularly when sending congratulations and celebrating achievements which are precisely positive emotional circumstances.

Extraversion Similar to conscientiousness, we find people high in extraversion dimension prefer positive emotion emojis such as 👍 (0.167), 😊 (0.136) and 🍻 (0.102). On the other hand, they seldom use emojis like 🙄 (-0.369), 😞 (-0.263) and 😏 (-0.252) which represent negative or ambiguous facial expressions. We think 🍻 (0.084) is often used at celebrations or parties where many people are gathered together. This conforms their extravert personality.

Agreeableness We can see many heart shape (or contain heart shape) emojis such as 💕 (0.206), ❤️ (0.186), 🧡 (0.165) and particularly 😊 (0.167) are positively correlated with Agreeableness. We consider people are showing “like”, “love” and other empathetic meanings with strong positive emotions and inner feelings which are highly representative for users with high agreeableness scores. On the opposite, 😞 (-0.174), 😏 (-0.150) and 🙄 (-0.121) are seldom used because of their negative and dislike emotions.

Neuroticism It is interesting to see that neuroticism people prefer to use exaggerated facial expressions with rich emotions. For example, “Loudly Crying Face” 😭 (0.245), “Weary Face” 😩 (0.234), “Upside Down Face” 🙄 (0.202) and “Face w/ Rolling Eyes” 🤪 (0.194). These emojis have little positive correlation with other personality traits which represents a unique emoji usage pattern for neuroticism users as well as their distinct emotional characteristics.

Predicting Emoji Usages

To validate our observations under a real application, we build a classifier to predict 19 most frequently used emojis based on users’ tweets texts data. Our goal is to see if fusing five personality traits as additional features would improve the classifier’s performance.

Our final prediction results are reported in table 3. Given that our test set is balanced, we use accuracy here for intuitive representation. It is clear to see that the prediction performance is improved when the classifier is fused with personality data. Each personality dimension increases prediction result. Openness, which is found to be hardly correlated with emoji usage patterns, has a little improvement. Conscientiousness, on the other hand, provides the most information as a personality trait. Our best prediction result is obtained when all five personalities are given to the classifier which indicates the combined power of five personalities. The improved prediction results not only validate our previous correlation findings but also show the potential of personality feature in personalized emoji recommendation systems.

To show the effects of personality traits, we report a sample result for each personality dimension in Table 3. The emojis are initially classified incorrectly by only using text information, but are correctly classified after fusing the corresponding row’s personality information. Concretely, for conscientiousness, extraversion, and neuroticism, the classifier correctly labeled the sample tweet with emojis in correct sentiments but fails to predict the exact emoji type for the user. For agreeableness, the classifier gives a reasonable prediction based on words success and celebrate, while the user who posted this tweet prefer to use ❤️. After fusing person-

| | Acc | Tweet Text | GD | Pred. 1 | Pred. 2 |
|------------|--------------|--|----|---------|---------|
| Text | 0.158 | - | - | - | - |
| Text + O | 0.160 | - | - | - | - |
| Text + C | 0.166 | Stay positive mate, know how you feel, everyday is a new adventure | 👍 | 😊 | 👍 |
| Text + E | 0.165 | Yay! I can't be in phoenix tonight, but can watch online | 👍 | 😊 | 👍 |
| Text + A | 0.162 | Going to post an entry to celebrate the success of the album gone gold | ❤️ | 👍 | ❤️ |
| Text + N | 0.163 | I'm working 10:30 - 9:00 today ... how did I get myself into | 😓 | 😊 | 😓 |
| Text + All | 0.172 | - | - | - | - |

Table 3: Accuracy of test data set for each model. Text: using text feature only. Text + ‘X’: using text feature and one of five personality dimensions. Text + All: using text feature and all 5 personality dimensions. Acc: Accuracy. GD: Ground Truth Label. Pred. 1: Prediction result using only text feature. Pred. 2: Prediction result using text feature and corresponding row’s personality trait.

ality traits, our classifier correctly labels all these examples by learning the personalized information from personality.

Discussion and Conclusion

In this paper, we investigate the hidden relationship between user personality and emoji usage patterns in Twitter tweets using a computational approach. Generally speaking, users with low extraversion scores use emojis most frequently while users with low neuroticism scores use emojis the least often. From another perspective, we find high correlation values and specific emoji usage in line with perceived user’s personality traits. However, some exceptions do exist. For example, Openness, as a sign of the wiliness to new objects, shows no obvious relationship with emoji usage. We consider emoji, itself as a pre-designed form of non-verbal communication tool, may suggest little information on openness characteristics. Finally, our prediction results show discriminant power of personalities in differentiating different emoji usage patterns. Best prediction performance is achieved by fusing all five personalities into the developed classifier. We believe our findings can benefit researchers and engineers in designing personalized emoji recommendation systems.

Future work can be directed at studying personality influences under more specific circumstances such as under different topics. Although personality is a human characteristic, users’ personality may change on different topics.

Acknowledgement

We thank the support of New York State through the Goergen Institute of Data Science.

References

- [Ai et al. 2017] Ai, W.; Lu, X.; Liu, X.; Wang, N.; Huang, G.; and Mei, Q. 2017. Untangling emoji popularity through semantic embeddings. In *ICWSM*, 2–11.
- [Barbieri, Ballesteros, and Saggion 2017] Barbieri, F.; Ballesteros, M.; and Saggion, H. 2017. Are emojis predictable? In *Lapata M, Blunsom P, Koller A, editors. 15th Conference of the European Chapter of the Association for Computational Linguistics; 2017 Apr 3-7; Valencia, Spain. Stroudsburg (PA): ACL; 2017. p. 105-11. ACL (Association for Computational Linguistics).*
- [Felbo et al. 2017] Felbo, B.; Mislove, A.; Søgaard, A.; Rahnwan, I.; and Lehmann, S. 2017. Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, 1616–1626.
- [Hu et al. 2017] Hu, T.; Guo, H.; Sun, H.; Nguyen, T.-v. T.; and Luo, J. 2017. Spice up your chat: The intentions and sentiment effects of using emoji. *arXiv preprint arXiv:1703.02860*.
- [John and Srivastava 1999] John, O. P., and Srivastava, S. 1999. The big five trait taxonomy: History, measurement, and theoretical perspectives. *Handbook of personality: Theory and research* 2(1999):102–138.
- [Liu et al. 2016] Liu, L.; Preotiuc-Pietro, D.; Samani, Z. R.; Moghaddam, M. E.; and Ungar, L. H. 2016. Analyzing personality through social media profile picture choice. In *ICWSM*, 211–220.
- [Marengo, Giannotta, and Settanni 2017] Marengo, D.; Giannotta, F.; and Settanni, M. 2017. Assessing personality using emoji: An exploratory study. *Personality and Individual Differences* 112:74–78.
- [Miller et al. 2016] Miller, H.; Thebault-Spieker, J.; Chang, S.; Johnson, I.; Terveen, L.; and Hecht, B. 2016. Blissfully happy or ready to fight: Varying interpretations of emoji. *Proceedings of ICWSM 2016*.
- [Miller et al. 2017] Miller, H.; Kluver, D.; Thebault-Spieker, J.; Terveen, L.; and Hecht, B. 2017. Understanding emoji ambiguity in context: The role of text in emoji-related miscommunication. In *11th International Conference on Web and Social Media, ICWSM 2017*. AAAI Press.
- [Pennebaker et al. 2015] Pennebaker, J. W.; Boyd, R. L.; Jordan, K.; and Blackburn, K. 2015. The development and psychometric properties of liwc2015. Technical report.
- [Wei et al. 2017] Wei, H.; Zhang, F.; Yuan, N. J.; Cao, C.; Fu, H.; Xie, X.; Rui, Y.; and Ma, W.-Y. 2017. Beyond the words: Predicting user personality from heterogeneous information. In *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*, 305–314. ACM.
- [Wijeratne et al. 2016] Wijeratne, S.; Balasuriya, L.; Sheth, A.; and Doran, D. 2016. Emojinet: Building a machine read-

able sense inventory for emoji. In *International Conference on Social Informatics*, 527–541. Springer.