

Decoding the New World Language: Analyzing the Popularity, Roles, and Utility of Emojis

Qiaozhu Mei
University of Michigan
Ann Arbor, Michigan
qmei@umich.edu

ABSTRACT

Emojis have quickly become a universal language that is used by worldwide users, for everyday tasks, across language barriers, and in different apps and platforms. The prevalence of emojis has quickly attracted great attentions from various research communities such as natural language processing, Web mining, ubiquitous computing, and human-computer interaction, as well as other disciplines including social science, arts, psychology, and linguistics.

This talk summarizes the recent efforts made by my research group and our collaborators on analyzing large-scale emoji data. The usage of emojis by worldwide users presents interesting commonality as well as divergence. In our analysis of emoji usage by millions of smartphone users in 212 countries, we show that the different preferences and usage of emojis provide rich signals for understanding the cultural differences of Internet users, which correlate with the Hofstede's cultural dimensions [4].

Emojis play different roles when used alongside text. Through jointly learning the embeddings and topological structures of words and emojis, we reveal that emojis present both complementary and supplementary relations to words. Based on the structural properties of emojis in the semantic spaces, we are able to untangle several factors behind the popularity of emojis [1].

This talk also highlights the utility of emojis. In general, emojis have been used by Internet users as text supplements to describe objects and situations, express sentiments, or express humor and sarcasm; they are also used as communication tools to attract attention, adjust tones, or establish personal relationships. The benefit of using emojis goes beyond these intentions. In particular, we show that including emojis in the description of an issue report on GitHub results in the issue being responded to by more users and resolved sooner.

Large-scale emoji data can also be utilized by AI systems to improve the quality of Web mining services. In particular, a smart machine learning system can infer the latent topics, sentiments, and even demographic information of users based on how they use emojis online. Our analysis reveals a considerable difference between female and male users of emojis, which is big enough for a machine learning algorithm to accurately predict the gender of a user. In Web services that are customized for gender groups, gender

inference models built upon emojis can complement those based on text or behavioral traces with fewer privacy concerns [2].

Emojis can be also used as an instrument to bridge Web mining tasks across language barriers, especially to transfer sentiment knowledge from a language with rich training labels (e.g., English) to languages that have been difficult for advanced natural language processing tasks [3]. Through this bridge, developers of AI systems and Web services are able to reduce the inequality in the quality of services received by the international users that has been caused by the imbalance of available human annotations in different languages.

In general, emojis have evolved from visual ideograms to a brand-new world language in the era of AI and a new Web. The popularity, roles, and utility of emojis have all gone beyond people's original intentions, which have created a huge opportunity for future research that calls for joint efforts from multiple disciplines.

CCS CONCEPTS

• Information systems → Data mining; • Computing methodologies → Natural language processing.

KEYWORDS

emojis; text mining; natural language processing

ACM Reference Format:

Qiaozhu Mei. 2019. Decoding the New World Language: Analyzing the Popularity, Roles, and Utility of Emojis. In *Companion Proceedings of the 2019 World Wide Web Conference (WWW '19 Companion)*, May 13–17, 2019, San Francisco, CA, USA. ACM, New York, NY, USA, 2 pages. <https://doi.org/10.1145/3308560.3316541>

BIOGRAPHY

Qiaozhu Mei is an associate professor in the School of Information and the Department of Electrical Engineering and Computer Science at the University of Michigan. His research focuses on large-scale data mining, machine learning, information retrieval, and natural language processing, with broad applications to social media, Web search, and health informatics. Qiaozhu is an ACM distinguished member (2017) and a recipient of the NSF Career Award (2011). His work has received multiple best paper awards and nominations at ICML, KDD, WWW, WSDM, and other major conferences. He has served as the General Co-Chair of SIGIR 2018 and on the editorial boards of multiple journals such as JMLR, TOIS, and TWEB. He is the founding director of the new master degree of applied data science at the University of Michigan.

Qiaozhu received his PhD degree in computer science from the University of Illinois at Urbana-Champaign and his Bachelor of science degree from Peking University.

This paper is published under the Creative Commons Attribution 4.0 International (CC-BY 4.0) license. Authors reserve their rights to disseminate the work on their personal and corporate Web sites with the appropriate attribution.

WWW '19 Companion, May 13–17, 2019, San Francisco, CA, USA

© 2019 IW3C2 (International World Wide Web Conference Committee), published under Creative Commons CC-BY 4.0 License.

ACM ISBN 978-1-4503-6675-5/19/05.

<https://doi.org/10.1145/3308560.3316541>

REFERENCES

- [1] Wei Ai, Xuan Lu, Xuanzhe Liu, Ning Wang, Gang Huang, and Qiaozhu Mei. 2017. Untangling emoji popularity through semantic embeddings. In *Eleventh International AAAI Conference on Web and Social Media*.
- [2] Zhenpeng Chen, Xuan Lu, Wei Ai, Huoran Li, Qiaozhu Mei, and Xuanzhe Liu. 2018. Through a gender lens: learning usage patterns of emojis from large-scale android users. In *Proceedings of the 2018 World Wide Web Conference*. International World Wide Web Conferences Steering Committee, 763–772.
- [3] Zhenpeng Chen, Sheng Shen, Ziniu Hu, Xuan Lu, Qiaozhu Mei, and Xuanzhe Liu. 2019. Emoji-powered representation learning for cross-Lingual sentiment classification. In *Proceedings of the 2019 World Wide Web Conference*.
- [4] Xuan Lu, Wei Ai, Xuanzhe Liu, Qian Li, Ning Wang, Gang Huang, and Qiaozhu Mei. 2016. Learning from the ubiquitous language: an empirical analysis of emoji usage of smartphone users. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 770–780.