Emojis As Indicators Of Spatial-Temporal-Thematic Developments In Geo-Social Media

by SAMANTHA LEVI

Social media platforms like Twitter play an increasingly large role in the lives of over half of the world’s population and present a unique opportunity for gaining insights into users’ thoughts and ideas. Emojis are increasingly used in social media to convey meaning and can be leveraged to identify current events and other topics that attract large numbers of users, such as election results and natural disasters [1]. This study seeks to determine whether emojis can be used to identify relevant topics and their evolution in a non-topic-specific dataset.

**METHODOLOGY**

This study used a combination of exploratory analysis and emoji-specific analysis to identify topically consistent emojis and their usage across time and space. The typicality measure was used to find the representativeness of an emoji for a given subset of the larger dataset. Typicality is the normalized difference of relative frequencies between a subset and total dataset [2].

**DATA**

Raw data was sourced using the Twitter API. The dataset was refined to include only tweets that were geotagged within Europe during the year 2020 and that contain at least one hashtag and one emoji. Notably, the dataset was not thematically filtered in any way. The HyperLogLog algorithm was used together with cryptographic hashing and spatial data aggregation to improve user privacy whenever possible.

**TOPICAL CONSISTENCY**

To test the semantic consistency of a selected emoji, a list of the 20 most frequently co-occurring hashtags was generated. Then, posts containing thematically similar hashtags were quantified and divided by the number of posts in the list. The result is the percent topical consistency of the emoji. The clapping hands emoji, for example, is 89% topical consistent regarding healthcare worker appreciation.

**TEMPORAL TYPICALITY**

Typicality calculations were conducted using monthly subsets of the data for over 100 different emojis. Charts of these measurements illustrate temporal trends in emoji usage.

**SPATIAL TYPICALITY**

For the calculation of typicality across space, a 100 by 100 kilometer grid was superimposed on the study area. One subset in this case consists of all tweets that were geotagged within one cell. The resulting maps display the typical spatial distribution of emoji use.

**RESULTS**

In addition to calculating spatial and temporal typicality on the top 100 emojis by user days, topical consistency and spatial-temporal typicality were calculated for 35 selected emojis. Relevant topics that were detected in the dataset fall into four main categories: the COVID-19 pandemic, healthcare worker appreciation, activist movements such as Black Lives Matter and LGBTQIA+, and leisure activities.

**CONCLUSION**

This research took a novel approach to the analysis of emojis in geo-social media by 1) using a dataset that was not thematically pre-filtered, 2) analyzing trends in emoji usage across space and time simultaneously, and 3) using these trends to identify local topics in a geo-social media dataset. Methods were created for detecting the topical consistency of emojis and visualizing their use over both time and space. Ultimately, changes in the use of topically consistent emojis were leveraged to reveal relevant topics within the dataset.

**SUPPLEMENTARY MATERIALS**

This master thesis was created within the Cartography M.Sc. programme – proudly co-funded by the Erasmus+ Programme of the European Union.

**REFERENCES**
