Facial expression emoji : visualization and analysis

Ngoc-Bich NGUYEN[1], Cheng CHENG[2]
Université Claude Bernard – Lyon 1

ABSTRACT

Research question: During a particular period of time, what are the mostly used emoji? How their emotions have been changing during this period?

Methodology: Using streaming twitter to process data in realtime, force-directed graph algorithm and shuffling algorithms using matrix diagrams to visualize emoji co-occurrence graph, matrix

Practical implications: knowledge of processing and visualizing data.

KEYWORDS: EMOJI, GLYPH, EMOJI FREQUENCY, TWITTER, INTERACTION GRAPHS.

1 INTRODUCTION

Emoji has become part of our daily lives. In fact, people can use emoji to express their complicated emotions, their fancy thoughts or even they can also plop one to reply someone while they are in a rush. This is convient in term of expression and economical in term of time.

In this work, we present the emoji ranking over a period of time with an emoji sentiment lexicon of 50 emoji. This lexicon is constructed from 10 millions tweets selected from 7 Jan. to 11 Jan. 2017. These informations are figured in a bar chart, a line graph, an emoji-cloud and an emoji co-occurrence matrix. We introduce co-occurence network of emoji that specifies the appearance of two or more emoji in the same tweet (with the emoji as the vertices and the edges representing the found relation between two emoji).

This paper first reviews related work on tracking emoji and on analysing sentiment of emoji. We then describe our project in detail from technical aspect to graphics interpretation. Finally, we discuss the results of our work and suggest improvement.

1 nguyenntngocbich@gmail.com
2 cheng.cheng.chg@gmail.com

2 RELATED WORKS

2.1 Analysing sentiment of emoji

Previous work on emoji analysis falls into analysing sentiment of emoji. This study of P. K. Novak et al.[5] indicates that the sentiment of emoji computed from the sentiment of the tweets in which they occur. Regarless of countries and culture, it is shown that twitteurs tend to use positive emoji, they use “joy”, “smiling face”, “smiling face with heart-shaped eyes”, “heart”... to a large extend.

2.2 Tracking emoji

In the closest related work, Matthew Rothenberg[4] uses Twitter Streaming API to collect tweets (online publications in Twitter) and calculates the frequency of emoji. He introduces a classification of 821 emoji glyphs by using emoji frequency. This classification is processed in realtime. The project increments the count for each emoji glyph contained in the tweet. The issue of the project is the website www.emojitracker.com that monitors the use of emoji in realtime.

3 PROJECT DESCRIPTION

3.1 Architecture

We track tweets with our desired emoji in public twitter streams and write them to mongoDB by a stream parser. We then run mapreduce regularly to update the frequency. Theses databases (tweet database and frequency dataset) are public that everyone can access with read privilege via a RESTful interface. Our web application queries data via the RESTful interface and receives them

in JSON format. This proces is described in figure 1.

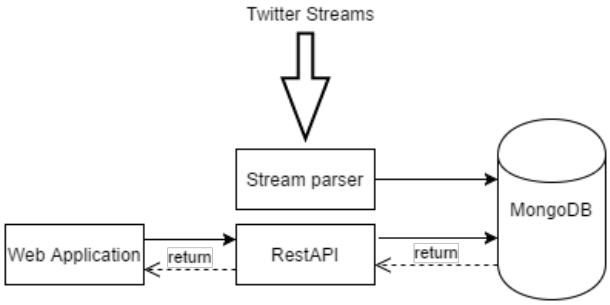

Figure 1: Processing data

3.2 Data Collection

Twitter Streaming APIs allows to create a set of criteria to monitor tweets in realtime, and then Twitter handles the work of pushing updates to developers whenever they occur over a single long-life socket connection. Every time a tweet with at least one of our desired emoji is detected, the stream parser will write it to mongoDB.

Before we go through the data structure, let's introduce some conceptions in MongoDB. MongoDB is a non-relational database management systems. A database is a set of collections. A collection is the unit of storing data in a MongoDB database, analogous to table in RDBMS(Relational Database Management System). Documents use JSON format for storing data. Therefore, we choose to use mongoDB since JSON structures can be transformed easily to JavaScript objects within the browser environment.

The dataset consists of tweets collected from 7 Jan. 2017 at nearly 10 million elements and is enriching everyday. Other reason helps us choosing MongoDB is that its map-reduce framework has good performance in processing large volume of data. Every document in mongoDB contains the tweet text, published time, user name, published location encoded in longitude and latitude, user location written in his profile. The statistics of emoji are calculated and updated regularly by the job map-reduce of mongoDB. Since the results of map-reduce functions are collections, it is accessible by sending http 'GET' requests. Thanks to the RESTful API, the documents are returned in JSON format.

3.3 Data visualization

At the customer'd end, the web application loads JSON format data by AJAX requests. Once the data is loaded, the web application begins to render graphs. The loading and rendering processes are written in javascript with D3.js library. D3.js (D3 for Data-Driven Documents) is a JavaScript library for producing dynamic, interactive data visualizations in web browsers. It makes use of the widely implemented SVG, HTML5, and CSS standards.

In this project, we have two main goals in the visualization: detecting the emotions changes of twitteurs by tracking the frequency of emoji and find the relevance between the emoji by calculating the co-occurrence of the emoji pairs. Thus, we expect that the issues of this works can reach general public as our target audience.

Figure 2: Emoji cloud

First, the most intuitive and effective way to visualize the frequency is to draw a tag cloud (we name it 'emoji-cloud' – figure 2). Tags are single emoji, and the importance of each tag is presented as font size. The more frequent the

emoji is used, the bigger it is rendered. This format is useful for quickly perceiving the most prominent terms and in case that precise results are not tricky required.

Secondly, we would like to build a bar chart figuring the precise number of times that emoji appear and comparing them (see figure 3 – Emoji frequency bar chart). The y-axis represents the total frequencies of emoji and the x-axis is labeled by emoji. Both the tag cloud and the bar chart show total frequencies over a period of time so that the audience could have a global view of the twitteurs' sentiment. However, it doesn't reflect the emotion changes during this period.

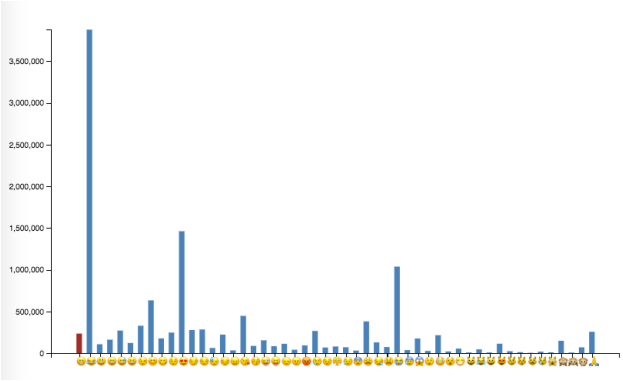

Figure 3: Emoji frequency bar chart

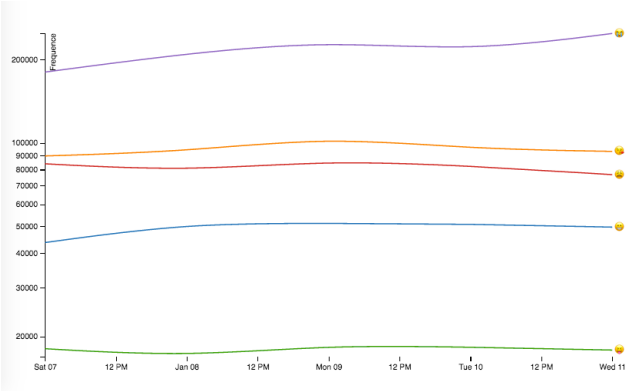

Figure 4: Emoji n-gram

Thirdly, we therefore design a N-gram chart also known as line chart, to visualize the daily frequencies (see figure 4). The N-gram chart is used by Google Books Ngram Viewer, which is an online search engine that analyzes the frequency of word use in their scanned books corpus over time. The y-axis represents the logarithmic scale of the daily frequencies and in x-axis represents dates. A logarithmic scale is a nonlinear scale which is used when there is a large range of quantities. In our case, it avoids lines with small gaps to crowd together in the graph.

Fourthly, we use a force-directed graph and co-occurrence matrix to visualize emoji co-occurrence (see figure 5). In the force-directed graph, the nodes of the graph are emoji, which are links by an edge whose frequency of the emoji pair is greater than a threshold. User can change the value of the threshold by dragging a slider. Meantime, in the co-occurrence matrix, user can use the drop-down menu to reorder the matrix either by emoji or by frequency (see figure 6)

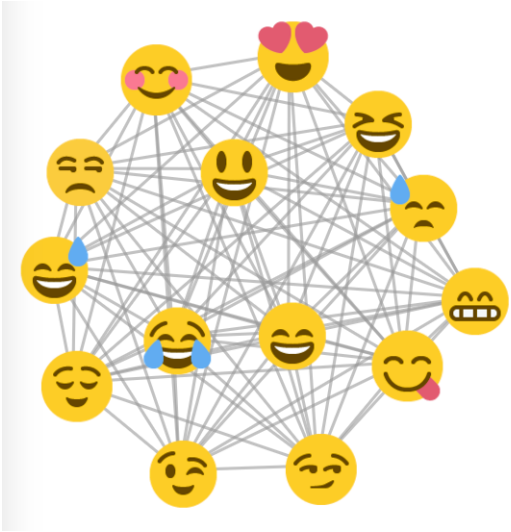

Figure 5: Force-directed graph

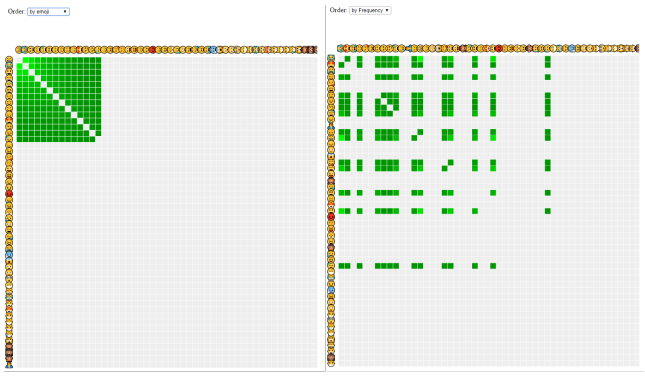

Figure 6: Matrix of emoji co-occurrence

3.4 Interaction technique

To hightlight the frequency of using emoji, it is important to inject and describe the interaction techniques of each graphic and their relation.

Some tools to investigate and modify graph:

Select: Select one or more emoji to explore emoji graphics. If we click on emoji in force-directed graph, all adjacent vertices (its relations) with be highlighted. Meanwhile, the associated bar of selected emoji(s) in bar char will also be hightlighed and line graph will show a line indicating the frequency of this emoji. Clicking again the emoji, vertices will be unhighlighted.

Drag: Use this option to move around one specific emoji (vertex) in force-directed graph.

Change parameter of drop-down menu: reorder the matrix

4 Discussion

Challenges. With the small set of data (data collected within one week) and the limited capacity of storage, we cannot have a fully picture of emoji from all Twitter publication as well as the global view of emotional analysis. Nonetheless, our project sets the first steps to observe the emotions of people all around the world before or after a specific event. It is described by the curves of frequency and the tendancy in combining different emoji.

The high number of emoji et theirs relations with other emoji require high computational power and complicate the rendering of the graph as well as the difficulty in storing database. Therefor, we decided to visualize only 51 emoji of face expressions. The same effet is applied if we select a bar in bar chart.

The possibility to relate 5 graphics and make them synchonize when user clicks on one part graphic increase the complexity of task.

Strength. Our database contains many features for further usage (ex: location, user name) and it is accessible for public.

5 Conclusion

Result. When we count all emoji together – "Face with tears of joy 😂", "smiling face with heart-shaped eyes 😍", "smiling face with smiling eyes 😘", "loudly crying face 😭" and "kissing face 😗" – are mostly used.

We have implemented a web application which collects tweets and visualizes the frequencies of 51 emoji. Data are collected from twitter public streams and stored in mongoDB. In the customer'd end, web application loads data and renders graphs.

We have implemented 5 graphics: emoji cloud, force-directed graph, bar chart, line graph and matrix of emoji co-occurrence.

The twitter users' general emotion all over the world seems to be cheerful because they use more positive emoji than negative ones. Generally, when the frequency of positive emoji increases, that of negative ones decreases. More specifically, the most popular emoji is 😂 (face with tears of joy) used about 3.5 million times in the 10 million tweets. And the most popular emoji pair is 😁😂 (grinning face with smiling eyes and face with tears of joy). The pair of 😂😃 and 😄😃 are in the second and third place with small diffence.

Future direction: From this work, we expect to go further impliment an other function that allows us to analyse the emotion of a specific person during a periode of time (using existing interactiongraphic)

We can envision that future developments of the analysis and visualization of emoji will be:

- Integrated tweets
- Visualized in realtime. Our current visualization allow to rank emoji and figure emoji co-occurence of static database. It would be interesting to visulize these emoji in realtime..

REFERENCES

[1] A. Cairo (2012), *Infographics and Visualizations as Tools For the Mind*, http://www.scribblelive.com/blog/2012/09/05/infographics-and-visualizations-as-tools-for-the-mind/, last visited 08 Jan. 2017

[2] Instagram Engineering, *Emojineering Part 1: Machine Learning for Emoji Trends*, published May 1, 2015, https://engineering.instagram.com/emojineering-part-1-machine-learning-for-emoji-trendsmachine-learning-for-emoji-trends-7f5f9cb979ad#.ug20ffvt3, last visited 10 Jan. 2017

[3] J. Stromberg (2013) *Your Tweets Can Predict When You'll Get the Flu*, www.smithsonianmag.com

[4] M. Rothenberg, *How I Built Emojitracker – Adventures in Unicode, Real-time Streaming, and Media Culture,* https://medium.com/@mroth/how-i-built-emojitracker-179cfd8238ac#.s7vj3dect, last visited 11 Jan. 2017

[5] P. K. Novak, J. Smailović, B. Sluban, I. Mozetič, *Sentiment of Emojis,* published December 7, 2015, http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0144296, last visited 10 Jan. 2017

[6] SwiftKey (2015) *SwiftKey PT. Most-used emoji revealed: Americans love skulls, Brazilians love cats, the French love hearts* http://swiftkey.com/en/blog/americans-love-skulls-brazilians-love-cats-swiftkey-emoji-meanings-report/, last visited 11 Jan. 2017

[7] The Next Web (2014) *"Emoji love: The science behind why we <3 emoticons."*, http://thenextweb.com/dd/2014/03/10/Emoji-love-science-behind/, last visited 2017