東京大学　関本研究室 / Sekimoto Lab. IIS, the University of Tokyo.

Mapping emotional geographical trends based on Japanese Tweets

Mingchen Liu, Yuya Shibuya, Yoshihide Sekimoto

Background

Nowadays, there are about 72% people in Japan are active SNS users. Analyzing the spatio-temporal data of SNS can help us understand how sentiment is related to Japanese user' behaviors. This research will use Natural Language Processing model to analyze the sentiment of Japanese Tweets with geo-tags, then explore the relationship between people behavior and spatial psychology.

Data source

・Tweets dataset
All Japanese geo-tagged Tweets in Japan, 2021
70,266,303 Tweets were collected in total.
・Pre-trained Japanese BERT model and tokenizer developed by Tohoku University.
・Tel Point data, 2021
・City boundary shape file, ESRI Japan, 2021

Methodology

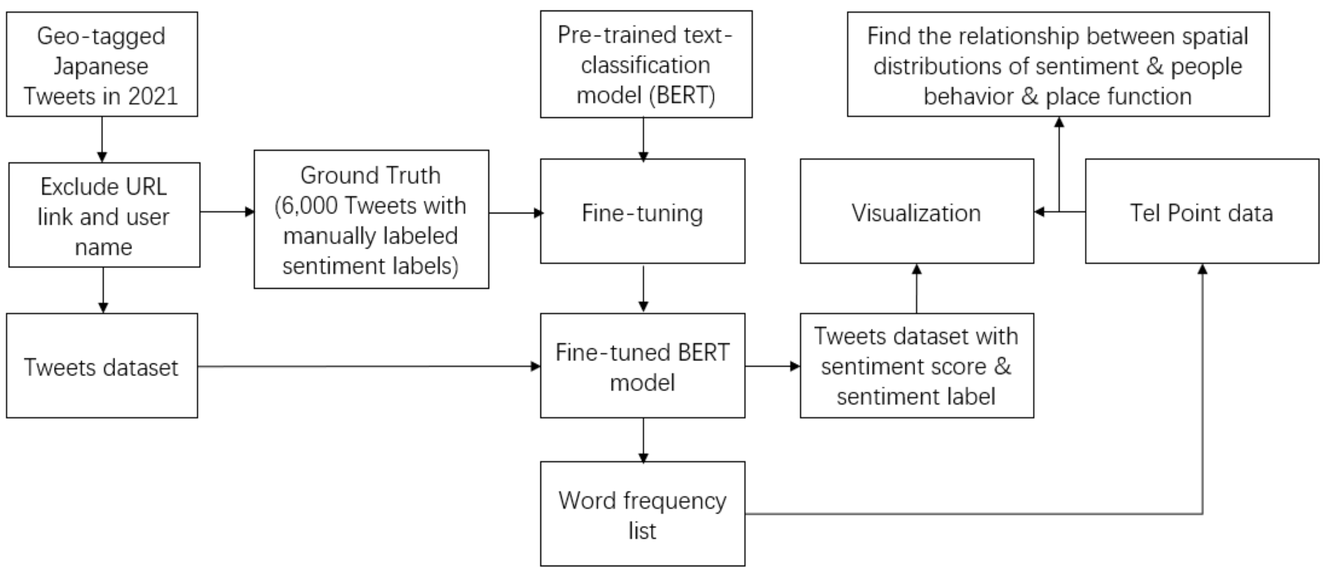

Fig.1. Research Outline

After data pre-processing, we chose BERT (Bidirectional Encoder Representations from Transformers), an unsupervised deep learning model to analyze the sentiment of Tweets. We randomly selected 6,000 Tweets from the whole dataset as the ground truth and labeled their sentiment manually. To improve the accuracy of BERT model, we then fine-tuned this model with ground truth.
Combining the Tel Point data with our Tweets data, then we can get the function of the POI location where users tagged. Then we can do the spatial analysis and examine the relationship between place function & sentiment & behavior.

Result & conclusion

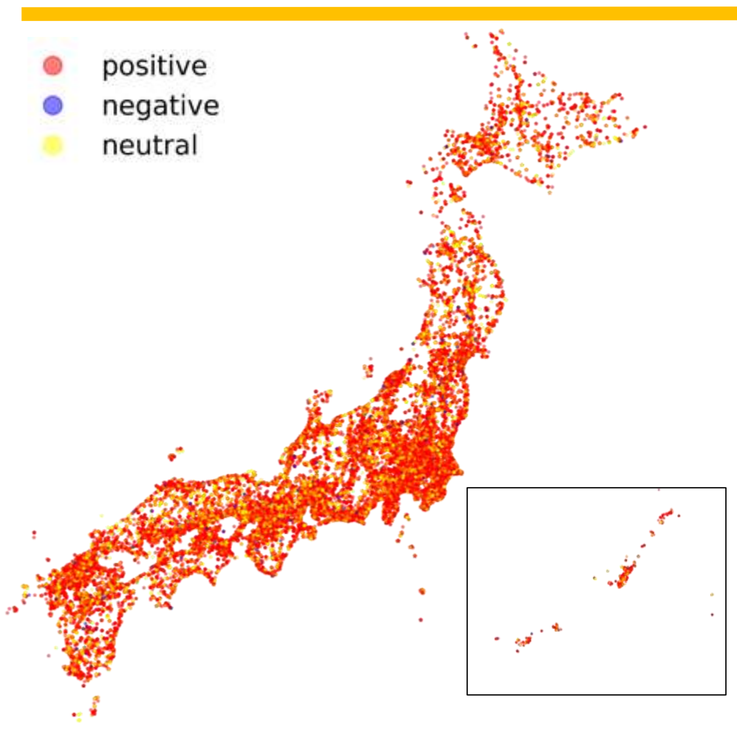

Fig.2. Sentiment color map

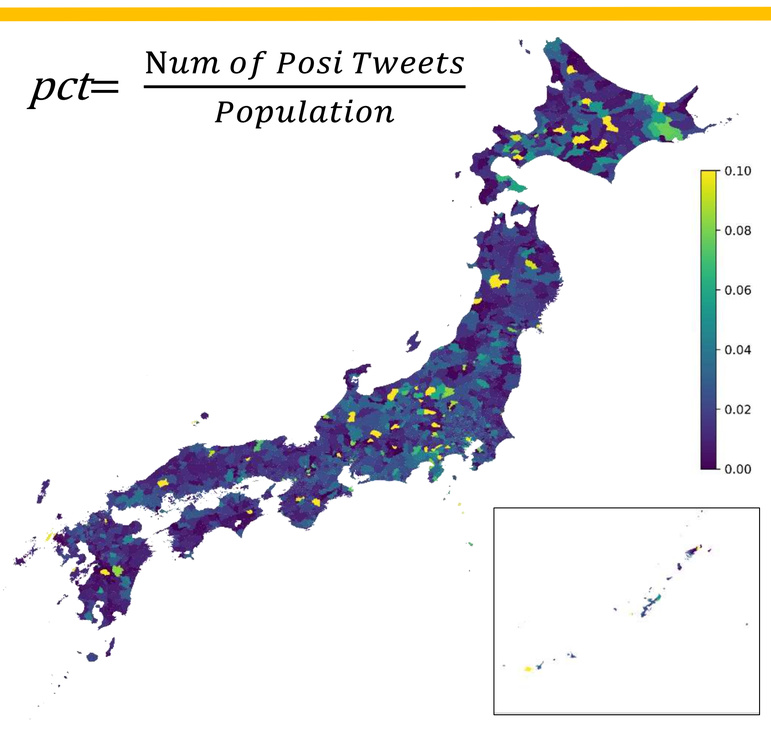

Fig.3. Percentage of positive Tweets per population

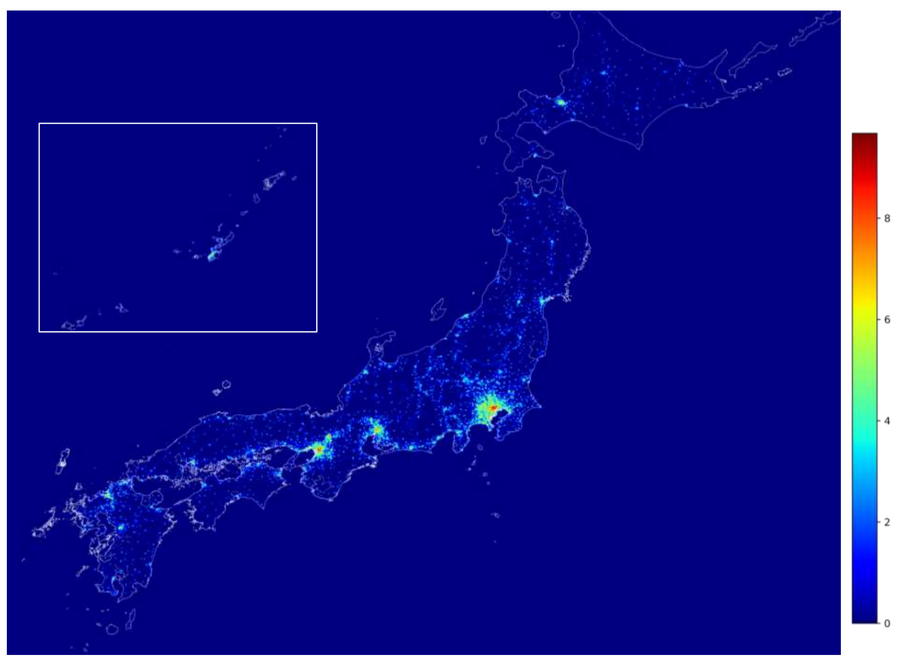

Fig.4. 2D Heat map of Positive Tweets

We used our model to analyze the Tweets dataset and got the sentiment score of each Tweet. If the score of the label is greater than 95, it is classified as the same label, otherwise it will be classified as a neutral label.
Then we visualized the sentiment scores. Most Tweets in our dataset are positive and distributed in densely populated cities. From Fig.3., it can be seen that the city with the most positive per capita is not the city with the largest population, but rather a tourist city with a smaller fixed population (The most positive city is 熊本県球磨郡五木村, with a percentage of 129.43%).
To find the specific emotional geographical trends in Japan, we still need to do the case study of a specific city and narrow the scope of it, and then the relationship between sentiment and people behavior can be found.