Neighborhood's Sentiment Cloud:
Visualization of leisure through text analysis

○ Danilo Eidy Miura[*1] and Shunsuke Miyahara[*2]

*1 Research Assistant, School of Engineering, University of Tokyo, M.Eng.
*2 Lecturer, College of Industrial Technology, Nihon University, Dr.Eng.

Keywords: Urban planning, Urban leisure, natural language processing, data visualization.

1. Introduction

In Urban planning, the experience of the future users and dwellers has been noticed as an important dimension for the attractiveness of a certain area. Factors such as Sensorial experience, leisure, quality of life indicates how users perceive their experience.
A problem for urban planners is the access to the subjective experience of users. But with the advent of Big Data and AI technologies, it becomes possible to assess the experiences through the history of users in a certain area.
This paper presents an application o sentiment analysis of users' reviews around the 5 busiest stations in the world.

2. Related Work

Previous studies show the relevance of leisure in urban areas to quality of life and sensorial experience of in a certain area [1]. In addition, it has its relevance to regional tourism and urban planning. Therefore, leisure constitutes one of the relevant factors for the urban regeneration.
Taking advantage of the grown Big Data, previous research explored the Sentiment Analysis of data from users of social networks, showing the potential use of Natural Language Processing to forecast the overall sentiment of the urban area [3,4,5,6].

2.2. Natural Language Processing and Data Visualization

Sentiment Analysis is a subfield of Natural Language Processing which aims to recognize sentiment, feelings and opinion from natural language text [7]. Sentiment Analysis has been developed in several different approaches [8,9] with shared techniques with text classification. In this current study, we approach the sentiment of users by the simple rule-based analysis, in which, basically, adjectives are classified by being positive, neutral or negative.

3. Research Method

This study aimed to demonstrate the analysis of sentiment of users of certain areas expressed on online service where users can post their reviews. After analysis of the textual data, the results are shown in visualization in form of heat map, which provides special sense on geographic map and high resolution of the sentiment, using heat marks.

3.1. Scope

We define leisure as the activity out of work, job hunting or housing activities. In this study, leisure in consideration are the activities at places around stations that users of Google Maps posted their reviews. The geographic locations were defined as the regions around the busiest stations of the world [11], Shinjuku, Shibuya, Zurich Hauptbahnhof, Gare du Nord (Paris), and Pennsylvania (New York).

3.2. Data Collection

In order to investigate the leisure activity in a certain area, we aimed the reviews data from users about places provided online. The data were collected through Google Maps API, which is an online service that gives access to data of Google Maps and reviews of places.
Places in the area of 2-kilometers radius around the busiest stations were identified by their names, coordinates. And for each place, we extracted total number of reviews and the contents of the latest 5 reviews.

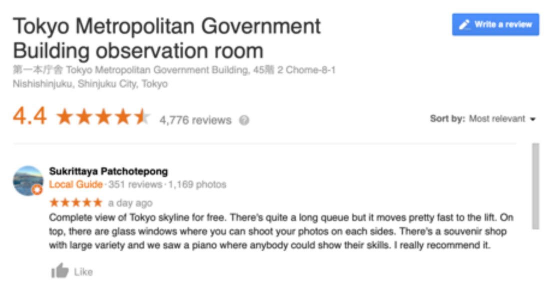


Figure 1 – Example of User's review

3.3. Procedures of Sentiment Analysis

We used the lexicon-based method to analyze the sentiment of the reviews. In this method, the text is split into words and each word is searched in a dictionary that classifies the word as positive, negative or neutral.
Normally, in the second step, the sentiment of the full text is calculated according to the sentiment of the words in the text. But instead of overall calculation, we kept the list of positive and negative words. We use the bags-of-words to make interpretation in a qualitative analysis.
With the purpose of qualifying the particular sentiment that reviewers have about each station, we divided the words according to their representativeness to each station. It means a word is assigned to a station if this word has higher relative frequency to this station than the others.

3.4. Data Visualization

Collected data were visualized in two different graphs. The number of reviews is visualized in Heat Map in geographic map to provide the sense of concentration of leisure in the area around the stations. And the words with sentimental change were visualized in Word Clouds, where the size of words represent their frequency, the larger it is more frequent in reviews.

日本建築学会情報システム技術委員会
第42回情報･システム･利用･技術シンポジウム論文集，78-81，2019年12月，東京
Proceedings of the 42nd Symposium on Computer Technology of Information，
Systems and Applications，AIJ，78-81，Dec.，2019，Tokyo

4. Results

4.1. Collected Data

The data shows a large number of places around the busiest stations in the radius of 2-kilometers, with the average of about 226 places per station, and the average of total number of reviews per station is about 100,600, with a discrepancy of the number of reviews around the Pennsylvania station, New York.

Table 1 - Total number of collected data per station

Station	No. of Places	Total no. of reviews
Shinjuku	194	53,761
Shibuya	249	59,338
Zurich	221	56,286
Gare du Nord	233	61,601
Pennsylvania	234	274,774

This discrepancy was treated before the visualization, defining parameters to allow a "fair" comparison of the resulted visualization.

4.2. Distribution of types of places

The reviewed places were classified in 5 categories: food, shopping, lodging, service, and others. Google Maps API provides the types of each place registered and it could be automatically assigned to the place.

Food: places that primarily offers food or drink as service.
Shopping: Commercial establishments.
Lodging: Hotels, hostels, and rental houses.
Service: Services except food and lodging.
Other: Any other than previous and unclassified places.

Table 2 – Distribution of types of places per area

	Shinjuku	Shibuya	Zurich	Paris	New York
food	113	114	37	30	21
shopping	27	87	84	30	27
lodging	23	6	32	119	60
service	14	17	31	29	55
other	15	25	28	22	67

We can consider food and shopping as main categories of leisure, while lodging just indicates travelers staying in the area, and service may not be considered as leisure activity, but needed service.

The representativeness of the distribution of types of places is shown in Figure 2. And the comparison reveals dominance of food, shopping and lodging in each area.
Shinjuku: Food dominant
Shibuya: Hybrid Food-Shopping dominant
Zurich: Shopping dominant
Gare du Nord: Lodging dominant
Pennsylvania: Lodging dominant (not considering other)

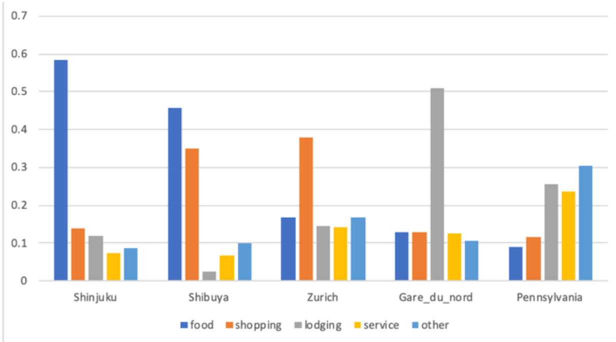


Figure 2 – Comparative representativeness of types of places

The dominance of the type of place in a certain area indicates what kind of review we expect to find more and the main purpose of the leisure activity. A possible explanation to the difference of representativeness of lodging type reviews is the predominance of leisure activity of tourists or locals.

4.3. Data parameterization

Due to the discrepancy of the number of reviews of around Pennsylvania station, which has about 5 times the average of the others, it was decided to parameterize with a higher parameter, defined by a scale of 20 points between minimum and maximum value of number of reviews of places around each station. Doing so, each visualization shows the maximum level of heat at 20 degrees and minimum of 1 degree. This treatment allows us to see the concentration of the leisure activity around the stations with the same parameter of colors of heat.
The higher parameters defined for the Heat Map visualization allow us to see the area of the 5 stations with the same color parameter but it limits the sense of accumulated reviews for comparison. In addition, Google Maps API provides the total number of reviews per place in the history without providing the period that that place is registered in the database. That may explain the discrepancy of number of reviews in New York, which may have been registered for longer time than the other stations, accumulating more reviews.

4.4. Positive and negative words

The words from the reviews were collected and classified as positive and negative sentiment charge. Neutral words were ignored.

Table 3 – Top 10 words with sentiment

Good words (frequency)			Bad words (frequency)	
Good	791		Bad	100
Great	548		Pay	76
Nice	504		Worst	50
Friendly	362		Hard	48
Like	317		Rude	46
Clean	313		Leave	45
Best	212		Problem	39
Recommend	197		Avoid	37
Well	175		Unfortunately	33
Helpful	167		Noisy	32

Table 3 shows the overall counting of good and bad words of all areas together. Positive words have superior frequency and variety of words (831 different positive words vs 621 different negative words).

5. Discussions

5.1. Limited Data

Google Maps API presented two limitations to the data collection. One is the number of places retrieved per query and the other is limited reviews per place.
Each query of places in Google Maps API returns the maximum of 60 places. To cope with it, the data was collected in a series of queries centered on the station with increasing radius. Starting the first query with 100-meter radius and ending with a query with 2000 meters radius. It implies in a higher chance to retrieve places closer to the station.
Each query of reviews' contents in Google Maps API returns only the last 5 reviews. It implies in a limited textual data that may not properly represent the common opinions of users of places around the stations in a long term and may change from time to time.

5.3. Comparison of leisure activity around the stations

The Heat Map shows the concentration of reviews around the stations.
We can make the observation of concentration-dispersion of review places around stations and compared than with the shape of the roads. We can infer the how the proportion of four-way intersections may interfere in the concentration of leisure activity.

Figure 3 – Heat Map of Shinjuku and New York (concentration)

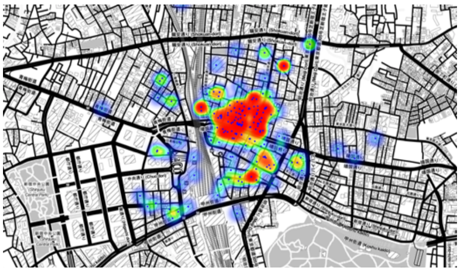
Observing the example of Shinjuku, where there is a large variety of intersections including dead ends, we see an intense concentration of reviewed places close to the station. In contrast, planned area around Pennsylvania shows relative dispersion of reviewed places.

5.4. Particularity of sentiments about the stations

The word clouds show the particular good and bad words for each station. These words reveal some of the relevant points for each area. The points may refer to some issues or specialty of the places which is reinforced by the common antonyms present in the pair of good-bad words. For example, issues such as better-worst in Shinjuku, recommended-disappointed in Shibuya, friendly-unfriendly in Zurich, clean-dirty in Paris, great-terrible in New York.

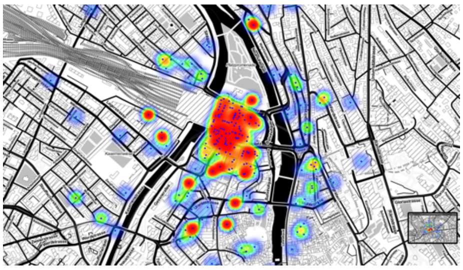
References

1) Lloyd, K., & Auld, C. (2003). Leisure, public space and quality of life in the urban environment. *Urban policy and research*, *21*(4), 339-356.
2) Lian, H., & Yue, P. (2019, August). A Review on Leisure City Evaluation Index Systems and Model. In *2019 5th International Conference on Social Science and Higher Education (ICSSHE 2019)*. Atlantis Press.
3) Weerdenburg, D. V. (2017). *Web Data and Urban Space. Investigating Urban Leisure Activities using Web based Semantic Enrichment* (Master's thesis).
4) SentiHood Targeted Aspect Based Sentiment Analysis Dataset for Urban Neighbourhoods
5) Kovacs-Gyori, A., Ristea, A., Havas, C., Resch, B., & Cabrera-Barona, P. (2018). # London2012: Towards Citizen-Contributed Urban Planning Through Sentiment Analysis of Twitter Data. *Urban Planning*, *3*(1), 75-99.
6) Bertrand, K. Z., Bialik, M., Virdee, K., Gros, A., & Bar-Yam, Y. (2013). Sentiment in new york city: A high resolution spatial and temporal view. *arXiv preprint arXiv:1308.5010*.
7) Mäntylä, M. V., Graziotin, D., & Kuutila, M. (2018). The evolution of sentiment analysis—A review of research topics, venues, and top cited papers. *Computer Science Review*, *27*, 16-32.
8) Mirończuk, M. M., & Protasiewicz, J. (2018). A recent overview of the state-of-the-art elements of text classification. *Expert Systems with Applications*, *106*, 36-54.
9) Araque, O., Corcuera-Platas, I., Sanchez-Rada, J. F., & Iglesias, C. A. (2017). Enhancing deep learning sentiment analysis with ensemble techniques in social applications. *Expert Systems with Applications*, *77*, 236-246.
10) Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams engineering journal*, *5*(4), 1093-1113.
11) 5 of the Busiest Train Stations in the World. Retrieved September, 1st, 2019 from https://www.triptrivia.com/5-of-the-busiest-train-stations-in-the-world/



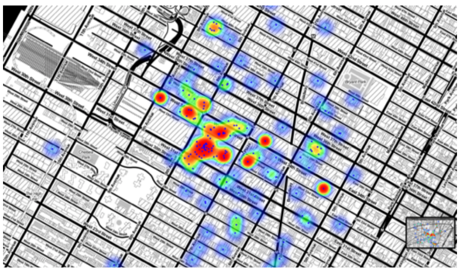
Shinjuku

Shibuya



Zurich

Gare du Nord



Pennsylvania

4.3. Compared results (Heatmap)

Shinjuku

Shibuya

Zurich

Gare du Nord

Pennsylvania

4.4. Compared Results (Word Cloud)

ANNEX – COMPARISON OF VISUALIZATIONS OF HEATMAPS AND WORDCLOUDS OF AREAS