Region-to-Region Similarity Analysis based on Foursquare Venue Database

Akira Ito[1] Yutaka Arakawa[2] Hirohiko Suwa[2] Akira Fukuda[1]

[1]Graduate School / Faculty of Information Science and Electrical Engineering, Kyushu University, Japan
[2]Nara Institute of Science and Technology, NAIST, Japan

1. INTRODUCTION

Value assessment of the place/region is important in a real estate market. Among various real estate, our target is a place for some shops and restaurants. These are affected by the dweller, visitors and flow of them. In other words, if we can find a new place where a similar people are living or visiting, the place can be recommended as a candidate of a new shop. Therefore, a real estate company have been tried to find it manually. But it depended on the personal experience and skills.

The method for regional analysis can be categorized into two types, one is crowdsourcing[1] and the other is using social data or open data[2]. Compared with the crowdsourcing approach, social data based approach is more scalable and adaptable for various regions. Therefore, we focus on the social data based approach and propose a region-to-region similarity analysis based on Foursquare venue database.

Through the experimental evaluation about 29 stations of Yamanote-line in Tokyo, we show that our system can measure the similarity of a certain region and can find out a similar region.

2. ANALYSIS

In this paper, region is the area represented by Geographic coordinates of (*latitude*, *longitude*) and radius of r[m].

2.1 Data

Venue data obtained from Foursquare Search API by giving latitude, longitude, and radius. Among the various response, we collect the total number of venues within a certain region (NOV), the number of venues included in each categories, and the total check in count in each categories. In this study, we use 2nd level categories among the hierarchical three levels of Foursquare venue database.

2.2 Method

We define the n-dimensional vector $\mathbf{c} = (c_1, c_2, ..., c_n)$, where c_i is the number of venues of category i within region r, and n is the total number of all the 2nd level categories.

Ratio-Cosine Method (RCM): First, normalizing vector $\mathbf{c}$ as $\frac{\mathbf{c}}{m}$. Then, calculating the similarity between regions based on normalized vectors using the cosine similarity method.

TFIDF-Cosine Method (TCM): When there are m regions to be analyzed, we calculate TF-IDF for each category in vector $\mathbf{c_1}, \mathbf{c_2}, ..., \mathbf{c_m}$. Then, calculating the similarity in the same way as RCM.

Check-in Ranking Method (CRM): First, ranking the number of check-in for each category in region r. Then, using check-in ranking top-k categories, we calculate the similarity in the same way as RCM or TCM.

3. EXPERIMENTS

We collected Foursquare venues from 29 stations of Yamanote-line in Tokyo, Japan. Venue data was collected up to November 20, 2014 using the Venue Search API of Foursquare. Our data set consisted of 8,505 venues and 436 categories across these stations. In this section, we show the assigned data to the API about 4 stations in Table 1. We also quantified the similarity between 29 stations. Table 2 shows the result of TCM for them. Looking at Table 2, you can figure out Shinjuku and Shibuya are similar, but, Shinjuku and Kanda are not similar. These results are subjectively true from the view point of the person who knows these stations.

Table 1: Assigned Data about 4 Stations

	Latitude	Longitude	Radius	NOV
Shinjuku	35.690833	139.700278	100	542
Shibuya	35.658611	139.701111		539
Shinbashi	35.666389	139.758056		603
Kanda	35.691667	139.770833		421

Table 2: Station Similarity by TCM

	Shinjuku	Shibuya	Shinbashi	Kanda
Shinjuku	-	**0.8132**	0.5618	0.3915
Shibuya	**0.8132**	-	0.6336	0.4502
Shinbashi	0.5618	0.6336	-	**0.8237**
Kanda	0.3915	0.4502	**0.8237**	-

4. ACKNOWLEDGMENTS

A part of this work was supported by SCOPE (Strategic Information and Communications R&D Promotion Programme).

5. REFERENCES

[1] Y.Chon, N.D.Lane, F.Li, H.Cha, and F.Zhao., Automatically Characterizing Places with Opportunistic Crowdsensing using Smartphones, In UbiComp'12, Pittsburgh, PA.s

[2] D.Preotiuc-Pietro, J.Cranshaw, and T.Yano, Exploring venue-based city-to-city similarity measures, In Proceedings of the 2nd ACM SIGKDD International Workshop on Urban Computing, ACM, No.16, 2013