Analysis of Activity Trends Based on Smart Card Data of Public Transportation

T. N. Maeda*[1], J. Mori[1], F. Toriumi[1], H. Ohashi[1]

[1]The University of Tokyo, 7-3-1 Hongo Bunkyo-ku, Tokyo, Japan
*Email: maeda@crimson.q.t.u-tokyo.ac.jp

Abstract

Smart card data of public transportation have no information on the purpose each individual has moved from a place to another place for. In previous studies, land uses are estimated based on time-series distributions of daily populations in each place. However, each place is not used for single kind of activity. It is rather necessary to regard that land use of each place is defined by a mixture of multiple kinds of activities. Our research aims to develop a method to grasp the trends of activities in each place on each day by decomposing time-series distribution of population. The method can be applied for monitoring and detecting changes or anomalies of activities in each place.

Keywords: Smart Card Data, Public Transportation, Land Use, Non-negative Matrix Factorization.

1. Introduction

Events and developments of commercial facilities can rapidly change urban human mobility and land uses. For example, when a commercial facility is opened in a residential area, the area will attract visitors for leisure from other places. It is beneficial for urban planners and commercial developers to grasp the trends in people's activity in each area.

Previous studies analyze what kind activities are popular in each place based on text data of from geolocational social media (Kurashima et al, 2013). Moreover, there are studies that grasp activities of each location based on POI data (Georgiev et al., 2014, Maeda et al. 2016).

Estimation of land uses without using text data and POI data has been conducted based on time-series distribution of population. Previous studies (Frias-Martinez et al, 2014, Nishi et al, 2014) characterize daily land use by analyzing time series distribution of the population on each day in each area. The methods of these studies classifies the land use of each area into a category such as business, nightlife and leisure. There are two problems regarding these studies. The first problem is that each place is used for multiple activities rather than single kind of activity. The second problem is that these methods cannot grasp population trends of each kind of activity. On the other hand, there is research that tries to infer trip purpose by using individual trajectory data (Gong et al., 2016). This method is also very useful to analyze changes of land use, but it is necessary to collect individual mobility data.

Our research aims to develop a method to grasp the trends of activities in each place on each day by decomposing time-series distribution of population. Our research applies non-negative matrix factorization for the smart card data of public transportation in the Kanasai area, Japan. Based on the data, we demonstrate the trends of activities in each place on each day, and we also show changes caused by development of a commercial facility.

2. Method

A time-series distribution of population in a day at each location can be regarded as a mixture of basic distributions, and each basic distribution can be regarded to represent movements of one purpose such as commuting and leisure (Figure. 1). Therefore, our research extracts each basic distribution by using non-negative matrix factorization that is also used for separation of mixed sounds (Virtanen, 2007).

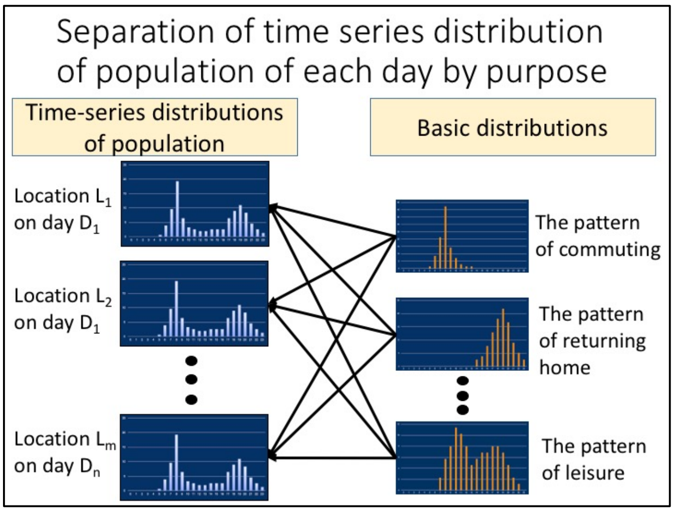

Figure 1. Decomposing time-series distribution of population

Our method firstly creates a matrix. Each row vector of the matrix denotes a time-series population of the people who get off at the station. We let the number of the columns be twenty-four, so that the columns denote the hourly numbers of people who get off the station.

We apply non-negative matrix factorization for this matrix. We can arbitrarily determine the number of basic components. We here define the number as ten.

3. Data

We analyze the smart card data of public transportation in the Kansai Area, Japan. The data elements we used are trip ID, passenger ID, card types, boarding time, alighting time, boarding station, and alighting station. We obtained those data from 6 railway companies. They anonymized those data before they provided those data to us. The period is from April 2013 to March 2014. The data contain about 970 million trips. Figure. 2 shows the locations of stations.

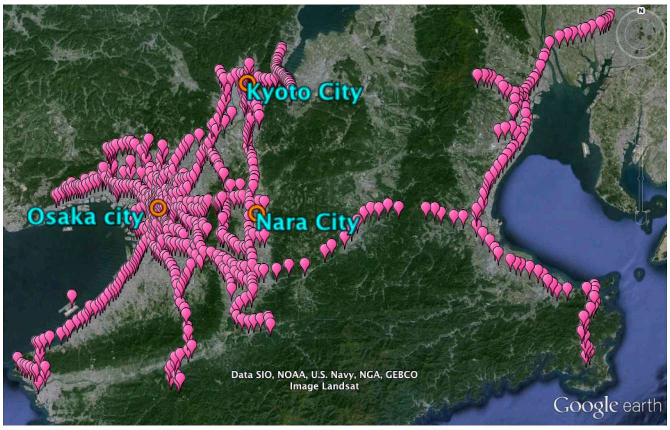

Figure 2. The location of the stations.

4. Results

Firstly, we show what kind of basic components we have obtained. We manually estimate the activity of each basic component based on the time-series distribution of the component. The activities of the basic components are as below:

1. Commuting
2. Going back home (Monday to Thursday)
3. Going back home (Friday)
4. Leisure after work
5. Events such as sport games and concerts 1
6. Events such as sport games and concerts 2
7. Events such as sport games and concerts 3
8. Leisure on a day without work 1
9. Leisure on a day without work 2
10. (Uncertain)

Next, we show what kind of activities are popular in places as examples. Figure. 3 shows the population trends by type of activity at a station. As shown in this figure, the number of commuters increased just before the opening date of a shopping mall, and it decreased after the date. The rapid increase is explained by staffs for the opening of the mall. On the other hand, the trend of leisure increased on the opening date of the shopping mall, and it keeps remaining flat.

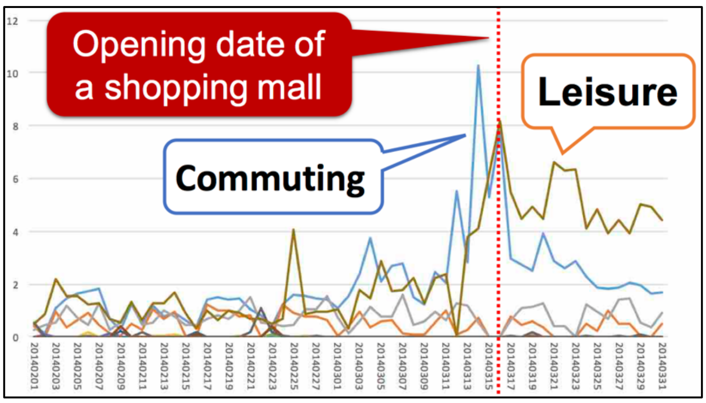

Figure 3. Population trend of each activity (a station near shopping mall)

Figure. 4 shows the population trends by type of activity at another station in business area. As this figure shows, the number of commuting only decreases during summer holiday period.

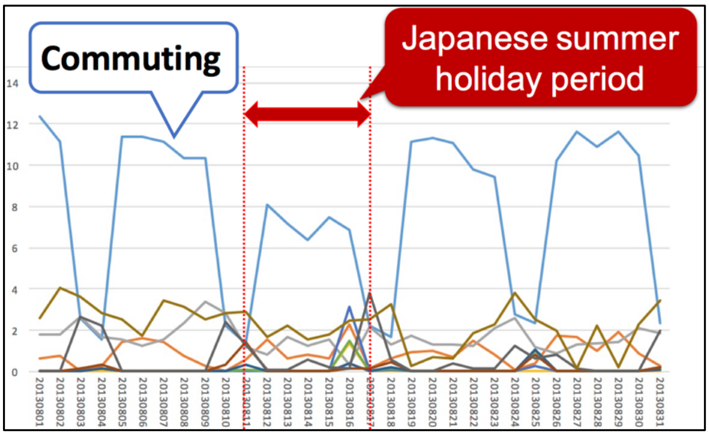

Figure 4. Population trend of each activity (business area)

Finally, Figure. 5 shows the population trends by type of activity at another station in residential area. There can be seen that the numbers of people going back home is the highest. The reason why the component of Monday to Thursday differs from that of Friday is that working people usually go back home late on Friday.

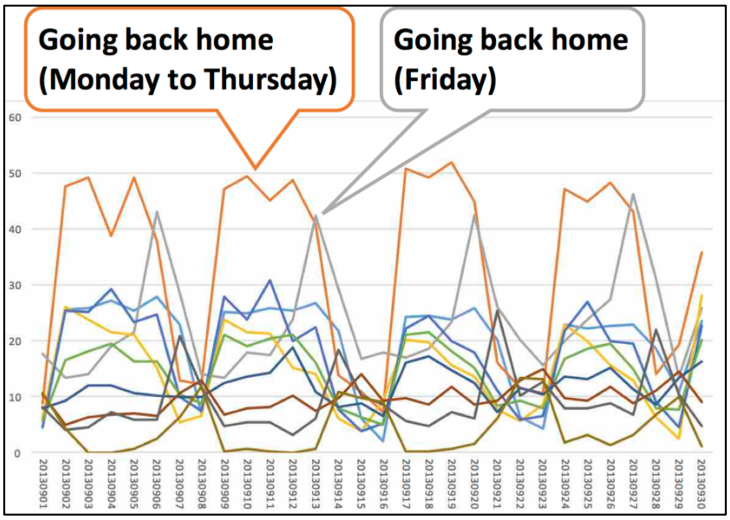

Figure 5. Population trend of each activity (residencial area)

5. Evaluation

We evaluate the outcome by using survey data conducted by Ministry of Land, Infrastructure, Transport and Tourism in Japan[1]. The data includes information about purposes, exit stations, and number of passengers in 2010. We compare the correlation between the survey data and the outcome of our method applied for the data in April, May and June in 2014, in regard to the purposes of commuting, leisure, and going back home.

[1] http://www.kkr.mlit.go.jp/plan/pt/ (Written in Japanese)

Table1 shows the result of evaluation. As shown in this table, the outcome of our method sufficiently reflects the outcome of the survey.

Table 1. Evaluation of our method.

purpose	day	correlation coefficient	p-value
commuting	weekday	0.877540053	5.32E-163
commuting	holiday	0.774460556	6.54E-96
leisure	weekday	0.857119820	6.93E-148
leisure	holiday	0.927275600	1.78E-218
going back home	weekday	0.812499757	3.83E-121
going back home	holiday	0.856006218	5.86E-148

6. Conclusion

This paper has proposed a method to grasp what kind of activity is popular in each place on each day. Further research will be needed to study the following two points. The first one is to study how to decide the most suitable number of basic component. The second one is to study how to detect anomalies and changes in activities in each place.

7. References

Frias-Martinez V, Soto V, Hohwald H and Frias-Martinez E, 2012, “Characterizing Urban Landscapes Using Geolocated Tweets”, International Conference on Privacy, Security, Risk and Trust and 2012 International Confernece on Social Computing

Nishi K, Tsubouchi K, and Shimosaka M, 2014, “Extracting land-use patterns using location data from smartphones”, The First International Conference on IoT in Urban Space

Kurashima T, Iwata T, Hoshide T, Takaya N and Fujimura K, 2013, “Geo topic model: joint modeling of user's activity area and interests for location recommendation”, The Sixth ACM International Conference on Web Search and Data Mining

Georgiev P, Noulas A and Mascolo C, 2014, “Where businesses thrive: Predicting the impact of the olympic games on local retailers through location-based services data”, The 8th International Conference on Weblogs and Social Media

Maeda NT, Yoshida M, Toriumi F and Ohashi H, 2016, “Decision Tree Analysis of Tourists' Preferences Regarding Tourist Attractions Using Geotag Data from Social Media”, The Second International Conference on IoT in Urban Space

Virtanen T, 2007, “Monaural sound source separation by non-negative matrix factorization with temporal continuity and sparseness criteria”, IEEE Transactions on Audio, Speech and Language Processing

Gong L, Liu X, Wu L, and Liu Y, 2016 "Inferring trip purposes and uncovering travel patterns from taxi trajectory data", Cartography and Geographic Information Science