A New Method of Estimating Locality of Industry Cluster Regions Using Large-scale Business Transaction Data

Yuki Akeyama and Yuki Akiyama and Ryosuke Shibasaki

Abstract

In an industry cluster many firms tightly coupled with each other via business transaction, which is important factor of economic activities in local or urban area. Furthermore, the Japanese government tries to understand industrial structure or expanse of transactions of local area for effective revitalization of local area, but previous studies of industrial area is limited to case study for difficulty of collecting firm data. Meanwhile, large-scale transaction data appear and we can extract industrial cluster and comprehensively analyze local area. Our research topic is new definition of local area based on industrial clusters and analysis of local area toward effective revitalization. As a result, we confirmed geospatial clustering extracts certain local areas and analyze industrial structure and locality all over the clusters. This framework of clustering and analyzing is an effective method of analyzing local area for regional revitalization.

Y. Akeyama (✉) • Y. Akiyama • R. Shibasaki
Department of Civil Engineering, University of Tokyo, 7-3-1, Hongo, Bunkyo-ku, Tokyo, Japan
Email: toake555@iis.u-tokyo.ac.jp

Y. Akiyama
Email: aki@ iis.u-tokyo.ac.jp

R. Shibasaki
Email: shiba@csis.u-tokyo.ac.jp

1 Introduction

1.1 Local Revitalization and an Industrial Cluster

One of the important keywords that the present Japanese government holds up is *Local Revitalization*. The prime concept of *Local Revitalization* is decision of the economical policy based on the characteristic of industrial structure of each region in order to stop heavy concentration on Tokyo and a decline of local economy. The Japanese government notes the two points as the important characteristics of industrial structure. One is what kind of industry and how many firms are concentrating in the area, which is an important element. Another is how far firms in the area have business transactions, which is also an important element that the government recently notes. Clarifying this two point against each area, they are able to find the prime or competitive industry and promote revitalization of the promising industry within a limited budget.

There are also many prior studies analyzing the characteristic of industrial structure of the specific area (Yamamura 2003; Kodama 2008; Sasaki 2010). For example, the previous study about manufacture industry in Yamagata Prefecture (Takeda et al. 2008) revealed that the hub-firms trading with many firms have a big influence on the network structure or deciding locations of trading firms.

In the meantime, these studies that analysis of industrial structure need detailed information of each firm in the area. However, the conventional method of collecting firm information such as a questionnaire needs much time or expense. Therefore, analysis of industrial structure is limited to a case study of the specific area and few studies comprehensively analyze all area in Japan.

Moreover, another problem of analysis of industrial structure is how we define each area. The general definition of each area is an administrative area that each government designates. Meanwhile, an industrial cluster is a kind of characteristic area that we should specially note considering revitalization of local economy. An industrial cluster that many firms are located in a small area and tightly coupled with each other via business transaction is one of spatial phenomena. In the industrial cluster firms and people actively trade deals such as the intermediate goods or final goods, therefore we can regard an industrial cluster as an area of local economy. Each industrial cluster also does not necessarily accord with administrative area. Furthermore, it is possible that we are able to reveal new industrial characteristics which have not appeared by analyzing each industrial cluster against all area in Japan. From the above reasons, the new definition of the local area based on industrial clusters is a meaningful way when the government makes effective plan toward local revitalization.

1.2 Usage of large-scale business transaction big data toward research

In recent years, big data recording the details of companies and their transactions over Japan is made available. Plotting firms' spatial information on the map, we can find the area where many firms aggregate, the industrial cluster. We can also analyze the firms network structure by extracting transaction data of each firm. There are several studies of analysis of firms network structure using large-scale transaction data, such as the estimating the effect of the economic shock wave through supply chain at the Great East Japan Earthquake (Carvalho et al. 2014).

1.3 Our study plan

Considering the background, our research topic is the new definition of local area based on industrial clusters and analysis of each local area toward effective revitalization of local area. Firstly, we newly defined "local cluster" as a kind of economic area with the quantitative clustering method. Secondly, in order to understand regional characteristics of each defined local cluster, we evaluated the industrial structure and the spatial expanse of firm transactions that each local cluster includes.

2 Definition of Local Cluster Based on Firms' Spatial Distribution

In this chapter, we describe the abstract of large-scale business transaction data, the method of newly defining local clusters, and the result.

2.1 Abstract of Large-scale Business Transaction Data

The large-scale business transaction data, which we use for our study, is provided by Teikoku Databank Ltd. (TDB). This big data consist of two databases, the transaction database and the firm database. The transaction database record transaction information between two firms that TDB has confirmed per year. The firm database record information per year about all firms whose transactions are recorded in the transaction database. Each database includes details of a firm or a transaction (Table 1). We used the 2008 – 2013 firm database for the definition of local clusters. The 2008 – 2013 firm database records about 1.7 million firms. As a reference, the number of all firms in Japan 2014 is 3.9 million according to The Small and Medium Enterprises Agency. Therefore, the firm database has very high comprehensiveness that has information about 40% of all Japanese firms.

Table 1. Details of the transaction database and the farm database

A kind	Transaction database	Firm database
The number of data (2008 - 2013)	About 22 million transactions	About 1.7 million firms
Detailed information	Owner ID Vender ID Contents of item Estimated sum of turnover And so on	Firm ID Sales The number of workers Address Types of industry The boss age And so on

2.2 Definition of Local Clusters

In this study, firstly we specified the latitude and longitude of each firm of the 2008 – 2013 firm database with geocoding. Secondly we defined local clusters by performing the two-dimensional clustering using the latitude and longitude of each firm. This trial that we extract industrial clusters all over Japan based on the extremely spatial indexes, latitude and longitude, using large-scale business transaction data has no precedent and has novelty.

The methods of clustering we used was K-means (Jain 2010) and K-means++ (Arthur et al. 2007). K-means is a popular method of clustering many samples such as big data because of low calculation cost. K-means++ is a more improved method about selection of initial clusters than K-means. The simple algorithm of K-means and K-means++ is as follows

2.2.1 K-means clustering (Jain 2010)

1. Select an initial partition with K clusters; repeat step 2 and 3 until cluster membership stabilizes.
2. Generate a new partition by assigning each pattern to its closest cluster centers.
3. Compute new cluster centers.

2.2.2 K-means++ clustering (Arthur et al. 2007)

We propose a specific way of choosing centers for the k-means algorithm. In particular, let $D(x)$ denote the shortest distance from a data point to the closest center we have already chosen. Then, we define the following algorithm, which we call k-means++.

1. Take one center c_1, chosen uniformly at random from χ .
2. Take a new center c_i, choosing $x \in \chi$ with probability $\frac{D(x)^2}{\sum_{x \in \chi} D(x)^2}$.
3. Repeat Step 1. Until we have taken k centers altogether.
4. Proceed as with the standard k-means algorithm.

We performed the geospatial clustering using K-means and K-means++, and compared the difference between their results. The number of clusters was conveniently set to 400.

2.3 Results of Geospatial Clustering

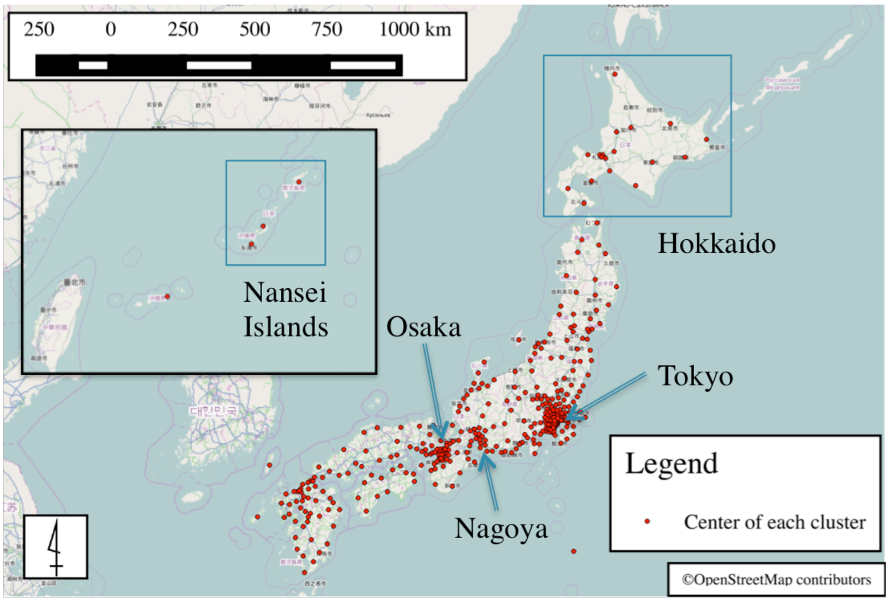

Fig. 1. The result of K-means clustering

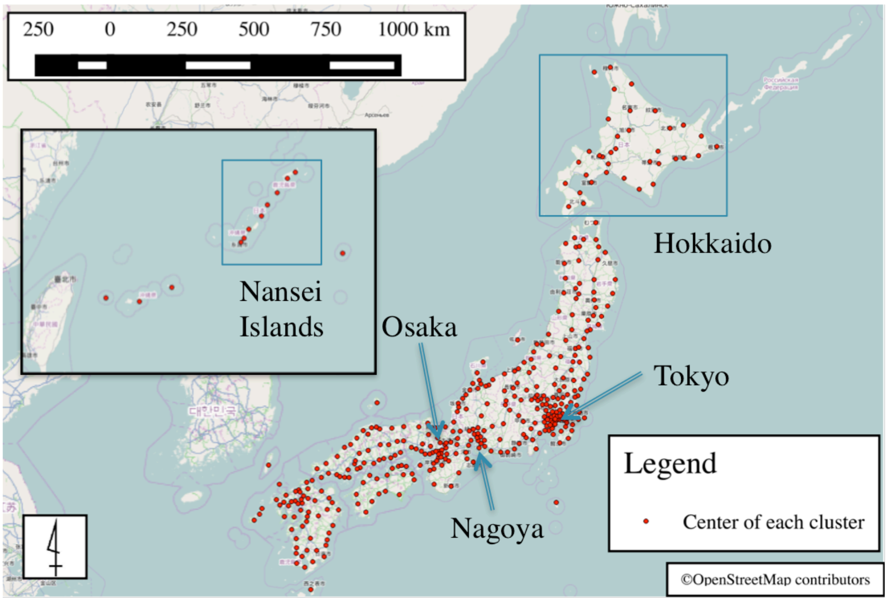

Fig. 2. The result of K-means++ clustering

Fig. 1 is visualization of Local clusters defined with K-means clustering (K-means clusters) and Fig. 2 is visualization of Local clusters defined with K-means++ clustering (K-means++ clusters).
Both figures show K-means clusters and K-means++ clusters aggregated in several large urban areas such as Tokyo, Osaka, or Nagoya. On the other hand, both figures show also there are even many local clusters in local area. Therefore, we recognize that the method of definition of local area with geospatial clustering is a certain persuasive method.

Moreover, K-means++ clusters more distributed all over Japan such as local area in Hokkaido or Nansei Islands than K-means cluster, and the sum of distance between the cluster's center and each sample of K-means++ clusters was lower than K-means clusters. The lower the value is, the more samples of each cluster aggregate together, therefore we conclude that K-means++ clustering is a more appropriate method than K-means clustering considering the concept that extracting industrial clusters.

3. Analysis of Characteristics of Local Clusters

At this chapter, we analyzed various characteristics of 400 local clusters defined K-means++ clustering in Chap. 3. As we have noted in Chap. 1, there are two prime points that characterize local industry. One point is characteristic of industrial structure that what industry aggregates there and another point is spatial characteristics of transactions that how far firms trade with their partners. Therefore, we defined the index explaining the industrial structure and the index explaining spatial condition of transactions against each local cluster.

3.1 Analysis of Industrial Structure of Local Clusters

A previous study has developed a method evaluating the industrial structure of each prefecture in Japan based on the Census of Manufactures (Gonda et al. 2001). The index of the industrial structure is defined as follows.

The index of Conversion of Regional Industrial Structure (*ICRIS*) for region *i* is:

$$ICRIS = \frac{1}{2}\sum_{r=1}^{R}\left|\frac{Ari}{Ani} - \frac{Ar}{An}\right| ,$$

Where summation is taken over all industry *R*.
Ari = number of firms (or employees, etc.) in industry *r* of region *i*.
Ar = total number of firms (or employees, etc.) in industry *r*.
Ani = total number of firms (or employees, etc.) from all industry *R* in region *i*.
An = total population of firms (or employees, etc.) in all regions.

ICRIS indicates how far the industrial structure of each region differs from the whole industrial structure. *ICRIS* value is from 0 to 1. The lower *ICRIS* is, the nearer the industry structure of the area approximates to the whole industrial structure. On the other hand, the higher *ICRIS* is, the more the industrial structure of the area differs from the whole industrial structure. Value *Ari* / *Ani* also indicates the concentrating condition of each industry and region.

We calculated *ICRIS* of employees against 89 types of industry that TDB originally classified as middle classification and 400 local clusters we defined.

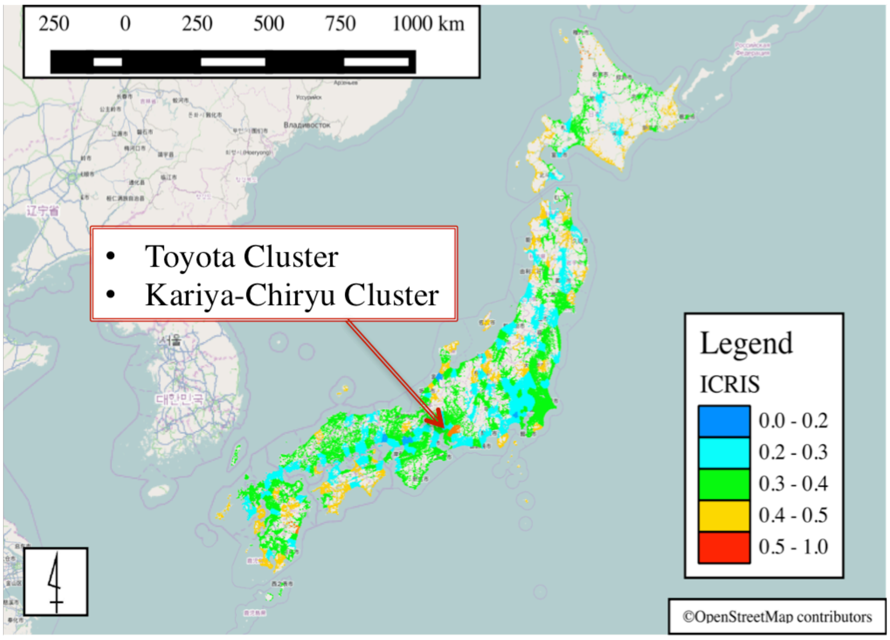

Fig. 3. Visualization of ICRIS of the number of employees

Fig. 3 is visualization of *ICRIS* of each local cluster. Cluster No. 343 and Cluster No. 371 specially took high *ICRIS* value. Cluster No. 343 is located on Toyota area and Cluster No. 371 is located on Kariya and Chiryu area, therefore we called Cluster No. 343 *Toyota Cluster* and Cluster No. 371 *Kariya-Chiryu Cluster*, and focused on the industrial structure of two clusters.

Table 2 and Table 3 are lists of higher *Ari* / *Ani* of Toyota Cluster and Kariya-Chiryu Cluster. Both tables show that transportation machinery industry extremely aggregates in the two clusters that from 40% to 60% of all employees in Toyota Cluster or Kariya-Chiryu Cluster work at firms of transportation machinery. Prime factories of Toyota Motor, a one of most famous firms in the world, are located in Toyota Cluster and Kariya-Chiryu Cluster, therefore it seem that Large-scale firms in the local cluster decide the industrial structure.

Table 2. industrial structure of *Toyota Cluster*

Rank	Type of industry (Middle classification)	*Ari / Ani*	*Ar / An* Whole average
1	Transportation machinery	0.556	0.027
2	General machinery	0.051	0.033
3	Professional service	0.049	0.021
4	Fabricated metal products	0.038	0.019
5	Road freight transport	0.033	0.044

Table 3. the industrial structure of *Kariya-Chiryu Cluster*

Rank	Type of industry (Middle classification)	*Ari / Ani*	*Ar / An* Whole average
1	Transportation machinery	0.388	0.027
2	Electrical machinery	0.180	0.041
3	General machinery	0.057	0.033
4	Road freight transport	0.042	0.044
5	Fabricated metal products	0.037	0.019

3.2 Analysis of Spatial Transactions Condition in Local Clusters

At this chapter we defined *Locality* as an index of spatial transactions condition in local clusters. *Locality* was defined as follow.

$$L_i = N_{ii} \Big/ \sum_j N_{ij} \qquad (1)$$

L_i : Locality of cluster *i*

N_{ij} : The number of transactions from cluster *i* to cluster *j*

Locality is the proportion of transaction of which owner and vender is located in same cluster to all transactions in the local cluster. The higher Locality of the cluster is, probably the more independent economic activities such as local-oriented retail or service industry are made in the cluster. On the other hand, local clusters with low Locality probably include global-oriented large firms or trade with other specific clusters. We calculated Locality of each local cluster using the 2013 year transaction database.

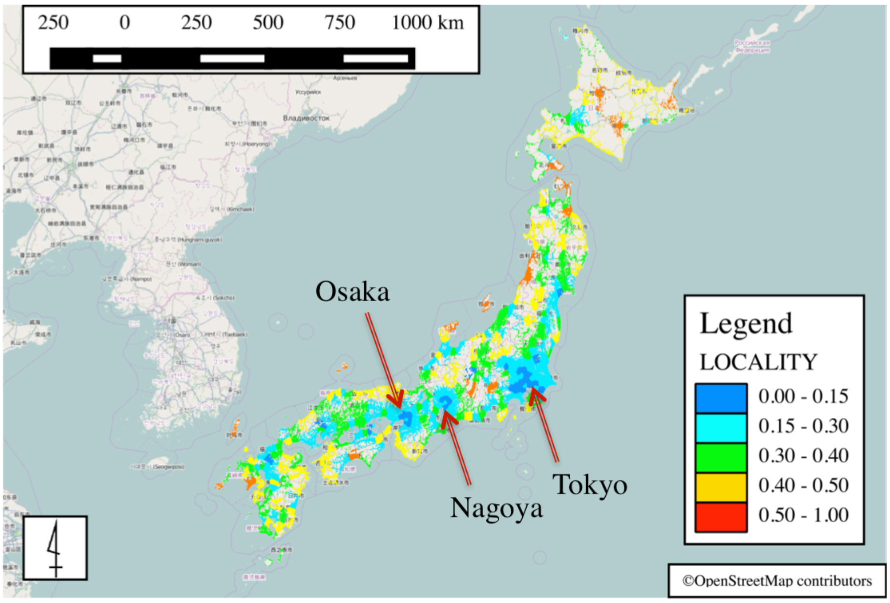

Fig. 4. Visualization of Locality all over clusters

Fig. 4 is visualization of Locality. Fig. 4 shows that geographically isolated clusters such as small islands, the tip of the peninsula, or the remote rural clusters from large urban area tend to take high Locality. On the other hand, the clusters around large urban area such as Tokyo, Osaka, or Nagoya tend to take low Locality. In order to understand this phenomenon that low Locality around big cities, we analyzed spatial transaction condition of clusters around Tokyo.

We defined *Dependency* as following.

$$D_{ij} = N_{ij} \Big/ \sum_j N_{ij} \qquad (2)$$

D_{ij} : Dependency from cluster i to cluster j

N_{ij} : The number of transactions from cluster i to cluster j

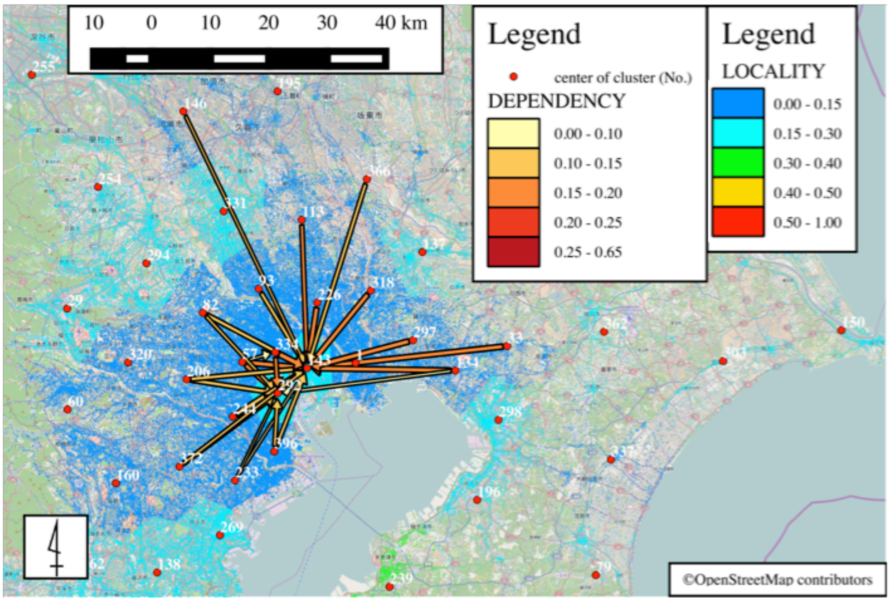

Fig. 5. Visualization of Locality and arrows of Dependency around Tokyo area

Fig. 5 is visualization of combination as transactions between two clusters. If Dependency was higher than Locality of each cluster, we depicted arrows from the vender cluster to the owner cluster on the map. Fig. 5 shows that the Tokyo central clusters such as Cluster No. 143 or Cluster No. 292 pull many transactions from around clusters and intermediate goods concentrate from around clusters on central clusters. This analysis reveals transaction communication between clusters, and probably gives knowledge that what cluster is best cooperative with each cluster.

4. Conclusion

In this paper, we focused on local revitalization which is one of the important tasks in Japan and considered that it is the most significant for local revitalization to extract industrial clusters appropriately and to understand the industrial characteristics of those local clusters.

Then we extracted regions where firms are concentrated by spatial index using a big data about distribution of firms and newly defined them as local clusters. As a result, it became clear that we can extract industrial clusters in each region properly to some extent while they are not evaluated strictly. Besides, we analyzed the characteristic of industrial structure and the spatiality of transactions in each local cluster defined. As a result, we appeared what clusters have special characteristics. These results not only show a new utilization of large-scale business transactions data for research activities, but also can improve cost-effectiveness of economic polices by giving a guideline selecting a combination of area and policy which maximizes an economic effect. Our future works are extraction of economic trends of each cluster with analysis of time series transaction data and modeling how characteristics of industrial structure impact development or decline of cluster.

References

Yamamura E (2003) Human capital, cluster formation, and international relocation: the case of the garment industry in Japan, 1968-98. *Journal of Economic Geography*, 3(2003), 37-56

Kodama T (2008) The role of intermediation and absorptive capacity in facilitating university-industry linkages-An empirical study of TAMA in Japan, *Research Policy*, 37(2008), 1224-1240

Sasaki M (2010) Urban regeneration through cultural creativity and social inclusion: Rethinking creative city theory through a Japanese case study, *Cities*, 27(2010), S3-S9

Takeda Y, Kajikawa Y, Sakata I, Matsushima K (2008) An analysis of geographical agglomeration and modularized industrial networks in a regional cluster: A case study at Yamagata prefecture in Japan, *Technovation*, 28(2008), 531-539

Carvalho V, Nirei M, Saito Y (2014) Supply Chain Disruptions: Evidence from the Great East Japan Earthquake, *RIETI Discussion Paper Series*, 14-E-035

Jain A, Data clustering: 50 years beyond K-means, *Pattern Recognition Letters*, 31(2010), 651-666

Arthur D, Vassilvitskii S (2007) k-means++: the advantages of careful seeding, *Society for Industrial and Applied Mathematics Philadelphia* (pp. 165-174). USA: PA

Gonda K, Kakizaki F (2001) Knowledge Transfer in Agglomerations: A regional Approach to Japanese Manufacturing Clusters, *Innovative Clusters Drives of National Innovation Systems* (pp. 289 - 301). OECD Publishing