様式６－３

Topic-Setting Program to Advance Cutting-Edge Humanities and Social Sciences Research

（Responding to Real Society）

Progress Report

（Summary of Final Report）

［**Development of indicators for sustainable livelihood in regional communities**］

Core-Researcher: Takaaki Ohnishi

Institution: The University of Tokyo

Academic Unit: Graduate School of Information Science and Technology

Position: Associate Professor

Research Period： FY2015- FY2018

1. Basic information of research project

Research Area	Building an anxiety-free care system in a shrinking society; Establishing a lifeline infrastructure
Project Title	Development of indicators for sustainable livelihood in regional communities
Institution	The University of Tokyo
Core-Researcher (Name, Academic Unit & Position)	Takaaki Ohnishi Graduate School of Information Science and Technology Associate Professor
Project Period	FY2015 - FY2018
Appropriations Plan (¥)	FY2015 2,330,000JPY
	FY2016 4,250,000JPY
	FY2017 3,150,000JPY
	FY2018 1,500,000JPY

2. Purpose of research

In underpopulated areas, regional city centers, and metropolitan suburbs that suffer from depopulation and population aging, concerns are being raised about growing life insecurities, "shopping refugees" (food deserts), the emergence of people who have difficulties moving (senior citizens without means of transportation), issues in maintaining and managing one's lifeline, and declining local disaster prevention capacities caused by the withdrawal of grocery stores, public institutions for medical care and welfare, and buses and other transportation facilities as well as labor scarcity. This study develops high-precision, high-frequency quantitative indicators based on empirical data, which link to proposals to avoid such this kind of social disorder. Spatio-temporal big data is applied to measure how far away stores, hospitals, schools, and shelters are from each and every resident in Japan, as well as to develop sustainability indicators of local infrastructure in local communities. Although maintaining healthy, safe, and secure living environments is an urgent matter for local governments, they cannot endlessly spend government funds to maintain communities and protect people's lifestyles. Municipalities often consider measures after the problems arise, and their responses toward solving problems tend to delay. This study scientifically supports the planning of policies and systems for sustainable local infrastructure, safety, security, and disaster prevention and mitigation by developing high-precision, high-frequency quantitative indicators based on empirical data, and by presenting and estimating issues, which are expected to surface for both municipalities and residents, as numerical values.

The essential social issue that this study aims to solve lies in the progression of population decline and the super-aging society; the issue is how to secure living space for residents to be able to live with a sense of security. The government is promoting the concept of a compact city to deal with this issue, where administration and commercial facilities are consolidated in some areas to raise efficiency. However, this concept has low feasibility, as credible scientific information about the positional relations between urban facilities and residents does not exist. To further complicate matters, this issue cannot be resolved simply from a perspective of economic efficiency. Consensus needs to be built between municipalities and residents to seek the most appropriate solution, while also respecting the local history, culture, and residents' wishes to stay in their current habitat. The sustainability indicators of local infrastructure in communities should represent the current situation quantitatively and help this decision-making process.

Diverse and detailed information of human activities is being recorded at a high frequency, due to the dramatic improvements of information technology (IT) and computer performance in recent years. However, the conventional framework of the humanities and social sciences tends to emphasize concept and theory rather than data; it requires restructuring from the viewpoint of proven science, based on detailed data observation and analysis. In order to analyze social and economic phenomena from the viewpoint of science by using big data, knowledge of analytical methods from mathematical engineering and statistical mathematics is required. In addition, computing technology to process big data accurately and efficiently as well as interpretation based on cases from actual society and existing information from the humanities and social sciences. Based on this, the present study is one of interdisciplinary cooperation between researchers in the fields of mathematical engineering, statistical mathematics, and the humanities and social sciences, as well as local government staff.

3. Outline of research (Including study member)

Core-Researcher: Takaaki Ohnishi (Associate Professor, Graduate School of Information Science and Technology, The University of Tokyo)
Study Member: Takayuki Mizuno (Associate professor, National Institute of Informatics)
Study Member: Yasuyuki Muradate (Policy Planning Specialist, Office for Econometric Analysis, Cabinet Office)

Dr. Yasuyuki Muradate, who also worked as a staff member of Chiba City during the first half of this study's duration, mediated our research and work duties. The following categories were implemented with their application in an administrative setting in mind. The direction of the study was adjusted throughout the course of its implementation, and views were exchanged as fit with Chiba City staff members.

(1) Surveying, Acquiring, and Consolidating Data
We surveyed what kind of spatio-temporal big data existed covering the whole of Japan on the extremely microscopic level, and selected and acquired big data useful for the development of the indicators. Time-series map data was created by combining numerous acquired data, which was consolidated so as to allow analysis processing. The acquired data's

comprehensiveness, spatio-temporal accuracy, time period, and sampling bias were verified, and data was selected for analysis.

(2) Statistical Analysis and Indicator Development
We investigated the data's basic statistical properties, examined the analysis results, and performed an analytical study that integrated numerous data. We measured how far away stores, hospitals, schools, and shelters were from each and every resident in Japan. This process involved an extensive amount of calculations fit for parallel computation; therefore, we applied massively parallel computing by using a supercomputer. This enabled the development of high-precision, high-frequency quantitative indicators to measure the sustainability of local infrastructure in communities for each and every ordinary citizen.

(3) Indicator Improvement and Risk Assessment
We improved the developed indicators to be more useful. As means of transportation and living environments vary greatly depending on the district, the distance it takes to reach stores and facilities depend heavily on the location. We therefore developed assessment methods to measure distance after adjusting for differences characteristic of the districts, rather than simply measuring the distance between two points. We devised a new physical quantity to better understand the sustainability of local infrastructure. We then examined the indicators that applied this physical quantity.

(4) Scientific Policy Support
We examined methods to apply the developed indicators to potential risks and assessments of the current situation. We provided empirical material that scientifically support planning measures for sustainable local infrastructure, safety, security, and disaster prevention and mitigation; we visualized sustainability indicators unto maps for the early detection of issues in local infrastructure, and to promote local governments to promptly deal with these issues.

4. Research results and outcomes produced

(1) Surveying, Acquiring, and Consolidating Data
We surveyed what kind of spatio-temporal data existed; selected and acquired the following by carefully examining the data's characteristics, such as comprehensiveness, from the perspective of applying them to indicator development; we then consolidated the following for analysis.

-The 100-Meter Estimated Mesh Data from the 2010 and 2015 National Censuses
This data is the totalized data from national census 500-meter mesh data, which has been converted into a 100-meter mesh precision by the number of households. The 100-meter mesh across Japan enables us to acquire the population by age group and number of senior citizens' households, thereby giving us a high-precision understanding of all ordinary citizens' residences.

-Summary Data of Mobile Spatial Statistics in Fiscal Year 2014

This data estimates the population by weekday or holiday, gender, and age group through a nationwide 500-meter mesh, using cell phone location information.

-Corporate Telephone Directory Database Telepoint with Coordinates(ZENRIN)

This data is from a telephone directory including industry information for every year between 2010 and 2016. A wide variety of companies from large to small businesses are recorded, enabling the identification of geospatial information from all stores and facilities nationwide (approximately seven million).

-CSV Data from Personal and Corporate Phone Directories

This data is from phone directories with industry information for every few months between November 2011 and the present. Although longitudinal and latitudinal information is not provided, this data enables analyses that are accurate for the months listed, as well as estimations of the population using personal phone directories.

-Mesh Data of Commercial Statistics in 2007 and 2014

Although this data gives an overview of wholesale and retail store information, the survey is conducted only once every few years, and data cannot be obtained for the same time period as the national censuses, or the latest data. Moreover, no information exists about industries not subject to this survey. This data was supplementally applied for verification.

-Database of National Shelters

This data provides accurate information of the longitude, latitude, and divisions of national shelters for three points in time: 2015, 2016, and 2017.

-People-Flow Data

Data of people's movement records can be created from ID information by using Twitter data with longitudinal, latitudinal, and ID information (nationwide, over 5 months, 40 million tweets, 1 million users) and point-type floating population data by Agoop Corporation (location information data of smartphone users during one year in Chiba City).

-POS Data with ID Information

This data shows information relating to who purchased what kind of product when and where. We calculated on a nationwide scale how far away from their residences ordinary citizens go shopping at stores such as supermarkets, convenience stores, and drugstores.

(2) Statistical Data Analysis and Indicator Development

-Basic Statistical Data Properties

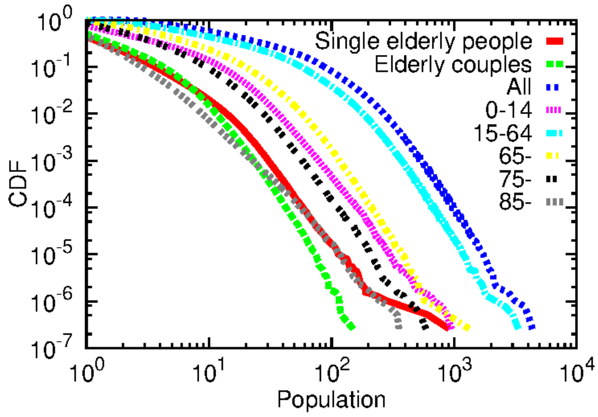

Figure 1: Complementary cumulative distribution of 100m Mesh population by age group

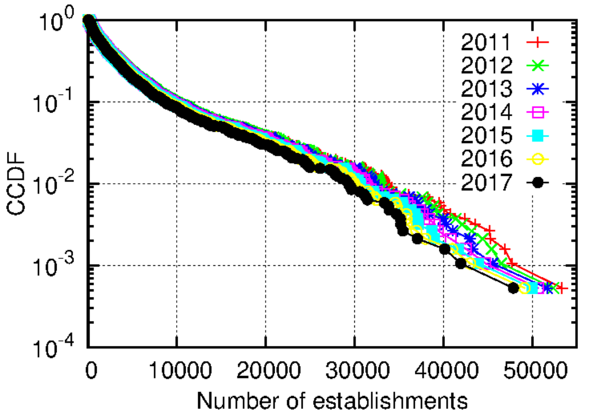

Figure 2: Complementary cumulative distribution of the number of shops and facilities in each municipality for each year

We examined the population on a 100-meter mesh, and verified that although most meshes include a population of only a small number of people, there were also a few meshes that include hundreds or thousands of people (Figure 1). We examined the total number of stores and facilities in every municipality recorded in the phone directory data. We verified that although most cities had stores and facilities on a scale of a few thousand or less, there were also big cities with tens of thousands of such facilities (Figure 2). Efficient calculations were required to determine how concentrated they are, as the population, number of stores, and facilities differ on a grand scale by mesh and city. We also examined how much the number of stores and facilities nationwide had increased in seven years by industry, and confirmed the data to show that numbers had increased significantly in industries such as welfare and nursing care, day care centers and nurseries, convenience stores, and funeral services, while numbers had decreased significantly in industries such as consumer money lending, magnetic tapes, discs, records, and CDs. As for the people-flow data, users who have the longitudinal and latitudinal information switched on tend to always have it on; by comparing this longitudinal and latitudinal data with the population's spatial distribution from the national census data, we found that Twitter users tend to tweet many times at night, and we needed to remove the day-time period in order to treat it as people-flow data.

-An Efficient Calculation Method of Distance

In order to efficiently calculate the distance between the population, stores, and facilities in relation to their concentration, we converted the longitudinal and latitudinal coordinates into first principal component and second principal component coordinates, which were found though a principal component analysis. By sorting the coordinates by the size of the first principal component's value, giving them serial numbers, and applying the distance calculated using only the first principal component, we developed an efficient calculation method using the Hubeny formula to determine the distance to the closest stores and facilities. We made calculations of stores and facilities in various industries using massively parallel computing through a supercomputer utilizing MPI.

-The Distance to the Closest Stores and Facilities

We calculated the distance to the closest stores and facilities by industry for all ordinary citizens who had responded in the national census (Figure 3). The median distance was approximately 200 meters for grocery stores, hospitals, and welfare facilities, and approximately 300 meters for shelters; while these are located close to residences, facilities such as police and fire stations are located a little further at approximately one kilometer away. Our calculations of distance by age group showed that senior citizens tended to live further away from stores and facilities than people of working age. A detailed examination showed that districts existed where the closest stores and facilities were more than a few kilometers away, and people aged sixty-five and above as well as aged households particularly tended to be in these districts. Furthermore, calculations of average distance stores and facilities are from households by prefecture showed more concentrated urban populations tended to live closer to stores and facilities, whereas shrinking countryside populations to live further away. We also calculated the distance to the closest stores and facilities by city, and created a list ranking the greatest average distance per industry. From this list, we can estimate the potential risk of so-called shopping refugees, who have no easy access to stores for daily necessities, and medical refugees, who have difficulties receiving medical care.

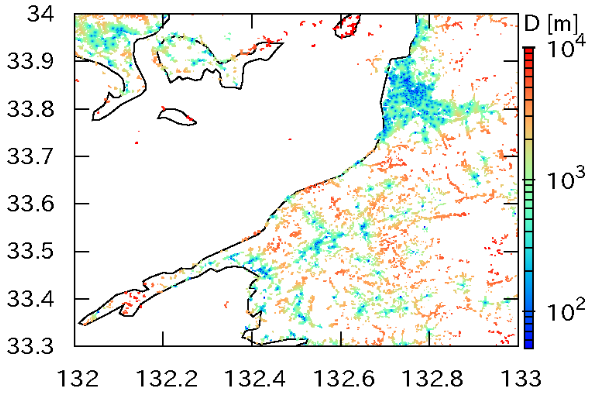

Figure 3: Distance to the nearest grocery store (Northwest of Ehime prefecture)

(3) Indicator Improvement and Risk Assessment

-A Comparison of Distance to the Closest Stores and Facilities

In order to examine whether senior citizens' living environments are good or bad compared to those of other age groups by prefecture of residence, while also considering the locality of their living environments, we divided the average number of residents aged sixty-five and above with the average number of residents of all ages, and calculated values for this (ratio of distance). For all industries, no differences were found between senior citizens and other age groups in urban districts; however, senior citizens strongly tended to live far away from their closest stores and facilities in the countryside. We calculated the distance ratio by city, and created a list ranking the greatest distance ratio for each industry by city. This enabled us to make a risk assessment while also considering factors characteristic of those districts.

-The Change Rate of the Distance to the Closest Stores and Facilities

We examined the change rate of the distance over five years from the median values of two points in time in 2010 and 2015, for which national census data exists; our findings showed that many industries tended to shorten their distance to residences during these five years—this tendency was more noticeable among working-age people than senior citizens. The distance between senior citizens and the closest stores and facilities are increasing compared to that of working-age people, raising concerns about the nationwide worsening of senior citizens' living environments. We calculated the proportion of the change rate in the distance over these five years for both senior citizens and all residents; analysis findings of the senior citizens' living environments when compared to other age groups showed that for grocery stores, the change rate proportion was great for Kyoto and Kagawa Prefectures. That is, we found that the distance between stores and facilities and senior citizens' residences was growing, and indicating a worsening tendency for their living environments. We also calculated the change rate for other industries and assessed whether the senior citizens' living environments were worsening or improving. Moreover, we calculated the change rate proportion by municipality, and ranked these proportions by municipality.

-Indicators of the Number of Stores and Facilities Suitable for the Population Using Scaling Exponents

Using data from phone directories, we analyzed the population X and the number of stores and facilities Y in all municipalities across Japan; findings showed that the number of stores and facilities are proportionate to the exponent of the population. Exponent is dependent on the type of number of stores and facilities; for example, if the population doubles, the number of dermatology clinics also doubles, but dental clinics multiply by 2.1, and obstetrician clinics by 1.8—therefore, nontrivial relations exist between the population and number of stores and facilities in various industries. By using industry-dependent scaling exponents (Figure 4), we developed indicators quantifying whether the number of stores and facilities truly correspond with the population by comparing them with other municipalities. By comparing the appropriate number of stores and facilities, which are calculated from our indicators, with the actual number of stores and facilities, we were able to quantitatively discuss which kinds of facilities each municipality has a shortage of, as well as which facilities and industries could potentially be considered superfluous. We also developed indicators for determining a municipality's risk of disappearing using corresponding values of the scaling relation's X-intercept, and quantitatively calculated the population size that is required to maintain elementary and junior high schools and hospitals in cities. We anticipate that these indicators will assist discussions about the sustainability of local infrastructure, as well as potential mergers of municipalities.

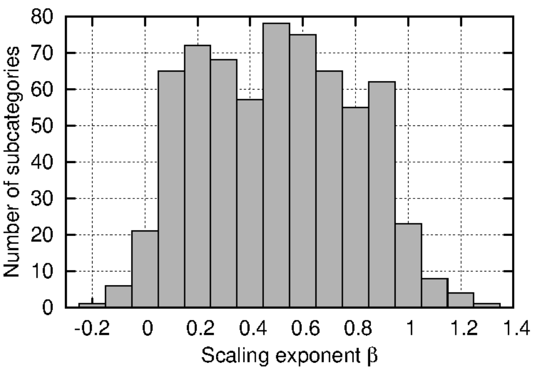

Figure 4: Distribution of scaling exponents for each industry

-Extracting the Locality of Local Infrastructure Using the Industrial Composition of Storesand Facilities

Locality exists in local infrastructure, as natural environments and industrial structures vary according to the district. We focused on the number of stores and facilities of 735 subcategorized industries existing in municipalities to extract the municipalities' characteristics from phone directories; we then characterized the cities along 735 vectors and conducted a cluster analysis (Figure 5). Through this, we were able to find municipalities with a similar proportion of industries for stores and facilities, even with greater geographic distances. The municipalities comprised of similar types of stores and facilities are highly likely to have similar living environments; referring to the municipality's current situation and policies can help local governments devise policies and predict risks.

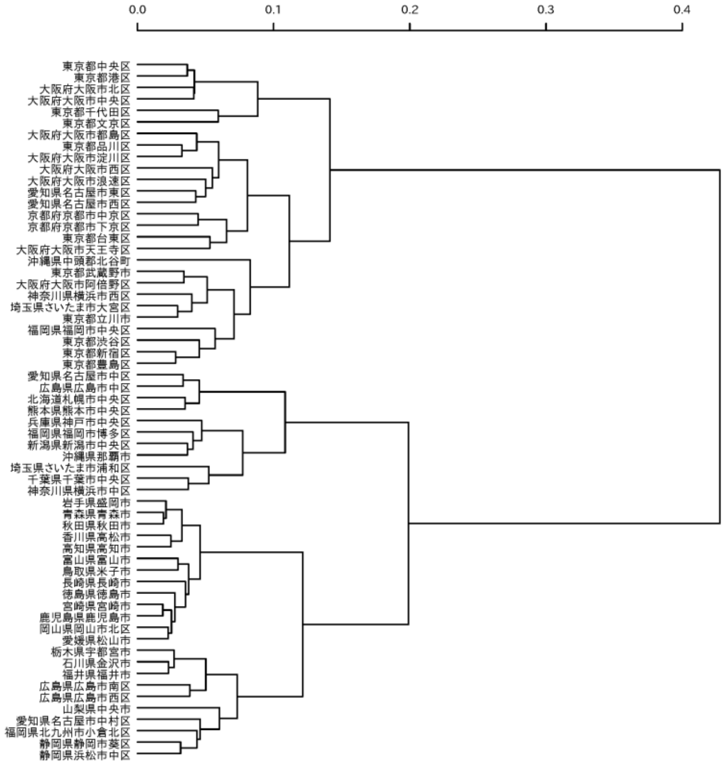

Figure 5: Dendrogram obtained by cluster analysis of municipalities in Japan

-A Substantial Estimation of Local Infrastructure Using People-Flow Data

We extracted the movement records of people in municipalities, and represented the relation between municipalities as a weighted network. We quantified the strength of these relations from the network, and developed indicators that demonstrate the relation between municipalities. We profiled user attributes from tweets on Twitter, and identified the fundamental statistical properties of people's movements according to attributes. We used the Hurst exponent and community analysis, to show that the diffusion coefficient for movement and size of the residents' livelihood zones differed by the residents' income, age, hobbies, and residence. We also estimated the number of people who are awake in each part of Japan during each time slot of the day by referring to tweets on Twitter. By comparing the population distribution of residences found in national censuses to the percentage of people who went to bed for each time slot—calculated from extensive life-logs—we developed a method based on Twitter data to estimate the floating population by attribute for each day-time time slot in every district in Japan. These findings can be utilized for detecting livelihood zones by residents' attributes, proposing optimal placements for administrative facilities, designing patrol routes for security and surveillance, maintaining roads and sidewalks, estimating working hours by district, extracting areas with potential high demand, attracting tourists, selecting administrative support, and estimating the attributes of facility visitors; they can also help in planning administrative policies. Present administrative divisions have been determined artificially and historically, and are not necessarily divided or bordered in a realistic and optimal way. By utilizing the estimated relation between municipalities, community structures, and regional livelihood zones to review administrative divisions, we hope to improve the sustainability of local infrastructure in communities.

(4) Scientific Policy Support

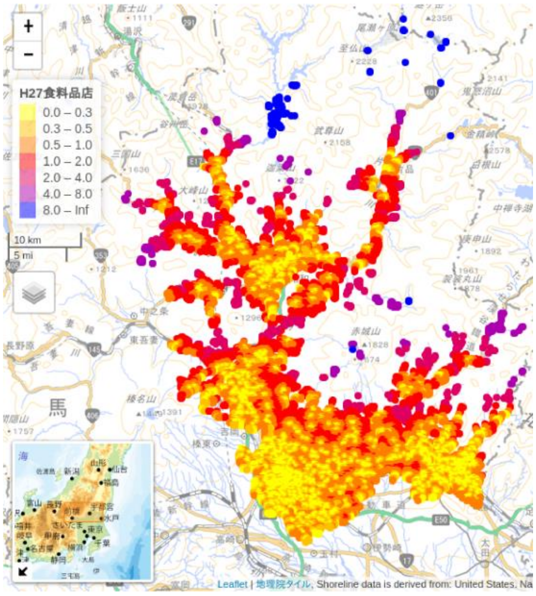

Figure 6: Project Web page

The distance, distance ratio, and change rate of distance to the closest stores and facilities, which were calculated by industry on a 100-meter municipal and prefectural scale; the indicators of the number of stores and facilities which corresponded to the population in every city using the scaling exponent; the extraction of the locality of local infrastructure using the industry structure of stores and facilities; and the actual estimation of local infrastructure using people-flow data developed in (2) and (3) can be utilized immediately as indicators to support planning administrative policies. Although questionnaire surveys tend to be conducted every few years towards specific districts to understand the communities' local infrastructure, it is important for administrations to understand the current situation and take measures from an early stage; therefore, methods with better precision and higher frequency are required. The developed sustainability indicator for local infrastructure is highly precise, both temporally and spatially on a nationwide scale. Administrative policies such as community-based integrated care systems need to be examined according to the district's specific characteristics; we can hope to utilize the information from the present study which was acquired by referring to spatio-temporal bigdata. We can specify the districts which are suspected of having potential risk by referring to the sustainability indicators, and efficiently utilize both the bigdata and questionnaire survey by conducting a questionnaire survey only to the extracted districts. We visualized the developed sustainability indicators on the map, and disclosed them onto the project webpage (Figure 6) to swiftly detect issues relating to the communities' local infrastructure, and to promote local governments to speedily resolve these issues. The present study produced helpful knowledge for planning policies, by referring to bigdata which had accumulated but were not fully utilized. We have proposed specific cases of bigdata utilization which lead to resolving social issues; this will lead to new efforts that scientifically guide policies.

The residents who had not responded to national censuses, and stores and facilities which were not recorded in phone directories have not been considered in calculating the sustainability indicators. Home delivery, moving stalls, home production and direct sales stores exist in rural areas, which need to be considered to estimate the risk of shopping refugees. Although we estimated the distance using one-line distance in the present study, in the strict sense, the distance should have gone along with the roads. The error in the distance measurement is considered to be significant in areas with rivers and seas. The issue for the future is to survey whether the calculated indicators are truly reflecting the actual situation, by observing from a different perspective. We must also survey whether the reasons of the trends acquired from these indicators are due to the withdrawal or advancement of stores and facilities, or due to the relocation, deaths, and births of people.

Also, although we mainly focused on age groups for people's attributes, other attributes for people requiring long-term care can also be considered for analysis. We should also examine how to combine the numerous developed indicators and lead them to a comprehensive conclusion. Future developments for this study are to devise an effective study method which combines both the bigdata analysis and previous questionnaire surveys (field surveys), and to aim for innovative research findings which cannot be discovered through studying questionnaire surveys alone. We feel that this form of research will become critical in various fields in the future.