ORIGINAL RESEARCH

Use of a text mining method for classifying citizen report data and analyzing the occurrence trend of local problems

Eiji Kano*[1,2], Kazuhiko Tsuda[2]

[1] Dept. of Research and Publication, Institute of Administrative Information Systems, Tokyo, Japan
[2] Graduate School of Business Sciences, University of Tsukuba, Tokyo, Japan

Received: August 30, 2019 **Accepted:** September 16, 2019 **Online Published:** September 30, 2019
DOI: 10.5430/air.v8n2p1 **URL:** https://doi.org/10.5430/air.v8n2p1

ABSTRACT

An important task of any municipality is the maintenance and improvement of the street-related living environment and traffic safety for citizens. For this, their department of street maintenance is expected to efficiently perform the maintenance and inspection of streets according to priority with limited human and budgetary resources. Recently, municipalities in various countries are adopting "the citizen report system," which is a system of reporting problems of streets, such as damaged streets, by citizens to their municipality, for citizens to perform part of street maintenance and inspection. It is possible that the data obtained by municipalities through the citizen report system can be utilized not only for early problem detection but also for prioritizing administrative measures by using it for analyzing the occurrence trend of problems. Problems reported by citizens, however, are classified by different methods from municipality to municipality, and thus the collection and comparative analysis of such data across municipalities is difficult. This study presents a method of commonly classifying such data, regardless of different classification standards, by analyzing the contents of citizen reports by using text mining. We then analyze the relationship between the trend of citizen reports and the occurrence trend of problems concerning the living environment and traffic safety, using the citizen report data of three large municipalities classified by this method, and infer the occurrence trend of problems. This study has confirmed that citizen report data possibly contributes to municipalities' prioritization of the maintenance and improvement of the living environment and traffic safety.

Key Words: Text mining, Citizen report, Road damage, City, Municipality

1. INTRODUCTION

In recent years, the environment for municipality administration is becoming severe more and more. While the finance of municipalities is becoming tight as a result of decreasing tax revenue and increasing cost of social security arising from an aging population, a declining birthrate and a decreasing working-age population, the shortage of local administrative officers is becoming serious. Thus, municipalities are expected to solve local problems according to priority with limited human and budgetary resources.[1] One of the main tasks of any municipality is the maintenance and improvement of the street-related living environment and traffic safety for their citizens. Especially, their department in charge of street maintenance is expected to efficiently perform the maintenance and inspection of streets according to priority.[2]

One of the measures for this problem is a system in which

*Correspondence: Eiji Kano; Email: kano@iais.or.jp; Address: Institute of Administrative Information Systems, Dept.of Research and Publication, Japan.

citizens report to their municipality problems such as street damage through their mobile phone applications (hereafter called "the citizen report system"), and it is rapidly adopted by municipalities in various countries around the world.

The introduction of the citizen report system allows citizens to easily access their municipality for quickly solving various problems. It is also expected that substituting the citizen report system for detecting street problems, which has been conventionally done by municipalities, allows them to efficiently operate their inspection of streets. In addition to this direct advantage, analyzing the occurrence trend of problems in the living environment and traffic safety using citizen report data provided by the citizen report system has a possibility of providing knowledge useful for prioritizing administrative measures. The reported problems, however, are classified in different methods among municipalities, it is difficult to collect and comparatively analyze such data across municipalities. In this study, we propose a method of commonly classifying citizen report data by analyzing the data by using a text mining method. Then, using the citizen report data classified by this method of three large cities, Sagamihara City, Chiba City and Hamamatsu City, we analyze the relationship between the trend of citizen reports and the occurrence trend of problems concerning the living environment and traffic safety and infer the occurrence trend of problems.

2. SIGNIFICANCE OF CITIZEN REPORTS

2.1 Streamlining the maintenance and inspection of streets by the citizen report system

The first citizen report system is FixMyStreet of the United Kingdom, whose service began in 2012, and since then similar systems have been introduced in various countries, as shown in Table 1.

Table 1. Examples of the citizen report system

Japan			Other countries	
MyCityReport	**FixMyStreet Japan**	**Original Systems**	**FixMyStreet**	**SeeClickFix**
Chiba City	Handa City	Otsu City	United Kingdom	Oakland(US)
Muroran City*	Beppu City	Sagamihara City	Norway	California(US)
Numazu City*	Koriyama City	Hamamatsu City	Canada	Detroit(US)
Hirosaki City*	Ikoma City	Ashiya City	NewZeaLand	Meatuses(US)
Adachi Ward*	Iwaki City	Kusatsu City	Brussel (Belgium)	Connecticut(US)
Sumida Ward*	Kumagaya City	Izumisano City	Zurich (Switzland)	Sonora(Mexico)
	+11 Cities and more	+10 Cities and more	etc.	etc.

*Under verification test

Citizen reports have contributed mainly to the streamlining of the maintenance and inspection of roads by municipalities. In Sagamihara City, for example, they could reduce the frequency of patrol for road inspection after the introduction of a citizen report system to less than half that before the introduction, as shown in Figure 1.

The introduction of the citizen report system does not substitute all maintenance and inspection operations. Chiba City, for example, has pointed out that their citizen report system does not fully cope with damage of major roads.[3] Therefore, while inspection patrol for major roads should be done by the municipality as before, the citizen report system is to be used mainly for minor roads, for which thorough inspection by the municipality is difficult.

In this case, it becomes necessary to determine in which areas minor roads should be inspected by the municipality and for which areas the citizen report system should be used. It seems possible to prioritize such inspections more properly by analyzing citizen report data, as shown in Figure 2.

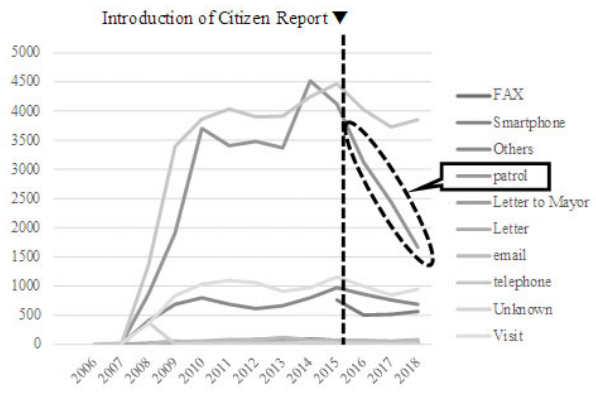


Figure 1. Composition of reporters in Sagamihara City

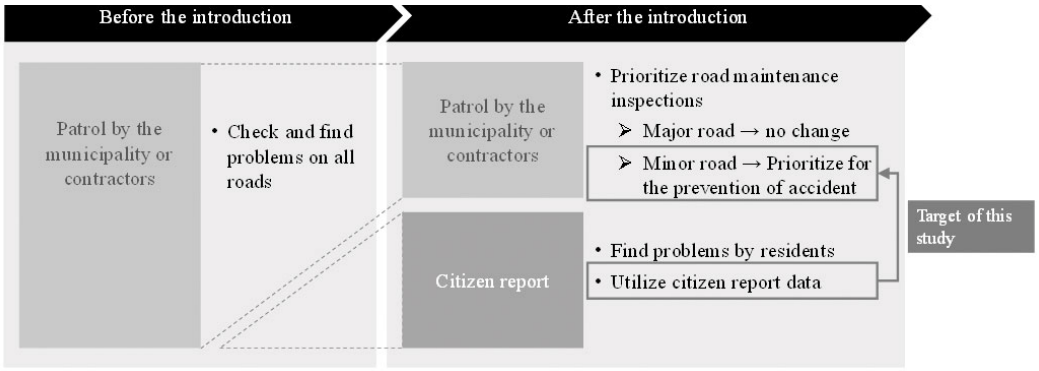


Figure 2. Significance of utilizing citizen report data to prioritize road maintenance and inspection

2.2 Previous Research on Citizen Report

Since FixMyStreet in UK has launched its service in 2012 as the forerunner, the challenge has been noted from the beginning as an innovative attempt to redefine the relationship between citizens and municipalities.[4] Over the years, various studies have been conducted to examine the significance of this approach, social functions, and proposals for new uses.[5,6] Recently, research has also been conducted to position the role of finding problems by citizens as a type of sensor and to try to explain its social significance.[7] In Japan, also, various considerations have been made regarding the mechanism, social significance, etc., of the citizen report system, mainly by organizations and stakeholders related to the operation of the citizen report.[8,9]

In analyses using the data of citizen report, there have been studies that analyzed the tendency of the number of registrants and used the data in workshops to analyze differences in the problem recognition of inhabitants.[10,11] Moreover, although it did not deal with citizen report, there was a study that tried to classify the data of residents' comments using machine learning as a related example.[12]

2.3 Subject of previous research

Previous studies had the following issues.

First, as described in 2.2, studies were conducted to elucidate the mechanism and significance of citizen report and to evaluate its achievements. On the other hand, none of the research analyzed the content of the citizen report for understanding regional issues. The content of regional issues can be understood only through a category set by the municipality as the service operator.

Second, in order to extract and classify local problems from citizen report, text mining can be an effective alternative to analyzing the report content.[13,14] However, no studies have applied this method to citizen report.

The significance of this study is to clarify the structure of regional issues by analyzing citizen report using text mining.

3. Construction of a Method of Classifying Citizen Reports

3.1 Data and programming environment used in this study

As shown in Table 2, the methods of classifying categories are different and inconsistent among municipalities. Some citizen report systems such as those specialized for roads do not set any categories from the beginning.

Among the municipalities, the citizen report system of Chiba City does not have smaller categories than "Roads" for road-related problems. Thus, while it is not possible to classify and grasp concrete local problems, the range of citizen report is not limited, and no category is set based on the contents of problems. It is likely, as a result, that the contents of citizen reports are not much biased by classification by the municipality and strongly reflect the subjective recognition of problems by citizens.

Therefore, we reclassify the citizen reports of Chiba City in a way suitable for the characters of the problems by applying text mining to the titles and contents of the reports and examine if the classification method can be applied to other municipalities.

In this study, we use 4,574 citizen reports of Chiba City covering from August 2014, when the operation of their citizen report system started, until March 2018.[13]

To conduct the analysis for the text separation in 3.3.1 and the dependency analysis in 3.3.2, we use Text Mining Studio ver. 6.0.3. As for the other parts, any other processes such as the experiments in 3.4 and 3.5 were performed manually while only using the Excel function because the rules were very simple.

Table 2. Examples of categories of citizen report

Chiba-City	Koriyama-City	Handa-City	Sendai-City	United Kingdom	
Road	Road	Problem of Water Channel	Road	Abandoned vehicles	Road traffic signs
Park	River	Problem of Weed	Gutter	Bus stops	Roads/highways
Garbage	Park	Problem of Traffic Safety	Guardrail	Car parking	Rubbish(refuse and recycling)
Others	Security Light	Problem of Park	Curve-mirror	Dog fouling	Street cleaning
	Garbage	Problem of Public Facility	Street Light	Flyposting	Street lighting
	Public Facility			Flytipping	Street nameplates
	Other			Graffiti	Traffic lights
				Parks/landscapes	Trees
				Pavements/footpaths	Other
				Potholes	
				Public toilets	

3.2 Text mining

Text mining is a method to extract useful information by analyzing the appearance frequency, modification structure, co-occurrence relation, etc.[14] In text mining, sentences are divided into morphemes, which are the smallest unit having meaning, and natural language processing such as morphological analysis is performed to identify the word class, etc.[15,16] Although there are various previous studies about classification using text mining,[17,18] there is no example of an application to citizen report, so this study constructs a new method to classify problems.

3.3 Classification using modification structure

The following steps were implemented to ensure appropriate problem classification in an objective and reproducible manner according to the content of citizen report:

3.3.1 *Dependency analysis*

Analysis of dependency was conducted as follows.

(1) Words are segmented, and "headword," that is the meaningful word for a sentence, is extracted from a morpheme or combined words.
(2) Headword is transformed into "replaced word" that is the original form before declension.
(3) A combination of words in which "replaced words" are in a dependency relationship in one sentence is extracted.

3.3.2 *Grouping using modification structure*

When multiple clauses are related as "subject and predicate" or "modifier and modified word" in one sentence, the combination of clauses is said to be in a dependency relationship. For example, in the sentence "The street light goes out.", the subject "light" and the predicate "go out" are in a dependency relationship. In this study, the subject or modifier "light" is called as "modifying part" and the predicate or modified word "go out" is called as the "modified part."

Words having the same modified phrases such as "go out" were regarded as synonymous when classifying words such as "electric lamp" and "street lamp" as shown in Figure 3. These were classified into the same group.

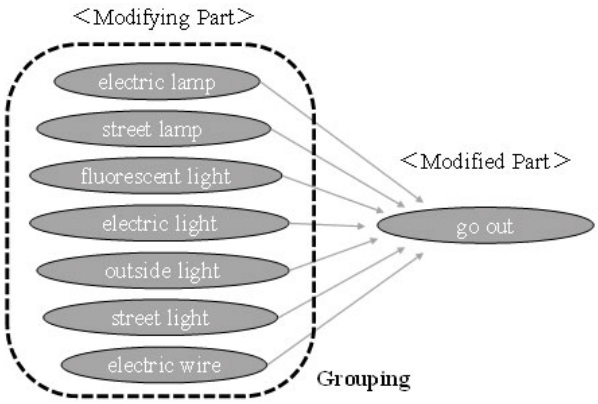


Figure 3. Grouping using modification structure

However, in order to exclude combinations of dependencies that are not useful for grouping, only combinations of the 706 cases that correspond to the following conditions are extracted:

(1) A combination of "noun-verb" or "noun of irregular conjugation of the line of 'sa'"
(2) A combination of two or more occurrences

3.3.3 *Removal of expressions in which problems are not identifiable*

In addition, combinations of the following expressions that are not useful for meaningful classification were excluded:

(1) Expressions that do not lead to the identification of the problem (e.g., rainfall, photograph-take, people-pass)
(2) Supplementary expression (e.g., do, please, think)

As a result, 397 combinations of dependencies were extracted. This included 32 modified words as shown in Figure 4 (hereafter, this group of words are called the "first group").

3.3.4 *Create categories*

In order to consolidate the first group into a smaller number of categories, modified words that have the same modifying word were classified into a group. When creating groups by modifying words that have two modified words, groups were formed that were considered to be well representative of the characteristics of local problems (hereafter, this group of modifying words are called the "second group"). When grouping words that have three or more modified words, useful groups were not formed because the generality of the meaning of the words became too strong. Finally, as a result of organizing the second group according to the dependency relationship, seven categories were derived.

Figure 4 shows the classification results. The categories are named based on their meanings: lamp breaks, road damage, facility defects, abandoned objects, overgrowth, clogging and invisibilities.

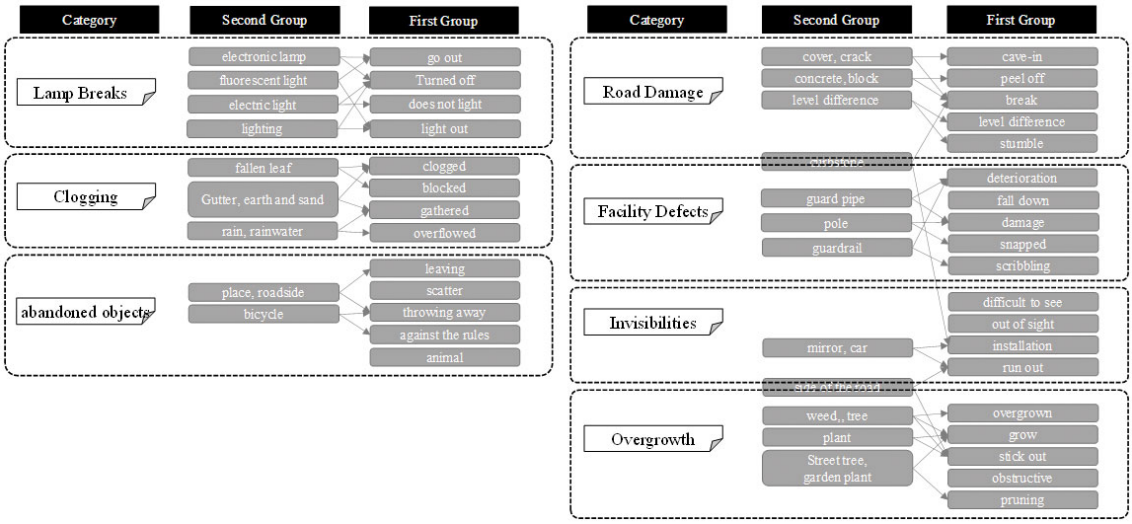


Figure 4. Category of citizen report built using the commonality of modified word

3.4 Set automatic classification rules

The category derivation in Chapter 3.3 was performed manually according to the procedure shown in the same chapter. Once categories are created, individual reports can be automatically classified according to the following procedure.

(1) When a modified word, modifying word, or synonym thereof belonged to a category included in a report, it will be classified into that category.

(2) If a report includes words belonging to multiple categories (e.g. in case a report includes multiple requests), it will be classified according to the following rules:

- Prioritize conformance to "theme" over conformance to "content"
  Like many other citizen reports, citizen report of Chiba City has two fields to fill in with text: theme and content. A reporter will fill in the theme field with the theme of the entire report, and the content field with the details of the problem such as how the road surface is damaged.If a word contained in a "theme" item and a word contained in a "content" item are classified in different categories, the priority is given to the category contained in "theme."
- Prioritize the word that appeared earlier
  If there are multiple words that belong to different categories in the "content" item in the same report, the word that appears first takes precedence.

(3) By processing the above steps, any report can be classified into a category automatically except when no word in a report is included in the category classification criteria or when a report is excluded because a particular reporter posts many reports repeatedly for the same issue and overlapping parts are removed as noise (hereinafter referred to as an "error"). In addition, the error ratio is limited to 12.4% (62 cases of 500 sample data mentioned later).

3.5 Verification of automatic classification rule

In order to verify the validity of the automatic classification rule constructed in 3.4, 500 cases were extracted from the oldest order, and the difference between the correct answer

data classified manually based on the categories in Figure 4 and the result of the automatic classification rules was verified. Consequently, the difference in the composition ratio of each category after error elimination was less than 5%, as shown in Figure 5.

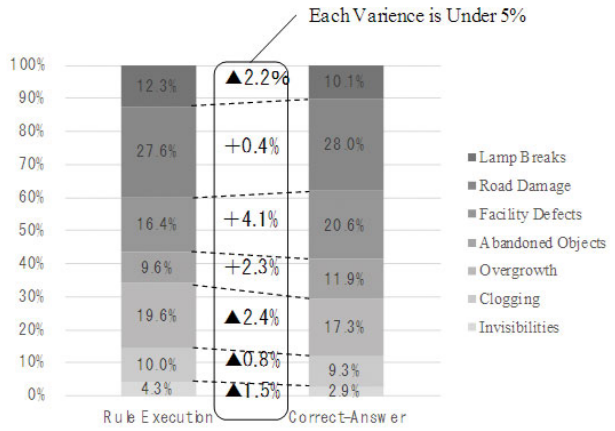


Figure 5. Verification result of automatic classification rule

From the above result, the automatic classification rule of section 3.4 was verified to be suitable for appropriately classifying citizen report.

3.6 Verification of applicability to other municipalities

We now verify 1) if the automatic classification rule of 3.4 is applicable to other municipalities and 2) if its result deviates from the result of classification done by the reporters themselves. For the verification, we use 1,434 citizen reports of Hamamatsu City starting from April 2015 until April 2019.[19] This is because in the citizen reports of the city the reporters themselves classify their reports and enter their contents and thus the data is suitable for the verification.

We first applied the automatic classification rule of 3.4 to the citizen report data of Hamamatsu City and classified it into the categories of Figure 4. We then created a mapping table relating the original categories of Hamamatsu City with the categories of Figure 4 and manually converted the classification done by the reporters themselves of Hamamatsu City into the classification of Figure 4. Finally, we examined the difference between the classification done by the reporters and the classification derived by applying the automatic classification rule. Figure 6 shows this process.

As a result, as shown in Figure 7, the difference after removing errors in the composition percentage of each corresponding category between before and after the operation was less than 5%.

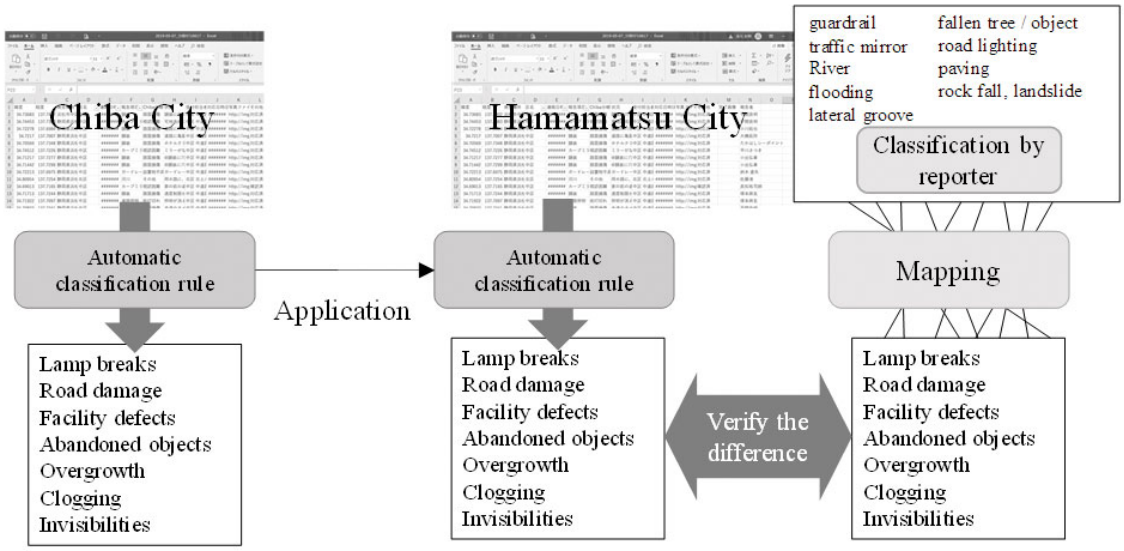


Figure 6. Application of the automatic classification rule to another city's citizen report classification

Thus, it was confirmed that 1) the automatic classification rule of 3.4 was applicable to data of other municipalities and 2) the result of the classification did not deviate from the classification done by the reporters themselves.

Therefore, it has been shown that it is possible to classify citizen report data of different municipalities with a common method, to collect data for each problem category and to analyze the data comparatively among the municipalities. Figure 8 shows the problem compositions, after removing errors, of the citizen reports of the three cities studied in this study obtained by the classification described above. It visualizes the difference of the problem compositions among the municipalities, which could not be compared before.

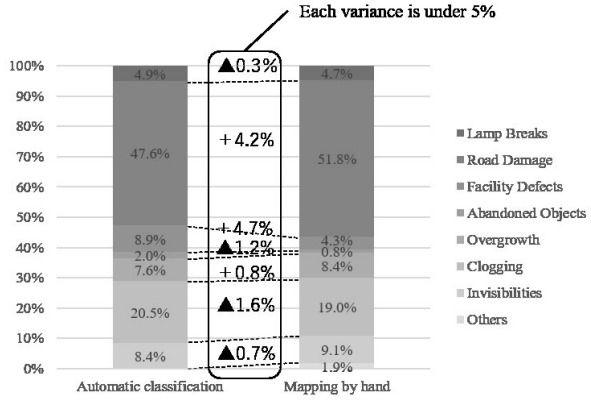


Figure 7. Result of applying the automatic classification rule

As a result of the above verification, we now call the problem categories derived from the automatic classification rule of 3.4, or lamp breaks, road damage, facility defects, abandoned objects, overgrowth, clogging and invisibilities, by the name "common categories."

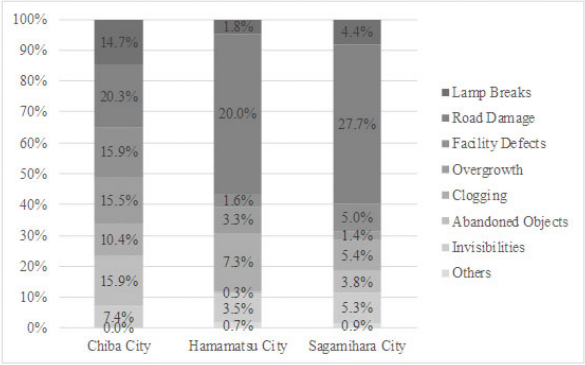


Figure 8. Difference in the problem composition of the citizen report systems among 3 cities

4. Analysis of the Relation between Citizen Reports and Citizens' Awareness

Through the citizen report system, many municipalities accept not only problems of road damage, for which the system was originally introduced, but also other problems included in the common categories. Consequently, it is expected for citizens that the use of the citizen report system leads to a quick detection and solving of a variety of problems related to the living environment and traffic safety in their area. On the other hand, if the municipality can grasp through the system the characteristics of the occurrence trend of problems in the area, they can possibly utilize the obtained data, for example, for prioritizing their measures. Especially in the areas experiencing large changes in population composition, such knowledge seems useful for quickly and flexibly taking measures against problems.

We thus examine in this chapter if the influence of population increase or decrease on the state of local living environment and traffic safety is reflected in the rate of citizen reporting of problems about the living environment and traffic safety. If it is verified, we confirm that citizen reports reflect the recognition of problems in the common categories by citizens. In addition, it would be possible to presume problems of the living environment and traffic safety from citizen reports.

It is noted that the municipalities studied here are Sagamihara City, Chiba City and Hamamatsu City. These cities have the following common characteristics and thus seem suitable for data collection and comparative analysis.[13, 19, 20]

- Has a population of similar size (all are large cities having a population of 730 to 970 thousand).
- Has the same legal authority and duty as an ordinance-designated city (a basic municipality of the largest class).
- Has been using a citizen report system for more than four years as an established system.

All the above cities have an accumulated total number of citizen reports of more than 1,000, which is among the largest in Japan.

4.1 Analysis of the influence of change in population composition: hypothesizing

The rate of population change is the sum of the rate of natural population change and the rate of social population change, which have different factors of change and different influences on citizens.[21] The rate of natural population change is the difference between the birth rate and the death rate. If needs for an improved child-rearing environment increase in an area with many births, the number of citizen reports should increase in many problem categories, since the child-rearing environment is related to the general living environment. In addition, the rate of social population change is the difference between the rate of moving-in and the rate of moving-out. In an area with many moving-in citizens, the traffic of vehicles increases, and therefore the number of citizen reports about unsafe roads should increase. These relations can be summarized as hypotheses given in Figure 9.

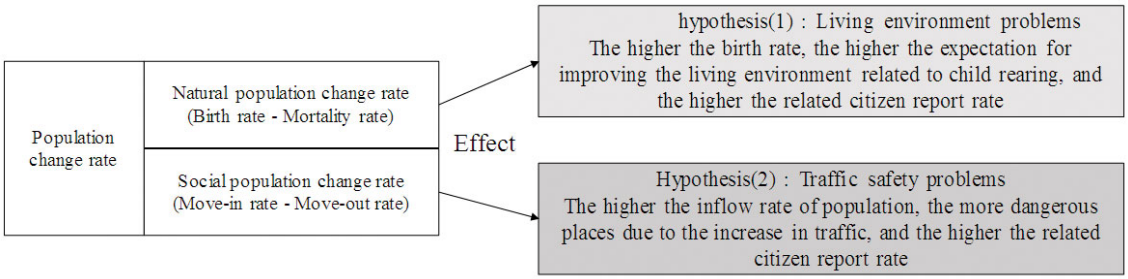
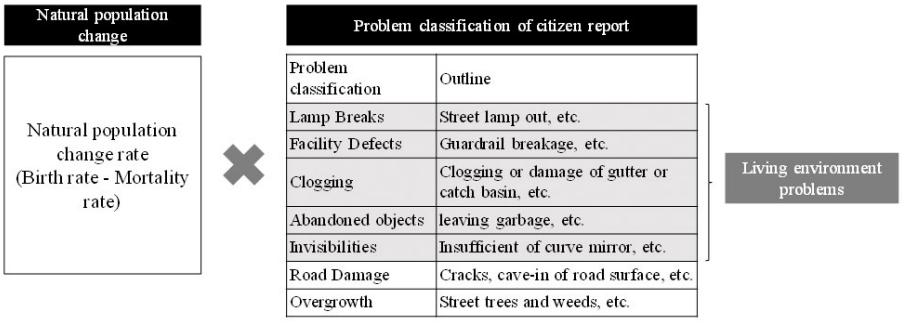


Figure 9. Hypotheses on the effect of population change on citizen reports

4.2 Analysis of the influence of natural population change on the recognition of problems in the living environment

While problems in the living environment are classified into all the common categories, the categories of those directly influenced by population change seem to be lamp breaks, facility defects, clogging, neglected objects and invisibilities, as shown in Figure 10 (all together called "living environment problems" hereafter). This is because the problems of road damage and overgrowth are not directly affected by population change.



Problem classification	Outline
Lamp Breaks	Street lamp out, etc.
Facility Defects	Guardrail breakage, etc.
Clogging	Clogging or damage of gutter or catch basin, etc.
Abandoned objects	leaving garbage, etc.
Invisibilities	Insufficient of curve mirror, etc.
Road Damage	Cracks, cave-in of road surface, etc.
Overgrowth	Street trees and weeds, etc.

Figure 10. Problem classification related to the living environment

If this hypothesis is correct, the rate of natural population change has a positive correlation with the rate of citizen reporting of living environment problems (a rate per unit population, hereafter called "the citizen report rate").

	Correlation coefficient
Lamp Breaks	0.41
Facility Defects	0.48
Clogging	0.55*
Abandoned Objects	0.44
Invisibilities	0.44
Road Damage	0.26
Overgrowth	0.07
Total	-0.03

Living environment problems

* *p<.01, *p<.05

Figure 11. Correlation coefficient between the citizen report rate and the rate of natural population change in each problem category

We calculated the rate of natural population change and the citizen report rate for each problem category for each of the 16 administrative districts composing the three cities and obtained the correlation coefficient between the two in each category, as shown in Figure 11. The rate of natural population change had a positive correlation coefficient with the citizen report rate of living environment problems in every category. On the other hand, it had no correlation with the citizen report rate in other problem categories.

If the correlation described above exists, it then seems possible to estimate the occurrence rate of living environment problems from the rate of natural population change. We thus carried out the simple regression analysis using the sum of the citizen report rates of living environment problems as the dependent variable and the rate of natural population change as the independent variable as shown in Table 3 and derived the regression equation shown in Figure 12.

Regression Statistics	
Multiple R	0.510771072
R Square	0.260887088
Adjusted R Square	0.208093309
Standard Error	10.08909507
Observations	16

(1) y=17.88+1410.59x
y=Citizen report rate (living environment related)
x=Natural population rate

ANOVA

	df	ss	MS	F	Significance F
Regression	1	503.0072685	503.0072685	4.941625524	0.043196552
Residual	14	1425.057752	101.7898394		
Total	15	1928.06502			

	Coefficients	Standard Error	t Stat	P-value	Lower95%	Upper95%
Intercept	17.87872534	3.014678301	5.930558275	3.66875E-05	11.41288346	24.34456723
Rainfall	1410.593009	634.551407	2.222976726	0.043196552	49.61559859	2771.57042

Figure 12. Results of simple regression analysis between the natural population change rate and the total value of citizen report rates for living environment problems

This result shows that the rate of natural population change positively influences the citizen report rate of living environment problems.

Table 3. Variables for the regression in Figure12

Districts		Citizen report rate (Living environment related)	Natural population change rate
Chiba City	Chuo Ward	39.78	-4.89E-05
	Hanamigawa Ward	12.70	-1.42E-03
	Inage Ward	19.33	-4.12E-04
	Wakaba Ward	8.12	-3.58E-03
	Midori Ward	34.70	9.85E-04
	Mihama Ward	32.20	8.98E-04
Hamamatsu City	Naka Ward	8.00	-2.54E-03
	Higashi Ward	7.58	-5.72E-04
	Nishi Ward	5.42	-3.51E-03
	Minami Ward	6.98	-2.11E-03
	Kita Ward	6.78	-5.34E-03
	Hamakita Ward	5.40	-1.72E-03
	Tenryu Wad	4.36	-1.67E-02
Sagamihara City	Midori Ward	9.96	-2.77E-03
	Chuo Ward	14.16	-1.45E-03
	Minami Ward	11.85	-1.38E-03

4.3 Analysis of the influence of the rate of social population change on the recognition of traffic safety problems

As shown in Figure 13, among the common classification, it is considered that the invisibilities is directly linked to the traffic safety problems (hereafter called “the traffic safety problem”). Therefore, the rate of social population change has a positive correlation with the citizen report rate of traffic safety problems.

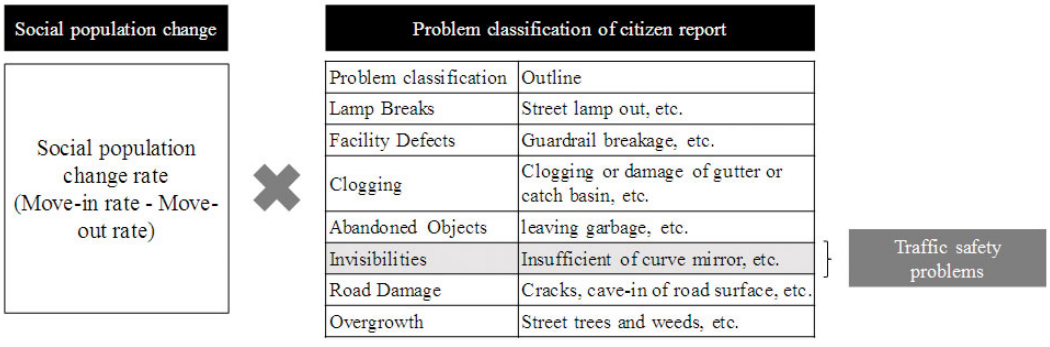


Figure 13. Problem classification related to traffic safety

	Correlation coefficient
Lamp Breaks	0.13
Facility Defects	0.25
Clogging	0.20
Abandoned Objects	0.24
Invisibilities	0.71**
Road Damage	-0.22
Overgrowth	0.20
Total	0.18

Traffic safety problems

**p<.01, *p<.05

Figure 14. Correlation coefficient between the citizen report rate and the social population change rate in each problem category

Using the same method as in 4.2, we calculated the social population change and the citizen report rate for each problem category and obtained the correlation coefficient between the two in each category, as shown in Figure 14. The rate of social population change had a strong positive correlation coefficient with the citizen report rate of traffic safety problems. On the other hand, it had no correlation with the citizen report rate in other problem categories.

Next, as in 4.2, we carried out the simple regression analysis using the sum of the citizen report rates of traffic safety problems as the dependent variable and the rate of social population change as the independent variable as shown in Table 4 and derived the regression equation shown in Figure 15.

Regression Statistics	
Multiple R	0.708028961
R Square	0.50130501
Adjusted R Square	0.465683939
Standard Error	0.83521655
Observations	16

(2) y=1.94+115.89x
y=Citizen report rate (traffic safety related)
x=Social population rate

ANOVA

	df	ss	MS	F	Significance F
Regression	1	9.817327013	9.817327013	14.07327178	0.002147282
Residual	14	9.766213597	0.697586686		
Total	15	19.58354061			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	1.938852418	0.22678332	8.549360749	6.26867E-07	1.452450571	2.425254265
Rainfall	115.8858538	30.89106545	3.751435962	0.002147282	49.63110788	182.1405998

Figure 15. Results of single regression analysis from the social population change rate to the citizen report rates for problems in traffic safety

Table 4. Variables for the regression in Figure15

Districts		Citizen report rate (Traffic safety related)	Social population change rate
Chiba City	Chuo Ward	3.33	0.0103
	Hanamigawa Ward	2.20	0.0023
	Inage Ward	5.58	0.0232
	Wakaba Ward	1.07	0.0057
	Midori Ward	2.89	0.0025
	Mihama Ward	1.82	-0.0036
Hamamatsu City	Naka Ward	1.36	-0.0030
	Higashi Ward	2.32	0.0009
	Nishi Ward	2.11	-0.0022
	Minami Ward	1.99	0.0012
	Kita Ward	1.29	0.0035
	Hamakita Ward	1.12	0.0074
	Tenryu Wad	1.09	-0.0072
Sagamihara City	Midori Ward	2.74	-0.0016
	Chuo Ward	3.09	0.0026
	Minami Ward	2.33	0.0035

This result shows that the rate of social population change positively influences the citizen report rate of traffic safety problems.

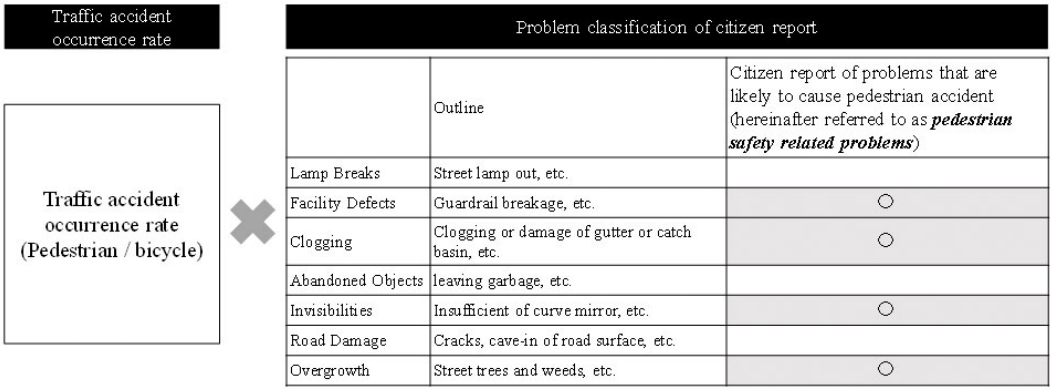
5. Analysis of the Relationship between Citizen Reports and the Incidence of Traffic Accidents

The result of the previous chapter confirms that citizen reports reflect the recognition of problems by citizens about living environment and traffic safety. In addition, citizen reports also seem to reflect various other factors.[22] If the relation between such factors and citizen reports is found, it should be possible to derive new knowledge from data of citizen reports.

ISSN 1927-6974 E-ISSN 1927-6982

For example, it is possible that matters which are precursors of traffic accidents are included in citizen reports by their reporters. If this hypothesis is correct, the incidence of traffic accidents should be high in areas where the citizen report rate of problems which could be causes of traffic accidents is high. In this chapter, we thus examine if there exists any correlation between the two.

It has been pointed out that the contents of citizen reports tend to be more about damage of sidewalks and community roads.[3] Therefore, it seems more possible that citizen reports contain precursors of pedestrian accidents and bicycle accidents. We thus extract from the seven problem categories of "the common categories" the categories which are likely to be causes of pedestrian and/or bicycle accidents, as shown in Figure 16.

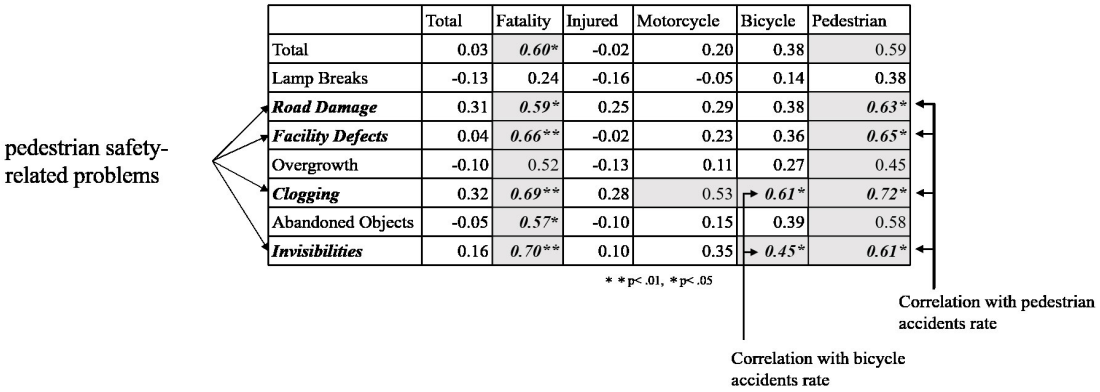


Problem classification of citizen report		
	Outline	Citizen report of problems that are likely to cause pedestrian accident (hereinafter referred to as ***pedestrian safety related problems***)
Lamp Breaks	Street lamp out, etc.	
Facility Defects	Guardrail breakage, etc.	○
Clogging	Clogging or damage of gutter or catch basin, etc.	○
Abandoned Objects	leaving garbage, etc.	
Invisibilities	Insufficient of curve mirror, etc.	○
Road Damage	Cracks, cave-in of road surface, etc.	
Overgrowth	Street trees and weeds, etc.	○

Figure 16. Problem classification related to pedestrian safety

In the same way as done in 4.2, we then calculated the correlation coefficient between the incidence of traffic accidents (hereafter called "the accident incidence") and the citizen report rate, and the result is shown in Figure 17. It is noted that because of limitation on obtaining data, data of some administrative districts was excluded or combined and thus the number of districts for analysis is not the original 16 but 11 for pedestrian accidents and 14 for bicycle accidents.

The incidence of pedestrian accidents was positively correlated with the citizen report rate in all problem categories of the "pedestrian safety related problems" as shown in Figure 17. In addition, the incidence of bicycle accidents was positively correlated with the problem occurrence rates of "clogging" and "invisibilities".



	Total	Fatality	Injured	Motorcycle	Bicycle	Pedestrian
Total	0.03	*0.60**	-0.02	0.20	0.38	0.59
Lamp Breaks	-0.13	0.24	-0.16	-0.05	0.14	0.38
Road Damage	0.31	*0.59**	0.25	0.29	0.38	*0.63**
Facility Defects	0.04	*0.66***	-0.02	0.23	0.36	*0.65**
Overgrowth	-0.10	0.52	-0.13	0.11	0.27	0.45
Clogging	0.32	*0.69***	0.28	0.53	*0.61**	*0.72**
Abandoned Objects	-0.05	*0.57**	-0.10	0.15	0.39	0.58
Invisibilities	0.16	*0.70***	0.10	0.35	*0.45**	*0.61**

Figure 17. Correlation coefficient between the citizen report rate and the accident incidence in each problem category

We thus carried out the simple regression analysis using the sum of the citizen report rates of pedestrian safety related problems as the dependent variable and the pedestrian accident rate as the independent variable as shown in Table 5 and derived the regression equation shown in Figure 18.

Regression Statistics	
Multiple R	0.700690323
R Square	0.490966929
Adjusted R Square	0.434407699
Standard Error	53.42986069
Observations	11

(3) y=19.72+3.45x
y=Citizen report rate (pedestrian safety related)
x=accident incidence of pedestrian

ANOVA

	df	*ss*	*MS*	*F*	*Significance F*
Regression	1	24780.88624	24780.88624	8.680580129	0.016319229
Residual	9	25692.75012	2854.750014		
Total	10	50473.63636			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	19.72326601	28.73208949	0.686454287	0.509723852	-45.27323603	84.71976804
Rainfall	3.453623993	1.172197194	2.946282425	0.016319229	0.801929713	6.105318273

Figure 18. Results of simple regression analysis between citizen report rate for pedestrian safety related problems to the accident incidence of pedestrian

Regression Statistics	
Multiple R	0.545663776
R Square	0.297748956
Adjusted R Square	0.239228036
Standard Error	128.6760453
Observations	14

(4) y=53.35+17.08x
y=Citizen report rate (clogging and invisibilities)
x=accident incidence of bicycle

ANOVA

	df	*ss*	*MS*	*F*	*Significance F*
Regression	1	84243.13308	84243.13308	5.087906254	0.043552159
Residual	12	198690.2955	16557.52462		
Total	13	282933.4286			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	53.35321548	59.70032072	0.8936839	0.389062942	-76.72260924	183.4290402
Rainfall	17.07770427	7.571116666	2.255638768	0.043552159	0.581658146	33.5737504

Figure 19. Results of single regression analysis from the citizen report rates for the problems of clogging and invisibility to the accident incidence of bicycle

Similarly, we carried out the simple regression analysis using the sum of the citizen report rates of clogging and invisibilities as the dependent variable and the rate of bicycle accident as the independent variable as shown in Table 6 and derived the regression equation shown in Figure 19.

Consequently, it has been confirmed that it is possible also to use the citizen report system for estimating from the citizen report rate the accident incidence, which is not the original objective of the system.

6. CONCLUSION

In this study, we have classified citizen reports according to their text contents by applying text mining to them. As a result, we have confirmed that, regardless of the categories set by each municipality, the classification of citizen reports can be automatically done according to their contents by using a common method. This classification method can be applied to other municipalities, and the obtained result was comparable to the classification done by the reporters themselves. Consequently, the method has made it possible to collect and comparatively analyze data across municipalities, which was difficult to perform before.

Furthermore, we have confirmed that citizen reports can be utilized not only for finding road damage but also for analyzing the occurrence trend of local problems concerning the

living environment and traffic safety of citizens. Concretely, the following trends have been found.

(1) The citizen report rate of problems about the living environment becomes higher as the rate of natural population change becomes higher.
(2) The citizen report rate of problems about traffic safety becomes higher as the rate of social population change becomes higher.

Table 5. Variables for the regression in Figure18

Districts		Pedestrian safety related problems	Accident incidence of pedestrian
Chiba City	Chuo Ward	32.73	171
	Wakaba Ward	8.99	70
	Inage/Mihama/ Hanamigawa Ward	58.50	210
	Midori Ward	21.96	42
Hamamatsu City	Naka Ward	17.02	207
	Higashi Ward	20.19	79
	Nishi Ward	12.78	54
	Minami Ward	10.67	52
	Kita Ward	17.98	47
	Hamakita Ward	10.09	47
	Tenryu Ward	12.34	9

Table 6. Variables for the regression in Figure19

Districts		Bicycle safety related problems	Bicycle incidence of pedestrian
Chiba City	Chuo Ward	8.46	196
	Wakaba Ward	3.20	71
	Inage/Mihama/ Hanamigawa Ward	21.96	420
	Midori Ward	7.11	54
Hamamatsu City	Naka Ward	5.74	497
	Higashi Ward	5.88	197
	Nishi Ward	4.05	133
	Minami Ward	5.98	124
	Kita Ward	4.74	92
	Hamakita Ward	3.67	95
	Tenryu Ward	2.90	10
Sagamihara City	Chuo Ward	5.99	66
	Minami Ward	5.01	304
	Midori Ward	5.54	29

In addition, it has also been confirmed that citizen reports can be used for presuming places of a high possibility of traffic accidents, which was also not initially expected.

It is possible that the use of the above knowledge will allow municipalities to perform more efficient operations, for example, by centralizing their measures such as patrolling into the areas of many citizen reports. On the other hand, the automatic classification rules established in this study are only for the purpose of deriving useful knowledge from past data of citizen report and are not intended to classify individual reports in real time for daily operations such as road maintenance. This is the current limit of this study. However, if the error rate of 12.4% shown in 3.4 can be reduced to a practically acceptable level by creating a dictionary for the classification, it can be used for real time classification in daily operations.

It is also possible that adding citizen report data of more municipalities to the existing data will make the data more versatile and that more conditions can be included as the total number of citizen reports increases. Consequently, it seems that the accuracy of estimation under each condition improves, making it possible to further increase their practicality.

While the data of traffic accidents used in this study was not enough, the National Police Agency is currently in the process of making some data of the prefectural police available to the public for free and secondary use in a common format throughout Japan.[23] If the government and municipalities take further measures for opening their data, it will be possible to obtain various data concerning the situations of the living environment and traffic safety, including data of traffic accidents. It will then possible to obtain more accurate estimations, and it is also expected that totally new knowledge will be obtained for solving various local problems.

The basic principle of this method is not premised on Japanese, so it can be applied to other languages and cultures. Specifically, although extracting the combination of dependency in 3.3.2 and removing noises in 3.3.3 are processed in the manner specific to Japanese, if they were replaced by a method suitable for other languages, this entire research method can be applied to any languages and cultures.

Such future work will make it possible to build more general classification methods that also takes into account problem recognition specific to language and culture, and make comparisons between countries and cultures using this method.

REFERENCES

[1] Ministry of Internal Affairs and Communications. The Second Report of the study group for designing local government strategy toward 2040. http://www.soumu.go.jp/menu_news/s-news/01gyosei04_02000068.html Accessed April 1, 2019.

[2] Hideto Mashita. Present state and future prospects of road structure inspection technology. Civil Engineering Resource. 2015; 57-8: 4-5.

[3] Chiba repo (Citizen Collaboration Report). 2018, https://scirex.grips.ac.jp/news/4128f99ed10dde4cadde000113f352d8649b66eb.pdf Accessed August 17, 2019.

[4] King SF, Brown P. Fix My Street or Else: Using the Internet to Voice Local Public Service Concerns. ICEGOV2007. December 2007; 10–13; 72–80.

[5] Burcu Baykurt. Redefining Citizenship and Civic Engagement: political values embodied in FixMyStreet.com, AoIR Selected Papers of Internet Research; IR 12, 2011.

[6] Nils Walravens. Validating a Business Model Framework for Smart City Services: The Case of FixMyStreet, paper presented at the 27th IEEE International Conference on Advanced Information Networking and Applications Workshops (Barcelona, Spain, 2013)

[7] Berntzen l, Johannessen MR, Böhm S, et al. Citizens as Sensors: Human Sensors as a Smart City Data Source. SMART 2018, The Seventh International Conference on Smart Cities, Systems, Devices and Technologies, Barcelona, Spain. 2018. p. 11–18.

[8] Hijikata H. Solve Problems in Community and Town by Smartphone: Data Driven Management via 'FixMyStreet Japan.' 2018 https://www.nttdata-strategy.com/aboutus/newsrelease/180705/ Accessed April 1, 2019.

[9] Kinzoku Yoshihiko. A New Communication Tool to Connect Citizens and Government. Administration & Information Systems. 2015; 51(2): 8–14.

[10] Honda M. Estimation of the Penetration Process of Administrative Application Based on the Registrant Information of 'Chiba Repo' Converted to Open Data. Journal of Information and Knowledge Society. 2016; 26(2): 187–194. https://doi.org/10.2964/jsik_2016_018

[11] Nakatogawa S, Seto T. The Possibilities of Citizen Participatory GIS from the View of Difference in Problem Recognition Based on Social Attributes: In the Case of "Chiba Repo." Proceedings of the Geography Society of Japan. 2016; S(0).

[12] Sano Y, Yamaguchi K, Mine T. Automatic Classification of Complaint Reports about City Park. Information Engineering Express International Institute of Applied Informatics. 2015; 1(4): 119–130.

[13] Chiba City. Chibarepo "Chiba Citizen Collaboration Report". https://www.city.chiba.jp/shimin/shimin/kohokocho/chibarepo.html Accessed April 1, 2019.

[14] Nasukawa T. Technology of Using/Making Text Mining: The Essence and Utilization Method Derived from Basic Technology and Application Cases, Tokyo Denki University Press; 2006.

[15] Kida M. Introduction of text mining -Usage in management research, Hakuto shobo; 2008.

[16] Shinya H, Katsuyuki H, Shinji T. Systematic analysis using text mining in organizational research. Journal of Japan Society for Information and Management. 2012; 32(3): 97-109.

[17] Akihiro S. Application of Textmining in Japan, The society for economic studies, The University of Kitakyushu, Working Paper Series; 2012. http://www.kitakyu-u.ac.jp/economy/study/pdf/2011/2011_11.pdf Accessed April 1, 2019.

[18] Masahiro M, Asako M. Text Mining for Humanities and Social Sciences [revised edition]. Seishin Shobo. 2014.

[19] Sagamihara City, Road reporting Application "Pattorun" operation report. http://www.city.sagamihara.kanagawa.jp/patrun/1012744/index.html August 17, 2019.

[20] Hamamatsu City. civil engineering report system "Icchao" <https://www.city.hamamatsu.shizuoka.jp/dourohozen/icchao.html> Accessed August 29, 2019.

[21] Cabinet Office of Japan. "Selected Future" Committee Report Commentary / Documents, Chapter 3, A8-Factors of population change https://www5.cao.go.jp/keizai-shimon/kaigi/special/future/sentaku/s3_1_8.html Accessed August 30, 2019.

[22] What kind of places are likely to encounter traffic accidents? https://www.jicobengo.com/knowledge/traffic-accident-where.html Accessed August 30, 2019.

[23] National Police Agency of Japan, About promotion of open data of crime occurrence information. Available from: https://www.kantei.go.jp/jp/singi/it2/senmon_bunka/data_ryutsuseibi/opendata_wg_dai7/odwg_siryou1-3.pdf Accessed August 30, 2019.