

Building Urban LOD for Solving Illegally Parked Bicycles in Tokyo

Shusaku Egami¹, Takahiro Kawamura^{1,2}, and Akihiko Ohsuga¹

¹ Graduate School of Informatics and Engineering,
The University of Electro-Communications, Tokyo, Japan
`egami.shusaku@ohsuga.lab.uec.ac.jp`
`ohsuga@uec.ac.jp`

² Japan Science and Technology Agency
`takahiro.kawamura@jst.go.jp`

Abstract. The illegal parking of bicycles is an urban problem in Tokyo and other urban areas. The purpose of this study was to sustainably build Linked Open Data (LOD) for the illegally parked bicycles and to support the problem solving by raising social awareness, in cooperation with the Bureau of General Affairs of Tokyo. We first extracted information on the problem factors and designed LOD schema for illegally parked bicycles. Then we collected pieces of data from Social Networking Service (SNS) and websites of municipalities to build the illegally parked bicycle LOD (IPBLOD) with more than 200,000 triples. We then estimated the missing data in the LOD based on the causal relations from the problem factors. As a result, the number of illegally parked bicycles can be inferred with 70.9% accuracy. Finally, we published the complemented LOD and a Web application to visualize the distribution of illegally parked bicycles in the city. We hope this raises social attention on this issue.

Keywords: Linked Open Data, Urban Problem, Illegally Parked Bicycles

1 Introduction

An increased awareness of health problems [1] and energy conservation [2] led to a 2.6-fold increase in bicycle ownership in Japan from 1970 to 2013. Consequently, illegally parked bicycles around railway stations have become an urban problem in Tokyo and other urban areas. In addition to the insufficient availability of bicycle parking spaces, inadequate public knowledge on bicycle parking laws has contributed to this urban problem. Illegally parked bicycles obstruct vehicles, cause road accidents, encourage theft, and disfigure streets.

In order to address this problem, we believe it would be useful to publish the distribution of illegally parked bicycles as Linked Open Data (LOD). For example, it would serve to visualize illegally parked bicycles, suggest locations for optimal bicycle parking spaces, and assist with the removal of illegally parked

bicycles. However, Open Data sets available for illegally parked bicycles are currently distorted, and it is difficult for services to utilize the data. In addition, other data concerning issues such as bicycle parking spaces and government statistics, have been published in a variety of formats. Hence, unification of the data formats and definition of schema for data storage are important issues that need to be addressed. Bischof et al. [3] proposed a method for integrating Open City Data as Linked Data and proposed methods for the complementation of missing values. The study improved the utilization of unreusable Open Data. However, more spatio-temporal data and factor data are necessary to develop services for combating illegally parked bicycles.

In this study, we first extracted domain requirements of illegally parked bicycles from articles on the Web and design LOD schema. Next, we collected data about illegally parked bicycles from Twitter and the data describing factors that affect the number of illegally parked bicycles. In order to reuse these data sets, which have different formats, we unify the data formats based on designed schema and publish the data on the Web as LOD. Moreover, we estimate the missing data (the number of illegally parked bicycles) based on the causal relations from the factors. Our predictions take into consideration factors such as time, weather, nearby bicycle parking information, and nearby points of interest (POIs). However, since there are cases that lack these factor values, the missing factor values are also complemented by searching similar observation data. We thus use Bayesian networks to estimate the number of illegally parked bicycles for data sets after complementation of the factors. These results are also incorporated to build LOD with a specified property. In addition, we develop a service that visualizes the illegally parked bicycles using the constructed LOD. This visualization service raises the awareness of the issue in local residents and prompts users to provide more information about illegally parked bicycles. Therefore, our contributions are as follows.

1. Proposal of a methodology for designing LOD schema for an urban problem
2. Collection of data from SNS and municipalities of Tokyo and other urban areas, and the building of illegally parked bicycle LOD (IPBLOD)
3. Development and evaluation of an approach for complementing the missing factor values and estimating the missing values
4. Development and publishing of a Web application for visualizing illegally parked bicycles in Tokyo and other urban areas

The remainder of this paper is organized as follows. In Section 2, related works of data collection and urban LOD are described. In Section 3, the methodology for designing the LOD schema and IPBLOD are presented. In Section 4, two approaches that complement the missing factors, and estimate the illegally parked bicycles using Bayesian networks, are described. Also, we evaluate our results and our findings. In Section 5, visualization of the IPBLOD is described. Finally, Section 6 concludes this paper with future works.

2 Related Work

In most cases, LOD sets have been built based on existing databases. However, there is little LOD available so far that describes urban problems. Thus, methods for collecting new data to build urban LOD are required. Data collection methods for building Open Data include crowdsourcing and gamification. A number of projects have employed these techniques. OpenStreetMap [4] is a project that creates an open map using crowdsourced data. Anyone can edit the map, and the data are published as Open Data. Similarly, FixMyStreet [5] is a platform for reporting regional problems such as road conditions and illegal dumping. Crowdsourcing to collect information in FixMyStreet has meant that regional problems are able to be solved more quickly than ever before. Zook et al. [6] reported a case, where crowdsourcing was used to link published satellite images with OpenStreetMap after the Haitian earthquake. A map of the relief efforts was created, and the data were published as Open Data. Celino et al. [8] proposed an approach for editing and adding Linked Data using a game with a purpose (GWAP) [7] and human computation. However, since the data concerning illegally parked bicycles are time-series data, it is difficult to collect data using these approaches. Therefore, new techniques are required. We propose a method to build urban LOD while complementing the missing data.

Also, there have been studies about building Linked Data for cities. Lopez et al. [9] proposed a platform that publishes sensor data as Linked Data. The platform collects streamed data from sensor and publishes Resource Description Framework (RDF) in real time using IBM InfoSphere Stream and C-SPARQL [10]. The system is used in Dublinked2³, which is a data portal of Dublin, Ireland, that publishes information of bus routes, delays, and congestion updated every 20 seconds. However, since embedding sensors is costly, this approach is not suitable for our study.

Furthermore, Bischof et al. [3] proposed a method for the collection, complementation, and republishing of data as Linked Data, as with our study. This method collects data from DBpedia [13], Urban Audit⁴, United Nations Statistics Division (UNSD)⁵, and U.S. Census⁶ and then utilizes the similarity among such large Open Data sets on the Web. However, we could not find the corresponding data sets and thus could not apply the same approach to our study.

3 Building LOD

In this study, we propose a method for sustainably building urban LOD and applying them to Tokyo and other urban areas. Managing urban problem data *joining* multiple tables in (distributed) RDBs is troublesome from the aspect of data interoperability and maintenance, since the urban problem is closely related

³ <http://www.dublinked.ie/>

⁴ <http://ec.europa.eu/eurostat/web/cities>

⁵ <http://unstats.un.org/unsd/default.htm>

⁶ <http://www.census.gov/>

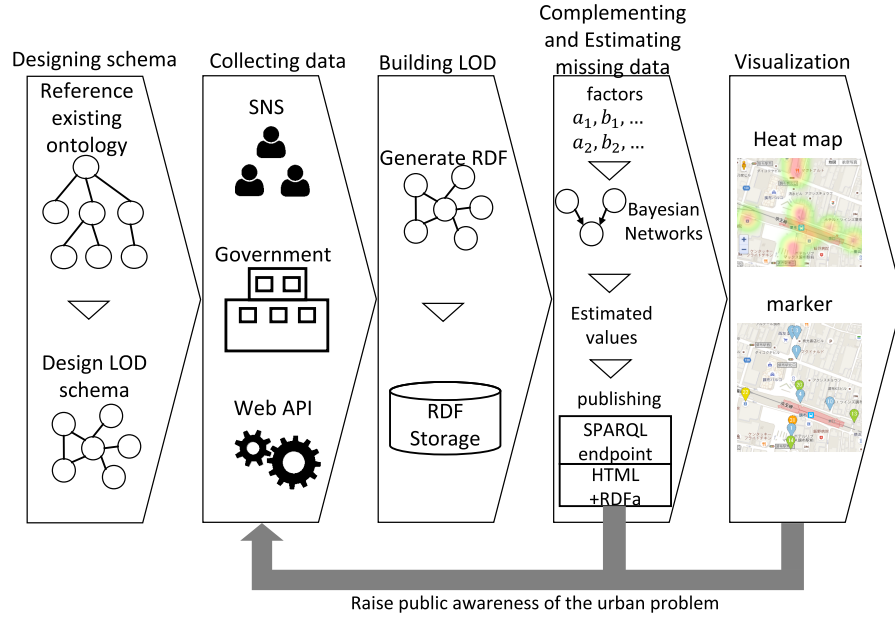


Fig. 1. Overview of this study

to multiple domains, such as government data, legal data, and social data as we already incorporated POIs and weather data in this application, and also those have different schemata. Thus, Linked Data is a suitable format as the data infrastructure of not only illegally parked bicycles, but also urban problems in general, since Linked Data can have advantages of flexible linkability and schema.

Figure 2 provides an overview of this study. This study is divided into the following five steps. Steps (2) to (5) are executed repeatedly as more input data become available.

1. Designing LOD schema
2. Collecting observation data and factor data
3. Building the LOD based on schema
4. Using Bayesian networks to estimate the missing number of illegally parked bicycles at each location
5. Visualizing illegally parked bicycles using LOD

3.1 A methodology for designing LOD schema

Illegally parked bicycles can be observed by social sensors, as it is difficult to install physical sensors in the streets. In our previous work [11], the schema for IPBLOD was based on the Semantic Sensor Network ontology. However, in order to address this urban problem using LOD, the LOD should not only

Table 1. Results of clustered keywords

Property	Category	Keyword
place and factor	POI	shopping street, large-scale retail stores, public office, school, library, public hall, department store, supermarket, bank, amusement, pachinko, commercial building
	bicycle parking	usability, safety, comforts, supply, fee
	train	train, distance from home to the station
	storage space	capacity
time and factor	time	weekday, holiday, hour
product	accident	congestion, blocking vehicle traffic
factor	objective	commutation, destination
	weather	rainy
	bicycle	price reduction, deterioration, cost of maintenance, variety

have the number of illegally parked bicycles, location and time information but also contain the factors related to illegally parked bicycles, such as POIs and weather. In this paper, we present an LOD schema, including the factors related to the illegally parked bicycles, and we propose a methodology for designing LOD schema for urban problems, such as illegally parked bicycles.

In the ontology study, the methodology for building ontology has been discussed. We propose the methodology for designing practical LOD schema in reference to Activity-First Method [12]. The schema for IPBLOD is designed based on this methodology, which consists of two steps as follows:

1. Extraction of domain requirements
 - a. Select an ontology that models the urban problem
 - b. Search for articles on the urban problem using a search engine
 - c. Extract keywords from the articles based on properties of the ontology
 - d. Cluster the keywords
2. Designing schema
 - a. Design classes based on the ontology
 - b. Design instances and properties based on the result of the clustering

First, the existing ontology is selected in order to build LOD based on the ontology. To serve as a source of this problem, it is necessary to consider the accessibility of the LOD as well as the semantic consistency. Thus, we select Event Ontology (EO)⁷ as a practical and intuitive structure wherein illegally parked bicycles can be considered as an event. In the EO, an event class has properties for place, time, agent, factor, and product.

⁷ <http://pur1.org/NET/c4dm/event.owl>

Next we search for articles on illegally parked bicycles using Google. Then we investigate the top 10 articles and their references and then manually extract the keywords based on the properties of the existing ontology. Specifically, keywords are extracted from sentences that describe the place, time, agent, factor, and product. Even if keywords that are not defined in the ontology appear to be important in the article, the keywords are also extracted.

The extracted keywords are clustered manually as in Table 1. Then the classes are designed based on the EO. The expression of Description Logic (DL) is as follows:

`IllegallyParkedBicycles` \sqsubseteq `Event`
`IllegallyParkedBicycles` \sqsubseteq \exists `place.SpatialThing`
`IllegallyParkedBicycles` \sqsubseteq \exists `time.TemporalEntity`
`IllegallyParkedBicycles` \sqsubseteq \exists `weather.WeatherState`
`IllegallyParkedBicycles` \sqsubseteq \exists `factor.Thing`
`IllegallyParkedBicycles` \sqsubseteq \exists `agent.Agent`
`IllegallyParkedBicycles` \sqsubseteq \exists `product.Thing`
`IllegallyParkedBicycles` \sqsubseteq \exists `value.Integer`

The `IllegallyParkedBicycles` class refers to a set of illegally parked bicycles, and it is a subclass of the `Event` class. `IllegallyParkedBicycles` contains the place, time, weather, agent, factor, product, and number of illegally parked bicycles. Weather and the number of illegally parked bicycles are not defined in the EO, but since these are considered to be important in the domain of illegally parked bicycles, we add them to the LOD schema.

Then we design their instances in reference to Table 1. Figure 2 shows an overview of LOD schema with the instances. In Table 1 the column of category refers to the instance, and the instance is linked with the EO property. Also, some instances are linked to other instances. However, it is difficult to obtain data on individuals, who park their bicycles illegally. Therefore, we omit the `event:agent` property. However, we use population statistics data instead of the data on individuals. We obtain population statistics from the portal site of official statistics of Japan called e-stat⁸, and we use “population density per 1 square kilometer of habitable area” and “the number of commuters who use trains.” In fact, there are a large number of illegally parked bicycles near the stations, which are located in densely populated areas in Japan. Many bicycles are illegally parked during the morning-commuting hours. Also, we omit the `event:product` property, since it is difficult to obtain data on accidents caused by illegally parked bicycles. In the same way the storage space, the objective of the person who parked the bicycle illegally, and the price of the bicycle are also omitted. Moreover, the POI, nearest bicycle parking, nearest train station, time, and weather are added to the LOD schema as factors related to illegally parked bicycles. The nearest train station is a resource of DBpedia Japanese⁹.

⁸ <http://www.e-stat.go.jp/SG1/estat/eStatTopPortalE.do>

⁹ <http://ja.dbpedia.org>

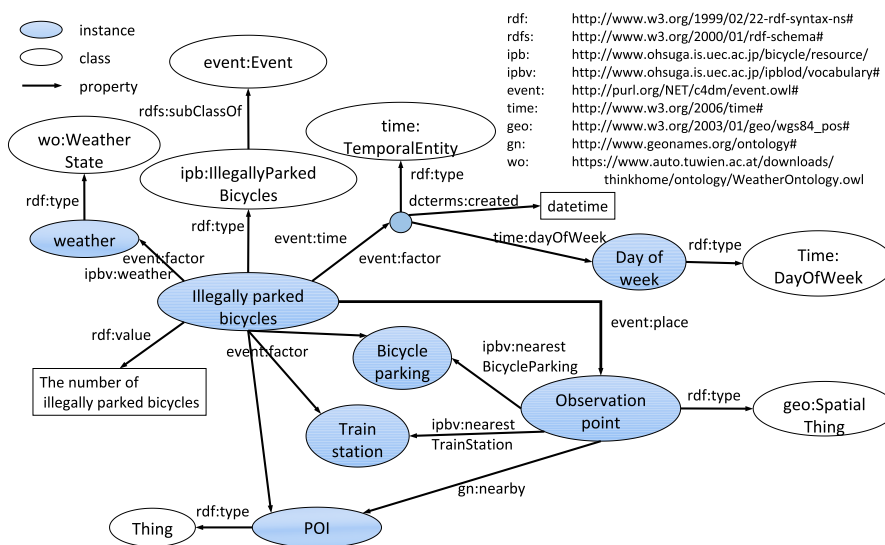


Fig. 2. LOD schema containing instances

3.2 Collection of observation data

We started this study by collecting tweets containing location information, pictures, hash-tags, and the number of illegally parked bicycles. However, obtaining the correct locations from Twitter was difficult, since mobile phones often attach incorrect location information. Mobile phones are equipped with inexpensive GPS chips, and it is known that the accuracy is often low due to weather conditions and GPS interference area [15]. To address this problem, we developed a Web application that enables users to post tweets on Twitter after correcting their location information, and we made an announcement asking public users to post tweets of illegally parked bicycles using our application. Figure 3 shows a screen shot of this application. After OAuth authentication, a form and buttons are shown. When the location button is pressed, a marker is displayed at the user's current location on a map. The marker is draggable, thus allowing users to correct their location information. When users add their location information, enter the number of illegally parked bicycles, take pictures, and submit them, then tweets including this information with a hashtag are posted.

Furthermore, we collected information on POI using Google Places API¹⁰ and Foursquare API¹¹. Also, we obtained bicycle parking information from websites of municipalities and in cooperation with the Bureau of General Affairs of Tokyo¹². The Bureau of General Affairs of Tokyo publishes Open Data on

¹⁰ <https://developers.google.com/places/?hl=en>

¹¹ <https://developer.foursquare.com/>

¹² <http://www.soumu.metro.tokyo.jp/30english/index-en.htm>



Fig. 3. Screenshot of the tweet application

bicycle parking areas as CSV. The data contain names, latitudes, longitudes, addresses, capacities, and business hours. More information was collected from municipalities, for example, monthly parking fees and daily parking fees. Also, we retrieved weather information from the website of the Japanese Meteorological Agency (JMA)¹³.

3.3 Building LOD based on designed schema

The collected data on illegally parked bicycles are converted to LOD based on the designed schema. Figure 4 shows the process of building IPBLOD. First, the server program collects tweets containing the particular hash-tags, the location information, and the number of illegally parked bicycles in real time. The number of illegally parked bicycles is extracted from the text of tweets using regular expressions.

Next, the server program checks whether there is an existing observation point within a radius of less than 30m by querying our endpoint¹⁴ using the SPARQL query. If there is no observation point on the IPBLOD, the point is added as a new observation point. In order to add new observation points, the nearest POI information is obtained using Google Places API and Foursquare API. The new observation point is generated based on the name of the nearest POI. It is possible to obtain the types of the POI from Google Places API and

¹³ <http://www.jma.go.jp/jma/indexe.html>

¹⁴ <http://www.ohsuga.is.uec.ac.jp/sparql>

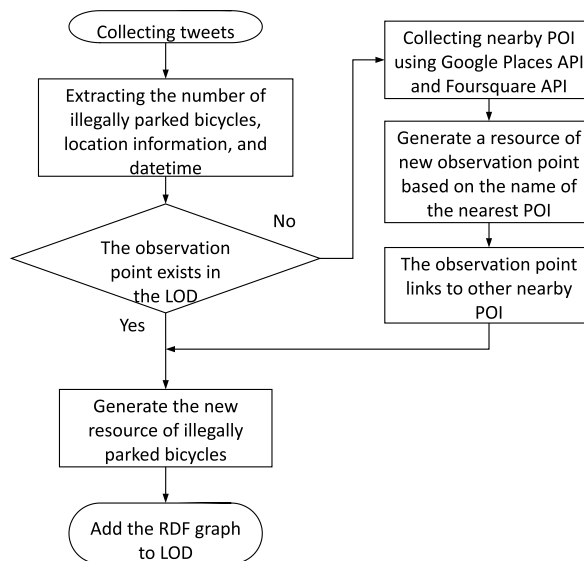


Fig. 4. Process of LOD building

Foursquare API. We map the types of POI to classes in LinkedGeoData [16]. Thus, the POI is an instance of classes in LinkedGeoData. However, some POIs do not have a recognized types. Therefore, their types are decided by a keyword search with the name of the POI.

Then the address, prefecture’s name, city name, and land lot name are obtained using Yahoo! reverse geocoder API, and then the links to GeoNames.jp¹⁵ are generated based on the obtained information. GeoNames.jp is a Japanese geographical database. Thus, data are collected and added to the IPBLOD using Web APIs in real time. Figure 5 shows part of the IPBLOD. The LOD are stored in Virtuoso¹⁶ Open-Source Edition. Also, the RDF data set is published with CC-BY license on our website¹⁷.

4 Complementing and Estimating Missing Values

Since we rely on the public to observe illegally parked bicycles, we do not have round-the-clock data for every place, and thus, missing data in the IPBLOD are inevitable. However, the number of the illegally parked bicycles should be influenced by several factors, thus we try to estimate these missing data using Bayesian networks. If the data is expanded in density through the estimation, it will serve, for example, as the suitable location of bicycle parking spaces, the

¹⁵ <http://geonames.jp/>

¹⁶ <http://virtuoso.openlinksw.com/>

¹⁷ <http://www.ohsuga.is.uec.ac.jp/bicycle/dataset.html>

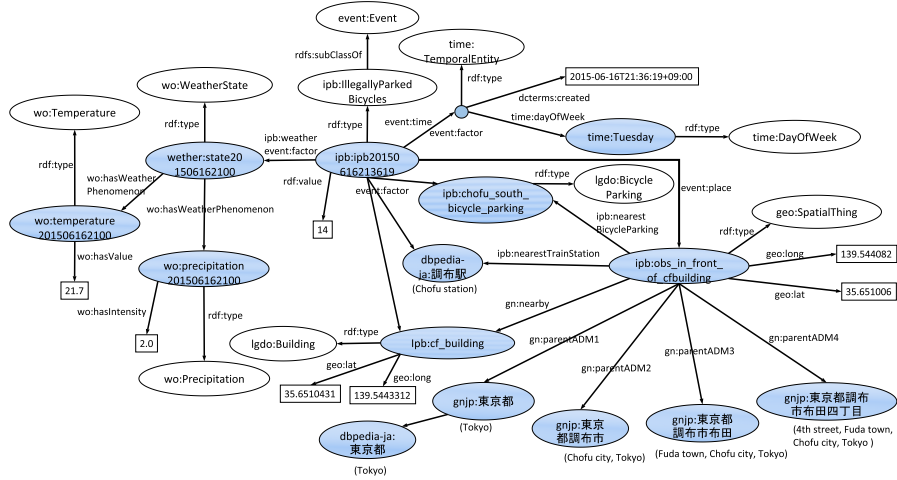


Fig. 5. Part of the integrated LOD

decision on variable prices of the parking fee and efficient timing of removal of illegally parked bicycles by the city, and part of the references for future urban design.

4.1 Complementing missing factors

As the factors (attributes), we use observation points, day of week, hours, precipitation, temperature, daily fee for the nearest bicycle parking, monthly fee for the nearest bicycle parking, “population density per 1 square kilometer habitable are,” “the number of commuters who use trains,” and types of POIs. We selected Building, Bank, Games, DepartmentStore, Supermarket, Library, Police, and School as the types of POIs based on Table 1. However, there are also missing factor values. We assume that the missing factor values are similar to the corresponding factor value in the similar observation data. Therefore, we used the factor values of the similar observation data as substitutes for actual values which cannot be obtained. The similar observation data are found using Jaccard coefficient. Suppose the aggregates of each factor are given by Location, Day={sun, mon,...,sat}, Hour={0,1,...,23}, Precipitation={0,1,...}, Temperature={...,-1,0,1,...}, DailyFee={0,1,...}, MonthlyFee={0,1,...}, Density={0,1,...}, Commuters={0,1,...}, Building, Bank, DepartmentStore, Games, Supermarket, Library, Police, School={0,1}, and Number (of illegally parked bicycles)={1,...,4}, then the observation data are stored as an aggregate O of vectors $o \in Location \times Day \times Hour \times Precipitation \times Temperature \times DailyFee \times MonthlyFee \times Density \times Commuters \times Building \times Bank \times DepartmentStore \times Games \times Supermarket \times Library \times Police \times School \times Number$. The number of illegally parked bicycles is classified into four classes by Jenks natural breaks

[14], which are often used in Geographic Information Systems (GISs). The range is 0 to 6, 7 to 17, 18 to 35, and 36 to 100. Therefore, the similarity of observation data o_1 and o_2 is $sim(o_1, o_2) = |o_1 \cap o_2| / |o_1 \cup o_2|$.

4.2 Estimating the number of illegally parked bicycles using Bayesian network

We then estimate the number of illegally parked bicycles, at observation points, where the number data are missing. The input dataset is the dataset complemented using the method described in Section 4.1. We use the Bayesian network tool Weka¹⁸ to estimate the unknown numbers of illegally parked bicycles. There are 897 observation data. The input data is a set O that consists of vectors with eight elements at first. We used HillClimber as a search algorithm, and also used Markov blanket classifier. The maximum number of parent nodes was two. As a result of 10-fold cross validation, we got 65.2% accuracy.

To raise the accuracy, we focused on types of POIs. We did not restrict the types of POIs when building the IPBLOD, but we restricted types to the POIs contained in Table 1 when estimating the number of illegally parked bicycles using Bayesian networks. However, other POIs could become factors related to illegally parked bicycles. Hence, we first used all POI types as factors, and the number of POI types became 68. However, the accuracy became relatively low due to too many factors. Thus, we used super classes in LinkedGeoData ontology for clustering those types. Since we mapped the POI types to classes of LinkedGeoData, it was possible to obtain their super classes by querying the LinkedGeoData. As the result, the number of POI types became 46, as follows.

Pharmacy , Park , Retail , Restaurant , Police , University ,
 FastFood , BusStation , Gym , Parking , Church , Florist , Cafe ,
 Supermarket , Hospital , Nightclub , Sport , Advertising , Casino ,
 Hairdresser , Doctor , Bar , Bakery , Bank , TakeAway , Amenity ,
 Dentist , EmergencyThing , Hall , Office , Residential ,
 PlaceOfWorship , School , CommunityCentre , Building , Spa ,
 CarRental , VideoRental , Hotel , Cinema , CoffeShop , Construction
 , Lawyer , HighwayThing , Shop , PublicTransportThing

Therefore, an observation datum became a vector $o \in Location \times Day \times Hour \times Precipitation \times Temperature \times DailyFee \times MonthlyFee \times Density \times Commuters \times Pharmacy \times \dots \times PublicTransportThing \times Number$, which resulted in 56 possible elements. Finally, the average estimation accuracy of ten times 10-fold cross validation became 70.9%. The maximum number of parent nodes was seven, after random sampling with a 90% rate. We estimated the number of illegally parked bicycles on unobserved dates using the above parameters. Specifically, we examined the observation data in each observation point from the first observation date to the last observation date. If there are no data at 9 am or 9 pm, we estimated and complemented the number of illegally parked bicycles. Then, we added the estimated number and its probability to IPBLOD as follows.

¹⁸ <http://www.cs.waikato.ac.nz/ml/weka/>

```

@prefix ipb: <http://www.ohsuga.is.uec.ac.jp/ipblod/
    vocabulary#>
@prefix bicycle: <http://www.ohsuga.is.uec.ac.jp/bicycle/
    resource/>
bicycle:ipb_{observation point}_{datetime}
    ipb:estimatedValue [ rdf:value "0-7" ;
        ipb:probability "0.772"^^xsd:double ] .

```

4.3 Evaluation and Discussion

The observational data were collected from January 2015 to April 2016. Eighteen users posted data on Twitter using our application in Figure 3. The number of triples included in the IPBLOD became more than 200,000. Table 2 shows statistics about the observation data. Observation points are places at which someone observed the bicycles, and the amount of observation data is the total number of submitted data (tweets), in which the same observation points may appear several times. As the result, Chofu City in Tokyo had the largest amount of the observation data. Since we posted promotional tweets on our Twitter accounts, contributors from our university increased.

There are 219 pieces of observation data that have missing factor values, and these values have been complemented using the method discussed in Section 4.1. The missing factors are found in the daily monthly fees of the nearest bicycle parking since the municipalities publish information on bicycle parking in different details. Also, the missing factors are found in precipitation and temperature values since there are not the source data in JMA.

Also, we found that LOD is also useful for constructing probabilistic networks for the estimation since possible nodes in the network can be obtained by following the properties like `event:factor` and ontological hierarchies. As the result of 10-fold cross validation repeated ten times, the precision became 69.9%, the recall became 70.9%, and the F-measure became 69.7%. The precision is the ratio of correct data in the estimated data. The recall (accuracy) is the correctly estimated data divided by correct data. The training data are 897 observation data and their attributes. The accuracy of the estimated data in this study was low for the following reasons. The number of observations was not very many, and it was also unbalanced. Also, Table 3 shows a confusion matrix. The amount of 36-100 data is few; thus, this class of data is not correctly estimated.

Also, since we did not define the range of observation points, there were differences of the range decisions for each person. It was found that some people tweeted many illegally parked bicycles at one times, while some people divided the illegally parked bicycles and tweeted individually. Thus, data for 0-6 and 7-17 were higher, and this fact affected the estimation accuracy. We plan to visualize a specified range of circles which indicate observation points in the tweet application. Also, we plan to add a selection button such as “low (less than 10)” and “high (greater than 30)” in order to reduce the work burden.

Table 2. Statistics for observation data

Area	# of observation points	Amount of observation data
Chofu City, Tokyo	19	673
Nerima City, Tokyo	4	96
Naka-ku, Yokohama City, Kanagawa	5	39
Fussa City, Tokyo	2	23
Fuchu City, Tokyo	7	20
Musashino City, Tokyo	4	16
Chuo-ku, Sapporo City, Hokkaido	9	14
Shinjuku City, Tokyo	3	4
Isogo-ku, Yokohama City, Kanagawa	2	3
Kokubunji City, Tokyo	2	3
Shibuya City, Tokyo	3	3
Kita-ku, Sapporo City, Hokkaido	2	2
Shinagawa City, Tokyo	1	1

Table 3. Estimated results

		0-6	7-17	18-35	36-100	Total
Estimated data	0-6	339	38	10	2	389
	7-17	62	158	19	2	241
	18-35	22	34	89	2	147
	36-100	2	16	5	5	28
TP		0.871	0.656	0.605	0.179	

5 Visualization of LOD

Data visualization enables people to intuitively understand data contents. Thus, it can possibly raise the awareness of an issue among local residents. Furthermore, it is expected that we shall collect more urban data. In this section our visualization application of the IPBLOD is described.

The IPBLOD are published on the Web, and a SPARQL endpoint¹⁹ is also available. Consequently, anyone can download and use IPBLOD as APIs via the SPARQL endpoint. As an example of the use of these data, we developed a Web application that visualizes illegally parked bicycles. The application can display time-series changes in the distribution of illegally parked bicycles on a map. Also, the application has a responsive design, so it is possible to use it on various devices such as PCs, smartphones, and tablets. When the start and end times are selected, and the play button is pressed, the time series changes of the distribution of the illegally parked bicycles are displayed. Figure 6.(a) and 6.(b) show screenshots of an Android smartphone, on which the Web application is displaying such an animation near Chofu Station in Tokyo using a heatmap and

¹⁹ <http://www.ohsuga.is.uec.ac.jp/sparql>

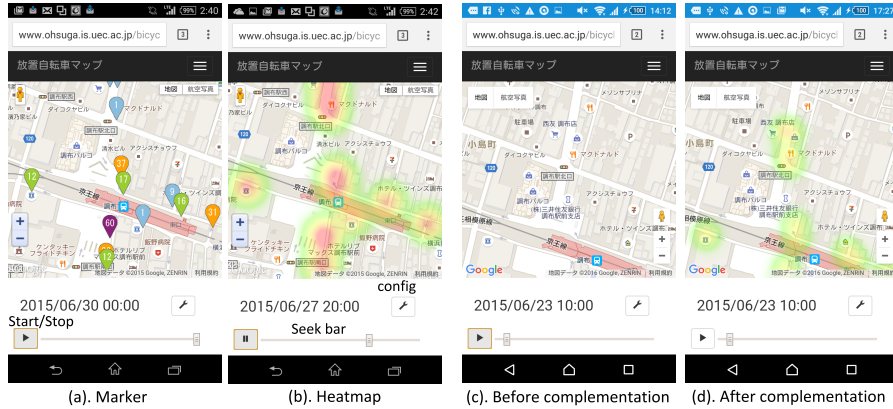


Fig. 6. Screenshots of the visualization application

a marker UI. This visualization application and the tweet application in Figure 3 are hosted on the same website, so it is possible to see the visualized information just after tweeting. Thus, users obtain the instant feedback on posting new data.

The IPBLOD contain not only the data collected from Twitter, but also the data estimated by Bayesian networks. Therefore, time-series changes in the distribution of illegally parked bicycles become smoother than before estimating the missing values. Figure 6.(c) and 6.(d) show the comparison between the before and after complementation. The time-series changes after complementation are successive, whereas the time-series changes before complementation are intermittent.

As another example, we can see the average number of illegally parked bicycles per hour using the short SPARQL query. Figure 7 shows a visualization of the result. We found from the result that there are more illegally parked bicycles at night rather than in the morning. In general, many bicycles are thought to be illegally parked during morning commuting hours. However, the opposite result was shown in this study.

The number of page views of the visualization application increased from January to April 2016 (in Figure 8). The number of page views of the visualization application was 187 in January 2016, and the number of page views gradually increased to 705 in April 2016. Also, the average session duration is 2 minutes and 32 seconds. Therefore, it was found that visitors are increasing and tend to use our application for longer periods of time.

6 Conclusion and Future Work

In this paper, the building and visualization of urban LOD was described as a solution for an illegally parked bicycles problem. The techniques proposed were a methodology for designing LOD schema from pages about an urban problem, data collection from Twitter with the exact location, a schema design of the

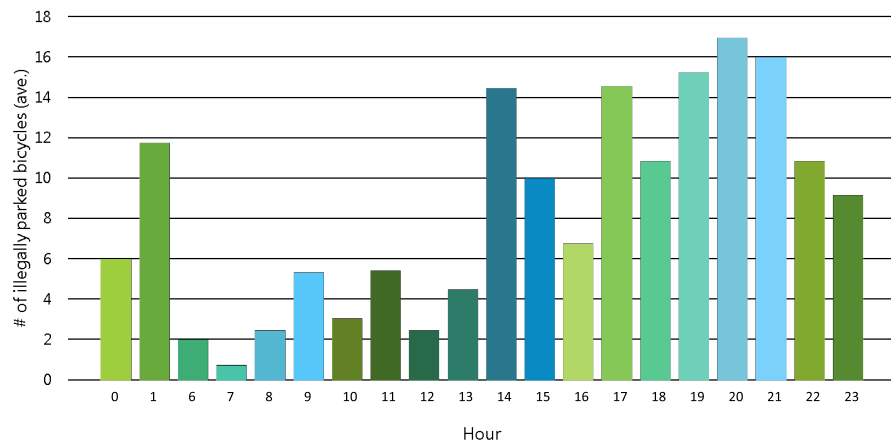


Fig. 7. The average number of illegally-parked bicycles per hour

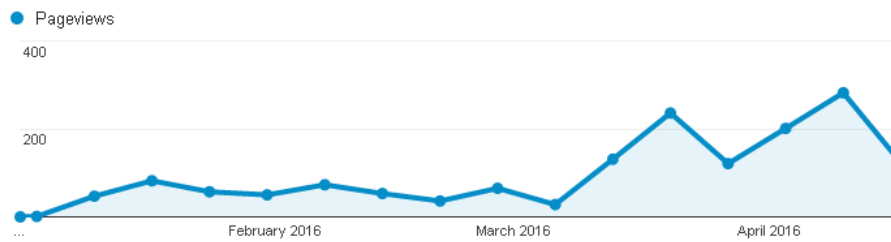


Fig. 8. Pageview of visualization application

illegally parked bicycles, complementation and estimation of the missing data, and then visualization of the LOD. We expect that this will increase the public awareness of local residents regarding the problem and encourages them to post more data.

In the future we will increase the amount of observation data and factors in order to improve the accuracy of the estimation. Also, we will collect more bicycle parking information and illegally parked bicycle data in cooperation with the Bureau of General Affairs of Tokyo and NPOs. Moreover, we will visualize more statistics of the IPBLD and clarify the problems caused by illegally parked bicycles in cooperation with local residents. We estimated and complemented the temporally missing values in this paper. However, there are also missing spatial data where bicycles might be illegally parked. In the future we will also estimate these missing spatial values. Then we will measure the growth rate of IPBLD in Tokyo.

Acknowledgments. This work was supported by JSPS KAKENHI Grant Numbers 16K12411, 16K00419, 16K12533.

References

1. Nishi, N.: The 2nd Health Japan 21: goals and challenges. *The Journal of the Federation of American Societies for Experimental Biology*, vol.28, no.1, 632-19 (2014)
2. Ministry of Internal Affairs and Communications.: Current bicycle usage and bicycle-related accident. http://www.soumu.go.jp/main_content/000354710.pdf (accessed: September 10, 2015) (Japanese)
3. Bischof, S., Martin, C., Polleres, A., Schneider, P.: Collecting, Integrating, Enriching and Republishing Open City Data as Linked Data. In: *Proc. The 14th International Semantic Web Conference*, pp.57-75 (2015)
4. Haklay, M., Weber, P.: Openstreetmap: User-generated street maps. *IEEE Pervasive Computing*, vol.7, no.4, pp.12-18 (2008)
5. King, S.F., Brown, P.: Fix my street or else: using the internet to voice local public service concerns. In: *Proc. The 1st international conference on Theory and practice of electronic governance*, pp.72-80 (2007)
6. Zook, M., Graham, M., Shelton, T., BA, Gorman, S.: Volunteered Geographic Information and Crowdsourcing Disaster Relief: A Case Study of the Haitian Earthquake, *World Medical & Health Policy*, vol.2, no.2, pp.7-33 (2010)
7. Ahn, L. V.: Games with a purpose. *IEEE Computer*, vol.39, no.6, pp. 92-94 (2006)
8. Celino, I., Cerizza, D., Contessa, S., Corubolo, M., Dell' Aglio, D., Valle, E. D., Fumeo, S., Piccinini, F.: Urbanopoly: Collection and Quality Assesment of Geospatial Linked Data via a Human Computation Game. In: *Proc. The 10th Semantic Web Challenge* (2012)
9. Lopez, V., Kotoulas, S., Sbodio, M. L., Stephenson, M., Gkoulalas-Divanis, A., Aonghusa, P. M.: QuerioCity: A Linked Data Platform for Urban Information Management. In: *Proc. The 11th International Semantic Web Conference*, pp.148-163 (2012)
10. Barbieri, D. F., Ceri, S.: C-SPARQL: SPARQL for continuous querying. In: *Proc. The 18th International Conference on World Wide Web*, pp.1061-1062 (2012)
11. Egami, S., Kawamura, T., Sei, Y., Tahara, Y., Ohsuga, A.: Visualization of Open Urban Data for Illegally Parked Bicycles. *Complexity: Technology for Complex Urban Systems in the 49th Hawaii International Conference on System Sciences* (2016)
12. Mizoguchi, R., Ikeda, M., Seta, K., Vanwelkenhuysen, J.: Ontology for Modeling the World from Problem Solving Perspectives. In: *Proc. the 1995 International Joint Conference on AI (IJCAI) Workshop on Basic Ontological Issues in Knowledge Sharing*, pp.1-12 (1995)
13. Auer, S., Bizer, C., Kobilarov, G., Lehmann, J., Cyganiak, R., Ives, Z.: DBpedia: A Nucleus for a Web of Open Data. In: *Proc. The 6th International Semantic Web Conference, 2nd Asian Semantic Web Conference (ISWC 2007 + ASWC 2007)*, pp.722-735 (2007)
14. Jenks, G. F.: The data model concept in statistical mapping. *International year-book of cartography*, vol.7, no.1, pp.186-190 (1967)
15. Hwang, S., Yu, D.: GPS Location Improvement of Smartphones Using Built in Sensors, *International Journal of Smart Home*, vol.6, no.3, pp.1-8 (2012)
16. Stadler, C., Lehmann, J., Höffner, K., Auer, S.: LinkedGeoData: A Core for a Web of SpatialOpen Data. *Semantic Web Journal*, vol.3, no.4, pp.333-354 (2012)