IEICE TRANS. INF. & SYST., VOL.E101–D, NO.1 JANUARY 2018

PAPER *Special Section on Semantic Web and Linked Data*

Temporal and Spatial Expansion of Urban LOD for Solving Illegally Parked Bicycles in Tokyo

Shusaku EGAMI[†a], **Takahiro KAWAMURA**[†,††b], ***Nonmembers*, *and*** **Akihiko OHSUGA**[†c], ***Member***

SUMMARY The illegal parking of bicycles is a serious urban problem in Tokyo. The purpose of this study was to sustainably build Linked Open Data (LOD) to assist in solving the problem of illegally parked bicycles (IPBs) by raising social awareness, in cooperation with the Office for Youth Affairs and Public Safety of the Tokyo Metropolitan Government (Tokyo Bureau). We first extracted information on the problem factors and designed LOD schema for IPBs. Then we collected pieces of data from the Social Networking Service (SNS) and the websites of municipalities to build the illegally parked bicycle LOD (IPBLOD) with more than 200,000 triples. We then estimated the temporal missing data in the LOD based on the causal relations from the problem factors and estimated spatial missing data based on geospatial features. As a result, the number of IPBs can be inferred with about 70% accuracy, and places where bicycles might be illegally parked are estimated with about 31% accuracy. Then we published the complemented LOD and a Web application to visualize the distribution of IPBs in the city. Finally, we applied IPBLOD to large social activity in order to raise social awareness of the IPB issues and to remove IPBs, in cooperation with the Tokyo Bureau.

key words: *linked open data, urban problems, 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 (IPBs) 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. IPBs 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 IPBs as Linked Open Data (LOD). For example, it would serve to visualize IPBs, suggest locations for optimal bicycle parking spaces and stations of sharing bicycles, and assist with the removal of IPBs. However, the Open Data sets available for IPBs 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 the schema for data storage are important issues that must 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 IPBs. The spatio-temporally enriched LOD make it possible to develop the services that visualize rich spatio-temporal information. The services lead to an increase of social awareness of IPBs, detection of new factors, and generation of solutions.

In this study, we proposed a method for sustainably building spatio-temporal urban LOD and applying them to Tokyo and other urban areas. Managing urban problem data joining multiple tables in (distributed) relational databases is troublesome from the aspect of data interoperability and maintenance, since the urban problem is closely related 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 schema. Thus, Linked Data are suitable for the data infrastructure of IPBs, since Linked Data have advantages of flexible linkability and schema. Also, we intend to integrate heterogeneous urban problems, such as littering and graffiti in the future; thus, adopting Linked Data format could be useful for further expansion to more comprehensive urban LOD.

Figure 1 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 IPBLOD based on schema
4. Estimating spatio-temporal missing data
    a. Using Bayesian networks to estimate the missing number of IPBs at each location
    b. Using computational fluid dynamics (CFDs) to estimate unobserved points where bicycles might be illegally parked
5. Visualizing IPBLOD to raise awareness of issues involving IPBs

We first extracted domain requirements of IPBs from articles on the Web and designed LOD schema. Next, we

Manuscript received February 16, 2017.
Manuscript revised July 1, 2017.
Manuscript publicized September 15, 2017.

†The authors are with Graduate School of Informatics and Engineering, The University of Electro-Communications, Tokyo, 182–8585 Japan.
††The author is with Japan Science and Technology Agency, Tokyo, 102–8666 Japan.
a) E-mail: egami.shusaku@ohsuga.lab.uec.ac.jp
b) E-mail: takahiro.kawamura@jst.go.jp
c) E-mail: ohsuga@uec.ac.jp

DOI: 10.1587/transinf.2017SWP0010

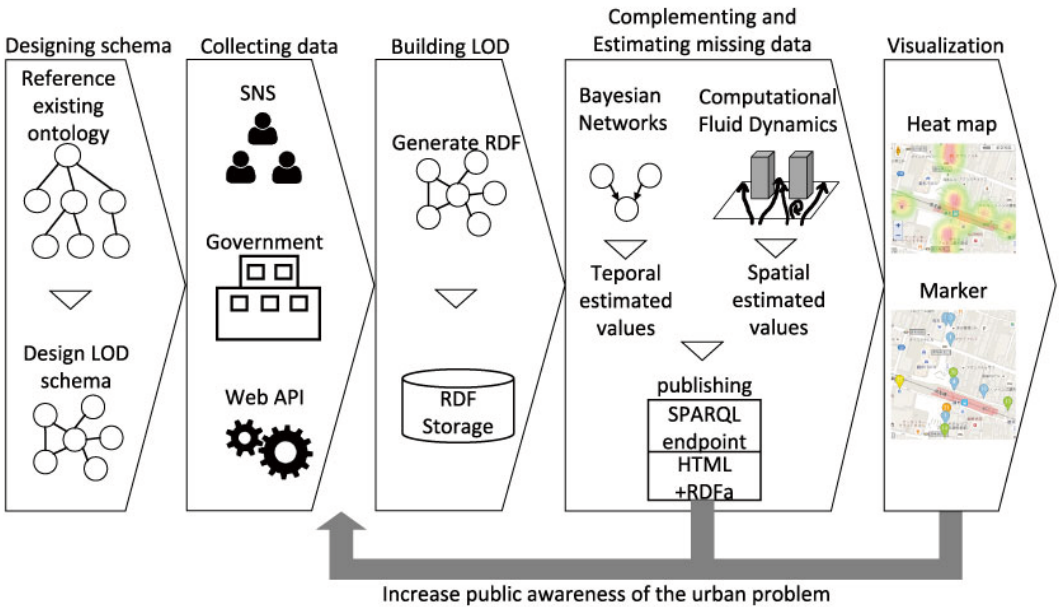


Fig. 1 Overview of this study

collected data about IPBs from Twitter and the data describing factors that affect the number of IPBs. In order to reuse these data sets that have different formats, we unified the data formats based on designed schema and published the data on the Web as LOD. Moreover, we estimated and complemented the spatio-temporal missing data. In terms of the temporal missing data (the number of IPBs when social sensors did not observe), we estimated the data 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 then use Bayesian networks to estimate the number of IPBs for data sets after complementation of the factors. In terms of the spatial missing data (unobserved points where bicycles might be illegally parked), we estimate the data using CFDs. Since we considered the flow of people as fluid, we estimated the spatial missing data in such a way as to find the stagnation points of the fluid. These results are also incorporated to build LOD with a specified property. In addition, we developed a service that visualizes the IPBs using the constructed LOD. This visualization service raises awareness of the issue in local residents and prompts users to provide more information about IPBs. Therefore, our contributions are as follows.

1. Collection of data from SNS and municipalities of Tokyo and other urban areas, and the building of illegally parked bicycle LOD (IPBLOD)
2. Development and evaluation of an approach for complementing the spatio-temporal missing data
3. Development and publishing of a Web application for visualizing IPBs in Tokyo and other urban areas
4. Increased social awareness of IPBs in cooperation with the Tokyo Bureau

The remainder of this paper is organized as follows. In Sect. 2, related works of data collection, urban LOD, and spatio-temporal data model are described. In Sect. 3, building IPBLOD are presented. In Sect. 4, estimation and complementation of temporal missing data using the Bayesian network is described. In Sect. 5, estimation and complementation of spatial missing data is described. In Sect. 6, visualization of the IPBLOD is described. In Sect. 7, application of IPBLOD and discussion are described. Finally, Sect. 8 concludes this paper with future works.

2. Related Work

In most cases, LOD sets have been built based on existing databases. However, so far there is little LOD available that describe 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. [7] proposed an approach for editing and adding Linked Data using a game with a purpose [8] and human computation. However, since the data concerning IPBs 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. [10] proposed a platform that publishes sensor data as Linked Data. The platform collects streamed data from the sensor and publishes the Resource Description Framework (RDF) in real time using IBM InfoSphere Stream and C-SPARQL [9]. The system is used in Dublinked2†, 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 static city data from DBpedia [10], Urban Audit††, the United Nations Statistics Division (UNSD)†††, and the U.S. Census††††, and then these data are integrated based on ontology. Furthermore, missing data are estimated using statistical regression methods. This approach is effective when there are some similar 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.

Consoli et al. [11] proposed a methodology for collecting heterogeneous urban data, integrating them based on the designed data model, and publishing them as LOD. The data are collected from different data sources; for example, GeoData from the Geographic Information System (GIS), public transport bus system, public lighting system of the city, database dump of the road conditions, Microsoft Excel file of municipal waste collection, and urban fault reporting. This paper also includes, user-friendly access services that allowing citizens, developers, and companies to access the produced data and ontology. Furthermore, an urban fault reporting Web platform was developed using the LOD [12]. The platform aims to control and monitor the complex process of urban fault reporting.

There are studies that model spatio-temporal data using semantic technologies. SOWL [13] is a spatio-temporal data model for Web Ontology Language. In SOWL, temporal information is modeled based on the 4D-fluents approach [14] and spatial information is modeled based on region connection calculus and cone-shaped directional relations. Also, L3 [15] is proposed as a spatio-temporal data model for next-generation GISs. The L3 can track the evolution of spatial entities throughout time, and it is useful in the area of land use/land cover change (LULCC). Sextant [16] is developed as a Web-based tool for visualization and exploration of time-evolving linked geospatial data. The front-end part is the graphical user interface to serve functions such as an exploration of linked spatio-temporal data using multiple SPARQL endpoints and collaborative editing of thematic maps, as with GISs. The back-end part is composed of five components: the endpoint connector, the ontology manager, the map registry, the time manager, and the Keyhole Markup Language translator. Sextant is utilized in the area of LULCC as well as L3. However, in the urban problem area, the data model should be easy to understand in order to facilitate the data use in Civic Tech events. Therefore, we must build a new data schema for urban problems.

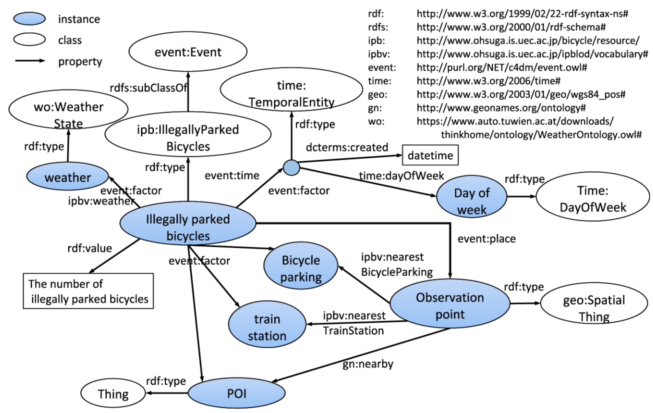


Fig. 2 LOD schema containing instances

Local governments are now addressing their own urban problems by several countermeasures. For example, Tokyo is distributing leaflets and installing signboards for warning the IPBs. Osaka is promoting "no littering" using the original character†††††. However, to the best of our knowledge, there is no social activity to solve the IPBs problem before this study.

3. Building IPBLOD

First, we designed IPBLOD schema based on the result of extracting domain requirements from Web articles related to IPBs. We used the methodology for designing LOD schema, that was presented at our previous work [17]. After following some part regarding the estimation of temporal missing data, we then extended our work to estimate the spatial missing data in this paper. Figure 2 shows an overview of IPBLOD schema.

3.1 Collecting Data

Next, we collected tweets containing location information, pictures, hash-tags, and the number of IPBs. Figure 4 shows the procedure of collecting data and building LOD. Users submit the number of IPBs with certain location information to Twitter through our Web application in Fig. 5. After the OAuth authentication of Twitter, the following form and buttons will appear, and the users can input the number

†http://www.dublinked.ie/
††http://ec.europa.eu/eurostat/web/cities
†††http://unstats.un.org/unsd/default.htm
††††http://www.census.gov/
†††††http://www.city.osaka.lg.jp/contents/wdu150/akanzukin/akanzukin_room.html

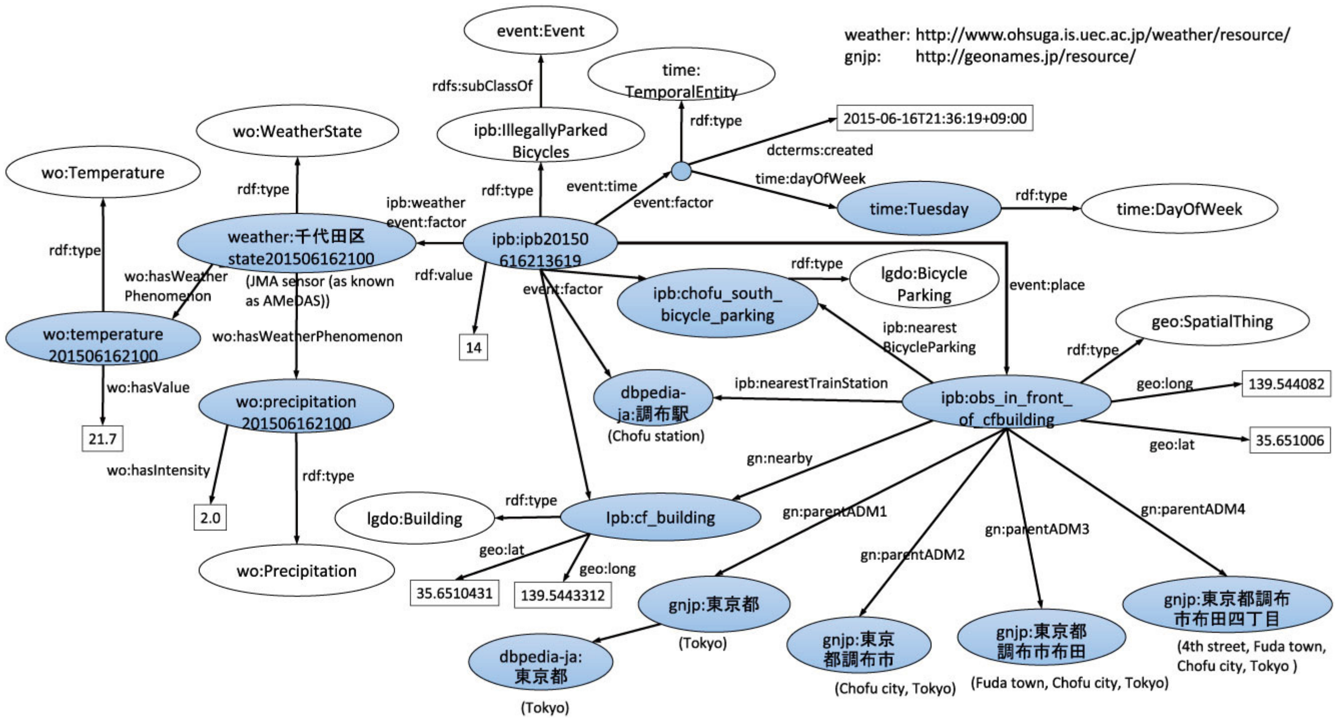


Fig. 3 Part of the integrated LOD

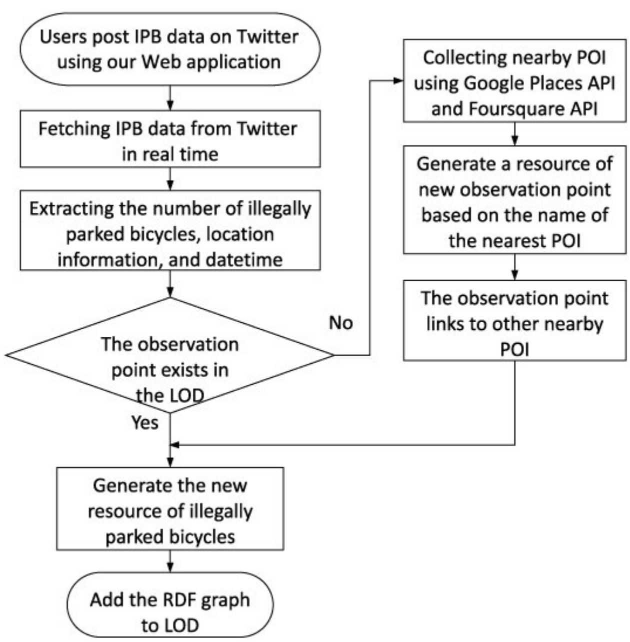


Fig. 4 Procedure of LOD building

Fig. 5 Screenshot of the tweet application

of IPBs. In addition, when the users tap the "location information" button, a marker will be displayed on the map. The users can then fix the location by dragging the marker. Users can also attach including the aboce information with the hashtah (#放置自転車マップ) are posted. A server is periodically collecting those tweets, and then extracts the number of IPBs, the location information, and corresponding date/time from the tweets.

Furthermore, we collected information on POI using Google Places API† and Foursquare API††. Also, we obtained bicycle parking information from websites of municipalities and in cooperation with the Office for Youth Affairs and Public Safety of the Tokyo Metropolitan Government (Tokyo Bureau)†††. Tokyo Bureau publishes Open Data on bicycle parking areas as CSV. The data contain names, latitudes, longitudes, addresses, capacities, and business hours.

†https://developers.google.com/places/?hl=en
††https://developer.foursquare.com/
†††http://www.seisyounen-chian.metro.tokyo.jp/english/

Fig. 6 Example of a promotion tweet

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)†.

3.2 Building LOD Based on Collected Data

The collected data on IPBs are converted to LOD based on the designed schema. We called for cooperation through a Twitter account of our project, such as in Fig. 6, and also we retweeted this tweet with our private accounts. There were 18 users in total. Students in our university were the most of them, but the other users who have a high interest in the IPBs problem also joined in our project. The server program checks whether there is an existing observation point within a radius of less than 30m by querying our endpoint†† 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 Foursquare API. We map the types of POI to classes in LinkedGeoData [18]. 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. Figure 3 shows part of the IPBLOD. The LOD are stored in Virtuoso††† Open-Source Edition. Also, the RDF data set is published with CC-BY license on our website††††.

According to the Sect. 10.5.d in the license of Google Maps API including Google Places API†††††, temporally caching of contents is allowed. Likewise, according to the Sect. 4 in the license of Foursquare API††††††, temporally caching of contents is allowed. In this study, we cached part of contents from these APIs, and then changed literal expressions in the contents. Numerical values are rounded. Then, we put some data in totally-different files in our invented RDF schema. Therefore, anyone cannot restore Google's and/or Foursquare's contents from our RDF files; thus, we believe that this study does not violate the license agreements and their copyrights.

†http://www.jma.go.jp/jma/indexe.html
††http://www.ohsuga.is.uec.ac.jp/sparql
†††http://virtuoso.openlinksw.com/
††††http://www.ohsuga.is.uec.ac.jp/bicycle/dataset.html
†††††https://developers.google.com/maps/terms
††††††https://foursquare.com/legal/api/platformpolicy

4. Estimation of Temporal Missing Data

Since we rely on the social sensors to observe IPBs, we do not have round-the-clock data for every place. Therefore, temporal missing data in the IPBLOD are inevitable. Thus, we assumed that the number of the IPBs should be influenced by several factors, and we tried to estimate these missing data using Bayesian networks. A Bayesian network is a probabilistic graphical model that represents a set of random variables and their conditional dependencies via a directed acyclic graph.

4.1 Complementing Missing Factors

As the factors (attributes), we used observation points, day of the 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 extracted domain requirements. However, there are also missing factor values. We assumed 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 that cannot be obtained. The similar observation data were found using the Jaccard coefficient. Suppose the aggregates of each factor are given by Location, Day = $\{sun, mon, \ldots, sat\}$, Hour = $\{0, 1, \ldots, 23\}$, Precipitation = $\{0, 1, \ldots\}$, Temperature = $\{\ldots, -1, 0, 1, \ldots\}$, DailyFee = $\{0, 1, \ldots\}$, MonthlyFee = $\{0, 1, \ldots\}$, Density = $\{0, 1, \ldots\}$, Commuters = $\{0, 1, \ldots\}$, Building, Bank, DepartmentStore, Games, Supermarket, Library, Police, School = $\{0, 1\}$, and Number (of IPBs) = $\{1, \ldots, 4\}$, and then the observation data were stored as an aggregate O of vectors (Eq. (1))

$$\begin{aligned} o \in\ & Location \times Day \times Hour \times Precipitation \\ & \times Temperature \times DailyFee \times MonthlyFee \\ & \times Density \times Commuters \times Building \\ & \times Bank \times DepartmentStore \times Games \\ & \times Supermarket \times Library \times Police \\ & \times School \times Number \end{aligned} \tag{1}$$

The number of IPBs was classified into four classes by Jenks natural breaks [19], which are often used in 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 was calculated as Eq. (2).

$$sim(o_1, o_2) = \frac{|o_1 \cap o_2|}{|o_1 \cup o_2|} \tag{2}$$

4.2 Estimation of the Number of IPBs Using Bayesian Network

We then estimated the number of IPBs using Bayesian networks at observation points, where the number data are missing. We empirically found that the number of IPBs is changing according to various factors, such as time, weather, place and its surroundings. Bayesian networks can represent various causal relations based on conditional probabilities of each factor; thus, it is suitable for this problem of estimating the number of IPBs. In addition, the Bayesian networks can estimate the number of IPBs with its probability; thus, there is also an advantage that it is possible to check the certainty of the estimation result. We considered many other learning algorithms to estimate the missing data, such as Support Vector Machine (SVM), Deep Neural Networks (DNN), and Hidden Markov Model (HMM). SVM works effectively in 2-class classifications. However, it is in general not good at multi-class classifications to which this study belong. DNN requires a large amount of training data, and thus it is not applicable this time. HMM is suitable for time-series data, such as speech, signal and sensor data. However, since the IPB data is very sparse, we found that HMM cannot estimate the missing data correctly in this study. Therefore, we adopted the Bayesian networks for estimating the temporal missing data.

Thus, we built a Bayesian network that is a kind of probabilistic network based on the factors obtained from IPBLOD. It is possible to obtain factors of IPBs using `event:factor` property, since the IPBLOD schema is designed based on EO after extracting factors of IPBs from Web articles.

The input data set is the data set complemented using the method described in Sect. 4.1. We used the Bayesian network tool Weka[†] to estimate the unknown numbers of IPBs. There are 897 observation data. The input data are a set O that consists of vectors with eighteen 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 received 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 extracted domain requirements when estimating the number of IPBs using Bayesian networks. However, other POIs could become factors related to IPBs. 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 a 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, CoffeeShop, Construction, Lawyer, Highway-Thing, Shop, and PublicTransportThing. Therefore, an observation datum became a vector (Eq. (3))

$$\begin{aligned}
o \in\ & Location \times Day \times Hour \times Precipitation\\
& \times Temperature \times DailyFee \times MonthlyFee\\
& \times Density \times Commuters \times Pharmacy\\
& \times \cdots \times PublicTransportThing \times Number,
\end{aligned}$$

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 IPBs 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 were no data at 9 am or 9 pm, we estimated and complemented the number of IPBs. Then we added the estimated number and its probability to IPBLOD as follows.

```
@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  "0-6" ;
        ipb:probability  "0.772"^^xsd:double ] .
```

4.3 Results and Discussion of Estimation of Temporal Missing Data

There are 219 pieces of observation data that have missing factor values, and these values have been complemented using the method discussed in Sect. 4.1. The missing factors were 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 were found in precipitation and temperature values since they are not the source data in JMA.

Also, we found that LOD are 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 a 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

[†] http://www.cs.waikato.ac.nz/ml/weka/

Table 1 Results of estimation of temporal missing data

		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
True Positive		0.871	0.656	0.605	0.179	

was not very many, and it was also unbalanced. Also, Table 1 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 in the range decisions for each person. It was found that some people tweeted many IPBs at one time, while some people divided the IPBs and tweeted individually. Thus, data for 0-6 and 7-17 were higher, and this fact affected the estimation accuracy. Also, the number of IPBs can be zero due to some events, such as roadblock and building construction. Thus, in future we will apply some logical rules to address this issue. Also, we consider applying the latest learning algorithm, which is assessed in the state-of-the-art studies, such as A* algorithm. We also consider adding regional statistics to factors, such as age distribution and salary, and filtering factors using information gain. Finally, we will promote this study to the general public for increasing social awareness in cooperation with Tokyo Metropolitan Government. We believe that such promotion increases the data amount, and then also lead to the improvement of the estimation accuracy.

5. Estimation of Spatial Missing Data

There are not only temporal missing data but spatial missing data (unobserved points where bicycles might be illegally parked) in IPBLOD since the data are collected from social sensors. Thus, we geographically expanded IPBLOD by estimating and complementing the spatial missing data using CFDs.

The reason why we adopted CFDs is that we assumed that there are some common spatial features of locations where bicycles might be illegally parked, such as building density and road width. Thus, we simulated the flow of people and examined whether stagnations correlate with the observation points of IPBs. There are some simulation methods such as physical model simulation and multi-agent simulation. We used physical model simulation, which does not have to decide the rules of agents, since we considered that building density and road width are the main spatial features of IPBs factors. Then we defined that stagnation points are places where bicycles might be illegally parked and searched stagnation points using airflow simulations around train stations. Furthermore, we filtered stagnation points using DBpedia Japanese, and we regarded these filtered points as new observation points. Therefore, in this section, we describe the hybrid approach using CFDs and Linked Data for estimating spatial missing data.

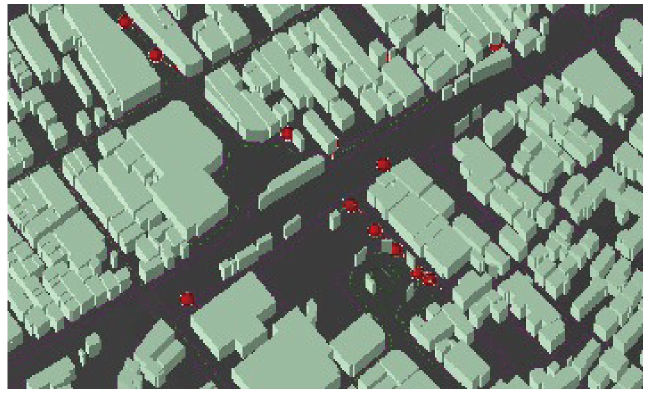


Fig. 7 The 3D map around of Chofu Sta. (Red markers are observation points.)

5.1 Finding Stagnation Points Using CFD

There would be some methods to estimate spatial missing data. However, applying machine learning methods had difficulty, since there were few observation points corresponding to the training data. We first considered the creation of heuristic rules, but finally, we found a commonality to places, in which airflows occasionally stagnate. Therefore, we adopted the airflow simulation as a kind of nature-inspired methods.

Also, there are wind tunnel tests and CFDs as the methods of airflow simulation. CFDs are methods that observe the movements of fluid using computer simulation. A wind tunnel test requires expensive and large equipment, but CFDs are easy to experiment with in different environments when using a computer. However, CFDs cannot produce exact copies of fluid movements, since CFDs use an approximate solution.

We first obtained maps of building from the Geospatial Information Authority of Japan†. These data consist of 2D polygons. We converted the 2D maps to 3D maps using ArcGIS for the Desktop††. Since we could not obtain information on the height of the buildings, we set the height of all buildings to 30 meters. Figure 7 shows the 3D map around Chofu Station in Tokyo. Red markers are observation points of IPBs. We obtained observation points in a CSV format from the SPARQL endpoint of the IPBLOD, and then we imported the CSV to the 3D map. Next, we simulated the airflow around the station using Airflow Analyst†††, which is a simulation software run on ArcGIS. Figure 8 shows the grid cells, which are set as the analysis range. We set the analysis range to include all observation points around Chofu Station. In Fig. 8, we selected 700 × 700 square meters around Chofu Station as the analysis range. Also, the node spacing is 5 meters, and the number of nodes is 10,000. In this study, we considered flows of people as

†http://www.gsi.go.jp/ENGLISH/index.html
††http://www.esri.com/software/arcgis/arcgis-for-desktop
†††http://www.airflowanalyst.com/en/index.php

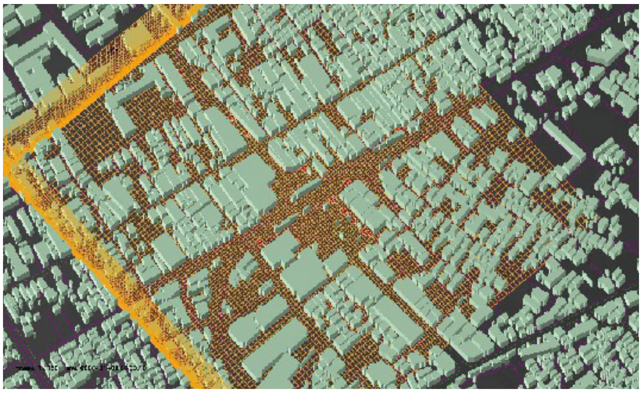


Fig. 8 The view of grid cells

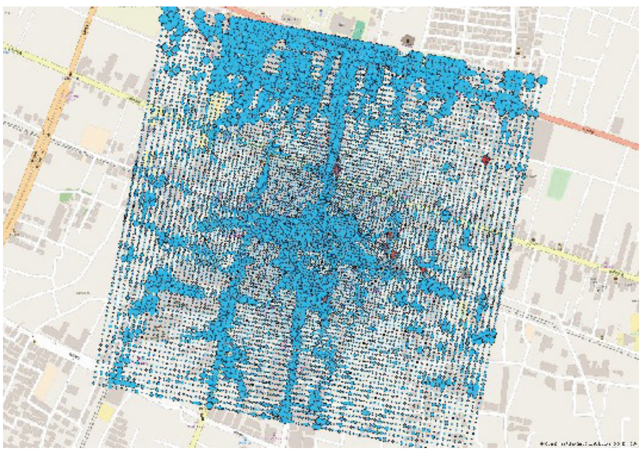


Fig. 9 The result of the airflow simulation. (The size of blue circle is the average wind velocity. Red markers are observation points.)

airflows. Also, since we assumed that people come to the nearest station along with main streets, we set the wind directions to be the same as the streets. Thus, in the case of Chofu city, we used the 11-degree, 109-degree, 190-degree, and 288-degree for the wind directions, from which people come to Chofu station.

Figure 9 shows the visualization of the average wind velocity based on the results of the simulation when the wind direction is 11-degree and the wind speed is 5 m/s. The size of the blue circle means the average wind velocity. We found stagnation points based on this numerical data. A stagnation point is a point where the velocity of the fluid is zero in the flow field. We tried to find stagnation points using patterns in Fig. 10. A black node is a node with an average wind velocity of $x > 0.1$. The white node has an average wind velocity of $x = 0$. The gray node has an average wind velocity of $0 < x \leq 0.1$. In general, a stagnation point is a white node under these conditions. However, white nodes became buildings in our experiment. Therefore, we defined gray nodes as stagnation points. Table 2 shows the total accuracy of the findings of stagnation points around Chofu Station, Fuchu Station, and Shinjuku Station using all the patterns. The precision is the ratio of stagnation points

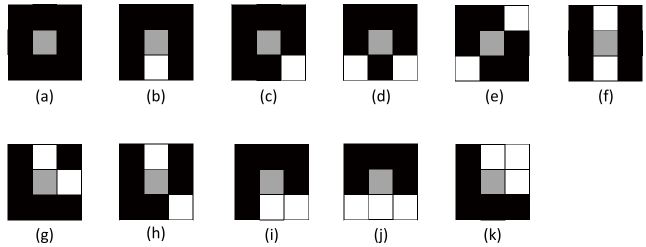


Fig. 10 Patterns of stagnation points

Table 2 Results of the findings stagnation points when we used patterns in Fig. 10

Pattern	Precision	Recall	F-measure
(a)	0.102	0.286	0.150
(b), (c)	0.0833	0.0357	0.0500
(d), (e), (f), (g), (h), (i)	0.000	0.000	0.000
(j)	0.0913	0.429	0.151
(k)	0.0746	0.107	0.0880

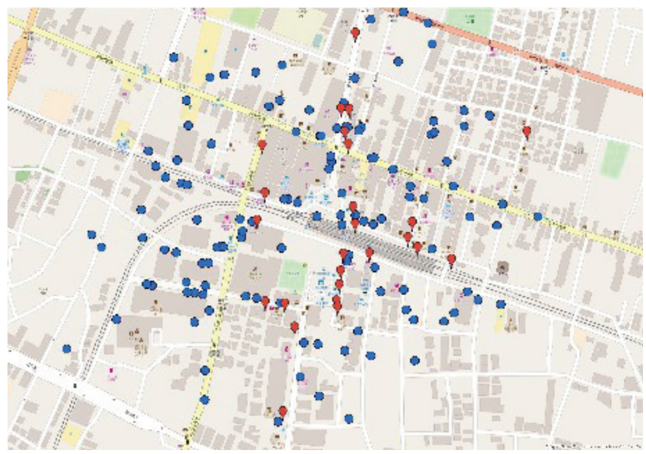


Fig. 11 The stagnation points around of Chofu Station

within a 20-meter radius from an observation point. The recall is the ratio of observation points that have stagnation points within a 20-meter radius. As a result, the F-measure, when we used the pattern (j), became the highest. Also, the F-measure, when we used the pattern (a), became the second highest. Hence, we used pattern (a) and (j) to find stagnation points in this study. Figure 11 shows the results of the findings' stagnation points around Chofu Station. This is the merged result of the simulation results of the four directions.

5.2 Filtering Stagnation Points by DBpedia

We found the stagnation points, but there were many noise points, as can be seen in Fig. 11. We assumed that bicycles tend to be parked illegally at stagnation points having nearby POIs whose popularity stakes are high. Therefore, we calculated the popularity stakes of the POIs around the stagnation points and then filtered the stagnation points using the popularity stakes. Figure 12 illustrates this method of calculating the popularity stakes of POI.

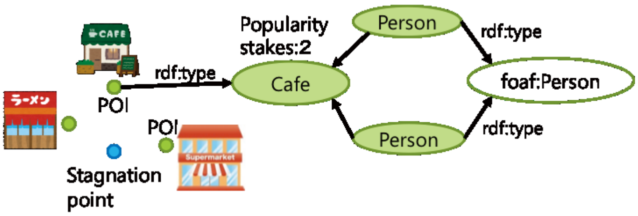


Fig. 12 Method for calculating popularity stakes of POI

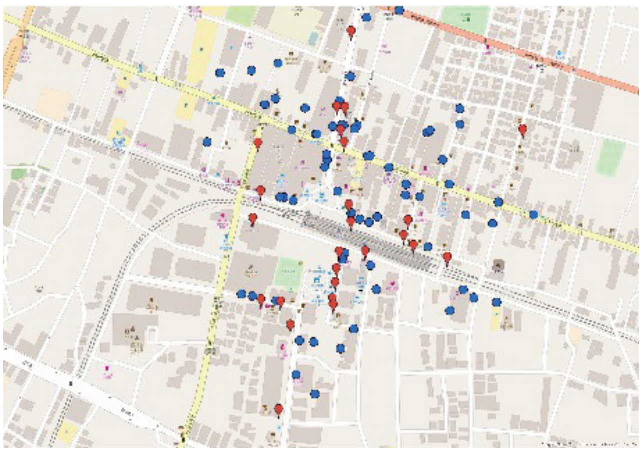


Fig. 13 The filtered stagnation points around of Chofu Station

We first obtained the POIs information within a 20-meter radius from the stagnation points using Google Places API. Then we calculated the number of links from person resources to POIs on DBpedia Japanese. Also, we mapped the types of POIs to DBpedia Japanese resources. We considered the number of inbound links from person resources as the popularity stakes, and we obtained the number of links from instances of `foaf:Person` to types of POIs. Then we calculated the sum of the popularity stakes of POIs, and we filtered stagnation points if the sum of the popularity stakes was less than the threshold. We varied the threshold from 100 to 1,000, and we could achieve the best results when the threshold was 200. Hence, we set the threshold to 200. Figure 13 shows the results of the filtering.

Next, we excluded stagnation points that are located in the buildings' premises, since the public and bicycles cannot enter the buildings' premises. The stagnation points were excluded if they were at a distance of 10 meters or more from the nearest road. We used the polygon data of the road, which are published by the Geospatial Information Authority of Japan. Figure 14 shows the result of excluding stagnation points that are located in the buildings' premises. Green circles are stagnation points that were excluded.

Furthermore, we added estimated data to IPBLOD separately from the real data as follows. The latitude and longitude were obtained from ArcGIS. Address information was obtained from Yahoo! Reverse Geocoder API[†]. The POIs

[†]http://developer.yahoo.co.jp/webapi/map/openlocalplatform/v1/reversegeocoder.html

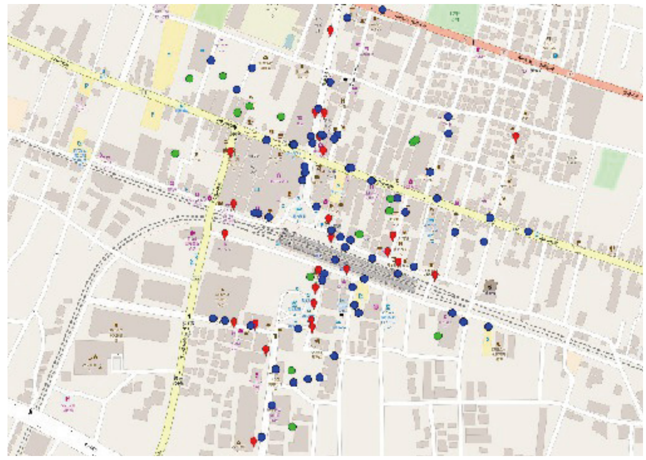


Fig. 14 The result of excluding stagnation points that are located in the buildings' premises (Green circles were excluded)

were also obtained from Google Places API.

```
@prefix ipb:
    <http://www.ohsuga.is.uec.ac.jp/ipblod/
        vocabulary#>
@prefix bicycle:
    <http://www.ohsuga.is.uec.ac.jp/bicycle/
        resource/>
@prefix geo:
    <http://www.w3.org/2003/01/geo/wgs84_pos#> .
@prefix ogcgs:
    <http://www.opengis.net/ont/geosparql#> .
@prefix ngeo:  <http://geovocab.org/geometry#> .
@prefix dcterms: <http://purl.org/dc/terms/> .
@prefix gn: <http://www.geonames.org/ontology#> .
@prefix gnjp:  <http://geonames.jp/resource/> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#>
    .
bicycle:estimated_obs_{timestamp}
    rdf:type     ipb:EstimatedObservationPoint ;
    geo:lat      "latitude"^^xsd:double ;
    geo:long     "longitude"^^xsd:double ;
    gn:parentADM         gnjp:{Prefecture}
    gn:parentADM2        gnjp:{City, Prefecture} ;
    gn:parentADM3        gnjp:{Town, City,
        Prefecture} ;
    gn:parentADM4        gnjp:{Land lot, Town,
        City, Prefecture} ;
    ngeo:geometry        [
        a ngeo:Geometry;
        ogcgs:asWKT "POINT(latitude,longitude)
            "^^<http://www.openlinksw.com/schemas
            /virtrdf#Geometry> . ] ;
    gn:nearby    bicycle:{POI name} ;
    dcterms:created      "datetime"^^xsd:dateTime
        .
```

5.3 Results and Discussion of Estimation of Spatial Missing Data

In this section, we describe the validation results as to whether there is a correlation of the data estimated from our approach and the observation points of IPBs and then discuss the evaluation of the utility of our approach.

We carried out the experiments on Chofu Station,

Fig. 15 The stagnation points of baseline

Table 3 Evaluation results of both baseline and stagnation point method

		Baseline	Stagnation point method
Chofu Sta.	Precision	0.049	0.124
	Recall	0.231	0.539
	F-measure	0.081	0.202
Fuchu Sta.	Precision	0.125	0.231
	Recall	0.222	0.444
	F-measure	0.160	0.304
Shinjuku Sta.	Precision	0.049	0.167
	Recall	0.190	0.667
	F-measure	0.078	0.267
Total	Precision	0.0550	0.150
	Recall	0.214	0.571
	F-measure	0.087	0.238

Fuchu Station, and Shinjuku Station, which have multiple observation points of IPBs. The total number of observation points was 56. We validated the utility of the proposed method by comparing the result of the baseline and the result of the proposed method. First, we compared the baseline and the stagnation point method, as described in Sect. 5.1. Figure 15 shows the result of the baseline for the Chofu Station. The baseline estimates the spatial missing data at regular intervals, as many as the number of stagnation points. Table 3 shows the accuracy of both the baseline and the stagnation point method. As a result, the precision, the recall, and the F-measure of the stagnation point method became higher than the result of the baseline. Also, we validated the utility of the stagnation point method using the chi-square test. The null hypothesis is that there is no difference between the result of the baseline and the result of the stagnation point method, and we used a standard level of significance $p < 0.05$. As a result, when we applied the chi-square test to the 2×2 confusion matrix, which is composed of true and false values, the p-value was 0.01591. Hence, we found that there is a significant difference between the result of the baseline and the result of the stagnation point method.

Next, we compared the baseline and the hybrid method (filtering stagnation points). Table 4 shows the accuracy of both the baseline and the hybrid method. As a result,

Table 4 Evaluation results of both baseline and hybrid method

		Baseline	Hybrid method
Chofu Sta.	Precision	0.0469	0.247
	Recall	0.115	0.462
	F-measure	0.0667	0.322
Fuchu Sta.	Precision	0.125	0.250
	Recall	0.222	0.333
	F-measure	0.160	0.286
Shinjuku Sta.	Precision	0.0493	0.211
	Recall	0.190	0.571
	F-measure	0.0784	0.308
Total	Precision	0.0559	0.228
	Recall	0.161	0.482
	F-measure	0.0829	0.310

the precision, the recall, and the F-measure of the hybrid method became higher than the result of the baseline. Also, as it is possible to see from Table 3 and Table 4, the precision and the F-measure of the hybrid method became higher than the result of the stagnation point method. Therefore, there is utility of the hybrid method using POIs and DBpedia Japanese. Also, we validated the utility of the hybrid method using the chi-square test. The null hypothesis is that there is no difference between the result of the baseline and the result of the hybrid method, and we used a standard level of significance, $p < 0.05$. As a result of the chi-square test, the p-value was 0.0000739. Hence, we found that there is a significant difference between the result of the baseline and the result of the hybrid method. Although the accuracy is not high, the data of estimated points were considered to help collect new data from social sensors.

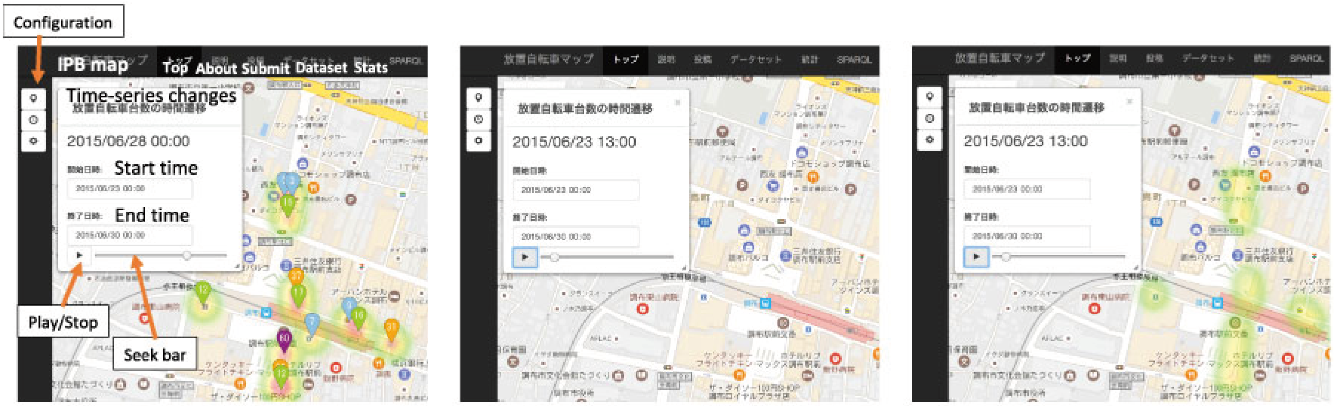
Furthermore, we validated whether the distribution of estimated points tends to have spatial homogeneity with distribution of observation points using the Mutual Nearest Neighbor (MNN) method. The MNN method, which calculates the nearest neighbor distance and validates the difference between the proposed method and a random distribution, was used in the area of spatial analysis. The average distance of MNN was calculated as follows:

$$D = \frac{1}{n_a + n_b}\left(\sum_{i=1}^{n_a} d_{ai} + \sum_{i=1}^{n_b} d_{bi}\right) \quad (3)$$

When there are distributions of A and B, the d_{ai} (or d_{bi}) is the distance from a point in A (or B) to the nearest point in B (or A). Also, the n_a (or n_b) is the number of points in A (or B). When A and B are randomly distributed, the excepted average distance of MNN is calculated as follows:

$$E[D] = \frac{1}{n_a + n_b}\left(\frac{n_a}{2\sqrt{\lambda_b}} + \frac{n_b}{2\sqrt{\lambda_a}}\right) \quad (4)$$

Therefore, we adopted the random distribution in this test, instead of the lattice points as in Fig. 15. The λ_a (or λ_b) is a density of A (or B). The value of $E[D]$ became 61.4, and the value of D of the proposed method became 46.9. The degree of freedom is 88, and a standard level of significance is 0.01. As a result of the t-test, the t-boundary value became 2.633 and the t-value became 4.76. Therefore, we found that there is a significant difference between



(a). Heat map and marker (b). Before temporal complementation (c). After temporal complementation

Fig. 16 Screenshots of the visualization application

the proposed method and the random method, and we also found that the estimated points tend to have spatial homogeneity with distribution of observation points. Therefore, there is a correlation of the estimated points and the observation points of IPBs.

The accuracy of the estimated data in this study was low for the following reasons. The number of observation points was less, and there was a possibility that new observation points would be found around the estimated points. Therefore, we should conduct a field survey in future work, and then we should evaluate our approach once again. Furthermore, a lack of time information can be considered as lowering accuracy. In this experiment, we simulated the airflow without considering time. However, the population of train stations and POIs increases and decreases with time. To increase the estimation accuracy of the spatial missing data, we consider incorporating data/time information to the airflow simulation in future. We will also investigate optimal parameters, such as wind speeds and temperatures.

6. Visualization of IPBLOD

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 endpoint† 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 IPBs. The application can display time-series changes in the distribution of IPBs on a map. The following is a sample SPARQL query that is actually executed when displaying time-series changes of the number of IPBs. This query obtains the number of IPBs from June 23 to June 30, 2015, and used in the visualization application.

†http://www.ohsuga.is.uec.ac.jp/sparql

```
@prefix ipb: <http://www.ohsuga.is.uec.ac.jp/bicycle/
     resource/>
@prefix event: <http://purl.org/NET/c4dm/event.owl#>
@prefix dcterms: <http://purl.org/dc/terms/>
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns
     #>
@prefix xsd: <http://www.w3.org/2001/XMLSchema#>
select ?s ?lat ?lon ?num ?date where {
?s a ipb:IllegallyParkedBicycles;
event:place ?place;
event:time [dcterms:created ?date] .
?place geo:lat ?lat;
geo:long ?lon .
optional {?s rdf:value ?num}
optional {?s <http://www.ohsuga.is.uec.ac.jp/ipblod/
     vocabulary#estimatedValue> [rdf:value ?num]}
filter(?date >= "2015-06-23"^^xsd:dateTime && ?date <=
     "2015-06-30"^^xsd:dateTime)} order by ASC(?date)
```

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 IPBs are displayed. Figure 16 (a) shows a screenshot of a PC, on which the Web application is displaying such an animation near Chofu Station in Tokyo using a heat map and a marker UI. It is possible to see the visualized information just after tweeting the number of IPBs. Thus, users obtain the instant feedback after posting new data.

The IPBLOD not only contain the data collected from Twitter but also the data estimated by Bayesian networks. Therefore, time-series changes in the distribution of IPBs become smoother than before in estimating the missing values. Figure 16 (b) and 16 (c) 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.

Furthermore, we implemented a function for mapping estimated points to the visualization application. Figure 17 shows the results of visualization of estimated points. This visualization is helpful to observe and submit new data.

As another example, we can see the average number of IPBs per hour using the short SPARQL query. Figure 18

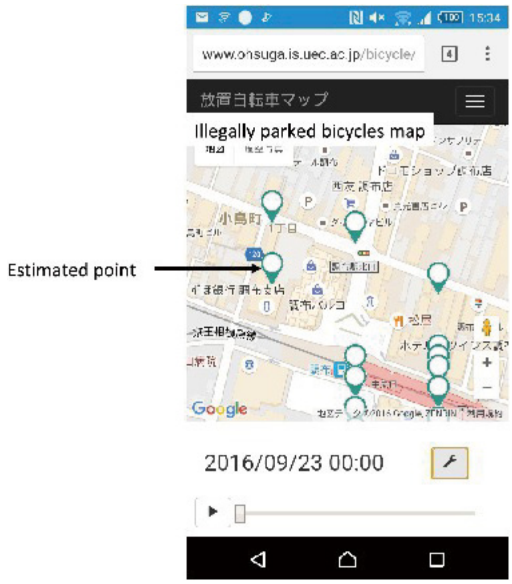


Fig. 17 Visualization of estimated points

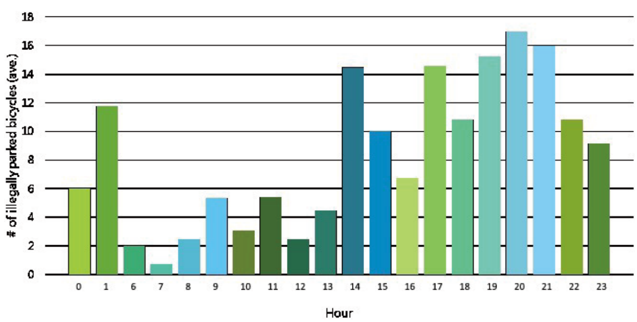


Fig. 18 The average number of illegally-parked bicycles per hour

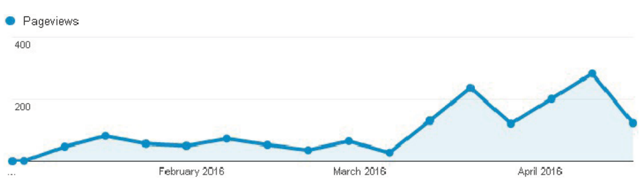


Fig. 19 Pageview of visualization application

shows a visualization of the result. We found from the result that there are more IPBs 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.

6.1 Results and Discussion of Visualizing IPBLOD

The number of page views of the visualization application increased from January to April 2016 (in Fig. 19). 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 number of unique users of the visualization application increased from 40 to

Table 5 Top 10 of the number of users by city

Oct. 2015 - Mar. 2016		Apr. 2016 - Sep. 2016	
City	# of users	City	# of users
Chofu	66	Shinjuku	129
Yokohama	61	Yokohama	127
Minato	48	Chofu	69
Shinjuku	39	Minato	64
Kodaira	35	Osaka	24
Osaka	20	New Delhi	21
Nagoya	19	Chuo	20
Shibuya	9	Fucu	14
Setagaya	8	Setagaya	14
Kawasaki	8	Kodaira	12

Fig. 20 Promotion poster of the 33rd campaign to clean up illegally parked bicycles around the station

118, and the average session duration is 2 minutes and 32 seconds. A session duration is a measure of a continuous period of browsing of a website by a particular user. Therefore, it was found that visitors are increasing and tend to use our application for longer periods of time.

Table 5 shows top 10 number of users by city. The time rage is from October 2015 to September 2016. It is found that the number of users from Yokohama and Shinjuku is larger than Chofu, although the amount of their observation data is less than Chofu. Also, the number of users increased more than before in the cases of Chofu, Shinjuku, Yokohama, Minato, Osaka, and Setagaya. Specifically, the number of users from Yokohama about doubled, and users from Shinjuku increased about three times. Furthermore, the number of cities where users were recorded increased from 68 to 118. Therefore, we consider that the IPBLOD and visualization application increased visibility and raised social awareness of the IPB problem.

7. Application of IPBLOD

Currently, we are discussing applications of IPBLOD and other Open Data for solving IPBs in cooperation with the Tokyo Bureau. Our visualization application was introduced on the website of Tokyo Bureau†. The Tokyo Metropolitan Government considers that IPB is a serious urban problem and implemented the "33rd campaign to clean up illegally parked bicycles around the station††" from Oct. 22, 2016 to Oct. 31, 2016. Figure 17 shows the promotional poster of this campaign†††. In the 32nd campaign held from Oct. 22, 2015 to Oct. 31, 2015, many measures are adopted; for example, distributing posters, distributing tissues with a leaflet, publishing official publications and public relations magazines, using a famous comedian for the promotional movie, broadcasting the promotional movie on electronic billboards and digital signages, and holding a promotional poster contest for children. We cooperated with this campaign by collecting IPBs data, publishing IPBLOD, and visualizing IPBLOD, and then we raised social awareness of the IPB problem and promoted a solution for the problem through the reuse of IPBLOD. In 2015, 44,600 posters were created by the Tokyo Bureau, 406 posters were created by municipalities, and 26,900 posters were created by concerned bodies. They were posted on train stations, buses, schools, and concerned bodies. Also, 390,000 leaflets were created by the Tokyo Metropolitan Government, and 31,599 leaflets were created by municipalities. Also, 10,785,770 official publications and public relations magazines were published. Furthermore, 13,879 workers who promoted this campaign and removed IPBs participated. Conclusively, 10,652 IPBs were removed during this campaign. The Tokyo Bureau collected these data from each municipality using Microsoft Word and then published them as a PDF. We are negotiating to obtain raw data in Microsoft Word/Excel format, which were collected by the Tokyo Bureau in the past, and then we incorporate those data to IPBLOD and visualize them in the campaign.

Also, we provided the IPBLOD to Tokyo crime prevention ideathon†††† as an available open data set. This event was held by Tokyo Bureau. The purpose of this event is thinking the safe of Tokyo through a workshop and a group work. As the result, the event generated several ideas of applications of IPBLOD for solving IPB issues†††††.

Moreover, we provide the IPBLOD to Linked Open Data Challenge Japan (LODC) 2016††††††. The LODC is the biggest application contest for Linked Open Data in Japan. The LODC has been held every year since 2011, and more than 1,000 applications have been submitted in total. The IPBLOD was the first place in the dataset category of LODC in 2015, and a number of applications based on IPBLOD could be expected in LODC 2016.

8. Conclusion and Future Work

In this paper, building and using IPBLOD were described as a solution for a IPBs problem. The techniques proposed included complementation and estimation of the temporal missing data using Bayesian networks, complementation and estimation of spatial missing data using CFDs and DBpedia, and then visualization of the LOD. We expect that this will increase the public awareness of local residents regarding the problem and encourage them to post more data.

We will increase the amount of observation data and factors. Also, we will collect more bicycle parking information and IPB data in cooperation with the Tokyo Bureau and NPOs. Also, we must evaluate a growth rate of IPBLOD and examine the statistics of IPBs on a long-term basis in order to validate whether IPBLOD have contributed to remove IPBs.

Regarding the estimation of temporal missing data, we will address the problem where some people divided the IPBs and tweeted individually in order to improve the estimation accuracy. We will visualize a specified range of circles that indicate observation points in the tweet application. Moreover, we will implement a selection button, such as "low (less than 10)" and "high (greater than 30)," in order to reduce the work burden.

Regarding the estimation of spatial missing data, we will use time information when simulating airflow around train stations in order to improve the estimation accuracy. Also, we will conduct a field survey and then evaluate our estimation method of spatial missing data once again.

In this study, we cooperated with the "33rd campaign to clean up illegally parked bicycles around the station," through applications of IPBLOD. After finishing this campaign, we will discuss with the Tokyo Bureau in terms of IPBLOD applications and future works. Furthermore, we will facilitate utilizing IPBLOD for solving IPBs by becoming a partner that serves data sets of Civic Tech events, hackathons, and ideathons. Also, we will visualize more statistics of the IPBLOD and clarify the problems caused by IPBs in cooperation with local residents. Furthermore, we will develop a querying interface in order to facilitate the use of IPBLOD.

We believe that our approaches can be recycled for other urban problems containing spatio-temporal information. We will apply our approach to other urban problems such as littering and graffiti.

†http://www.seisyounen-chian.metro.tokyo.jp/kotsu/kakusyutaisaku/jitensha/seisaku-jyourei/churinjou/index.html
††http://www.seisyounen-chian.metro.tokyo.jp/kotsu/kakusyutaisaku/jitensha/houchi/special/english/index.html
†††http://www.seisyounen-chian.metro.tokyo.jp/kotsu/kakusyutaisaku/jitensha/houchi/jitensha-campaign01/jitensha-campaign02/index.html
††††https://ssl.bouhan.metro.tokyo.jp/05_guide/03_apply/ideason/ideason.html
†††††http://www.bouhan.metro.tokyo.jp/90_archive/topic/report_2016/11/p1106.html
††††††http://2016.lodc.jp

References

[1] N. Nishi, "The 2nd health japan 21: goals and challenges," The Journal of the Federation of American Societies for Experimental Biology, vol.28, no.1, p.632.19, 2014.
[2] M. of Internal Affairs and Communications, "Current bicycle usage and bicycle-related accident." http://www.soumu.go.jp/main_content/000354710.pdf. Accessed: 2015-09-10.
[3] S. Bischof, C. Martin, A. Polleres, and P. Schneider, "Collecting, in-

tegrating, enriching and republishing open city data as linked data," Proceedings of the 14th International Semantic Web Conference (ISWC), vol.9367, pp.57–75, 2015.
[4] M. Haklay and P. Weber, "Openstreetmap: User-generated street maps," IEEE Pervasive Comput., vol.7, no.4, pp.12–18, 2008.
[5] S.F. King and P. Brown, "Fix my street or else: using the internet to voice local public service concerns," Proceedings of the 1st international conference on Theory and practice of electronic governance, pp.72–80, 2007.
[6] M. Zook, M. Graham, T. Shelton, and S. Gorman, "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] I. Celino, D. Cerizza, S. Contessa, M. Corubolo, D. Dell'Aglio, E.D. Valle, S. Fumeo, and F. Piccinini, "Urbanopoly: Collection and quality assesment of geo-spatial linked data via a human computation game," Proceedings of the 10th Semantic Web Challenge, pp.148–163, 2012.
[8] L.V. Ahn, "Games with a purpose," IEEE Computer, vol.39, no.6, pp.92–94, 2006.
[9] D.F. Barbieri, D. Braga, S. Ceri, E.D. Valle, and M. Grossniklaus, "C-sparql: Sparql for continuous querying," Proceedings of the 18th International Conference on World Wide Web (WWW), pp.1061–1062, 2012.
[10] S. Auer, C. Bizer, G. Kobilarov, J. Lehmann, R. Cyganiak, and Z. Ives, "Dbpedia: A nucleus for a web of open data," Proceedings of the 6th International Semantic Web Conference (ISWC), vol.4825, pp.722–735, 2007.
[11] S. Consoli, M. Mongiví, A.G. Nuzzolese, S. Peroni, V. Presutti, D.R. Recupero, and D. Spampinato, "A smart city data model based on semantics best practice and principles," Proceedings of the 24th International Conference on World Wide Web (WWW), pp.1395–1400, 2015.
[12] S. Consoli, D.R. Recupero, M. Mongiví, V. Presutti, G. Cataldi, and W. Patatu, "A urban fault reporting and management platform for smart cities," Proceedings of the 24th International Conference on World Wide Web (WWW), pp.535–540, 2015.
[13] S. Batsakis and E.G.M. Petrakis, "Sowl: A framework for handling spatio-temporal information in owl 2.0," Proceedings of the 5th International Symposium on Rules: Research Based and Industry Focused (RuleML), vol.6826, pp.242–249, 2011.
[14] C. Welty and R. Fikes, "A reusable ontology for fluents in owl," Frontiers in Artificial Intelligence and Applications, vol.150, pp.226–236, 2006.
[15] B. Harbelot, H. Arenas, and C. Cruz, "Lc3: A spatio-temporal and semantic model for knowledge discovery from geospatial datasets," Journal of Web Semantics: Science, Services and Agents on the World Wide Web, vol.35, pp.3–24, 2015.
[16] C. Nikolaou, K. Dogani, K. Bereta, G. Garbis, M. Karpathiotakis, K. Kyzirakos, and M. Koubarakis, "Sextant: Visualizing time-evolving linked geospatial data," Journal of Web Semantics: Science, Services and Agents on the World Wide Web, vol.35, pp.35–52, 2015.
[17] S. Egami, T. Kawamura, and A. Ohsuga, "Building urban lod for solving illegally parked bicycles in tokyo," Proceedings of the 15th International Semantic Web Conference (ISWC), vol.9982, pp.291–307, 2016.
[18] C. Stadler, J. Lehmann, K. Höffner, and S. Auer, "Linkedgeodata: A core for a web of spatialopen data," Semantic Web Journal, vol.3, no.4, pp.333–354, 2012.
[19] G.F. Jenks, "The data model concept in statistical mapping," International yearbook of cartography, vol.7, no.1, pp.186–190, 1967.

Shusaku Egami received his master's degree in computer science from University of Electro-Communications in 2016. He is a Ph.D. student in University of Electro-Communications. His research interests include the Semantic Web and Open Data.

Takahiro Kawamura received his Ph.D. degree in computer science from Waseda University in 2001. He was a visiting researcher in Carnegie Mellon University, US, from 2001 to 2002. He is a senior researcher in Japan Science and Technology Agency. He is also an visiting associate professor in the Graduate School of Information Systems at the University of Electro-Communications since 2001, and a part-time lecturer in the Graduate School of Engineering at Osaka University since 2007. His research interests include the Semantic Web, open data, and software agent. He was a board member of the Japanese Society for Artificial Intelligence (JSAI) from 2012 to 2013. He received the 10-Year Award from International Semantic Web Conference 2012.

Akihiko Ohsuga received his Ph.D. degree in computer science from Waseda University in 1995. From 1981 to 2007 he was with Toshiba Corporation. He joined the University of Electro-Communications in 2007, and is currently a professor in the Graduate School of Informatics and Engineering, and is also a dean of the Graduate School of Information Systems. He is also a visiting professor at National Institute of Informatics. His research interests include agent technologies, web intelligence, and software engineering. He is a member of IEEE Computer Society (IEEE CS), Information Processing Society of Japan (IPSJ), Institute of Electronics, Information and Communication Engineers (IEICE), Japanese Society for Artificial Intelligence (JSAI), Japan Society for Software Science and Technology (JSSST), and Institute of Electrical Engineers of Japan (IEEJ). He was a chair of IEEE CS Japan Chapter. He was a member of the board of directors of JSAI and JSSST. He received the IPSJ Best Paper Awards in 1987 and 2017.