Estimation of Spatial Missing Data for Expanding Urban LOD

Shusaku Egami[1], Takahiro Kawamura[1,2], and Akihiko Ohsuga[1]

[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, Tokyo, Japan
takahiro.kawamura@jst.go.jp

Abstract. The illegal parking of bicycles has been an urban problem in Tokyo and other urban areas. We have sustainably built a Linked Open Data (LOD) relating to the illegal parking of bicycles (IPBLOD) to support the problem solving by raising social awareness. Also, we have estimated and complemented the temporal missing data to enrich the IPBLOD, which consisted of intermittent social-sensor data. However, there are also spatial missing data where a bicycle might be illegally parked, and it is necessary to estimate those data in order to expand the areas. Thus, we propose and evaluate a method for estimating spatial missing data. Specifically, we find stagnation points using computational fluid dynamics (CFD), and we filter the stagnation points based on popularity stakes that are calculated using Linked Data. As a result, a significant difference in between the baseline and our approach was represented using the chi-square test.

Keywords: Linked Open Data, Urban problem, Illegally parked bicycles

1 Introduction

The illegal parking of bicycles have been an urban problem in Tokyo and other urban areas since the number of bicycles owned in Japan is large. Illegally parked bicycles (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 illegally parked bicycles (IPBs) as Linked Open Data (LOD). For example, it would serve to visualize IPBs, suggest locations for optimal bicycle parking spaces, assist with the removal of IPBs, and assist with the urban design. Thus, we built the illegally parked bicycle LOD (IPBLOD)[3] based on social data after designing LOD schema [1]. However, there are spatial missing data where bicycles might be illegally parked. It is necessary to complement the spatial missing data in order to apply IPBLOD to various urban areas.

[3] http://www.ohsuga.is.uec.ac.jp/bicycle/dataset.html

However, it is not satisfied merely by social sensors when collecting observation points of IPBs.

In this paper, we propose the method for geographically expanding LOD by estimating spatial missing data. We thought that observation points of IPBs have spatial or geographic features common such as road width and building density. Thus, we first simulated airflow in urban area using computational fluid dynamics (CFD) and found stagnation points. Next, we collected POI data around each of the stagnation points and calculate popularity stakes of the POIs using DBpedia Japanese. Then, we filtered stagnation points if their sum of the popularity stakes of POIs is less than the threshold. We considered the filtered stagnation points as estimated data.

2 IPBLOD and Related Work

We have sustainably built IPBLOD and applied them to Tokyo and other several urban areas . We collected tweets containing location information, pictures, hashtags, and the number of IPBs from Twitter. Also, we collected information on POI using Google Places API[4] and Foursquare API[5]. Also, we obtained bicycle parking data and weather data from websites of municipalities. These data were used as factor data. Finally, the collected data were converted to LOD.

Bischof et al. [2] proposed a method for the collection, complementation, and republishing of data as Linked Data, as with our study. This method collects data from open city data such as Urban Audit[6] and United Nations Statistics Division (UNSD)[7], 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. Therefore, we estimated spatial missing data using CFD and DBpedia Japanese.

3 Estimation of Spatial Missing Data

We consider the flow of people to the fluid, and we find stagnation points of areas around train stations by airflow simulation using 3D maps and CFD. Also, we filter stagnation points using DBpedia Japanese, and we regard these filtered points as new observation points.

3.1 Finding stagnation points using CFD

We simulated the airflow around the station using Airflow Analyst[8], which is a simulation software run on ArcGIS. The wind direction is set as being parallel to a road from a train station, since it is considered that people come to the station along with the roads.

[4] https://developers.google.com/places/?hl=en
[5] https://developer.foursquare.com/
[6] http://ec.europa.eu/eurostat/web/cities
[7] http://unstats.un.org/unsd/default.htm
[8] http://www.airflowanalyst.com/en/index.php

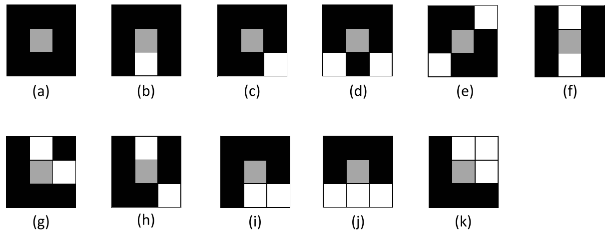

Fig. 1. Patterns of stagnation points

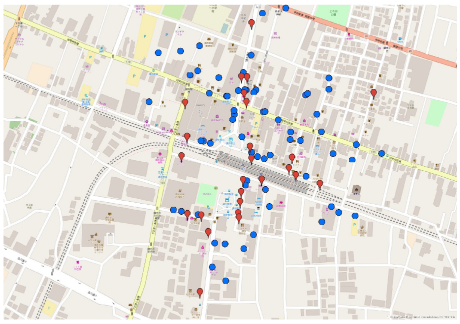

Fig. 2. The filtered stagnation points around of Chofu Station

	Baseline	Proposed method
Precision	0.0559	0.129
Recall	0.161	0.393
F-measure	0.0829	0.194

Table 1. Evaluation results of both baseline and proposed method

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 Figure 1. A black node is a node with average wind velocity $x > 0.1$. The white node is the node which is $x = 0$. The grey node is the node that $0 < x \leq 0.1$. In general, a stagnation point is a white node under these conditions. However, white nodes became buildings in our experiment. Therefore, we defined grey nodes as stagnation points. The total accuracy of the findings of stagnation points around Chofu Station, Fuchu Station, and Shinjuku Station became the highest when we use pattern (j). Hence, we use pattern (j) to find stagnation points in this paper.

3.2 Filtering stagnation points using DBpedia Japanese

We found the stagnation points, but, there were many noise points. 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 of the stagnation points and then filtered the stagnation points using the popularity stakes.

We first obtained the POIs data 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. All types of POIs were manually mapped to resources of DBpedia Japanese. 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. For example, the popularity stakes of the bar resource of DBpedia Japanese (http://ja.dbpedia.org/resource/居酒屋) became 47. Then, we calculated the sum of the popularity stakes of POIs, and we filtered stagnation points if the sum of the popularity stakes is less the threshold. We set the threshold to 200. Figure 2 shows the results of the filtering. Red markers are observation points of illegally parked bicycles. Blue circles are estimated points.

3.3 Evaluation and Discussion

We carried out the experiments on Chofu Station, Fuchu Station, and Shinjuku Station which have multiple observation points of IPBs. The total number of observation points was 56. The baseline estimates the spatial missing data at regular intervals, as many as the number of stagnation points. Table 1 shows the accuracy of both the baseline and the proposed method. As the result, the precision, the recall, and the F-measure of the proposed method became higher than the result of the baseline. Therefore, there is the utility of the proposed method. Also, we validated the utility of the proposed 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 proposed method, and we used a standard level of significance $p < 0.05$. As the result, the p-value of precision was 7.393e-05, and the p-value of recall was 2.244e-06. Hence, we found that there is a significant difference between the result of the baseline and the result of the proposed method.

The accuracy of the estimated data in this study was low for the following reasons. The number of observation points was less. There is a possibility that new observation points are found around the estimated points.

4 Conclusion and Future Work

In this paper, we described geographically expansion of IPBLOD by estimating the spatial missing data. The mainly technical contribution is the proposal of a hybrid method using CFD and DBpedia Japanese for estimating the spatial missing data in LOD. In the future, we will estimate spatial missing data in more urban areas, and we will check true-false results to go to estimated points. Furthermore, we will visualize estimated observation points and will design incentive for social sensors (workers of crowdsourcing), in order to collect more data related to IPBs.

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

References

1. Egami, S., Kawamura, T., Ohsuga, A.: Building Urban LOD for Solving IPBs in Tokyo. In:Proc. The 15th International Semantic Web Conference, pp.291-307 (2016)
2. 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)