Station-level demand prediction for bike-sharing systems planning with graph convolutional neural networks

Nikolaos Staikos*, Pengxiang Zhao*, Ali Mansourian*

* Department of Physical Geography and Ecosystem Science, Lund University, Lund, Sweden

Abstract. Accurately predicting bike-sharing demand at the station level is of paramount importance to facilitate station planning and enhance the efficiency of bike-sharing systems. In this study, we develop graph convolutional neural networks (GCNN) to predict station-level bike-sharing demand by modeling the spatial dependence of stations in two ways, namely trips and nearest neighbor, and compare the prediction performance with three machine learning models, including multiple linear regression, multilayer perceptron (MLP) and random forest. The two GCNNs and three machine learning models were implemented and evaluated using a bike-sharing trip dataset in Zurich, Switzerland. The results show that the GCNN model based on the graph structure built by k nearest neighbor achieves the best prediction performance. The way of modeling spatial dependence of bike-sharing stations presents an influence on the prediction. This research is beneficial for decision-making in establishing new stations to support bike-sharing systems planning.

Keywords. Bike-sharing demand prediction, graph convolutional neural networks, spatial dependence, machine learning

1. Introduction

Bike-sharing systems (BSS) have been operated and popularized in many cities worldwide, aiming to reduce transport-related carbon emissions and promote sustainable urban mobility (e.g., Li et al., 2021). Although the adoption of BSS brings environmental and social benefits, unbalanced usage of bike-sharing raises concerns (e.g., inefficient resource allocation). Demand prediction plays a crucial role in BSS planning by helping opera-

Published in "Proceedings of the 18th International Conference on Location Based Services (LBS 2023)", edited by Haosheng Huang, Nico Van de Weghe and Georg Gartner, LBS 2023, 20-22 November 2023 Ghent, Belgium.

This contribution underwent single-blind peer review based on the paper.
https://doi.org/10.34726/5745

tors optimize their resources, improve customer satisfaction, and ensure efficient operations (Boufidis et al., 2020). Predicting demand accurately allows BSS to allocate bikes and docking spaces effectively, thus reducing instances of overcrowding or empty stations and improving users' satisfaction.

The current bike-sharing demand prediction is mainly implemented at three levels of research units, namely cluster, area, and station (Xiao et al., 2021). In this paper, we focus on station-level bike-sharing demand prediction. As a typical spatial regression problem that relies heavily on observed variables across geographic space, station-level demand prediction has attracted notable attention in the fields of GIS and transportation (e.g., Lin et al., 2018; Yang et al., 2020). Since bike-sharing stations are usually distributed irregularly across the space, it is necessary to model and understand the spatial dependence effects of stations on demand prediction. Graph-based deep learning that explicitly models the relationships among connected locations can be an effective tool to deal with the problem (Zhu et al., 2021). Although graph convolutional neural networks have been used to predict bike-sharing demand, the existing studies are mainly concentrated on short-term prediction using historical demand time series (e.g., Lin et al., 2018; Xiao et al., 2021). In addition, little attention has been paid to examining the influence of graph structure (e.g., neighbor-based, distance-based) on demand prediction with GCNNs.

To fill the gap, this study is dedicated to systematically modeling the spatial relationships of bike-sharing stations and developing GCNNs in different graph structures to predict station-level bike-sharing demand based on influencing factors. The developed GCNNs are tested using the bike-sharing trip dataset in Zurich, Switzerland. Since the proposed method is not constrained by historical demand data, it could serve as a planning tool for municipal urban planners and operators to establish new stations in bike-sharing systems.

2. Data and methods

2.1. Data

The public bike-sharing (PBS) trip dataset is collected from a service provider company in Zurich, from May 27 to July 7, 2022. Each trip records the following attributes: id, start/end time, start/end longitude, start/end latitude, start/end station id, distance, and duration. In addition, land use, points of interest (POI), public transport stops, population density, and employment density datasets are also collected from open geoportals to calculate the influencing factors.

2.2. Variables

Previous studies (e.g., Eren and Uz, 2020) have indicated that bike-sharing studies are related to various influencing factors, such as built environment, public transport, socio-demographic factors, and weather conditions (e.g., temperature and precipitation), etc. In this study, 16 variables are calculated to describe the influencing factors and used to train GCNNs and machine learning models, as shown in Table 1. To estimate the variables at the station level, a 300-meter buffer is created for each bike-sharing station based on prior knowledge.

Categories	Variables
Urban built environment	Bike lane density, Land use mixture, Education density, Tourism density, Healthcare density, Entertainment density, Sports density, Accommodation density, Errand density, Dining density
Public transport	Public transport stop density, Closest distance to bus stops, Closest distance to railway stations, Closest distance to tram stops
socio-demographic factors	Population density, employment density

Table 1. Variables for influencing factors.

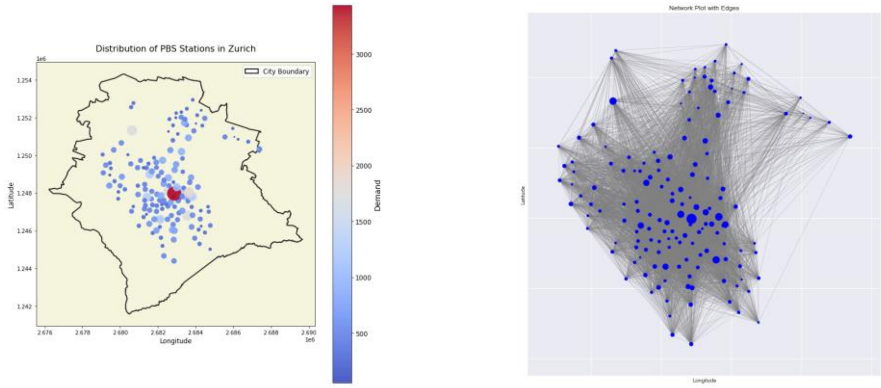
2.3. Model description

A Graph Convolutional Neural Network (GCNN) is a deep learning technique designed for the analysis and processing of data structured as graphs (Kipf and Welling, 2017). In a graph, entities are represented as nodes, and the relationships between entities are captured by edges connecting these nodes. A GCNN leverages the inherent graph structure to perform convolutions and learn hierarchical representations directly from the graph data. Zhu et al. (2021) further proposed the use of spatial regression graph convolutional neural networks (SRGCNN) as a deep learning paradigm in spatial regression analysis.

As an invariant of convolutional neural networks (CNNs), GCNN applies convolutional operations to node features and adapts these operations to the irregular graph structure compared with grid-like data in CNNs. Therefore, graph structure has a significant influence on the prediction performance. In this study, we attempt to model the spatial dependency effects among bike-sharing stations and thus construct graph structures in two ways. The first graph structure is constructed based on whether trips occurred between two stations. One edge is added to the corresponding nodes if there are trips occur between two stations. The number of trips between them can be the weight of the edge. The second graph structure is built based on k nearest neighbor inspired by the study (Zhu et al., 2021).

3. Experiment and results

The GCNNs are developed based on the two graph structures and tested using the station-level demand and influencing factors in Zurich. Figure 1 (a) shows the spatial demand distribution across bike-sharing stations. The stations with high demand are mainly concentrated in the city center. Figure 1(b) displays the graph structure based on whether trips occurred between two stations.



(a) Demand distribution (b) Graph structure

Figure 1. Demand distribution across stations and the graph structure.

The two GCNNs (i.e., $GCNN_{trip}$ and $GCNN_{neighbor}$) are implemented and validated by comparing with three machine learning models, including multiple linear regression, MLP regressor, and random forest regressor, in terms of Mean Square Error (MSE), Root Mean Square Error (RMSE) and R-squared. In this study, the entire dataset is split into 80% training data and 20% percent test data. Table 2 shows the prediction performance comparison of the five models in terms of the three evaluation metrics. It can be observed that the $GCNN_{neighbor}$ model achieves the best performance. It implies that the graph structures that model the spatial dependence of bike-sharing stations have a significant influence on the prediction using GCNN.

Models	**Evaluation metrics**		
	MSE	RMSE	R^2
$GCNN_{trip}$	328452.56	573,11	0.58
$GCNN_{neighbor}$	**306828.62**	**553.92**	**0.60**
Multiple linear regression	353501.13	594.56	0.03
MLP	342384.41	585.14	0.06
Random forest	320748.70	566.35	0.12

Table 2. Prediction performance comparison of different models.

4. Conclusion

In this study, we developed the GCNNs to predict the station-level bike-sharing demand by modeling the spatial dependence of stations in two ways: trips and nearest neighbors. Using the datasets in Zurich for evaluation, the results indicate that the graph structure that models the spatial dependence of bike-sharing stations influences the prediction. Overall, the developed GCNNs present potential for accurately predicting demand. For our future work, distance-based spatial dependence modeling will be exploited in the prediction of GCNNs. In addition, spatial analysis of errors will be conducted to examine how the underestimated or overestimated errors are distributed across space, thereby further improving the prediction performance of the models. Finally, the generalizability and transferability of the GCNN models will be investigated by applying them to other cities.

References

Boufidis, N., Nikiforiadis, A., Chrysostomou, K. and Aifadopoulou, G., 2020. Development of a station-level demand prediction and visualization tool to support bike-sharing systems' operators. Transportation Research Procedia, 47, pp.51-58.

Eren, E. and Uz, V.E., 2020. A review on bike-sharing: The factors affecting bike-sharing demand. Sustainable cities and society, 54, p.101882.

Kipf, T.N. and Welling, M., 2016. Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907.

Li, A., Gao, K., Zhao, P., Qu, X. and Axhausen, K.W., 2021. High-resolution assessment of environmental benefits of dockless bike-sharing systems based on transaction data. Journal of Cleaner Production, 296, p.126423. doi: 10.1016/j.jclepro.2021.126423

Lin, L., He, Z. and Peeta, S., 2018. Predicting station-level hourly demand in a large-scale bike-sharing network: A graph convolutional neural network approach. Transportation Research Part C: Emerging Technologies, 97, pp.258-276.

Xiao, G., Wang, R., Zhang, C. and Ni, A., 2021. Demand prediction for a public bike sharing program based on spatio-temporal graph convolutional networks. Multimedia Tools and Applications, 80, pp.22907-22925.

Yang, Y., Heppenstall, A., Turner, A. and Comber, A., 2020. Using graph structural information about flows to enhance short-term demand prediction in bike-sharing systems. Computers, Environment and Urban Systems, 83, p.101521.

Zhu, D., Liu, Y., Yao, X. and Fischer, M.M., 2021. Spatial regression graph convolutional neural networks: A deep learning paradigm for spatial multivariate distributions. GeoInformatica, pp.1-32.