Spatial Attention Based Grid Representation Learning For Predicting Origin–Destination Flow

Mingfei Cai
Department of Civil Engineering
The University of Tokyo
Meguro, Tokyo, Japan
mfcai@iis.u-tokyo.ac.jp

Yanbo Pang
Center for Spatial Information Science
The University of Tokyo
Meguro, Tokyo, Japan
pybdtc@csis.u-tokyo.ac.jp

Yoshihide Sekimoto
Center for Spatial Information Science
The University of Tokyo
Meguro, Tokyo, Japan
sekimoto@csis.u-tokyo.ac.jp

***Abstract*—Origin–destination (OD) flow data are critical for urban planning and traffic system design. Such data are suitable for describing movement at the macroscopic level. However, collecting them on a large scale, such as in a city, is challenging. Their form incompatibility makes using them for other tasks difficult. Therefore, we propose a deep model to learn meaningful OD information on grids within a city to address these problems. We collected multimodal characteristics of regions, such as road network densities and facility distributions, from several open-source datasets and used them as grid signals. We then constructed a spatial attention-based deep graph network to generate grid embeddings and used them to predict the OD volumes. The proposed method was evaluated against a set of baseline approaches using a real-world dataset in Japan. The analysis indicated that our model can extract more accurate latent topographical information from OD graphs and produce reasonable grid embeddings; these representations apply to other downstream tasks.**

***Index Terms*—OD flow, graph attention network, grid embedding**

I. Introduction

The estimation of people flow is essential for urban planning. The government can allocate resources to specific areas such that more people can benefit from this arrangement. There are numerous types of information to describe people flow, and the origin–destination (OD) matrix is an aggregated data form that can depict the phenomenon of the crowd [1]. It can uncover the trend of massive movement on a large scale compared with other data, such as individual trajectories. However, collecting OD volume data within a large scope is not trivial, regardless of whether traditional censuses or telecom records are used [2]. Creating a prediction model that can generate OD matrices based on the indicators of the accessible auxiliary region is advantageous. Many methodologies have been proposed to predict OD matrices. Given that the road network naturally maintains a topological structure, graph representation learning has recently gained popularity as a method for estimating urban flow. Implicit geometrical information can be effectively implemented by modeling the network as a relational structure and using graph convolutions to represent complicated interactions. Graph learning can reflect urban components as dense embeddings and conduct a similarity analysis between them. Such a representation is significant for learning the semantics of urban areas and applying the model to various cities. Therefore, the construction of OD graphs with the by-product of region embeddings is an appropriate way to describe the urban flow and municipal functions.

However, generating accurate OD graphs is difficult. First, the structure of an OD graph is implicit and can be defined in different ways. Most studies use a geographical graph, which means that neighboring regions on the map are connected by an edge. This design neglects trips covering large distances because long-distance trips require messages passing through numerous hops, which is also equal to the number of the convolutional layers. Some studies have considered the concept of semantic neighbors, which means that two regions are connected if there are OD flows between them. Such settings can leverage data-driven knowledge and consider complex urban movement patterns. However, geographic patterns are disregarded, which results in the loss of network information. In this study, we propose a spatial attention-based graph construction. First, we construct a graph structure from historical OD data. Furthermore, we incorporate spatial patterns into the attention mechanism so that the network can elucidate the relationship between different areas on both the spatio-structural and semantic sides.

Second, OD volume prediction is more difficult than tasks such as traffic speed prediction. Predicting a general OD flow with multiple modes of transportation is more difficult than it is with a single mode, such as a taxi or subway. The distribution of the data tends to be unbalanced, which means that extreme values are more common. Some OD pairs may have a considerable traffic volume, whereas others may have no volume. Consequently, a simple model structure cannot provide accurate predictions for several cities simultaneously. Additionally, given the same OD pair, the bidirectional people flow diversifies significantly. All aforementioned cases lead to the need for a more powerful model to achieve the purpose of OD volume prediction.

Consequently, traditional regression approaches cannot address the complicated latent relationship between regional indicators and OD flows. The implicit spatial structure even entangles clues for the model construction. Thus, we propose a graph convolutional model to learn OD semantics over a city. Figure 1 provides an overview of the model structure.

978-1-6654-8045-1/22/$31.00 ©2022 IEEE

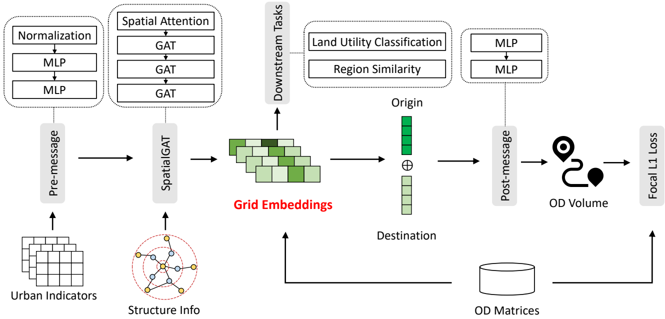


Fig. 1: Overview of this study.

Our approach relies on graph attention networks (GAT) and integrates the spatial features among areas. GAT is a spatial-based approach; hence, dealing with dynamic structures and new graphs to achieve application in other target areas is facile. We constructed a graph employing the units of the grid. The GAT performs as an encoder model and outputs a dense embedding for each grid. The origin and destination grids are encoded to create edge representations. We used a multilayer perceptron as the decoder to calculate the volume of the OD pairs. Our proposed model can extract information from open-source data to illustrate the combination flow from different transportation sources considering multimodal people flow.

Furthermore, the aim of this study is not limited to the accurate prediction of OD volume. Another initiative is to use generated location representations. We conducted two downstream tasks, namely cross-city similar areas extraction [3] and land utility classification [4], to elucidate embeddings containing valuable OD semantics that can be used in other related urban tasks.

The main contributions of our work are as follows:

- We designed a deep graph-learning model to reconstruct the OD matrices. The model can simultaneously consider OD semantics for historical data and geographical structures to build OD graphs for effective message passing.
- We used multiple public census data as grid indicators to learn a set of meaningful embedding containing OD volume information.
- We designed two downstream urban tasks to validate the usage of generated location representations.

Extensive experiments in six Japanese cities with different attributes showed that the proposed model can achieve satisfactory performance. The remainder of this paper is organized as follows. Section II derives definitions of key concepts and the problem formulation. Section III introduces the methodology. Section IV provides the experimental details as well as the results and discussion. Section V reviews the literature. Section VI concludes the paper and discusses future work.

II. Preliminary

In this section, we first present the definitions of the key concepts related to our model. Then, we provide the problem formulation. In this study, we considered the unit of the mesh grid.

Definition 1 (Mesh grid). The target region is divided into N subareas of equal size (e.g., $500\,m \times 500\,m$), denoted by $\mathcal{M} = \{m_1, m_2, \ldots, m_N\}$. Figure 3 shows the mesh grid division in Fuji City, Shizuoka Prefecture. Each grid has a code calculated from the longitude and latitude of its central point. We adopted a division approach from the Statistics Bureau of Japan [5].

Definition 2 (Urban indicators). Each mesh grid has unique characteristics that can be summarized using multisource information. The matrix $\mathcal{U} \in \mathbb{R}^{N \times D}$ describes the D dimensional attributes of the mesh grids, which we refer to as the urban indicator. The source of indicators includes several aggregated open data, such as facility distributions and night population.

OD matrices efficiently determine macroscopic human mobility. Use of such data avoids the leakage of private information and enables a grasp of the overall knowledge of people flow of population on a large scale.

Definition 3 (OD matrix). The OD matrix is a data format used to describe the flow of massive numbers of people. In the OD matrix $\mathcal{A} \in \mathbb{R}^{N \times N}$, the row denotes the origin, and the column denotes the destination. The origin and destination are represented by mesh grids in the aforementioned definition. The value of each cell $\mathcal{A}_{i,j}$ indicates the number of people who travel from the origin m_i to the destination m_j.

We can construct OD graphs from OD matrices to simultaneously consider the semantics of the regions and topological structure information.

Definition 4 (OD graph). Given an OD matrix $\mathcal{A}$, the OD graph is formulated as a directed graph $\mathcal{G} = (V, E)$, where $V = \{v_1, v_2, \ldots, v_N | \mathcal{M}, \mathcal{U}\}$ represents the vertex, and $E = \{e_{i,j} | \mathcal{A}_{i,j} > 0\}$ is the edge connecting two vertices. We can generate meaningful grid embeddings using OD graphs to obtain multisource information for diversified downstream tasks.

Definition 5 (Grid embedding). Given an OD graph $\mathcal{G}$, the grid embedding learns a function $\mathcal{F} : v_i \rightarrow \mathbb{R}^d$ that maps the urban indicator to the low-dimensional representation of each grid, where d is the dimension of the grid embedding.

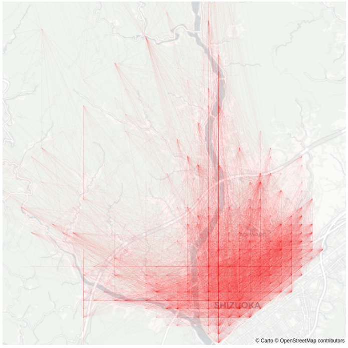
A. Problem Formulation

We investigated the OD volume prediction as a link regression problem. The goal is to learn the mesh grid embedding $\mathcal{F}$ from the OD graph $\mathcal{G}$, given the mesh grid of target area $\mathcal{M}$ with urban indicators $\mathcal{U}$. The model then predicts the values of all cells in OD matrix $\mathcal{A}$ using the abovementioned embeddings.

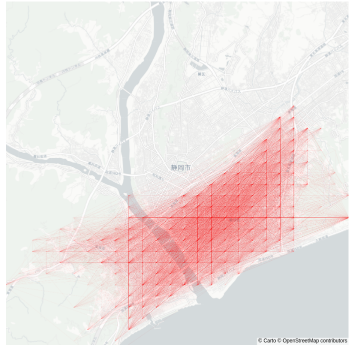
Note that the OD matrix is asymmetric, which means that $|\mathcal{A}_{i,j}| \neq |\mathcal{A}_{j,i}|$, and the edge representation from grid embeddings needs to be order-sensitive for the start and end areas. Therefore, we concatenate the origin and destination embeddings to generate an edge vector,

$$e_{i,j} = h_i \oplus h_j \tag{1}$$

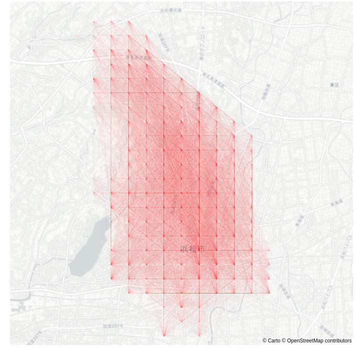
Then, the edge representation will be regressed to calculate the final OD flow value $A_{i,j}$.



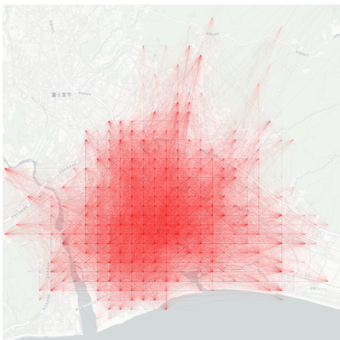
(a) Aoi Ward, Shizuoka City



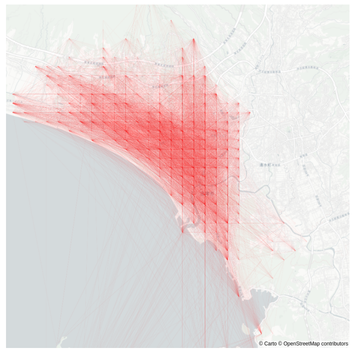
(b) Suruga Ward, Shizuoka City



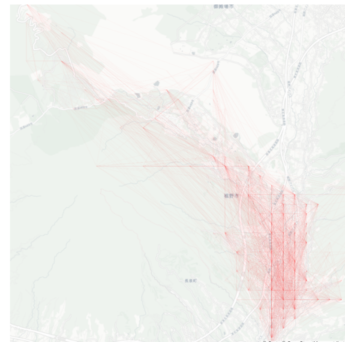
(c) Naka Ward, Hamamatsu City



(d) Fuji City



(e) Numazu City



(f) Susono City

Fig. 2: OD graph of six target areas. We performed the experiment in Shizuoka Prefecture, located in central Japan. We selected six areas with different characteristics. Aoi Ward, Suruga Ward, and Naka Ward are typical city areas of relatively large, medium, and small size. Similarly, Fuji City, Numazu City, and Susono City are local areas of different sizes.

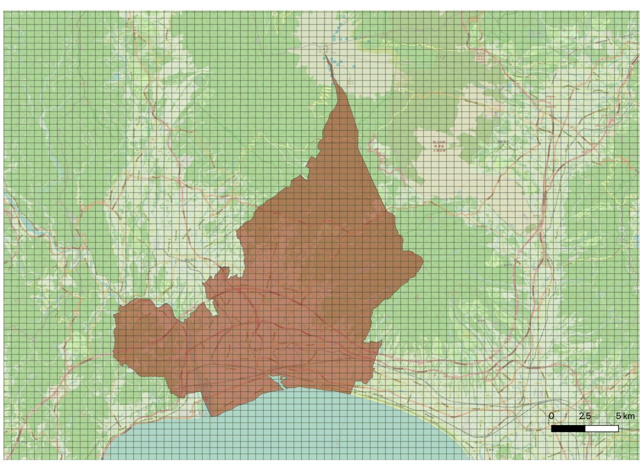


Fig. 3: The $500\,m$ grid system in Fuji City, Shizuoka Prefecture. The target city area is divided into $500\,m \times 500\,m$ squares. Each square is called a grid and assigned a code calculating from the central point of this grid.

III. Methodology

In this section, we provide the details of the proposed framework. Our motivation was to leverage multisource open data to describe grids and train a model to grasp meaningful embeddings from the original urban signals.

A. GAT Encoder–Decoder Framework

We maintained an encoder–decoder framework for OD volume prediction. All the parts were trained jointly to learn the parameters in an end-to-end manner. Algorithm 1 shows the proposed framework for the OD graph construction.

1) Generation of Message: Each grid was initialized with a descriptive signal vector, as mentioned in Section II. This vector concatenates several numerical grid attributes from different open-source census data. The range of data varies significantly owing to the multiple data sources. Thus, we first normalized the indicator to be in a Gaussian distribution across each indicator category to deal with multisource data.

$$\bar{x} = \frac{x - \mu}{\sigma} \tag{2}$$

where μ is the mean and σ is the standard deviation of data.

The normalized signal vector is then passed into a pre-message layer to generate the initial message. The pre-message layer is a simple multilayer perceptron (MLP).

2) Generation of Grid Embedding: The GAT-encoder utilizes the initial message to generate the mesh-grid embeddings. The encoder considers both mesh grid semantics and spatial structural information to construct an OD graph using the attention mechanism. We adopted GATv2 for the convolutional layers, as proposed by [6], to enable a dynamic attention mechanism. The attended nodes are conditioned by the query nodes such that every node can consider any other node when calculating attention weights compared with the original GAT. The attention weight was calculated as follows:

$$\alpha_{i,j} = \frac{exp(\mathbf{a}^T LeakyReLU(\Theta[\mathbf{x_i} \| \mathbf{x_j} \| \mathbf{e_{i,j}}]))}{\sum_{k \in \mathcal{N}(i) \cup \{i\}} exp(\mathbf{a}^T LeakyReLU(\Theta[\mathbf{x_i} \| \mathbf{x_k} \| \mathbf{e_{i,k}}]))} \tag{3}$$

We combined the initial message of the target node and its neighboring nodes for the mesh grid semantics part. Conversely, we used the inverse of the distance between center points for two mesh grids to reflect the geographic structure regarding the spatial structure information. As we adopted the grid system for the division of areas, such a setting was effective for the inclusion of geographical characteristics. These two parts were concatenated and passed through linear and nonlinear transformations.

After calculating the attention weight, the message for the node v_i can be computed as

$$\mathbf{x_i'} = \alpha_{i,i}\Theta\mathbf{x_i} + \sum \alpha_{i,j}\Theta\mathbf{x_j} \tag{4}$$

Moreover, residual blocks were added across the convolutional layers to increase the expressive ability of the network. Skipping connections between different layers can alleviate the over-smoothing problems. Deeper graph convolutional models tend to be poorer at differentiating nodes. All messages become the same so that the model loses the ability to make accurate predictions, known as over-smoothing problems. Adding skip connections can underscore the impact of earlier layers to solve this problem. It is also reasonable to consider the background of the OD volume prediction problems for others. Commonly, trips covering a long distance consist of several sub-trips that generate an OD sequence. It is essential to elaborate on the relationship between the initial origin and the last destination for a better understanding of the trip, rather than emphasizing the stay points in the middle. The residual blocks can build a direct connection between two mesh grids that are far from each other to overcome the limitations of the local neighborhood.

Note that we did not separately train the inflow and outflow embeddings. We used only two different linear layers to interpret the mesh grid embeddings as the origin and destination reflections. This is because the inflow and outflow are correlated rather than independent. Furthermore, different embedding layers for inflow and outflow hamper the usage of generated representations in other tasks. Therefore, the model produces a general mesh grid vector for the inward and outward OD flows.

3) Generation of OD Volume: The decoder calculates the volume of people flow for OD pairs from the mesh grid embeddings. The origin and destination embeddings were concatenated to create edge representations after generating the grid embeddings. The edge representations were passed into the post-message layers to calculate the flow volume of OD pairs. The post-message layer was a two-layer MLP, which was the same as the pre-message layer.

B. Training

OD data can have a severely skewed distribution as we deal with the general OD volume prediction, which considers all transportation modes. We implemented a regression version of the focal loss adapted by [7] to tackle the unbalanced data distribution problem. The focal L1 loss function assigns larger weights to the more difficult data points.

$$\mathcal{L} = \frac{1}{n}\sum_i \sigma(|\beta e_i|^\gamma)e_i \tag{5}$$

where n denotes the number of edges. β and γ are the hyperparameters. e_i is the difference between the prediction and the ground-truth labels of edge i, and $\sigma(\cdot)$ is the sigmoid function.

We split the entire graph into training, validation, and test datasets. First, the proposed model was trained to generate an embedding for each grid in the training dataset. The model learns the transformation function from the initial grid indicators to the grid embeddings. The same aggregation parameters were shared for all nodes to ensure inductive capability. The learned shared weight can be used to deal with indicators in the same form from the validation and test datasets.

Algorithm 1: Generation of OD graphs

Input: Historical OD Lists, Region Indicators
Output: OD Volume Matrices, Region Embeddings
$N_{node} \leftarrow number\ of\ nodes(grids)$
$N_{edge} \leftarrow number\ of\ edges(OD\ pairs)$
initialize the graph $\mathcal{G}$
for $i\ in\ 1 : N_{edge}$ **do**
  add the edge E_i to the graph $\mathcal{G}$
  set OD volume of E_i to 0
for $i\ in\ 1 : N_{node}$ **do**
  initialize urban indicators u_i
pre-message layer: generate the message $\mathbf{x}'$
while $\mathcal{L} \geq$ *threshold* **do**
  GAT message passing
  generate mesh grid embeddings $\mathcal{F}$
  for $i\ in\ 1 : N_{edge}$ **do**
    calculate the edge vector e_i
  post-message layer: calculate OD volume $\mathcal{A}$
  calculate the loss $\mathcal{L}$
  go to 10

C. Downstream Tasks

Embeddings are often used in diverse downstream tasks. We utilized trained embeddings in two urban downstream tasks: region similarity calculation and land utility prediction, to test the usage of our proposed grid embeddings.

1) Region Similarity Analysis: Embeddings are representations of low dimensions for complex semantics with high dimensions. Thus, the generated grid embeddings maintained spatial attributes in their spaces. This property can be used to compare the similarities between different grids in different areas. This will help us understand the essence of grid function distribution, regardless of diversified urban layouts. In practice, we selected one typical grid in one region and calculated the similarity with all grids in the other city.

Specifically, we used the cosine similarity to illustrate the closeness of the two grids. The range of the similarity is $[-1, 1]$, where a higher value indicates higher similarity.

$$\mathcal{S}\langle \boldsymbol{a}, \boldsymbol{b} \rangle = \frac{\boldsymbol{a} \cdot \boldsymbol{b}}{\|\boldsymbol{a}\| \cdot \|\boldsymbol{b}\|} \tag{6}$$

2) Land Utility Classification: Land utility classification is another appropriate downstream task of mesh grid embeddings for the OD volume. The functionality of the mesh grid determines the purpose of the travel, which is reflected in the OD volume distribution map. For instance, for residential areas, the outward OD flow on weekdays tends to be large in the morning and small in the evening, which matches the patterns of commuting trips. Thus, land utility classification using mesh grid embeddings can validate the accuracy of the proposed model. We directly applied k-means clustering to the generated embeddings and compared the clustering results with the ground-truth data.

IV. Experiments

In this section, we present the details of the experiment, including the datasets used, experimental settings, and baseline models.

A. Dataset

In this subsection, we present the details of the datasets used in the experiment. The datasets consist of the OD volume dataset from SoftBank Group Corporation and urban indicator datasets from multiple open-source censuses.

1) OD Volume: We used National Move Statistics as the OD volumes source data, provided by SoftBank Corporation, one of the biggest telecom companies in Japan. The SoftBank Corporation aggregates location-based service data from mobile base stations in Japan. The raw data are upsampled to cover the entire population rather than just the SoftBank mobile users. The data resolution was $500\,m$, the same as the previous definition in Section II. Each row of data indicates the volume of OD flow for one specific OD pair within a day. Additionally, the number of OD flows that are less than the threshold is hidden to protect privacy. We utilized data within one week (October 13–19, 2019) for the ground-truth labels of the predictions.

2) Grid Indicators: The key to achieving a good model performance is the use effective features for nodes in graphs. We leveraged multisource features and collected data from several open datasets [8] [9] [10] [11]. Table II provides a summary of the grid indicators used in this study.

First, the night population [10] indicates the number of residents in the grid. Intuitively, numerous OD flows come from people living in the area. Thus, the night population indicator has a strong relationship with the OD flow volume. Second, the density of a road network implies traffic convenience. The more convenient the road network, the larger is the people flow of the grid. We combined some categories in the original road density dataset to obtain a more compact signal vector to improve efficiency. Third, annual passengers of the railway station can provide information on long trips, as the railway connects two grids over a long distance. We used the passenger number of railway stations in 2020, offered by Ministry of Land, Infrastructure, Transport and Tourism (MLIT), Japan. The number of passengers was added if there were several stations in one grid. The number was divided equally and shared with each grid if one station covered several grids. Furthermore, the distribution of the point of interest (POI) was used to illustrate grid attraction. The data comprised 40 categories. This indicator includes the number of specific types of POI as well as employees and assigns different signals for trips for different purposes.

B. Experimental Setup

In this subsection, we provide the settings for conducting different experiments with the baseline and proposed models used for comparison.

We performed an experiment in Shizuoka Prefecture, which is located in central Japan. We selected six areas with different characteristics: Aoi Ward in Shizuoka City, Suruga Ward in Shizuoka City, Naka Ward in Hamamatsu City, Numazu City, Fuji City, and Susono City. Specifically, Aoi Ward, Suruga Ward, and Naka Ward are typical city areas of relatively large, medium, and small sizes, respectively. Similarly, Fuji city, Numazu City, and Susono City are local areas of different sizes. Table III summarizes the details of the six target areas. Figure 2 illustrates OD graphs of the six target areas.

The ratio of training, validation, and test sets was 8:1:1. The model was constructed using the PyTorch geometric library [12]. Regarding the related hyperparameters, we set the embedding dimensions to 128 and the number of hidden channels to 256. We also implemented multi-head attention over graph convolution to stabilize the learning process. Specifically, we set the number of heads to four. Furthermore, we used the Adam optimizer for training. The computation was accelerated using a GPU. We conducted experiments on the Amazon p2.xlarge instance with one NVIDIA K80 GPU, 4vCPUs, and 61 GiB of host memory.

C. Baseline Models

We compared our model with several baseline models. We chose models ranging from traditional models to state-of-the-art deep learning models.

The gravity model is a traditional model that considers static grid indicators for mutual effects between the two areas [13]. Decision tree and random forest models are tree-based models often used for regression problems. The gradient boosting model is a traditional machine learning technique used for regression tasks that uses an ensemble of several weak models for better predictions. A naïve multilayer perceptron model is a

Model	Area								
	Aoi Award, Shizuoka City			Suruga Ward, Shizuoka City			Naka Ward, Hamamatsu City		
	RMSE	**MAE**	**COR**	**RMSE**	**MAE**	**COR**	**RMSE**	**MAE**	**COR**
Gravity Model	153.39	68.93	0.04	149.39	121.21	0.09	438.66	406.93	0.15
Decision Tree	191.15	50.84	0.13	146.51	50.38	0.15	184.83	65.30	0.29
Random Forest	137.74	42.45	0.09	109.05	43.22	0.13	128.31	59.28	0.22
Gradient Boosting	137.84	42.14	0.19	106.00	41.30	0.27	119.97	55.46	0.39
2-Layer MLP	146.61	58.39	0.15	128.67	93.27	0.15	127.74	63.88	0.23
GraphSAGE	75.31	22.23	0.48	99.44	33.62	0.35	**75.30**	**21.82**	0.47
SpatialGAT	**67.34**	**21.40**	**0.64**	**79.18**	**22.95**	**0.58**	91.64	35.33	**0.72**
Model	**Area**								
	Numazu City			Fuji City			Susono City		
	RMSE	**MAE**	**COR**	**RMSE**	**MAE**	**COR**	**RMSE**	**MAE**	**COR**
Gravity Model	229.19	166.11	0.11	87.78	77.23	0.04	123.22	93.25	0.27
Decision Tree	189.47	49.07	0.11	86.56	29.18	0.08	148.91	46.87	0.06
Random Forest	106.88	40.31	0.09	57.27	21.93	0.12	97.45	39.03	0.20
Gradient Boosting	111.47	39.75	0.13	56.65	21.47	0.18	94.75	36.25	0.29
2-Layer MLP	113.69	52.21	0.22	75.49	56.95	0.11	99.02	56.22	0.28
GraphSAGE	**50.65**	**22.00**	0.38	50.59	22.97	0.39	88.15	30.12	0.39
SpatialGAT	108.80	24.45	**0.48**	**45.26**	**14.31**	**0.58**	**69.53**	**25.89**	**0.62**

TABLE I: Result of baseline models and the proposed model.

Feature Categories	#Features	Contents
Road Densities	24	Road number and density of different widths for mesh in 2010 [8]
Facilities (POIs)	40	Number of facilities and employees in different industrial categories in 2016 [9]
Grid Population	1	Night population distribution in 2015 [10]
Railway Users	1	Number of annual railway station users in 2019 [11]
Total	66	

TABLE II: Grid indicators summary.

Name	Area /$\mathbf{km^2}$	Population	#Grids	#OD Pairs
Aoi	1073.75	246221	366	17285
Suruga	73.06	211635	191	12570
Naka	44.34	234013	163	11906
Numazu	186.96	186169	280	16095
Fuji	244.95	242890	401	30824
Susono	138.12	49604	146	3375

TABLE III: Summary of target areas.

common approach for illustrating the complicated relationship between features and targets. GraphSAGE is a spatial-based graph learning model that is similar to GAT, but it treats all neighbors under equal importance.

D. Evaluation Metrics

We measure performance of the regression problem with the following metrics:

- Root Mean Squared Error (RMSE)

$$\mathbf{RMSE}(\hat{y_i}, y_i) = \sqrt{\frac{\Sigma_{i=1}^{N}(\hat{y_i} - y_i)^2}{N}} \tag{7}$$

- Mean Average Error Ratio (MAE)

$$\mathbf{MAE}(\hat{y_i}, y_i) = \frac{\Sigma_{i=1}^{N}|\hat{y_i} - y_i|}{N} \tag{8}$$

- Pearson Correlation Coefficient (COR)

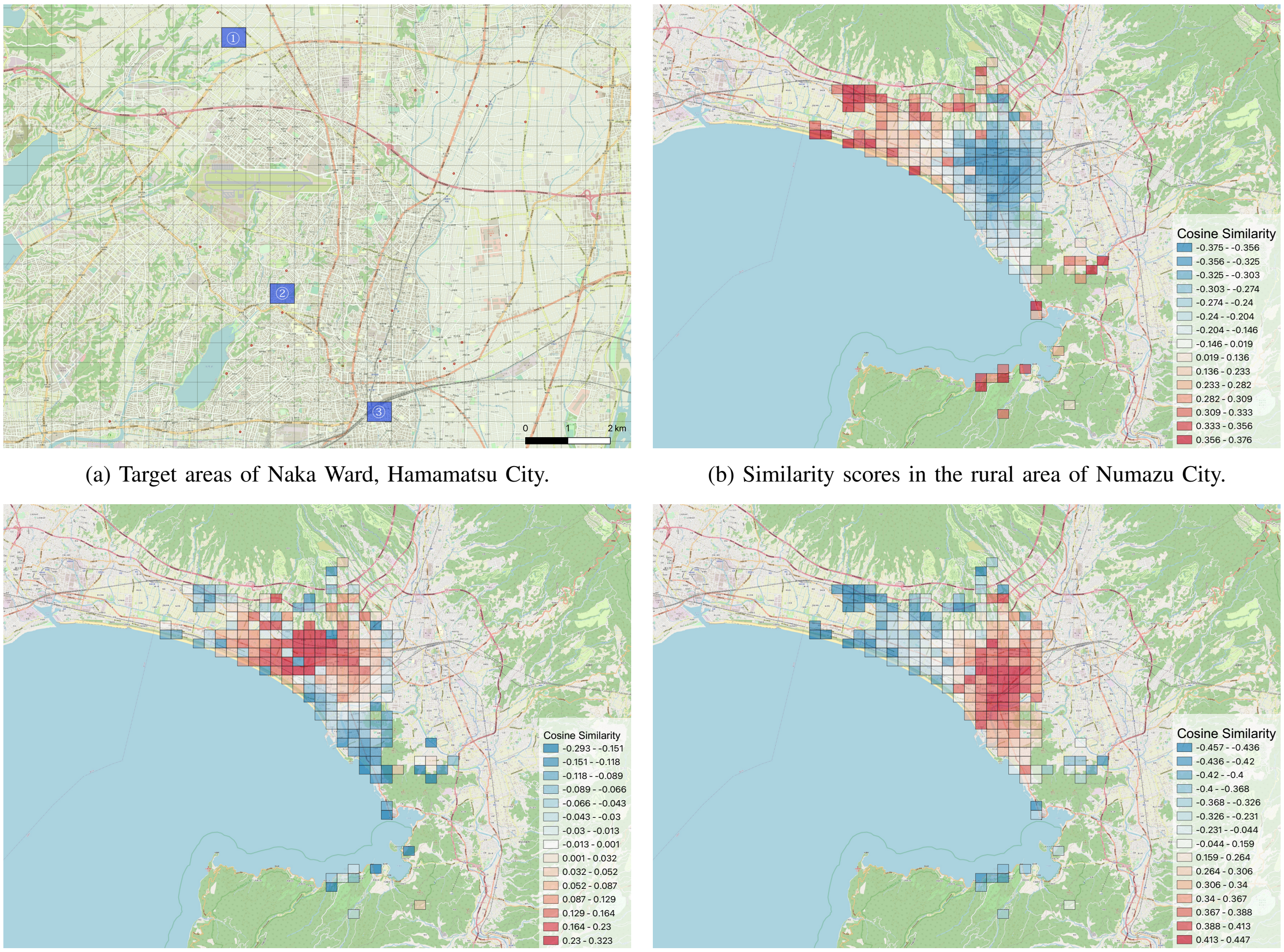
$$\mathbf{COR}(\hat{y_i}, y_i) = \frac{\Sigma xy - \Sigma x \Sigma y}{\sqrt{[\Sigma x^2 - (\Sigma x)^2][\Sigma y^2 - (\Sigma y)^2]}} \tag{9}$$

RMSE and MAE emphasize the individual mesh grid prediction, whereas COR focuses on the overall prediction. We depicted the performance more thoroughly using these three metrics.

E. Result and Discussion

1) OD Volume Prediction: We conducted experiments using all baseline models and the proposed model. Table I exhibits the results. Regarding all the metrics mentioned in Section IV-D, our proposed model achieved the best performance in almost all areas.

Traditional models, such as the gravity model, only consider simple socioeconomic factors; thus, they are insensitive to multimodal indicators and cannot simulate complicated interactions between different OD pairs. Tree-based models can predict accurate results in one-feature-to-one-target problems while performing worse in problems like OD volume predictions. This is because such a problem is a pair-feature-to-one problem, as one embedding needs to meet all related OD pairs to provide an accurate prediction of OD volume, which is much more difficult than the one-feature-to-one-target one. GraphSAGE treats all neighbors the same; the model may overlook some more essential neighbors but underscore meaningless ones, rendering the prediction worse.



(a) Target areas of Naka Ward, Hamamatsu City.

(b) Similarity scores in the rural area of Numazu City.

(c) Similarity scores in the residence area of Numazu City.

(d) Similarity scores in the station area of Numazu City.

Fig. 4: Mesh grid similarity analysis between Naka Ward, Hamamatsu City and Numazu City. We selected three typical mesh grids of the rural (mesh grid one), residence (mesh grid two), and station area (mesh grid three) in Naka Ward. Then, we calculated the cosine similarity of all grids in Numazu City. The similarity analysis shows that the generated embedding can illustrate similarity between areas with similar functions and attributes.

2) Grid Similarity and Urban Utility Classification: Figure 4 shows the result between the target grid in Fuji City and all grids in Naka Ward, Hamamatsu City. The target grid was one grid near the largest train station in Fuji City. We can see that grids near the train station show a higher similarity from the heat map of Naka Ward in Figure 4, whereas grids in the outskirts of the city show extremely low similarity.

Figure 4 shows the result between the target grids in Naka Ward, Hamamatsu City, and all meshes in Numazu City. We selected three target mesh grids in numerous types for analysis. First, target mesh grid one was in the local area of Naka Ward. In this area, there were nearly no tenant buildings, only several factories. In the heat map of similarity in Numazu city in Figure 4b, we can see those mesh grids on the border of the city illustrate high similarity compared with the downtown area. The most similar grids were near the mountain areas without urban facilities, the same as Naka Ward. Second, mesh grid two was in the residence area. This grid was in the center of a residential block, and the road network was not complicated around the mesh. The tenant areas in Numazu City indicate the highest similarity from the result in Figure 4c. Furthermore, similar mesh grids are also more scattered, as residence areas tend to disperse around the city. Third, mesh grid three was a grid near the largest train station, Hamamatsu station, in Naka Ward. We can see that grids near the train station show a higher similarity from the heat map of Numazu City in Figure 4d, whereas grids in the outskirts of the city show extremely low similarity.

Note that this similarity analysis is a natural comparison of generated embeddings. Further mining of the proposed embeddings dealing with specific functions can be performed using postprocessing.

Figure 5 shows the result of the land utility classification in Aoi Ward, Shizuoka City. We applied the unsupervised k-

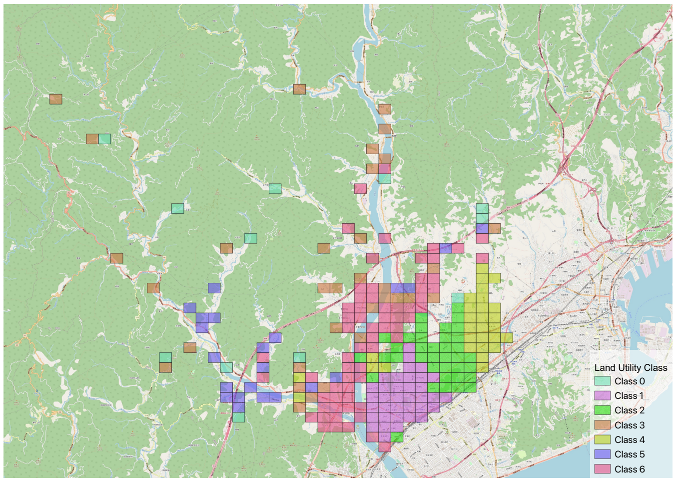


Fig. 5: Land utility classification using unsupervised clustering in Aoi Ward, Shizuoka City. The classification is extremely accurate for the differentiation between rural and downtown areas.

means clustering method to the generated embeddings of Aoi Ward. Specifically, we chose the number of classes as seven. The result implies that the generated embeddings naturally own the semantics for the land utility. The classification is accurate for the differentiation between rural and downtown areas. However, it cannot perform well in the central city, where the land utility is complicated. Note that we apply the naïve unsupervised clustering approach to uncover the mesh grid representations. The application of other supervised methods to our proposed embeddings can give better predictions in populated areas.

3) Indicator Importance Analysis: Understanding which indicator plays a more important role is necessary because the model utilizes multimodal urban indicators. Therefore, we conducted an indicator importance analysis in this study. Specifically, we measured the gradient change of all indicators, given the same initial value. The greater the gradient changes concerning the initial unit value, the more sensitive the model is to the urban indicator. We used the integrated gradients algorithm to calculate the significance of each urban indicator. The algorithm used sensitivity and implementation invariance as two fundamental axioms to approximate the feature attribution [14]. We conducted the analysis using the Captum library [15]. The result can be seen in Figure 6. Note that we normalized the indicator importance between zero and one to better display the result.

The road density in the mesh grid dominated the feature contribution. Such outcomes were reasonable, as the traffic conditions influence the OD flow directly. The vehicle is one of the most essential means of transportation in areas except super cities like Tokyo in Japan. Additionally, roads with widths between 3 to $13\,m$ illustrated considerable importance. Such road networks are the most common urban arterial roads, which assume the responsibility for the large amounts of the city traffic volumes. The wider roads, such as highways, and the narrower roads, such as branch lanes, are of relatively less importance but are still exigent.

In both six cities, the night population contributed to the OD volume calculation, which met the common knowledge. The populated areas should have greater traffic flows because the people are the source of the OD flow. Nonetheless, the POI distribution signal indicated a relatively small attribution to the proposed result. The reason may be that the classification of municipal facilities cannot differentiate the purpose of travel. The number of POI can influence the destination choice of people only in several downtown areas.

The signal of railway users had a different effect in each region. We can determine that the railway indicator performs clearly in Aoi Ward, Numazu City, and Fuji City, whereas the impact can be neglected in other areas. Most railway users are from passing traffic instead of the intra-city flow. One instance is Naka Ward, as Hamamatsu station is a middle station in Tokaido-Sanyo Shinkansen, one of the busiest Japanese high-speed rail lines. Even though it has many railway users, the traffic has nothing to do with the OD flow within the region. The neglectable effect in local areas like Susono City can be explained by the primitive local railway system.

4) Discussion: To a certain extent, the graph convolution model can elaborate on the interdependence between multimodal features and intrigue latent correlation to generate reasonable embeddings. We introduced the notion of grid representation and used a graph structure to capture spatial patterns. Large amounts of temporal data will not suffice to highlight the key attributes of the OD flows. Furthermore, we developed an urban analysis and studied the latent characteristic of the cities. It is also imperative to distinguish between entities with different identities.

V. Related Work

This section presents a literature review of OD matrix prediction and the corresponding techniques. The OD matrix is a significant indicator of human mobility conditions, which grasps the overall trend of people flow. Addressing a complicated urban system with several unpredictable variables is challenging. This section describes the state-of-the-art methods that utilize a deep learning model to predict the OD matrix.

A. OD Flow Prediction

A city is a complicated and dynamic system that is difficult to simulate. The OD flow matrix is a key signal to illustrate city-wide people movement, while predicting human mobility is a high-dimensional and multimodal problem [16]. OD volume prediction over a large spatial scale is difficult owing to the complex mutual relationships because the number of possible OD pairs is quadratic to the number of grids.

General OD matrices contain volumes from multiple means of transportation, which makes them considerably more challenging to predict. Previous studies have often used taxi trajectories [17] or subway user history datasets [18] to construct OD tensors. However, the coverage of such data is limited, and

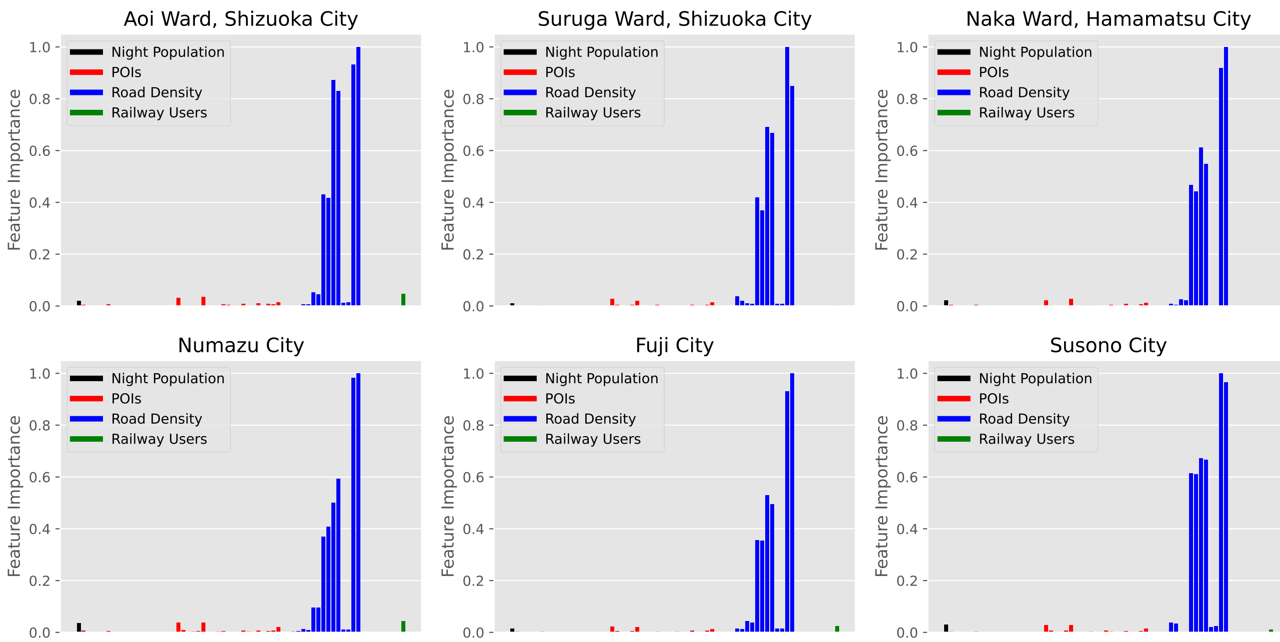


Fig. 6: Analysis of importance for urban indicators in six target areas. The road density in the mesh grid dominated the feature contribution, whereas other signals had different attributions to the generated OD flow volumes.

sparsity may render them useless. Additionally, the dimension of the OD matrix, which is the total number of grids in the analysis, is not large. Thus, this method may not be suitable for large-scale predictions.

Furthermore, many prior studies have focused on spatiotemporal OD matrix predictions [19] [20]. However, we consider all citywide trips rather than those only from taxis or subways, which include multiple transportation modes. The problem setting is similar to that in [21]. This type of data is easier to access and does not cause privacy concerns. Nonetheless, we focus on the overall potential features of movement patterns. [21] implemented multitask learning for commuting flows. They considered the volume of the OD pair and the total inflow and outflow for each node. Moreover, the study used two GAT layers to consider inflow and outflow separately. However, such a division may neglect the interaction between the origin and destination semantics.

B. Location Representation Learning

Location embedding is a popular topic in the academic domain. The goal is to encode locations such that the similarity measured in the embedding space approximates that in the original world [22]. Specifically, we can use dense vectors in the low dimension to describe places [23]. On the one hand, distributed representation resolves the sparse matrix problems that lead to the expensive computation of features. On the other hand, location representation learning alleviates the need for feature engineering every time. With embedding learning, there is no need to perform labor-intensive feature engineering for every application because the model automatically learns the feature distribution.

Location-representation learning is important and efficient for urban computing-related problems. The prediction heads differ from each other as an urban system has many different components. If the location representations, which contain considerable information from different data, are prepared, the task can utilize such pre-trained embeddings and place them into diversified downstream tasks to make improved predictions.

C. Graph Convolutional Network

Graph convolutional network (GCN) is a powerful technique to handle diversified data structures with an arbitrary shape other than sequences and grids. It can capture the explicit or implicit topographical information through the message passing procedure, in which the central node will aggregate information from its neighbor nodes. Some networks, such as GCN, are spectral-based. They operate on the whole adjacent matrix. In contrast, networks like GAT [24] and GraphSAGE [25] are spatial-based, which can form computation graphs without considering the whole structure all the time. Furthermore, networks like GAT have a fixed number of parameters, irrespective of the graph size. Thus, it is suitable for graphs of different sizes.

The neighborhood aggregation function, which is a shared edge-wise mechanism and is independent of the global graph adjacency matrix, is to be learned by the network. It enables inductive learning, i.e., the network can easily consider unseen nodes or graphs and can generate new embeddings on the fly for generalization, which is broadly applicable. The attention mechanism has been at the leading edge for sequence-based deep learning tasks. It enables the network to focus on the most relevant section of inputs to make the decision. It can also benefit graph-structured data through the usage of GAT. The model can take advantage of relational structures and handle dynamic node features in GAT as the central node attends to each neighbor node in the network.

Recent studies have shown that utilizing graph learning for data with spatial and temporal attributes is possible. Human mobility maintains the graph-related attributes. Consequently, the graph neural network (GNN) is an efficient method for unraveling complicated graph structures. GNN-based approaches are commonly used in traffic speed forecasting because road links and traffic monitors can be seen as edges and nodes in the graph. Additionally, several studies consider estimated time of arrival. They view road segments as nodes and the connectivity between road segments as edges.

VI. Conclusion

We derived a spatially weighted GAT for OD volume prediction. The results illustrate that our proposed model can capture the major patterns of daily OD flow, and the results can be utilized in other downstream tasks. Complicated feature distributions can be extracted by explicitly modeling the relationship.

The government can easily apply our proposed model in other cities, as our model owns the inductive characteristics. All the necessary urban indicators are from open-source data, and the generated OD volumes are valuable. Furthermore, the by-product grid embeddings contain considerable meaningful information, which is suitable for diversified urban downstream tasks. Such usage can avoid multiple inclusion of the same data in the urban analysis, which is advantageous for efficiency and limited budgets.

In future studies, we intend to improve the inductive capacity of the model to enable few-shot learning in other cities. We will also measure other factors that affect human mobility and clarify the heterogeneous mutual effects.

Acknowledgment

We sincerely thank the SoftBank Corporation for providing the OD volume data used in this study.

References

[1] F. Simini, M. C. González, A. Maritan, and A.-L. Barabási, "A universal model for mobility and migration patterns," *Nature*, vol. 484, no. 7392, pp. 96–100, 2012.

[2] T. Kashiyama, Y. Pang, and Y. Sekimoto, "Open pflow: Creation and evaluation of an open dataset for typical people mass movement in urban areas," *Transportation research part C: emerging technologies*, vol. 85, pp. 249–267, 2017.

[3] Y. Pang, K. Tsubouchi, T. Yabe, and Y. Sekimoto, "Intercity simulation of human mobility at rare events via reinforcement learning," in *Proceedings of the 28th International Conference on Advances in Geographic Information Systems*, 2020, pp. 293–302.

[4] Z. Yao, Y. Fu, B. Liu, W. Hu, and H. Xiong, "Representing urban functions through zone embedding with human mobility patterns," in *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence (IJCAI-18)*, 2018.

[5] Ministry of Internal Affairs and Communications of Japan, "Characteristics and history of regional mesh statistics. (in Japanese)," 1996-2022, accessed: 2022-08-09. [Online]. Available: https://www.stat.go.jp/data/mesh/pdf/gaiyo1.pdf

[6] S. Brody, U. Alon, and E. Yahav, "How attentive are graph attention networks?" in *International Conference on Learning Representations*, 2022. [Online]. Available: https://openreview.net/forum?id=F72ximsx7C1

[7] Y. Yang, K. Zha, Y. Chen, H. Wang, and D. Katabi, "Delving into deep imbalanced regression," in *International Conference on Machine Learning*. PMLR, 2021, pp. 11 842–11 851.

[8] MLIT of Japan, "Road density statistics for regional meshes. (in Japanese)," 1978-2022, accessed: 2022-08-09. [Online]. Available: https://nlftp.mlit.go.jp/ksj/gml/datalist/KsjTmplt-N04.html

[9] National Statistics Center of Japan, "Categorical industries statistics for regional meshes. (in Japanese)," 2015-2022, accessed: 2022-08-09. [Online]. Available: https://www.e-stat.go.jp/gis/statmap-search?page=1&type=1&toukeiCode=00200553&toukeiYear=2016&aggregateUnit=H&serveyId=H002005112016&statsId=T000918

[10] MLIT of Japan, "Night population statistics for regional meshes. (in Japanese)," 2018-2022, accessed: 2022-08-09. [Online]. Available: https://nlftp.mlit.go.jp/ksj/gml/datalist/KsjTmplt-mesh500h30.html#prefecture22

[11] MLIT of Japan, "Railway statistics for regional meshes. (in Japanese)," 2018-2022, accessed: 2022-08-09. [Online]. Available: https://nlftp.mlit.go.jp/ksj/gml/datalist/KsjTmplt-S12-v2_7.html

[12] M. Fey and J. E. Lenssen, "Fast graph representation learning with PyTorch Geometric," in *ICLR Workshop on Representation Learning on Graphs and Manifolds*, 2019.

[13] J. E. Anderson, "The gravity model," *Annu. Rev. Econ.*, vol. 3, no. 1, pp. 133–160, 2011.

[14] M. Sundararajan, A. Taly, and Q. Yan, "Axiomatic attribution for deep networks," in *International conference on machine learning*. PMLR, 2017, pp. 3319–3328.

[15] N. Kokhlikyan, V. Miglani, M. Martin, E. Wang, B. Alsallakh, J. Reynolds, A. Melnikov, N. Kliushkina, C. Araya, S. Yan *et al.*, "Captum: A unified and generic model interpretability library for pytorch," *arXiv preprint arXiv:2009.07896*, 2020.

[16] M. S. Iqbal, C. F. Choudhury, P. Wang, and M. C. González, "Development of origin–destination matrices using mobile phone call data," *Transportation Research Part C: Emerging Technologies*, vol. 40, pp. 63–74, 2014.

[17] Y. Wang, H. Yin, H. Chen, T. Wo, J. Xu, and K. Zheng, "Origin-destination matrix prediction via graph convolution: a new perspective of passenger demand modeling," in *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, 2019, pp. 1227–1235.

[18] W. Jiang, Z. Ma, and H. N. Koutsopoulos, "Deep learning for short-term origin–destination passenger flow prediction under partial observability in urban railway systems," *Neural Computing and Applications*, vol. 34, no. 6, pp. 4813–4830, 2022.

[19] R. Zhang, L. Han, B. Liu, J. Zeng, and L. Sun, "Dynamic graph learning based on hierarchical memory for origin-destination demand prediction," 07 2022, pp. 2358–2364.

[20] H. Shi, Q. Yao, Q. Guo, Y. Li, L. Zhang, J. Ye, Y. Li, and Y. Liu, "Predicting origin-destination flow via multi-perspective graph convolutional network," in *2020 IEEE 36th International Conference on Data Engineering (ICDE)*. IEEE, 2020, pp. 1818–1821.

[21] Z. Liu, F. Miranda, W. Xiong, J. Yang, Q. Wang, and C. Silva, "Learning geo-contextual embeddings for commuting flow prediction," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 34, no. 01, 2020, pp. 808–816.

[22] G. Mai, K. Janowicz, Y. Hu, S. Gao, B. Yan, R. Zhu, L. Cai, and N. Lao, "A review of location encoding for geoai: methods and applications," *International Journal of Geographical Information Science*, vol. 36, no. 4, pp. 639–673, 2022.

[23] A. Dadoun, R. Troncy, O. Ratier, and R. Petitti, "Location embeddings for next trip recommendation," in *Companion Proceedings of The 2019 World Wide Web Conference*, 2019, pp. 896–903.

[24] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio, "Graph attention networks," in *International Conference on Learning Representations*, 2018. [Online]. Available: https://openreview.net/forum?id=rJXMpikCZ

[25] W. L. Hamilton, R. Ying, and J. Leskovec, "Inductive representation learning on large graphs," ser. NIPS'17. Red Hook, NY, USA: Curran Associates Inc., 2017, p. 1025–1035.