Modeling Heterogeneous Relations across Multiple Modes for Potential Crowd Flow Prediction

Qiang Zhou[1], Jingjing Gu[1,*], Xinjiang Lu[2], Fuzhen Zhuang[3,4], Yanchao Zhao[1], Qiuhong Wang[1], Xiao Zhang[5]

[1]Nanjing University of Aeronautics and Astronautics,
[2]Business Intelligence Lab, Baidu Research,
[3]Key Lab of Intelligent Information Processing of Chinese Academy of Sciences (CAS), ICT, Beijing, China,
[4]University of Chinese Academy of Sciences, Beijing, China,
[5]Shandong University
{zhouqnuaacs, gujingjing}@nuaa.edu.cn, luxinjiang@baidu.com, zhuangfuzhen@ict.ac.cn,
{yczhao, baoh9491}@nuaa.edu.cn, xiaozhang@sdu.edu.cn

Abstract

Potential crowd flow prediction for new planned transportationsites is a fundamental task for urban planners and administrators. Intuitively, the potential crowd flow of the new comingsite can be implied by exploring the nearby sites. However, the transportation modes of nearby sites (e.g. bus stations,bicycle stations) might be different from the target site (e.g. subway station), which results in severe data scarcity issues. To this end, we propose a data-driven approach, named MOHER, to predict the potential crowd flow in a certain mode for a new planned site. Specifically, we first identify the neighbor regions of the target site by examining the geographical proximity as well as the urban function similarity. Then, to aggregate these heterogeneous relations, we devise a cross-mode relational GCN, a novel relation-specific transformation model, which can learn not only the correlations but also the differences between different transportation modes. Afterward, we design an aggregator for inductive potential flow representation. Finally, an LTSM module is used for sequential flow prediction. Extensive experiments on real-world data sets demonstrate the superiority of the MOHER framework comparedwith the state-of-the-art algorithms.

Introduction

Public transportation site selection acts as a pivotal part in urban planning. For example, is it appropriate to build a subway station in the target site/region when promoting urban infrastructure construction. To answer this question, existing methods, like (Lin, Huang, and Xu 2020; Jelokhani-Niaraki and Malczewski 2015), mainly focus on how to balance the trade-offs between efficiency and cost, only a few studies (Chen et al. 2015; Gong et al. 2020) focus on site selection through predicting future demands. How to fulfil the traffic demands in the future is the key to assess the quality of a site selection, especially for a transportation site. Therefore, potential crowd flow prediction of new planned sites is one of the most fundamental and important tasks for the city planners and administrators.
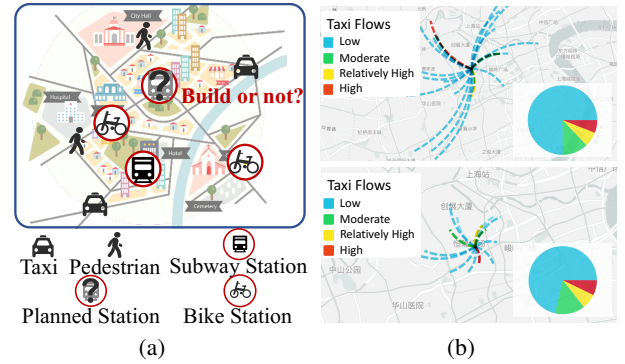
Figure 1: (a): An example of the subway station placement problem. (b): From the same origin, the outgoing crowd flows in different modes are different and similar: 1) different travel distances and destinations; and 2) similar traffic volume distributions depicted in the pie charts.

In general, the potential crowd flow of a new planned site can be estimated by examining nearby sites in a collective way. For example, (Gong et al. 2020) studies the potential flow prediction of a new subway station with neighbor subway stations. However, as the transportation sites in the same mode often locate far from each other, the transportation modes of nearby sites might be different from the planned target site as illustrated in Fig. 1(a), which leads to severe data scarcity issues.

Recently, advanced matrix/tensor completion approaches, e.g. (Babu, Sure, and Bhuma 2020; Mutinda et al. 2019; Takeuchi, Kashima, and Ueda 2017), have been proposed to solve the Kriging problems of unrecorded samples. Some researchers (Li et al. 2017; Xie et al. 2019; Chen, Yang, and Sun 2020) have also put many efforts to solve the data sparsity problem in traffic completion and prediction. However, these methods are essentially transductive and could hardly be applied for new sites/nodes directly using trained models. Besides, it is not worthy to retrain the model from scratch only for a single site.

In this paper, we study the problem of potential crowd flow prediction for new planned transportation sites. Specifically, we propose a data-driven approach to MOdeling HEterogeneous Relations across multiple transportation modes for potential crowd flow prediction, named MOHER. We detail the MOHER framework in the following three aspects.

First, the relations between nearby sites in different transportation modes are highly dependent, dynamic and heterogeneous. As graph convolutional network (GCN) has shown its capability in characterizing complex dynamic spatio-temporal dependencies and the generalization potential to unseen nodes (Hamilton, Ying, and Leskovec 2017; Zhang and Chen 2019; Yuan et al. 2020), we devise a cross-mode message passing network to model the cross-mode heterogeneous relations. We then leverage a neighbor aggregator to learn the inductive representations for each new site and further predict the potential crowd flows instead of retraining at every turn.

Second, the diversion phenomenon cannot be neglected when predicting the future crowd flow for a new planned site through the nearby sites. For instance, the new subway station will absorb a bulk of passengers who used to take buses. Besides, the new subway station will definitely attract more urban functions, which will compensate for the crowd flows of buses and bicycles. Along this line, we construct localized graphs centered by the target site/region to measure the geographical proximity and the functional similarity (characterized using surrounding Points Of Interest (POIs) for each category) simultaneously. Such that, the flow diversion and potential crowd flows can be captured automatically.

Third, apart from the diversion effect on the neighbor sites/regions, the planned new site will also have the diversion effect on other traffic modes (Liu et al. 2020), where similarities and differences co-exist as illustrated in Fig. 1(b). Intuitively, contiguous sites should share similar travel patterns in terms of time and volume. On the other hand, due to the intrinsic characteristics (e.g. price and travel distance) of different transportation modes, the difference between crowd flows in different modes could be very large. Thus, we introduce a novel relation-specific transformation model that can extract the similarity and difference of heterogeneous relations simultaneously.

To sum up, our contributions in this paper are as follows:

- We propose the inductive potential crowd flow prediction framework MOHER, which can naturally generalize to future new planning sites after training with only the historical flow data of the existing sites.

- To capture the flow diversion and potential crowd flows of a new site, we respectively encode the geographical proximity and functional similarity among the nearby sites in different transportation modes.

- We deeply explore the correlations and differences between cross-mode crowd flows and propose the Cross-Mode Relational GCN (CMR-GCN) to explicitly model each type of relations.

- We conduct extensive experiments on real-world datasets, which demonstrate the effectiveness of MOHER for potential crowd flow prediction.

Notations	Description
$p \in \mathcal{P}$	The transportation mode and its set.
$v_p \in \mathcal{V}_p$	The site/region belonging to the mode p and its set.
p_0, v_0	The target transportation mode, the newly planned target site/region.
$v^{\mathrm{Re}}, v^{\mathrm{Un}}$	The region with/without historical crowd flow records.
$\mathcal{X}_p^{(t)} = \{x_{v_p}^{(t)}\}$	The historical flow data of region v_p at time t.
$r_{\mathrm{GEO}}, r_{\mathrm{POI}}$	The cross-mode edge type computed by the Geo-proximity/POI-similarity.
$\epsilon: (v_i, r, v_j)$	The edge attribute between node (region/site) i & j with edge type r. r can be r_{GEO} or r_{POI}.

Table 1: Notations of symbols used in this work.

Preliminaries

We first introduce some important definitions and then formalize our problem. For brevity, we show notations and corresponding descriptions in Table 1.

Definition 1 *Cross-mode Flow Relation Graph (CFRG).* *A cross-mode flow relation graph centered by v_0 at time t is a heterogeneous undirected weighted graph $G_{v_0}^{(t)} = (\mathcal{V}, \varepsilon, \mathcal{R})$, where $\mathcal{V} = \{\mathcal{V}_p, p \in \mathcal{P}\}$ is a set of heterogeneous nodes (regions/sites) from all available transportation modes, ε is the set of edge attributes of multiple relations, $\mathcal{R} = \{r_{\mathrm{GEO}}\} \cup \{r_{\mathrm{POI}}\}$ is the set of heterogeneous edge types. Hence, there may be more than one relations between two cross-mode nodes.*

Definition 2 *Flow Prediction of New Planned Sites. Assuming our new planned target region/site is v_0 with no data records belonging to the crowd flow p_0, we aimed at predicting the potential future flows $x_{v_0}^{(t+1)}$ of the target v_0 with p_0. Therefore, the crowd flow prediction of new planned sites is formulated as a spatio-temporal prediction given a fixed temporal length t':*

$$\{(\mathcal{X}, G_{v_0})^{(t-t'+1)}, (\mathcal{X}, G_{v_0})^{(t-t'+2)}, \\ \cdots, (\mathcal{X}, G_{v_0})^{(t)}\} \xrightarrow{f(\cdot)} x_{v_0}^{(t+1)}, \quad (1)$$

where $\mathcal{X} = \{\mathcal{X}_p, p \in \mathcal{P}\}$ is the set of all cross-mode historical flow data. Eq. 1 is to learn a function $f(\cdot)$ that maps the t' historical target-related cross-mode flow features to the target region flows in the next timestamp. To the best of our knowledge, we provide the first attempt on inductive potential crowd flow prediction of new planned sites with cross-mode flow data.

Methodology

In this section, we introduce our framework for inductive potential crowd flow prediction of new planned sites/regions by modeling heterogeneous relations across multiple transportation modes. The overview of MOHER framework is
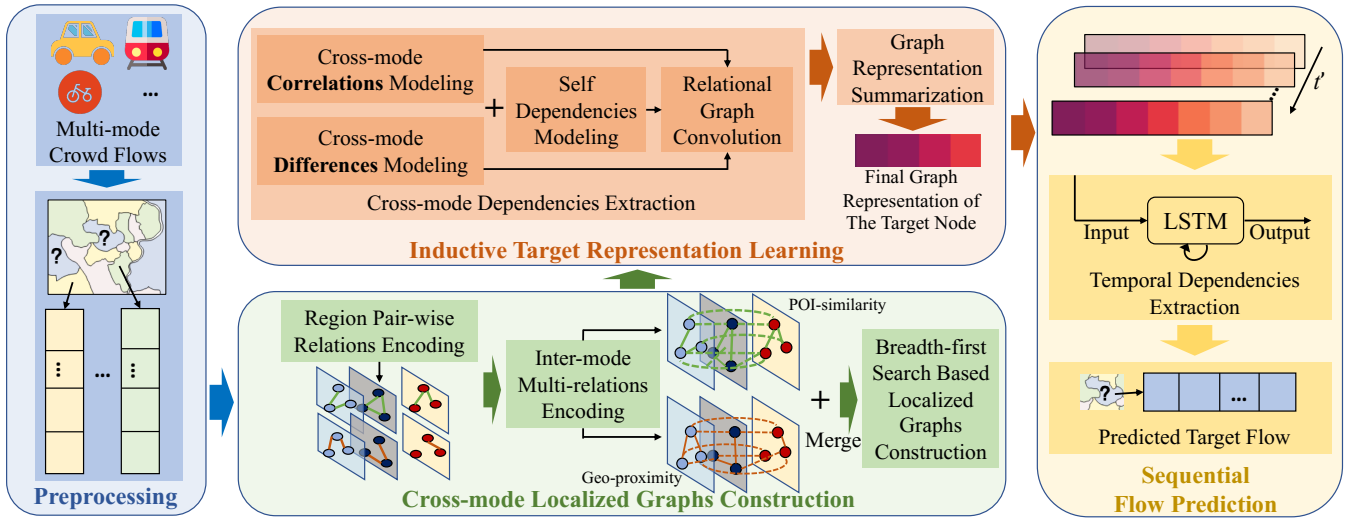
Figure 2: An overview of the framework.

shown in Fig. 2, which consists of four major parts. (1) Pre-processing: collect a large number of multi-mode flow records and construct the input $\mathcal{X}$. (2) Cross-mode Localized Graphs Construction: explore the potential crowd flows from nearby sites/regions via encoding both intra- and inter-mode relations according to the Geo-proximity and the POI-similarity, and construct localized graphs with the target node as the input. (3) Inductive Target Representation Learning: develop the cross-mode relational GCN for extracting the correlations and differences simultaneously between modes, then learn the inductive target node representation. (4) Sequential Flow Prediction: build an LSTM module to conduct sequential flow prediction of the target site/region.

Cross-mode Localized Graphs Construction

Multiple pair-wise relations encoding Potential crowd flows are hidden in neighbor sites/regions of multiple transportation modes. However, aggregating inadequate information with only geographical distance could hardly characterize the flow features of new sites. The new construction of transportation sites can inspire the local functionality, which may cause incremental crowd flows of the new urban functional zone. Therefore, we encode two types of relations among sites/regions:

- **Geographical proximity.** Two sites/regions of multiple transportation modes near each other are affected by the same varying of travel demands. We choose geographic centers or station coordinates as the representatives of regions or sites, and then treat the normalized distance $\widetilde{\text{dis}}_{ij}$ between arbitrary two regions i and j as inputs. The Geo-proximity is encoded as Eq. 2.

$$\epsilon_{ij,r_{\text{GEO}}} = \begin{cases} \exp(-\widetilde{\text{dis}}_{ij}^2), & \text{dis}_{ij} \leq \gamma \\ 0, & \text{dis}_{ij} > \gamma \end{cases} \quad (2)$$

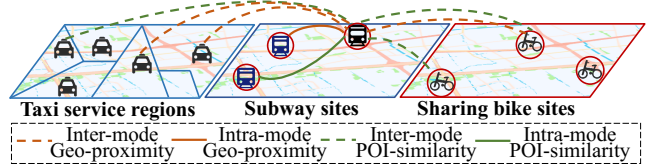in which γ is a threshold for tuning the sparsity of $\epsilon_{r_{\text{GEO}}}$.

Figure 3: The diagrammatic sketch of establishing multiple heterogeneous relations for one region.

- **POI similarity.** Since POIs can indicate the region functionality, it is well instructive to refer to other regions with similar POI distribution for capturing the potential incremental crowd flows of new planned sites. In order to deal with the varying shapes and sizes of cross-mode nodes, we normalize the POI vectors F of nodes and calculate the cosine distance between them using Eq. 3.

$$\epsilon_{ij,r_{\text{POI}}} = \begin{cases} \mathcal{N}(\text{Cos}(\widetilde{F}_i, \widetilde{F}_j)), & \text{Cos}(\widetilde{F}_i, \widetilde{F}_j) \geq \beta \\ 0, & \text{Cos}(\widetilde{F}_i, \widetilde{F}_j) < \beta \end{cases} \quad (3)$$

in which β is a threshold for controlling the sparsity of $\epsilon_{r_{\text{POI}}}$. After that, the results greater than zero are renormalized by $\mathcal{N}(\cdot)$ for more uniform values.

Inter-mode relations encoding Potential crowd flows of new planned sites are not only involved in the target transportation mode, but we need to explore crowd flows from related sites/regions of multiple modes comprehensively. Therefore, we also encode the inter-mode relations by resorting to both **Geo-proximity** and **POI-similarity**.

An example is shown in Fig. 3. We establish multiple relations for one site/region with both intra-mode (e.g. subway stations) and inter-mode (e.g. bike stations/taxi service regions) sites/regions. Due to the diversities of cross-mode flow patterns, we consider each different relation across multiple transportation modes is heterogeneous.

Algorithm 1 Breadth-first search based localized graph construction for the target node

---
1: **Input:** M, the target node v_0, the historical recorded nodes v^{Re}, the timeslot t

2: **Output:** the $(M+1)$-nodes localized graph $G_{v_0}^{(t)}$

3: **Init** size = 0, seq = $\{v_0\}$;

4: **while** size$< M+1$ and seq!=$\{\}$ **do**

5:     v_i = seq.dequeue();

6:     add v_i into $G_{v_0}^{(t)}$, add relations of v_i into $G_{v_0}^{(t)}$;

7:     size = size + 1;

8:     **for** v_j in v^{Re} **do**

9:        calculate $\epsilon_{ij,r_{\text{GEO}}}$ with Eq.2;

10:     **end for**

11:     enqueue v_j into seq by the descending order of $\epsilon_{ij,r_{\text{GEO}}}$ with $\epsilon_{ij,r_{\text{GEO}}} > 0, v_j \notin G_{v_0}^{(t)} \cup$ seq;

12:     **for** v_j in v^{Re} **do**

13:        calculate $\epsilon_{ij,r_{\text{POI}}}$ with Eq.3;

14:     **end for**

15:     enqueue v_j into seq by the descending order of $\epsilon_{ij,r_{\text{POI}}}$ with $\epsilon_{ij,r_{\text{POI}}} > 0, v_j \notin G_{v_0}^{(t)} \cup$ seq;

16:**end while**

17:**return** $G_{v_0}^{(t)}$;

Generally, we can model more types of relations with an extension of the above method. In this case, assuming we model ρ ($\rho = 2$ in this work) types of relations among $|\mathcal{P}|$ transportation modes, there will be total $\frac{|\mathcal{P}| \times (|\mathcal{P}|+1) \times \rho}{2}$ types of heterogeneous edges. For cross-mode relations, the $\epsilon_{r_{\text{GEO}}}$ and $\epsilon_{r_{\text{POI}}}$ are larger as the overlapping area is larger, so that MOHER can deal with the complex compositions of multiple transportation modes like Fig. 1(a).

Breadth-first search based localized graphs construction
Due to the continuous developments of urban infrastructure (e.g., new planned sites), the urban regional composition and functionality is evolving. Therefore, a static graph could hardly meet the demand of potential crowd flow prediction of new planned sites. In this paper, we extract M neighbor nodes of the target region v_0 to construct localized graphs $\{G_{v_0}^{(t-t'+1)}, \cdots, G_{v_0}^{(t)}\}$ given timeslots $\{t - t' + 1, \cdots, t\}$ via breadth-first search. Particularly, the hyperparameter M should be approximately equal to the size of the first and second-order neighborhood of target regions for involving most of the related information (Hamilton, Ying, and Leskovec 2017). As a result, our method can focus on the localized information and apply to the evolving graph structure for inductive regional flow prediction.

Inductive Target Representation Learning

Cross-mode dependencies extraction Given heterogeneous graphs constructed by Algorithm 1, we should explicitly model each kind of the heterogeneous cross-mode relations to extract flow embeddings and then map the localized graphs information into the potential flow patterns of the new planned site. Accordingly, we propose the Cross-Mode Relational Graph Convolutional Network (CMR-GCN) as the message passing module for modeling **correlations** and **differences** of heterogeneous relations simultaneously.

Specifically, for each node $v_i, 1 \leq i \leq M$ in a localized graph except the planned target node, we extract the regional **correlations** $\{\mathcal{C}_{i,r}^{l+1} | r \in \mathcal{R}\}$ at layer $l+1$ of each kind of heterogeneous relations via (4):

$$\mathcal{C}_{i,r}^{l+1} = \sum_{j \in \mathcal{N}_r(i)} \sigma\left(\frac{\epsilon_{r,ij}}{|\mathcal{N}_r(i)|} \boldsymbol{x}_j^l \boldsymbol{W}_{r,c}^l + \boldsymbol{b}_{r,c}^l\right), \quad (4)$$

where $\boldsymbol{x}_j^l$ denotes the features of node $j, 1 \leq j \leq M$ at layer $l, \{\boldsymbol{W}_{r,c}^l | r \in \mathcal{R}\}$ and $\{\boldsymbol{b}_{r,c}^l | r \in \mathcal{R}\}$ are learnable parameters, $\mathcal{N}_r(i)$ denotes the set of i's neighbors that connect to i with relation category r. σ denotes the ReLU activation function.

In addition to the correlations, there are definite differences between cross-mode regional flow features. If the information of differences is ignored, there may even produce a negative transfer effect. Hence, we also model the **differences** $\{\mathcal{D}_{i,r}^{l+1} | r \in \mathcal{R}\}$ between cross-mode node features:

$$\mathcal{D}_{i,r}^{l+1} = \sum_{j \in \mathcal{N}_r(i)} \omega\left(\frac{\epsilon_{r,ij}}{|\mathcal{N}_r(i)|} |\boldsymbol{x}_j^l - \boldsymbol{x}_i^l| \boldsymbol{W}_{r,d}^l + \boldsymbol{b}_{r,d}^l\right), \quad (5)$$

where $|\boldsymbol{x}_j^l - \boldsymbol{x}_i^l|$ represents the natural gap between node i and neighbor node j, and ω is set as $tanh$. An intuitive explanation of Eq. 5 is that we construct a shadow relation r_s for each kind of relation r to indicate the differences between two original nodes connected by r. Consequently, we aggregate the shadow node j_s's information which is D-value between node features into $\mathcal{D}_{i,r}^{l+1}$ via r_s. The correlations and differences of each kind of heterogeneous relations are accumulated on the transformed self-features, then the forward-pass-update formula of $\boldsymbol{x}_i^{l+1}$ is given by:

$$\boldsymbol{x}_i^{l+1} = \sigma\left(\boldsymbol{x}_i^l \boldsymbol{W}_p^l + \sum_{r \in \mathcal{R}} (\mathcal{C}_{i,r}^{l+1} + \mathcal{D}_{i,r}^{l+1})\right). \quad (6)$$

Finally, assuming there are L message passing layers CMR-GCNs, the final node representation $\boldsymbol{g}_i$ of node v_i is concatenated by the outputs of layer 1 to layer L:

$$\boldsymbol{g}_i = \text{concat}(\boldsymbol{x}_i^1, \boldsymbol{x}_i^2, \cdots, \boldsymbol{x}_i^L) \quad (7)$$

Regularization for modeling heterogeneous relations
With both correlations and differences modeling, the total types of heterogeneous relations will be double to $|\mathcal{P}| \times (|\mathcal{P}| + 1) \times \rho$ plus the number of shadow relations $\{r_s | r \in \mathcal{R}\}$, which will lead to a surge in the number of parameters at the same time. To reduce the complexity of our model in case of overfitting, we apply matrix reconstruction to each of the learnable parameters:

$$\boldsymbol{W}_{r,c}^l = \sum_{q=1}^{Q} a_{\boldsymbol{W}_{r,c},q}^l B_{\boldsymbol{W},q}^l, \ \boldsymbol{W}_{r,d}^l = \sum_{q=1}^{Q} a_{\boldsymbol{W}_{r,d},q}^l B_{\boldsymbol{W},q}^l$$

$$\boldsymbol{b}_{r,c}^l = \sum_{q=1}^{Q} a_{\boldsymbol{b}_{r,c},q}^l B_{\boldsymbol{b},q}^l, \ \boldsymbol{b}_{r,d}^l = \sum_{q=1}^{Q} a_{\boldsymbol{b}_{r,d},q}^l B_{\boldsymbol{b},q}^l$$
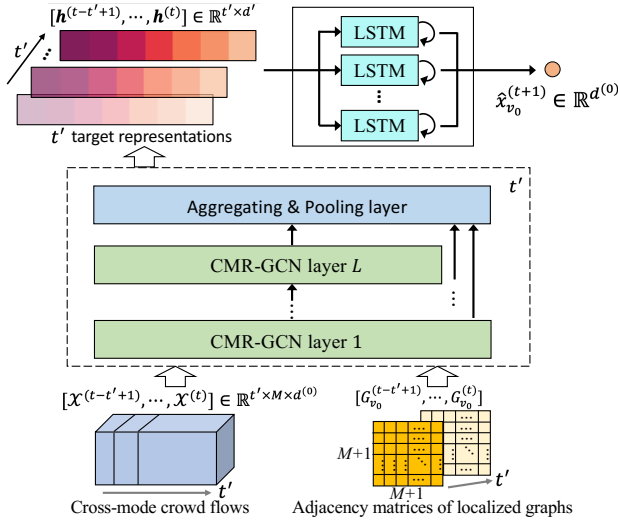
$$(8)$$

Figure 4: The network structure of our MOHER, where $d^{(0)}$ and d' are the dimensions of origin and final flow representations. The shared network across different timeslots in the dotted box is applied to learn inductive cross-mode target representation.

Accordingly, each learnable parameter is reconstructed by basis B and corresponding coefficients a. Along such a reconstruction process, all the parameters that model the correlations and differences share the same basis matrices respectively, which can also mitigate overfitting when some kinds of relations are uncommon in the training set.

Graph representation summarizing After the L layers of regularized CMR-GCNs, each node v_i except the target v_0 gets its flow embedding g_i. Inspired by max-pooling operation, we propose a nearest neighbor pooling aggregator that selects the most related nodes connected to target v_0 by each existing type of relations, which can directly reduce the redundancy of features. Thus, the representation of target v_0 is constructed as:

$$h = \text{concat}(\epsilon_{v_0,r} \cdot g_r | r \in \mathcal{R}_{v_0}), \qquad (9)$$

where $\mathcal{R}_{v_0}$ is a subset of $\mathcal{R}$ which contains the relations connected to v_0, g_r is the nearest neighbor of v_0 connected by relation r, and $\epsilon_{v_0,r}$ is the value of the edge. Here h is the flow representation of the target node v_0. Finally, we adopt LSTM networks to extract temporal dependencies for predicting the $t+1$ target regional or station-level flows. The network structure is shown in Fig. 4. Note that our proposed MOHER is inductive, it can generalize to future planned sites along with the development of the city.

Model Training

In the training phase, our model is optimized by minimizing the loss of mean square error (MSE) using Adam optimizer.

It is difficult to retrieve the historical data before the construction of existed sites. Therefore, we propose a simulation strategy to approximate real historical flows. Specifically, we randomly select existed sites to simulate planned

sites, and we distribute the historical flow of them to the first-order Geo-proximate neighbors according to the product of the flow proportions and the connected edge weights. These selected sites are used as training targets.

Experiments

Experiment Settings

Datasets We construct a benchmark with three real-world flow datasets collected from NYC OpenData. Two of them contain the trip records of the taxi and bike in NYC. The other contains the turnstile usage condition of each subway station in NYC. All of them are generated in the time period from January 1st, 2018 to December 1st, 2019 (700 days).

- **NYC Turnstile Usage of Subway Stations:** NYC Metropolitan Transportation Authority provides the turnstile usage counts of NYC subway stations. There are about 5042 turnstiles used as entry/exit registers at 425 subway stations. The value records per 4 hours consist of following information: station-ID, turnstile-ID, record time, Entries value, Exits value, etc.

- **NYC Citi Bike:** NYC Bike Sharing System generates the Citi Bike orders including 38 million and 100 thousand transaction records of 647 stations in the selected time period. This data set contains following information: bike pick-up station, bike drop-off station, bike pick-up time, bike drop-off time, trip duration.

- **NYC Taxi:** NYC Taxi consists of about 187 million taxicab trip records. On average, there are about 267 thousand records in 262 regions every day. The records consist of following information: pick-up time, drop-off time, pick-up region, drop-off region, trip distance, etc.

Baselines We compare our framework with the following methods:

- **NA-HA:** Neighbor Average and Historical Average (Kamarianakis and Prastacos 2003). We model the spatial similarity and use the average of neighbor sites of the same transportation modes as the historical flows. Then, we predict the target crowd flow using the mean value over time slots.

- **MGCN-MLP:** From Multiple Graph Convolutional Network (Geng et al. 2019), we aggregate information for modeling spatial dependency by multiple graphs. Then, a Multi-Layer Perceptron is used for mapping graph representation to the prediction result.

- **NA-LSTM:** The feature representation of historical flows is similar to NA-HA, then the Long-Short-term-Memory method (Hochreiter and Schmidhuber 1997) is introduced for temporal dependency modeling.

- **LP-GLP:** We use Label Propagation to complete the historical flows and construct a semi-supervised learning problem on graphs. Generalized Label Propagation (Li et al. 2019a) is a semi-supervised learning method that produces flow representations of different times by exploring graph relations. Then, a Feedforward Neural Network (FNN) is used for prediction.

- **STMGCN:** STMGCN leverages complete convolutional structures to model both temporal and multiple spatial dependencies. MGCN is applied in the ST-Conv blocks (Yu, Yin, and Zhu 2018) for modeling multiple relations.

- **MLC-PPF:** The Multi-view Localized Correlation learning model(Gong et al. 2020) for subway Potential Passenger Flow uses the economic statistics as cross-domain knowledge, which just uses one mode flow.

Metrics We evaluate performances based on two popular metrics, which are Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE) (Geng et al. 2019; Yu, Yin, and Zhu 2018; Ye et al. 2019; Deng et al. 2016).

Implementations We aggregate crowd flows into 4 hours windows and apply Z-score normalization. We randomly selected 20% sites for validation and 10% for testing. They are deleted from the existing sites, and we construct the train set based on the remained 70% site with flow data of the previous 60% time (420 days). The validation and test set are constructed based on the respectively selected sites with flow data of the following 40% time (280 days). We tune the hyperparameters using grid search based on the performance on the validation set, and report the average performances on the test set over 10 runs. Due to the unimode flow, NA-HA, LP-GLP, and MLC-PPF use the data just from the target mode. And we use 1 layer of MGCN and 2 layers of MLP with 64 hidden units and a dropout rate of 0.2 for MGCN-MLP. The modules in STMGCN is the same with MGCN-MLP but it is extended with Gated Linear Units of $K_t = t'/2$ and $K_t = t'/2 + 1$. The statistics used in MLC-PPF are the number of various kinds of commercial POIs. Other hyperparameters of baselines are the same as the grid searched hyperparameters of our method. All experiments are run on a PC with an NVIDIA GeForce RTX 2080Ti GPU, and 64 GB RAM running the 64-bit Ubuntu 16.04 system, python3.6, TensorFlow 2.0.

Experimental Results

Table 2 shows the test errors comparison of different methods for potential crowd flow prediction of new planned sites.

Results summary We can observe: (1) the models which can synthesis both spatial and temporal dependencies (e.g., LP-GLP, STMGCN, and proposed MOHER) achieve better performances. The proposed MOHER which can sufficiently model each type of cross-mode relations outperforms other baselines. Compared with the second-best result, MOHER decreases the errors by (9.1%, 6.8%) under RMSE and (37.0%, 13.3%) under MAPE. (2) The methods which model the multiple relations perform better than NA-HA or NA-LSTM. Meanwhile, the methods which especially consider the temporal dependency achieve better performances than NA-HA or MGCN-MLP, especially on the NYC Citi Bike Dataset. (3) MLC-PPF performs worse than our MOHER for the prediction of the long-term evolving targets due to the lack of cross-mode information and spatial dependency modeling. As the sharing bike is often used as an alternative transportation mode, the influence of cross-mode crowd flows on bikes is much more important than that on

Methods	Subway		Bike	
	RMSE	MAPE	RMSE	MAPE
NA-HA	2182.2	3.942	20.967	1.828
MGCN-MLP	2060.6	3.185	23.020	1.528
NA-LSTM	1947.1	1.955	18.177	1.015
LP-GLP	1986.0	1.799	19.021	0.826
STMGCN	1821.2	1.784	19.230	0.988
MLC-PPF	1757.5	1.931	22.612	1.532
MOHER	**1597.3**	**1.124**	**16.767**	**0.716**

Table 2: Potential Crowd flow forecasting error of new planned sites given by RMSE and MAPE on NYC Subway Turnstile Usage Dataset and NYC Citi Bike Dataset.
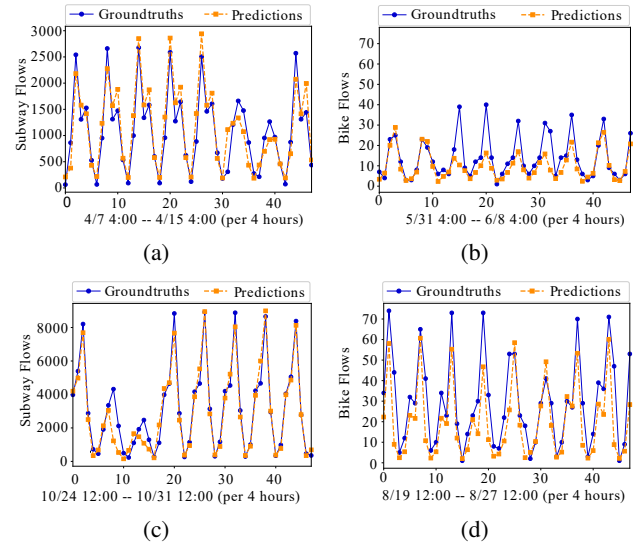

Figure 5: (a) 'Penn' subway station. (b) '21th St-49th Ave' bike station. (c) '40 Street - Lowery St' subway station. (d) '24th St-41st Ave' bike station.

the subway, which may result in the worse performances of MLC-PPF on the bike dataset.

Case Study We display result insights of four sampled stations. The selected timeslots are evenly distributed among the testing set. We can observe from Fig. 5: (1) 'Penn' subway station is downtown of Manhattan and is around with abundant multi-mode crowd flows. Our method fits the curve well but may overestimate the potential flows. (2) In contrast, the bike station in (b) is sited just next to a large railway area without adequate crowd flows and POIs nearby. Our method tends to underestimate the potential flows. (3) the stations of (c)(d) are located in ordinary residential areas with regular crowd travel demands, and our method can model the diversion effects well. To summarise, even in the two extremes in (a)(b), our method can generate stable predictions of potential crowd flows for unseen new planned sites of different time slots in the long-range test sets.

Component	RMSE($\times 10^3$)	MAPE
w/o cross-mode relations	1.776	1.499
w/o functional similarity	1.689	1.247
w/o differences modeling	1.861	1.346
w/o regularization	1.676	1.471
MOHER	**1.597**	**1.124**

Table 3: Effect of our method without each component on NYC Subway Turnstile Usage Data. Removing any component will result in a statistically significant error increase.
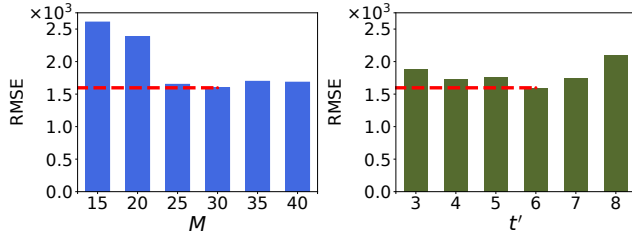

Figure 6: The Effect of Model Parameters

Ablation Study

To verify the contribution of different components in MO-HER to the performance gain, we further implement three simplified versions of our model to conduct ablation tests. The results of the degraded versions of MOHER on NYC Subway Turnstile Usage Dataset are shown in Table 3. We summarize the effects of three different components:

Effect of cross-mode flow data   To investigate the effect of cross-mode dependency, we construct the graph with the geographical proximity and POI similarity only on the target mode (i.e., the subway). With the synthesis of the two metrics in Table 3, we can find that removing the cross-mode relations increases the error which justifies the importance of exploring the cross-mode flow data. Thus, the potential crowd flows of target new sites are implied in the various nearby transportation modes.

Effect of modeling functional similarity   The regions with similar urban functions are considered to share similar flow patterns, so that we can infer the potential flow increments of new planned sites. The increased error demonstrates the importance of prior knowledge in functional similarity relations.

Effect of modeling differences among multi-mode   Here we only model the correlations for cross-mode spatial dependencies extraction. The insufficient modeling of cross-mode relations results in a significant increase in error, which validates our intuition and observation.

Effect of matrix reconstruction regularization   To evaluate the importance of parameter regularization, we directly leverage the original FNN to model heterogeneous relations. Because the high model complexity leads to overfitting, the degraded framework performs worse than MOHER.

Effect of Model Parameters

To study the effects of hyperparameters in our proposed framework, we present the performance on the Subway Dataset of two main hyperparameters for spatiotemporal prediction in Fig. 6, i.e., the number of neighbor nodes in CFRGs M and the length of time intervals t'. M is the hyperparameter to control the boundary of modeling spatial dependency, and t' controls the range of modeling temporal dependency. We observe that both the errors first decrease and then increase with the increase of M and t', because too small M or t' will both cause the lack of essential information for prediction. Meanwhile, larger M will bring some unnecessary noise, and larger t' may exceed the knowledge capacity of the hidden state in LSTM.

Related Work

How to mine the spatiotemporal information has become a long-standing problem for flow prediction (Lin et al. 2019; Yao et al. 2019). Some early works (Pan et al. 2019; Wang et al. 2019) used geographical distance to model the relationship. Recently, there have been some researches (Li et al. 2018; Zheng et al. 2020) on mining the complex relations of traffic flows. In particular, multiple heterogeneous relations have been explored as prior-knowledge for prediction (Chai, Wang, and Yang 2018; Geng et al. 2019; Sun et al. 2020; Zhou et al. 2020). (Li et al. 2019b) leveraged the graph embedding on heterogeneous spatial-temporal graphs for station-level demand prediction. (Ye et al. 2019) combined the bike and taxi data for demand co-prediction. However, there are only a few works towards unrecorded or potential flow prediction. Most of them are based on transductive tensor completion (Li et al. 2019c) to predict the historical missing data. In (Wu et al. 2020), the authors learned an inductive GNN for flows, which indicates the effect of GNN for unrecorded flow prediction. Moreover, (Gong et al. 2020) intends to predict the potential passenger flows of planned subway stations based on the flow completion problem. Nevertheless, this method ignores the cross-mode crowd flows, thus having a diversion effect.

Conclusion

In this paper, we proposed MOHER, an inductive potential crowd flow prediction framework which can naturally generalize to future new planned sites. We encoded the heterogeneous relations among the target and its cross-mode neighbor sites/regions by measuring the geographical proximity and the functional similarity to capture the flow diversion and potential crowd flows. For explicitly modeling cross-mode heterogeneous relations, we developed a novel cross-mode relational GCN to learn the correlations and the differences between multiple transportation modes. Experimental results on real-world datasets demonstrated that the proposed MOHER framework outperforms the compared state-of-the-art methods with satisfactory margins.

Acknowledgments

References

Babu, C. N.; Sure, P.; and Bhuma, C. M. 2020. Sparse Bayesian Learning Assisted Approaches for Road Network Traffic State Estimation. *IEEE Transactions on Intelligent Transportation Systems* .

Chai, D.; Wang, L.; and Yang, Q. 2018. Bike flow prediction with multi-graph convolutional networks. In *Proceedings of the 26th ACM SIGSPATIAL international conference on advances in geographic information systems*, 397–400.

Chen, L.; Zhang, D.; Pan, G.; Ma, X.; Yang, D.; Kushlev, K.; Zhang, W.; and Li, S. 2015. Bike sharing station placement leveraging heterogeneous urban open data. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 571–575.

Chen, X.; Yang, J.; and Sun, L. 2020. A nonconvex low-rank tensor completion model for spatiotemporal traffic data imputation. *Transportation Research Part C: Emerging Technologies* 117: 102673.

Deng, D.; Shahabi, C.; Demiryurek, U.; Zhu, L.; Yu, R.; and Liu, Y. 2016. Latent space model for road networks to predict time-varying traffic. In *ICDM'16*, 1525–1534.

Geng, X.; Li, Y.; Wang, L.; Zhang, L.; Yang, Q.; Ye, J.; and Liu, Y. 2019. Spatiotemporal multi-graph convolution network for ride-hailing demand forecasting. In *AAAI'19*, volume 33, 3656–3663.

Gong, Y.; Li, Z.; Zhang, J.; Liu, W.; and Yi, J. 2020. Potential Passenger Flow Prediction: A Novel Study for Urban Transportation Development. In *AAAI'20*, 4020–4027.

Hamilton, W.; Ying, Z.; and Leskovec, J. 2017. Inductive representation learning on large graphs. In *Advances in neural information processing systems*, 1024–1034.

Hochreiter, S.; and Schmidhuber, J. 1997. Long short-term memory. *Neural computation* 9(8): 1735–1780.

Jelokhani-Niaraki, M.; and Malczewski, J. 2015. Decision complexity and consensus in Web-based spatial decision making: A case study of site selection problem using GIS and multicriteria analysis. *Cities* 45: 60–70.

Kamarianakis, Y.; and Prastacos, P. 2003. Forecasting traffic flow conditions in an urban network: Comparison of multivariate and univariate approaches. *Transportation Research Record* 1857(1): 74–84.

Li, Q.; Wu, X.-M.; Liu, H.; Zhang, X.; and Guan, Z. 2019a. Label efficient semi-supervised learning via graph filtering. In *CVPR'19*, 9582–9591.

Li, W.; Nie, D.; Wilkie, D.; and Lin, M. C. 2017. Citywide estimation of traffic dynamics via sparse GPS traces. *IEEE Intelligent Transportation Systems Magazine* 9(3): 100–113.

Li, Y.; Yu, R.; Shahabi, C.; and Liu, Y. 2018. Diffusion Convolutional Recurrent Neural Network: Data-Driven Traffic Forecasting. In *ICLR'18*.

Li, Y.; Zhu, Z.; Kong, D.; Xu, M.; and Zhao, Y. 2019b. Learning heterogeneous spatial-temporal representation for bike-sharing demand prediction. In *AAAI'19*, volume 33, 1004–1011.

Li, Z.; Sergin, N. D.; Yan, H.; Zhang, C.; and Tsung, F. 2019c. Tensor completion for weakly-dependent data on graph for metro passenger flow prediction. *arXiv preprint arXiv:1912.05693* .

Lin, M.; Huang, C.; and Xu, Z. 2020. MULTIMOORA based MCDM model for site selection of car sharing station under picture fuzzy environment. *Sustainable cities and society* 53: 101873.

Lin, Z.; Feng, J.; Lu, Z.; Li, Y.; and Jin, D. 2019. Deepstn+: Context-aware spatial-temporal neural network for crowd flow prediction in metropolis. In *AAAI'19*, volume 33, 1020–1027.

Liu, H.; Tong, Y.; Han, J.; Zhang, P.; Lu, X.; and Xiong, H. 2020. Incorporating Multi-Source Urban Data for Personalized and Context-Aware Multi-Modal Transportation Recommendation. *IEEE Transactions on Knowledge and Data Engineering* .

Mutinda, F.; Nakashima, A.; Takeuchi, K.; Sasaki, Y.; and Onizuka, M. 2019. Time series link prediction using NMF. *Journal of Information Processing* 27: 752–761.

Pan, Z.; Liang, Y.; Wang, W.; Yu, Y.; Zheng, Y.; and Zhang, J. 2019. Urban traffic prediction from spatio-temporal data using deep meta learning. In *KDD'19*, 1720–1730.

Sun, J.; Zhang, J.; Li, Q.; Yi, X.; Liang, Y.; and Zheng, Y. 2020. Predicting citywide crowd flows in irregular regions using multi-view graph convolutional networks. *IEEE Transactions on Knowledge and Data Engineering* .

Takeuchi, K.; Kashima, H.; and Ueda, N. 2017. Autoregressive tensor factorization for spatio-temporal predictions. In *ICDM'17*, 1105–1110. IEEE.

Wang, Y.; Yin, H.; Chen, H.; Wo, T.; Xu, J.; and Zheng, K. 2019. Origin-destination matrix prediction via graph convolution: a new perspective of passenger demand modeling. In *KDD'19*, 1227–1235.

Wu, Y.; Zhuang, D.; Labbe, A.; and Sun, L. 2020. Inductive Graph Neural Networks for Spatiotemporal Kriging. *arXiv preprint arXiv:2006.07527* .

Xie, K.; Li, X.; Wang, X.; Xie, G.; Wen, J.; and Zhang, D. 2019. Active sparse mobile crowd sensing based on matrix completion. In *Proceedings of the 2019 International Conference on Management of Data*, 195–210.

Yao, H.; Tang, X.; Wei, H.; Zheng, G.; and Li, Z. 2019. Revisiting spatial-temporal similarity: A deep learning framework for traffic prediction. In *AAAI'19*, volume 33, 5668–5675.

Ye, J.; Sun, L.; Du, B.; Fu, Y.; Tong, X.; and Xiong, H. 2019. Co-prediction of multiple transportation demands based on deep spatio-temporal neural network. In *KDD'19*, 305–313.

Yu, B.; Yin, H.; and Zhu, Z. 2018. Spatio-temporal graph convolutional networks: a deep learning framework for traffic forecasting. In *IJCAI'18*, 3634–3640.

Yuan, Z.; Liu, H.; Liu, Y.; Zhang, D.; Yi, F.; Zhu, N.; and Xiong, H. 2020. Spatio-Temporal Dual Graph Attention Network for Query-POI Matching. In *SIGIR'20*, 629–638.

Zhang, M.; and Chen, Y. 2019. Inductive Matrix Completion Based on Graph Neural Networks. In *International Conference on Learning Representations*.

Zheng, C.; Fan, X.; Wang, C.; and Qi, J. 2020. Gman: A graph multi-attention network for traffic prediction. In *AAAI'20*, volume 34, 1234–1241.

Zhou, Q.; Gu, J.-J.; Ling, C.; Li, W.-B.; Zhuang, Y.; and Wang, J. 2020. Exploiting Multiple Correlations Among Urban Regions for Crowd Flow Prediction. *J. Comput. Sci. Technol* 35(2): 338.