

# Potential Passenger Flow Prediction: A Novel Study for Urban Transportation Development

Yongshun Gong,<sup>1</sup> Zhibin Li,<sup>1</sup> Jian Zhang,<sup>1</sup> Wei Liu,<sup>1</sup> Jinfeng Yi<sup>2</sup>

<sup>1</sup>Faculty of Engineering and IT, University of Technology Sydney, Sydney, Australia

<sup>2</sup>JD AI Research, Beijing, China

{yongshun.gong, zhibin.li}@student.uts.edu.au; {jian.zhang, wei.liu}@uts.edu.au; yijinfeng@jd.com

## Abstract

Recently, practical applications for passenger flow prediction have brought many benefits to urban transportation development. With the development of urbanization, a real-world demand from transportation managers is to construct a new metro station in one city area that never planned before. Authorities are interested in the picture of the future volume of commuters before constructing a new station, and estimate how would it affect other areas. In this paper, this specific problem is termed as potential passenger flow (PPF) prediction, which is a novel and important study connected with urban computing and intelligent transportation systems. For example, an accurate PPF predictor can provide invaluable knowledge to designers, such as the advice of station scales and influences on other areas, etc. To address this problem, we propose a multi-view localized correlation learning method. The core idea of our strategy is to learn the passenger flow correlations between the target areas and their localized areas with adaptive-weight. To improve the prediction accuracy, other domain knowledge is involved via a multi-view learning process. We conduct intensive experiments to evaluate the effectiveness of our method with real-world official transportation datasets. The results demonstrate that our method can achieve excellent performance compared with other available baselines. Besides, our method can provide an effective solution to the cold-start problem in the recommender system as well, which proved by its outperformed experimental results.

## Introduction

With the growth of intelligent transportation systems, passenger flow prediction models concentrate on discovering the volume of crowds and mobility patterns that best serve people’s daily life (Pan et al. 2019; Zhang et al. 2019). Recent advances in passenger flow prediction are focusing mainly on next time interval flow conditions with time evolves (Gong et al. 2018; Sun et al. 2015). If a brand-new metro station is inserted into the original metro network, existing predictors have to collect a large amount of latest transactional data to ensure normal operation. However, a real-world requirement from transportation authorities is

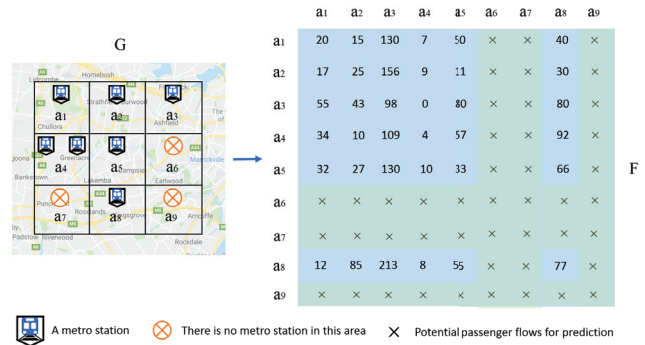


Figure 1: The example of PPF prediction problem. We aim to forecast the passenger flows of target areas (e.g.,  $a_6$ ,  $a_7$ ,  $a_9$ ) across the entire city network.

that they want to obtain the potential passenger flows (PPF) of a planned city area in advance (i.e., before constructing a station in this area). It is significant for the urban traffic development and transportation management, as it can provide insights for the site selection of stations and analysis of passenger movement patterns, as well as give the potential crowd warning.

In the PPF prediction task, concentrating solely on the entrance and exit potential flows does not provide adequate information, authorities also desperately want to master the distribution of predicted PPF, i.e., forecast the number of potential passengers moving to different destinations. It is utmost important to find how will the new station affect other areas. For instance, Figure 1 illustrates an example of the PPF prediction problem. A city region is partitioned into nine areas<sup>1</sup>, six of them have metro stations (termed as known areas), and three have not constructed yet (termed as target areas). The right part of Figure 1 presents an origin-destination (OD) matrix (each row point is the origin area and column points are destinations), e.g.,  $F(a_1, a_3) = 130$  indicates that there are 130 passengers departure from  $a_1$  and are going to the  $a_3$ . PPF task aims to make an accurate

Copyright © 2020, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

<sup>1</sup>We use grids for clear and simple illustration, the real partition standard is explained in the section of data description.

prediction for the target areas in one period (e.g., rush hours) that completes the crowd flows between them and known areas.

To date, limited studies considered the OD passenger flow prediction problem (Gong et al. 2018; Wang et al. 2019), and to the best of our knowledge, none of existing techniques can forecast PPF across the entire city. It is a novel problem and a real urban developing demand that faces several major challenges: (1) Considering the number of passenger flows and their final destinations simultaneously. (2) Analogously to the cold-start problem in the recommender system (Lam et al. 2008), it is hard to infer the preference of a new user from the known data. In our problem, a new station in the target area can be similarly regarded as a new user. (3) Since the PPF is a spatial-temporal mining problem, spatial and temporal information should be taken into account appropriately.

To resolve this novel and significant problem, in this paper, we devise a multi-view localized correlation learning model for the PPF prediction (MLC-PPF for short). To leverage the spatial information, we first construct a localized similarity matrix which associates with the real geographical neighbors and regional properties (e.g., business or residential regions). The intuition behind this strategy is from the First Law of Geography (Tobler 1970), i.e. *“Everything is related to everything else, but near things are more related than distant things”*. Second, a novel weighted correlation learning strategy is proposed. At last, to improve the prediction accuracy and well handle the cold-start challenge, we draw the side information from urban statistical data, where each area has a multi-view features to guide the learning process. In summary, our main contributions are shown as follows:

- We formulate the PPF prediction problem and provide the first attempt on forecasting passenger flows for urban transportation development.
- We propose a multi-view localized correlation learning method to provide a solution for the PPF prediction that can learn localized correlations via a multi-view learning process.
- We show that our method can be transferred to the classic cold-start problem in the recommender system. It achieves a superior result that gives a new perspective for relevant tasks.
- We conduct extensive prediction experiments on a large real-world transactional dataset and show that our model outperforms other available algorithms.

## Related Work

### Passenger Flow Prediction

Most existing passenger flow prediction models focused on forecasting entrance/exit flows at certain stations or areas, neglecting the crowd flows across different stations. (Chen and Wei 2011; Wei and Chen 2012; Ni, He, and Gao 2017). Wei et al. (Chen and Wei 2011) developed an effective short-term passenger flow prediction model to explore the time variants and capture dynamic patterns on a single subway

line. Subsequently, a modified approach is proposed based on the neural network, which aims to solve the same entrance/exit crowd flow prediction task in a few metro lines (Wei and Chen 2012). Ni et al. (Ni, He, and Gao 2017) used auxiliary information, such as social media events, to improve the forecast performance.

One of the research hotspots is named the city network-wide crowd flow prediction, which is a significant task for the modern transportation management (Deng et al. 2016). Nowadays, some of methods were focusing on forecasting the citywide crowd flows. (Ma et al. 2016) devised a series of visualization approaches to show the flows’ dynamic changes in the networks. Zhang et al. proposed the deep learning models based on the ResNet to predict crowd in-flows and outflows of the entire city regions. (Zhang et al. 2018; 2019). The Probabilistic model is an effective approach to estimate the traffic speed. For example, (Zhan et al. 2016) and (Lin et al. 2017) used trajectory data to estimate citywide traffic volume via probabilistic graphical models. (Gong et al. 2018) proposed an effective method based on online latent space learning to predict the crowd flow distribution, i.e., forecast the OD pairs and the quantity of passenger flows simultaneously. To the best of our knowledge, none of existing crowd flow prediction methods considered the PPF problem studied in this paper.

Other relevant studies, such as (Hsieh, Lin, and Zheng 2015), are point-based prediction model, not in a matrix formulation. (Hsieh, Lin, and Zheng 2015) selects  $k$  points to predict  $k$  values. But in our task,  $k$  target areas are required  $nk$  prediction values, where  $n$  is the number of known areas. It is because we also need to consider the crowd flows between each area.

### Multi-view Learning

Traditional passenger flow mining usually deals with data from a single view. Recently, there exists a diversity of datasets from different sources in various domains with multiple views (Zheng 2015). The multi-view learning algorithm is widely recognized as an effective way of solving the cross-domain problem, that features from different views can be served for the target domain learning process (Singh and Gordon 2008; Xu, Tao, and Xu 2013; Elkahky, Song, and He 2015). (Xu, Tao, and Xu 2015) proposed a matrix co-factorization based method (MVL-IV) to embed different views into a shared subspace, such that the incomplete views can be estimated by the information on observed views. To connect multiple views, MVL-IV assumed that different views have distinct ‘feature’ matrices (i.e.,  $\{\mathbf{H}_i\}_{i=1}^m$ ), but correspond to the same coefficient matrix (i.e.,  $\mathbf{W}$ ). The tensor-based methods, such as (Hu et al. 2013), (Hu et al. 2016) (Taneja and Arora 2018) were proposed to address the cross-domain recommendation problem. They devised a cross-domain triadic factorization model to learn the triadic factors for user, item and domain, where the item dimensionality varies with domains. The above approaches and other similar methods (Rendle et al. 2009; Xiong et al. 2010) cannot address our PPF prediction problem directly because they are not formulated for the passenger flow prediction task. However, since they can handle the

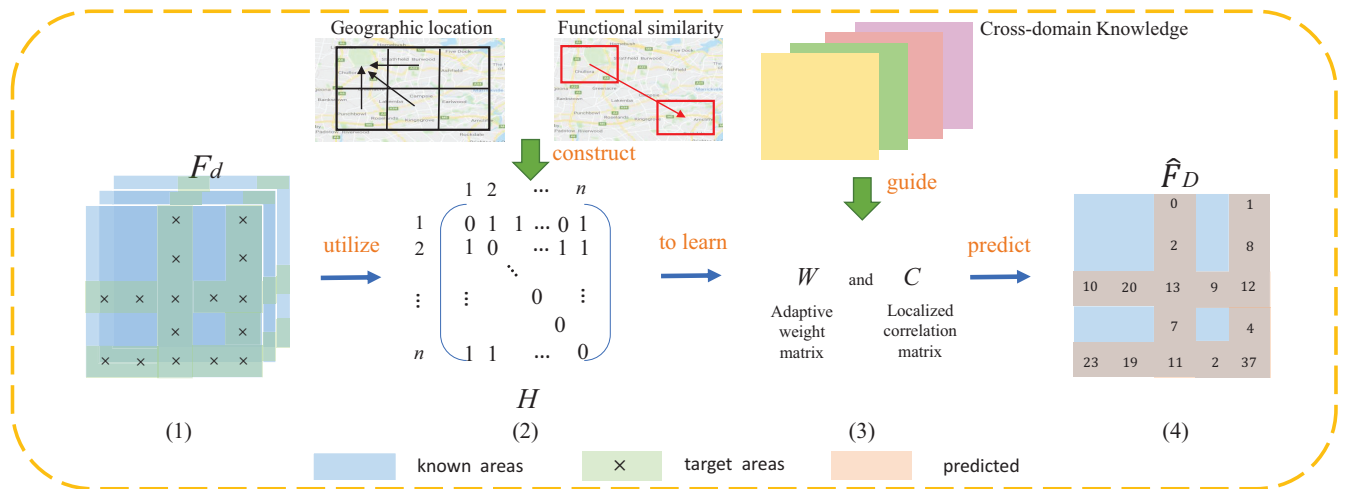


Figure 2: The flowchart of our proposed model. In the learning process, given a set of previous PPF matrices  $\{F_d\}$ , MLC-PPF learns the localized correlation matrix  $C$  and adaptive-weight  $W$  via a  $k$ -nearest indicator matrix  $H$ . The cross-domain knowledge is utilized to guide the updating of  $C$ . Then, the target prediction can be inferred by Algorithm 1.

cold-start problem by utilizing the cross-domain knowledge, an illuminative clue is educated.

In conclusion, none of relevant studies can solve the PPF prediction problem directly. Accordingly, this paper aims to design a reliable approach for PPF prediction with cross-domain knowledge involved.

### Problem Statement

Focusing on the PPF prediction problem, every origin-destination among areas needs to be recorded. We formulate the OD passenger flow network as a fully connected graph  $G = (A, E)$ , where  $A$  is a set of vertexes and  $E$  is the set of edges.  $a_i \subseteq V$  records the  $i$ -th origin or destination area, and an edge  $e(v_i, v_j)$  denotes an origin-destination flow from area  $a_i$  to  $a_j$ . The value of each edge  $e(a_i, a_j)$  is associated with the observed flow  $f(a_i, a_j)$ , i.e.,  $f(a_i, a_j)$  is the total number of passengers that departure from  $i^{th}$  area and are going to the  $j^{th}$  area. Then,  $G$  can be represented by PPF matrix  $F = (f_{ij})$ , where  $f_{ij} = f(a_i, a_j)$ . The example of  $G$  and  $F$  is shown in Figure 1.  $f_{31} = 55$  means that 55 passengers leave from  $a_1$  and their destination is  $a_2$ .

In the real-world scenario, one area may have several stations. In this case, we calculate the passenger flows of these stations together to present the total flows of the area. We consider three specific and useful time periods to predict PPF, which will help the authorities to do a better temporal analysis of transportation development. The three periods are morning rush hour (7.00 AM - 9.00 AM), afternoon rush hour (5.00 PM - 7.00 PM), and non-rush hour (2.00 PM - 4.00 PM).

Furthermore, traffic periodicity is a very important factor for relevant studies. Crowd flows also represent the stable and daily periodic properties, especially on weekdays. To extract the temporal information and make a more general prediction, we consider a series of previous daily PPF matrices  $(F_1, F_2, \dots, F_D)$  in the same time period to predict the

PPF matrix of target areas ( $\hat{F}_D$ ) for the day  $D$ . Note that, the prediction is not limited in the  $D$ -th day, the target can be changed easily based on the real requirement.

To best simulate the crowd flow changes when picking the target areas, in this paper, we tracked all trajectories of passengers, from origins to their destinations. For example, if area  $a_1$  is selected as a target area, all the departure crowd flows from  $a_1$  will add to its closest area (e.g.,  $a_2$ ) to best simulate the people's choice. In this way, the PPF is learned by the crowd flows under the assumption that the original passengers from  $a_1$  will departure from its closest neighbor  $a_2$ .

### The Proposed Method

In this section, we propose our PPF prediction model MCL-PPF. We will describe the strategy of localized correlation learning with adaptive-weight, and how to leverage the cross-domain multi-view information to improve our work. Figure 2 shows the flowchart of our model.

#### Localized Correlation Learning

PPF prediction problem is a spatially related task that the more similar between two areas, the more correlations of passenger flow condition they have. Assuming that a city is partitioned into  $n$  areas, including known and target areas.  $F \in \mathbb{R}^{n \times n}$  presents the PPF matrix, and  $F^k \in \mathbb{R}^{n \times n}$  presents the localized flow matrix of  $F$ , where the  $i$ -th row of  $F^k$  is a combination of its  $k$ -nearest area passenger flows. In that way, we formulate the function to learn the localized correlation, which can be expressed as:

$$\begin{aligned} \min_C \quad & \frac{1}{2} \sum_{d=1}^D \|F_d - F_d^k C\|_F^2 \\ \text{s.t.} \quad & P_\Omega(F_d) = P_\Omega(F_d^k) \end{aligned} \quad (1)$$

where  $C \in \mathbb{R}^{n \times n}$  is the localized correlation matrix that learns the transformation from  $F_d^k$  to  $F_d$  in each period of day  $d$ ,  $D$  is the total number of days;  $P_\Omega(\cdot)$  stands for the projection of all observed passenger flows from the known area set  $\Omega$ ;  $\|\cdot\|_F$  is Frobenius norm of matrix.

Now, we will discuss how to build  $F_d^k$  of one period in a day. The physical distances among areas need to be considered first. Moreover, the development of a city gradually fosters different functional regions, such as business and educational areas, where the areas belonging to the same functional region will have strong connections with their properties (Zheng et al. 2014).

Thus, the similarities among areas should take into account the above two standards. To this end, we build two distance metrics from the real geographic location and regional similarity. The distance metric between  $i$ -th and  $j$ -th areas is shown as follows:

$$s_{i,j} = 2 - \left( \frac{dist_{i,j}}{\max(dist_{i,:})} + \frac{dist'_{i,j}}{\max(dist'_{i,:})} \right) \quad (2)$$

where  $dist_{i,j}$  is the geographic distance between  $i$ -th and  $j$ -th areas; and  $dist'_{i,j}$  presents the Euclidean distance which is calculated by intrinsic features of areas (e.g., point of interest attributes).

After having gotten the  $s_{i,j}$ , the  $k$  neighbors of the  $i$ -th area can be picked. Then, we construct an indicator matrix  $H$  for the  $k$ -nearest neighbors of all areas where each row indicates the position of its  $k$ -nearest known areas. For example, in the stage 2 of Figure 2, the first row of  $H$  illustrates that  $a_2, a_3$ , and  $a_n$  are the  $k$ -nearest areas of  $a_1$  if  $k$  is setting to 3. Accordingly, for each day  $d$ , the localized flow matrix  $F_d^k$  can be represented as  $F_d^k = (H \odot W)F_d$ , where  $W$  is an adaptive-weight matrix that learns the different weights of  $k$ -nearest areas. To this stage, the localized correlation learning process is shown as Eq. (3).

$$\min_{C,W} \frac{1}{2} \sum_{d=1}^D \|F_d - (H \odot W)F_d C\|_F^2 \quad (3)$$

s.t.  $P_\Omega(F_d) = P_\Omega(F_d^k)$

where  $\odot$  is the entrywise product; The loss function aims to learn the localized correlations matrix  $C$  and weight matrix  $W$  simultaneously.

### Improvement by Cross-domain Learning Process

As mentioned above, there are various functional regions of a city. Thanks to the urban statistical data, the passenger flow similarities among different areas can be reviewed from this cross-domain perspective. Based on the phenomenon that the similar functional regions have the similar passenger flows (e.g., the business regions have a large number of entrance flows during the morning rush hour, and people leave from residential areas in the same time span), we leverage such information to guide the localized correlation learning process.

The statistical data have multiple views to record the differences between areas. For example, the economy

view reveals the economic features, such as the number of industries and employee statistics; and the population view consists of detailed population information. Let  $\{X_1, X_2, \dots, X_V\}$  denote the multi-fold views of statistical data, where  $X_v \in \mathbb{R}^{n \times m_v}$ , the row of  $X_v$  denotes the area and column denotes the feature. To improve the prediction performance, cross-domain knowledge is involved as guidance, which can be formulated as:

$$\min_{C,W} \frac{1}{2} \sum_{d=1}^D \|F_d - (H \odot W)F_d C\|_F^2 + \frac{\lambda}{2} \sum_{v=1}^V \|X_v - C X_v^k\|_F^2$$

s.t.  $P_\Omega(F_d) = P_\Omega(F_d^k)$  (4)

where  $\lambda$  is the regularization parameter;  $X_v^k = H X_v$  denotes the localized matrix of  $X_v$ .

After solving Eq. (4), the learned matrices  $W$  and  $C$  can be used to make the prediction. The predicted PPF of target areas in  $D$ -th day is:

$$\hat{F}_D = (\mathbf{1} - Y) \odot ((H \odot W)F_D C) \quad (5)$$

where  $Y$  is an indicator matrix whose entry  $(i, j)$  is one if  $F(i, j)$  is observed and zero otherwise.

To this stage, the OD passenger flows in the target areas are learned by the above processes, i.e., predict each row of target areas. However, the column of target areas revealing how much crowds arrived at these areas needs to be predicted with a slight modification. That is, replace  $F_d$  with  $F_d^\top$  in Eq. (4) to learn the localized correlation from the other side. It can be solved in a likewise manner. Thus we only presented the optimization strategy of Eq. (4) due to the page limitation.

### Learning and Prediction

Eq. (4) is a complex non-convex problem. But the loss function associated with Eq. (4) is convex regarding  $C$  with fixed  $W$  and vice versa. We can optimize them alternatively until convergence (e.g., alternating least squares (ALS)). A straightforward way to minimize the loss function is through the gradient method.

Considering  $C$  while  $W$  is fixed, Eq. (4) can be rewritten as follows:

$$\mathcal{L} = \frac{1}{2} \sum_{d=1}^D Tr((F_d - (H \odot W)F_d C)(F_d - (H \odot W)F_d C)^\top) + \frac{\lambda}{2} \sum_{v=1}^V Tr((X_v - C X_v^k)(X_v - C X_v^k)^\top) \quad (6)$$

Taking the derivative of  $\mathcal{L}$  with respect to  $C$ , we can get gradient  $gC$ :

$$gC = \sum_{d=1}^D ((H \odot W)F_d)^\top ((H \odot W)F_d C - F_d) + \lambda \sum_{v=1}^V (C X_v X_v^{k\top} - X_v X_v^{k\top}) \quad (7)$$

Analogously, the derivative of  $\mathcal{L}$  with respect to  $W$  is:

$$gW = \sum_{d=1}^D (H \odot W) F_d C C^\top F_d^\top - H \odot (F_d C^\top F_d^\top) \quad (8)$$

Let  $\alpha$  and  $T$  be the step-size and number of iterations. In each stage, we adopt the following update rules:

$$C_{t+1} = C_t - \alpha \frac{gC_t}{\|gC_t\|_F} \quad (9)$$

$$W_{t+1} = W_t - \alpha \frac{gW_t}{\|gW_t\|_F} \quad (10)$$

$$F_{d(t+1)} = Y \odot F_{d(t)} + (\mathbf{1} - Y) \odot ((H \odot W_t) F_{d(t)} C_t) \quad (11)$$

where  $t = 1, 2, \dots, T$ .

Based on the above update equations, the iterative learning and prediction process for MLC-PPF are summarized in Algorithm 1.

---

#### Algorithm 1: MLC-PPF

---

**Input:** PPF matrices  $[F_1, \dots, X_D]$ ; Mutiple views of statistical data  $[X_1, \dots, X_V]$ .

**Output:** Prediction  $F_D$

- 1 Initialize  $C$ :  $C \leftarrow (Y \odot (H \odot W) F_D)^\dagger (Y \odot F_D)$  by solving Eq. (3), where  $\dagger$  the pseudo-inverse of matrix;
  - 2 Initialize  $W$ :  $W \leftarrow S$ , where  $S$  is built by Eq. (2)
  - 3 Construct  $H$  by the real geographic location and regional similarity.
  - 4 **for**  $t = 1$  to  $T$  **do**
  - 5     **if**  $|\mathcal{L}_t - \mathcal{L}_{t+1}| / \mathcal{L}_t \geq \varepsilon$  **then**
  - 6         update  $C_{t+1}$  **By** Equation 9
  - 7         update  $W_{t+1}$  **By** Equation 10
  - 8         update  $F_{t+1}$  **By** Equation 11
  - 9     **else**
  - 10         Break
  - 11 Return  $\hat{F}_D$  **By** Equation 5.
- 

## Experiments

### Data Description

- We describe the transactional dataset used in this paper, which is a large-scale, real-world dataset provided by NSW Sydney Transport. After data cleaning<sup>2</sup>, the dataset contains above 35 million transactional records covering 194 stations including the city train and ferry stations between 7 Nov 2016 and 11 Dec 2016. We pick the data between 7 Nov. 2016 and 20 Nov. 2016 as the training and validation sets (used to tune parameters); the remaining data are used as the test set.

<sup>2</sup>We removed the recording errors and UNKNOWN trips, etc.

- The urban statistical data are collected from Australian Bureau of Statistics-2016 (ABS) with four views; those are Economy, Family, Income, and Population. The numbers of the di-mension of four views are 43, 44, 50, 97, respectively.
- All the transactional dataset across the transport network are mapped into 117 areas to build the flow matrices  $F_d$ ,  $d = 1, \dots, D$ . The designation of areas is based on the Australian Statistical Geography Standard for the best practical value.

### Methods and Metrics

We use the following five baselines which can learn the flow data by the cross-domain knowledge guidance. Among them, CDTF and WITF are two tensor-factorization-based (TF) methods that can solve the cold-start problem. For NMF, we concatenate the flow matrix with the statistical data. All parameters used in baselines and our method are picked by a grid search approach.

- **NMF**: Predict the PPF by the non-negative matrix factorization, which concatenates the flow matrix and the statistical data (Lee and Seung 2001).
- **MVL-IV**: A state-of-the-art multi-view learning method based on the matrix co-factorization, it learns the same coefficient matrix to connect multiple views (Xu, Tao, and Xu 2015). In this method, we set the flow matrix as one of the views, other views are from the ABS dataset.
- **LS-KNN**: Latent similarity  $k$ -nearest neighbors. After calculating the latent similarities among areas by Eq. (2), we pick  $k$ -nearest neighbors of the target areas, and use average crowd flows of these neighbors as an estimate ( $k=4$ ).
- **CDTF**: A state-of-the-art TF method to learn the cross-domain knowledge (Hu et al. 2013).
- **WITF**: A weighted irregular TF method which is similar as the CDTF (Hu et al. 2016). For CDTF and WITF, we leverage the passenger flow and ABS data to construct the tensor.

**Metrics.** We used the two most widely used evaluation metric to measure the PPF prediction quality. They are *Mean Absolute Error* (MAE) and *Normalized Root Mean Square Error* (NRMSE).

$$MAE = \frac{\sum_{i,j=1}^M |f_{ij} - \hat{f}_{ij}|}{M}$$

$$NRMSE = \frac{100\%}{nval} \sqrt{\frac{1}{M} \sum_{i,j=1}^M (f_{ij} - \hat{f}_{ij})^2}$$

where  $\hat{f}_{ij}$  is a forecasting passenger flow from  $i$ -th area to  $j$ -th; and  $f_{ij}$  is the ground truth;  $M$  is the number of predictions;  $nval = \max(f_{ij}) - \min(f_{ij})$ .

Table 1: Comparisons with different time periods. We report the average mean absolute errors (MAE) and Normalized Root Mean Square Error (NRMSE) among various methods. The target areas occupied 20% of the total set. Best results are bold.

Methods	Morning Rush Hour		Afternoon Rush Hour		Non-rush Hour		Average	
	MAE	NRMSE	MAE	NRMSE	MAE	NRMSE	MAE	NRMSE
NMF	124.50	30.78%	117.92	37.13%	89.11	28.44%	110.51	32.12%
MVL-IV	108.31	29.50%	101.55	29.78%	92.05	27.54%	100.64	28.94%
CDTF	75.15	22.43%	84.02	25.93%	67.78	19.37%	75.65	22.58%
WITF	69.30	18.73%	72.06	19.45%	62.57	17.26%	67.98	18.48%
LS-KNN	19.89	5.42%	20.20	7.67%	23.51	7.94%	21.20	7.01%
MLC-PPF	<b>9.84</b>	<b>2.30%</b>	<b>11.47</b>	<b>3.12%</b>	<b>8.22</b>	<b>1.21%</b>	<b>9.84</b>	<b>2.21%</b>

Table 2: Comparisons with different removing ratios. We report MAE and NRMSE through all test data.

Methods	5%		10%		15%		25%	
	MAE	NRMSE	MAE	NRMSE	MAE	NRMSE	MAE	NRMSE
NMF	88.11	20.70%	90.23	21.66%	107.44	26.52%	128.70	30.11%
MVL-IV	92.75	20.36%	90.40	20.75%	99.01	23.79%	101.93	32.40%
CDTF	59.25	18.90%	61.25	19.26%	68.07	22.00%	79.06	26.35%
WITF	60.41	18.23%	60.72	19.77%	61.18	19.23%	71.07	20.73%
LS-KNN	13.69	4.72%	16.45	5.01%	18.51	5.44%	23.99	9.25%
MLC-PPF	<b>8.64</b>	<b>1.37%</b>	<b>8.80</b>	<b>1.20%</b>	<b>9.73</b>	<b>1.90%</b>	<b>11.07</b>	<b>2.34%</b>

### Comparisons on Different Time Periods

The first set of experiments is designed to assess the performance on different time periods. We randomly removed 20% areas as the target set, and the remaining 80% areas as the known set. The learning step  $\alpha$  is fitted to  $10^{-2}$  and there are only two hyper-parameters used in our method, where  $k$  and  $\lambda$  are chosen from  $\{1,2,3,\dots,10\}$  and  $\{10^{-5}, 10^{-4}, \dots, 10^5\}$  respectively. We repeat the experiment 20 times with random initialization and report the average results.

Experimental results are presented in Table 1. Compared with other approaches, our method achieved the best prediction accuracies on both three time periods. None of the multi-view and cross-domain methods work well because it is hard to capture the relationships between statistical data and the passenger flows. The approach LS-KNN performs better than other baselines, which illustrates that the PPF prediction problem has a strong spatial correlation property. In summary, the proposed method is a well-designed model for PPF prediction, which outperforms the other available baselines because it considers the localized correlations and the cross-domain knowledge simultaneously.

### Comparisons on Various Missing Ratios

In this experiment, we evaluate how performance will change with varied number of target areas. We randomly pick 5%, 10%, 15%, 25% areas as the target areas, and run 20 times to report the average errors. The test period is in the morning rush hours. The performances of different methods are summarized in Table 2.

It is apparent that the experimental results lead to similar conclusions to the first comparison. Our model, MLC-PPF, significantly outperforms all other comparative methods over all testing sets. The performances of MLC-PPF in

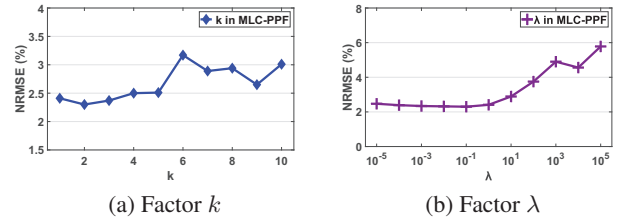


Figure 3: Effect of Parameters

5% dataset are very close to that of in 10%, which illustrates that the 90% remaining area set can learn a satisfied localized correlation and make accurate PPF predictions. In the real-world application, the proportion of target areas is usually small since only a few areas are suitable for constructing a new station.

### Parameter Analysis

In this section, we analyze the effects of two hyper-parameters used in this paper, where  $k$  is the number of nearest neighborhoods, and the regularization factor  $\lambda$  controls the strength of guidance from *ABS* data.

Figure 3(a) shows the different performances with a varying setting for  $k$ . For each area, the correlation matrix  $C$  only learns the transform from these neighborhoods. As can be shown in the results,  $k=2$  is the best choice for our method, and  $k$  is not sensitive in the range of 1 and 5.

Figure 3(b) represents various results by varying  $\lambda$ .  $\lambda=2^{-1}$  achieves the best results for our method, and the performances are stable when choosing between  $[10^{-5}, 10^0]$ .

In a summary, the parameters used in this paper are benefit to the improvement of our models. MLC-PPF is stable because it is insensitive to parameters.

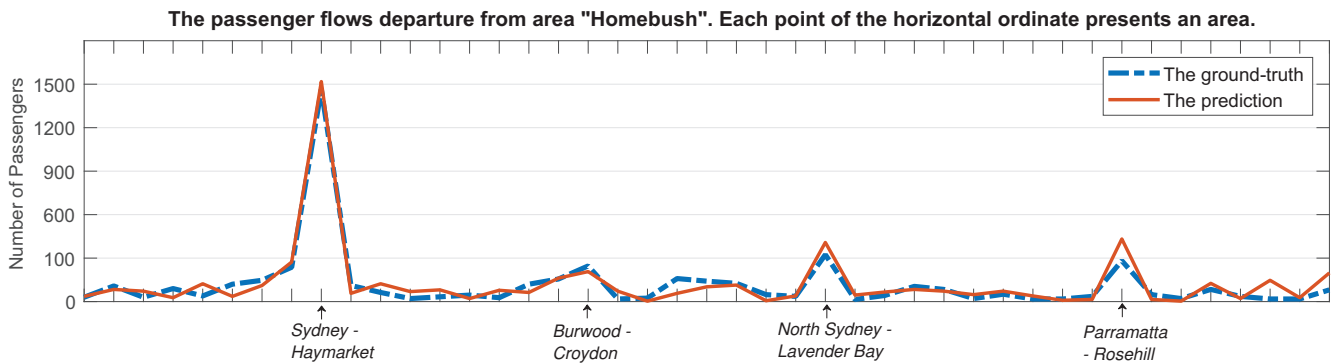


Figure 4: The case study. This figure shows the passenger flow prediction that departure from “Homebush” to other areas. To keep figure clear, we only draw our method and the ground-truth because other baselines perform far worse than the MLC-PPF.

### Case Study

We display a PPF prediction result of one target area in this section. In this case, the area “Homebush” is treated as the target area. For better visualization, we only remain the areas where the number of arrived passengers is greater than 5.

As shown in Figure 4, our model yields a great prediction result compared with the ground-truth, especially in some main areas of Sydney, such as the central area “Sydney-Haymarket”, “Burwood-Croydon”, “North Sydney-Lavender Bay” and “Parramatta-Rosehill”. The case study demonstrates the effectiveness of our method for the PPF prediction.

### Transfer to the Cold-start Problem

As we have emphasized, our strategy can provide a new perspective to address the classic cold-start problem in the recommender system. This set of experiments is designed to assess the transferability of our model.

We choose a very famous dataset from Amazon to do the evaluation, in which the dataset contains 1,555,170 users and 1-5 scaled ratings over 548,552 different products covering four domains: books, music CDs, DVDs and videos (Hu et al. 2013). We randomly remove the 20% users from target domains to simulate the cold-start situation. Three baselines are used in this comparison. CMF is an effective method based on the collective matrix factorization which couples rating matrices for all domains on the User dimension (Singh and Gordon 2008). CDTF and WITF are two tensor-factorization-based cross-domain recommendation methods, they devise a strategy to transform original data into a cubical tensor that can better capture the interactions between user factors and item factors (Hu et al. 2013; 2016). In this experiment, we leverage the information excluding target domain to build the  $k$ -nearest indicator matrix  $H$ .

Table 3 shows the results of our methods together with some state-of-the-art approaches. MLC-PPF can achieve the greatest accuracies for the target domains, which illustrates that our method is able to solve the unacquainted world phenomenon and give inspiration for relevant tasks. Despite the effectiveness of our methods, we should admit that there is

Table 3: Transfer to the cold-start problem. We report the MAE of all test methods.

Target Domain	CMF	CDTF	WITF	MLC-PPF
Book	0.834	0.755	0.740	<b>0.396</b>
Music	0.847	0.779	0.716	<b>0.582</b>

a limitation of MLC-PPF. MLC-PPF only can make the prediction when its  $k$ -neighbors have ratings. However, based on the test results, the predicted ratings are reliable and able to make the recommendation.

### Conclusion

In this paper, we proposed an effective method for the potential passenger flow prediction, which is a novel study that brings benefits to the urban transportation development. To address this spatio-temporal problem, we design a multi-view localized correlation learning model (MLC-PPF) for the PPF prediction. The  $k$ -nearest indicator matrix  $H$  is constructed by the real geographical neighbors and regional properties. MLC-PPF can learn the correlations between each known area and its  $k$ -nearest neighbors with the cross-domain knowledge guidance. We evaluate the performance of our method with a set of well-designed experiments. All empirical results not only demonstrate that the proposed model outperforms all the other methods in the PPF prediction task, but also represent the capability of tackling the cold-start problem in recommender system.

### References

- Chen, M.-C., and Wei, Y. 2011. Exploring time variants for short-term passenger flow. *Journal of Transport Geography* 19(4):488–498.
- 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 *Proceedings of the 22nd ACM SIGKDD*, 1525–1534.
- Elkahky, A. M.; Song, Y.; and He, X. 2015. A multi-view deep learning approach for cross domain user modeling in

- recommendation systems. In *Proceedings of the 24th International Conference on World Wide Web*, 278–288. International World Wide Web Conferences Steering Committee.
- Gong, Y.; Li, Z.; Zhang, J.; Liu, W.; Zheng, Y.; and Kirsch, C. 2018. Network-wide crowd flow prediction of sydney trains via customized online non-negative matrix factorization. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, 1243–1252. ACM.
- Hsieh, H.-P.; Lin, S.-D.; and Zheng, Y. 2015. Inferring air quality for station location recommendation based on urban big data. In *the 21th ACM SIGKDD International Conference*.
- Hu, L.; Cao, J.; Xu, G.; Cao, L.; Gu, Z.; and Zhu, C. 2013. Personalized recommendation via cross-domain triadic factorization. In *Proceedings of the 22nd international conference on World Wide Web*, 595–606. ACM.
- Hu, L.; Cao, L.; Cao, J.; Gu, Z.; Xu, G.; and Yang, D. 2016. Learning informative priors from heterogeneous domains to improve recommendation in cold-start user domains. *ACM Transactions on Information Systems (TOIS)* 35(2):13.
- Lam, X. N.; Vu, T.; Le, T. D.; and Duong, A. D. 2008. Addressing cold-start problem in recommendation systems. In *Proceedings of the 2nd international conference on Ubiquitous information management and communication*, 208–211. ACM.
- Lee, D. D., and Seung, H. S. 2001. Algorithms for non-negative matrix factorization. In *Advances in neural information processing systems*, 556–562.
- Lin, L.; Li, J.; Chen, F.; Ye, J.; and Huai, J. 2017. Road traffic speed prediction: a probabilistic model fusing multi-source data. *IEEE Transactions on Knowledge and Data Engineering* 30(7):1310–1323.
- Ma, Y.; Lin, T.; Cao, Z.; Li, C.; Wang, F.; and Chen, W. 2016. Mobility viewer: An eulerian approach for studying urban crowd flow. *IEEE Transactions on Intelligent Transportation Systems* 17(9):2627–2636.
- Ni, M.; He, Q.; and Gao, J. 2017. Forecasting the subway passenger flow under event occurrences with social media. *IEEE Transactions on Intelligent Transportation Systems* 18(6):1623–1632.
- 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 *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 1720–1730. ACM.
- Rendle, S.; Balby Marinho, L.; Nanopoulos, A.; and Schmidt-Thieme, L. 2009. Learning optimal ranking with tensor factorization for tag recommendation. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, 727–736. ACM.
- Singh, A. P., and Gordon, G. J. 2008. Relational learning via collective matrix factorization. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, 650–658. ACM.
- Sun, L.; Lu, Y.; Jin, J. G.; Lee, D.-H.; and Axhausen, K. W. 2015. An integrated bayesian approach for passenger flow assignment in metro networks. *Transportation Research Part C: Emerging Technologies* 52:116–131.
- Taneja, A., and Arora, A. 2018. Cross domain recommendation using multidimensional tensor factorization. *Expert Systems with Applications* 92:304–316.
- Tobler, W. R. 1970. A computer movie simulating urban growth in the detroit region. *Economic geography* 46(sup1):234–240.
- 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 *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 1227–1235. ACM.
- Wei, Y., and Chen, M.-C. 2012. Forecasting the short-term metro passenger flow with empirical mode decomposition and neural networks. *Transportation Research Part C: Emerging Technologies* 21(1):148–162.
- Xiong, L.; Chen, X.; Huang, T.-K.; Schneider, J.; and Carbonell, J. G. 2010. Temporal collaborative filtering with bayesian probabilistic tensor factorization. In *Proceedings of the 2010 SIAM International Conference on Data Mining*, 211–222. SIAM.
- Xu, C.; Tao, D.; and Xu, C. 2013. A survey on multi-view learning. *arXiv preprint arXiv:1304.5634*.
- Xu, C.; Tao, D.; and Xu, C. 2015. Multi-view learning with incomplete views. *IEEE Transactions on Image Processing* 24(12):5812–5825.
- Zhan, X.; Zheng, Y.; Yi, X.; and Ukkusuri, S. V. 2016. Citywide traffic volume estimation using trajectory data. *IEEE Transactions on Knowledge and Data Engineering* 29(2):272–285.
- Zhang, J.; Zheng, Y.; Qi, D.; Li, R.; Yi, X.; and Li, T. 2018. Predicting citywide crowd flows using deep spatio-temporal residual networks. *Artificial Intelligence* 259:147–166.
- Zhang, J.; Zheng, Y.; Sun, J.; and Qi, D. 2019. Flow prediction in spatio-temporal networks based on multitask deep learning. *IEEE Transactions on Knowledge and Data Engineering*.
- Zheng, Y.; Capra, L.; Wolfson, O.; and Yang, H. 2014. Urban computing: concepts, methodologies, and applications. *ACM Transactions on Intelligent Systems and Technology (TIST)* 5(3):38.
- Zheng, Y. 2015. Methodologies for cross-domain data fusion: An overview. *IEEE transactions on big data* 1(1):16–34.