Analysis and Prediction of Regional Mobility Patterns of Bus Travellers Using Smart Card Data and Points of Interest Data

Geqi Qi, Ailing Huang, Wei Guan, and Lingling Fan

***Abstract*— Mobility patterns at region level can provide more macroscopic and intuitive knowledge on how people gather in or depart from the region. However, the analysis and prediction of regional mobility patterns have yet to be effectively addressed. In light of this, using smart card data (SCD) and points of interest (POI) data, a multi-step methodology which integrates the inner-restricted fuzzy C-means clustering, nonnegative tensor factorization and artificial neural network are proposed and implemented in this paper. It overcomes the difficulties in region division, pattern extraction, and prediction. The bus SCD and POI data in Beijing city are utilized for proving the usefulness of the methodology. The regional mobility patterns of bus travellers in Beijing city are extracted from the third-order tensors involving 1110 regions, 34 time slots, and 7 days of the week. The analyzed results show that the proposed methodology has a good performance on predicting the regional mobility patterns based on the regional properties. Furthermore, by considering both of the regional boarding and alighting patterns, the predictions of the regional aggregation pattern can also be achieved. These research achievements can not only provide a deep insight on the human mobility patterns at region level, but also support the evidence-based and forward-looking urban planning and intelligent transportation management.**

***Index Terms*— Regional mobility pattern, smart card data, points of Interest, inner-restricted fuzzy C-means clustering, nonnegative tensor factorization, artificial neural network.**

I. Introduction

With the rapid development of urbanization and motorization, the aggregation of urban travellers induce many problems such as road traffic congestion [1], environmental pollution [2], crowding [3] and disease spreading [4], especially in big cities and mega cities. Therefore, the mobility of urban travellers has attracted more and more attention and the researches on mobility patterns have emerged one after another due to the increasing availability of big data sources [5]–[8]. However, two major issues that should be concerned about have not been properly solved in existing studies. They are as follows:

(1) Less attention has been focused on the mobility patterns at region level comparing to the mobility patterns at individual, stop or route level. In fact, the mobility patterns at region level can provide more macroscopic and intuitive knowledge about the travellers' collective moving characteristics with regard to the regions, which is vital to the urban planners and managers to conduct regional arrangement and control.

(2) Most of the previous studies analysed the stable mobility patterns of urban travellers, which are historically consistent in the near future. However, in a longer time window, the mobility patterns of urban travellers can be changed when the land use structure is altered. This may cause unexpected change and enhance the overloading risk of the urban system in some cases. So, an effective mobility pattern prediction method is needed to assess the long-term effect of changing the land use structure of a region.

To deal with these issues, a multi-step methodology using smart card data and points of interest data is proposed and implemented to analyze and predict regional mobility patterns of bus travellers. In this paper, we define the regional mobility patterns as the patterns of travellers' collective movements at region level, which indicate how people gather in or depart from the region. The following questions plaguing urban planners and managers can be answered by the proposed methodology: how many bus travellers will get in or get out of the region in a morning peak on Monday? Are there any common rules to follow? What will happen if the land use structure of the region is changed in the future? The study is of great value to provide a deep insight on the human mobility patterns at region level and can effectively support evidence-based and forward-looking urban planning and intelligent transportation management.

There are three primary contributions in our work:

- *Region Division:* an inner-restricted fuzzy c-means clustering method is proposed to properly divide the urban space into the meaningful regions by considering with the service coverage of bus stops;
- *Pattern Extraction:* we model the spatial-temporal mobility data (obtained from smart card data) with the three dimensional tensors and introduce the nonnegative tensor factorization method to extract the regional mobility patterns;
- *Prediction:* the artificial neural networks are built up by linking the regional properties (obtained from points of interest data) to the regional mobility patterns. In this way,

Manuscript received November 15, 2017; revised March 13, 2018 and May 16, 2018; accepted May 22, 2018. This work was supported by the National Natural Science Foundation of China under Grants 71621001, 91746201, and 71501013. The Associate Editor for this paper was N. Geroliminis. *(Corresponding author: Wei Guan.)*

The authors are with the MOE Key Laboratory of Urban Transportation Complex System Theory and Technology, Beijing Jiaotong University, Beijing 100044, China (e-mail: weig@bjtu.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TITS.2018.2840122

we achieve the prediction of regional mobility patterns based on the land use structure of the regions.

The rest of the paper is organized as follows: section II introduces the related works of the study; section III details the multi-step methodology for analyzing and predicting regional mobility patterns of bus travellers; section IV uses the empirical data of Beijing and analyses the application results of the proposed methodology; and section V gives conclusion to the paper and proposes the future work.

II. Related Works

Smart card data (SCD) [9] which record individual mobility information including passenger ID, boarding time, boarding stop ID, alighting time, alighting stop ID, are originally used in fare collection of bus transit and subway [10], and have been gradually accepted and widely used for data mining [11]–[17].

In the previous studies, a considerable amount of research has been performed on analysing the mobility patterns of bus travellers using smart card data. Some researches mainly focused on the temporal information of smart card data. For example, Agard *et al.* [18] divided the public transport users into four behavioral patterns according to the similar trip habits of travellers using K-mean clustering and Hierarchical Ascending Clustering (HAC) method on the boarding records (each boarding record is divided into 20 attributes informing whether the card holders travel in 5 weekdays and 4 periods per day). Medina [19] recognized the weekly mobility patterns by clustering a 14-dimension vector which is composed by start time and duration of mobility during 7 days in a week. In contrast, some researches put more efforts on analyzing the spatial information of smart card data. For example, Kieu *et al.* [20] proposed a new algorithm to detect the spatial travel pattern according to the number of repeated journeys from smart card data. However, the patterns discovered in these studies are incomplete, which only consider time or space in isolation from another.

To discover more comprehensive knowledge of mobility patterns, more researches took into account both of time and space characteristics. For example, Chen and Yang [21] studied the spatial-temporal characteristics of commuters travel and obtained commuters Origin-Destination (OD) distribution using public transportation routes data. Goulet-Langlois *et al.* [22] measured the regularity of individual travel patterns by considering the order in travel sequences. Chu and Chapleau [23] conducted a study on smart card boarding transactions and revealed the transfer patterns of travellers in their studies on transit demand modeling. Ma *et al.* [24] developed an effective data mining procedure for describing the travel patterns of transit riders and five clusters of regularity were obtained according to the number of travel days, number of similar first boarding times, number of similar route sequences, and number of similar stop ID sequences. Zhong *et al.* [25] measured the variability of mobility patterns using multiday smart card data and found out that mobility patterns varies from day to day. Tao *et al.* [26] demonstrated a multi-step methodology to analyse the aggregate flow patterns over space and time and visualised them by creating flow-comaps. Long and Thill [27] identified job-housing locations and analyzed commuting patterns in terms of duration and distance by combining smart card data and household travel survey. Similarly, Ma *et al.* [28] proposed a series of data mining methods to identify and understand commuting patterns using smart card data and confirmed the existence of job-house imbalance in the city. El Mahrsi *et al.* [29] extracted the mobility patterns at stop level and individual level by clustering smart card data. However, the aforementioned studies mainly concentrated on analyzing the mobility patterns at individual [22], [24], [27]–[29], stop [25], [29] or route level [21], [23], [26], and failed to discover more macroscopic knowledge at region level.

In view of this, Yu and He [30] divided the study area into heat maps with 1000*1000 grids and revealed the spatial-temporal patterns of bus travel demand at region level by extracting the features of the heat maps using principal component analysis. Nevertheless, there is still room for improvement: firstly, the regions were simply segmented by grids and failed to consider the aggregation of bus stops whose spatial distribution is discrete and inhomogeneous; secondly, the results of traditional principal component analysis are hard to interpret, because the value of principal components can be negative, which makes their physical meanings unclear.

Furthermore, even though some researches put little focus on the regional mobility patterns, their attempts on pattern extraction are still insightful. Gong *et al.* [31] have adopted an eigen decomposition method to capture the spatiotemporal patterns of passengers and uncovered the relationship between travel behaviur patterns and land use structures. Han and Moutarde [32] used nonnegative tensor factorization not only to enable pattern extraction of traffic flow but also to achieve prediction separately. Wang *et al.* [33] incorporated POI data into tensor decomposition model to further predict travel time of a path. These studies provide possible ways to extract the regional mobility patterns from the mobility data.

As alluded in most of the previous studies, the mobility patterns of urban travellers are historically consistent and the prediction work is not necessary in the near future. However, mobility patterns are always influenced by the land use structure [34]. In a longer time window, the mobility patterns of urban travellers can be changed when the land use structure is altered. Because the attractiveness of a region to the passengers could be changed when the functional properties of the region are altered, for example, removing a park from the region or adding a new office building in the region. The change in land use structure may happen spontaneously with the long-term urban development. Meanwhile, unbalanced urban development like job-house imbalance has also been found [28] and being solved by government through intentionally reorganizing the land use structure. In such conditions, the regional mobility patterns are no longer consistent with the past and need a new way to predict. Lee and Holme [34] proposed a prediction model of the human mobility patterns based on land-use maps. They separated the urban space into several regions according to the categories of land-use maps. Nevertheless, the functional properties of a region are always mixed and difficult to tell even with high-precision land-use

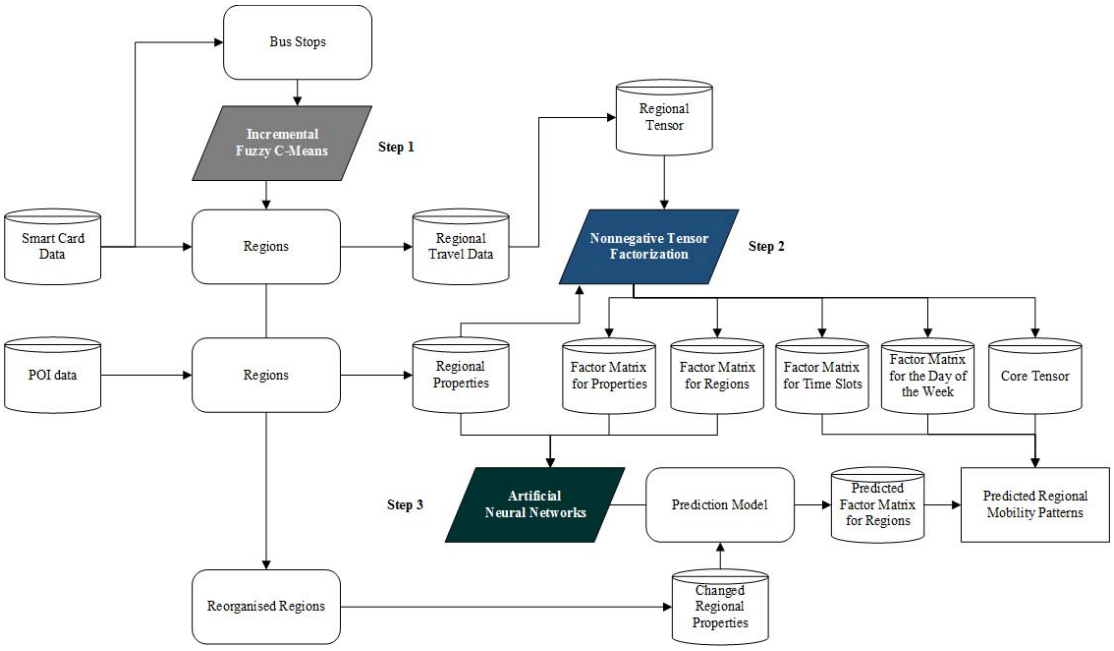


Fig. 1. The flow chart of three-step method for analyzing and predicting regional mobility patterns.

maps. For example, as a point in the land-use map, a high building may have multiple functional properties including companies, restaurants and supermarkets. With respect to the land-use maps, points of interest (POI) data, which include quantitative information of the regional properties such as the number of companies, hospitals and residences in the regions, can provide a more detailed and up-to-date land use structure information. However, to our knowledge, few studies have investigated the prediction of regional mobility patterns based on the regional properties provided by the POI data.

Overall, even though the previous studies has made great achievements in analysing mobility patterns of bus travellers from the perspective of behavioural type, regularity and OD demand, there are still some limitations:

i) regional mobility patterns of bus travellers, which can provide more macroscopic and intuitive knowledge beyond the bus network, has not been fully understood and analyzed;

ii) the extraction of regional mobility patterns is still defective, especially in dividing the regions and enhancing the interpretability of the patterns;

iii) the general approaches for predicting regional mobility patterns based on the detail regional properties have not been developed yet. In view of these limitations, a multi-step methodology is proposed and introduced in the following section.

III. Methodology

In this section, a multi-step methodology is proposed to analyze and predict the regional mobility patterns of bus travellers. There are three steps in the proposed methodology. The flow chart of three-step method for analyzing and predicting regional mobility patterns is shown in Fig.1. In the first step, a clustering method (Inner-restricted Fuzzy C-Means) is put forward to convert the analysis mode from stops to regions. In the second step, a regional tensor is created to gather historical travel information of the extracted regions; nonnegative tensor factorization is implemented on the created tensor to extract the core tensor and factor matrices for properties, regions, time slots and the day of the week, which are the expected parameters of the regional mobility patterns. In the third step, the regional properties, factor matrix for properties and factor matrix for regions are used to build up and train an artificial neural network; then, by combining the results of the nonnegative tensor factorization and the results of the established prediction model (predicted factor matrix for regions), the regional mobility patterns of bus travellers can be effectively predicted based on the regional properties.

A. Inner-Restricted Fuzzy C-Means

Boarding and alighting geographic information of smart card data are always at the stop level. There are two ways to convert them into the information at region level: (i) dividing the study area into grids with same size; (ii) clustering the bus stops. Apparently, the grid method is simpler than the latter one and the regional data can be directly obtained from stop information in the grids. However, natural groupings of points potentially represent more meaningful regions [35]. In this study, the spatial distribution of bus stops is always discrete and inhomogeneous. The grid method can hardly consider agglomeration characteristics of bus stops in practice, which may lead to unreasonable classification. Therefore, to extract proper regions for the following analysis, an Inner-restricted Fuzzy C-Means clustering method (IFCM) is proposed in this paper.

The well-known Fuzzy C-Means clustering method (FCM) [36], [37] has been widely accepted and applied in numerous applications due to its stable and excellent performance in the pattern recognition and data mining. By using

FCM, the target data can be clustered into several groups which have similar characteristics in feature space. The IFCM proposed in this paper, is composed by many times of FCM, which changes the cluster number iteratively with the limitation on clustering range.

The objective function of FCM used in the study can be written as

$$J = \sum_{k=1}^{n}\sum_{i=1}^{c}(u_{ki})^m ||x_k - v_i||^2 \tag{1}$$

in which $v_i (i = 1, 2, \ldots, c)$ are the centres of regions; $x_k (k = 1, 2, \ldots, n)$ are the positions of bus stop; u_{ki} $(i = 1, 2, \ldots, c; k = 1, 2, \ldots, n)$ are the membership degree of bus stop k with respect to region i; m is the fuzzy factor. The best choice of m is in the interval of [1.5, 2.5] and we choose $m = 2.0$ in this paper as many users preferred [38]. And u_{ki} $(i = 1, 2, \ldots, c; k = 1, 2, \ldots, n)$ should satisfy equation (2).

$$\sum_{i=1}^{c} u_{ki} = 1, \quad k = 1, 2, \ldots, n \tag{2}$$

To minimize the objective function under the constraint of (2), let the gradient of its LaGrangian equal to zero, and thereby get the following equations as the updated membership and centres. With random initial value of the region centres $v_i (i = 1, 2, \ldots, c)$, the updated membership and centres in (3) and (4) can be calculated iteratively until they converge to stable values.

$$u_{ki} = \frac{||x_k - v_i||^{\frac{2}{1-m}}}{\sum_{j=1}^{c} ||x_k - v_j||^{\frac{2}{1-m}}}, \quad k = 1, 2, \ldots, n \quad i = 1, 2, \ldots, c \tag{3}$$

$$v_i = \frac{\sum_{k=1}^{n} (u_{ik})^m x_k}{\sum_{k=1}^{n} (u_{ik})^m}, \quad i = 1, 2, \ldots, c \tag{4}$$

To convert the smart card data from stop level to region level, proper assembling method is needed. For a bus stop, the passenger proportion that reaches or leaves the different regions, are closely correlated with the membership degrees calculated from equation (3), which describe how much the bus stops belong to a region. Taking Fig.2 as an example, the passengers who get on or get off at the bus stop x_1 may come from or go to region v_1 with the proportion of $u_{11}/(u_{11} + u_{12} + u_{13})$. u_1, u_2 and u_3 are the membership degrees of bus stop x_1 with respect to region v_1, v_2 and v_3. Therefore, the boarding and alighting passenger number of the regions can be achieved according to equation (5). sN and rN represent the passenger number at the stop level and region level respectively.

$$rN_i = \sum_{k=1}^{n} [(u_{ki} \Big/ \sum_{j=1}^{c} u_{kj}) \cdot sN_k], \quad i = 1, 2, \ldots, c \tag{5}$$

Even though FCM is one of the most popular and effective techniques for data clustering, its practical performance is not always satisfactory. For example, in traditional FCM,

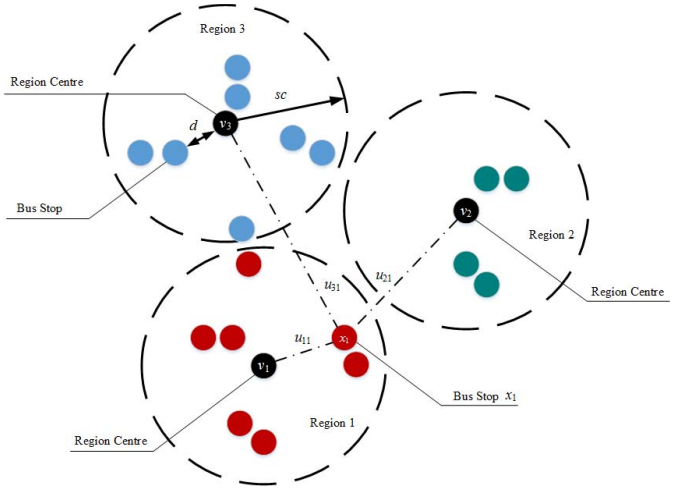


Fig. 2. The relationship between bus stops and regions.

the number of the clusters is unknown and should be determined manually according to different clustering requirements. Many research works have been carried out to automatically determine the number of the clusters [39]–[41]. Furthermore, many modified FCM algorithms have been proposed to deal with the other various problems occurred in practical application [42]–[47]. However, very few attempts have been made to restrict the influential range of clusters, which would affect the rationality of region extraction. Because, in the context of this paper, a region may include several bus stops which locate in different positions, while any bus stop in a region has its service coverage. Therefore, if the extracted region is represented as a clustered center, then the distance between the clustered center and the bus stops in the region should be limited. In other words, the distance between the clustered center and any bus stops in the region should be within an acceptable threshold, which is service coverage of bus stops. As shown in Fig.2, the distance (marked as d) between the region centers (marked as v_1, v_2 and v_3) and the bus stops should be smaller than the service coverage (marked as sc).

Compared with the traditional FCM, the IFCM proposed in this paper can automatically determine the number of the clusters, and more importantly, can restrict the influential range of the clusters. In IFCM, the number of the regions can be incrementally added if the range of the extracted regions is not within the predefined service coverage. The process of the IFCM is given as following:

(i) Initialize the cluster number $c = 2$, distance limitation sc and cluster centres $v_i^{(0)}$ $(i = 1, 2, \ldots, c)$.

(ii) Set fuzzy factor m, termination condition ε, iteration number $t = 1$ and maximum iteration number T.

(iii) Calculate the updated membership degrees $u_{ki}^{(t)}$ $(i = 1, 2, \ldots, c; \ k = 1, 2, \ldots, n)$ in iteration t according to equation (3).

(iv) Calculate the updated centres $v_i^{(t)}$ $(i = 1, 2, \ldots, c)$ in iteration t according to equation (4).

(v) If $\left\| v^{(t)} - v^{(t-1)} \right\| < \varepsilon$ or $t = T$, let $t_{\text{end}} = t$ and continue to next step; if not, let $t = t + 1$ and go back to step (iii).

(vi) Find out the maximum distance between the cluster centers and the data points belonged to the corresponding

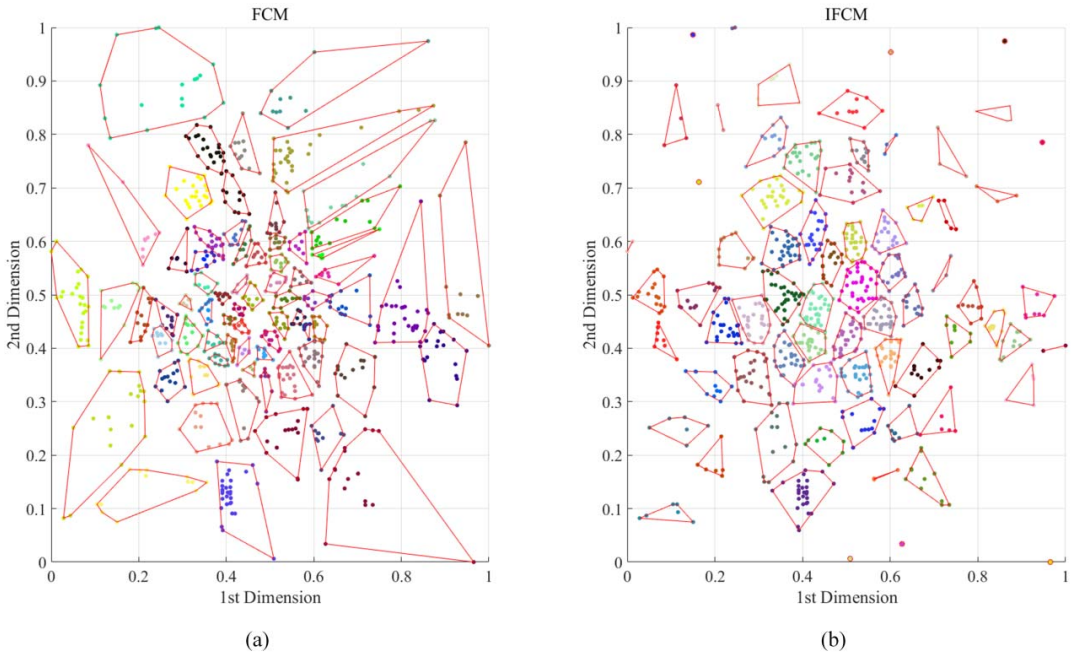


Fig. 3. The clustering results on a synthetic dataset.

clusters, according to equation (6). i_f and k_f are the sequence numbers of the cluster centres and data point with maximum distance.

$$[i_f, k_f] = \underset{i=1,2,\ldots,c;k=1,2,\ldots,n}{\arg\max} \times \left[I\left(i = \underset{i'=1,2,\ldots,c}{\arg\max}\ u_{ki'}\right) \cdot \left\| x_k - v_i^{(t_{end})} \right\| \right] \quad (6)$$

(vii) If $\left\| x_{k_f} - v_{i_f} \right\| < sc$, then exit and output the stable cluster centres $v_i^{(t_{\text{end}})}$ $(i = 1, 2, \ldots, c)$; if not, let $c = c + 1$, $v_i^{(0)} = v_i^{(t_{\text{end}})} (i = 1, 2, \ldots, c)$, choose the data point as the new added centre $(v_{c+1}^{(0)} = x_{k_f})$ and go back to step (ii).

The convergence of IFCM is proven as follows through referring to the previous studies on the convergence of the FCM family of algorithms [48]–[50]. Let

$$\mathbf{U} = \{u_{ki}; k = 1, 2, \ldots n; i = 1, 2, \ldots c\},$$
$$\mathbf{V} = \{v_i; i = 1, 2, \ldots c\}$$

Theorem 1: Let $\mathbf{V}$ be fixed, then $\mathbf{U}$ is a local minimum of J in equation (1) if and only if $\mathbf{U}$ is computed from equation (3).

Proof: The first derivative of the Lagrangian function of $\mathbf{U}$ confirms the only-if part. To prove it is sufficient, the Hessian of the Lagrangian function of $\mathbf{U}$ is inspected, as shown in equation (7).

$$h_{ij,ab}(\mathbf{U}) = \frac{\partial}{\partial u_{ij}} \left(\frac{\partial J(\mathbf{U})}{\partial u_{ab}} \right) = \begin{cases} \text{if } i = a, j = b\ m(m-1)\left((u_{ij})^{m-2} ||x_j - v_i||^2\right) \\ \text{otherwise } 0. \end{cases} \quad (7)$$

Since $m > 1$, $u_{ij} > 0$ and $||x_j - v_i||^2 > 0$, the Hessian of the Lagrangian function of $\mathbf{U}$ is positive definite. Thus, equation (3) is sufficient and can achieve the local minimum of the objective function.

Theorem 2: Let $\mathbf{U}$ be fixed, then $\mathbf{V}$ is a local minimum of J in equation (1) if and only if $\mathbf{V}$ is computed from equation (4).

Proof: The derivation of the Lagrangian function of $\mathbf{V}$ confirms the only-if part. Meanwhile, as the Lagrangian function of $\mathbf{V}$ is unconstrained and strictly convex function, the local minimum can be obtained through first derivative computation.

According to Theorems 1 and 2, it can be proved that the objective function is a decreasing function and the IFCM algorithm is locally convergent as it iteratively computes the membership degrees and the updated centres using equation (3) and (4) in step (iii) and (iv).

The clustering results on a synthetic dataset using traditional FCM and IFCM are shown in Fig.3(a) and Fig.3(b). The synthetic dataset is with 1000 points which are sampled from the bus stop dataset of Beijing after standardization. Both of the methods cluster the data into 70 regions. In traditional FCM, the initial cluster number is set to 70 and there is no limitation on the range of the clusters. In IFCM, the cluster number is incrementally increased from 2 and the service coverage is set to 0.085 in this example. By analyzing the clustering results, it can be seen that there are no clusters whose range exceeds the service coverage in IFCM, but 18 in traditional FCM. As we are looking for the regions which obey the service coverage of bus transit (as shown in Fig.2), the results of the proposed clustering method are more sensible and practical in this research. In the region division methods, the grid method and traditional FCM shoulder much lower computation burden than the IFCM. The computation time complexity of the IFCM is $O(c^2nt)$, where c is the number of cluster centers, n is the number of data and t is the iteration number. However, the

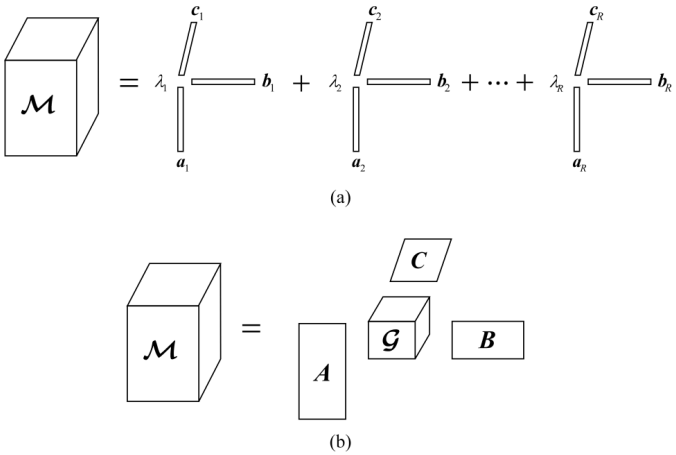


Fig. 4. Tensor decomposition methods: (a) $\boldsymbol{a}_r$, $\boldsymbol{b}_r$ and $\boldsymbol{c}_r$ $(r = 1, 2, \ldots, R)$ are the factorized vectors, which can compose the original tensor by using the weighted sum of their outer product; (b) $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{C}$ are the factor matrices, which can compose the original tensor by using their mode-n product with the core tensor.

grid method fails to consider the agglomeration characteristics of bus stops, and the traditional FCM ignores the service coverage limitation which may lead to improper clustering results. Comparatively, the IFCM can simultaneously take into account the agglomeration characteristics and service coverage of bus stops, which can achieve more natural and practical region division results. Thus, these advantages of the IFCM could probably justify its much larger computation time than other methods.

B. Nonnegative Tensor Factorization

Tensors are generalizations of matrices, and a tensor can be seen as a multi-dimensional array [51]. For instance, a 1st-order tensor is a vector and a 2nd-order tensor is a matrix. Nonnegative Tensor Factorization (NTF) is a generalization of Nonnegative Matrix Factorization (NMF) [52], which has many advantages such as strong interpretability (due to non-negative factors) and small storage space, and is widely used in the field of data mining [53].

According to the different research needs, NTF is divided into different decomposition methods, such as CP (CANDECOMP/PARAFAC) decomposition [54], [55] and Tucker decomposition [56]. As shown in Fig.4(a), CP decomposition decomposes original tensor into the sum of 1st-order tensor which can imply features of each dimension [57], and its objective function is given in equation (8). Tucker decomposition decomposes original tensor in the mode-n product form of core tensor and factor matrices as shown in Fig.4(b), and its objective function is given in equation (9). The main difference between two decomposition methods is the presence of core tensor in Tucker decomposition which allows every combination of column vectors of mode matrices [58]. Therefore, Tucker decomposition is utilized in this paper due to its flexibility and accuracy.

$$J = \frac{1}{2}||\boldsymbol{\mathcal{M}} - \sum_{r=1}^{R} \lambda_r \boldsymbol{a}_r \circ \boldsymbol{b}_r \circ \boldsymbol{c}_r||^2$$
$$\boldsymbol{\mathcal{M}} \in \mathbb{R}^{d_1 \times d_2 \times d_3}, \boldsymbol{a}_r \in \mathbb{R}^{d_1}, \boldsymbol{b}_r \in \mathbb{R}^{d_2}, \boldsymbol{c}_r \in \mathbb{R}^{d_3} \quad (8)$$

$$J = \frac{1}{2}||\boldsymbol{\mathcal{M}} - \boldsymbol{\mathcal{G}} \times_1 \boldsymbol{A} \times_2 \boldsymbol{B} \times_3 \boldsymbol{C}||^2$$
$$\boldsymbol{\mathcal{M}} \in \mathbb{R}^{d_1 \times d_2 \times d_3}, \boldsymbol{\mathcal{G}} \in \mathbb{R}^{r_1 \times r_2 \times r_3}, \boldsymbol{A} \in \mathbb{R}^{r_1 \times d_1},$$
$$\boldsymbol{B} \in \mathbb{R}^{r_2 \times d_2}, \boldsymbol{C} \in \mathbb{R}^{r_3 \times d_3} \quad (9)$$

In above equations, $\boldsymbol{\mathcal{M}}$ is the original third-order tensor; d_j $(j = 1, 2, 3)$ represent the dimensions of the original tensors; $\boldsymbol{a}_r$, $\boldsymbol{b}_r$ and $\boldsymbol{c}_r$ are the factorized three first-order tensors; R is a specified order; λ_i $(i = 1, 2, \ldots, R)$ are the normalized coefficients; $\boldsymbol{\mathcal{G}}$ is the core tensor with specified orders r_1, r_2 and r_3; $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{C}$ are the factor matrices; $\circ$ and $\times_n$ represent outer product and mode-n product respectively.

To compute the tucker decomposition effectively, many algorithms such as HOSVD [59] and HOOI [60] have been proposed by researchers. In this paper, a Block Coordinate Descent (BCD) method proposed by Xu and Yin [61] is adopted due to its outstanding performance in both speed and solution quality. In BCD method, if the optimization problem is given as equation (10), then the prox-linear updates with extrapolation at each iteration step can be written as (11).

$$\min_{\mathbf{x} \in \mathcal{X}} F(\mathbf{x}_1, \ldots, \mathbf{x}_p) = f(\mathbf{x}_1, \ldots, \mathbf{x}_p) + \sum_{i=1}^{p} e_i(\mathbf{x}_i) \quad (10)$$

where variable $\mathbf{x}$ is decomposed into p blocks $\mathbf{x}_1, \ldots, \mathbf{x}_p$; f is a differentiable function; $e_i (i = 1, \ldots, p)$ are the extended-value functions which can be used to include individual constraints of $\mathbf{x}_1, \ldots, \mathbf{x}_p$; if $\mathbf{x}_i$ break the constraint, then $e_i(\mathbf{x}_i) = \infty$.

$$\mathbf{x}_i^k = \arg\min \langle \hat{\boldsymbol{g}}_i^{k-1}, \mathbf{x}_i - \hat{\boldsymbol{x}}_i^{k-1} \rangle + \frac{L_i^{k-1}}{2} ||\mathbf{x}_i - \hat{\boldsymbol{x}}_i^{k-1}||^2 + e_i(\mathbf{x}_i) \quad (11)$$

where $\mathbf{x}_i^k$ denote the value of $\mathbf{x}_i$ in iteration k; $\hat{x}_i^{k-1}$ are the extrapolated values which are calculated by equation (12); $\hat{g}_i^{k-1} = \nabla f(\hat{x}_i^{k-1})$ is the block-partial gradient of f at $\hat{x}_i^{k-1}$ after $k-1$ updates; L_i^{k-1} is Lipschitz constant and can be assigned as L_2-norm of the second-order partial derivatives of $\mathbf{x}_i$; $\langle , \rangle$ represents the inner product and $\|\bullet\|$ calculates the L_2-norm.

$$\hat{\boldsymbol{x}}_i^{k-1} = \mathbf{x}_i^{k-1} + \omega_i^{k-1}(\mathbf{x}_i^{k-1} - \mathbf{x}_i^{k-2}) \quad (12)$$

where ω_i^{k-1} is the extrapolation weight and its calculation can be referred to [61].

In nonnegative tensor decomposition, the updated $\mathbf{x}_i^k$ can be further derived and written as equation (13).

$$\mathbf{x}_i^k = \max\left(0, \hat{\boldsymbol{x}}_i^{k-1} - \hat{\boldsymbol{g}}_i^{k-1}/L_i^{k-1}\right) \quad (13)$$

In our study, the regional tensors with the three dimensions respectively standing for regions, time slots and the days of the week are constructed. Two kinds of regional tensors considering the boarding and alighting activity can be built up. Each entry of the tensor $\boldsymbol{\mathcal{M}}_{ijk}$ denotes the boarding or alighting passenger number of the i-th region in the j-th time slot of the k-th day of the week. The entries of the regional tensors can be obtained by converting the mobility data at stop level to region level according to equation (5).

The optimization problem of the study is proposed as equation (14), which adds a collaborative function on the basis of the regional tensor decomposition problem. The collaborative function is to decompose the regional properties matrix $\boldsymbol{Y}$, which gives the properties of each region, into factor matrices for regions ($\boldsymbol{R}$) and properties ($\boldsymbol{F}$). The idea of using collaborative function is mainly inspired by research [33] which incorporated POI data into tensor decomposition model. In the previous research, the relationship between the time cost and the properties of the road segments was considered in the analysis. Comparatively, there are two main differences in our work. First, the research object is changed from analyzing time cost of the road segments to analyzing boarding/alighting passenger number of the regions. Therefore, the regional boarding/alighting tensor and the regional properties matrix are jointly decomposed by sharing same factor matrix for regions ($\boldsymbol{R}$) between $\boldsymbol{\mathcal{M}}$ and $\boldsymbol{Y}$, which is one of the original contributions of this study. Joint NTF and matrix factorization can be helpful in reducing the decomposition error and increasing the reliability of the extracted mobility patterns. Intuitively, this setting is also reasonable in reality, since the regions with similar properties are more likely to have similar mobility patterns. Second, the value of λ was manually specified in the previous research, which can hardly promise the proper decomposition results. In this paper, the value of λ controlling the contribution of the collaborative function is chosen by using a grid-search method, which can maximize the average reconstructing rate of NTF and matrix factorization.

$$\begin{aligned}\min f(\boldsymbol{\mathcal{G}}, \boldsymbol{R}, \boldsymbol{T}, \boldsymbol{D}, \boldsymbol{F}) = &\frac{1}{2}||\boldsymbol{\mathcal{M}} - \boldsymbol{\mathcal{G}} \times_1 \boldsymbol{R} \times_2 \boldsymbol{T} \times_3 \boldsymbol{D}||^2 \\ &+ \frac{\lambda}{2}||\boldsymbol{Y} - \boldsymbol{R}\boldsymbol{F}||^2 \\ \text{subject to } \boldsymbol{\mathcal{G}} \in \mathbb{R}_+^{J_1 \times J_2 \times J_3}&, \boldsymbol{R} \in \mathbb{R}_+^{I_R \times J_1}, \boldsymbol{T} \in \mathbb{R}_+^{I_T \times J_2}, \\ \boldsymbol{D} \in \mathbb{R}_+^{I_D \times J_3}&, \boldsymbol{F} \in \mathbb{R}_+^{J_1 \times I_F}\end{aligned} \tag{14}$$

where $\boldsymbol{\mathcal{M}} \in \mathbb{R}_+^{I_R \times I_T \times I_D}$ is the analysed regional tensor; J_1, J_2 and J_3 are the specified orders; $\boldsymbol{\mathcal{G}}$ is the decomposed core tensor; $\boldsymbol{R}$, $\boldsymbol{T}$ and $\boldsymbol{D}$ are the factor matrices for regions, time slots and the days of the week respectively; λ is the parameter controlling the contribution of the collaborative function; $\boldsymbol{Y} \in \mathbb{R}_+^{I_R \times I_F}$ is the regional properties matrix and further introduced in section III.C; $\boldsymbol{F}$ is the factor matrix for properties.

The given regional tensor is with very high dimensions and hard to interpret the mobility differences. Comparatively, the factor matrices for regions, time slots and the days of the week provide a more concise and meaningful way to understand the mobility. The column factors in the factor matrices for regions, time slots and the days of the week are respectively called region factors, time factors and day factors in this paper, which provide the concise forms of the coefficients of describing regional mobility patterns. Moreover, the decomposed factors are only associated with one related dimension in the factor matrix, which can help to analyze each dimension separately. For example, the region factors only describe the region-related differences in the mobility data. This knowledge can be used not only to specify the different types of regions, but also to incorporate the regional properties into the prediction of regional mobility patterns.

To solve the optimization problem in equation (14) using BCD method, the first-order partial derivatives of the unknown variables $\hat{\mathbf{g}}_i^{k-1} = \nabla f(\hat{\mathbf{x}}_i^{k-1})$ is calculated and the corresponding Lipschitz constant L_i^{k-1} is taken, according to equation from (15) to (24). See researches [61]–[63] for further details on matrix calculation.

$$\begin{aligned}\nabla_{\boldsymbol{\mathcal{G}}} f &= \boldsymbol{\mathcal{G}} \times_1 (\boldsymbol{R}^T\boldsymbol{R}) \times_2 (\boldsymbol{T}^T\boldsymbol{T}) \times_3 (\boldsymbol{D}^T\boldsymbol{D}) \\ &\quad - \boldsymbol{\mathcal{M}} \times_1 \boldsymbol{R}^T \times_2 \boldsymbol{T}^T \times_3 \boldsymbol{D}^T \qquad (15)\\ L_{\boldsymbol{\mathcal{G}}} &= ||\boldsymbol{R}^T\boldsymbol{R}|| \cdot ||\boldsymbol{T}^T\boldsymbol{T}|| \cdot ||\boldsymbol{D}^T\boldsymbol{D}|| \qquad (16)\\ \nabla_{\boldsymbol{R}} f &= \boldsymbol{R}\boldsymbol{G}_{(1)}\left((\boldsymbol{D}^T\boldsymbol{D}) \otimes (\boldsymbol{T}^T\boldsymbol{T})\right)\boldsymbol{G}_{(1)}^T \\ &\quad - \boldsymbol{M}_{(1)}\left[(\boldsymbol{D} \otimes \boldsymbol{T})\,\boldsymbol{G}_{(1)}^T\right] + \lambda\left[\boldsymbol{R}(\boldsymbol{F}\boldsymbol{F}^T) - \boldsymbol{Y}\boldsymbol{F}^T\right] \qquad (17)\\ L_{\boldsymbol{R}} &= ||\boldsymbol{G}_{(1)}\left((\boldsymbol{D}^T\boldsymbol{D}) \otimes (\boldsymbol{T}^T\boldsymbol{T})\right)\boldsymbol{G}_{(1)}^T|| + ||\lambda\boldsymbol{F}\boldsymbol{F}^T|| \qquad (18)\\ \nabla_{\boldsymbol{T}} f &= \boldsymbol{T}\boldsymbol{G}_{(2)}\left((\boldsymbol{D}^T\boldsymbol{D}) \otimes (\boldsymbol{R}^T\boldsymbol{R})\right)\boldsymbol{G}_{(2)}^T \\ &\quad - \boldsymbol{M}_{(2)}\left[(\boldsymbol{D} \otimes \boldsymbol{R})\,\boldsymbol{G}_{(2)}^T\right] \qquad (19)\\ L_{\boldsymbol{T}} &= ||\boldsymbol{G}_{(2)}\left((\boldsymbol{D}^T\boldsymbol{D}) \otimes (\boldsymbol{R}^T\boldsymbol{R})\right)\boldsymbol{G}_{(2)}^T|| \qquad (20)\\ \nabla_{\boldsymbol{D}} f &= \boldsymbol{D}\boldsymbol{G}_{(3)}\left((\boldsymbol{T}^T\boldsymbol{T}) \otimes (\boldsymbol{R}^T\boldsymbol{R})\right)\boldsymbol{G}_{(3)}^T \\ &\quad - \boldsymbol{M}_{(3)}\left[(\boldsymbol{T} \otimes \boldsymbol{R})\,\boldsymbol{G}_{(3)}^T\right] \qquad (21)\\ L_{\boldsymbol{D}} &= ||\boldsymbol{G}_{(3)}\left((\boldsymbol{T}^T\boldsymbol{T}) \otimes (\boldsymbol{R}^T\boldsymbol{R})\right)\boldsymbol{G}_{(3)}^T|| \qquad (22)\\ \nabla_{\boldsymbol{F}} f &= \lambda(\boldsymbol{R}^T\boldsymbol{R})\boldsymbol{F} - \boldsymbol{R}^T\boldsymbol{Y} \qquad (23)\\ L_{\boldsymbol{F}} &= ||\lambda\boldsymbol{R}^T\boldsymbol{R}|| \qquad (24)\end{aligned}$$

where $\boldsymbol{M}_{(n)}$ and $\boldsymbol{G}_{(n)}$ denote the mode-n matricization of $\boldsymbol{\mathcal{M}}$ and $\boldsymbol{\mathcal{G}}$; $\otimes$ represents the Kronecker product.

Algorithm 1 summarizes the process of nonnegative tensor factorization used in this paper. As the outputs of the algorithm, the core tensor ($\boldsymbol{\mathcal{G}}$) and the factor matrices for time slots and the days of the week ($\boldsymbol{T}$ and $\boldsymbol{D}$) are saved as the constant parameters, while the factor matrices for regions and properties ($\boldsymbol{R}$ and $\boldsymbol{F}$) are further used to train the neural networks introduced in section III.*C* for predicting the regional mobility patterns.

C. Artificial Neural Networks

Regional mobility patterns can be described by the parameters achieved in section III.*B* including the core tensor ($\boldsymbol{\mathcal{G}}$) and the factor matrices ($\boldsymbol{R}$, $\boldsymbol{T}$ and $\boldsymbol{D}$). However, in a longer time window, the regional mobility patterns can be changed when the land use structure is altered. It should be noted that the change of the land use structure is affected by the urban long-term dynamic mechanism which is always influenced by many factors, such as culture, policy, population growth, traffic network and so on. Building up a model explaining such complicated mechanism needs more data (maybe huge in a long-term) and more comprehensive knowledge, neither of which is available to this research. Therefore, instead of discussing the urban long-term dynamic mechanism, we tried to build up a link between land use structure (properties of the regions) and regional mobility patterns in this paper.

Algorithm 1 Algorithm for Solving (14)

1: **Input**: Nonnegative Tensor $\mathcal{M}$, nonnegative matrix $\boldsymbol{Y}$ and the specified orders $\boldsymbol{J}$
2: **Output**: Nonnegative core tensor $\mathcal{G}$ and factor matrices $\boldsymbol{R}$, $\boldsymbol{T}$, $\boldsymbol{D}$, $\boldsymbol{F}$
3: **Initialization**: randomize $\mathcal{G}^{-1} = \mathcal{G}^0$, $\boldsymbol{R}^{-1} = \boldsymbol{R}^0$, $\boldsymbol{T}^{-1} = \boldsymbol{T}^0$, $\boldsymbol{D}^{-1} = \boldsymbol{D}^0$, $\boldsymbol{F}^{-1} = \boldsymbol{F}^0$
4: **for** $k = 1, 2, \ldots$ **do**
5: Compute $\hat{\mathcal{G}}^{k-1}$, $\hat{\boldsymbol{R}}^{k-1}$, $\hat{\boldsymbol{T}}^{k-1}$, $\hat{\boldsymbol{D}}^{k-1}$, $\hat{\boldsymbol{F}}^{k-1}$ according to (11)
6: Update $\mathcal{G}^k$ according to (15), (16) and (13)
7: Update $\boldsymbol{R}^k$ according to (17), (18) and (13)
8: Update $\boldsymbol{T}^k$ according to (19), (20) and (13)
9: Update $\boldsymbol{D}^k$ according to (21), (22) and (13)
10: Update $\boldsymbol{F}^k$ according to (23), (24) and (13)
11: **if** $f(\mathcal{G}^k, \boldsymbol{R}^k, \boldsymbol{T}^k, \boldsymbol{D}^k, \boldsymbol{F}^k) > f(\mathcal{G}^{k-1}, \boldsymbol{R}^{k-1}, \boldsymbol{T}^{k-1}, \boldsymbol{D}^{k-1}, \boldsymbol{F}^{k-1})$ **then**
12: Reupdate $\hat{\mathcal{G}}^{k-1} = \mathcal{G}^{k-1}$, $\hat{\boldsymbol{R}}^{k-1} = \boldsymbol{R}^{k-1}$, $\hat{\boldsymbol{T}}^{k-1} = \boldsymbol{T}^{k-1}$, $\hat{\boldsymbol{D}}^{k-1} = \boldsymbol{D}^{k-1}$, $\hat{\boldsymbol{F}}^{k-1} = \boldsymbol{F}^{k-1}$
13: Go back to step 6
14: **end if**
15: **if** stopping criterion is satisfied **then**
16: Return $\mathcal{G}^k$, $\boldsymbol{R}^k$, $\boldsymbol{T}^k$, $\boldsymbol{D}^k$, $\boldsymbol{F}^k$
17: **end if**
18: **end for**

The most direct influence of the new land use structure is changing the factor matrix for regions ($\boldsymbol{R}$), which quantitatively describes the characteristics of the regions in regional mobility patterns. The linear prediction of the factor matrix for regions ($\boldsymbol{R}_l$) can be achieved according to equation (25) using the factor matrix for properties ($\boldsymbol{F}$) and the properties of the regions ($\boldsymbol{Y}$), but not enough to be effective in the nonlinear prediction.

$$\boldsymbol{R}_l = \boldsymbol{Y}\boldsymbol{F}^{-1}$$
$$\boldsymbol{R}_l \in \mathbb{R}_+^{I_R \times J_1}, \boldsymbol{Y} \in \mathbb{R}_+^{I_R \times I_F}, \boldsymbol{F} \in \mathbb{R}_+^{J_1 \times I_F} \tag{25}$$

where I_R is the number of the regions; I_F is the dimension of properties; J_1 is the specified order; $\boldsymbol{F}^{-1}$ represents the pseudo inverse of the factor matrix for properties.

Artificial Neural Networks (ANN) is widely used in various fields, such as image recognition [64], classification [65], automatic control [66], numerical prediction [67] and so on. Furthermore, artificial neural networks have also been used in the traffic speed forecasting, traffic flow prediction and traffic signal control [68]–[70]. Due to strong nonlinear fitting ability and simple learning rules, the multilayer perceptron model [71], which is a class of feedforward artificial neural network, is utilized for predicting the regional mobility patterns in the third step of the proposed methodology. The multilayer perceptron model consists of three or more layers, which include input layer, output layer and one or more hidden layers. More details about the model can be found in [72].

In the training process of ANN models, the weight values linking the artificial neurons in different layers are adjusted to reduce the overall prediction error. As an important method for training the multilayer perceptron model, Backpropagation (BP) algorithm [73], which updates the value of the weights by iteratively computing the gradients of each layer based on the loss function, is always used. The loss function of the ANN model (which represents the multilayer perceptron model in the following text) is given in equation (26). The regularization term in the cost function can help to avoid overfitting and obtain better generalization by minimizing the weight magnitude. In the training process of BP algorithm, the change in weight is the product of the learning rate and the gradient, as shown in equation (27). The negative sign in equation (27) is to update the weights in the direction of minimizing the loss function. The gradient, which is the partial derivative of the loss function with respect to the weight, can be achieved by using chain rules, when the corresponding neuron units are not directly connected with output units. The updated function of the weights is shown in (28). It should be noted that, in order to achieve better training effect, the method of momentum [74] can be considered, which makes the weight adjustment depend on both of the current and last change of the weight. By doing so, the BP algorithm which uses gradient descent method, can gain faster convergence and reduced oscillation [75]. When the total error between the predicted outputs and correct outputs reaches the threshold value, or the number of training iterations exceeds the preset value, the BP algorithm will end the training of the ANN model.

$$E(\mathbf{w}) = \frac{1}{2}\sum_{k=1}^{n}(o_k - y_k)^2 + \gamma\frac{1}{2}\|\mathbf{w}\| \tag{26}$$

$$\Delta w_{ij} = -\eta g_{ij} = -\eta\frac{\partial E(\mathbf{w})}{\partial w_{ij}} \tag{27}$$

$$w_{ij}^{(t+1)} = w_{ij}^{(t)} + \Delta w_{ij}^{(t)} + \alpha\Delta w_{ij}^{(t-1)} \tag{28}$$

in which o_k $(k = 1, \ldots, n)$ and y_k $(k = 1, \ldots, n)$ are the predicted output value and the correct output value respectively; γ is the regularization parameter; $E(\mathbf{w})$ denote the loss function; w_{ij} and g_{ij} respectively represent the linked weights and gradients between neuron unit i and j; η is the learning rate and α is the momentum coefficient.

In this paper, we construct the artificial neural networks with multiple inputs, multiple outputs and one hidden layer. The outputs of the ANN model are the updated factor matrix for regions ($\boldsymbol{R}_u$), and the inputs of the ANN model include the properties of the regions ($\boldsymbol{Y}$) and linear prediction of the factor matrix for regions ($\boldsymbol{R}_l$) according to equation (25). In the training process of the ANN models, the output data of the training dataset are provided by the extracted factor matrix for regions in section III.B, while the input data are not directly given. The points of interest (POI) data collected from websites, twitters and smart phones, are employed as the influencing factors of regional mobility patterns in this study. Thus, the properties of a region are defined as the collection of the numbers of POI categories. As shown in Fig.5, different POI categories, such as company, school, park, hospital, entertainment and so on, are always spatially distributed in different locations. To obtain the properties of the regions, the points of interest data at point level have to be converted to the property information at region level. The properties of

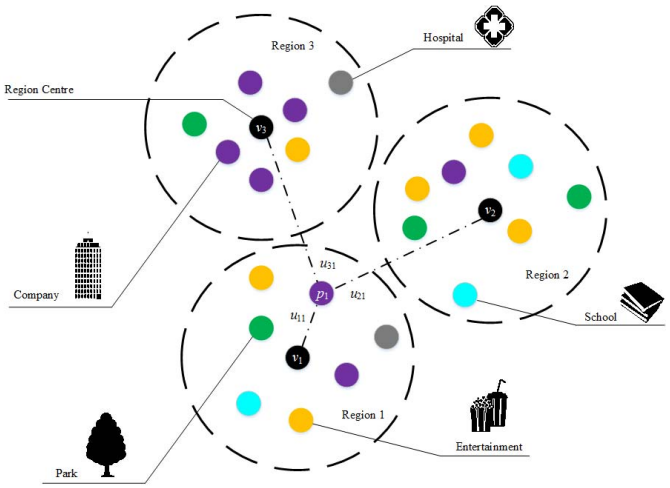


Fig. 5. POI categories in the regions.

the regions ($\boldsymbol{Y}$) can be collectively calculated from the points of interest data according to equation (29). It can be seen that the property of a points of interest is partially assigned to different regions according to the assigned weights of the point with respect to the regions. The assigned weights are related with the relative locations of points of interest data with respect to the region centers. More precisely, the assigned weight is larger or smaller when the point of interest is closer to or farther from the center of the region. The centers of the extracted regions have been achieved in section III.*A*, and the assigned weight of each point with respect to each region can be obtained according to equation (3).

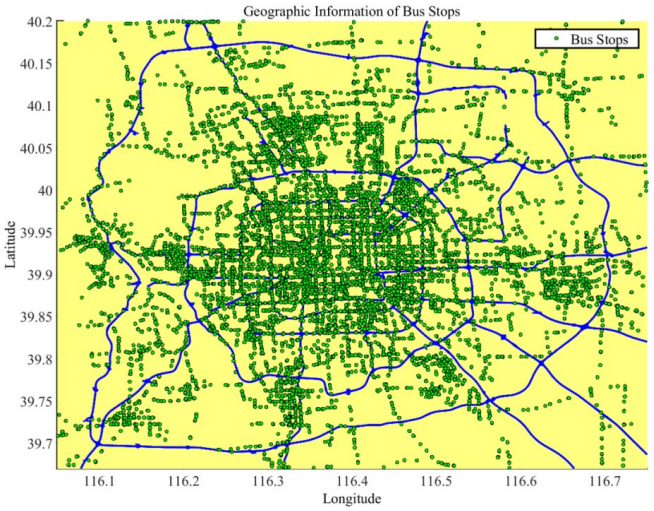
$$\boldsymbol{Y} = \left\{ Y_{ik} = \sum_{j=1}^{n} u_{ij} I(\varphi(p_j) = k), \right. \\ \left. k = 1, 2 \ldots, K; i = 1, 2 \ldots, c \right\} \quad (29)$$

in which $\boldsymbol{Y}$ are the properties of the regions; u_{ij} is the assigned weight of point p_j with respect to region i; c, n and K are the number of regions, points of interest data and POI categories respectively; function $\varphi(p_j)$ returns the category of point p_j; and only if the value of $\varphi(p_j)$ equals the category k, then the value of $I(\varphi(\mathrm{p_j}) = k)$ equals 1, otherwise equals 0.

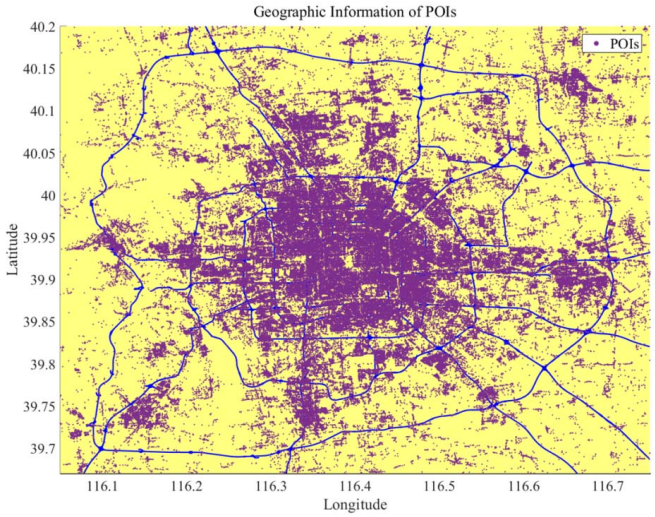
As the third step of the proposed methodology in this paper, the ANN model is established and trained for predicting the factor matrix for regions based on the regional properties. At last, by combining the results of the NTF model and the results of the ANN model, the parameters describing the regional mobility patterns, including the core tensor ($\mathcal{G}$), factor matrix for time ($\boldsymbol{T}$), factor matrix for the day of the week ($\boldsymbol{D}$) and the factor matrix for regions ($\boldsymbol{R}_u$), can be obtained, and the prediction of the regional mobility patterns of bus travellers based on the regional properties can be achieved.

IV. Results and Analysis

In order to prove the feasibility of the proposed methods, the bus smart card data within the sixth ring road of Beijing city from February 6th, 2017 to February 19th, 2017 are employed for analysis. The transfers in the bus smart card



(a)



(b)

Fig. 6. Geographic information of bus stops and POIs.

data have been pre-processed, so each record of the bus smart card data can provide the origin and destination of one journey. The total data records exceed 40 million, involve over 40 thousand bus stops and cover an area of 2267 square kilometres. Besides, the POI dataset of Beijing is also used for providing the basic information about the properties of the regions. The geographic information of bus stops and POIs are shown in Fig.6(a) and Fig.6(b). The distribution of journeys in the whole urban space during the week from February 6 to 12 is given in Fig.7 based on the simple statistical analysis, in which the numbers of the journeys happened in each time slot (half an hour) are accumulated. It can be seen that the mobility patterns of bus travellers are different between weekdays and weekends in the whole urban space. Meanwhile, two peaks can be observed respectively in the morning and evening.

The enriched data give us the opportunity to extract and analyze the regional mobility patterns using the proposed multi-step methodology. First of all, the IFCM method is applied on

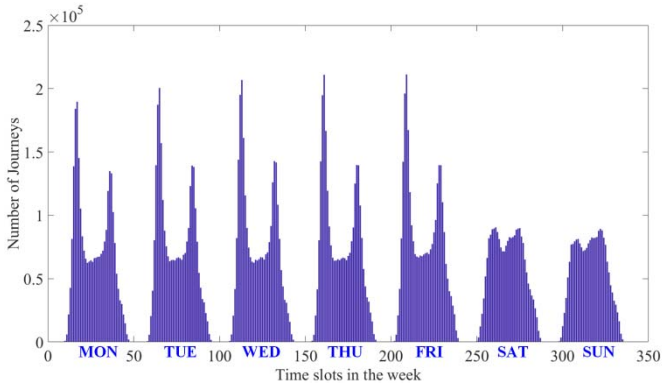


Fig. 7. Distribution of journeys during the week from February 6 to 12.

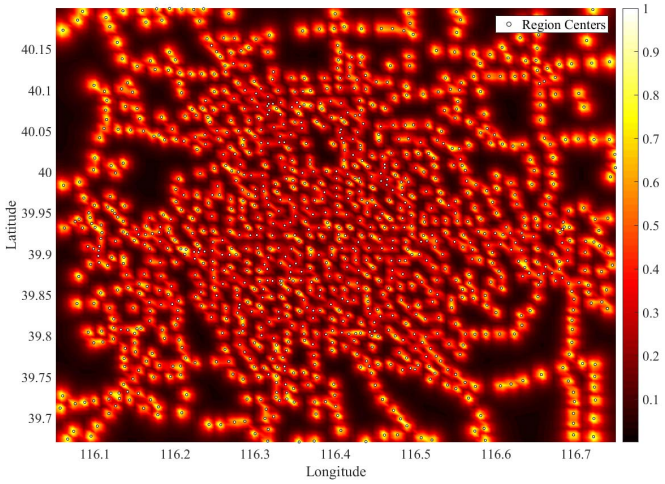


Fig. 8. Extraction of regions using IFCM.

the geographic information of bus stops. In Transit Capacity and Quality of Service Manual (TCQSM), the service coverage of bus stop is defined as 800 metres [76]. Nevertheless, with wide spread use of e-bike share in Beijing, the range and accessibility of passengers are increased and the range could be from 25% to 100% greater than private bike [77]. For this reason, the service coverage of bus stop is set to 1000 metres (800 + 800*25%) in this paper. As it is shown in Fig.8, based on IFCM, 1110 valid regions are extracted and brighter colors represent the range of the regions with larger membership degree.

Bus smart card data can provide the time of the activity and bus stop ID with respect to each individual. By aggregating the individual data, the passenger number at specific bus stop during specific time period can be obtained. Using the membership degree of each stop with respect to each region obtained from IFCM, the boarding and alighting passenger number of the extracted regions can be calculated based on the boarding and alighting passenger number at bus stops according to equation (5). Then, two third-order regional tensors for the boarding and alighting activity, whose three dimensions stand for the regions, time slots and the day of the week, can be built up. The entry of the regional tensors represents the passenger number get into or get out of the region in the specific time slots and the day of the week. The sizes of the tensors are set to $\mathbf{1100} \times \mathbf{34} \times \mathbf{7}$ (**1100** extracted regions in IFCM, **34** half-hour time slots from 6 AM to 11 PM, **7** days of the week).

Subsequently, the nonnegative tensor factorization method, which uses the decomposition algorithm 1 given in section III.B, is implemented on the constructed regional boarding and alighting tensor. In this paper, we take $J_1 = J_2 = J_3 = 5$ as the specified orders for both of the tensors, and $\lambda = 4.96 \times 10^6$ (boarding), 4.72×10^6 (alighting) as the controlling parameters in the decomposition process (the large value is due to the different dimensions involved in NTF and matrix factorization). After computation, the decomposition results can reconstruct 88.85% and 89.28% of the regional boarding and alighting tensors respectively. The decomposed factor matrix for time slots ($\boldsymbol{T} \in \boldsymbol{R}_{+}^{34\times5}$), factor matrix for the day of the week ($\boldsymbol{D} \in \boldsymbol{R}_{+}^{7\times5}$) and the core tensor ($\mathcal{G} \in \boldsymbol{R}_{+}^{5\times5\times5}$) are shown in Fig.9 and Fig.10, for two tensors. The larger value in the factor matrices suggests that the analysed time slots or the day of the week has a greater impact in the corresponding factor. For example, as shown in Fig.9(a), the time slots from 7:30 AM to 8:30 AM have a greater impact to time factor 4 of the regional boarding tensor. It means that the boarding activities are more likely to happen in these time slots in time factor 4. Similarly, as shown in Fig.9(b), Saturday and Sunday have a greater impact to day factor 3 of the regional boarding tensor. It means that many boarding activities tend to happen on weekends in day factor 3. Furthermore, the extracted core tensor can depict the relationship between the three dimensions. As shown in Fig.9(c), with respect to region factor 2, time factor 4 is more related with other day factors except day factor 3. It indicates that for region factor 2, time factor 4 (boarding activities from 7:30 AM to 8:30 AM) and day factor 3 (boarding activities in weekends) are unlikely to jointly happen. The decomposed results of the regional alighting tensor shown in Fig.10 can also be analysed in this way, so no more examples are given here.

The achieved factor matrix for time slots $\boldsymbol{T}$, factor matrix for the day of the week $\boldsymbol{D}$ and core tensor $\mathcal{G}$ are saved as the constant parameters of the regional mobility patterns. These constant parameters can be used to build up a constant tensor $\mathcal{H} \in \boldsymbol{R}_{+}^{5\times34\times7}$ according to equation (30). The constant tensor $\mathcal{H}$ can provide the relationship between the region factors and temporal characteristics including time slots and the day of the week, which can provide more intuitive understanding of the region factors. As shown in Fig.11, larger value of the constant tensor denotes that more passengers get into or get out of the regions with the corresponding region factor at the specified time slot and the day of the week. With the help of the constant tensor, the semantic meanings of the region factors can be further inferred. For example, as region factor 4 in the boarding tensor are more related with the time slots from 6:00 AM to 8:00 AM in weekdays, the regions with larger region factor 4 are more likely to be the home of the commuting passengers who get on the bus for commuting in the early morning of the weekdays. Similarly, region factor 5 in the alighting tensor are more related with the time slots from 6:00 AM to 8:00 AM

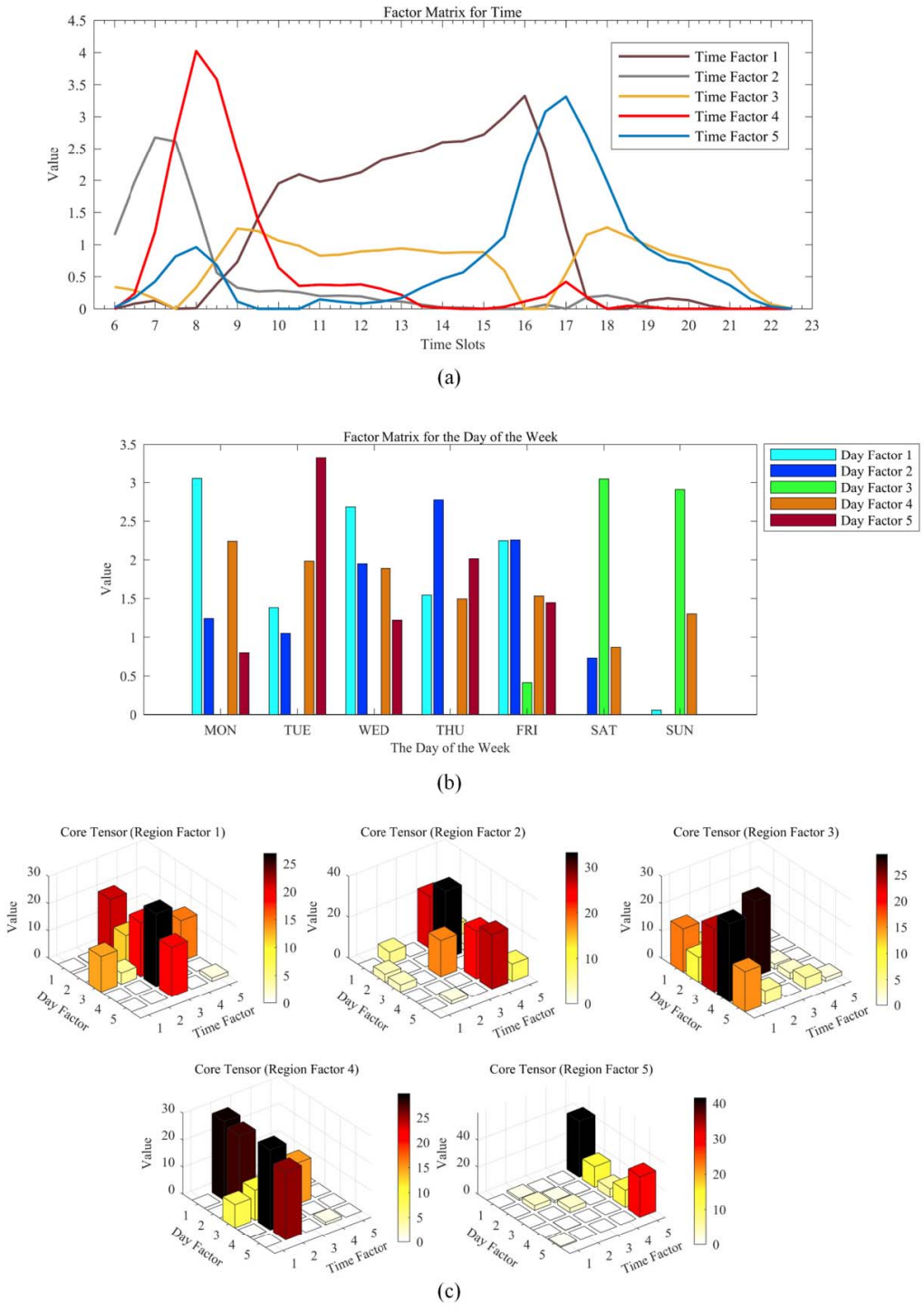


Fig. 9. Decomposition results of regional boarding tensor.

in weekdays, and thus the regions with larger region factor 5 are more likely to be the working place of the commuting passengers who get off the bus after commuting in the early morning of the weekdays.

$$\mathcal{H} = \mathcal{G} \times_2 \boldsymbol{T} \times_3 \boldsymbol{D} \tag{30}$$

The factor matrix for regions $\boldsymbol{R}$ is also obtained after nonnegative tensor factorization. The factor matrix for regions $\boldsymbol{R} \in \mathbb{R}_+^{1110\times 5}$ denotes the distributions of the region factors on the extracted regions. Each column of $\boldsymbol{R}$ represents the scores of the specific region factor among all of the regions. For the studies using heat maps, we need to review all heat maps in different time slots and different days, and many comparisons and computations are needed to extract the hidden patterns. By applying the nonnegative tensor factorization method, the region factors only describe the region-related differences in the mobility data, which can largely lighten the burden of understanding the differences among the regions. Also,

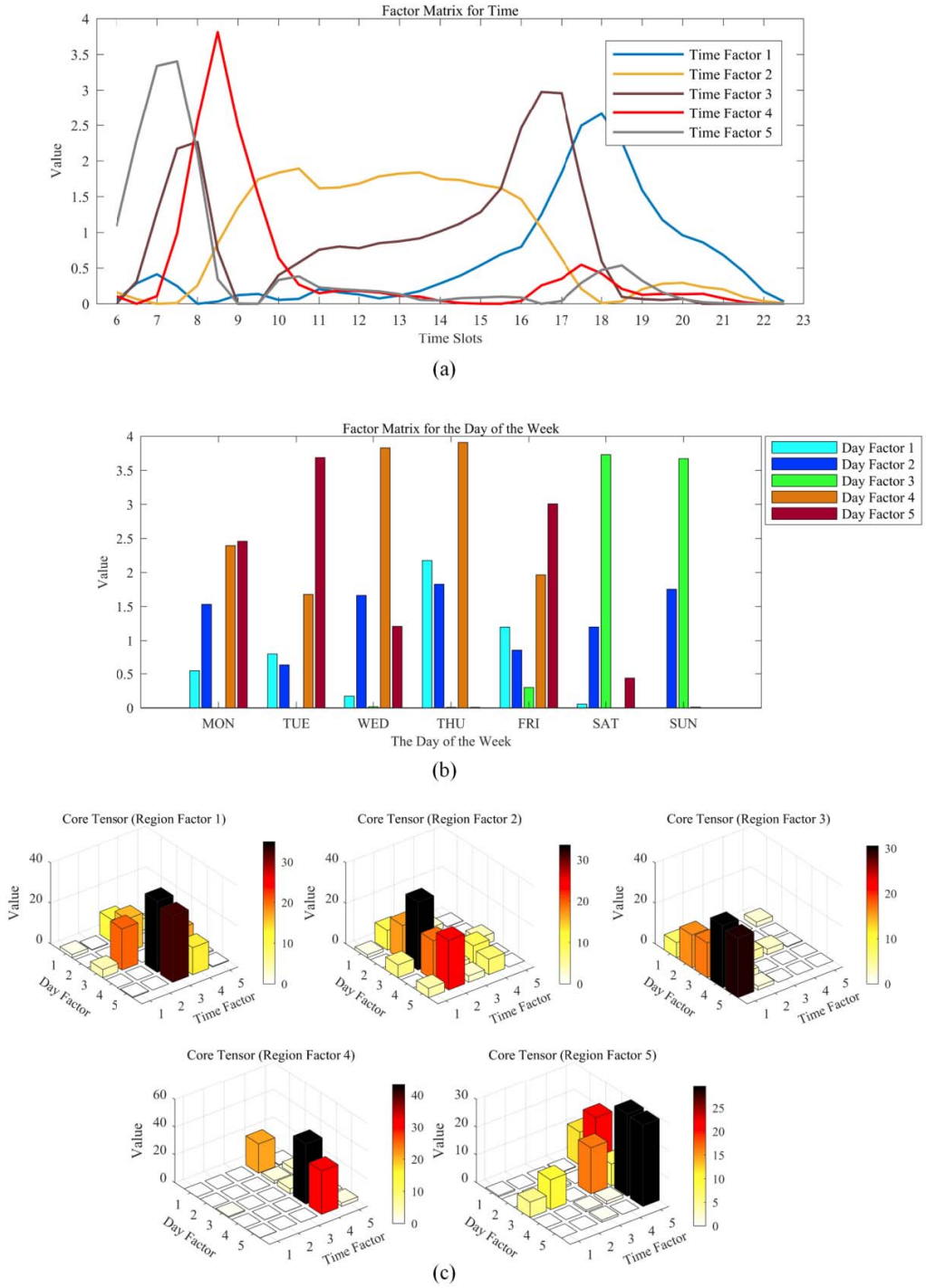


Fig. 10. Decomposition results of regional alighting tensor.

the region factors are more appropriate to incorporate with the regional properties for predicting regional mobility patterns.

Due to the length limitation, only the region factor 4 in the boarding tensor and region factor 5 in the alighting tensor, are shown in Fig.12(a) and Fig.12(b). The centers of the circles represent the centers of the extracted regions and larger circle means higher score of the region factor on the region. According to the detailed temporal mobility characteristics corresponding to each region factor shown in Fig.11(a) and Fig.11(b), the regions with larger region factor 4 in the boarding tensor can be seen as the origins of the commuting passengers and the regions with larger region factor 5 in the alighting tensor can be deemed as the destinations of the commuting passengers. It can be seen that most of the early commuting origins are distributed throughout the city, while the early commuting destinations are more concentrated in the city center. Such spatial knowledge discovered from the analysis of the factor matrix for regions can be further

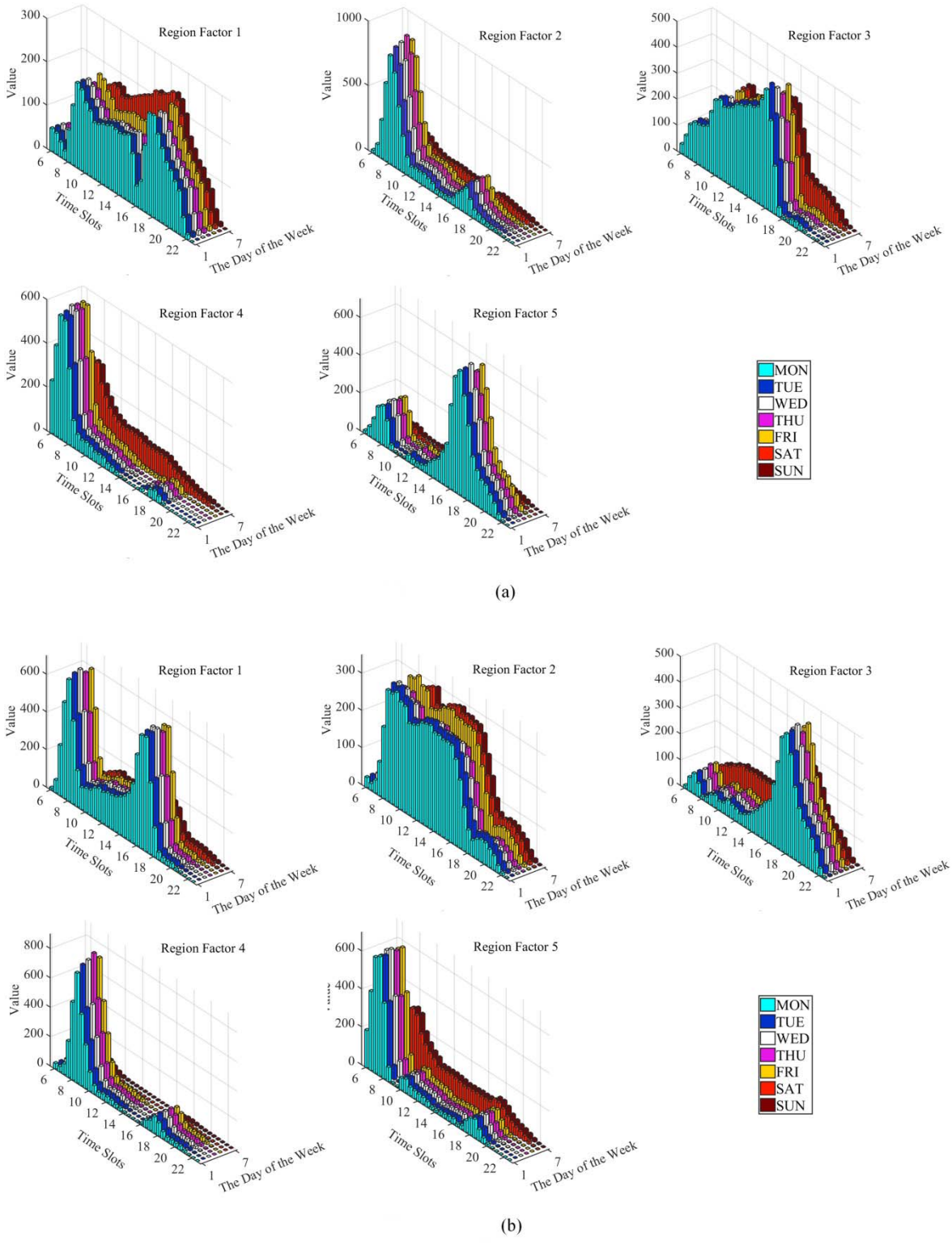


Fig. 11. Constant tensor for boarding and alighting. (a) Constant tensor for boarding. (b) Constant tensor for alighting.

used to analyze the differences among the regions and assess the current urban state, for example, evaluating the job-house imbalance problem of the city.

Moreover, the relationship between the region factors and regional properties can also be found. In this paper, the regional properties are represented as the vector which collects the numbers of POI categories in the region. The POI data has 18 main categories and 122 subcategories. In this paper, we only use the main categories including school, hospital, company, government, residence, entertainment, attraction and so on. Staying with the aforementioned examples, the regional properties of the regions with larger region

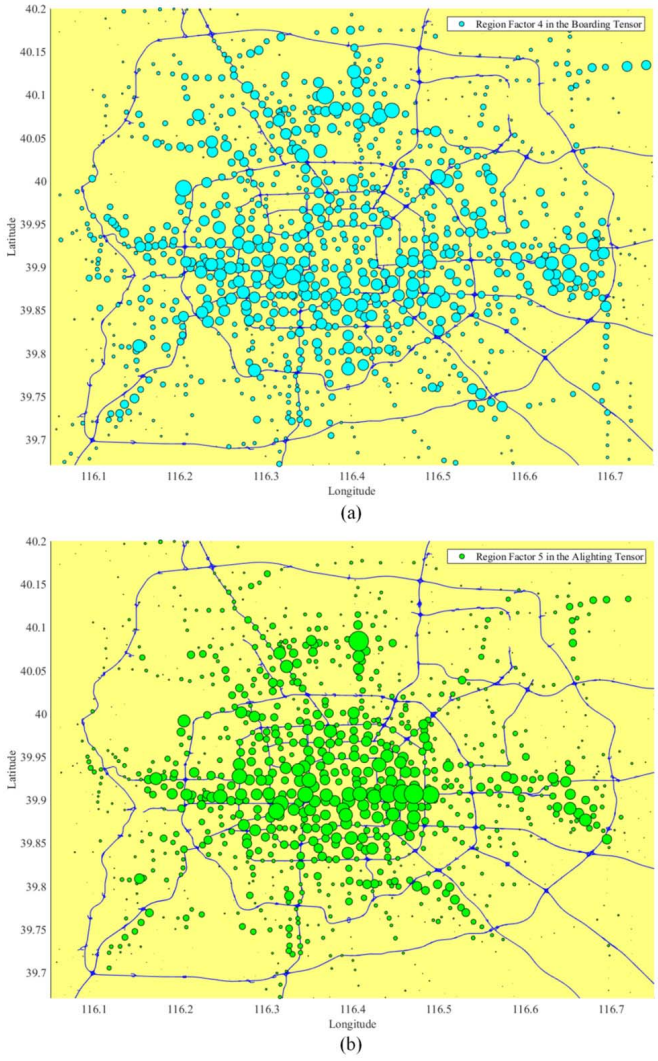


Fig. 12. Factor matrix for regions.

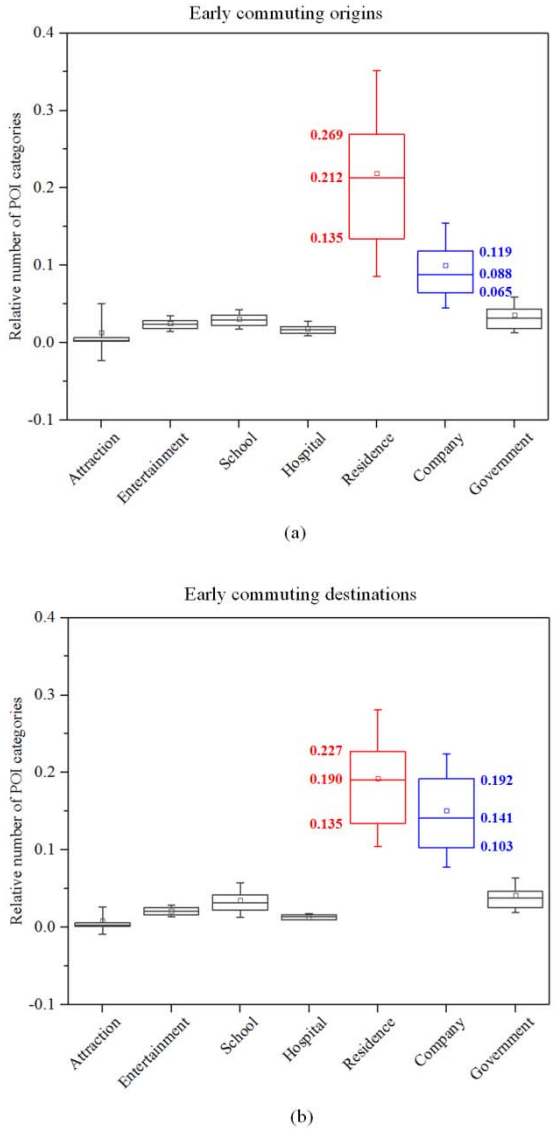


Fig. 13. The regional properties of the early commuting origins and destinations.

factor 4 in the boarding tensor (early commuting origins) and larger region factor 5 in the alighting tensor (early commuting destinations) are respectively presented in Fig.13(a) and Fig.13(b). For clarity, only a few POI categories are shown on the horizontal ordinate and the ordinate gives the relative number of POI categories (number of the POI category in the region /total number of the POI in the region). By comparing the median values in Fig.13, it can be seen that early commuting destinations have 11.6% less residences and 60.2% more companies than early commuting origins.

If the properties of the analyzed regions are fixed, the achieved factor matrix for regions $\boldsymbol{R}$ is historically consistent and the regional mobility patterns can be directly achieved from the results of nonnegative tensor factorization. However, in a longer time window, the regional properties can be changed due to the self development of the city and intentional reorganization of the government. Meanwhile, according to the analysis in Fig.13, it is noted that the regional properties influence the boarding and alighting activities differently. Therefore, two ANN models with respect to the boarding and alighting activities are respectively constructed for predicting the factor matrix for regions based on the regional properties. Both of the two ANN models are with one hidden layer. There are 39 (boarding) and 30 (alighting) hidden neuron units for the two ANN models. The number of the hidden neuron units is determined by looking for the minimum overall error from the loop computation with 2:50 hidden neuron units. Empirically, we take $\eta = 0.1$ as the learning rate and $\alpha = 0.9$ as the momentum coefficient for the ANN models. Readers can refer to section III.*C* for more details about the ANN models. Actually, the prediction using the observed data in the future can be more helpful and meaningful to the analysis. However, neither new properties of the regions nor the mobility data in the future are available to the current research. Thus, we separated 90% of the current data for training the model and 10% of the current data for testing the model. The testing data can be deemed as the

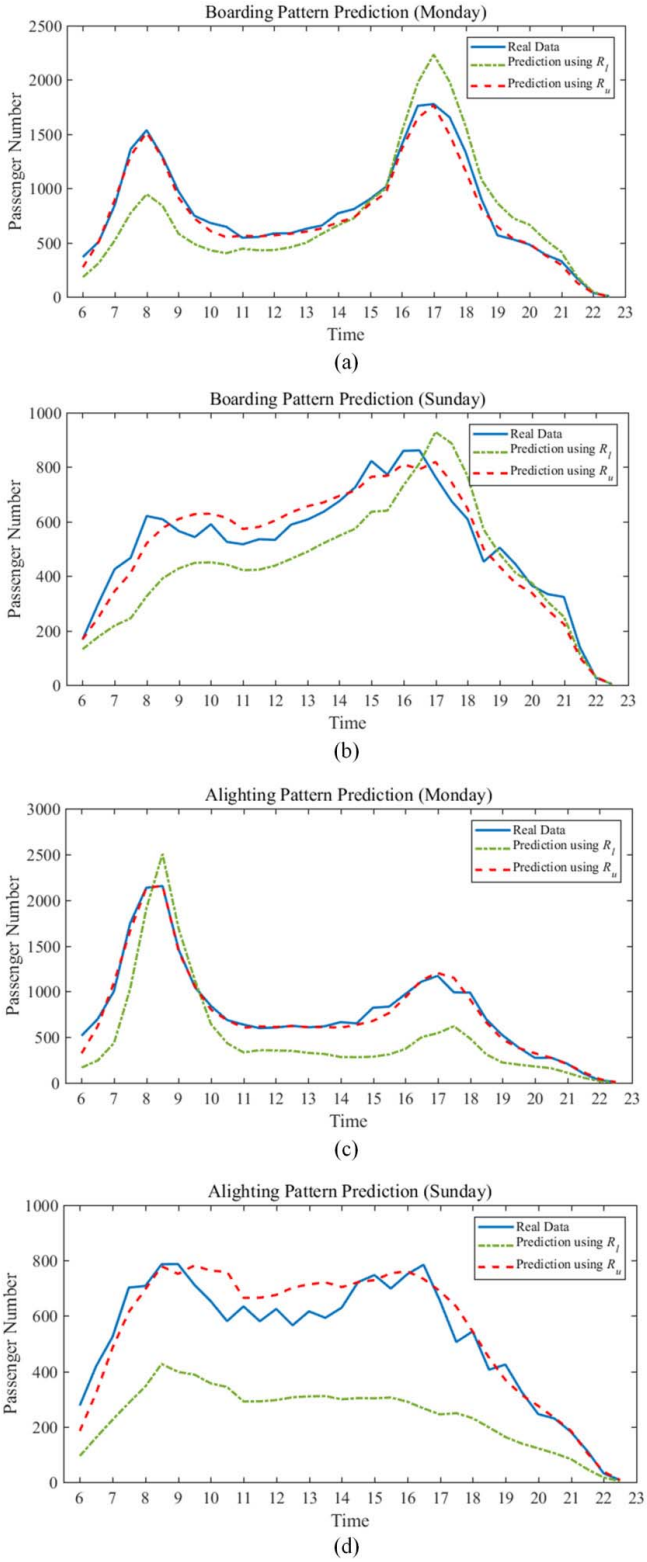


Fig. 14. Predictions of the boarding and alighting passenger number with regard to a testing region.

extreme cases in the future, in which the properties of the region are changed from nothing to current values. Through verification, the overall mean square errors for two ANN models are 0.00225 (boarding) and 0.00204 (alighting). Then,

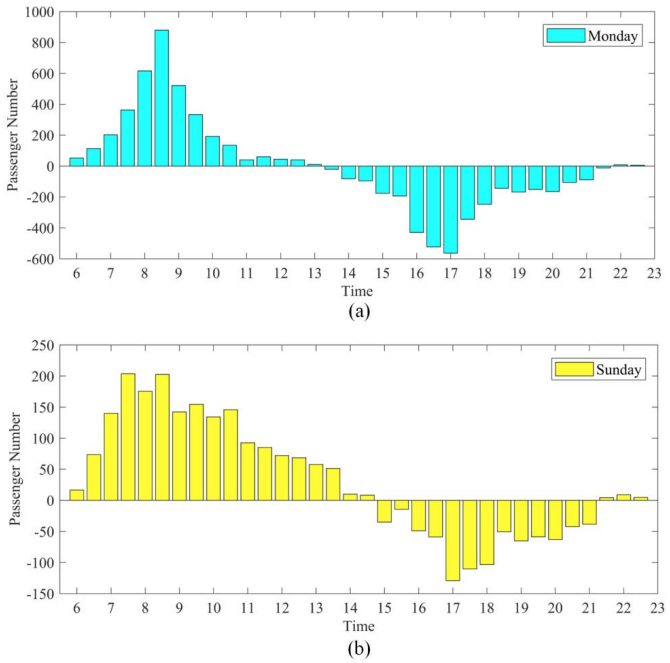


Fig. 15. Predictions of the regional aggregation pattern.

by combining the predicted factor matrix for regions $\boldsymbol{R}_u$ and the constant tensor $\mathcal{H}$ (obtained in equation (30)) according to equation (31), the boarding and alighting passenger number in the regions at any time slots of any day of the week, can be determined. For example, the predictions of the boarding and alighting passenger number with regard to a testing region on Monday and Sunday are shown in Fig.14. It can be seen that the prediction using $\boldsymbol{R}_u$ (ANN models) has a better performance compared with the prediction using $\boldsymbol{R}_l$ (linear model prediction according to equation (25)). T-test method and F-test method are used to evaluate the effectiveness of the predictions. For the real data and the linear prediction using $\boldsymbol{R}_l$, the T-test result ($p = 0.0008, \alpha = 0.05$) rejects the null hypothesis and the F-test ($p = 0.281, \alpha = 0.05$) result fails to reject the null hypothesis with small p-value. This can indicate that the predictions of the linear model using $\boldsymbol{R}_l$ are with lower accuracy. Comparatively, with respect to the real data and the prediction using $\boldsymbol{R}_u$ (outputs from ANN model), both of the T-test result ($p = 0.859, \alpha = 0.05$) and F-test ($p = 0.862, \alpha = 0.05$) result fail to reject the null hypothesis. It means that no significant differences exist between the real data and the prediction using $\boldsymbol{R}_u$. Therefore, it can be proved that the proposed methodology has a better performance on predicting the regional mobility patterns based on the regional properties.

$$\mathcal{M}_p = \mathcal{H} \times_1 \boldsymbol{R}_u \tag{31}$$

in which $\mathcal{M}_p$ represents the predicted regional tensor; $\mathcal{H}$ is the constant tensor; $\boldsymbol{R}_u$ is the predicted factor matrix for regions.

Furthermore, by providing the increasing and decreasing characteristics of travellers' number at region level, the regional alighting and boarding patterns can further

indicate the travellers' aggregation patterns in the regions. By subtracting the boarding passenger number from the alighting passenger number according to equation (32), the passenger number gathered in the region can also be achieved, as shown in Fig.15. The larger positive value of the gathered passenger number means that there are more passengers getting into the region, while larger negative value means that there are more passengers leaving the region. These analysis and prediction could be helpful to the future urban planning and management facing to the regions without enough data, such as newly-built urban area and reorganized regions with new land use structure.

$$P_r = P_a - P_b \tag{32}$$

in which P_r is the passenger number gathered in the region; P_a and P_b are the regional alighting and boarding passenger number respectively.

V. Conclusions

To analyse and predict the regional mobility patterns of bus travellers, the multi-step methodology using bus smart card data and points of interest data, which incorporates the inner-restricted fuzzy C-means clustering, nonnegative tensor factorization and artificial neural network, is proposed in this paper. Rather than analyzing the mobility patterns at individual, stop or route level [21]–[29], this study puts more focus on the regional mobility patterns, which can provide more macroscopic and intuitive knowledge at region level. In the analysis of the regional mobility patterns, the inner-restricted fuzzy C-means clustering method and nonnegative tensor factorization can provide more meaningful region division and higher interpretability of the extracted patterns. Moreover, to assess the long-term effect of changing the land use structure of the regions, the prediction of the regional mobility patterns based on the regional properties is also achieved. For the urban planners and managers, understanding and predicting the regional mobility patterns can effectively support evidence-based and forward-looking planning and intelligent transportation management to achieve the healthy development and effective operation of urban cities.

The proposed methods show good performance on a case study in Beijing city. However, there is still room for improvements. The future improvements on the study can include four aspects. First, we only build up the third-order regional tensors in the paper, but the influence of the weather, events and festivals are not considered yet. Moreover, the analysis of the bus users' mobility is not enough to fully describe the mobility patterns of multiple travel modes. Therefore, more comprehensive data sources should be covered in future study for increasing the accuracy of the methods. Second, we only consider the land use structure as the main planning and design variables of influencing the mobility patterns, but ignore the other variables like bus network and frequency, which should be further analyzed in the future study. Third, in the nonnegative tensor factorization, the decomposing orders are specified instead of looking for the optimal value. In light of this, it is possible to consider combining the heuristic algorithm like genetic algorithm [78] with the current model for improving the selection of the decomposing orders. Fourth, we use a simple artificial neural network with one hidden layer in the paper. In the future study, the implementation of the deep learning method [79] may lead to better prediction results.

Acknowledgment

The authors would like to thank Dr Ming Xu (Tsinghua University), Dr Yuhan Jia (Tsinghua University) and Terigele (Dalian Medical University) for their valuable insights on the research and kind support in the preparation of the paper.

References

[1] J. Y. Wang, Y. Mao, J. Li, Z. Xiong, and W.-X. Wang, "Predictability of road traffic and congestion in urban areas," *PLoS ONE*, vol. 10, no. 4, p. e0121825, 2015.

[2] M. Rosenlund *et al.*, "Comparison of regression models with land-use and emissions data to predict the spatial distribution of traffic-related air pollution in Rome," *J. Exposure Sci. Environ. Epidemiol.*, vol. 18, no. 2, pp. 192–199, 2008.

[3] A. S. Rao, J. Gubbi, S. Marusic, and M. Palaniswami, "Crowd event detection on optical flow manifolds," *IEEE Trans. Cybern.*, vol. 46, pp. 1524–1537, Jul. 2016.

[4] I. M. Longini, Jr., *et al.*, "Containing pandemic influenza at the source," *Science*, vol. 309, no. 5737, pp. 1083–1087, 2005.

[5] K. Qing-Jie *et al.*, "An approach to urban traffic state estimation by fusing multisource information," *IEEE Trans. Intell. Transp. Syst.*, vol. 10, no. 3, pp. 499–511, Mar. 2009.

[6] H. Senaratne *et al.*, "Urban mobility analysis with mobile network data: A visual analytics approach," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 5, pp. 1537–1546, May 2018.

[7] Y. Ma, T. Lin, Z. Cao, C. Li, and W. Chen, "Mobility viewer: An Eulerian approach for studying urban crowd flow," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 9, pp. 2627–2636, Sep. 2016.

[8] P. A. Laharotte, R. Billot, E. Come, L. Oukhellou, A. Nantes, and N. E. E. Faouzi, "Spatiotemporal analysis of bluetooth data: Application to a large urban network," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 3, pp. 1439–1448, Jun. 2015.

[9] M.-P. Pelletier, M. Trépanier, and C. Morency, "Smart card data use in public transit: A literature review," *Transp. Res. C, Emerg. Technol.*, vol. 19, no. 4, pp. 557–568, 2011.

[10] S. Hasan, C. Schneider, S. V. Ukkusuri, and M. C. González, "Spatiotemporal patterns of urban human mobility," *J. Statist. Phys.*, vol. 151, nos. 1–2, pp. 304–318, Apr. 2013.

[11] M. Bagchi and P. R. White, "The potential of public transport smart card data," *Transp. Policy*, vol. 12, no. 5, pp. 464–474, 2005.

[12] T. Kusakabe and Y. Asakura, "Behavioural data mining of transit smart card data: A data fusion approach," *Transp. Res. C, Emerg. Technol.*, vol. 46, pp. 179–191, Sep. 2014.

[13] A. Ali, J. Kim, and S. Lee, "Travel behavior analysis using smart card data," *KSCE J. Civil Eng.*, vol. 20, no. 4, pp. 1532–1539, 2016.

[14] J. Zhao, C. Tian, F. Zhang, C. Xu, and S. Feng, "Understanding temporal and spatial travel patterns of individual passengers by mining smart card data," in *Proc. IEEE 17th Int. Conf. Intell. Transp. Syst. (ITSC)*, Oct. 2014, pp. 2991–2997.

[15] L. Li, J. Wang, Z. Song, Z. Dong, and B. Wu, "Analysing the impact of weather on bus ridership using smart card data," *IET Intell. Transp. Syst.*, vol. 9, no. 2, pp. 221–229, 2015.

[16] L. M. Kieu, A. Bhaskar, and E. Chung, "Passenger segmentation using smart card data," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 3, pp. 1537–1548, Jun. 2015.

[17] Y. Zhou, L. Yao, Y. Chen, Y. Gong, and J. Lai, "Bus arrival time calculation model based on smart card data," *Transp. Res. C, Emerg. Technol.*, vol. 74, pp. 81–96, Jan. 2017.

[18] B. Agard, C. Morency, and M. Trépanier, "Mining public transport user behaviour from smart card data," *IFAC Proc. Volumes*, vol. 39, no. 3, pp. 399–404, 2006.

[19] S. A. O. Medina, "Inferring weekly primary mobility patterns using public transport smart card data and a household travel survey," *Travel Behaviour Soc.*, Dec. 2016, doi: 10.1016/j.tbs.2016.11.005.

[20] L.-M. Kieu, A. Bhaskar, and E. Chung, "A modified density-based scanning algorithm with noise for spatial travel pattern analysis from smart card AFC data," *Transp. Res. C, Emerg. Technol.*, vol. 58, pp. 193–207, Sep. 2015.

[21] C. Jun and Y. Dongyuan, "Estimating smart card commuters origin-destination distribution based on APTS data," *J. Transp. Syst. Eng. Inf. Technol.*, vol. 13, no. 4, pp. 47–53, Aug. 2013.

[22] G. Goulet-Langlois, H. N. Koutsopoulos, Z. Zhao, and J. Zhao, "Measuring regularity of individual travel patterns," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 5, pp. 1583–1592, May 2018.

[23] K. K. A. Chu and R. Chapleau, "Enriching archived smart card transaction data for transit demand modeling," *Transp. Res. Rec.*, no. 2063, pp. 63–72, 2008.

[24] X. Ma, Y.-J. Wu, Y. Wang, F. Chen, and J. Liu, "Mining smart card data for transit riders' travel patterns," *Transp. Res. C, Emerg. Technol.*, vol. 36, pp. 1–12, Nov. 2013.

[25] C. Zhong, E. Manley, S. M. Arisona, M. Batty, and G. Schmitt, "Measuring variability of mobility patterns from multiday smart-card data," *J. Comput. Sci.*, vol. 9, pp. 125–130, Jul. 2015.

[26] S. Tao, D. Rohde, and J. Corcoran, "Examining the spatialâŁ"temporal dynamics of bus passenger travel behaviour using smart card data and the flow-comap," *J. Transp. Geogr.*, vol. 41, pp. 21–36, Dec. 2014.

[27] Y. Long and J.-C. Thill, "Combining smart card data and household travel survey to analyze jobsâŁ"housing relationships in Beijing," *Comput., Environ. Urban Syst.*, vol. 53, pp. 19–35, Sep. 2015.

[28] X. Ma, C. Liu, H. Wen, Y. Wang, and Y. Wu, "Understanding commuting patterns using transit smart card data," *J. Transp. Geogr.*, vol. 58, pp. 135–145, Jan. 2017.

[29] M. K. El Mahrsi, E. Côme, L. Oukhellou, and M. Verleysen, "Clustering smart card data for urban mobility analysis," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 3, pp. 712–728, Mar. 2017.

[30] C. Yu and Z.-C. He, "Analysing the spatial-temporal characteristics of bus travel demand using the heat map," *J. Transp. Geogr.*, vol. 58, pp. 247–255, Jan. 2017.

[31] Y. Gong, Y. Lin, and Z. Duan, "Exploring the spatiotemporal structure of dynamic urban space using metro smart card records," *Comput., Environ. Urban Syst.*, vol. 64, pp. 169–183, Jul. 2017.

[32] Y. Han and F. Moutarde, "Analysis of large-scale traffic dynamics in an urban transportation network using non-negative tensor factorization," *Int. J. Intell. Transp. Syst. Res.*, vol. 14, no. 1, pp. 36–49, Jan. 2016.

[33] Y. Wang, Y. Zheng, and Y. Xue, "Travel time estimation of a path using sparse trajectories," in *Proc. ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2014, pp. 25–34.

[34] M. Lee and P. Holme, "Relating land use and human intra-city mobility," *PLoS ONE*, vol. 10, no. 10, p. e0140152, 2015.

[35] D. Guo, X. Zhu, H. Jin, P. Gao, and C. Andris, "Discovering spatial patterns in origin-destination mobility data," *Trans. GIS*, vol. 16, pp. 411–429, May 2012.

[36] J. C. Bezdek, *Pattern Recognition With Fuzzy Objective Function Algorithms*, 1st ed. New York, NY, USA: Plenum, 1981.

[37] J. C. Dunn, "A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters," *J. Cybern.*, vol. 3, no. 3, pp. 32–57, 1973.

[38] N. R. Pal and J. C. Bezdek, "On cluster validity for the fuzzy c-means model," *IEEE Trans. Fuzzy Syst.*, vol. 3, no. 3, pp. 370–379, Aug. 1995.

[39] Y. Wang, C. Li, and Y. Zuo, "A selection model for optimal fuzzy clustering algorithm and number of clusters based on competitive comprehensive fuzzy evaluation," *IEEE Trans. Fuzzy Syst.*, vol. 17, no. 3, pp. 568–577, Jun. 2009.

[40] Y. Yejun, "Optimization of the clusters number of an improved fuzzy C-means clustering algorithm," in *Proc. 10th Int. Conf. Comput. Sci. Edu. (ICCSE)*, Jul. 2015, pp. 931–935.

[41] M. Ren, P. Liu, Z. Wang, and J. Yi, "A self-adaptive fuzzy C-means algorithm for determining the optimal number of clusters," *Comput. Intell. Neurosci.*, vol. 2016, Oct. 2016, Art. no. 2647389, doi: 10.1155/2016/2647389.

[42] F.-F. Guo, X.-X. Wang, and J. Shen, "Adaptive fuzzy c-means algorithm based on local noise detecting for image segmentation," *IET Image Process.*, vol. 10, no. 4, pp. 272–279, Apr. 2016.

[43] T. C. Havens, J. C. Bezdek, C. Leckie, L. O. Hall, and M. Palaniswami, "Fuzzy c-means algorithms for very large data," *IEEE Trans. Fuzzy Syst.*, vol. 20, no. 6, pp. 1130–1146, Dec. 2012.

[44] O. Kesemen, Ö. Tezel, and E. Özkul, "Fuzzy c-means clustering algorithm for directional data (FCM4DD)," *Expert Syst. Appl.*, vol. 58, pp. 76–82, Oct. 2016.

[45] A. Stetco, X.-J. Zeng, and J. Keane, "Fuzzy C-means++: Fuzzy C-means with effective seeding initialization," *Expert Syst. Appl.*, vol. 42, no. 21, pp. 7541–7548, Nov. 2015.

[46] Y. Liu, T. Hou, and F. Liu, "Improving fuzzy c-means method for unbalanced dataset," *Electron. Lett.*, vol. 51, no. 23, pp. 1880–1882, Nov. 2015.

[47] A. W. Dougherty and J. You, "A kernel-based adaptive fuzzy c-means algorithm for M-FISH image segmentation," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, May 2017, pp. 198–205.

[48] J. C. Bezdek, "A convergence theorem for the fuzzy ISODATA clustering algorithms," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-2, no. 1, pp. 1–8, Jan. 1980.

[49] S. Z. Selim and M. A. Ismail, "K-means-type algorithms: A generalized convergence theorem and characterization of local optimality," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-6, no. 1, pp. 81–87, Jan. 1984.

[50] H. Shen, J. Yang, S. Wang, and X. Liu, "Attribute weighted mercer kernel based fuzzy clustering algorithm for general non-spherical datasets," *Soft Comput.*, vol. 10, no. 11, pp. 1061–1073, 2006.

[51] A. Cichocki, R. Zdunek, A. H. Phan, and S. Amari, *Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multi-way Data Analysis and Blind Source Separation*. Hoboken, NJ, USA: Wiley, 2009.

[52] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, Oct. 1999.

[53] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM Rev.*, vol. 51, no. 3, pp. 455–500, 2009.

[54] J. D. Carroll and J.-J. Chang, "Analysis of individual differences in multidimensional scaling via an n-way generalization of 'Eckart–Young' decomposition," *Psychometrika*, vol. 35, no. 3, pp. 283–319, Sep. 1970.

[55] R. A. Harshman, "Foundations of the PARAFAC procedure: Models and conditions for an 'explanatory' multi-modal factor analysis," UCLA Working Papers Phonetics, Tech. Rep., 1970, vol. 16, pp. 1–84.

[56] L. R. Tucker, "Some mathematical notes on three-mode factor analysis," *Psychometrika*, vol. 31, no. 3, pp. 279–311, Sep. 1966.

[57] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM Rev.*, vol. 51, no. 3, pp. 455–500, 2009.

[58] Y.-D. Kim and S. Choi, "Nonnegative tucker decomposition," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit. (CVPR)*, Jun. 2007, pp. 1–8.

[59] L. De Lathauwer, B. De Moor, and J. Vandewalle, "A multilinear singular value decomposition," *SIAM J. Matrix Anal. Appl.*, vol. 21, no. 4, pp. 1253–1278, 2000.

[60] L. De Lathauwer, B. De Moor, and J. Vandewalle, "On the best rank-1 and rank-(R1, R2, ..., RN) approximation of higher-order tensors," *SIAM J. Matrix Anal. Appl.*, vol. 21, no. 4, pp. 1324–1342, 2013.

[61] Y. Xu and W. Yin, "A block coordinate descent method for regularized multiconvex optimization with applications to nonnegative tensor factorization and completion," *SIAM J. Imag. Sci.*, vol. 6, no. 3, pp. 1758–1789, 2013.

[62] T. G. Kolda, "Multilinear operators for higher-order decompositions," Sandia Nat. Lab., Albuquerque, NM, USA, Tech. Rep. SAND2006-2081, 2006.

[63] P. L. Fackler. (2005). *Notes on Matrix Calculus*. [Online]. Available: http://www4.ncsu.edu/~pfackler/MatCalc.pdf

[64] C. L. Wan and W. K. Dickinson, "Computer vision and neural networks for traffic monitoring," in *Proc. Road Traffic Monit.*, 1992, p. 143.

[65] P. Singhal and A. Yadav, "Congestion detection in Wireless sensor network using neural network," in *Proc. Int. Conf. Convergence Technol.*, Apr. 2014, pp. 1–4.

[66] D. A. Pomerleau, "Progress in neural network-based vision for autonomous robot driving," in *Proc. Intell. Vehicles Symp.*, Jun./Jul. 1992, pp. 391–396.

[67] H.-F. Yang, T. S. Dillon, and Y.-P. P. Chen, "Optimized structure of the traffic flow forecasting model with a deep learning approach," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 10, pp. 2371–2381, Oct. 2016.

[68] T. Xu, X. Sun, Y. Wu, Y. He, and C. Xie, "Artificial neural network and wavelet analysis application in speed forecast of Beijing urban freeway," in *Proc. IEEE Intell. Vehicles Symp.*, Jun. 2009, pp. 1004–1008.

[69] H. Yin, S. C. Wong, J. Xu, and C. K. Wong, "Urban traffic flow prediction using a fuzzy-neural approach," *Transp. Res. C, Emerg. Technol.*, vol. 10, no. 2, pp. 85–98, 2002.

[70] D. Zhao, Y. Dai, and Z. Zhang, "Computational intelligence in urban traffic signal control: A survey," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 42, no. 4, pp. 485–494, Jul. 2012.

[71] F. Rosenblatt, *Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms*. Washington, DC, USA: Spartan Books, 1962.
[72] A. K. Jain, J. Mao, and K. M. Mohiuddin, *Artificial Neural Networks: A Tutorial*. San Francisco, CA, USA: Academic, 1996.
[73] D. E. Rumelhart and J. L. McClelland, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*. Cambridge, MA, USA: MIT Press, 1986.
[74] N. Qian, "On the momentum term in gradient descent learning algorithms," *Neural Netw.*, vol. 12, no. 1, pp. 145–151, 1999.
[75] S. Ruder. (2016). *An Overview of Gradient Descent Optimization Algorithms*. http://sebastianruder.com/optimizing-gradient-descent/index.html
[76] *National Research Council, TCRP Report 100: Transit Capacity and Quality of Service Manual*, 2nd ed., Nat. Acad. Press, Washington, DC, USA, 2004.
[77] A. A. Campbell, C. R. Cherry, M. S. Ryerson, and X. M. Yang, "Factors influencing the choice of shared bicycles and shared electric bikes in Beijing," *Transp. Res. C, Emerg. Technol.*, vol. 67, pp. 399–414, Jun. 2016.
[78] Y. Norouzi, M. Kaffashpour-Yazdi, S. Araghi, and A. A. Shams-Baboli, "Fusion of Genetic Algorithm with tensor based algorithms for face recognition," in *Proc. Iranian Conf. Mach. Vis. Image Process.*, Nov. 2015, pp. 96–99.
[79] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, May 2015.

Geqi Qi was born in Inner Mongolia, China, in 1987. He received the B.S. and M.S. degrees in electrical engineering and automation from the College of Information and Electrical Engineering, China Agricultural University, Beijing, China, in 2009 and 2011, respectively, and the Ph.D. degree in civil engineering from Tsinghua University, Beijing, in 2016. He currently holds a post-doctoral position at the School of Traffic and Transportation, Beijing Jiaotong University, Beijing. His research interests include machine learning, data mining, behavior analysis, human factors, and intelligent transport systems.

Ailing Huang received the B.S. degree in traffic and transportation, and the M.S. and Ph.D. degrees in systems engineering from Beijing Jiaotong University, Beijing, China, in 2000, 2003, and 2014, respectively. She is currently an Associate Professor with the School of Traffic and Transportation, Beijing Jiaotong University. She has authored and co-authored over 30 articles in refereed journals, book chapters, and conference proceedings. Her current research interests include systems engineering and traffic network modeling.

Wei Guan received the B.S., M.S., and Ph.D. degrees in systems engineering from Tianjin University, China, in 1990, 1993, and 1997, respectively. From 1997 to 1999, he was a Post-Doctoral Researcher with Beijing Jiaotong University, Beijing, China.

He is currently a Professor with the School of Traffic and Transportation, Beijing Jiaotong University. He has authored and co-authored over 170 articles in refereed journals, book chapters, and conference proceedings. His current research interests include systems engineering and intelligent transportation systems.

Lingling Fan received the B.E. and M.E. degrees from the Department of Traffic and Transportation, Beijing Jiaotong University, Beijing, China, in 2002 and 2005, respectively, and the Ph.D. degree from the School of Mathematical Sciences, Graduate University, Chinese Academy of Sciences, Beijing, in 2008. From 2011 to 2013, she was a Post-Doctoral Researcher in transportation engineering with Beijing Jiaotong University, where she is currently an Assistant Professor with the School of Traffic and Transportation. Her research interests are in intelligent transportation, image processing, and pattern recognition.