IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received June 13, 2019, accepted August 16, 2019, date of publication August 20, 2019, date of current version September 3, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2936443

Detecting Travel Modes Using Rule-Based Classification System and Gaussian Process Classifier

GUANGNIAN XIAO[1], QIN CHENG[1], AND CHUNQIN ZHANG[2]
[1]School of Economics and Management, Shanghai Maritime University, Shanghai 201306, China
[2]School of Civil Engineering and Architecture, Zhejiang Sci-Tech University, Hangzhou 310018, China

Corresponding author: Guangnian Xiao (gnxiao@shmtu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 51478266, in part by the Ministry of Education in China (MOE) Project of Humanities and Social Sciences under Grant 19YJA630058 and Grant 19YJCZH238, and in part by the Natural Science Foundation of Zhejiang Province, China, under Grant LQ18E080010.

ABSTRACT Travel modes are generally derived from Global Positioning System (GPS) data on the basis of either a rule-based or machine learning classification method. The rule-based classification approach is generally easy to understand, whereas the machine learning classification method has better generalization. However, studies that jointly explore both methods are limited. The present research proposes a two-stage method that aims to impute travel modes from GPS trajectory data. In the first stage, rules are employed to detect subway modes. In the second stage, a Gaussian process classifier based on sequential forward selection methods is utilized to derive the remaining travel modes. On the basis of the four selected features constituting the feature set (i.e., average speed, average acceleration, heading change, and low-speed point rate), over 97% of the samples with subway modes are correctly identified and 93.04% of segments in the walk-based balanced test subset are accurately detected. Over 92% of the car and bus samples are correctly detected for the training and test datasets. Results provide a new perspective in selecting classification methods for the detection of travel modes and other travel characteristics from GPS trajectory data. Furthermore, high differentiation is achieved between the bus and car modes without the bus transit geographic information system sources of bus networks. Therefore, reasonable extracted features contribute to the detection of travel modes, particularly between bus and car modes.

INDEX TERMS Gaussian processes, classification algorithms, global positioning system.

I. INTRODUCTION

In the past few decades, dedicated Global Positioning System (GPS) devices and smartphones have been increasingly applied to gather location-based data in GPS-based travel surveys, which are widely considered promising alternatives to conventional travel surveys. This type of travel survey method is advantageous because of its minimal burden on respondents [1], high reliability of GPS data, and valuable information obtained from such data [2]–[4]. Nevertheless, some key travel characteristics, including trip rates, trip purposes, and travel modes, may not be directly derived from GPS data. Thus, studying the travel behavior using GPS data is difficult. Researchers generally use rules to flag trip ends and employ geographic information system (GIS) data to detect trip purposes [5]–[9]. By contrast, transportation modes are detected with either rule-based [10]–[12] or machine learning classification methods [13]–[16]. Several studies on the detection of travel modes exploit GIS data [17], whereas others utilize GPS data alone [18] or an application programming interface (API) [19]. No consensus has been reached with regard to the type of detection method preferred.

The majority of the existing rule-based algorithms that detect travel modes use speed-related features, including percentile speed, acceleration, and average speed. Stopher *et al.* [20] adopted three steps to flag a single-mode segment as walk, bicycle, car, or bus modes. First, a segment with average and maximum speeds of below 6 km/h and under 10 km/h, respectively, was flagged as a walk mode. Second, a segment that matched bus networks, exhibited

The associate editor coordinating the review of this article and approving it for publication was Bora Onat.

periodical stops, and had an average speed of 10-40 km/h, was detected as a bus mode. Third, a segment was imputed as a bicycle mode if its average speed was below 40 km/h; otherwise, it was flagged as a car mode. According to these rules, they properly imputed approximately 95% of the segments. Therefore, rule-based algorithms are easily understandable and highly effective in certain situations.

Machine learning classification methods, including fuzzy logic regression [6], [21], decision tables [22], random forest [23], [24], deep learning [25], and other machine learning approaches [26]–[28], are also frequently used in deriving travel modes. Byon *et al.* [29] extracted several features, such as the average accuracy of the horizontal dilution of precision, the average number of satellites in view, speed, and acceleration, and differentiated four travel modes with neural networks. To alleviate the interference of GPS positioning errors on travel mode detection, Joseph *et al*. [30] adopted the 95th percentile acceleration, median speed, and 95th percentile speed instead of the maximum speed. In addition, they used a multinomial logit (MNL) model to distinguish the four travel modes, and achieved a preferable accuracy of 87.5%. Feng and Timmermans [31] recently compared the rule-based and machine learning classification methods and found that the latter had higher flexibility and robustness than the former in classifying travel modes.

Another aspect is to determine the necessity of GIS data in the detection of travel modes. Joseph *et al*. [30] imputed travel modes with two different inputs, namely, GPS data alone and GPS data combined with bus networks. They found that the classification accuracy increased from 87.5% to 89.6% after including bus networks. Zheng *et al*. [32] constructed several representative features in the case of the non-availability of GIS data, including speed change rate, stop rate, and heading change rate, because non-motorized modes change headings more frequently than motorized modes and bus modes display periodic stops. Furthermore, they employed decision trees to distinguish five travel modes and achieved a promising classification result. Given that no GIS data were involved, they developed an online application with more convenience. Similarly, Schuessler and Axhausen [6] developed algorithms to impute travel modes from GPS data alone, and achieved a preferable classification performance. Zhu and Gonder [19] used the Google Maps Direction API as basis to (a) propose a novel mode detection method to obtain a best-matched API route for each actual route (segment), (b) calculate the similarity scores for each pair of routes, and (c) feed them into a logistic regression model to differentiate the car and non-car modes from each other. Consequently, they achieved an overall detection accuracy of approximately 89%, which was better than that of the logistic regression model based on the raw GPS data.

In summary, a rule-based classification method provides an intuitive explanation for travel mode imputation, but such a classification method may result in relatively high confusion among "similar" travel modes. For example, a bicycle mode with low speed may be mistakenly flagged as a walk mode if the average speed is a key feature to distinguish travel modes. By contrast, machine learning classification methods provide an opportunity to deal with this issue. For example, kernel-based classifiers are widely considered powerful when used to deal with classification issues [33]. Support Vector Machine (SVM) is one of the most prevalent kernel-based classifiers and constructed on the basis of the margin maximization principle, which contributes to their favorable generalization. The Gaussian process classifier (GPC), which is derived from GP, is another kernel-based classifier.

GP is a promising statistical model because it allows for a complete Bayesian treatment of classification and regression problems. GP applies a probabilistic and practical approach to learning in kernel machines, resulting in advantages over its competitors in the interpretation of model architecture and the integrated treatment of learning and model selection.

In contrast to other commonly used classifiers, GPC has three main advantages. First, GPC can handle high-dimensional and nonlinear issues, which are encountered in travel mode detection. Second, GPC offers probabilistic outputs instead of determinant classification results, thereby accounting for the model uncertainty inherent in travel mode detection. Third, GPC belongs to a nonparameterized model and can tune hyperparameters directly on the basis of the training data. GPC can also use evidence to implement a model selection process in a fully automatic manner [34]. GPC has been successfully employed in the transportation domain [34], [35]. The main disadvantage of GPC lies in the relatively high computation cost owing to numerous trials. However, the current study does not require real-time travel mode detection, but focuses on the improvement of the classification accuracy. Therefore, GPC is applied in this research.

However, the possibility of jointly using rule-based approaches and machine learning classification methods to derive travel modes has rarely been explored. Therefore, the present study proposes a two-stage method to impute travel modes. A rule-based approach is suited for inferring a travel mode that is highly different from other modes in terms of one or multiple features. In addition, GPS signal is frequently blocked in subway segments, thereby resulting in difficulty to compute the features fed in classifiers. Hence, we identify subway modes with a rule-based algorithm in the first stage because subway modes generally match the metro network. Moreover, the majority of subway trips are characterized by severe signal loss. In the second stage, we employ GPC to detect the remaining modes. Instead of bus transit GIS data, several targeted features are extracted to increase the classification accuracy. Thus, the two basic objectives of this study are as follows: (a) to decide whether the two-stage method is superior to other frequently used classifiers in terms of classification accuracy and (b) to decide whether bus transit GIS data are indispensable in imputing travel modes. The preliminary results have been described by TRB 2015 [36] and consist of two aspects. On the one hand, a rule-based approach is preferable to detect subway

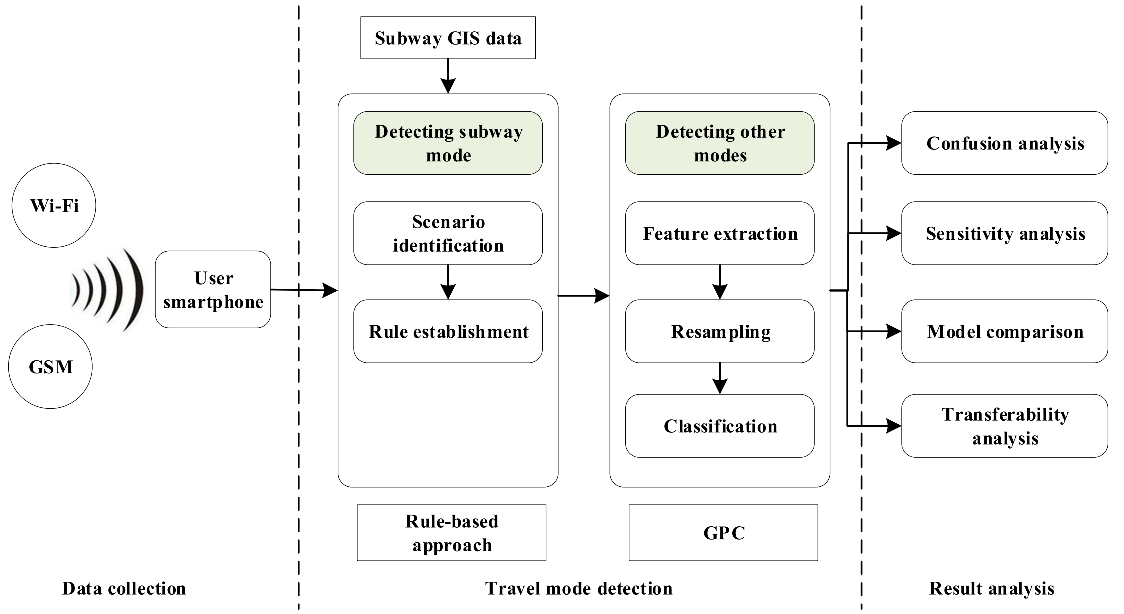


FIGURE 1. Flowchart of the current study.

modes, because the characteristic of positioning data for subway trips is evidently different from other modes. On the other hand, feature selection is important for detecting travel modes because a reasonable feature combination contributes to promising classification results.

The flowchart of this study is shown in Fig. 1. The remainder of this paper is organized as follows. Section 2 describes the data profile. Section 3 elaborates the theoretical and methodological backgrounds of the two-stage method. Section 4 presents the results of the travel mode detection. Lastly, Section 5 provides the main conclusions and future research directions.

II. DATA DESCRIPTION

The GPS data were acquired in a smartphone-based travel survey administered in three waves in Shanghai from mid-October 2013 to mid-July 2014. This survey should handle privacy issues reasonably owing to the collection of real-time positioning data. On the one hand, the respondents may examine the authorities achieved by the application and check whether the application collected unnecessary information. On the other hand, the respondents may receive a commitment letter explaining that the collected data is confined to scientific research by our group alone and that the data collected would be kept anonymous and confidential. In the survey, the travel mode of each single-mode segment (hereafter referred to as "segment") was recorded instead of the main mode of a trip. The validated travel modes were taken as the "ground truth" for the evaluation of the classification result. A total of 1,512 person-day GPS data from 203 respondents were acquired. After the data cleaning based on integrity and logicality check, 1,248 person-day data were reserved for travel mode inference. The data may

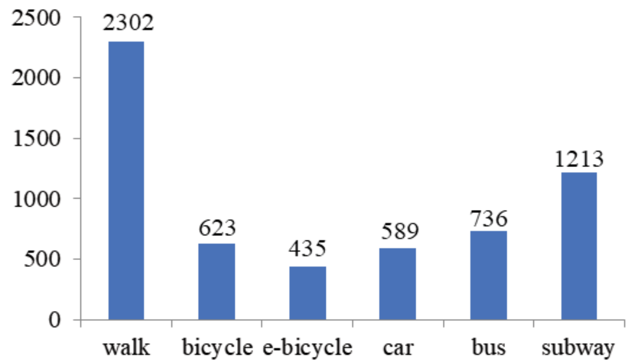


FIGURE 2. The number of segments with six reported travel modes.

have periodic characteristics because multi-day data were collected for each respondent.

The walk, bicycle, e-bicycle, car, bus, and subway modes were distinguished in the travel survey. The trajectory data were divided into segments in accordance with the validated travel modes. Exactly 5,898 segments were collected from the GPS data (see Fig. 2). Walk accounted for the largest proportion because walk modes often connect two other modes during a trip. In Shanghai, subways have increasingly gained popularity since Line 1 began operation in 1993. As of February 2019, 17 lines (including the Shanghai Maglev Train and Shanghai Metro Pujiang Line) with 330 stations are operating, with the line length reaching 705 km.

III. METHODOLOGY

Travel mode detection consists of two steps, namely, the identification of subway segments with a rule-based classification system and detection of segments with the remaining transportation modes using GPC. When GPC is employed, we utilize several feature selection methods to achieve the optimal feature set from the six extracted features.

A. DETECTING SUBWAY SEGMENTS

Subway segments can be detected using an appropriate rule-based classification system. In the majority of cases, the Shanghai subway network does not overlap with highways. Furthermore, the quality of GPS signals for subway modes is worse than that for other modes. The number of GPS points in subway segments is also relatively few owing to signal blockage. These unique characteristics improve the distinction between the subway and other modes. Two possible scenarios, namely, segments with incomplete GPS signal and those without GPS signal, are considered for the subway mode detection. As shown in Fig. 3, two trips sequentially consist of a walk segment, a subway segment, and another walk segment.

Subway segments without GPS signal are relatively common. Accordingly, receiving sufficient satellite signals (in most cases, four satellites are sufficient) for positioning is difficult when smartphones are in a subway train. Moreover, a smartphone may not record any point when it is in a subway train operating underground. Therefore, most trips involving subway modes are displayed (see Fig. 3 (a)), in which no GPS points are collected for a subway segment. A partially enlarged subway segment shown in Fig. 3 (c) indicates that the GPS signal recovers after the traveler walks out of Exit 1. Another issue is the detection of a subway segment with incomplete GPS signal. Fig. 3 (b) demonstrates that the GPS signal does not vanish immediately when the subway train starts to run. All GPS points collected for the subway segment are along both sides of the line, and the distance between the line and these points differs because of the positioning error (see Fig. 3 (d)).

The grid search algorithm is appropriate for establishing the rules for data mining because it tests all given parameter combinations and outputs the optimal parameter combination according to a predefined evaluation criterion. Grid search aims to perform parameter optimization, which finds the best parameter combinations with an exhaustive search through a manually specified subset of the parameter space [37], [38]. In addition, this method involves low computation cost when the number of parameters is low. To establish rules for detecting subway segments, the critical values of relevant parameters should be determined. These parameters consist of the critical duration, maximum speed of GPS points, and distance between GPS points and subway stations or lines [21]. The candidate critical values of the parameters are determined according to the statistical analysis of the smartphone-based travel survey (see Table 1). We randomly choose 75% of the full sample as the training dataset to establish the rules [31].

TABLE 1. Parameters and candidate critical values for detecting subway segments.

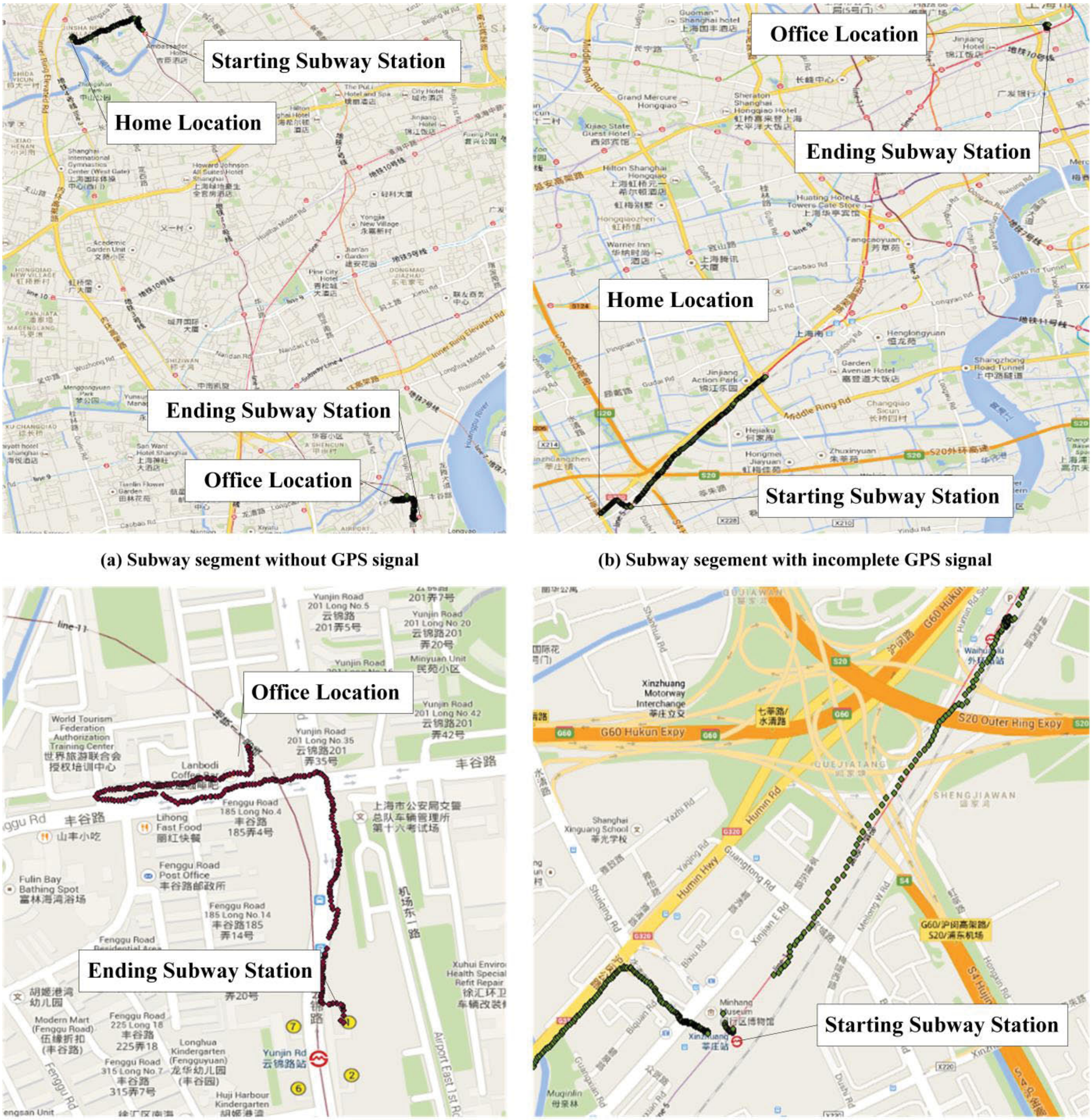
Parameters	Description	Candidate critical values
min_dur	Minimum duration	5, 10, 15, 20, 25 min
max_dur	Maximum duration	120, 150, 180, 210, 240 min
max_spe	Maximum speed of points	60, 70, 80, 90, 100 km/h
max_dis_sta	Maximum distance from the start point of the segment to the nearest entrance/exit	50, 100, 150, 200, 250 m
max_dis_end	Maximum distance from the end point of the segment to the nearest entrance/exit	50, 100, 150, 200, 250 m
max_dis_line	Maximum distance from each point of the segment to the nearest subway line	10, 20, 30, 40, 50 m

B. FEATURE DESCRIPTION

Existing studies have indicated that traveled distance, average speed, average acceleration, and 95th percentile speed are chosen to derive travel modes except subway in this study. Nevertheless, distinguishing car modes from bus modes will be difficult if only these speed-related features are used [20], [39]. In general, the introduction of a bus network may significantly decrease the confusion between car modes and bus modes [30]. However, bus transit GIS data are difficult to obtain. In addition, bus networks are updated monthly or even daily in megacities, including Shanghai. Thus, a timely update must be carried out to effectively infer bus segments. In this case, maintaining the latest version of the bus transit GIS layer becomes costly. Such a situation motivates us to extract a novel feature called ''low-speed point rate,'' i.e., the rate of GPS points with a speed of below 1 m/s, which represents the periodic stops of buses [40]. Four different critical speeds, namely, 0.5, 1.0, 1.5, and 2.0 m/s, are investigated for constructing the feature. Consequently, 1 m/s is employed because it can best distinguish bus segments from car segments. Another obstacle that potentially increases the confusion among travel modes results from the high uncertainty of speed-related features [41]. For example, a bus or car segment with a relatively low speed may be mistakenly flagged as an e-bicycle segment. Notably, a motorized vehicle may only drive on motorways and along a specific direction in most cases. By contrast, a bicycle traveler may stop or overtake others casually. Therefore, we construct of the ''heading change rate'' feature, which is the average heading change between two consecutive points. In summary, two additional features are used to further distinguish car segments from the bus segments and motorized segments from non-motorized segments.

C. DETECTING TRAVEL MODES EXCEPT SUBWAY

It is a classification problem to detect travel modes except subway based on the previously mentioned features. In particular, the input of GPC for each segment comprises six features (i.e., traveled distance, average speed, average acceleration, 95th percentile speed, low-speed point rate, and heading change rate) and the output is one of five travel modes (i.e., walk, bicycle, e-bicycle, car, and bus). GPC is applied



(a) Subway segment without GPS signal

(b) Subway segement with incomplete GPS signal

(c) Partially enlarged subway segment without GPS signal

(d) Partially enlarged subway segment with incomplete GPS signal

FIGURE 3. Subway segments with incomplete GPS signal and without GPS signal.

to detect travel modes because of two main advantages. On the one hand, each segment is assigned to a travel mode with the maximum probability, thereby accommodating the model uncertainty. Such an uncertainty is frequently seen in travel mode detection, particularly between bus and car modes. On the other hand, GPC is a nonparameterized model, in which the values of hyperparameters are not predefined by researchers, but automatically determined by the model itself according to the training data. In addition, existing studies have found that GPC competes with other commonly used classifiers [42], [43]. GPC is described in detail in the next section.

Data balance should be addressed before the travel mode detection. The walk takes a much higher percentage than

TABLE 2. The number of samples in the training and datasets.

		Subway	Walk	Bicycle	E-bicycle	Bus	Car	Total
Training datasets	Unbalanced training dataset	910	1,727	467	326	552	442	3,514
	Walk-based balanced training subset	0	1,727	1,727	1,727	1,727	1,727	8,635
	Bus-based balanced training subset	0	552	552	552	552	552	2,760
	E-bicycle-based balanced training subset	0	326	326	326	326	326	1,630
Test datasets	Unbalanced test dataset	303	575	156	109	184	147	1,171
	Walk-based balanced test subset	0	575	575	575	575	575	2,875
	Bus-based balanced training subset	0	184	184	184	184	184	920
	E-bicycle-based balanced training subset	0	109	109	109	109	109	545

other modes and walk segments are five times as many as e-bicycle segments. This imbalanced scenario may result in a low detection precision for the e-bicycle mode. Thus, resampling is used to rebalance the training and test dataset to alleviate the effect of the skewed class distribution on the learning process. Compared with other the two other commonly used rebalancing methods (i.e., cost-sensitive learning and ensemble learning), the resampling technique is more versatile because it is independent of the selected classifier [44]. To acquire the expected number of segments for each mode, the resampling technique is applied by randomly eliminating majority mode segments (under-sampling) and creating synthetic minority mode segments (over-sampling) with SMOTE [45]. Compared with the over-sampling based on replacement, SMOTE has the potential to significantly improve the precision of minority modes. To evaluate the effect of the expected number on the detection accuracy, we take the number of segments of walk, bus, and e-bicycle modes as three expected numbers of each mode. The unbalanced test dataset consists of 25% of the samples that are left after a random draw of the unbalanced training dataset. The number of samples in training and test datasets are denoted in Table 2.

D. APPLICATION OF GPC TO TRAVEL MODE DETECTION

To detect travel modes in this study, GPC works as follows:

(1) Define the prior of latent variables for travel modes

Given a training set $D = (\mathbf{X}, \mathbf{y})$ with n segments, in which $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n)^{\mathrm{T}}$ and $\mathbf{y} = (\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_n)^{\mathrm{T}}$. For each vector $\mathbf{x}_i \in \Re^d (i = 1, 2, ..., n)$, a certain travel mode $c \in \{1, 2, 3, 4, 5\}$ is matched, with each number corresponding to a travel mode. Each $\mathbf{y}_i$ with travel mode c has a length of five, with an entry of 1 for the cth element and 0 for others. Thus, $\mathbf{y}$ is defined as follows:

$$\mathbf{y} = (y_1^1, ..., y_n^1, y_1^2, ..., y_n^2, ..., y_1^5, ..., y_n^5)^{\mathrm{T}}. \tag{1}$$

The vector for the latent function values of all n training segments is defined as follows:

$$\mathbf{f} = (f_1^1, ..., f_n^1, f_1^2, ..., f_n^2, ..., f_1^5, ..., f_n^5)^{\mathrm{T}}. \tag{2}$$

The prior of $\mathbf{f}$ is assumed to be normally distributed with $\mathbf{f}|\mathbf{X} \sim N(0, \mathbf{K})$ and the five latent processes are assumed to be uncorrelated. The covariance matrix $\mathbf{K}$ is block-diagonal with five identical block matrices. We employ a well-known kernel form called the Gaussian radial basis function, which can be written as follows:

$$k(\mathbf{x}_i, \mathbf{x}_j) = \theta_0 \exp(-\frac{|\mathbf{x}_i - \mathbf{x}_j|^2}{2l^2}), \tag{3}$$

where θ_0 is the process variance and l denotes the length scale, both of which constitute the hyperparameter set $\boldsymbol{\theta}$. All five classes share the same hyperparameter values in the current study.

(2) Define the likelihood function of the travel modes

Let π_i^c be the probability of detecting training segment i as travel mode c. In this case, π_i^c is the sigmoid function of the corresponding latent variables of travel modes as follows:

$$\pi_{\mathrm{i}}^c = \frac{\exp(f_i^c)}{\sum_{c'} \exp(f_i^{c'})}. \tag{4}$$

Consequently, the likelihood function is defined as follows:

$$p(\mathbf{y}|\mathbf{f}) = \prod_{i=1}^{n}\prod_{c=1}^{5} y_i^c \pi_i^c = \prod_{i=1}^{n}\prod_{c=1}^{5} y_i^c \frac{\exp(f_i^c)}{\sum_{c'} \exp(f_i^{c'})}. \tag{5}$$

(3) Compute the posterior of the latent variable for travel modes

According to Bayes' rule, the posterior of the latent variable $p(\mathbf{f}|\mathbf{X}, \mathbf{y}) = p(\mathbf{y}|\mathbf{f})p(\mathbf{f}|\mathbf{X})/p(\mathbf{y}|\mathbf{X})$, where $p(\mathbf{y}|\mathbf{X})$ is independent of $\mathbf{f}$. Given that $p(\mathbf{f}|\mathbf{X}, \mathbf{y})$ is not Gaussian, Laplace approximation is used to approximate it with the Gaussian $q(\mathbf{f}|\mathbf{X}, \mathbf{y})$. With a second-order Taylor expansion, the following Gaussian approximation is obtained:

$$q(\mathbf{f}|\mathbf{X}, \mathbf{y}) = N(\mathbf{f}|\widehat{\mathbf{f}}, \mathbf{A}^{-1}) \propto \exp(-\frac{1}{2}(\mathbf{f} - \widehat{\mathbf{f}})^T \mathbf{A}(\mathbf{f} - \widehat{\mathbf{f}})), \tag{6}$$

where $\widehat{\mathbf{f}} = \arg\max_{\mathbf{f}} p(\mathbf{f}|\mathbf{X}, \mathbf{y})$ is the maximum of the posterior and $\mathbf{A} = -\nabla\nabla \log p(\mathbf{f}|\mathbf{X}, \mathbf{y})|_{\mathbf{f}=\hat{\mathbf{f}}}$ denotes the Hessian of the negative log posterior.

(4) Determine the travel modes of the test samples

According to the balanced training set D, the travel mode of a test segment $\mathbf{x}_*$ is predicted by computing the travel mode posterior distribution $p(\mathbf{y}_*|D, \mathbf{x}_*)$. To derive this distribution, the vector of the latent function values is evaluated for all n training segments and five classes.

Thereafter, $\boldsymbol{\pi}$ consists of each element π_i^c and denotes a vector with the same size as $\mathbf{f}$. The distribution of a latent variable corresponding to a test segment is denoted as follows:

$$p(\mathbf{f}_*|\mathbf{X}, \mathbf{y}, \mathbf{x}_*) = \int p(\mathbf{f}_*|\mathbf{X}, \mathbf{x}_*, \mathbf{f})p(\mathbf{f}|\mathbf{X}, \mathbf{y})d\mathbf{f}, \tag{7}$$

where $p(\mathbf{f}|\mathbf{X}, \mathbf{y})$ denotes the posterior distribution of the latent variables and $\mathbf{f}(x_*) \triangleq \mathbf{f}_* = (f_*^1, \ldots, f_*^C)^T$.

To predict the latent function value of a given test segment $\mathbf{x}_*$, we calculate the posterior distribution $q(\mathbf{f}_*|\mathbf{X}, \mathbf{y}, \mathbf{x}_*)$ as follows:

$$q(\mathbf{f}_*|\mathbf{X}, \mathbf{y}, \mathbf{x}_*) = \int p(\mathbf{f}_*|\mathbf{X}, \mathbf{x}_*, \mathbf{f})q(\mathbf{f}|\mathbf{X}, \mathbf{y})d\mathbf{f}. \tag{8}$$

Given that $p(\mathbf{f}_*|\mathbf{X}, \mathbf{x}_*, \mathbf{f})$ and $q(\mathbf{f}|\mathbf{X}, \mathbf{y})$ are Gaussian, $q(\mathbf{f}_*|\mathbf{X}, \mathbf{y}, \mathbf{x}_*)$ is also Gaussian. Therefore, its mean is as follows:

$$\mathrm{E}_q[\mathbf{f}(\mathbf{x}_*|\mathbf{X}, \mathbf{y}, \mathbf{x}_*)] = Q_*^{\mathrm{T}}\mathbf{K}^{-1}\hat{\mathbf{f}} = Q_*^{\mathrm{T}}(\mathbf{y} - \boldsymbol{\pi}), \tag{9}$$

where

$$Q_* = \begin{pmatrix} \mathbf{k}_1(\mathbf{x}_*) & 0 & \cdots & 0 \\ 0 & \mathbf{k}_2(\mathbf{x}_*) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{k}_5(\mathbf{x}_*) \end{pmatrix}, \tag{10}$$

where $\mathbf{k}_c(\mathbf{x}_*)$ denotes the covariance matrix between the test segment and each training segment pertaining to travel mode c. The covariance matrix is as follows:

$$\mathrm{cov}_q(\mathbf{f}_*|\mathbf{X}, \mathbf{y}, \mathbf{x}_*) = \sum + Q_*^{\mathrm{T}}\mathbf{K}^{-1}(\mathbf{K}^{-1} + \mathbf{W})^{-1}\mathbf{K}^{-1}Q_*, \tag{11}$$

where $\sum$ is a five-order diagonal matrix with $\Sigma = k_c(\mathbf{x}_*, \mathbf{x}_*) - \mathbf{k}_c^{\mathrm{T}}(\mathbf{x}_*)\mathbf{K}_c^{-1}\mathbf{k}_c(\mathbf{x}_*)$. Thereafter, we take trials from $p(\mathbf{f}_*|\mathbf{X}, \mathbf{x}_*, \mathbf{f})$, softmax them according to equation (4), and average them to derive the travel mode for each test segment [46]. The number of trials is determined for each test segment by taking trials repeatedly until the observed probability fluctuation of travel modes has "stabilized". The marginal likelihood $\log p(\mathbf{y}|\mathbf{X}, \theta)$ may be represented as follows by using Laplace approximation:

$$\begin{aligned} \log p(\mathbf{y}|\mathbf{X}, \theta) &\approx \log \mathrm{q}(\mathbf{y}|\mathbf{X}, \theta) \\ &= -\frac{1}{2}\hat{\mathbf{f}}^{\mathrm{T}}\mathbf{K}^{-1}\hat{\mathbf{f}} - \sum_{i=1}^{n}\log(\sum_{c=1}^{5}\exp\hat{f}_i^c) \\ &\quad - \frac{1}{2}\log|\mathbf{I}_{5n} + \mathbf{W}^{\frac{1}{2}}\mathbf{K}\mathbf{W}^{\frac{1}{2}}|. \end{aligned} \tag{12}$$

The marginal likelihood is maximized to tune the hyperparameters in the covariance function (see equation (3)).

IV. RESULTS AND DISCUSSIONS

The classification performance of the two-stage method is first evaluated with a confusion matrix. A local parameter sensitivity analysis (SA) is employed thereafter, followed by the comparison of the two-stage method with other commonly used classifiers. Lastly, spatial and temporal transferability is assessed with two metrics.

TABLE 3. Rules used in the subway mode detection.

Part 1. Segments without GPS signal
1a. The duration of the segment is between 5 min and 150 min
1b. The speed of the segment is below 90 km/h
1c. The distance from the starting point of the segment to the nearest entrance is below 100 m
1d. The distance from the ending point of the segment to the nearest exit of a different station is under 250 m
Part 2. Segments with incomplete GPS signal
2a. All the requirements in Part 1 are satisfied
2b. The distance from all points except the starting/ending point of the segment to the nearest subway line is below 30 m
2c. The maximum speed of points during the segment is below 90 km/h

A. CLASSIFICATION RESULTS

The results are obtained according to the following steps. First, rules trained from the training dataset are applied to identify the subway segments from the test dataset. Second, the three rebalanced training subsets are used to train GPC and the GPC that performs best is derived. Third, the derived GPC is utilized to detect the travel modes from the corresponding balanced test subset.

Rules that achieve the highest average of recall and precision are different for subway segments with incomplete GPS signal and those without GPS signal (see Table 3). The critical value of max_dur is associated with Shanghai subway network, which is so extensive that one might spend 150 min to take a subway segment. In addition, the maximum speed of Shanghai subway trains is the same as the critical value of max_spe. Tsui and Shalaby [21] set 630 m as the critical distance from both the starting point and ending point to the nearest entrance/exit. We collected the coordinates of all the entrances/exits for all stations and kept updating the data in time. Therefore, the critical distance from the starting/ending point of a segment to the nearest entrance/exit significantly decreases, but an excellent classification performance is maintained.

The evaluation criterion of GPC is the overall accuracy represented as the number of correctly detected segments in the balanced training and test subsets. The walk-based rebalanced training and test subsets are used for subsequent analysis because it achieves the highest overall accuracy. Average speed is first included in the feature set and the travel modes of over half of the segments are correctly detected with this feature alone (see Table 4). This prominent result demonstrates the irreplaceability of average speed in travel mode detection. The second feature included in the feature set is average acceleration, which increases the classification accuracy by over 30%. The heading change rate is the next feature included in the feature set. Unfortunately, the overall accuracy decreases after the inclusion of traveled distance. Therefore, the SFS procedure terminates with four features constituting the optimal feature set and the overall accuracy reaching 93.43%.

It facilitates a comprehensive evaluation for the two-stage method to analyze the confusion matrix of the training and

TABLE 4. The process of sequential forward selection.

Serial ID	Feature added	Accuracy of training subset	Accuracy of test subset	Overall accuracy	Stop the procedure ?
1	Average speed	51.30%	49.81%	50.56%	False
2	Average acceleration	83.30%	82.61%	82.96%	False
3	Low-speed point rate	90.21%	89.18%	89.70%	False
4	Heading change rate	93.82%	93.04%	93.43%	False
5	Traveled distance	94.22%	87.03%	90.61%	True

test datasets (see Table 5 and Table 6). On the one hand, subway segments are imputed with the highest precision of over 97% for the training and the test datasets. The detection precision of the subway segments of the training dataset is analogous to that of the test dataset. On the other hand, the precision of all travel modes exceeds 89%. In addition, the travel mode imputation of the test subset is nearly as accurate as the training dataset. These prominent results demonstrate the preferable classification ability and generalization of the proposed method.

Further analysis of the detection precision of the travel modes (except subway mode), contributes to an understanding of the advantages of the two-stage method and ways to improve classification accuracy in future studies. Over 94% of the bicycle segments are correctly flagged possibly because of the inclusion of heading change rate, which can effectively distinguish non-motorized and motorized segments. The detection recall for car and bus modes is above 92%. Meanwhile, the majority of the falsely detected bus segments are flagged as e-bicycle modes rather than car modes, and vice versa. The low confusion between car and bus segments may be associated with the introduction of the low-speed point rate.

B. SA

SA is generally used to analyze the quantitative and qualitative effects of uncertainty in the inputs on the uncertainty in the output. In the present research, a local SA is applied to evaluate the contribution of the parameters to the subway classification performance, because the values of hyperparameters in GPC are optimized by maximizing the marginal likelihood in equation (3). That is, only one parameter varies, while the others maintained constant at each time. A local SA requires minimal computation compared with a global SA. We use a normalized sensitivity index to explore the relative change in the model output caused by the change in the parameter value. The effect of a certain change in a given parameter is defined as follows [47]:

$$S_{(p)}(i) = \frac{(V_{(p)} - V_s)/V_s}{(p(i) - p_s)/p_s}, \tag{13}$$

where all the variables with a subscript S indicate reference values and those with a subscript P are calculated under alternative parameter values. p_s indicates the reference value of the target parameter, V_p is the output value under the alternative case. $p(i)$ denotes the ith element of the vector $p = \{0.8p_s, 0.9p_s, 1.1p_s, 1.2p_s\}$.

The sensitivity index determines the degree of model sensitivity toward any given parameter and is defined as follows:

$$S_{(p)} = \sum_{i=0}^{I} |S_{(p)}(i)|/I, \tag{14}$$

where I is 4 and $|\bullet|$ denotes the absolute value. A parameter is defined as not sensitive when $|S| \leq 0.1$, sensitive when $0.1 < |S| \leq 1$, highly sensitive when $1 < |S| \leq 10$, and extremely sensitive when $|S| > 10$. The output value is the detection accuracy of the subway segments, while the input consists

TABLE 5. Detection results of the training dataset.

		Identified						Total	Recall (%)
		Subway	Walk	Bicycle	E-bicycle	Bus	Car		
Reported	Subway	897	0	2	3	2	6	910	98.57
	Walk	1	1,661	37	15	8	5	1,727	96.18
	Bicycle	6	7	1,649	44	15	6	1,727	95.48
	E-bicycle	24	5	56	1,560	43	39	1,727	90.33
	Bus	14	6	20	36	1,623	28	1,727	93.98
	Car	28	8	18	37	28	1,608	1,727	93.11

TABLE 6. Detection results of the test dataset.

		Identified						Total	Recall (%)
		Subway	Walk	Bicycle	E-bicycle	Bus	Car		
Reported	Subway	295	0	1	2	0	5	303	97.36
	Walk	0	549	17	5	3	1	575	95.48
	Bicycle	4	4	544	15	4	4	575	94.61
	E-bicycle	5	0	21	512	26	11	575	89.04
	Bus	6	0	9	9	538	13	575	93.57
	Car	12	4	0	12	15	532	575	92.52

TABLE 7. Results of SA.

Parameter	Reference value	Effects of a certain change				Sensitivity index
		-20%	-10%	+10%	+20%	
min_dur	5 min	0.0718	0.0933	0.1508	0.0969	0.1032
max_dur	150 min	0.1149	0.1866	0.1005	0.0826	0.1211
max_spe	90 km/h	0.4415	0.3230	0.1651	0.1687	0.2746
max_dis_sta	100 m	0.3446	0.2225	0.1005	0.1220	0.1974
max_dis_end	250 m	0.5205	0.4092	0.2225	0.2692	0.3553
max_dis_line	30 m	0.3338	0.2513	0.2225	0.2836	0.2728

TABLE 8. Comparison of detection accuracy (without subway).

Methods	Walk-based training subset			Walk-based test subset		
	Sample size	# of segments correctly detected	Detection accuracy (%)	Sample size	# of segments correctly detected	Detection accuracy (%)
SVM	8,635	7,699	89.16	2,875	2,500	86.96
MNL	8,635	7,089	82.10	2,875	2,282	79.37
BN	8,635	7,806	90.40	2,875	2,533	88.10
ANN	8,635	8,065	93.40	2,875	2,608	90.71
GPC	8,635	8,101	93.82	2,875	2,675	93.04

of six parameters. The sensitivity index of each parameter is shown in Table 7.

The model output is sensitive toward the change in all parameters. When a parameter is the maximum critical value, the effect of a negative perturbation (-10% or -20%) is generally stronger than that of a positive one ($+10\%$ or $+20\%$). This finding may be explained by the asymmetrical impact of a perturbation along different directions. However, when the maximum critical value is increased, the number of segments falsely detected as a subway mode does not increase substantially. This result is associated with the idea that a segment is flagged as a subway mode only when all requirements are met. Moreover, the sensitivity index of any parameter is not classified as ''not sensitive'' or ''extremely sensitive,'' thereby indicating that the parameter selection is reasonable.

C. COMPARISON BETWEEN CLASSIFIERS

We choose four popular methods to compare the classification accuracy of GPC with that of other frequently used classifiers. These classifiers are used to differentiate five travel modes (subway excluded) to evaluate the relative performance of GPC based on the walk-based balanced training and test subsets. The results are shown in Table 8.

The four classifiers are SVM, MNL, Bayesian network (BN), and artificial neural network (ANN). The SVM classifier is trained and tested with a Gaussian kernel and one-versus-rest mode in the multiclass problem. The MNL model is employed, and all features are regarded as alternative-specific attributes. To train the BN classifier, the structure is constructed with a K2 algorithm, and the conditional probability tables are calculated with maximum likelihood methods. The ANN is constructed on the basis of a three-layer neural network, in which 1 to 20 neurons are tested in the hidden layer and the highest detection accuracy is achieved with 15 neurons.

Table 8 shows that GPC achieves the highest accuracy, which may be partially explained as follows. Although SVM possesses considerable distinction ability by mapping a low-dimensional space to a high-dimensional space, it generates non-probabilistic output of class determination, thereby implying an inability to deal with the ambiguity involved in travel mode imputation. MNL is widely used travel behavior analysis, but it does not necessarily apply to this issue because the requirement of the independence of irrelevant alternatives may not be met. ANN can implicitly represent complex nonlinear relationships between dependent and independent variables, but it is prone to overfitting. Although the BN classifier overcomes the aforementioned disadvantages, it cannot employ different methods to detect subway segments and segments with other modes, which have different characteristics in terms of their features. This issue is also encountered by the three other classifiers.

D. MODEL TRANSFERABILITY

Transferability is generally evaluated in the space and time aspects. One of the most important motivations for studying mode transferability is the opportunity to substantially reduce the amount of data required to estimate a model in a different area or time. Two contexts are used to assess the transferability of the two-stage method in the spatial and temporal dimensions. The context from which a model is transferred is the base context, whereas the context to which the model is transferred is the local context. Root mean square error (*RMSE*) and relative aggregate transfer error (*RATE*) are used to evaluate transferability. The formula of *RMSE* is as follows:

$$RMSE = \left(\sum_k P_k \times \left(\frac{P_k - O_k}{O_k} \right)^2 \right)^{1/2}, \qquad (15)$$

where P_k and O_k are the aggregate predicted and observed/reported shares, respectively, for travel mode k. The formula of *RATE* is as follows:

$$RATE = \frac{RMSE_b(l)}{RMSE_l(l)}, \tag{16}$$

where $RMSE_b(l)$ is the *RMSE* of a model applied in the local context but developed in the base context, and $RMSE_l(l)$ denotes *RMSE* of the locally estimated model.

The dataset used to assess spatial transferability was collected in Changchun City, Jilin Province from November to December 2013. A total of 254 segments are derived from the GPS data of 13 respondents in a travel survey administered via smartphones. According to Fig. 4, the $RMSE_b(l)$ and *RATE* values are 0.1648 and 1.2482, respectively. These values are substantially less than those of activity-based travel forecasting models for inter-state transfer between California and Florida [48]. This comparison reveals that the two-stage method has the acceptable spatial transferability.

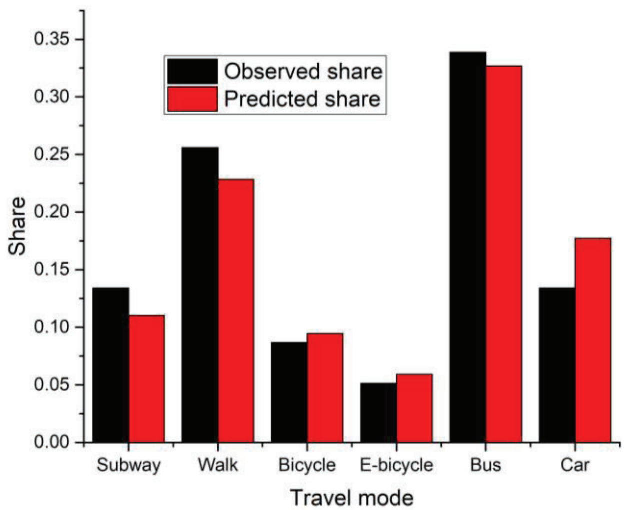


FIGURE 4. Spatial transferability analysis.

Temporal transferability is evaluated with another dataset, which consists of 953 segments extracted from the GPS data collected by 43 respondents with dedicated GPS devices in Shanghai City from October 2012 to December 2012. The format of the GPS data recorded by the dedicated GPS devices is similar to those obtained with smartphones. According to Fig. 5, $RMSE_b(l)$ and *RATE* values are 0.0643 and 1.1202, respectively. These values are substantially less than those for the spatial dimension. Thus, the temporal transferability of the two-stage method is better than its spatial transferability. This result may be explained by the considerable similarity in the socio-demographic characteristics, built environment, land use, and mode preferences of individuals in different periods of a city compared with those in different cities.

V. SUMMARY AND CONCLUSIONS

The contribution of this study is twofold. From a methodological perspective, we proposed a novel two-stage method to

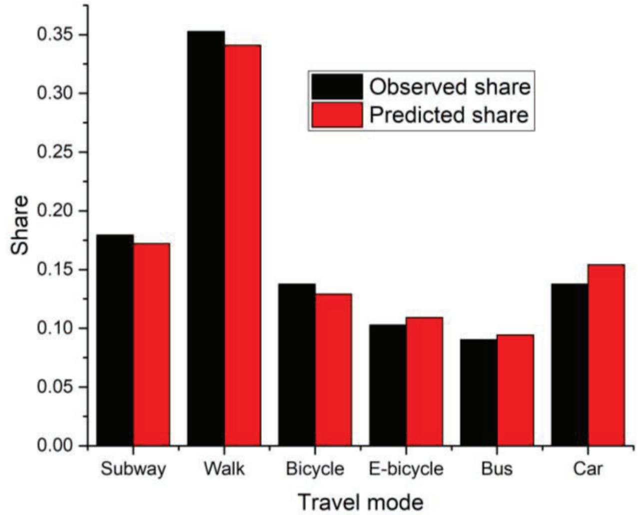


FIGURE 5. Temporal transferability analysis.

impute travel modes from GPS data. In the first stage, a rule-based classification system is employed to detect subway segments. In the second stage, GPC with SFS procedures is used to flag the remaining modes. The two-stage method maximizes the high understandability and interpretability of the rule-based approach and the preferable flexibility and robustness of machine learning classification methods. GPC was chosen on the basis of its apple-to-apple comparison with commonly used classifiers. In addition, the spatial and temporal transferability of the proposed method was proven to be acceptable. From an empirical perspective, reasonably extracted features may evidently improve the accuracy of the two-stage method. The reason is that this method achieved a high distinction between bus and car modes when no road and bus networks were included. That is, extracted features contribute to the differentiation between the bus and car modes, which depends on the bus transit GIS sources in existing studies. Note that the importance of the extracted features is emphasized on the basis of the improved classification accuracy, rather than their relative advantages over GIS data. These favorable results motivate researchers to develop additional features to represent bus transit GIS sources.

The two-stage method has rarely been included in existing studies. Nevertheless, this method helps to handle travel mode detection and other related issues in GPS travel surveys, including trip purpose detection. Compared with other methods, the two-stage method acquires the highest classification accuracy and presents a favorable generalization. From the perspective of empirical studies, the two-stage method also gains an advantage over those in related studies. Zheng *et al.* [32] applied decision trees to distinguish the walk, bicycle, car, and bus modes from one another and achieved an accuracy of 75.6%. Tsui and Shalaby [21] used the fuzzy logic method to detect these four travel modes and achieved a high accuracy of 91%. Zong *et al.* [49] used SVM combined with a genetic algorithm to recognize the walk, bicycle, subway, bus, and car modes and correctly detected 92.2% of their samples.

The proposed method is highly suited for imputing travel modes in a GPS travel survey conducted in a city where subway segments account for a relatively high percentage. Given that large-scale subway networks are gradually constructed and operated in many countries, the proposed method is expected to be applied extensively in the coming years. For cities without subway lines, travel modes for all segments should be detected with GPC because no significant ambiguity exists between segments with travel modes other than subway modes. In a future study, features that achieve high distinctiveness between e-bicycle segments and segments with other travel modes should be extracted because of the relatively low precision in detecting e-bicycle segments.

REFERENCES

[1] C. Stecher, S. Chesebro, and Y. Zhang, "Herding tigers: Lessons learned from the california household travel survey," presented at the Transp. Res. Board 93rd Annu. Meeting, Washington, DC, USA, 2014.

[2] A. C. Prelipcean, G. Gidófalvi, and Y. O. Susilo, "MEILI: A travel diary collection, annotation and automation system," *Comput., Environ. Urban Syst.*, vol. 70, pp. 24–34, Jul. 2018.

[3] H. Gjoreski, M. Ciliberto, L. Wang, F. J. O. Morales, S. Mekki, S. Valentin, and D. Roggen, "The University of Sussex-Huawei locomotion and transportation dataset for multimodal analytics with mobile devices," *IEEE Access*, vol. 6, pp. 42592–42604, 2018.

[4] P. Xiao, M. Ang, Z. Jiawei, and W. Lei, "Approximate similarity measurements on multi-attributes trajectories data," *IEEE Access*, vol. 7, pp. 10905–10915, 2019.

[5] J. Du and L. Aultman-Hall, "Increasing the accuracy of trip rate information from passive multi-day GPS travel datasets: Automatic trip end identification issues," *Transp. Res. A, Policy Pract.*, vol. 41, no. 3, pp. 220–232, 2007.

[6] N. Schuessler and K. Axhausen, "Processing raw data from global positioning systems without additional information," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2105, no. 1, pp. 28–36, 2009.

[7] D. S. Procter, A. S. Page, A. R. Cooper, C. M. Nightingale, B. Ram, A. R. Rudnicka, P. H. Whincup, C. Clary, D. Lewis, S. Cummins, A. Ellaway, B. Giles-Corti, D. G. Cook, and C. G. Owen, "An open-source tool to identify active travel from hip-worn accelerometer, GPS and GIS data," *Int. J. Behav. Nutrition Phys. Activity*, vol. 15, no. 1, 2018, Art. no. 91.

[8] A. Ermagun, Y. Fan, J. Wolfson, G. Adomavicius, and K. Das, "Real-time trip purpose prediction using online location-based search and discovery services," *Transp. Res. C, Emerg. Technol.*, vol. 77, pp. 96–112, Apr. 2017.

[9] J. van Dijk, "Identifying activity-travel points from GPS-data with multiple moving windows," *Comput., Environ. Urban Syst.*, vol. 70, pp. 84–101, Jul. 2018.

[10] J. Wolf, R. Guensler, and W. Bachman, "Elimination of the travel diary: Experiment to derive trip purpose from global positioning system travel data," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 1768, no. 1768, pp. 125–134, 2001.

[11] W. Bohte and K. Maat, "Deriving and validating trip purposes and travel modes for multi-day GPS-based travel surveys: A large-scale application in The Netherlands," *Transp. Res. C, Emerg. Technol.*, vol. 17, no. 3, pp. 285–297, 2009.

[12] P. R. Stopher and L. Wargelin, "Conducting a household travel survey with GPS: Reports on a pilot study," presented at the 12th World Congr. Transp. Res., Lisbon, Portugal, 2010.

[13] L. Liao, D. J. Patterson, D. Fox, and H. Kautz, "Learning and inferring transportation routines," *Artif. Intell.*, vol. 171, nos. 5–6, pp. 311–331, 2007.

[14] C. Rudloff and M. Ray, "Detecting travel modes and profiling commuter habits solely based on GPS data," presented at the Transp. Res. Board 89th Annu. Meeting, Washington, DC, USA, 2010.

[15] G. Xiao, Q. Cheng, and C. Zhang, "Detecting travel modes from smartphone-based travel surveys with continuous hidden Markov models," *Int. J. Distrib. Sensor Netw.*, vol. 15, no. 4, pp. 1–15, 2019.

[16] T. Feng and H. J. P. Timmermans, "Integrated imputation of activity-travel diaries incorporating the measurement of uncertainty," *Transp. Planning Technol.*, vol. 42, no. 3, pp. 274–292, 2019.

[17] H. Gong, C. Chen, E. Bialostozky, and C. T. Lawson, "A GPS/GIS method for travel mode detection in New York City," *Comput., Environ. Urban Syst.*, vol. 36, no. 2, pp. 131–139, 2012.

[18] T. Feng and H. J. P. Timmermans, "Transportation mode recognition using GPS and accelerometer data," *Transp. Res. C, Emerg. Technol.*, vol. 37, pp. 118–130, Dec. 2013.

[19] L. Zhu and J. D. Gonder, "A driving cycle detection approach using map service API," *Transp. Res. C, Emerg. Technol.*, vol. 92, pp. 349–363, Jul. 2018.

[20] P. R. Stopher, Q. Jiang, and C. FitzGerald, "Processing GPS data from travel surveys," presented at the 2nd Int. Colloqium Behavioural Found. Integr. Land-Use Transp. Models, Frameworks, Models Appl., Toronto, ON, Canada, 2005.

[21] A. Y. A. Tsui and A. S. Shalaby, "Enhanced system for link and mode identification for personal travel surveys based on global positioning systems," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 1972, no. 1, pp. 38–45, 2006.

[22] Y. Zheng, L. Liu, L. Wang, and X. Xie, "Learning transportation mode from raw GPS data for geographic applications on the Web," in *Proc. 17th Int. Conf. World Wide Web*, Beijing, China, 2008, pp. 247–256.

[23] M. A. Shafique and E. Hato, "Travel mode detection with varying smartphone data collection frequencies," *Sensors*, vol. 16, no. 5, pp. 716–740, 2016.

[24] B. Wang, L. Gao, and Z. Juan, "Travel mode detection using GPS data and socioeconomic attributes based on a random forest classifier," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 5, pp. 1547–1558, May 2018.

[25] S. Dabiri and K. Heaslip, "Inferring transportation modes from GPS trajectories using a convolutional neural network," *Transp. Res. C, Emerg. Technol.*, vol. 86, pp. 360–371, Jan. 2018.

[26] J. Hagenauer and M. Helbich, "A comparative study of machine learning classifiers for modeling travel mode choice," *Expert Syst. Appl.*, vol. 78, pp. 273–282, Jul. 2017.

[27] B. Assemi, H. Safi, M. Mesbah, and L. Ferreira, "Developing and validating a statistical model for travel mode identification on smartphones," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 7, pp. 1920–1931, Jul. 2016.

[28] X. Su, H. Caceres, H. Tong, and Q. He, "Online travel mode identification using smartphones with battery saving considerations," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 10, pp. 2921–2934, Oct. 2016.

[29] Y.-J. Byon, B. Abdulhai, and A. S. Shalaby, "Impact of sampling rate of GPS-enabled cell phones on mode detection and GIS map matching performance," presented at the Transp. Res. Board 86th Annu. Meeting, Washington, DC, USA, 2007.

[30] B. Joseph, N. W. McNeil, and D. Jennifer, "Travel mode imputation using GPS and accelerometer data from a multi-day travel survey," presented at the Transp. Res. Board 93rd Annu. Meeting, Washington, DC, USA, 2014.

[31] T. Feng and H. J. P. Timmermans, "Comparison of advanced GPS data imputation algorithms for detection of transportation mode and activity episode," presented at the Transp. Res. Board 93rd Annu. Meeting, Washington, DC, USA, 2014.

[32] Y. Zheng, Y. Chen, Q. Li, X. Xie, and W.-Y. Ma, "Understanding transportation modes based on GPS data for Web applications," *ACM Trans. Web*, vol. 4, no. 1, 2010, Art. no. 1.

[33] M. Allahviranloo and W. Recker, "Daily activity pattern recognition by using support vector machines with multiple classes," *Transp. Res. B, Methodol.*, vol. 58, pp. 16–43, Dec. 2013.

[34] S. Sun and X. Xu, "Variational inference for infinite mixtures of Gaussian processes with applications to traffic flow prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 2, pp. 466–475, Jun. 2011.

[35] J. Q. Retherford and M. McDonald, "Estimation and validation of Gaussian process surrogate models for sensitivity analysis and design optimization: Based on the mechanistic–empirical pavement design guide," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2226, no. 1, pp. 119–126, 2011.

[36] G. Xiao, Z. Juan, and J. Gao, "Detecting travel modes using particular rules combined with a Naïve Bayesian classifier," presented at the Transp. Res. Board 93rd Annu. Meeting, Washington, DC, USA, 2014.

[37] F. J. Pontes, G. F. Amorim, P. P. Balestrassi, A. P. Paiva, and J. R. Ferreira, ''Design of experiments and focused grid search for neural network parameter optimization,'' *Neurocomputing*, vol. 186, pp. 22–34, Apr. 2016.
[38] J. Verwaeren, P. Van Der Weeën, and B. De Baets, ''A search grid for parameter optimization as a byproduct of model sensitivity analysis,'' *Appl. Math. Comput.*, vol. 261, pp. 8–27, Jun. 2015.
[39] P. A. Gonzalez, J. S. Weinstein, S. J. Barbeau, M. A. Labrador, P. L. Winters, N. L. Georggi, and R. Perez, ''Automating mode detection for travel behaviour analysis by using global positioning systems-enabled mobile phones and neural networks,'' *IET Intell. Transp. Syst.*, vol. 4, no. 1, pp. 37–49, 2010.
[40] C. Zhang, Y. Liu, W. Lu, and G. Xiao, ''Evaluating passenger satisfaction index based on PLS-SEM model: Evidence from Chinese public transport service,'' *Transp. Res. A, Policy Pract.*, vol. 120, pp. 149–164, Feb. 2019.
[41] C. Xu, M. Ji, W. Chen, and Z. Zhang, ''Identifying travel mode from GPS trajectories through fuzzy pattern recognition,'' presented at the 7th Int. Conf. Fuzzy Syst. Knowl. Discovery (FSKD), Yantai, China, 2010.
[42] Y. Bazi and F. Melgani, ''Gaussian process approach to remote sensing image classification,'' *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 1, pp. 186–197, Jan. 2010.
[43] M. Kim and F. De la Torre, ''Multiple instance learning via Gaussian processes,'' *Data Mining Knowl. Discovery*, vol. 28, no. 4, pp. 1078–1106, 2014.
[44] G. Haixiang, L. Yijing, J. Shang, G. Mingyun, H. Yuanyue, and G. Bing, ''Learning from class-imbalanced data: Review of methods and applications,'' *Expert Syst. Appl.*, vol. 73, pp. 220–239, May 2017.
[45] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, ''SMOTE: Synthetic minority over-sampling technique,'' *J. Artif. Intell. Res.*, vol. 16, no. 1, pp. 321–357, 2002.
[46] C. E. Rasmussen, ''Gaussian processes in machine learning,'' in *Advanced Lectures on Machine Learning*. Berlin, Germany: Springer, 2004, pp. 63–71.
[47] M. J. R. Fasham, H. W. Ducklow, and S. M. McKelvie, ''A nitrogen-based model of plankton dynamics in the oceanic mixed layer,'' *J. Mar. Res.*, vol. 48, no. 3, pp. 591–639, 1990.
[48] J. L. Bowman and M. Bradley, ''Testing spatial transferability of activity-based travel forecasting models,'' *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2669, no. 1, pp. 62–71, 2017.
[49] F. Zong, Y. Bai, X. Wang, Y. Yuan, and Y. He, ''Identifying travel mode with GPS data using support vector machines and genetic algorithm,'' *Information*, vol. 6, no. 2, pp. 212–227, 2015.

GUANGNIAN XIAO received the B.S. degree in industrial engineering and the M.S. degree in management science and engineering from Jiangnan University, Wuxi, China, in 2009 and 2012, respectively, and the Ph.D. degree in management from Shanghai Jiao Tong University, Shanghai, China, in 2016. He is currently an Assistant Professor with the School of Economics and Management, Shanghai Maritime University, Shanghai. His research interests include travel behavior modeling and pattern recognition issues related to GPS-based travel surveys.

QIN CHENG received the B.S. degree in information management and information system from the Anhui University of Finance and Economics, Bengbu, China, in 2018. She is currently pursuing the degree with the School of Economics and Management, Shanghai Maritime University, Shanghai, China. Her research interests include travel behavior modeling and shipping network modeling.

CHUNQIN ZHANG received the B.S. degree in traffic engineering from Shandong Jiao Tong University, Jinan, China, in 2008, the M.S. degree in transportation information engineering and control from Jilin University, Changchun, China, in 2011, and the Ph.D. degree in management from Shanghai Jiao Tong University, Shanghai, China, in 2015. She is currently an Assistant Professor with the School of Civil Engineering and Architecture, Zhejiang Sci-Tech University, Hangzhou, Zhejiang, China. Her research interests include travel behavior modeling and evaluation of public transport service.