IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received September 11, 2020, accepted September 17, 2020, date of publication September 25, 2020, date of current version October 28, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3026685

Trip Purpose Identification of Docked Bike-Sharing From IC Card Data Using a Continuous Hidden Markov Model

WENHAO LI[1,2], YANJIE JI[1], XIANQI CAO[3], AND XINYI QI[1]

[1]School of Transportation, Southeast University, Nanjing 211189, China
[2]School of Informatics, Computing, and Cyber Systems, Northern Arizona University, Flagstaff, AZ 86011, USA
[3]China Railway Siyuan Survey and Design Group Company, Ltd., Wuhan 430063, China

Corresponding author: Yanjie Ji (jiyanjie@seu.edu.cn)

This work was supported in part by the National Key Research and Development Program of China under Grant 2018YFE0120100, in part by the Less Developed Regions of the National Natural Science Foundation of China under Grant 71861006, in part by the Postgraduate Research&Practice Innovation Program of Jiangsu Province under Grant KYCX20_0137, and in part by the Fundamental Research Funds for the Central Universities under Grant 2242020K40063.

ABSTRACT It is different from the previous supervised learning algorithm based on personal travel questionnaire, the aim of this study is to develop an unsupervised learning methodology to estimate the docked bike-sharing users' trip purposes using IC card data, which trip purposes were unknown from the dataset. The present study is able to extract the trip-chains, which is used to understand the complete individual trip process. A rigorous method is then proposed to interpret the purpose of each leg of the trip-chain using a continuous hidden Markov model (CHMM). This method effectively combines the Gaussian mixture model and the hidden Markov model, and realizes the inference based on trip-chains. It is intended to enhance the understanding of docked bike-sharing users' transfer intention, which is different from most trip motivation recognition methods. The Gaussian mixture layer uses the feature space constructed by the spatial and temporal information on trip-chains from the IC card data, as well as the land-use characteristics of the docked bike-sharing docking stations to complete the transfer of the trip-chains to the trip modes. The hidden Markov structure can realize the process from the trip modes to the trip purposes. The IC card data of docked bike-sharing usage in Nanjing, China is used to interpret the specific steps of the proposed model. A questionnaire survey is conducted to obtain the real trip purposes, which is compared with the estimated results from the model to verify the effectiveness of the model. The results show that the accuracies of single trip recognition and chain trip recognition are 0.770 and 0.756, respectively. Compared with the baseline algorithm, the model also shows good performance. Therefore, the proposed approach can be used to discover and interpret the trip purpose using the IC card data.

INDEX TERMS Continuous hidden Markov model, IC card, docked bike-sharing, trip-chain, trip purpose.

I. INTRODUCTION

The docked bike-sharing system is a product of the city's promotion of "green and low-carbon transportation". It has the advantages of low-carbon environmental protection and the improvement of residents' awareness of green travel [1]. At the same time, the development of docked bike-sharing system is conductive to enhance the accessibility of road network, improving the accessibility of public transport, reducing the load of road network, effectively solving the problem of "the last kilometer" of public travel, easing traffic congestion, and meeting the short-distance travel needs of residents [2]–[4].

Since the docked bike-sharing project started in 1960s, after 40 years of continuous improvement, it has developed rapidly all over the world, and the research on the docked bike-sharing system has formed a certain theoretical basis [2], [5], [6]. Benefiting from the development of information and data technology, the urban intelligent transportation system can play a greater role [7]–[9]. At present, IC cards are used to rent and return docked bike-sharings. This provides a new way of collecting and analyzing data

The associate editor coordinating the review of this manuscript and approving it for publication was Chi-Hua Chen.

for research purposes [2]. Many scholars have used IC card data to study the travel characteristics of docked bike-sharings. For example, document [10] used the data of Beijing docked bike-sharing usage to understand the frequency and turnover rate of docked bike-sharings in different administrative regions to provide a basis for station optimization, and calculate the service radius of the station. Through a large amount of data mining and analysis of the docked bike-sharing usage, document [11] counted the frequency of docked bike-sharings borrowing and returning in different periods of weekdays and weekends, and studied the activity patterns of different types of stations combined with their spatial distribution. These studies are closely related to the travel characteristics and user preferences of users. In recent years, travel behavior analysis is an important research topic [12], [13]. As we all know, travel related data is an important and valuable source of comprehensive and in-depth understanding of travel behavior. By analyzing these data, urban planners and policy makers can improve their ability to solve urban planning, management and operation problems. Docked bike-sharing cycling activities has obvious space-time characteristics. For example, commuting and transfer activities have relatively specific use time and use area. If we can know the trip purpose of docked bike-sharing users in advance, we can effectively judge the travel demand of docked bike-sharing and solve the scheduling problem of different docking stations.

The research on the inference of trip purpose is based on the personal travel questionnaire, which requires respondents to record their trip purpose synchronously when filling in the trip survey, and the mathematical models use supervised learning methods to provide a clear trip purpose label. For example, document [14] developed a personal travel analysis system based on GIS. This system can collect travel trajectory information, and uses the prior knowledge of GIS to get the possible trip activity mode. Document [15] discussed the possibility of using GPS system to replace trip investigation completely, and proposed a model to extract a large number of trip data from GPS data, including trip purposes. All of these methods have several limitations, including short time and space coverage, high survey costs, heavy burden on respondents, and underreporting of trips (inaccurate) [16]. With the rapid development of big data technology, the passive acquisition of large-scale location data using time stamp (trajectory data) becomes easy to achieve both technically and economically. For example, GPS based trajectory data records the physical coordinates of moving objects, and smart card data records the location in the docking station. These data can clearly contain travel information, but generally lack a clear understanding of individual travel intention. In other words, although there are such unlabeled data, there is a lack of semantic tags for trip purposes. To solve this problem, some studies [17]–[20] believe that the trip purpose is limited by space, time, sequence of activities and the activity rules of other interviewees. Considering the unconventional behavior, these constraints are defined as soft constraints and used to infer the trip purpose. The feature of this method is to extract high-level semantics from the raw data and further use them to better understand the potential meaningful motion behavior. Quite a lot of technical terms are used to explain the trip purpose from activities after travel. These technologies mainly include deterministic and heuristic rules [21], clustering analysis [22]–[24] and activity-travel behavior analysis model [25]. However, these studies have the following two problems: 1) The research focuses on explaining the trip purpose at the overall level such as the city scale. In contrast, there are relatively few studies on inferring the trip purpose at the individual level. 2) Most studies are based on a single trip to make inferences and do not have a full-day tracking investigation of activity trajectories and correlation analysis between various activities during the trip, resulting in the separation between the various travel activities in the user's one-day trip-chain.

In response to the above research goals and challenges, the main contributions of this paper are twofold.

1) This paper overcomes the shortcomings of supervised learning methods that require additional collection of users' accurate travel purpose data. A reliable unsupervised learning method using only IC card data is proposed to identify the trip purpose of docked bike-sharing usage. Simultaneously compared with the traditional questionnaire survey, it is of great significance to propose this data-driven trip purpose identification method, which can reduce the manpower and material consumption in the process of data collection and improve the efficiency of urban traffic information collection resources. By introducing unsupervised learning mechanism, this method can infer the probability that a traveler will have a certain trip purpose under certain conditions without having to obtain an accurate activity label in advance.
2) The proposed methodology is different from the existing trip motivation recognition methods. This model takes the trip-chain as the inference unit. For docked bike-sharing users, there is a high correlation between multiple riding activities in a single day. The stitching of each trip record of docked bike-sharing users into a docked bike-sharing trip-chain is beneficial to further reveal the travel characteristics of docked bike-sharing users. Previous studies have focused on inferring the type of activity of a specific traveler's single trip, while this study focuses on predicting the next activity when the current activity is known. Inferring through the correlation between activities may have better results. Besides, in one dimension of traffic planning and policy making, it is meaningless to estimate the exact trip purpose of a specific individual. It is enough to know the probability that the traveler has a specific purpose under certain conditions. Compared with the single activity result, this method can clearly see the dynamic process of travelers from one activity to another, which is more practical.

This paper is formatted as follows. Section II introduces the basic concept of docked bike-sharing trip-chain and how to obtain trip-chains through IC card data. Section III presents the continuous hidden Markov model and its application in estimating the trip purpose behind the trip-chain, and the algorithm for solving the model. Section IV demonstrates the process of estimating model parameters through specific data explains the results, and verifies the training effect of the model through the data of the questionnaire. Finally, in the last section, a conclusion and further research on the model deficiency are provided.

II. DOCKED BIKE-SHARING TRIP-CHAIN EXTRACTION FROM IC CARD DATA

A. DEFINITION OF DOCKED BIKE-SHARING TRIP-CHAIN

Traditional trip-chain refers to multi-objective continuous activities in a certain time range [26]–[28]. Considering each borrow and return of a docked bike-sharing as an activity, then it is reasonable to arrange all related activities in chronological order to form a closed-loop round trip (trip-chain) as a complete travel activity. The docked bike-sharing trip-chain mentioned in this paper refers to the chain records of a series of trips completed by users using docked bike-sharings in a day, which starting at one docking station and ending at the same or a different station. The trip purpose (i.e., commuting chain, entertainment chain, and mixed chain) and complexity of the chain (i.e., single-circle chain and multiple-circle chain) are the main factors that divide trip-chain into different categories [29]. Considering that IC card data cannot provide users' travel intention, this paper classifies the trip-chain into five categories based on the relationship between borrowing docking station and returning docking station and trip-chain mode. We use symbols to represent the trip-chain as follows, where "O" denotes the borrowing docking station, "D" denotes the returning docking station, "sD" denotes a trip-chain contains only one borrowing behavior stopping the journey, "mD" denotes a trip-chain contains multiple borrowing behaviors stopping the journey:

O-O: users borrow and return their docked bike-sharings at the same bike-sharing docking station

O-sD-O: users borrow their docked bike-sharings from a bike-sharing docking station and return the bike to the same docking station, but there is one borrowing and returning behavior from a different origin on the trip-chain.

O-mD-O: users borrow their docked bike-sharings from a bike-sharing docking station and return the bike to the same docking station, but there is more than one borrowing and returning behavior from a different origin on the trip-chain.

O-sD: users borrow public bikes from one bike-sharing docking station and return them to another

O-mD: users borrow public bikes from one bike-sharing docking station and then return them to another docking station, but there is more than one borrowing and returning behavior from a different origin on the trip-chain.

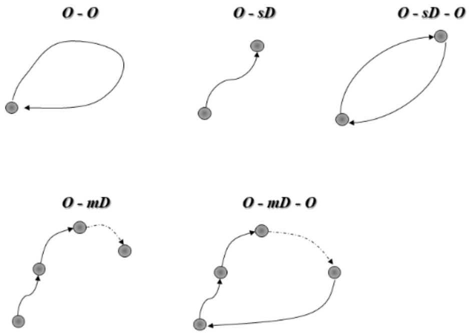

FIGURE 1. The types of docked bike-sharing trip-chain.

B. THE EXTRACTION METHOD OF DOCKED BIKE-SHARING TRIP-CHAIN

In the original IC card data, a complete docked bike-sharing record includes 10 parts: card number, docked bike-sharing number, borrowing time, borrowing docking station number, borrowing docking station name, borrowing docks number, returning time, returning docking station number, returning docking station name and returning docks number. According to the data structure of IC card and the previous definition of docked bike-sharing trip-chain, this paper proposes a scientific and effective method to extract the trip-chains of docked bike-sharing, which is described as follows.

Step I. Data cleaning and feature extraction. Clean out the IC card records with incomplete information, and then extract the effective information needed for docked bike-sharing trip-chain, including: card number, traveler's personal attribute information, borrowing time, borrowing docking station information, returning time, and returning docking station information.

Step II. Data grouping. Arrange the records of each card in chronological order, count the number of swiping times of each card in a day, and group them according to the number of swiping times. n denotes the number of swiping times. d_n denotes the group with n swipes.

Step III. Trip-chain split. There are two ways to define the disconnection of trip records in advance: (1) the borrowing docking station of any record is the same as the returning docking station. (2) the returning docking station of the previous record is different from the borrowing docking station of the latter record, and then establish the pointer variables P_1 and P_2, in which P_1 points to the first record of the pending record, P_2 points to the second record of the pending record. If the record corresponding to P_2 is not disconnected from the previous record, move P2 to the next record. If the record corresponding to P_2 is disconnected from the previous record, move P_2 back to the previous record, and takes all the records between P_1 and P_2 as trip-chain.

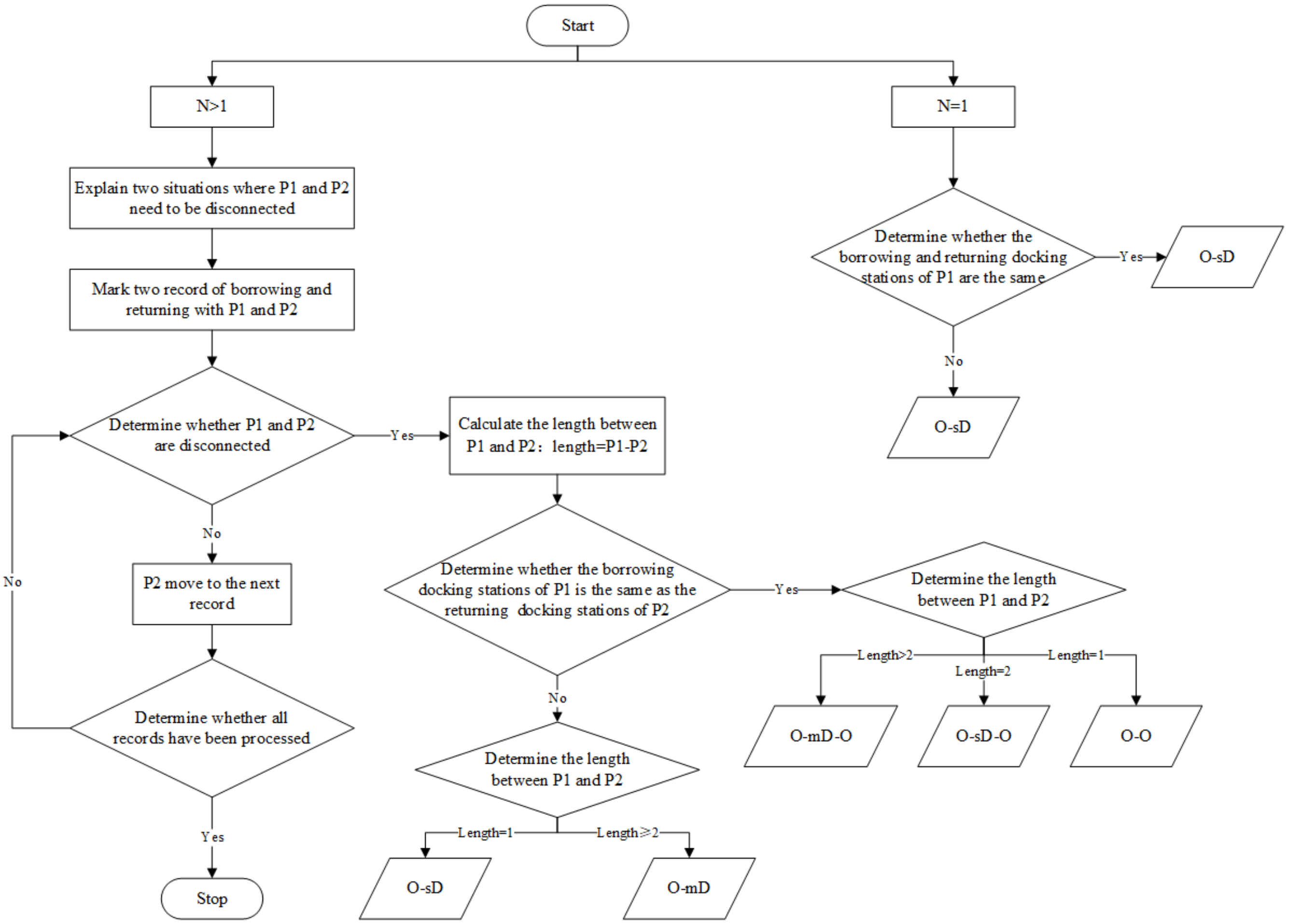
Start
N>1
N=1
Explain two situations where P1 and P2 need to be disconnected
Mark two record of borrowing and returning with P1 and P2
Determine whether P1 and P2 are disconnected
Yes
Calculate the length between P1 and P2: length=P1-P2
No
P2 move to the next record
No
Determine whether all records have been processed
Yes
Stop
Determine whether the borrowing docking stations of P1 is the same as the returning docking stations of P2
Yes
Determine the length between P1 and P2
Length>2
Length=2
Length=1
O-mD-O
O-sD-O
O-O
No
Determine the length between P1 and P2
Length=1
Length⩾2
O-sD
O-mD
Determine whether the borrowing and returning docking stations of P1 are the same
Yes
O-sD
No
O-sD

FIGURE 2. The extraction process of docked bike-sharing trip-chain.

Step IV. Trip-chain discrimination. The type of trip-chain is determined according to the number of records of the trip-chain and whether the borrowing docking station of P_1 is the same as the returning docking station of P_2. The specific process is shown in Figure 2.

C. THE EXTRACTION PROCESS OF DOCKED BIKE-SHARING TRIP-CHAIN

The data used in this study is the IC card data of docked bike-sharing usage in the main city zone of Nanjing, China on March 2, 2016, and was provided by Nanjing docked bike-sharing company. After eliminated the incomplete records, blank records and dispatcher's records, a total of 91739 records were taken as the original data set, which was processed by the trip-chain extraction method described earlier. A total of 70568 docked bike-sharing trip-chains were obtained for analysis and the quantity information on docked bike-sharing trip-chain types is shown in Table 1.

The *O-D* trip-chain (*O-sD* and *O-mD*) are the most common types that take up 79.37% of all trip-chains. The high ratings of *O-sD* trip-chains (68.69%) are mainly due to the fact that there are many "improvised-bike" and "one-way" bike-sharing users, but other reasons, such as the shortage of "D" bicycles, may also be caused by this. The regular bike-sharing trip-chains *O-sD-O* takes up 12.81% of all chains, whilst the complex *O-mD-O* trip-chain presents only 1.05% of all chains in our data. Finally, the *O-O* trip-chain account for 6.76% of all trip-chains. Even if we limit the travel time to between 2 minutes and 120 minutes, the results of the *O-O* trip-chain are still reasonable, because some users may use docked bike-sharing for exercise (or other purposes) in daily life, which can be regarded as flexible cycling activities. Furthermore, we choose representative O-O chain and o-sd chain for spatio-temporal analysis. The results are shown in Figure 3 and Figure 4. Figure 3 shows the travel impact range of different docking stations. The docking stations with wide coverage represent more frequency of use, and there are obvious traffic attraction points in the areas where the docking stations are frequently used. To some extent, this has a strong correlation with the land use of the region. From the OD flow reflected in Figure 4, it can be seen that the positive and negative flow of cycling between some adjacent docking stations has obvious symmetry. It can be inferred that there

TABLE 1. Description of docked bike-sharing trip-chain types.

Type of trip-chain	Frequency	Percentage
O-O	4772	6.76
O-sD-O	9041	12.81
O-mD-O	739	1.05
O-sD	48476	68.69
O-mD	7540	10.68
Overall	70568	100

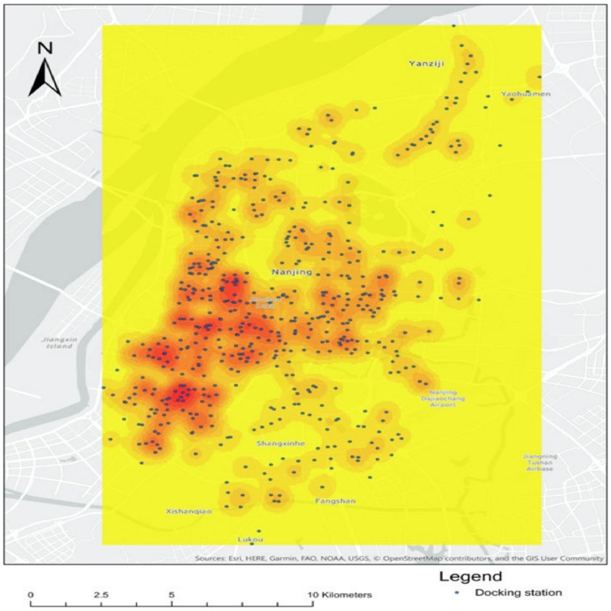

FIGURE 3. **The Influence scope of docking station where *O-O* trip-chains occurred.**

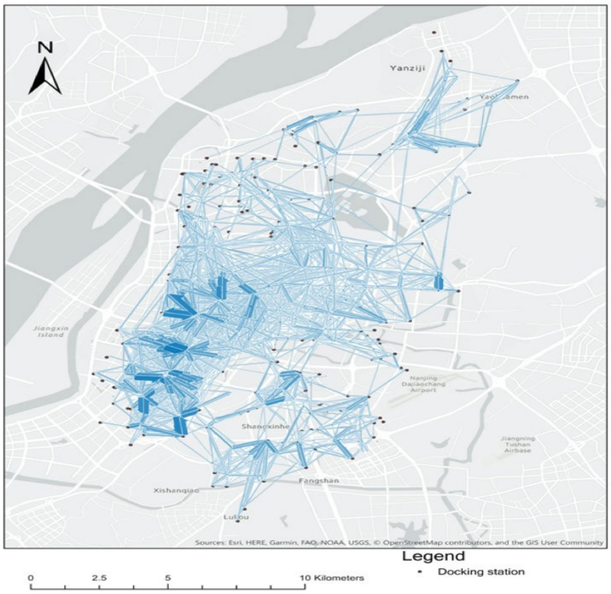
N
Yanziji
Yaohuamen
Jiangxin Island
Nanjing Dajiaochang Airport
Shangxinhe
Jiangning Tushan Airbase
Xishanqiao
Fangshan
Lukou
Sources: Esri, HERE, Garmin, FAO, NOAA, USGS, © OpenStreetMap contributors, and the GIS User Community
Legend
Docking station
0
2.5
5
10 Kilometers

FIGURE 4. **The visual analytics of *O-sD* trip-chains by the number of docked bike-sharing borrowed and returned at different docking station.**

may be some activity connection between adjacent docking stations to promote the bidirectional movement of cycling. Compared with the isolation of single activity, mining these chain rules based on trip-chains may improve the accuracy of activity reasoning.

III. CONTINUOUS HIDDEN MARKOV MODEL (CHMM) FOR ACTIVITY PURPOSE IDENTIFICATION

A. MODELING OF CHMM

The Hidden Markov Model (HMM) [30], [31] is a statistical model that is used to describe a Markov chain with hidden unknown states. Its hidden state cannot be observed directly, but can be inferred from the observation sequence. Each observation vector is represented by the probability density distribution of state members. Considering the characteristics of sequence reasoning of hidden Markov model, it is used as trip-chain inference. Its feature is used to mine the mapping relationship between the trip-chains extracted from the docked bike-sharing IC card data and its trip purpose. However, the probability distribution of trip chain characteristics is continuous, and HMM model is only suitable for discrete probability distribution. In order to overcome this problem, the HMM and Gaussian mixture model [32] are fused together to form a continuous hidden Markov model (CHMM) in this paper. The innovation of the model is to realize the discretization of trip-chain from continuous feature to travel mode by Gaussian mixture model, and construct Gaussian mixture model library of different travel modes to represent the space-time characteristics of trip-chain. The CHMM is assumed that each observation sequence is generated from its hidden state in regular order, but there is a transition layer between the observation layer and the state layer, which has a specific meaning and is a clustering layer of the observation sequence feature space, forming a three-layer hierarchical structure as shown in Figure 5. This model extracts trip-chains (observation sequence) from IC card data, including travel time, origin-destination information, traveler attributes and other characteristics. The docked bike-sharing riding pattern (hidden cluster) is deduced from the feature space constructed from these variables, and the observation sequence is transformed into the riding pattern sequence. This process is implemented by a Gaussian mixture model. For CHMM, the Gaussian mixture model is responsible for the input of the Markov chain, and the hidden state in each active mode has the output probability belonging to a specific feature space cluster [17]. Upon completing the above process, the hidden trip purpose sequence (hidden state) behind the generated riding pattern sequence is identified. Next, we will describe the detailed modeling process.

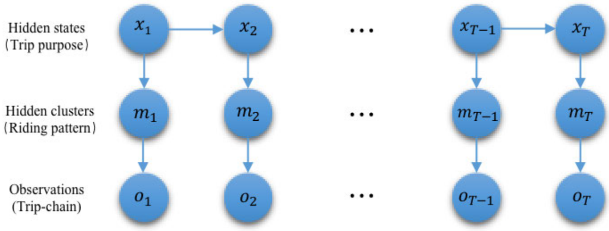

FIGURE 5. The delivery process of determining states and observations behind a CHMM.

The number of possible values of a state variable is determined before modeling. Eq. (1) denotes the probability set that an initial state belongs to a travel activity.

$$\pi = \{\pi_i\} = \{p(x_1 = i)\}, \quad i = 1, 2, \cdots, N \qquad (1)$$

where π is an initial probability vector, and π_i is the probability that the first state is the i^{th} activity purpose. x_1 denotes the initial state variable of the activity purpose sequence of a trip-chain. N is the number of possible activities obtained from the hidden state.

In this model, the trip-chain is transformed into a sequence of states. In other words, the state within the trip-chain is assumed to depend only on the previous state, following the Markov process. Eq. (2) denotes the transition probability matrix between two continuous states.

$$A = \left\{a_{ij}\right\} = \{P(x_t = j \mid x_{t-1} = i)\}, \quad i = 1, 2, \cdots N, \ j = 1, 2, \cdots N \qquad (2)$$

where A is a $N \times N$ matrix with transition probability. x_t represents the t^{th} state of the activity purpose sequence of the trip-chain. a_{ij} is the transition probability that the t^{th} state selects activity purpose j when the $(t-1)^{th}$ state is given as activity purpose i.

A Gaussian mixture model is a simple weighted sum of K Gaussian densities, each represents a component that corresponds to the output probability $b_i(o_t)$ in the CHMM mode as shown in Eq. (3).

$$b_i(o_t) = \sum_{k=1}^{K} g_{ik} f(o_t \mid \mu_{ik}, \delta_{ik}) = \sum_{k=1}^{K} g_{ik} \frac{1}{\sqrt{2\pi}\delta_{ik}} e^{\frac{-(\theta_i-\mu_k)^2}{2\delta_k^2}}, \quad i = 1, 2, \cdots N \qquad (3)$$

where g_{ik} is the probability that an observation belongs to cluster k when the state is activity purpose i. $f\left(o_t \mid \mu_{ik}, \delta_{ik}^2\right)$ is the Gaussian probability density formula. o_t is an observation of the t^{th} state in the sequence. μ_{ik} is the mean of the eigenvectors of the kth cluster of activity purpose i. δ_{ik} is the variance-covariance matrix of the kth cluster of the activity purpose i. K is the number of hidden clusters in a feature space.

Behind this simple form, there are hidden feature variables for each observation. In this study, the feature space is composed of 7 feature variables: age of docked bike-sharing users, travel time of activity, type of borrow/return and 4 land-use characteristics of docked bike-sharing station. The cluster of each observation is latent in the model, i.e. an observation belongs to which component is unknown. The N × K weight matrix (G) can represent the probability matrix of hidden clusters, and the formula is as follows.

$$G = \{g_{ik}\} = \{p(m_t = k | x_t = i)\} \quad i, k = 1, 2, \cdots, N \qquad (4)$$

where m_t represents the hidden cluster in the characteristic space of the t^{th} hidden state.

Eq. (5) stands for the probability density of observations, given that its latent cluster is known. In order to simplify the model, each state shares a common cluster [see Eq. (6)] was assumed. Integrating the three formulas [Eqs. (4)-(6)] for all possible values of o_t leads back to Eq. (3).

$$p(o_t \mid m_t) = f(o_t \mid \mu_{ik}, \delta_{ik}) \quad i, k = 1, 2, \cdots, N \qquad (5)$$

$$\mu_{ik} = \mu_k \ and \ \delta_{ik} = \delta_k \quad i, k = 1, 2, \cdots, N \qquad (6)$$

The delivery process of determining states and observations behind a CHMM is introduced as follows. First, according to the transition probabilities based on the previous state, the state sequence of activity purpose is inferred for a trip-chain. The second step is to select a hidden cluster for the state according to the transfer probability between cluster and state. Finally, observations are drawn based on the determined cluster from the Gaussian probability distribution with the cluster's mean and variance-covariance matrix. This procedure is shown in Figure 3.

As mentioned above, the Gaussian mixture model classifies the sample trip-chain into predefined hidden clusters, and then the CHMM model parameters are estimated by establishing the likelihood function of the observation sequence of trip-chain, which is represented by a vector set $\lambda = (\pi, A, G)$. Next, we build the corresponding maximum likelihood function.

$$\begin{aligned} L(\lambda) &= P(o_1 \cdots o_T \mid \lambda) \\ &= \sum_{all\ possible\ x_1 \cdots x_T} P(o_1 \cdots o_T | x_1 \cdots x_T, \lambda) p(x_1 \cdots x_T | \lambda) \\ &= \sum_{all\ possible\ x_1 \cdots x_T} \left(\prod_{t=1}^{T} p(o_t | x_t, G)\right)\left(\prod_{t=1}^{T} p(x_t | x_{t-1}, A)\right) \\ &= \sum_{all\ possible\ x_1 \cdots x_T} \left(\prod_{t=1}^{T}\left(\sum_{k=1}^{K} g_{x_t k} f(o_t \mid \mu_k, \delta_k)\right)\right) \\ &\quad \times \left(\prod_{t=1}^{T} a_{x_{t-1} x_t}\right) \end{aligned} \qquad (7)$$

where T is the length of the trip-chain activity purpose sequence.

According to Eq. (7), the likelihood function is extended to adapt the parameter estimation of multiple trip-chains of different lengths. The extended function as shown below.

$$\begin{aligned}\hat{L}(\lambda) &= P\left(o_1 \cdots o_T \mid \lambda\right)\\ &= \prod_{l=1}^{M}\left\{\sum_{all\,possible\,x_1\cdots x_T}\left(\prod_{t=1}^{T^l}\left(\sum_{k=1}^{K} g_{x_t k} f\left(o_t^l \mid \mu_k, \delta_k\right)\right)\right)\right.\\ &\qquad \left.\times\left(\prod_{t=1}^{T^l} a_{x_{t-1}x_t}\right)\right\} \end{aligned} \tag{8}$$

where o_t^l is an observation of the tth state in the activity purpose sequence of the lth trip-chain. T^l is the length of the activity purpose sequence of the lth trip-chain.

B. SOLUTION OF MODEL

In this model, the following questions need to be answered. How to estimate the best parameters through a series of observations or multiple observed series (multiple trip-chains)? How to derive the most likely sequence of hidden state, given the characteristics of observation sequence and model parameters? Regarding a CHMM, there has been mature technology to answer these two questions. The present study adopted the Baum-Welch algorithm and forward-backward algorithm to estimate the parameters of a CHMM, and used the Viterbi algorithm to infer the sequence of borrowing and returning activities for a trip-chain based on both the observations and the estimated parameters.

Solving the extended likelihood function established above is extremely difficult. It contains the sum of the individual likelihood on each possible assignment of state, which makes it impossible to use the traditional method such as the Newton-Raphson. The Baum-Welch algorithm, a training methodology for a CHMM, is developed as a powerful tool for solving a maximization problem when latent variables are involved. The algorithm transforms the original maximization of the log-likelihood function of multiple integrals (or sums) containing all possible potential variables into a simple recursive process. Each iteration of the Baum-Welch algorithm consists of two steps: Expectation and Maximization. A proxy function $Q\left(\lambda, \lambda^i\right)$ is instead of the original log-likelihood function in the Expectation step. $Q\left(\lambda, \lambda^i\right))$ is the expectation of conditional probability distribution $P\left(x \mid \boldsymbol{o}, \lambda^{(i)}\right)$ of the likelihood function of complete data $log\left[P\left(\boldsymbol{o}, x \mid \lambda\right)\right]$ with respect to the unmeasured data x, which is restricted to a premeasured data $\boldsymbol{o}$ and current parameter $\lambda^{(i)}$. In practice, the $P\left(x \mid \boldsymbol{o}, \lambda^i\right)$ is transformed to $P\left(\boldsymbol{o}, x \mid \lambda^i\right)$ using the Bayes' theorem $P\left(X \mid Y\right) = P\left(X, Y\right)/P\left(Y\right) \propto P\left(X, Y\right)$. The expression of the $Q\left(\lambda, \lambda^i\right)$ function is as follows, in which Eqs. (9) is for a single trip-chain, and Eqs. (10) is for multiple trip-chains. In the Maximization step, we find the λ that maximizes the $Q\left(\lambda, \lambda^{(i)}\right)$ function, and determine the parameter estimated value $\lambda^{(i+1)}$ of the of $i+1$ iteration. The expression is shown in Eqs. (11).

$$\begin{aligned}Q\left(\lambda, \lambda^i\right) = \sum_{all\,possible\,x_1\cdots x_T} & log\left[P\left(\boldsymbol{o}_1 \cdots \boldsymbol{o}_T, x_1 \cdots x_T \mid \lambda^i\right)\right.\\ &\left.\times P\left(\boldsymbol{o}_1 \cdots \boldsymbol{o}_T, x_1 \cdots x_T \mid \lambda\right)\right]\end{aligned} \tag{9}$$

$$\begin{aligned}Q\left(\lambda, \lambda^i\right) = \sum_{j=1}^{M}\sum_{all\,possible\,x_1\cdots x_T} & log\Big[P\Big(\boldsymbol{o}_1^j \cdots \boldsymbol{o}_{T^j}^j, x_1 \cdots x_{T^j} \mid \lambda^i\Big)\\ &\times P\left(\boldsymbol{o}_1^j \cdots \boldsymbol{o}_{T^j}^j, x_1 \cdots x_{T^j} \mid \lambda\right)\Big]\end{aligned} \tag{10}$$

$$\lambda^{i+1} = \arg\max_{\lambda} Q\left(\lambda, \lambda^i\right) \tag{11}$$

Baum-Welch algorithm [31] is the concrete implementation of EM algorithm in CHMM model. Compared with EM algorithm, it can complete the constraint of estimation parameters. In our model, the sum of the elements of the vector, each column of the transition matrix and each row of the membership matrix should be 1. Therefore, in the Maximization step, the Lagrangian relaxation of the original proxy function is required as shown in Eqs. (12).

$$\begin{aligned}L\left(\lambda, \lambda^i\right) = Q\left(\lambda, \lambda^i\right) - \mu_\pi \cdot \left(\sum_{i=1}^{N}\pi_i - 1\right) - \sum_{i=1}^{N}\mu_i^A\\ \cdot\left(\sum_{j=1}^{N}a_{ij} - 1\right) - \sum_{i=1}^{N}\mu_i^G \cdot \left(\sum_{k=1}^{K}g_{ik} - 1\right)\end{aligned} \tag{12}$$

The first-order condition of Lagrangian function [Eqs. (18)] is that the derivative of the function to each original parameter $(\pi_i, a_{ij}, g_{ik}, \mu_{kd}, \sigma_{kd_1d_2})$ and to each Lagrangian multiplier should be zero, which provides a ready-made parameter solution set in each iteration of Baum-Welch algorithm. The best parameters $(\hat{\pi}_i, \hat{a}_{ij}, \hat{g}_{ij}, \hat{\mu}_{kd}, \hat{\sigma}_{kd_1d_2})$ in the iteration of the Baum Welch algorithm derived from the first-order condition of the Lagrangian function is summarized as follows.

$$\hat{\pi}_i = \sum\nolimits_{j=1}^{M} P\left(x_1 = i \mid \boldsymbol{o}_1^j \cdots \boldsymbol{o}_{T^j}^j, \lambda^i\right) \Big/ M, \quad for\ i = 1, \cdots, N \tag{13}$$

$$\hat{a}_{ij} = \frac{\sum_{j=1}^{M}\sum_{t=1}^{T^j-1} P\left(x_t = i, x_{t+1} = j \mid \boldsymbol{o}_1^j \cdots \boldsymbol{o}_{T^j}^j, \lambda^i\right)}{\sum_{j=1}^{M}\sum_{t=1}^{T^j-1} P\left(x_t = i \mid \boldsymbol{o}_1^j \cdots \boldsymbol{o}_{T^j}^j, \lambda^i\right)}, \quad for\ i = 1, \cdots, N\ and\ j = 1, \cdots, N \tag{14}$$

$$\hat{g}_{ik} = \frac{\sum_{j=1}^{M}\sum_{t=1}^{T^j-1} P\left(x_t = i, m_t = k \mid \boldsymbol{o}_1^j \cdots \boldsymbol{o}_{T^j}^j, \lambda^i\right)}{\sum_{j=1}^{M}\sum_{t=1}^{T^j-1} P\left(x_t = i \mid \boldsymbol{o}_1^j \cdots \boldsymbol{o}_{T^j}^j, \lambda^i\right)}, \quad for\ i = 1, \cdots, N\ and\ k = 1, \cdots, K \tag{15}$$

$$\hat{\mu}_{kd} = \frac{\sum_{j=1}^{M}\sum_{t=1}^{T^j-1} P\left(x_t = i, m_t = k \mid \boldsymbol{o}_1^j \cdots \boldsymbol{o}_{T^j}^j, \lambda^i\right) \cdot o_{td}^j}{\sum_{j=1}^{M}\sum_{t=1}^{T^j-1} P\left(x_t = i, m_t = k \mid \boldsymbol{o}_1^j \cdots \boldsymbol{o}_{T^j}^j, \lambda^i\right)}, \quad for\ k = 1, \cdots, K\ and\ d = 1, \cdots, D \tag{16}$$

where $\boldsymbol{o}_t^j = \left(\boldsymbol{o}_{t1}^j, \cdots, \boldsymbol{o}_{tD}^j\right)'$, $\hat{\mu}_k = \left(\hat{\mu}_{k1}, \cdots, \hat{\mu}_{kD}\right)'$, and D denotes different observed feature vector (17), as shown at the bottom of the next page.

The covariance matrix of different feature vector can be expressed as:

$$\hat{\boldsymbol{\Sigma}}_{\boldsymbol{k}} = \begin{bmatrix} \hat{\sigma}_{K1}^2 & \cdots & \hat{\sigma}_{kd_1d_2} \\ \vdots & \ddots & \vdots \\ \hat{\sigma}_{kd_1d_2} & \cdots & \hat{\sigma}_{KD}^2 \end{bmatrix} \tag{18}$$

Eqs. (13)-(18) contains three probability terms ($P\left(x_t = i, x_{t+1} = j \mid \boldsymbol{o}_1^j \cdots \boldsymbol{o}_{Tj}^j, \lambda^i\right)$, $P\left(x_t = i \mid \boldsymbol{o}_1^j \cdots \boldsymbol{o}_{Tj}^j, \lambda^i\right)$, and $P\left(x_t = i, m_t = k \mid \boldsymbol{o}_1^j \cdots \boldsymbol{o}_{Tj}^j, \lambda^i\right)$) which cannot be calculated directly when the length of the active sequence or the number of possible hidden states becomes large. The forward-backward algorithm [33] can solve this problem. By establishing the forward and backward variables ($\alpha_t(j)$ and $\beta_t(i)$), we can get a simple method to calculate $P(x_t, x_{t+1} \mid \boldsymbol{o}_1 \cdots \boldsymbol{o}_T, \lambda)$. The calculation methods of the three probability terms mentioned above are as follows Eqs. (19) - (21).

$$P\left(x_t = i, x_{t+1} = j \mid \boldsymbol{o}_1^j \cdots \boldsymbol{o}_{Tj}^j, \lambda^i\right) = \alpha_t^j(i)\, a_{ij}\beta_{t+1}^j(j) b_j\left(\boldsymbol{o}_{t+1}^j\right) \Big/ \sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_t^j(i)\, a_{ij}\beta_{t+1}^j(j) b_j\left(\boldsymbol{o}_{t+1}^j\right) \tag{19}$$

$$P\left(x_t = i \mid \boldsymbol{o}_1^j \cdots \boldsymbol{o}_{Tj}^j, \lambda^i\right) = \alpha_t^j(i)\, \beta_t^j(i) \Big/ \sum_{i=1}^{N} \alpha_t^j(i)\, \beta_t^j(i) \tag{20}$$

where $\alpha_t^j(i)$ denotes forward variables, and $\beta_t^j(i)$ denotes backward variables both computed for the jth trip-chain. The forward and backward variables stand for $P\left(\boldsymbol{o}_{t+1} \cdots \boldsymbol{o}_T x_t = j \mid x_t = i, \lambda\right)$ and $P(x_t = j \mid \boldsymbol{o}_1 \cdots \boldsymbol{o}_T, \lambda)$, respectively.

$$P\left(x_t = i, m_t = k \mid \boldsymbol{o}_1^j \cdots \boldsymbol{o}_{Tj}^j, \lambda^i\right) = P\left(x_t = i \mid \boldsymbol{o}_1^j \cdots \boldsymbol{o}_{Tj}^j, \lambda^i\right) \cdot g_{ik} f\left(\boldsymbol{o}_t^j \mid \mu_k, \sigma_k\right) \Big/ b_i\left(\boldsymbol{o}_t^j\right) \tag{21}$$

The following describes the complete calculation process of Baum-Welch algorithm.

Step I. For $\mathrm{n} = 0$, choose $\pi_{\mathrm{i}}^{(0)}$, $\mathrm{a}_{\mathrm{ij}}^{(0)}$, $\mathrm{g}_{\mathrm{ik}}^{(0)}$ randomly.

Step II. Recursion for $\mathrm{n} = 1, 2, \cdots$ for EM algorithm

(1) compute $\alpha_{\mathrm{t}}^{\mathrm{j}}(\mathrm{i})$ and $\beta_{\mathrm{t}}^{\mathrm{j}}(\mathrm{i})$ for each trip-chain though forward-backward algorithm

(2) compute $\mathrm{P}\left(\mathrm{x_t} = \mathrm{i}, \mathrm{x_{t+1}} = \mathrm{j} \mid \boldsymbol{o}_1^{\mathrm{j}} \cdots \boldsymbol{o}_{\mathrm{Tj}}^{\mathrm{j}}, \lambda^{\mathrm{i}}\right)$, $\mathrm{P}\left(\mathrm{x_t} = \mathrm{i} \mid \boldsymbol{o}_1^{\mathrm{j}} \cdots \boldsymbol{o}_{\mathrm{Tj}}^{\mathrm{j}}, \lambda^{\mathrm{i}}\right)$, and $\mathrm{p}\left(\mathrm{x_t} = \mathrm{i}, \mathrm{m_t} = \mathrm{k} \mid \boldsymbol{o}_1^{\mathrm{j}} \cdots \boldsymbol{o}_{\mathrm{Tj}}^{\mathrm{j}}, \lambda^{\mathrm{i}}\right)$ using Eqs. (19)-(21).

(3) compute the incumbent parameters $\pi_{\mathrm{i}}^{(\mathrm{n})}$, $\mathrm{a}_{\mathrm{ij}}^{(\mathrm{n})}$, $\mathrm{g}_{\mathrm{ik}}^{(\mathrm{n})}$, $\mu_{\mathrm{kd}}^{(\mathrm{n})}$, and $\sigma_{\mathrm{kd_1d_2}}^{(\mathrm{n})}$ using Eqs. (13)-(17).

(4) convergence test: If the difference between each estimated parameter in the current iteration and that in the previous iteration is within the threshold value of 10^{-7}, escape the loop and stop the procedure. Otherwise, repeat the loop until convergence.

Once the learning process of CHMM is completed with Baum Welch algorithm, Viterbi algorithm [34] can be used to derive the most likely sequence of trip-chain purpose from the estimated parameters and observed characteristic data. In fact, Viterbi algorithm uses dynamic programming to find the maximum probability path (the optimal path), in which each path corresponds to a state sequence. The maximum probability value of all single paths $(x_1, x_2, \cdots, x_t)$ with state i at time t is defined as $\delta_t(i)$, and the $t-1$ node of all single paths $(x_1, x_2, \cdots, x_{t-1}, x)$ with state i at time t is defined as $\psi_t(i)$. The expression of the two variables is as follows.

$$\delta_t(i) = \max_{x_1, \cdots, x_{t-1}} P\left(x_t = i, x_{t-1} \cdots x_1, \boldsymbol{o}_t \cdots \boldsymbol{o}_1 \mid \lambda\right), \quad i = 1, \cdots, N \tag{22}$$

$$\begin{aligned} \delta_{t+1}(i) &= \max_{x_1, \cdots, x_t} P\left(x_{t+1} = i, x_{t-1} \cdots x_1, \boldsymbol{o}_t \cdots \boldsymbol{o}_1 \mid \lambda\right) \\ &= \max_{1 \le j \le N} \left[\delta_t(j) \cdot a_{ji}\right] \cdot b_i\left(\boldsymbol{o}_{t+1}\right), \\ & \qquad i = 1, \cdots, N, t = 1, \cdots, T-1 \end{aligned} \tag{23}$$

$$\psi_t(i) = arg \max_{1 \le j \le N} \left[\delta_{t-1}(j) \cdot a_{ji}\right], i = 1, \cdots, N \tag{24}$$

The following describes the full calculation process of Viterbi algorithm.

Step I. Initialization

$$\delta_1(\mathrm{i}) = \pi_{\mathrm{i}} \mathrm{b}_{\mathrm{i}}(\boldsymbol{o}_1), \quad \text{for i} = 1, 2, \cdots, \mathrm{N}$$
$$\psi_1(\mathrm{i}) = 0, \quad \text{for i} = 1, 2, \cdots, \mathrm{N}$$

Step II. Recursion for all $\mathrm{t} = 2, \cdots, \mathrm{T}$ and all $\mathrm{i} = 1, \cdots, \mathrm{N}$

$$\delta_{\mathrm{t}}(\mathrm{i}) = \max_{1 \le \mathrm{j} \le \mathrm{N}} \left[\delta_{\mathrm{t-1}}(\mathrm{j})\, \mathrm{a_{ij}}\right] \mathrm{b_i}(\boldsymbol{o}_{\mathrm{t}}), \quad \text{for i} = 1, 2, \cdots, \mathrm{N}$$
$$\psi_{\mathrm{t}}(\mathrm{i}) = \arg\max_{1 \le \mathrm{j} \le \mathrm{N}} \left[\delta_{\mathrm{t}}(\mathrm{j})\, \mathrm{a_{ij}}\right], \quad \text{for i} = 1, 2, \cdots, \mathrm{N}$$

Step III. Termination

$$\mathrm{P}^*\left(\boldsymbol{o}_1 \cdots \boldsymbol{o}_{\mathrm{T}} \mid \lambda\right) = \max_{1 \le \mathrm{j} \le \mathrm{N}} \delta_{\mathrm{T}}(\mathrm{i})$$
$$\mathrm{x}_{\mathrm{T}}^* = \arg\max_{1 \le \mathrm{j} \le \mathrm{N}} \delta_{\mathrm{T}}(\mathrm{i})$$

Step IV. Backtracking of optimal state sequence

$$\mathrm{x}_{\mathrm{t}}^* = \psi_{\mathrm{t+1}}\left(\mathrm{x}_{\mathrm{t+1}}^*\right), \quad \text{for t} = \mathrm{T} - 1, \cdots, 1$$

$$\hat{\sigma}_{kd_1d_2} = \frac{\sum_{j=1}^{M}\sum_{t=1}^{Tj-1} P\left(x_t = i, m_t = k \mid \begin{matrix} \boldsymbol{o}_1^j \\ \cdots \boldsymbol{o}_{Tj}^j, \lambda^i \end{matrix}\right) \cdot \left(o_{td_1}^j - \hat{\mu}_{kd_1}\right) \cdot \left(o_{td_2}^j - \hat{\mu}_{kd_2}\right)}{\sum_{j=1}^{M}\sum_{t=1}^{Tj-1} P\left(x_t = i, m_t = k \mid \boldsymbol{o}_1^j \cdots \boldsymbol{o}_{Tj}^j, \lambda^i\right)}, \quad \textit{for } k = 1, \cdots, K \textit{ and } d_1, d_2 = 1, \cdots, D \tag{17}$$

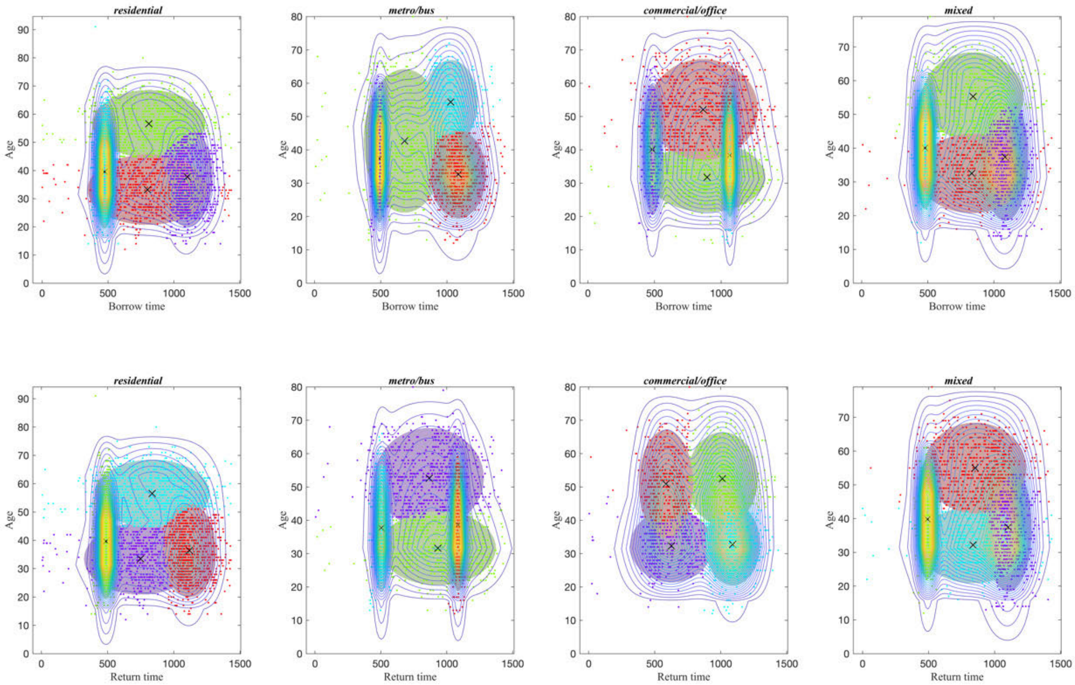

FIGURE 6. Contour map of Gaussian mixture model with arrangement and combination of dimensions in feature space.

IV. RESULTS AND DISCUSSION

A. MODEL TRAINING RESULT

The input variables of the model come from the docked bike-sharing trip-chains extracted in the second chapter of this paper. The *O-O* trip-chain means the origin and destination of the trip are the same, so the purpose of this type is considered as flexible cycling activities. In this study, four other types of trip-chain prototype were used to train the CHMM model. The observation sequence is characterized by the characteristic space established by the types of borrowing/returning, travel time, age of docked bike-sharing users and 4 land-use characteristics around docking station (Mixed, Commercial/office, Residential, and Metro/bus), which are represented by the mean vector and variance covariance matrix. Since types of borrowing/returning and land use characteristics are classified variables, hence each combination is discussed separately. Through the method of violence search, we find that four clusters can obtain stable Gaussian mixture model to represent each group of data. According to the arrangement and combination of these four variables, 32 groups of virtual clusters are obtained. Figure 6 and Figure 7 show the contour map and surface map of each group of Gaussian mixture model, respectively. These two figures visualize the Gaussian mixture clustering results of the observation sequence feature space. This study changes the number of clusters to examine how well the results match previous expectations and/or common sense. These hypothetical clusters can be regarded as regular cycling models for induction. As a result, 14 cycling models were used as the most reasonable cluster number. The number of hidden states may differ from the number of clusters. After investigating 14 clusters, 4 hidden states (trip purpose: transfer to metro/bus, flexible cycling activity, home-based commute and work-based commute) were found to be appropriate. Table 2 shows the connection between cycling mode and its possible trip purpose. According to the determined CHMM structure, the Baum Welch algorithm was further implemented based on the sample data, and the estimated values of three parameter sets are obtained: the probability of initial state, the transition probability, the membership probability.

The membership probability matrix matches the cycling mode (or derived clusters) extracted from the feature space of observation sequence with the hidden activities (or states). Table 3 shows the membership probability of each activity. For clarity, the dominant membership probability is highlighted in bold font. It can be seen from the above table that (1) Activity 1 (metro/bus transfer activity) only occurs in Cluster 1 (borrow or return bicycle around the subway in peak hour) and Cluster 2 (borrow or return bicycle around the subway in off-peak hour), because metro/bus transfer activity can only occur at stations with land-use type of Metro/bus, and the main metro/bus transfer activity occurs in the morning and evening peak hours. (2) Activity 2 (flexible cycling activity) is related to Cluster 2 (borrow or return bicycle around the subway during off-peak hours), Cluster 4 (borrow or return bicycle in mixed land-use area during off-peak hours), Cluster 9 (borrow or return bicycle around residential area during off-peak hours) and Cluster 14 (borrow or return bicycle

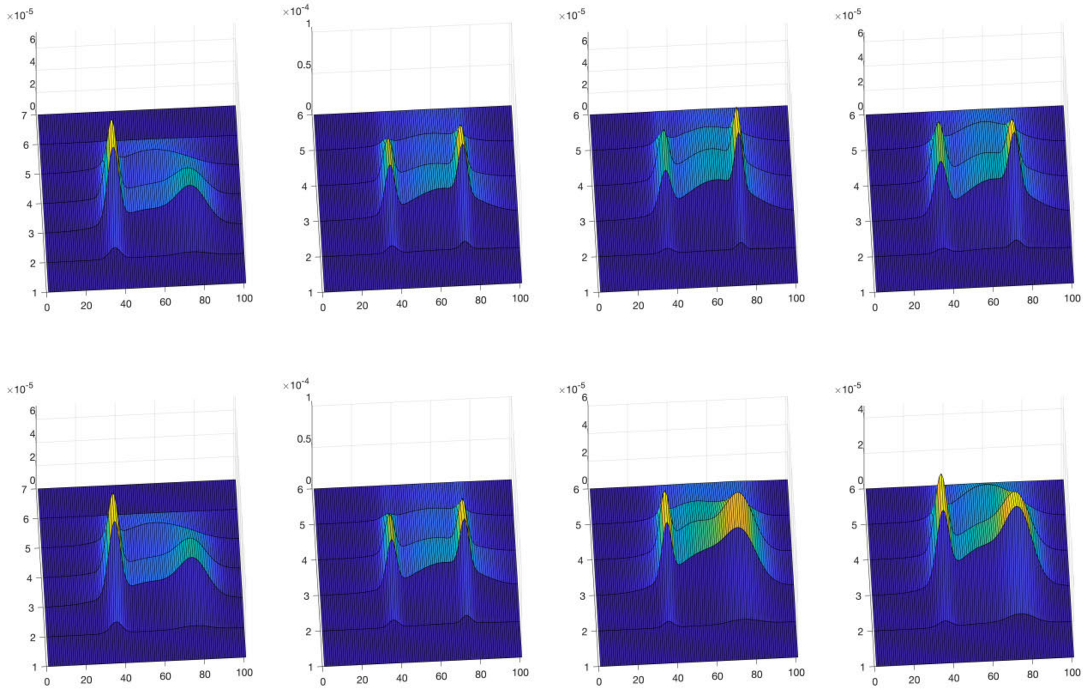

FIGURE 7. Surface graph of Gaussian mixture model with arrangement and combination of dimensions in feature space.

around commercial/office area during off-peak hours), which mainly occurs in off-commuting peak hours of all kinds of docking stations. These stations are used to borrow or return bicycles around the residential area during off-peak hours. (3) Activity 3 (home-based commute activity) is linked to Cluster 3 (borrow or return bicycle in the mixed land-use area in peak hour) and Cluster 5 (borrow bicycle around the residential area in morning peak hour), which mainly occurs in residential areas and mixed land-use areas during peak hours, and it is noted that home-based commuting activities will also occur in docked bike-sharing docking stations with the land-use of commercial/office. This is because the land-use of the docking station does not fully represent the use function of the docking station. (4) Activity 4 (work-based commute activity) is associated with Cluster 3 (borrow or return bicycle in the mixed land-use area during peak hours), Cluster 10 (return bicycle around commercial/office area during morning peak hours) and Cluster 11 (borrow bicycle around commercial/office area during evening peak hours), which mainly occurs in returning bicycle around commercial/office area during morning peak hours and borrowing bicycle around commercial/office area during evening peak hours. A small probability of work-based commute activities may occur at docked bike-sharing docking station in residential areas.

The transition probability matrix shows how the hidden trip purpose sequence for a docked bike-sharing trip-chain is generated. Table 4 shows the transition matrix with a possible title for each activity. We can get the following information from Table 4: (1) if the current activity is a metro/bus transfer activity, then the next activity is most likely to be a flexible cycling activity; (2) if the current activity is a flexible cycling activity, and the next activity is most likely to be a flexible cycling activity: (3) if the current activity is a home-based commute activity, then the next activity is most likely to be a work-based commute activity, or a metro/bus transfer activity, e.g. metro/bus transfer activity as part of the commute activity; (4) if the current activity is a work-based commute activity, the next activity is most likely to be a home-based commute activity, followed by a flexible cycling activity, e.g. docked bike-sharing users carrying on with their trips for leisure and entertainment activities after work.

After the CHMM model was trained, the observed docked bike-sharing feature vectors were inputted, and the Viterbi algorithm was employed to infer the most likely trip purpose sequence of the trip-chains. Table 5 shows the results of the trip purpose imputation of trip-chains in the sample. We can get some interesting information by analyzing the recognition results of each classified trip-chains. (1) For the O-sD trip-chains, about half are flexible cycling activities, followed by home-based commute activity to work-based commute activity, accounting for about 20% of the total, and thirdly home-based commute activity to metro/bus transfer activity and metro/bus transfer activity to home-based commute activity, each accounting for about 10% of the total. (2) For O-sD-O trip-chains, it is mainly composed of flexible cycling activities. The second is commuting trips to and from home, most likely the regular trip of starting from home in the morning peak hour and returning home in the evening peak

TABLE 2. Description of the resultant clusters based on Gaussian mixture model.

Cluster description	Cluster centroid				Most likely activity purpose
	Age	Travel time	Borrow /Return	Bike-sharing docking station type	
Cluster 1: borrow or return bicycle around the subway in peak hour	All	Peak period	Borrow or return	Metro/bus	Transfer to metro/bus
Cluster 2: borrow or return bicycle around the subway in off-peak hour	All	Off-peak period	Borrow or return	Metro/bus	Transfer to metro/bus
Cluster 3: borrow or return bicycle in mixed land-use area in peak hour	All	Peak period	Borrow or return	Mixed	Flexible cycling activity
Cluster 4: borrow or return bicycle in mixed land-use area in off-peak hour	All	Off-peak period	Borrow or return	Mixed	Flexible cycling activity
Cluster 5: borrow bicycle around residential area in morning peak hour	Commuter	Morning peak	Borrow	Residential	Home-based commute
Cluster 6: return bicycle around residential area in evening peak hour	Commuter	Evening peak	Return	Residential	Home-based commute
Cluster 7: Non-commuters borrow bicycle around residential area in morning peak hour	Non-commuter	Morning peak	Borrow	Residential	Flexible cycling activity
Cluster 8: Non-commuters return bicycle around residential area in evening peak hour	Non-commuter	Evening peak	Return	Residential	Flexible cycling activity
Cluster 9: borrow or return bicycle around residential area in off-peak hour	All	Off-peak period	Borrow or return	Residential	Flexible cycling activity
Cluster 10: return bicycle around commercial/office area in morning peak hour	Commuter	Morning peak	Return	Commercial/office	Work-based commute
Cluster 11: borrow bicycle around commercial/office area in evening peak hour	Commuter	Evening peak	Borrow	Commercial/office	Work-based commute
Cluster 12: Non-commuters borrow bicycle around commercial/office area in morning peak hour	Non-commuter	Morning peak	Return	Commercial/office	Flexible cycling activity
Cluster 13: Non-commuters return bicycle around commercial/office area in evening peak hour	Non-commuter	Evening peak	Borrow	Commercial/office	Flexible cycling activity
Cluster 14: borrow or return bicycle around commercial/office area in off-peak hour	All	Off-peak period	Borrow or return	Commercial/office	Flexible cycling activity

TABLE 3. Estimated probabilities ($\hat{g}_{ik}$) of hidden clusters.

Land-use	Metro/bus	Metro/bus	Mixed	Mixed	Residential	Residential	Residential	Residential	Residential	Commercial/office	Commercial/office	Commercial/office	Commercial/office	Commercial/office
	Cluster 1 (Transfer to metro/bus)	Cluster 2 (Transfer to metro/bus)	Cluster 3 (Flexible cycling activity)	Cluster 4 (Flexible cycling activity)	Cluster 5 (Home-based commute)	Cluster 6 (Home-based commute)	Cluster 7 (Flexible cycling activity)	Cluster 8 (Flexible cycling activity)	Cluster 9 (Flexible cycling activity)	Cluster 10 (Work-based commute)	Cluster 11 (Work-based commute)	Cluster 12 (Flexible cycling activity)	Cluster 13 (Flexible cycling activity)	Cluster 14 (Flexible cycling activity)
Activity 1	**0.667**	0.333	-	-	-	-	-	-	-	-	-	-	-	-
Activity 2	-	**0.2**	0.006	**0.213**	-	-	0.018	0.016	**0.332**	-	-	0.009	0.008	0.198
Activity 3	-	-	**0.256**	-	**0.395**	0.138	-	-	-	0.111	0.101	-	-	-
Activity 4	-	-	**0.236**	-	0.13	0.175	-	-	-	**0.259**	**0.2**	-	-	-

hour. At the same time, it is noted that 10% of trip are home-based commute activities to work-based commuting activities, followed by using docked bike-sharings for flexible cycling activities. This kind of trip-chains should be that the docked bike-sharings were used for commuting to workplace during the morning peak hours, then for work-related trips during work hours, and for finally returning to workplace. (3) For O-mD trip-chains, it is also composed of flexible cycling activities, consistent with the actual situation. Secondly, it is to carry out home-based commute activities, work-based commute activities, followed by a series of flexible cycling activities. This situation may be that this part of commuters use docked bike-sharings for commute activities in the morning peak hour, and then use docked bike-sharings for other recreational activities before returning home. (4) For O-mD-O trip-chains, the total amount of them is less, and the

TABLE 4. Estimated probabilities ($\hat{a}_{ij}$) of activity purpose.

Transition probabilities	Activity 1 (Transfer to metro/bus)	Activity 2 (Flexible cycling activity)	Activity 3 (Home-based commute)	Activity 4 (Work-based commute)
Activity 1 (Transfer to metro/bus)	0.143	0.501	0.233	0.123
Activity 2 (Flexible cycling activity)	0.027	0.919	0.000	0.054
Activity 3 (Home-based commute)	0.317	0.050	0.000	0.632
Activity 4 (Work-based commute)	0.125	0.229	0.518	0.128

TABLE 5. Distribution of the most probable activity purpose sequences in the training sample.

Type	Activity purpose sequence	Counts
O-sD	2-2	24784 (51%)
	3-4	10008 (21%)
	3-1	4330 (9%)
	1-3	3202 (7%)
	4-1	1127 (2%)
	4-3	1121 (2%)
	1-4	1119 (2%)
O-sD-O	2-2-2-2	3616 (47%)
	3-4-4-3	890 (11%)
	3-4-2-2	741 (10%)
	3-1-1-3	544 (7%)
	1-3-4-1	450 (6%)
	3-1-2-2	430 (5%)
	2-2-4-4	362 (4%)
	1-4-2-2	163 (2%)
O-mD (m=2)	2-2-2-2	2649 (43%)
	3-4-2-2	461 (8%)
	3-4-4-4	439 (7%)
	2-2-4-4	280 (5%)
	1-3-4-4	177 (3%)
	3-1-2-2	175 (3%)
	1-4-2-2	154 (3%)
O-mD (m=3)	2-2-2-2-2-2	410 (43%)
	3-4-2-2-2-2	85 (9%)
	2-2-2-2-4-1	46 (5%)
	1-4-2-2-2-2	36 (4%)
	2-2-2-2-2-1	30 (3%)
	3-1-1-2-2-2	30 (3%)
	2-2-4-4-2-2	19 (2%)
O-mD (m>3)	2-2-2-2-2-2-2-2	120 (23%)
	2-2-2-2-2-2-2-2-2-2	37 (7%)
	2-2-2-2-2-2-4-1	19 (4%)
	3-4-2-2-2-2-2-2	18 (4%)
	3-4-2-2-2-2-4-4	16 (3%)
	2-2-2-2-2-2-2-2-2-2-2-2	11 (2%)
	4-2-2-2-2-2-2-2	9 (2%)
O-mD-O	2-2-2-2-2-2	260 (46%)
	2-2-2-2-2-2-2-2	55 (10%)
	3-4-4-4-4-3	35 (6%)
	3-4-2-2-2-2	23 (4%)
	3-1-2-2-2-2	21 (4%)
	2-2-2-2-4-4	18 (3%)
	3-4-4-4-2-2	18 (3%)
	3-4-2-2-2-2-4-3	12 (2%)
	4-2-2-2-2-2	11 (2%)

results of trip purpose identification show that more than half of the trip-chains is flexible cycling activities, Secondly, there are more symmetrical commuting trips, from home to work place in the morning peak hour, and back home from work place in the evening peak hour. In the middle of the day, there are borrowing and returning docked bike-sharing activities at

TABLE 6. Trip purpose results of questionnaire survey.

Activity purpose sequence	Counts
1-2	2
2-1	28
2-2	84
3-1	5
3-4	13
2-2-2-1	3
2-2-2-2	12
3-4-4-3	5
Total	152

work places, which may be that travelers do not return home during lunch break, but using the docked bike-sharings to get around the area of work places.

B. MODEL VALIDATION RESULTS

The CHMM applied in this paper is an unsupervised machine learning tool, which does not need to be trained by labeled data. However, the result of target recognition could not be verified due to the real trip purposes of docked bike-sharing usage was missing from the IC card data set. In April 2016, a questionnaire survey on docked bike-sharing users' travel activities was conducted. The survey data was used to verify the trip purpose recognition results of the CHMM model. If the trained CHMM model could effectively identify the trip purpose of the respondents, it would verify the proposed method. The questionnaire design adopted the method of recall survey. Firstly, the respondents needed to recall the experience of using docked bike-sharing on the previous day, and then filled in the times of borrowing and returning, the name(s) of the docking station(s), and the trip purpose(s) every time. In this survey, 166 questionnaires were collected, 152 were valid, and travel chains were extracted, including 132 O-sD trip-chains, 15 O-mD type trip-chains and 5 O-sD-O trip-chains. Table 6 shows the trip purpose of docked bike-sharing directly obtained through the questionnaire.

The trip purposes of the respondents' trips were identified by the trained CHMM model trained first, and then compared with the real trip purposes reported by the respondents, so as the accuracy of the proposed method were verified. Since the label variables were added into the questionnaire, the reliability evaluation of the model could be conducted according to the multi-classification problems [35]. Each activity in the trip-chains was regarded as a sample for metrics calculation, so that more calculation samples were used during the evaluation. Some metrics were essentially defined for binary classification tasks. In extending a binary metric to multiclass or multilabel problems, the data were treated as a collection of binary problems, one for each class. There are then a number of ways to average binary metric calculations across the set of classes. For example, "macro" simply calculates the mean of the binary metrics, giving equal weight to each class. "weighted" accounts for class imbalance by computing the average of binary metrics in which each class's score was weighted by its presence in the true data sample. We calculated precision (P), recall (R), f1-score (F1) for each class [36], gave the accuracy of the model, and calculated the overall macro-precision (macro-P), macro-recall (macro-R), macro-f1 (macro-F1), weighted-precision (weighted-P), weighted-recall (weighted-R), weighted-f1 (weighted-F1) according to the "weighted" and "macro" criteria. The calculation results are given in Table 7. As a large skew in the sample is presented and may affect our judgment on the results. Therefore, the classification effect of each class was judged by the normalized confusion matrix. Such normalization can be interesting in case of class imbalance to have a more visual interpretation of which class is being misclassified. The normalized confusion matrix [37] of the model used in this study is shown in Figure 8. Further, we note that the result of CHMM model is a series of trip purposes, including the trip purpose of multiple trips of docked bike-sharing users. Although the identification results of some docked bike-sharing trip-chains are not completely in line with the respondents', there are still some stages of trip-chains were successful identified. Therefore, a simple method was proposed to calculate the recognition accuracy of the docked bike-sharing trip purpose sequence. The expression is as follows, and the specific results are shown in Table 8.

$$\begin{aligned} &sequence\ accuracy \\ &\quad = \frac{Number\ of\ correct\ identification\ in\ a\ trip-chain}{Number\ of\ trips\ in\ the\ a\ trip-chain} \end{aligned} \tag{25}$$

As can be seen from Table 7 and Table 8, for the questionnaire sample, the overall unit accuracy and sequence accuracy of the model reached 0.770 and 0.756, respectively, both exceeding 70%, indicating that the model has good recognition efficiency and can accurately identify most of the trip purpose of docked bike-sharing trip-chains. The F1-score is a compromise between recall (R) and precision (P). The higher the F1-score, the higher the recognition rate. Although the F1-score of the model is only 0.56 in the case of equal weights in each category, the F1-score weighted according to the sample ratio reaches 0.79. It means that some categories have not achieved good recognition and intensive due to the small sample. Next, we specifically analyze the recognition of each category. The method presented in this paper has a large

TABLE 7. A report of the main identify performance metrics.

	Sample proportion	Precision	Recall	F1-score
Transfer to metro/bus	9.7%	0.62	0.42	0.50
Flexible cycling activity	83.9%	0.92	0.79	0.85
Home-based commute	4.0%	0.31	1.00	0.48
Work-based commute	2.4%	0.27	1.00	0.43
Accuracy	100%			0.77
Macro avg	100%	0.53	0.80	0.56
Weighted avg	100%	0.85	0.77	0.79

TABLE 8. Recognition effect of trip purpose sequence.

Activity purpose sequence	Number of fully-identified trip-chains	Sequence accuracy
1-2	0	0.500
2-1	13	0.500
2-2	65	0.798
3-1	4	0.800
3-4	11	0.923
2-2-2-1	0	0.750
2-2-2-2	10	0.833
3-4-4-3	1	0.750
Total	105	0.756

TABLE 9. Model evaluation on weekends and weekdays.

	Accuracy	Weighted-P	Weighted-R	Weighted-F1
Weekdays	0.79	0.87	0.76	0.81
Weekends	0.76	0.83	0.75	0.79

difference in the recognition performance of different travel destination categories. The minimum value of the F1-score is 0.43 and the maximum is 0.85. The recognition errors tend to be the categories with fewer samples, and there is a phenomenon of category imbalance. The diagonal elements of the confusion matrix represent the number of points for which the predicted label is equal to the true label, while off-diagonal elements are those that are mislabeled by the classifier. The higher the diagonal values of the confusion matrix indicating a larger number of correct prediction. From the regularized confusion matrix, one can see that Activity 2 (flexible cycling activity), Activity 3 (home-based commute activity), and Activity 4 (work-based commute activity) have better recognition effects. The accuracy is low for Activity 1 (metro/bus transfer activity) as it is easy to be identified as Activity 2 (flexible cycling activity) by mistake due to the law of Activity 1 (metro/bus transfer activity) being more difficult to find from the IC card records of docked bike-sharing usage. The feature space we established is less sensitive to these two types of activities than others.

Considering that there may be a large gap between the travel characteristics of docked bike-sharing users on weekdays and weekends, we try to compare the performance of the model on weekdays and weekends to test the reliability and generality of the model. We select data training models on weekends and weekdays respectively, and complete model testing based on their questionnaire data. The evaluation indicators still select accuracy, precision, recall, f1-score, and the result is shown in Table 9. It can be seen from the table that the performance difference between the model on weekends and workdays is small, reflecting the model's good adaptability. The overall performance of the model on weekdays is slightly better, perhaps due to the large amount of data on weekdays, the model has more learning samples, on the other hand, it may be that the travel characteristics shown on weekdays are simpler than those on weekends.

Furthermore, we compare our method with the two baseline algorithms. The details are as follows:

Nearest [38]. The idea of *Nearest* algorithm comes from k-nearest neighbor, which simply sets the POI nearest to the exit position as the final destination of docked bike-sharing users, without considering other factors, such as drop-off time. Therefore, the trip purpose is predicted to be an activity associated with the POI category.

Bayes' rule [39], [23]. This method is derived from a patented technology of my research group. The probability of trip purpose is modeled by Bayesian rule, which considers space and time constraints. The difference between this method and the model in this paper is that it needs to use the trip purpose in real situation, and deduces the optimal probability of trip purpose under space-time constraints through the prior probability.

The theoretical basis of our algorithm evaluation is: if the trip purpose distribution inferred by the proposed method

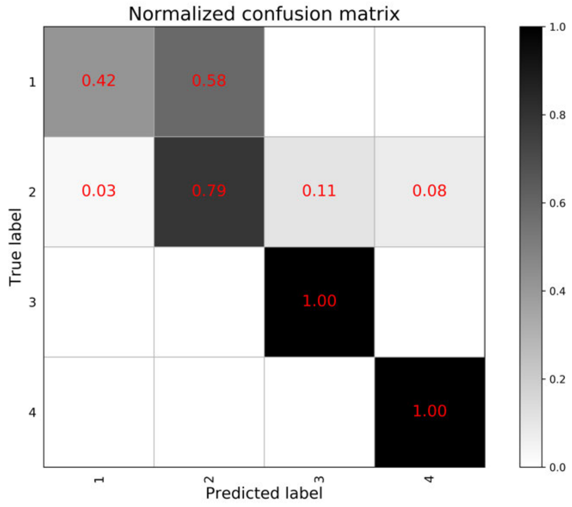

FIGURE 8. Normalized confusion matrix.

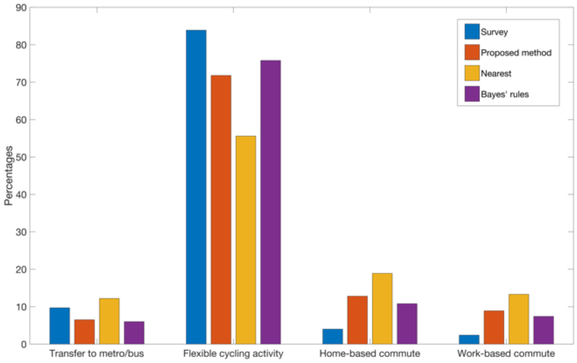

FIGURE 9. Comparison results to baseline algorithms and survey data.

is close to the statistical distribution of survey data on the regional scale, then the proposed method should be reliable. In Figure 9, we compare the travel survey data for the inferential results of our proposed method. In addition, we plot the results of two other baselines for comparison. It is easy to understand that the closer the percentage ratio of each class is to the corresponding survey data value, the better the performance of the algorithm is. From the results, we can see that the performance of our proposed algorithm is similar to that of *Bayes' rule,* but slightly lower than the latter, while *Nearest* algorithm achieves the worst performance. To some extent, our proposed algorithm does not achieve the best inference effect, but its performance is similar to that of Bayesian rules. At the same time, *Bayes' rule* need to rely on trip purpose survey data. Our algorithm belongs to unsupervised learning method and does not need the tag value of real trip purpose, which makes the implementation cost of the algorithm lower.

V. CONCLUSION

This paper presented a method to identify docked bike-sharing trip purposes based on the CHMM model, and introduced the modeling process and algorithm implementation of the model. The CHMM model was implemented through Matlab software. Taking the IC card data of Nanjing docked bike-sharing usage on March 2, 2016, as a research sample, we calibrated the parameters of the CHMM model and employed the parameter estimation results to identify trip purposes. Finally, the questionnaire survey was conducted to collect the data of docked bike-sharing trip purposes, which were used to verify the reliability and accuracy of the model.

Deficiencies in the model of this study are noted and improvements are needed in future studies. In the model, although the defined trip purpose category covers the travel needs of most docked bike-sharing users, the definition of trip purposes in the hidden layer is arbitrary and depends on the land-use attribute excessively. The composition of the feature space needs to be studied further. The clustering results of the observation variables based on the feature variables might have affected the final recognition of the trip purposes. The recognition quality of some categories, which was largely derived from the structure of the feature space, needs to be improved. At the same time, the research object of this model is the docked bike-sharing, which has a relatively fixed geographical distribution. At present, dockless bike-sharing is developing rapidly, it will become our next research focus. Because of its more complex space-time distribution characteristics, it has a greater challenge. We will also use the travel data of dockless bike-sharing in different cities to analyze the travel characteristics of users.

REFERENCES

[1] Y. Ji, X. Ma, M. He, Y. Jin, and Y. Yuan, "Comparison of usage regularity and its determinants between docked and dockless bike-sharing systems: A case study in Nanjing, China," *J. Cleaner Prod.*, vol. 255, May 2020, Art. no. 120110.

[2] Y. Ji, Y. Fan, A. Ermagun, X. Cao, W. Wang, and K. Das, "Public bicycle as a feeder mode to rail transit in China: The role of gender, age, income, trip purpose, and bicycle theft experience," *Int. J. Sustain. Transp.*, vol. 11, no. 4, pp. 308–317, Apr. 2017.

[3] X.-H. Yang, Z. Cheng, G. Chen, L. Wang, Z.-Y. Ruan, and Y.-J. Zheng, "The impact of a public bicycle-sharing system on urban public transport networks," *Transp. Res. A, Policy Pract.*, vol. 107, pp. 246–256, Jan. 2018.

[4] J. Yin, L. Qian, and A. Singhapakdi, "Sharing sustainability: How values and ethics matter in consumers' adoption of public bicycle-sharing scheme," *J. Bus. Ethics*, vol. 149, no. 2, pp. 313–332, 2018.

[5] F. Scalone, P. Agati, A. Angeli, and A. Donno, "Exploring unobserved heterogeneity in perinatal and neonatal mortality risks: The case of an italian sharecropping community, 1900–39," *Population Stud.*, vol. 71, no. 1, pp. 23–41, Jan. 2017.

[6] X. Shi, Z. Yu, J. Chen, H. Xu, and F. Lin, "The visual analysis of flow pattern for public bicycle system," *J. Vis. Lang. Comput.*, vol. 45, pp. 51–60, Apr. 2018.

[7] X. Fu and W. H. K. Lam, "Modelling joint activity-travel pattern scheduling problem in multi-modal transit networks," *Transportation*, vol. 45, no. 1, pp. 23–49, Jan. 2018.

[8] Y. Liu, Z. Liu, and R. Jia, "DeepPF: A deep learning based architecture for metro passenger flow prediction," *Transp. Res. C, Emerg. Technol.*, vol. 101, pp. 18–34, Apr. 2019.

[9] C. Wang, C. Xu, and Y. Dai, "A crash prediction method based on bivariate extreme value theory and video-based vehicle trajectory data," *Accident Anal. Prevention*, vol. 123, pp. 365–373, Feb. 2019.

[10] Y. Bian, D. Wu, S. Shu, J. Rong, and Y. Tang, "Study on travel characteristics of public bicycles in Beijing," in *CICTP 2014: Safe, Smart, and Sustainable Multimodal Transportation Systems*. Reston, VA, USA: American Society of Civil Engineers, 2014, pp. 3331–3343.

[11] P. Vogel, T. Greiser, and D. C. Mattfeld, "Understanding bike-sharing systems using data mining: Exploring activity patterns," *Procedia-Social Behav. Sci.*, vol. 20, pp. 514–523, Jan. 2011.

[12] C. Kang, X. Ma, D. Tong, and Y. Liu, "Intra-urban human mobility patterns: An urban morphology perspective," *Phys. A, Stat. Mech. Appl.*, vol. 391, no. 4, pp. 1702–1717, Feb. 2012.

[13] M. Li, S. Gao, F. Lu, and H. Zhang, "Reconstruction of human movement trajectories from large-scale low-frequency mobile phone data," *Comput., Environ. Urban Syst.*, vol. 77, Sep. 2019, Art. no. 101346.

[14] 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.*, vol. 1972, no. 1, pp. 38–45, 2006.

[15] 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.*, vol. 1768, no. 1, pp. 125–134, 2001.

[16] C. M. Krause and L. Zhang, "Short-term travel behavior prediction with GPS, land use, and point of interest data," *Transp. Res. B, Methodol.*, vol. 123, pp. 349–361, May 2019.

[17] G. Han and K. Sohn, "Activity imputation for trip-chains elicited from smart-card data using a continuous hidden Markov model," *Transp. Res. B, Methodol.*, vol. 83, pp. 121–135, Jan. 2016.

[18] S. G. Lee and M. Hickman, "Trip purpose inference using automated fare collection data," *Public Transp.*, vol. 6, nos. 1–2, pp. 1–20, Apr. 2014.

[19] L. Liao, D. Fox, and H. Kautz, "Location-based activity recognition using relational Markov networks," in *Proc. Int. Joint Conf. Artif. Intell.*, 2005, pp. 773–778.

[20] P. Wang, Y. Fu, G. Liu, W. Hu, and C. Aggarwal, "Human mobility synchronization and trip purpose detection with mixture of hawkes processes," in *Proc. 23rd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2017, pp. 495–503.

[21] Z. Deng and M. Ji, "Deriving rules for trip purpose identification from GPS travel survey data and land use data: A machine learning approach," in *Traffic and Transportation Studies*. Reston, VA, USA: American Society of Civil Engineers, 2010, pp. 768–777.

[22] B. Furletti, P. Cintia, C. Renso, and L. Spinsanti, "Inferring human activities from GPS tracks," in *Proc. 2nd ACM SIGKDD Int. Workshop Urban Comput. (UrbComp)*, 2013, pp. 1–8.

[23] L. Gong, X. Liu, L. Wu, and Y. Liu, "Inferring trip purposes and uncovering travel patterns from taxi trajectory data," *Cartogr. Geograph. Inf. Sci.*, vol. 43, no. 2, pp. 103–114, Mar. 2016.

[24] Z. Zhu, U. Blanke, and G. Tröster, "Inferring travel purpose from crowd-augmented human mobility data," in *Proc. 1st Int. Conf. IoT Urban Space*, 2014, pp. 44–49.

[25] Y. Lin, H. Wan, R. Jiang, Z. Wu, and X. Jia, "Inferring the travel purposes of passenger groups for better understanding of passengers," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 1, pp. 235–243, Feb. 2015.

[26] S. Krygsman, T. Arentze, and H. Timmermans, "Capturing tour mode and activity choice interdependencies: A co-evolutionary logit modelling approach," *Transp. Res. A, Policy Pract.*, vol. 41, no. 10, pp. 913–933, Dec. 2007.

[27] H. Timmermans, P. van der Waerden, M. Alves, J. Polak, S. Ellis, A. S. Harvey, S. Kurose, and R. Zandee, "Spatial context and the complexity of daily travel patterns: An international comparison," *J. Transp. Geogr.*, vol. 11, no. 1, pp. 37–46, Mar. 2003.

[28] C.-H. Wen and F. S. Koppelman, "A conceptual and methdological framework for the generation of activity-travel patterns," *Transportation*, vol. 27, no. 1, pp. 5–23, 2000.

[29] J. Zhao, J. Wang, and W. Deng, "Exploring bikesharing travel time and trip chain by gender and day of the week," *Transp. Res. C, Emerg. Technol.*, vol. 58, pp. 251–264, Sep. 2015.

[30] H. N. Nagaraja, "Inference in hidden Markov models," *Technometrics*, vol. 48, no. 4, pp. 574–575, Nov. 2006.

[31] M. Zraiaa, "Hidden Markov models: A continuous-time version of the Baum-Welch algorithm," Dept. Comput., Imperial College London, London, U.K., 2010.

[32] H. Peng, T. Huang, and K. Zhang, "Model selection for Gaussian mixture models," *Statist. Sinica*, vol. 27, no. 1, pp. 147–169, 2017, doi: 10.5705/ss.2014.105.

[33] E. Chouzenoux, J.-C. Pesquet, and A. Repetti, "A block coordinate variable metric forward–backward algorithm," *J. Global Optim.*, vol. 66, no. 3, pp. 457–485, Nov. 2016.

[34] A. Viterbi, "Error bounds for convolutional codes and an asymptotically optimum decoding algorithm," *IEEE Trans. Inf. Theory*, vol. IT-13, no. 2, pp. 260–269, Apr. 1967.

[35] V. Santhanam, V. I. Morariu, D. Harwood, and L. S. Davis, "A non-parametric approach to extending generic binary classifiers for multi-classification," *Pattern Recognit.*, vol. 58, pp. 149–158, Oct. 2016.

[36] R. Arora, C.-T. Tsai, K. Tsereteli, P. Kambadur, and Y. Yang, "A semi-Markov structured support vector machine model for high-precision named entity recognition," in *Proc. 57th Annu. Meeting Assoc. Comput. Linguistics*, 2019, pp. 5862–5866.

[37] F. Jia, Y. Lei, N. Lu, and S. Xing, "Deep normalized convolutional neural network for imbalanced fault classification of machinery and its understanding via visualization," *Mech. Syst. Signal Process.*, vol. 110, pp. 349–367, Sep. 2018.

[38] C. Chen, S. Jiao, S. Zhang, W. Liu, L. Feng, and Y. Wang, "TripImputor: Real-time imputing taxi trip purpose leveraging multi-sourced urban data," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 10, pp. 3292–3304, Oct. 2018.

[39] J. Yanjie, M. Xinwei, Y. Mingyuan, J. Yuchuan, T. Xu, and C. Xianqi, "Travel purpose prediction method of public bicycle users based on Bayesian probability model," Southeast Univ., Jiangsu, China, Tech. Rep. CN107368916A, 2017.

WENHAO LI received the M.S. degree in transportation engineering from the Guilin University of Electronic Technology, Guilin, China, in 2019, and the M.S. degree in computer information technology from Northern Arizona University, Flagstaff, AZ, USA. He is currently pursuing the Ph.D. degree in transportation engineering with the School of Transportation, Southeast University, Nanjing, China. His research interests include transportation planning and control, transport data analytics, and intelligent transport system.

YANJIE JI received the Ph.D. degree from Southeast University, in 2007. She is currently an Associate Professor with the School of Transportation, Southeast University. Her research interests include transportation planning and control, intelligent transport systems, travel behavior analysis, and data-driven transit operations.

XIANQI CAO received the B.S. and M.S. degrees in transportation engineering from Southeast University, Nanjing, China, in 2017.

From 2015 to 2017, he was a Research Assistant with the School of Transportation, Southeast University. He is currently working with the China Railway Siyuan Survey and Design Group Company, Ltd. His research interests include public transport planning and public bicycle planning using machine learning technology.

XINYI QI received the B.S. degree in transportation engineering from Southeast University, Nanjing, China, in 2020, where she is currently pursuing the master's degree in transportation engineering. Her research interests include transportation planning and management, and traffic safety.