IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received July 11, 2020, accepted August 3, 2020, date of publication August 25, 2020, date of current version September 9, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3019253

Individual Commute Time Recognition Based on the Hierarchical Semantic Model

WEI LIU, TAO WENXUAN, ZHONG ZIMING, AND JIANQIN YIN, (Member, IEEE)
Automation School, Beijing University of Posts and Telecommunications, Beijing 100876, China

Corresponding author: Wei Liu (twhlw@163.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 61673192, in part by the Fund for Outstanding Youth of Shandong Provincial High School under Grant ZR2016JL023, and in part by the Basic Scientific Research Project of Beijing University of Posts and Telecommunications under Grant 2018RC31.

ABSTRACT Individual commute time recognition is essential for traffic demand management. However, this problem has yet to be studied. In this study, we propose a hierarchical semantic model (HSM) to recognize individual commute time. To the best of our knowledge, this work is the first to integrates large scale travellers commute time prediction at an individual level. HSM consists of a low and a high semantic layer. The low semantic layer models spatial, temporal and environmental information, whereas the high semantic layer recognises commute time using the hidden Markov model on the basis of the low semantic layer outputs. Experimental results demonstrate the effectiveness of our proposed model for individual commute time recognition.

INDEX TERMS Commute time recognition, hierarchical semantic model, traffic demand management.

I. INTRODUCTION

The recognition of individual commute time can improve the efficiency of transportation systems. This topic is crucial to the study of traffic demand management.

Commuters have diverse commuting behaviour. Some of them have a fixed commuting mode, as shown in Fig. 1. The commute time of some workers, such as civil servants, doctors and comapny employees, does not change in a short period. Other workers, such as couriers, have no apparent characteristics. In this study, we aim to recognise the commute time of people with a fixed commuting pattern which have a high degree of confidence in traffic demand forecasting at an individual level.

Datasets used in related research are concentrated on commuting behaviour at the group level instead of the individual level. For example, Toole *et al.* [1] used call detail records from mobile phones in conjunction with open and crowd sourced geospatial data, census records and surveys to estimate travel demand and infrastructure use. To predict the behaviour of travellers accurately, we use mobile phones as carriers and form our dataset by collecting the sensor data of the subjects for up to a month. We call this dataset commuting data from phone sensors (CPS) and apply it to our research.

The associate editor coordinating the review of this manuscript and approving it for publication was Jenny Mahoney.

From the sensor data, we can obtain low semantic data which include the basic actions and frequent locations of individuals. High semantic data, which consist of home and work locations and meaningful commute-related indoor and outdoor transition, are extracted. We use hierarchical clustering to obtain frequent visiting sites and use the Gaussian mixture model(GMM) and the hidden Markov model (HMM) to identify human actions simultaneously.

The main contributions of this study are as follows:

(1) We collect an utterly new dataset named CPS for individual commuting analysis based on phone sensor data.

(2)To the best of our knowledge, this study is the first to propose a hierarchical semantic model for commute time recognition at an individual level.

(3) An end-to-end framework is proposed to recognize an individual's commute time.

The rest of this paper is organized as follows: In Section II, related work is reviewed. In Section III and Section IV, we present our proposed method. In Section V, we discuss our data and experiments in detail. Lastly, we summarize and conclude the study in Section VI.

II. RELATED WORK

In the literature, a common method for predicting commute time is to study the major factors that affect commute time in a specific commute scenario and predict the average commute

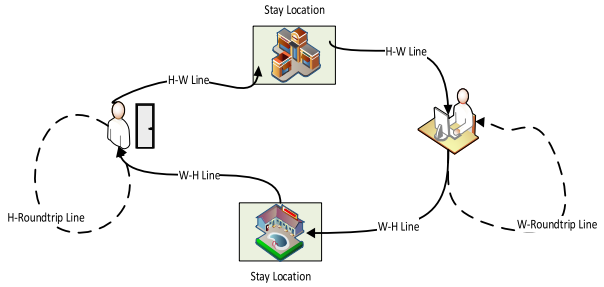


FIGURE 1. Daily commute activity travel pattern. 'H' means home, whereas 'W' means workplace. 'Stay location' represents the locations where the individual may stay during the commute. 'Round trip Line' indicates the out-of-town situation that an individual may have when staying at home or in the workplace. Individual commuting behaviour is dynamic. For most commuters, 'Stay Location' is a non-option because the home and workplace are fixed. But our method can recognise the locations where the individual stays during the commute.

time of the corresponding group by modelling these factors. For example, Ford *et al.* [3] predicted commute time by studying the major factors affecting the commute time of people who use shared bicycles in Wall Street in New York City. This method dicovers the major factors that affect the commute time of the group; thus, efficient traffic management measures are formulated on the basis of the group's perspective. However, this traditional method of prediction and recognition ignores the individual needs of commuters and cannot achieve commute time identification at the individual level. In our method, we overcome this limitation by modelling individual sensors data.

This work builds upon the CPS dataset described in Section V because the data and method of commute time recognition have not been studied yet. Nevertheless, four topics are related to this study:(1) related datasets, (2) activity recognition, (3) GMM-HMM model and (4) commute location recognition. Therefore, we review them separately.

A. DATA

To the best of our knowledge, the most similar datasets to CPS are the Actitracker dataset published by WISDM Labs [3] and the Har dataset of UCI [4]. The Actitracker dataset collected accelerator data from 36 mobile phone users at a frequency of 20 Hz, totalling 1,098,207 records. Meanwhile, the Har dtatset involved 30 subjects aged 19–48, thus having a total of 10,299 records. CPS is different in two major aspects. Firstly, CPS is an unlabeled dataset which simulates real-life and samples in open scenarios, whereas the Actitracker and Har are both labelled, and the mobile phones for sampling are required to be fixed in a part of the subject's body. Secondly, CPS samples 13 kinds of sensors, thereby providing a data foundation to identify the subjects' environment; by contrast, Actitracker only samples accelerometer data, and Har samples accelerometer and gyroscope data. Tbale 1 shows the composition and relation information of the Actitracker, Har and CPS datasets.

Although no similar work has been conducted on commute time recognition, many datasets have already included relevant information, such as travel survey databases of residents and public transportation datasets. The residents' travel survey databases which comprise family, personal and travel information tables, are frequently used to construct time assignment models in travel demand management. Alexander *et al.* [5] generated a user travel matrix using the private survey data of the United States for the analysis and modelling of traffic planning and investment; Toole *et al.* (2015) combined CDR and other datasets to estimate travel demand and infrastructure usage. As one of the main commute modes for commuters, public transportation datasets provide the possibility to identify passengers' commute time. Chen *et al.* [6] used ATPS to analyse commuters' travel demand and optimise bus management strategies. CPS differs from the above datasets in the sense that (1) phone sensor data are easy to obtain, and (2) data are flexible and can reflect the users' environment in detail. Although the microphone information of the user is collected and analysed in the CPS dataset, we only extract the loudness data of the environment, and do not save the microphone recording; thus, we do not violate the code of ethics. Moreover, we have informed the subjects of relevant information before the experiment and obtained their consent.

B. ACTIVITY RECOGNITION

Sensor-based activity recognition (AR) uses various sensors to measure the user's activity status, and the available hardware devices can be divided into three categories: (1) mobile phones, (2) wearable devices and (3) special sensors. San-Segundo *et al.* [7] used multiple sensors in mobile phones to identify six kinds of behavioural patterns, such as running and walking. Takahiro *et al.* [8] improved the accuracy of human activity recognition using ensemble learning based on single inertial measurement unit sensors. Ghosh *et al.* [9] used ultrasound sensor arrays to identify human activities. In addition to the expansion of sophisticated sensors, the authors in [10] developed a new strain sensor with excellent biocompatibility; the sensor can detect various human movements, including that of the wrist and fingers, breathing, speaking and swallowing.

Sensor-based AR algorithms have been continuously improved. Triboan *et al.* [11] proposed a semantics-based segmentation method for sensor data sequence to identify the complex activities of the elderly. In 2018, Hussein *et al.* [12] extracted features from mobile phone accelerometers worn by subjects, and further classified the features using random forest classifiers to identify human activities. Similarly, Reunanen *et al.* [13] presented a computational graph structure of human activity detection, which used accelerometers. Honghe *et al.* [14] proposed a wavelet tensor fuzzy clustering method for multi sensor activity recognition for human activity recognition. Hsu *et al.* [15] described an activity recognition algorithm based on wearable devices, extracted relevant features after productive human activities and then used PCA to recognize social activities. Moreover, a multi sensor classification and multi layer fusion model based on

TABLE 1. Comparison of the Actitracker, Har, and CPS datasets. Three things should be noted: (1) Actitracker and Har data records are labelled, whereas CPS belongs to unlabelled datasets; (2) Data sampling in CPS datasets depends on the sampling frequency of mobile phone sensors, and different sensors have different sampling frequencies (details are described in Section V); (3) CPS sampling scenarios are open, and subjects have no special requirement because they use their mobile phones in accordance with their daily habits.

Dataset		Actitracker	Har	CPS
Sensor type	Acclerate	✓	✓	✓
	Gproscope	-	✓	✓
	Rotation vector	-	-	✓
	WIFI	-	-	✓
	GPS	-	-	✓
	GPS state	-	-	✓
	Mic	-	-	✓
Attributes (Num)		6	6	Unlabelled
Sampling frequency (Hz)		50	50	Unfixed
Examples (Num)		1,098,207	10,299	1,428,536,715
Subjects (Num)		36	30	10
Experimental scene		Keep the phone in the subject's pocket	Fix the phone at the subject's waist	No limitations to the phone's location

entropy weight were proposed by Ming *et al.* [16] for human motion recognition of wearable devices.

In our work, we refer to the method of extracting behavioural patterns in [7] and identify six actions, namely, running, walking, climbing up, climbing down, static and unknown.

C. GMM-HMM MODEL

Our method uses the GMM-HMM model to calculate the sequence of indoor and outdoor states of individuals with time stamps and place labels.

The GMM-HMM model was launched in the 1990s and is mainly used in the field of speech recognition. [17] first combined the output probability of the normalised neural network with the transmission probability of the HMM model. Furthermore, Renals *et al.* [18] presented a combination method of scale similarity generated by two-layer MLP and GMM-HMM. Zhenjun *et al.* [19] proposed a linear spectrum frequency conversion method based on HMM-HMM, which can improve the naturalness and clarity of speech conversion. Zweig *et al.* [20] combined PLP-GMM and hybrid systems using the SCARF framework. Pang *et al.* [21] improved system performance by embedding a competitive penalty learning mechanism under the hidden Markov state during model training. Yan *et al.* [22] used DNN for feature extension based on the GMM-HMM model. Wang *et al.* [23] used the GMM-HMM model for single-channel speech separation. Zimmermann *et al.* [24] applied the features learned by neural networks to the GMM-HMM speech recognition system, and their work improved the accuracy of speech recognition. Heck *et al.* [25] proposed the DPFMM–HMM acoustic unit recogniser to enhance the performance of the language model.

In addition to its application in speech recognition, GMM-HMM models are used in capturing remarkable changes in the state of the mobile robot's motion [26], and segmenting human activities.

D. COMMUTER LOCATION RECOGNITION

Studies related to commuter location recognition are lacking. Ahas *et al.* [27] proposed the use of the passive location dataset provided by Estonia's mobile operator EMT in 2010 to determine the home and work location of users. They used the geographic coordinates of the data to determine the anchor points for the home and workplace on the basis of the standard deviation of the call's average-start-time and call-start-time to arrive at a specific location. In [6], the author deduced the passengers' home and workplace by calculating the passenger's ride frequency and spatially clustering the passenger's bus stop coordinates based on the public bus system data—ATPS.

However, our method is different from the above studies in the following aspects: (1) The GPS data we used is active positioning data which can reflect user's position accurately and timely. (2) To the best of our knowledge, our work is the first to suggest to cluster the GPS discrete points and construct the time-place distribution map to determine the user's home and workplace.

III. HIERARCHICAL SEMANTIC MODEL

Before presenting the proposed model, we first formalize our task. The input is a time-stamped sequence of phone sensor datasets of length T, $\{G, M, A, G_y, R\}_T^{t=1}$, whereas (G, M, A, G_y, R) represent GPS, mic, accelerometer, gyroscope, and rotation vectors, respectively. The output is a sequence of indoor and outdoor states of individuals with time stamps and place labels, $\{I_t^*\}$. By calculating the time interval between home and work location in $\{I_t^*\}$, we can obtain the individual commute time on the day.

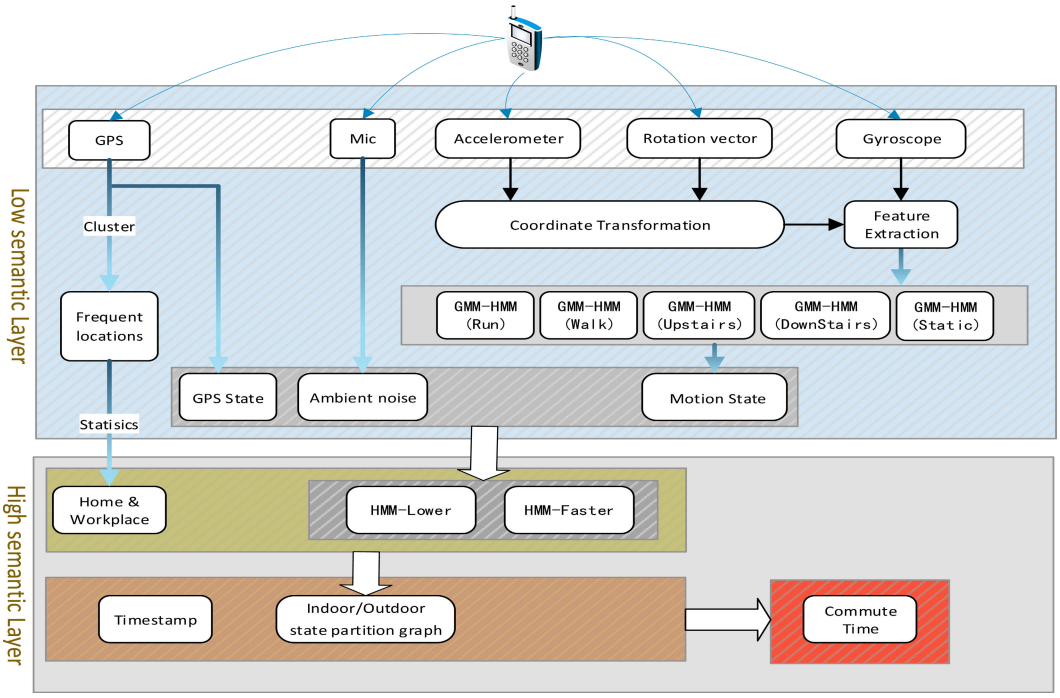


FIGURE 2. Hierarchical semantic structural diagram. An end-to-end framework whose input is sensor data and output is individual commuting time.

As illustrated in Figure 2, the hierarchical semantic model is composed of a low semantic layer and a high semantic layer. The task of the low semantic layer is to recognise the individual's action state and relevant locations in accordance with the input sequence, and the high semantic layer fuses the two to obtain $\{I_t^*\}$.

A. LOW SEMANTIC LAYER

In this work, the function of the low semantic layer lies in data processing, individual activity recognition and regularly visited place recognition.

We divide the GPS state into a strong state and weak state in accordance with the number of satellites captured by GPS. Moreover, ambient noise is similarly divided into two states. Given that the original accelerometer data are related to the postition of the mobile phone, we use the quaternion method to convert raw data so that it can fit the earth's coordinate system.

The recognition of individual activities can be described into three parts. Firstly, we segment the processed accelerometer and gyroscope data sequence using a sliding window and then form feature vectors by extracting features from each window. Considering the high density and serialisation of input data and inspired by the speech recognition algorithm, we select the GMM–HMM model to recognise individual activities in accordance with the eigenvector group. The advantage of GMM-HMM over the HMM model is that the sample points projected by GMM are not a definite classification marker, but a classification probability, which can predict the action state accurately.

The locations that individuals visited are obtained by clustering GPS trajectory, and an individual's regularly visited places are recognised by counting those locations. The beginning and the end of the path can reflect an individual's departure and destination; thus, we intercept the head and tail parts of the GPS track for clustering. Each caught part is treated as a set of points, and our goal is to find a centre point that can represent the set. To solve this task, we choose the AGglomerative NESting algorithm (AGNES) [28] in the hierarchical clustering method to cluster the points in this set and, the final clustered point is considered as the recognised location. The reason we choose the hierarchical clustering method instead of other methods is that this method artificially designates central points, which may cause subjective effects, before clustering.

B. HIGH SEMANTIC LAYER

The high semantic layer fuses the information of the individual's basic actions and frequent locations, which are mined from phone sensors data in the low semantic layer. In this layer, basic motions and the state of GPS/noise (ambient noise) construct a hidden Markov group to recognise the indoor or outdoor states of individuals. Afterward, the status' sequence is marked with location information at a time scale, whereas the location of home and work is recognised in accordance with the time distribution of sites obtained in the low semantic layer.

IV. METHODS

A. LOW SEMANTIC INFORMATION RECOGNITION

The low semantic level focuses on individuals' basic activities and frequent location recognition.

1) INDIVIDUAL BASIC ACTION RECOGNITION

The achievement of individual activity recognition is divided into three parts: pre processing of sensor data, feature extraction and activity recognition.

a: DATA PRE PROCESSING

The sensor' data used for recognising individuals' activities contain three kinds of sensor data: accelerometer, gyroscope and rotation vector sensor data. Before the system works, we use the quaternion method [29]–[31] to transform the accelerometer sensor data.

Let the quaternion be given by the rotate sensor: (p_1, p_2, p_3, λ), where p_1, p_2, p_3 represent the rotation vectors along the coordinate axis x, y, z respectively, and λ represents the value of the rotation vector. Accelerometer sensor raw data composed of the component of acceleration on each axis can be represented by the vector $[a_x, a_y, a_z]$. We can obtain the unaffected coordinates using Formula (1).

$$\left[a_x^{'} \; a_y^{'} \; a_z^{'} \right]^T = M \left[a_x \; a_y \; a_z \right]^T, \tag{1}$$

where M is a transformation matrix used in the quaternion method, and it can be represented as Formula (2), as shown at the bottom of the next page. We use matrix M to transform the coordinate system and convert the accelerometer data of mobile phones into carrier coordinates. The coefficient derivation of matrix M can be referred to Henderson *et al.* [32].

b: FEATURE EXTRACTION

The converted vector $[a_x^{'}, a_y^{'}, a_z^{'}]$ and the gyroscope sensor data are sampled at a 50 Hz rate and filtered for noise reduction [7]. A Butterworth low pass filter with a cut-off frequency of 0.3 Hz is used to separate the gravitational and body motion components included in the sensor acceleration signals. The rate is sufficient for capturing human body motion, that is, than 95% of its energy is contained below 15 Hz [33]. Then, the processed sequences are grouped into frames by fixed-width windows of 2.56 s and 50% overlap (128 samples per frame with an overlay of 64 samples).

We use the method in [34] for feature extraction. Each frame extracts a feature vector by computing measurements from the time and frequency domains of inertial signals. The feature vector consists of 561 features, including wellknown standard measures [35], such as mean, correlation, signal magnitude area (SMA) and autoregression coefficients[36]. In [34], new features were included: energy of different frequency bands, frequency skewness, and the angle between vectors (e.g. mean body acceleration and vector). Further details are provided in [34].

c: ACTION STATE RECOGNITION

Complex action recognition is difficult; in general, the more basic actions are recognised, the higher the accuracy obtained. We identify six kinds of basic actions (running, static, walking, climbing up, climbing down and unknown) using the GMM-HMM model [7]. Similar to the literature [7], we construct five models to recognise five basic actions except for unknown actions. Frames described by 561 features are represented as feature vectors divided into minutes, in alignment with the division of an individual's indoor/outdoor states. The mixed Gaussian distribution generated by feature vectors is processed using GMM models. Moreover, it is used as an input sequence of HMM models, and the probabilities of input action corresponding to models are obtained separately by calculating the joint probability of the sequence path of each model. Moreover, the probabilities of unknown actions are recognised as an infinite negative value.

The probabilities obtained from GMM-HMM models can be represented as $\{P_{run}, P_{motionlessness}, P_{walk}, P_{up}, P_{down}\}$, and the maximum value is expressed as follows:

$$P = \max\{P_{run} \quad P_{motionlessness} \quad P_{walk} \quad P_{up} \quad P_{down}\} \tag{3}$$

We recognise the action class label on the basis of the maximum value. However, if

$$P - T_{unknown} < 0, \tag{4}$$

where $T_{unknown}$ is set as a threshold to recognise the unknown actions, the action will be recognized of unknown actions. The output value of unknown actions becomes meagre (almost zero) after passing through the model group, whereas the probability of the corresponding action becomes high. We classify running and climbing up and down into other actions, and the six actions recognised are further classified into four actions: static, walking, other actions, and unknown actions.

2) FREQUENT LOCATION RECOGNITION

We use GPS to recognise the locations where the individual frequently visits. The frequency of GPS data downsampling to 1 Hz. Moreover, we use AGNES hierarchical cluster algorithm to understand the sites that individuals often visit.

Step 1: We token the first r points and the last r points of each trajectory. Then, the chronologically arranged points are represented as a sequence $N_1, N_2, \cdots, N_n$, and an empty set D is used for indoor locations coordinates.

Step 2: Traversing coordinate sequence $N_1, N_2, \cdots, N_n$. For each point N_i, if set D is empty, then the indoor location coordinates $d_1 = N_i$ are added to set D; otherwise, the indoor location coordinates $d_1, d_2, \cdots, d_m$ of set D is traversed, and the Euclidean distance is calculated for each indoor location d_j using Formula (5).

$$l_j = \sqrt{\left(x_{N_i} - x_{d_j}\right)^2 + \left(y_{N_i} - y_{d_j}\right)^2} \tag{5}$$

Step 3: The minimum distance $l = \min\limits_{j \in \{1, \cdots, m\}} l_j$ is taken, and then the corresponding indoor location is d_s when the distance is within the confidence range, thus updating the coordinates using Formula (6).

$$d_{s_{new}} = \frac{(d_s * k + N_i)}{(k+1)}, \tag{6}$$

where k is the number of GPS coordinate points N_i which is used in indoor location coordinate d_s. Then, the indoor location coordinates are added to set D.

Step 4: Afterwards, the coordinates are updated until set $D = \{d_1, d_2, \cdots, d_3\}$ is obtained.

In conclusion, set D contains the locations where individuals frequently visit during the sampling period.

B. HIGH SEMANTIC INFORMATION RECOGNITION

1) INDOOR AND OUTDOOR STATE RECOGNITION

Indoor and outdoor state division is an essential part of our work. The locations' time distribution obtained from the state sequence can further recognise the home and workplace. The hidden Markov model (HMM) [37] and the Viterbi algorithm [38] are used for indoor and outdoor state recognition. Moreover, two HMM models are used to make joint decisions and obtain a sequence with high accuracy.

HMM. As described in Section 4.1, the HMM model's observable state is composed of four basic actions (motionlessness, walking, other and unknown actions) and two states (strong and weak) of GPS and ambient noise. We describe the observable state as a triple elements group (A, G, V).

Our work focuses on the division of indoor and outdoor states, and the Viterbi algorithm is used to obtain the optimal prediction sequence of indoor and outdoor states. Concretely, we have

$$\delta_t(i) = \max_{1\le j\le N}\left[\delta_{t-1}(j)a_{ji}\right]b_i(o_t), \quad i = 1, 2, \ldots, N \tag{7}$$

$$\delta_1(i) = \pi_i b_i(o_1), \quad i = 1, 2, \ldots, N \tag{8}$$

$$\psi_t(i) = \max_{1\le j\le N}\left[\delta_{t-1}(j)a_{ji}\right], \quad i = 1, 2, \ldots, N \tag{9}$$

$$\psi_1(i) = 0, \quad i = 1, 2, \ldots, N \tag{10}$$

$$a_{ij} = P\left(i_{t+1} = q_j \mid i_t = q_j\right), \quad i = 1, 2; j = 1, 2 \tag{11}$$

$$b_j(k) = P\left(o_t = v_k \mid i_t = q_j\right), \quad k = 1, 2, \ldots, 16;\ j = 1, 2 \tag{12}$$

where a_{ij} is the state transition probability, and $b_j(k)$ is the observation probability. Given no prior condition, we set the initial state probability vector $\pi = (0.5, 0.5)$.

When the sequence computation terminates, we can obtain the probability P^* of the optimal path and the corresponding terminal i_T^* under the path.

$$P^* = \max_{1\le j\le N}\delta_T(i) \tag{13}$$

$$i_T^* = \underset{1\le i\le N}{argmax}\left[\delta_T(i)\right] \tag{14}$$

Then, we retrospect the optimal path to obtain the sequence of the path which can represented as $I^* = \left(i_1^*, i_2^*, \ldots, i_T^*\right)$.

$$i_t^* = \Psi_{t+1}\left(i_{t+1}^*\right) \tag{15}$$

$I^* = \left(i_1^*, i_2^*, \ldots, i_T^*\right)$ is the optimal prediction sequence of indoor and outdoor states.

Misjudgment may occur in the models, and it happens during state transitions; thus, we simultaneously use two HMM models with different state transfer matrices A to pursue high accuracy. Two sequences from the models are combined with the WiFi and location information, which are provided by the mobile phone's WiFi module and GPS data separately. WiFi name is the first condition for discriminating a change in state. If the WiFi name is not changed, then we deem the state of the individual as not altered, and vice versa. Supposing an individual's phone has no WiFi connection, we calculate the Euclidean distance between the actual and clustered locations to perceive the change of the individual's location; GPS data can obtain the real site, whereas the cluster locations are recognised by the genealogical cluster algorithm in Section 4.1. In summary, the sequences synthetically judged are the final status sequences that will be used to recognise commute time.

2) LOCATION RECOGNITION OF HOME AND WORK

The locations obtained in the sampling cycle in Section 4.1 are be counted. We recognise the home and workplace based on the basis of the general situation wherein home and commute locations are the most common places for most people. The subjects we studied had distinct characteristics of day work and night rest. Therefore, we conduct time interval distributionon the basis of the two locations with the most statistics (as shown in Fig.8). In addition, the place where time concentrates in the daytime is recognised as the commute location. Similarly, the site where time focuses in the evening will be remembered as the location of the home.

3) COMMUTE TIME RECOGNITION

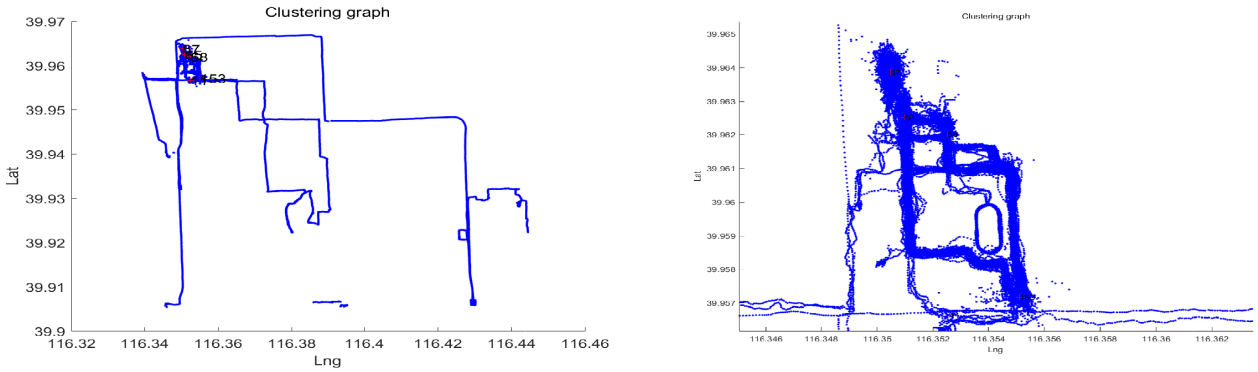
The indoor/outdoor state sequence with time information is marked with locations in set D (Section 4.1), and the home and commute location information is added. The time used between home and commute location is recognised as commute time.

V. EXPERIMENTS AND DISCUSSION

A. DATASET

We provide a publicly available dataset CPS. Table 1 shows details of the CPS. CPS contains ten individual data with an average of 27 days, and the subjects include six graduate students and four young workers aged 22 to 30 years. We used Huawei Mate 8 with the Android 6.0 system as experimental device. The participants were tasked to collect data for more than three weeks. Throughout the experiment process, WiFi and GPS remained open. We developed an

$$M = \begin{bmatrix} \lambda^2 + {p_1}^2 - {p_2}^2 - {p_3}^2 & 2(p_1p_2 + \lambda p_3) & 2(p_1p_3 - \lambda p_2) \\ 2(p_1p_2 - \lambda p_3) & \lambda^2 + {p_2}^2 - {p_1}^2 - {p_3}^2 & 2(p_2p_3 + \lambda p_1) \\ 2(p_1p_3 + \lambda p_2) & 2(p_2p_3 - \lambda p_1) & \lambda^2 + {p_3}^2 - {p_1}^2 - {p_2}^2 \end{bmatrix} \tag{2}$$



(a) Cluster map of an individual's frequent locations during the sampling period.

(b) Local magnification of (a).

FIGURE 3. **Cluster map of a monthly travel locations.**

app to collect relevant sensors data automatically. Moreover, no restrictions were imposed on the usage of the provided smartphone. The website for obtaining the dataset is https://pan.baidu.com/s/12jKE18tpO4u2ihie4AwiNw. Please contact us to obtain the extraction code if you need to use the dataset.

A prior dataset (collected from subjects) was built for training the GMM-HMM and HMM models. The previous dataset consists of two parts: five kinds of physical activities (i.e., static, walk, run, upstairs, and downstairs) and three day phone sensor data for each subject. The activity data for GMM-HMM pretraining and daily data were further divided into two segments for HMM training.

Two places are needed to set thresholds manually; thus, we used heuristic rules to fix them: (1)The division of states of GPS and ambient noise. The state of GPS determined by the number of satellites (when the number of satellites exceeds 8, it is judged as a strong state, vice versa). Similarly, the state of ambient noise is determined by a threshold of 60 db; (2)Sequence length of location clustering for trajectory interception. In this study, the value of r is set to 10. Usually, 10 points are enough to determine the location. Too many points affect the clustering centre, whereas too few points may result in noise. In addition, the length of the input and output sequence of HMM is 1,440 because a day is divided into 1,440 min.

B. COMMUTE TIME RECOGNITION

The experiment was conducted on the dataset described in Section 5.1. We present our experimental results in accordance with the structure of the hierarchical semantic model.

1) LOW SEMANTIC LAYER

The role of the lower layer is data fusion. We extract the locations information and recognise the individual's actions from the sensor data.

a: FREQUENT LOCATION RECOGNITION

On the basis of the method described in Section IV (frequent location recognition), the location information about the frequently visited places of individuals is obtained by clustering GPS trajectory.

Fig. 3 is the cluster map of an individual's frequent locations during the sampling period, and the map on the right is an enlarged local map of the main activity area of the individual. The frequent sites are marked with statistical times in the maps, and the latitudes and longitudes in blue colour represent the locus of the individual. Locations with the two most statistics are used for further time distribution analysis. In this map, these statistics are the points with statistical values of 153 and 67.

According to the cluster location, Fig. 4 shows the quantitative distribution of the subject's recognised frequently visited sites.

We compare the identified location with the standard map coordinates of the location. Considering that the experimental and standard coordinates of the location are latitude and longitude coordinates, we use the Haversine formula ([Formula(16), as shown at the bottom of the next page]) to calculate the coordinate deviation.

In the Formula (16):

r is the radius of the sphere. (r in this study refers to the radius of the earth.)

d is the distance between the two points.

ϕ_1, ϕ_2: denotes the latitude of points 1 and 2.

λ_1, λ_2: denotes the longitude of pints 1 and 2.

Table 2 details the experimental data. One point to be explained here is that the data deviation of some places in the table is relatively large, because of two reasons: 1) After the subject enters the building, the GPS signal may be weak. In this case, we can only locate the periphery of the building, but the standard landmark is the core point coordinate of the building. At this time, the main factor of the location identification error is the size of the building. Generally, within 100 meters (D-deviation less than 0.1) is within the normal error range. In some extreme cases, relatively large errors may occur. For example, the actual location is a terminal of the airport, and a coordinate point around the terminal of the location is identified. At this time, the identification error reaches the order of 100 m. 2) The

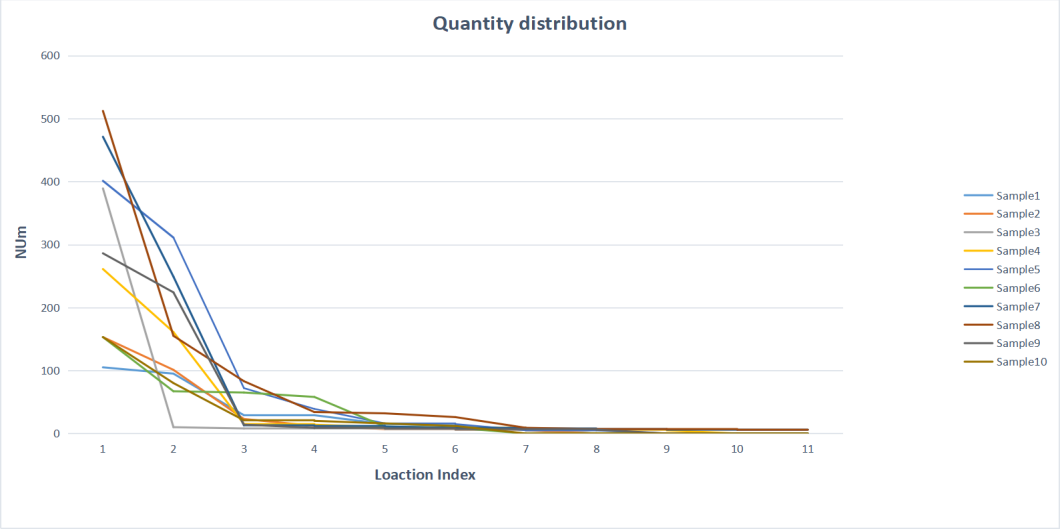


FIGURE 4. Location statistics.

TABLE 2. Location recognition deviation. 'D-deviation' means 'Distance deviation'.

		1	2	3	4	5	6	7	8	9	10	11	AVG	AVG(1-2)
Sample 1	Num	105	95	29	29	15	13	6	-	-	-	-	41.7	100.0
	D-deviation(km)	0.001	0.037	0.040	0.017	0.036	0.112	0.486	-	-	-	-	0.104	0.019
Sample 2	Num	153	101	23	12	7	6	5	-	-	-	-	43.9	127.0
	D-deviation(km)	0.017	0.056	0.009	0.015	0.144	0.055	0.010	-	-	-	-	0.044	0.037
Sample 3	Num	389	10	8	8	7	6	5	5	-	-	-	54.8	199.5
	D-deviation(km)	0.000	0.012	0.439	0.088	0.032	0.009	0.051	0.030	-	-	-	0.083	0.006
Sample 4	Num	261	161	15	14	10	9	8	6	5	-	-	54.3	211.0
	D-deviation(km)	0.053	0.000	0.009	0.101	0.531	0.014	0.014	0.006	0.000	-	-	0.081	0.027
Sample 5	Num	401	311	72	39	16	15	5	5	-	-	-	108.0	356.0
	D-deviation(km)	0.003	0.000	0.011	0.011	0.015	0.001	0.016	0.006	-	-	-	0.008	0.002
Sample 6	Num	153	67	65	58	11	9	-	-	-	-	-	60.5	110.0
	D-deviation(km)	0.000	0.011	0.001	0.003	0.005	0.000	-	-	-	-	-	0.003	0.006
Sample 7	Num	471	249	13	12	11	9	8	7	6	6	6	72.6	360.0
	D-deviation(km)	0.002	0.002	0.001	0.002	0.002	0.006	0.005	0.016	0.008	0.007	0.015	0.006	0.002
Sample 8	Num	512	155	83	34	32	26	9	7	7	6	5	79.6	333.5
	D-deviation(km)	0.012	0.001	0.003	0.001	0.015	0.002	0.003	0.001	0.002	0.002	0.001	0.004	0.007
Sample 9	Num	286	224	14	9	8	7	7	6	-	-	-	70.1	255.0
	D-deviation(km)	0.008	0.001	0.007	0.014	0.000	0.002	0.001	0.004	-	-	-	0.006	0.005
Sample 10	Num	153	80	21	20	16	12	-	-	-	-	-	50.3	116.5
	D-deviation(km)	0.014	0.008	0.001	0.016	0.140	0.029	-	-	-	-	-	0.035	0.011

system may recognise the endpoint of the track before GPS misses as a location, and the error between the target location and the recognition location is relatively large. Nevertheless, the purpose of location recognition is to identify the location information of the home and workplace, and the other locations are secondary information, whose accuracy does not affect the recognition accuracy of the last commuting time.

$$d = 2 * r * arcsin\left(\sqrt{haversin(\phi_2 - \phi_1) + cos(\phi_1) * cos(\phi_2) * havesin(\lambda_2 - \lambda_1)}\right) \tag{16}$$

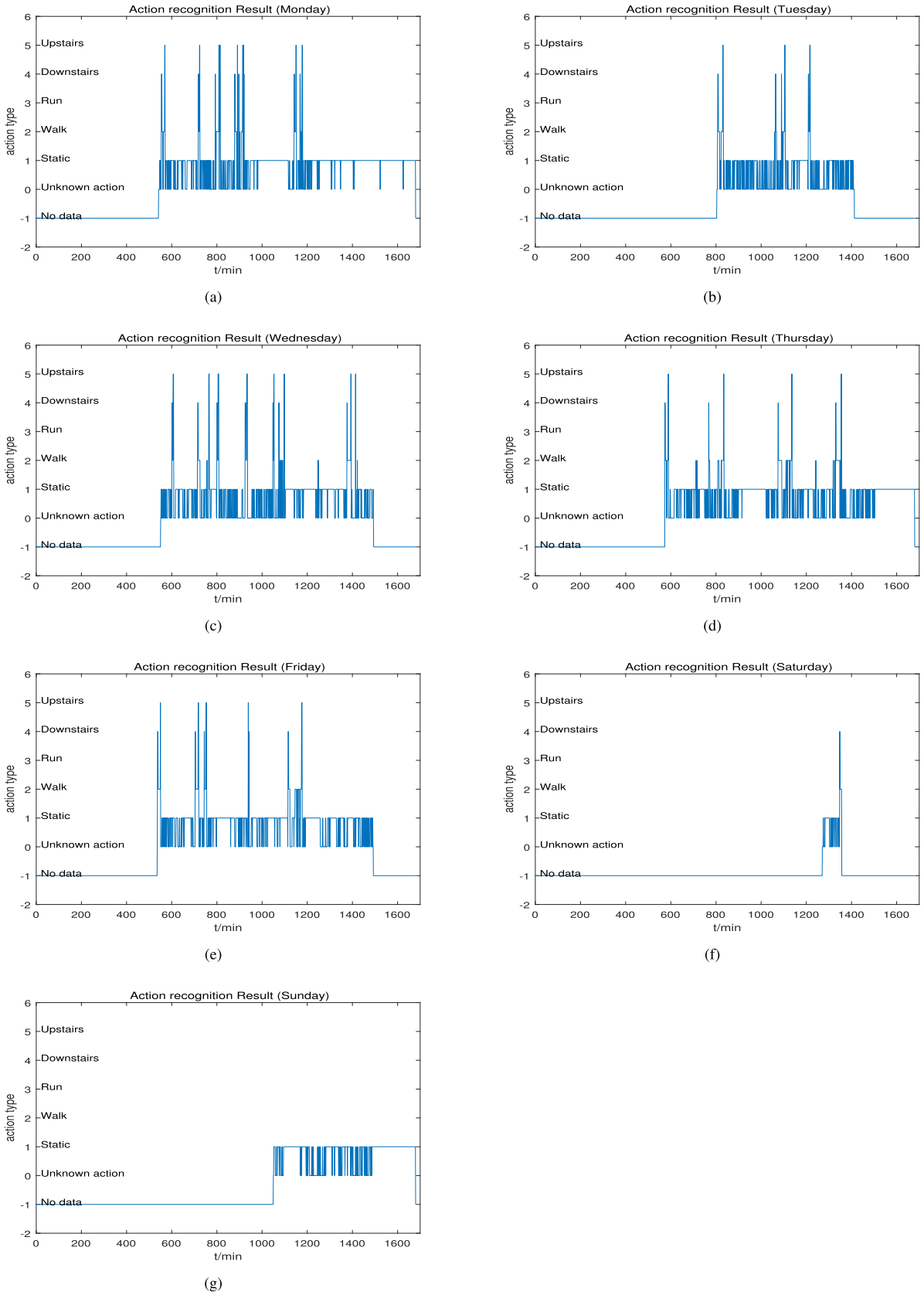
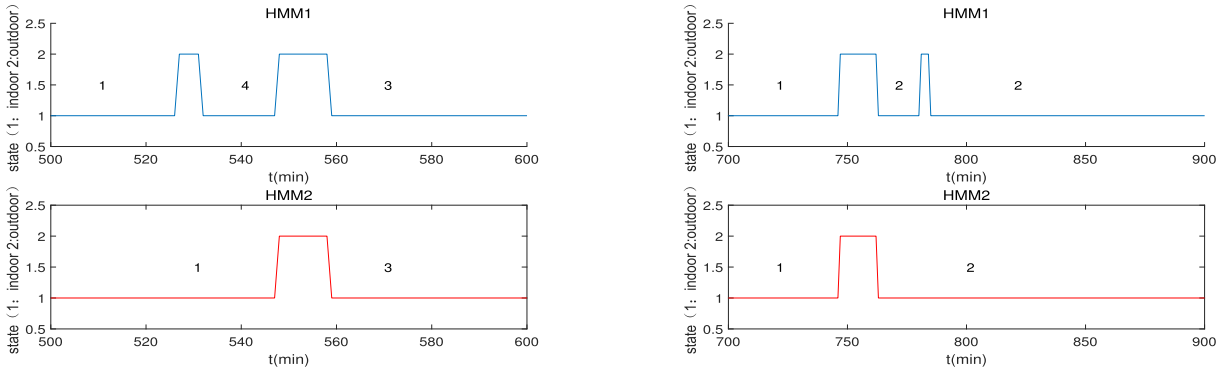
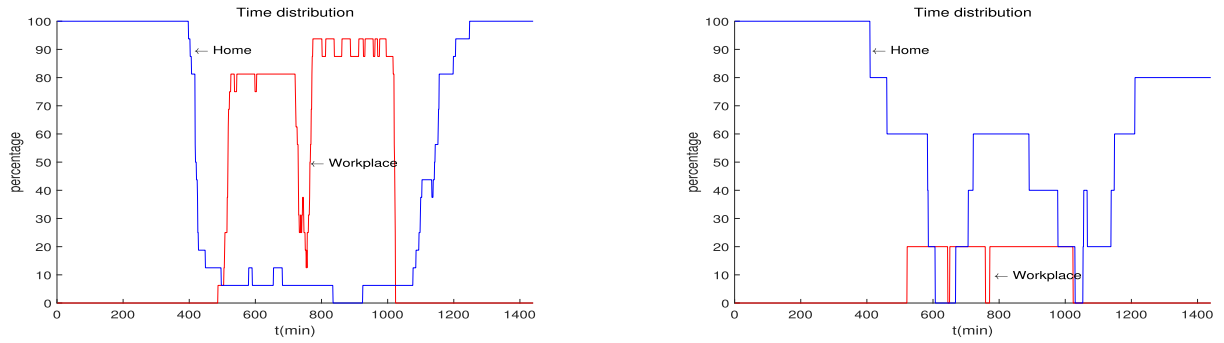


FIGURE 5. one week's action state recognition sequence of a subject.



(a) (b)

FIGURE 6. Indoor/outdoor state recognition result via two HMM models with different transform rates.



(a) Weekday (b) Weekend

FIGURE 7. Time distribution at home and workplace.

We use the two places with the most statistics as the home and wrokplace points. As shown in Table 2, we average the identification errors of the two points. The maximum average error of the experiment is 37 m, and the minimum average error is 2 m, which belongs to the normal error range (within 50 m).

b: ACTION STATE RECOGNITION

By processing the data of the cell phone accelerometer, rotation vector sensor and gyroscope, we can identify the action of the subjects and generate the action sequence. The length of the action sequence is 1440, indicating that 1440 action states are generated for each individual per day. Fig. 5 shows the action recognition results of a certain week for a subject.

2) HIGH SEMANTIC LAYER

In this layer, we process the data output from the lower semantic level to identify the indoor and outdoor state sequence of individuals. By combining the sequence with time stamp and the location data, we can derive the location semantic information of an individual's home and workplace and further identify the commuting time of individuals.

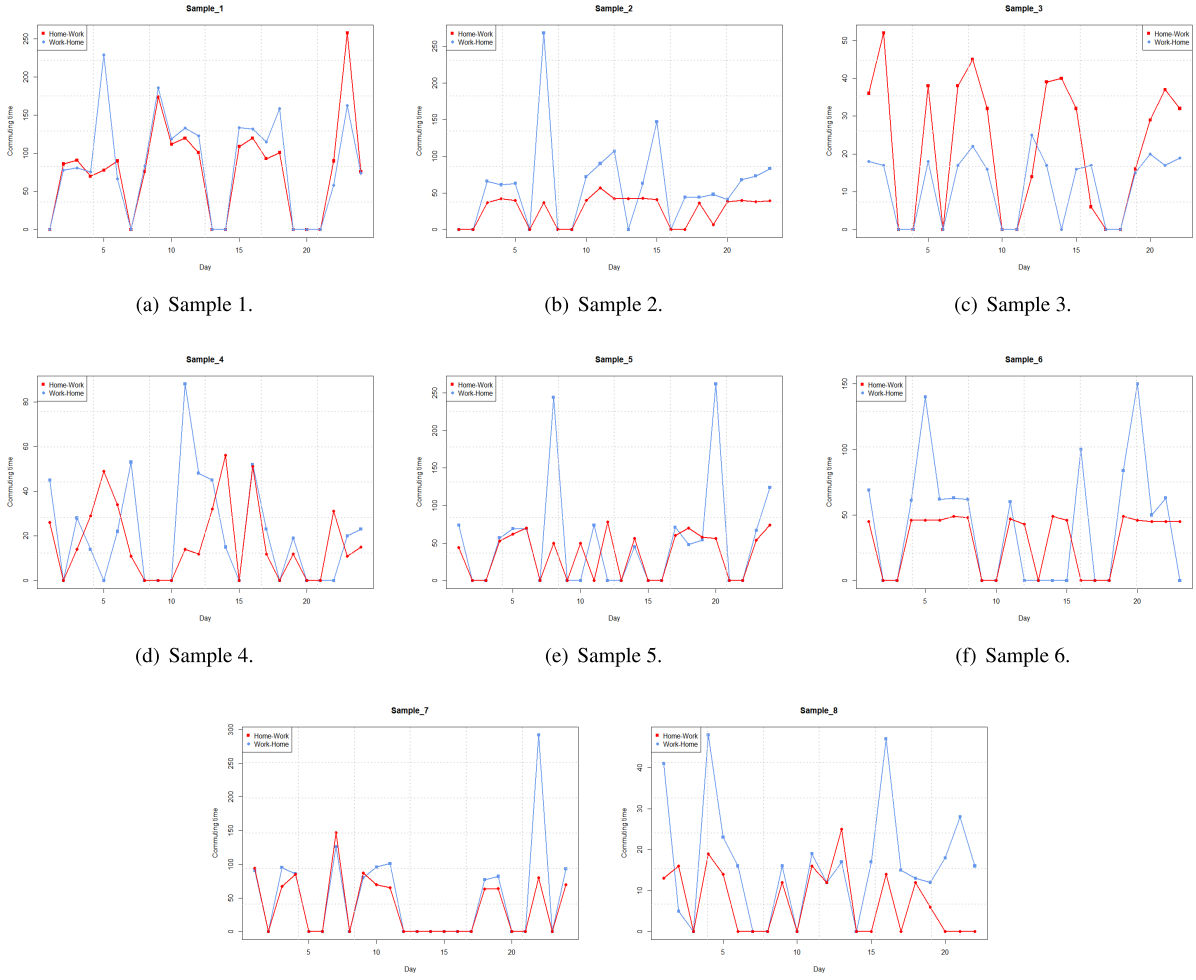
a: INDOOR AND OUTDOOR STATE RECOGNITION

The HMM group is used to recognise the indoor/outdoor state of the individuals after location recognition. Fig. 6 depicts a section of two different scenarios' recognition result via two HMM models; HMM 1, which has a critical transfer matrix is sensitive to the changes in indoor/outdoor states. By contrast, HMM2 with a smaller transfer matrix is more insensitive to the changes. The sequence of the indoor/outdoor state is divided into minutes. The indoor state is represented by value '1', and outdoor state is represented by value '2'. Moreover, the number marked represents the clustering locations obtained in Section 4.1. Different models obtain different results in some special situations, and the figure shows two different scenes that often happen in our daily life. In the left picture, the outdoor state transition is ignored by HMM2, resulting in the unrecognised location change from 1 to 4 at the 500th minute of the sequence. Similarly, the right graph shows the misidentified caused by the faster transform rate of HMM1.

b: LOCATION RECOGNITION

The two places with the most statistics are taken as the candidates for home and workplace. Given that most commuters are in commute locations by day and at home by night, two different sites are given the semantics of the home and workplace in accordance with their time distribution.

We divide each day into two sections; Section A is from the current day's 20:00 to the next day's 8:00, and section B is from the current day's 8:00 to 20:00. By comparing the time



(a) Sample 1. (b) Sample 2. (c) Sample 3.

(d) Sample 4. (e) Sample 5. (f) Sample 6.

(g) Sample 7. (h) Sample 8.

FIGURE 8. **Commute time recognition.**

proportion of two locations in the A/B period, we provide semantic information on home (A period accounts for more time) and workplace (B period accounts for more time). Fig. 7 describes the time distribution of the two locations using the statistical proportions of sites over time. Fig. 7(a) shows the time distribution of the two places on working days, wheres Fig. 7(b) shows that for the rest of the day. As is shown in Fig. 7, the time distribution of two places is remarkable different. One of the two locations is more concentrated in the evening, whereas the other location is concentrated during daytime.

AVG (1-2) in Table 2 shows the recognition deviation of home and workplace locations. We identified the home and commute locations of eight partcipants successfully; however, those of the other two subjects were not identified because their commute locations were not fixed. As shown in Table 3, our method is suitable for identifying fixed sites.

c: COMMUTE TIME RECOGNITION

By combining the recognition results of indoor/outdoor state sequence and workplace and home locations, the time used between home and commute location is recognised as commute time. Fig. 8 shows the results of commute time recognition for different subjects, the commute time spent from home to work are drawn in red, whereas that for the workplace to home is drawn in blue; the graph shown is not

TABLE 3. **Semantic inference of indoor locations.**

ID	Days	Indoor Locations (Num)	Home Recognition	Workplace Recogition
1	23	12	✓	✓
2	30	22	✓	×
3	28	13	✓	✓
4	29	10	✓	✓
5	30	8	✓	✓
6	30	12	✓	✓
7	29	13	✓	×
8	25	14	✓	✓
9	25	9	✓	✓
10	27	9	✓	✓

symmetrical. As shown in Fig. 8, many factors affect commute time. Although individuals have considerable differences, commute time fluctuates within a specific range individually. In addition, commute time is recognised as 0 min when the subjects go to work without the phone.

We thus provide a method to recognise the individual's commute time; this method essentially identifies the home and the workplace and calculates the time spent in between them. The commute time of commuters on regular working days is stable, and our experiment shows the same results.

Results show that our method is effective for commute time recognition.

VI. CONCLUSION

We propose a method for commute time recognition by using a hierarchical semantic model, which consists of two layers: low and high semantic layers. As a preliminary information fusion layer, the low semantic layer recognises the individual's basic actions and frequent locations. On this basis, we further identify home and commute locations to recognise meaningful commute-related indoor/outdoor state transition and commute time.

Our experiments show that the method is practical; however, many problems have yet to be solved. Our approach is only applicable to people who commute regularly, and it is not valid for people who work in an unstable workplace, with unpredictable working hours or whose working hours are mainly distributed in the evening. Moreover, the dataset only contains a small range; thus, whether age, ethnicity and other factors affect the of the GMM-HMM models' motion recognition effect, which leads to commute time recognition, remains uncertain. To solve the above problems, we will conduct further research to achieve the commute time recognition in complex situations.

REFERENCES

[1] J. L. Toole, S. Colak, B. Sturt, L. P. Alexander, A. Evsukoff, and M. C. González, "The path most traveled: Travel demand estimation using big data resources," *Transp. Res. C, Emerg. Technol.*, vol. 58, pp. 162–177, Sep. 2015.

[2] W. Ford, J. W. Lien, V. V. Mazalov, and J. Zheng, "Riding to wall street: Determinants of commute time using Citibike," *Int. J. Logistics*, vol. 22, no. 5, pp. 473–490, Feb. 2019.

[3] J. W. Lockhart, G. M. Weiss, J. C. Xue, S. T. Gallagher, A. B. Grosner, and T. T. Pulickal, "Design considerations for the WISDM smart phone-based sensor mining architecture," in *Proc. 5th Int. Workshop Knowl. Discovery From Sensor Data SensorKDD*, 2011, pp. 25–33.

[4] D. Anguita, A. Ghio, L. Oneto, X. Parra, and J. L. Reyes-Ortiz, "A public domain dataset for human activity recognition using smartphones," in *Proc. 21th Int. Eur. Symp. Artif. Neural Netw., Comput. Intell. Mach. Learn.*, 2013, pp. 437–442.

[5] L. Alexander, S. Jiang, M. Murga, and M. C. González, "Origin–destination trips by purpose and time of day inferred from mobile phone data," *Transp. Res. C, Emerg. Technol.*, vol. 58, pp. 240–250, Sep. 2015.

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

[7] R. San-Segundo, J. Lorenzo-Trueba, B. Martínez-González, and J. M. Pardo, "Segmenting human activities based on HMMs using smartphone inertial sensors," *Pervas. Mobile Comput.*, vol. 30, pp. 84–96, Aug. 2016.

[8] T. Ishikawa, H. Hayami, and T. Murakami, "Comprehensive evaluation of human activity classification based on inertia measurement unit with air pressure sensor," in *Proc. 24th Int. Conf. Mechatronics Mach. Vis. Pract. (M2VIP)*, Nov. 2017, pp. 1–6.

[9] A. Ghosh, A. Sanyal, A. Chakraborty, P. K. Sharma, M. Saha, S. Nandi, and S. Saha, "On automatizing recognition of multiple human activities using ultrasonic sensor grid," in *Proc. 9th Int. Conf. Commun. Syst. Netw. (COMSNETS)*, Jan. 2017, pp. 488–491.

[10] S.-H. Zhang, F.-X. Wang, J.-J. Li, H.-D. Peng, J.-H. Yan, and G.-B. Pan, "Wearable wide-range strain sensors based on ionic liquids and monitoring of human activities," *Sensors*, vol. 17, no. 11, p. 2621, Nov. 2017.

[11] D. Triboan, L. Chen, F. Chen, and Z. Wang, "Semantic segmentation of real-time sensor data stream for complex activity recognition," *Pers. Ubiquitous Comput.*, vol. 21, pp. 411–425, Jun. 2017.

[12] R. Hussein, J. Lin, K. Madden, and Z. J. Wang, "Robust recognition of human activities using smartphone sensor data," in *Proc. Int. Conf. Frontiers Adv. Data Sci. (FADS)*, Oct. 2017, pp. 92–96.

[13] R. Niko, V. Könönen, H. Hälvä, J. Mäntyjärvi, A. Lämsä, and J. Liikka, "Computational graph approach for detection of composite human activities," Tech. Rep., Dec. 2018.

[14] H. He, Y. Tan, and W. Zhang, "A wavelet tensor fuzzy clustering scheme for multi-sensor human activity recognition," *Eng. Appl. Artif. Intell.*, vol. 70, pp. 109–122, Apr. 2018.

[15] Y.-L. Hsu, S.-C. Yang, H.-C. Chang, and H.-C. Lai, "Human daily and sport activity recognition using a wearable inertial sensor network," *IEEE Access*, vol. 6, pp. 31715–31728, 2018.

[16] M. Guo, Z. Wang, N. Yang, Z. Li, and T. An, "A multisensor multiclassifier hierarchical fusion model based on entropy weight for human activity recognition using wearable inertial sensors," *IEEE Trans. Human-Mach. Syst.*, vol. 49, no. 1, pp. 105–111, Feb. 2019.

[17] C. Dugast, L. Devillers, and X. Aubert, "Combining TDNN and HMM in a hybrid system for improved continuous-speech recognition," *Speech Audio Process., IEEE Transac-tions on*, vol. 2, no. 1, p. 217 –223, Jan. 1994.

[18] S. Renals, N. Morgan, H. Bourlard, M. Cohen, and H. Franco, "Connectionist probability estimators in HMM speech recog-nition," *IEEE Trans. Speech Audio Process.*, vol. 2, no. 1, pp. 161–174, Jan. 1994.

[19] Z. Yue, X. Zou, and H. Wang, "Voice Conversion with the Combination of HMM and GMM," *J. Data Process.*, vol. 24, no. 3, pp. 285–289, 2009.

[20] G. Zweig and P. Nguyen, "Scarf: A segmental conditional ran-dom field toolkit for speech recognition," in *Interspeech*, 2010, pp. 2858–2861.

[21] Z. Pang, S. Tu, D. Su, X. Wu, and L. Xu, "Discriminative training of GMM-HMM acoustic model by RPCL learning," *Frontiers Electr. Electron. Eng. China*, vol. 6, no. 2, pp. 283–290, Jun. 2011.

[22] Z. J. Yan, Q. Huo, and J. Xu, "A scalable approach to using DNN-derived features in GMM-HMM based acoustic modeling for LVCSR," in *Interspeech*, 2013, pp. 104–108.

[23] Q. Wang, W. L. Woo, S. S. Dlay, C. S. Chin, and B. Gao, "Informed single channel speech separation with time-frequency exemplar GMM-HMM model," in *Proc. IEEE Int. Conf. Digit. Signal Process. (DSP)*, Jul. 2015, pp. 1130–1134.

[24] M. Zimmermann, M. M. Ghazi, H. K. Ekenel, and J.-P. Thiran, "Visual speech recognition using PCA networks and LSTMs in a tandem GMM-HMM system," in *Computer Vision ACCV 2016 Workshops* (Lecture Notes in Computer Science), vol. 10117, 2017, pp. 264–276.

[25] M. Heck, S. Sakti, and S. Nakamura, "Iterative training of a DPGMM-HMM acoustic unit recognizer in a zero resource scenario," in *Proc. IEEE Spoken Lang. Technol. Workshop (SLT)*, Dec. 2016, pp. 57–63.

[26] N. Vuković, M. Mitić, and Z. Miljković, "Trajectory learning and reproduction for differential drive mobile robots based on GMM/HMM and dynamic time warping using learning from demonstration framework," *Eng. Appl. Artif. Intell.*, vol. 45, pp. 388–404, Oct. 2015.

[27] R. Ahas, S. Silm, E. Saluveer, and O. Järv, "Modelling home and work locations of populations using passive mobile positioning data," in *Location Based Services and TeleCartography II*, Jan. 2009, pp. 301–315.

[28] L. Kaufman and P. J. Rousseeuw, "Finding groups in data ǁ agglomerative nesting (Program AGNES)," in *Wiley Series in Probability and Statistics*, 1990, pp. 199–252.

[29] T. Watanabe, T. Miyazawa, and J. Shibasaki, "A study on IMU-based stride length estimation for motor disabled subjects: A comparison under different calculation methods of rotation matrix," presented at the BHI Mar. 2018.

[30] L. Krieger and F. Grigoli, "Optimal reorientation of geophysical sensors: A quaternion-based analytical solution," *Geophysics*, vol. 80, no. 2, pp. F19–F30, Mar. 2015.

[31] J. Wu, Z. Zhou, J. Chen, H. Fourati, and R. Li, "Fast complementary filter for attitude estimation using low-cost MARG sensors," *IEEE Sensors J.*, vol. 16, no. 18, pp. 6997–7007, Sep. 2016.
[32] D. M. Henderson, "Euler angles, quaternions, and transformation matrices for space shuttle analysis," *Astronautics*, to be published.
[33] D. M. Karantonis, M. R. Narayanan, M. Mathie, N. H. Lovell, and B. G. Celler, "Implementation of a real-time human movement classifier using a triaxial accelerometer for ambulatory monitoring," *IEEE Trans. Inf. Technol. Biomed.*, vol. 10, no. 1, pp. 156–167, Jan. 2006.
[34] D. Anguita, A. Ghio, L. Oneto, X. Parra, and J. R. Reyes-Ortiz, "Energy efficient smartphone-based activity recognition using fixed-point arithmetic," *J. Universal Comput. Sci.*, vol. 19, no. 9, pp. 1295–1314, 2013.
[35] J.-Y. Yang, J.-S. Wang, and Y.-P. Chen, "Using acceleration measurements for activity recognition: An effective learning algorithm for constructing neural classifiers," *Pattern Recognit. Lett.*, vol. 29, no. 16, pp. 2213–2220, Dec. 2008.
[36] A. M. Khan, Y.-K. Lee, S. Y. Lee, and T.-S. Kim, "Human activity recognition via an Accelerometer-Enabled-Smartphone using kernel discriminant analysis," in *Proc. 5th Int. Conf. Future Inf. Technol.*, 2010, pp. 1–6.
[37] L. Rabiner and B. Juang, "An introduction to hidden Markov models," *IEEE ASSP Mag.*, vol. 3, no. 1, pp. 4–16, Jan. 1986.
[38] 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.

WEI LIU received the Ph.D. degree from the Beijing University of Aeronautics and Astronautics, Beijing, China, in 2003. He is currently a Professor with the Automation School, Beijing University of Posts and Telecommunications, Beijing. His research interests include human factors, human cognitive process, user's behavior analysis, and prediction.

TAO WENXUAN is currently pursuing the master's degree with the Automation School, Beijing University of Posts and Telecommunications, Beijing, China. Her research interests include pattern recognition and machine learning.

ZHONG ZIMING received the master's degree from the Beijing University of Posts and Telecommunications, Beijing, China, in 2018. His research interests include pattern recognition and machine learning.

JIANQIN YIN (Member, IEEE) received the Ph.D. degree from Shandong University, Jinan, China, in 2013. She is currently a Professor with the Automation School, Beijing University of Posts and Telecommunications, Beijing, China. Her research interests include service robot, pattern recognition, machine learning, and image processing.