

A Deep Learning Approach to Infer Employment Status of Passengers by Using Smart Card Data

Yang Zhang^{ID} and Tao Cheng

Abstract—Understanding the employment status of passengers in public transit systems is significant for transport operators in many real applications such as forecasting travel demand and providing personalized transportation service. This paper develops a deep learning approach to infer a passenger’s employment status by using smart card data (SCD) with a household survey. This paper first extracts an individual passenger’s weekly travel patterns in different travel modes from the raw SCD as a three-dimensional image. A deep learning architecture, called a thresholding multi-channel convolutional neural network, was developed to predict an individual’s employment status. The approach proposed here solves two critical problems of using the SCD for employment status studies. First, it automatically incorporates learning temporal features in different travel modes without the need for handcrafted travel feature design. Second, it considers the class-imbalance problem by leveraging the ensemble of oversampling and thresholding techniques. By applying our approach to a real dataset collected from the metropolitan area of London, U.K., about 72% of passengers were correctly categorized into six types of employment statuses. The promising results show the tight correlation between temporal travel behavior, mode choice, and social-demographic roles. To the best of our knowledge, this is the first paper to infer employment status by using the SCD.

Index Terms—Deep learning, employment status inference, travel mode choice, smart card data, temporal travel behavior.

I. INTRODUCTION

NOWADAYS, public transit (PT), e.g. bus and tube, has become one of the most preferable methods of daily travel. The modern PT network has been widely equipped with the Automated Fare Collection (AFC) systems, which collect massive smart card data (SCD) from a large population. This valuable mobility data have been intensively utilized for travel pattern analysis [1]–[4]. However, researchers have pointed out that the lack of social-demographic data (e.g. age, gender, and employment status) makes it difficult to conduct detailed travel behavior understanding.

Social-demographic information has a significant influence on individuals travel behavior and travel mode choice [5], [6].

Manuscript received April 8, 2018; revised August 20, 2018; accepted January 23, 2019. Date of publication February 20, 2019; date of current version February 3, 2020. This work was supported in part by the Consumer Data Research Centre, Economic and Social Research Council, U.K., under Grant ES/L011840/1. The work of Y. Zhang was supported in part by the China Scholarship Council under Grant 201603170309 and in part by the Dean’s Prize from University College London. The Associate Editor for this paper was C. Wu. (Corresponding author: Yang Zhang.)

The authors are with SpaceTimeLab for Big Data Analytics, University College London, London WC1E 6BT, U.K. (e-mail: yang.zhang.16@ucl.ac.uk; tao.cheng@ucl.ac.uk).

Digital Object Identifier 10.1109/TITS.2019.2896460

It is useful for transport planning, urban design, and personalized service [7], [8]. Among various socio-demographic characteristics, the ability to characterize the employment status of passengers based on SCD has numerous applications in modern PT systems and other areas. For examples:

- *Travel Demand Management (TDM)*: TDM is an important area of concern in the future intelligent transport systems. In activity-based travel demand analysis, employment status is essential for travel behavior interpretation and modeling [5]. Many reports have shown that the demographic shifts occurring in urban areas have an impact on the demand for transportation and travellers’ behavior [9], [10]. For example, Transport for London (TfL), a local government body in London, UK, takes into account the employment growth and other factors to forecast the increasing travel demand [11]. Therefore, a better understanding of the employment status composition of travellers can be leveraged to identify improvements to TDM tools and to help predict future travel demand.
- *Transport Policy Development*: From the perspective of transport policy-making, employment status information can aid in the design of optimal urban transport strategies, such as transport service levels and fares [12]. For instance, operators in South East Queensland, Australia, group passengers into six types according to age and occupation in order to design different fare collection schemes. This may help to attract more passengers to use PT.
- *Customer Service Improvement*: Realizing the employment status of passengers can help operators optimize transport planning, develop transportation service tools, and provide targeted services to improve travellers’ journey experience and access to PT. For example, transport planners can identify the spatial mismatch between distributions of PT service level and the unemployment rate in order to improve the regional public transit service and accessibility, which can help the unemployed gain employment [13].
- *Commercial Applications*: Employment status inference based on SCD can provide the spatial distribution of employment status within the neighborhood of tube/bus stations. It may help with the design and implementation of business settlements to enhance the convenience and attractiveness of the PT system. For example, Páez *et al.* [14] used SCD with survey data to identify potential business partnerships around metro stations.

The traditional way to obtain travellers' employment status is via census or travel survey, which is quite time-consuming and expensive. Fortunately, the increasing availability of SCD brings an opportunity to gain insight into the relationships between passengers' travel behavior and employment status. However, employment status inference from SCD is not a trivial task, due to several challenges. First, the raw SCD cannot be fed to a supervised learning model. The most straightforward way to depict passengers' travel behavior is to manually identify and extract a set of features. However, the hand-crafted features may not necessarily distinguish between various types of employment status. In contrast to standard classification problems, employment status prediction always suffers from the class-imbalance issue, because the majority of the population is employed and the government must control the unemployment rate. The class-imbalance issue may have a detrimental impact on classification performance.

To tackle the above-mentioned problems, a deep learning (DL) framework for employment status prediction was developed to demonstrate the possibility of inferring employment status using SCD with a household survey. Deep learning techniques, such as convolutional neural network (CNN), can extract useful features automatically. To take the advantage of DL, the raw SCD had to be transformed into a format that is acceptable for the DL architecture and efficient enough to represent the travel characteristics for employment status inference.

In the proposed methodology, passengers labeled with survey data were used to train a supervised learning model. As suggested, employment status is closely related to temporal travel behaviors and travel mode [5], [15], so an SCD representation method was used to characterize the two types of features. Each individual's temporal profiles in different travel modes were structured into a 3D image of size $N \times M \times D$, where N indicates the seven days of a week, M is the number of time slots in the day and D is the number of the travel mode. Each $N \times M$ matrix depicts the travel time distribution across each day of the week for a certain travel mode (e.g., bus or tube). It was then natural to transfer the employment inference problem to an image classification problem. Inspired by image processing, a thresholding multi-channel CNN (TMC-CNN) model was proposed to predict a passenger's employment status. The CNN-based structure obviated hand-crafted feature design and the 'multi-channel' configuration enabled it to automatically learn different temporal features in different travel modes for accurate employment status inference. Furthermore, the proposed model alleviated the class-imbalance problem via the structure of 'thresholding'.

The remainder of this paper is organized as follows. First, related work is discussed in Section II. The dataset used in this paper along with a pre-processing step are discussed in Section III. The approach to predicting passengers' employment status based on SCD is introduced in Section IV. To demonstrate its capability, the approach was tested using a real dataset and experimental results are presented and discussed in Section V. Finally, Section VI summarizes the major findings and directions for further research.

II. RELATED WORK

Employment status is a key demographic characteristic for travel behavior studies. Several works have attempted to infer the demographics from human mobility by using GPS trajectories [8], [16] or check-ins [17]. Both data sources can reveal the frequently visited locations, which appear a high degree of temporal and spatial regularity with significant probability [18]. For example, Zhu *et al.* [8] used an 18-month GPS dataset of 275 households to predict individuals' social-demographic information, including age group, gender, and employment status. Home-based tours were first detected from GPS trajectories and features and variability were then extracted, which were fed into the Support Vector Machine (SVM) and logistic regression models. The class-imbalance problem of the dataset was reported. As a result, the multi-class (six types of employment status) prediction task was converted into a binary class problem; i.e. full-time employee and others. Although the prediction accuracy of employment status was 95%, students, the retired, etc. could not be distinguished. In sum, the shortness of this kind of work is three folds. First, the sample size of the number of individuals was too small to be representative of a city population. Second, the manual feature extraction process was both difficult and expensive. Third, the coarse classification of employment status was not very useful for detailed transport planning and personalized service.

By leveraging limited travel survey data, researchers have attempted to study the correlation between human mobility and employment status using SCD. The most common way is to extract travel behavior indicators (e.g. trip accounts, number of travel days and daily travel time) and then compare among different passenger groups using statistical methods [19]–[21]. For example, Lu and Pas [5] compared the travel behavior of full-time employees with others' and demonstrated the significant difference in travel time. Scheiner and Holz-Rau [22] showed that employment status strongly affects travel distances as well as travel mode. Similar work can be also seen in [23] and [24]. The shortcomings of these works are twofold. First, while these studies highlight the correlation between travel behavior and employment status, the approaches are limited in the manual feature extraction process, which is both difficult and expensive. In addition, the extracted travel features may be not able to capture the full scale of the travel behaviors, then they may not necessarily distinguish between different employment statuses. Second, these studies stay at the step of measuring the statistical significance of these travel indicators across distinct groups of passengers. They prove the existence of the correlation but never explore how possible the travel behavior can indeed characterize a person. The previous study has stated that higher significance cannot automatically imply stronger predictivity [25]. These limitations inspire a rethink on an emerging problem: given an individual's SCD, can we possibly infer his/her employment status?

In this study, SCD is combined with household survey data to first explore the discriminative power of SCD for employment status prediction, which can be formulated as a classification problem. As opposed to the existing literature

TABLE I
LONDON'S OYSTER CARD DATA

ID	Date	Start time	End time	Boarding station	Exit station	Journey mode
100501541	2013-01-22	13:02	13:26	King's Cross	Camden Town	Tube
100501542	2013-01-22	14:50	--	29	--	Bus

that mainly mines knowledge from handcrafted travel features using traditional classifiers, this paper represents the raw SCD as a 3D image structure to capture the passengers' travel behavior. Benefit from that, the employment status inference task could then be undertaken using a CNN-based model. CNN is a popular deep learning technique that has achieved great success in the field of image classification [26]. A few recent studies have attempted to use CNN on mobility data, such as converting GPS logs into an image for travel mode detection [27]. Based on the standard CNN, a multi-channel CNN framework is proposed in this paper. The concept 'channel' was inspired by three-channel color image processing [26], [28]. Different from the standard CNN, the 'multi-channel' configuration allows different filters to be used to capture proper features from different channels and it is more robust than the standard model [29]. However, when using CNN for image classification, the task always faces the inevitable class-imbalance problem, so does the employment status prediction. Buda *et al.* [30] provided the first systematic study of the class imbalance problem in CNN and suggested a solution by combining the thresholding and oversampling. In this paper, similar techniques are applied to the proposed model to improve the performance of employment status prediction.

In summary, this paper proposes a thresholding multi-channel CNN model to predict employment status using SCD with a household survey. Different from existing works, the contributions of this work are summarized into three aspects below.

- From a *methodological perspective*, a novel temporal behavior profiling method was presented and a CNN-based architecture was developed. The proposed TMC-CNN model was used to predict passengers' employment status by leveraging SCD. This model is superior to many traditional supervised learning methods and the basic CNN model and it can successfully alleviate the class-imbalance problem.
- From an *empirical perspective*, the approach was applied to a case study of PT users in London, UK. The effectiveness of the method was validated and a detailed semantic analysis of the association between the discovered temporal travel patterns and employment status was provided. Some interesting findings are presented in the paper. It is important to understand passenger's temporal travel patterns and the choice of transportation changing with the employment status.
- From an *application* point of view, to the best of our knowledge, this is the first work to explore the possibility of using SCD as the dataset to infer employment status. This study mainly benefits the PT operators, agencies and private firms.

III. DATASETS

The dataset used in this study was provided by Transport for London (TfL), which is the local functional department responsible for the transport system in Greater London, UK. It consists of the Oyster card data (OCD) and employment working status information from the London Travel Demand Survey (LTDS), which provided the samples for training and testing the proposed model.

A. London's Oyster Card Data

The original OCD contained around 2.72 million transaction records of 9188 passengers, collected from London's public tube and bus system during 2013. After data cleaning, the rest of the dataset was made up of 33.9% tube journeys and 66.7% bus. In the OCD, each transaction record consists of the following fields: (1) the unique user ID, (2) transaction date, (3) start time, (4) end time, (5) boarding station, (6) exiting station, (7) journey mode (bus or tube). Note that in bus trip records the boarding station indicates the bus line number but not precise locations, and the exit station, and end time are unavailable. Table I shows trips records stored in the London's Oyster card.

B. Employment Status Data

The employment status information was obtained via LTDS, which is a consecutive annual survey covering the 32 London boroughs and the City of London. Each year, TfL randomly selects 8000 household in London and all members of the household aged five and over are asked to complete the individual questionnaire. This questionnaire includes demographics and travel-related information such as employment status and PT tickets held. In the original LTDS, the employment status of London residents aged 16 and over was categorized into 12 types, listed in the first column in Table II. In order to ensure the sample size in each category while preserving enough details, the original 12 categories were merged into six types: (1) full-time employee; (2) part-time employee; (3) student; (4) unemployment; (5) disability (not in employment); (6) retired.

C. Data Pre-Processing

As interviewees in the LTDS also voluntarily provide their unique Oyster card IDs, the two datasets can be matched for academic research purposes. For privacy concerns, all interviewees' card IDs were encrypted. After data matching, occasional PT users were discarded based on frequency using the 3-sigma rule [31], in order to ensure the OCD reveals a relatively complete travel diary of a user. The 6354 users' OCD was then taken as the primary sample. The proportion

TABLE II
EMPLOYMENT STATUS INFORMATION IN LTDS

Original Categories	Merged Categories	Percentage
Full-time employment: 37.24% Full-time self-employment: 6.45% Regular unpaid voluntary work: 0.97%	1. Full-time employment	44.66%
Part-time employment: 8.69% Part-time self-employment: 2.80%	2. Part-time employment	11.49%
Student: 6.97%	3. Student	6.97%
Unemployment and looking for job: 4.36% Unemployment and Waiting to take up a job: 0.51% Looking after home or family: 5.61% Other non-working: 0.38%	4. Unemployed	10.86%
Disability: 2.58%	5. Disability	2.58%
Retired: 23.44%	6. Retired	23.44%

of each employment status is shown in Table II. The dataset had a severe imbalance problem, e.g. the full-time employees are the majority class and only about 2.58% of people are on disability.

The exact alighting time of a bus trip was unavailable in OCD. To deal with it, the method proposed by Zhao *et al.* [32] was used to estimate the alighting time of the bus journey. This approach uses GIS technology without automatic vehicle location data to infer the alighting time and location of bus journeys. Approximately 71% of bus journeys can be successfully inferred. Since we mainly focus on people's temporal travel behavior, for bus journeys where the destination cannot be inferred, if the successive journey occurs within the transfer time threshold recommended by Seaborn *et al.* [33], we suggest using the time between the two successive tap-in timestamps as the travel time. Otherwise, we assume using 28 min, which is the average time taken to work reported in Transport Statistics Great Britain 2013 [34], as a proxy for the true travel time. This assumption had no significant impact on the results.

IV. METHODOLOGY

The aim of the study is to infer the employment status of passengers from SCD. Intuitively, the discovery can help in understanding the relationship between travel behavior and social-demographic roles profoundly and characterizing the travel demand accordingly. To begin, each passengers' temporal profiles are reconstructed from the SCD. A thresholding multi-channel convolutional neural network (TMC-CNN) model is then proposed to predict passengers' employment status. Finally, the metrics for performance evaluation are introduced.

A. Temporal Profile Representation as a 3D Image

The raw SCD cannot be fed directly into the supervised models. As suggested by previous works, employment status is closely related to temporal travel features and travel mode

choice [5], [15]. Therefore, how to represent the temporal behavior in different travel modes from SCD is fundamental for employment status prediction. Considering the 'week' is the most common time unit for the work cycle, passengers' weekly temporal travel profiles are built. In previous studies, the temporal profile is commonly set up by counting trips within predefined time slots according to the tap-in timestamp [3], [35], [36] to represent the travel behaviors across the day of the week. There are two drawbacks to this. First, this representation method cannot fully capture the temporal travel similarity between individuals. For example, as shown in Fig. 1, assume that the journeys A, B, and C occurred at 8:50 am, 9:00 am, and 9:50 am, respectively, and each journey lasted one hour. Considering a one-hour binning, the three journeys can be represented as the first row in Fig. 1, respectively. From a temporal standpoint, journeys B and C are more similar than journeys A and B using this representation method, which is misinterpreted the temporal similarity. Second, previous work never considered the differences in temporal behavior between different travel modes.

To tackle the abovementioned problems, a 3D temporal profiling method is proposed. Fig. 2 gives an example of the 3D temporal profile considering a one-hour binning. A 3D temporal profile is an $N \times M \times D$ array of numbers, where N is the number of days (i.e. seven days of the week), M is the number of predefined time slots, and D is the number of the travel mode. The value of each time slot is proportioned to the accumulated travel time spent within it. The 3D temporal profile describes the distribution of a passenger's travel time over each day of the week. Specifically, to compute the travel time distribution in a certain travel mode, the total weight of each trip is defined as one, and a trip's weight is assigned to each time slot, which is the travel time that falls into the slot divided by the total travel time. Finally, the value of a time slot is the sum of all assigned weights. For example, in Fig. 1, journey A represents a passenger travelling on a Monday from 8:50am to 10:30am (across three time-slots from 8 to 10). The total travel time is 100 min and the travel time spent in

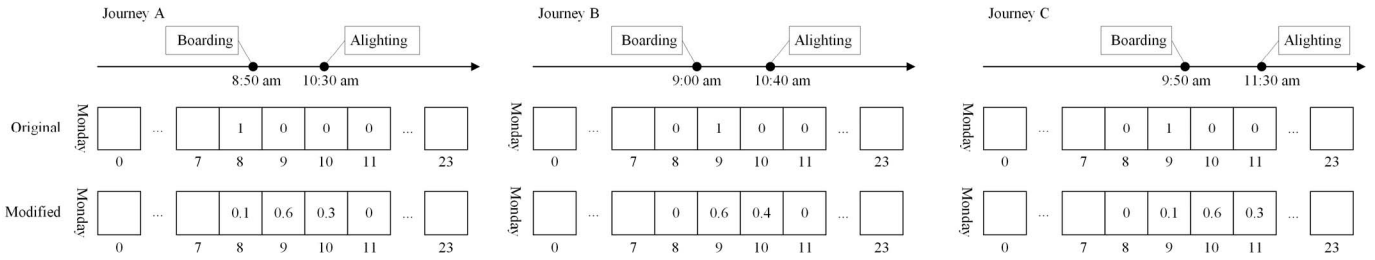


Fig. 1. The comparison of the original and modified temporal profile representation method. Here a grid is a one-hour time slot on Monday.

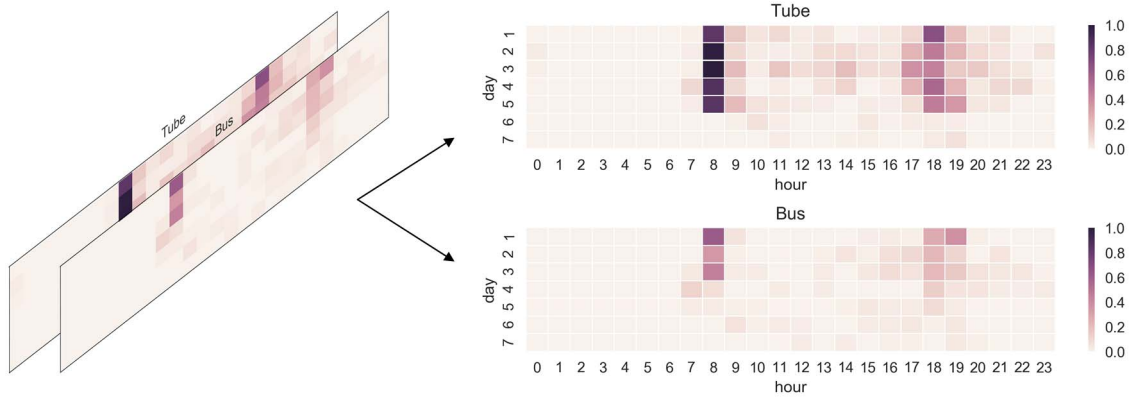


Fig. 2. A $7 \times 24 \times 2$ profile generated from a passenger’s SCD. It describes the travel time distribution over one-hour time slots of each day of the week by bus and tube. The colour of each grid is proportion to the travel time spent in it, the darker the higher.

each time slot was 10, 60, and 30 min, respectively. Then, the corresponding increments of the three slots were 0.1, 0.6, and 0.3. The temporal similarity of the three journeys were then measured, and journey A was more similar to B than journey B was with C. In this way, a user’s temporal travel behavior is represented as two 2D matrices, one in tube mode and the other in bus mode, just like a 3D image with two color channels. For standardization, each 3D temporal profile is scaled by its maximum, such that the maximum value of all time slots is one. Feature scaling aims to have all data on the same scale in order to ensure the learning ability of the model. According to [5], the employment status does not have a direct effect on the number of trips. In this way, the temporal profile can keep the original travel time distribution while eliminating the impact of the different scale of the number of trips between users. Therefore, it is reasonable to scale the temporal profiles using the individual maximum value.

B. A TMC-CNN Based Classification

After extracting temporal profiles of individuals, the employment status can be inferred using classifiers. A wide range of traditional classifiers has been applied for performing the classification task, including support vector machine, decision tree, and Naïve Bayes [8], [37]. Such classifiers usually need hand-crafted features as input for training. In this study, the proposed temporal profile has a 3D image structure, which provides an opportunity for classification without manual feature engineering processes by using a deep learning architecture. We do so by training a

model called thresholding multi-channel convolutional neural network (TMC-CNN) from the data.

The entire framework of the proposed TMC-CNN model is displayed in Fig. 3. In this paper, the architecture consists of two parallel channels. Each channel is applied to each temporal profile (a two-dimensional matrix $\mathbf{x}^{(i)} \in \mathbb{R}^{N \times M}$ ($i = 1, 2$)) in a different travel mode. This architecture can be extended to include other travel modes if required. Each channel has convolutional layers and average-pooling layers. The outputs of the two channels are concatenated and fed to several fully-connected layers. The configuration is finally completed by adding a thresholding layer for class-imbalance compensation.

The novelties of TMC-CNN are in two aspects:

- *Convert the standard CNN structure into multi-channel:* Considering the proposed 3D temporal profile is made up of two 2D matrices, representing the different temporal travel behaviors in bus or tube mode, the original CNN is modified by adding a structure of parallel channels that consists of several convolutional and pooling layers. The reason for this modification is that the temporal travel pattern in distinct travel mode may vary greatly. The standard CNN combines the two temporal patterns before being fed to the model, i.e., pixel-level information fusion. However, the multi-channel architecture enables different sets of filters to be used in each channel so that proper temporal features can be automatically extracted from different travel modes. This is a feature-level information fusion architecture and it can use richer information obtained from different channels than the standard CNN.

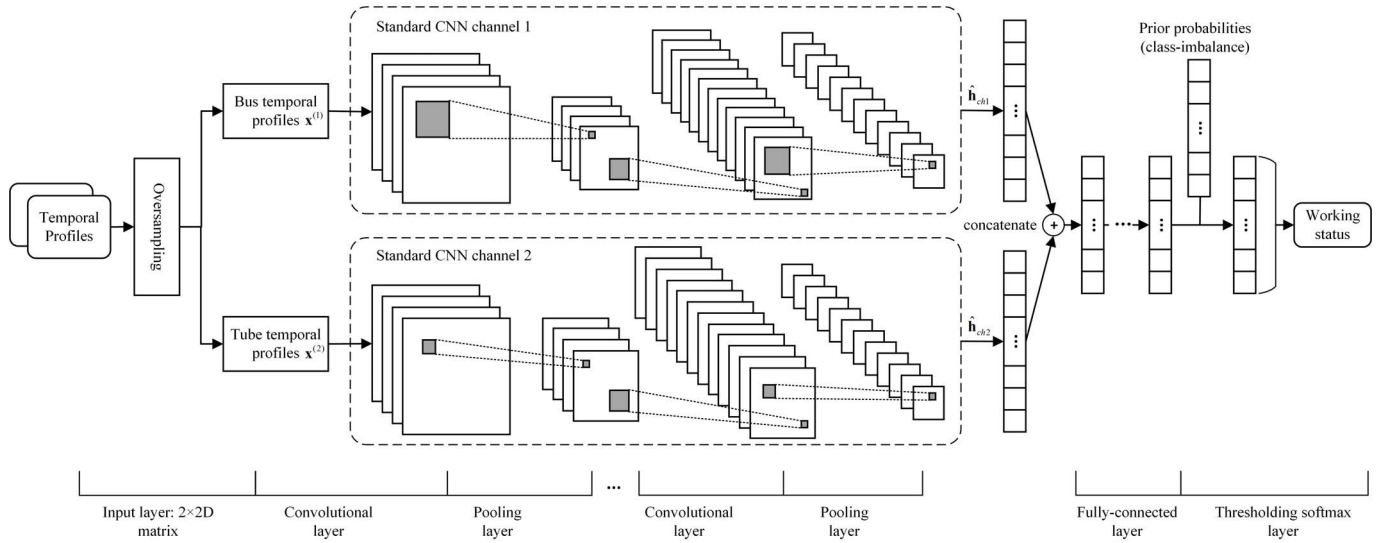


Fig. 3. Thresholding Multi-Channel Convolutional Neural Network (TMC-CNN) architecture. The architecture consists of two parallel channels, and each channel has several convolutional layers, and pooling layers. Then, the outputs of the two channels are concatenated and fed to the fully-connected layers. Finally, the configuration is completed by adding a thresholding softmax layer for class-imbalance compensation and classification.

- *Combine oversampling with thresholding to solve the class-imbalance problem:* This paper uses the ensemble of oversampling and thresholding techniques to address the class-imbalance problem. First, before training the model, the training dataset is oversampled at a certain rate. Then, a thresholding layer is used to compensate for prior class probabilities. This ensemble method reduces the error caused by a class imbalance and has no significant impact on the model's training speed.

In the following subsections, more details of the proposed TMC-CNN configuration are provided.

1) *Input Layer:* In TMC-CNN, the input of each channel is a 2D matrix $\mathbf{x}^{(i)} \in \mathbb{R}^{N \times M}$. In the input layer, the oversampling technique is applied to the original training data to partially solve the class-imbalance issue. Research shows that oversampling the minority by equalizing the size of all classes is not the optimal way to solve the problem [38]. The second drawback of full oversampling is that it greatly enlarges the sample size, which significantly slows down the training process. Therefore, in this paper, rather than simply oversampling all classes to full balance, only a 50% oversampling rate is used. For example, if the original dataset contains 100 positive and 10 negative samples, the oversampled dataset with 50% rate contains $10 + (100 - 10) \times 50\% = 55$ positive samples and 100 negative samples.

2) *Convolutional Layers:* The convolutional layer in each channel has several trainable convolution filters. The convolution operation is performed by sliding the filter over the input matrix. At any location, the region of the input covered by the filter is called the receptive field. Each filter strides around the input matrix to compute a dot product between the entries of the filter and the receptive field to generate a feature map. The filter of the i -th channel is denoted as $\omega_j^{(i)} \in \mathbb{R}^{p \times q}$, where $j = 1 : J$ is the number of filters. Given an input $\mathbf{x}^{(i)}$, the

feature map can be formulated as:

$$\mathbf{h}^{(i)} = f(\omega_j^{(i)} \otimes \mathbf{x}^{(i)} + \mathbf{b}^{(i)}), \quad (1)$$

where \otimes is the convolution operator, $\mathbf{b}^{(i)} \in \mathbb{R}^N$ is a bias term in the i -th channel and f is a non-linear activation function. Here, the Rectified Linear Unit (ReLU) function is used as the activation function.

The convolution operation can preserve the spatial relationship between pixels. In this case, the 'spatial' relationship in the temporal profile can be regarded as the travel time distribution characteristics across the time bins over the day of the week.

3) *Pooling Layers:* Pooling, also call subsampling, is used to subsample the feature map extracted by the convolutional layers to progressively reduce its dimensionality, decreasing the number of parameters and the computation cost in the network. It is commonly inserted between successive convolutional layers, making the model robust to small variations and avoiding overfitting. Here, one of the most commonly used pooling methods is used, i.e. average pooling, which keeps the average value within an extracted subarea of the feature map and discards all other values.

4) *Fully-Connected Layers:* The outputs of the convolution-pooling layers are flattened and concatenated as a single vector which is passed to the fully-connected layers. These layers connect the neurons in one layer to those in the successive layer. In principle, this is the same as the traditional multi-layer perceptron neural network (MLP). In fully-connected layers, the dropout technique [39] is adopted as the regularization method. It randomly drops neurons with a certain probability along with the connections during training, which can significantly reduce overfitting. The fully-connected layer in the TMC-CNN model follows the equation:

$$\hat{\mathbf{h}} = \mathbf{w} \cdot (\hat{\mathbf{h}}_{ch1} \oplus \hat{\mathbf{h}}_{ch2}) + \mathbf{b}, \quad (2)$$

where \mathbf{w} is the weight vector, $\hat{\mathbf{h}}_{chi}$ is the output of the last layer of the i -th channel after dropout operation, \mathbf{b} is the bias term, and \oplus is the concatenation operator.

5) *Thresholding Softmax Layer*: To tackle the class-imbalance problem, besides partially oversampling the original dataset, a thresholding layer is added after the fully-connected layer. Thresholding, also known as threshold moving, adjusts the decision boundary of a classifier. Generally, the adjusted decision boundary can be set to minimize the overall cost of errors. One simple way to implement the thresholding is to compensate for prior class probabilities [40], which are computed for each class after the partial oversampling has been applied. The neural network can then estimate Bayesian posterior probabilities as follows. For a given data point \mathbf{x} (input feature), their output for employment status e_i is:

$$P(e_i|\mathbf{x}) = \frac{P(e_i) \cdot P(\mathbf{x}|e_i)}{P(\mathbf{x})} \quad (3)$$

where $P(e_i)$ is the prior probability of employment status e_i , $P(\mathbf{x}|e_i)$ is the output of the last fully connected layer and $P(\mathbf{x})$ is common to all classes and usually omitted. Finally, the softmax function is used as the activation function for the multiclass classification.

C. Performance Metrics

With regard to the class-imbalance problem (test data is also imbalanced), the performance of prediction is evaluated by the weighted Precision (*prec*), Recall (*rec*), and *F1*-score (*F1*). In a multi-class classification task, the notions of precision, recall, and *F1*-score can be applied to each class independently. To make the definition more explicit, let L denote the set of labels, y_l and \hat{y}_l represent the set of *true* and *predicted* samples with label l , respectively. Then the three metrics are calculated as follows:

1) *Weighted Precision*: Evaluate the percentage of the corrected predicted samples in the set of all samples that are predicted to belong to the targeted class, defined as

$$prec = \frac{1}{\sum_{l \in L} |\hat{y}_l|} \sum_{l \in L} |\hat{y}_l| prec_l \quad (4)$$

where $prec_l = |y_l \cap \hat{y}_l| / |\hat{y}_l|$.

2) *Weighted Recall*: Evaluate the percentage of the corrected predicted samples in the set of all samples that truly belong to the targeted class, defined as

$$rec = \frac{1}{\sum_{l \in L} |\hat{y}_l|} \sum_{l \in L} |\hat{y}_l| rec_l \quad (5)$$

where $rec_l = |y_l \cap \hat{y}_l| / |y_l|$.

3) *Weighted F1-Score*: The trade-off between *prec* and *rec*, A *F1*-score reaches its best value at 1 and its worst value at 0, defined as

$$F1 = \frac{1}{\sum_{l \in L} |\hat{y}_l|} \sum_{l \in L} |\hat{y}_l| F1_l \quad (6)$$

where $F1_l = 2prec_l \times rec_l / (prec_l + rec_l)$.

V. RESULTS AND DISCUSSION

In this section, the proposed framework is evaluated for inferring employment status using a real dataset from London, UK. The effectiveness and robustness of the modified temporal profiling method is first evaluated. Results of the proposed model compared to other baselines are then presented. The influence of the training percentage on the prediction performance will be discussed. Finally, an intensive semantic analysis of the temporal profiles of different employment status groups is provided to further interpret this work.

A. Determination of the Structure of a TMC-CNN

With regard to the structure of a TMC-CNN model, hyperparameters of the TMC-CNN must be determined, including the number of layers, the order of the layers and the number and sizes of the filters in each convolutional layer of each channel. Parameter settings have a great impact on the experimental results. To determine the best hyperparameters, the basic way is to change one of the parameters while the other parameters remain unchanged, which is called a grid search. For each channel of the TMC-CNN, the number of a combination of the convolutional and average pooling layers was chosen from 1 to 4, and each layer's number of filters was selected from [4], [8], [16], [32]. In addition, to the best of our knowledge, there is no rule of thumb to choose the size of the filters. Therefore, the filter size in each convolutional layer was chosen from $[(1 \times 2), (1 \times 3), (2 \times 2), (2 \times 3)]$. To simplify the grid search process, we keep the parameters of one channel unchanged and optimize the other's parameters. The filter size in each average pooling layer was fixed to be (1×2) , the filter's stride step in all layers was set to be 1, and the fraction rate of dropped units (dropout probability) in each fully-connected layer was 0.2. Finally, multi-channel configurations were completed by adding one to three fully-connected layers. Except for the last fully-connected layer, the outputs of all convolutional and fully-connected layers are activated by ReLU.

In the training process, five-fold cross validation over the training dataset and grid search were conducted to find the optimal combination of parameter settings. The categorical cross-entropy loss was used as the training objective function and the Adam optimizer [41] with the default settings was utilized to update the parameters for minimizing the loss. The batch size for training was 128 and the number of epochs was 50.

The optimal configuration of the TMC-CNN could then be determined. The optimal results presented in this paper were obtained by the TMC-CNN in the following configuration. For each channel, it consisted of three combinations of a convolutional layer and an average pooling layer. The filter size of the convolutional layers in the tube channel was (1×2) and in the bus channel was (1×3) . The filter number in each convolution layer in both channels was 8, 8, and 16. Finally, two fully-connected layers were applied in the TMC-CNN model. Note that the optimal configuration was achieved when using different kernel size in different channel. In addition, the kernel size in the bus channel is larger than

TABLE III
RESULTS OF SENSITIVITY AND EFFECTIVENESS EVALUATION
OF THE MODIFIED TEMPORAL PROFILES

Temporal Profiles	<i>prec</i>	<i>rec</i>	<i>F1</i>
30 mins	0.725	0.719	0.715
60 mins	0.722	0.716	0.710
120 mins	0.619	0.590	0.591
240 mins	0.511	0.504	0.507
30 mins (without modes)	0.472	0.510	0.489

in the tube channel. This may indicate that the temporal travel pattern by bus was more diffused than by tube.

B. Experiment Results

1) *The Sensitivity and Effectiveness of Temporal Profiles:* Regarding the profiling method, the passenger temporal travel behavior is discretized into a series of finite time-bins. It is necessary to carry out a sensitivity analysis for the discretization threshold, that is, the length of the time-bin. In the existing literature, the discussion on the threshold was extremely insufficient. This analysis investigates how the time-bin length affects the temporal behavior representation, and consequently the employment status inference. The length of the time-bin ranges as [30, 60, 120, 240] min. In addition, to validate the effectiveness of the proposed temporal profiling method, we also test the temporal profile constructed by counting the trips using tap-in timestamp within each 30-min time slots but ignoring the travel mode choice, denoted as ‘30mins without modes’ in Table III.

In this experiment, 80% of the labelled data was used for training and the remaining 20% for testing. The employment status prediction results using different temporal profiles are shown in Table III. It can be observed that time-bin granularity can impact the prediction performance. The three metrics of the first two temporal profiles are quite approximate, but the performance decreases obviously when the length of the time-bin extends to 240 minutes. In addition, comparing the first and the last rows of Table III, it shows the prediction performance using the proposed profiling at 30-min binning method is much better on employment status inference task than that ignoring the travel modes at same time-bin. This proves the travel mode preference is necessary for employment status inference.

2) *Prediction Performance Comparison With Other Classifiers:* In this part, the prediction performance obtained by the proposed model was compared with a number of baselines. The same temporal profiles with 30-min time-bins were used. The standard models include Random Forest (RF), Logistic Regression Classifier (LR), Naïve Bayes (NB), and Multi-Layer Perceptron (MLP). These models are selected as the baselines because each has been widely adopted for demographic prediction in existing literature [8], [42]–[44]. To verify the effectiveness of the modification on the configuration, the performance of CNN, multi-channel CNN without

TABLE IV
PREDICTION PERFORMANCE COMPARISON BETWEEN
DIFFERENT MODELS BY USING THREE METRICS

Model	<i>prec</i>	<i>rec</i>	<i>F1</i>
RF	0.302	0.432	0.350
LR	0.276	0.305	0.286
NB	0.355	0.366	0.360
MLP	0.270	0.239	0.253
CNN	0.472	0.597	0.529
MC-CNN	0.522	0.717	0.597
TMC-CNN	0.725	0.719	0.715

thresholding layer (MC-CNN), and the proposed TMC-CNN model were compared. In this experiment, 80% of data was used for training and the rest for testing. The comparison results are shown in Table IV.

According to Table IV, it proves that all CNN based models effectively improve the prediction results compared with the first five traditional machine learning models. This may be because the CNN model can capture the temporal regularity and variation across the days of the week from the temporal profiles. In addition, comparing the results between CNN and TMC-CNN, it proves that the modification of the multi-channel configuration performs better than the standard CNN, in terms of the three matrices. Finally, comparing with MC-CNN, using the proposed TMC-CNN model, the best *prec* result was obtained at 72.5% and *F1* result at 71.5%, which is dramatically better than the MC-CNN’s (52.2% and 59.7%, respectively). With a similar value of *rec*, we can conclude that when taking the class-imbalance into account, the prediction using the TMC-CNN is much more accurate.

In summary, the success of the multi-channel configuration indicates that people with different employment statuses may exhibit distinct travel behaviors in different travel modes. In addition, oversampling combined with thresholding technique can effectively avoid over-weighted categories biasing the inference results. The promising results show the close relationship between people’s employment status and their temporal habits, as well as the travel mode choice.

To further analyses the TCM-CNN’s performance, we continue to dive into the prediction details. In Fig. 4, the overall prediction performance using the TMC-CNN model is demonstrated in the confusion matrix. The colour of each grid indicates the probability of a category in the row being estimated as the one in the column. In most cases, the probability of a correct prediction is larger than that of misclassification, except for the inference of part-time employees. About 41% of part-time employees were misclassified as full-time employees while about 19% of them are misidentified as unemployed. This phenomenon may suggest that part-time employees have more flexible and diverse temporal mobility patterns because of the various work schemes for part-time jobs. For disabled people, the ratio of misclassification as unemployed was

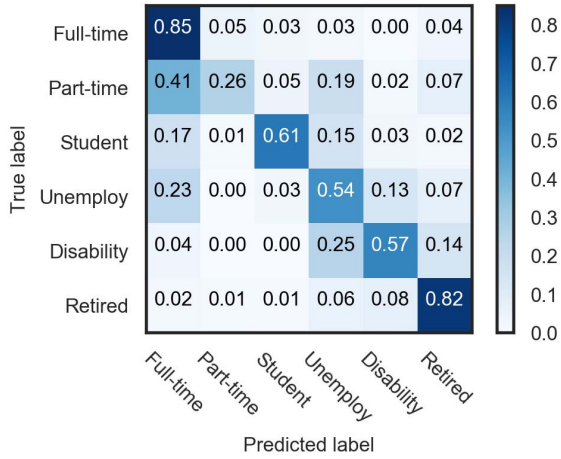


Fig. 4. Confusion matrix for employment status inference.

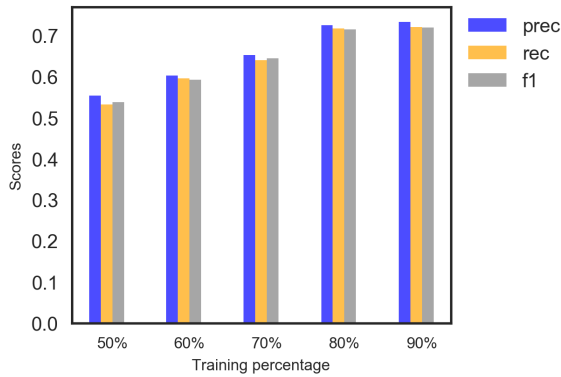


Fig. 5. Performance of employment status inference with different training percentage.

higher, but considering both categories are not under employment, the results are acceptable. The model performs best on the inference of the elderly and full-time employees. This indicates that the two groups of passengers are of high temporal stability and regularity. Moreover, the ratio of misinterpreting students, or the unemployed as full-time employees was higher than others. This may be because the class-imbalance problem has not been completely eliminated. To some extent, the temporal behavior in combination with mode choice inevitably has limits to employment status prediction.

3) *Influence of Training Percentage:* Fig. 5 shows the prediction performance when varying the ratio of the training dataset. The three matrices increase as the training percentage increases from 50% to 90%. This indicates that the training data size has a positive impact on inference performance. However, when the percentage reaches 80%, no obvious improvement beyond could be obtained on employment status inference. It means the prediction performance converges at 80%.

C. Semantic Analysis

To discover why temporal profiles can be used for employment status inference, a semantic interpretation of the results is presented. First, the aggregated temporal profiles of all

correct predictions in the test dataset are provided to explore the similar travel behaviors of passengers with the same employment status. The aggregated temporal profiles of several misclassified cases are then analyzed, from which the limitations of the proposed approach can be determined. For convenience, the six categories of passengers in Table II are denoted as group 1 to 6.

1) *Aggregated Temporal Profiles of Correct Predictions:* For a more direct understanding of the discovered differences between the distinct employment status, the temporal patterns of correctly predicted passengers with the same employment status are aggregated using the test dataset. It indicates that the temporal pattern that can be learned by the proposed model. The collective temporal patterns of the six categories are visualized in Fig. 7.

According to Fig. 7, group 1 presents a typical work-home commuting pattern with twin peaks in tube and bus mode. The first peak of travel demand appears between 7 and 8am and the second between 5 and 6pm on weekdays. In line with the confusion matrix, approximate 85% of full-time employees can be correctly inferred, which suggests that most of the full-time employees have extremely fixed work time and use PT exclusively for commuting purposes. As for travel mode choice, the tube is more popular than the bus for full-time employees. The reason may be twofold. First, full-time employees usually have steady incomes, therefore affording the expensive prices of tube trips. Second, the tube is faster than the bus and seldom delayed by traffic congestion, which is more reliable for daily commuters.

Temporal patterns of part-time employees and students are quite similar. Both exhibit a less dramatic morning peak and a more diffuse evening peak of usage during weekdays. Comparing with the collective temporal pattern of group 1, there are several obvious differences. First, the evening peak shifts can be seen clearly, especially among bus riders. Second, a much more diffuse usage can be observed between the two peaks. This implies the flexible work schedules of the two groups of passengers. Contrary to group 1, the proportion for bus trips of groups 2 and 3 is slightly higher than that of tube trips. Considering the groups 2 and 3, one of the main differences is the time of the second usage summit. Specifically, the evening usage peak of part-time employees occurred from 2 to 3pm whereas that of students appeared from 3 to 4pm. For part-time employees, it indicates that the main part-time work schedule that can be detected by the model is from the early morning to midday on weekdays. For students, one of the obvious features is that they are more likely to make trips late into the night, especially on weekends. These activities are most likely for entertainment purposes because 75.5% of the students interviewed in LTDS are between 16 to 25 years old.

Similar temporal patterns are also apparent in the last three groups, where all passengers are not employed. The most typical attribute is that bus trips make up a much larger proportion of the total trips. Concretely, the bus trip ratio for the disabled is the highest, accounting for 88.64% of the total trips, whereas that of the unemployed and the retired is about 73.96% and 80.16%, respectively. In addition to

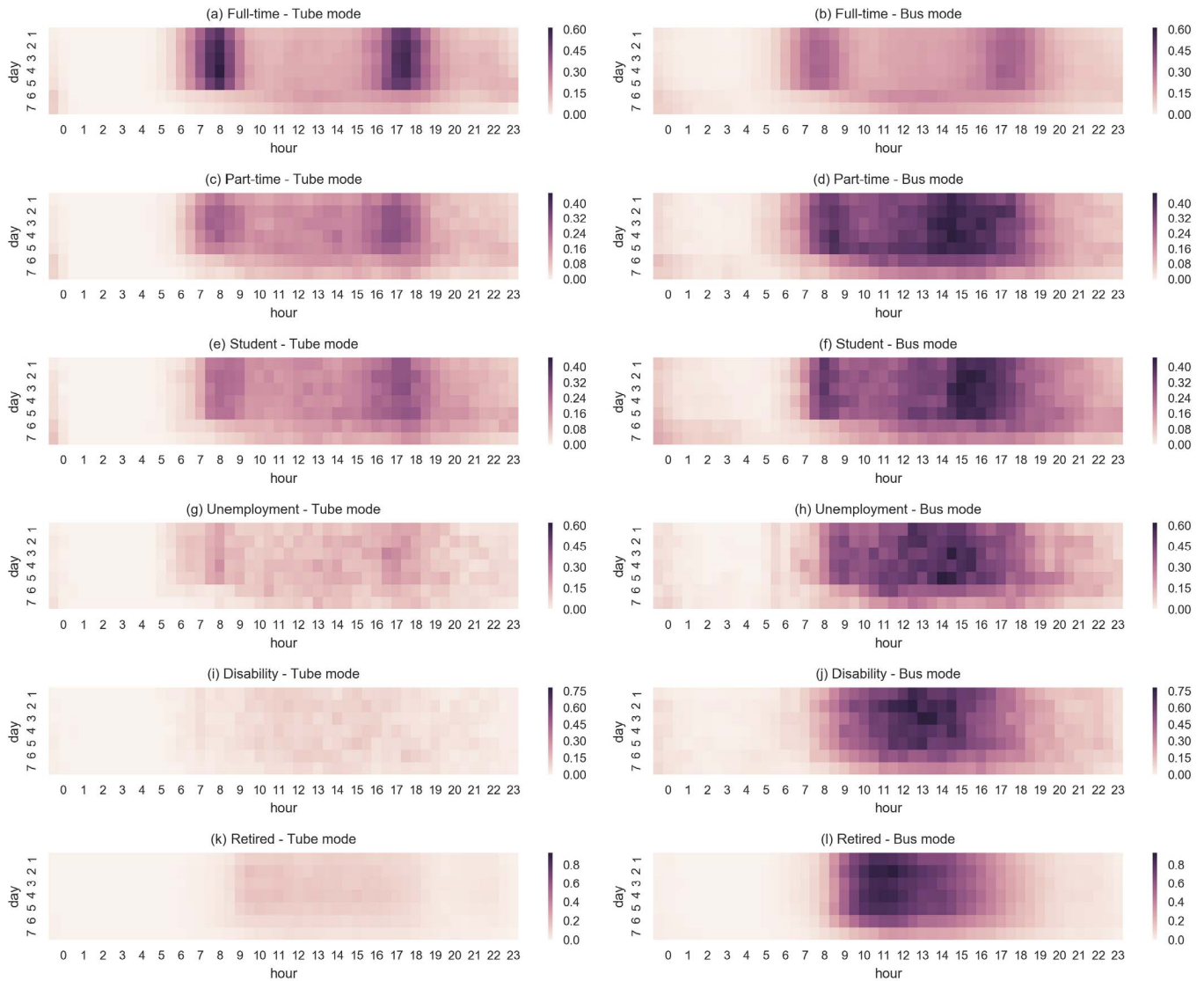


Fig. 6. Collective temporal profiles of six employment statuses under tube (left) and bus (right) mode, respectively. The heatmap shows the travel time distribution across the week. The colour of each grid indicates the percentage of travel time within in it and the two matrices of each category are normalized to sum to 100%.

the aforementioned economic reason, for the disabled and the retired passengers, low accessibility of the tube system may be another notable barrier to tube usage. For example, according to the TfL's report [11], for disabled passengers in London, especially wheelchair users, travelling on the tube can take a lot longer because of a shortage of step-free access to the underground. Besides the highly biased travel mode preference, another characteristic is the diffuse usage of PT during different periods of the daytime. For example, the bus trips made by the retired were centralized in the late morning, whereas the disabled travelled more in the early afternoon. For unemployed people, the diffuse usage started at about 8am and spanned until the evening peak time. Overall, people without jobs tend to have lower travel rates by tube, with much less commuting.

2) *Misclassification Analysis*: Fig. 7 displays three typical misclassified cases with relatively high misclassification rates.

According to Fig. 4, around 41% of part-time employees were misclassified as full-time. Observing the Fig. 7 (a), such misclassified part-time employees did exhibit very similar temporal travel behaviors and travel mode preferences to those employed full-time. It is likely because there are various work schedules of part-time jobs in the UK, such as job-sharing arrangements and term-time working arrangements. For instance, term-time working is designed primarily to help parents of school-age children, where an employ can work a particular number of weeks per year on either a full or part-time basis and take time off during the summer, Christmas and Easter holidays. Thus, the aggregated weekly temporal profiles of people with term-time working arrangements are quite similar to the full-time employees'. Such an intricate part-time scheme cannot be entirely learned by the proposed model.

In addition, 19% of part-time employees were misclassified as the unemployed while 23% of the unemployed were

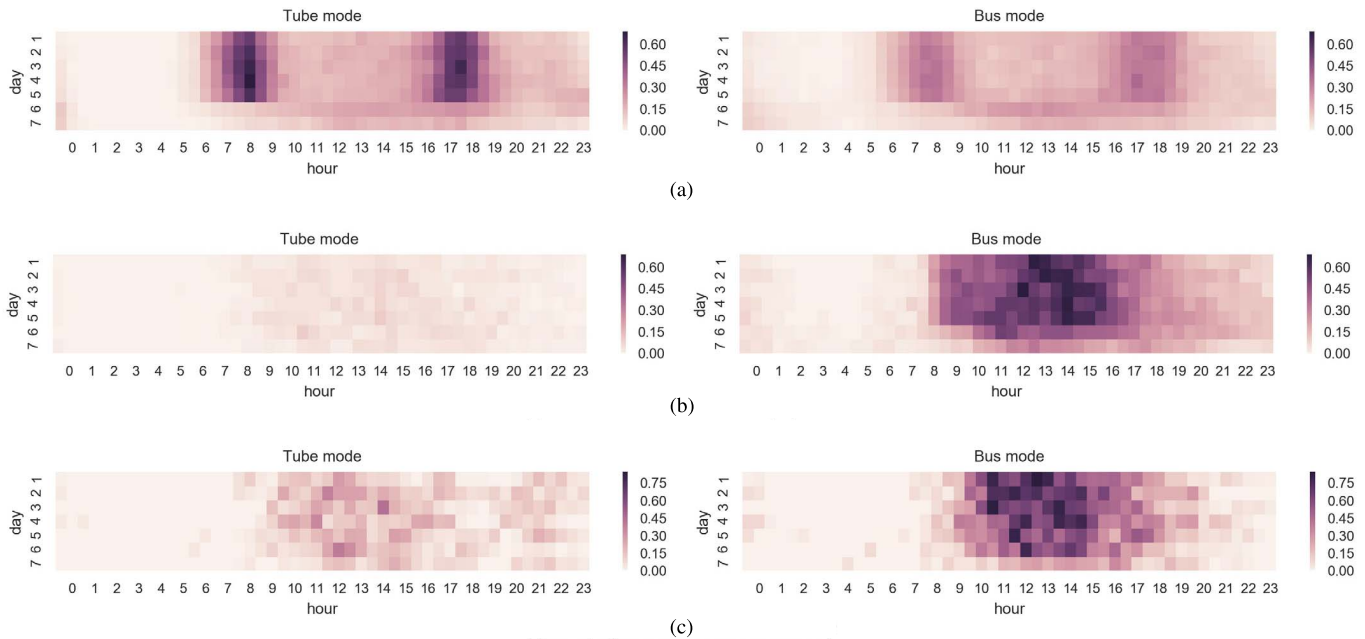


Fig. 7. Aggregated temporal profiles of misclassification examples. The heatmap shows the travel time distribution across the week. The colour of each grid indicates the percentage of travel time within in it and the two matrices of each category are normalized to sum to 100%. (a) Part-time employees are misclassified as the full-time. (b) Part-time employees are misclassified as the unemployment. (c) Unemployed people are misclassified as the disabled.

misclassified as disabled, as shown in Fig. 7 (b) and (c), respectively. The potential reasons for this phenomenon are threefold. First, besides the public bus and tube, many people also use other modes of transport. Although extremely occasional PT users have been excluded from the samples, the travel data obtained from the SCD cannot record an individual's entire travel activities. Second, SCD is incomplete. As abovementioned, the origin location and boarding information are missing from the SCD, and data about travel purposes are also unavailable. In some cases, the similar temporal patterns and mode choice are undistinguished for employment status inference. Third, travel behaviors of individuals are diverse and there are many other social-demographic factors (e.g. age, gender, and income) that have an impact. Therefore, it is understandable that people with distinct employment status may also have similar temporal patterns and travel mode choice.

VI. CONCLUSIONS AND FUTURE WORK

The relationship between individual travel behaviors and social-demographic roles was not entirely understood in previous works. In this paper, the possibility of using SCD to determine an individual's employment status was explored. We introduced a framework of employment status inference based on the travel temporal profiles from SCD. An approach to capture traveller's weekly temporal travel pattern and mode choices was proposed by representing the raw SCD as a 3D image. The similarity of individuals' temporal behavior can be captured in a better-refined way. The 3D temporal profiles were then fed into the proposed TMC-CNN model for employment status prediction. In TMC-CNN, different temporal behaviors within distinct travel modes can be learned.

And the class-imbalance problem was tackled via the ensemble of oversampling and thresholding techniques.

The study applies to a one-year London's Oyster Card dataset in combination with LTDS data. In the experiment, the effectiveness and robustness of the SCD representation method were first studied. Results showed that the prediction accuracy could be improved by about 20% after modifying the profiling method. The 30-minute or one-hour time slot was the preferable threshold for temporal behavior discretization. By comparison to the various standard models, the proposed model outperformed the baselines, and achieved 72.5%, 71.9% and 71.5% of *prec*, *rec* and *F1* in employment status prediction. In addition, we also analyzed how the training percentage impacted prediction accuracy. The prediction performance converged at 80%. Finally, a semantic analysis further revealed the temporal pattern and travel mode preference of different groups of passengers discovered by the model.

The results of this effort were quite interpretable and provided some new insights into the understanding of temporal travel patterns, mode choice and the relationships to the individuals' employment status. This suggests that the individual's travel behaviors are tightly bounded to employment status. For example, except for full-time employees, other groups of passengers averagely prefer to use bus rather than the tube and people without jobs travel more around midday. Overall, the evening usage peak of part-time employees and students was earlier than that of full-time employees. We believe these studies can be helpful for public transport agencies and operators to better understand who travels and when in the network and help to improve transport planning and personalized service.

However, there are some limitations to the current study. Future work can be conducted based on the work presented herein. First, in the SCD representation method, we only investigate the aggregated weekly temporal patterns. The temporal behavior may change from month to month, such as for the aforementioned part-time employees with the term-time working arrangements. Second, besides bus and tube, people may use other transportation methods for daily travel. In addition, the SCD is always incomplete. While boarding information is available for most AFC systems, alighting information is often missing. Consequently, we only consider the temporal pattern and mode choice but ignoring the spatial travel behaviors. Thus, data fusion of SCD and other data sources (e.g. GPS trajectories) could be leveraged to construct complete spatial-temporal behaviors of individuals for employment status prediction. Furthermore, only prediction of employment status was conducted in this study. The diverse travel patterns of individuals are dependent on a number of other social-demographic factors, such as age group, income level and car ownership. In the future, the approach should be extended to a multi-task demographic prediction model in order to include these aspects.

ACKNOWLEDGMENT

Transport for London is acknowledged for sharing the London's Oyster card data and LTDS data, which are used in this study. The authors declare no conflict of interest.

REFERENCES

- [1] X. Ma, Y.-J. Wu, Y. Wang, F. Chen, and J. Liu, "Mining smart card data for transit riders' travel patterns," *Transp. Res. C, Emerg. Technol.*, vol. 36, pp. 1–12, Nov. 2013.
- [2] G. G. Langlois, H. N. Koutsopoulos, and J. Zhao, "Inferring patterns in the multi-week activity sequences of public transport users," *Transp. Res. C, Emerg. Technol.*, vol. 64, pp. 1–16, Mar. 2016.
- [3] M. K. El Mahrsi, E. Côme, L. Oukhellou, and M. Verleysen, "Clustering smart card data for urban mobility analysis," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 3, pp. 712–728, Mar. 2017.
- [4] J. Zhao, Q. Qu, F. Zhang, C. Xu, and S. Liu, "Spatio-temporal analysis of passenger travel patterns in massive smart card data," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 11, pp. 3135–3146, Nov. 2017.
- [5] X. Lu and E. I. Pas, "Socio-demographics, activity participation and travel behavior," *Transp. Res. A, Policy Pract.*, vol. 33, pp. 1–18, Jan. 1999.
- [6] K. Xie, H. Xiong, and C. Li, "The correlation between human mobility and socio-demographic in megacity," in *Proc. IEEE Int. Smart Cities Conf. (ISC2)*, Sep. 2016, pp. 1–6.
- [7] M. Tovey, A. Woodcock, and J. Osmond, *Designing Mobility and Transport Services: Developing Traveller Exper. Tools*, New York, NY, USA: Taylor & Francis, 2016.
- [8] L. Zhu, J. Gonder, and L. Lin, "Prediction of individual social-demographic role based on travel behavior variability using long-term GPS data," *J. Adv. Transp.*, vol. 2017, Aug. 2017, Art. no. 7290248.
- [9] S. Bricka, M. Moran, K. Miller, and J. Hudson, "The future of TDM: Technology and demographic shifts and their implications for transportation demand management," Texas A&M Transp. Inst., Houston, TX, USA, Tech. Rep. PRC 15-25F, 2015.
- [10] *Drivers of Demand for Travel in London: A Review of Trends in Travel Demand and Their Causes*, Transp. London, London, U.K., 2014.
- [11] *Travel in London Report 9*, Transp. London, London, U.K., 2016.
- [12] S. P. Shepherd, X. Zhang, G. Emberger, M. Hudson, A. D. May, and N. Pailley, "Designing optimal urban transport strategies: The role of individual policy instruments and the impact of financial constraints," *Transport Policy*, vol. 13, pp. 49–65, Jan. 2006.
- [13] C. Yi, "Impact of public transit on employment status: Disaggregate analysis of Houston, Texas," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 1986, pp. 137–144, Jan. 2006.
- [14] A. Páez, M. Trépanier, and C. Morency, "Geodemographic analysis and the identification of potential business partnerships enabled by transit smart cards," *Transp. Res. A, Policy Pract.*, vol. 45, pp. 640–652, Aug. 2011.
- [15] T. Bantis and J. Haworth, "Who you are is how you travel: A framework for transportation mode detection using individual and environmental characteristics," *Transp. Res. C, Emerg. Technol.*, vol. 80, pp. 286–309, Jul. 2017.
- [16] P. Wang, F. Sun, D. Wang, J. Tao, X. Guan, and A. Bifet, "Inferring demographics and social networks of mobile device users on campus from AP-trajectories," in *Proc. 26th Int. Conf. World Wide Web Companion*, 2017, pp. 139–147.
- [17] Y. Zhong, N. J. Yuan, W. Zhong, F. Zhang, and X. Xie, "You are where you go: Inferring demographic attributes from location check-ins," in *Proc. 8th ACM Int. Conf. Web Search Data Mining*, 2015, pp. 295–304.
- [18] M. C. González, C. A. Hidalgo, and A.-L. Barabási, "Understanding individual human mobility patterns," *Nature*, vol. 453, no. 7196, pp. 779–782, 2008.
- [19] A. Siren and L. Hakamies-Blomqvist, "Private car as the grand equaliser? Demographic factors and mobility in Finnish men and women aged 65+," *Transp. Res. F, Traffic Psychol. Behav.*, vol. 7, pp. 107–118, Mar. 2004.
- [20] P. van den Berg, T. Arentze, and H. Timmermans, "A path analysis of social networks, telecommunication and social activity-travel patterns," *Transp. Res. C, Emerg. Technol.*, vol. 26, pp. 256–268, Jan. 2013.
- [21] Y. Zhang and T. Cheng, "Feature extraction for long-term travel pattern analysis," in *Proc. 25th GISRUK Conf.*, Manchester, U.K., 2017, pp. 39–42.
- [22] J. Scheiner and C. Holz-Rau, "Travel mode choice: Affected by objective or subjective determinants?" *Transportation*, vol. 34, pp. 487–511, Jul. 2007.
- [23] M. Giallonardo, "The impact of employment status on the travel behaviour of millennials," M.S. thesis, Dept. Urban Regional Planning, Queen's Univ., Kingston, ON, Canada, 2017.
- [24] Y. Zhang and T. Cheng, "Inferring social-demographics of travellers based on smart card data," in *Proc. 2nd Int. Conf. Adv. Res. Methods Anal.*, Valencia, Spain, 2018, pp. 55–62.
- [25] A. Lo, H. Chernoff, T. Zheng, and S.-H. Lo, "Why significant variables aren't automatically good predictors," *Proc. Nat. Acad. Sci.*, vol. 112, no. 45, pp. 13892–13897, 2015.
- [26] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.
- [27] S. Dabiri and K. Heaslip, "Inferring transportation modes from GPS trajectories using a convolutional neural network," *Transp. Res. C, Emerg. Technol.*, vol. 86, pp. 360–371, Jan. 2018.
- [28] D. Ribeiro, G. Carneiro, J. C. Nascimento, and A. Bernardino, "Multi-channel convolutional neural network ensemble for pedestrian detection," in *Proc. Iberian Conf. Pattern Recognit. Image Anal.*, 2017, pp. 122–130.
- [29] H. Shi, T. Ushio, M. Endo, K. Yamagami, and N. Horii, "A multichannel convolutional neural network for cross-language dialog state tracking," in *Proc. IEEE Spoken Lang. Technol. Workshop (SLT)*, Dec. 2016, pp. 559–564.
- [30] M. Buda, A. Maki, and M. A. Mazurowski. (2017). "A systematic study of the class imbalance problem in convolutional neural networks." [Online]. Available: <https://arxiv.org/abs/1710.05381>
- [31] F. Pukelsheim, "The three sigma rule," *Amer. Statist.*, vol. 48, no. 2, pp. 88–91, 1994.
- [32] J. Zhao, A. Rahbee, and N. H. M. Wilson, "Estimating a rail passenger trip origin-destination matrix using automatic data collection systems," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 22, pp. 376–387, Jul. 2007.
- [33] C. Seaborn, J. Attanucci, and N. H. M. Wilson, "Using smart card fare payment data to analyze multi-modal public transport journeys in London," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2121, pp. 55–62, 2009.
- [34] *Transport Statistics Great Britain 2013*, Department for Transport, London, U.K., 2013.
- [35] M. K. El Mahrsi, E. Côme, J. Baro, and L. Oukhellou, "Understanding passenger patterns in public transit through smart card and socioeconomic data: A case study in Rennes, France," in *Proc. 3rd Int. Workshop Urban Comput. (UrbComp)*, New York, NY, USA, Aug. 2014, pp. 1–9.
- [36] J. Zhao, C. Tian, F. Zhang, C. Xu, and S. Feng, "Understanding temporal and spatial travel patterns of individual passengers by mining smart card data," in *Proc. 17th Int. IEEE Conf. Intell. Transp. Syst. (ITSC)*, Oct. 2014, pp. 2991–2997.

- [37] R. S. Sokasane and K. Kyungbaek, "Predicting user attitude by using GPS location clustering," *IEICE Trans. Inf. Syst.*, vol. 98, pp. 1600–1603, Aug. 2015.
- [38] A. Estabrooks and N. Japkowicz, "A mixture-of-experts framework for learning from imbalanced data sets," in *Proc. Int. Symp. Intell. Data Anal.*, 2001, pp. 34–43.
- [39] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.
- [40] M. D. Richard and R. P. Lippmann, "Neural network classifiers estimate Bayesian a posteriori probabilities," *Neural Comput.*, vol. 3, pp. 461–483, Dec. 1991.
- [41] D. P. Kingma and J. Ba. (2014). "Adam: A method for stochastic optimization." [Online]. Available: <https://arxiv.org/abs/1412.6980>
- [42] S. Seneviratne, A. Seneviratne, P. Mohapatra, and A. Mahanti, "Predicting user traits from a snapshot of apps installed on a smartphone," *ACM SIGMOBILE Mobile Comput. Commun. Rev.*, vol. 18, no. 2, pp. 1–8, 2014.
- [43] S. Zhao *et al.*, "Mining user attributes using large-scale APP lists of smartphones," *IEEE Syst. J.*, vol. 11, no. 1, pp. 315–323, Mar. 2017.
- [44] Z. Qin, Y. Wang, H. Cheng, Y. Zhou, Z. Sheng, and V. C. M. Leung, "Demographic information prediction: A portrait of smartphone application users," *IEEE Trans. Emerg. Topics Comput.*, vol. 6, no. 3, pp. 432–444, Jul./Sep. 2017.



Yang Zhang received the B.Eng. degree in system engineering and the master's degree in management science and engineering from the National University of Defense Technology, China, in 2010 and 2016, respectively. She is currently pursuing the Ph.D. degree with SpaceTimeLab for Big Data Analytics, University College London, U.K. Her research interests include spatial-temporal data mining, urban computing, deep learning, data fusion techniques, public transportation systems, and artificial intelligence.



Tao Cheng is currently a Professor in geoinformatics and the Director of SpaceTimeLab for Big Data Analytics, University College London. Her research interests span network complexity, geocomputation, integrated spatial-temporal analytics, and big data mining (modeling, prediction, clustering, and simulation), with applications in transport, crime, health, social media, and environmental monitoring.