Enriching Large-Scale Trips With Fine-Grained Travel Purposes: A Semi-Supervised Deep Graph Embedding Framework

Chengwu Liao, Chao Chen, *Senior Member, IEEE*, Suiming Guo, Leye Wang, *Member, IEEE*, Fuqiang Gu, *Member, IEEE*, Jing Luo, and Ke Xu, *Senior Member, IEEE*

***Abstract*— Knowing why people travel is meaningful for human mobility understanding and smart services development. Unfortunately, in real-world scenarios, trip purpose cannot be automatically collected on a large scale, thus calling for effective prediction models. Nevertheless, since passengers' trip purposes in the city are diverse and complicated, the prediction is very difficult especially at a fine-grained level. Worse still, the informative data sources and real purpose-labels about trips are commonly limited for model learning. To resolve the dilemma, we propose a semi-supervised deep embedding framework for predicting fine-grained trip purposes on a large scale. Specifically, we first derive augmented trip contexts from the vehicle's GPS trajectory and public POI check-in data, then convert POI contexts into the graph structure. We further establish a Dual-Attention Graph Embedding Network with Autoencoder architecture (*DAGE-A*) to accomplish prediction and reconstruction simultaneously, in which category-aware graph attention networks are devised to model the POI semantics at trip's origin/destination and extract complementary knowledge from unlabeled trips; and soft-attention is employed to aggregate different trip semantics appropriately for the final prediction. We conduct extensive experiments in Beijing and Shanghai, and results show our framework outperforms state-of-the-arts and could reduce labelling efforts by up to 20%. We also find that our model is generalized at different times and locations, and the performance varies for different trip purposes.**

***Index Terms*— Trip purpose, semi-supervised learning, graph embedding, GPS trajectory, check-in data.**

Manuscript received 27 May 2022; revised 11 August 2022; accepted 29 August 2022. This work was supported in part by the National Natural Science Foundation of China under Grant 62172066, Grant 61872050, Grant 62002135, Grant 61972008, and Grant 42174050; in part by the National Science Foundation for Distinguished Young Scholars of China under Grant 61825204; in part by the Beijing Outstanding Young Scientist Program under Grant BJJWZYJH01201910003011; in part by the Science and Technology Projects in Guangzhou under Grant 202102021164; and in part by the Chongqing Innovative and Entrepreneurial Program of Returned Overseas Chinese Scholars under Grant cx2021047. The Associate Editor for this article was H. Lv. *(Chengwu Liao and Chao Chen are co-first authors.) (Corresponding authors: Chao Chen; Suiming Guo.)*

Chengwu Liao, Chao Chen, and Fuqiang Gu are with the Key Laboratory of Dependable Service Computing in Cyber Physical Society, Ministry of Education, and the College of Computer Science, Chongqing University, Chongqing 400044, China (e-mail: lcw@cqu.edu.cn; cschaochen@cqu.edu.cn; gufq@cqu.edu.cn).

Suiming Guo is with the College of Information Science and Technology, Jinan University, Guangzhou 510632, China (e-mail: guosuiming@email.jnu.edu.cn).

Leye Wang is with the School of Electronics Engineering and Computer Science, Peking University, Beijing 100871, China (e-mail: leyewang@pku.edu.cn).

Jing Luo is with the CRSC Research Laboratory for Smart City, China Railway Signal & Communication Company Ltd. (CRSC), Beijing 100070, China (e-mail: luojing@crsc.cn).

Ke Xu is with the Department of Computer Science, Tsinghua University, Beijing 100084, China (e-mail: xuke@tsinghua.edu.cn).

Digital Object Identifier 10.1109/TITS.2022.3203464

I. Introduction

TRIP purpose is the semantic information answering *why* people travel in the city, and has been identified as an important yet under-explored aspect in travel behavior analysis [14], [36]. Knowing trip purposes in cities is meaningful for human mobility understanding and smart services development [7], [21], [24]. For example, knowing the passenger's trip purpose could enable personalized in-car advertising/recommendation, and knowing the city-wide trip purposes could help public transportation planning (e.g., new bus routes for "Working" purposes). Unfortunately, unlike GPS trajectories, trip purposes cannot be automatically collected by sensing devices in real-world scenarios, thus calling for effective prediction models.

In general, the accurate prediction of trip purpose is very challenging due to the complexity of human mobility nature [2], [16], [22]. On one hand, trip purposes refer to the activities that passengers take after being dropped off, while human activities are *realistically diverse and varying among people*. Thus passengers' trip purposes are essentially with great uncertainty. On the other hand, for passengers in the city, their trip purposes (activities) are *implicitly influenced by many factors* (e.g., time, space, personality, and nearby land-use configuration) [23], so that the prediction would be very complicated.

In recent years, with the proliferation of Information and Communication Technologies (ICTs), Internet of Things (IoT) in daily life, many aspects of human behaviors are able to be recorded in the cyber space [26], [32]. Hence, one promising solution for the trip purpose prediction is to understand the passengers' activity semantics with multi-sourced information [21]. In this regard, a few works in the literature have achieved over 90% trip purpose prediction accuracy [10], [38]. However, many of them are in a personalized manner by employing some sensitive information (e.g., employment, activity duration) to model the respondent's preference/life-circles for prediction. As a matter of fact, with the prevalence of data protection laws and regulations (e.g., GDPR [35]), people have the rights to delete and not reveal their personal information to a service provider, thus whether this kind of personalized methods can still be practically applied becomes doubtful. To make matters worse, such methods mainly work on a small scale of trips contributed by a fraction of respondents, and return a coarse granularity of trip purpose. Consequently, the generalizability to support large-scale and diversified urban services in real-world scenarios is restricted.

To enable more pervasive and privacy-friendly services, in this paper, we aim at offering a *context-aware* prediction approach that is able to automatically predict the fine-grained trip purposes (i.e., up to 9 kinds of trip purposes in total) for the large-scale trips contributed by city-wide passengers. Note that the target application scenarios are door-to-door ride services like taxi trips. Nevertheless, enabling trip purpose prediction in such scenarios still faces two critical challenges: i) there are very limited useful information to depict the activity semantics for the accurate prediction; ii) labelling efforts (e.g., surveys) are very high-cost and with uncontrollable quality, so that in most cases, there are just quite limited labeled trips available for algorithms to learn how to predict.

In light of the aforementioned challenges, the contributions of our work can be summarized as follows.

- We propose a novel semi-supervised deep embedding framework (*DAGE-A*) for predicting fine-grained trip purposes in a context-aware manner. It leverages pervasive data to enable the large-scale yet fine-grained prediction, and is also effective with limited labeled training data, thus making it more applicable for real-world urban services.
- We integrate vehicle's GPS trajectory (for revealing the trip's time and space) and public POI check-in data (for characterizing human activities) to derive trip contexts with semantic meanings. We are among pioneers of using graph structure to represent localized POI contexts.
- We establish a dual-attention graph embedding network (i.e., category-aware graph attention networks and soft-attention) with autoencoder architecture, to model the higher-level activity semantics from trip contexts. It accomplishes the prediction and reconstruction simultaneously, so as to improve the performance by incorporating the complementary knowledge from unlabeled trips.
- We conduct a group of experiments with large-scale datasets in Beijing and Shanghai. The results show that our approach achieves a considerable improvement compared with the state-of-the-art baselines, and our semi-supervised framework could reduce labelling efforts by up to 20%. Moreover, we investigate the performance of trip purpose prediction at different times and locations.

The rest of this paper is organized as follows. In Section II, we introduce a few definitions and the problem we resolved in this paper. In Section III, we elaborate on the methodology of our proposed trip purpose prediction framework. After that, we present a group of experiments in Section IV. In Section V, we briefly introduce the related works. Finally, we conclude our work and outlook the future research directions in Section VI.

II. Preliminaries

A. Definitions

Definition 1 (Trip): Each trip is represented by its origin-destination pair, which consists of two GPS points with timestamps collected by the vehicle at the pick-up and drop-off locations, i.e., $tr = [(l_o, t_o), (l_d, t_d)]$*. Such information is used to reveal the original time and space of the passenger's trip.*

Definition 2 (Point of Interest): A POI refers to a place that is the very basic unit of taking human activities. POIs are usually represented by their positions and POI category information. Note that the POI category directly indicates the type of potential human activities at these POIs.

Definition 3 (Check-in Data): The Check-in data CI *is generated when users checked-in at POIs using LBSN platforms. A check-in record commonly contains information about the user's identity, the check-in time and the corresponding POI venue. Generally, the number of check-ins could reveal the popularity of a POI [5], [39].*

In this paper, the POIs and the corresponding check-in data are used to reveal the characteristics of human activities at passenger's origin and destination locations. The adopted 9 POI categories and check-in data are from a Chinese LBSN called *Jiepang* [20]. The mapping of POI categories and the corresponding human activities (i.e., trip purpose) can be found in Tab. I.

B. Problem Statement

With the aforementioned definitions, the problem of predicting trip purpose with limited labels can be viewed as a classification problem in semi-supervised learning.

Given:

1) Two sets of trips in a city (both labeled and unlabeled trips), i.e., TR_l and TR_u.
2) A set of POIs and the corresponding historical check-in records CI in the designated city.
3) A set of candidate trip purposes, i.e., $\bar{A}$ in Tab. I.

Train a semi-supervised prediction model with labeled and unlabeled trip data simultaneously.

Predict the probabilities of candidate trip purposes for an unseen trip tr by using the trained model, i.e., $\hat{p}\left(y = \bar{a}|tr, CI\right), \bar{a} \in \bar{A}$.

III. Methodology

A. Overview

Figure 1 shows our context-aware semi-supervised trip purpose prediction framework. The inputs consist of labeled trips, unlabeled trips, and public POI check-ins. The first stage is *trip context augmentation*, in which the public check-ins are aggregated with labeled and unlabeled trips respectively, to derive augmented trip contexts, including the OD (origin and destination) POI contexts and spatiotemporal context.

TABLE I

POI CATEGORY AND THE CORRESPONDING TRIP PURPOSE

k	POI Category	Trip Purpose
1	Recreation and Culture Facilities	Recreation
2	Outdoors and Sightseeing Places	Outdoors
3	Shop and Service Facilities	Shopping
4	Restaurant	Dining
5	School and Educational Facilities	Education
6	Transportation Facilities	Transportation
7	Apartment and Residence	Homing
8	Hospital and Clinic	Health
9	Office and Business Buildings	Working

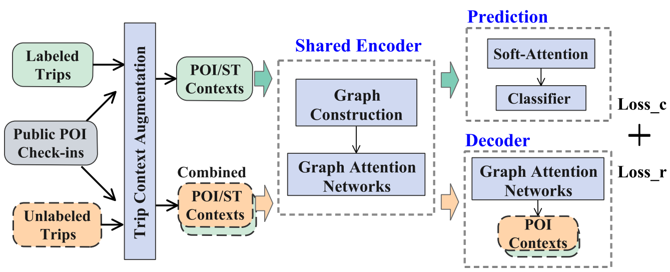


Fig. 1. Our semi-supervised trip purpose prediction framework.

The second stage is the *dual-attention neural network* with autoencoder architecture (in the three dotted boxes). Generally, there are two parallel tasks in the working flow, namely *supervised prediction with labeled trips*, and *unsupervised reconstruction with combined trips*. Note that the combined trips include all the labeled and unlabeled trips, so as to provide relatively completed data distribution of all possible trips in model training. The encoder is a shared component, in which the augmented OD POI contexts are first converted into the graph structure, then graph attention networks are used to extract the latent OD activity semantics. After that, the labeled trips are sent to the prediction component, in which a soft-attention mechanism and a classifier are used to aggregate the activity semantics and obtain the prediction loss (i.e., $Loss\_C$). Meanwhile, the combined trips are sent to the decoder component, in which a reversed encoder network is used to obtain the reconstruction loss (i.e., $Loss\_R$). Finally, these two kinds of losses are combined to train the semi-supervised neural network.

B. Trip Context Augmentation

In general, *when* and *where* one taking a trip are two foremost significant clues for the trip purpose prediction. In this component, we aggregate the trip and public POI check-in data to augment the semantic meaning of trip's spatiotemporal (ST) context, and OD POI contexts.

1) Spatiotemporal Context C_{st}: For a trip tr, we extract three kinds of temporal contexts from the vehicle's GPS trajectory, including the type of day (i.e., workday or non-workday) $TYP\,(tr)$ and the hour time $H\,(t)$ when this trip started and ended, and the travel time t_d-t_o. Particularly, in order to maintain the time similarity between 00:00 and 23:00, t is first converted to the radian of a unit circle, i.e., θ=2π $(t/24)$, then represented by $H\,(t)$=$(\cos\theta, \sin\theta)$. Together with the travel time, the spherical distance between the origin and destination l_d-l_o is used as the spatiotemporal cost of this trip. Finally, the spatiotemporal context C_{st} can be obtained as Eq. 1.

$$C_{st}\,(tr) = [TYP\,(tr)\,, H\,(t_o)\,, H\,(t_d)\,, t_d - t_o, l_d - l_o] \quad (1)$$

2) POI Contexts C_{poi}: To depict the activity conditions at the origin/destination location, for each POI category, we extract *static* and *dynamic* features from the nearby POI check-in data within a radius of r meters. In this study, r is set to 250 meters according to the studies of land-use buffer for human trips [6]. The static features refer to POI distributions, i.e., *distance* and *uniqueness*. Specifically, the *distance* feature is the ratio of the minimum distance among the k-th category of POIs at the drop-off point l_d:

$$Dist(k) = -\log_2\left(\frac{\min(distance(POIs^k, l_d))}{r}\right) \quad (2)$$

In addition, the *uniqueness* is adopted to reveal the ratio of the k-th category of POIs:

$$Uniq(k) = -\log_2\left(\frac{|POIs^k|}{\sum_{i\in K}|POIs^i|}\right) \quad (3)$$

In terms of dynamic features, we extract *period popularity* $Popu$ to reveal the time-variant attractiveness of different POIs. Specifically, for the k-th POI category, we compute the total check-in times from the check-in data CI during a given time period $\mathcal{T}$, i.e. $|CI|_k^{\mathcal{T}}$. Then, $Popu(k)$ is formulated as:

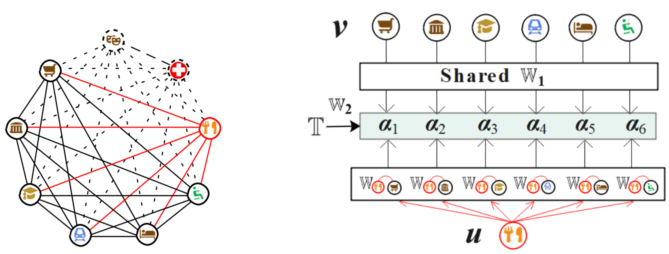
$$Popu(k) = -\log_2\left[1 - \left(\frac{|CI|_k^{\mathcal{T}}}{\sum_{i\in K}|CI|_i^{\mathcal{T}}}\right)\right] \quad (4)$$

where K denotes the number of all POI categories (i.e., $K = 9$). In particular, for the origin location, we set $\mathcal{T}$ to $[t_o - 2, t_o]$, i.e., $2h$ before the trip starts. For the destination location, we set $\mathcal{T}$ to $[t_d, t_d + 2]$, i.e., $2h$ after the trip ends.

C. Semi-Supervised Graph Embedding Model

In this section, we elaborate on the proposed semi-supervised neural network for trip purpose prediction with augmented trip contexts. Generally, the network models the passenger's activity semantics in a graph embedding manner, and uses an autoencoder framework to ingest unlabeled data in the model training. Such a semi-supervised way could help the model to abstract latent representations that capture the semantic meaning of all available samples, thus improving the model's performance.

1) Autoencoder With Graph Embedding: When extracting activity semantics from POI contexts, it's important to model the *inherent correlations between different POI categories*, since human activities at a location are often associated with each other, e.g., "Dinning" and "Shopping". Generally, the POI categories, POI features, and correlations between POI categories in the POI context are naturally analogous to the nodes, nodes' features, and edges in a graph. Moreover, the graph structure could maintain the concept of node in hidden layers, so that the modelling of node correlations is more intuitive. In this sense, we first convert OD POI contexts into the graph structure, then employ the graph attention network



(a) POI Graph (b) Category-aware GAT

Fig. 2. Illustration of our graph construction and the attention coefficients computation in the category-aware GAT, i.e., Equation 5.

to extract the neighboring activity semantics for each POI category.

Graph construction. As shown in Fig. 2 (a), an undirected completed POI graph is defined as $G = (V, E)$ to represent the POI context at the origin/destination. V is a set of nodes representing the existent POI categories (drawn with solid circles) in the origin/destination POI context, and E is a set of edges representing their inherent correlations. Note that the two nodes drawn with dotted lines in the figure refer to the nonexistent POI categories. In addition, each node contains 3 kinds of augmented features $h \in \mathbb{R}^F$ (i.e., *distance*, *uniqueness*, and *period popularity*). Consequently, the OD POI contexts are represented by G_o and G_d, respectively.

Encoder for POI semantics extraction. In this study, G_o and G_d representing POI contexts are arbitrarily structured graphs in reality (some nodes may not exist), since a location cannot always have all 9 categories of POIs nearby. In this regard, we adopt the *graph attention network (GAT)* [34] to model the non-identical correlations of neighboring POI categories, and accordingly extract the high-level POI semantics. The normal GAT adopts the attention mechanism to learn attention coefficients between a central node u and its neighboring nodes N_u. Note that for a central node, the attention coefficients to different neighbors are computed with same parameters. However, human activities usually have different inherent correlations. For example, for the "Dining" activity, its correlation with "Recreation" is stronger than "Health", since "Dining" and "Recreation" are more likely to be associated in people's daily life.

To narrow the gap, we propose a new category-aware GAT to further consider the neighboring nodes' inherent differences, i.e., differentiate the inherent correlations between specific nodes in the POI graph. In such a manner, the extracted neighboring semantics is also category-aware in the latent space. Additionally, we also consider the time features $\mathbb{T}$ (i.e., day type and hour time feature) in the computation, since the correlations between different activities also demonstrate to be time-dependent. Hence, the coefficient α_{uv} between u and a neighbor $v \in N_u$ can be obtained by:

$$\alpha_{uv} = \frac{\exp\left(g\left(\mathbb{W}_{\mathbf{uv}}{}^{T}\mathbf{W}h_u + \mathbb{W}_{\mathbf{1}}{}^{T}\mathbf{W}h_v + \mathbb{W}_{\mathbf{2}}{}^{T}\mathbb{T}\right)\right)}{\sum_{n\in N_u}\exp\left(g\left(\mathbb{W}_{\mathbf{un}}{}^{T}\mathbf{W}h_u + \mathbb{W}_{\mathbf{1}}{}^{T}\mathbf{W}h_n + \mathbb{W}_{\mathbf{2}}{}^{T}\mathbb{T}\right)\right)} \tag{5}$$

where $\mathbf{W} \in \mathbb{R}^{F' \times F}$ is a shared weight matrix for linearly transforming the input node features h into the latent space. $\mathbb{W}_{\mathbf{uv}} \in \mathbb{R}^{F'}$ *is a unique weight matrix of the center node u towards a specific neighbor v*. Different from that, $\mathbb{W}_{\mathbf{1}} \in \mathbb{R}^{F'}$ and $\mathbb{W}_{\mathbf{2}} \in \mathbb{R}^{|TYP|+|H|}$ are *shared* attention weight matrices for different neighbors and time features. Figure 2 (b) illustrates an example of our category-aware attention coefficients computation for the "Dining" POIs.

To obtain the neighboring POI features of u, the attention coefficients are used to combine the neighbors' features in a weighted sum manner. We also adopt the multi-head mechanism to extract node's neighboring features from multiple perspectives, then the multi-head features are concatenated and transformed into the final neighboring feature $\vec{h}_u$:

$$\vec{h}_u = \mathbf{W}'\left(\Big\Vert_{m=1}^{M}\sigma\left(\sum_{v\in N_u}\alpha_{uv}^{m}\mathbf{W}^{m}h_v\right)\right) \tag{6}$$

where M is the number of multi-head attentions, and α_{uv}^m and W^m are the attention coefficient and linear transformation weight matrix of the m-th attention. σ is a nonlinear function. $\mathbf{W}' \in \mathbb{R}^{F' \times MF'}$ is a weight matrix which transforms the concatenated features into F' dimension. Besides, in this graph embedding network, we stack two multi-head category-aware GATs to enhance the learning capability. As a result, we can obtain the higher-level OD POI contexts with neighboring activity semantics (i.e., $\vec{G_o}$, $\vec{G_d}$).

Besides, for the prediction part, we further aggregate each node's own augmented features and neighboring features as the twofold comprehensive POI semantics (i.e., $h'_u = [h_u \Vert \vec{h}_u]$, $h'_u \in \mathbb{R}^{F+F'}$), obtaining G'_o and G'_d.

Decoder for reconstruction. After the shared encoder, the features of combined trips (i.e., $\vec{G_o}$ and $\vec{G_d}$) are sent to the decoder component. The decoder is used to perform the inverse operations of the aforementioned encoder, to reconstruct the original features in POI contexts. Since the encoder mainly employs GATs to extract the neighboring features of POI contexts, we adopt a new group of GATs (i.e., Equations 5, 6) with inverse feature dimensions in the decoder. The loss function for reconstruction is the *mean squared error*:

$$L\_r = \sum_i (x_i^o - \hat{x}_i^o)^2 + \sum_j (x_j^d - \hat{x}_j^d)^2 \tag{7}$$

where x_i^o and x_i^d are the elements of G_o and G_d, respectively. $\hat{x}_i^o$ and $\hat{x}_i^d$ are the corresponding reconstructed vectors.

The reconstruction loss would be used to train the encoder together with prediction loss. In this manner, the activity semantics extraction in encoder would be trained with both the labeled and unlabeled trips (i.e., large-scale trip data), thus being more generalized to unseen trips in the city.

2) Prediction: In this section, we first aggregate the semantics of three kinds of trip contexts, then use a classifier to predict the possibilities of candidate activities being the passenger's trip purpose.

Soft-attention for trip semantics aggregation. Soft-attention can be described as mapping a *query* and a set of *key-value* pairs to an output [33]. The output is the weighted sum of *values*, where the weights are computed by using a compatibility function on the *query* and a specific *key*. In this study, passenger's activity at the destination location can be viewed as the response to a special query (i.e., a trip with specific origin and time). Hence, the origin activity semantics G'_o and trip's spatiotemporal cost C_{st} are viewed as the *query*. Since POIs are basic units of human activities and the destination is where a passenger takes the final activity, each category of POI semantics in the destination is used as the *key* and *value*, i.e., $keys = values = h'_{(D)}$. By modelling such dependencies on trip purpose, the aggregation of trip contexts could be more rational for trip purpose prediction.

We first combine G'_o and C_{st} with a full connected (FC) layer to serve as the *query* h_{ost}. Then, we establish a multi-head soft-attention with a feed-forward network as the compatibility function. Then, the coefficient $\widehat{\alpha}_u$ for a POI category $u \in G'_d$ and the final combined trip activity semantics $\mathscr{H}$ can be computed as follows:

$$\widehat{\alpha}_u = \frac{\exp\left(\tanh\left({\mathbb{W}_q}^T h_{ost} + {\mathbb{W}_k}^T h'_u + b\right)\right)}{\sum_{s\in V}\exp\left(\tanh\left({\mathbb{W}_q}^T h_{ost} + {\mathbb{W}_k}^T h'_s + b\right)\right)}$$
$$\mathscr{H} = \mathbf{W}''\left(\Big\|_{m'=1}^{M'} \sigma\left(\sum_{u\in V} \widehat{\alpha}_u^{m'} h'_u\right)\right) \tag{8}$$

where $\mathbb{W}_q$, $\mathbb{W}_k$ and b are parameters of the compatibility function. M' denotes the number of attention heads, and $\mathbf{W}'' \in \mathbb{R}^{|h'_u|\times M'|h'_u|}$ is a parameter matrix that transforms the concatenated multi-head features into $|h'_u|$ dimensions.

Softmax classifier. A FC layer with *softmax* function is adopted as the classifier to output the probabilities of candidates. The FC layer contains $|\bar{A}|$ neurons representing the candidate trip purposes $\bar{A}$. Then, the probability $\hat{p}$ of the i-th candidate $\bar{a}_i$ being the purpose of a trip tr, can be obtained as follows:

$$\hat{p}\left(y = \bar{a}_i | tr, CI\right) = \frac{exp\left(z_i\right)}{\sum_{j=1}^{|\bar{A}|} exp\left(z_j\right)}, (z_i, z_j) \in FC(\mathscr{H})$$
$$\hat{y} = \arg\max_i \hat{p}\left(y = \bar{a}_i | tr, CI\right) \tag{9}$$

The prediction result is the candidate $\hat{y}$ with the highest probability. The prediction loss function is based on the cross-entropy:

$$L\_c = -\sum_{i=1}^{|\bar{A}|} y^{(i)} \log\left(\hat{p}^{(i)}\right) \tag{10}$$

where $y^{(i)}$ and $\hat{p}^{(i)}$ denotes to the actual and predicted probability of the i-th candidate.

At last, the overall loss function of our semi-supervised model is the weighted summarization of prediction and reconstruction, namely $\mathbb{L} = L\_c + \lambda \cdot L\_r$.

IV. Experiments

A. Dataset Description

We conduct a group of experiments in Beijing and Shanghai, based on two kinds of real-world datasets, namely UCar trip data and POI check-in data. Note that Shenzhou UCar (a ride-on-demand service) is one of the door-to-door ride services and further possesses the information of passengers' trip purposes. Hence, we employ the large-scale UCar data as the labeled door-to-door ride trips to evaluate our trip purpose prediction framework.

UCar Trip Data. The data is composed of vehicle trips generated by arbitrary passengers with Shenzhou UCar in China, in November and December 2015. Each record contains the GPS information of the pick-up&drop-off locations (i.e., longitude, latitude, and timestamp), and the name of passenger's target POI for this ride (e.g., *Beijing Restaurant*). The activity type of this POI (e.g., "Dining") is served as the trip purpose (i.e., ground truth). Such a mapping process is automatically accomplished by using a pre-trained NLP model (i.e., ERNIE). The model is fine-tuned with large-scale POI descriptions and the corresponding POI categories from the Jiepang dataset, and could achieve over 99% prediction accuracy. More details of the mapping process can be found in our previous study [21]. Finally, we select $366,783$ purpose-labeled trips within the Five-Ring of Beijing, and $270,943$ purpose-labeled trips within the central area of Shanghai.

Jiepang POI Check-in Data. It contains $511,133$ *Jiepang* check-ins in Beijing and $712,305$ check-ins in Shanghai from August 2011 to September 2012. Each record contains an anonymous user ID, a check-in timestamp and the corresponding POI information (i.e., POI description and POI category). Note that the POIs and human activities are relatively stable in developed cities [37], [41], so that the time inconsistency problem has less impact on our study.

Besides, both the datasets used in our experiments are anonymized. During the prediction in real-life scenarios (e.g., taxi trips), our model would merely employ the *vehicle's GPS trajectory* from the trip, and it has no connection with the passenger in the digital space, thus it does not record or involve any personal information. In short, the use of data in this study is privacy-friendly. Moreover, the utilized data sources are relatively pervasive in door-to-door ride scenarios. Thus, we believe our framework could be generalized to unseen people and similar ride services.

B. Baselines and Evaluation Metrics

1) Baselines: We compare our prediction framework with various baseline models in existing trip purpose prediction works. Note that in order to compare the models' performance in the same pervasive scenario as ours, all baselines are omitting the privacy-involved information from their original studies.

- *Nearest* [3]: Trip purpose is the activity type of a POI that is closest to the passenger's drop-off location.
- *Bayes's Rule* [13]: Based on a set of spatial and temporal rules, trip purpose is the activity type of the most likely to be visited POI near the destination.
- *Artificial Neural Network (ANN)* [38]: A neural network with two hidden layers, and the prediction is based on the day type and the land-use of trip's end (binary codes of nearby POI categories).
- *Random Forest (RF)* [11]: The input variables include the nearby place characteristics (i.e., proportions of different

TABLE II
PREDICTION RESULTS OF DIFFERENT MODELS IN BEIJING

Models	Proportion of labeled trips in the training data																	
	10%		20%		30%		40%		50%		60%		70%		80%		90%	
	Acc	M_F_1	Acc	M_F_1	Acc	M_F_1	Acc	M_F_1	Acc	M_F_1	Acc	M_F_1	Acc	M_F_1	Acc	M_F_1	Acc	M_F_1
Nearest	26.08	24.76	26.08	24.76	26.08	24.76	26.08	24.76	26.08	24.76	26.08	24.76	26.08	24.76	26.08	24.76	26.08	24.76
Bayes' rule	35.04	33.03	35.04	33.03	35.04	33.03	35.04	33.03	35.04	33.03	35.04	33.03	35.04	33.03	35.04	33.03	35.04	33.03
ANN	37.05	23.22	37.20	23.42	38.26	24.43	38.40	24.88	38.63	25.71	38.93	26.63	39.21	26.91	39.27	27.18	39.49	27.60
RF	49.62	44.82	53.09	47.56	55.13	50.82	57.03	52.35	57.54	53.60	58.30	54.35	58.77	54.84	59.26	55.42	60.01	56.62
DAGE	52.51	47.76	55.83	51.17	59.99	55.04	60.29	55.63	61.15	57.17	61.85	58.07	63.21	58.72	63.29	58.99	63.49	59.68
DAGE-P	53.35	48.51	57.18	52.12	60.34	55.74	60.95	56.58	61.61	57.40	62.21	58.31	63.22	58.86	63.34	58.92	64.26	59.97
DAGE-A	**54.86**	**49.52**	**58.25**	**53.90**	**60.84**	**56.09**	**61.43**	**56.76**	**62.51**	**58.43**	**62.70**	**58.52**	**63.78**	**59.86**	**64.07**	**60.21**	**64.51**	**60.59**

POI categories) and time features (i.e., day type and time period of a day).

In addition, we also establish two variations of our *DAGE-A* to evaluate the effectiveness of our semi-supervised framework in trip purpose prediction and labelling reducing.

- *Dual-Attention Graph Embedding (DAGE)*: The supervised version of our dual-attention graph embedding network, i.e., without the unsupervised reconstruction task.
- *DAGE with Pseudo Label (DAGE-P)*: Based on the *DAGE*, the Pseudo-Labels of unlabeled data are used to compute complementary losses for the model training. Details about the Pseudo-Labels techniques can be found in [18].

Only the *DAGE-A* and *DAGE-P* models are trained in a semi-supervised manner with both the labeled and unlabeled data.

2) Metrics: We adopt *Accuracy* and *macro F_1-score* to evaluate the cross-type overall prediction performance, and adopt *F_1-score* to evaluate the category-specific performance on each kind of trip purpose. As shown in Eq. 11, F_1-score is the harmonic mean of precision and recall for the i-th class, and *macro F_1-score* is the arithmetic mean of class-wise F_1-score to evaluate the overall performance.

$$F_1\text{-}score_i = \frac{2 * Precision_i * Recall_i}{Precision_i + Recall_i},$$
$$M\_F_1\text{-}score = \frac{\sum_i^N F_1\text{-}score_i}{N} \tag{11}$$

C. Evaluation Environment and Settings

We implement *DAGE-A* using Python 3.7 with *TensorFlow-2.5*, on a PC with 4 NVIDIA GeForce RTX 2080 Ti GPU and 192 GB RAM. The hyperparameters of models are selected by comparing the performance of different groups of settings. Specifically, we employ *Adam* to optimize the loss function with a learning rate l_r of 0.0001. The batch size and $L2$ regularizer parameter are set to 128 and 0.0001, respectively. Additionally, F' in *GATs* is set to 50, and the number of heads for *GATs* and *soft-attention* (M, M') is set to (20, 30), respectively. The settings of the encoder and decoder are the same. The dimension of fused origin POI context and spatiotemporal context is set to 50. Besides, through a sufficient number of tests, the loss combination λ is set to 1.

We divide the trip data into the training, validation and test datasets at a ratio of 6 : 1 : 1. In particular, we assume the training data is all of the available trips in reality, i.e., the combination of labeled data and unlabeled data. To evaluate the performance of different models with limited training labels, we conduct a group of experiments in which models are sequentially trained with increasing labeled samples (from 10% to 90%) from the training data. *Such a manner can be viewed as a simulation of increasing labelling efforts. Accordingly, for supervised models, the training is only based on the labeled samples, while for the semi-supervised models, the training is based on all the samples.*

D. Effectiveness of Our Framework

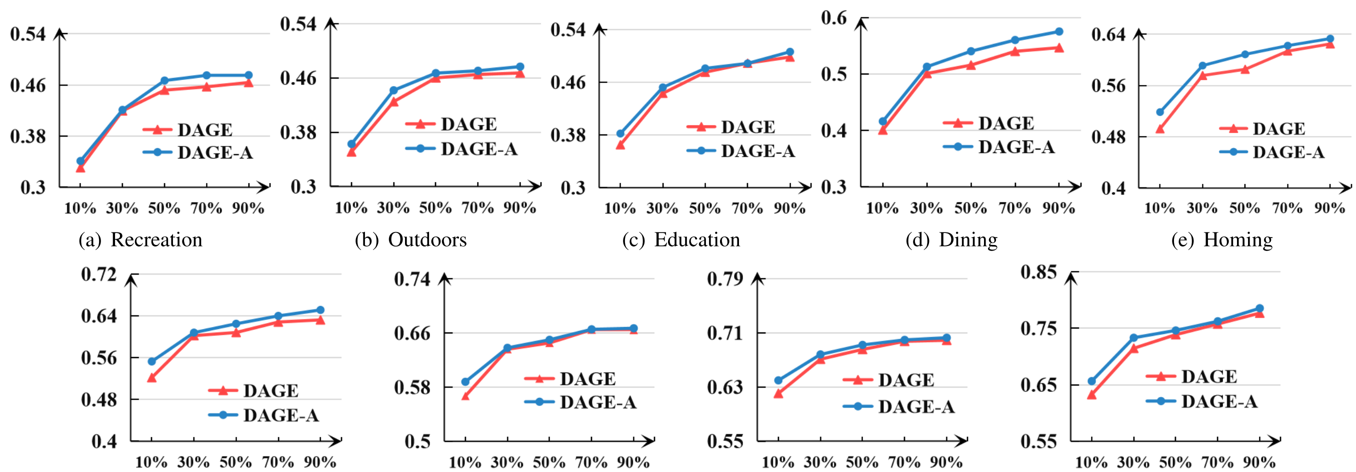
Tab. II and Tab. III show the overall performance of different models with various proportions of labeled samples in Beijing and Shanghai, respectively. Since *Nearest* and *Bayes' rule* are not sensitive to the training data, their performance is unchanged on the test data. Generally, this group of comparison experiments could show us the following insights.

Our modeling of trip semantics is superior. Among the first five supervised models, our *DAGE* outperforms other baselines under all the data proportions in both cities. Particularly, in Beijing, with 30% labeled data, our model achieves 4.86% improvement in accuracy and 4.22% improvement in macro F_1-score compared with the state-of-the-art *RF*. Such leads are stably maintained at 3% $\sim$ 4% in the rest data proportion settings. Generally, the prediction in Shanghai is more accurate but the improvements of *DAGE* are not as significant as that in Beijing, which indicates the trip purpose prediction in Beijing is more complicated yet our *DAGE* could perform much better than others. Additionally, although both *ANN* and our *DAGE* are neural networks, *ANN* performs much worse. It is because *ANN* simply aggregates all inputs in hidden layers, while our *DAGE* carefully models the correlations of features with two attention mechanisms in the latent space.

Semi-supervised learning is necessary. As shown in Tab. II and Tab. III, with the increase of labeled data, all the models

TABLE III
PREDICTION RESULTS OF DIFFERENT MODELS IN SHANGHAI

Models	Proportion of labeled trips in the training data																	
	10%		20%		30%		40%		50%		60%		70%		80%		90%	
	Acc	M_F$_1$	Acc	M_F$_1$	Acc	M_F$_1$	Acc	M_F$_1$	Acc	M_F$_1$	Acc	M_F$_1$	Acc	M_F$_1$	Acc	M_F$_1$	Acc	M_F$_1$
Nearest	26.76	23.94	26.76	23.94	26.76	23.94	26.76	23.94	26.76	23.94	26.76	23.94	26.76	23.94	26.76	23.94	26.76	23.94
Bayes' rule	39.08	32.67	39.08	32.67	39.08	32.67	39.08	32.67	39.08	32.67	39.08	32.67	39.08	32.67	39.08	32.67	39.08	32.67
ANN	41.23	24.77	41.69	25.14	41.87	25.38	42.27	26.62	42.34	26.80	42.43	27.22	42.59	27.68	42.71	27.48	42.78	27.24
RF	53.55	44.60	57.61	49.89	60.30	52.23	61.21	54.20	62.31	55.06	62.95	56.10	63.16	56.23	64.16	57.27	64.28	57.48
DAGE	55.67	46.66	59.97	51.58	62.12	52.33	64.05	55.33	65.03	56.53	65.92	57.79	66.63	58.46	67.01	58.63	68.01	60.54
DAGE-P	56.97	47.88	60.16	52.33	62.27	54.36	64.40	56.42	65.13	57.53	66.45	58.61	66.96	58.74	67.85	60.14	68.63	61.33
DAGE-A	**58.35**	**49.43**	**62.87**	**55.00**	**64.03**	**56.13**	**66.20**	**58.54**	**66.89**	**59.58**	**67.43**	**60.34**	**67.77**	**61.30**	**68.29**	**61.16**	**69.05**	**62.02**



(a) Recreation (b) Outdoors (c) Education (d) Dining (e) Homing
(f) Transportation (g) Shopping (h) Working (i) Health

Fig. 3. Category-specific performance of *DAGE* and *DAGE-A*.

achieve performance improvements. When increasing from 10% to 30%, our *DAGE* in both cities improves around 7% in accuracy and macro F$_1$-score. However, when increasing from 30% to 90%, the improvements are less than 5% in Beijing and around 7% in Shanghai. It means that for the problem of trip purpose prediction, the benefits from the high-cost labelling efforts would degrade at the early stage and the situation is more obvious in Beijing. Thus, it would be meaningful and necessary to establish a semi-supervised framework to improve the model's performance with unlabeled samples.

Our semi-supervised framework is effective and better. As shown in the last two rows, when using the semi-supervised learning framework, the performance of *DAGE* can be further improved. Moreover, *DAGE-A* (with autoencoder architecture) consistently outperforms *DAGE-P* (with pseudo-labels). For example, with 20% labeled data, *DAGE-P* in Beijing achieves 1.35% improvement in accuracy while our *DAGE-A* achieves 2.42%, and in Shanghai the improvements are 0.19% and 2.9%, respectively. Such results not only demonstrate the latent knowledge extracted from the unlabeled data is useful in the model training, but also show that our semi-supervised learning framework is more effective in capturing such latent knowledge. In particular, we find that with 70% labeled data, *DAGE-A* is even better than *DAGE* with 90% labeled data in Beijing. Such a result shows that our semi-supervised framework could reduce the labelling efforts by up to 20%.

E. Performance on Different Trip Purposes

In addition to the evaluation of overall performance, we also examine our model's performance on specific trip purposes. To save the place, we only present the detailed analysis in Beijing.

1) Category-Specific Performance Analysis: Figure 3 presents the category-specific results of *DAGE* and *DAGE-A* models, with the detailed analysis as follows. We obtain the following differences and insights regarding the 9 trip purpose categories.

- **The prediction difficulty is different.** For example, when using 10% labeled data, *DAGE* can achieve over 60% in F_1*-score* for the prediction of "Working" and "Health", while for "Recreation" and "Outdoors" that are around 35%. Such differences might be because that: i) the POI configurations near the "Working" and "Health" activities are usually simpler, thus they are easier to identify; ii) "Recreation" and "Outdoors" activities are often associated with other activities at time and space, e.g., "Dining". Moreover, we can find that even with 90% labeled data, the F_1-*scores* of "Recreation" and "Outdoors" are still less than 50%, thus the predictions of these two trip purposes are very difficult. Besides, "Health" is the most predictable one among 9 trip purposes, for which our models can achieve nearly 80% in F_1-*score*.

- **The improvements from labeled data are different.** When increasing the labeled data from 10% to 90%, the 9 trip purposes also show considerable differences in their performance improvements. For example, with our semi-supervised *DAGE-A*, "Dining" is improved by nearly 20%, while "Working" is improved by less than 10%. Such results indicate the latent distributions of all the possible "Dining" trips are more complicated, thus the increasing labeled data could consistently bring useful information for the model training. Similarly, the "Recreation" also achieves over 15% improvement.
- **The improvements from our semi-supervised framework are different.** From the standpoint of our semi-supervised *DAGE-A*, it is generally effective in improving the model's performance on each kind of trip purpose compared with the supervised *DAGE*. However, the improvements show two different trends with the increase of labeled data, i.e., from significant to negligible and the reverse. The first case is the majority of 9 trip purposes. It indicates that when the labeled data is sparse, the revealed data distributions are also limited thus the complementary knowledge from the unlabeled data could significantly enhance the performance. Then, with the increase of labeled data, the revealed distributions tend to be completed, so that the improvement goes down gradually. The second case occurs at "Recreation" and "Dining" purposes. According to the aforementioned observations, they are very difficult to predict and with complicated data distributions. Thus, at the beginning, the model would easily be over-fitting to the very limited labeled training data, so that the effects of unlabeled data are slight in the model training.
- **Our semi-supervised framework trained with** 70% **labeled data performs well.** For most trip purpose categories, when the labeled data comes over 70%, the improvements of our semi-supervised *DAGE-A* are not significant, but the performance still outperforms the *DAGE* with 100% labeled data. Hence, from the standpoint of real-world applications, it is effective and cost-efficient to employ 70% labeled data for the model training with our semi-supervised framework. *In the following, we will evaluate more aspects of the* DAGE-A *trained with* 70% *labeled data.*

2) Confusion Matrix: Figure 4 illustrates the normalized confusion matrix of *DAGE-A* trained with 70% labels in Beijing. Each row represents a set of trips corresponding to the same true purpose, and each column represents a set of trips with the same predicted purpose. The matrix is normalized at each row, thus the numbers in each row indicate the proportions of trips with different predicted labels, and the diagonal elements in this matrix are *recall* values.

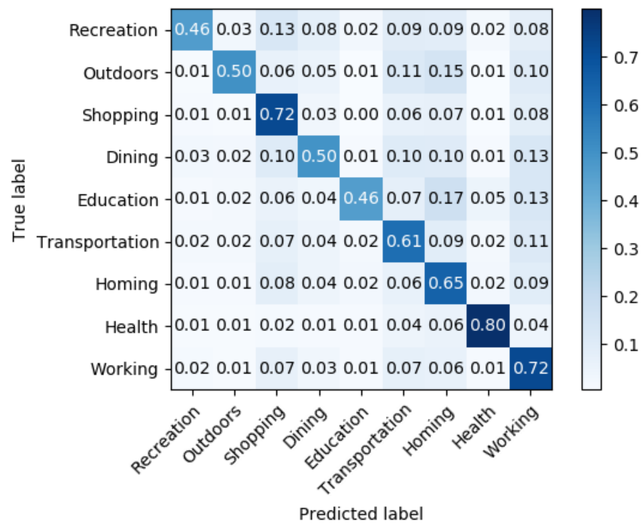


Fig. 4. Normalized confusion matrix of *DAGE-A* trained with 70% labels.

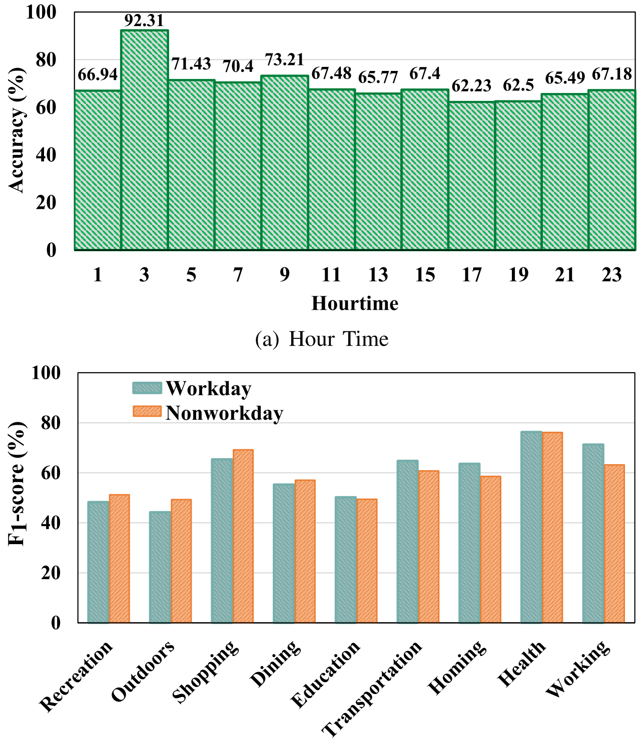
We find that with our prediction model, the recalls of "Shopping" and "Working" are over 70% and "Health" achieves 80%. While the recalls of " Recreation" and "Education" are less than 50%. One reason for such gaps lies in the differences of their nearby POI configurations. For example, the POI configurations near the "Health" activities are usually simpler than "Recreation". Thus it would be easier for our prediction model to identify the "Health" purposes through the POI check-in data. Besides, in reality, some human activities are usually associated with each other at time and space, so that our model can not distinguish them very well. For example, about 13% percent of "Recreation" purposes are predicted as "Shopping", and 11% percent of "Transportation" purposes are wrongly predicted as "Working".

F. Performance at Different Times and Locations

In this section, we investigate the performance of our *DAGE-A* trained with 70% labels at different times and locations in Beijing.

1) Temporal Dimension: Figure 5 (a) illustrates the prediction accuracies of our model at different hour times in a day. As we can see, for the most hour times, the performance is mainly floating around 60% and 70%, indicating the model is generalized to different times in a day. Moreover, the performance in the morning (from 5 am to 11 am) is somewhat better than that in the afternoon and evening. Such results show that trip purposes in the morning are less complicated. Besides, around 3 am, the prediction accuracy achieves 92.31%. It is because most activities are inactive at midnight, so that the potential candidates for prediction are much fewer.

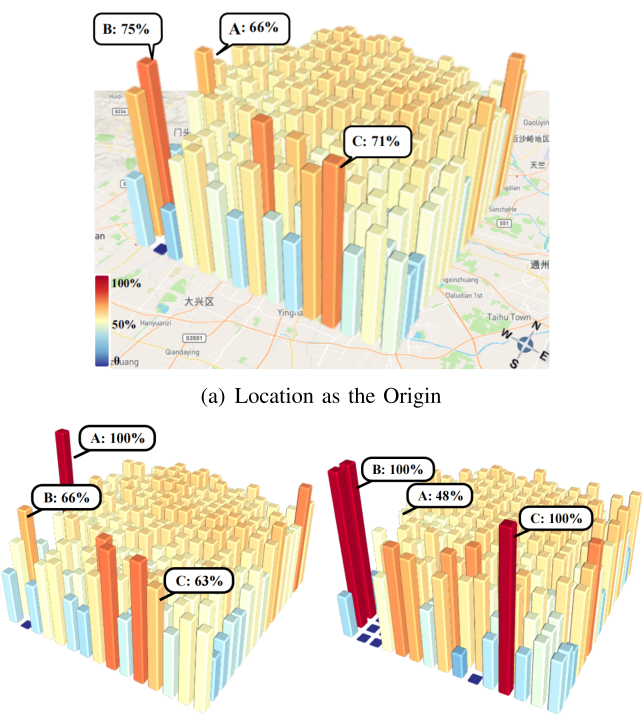
Figure 5 (b) presents the category-specific performance comparison from the perspective of day type. We can find that our model shows different performance on workday and non-workday. Specifically, the "Working", "Homing", "Transportation" and "Education" are more predictable on workday, while "Recreation", "Outdoors", "Shopping" and "Dining" are more predictable on non-workday. Such results are generally consistent with common sense that these activities have certain



(a) Hour Time

(b) Day Type

Fig. 5. Performance of *DAGE-A* at different times.



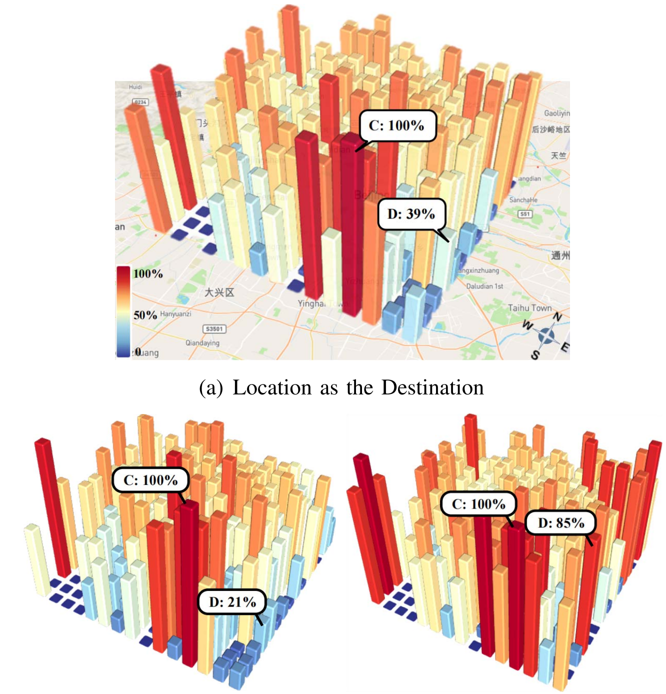
(a) Location as the Origin

(b) Origin_day (c) Origin_night

Fig. 6. Prediction accuracies of *DAGE-A* at different locations (as the origins of trips) in Beijing.

regularities at specific day types, accordingly demonstrating the effectiveness of our model.

2) Spatial Dimension: In order to investigate the performance at the spatial dimension, we divide the Five-Ring of Beijing city into 15 $\times$ 15 square grid cells with a width of 2 km. Figure 6 and Figure 7 show the overall performance of locations as the origins and destinations of trips respectively, and show the performance during the day and night as well. The 3D bar illustrates the corresponding prediction accuracy at each location.



(a) Location as the Destination

(b) Destination_day (c) Destination_night

Fig. 7. Prediction accuracies of *DAGE-A* at different locations (as the destinations of trips) in Beijing.

In Fig. 6, we can find that most locations as the origins are with around 60% prediction accuracies and don't change much from day to night. However, a few locations show very different patterns. For example, the overall accuracy of the location B is 75% in Fig. 6 (a), but the accuracy is only 66% during the day in Fig. 6 (b) while it achieves 100% during the night in Fig. 6 (c). Similarly, the location C is also with much higher accuracy during the night. By examining the prediction results, we find trips departed from these two locations are mainly for the "Homing" and "Transportation" purposes. On the contrary, the location A is with much higher accuracy during the day for "Dining". Hence, the prediction performance at locations is highly related to types of posed trip purposes.

In Fig. 7, we can find that a lot of locations are with higher accuracies as destinations than as the origins. It means trips ended at these locations are for naive (i.e., more predictable) purposes while trips departed from these locations are usually for complicated purposes. In other words, these locations are often connected to diverse activities and locations by trips. Besides, we find when the location C is as the destinations of trips, the accuracy is 100% during day and night consistently. It is because this location is mainly composed of residences and companies and the human activities in this location are simple and highly time-dependent, so that the trip purposes are easy to predict. There are also some destinations show very different patterns between day and night. For example, the location *D* is a place composed of various kinds of POIs. During the day, the prediction accuracy is only 21%, for which there are 5 frequently posed trip purpose categories. While during the night, the accuracy achieves 85%, and the frequent purposes are mainly "Homing". Such results indicate that for locations with complicated POI configurations, the prediction performance is more likely to show significant differences between day and night.

V. Related Work

A. Feature Engineering in Trip Purpose Prediction

Since human activities are influenced by various factors in reality, feature engineering is crucial in trip purpose prediction. *Geography characteristics* are widely used to depict the static activity-related characteristics of passenger's drop-off location, such as polygon-based information, POI configuration and street map [3], [11], [38]. *Trip and activity characteristics* (e.g., travel mode, activity duration) are also effective in identifying trip purposes [28], [38], [40], since human activities often show strong regularity at time and space. *Demographics characteristics* are used to reveal the respondents' preferences for activities or their travel patterns (e.g., age, gender, employment, and family structure) [10], [12], [17]. In real-life scenarios like taxis, many features cannot be obtained (e.g., passenger's activity duration and family structure), thus methods may lack pervasiveness.

B. Machine Learning Models in Trip Purpose Prediction

In recent years, machine learning algorithms are emerging in the prediction of trip purpose [8], [27], [30]. For example, since 2014, *Random Forest (RF)* [4] is widely adopted in trip purpose prediction [11], [12], [29]. Based on the Bayes model, the work in [7] takes both the fine-grained spatial and temporal patterns of human behaviors into consideration to impute the most likely trip purpose at the passenger's drop-off location. Besides, owing to the effectiveness in nonlinear regression, neural networks also show impressive performance in identifying trip purpose with complex input features [10], [28], [38]. Topic model (i.e., LDA) is used to infer trip purposes with the cellular network and POI data [43], where trips and users are regarded as words and documents respectively. In an unsupervised manner, autoencoder and a clustering algorithm are used to extract and cluster latent trip features from the GPS and POI data, then trip purposes are obtained by interpreting cluster centers [8]. Different from existing models, we are the first to: i) carefully model the correlations of features in the latent space; 2) establish a semi-supervised framework to improve the model's performance with a plenty of unlabeled trip data.

C. Semi-Supervised Learning in Mobile Computing

Although the ear of big data and IoT has opened up substantial opportunities for mobile computing, many researches still suffer from the problem of label shortage [9], [19]. In this regard, the semi-supervised learning techniques are widely used to enhance the model's performance with unlabeled data [1], [9], [25], [31], [42]. To name a few, the work in [31] present a semi-supervised framework for traffic anomaly detection at the edge of the mobile network, which only needs one class of samples (normal traffic) to train the model. In addition, a hierarchical semi-supervised training method is proposed in [1] for the intrusions detection in IoT networks, which takes into account the sequential characteristics of the unlabeled IoT traffic data during training. In [42], mean teacher semi-supervised learning is integrated with federated learning for the crowdsourced transportation mode identification, so as to utilize the sensed (unlabeled) data from distributed workers in the model training. In [9], a pseudo-label based semi-supervised framework is used to improve the performance of the graph representation model in identifying fine-grained driving style with the large-scale unlabeled GPS trajectory data. Note that the study in this paper is the first to adopt the semi-supervised learning in the trip purpose prediction.

VI. Conclusion and Future Work

In this paper, we propose a context-aware semi-supervised framework (*DAGE-A*) for predicting large-scale yet fine-grained trip purposes. It is based on pervasive data sources and is also effective with limited labeled training data, thus making it more applicable in real-world scenarios. Specifically, we employ the vehicle's GPS trajectory and public POI check-in data to reveal different trip contexts, then propose a dual-attention graph embedding network with autoencoder architecture to extract the higher-level activity semantics for trip purpose prediction. Moreover, our semi-supervised framework could improve the model's performance by incorporating the complementary knowledge from large-scale unlabeled data. Extensive experiments in Beijing and Shanghai demonstrate that the proposed framework significantly outperforms baseline models, and could reduce labelling efforts by up to 20%. We also demonstrate the great generalizability of our model at different times and locations in a city.

In the future, we will broaden and deepen this work in several directions. Specifically, we plan to explore more urban data sources to enrich the travel semantics, like the real-time social event data. Additionally, in some cities with low development of economy and infrastructure, it may be infeasible to collect sufficient trip data for model training, i.e., cold-start problem [15]. For this problem, since our trip purpose prediction is based on modelling the high-level human activity semantics (i.e., lifestyle) that may be similar across cities, we plan to leverage the prediction knowledge learned from a data-rich city to enable the prediction in data-scarce cities (i.e., transfer learning).

References

[1] M. Abdel-Basset, H. Hawash, R. K. Chakrabortty, and M. J. Ryan, "Semi-supervised spatiotemporal deep learning for intrusions detection in IoT networks," *IEEE Internet Things J.*, vol. 8, no. 15, pp. 12251–12265, Aug. 2021.

[2] L. Alessandretti, U. Aslak, and S. Lehmann, "The scales of human mobility," *Nature*, vol. 587, no. 7834, pp. 402–407, Nov. 2020.

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

[4] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[5] H. Cao, Z. Chen, F. Xu, Y. Li, and V. Kostakos, "Revisitation in urban space vs. Online: A comparison across POIs, websites, and smartphone apps," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 2, no. 4, pp. 1–24, Dec. 2018.

[6] C. Chen, H. Gong, C. Lawson, and E. Bialostozky, "Evaluating the feasibility of a passive travel survey collection in a complex urban environment: Lessons learned from the New York city case study," *Transp. Res. A, Policy Pract.*, vol. 44, no. 10, pp. 830–840, Dec. 2010.

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

[8] C. Chen, C. Liao, X. Xie, Y. Wang, and J. Zhao, "Trip2Vec: A deep embedding approach for clustering and profiling taxi trip purposes," *Pers. Ubiquitous Comput.*, vol. 23, no. 1, pp. 53–66, 2019.

[9] C. Chen, Q. Liu, X. Wang, C. Liao, and D. Zhang, "Semi-Traj2Graph: Identifying fine-grained driving style with GPS trajectory data via multi-task learning," *IEEE Trans. Big Data*, early access, Mar. 2, 2021, doi: 10.1109/TBDATA.2021.3063048.

[10] Y. Cui, C. Meng, Q. He, and J. Gao, "Forecasting current and next trip purpose with social media data and Google places," *Transp. Res. C, Emerg. Technol.*, vol. 97, pp. 159–174, Dec. 2018.

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

[12] L. Gong, R. Kanamori, and T. Yamamoto, "Data selection in machine learning for identifying trip purposes and travel modes from longitudinal GPS data collection lasting for seasons," *Travel Behav. Soc.*, vol. 11, pp. 131–140, Apr. 2018.

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

[14] L. Gong, T. Morikawa, T. Yamamoto, and H. Sato, "Deriving personal trip data from GPS data: A literature review on the existing methodologies," *Proc.-Social Behav. Sci.*, vol. 138, pp. 557–565, Jul. 2014.

[15] Y. Hu, J. Wang, B. Wu, and S. Helal, "RL-Recruiter+: Mobility-predictability-aware participant selection learning for from-scratch mobile crowdsensing," *IEEE Trans. Mobile Comput.*, early access, 2021, May 5, 2021, doi: 10.1109/TMC.2021.3077636.

[16] X. Kong, K. Wang, M. Hou, F. Xia, G. Karmakar, and J. Li, "Exploring human mobility for multi-pattern passenger prediction: A graph learning framework," *IEEE Trans. Intell. Transp. Syst.*, early access, Feb. 9, 2022, doi: 10.1109/TITS.2022.3148116.

[17] C. M. Krause and L. Zhang, "Short-term travel behavior prediction with GPS, land use, and point of interest data," *Transp. Res. B, Methodol.*, vol. 123, pp. 349–361, May 2018.

[18] D.-H. Lee *et al.*, "Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks," in *Proc. Workshop Challenges Represent. Learn. (ICML)*, vol. 3, 2013, p. 896.

[19] X. Li, H. Liu, W. Wang, Y. Zheng, H. Lv, and Z. Lv, "Big data analysis of the Internet of Things in the digital twins of smart city based on deep learning," *Future Gener. Comput. Syst.*, vol. 128, pp. 167–177, Mar. 2022.

[20] D. Lian, Y. Zhu, X. Xie, and E. Chen, "Analyzing location predictability on location-based social networks," in *Proc. Pacific-Asia Conf. Knowl. Discovery Data Mining*, 2014, pp. 102–113.

[21] C. Liao *et al.*, "Wheels know why you travel: Predicting trip purpose via a dual-attention graph embedding network," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 6, no. 1, pp. 1–22, Mar. 2022.

[22] C. Liao, C. Chen, C. Xiang, H. Huang, H. Xie, and S. Guo, "Taxi-Passenger's destination prediction via GPS embedding and attention-based BiLSTM model," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 5, pp. 4460–4473, May 2022.

[23] D. Liao, W. Liu, Y. Zhong, J. Li, and G. Wang, "Predicting activity and location with multi-task context aware recurrent neural network," in *Proc. 27th Int. Joint Conf. Artif. Intell.*, 2018, pp. 3435–3441.

[24] H. Liu, T. Li, R. Hu, Y. Fu, J. Gu, and H. Xiong, "Joint representation learning for multi-modal transportation recommendation," in *Proc. AAAI Conf. Artif. Intell.*, vol. 33, pp. 1036–1043, Jul. 2019.

[25] M. Lv, L. Chen, T. Chen, and G. Chen, "Bi-view semi-supervised learning based semantic human activity recognition using accelerometers," *IEEE Trans. Mobile Comput.*, vol. 17, no. 9, pp. 1991–2001, Sep. 2018.

[26] Z. Lv, L. Qiao, and A. K. Singh, "Advanced machine learning on cognitive computing for human behavior analysis," *IEEE Trans. Computat. Social Syst.*, vol. 8, no. 5, pp. 1194–1202, Oct. 2021.

[27] S. Lyu and T. Kusakabe, "Graph-aware chained trip purpose inference," in *Proc. IEEE Int. Intell. Transp. Syst. Conf. (ITSC)*, Sep. 2021, pp. 3691–3697.

[28] C. Meng, Y. Cui, Q. He, L. Su, and J. Gao, "Travel purpose inference with GPS trajectories, POIs, and geo-tagged social media data," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Dec. 2017, pp. 1319–1324.

[29] L. Montini, N. Rieser-Schüssler, A. Horni, and K. W. Axhausen, "Trip purpose identification from GPS tracks," *Transp. Res. Rec.*, vol. 2405, no. 1, pp. 16–23, Jan. 2014.

[30] M. G. S. Oliveira, P. Vovsha, J. Wolf, and M. Mitchell, "Evaluation of two methods for identifying trip purpose in GPS-based household travel surveys," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2405, no. 1, pp. 33–41, Jan. 2014.

[31] A. Pelati, M. Meo, and P. Dini, "Traffic anomaly detection using deep semi-supervised learning at the mobile edge," *IEEE Trans. Veh. Technol.*, vol. 71, no. 8, pp. 8919–8932, Aug. 2022.

[32] J. A. Stankovic, "Research directions for the Internet of Things," *IEEE Internet Things J.*, vol. 1, no. 1, pp. 3–9, Feb. 2014.

[33] A. Vaswani *et al.*, "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 5998–6008.

[34] P. Velikovi, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio, "Graph attention networks," in *Proc. Int. Conf. Learn. Represent.*, 2017, pp. 1–12.

[35] P. Voigt and A. Von dem Bussche, "The EU general data protection regulation (GDPR)," in *A Practical Guide*, vol. 10, 1st, ed. Cham, Switzerland: Springer, 2017, Art. no. 3152676.

[36] P. Wang, G. Liu, Y. Fu, Y. Zhou, and J. Li, "Spotting trip purposes from taxi trajectories: A general probabilistic model," *ACM Trans. Intell. Syst. Technol.*, vol. 9, no. 3, pp. 1–26, May 2018.

[37] M. Wu *et al.*, "Roles of locational factors in the rise and fall of restaurants: A case study of Beijing with POI data," *Cities*, vol. 113, Jun. 2021, Art. no. 103185.

[38] G. Xiao, Z. Juan, and C. Zhang, "Detecting trip purposes from smartphone-based travel surveys with artificial neural networks and particle swarm optimization," *Transp. Res. C, Emerg. Technol.*, vol. 71, pp. 447–463, Oct. 2016.

[39] L. Xu, C. Jiang, N. He, Y. Qian, Y. Ren, and J. Li, "Check in or not? A stochastic game for privacy preserving in point-of-interest recommendation system," *IEEE Internet Things J.*, vol. 5, no. 5, pp. 4178–4190, Oct. 2018.

[40] Y. Yang, P. Widhalm, S. Athavale, and M. C. González, "Mobility sequence extraction and labeling using sparse cell phone data," in *Proc. 13th AAAI Conference on Artificial Intelligence*, 2016, pp. 1–2.

[41] J. Yuan, Y. Zheng, and X. Xie, "Discovering regions of different functions in a city using human mobility and POIs," in *Proc. ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2012*, pp. 186–194.

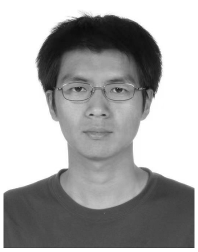
[42] C. Zhang, Y. Zhu, C. Markos, S. Yu, and J. James, "Towards crowd-sourced transportation mode identification: A semi-supervised federated learning approach," *IEEE Internet Things J.*, vol. 9, no. 14, pp. 11868–11882, Jul. 2021.

[43] X. Zhao, Z. Li, Y. Zhang, and Y. Lv, "Discover trip purposes from cellular network data with topic modeling," *IEEE Intell. Transp. Syst. Mag.*, vol. 14, no. 4, pp. 37–46, Jul. 2022.

Chengwu Liao received the B.Sc. degree from the College of Information Engineering and Automation, Kunming University of Science and Technology, Yunnan, China, in 2017. He is currently pursuing the Ph.D. degree with the College of Computer Science, Chongqing University, China. His research interests include intelligent transportation systems, spatiotemporal trajectory mining, and urban data visualization.

Chao Chen (Senior Member, IEEE) received the B.Sc. and M.Sc. degrees in control science and control engineering from Northwestern Polytechnical University, Xi'an, China, in 2007 and 2010, respectively, and the Ph.D. degree from Pierre and Marie Curie University and Institut Mines-Télécom/Télécom SudParis, France, in 2014. He is currently a Full Professor at the College of Computer Science, Chongqing University, Chongqing, China. He has published over 100 papers including 30 ACM/IEEE TRANSACTIONS. His research interests include pervasive computing, mobile computing, urban logistics, data mining from large-scale GPS trajectory data, and big data analytics for smart cities. His work on taxi trajectory data mining was featured by IEEE SPECTRUM in 2011, 2016, and 2020, respectively. He is a Senior Member of CCF. He was a recipient of the Best Paper Runner-Up Award at MobiQuitous in 2011.

Suiming Guo received the Ph.D. degree from The Chinese University of Hong Kong. He is currently an Associate Professor with the College of Information Science and Technology, Jinan University, Guangzhou, China. His research interests include data mining, urban computing, pervasive computing, and smart cities studies.



Leye Wang (Member, IEEE) received the Ph.D. degree in computer science from TELECOM SudParis and University Paris 6, France, in 2016. He was a Post-Doctoral Researcher with The Hong Kong University of Science and Technology. He is currently an Assistant Professor at the Key Laboratory of High Confidence Software Technologies (Peking University, China), MOE, and the School of Computer Science, Peking University. His research interests include ubiquitous computing, mobile crowdsensing, and urban computing.

Fuqiang Gu (Member, IEEE) received the Ph.D. degree from the University of Melbourne. He is currently a Professor at the College of Computer Science, Chongqing University, China. Before joining the Chongqing University, he has successively worked as a Researcher at RWTH-Aachen University, the University of Toronto, and the National University of Singapore. His main research interests are positioning and navigation, robotics, autonomous systems, and machine learning. He is a member of ACM and CCF.

Jing Luo received the Ph.D. degree in map and GIS from the Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, in 2007. She was a Post-Doctoral Researcher at Tsinghua University from 2007 to 2009. She is currently the CIO at China Railway Signal & Communication Company Ltd. (CRSC), and the Director of CRSC Research Laboratory for Smart City. Her research interests are in the areas of video analysis and processing, ubiquitous computing, smart city big data, intelligent transportation, and space-time cloud computing.

Ke Xu (Senior Member, IEEE) received the Ph.D. degree from the Department of Computer Science and Technology, Tsinghua University, Beijing, China. He currently serves as a Full Professor at Tsinghua University. He has published more than 200 technical articles and holds 11 U.S. patents in the research areas of next-generation internet, blockchain systems, the Internet of Things, and network security. He is a member of ACM. He served as the Steering Committee Chair for IEEE/ACM IWQoS. He is an Editor of IEEE INTERNET OF THINGS JOURNAL. He has guest edited several special issues in IEEE and Springer journals.