

Wheels Know Why You Travel: Predicting Trip Purpose via a Dual-Attention Graph Embedding Network

CHENGWU LIAO*, Chongqing University, China
 CHAO CHEN*†, Chongqing University, China
 SUIMING GUO, Jinan University, China
 ZHU WANG, Northwestern Polytechnical University, China
 YAXIAO LIU, AWS China, China
 KE XU, Tsinghua University, China
 DAQING ZHANG, Institut Polytechnique de Paris, France

Trip purpose - i.e., why people travel - is an important yet challenging research topic in travel behavior analysis. Generally, the key to this problem is understanding the activity semantics from trip contexts. However, most existing methods rely on passengers' *sensitive* information - e.g., daily travel log or home address from surveys - to achieve accurate results, and could thus be hardly applied in real-life scenarios. In this paper, we aim to predict the passenger's trip purpose in the scenarios of door-to-door ride services (e.g., taxi trips) by only using the vehicle's GPS trajectory on roads, for which "wheels" is used as a metaphor. Specifically, we propose a novel dual-attention graph embedding model based on the vehicle's trajectory and public POI check-in data. Firstly, both data are aggregated to augment the activity semantics of trip contexts, including the spatiotemporal context and POI contexts at the origin and destination, which are important clues. Based on that, *graph attention networks* and *soft-attention* are employed to model the dependency of different contexts on the trip purpose, so as to obtain the trip's comprehensive activity semantics for the final prediction. Extensive experiments are conducted based on the large-scale labeled datasets in Beijing. The prediction results show a considerable improvement compared to state-of-the-arts. A case study demonstrates the feasibility of our study.

CCS Concepts: • **Human-centered computing** → **Ubiquitous computing**; • **Computing methodologies** → *Machine learning algorithms*.

Additional Key Words and Phrases: trip purpose, GPS trajectory, POI check-in data, graph embedding, attention mechanism

ACM Reference Format:

Chengwu Liao, Chao Chen, Suiming Guo, Zhu Wang, Yaxiao Liu, Ke Xu, and Daqing Zhang. 2022. Wheels Know Why You Travel: Predicting Trip Purpose via a Dual-Attention Graph Embedding Network. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 6, 1, Article 22 (March 2022), 22 pages. <https://doi.org/10.1145/3517239>

*Both authors contributed equally to this work and share the first authorship.

†This is the corresponding author.

Authors' addresses: Chengwu Liao, lcw@cqu.edu.cn, Chongqing University, Chongqing, China, 400044; Chao Chen, cschaochen@cqu.edu.cn, Chongqing University, Chongqing, China, 400044; Suiming Guo, Jinan University, Guangzhou, China, guosuming@email.jnu.edu.cn; Zhu Wang, Northwestern Polytechnical University, Xi'an, China, wangzhu@nwpu.edu.cn; Yaxiao Liu, AWS China, Beijing, China, rootliu@gmail.com; Ke Xu, Tsinghua University, Beijing, China, xuke@mail.tsinghua.edu.cn; Daqing Zhang, Institut Polytechnique de Paris, France, daqing.zhang@telecom-sudparis.eu.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2022 Association for Computing Machinery.

2474-9567/2022/3-ART22 \$15.00

<https://doi.org/10.1145/3517239>

1 INTRODUCTION

Travel behavior analysis serves as the foundation and becomes a long-standing topic in smart mobility and urban applications, including transportation, urban planning, epidemic controlling, and so forth [7, 18, 21, 25]. In the past decade, with the wide availability of GPS trajectory data, a lot of achievements have been made on revealing the spatiotemporal patterns of travel behaviors [1, 5]. However, relatively few efforts were devoted to uncovering people's intention of travel behaviors, i.e., trip purposes. Different from trajectories explicitly telling *when* and *where* people move, trip purposes are the semantic information answering *why* people travel in the city. Predicting such knowledge could significantly benefit several parties in the city. Specifically, for travellers, the prediction could enable personalized in-car advertising/recommendation, meanwhile, the advertising could increase taxi companies' extra income. Besides, for urban planners, knowing the city-wide distribution of trip purposes could help the public transportation planning (e.g., set up a new bus route for travels with the "Working" purpose). Thus, in recent years, trip purpose has been recognized as an important aspect in travel behavior analysis [7, 38].

Although trip purpose is not embedded in the GPS trajectory data, it can be revealed by the activity semantics from trip contexts. Practically, trip purpose refers to the activity that a passenger takes after being dropped off, so that it cannot be directly sensed by the vehicle. While human activities in the city usually show strong regularity at time and space [23]. For example, the passenger would probably take the most popular activity near the drop-off location during a specific time period, such as the "Dining" activity in a dining area at noon. In this sense, those activity semantics from trip contexts could be used to identify the passenger's trip purpose. With the proliferation of information and communication technologies (ICTs) in daily life, human behaviors including *travel* and *activities* are able to leave behind substantial urban data at the cyber space. For example, people's travel information can be recorded by ride-on-demand (RoD) services like *Uber*, and activities can be shared by users at location-based social networks (LBSNs). Hence, such urban data sources grant us valuable opportunities to perceive the passenger's trip contexts, and further understand the travel-related activity semantics for trip purpose prediction.

However, existing trip purpose prediction methods are mainly restricted in practice due to the delimited data sources. Specifically, most of studies rely on *sensitive data sources* like household or travel surveys to characterize the passenger's preference on activities [13, 41]. We argue that although such an approach is effective in achieving accurate results, these methods could be hardly applied at a large scale in real-life scenarios. On one hand, most applications just cannot get those data sources in real situations [11]. On the other hand, using individual's *sensitive information* might cause a serious privacy issue. In this study, the *sensitive information* generally refers to any private data from the passenger, such as personal profile (e.g., identity, home address, etc.) and smartphone data. Recently, the use of these data sources is suggested to be cautious in smart city applications [43].

In this paper, for real-life applications, we aim to provide a more *ubiquitous* and *applicable* trip purpose prediction approach with *passenger's insensitive information*. Specifically, the target application scenarios are door-to-door ride services like taxi trips. Note that the trip in such a scenario is isolated, i.e., *there is no any other passenger's historical trip information and no connections between different trips*. Moreover, the system is meant to be deployed in vehicles and also has no digital connection with passengers, so that *it would not record or use any individual data*. With these in mind, we propose a novel trip purpose prediction approach by merely employing *the vehicle's GPS trajectory* from this trip, which can be viewed as using wheels on the ride ("wheels" is a metaphor for the trajectory).

As mentioned before, GPS trajectory only reveals when and where a vehicle (with the passenger) is moving, but without semantic meaning regarding human activities. To narrow the gap, we further employ the public POI check-in data from LBSNs which is contributed by *anonymous users* in the city (also being passenger-insensitive). As is well known, point-of-interest (POI) is the basic unit of human activity, and check-in records are generated when users visit it. In particular, POI context derived from the POI check-in data is able to reveal both the static

and dynamic characteristics of human activities in an area. In terms of the static characteristics, it conveys activities' types (e.g., "Dining") and geographic distribution. In terms of the dynamic characteristics, it reveals the time-varying popularity of activities in POIs. Hence, combining the GPS trajectory with public POI check-in data is a promising way to understand the travel-related activity semantics for trip purpose prediction.

Nevertheless, it is still non-trivial to effectively extract the trip's activity semantics. Firstly, the raw POI check-in numbers may not reveal the actual condition of human activities, since different activities usually have uneven sampling rates in LBSNs. What's more, passenger's activity at the destination location has complicated relationships with many factors. To be more specific, based on the trajectory and POI check-in data, the challenges for trip purpose prediction include: 1) Find the discriminative POI features related to the passenger's trip. 2) Model the correlation between neighboring POIs, since some kinds of human activities are usually associated with each other at time and space (e.g., "Recreation" and "Shopping"). 3) Consider the dependency of activity semantics from the origin location, since passenger's activities before and after the trip have a sort of inherent relationship. For example, after the "Working" activity, the passenger is probably going for "Homing" rather than "Working" again. 4) Consider the dependency of spatiotemporal context (including time and travel cost), since human activities demonstrate strong spatial and temporal regularity [23]. For example, the passenger may not take a long trip for the "Dining" activity at 3 PM.

To resolve the aforementioned challenges, we propose a dual-attention graph embedding model. Specifically, we first augment the semantic meaning of three trip contexts. In terms of the POI contexts at origin and destination locations (OD POI contexts for short), we extract three discriminative features for each POI category from the check-in data, namely *period popularity*, *distance* and *uniqueness*. In terms of the spatiotemporal context, we extract *day type*, *hour time*, *travel time* and *travel distance* from the GPS trajectory. We then convert the OD POI contexts into the graph structure. Based on that, *graph attention networks (GATs)* are employed to capture the activity semantics of each POI category by modelling its correlations with neighboring POIs. Next, *soft-attention* is used to extract the comprehensive activity semantics from three kinds of trip contexts, by modelling their dependency on the trip purpose. Finally, the extracted activity semantics is used to predict the probabilities of candidate trip purposes.

In short, the main contributions of this paper can be summarized as follows:

- We present a novel dual-attention graph embedding model for predicting the trip purpose in the scenarios of door-to-door ride services (e.g., taxi trips). The model extracts the activity semantics from passenger's insensitive information and trained with large-scale urban datasets, which is more ubiquitous and applicable for real-life applications.
- Based on the augmented trip contexts, category-aware GATs are employed to capture the neighboring activity semantics of each POI category, by modelling the correlation between neighboring POIs; *Soft-attention* is used to extract the comprehensive activity semantics of passenger's trip by modelling the dependency of different trip contexts on the trip purpose.
- Among studies on the trip purpose prediction, our model is the first neural network that performs the localized POI semantics extraction in a graph structure and carefully models the inherent correlations of features in the latent space. Moreover, the modified GAT is category-aware in extracting neighboring POI semantics.
- We conduct extensive experiments including a comparison study, an ablation study and a case study to evaluate the effectiveness of our model. Results show that our model outperforms baseline algorithms. It can achieve 65.56% and 79.76% accuracies on the 9-class and 4-class trip purposes respectively. The results also demonstrate the effectiveness of each component in our model.

The remainder of this paper is organized as follows. Section 2 presents the related work, and Section 3 introduces a few definitions and problem statement of this paper. Section 4 elaborates details on the trip context

augmentation and dual-attention graph embedding network. Section 5 presents results on a group of experiments together with a case study. Section 6 concludes the paper and discusses future directions.

2 RELATED WORK

In recent years, trip purpose prediction has attracted continuous attention in the urban computing society. Existing studies mainly concentrate on investigating it from two perspectives, namely feature engineering and prediction algorithms, detailed as follows.

2.1 Feature Engineering for Trip Purpose Prediction

Since human activities are influenced by various factors in reality, feature engineering becomes a crucial procedure in predicting trip purpose. Generally, researchers usually employ more than two kinds of data sources, so as to depict passenger's activity semantics from multiple perspectives. The commonly adopted data sources include sensing devices, LBSNs, travel surveys, and location-based services [30].

Geography characteristics are the most frequently employed features, such as polygon-based information, POIs configuration and street map [3, 10, 11, 41]. This kind of information is commonly used to depict the static activity-related characteristics of passenger's drop-off location. For example, polygon-based information and POIs are often used to determine the land-use type of trip end for trip purpose prediction [3, 10, 41]. In particular, the distance between trip end and nearby POIs (activity units) is identified as an important clue [7, 15, 29].

Trip and activity characteristics are also effective in identifying trip purposes [10, 11, 28, 41], since human activities often show strong regularity at time and space. Specifically, trip characteristics can be derived from sensing devices like GPS, including the time, travel mode and travel cost [13, 41]. The study in [34] achieves 81% prediction accuracy by using a few trip features (i.e., speed, acceleration, weekday and period of the day). However, its dataset is only composed of 19 respondents' daily travel logs, while our trip dataset was generated by more than 200,000 arbitrary passengers in Beijing, which is with much better data diversity. The activity characteristics of each individual (e.g., duration and activity history) are often obtained from travel surveys [10, 11, 26, 33]. Note that the activity duration is demonstrated to be the most effective feature in the trip purpose prediction [11]. Besides, the activity characteristics of trip ends are usually derived from the POI check-in data in LBSNs, e.g., *Foursquare*, *Twitter*, *Google Places* [7, 10, 12, 28]. However, many studies focus on the static POI context like land-use/functionality [10, 12], which cannot capture the dynamic characteristics of human activities. To narrow this gap, we further consider the *period popularity* to reveal the dynamic POI context.

Demographics characteristics (individual and household) are widely used in existing studies [10, 11, 15, 22, 41]. Such information is mainly collected by surveys (e.g., age, gender, employment, monthly income and family structure). Generally, demographics characteristics are used to reveal the respondents' preference on activities or their travel patterns, so as to narrow down the candidate trip purposes.

By reviewing existing studies, we find that considering features from different perspectives is effective in the trip purpose prediction. However, some features are either sensitive for passengers or cannot be obtained in real-life scenarios (e.g., the activity duration and family structure). As a result, models may cause privacy issues and lack pervasiveness.

2.2 Algorithms for Trip Purpose Prediction

In existing studies, algorithms for trip purpose prediction can be broadly categorized into three groups, namely rule-based, probability-based and machine learning algorithms.

Deterministic rules are the earliest algorithms adopted in the trip purpose prediction [3, 33, 39]. Such algorithms match the revealed information (e.g., geography and trip characteristics) with a series of predefined heuristic

rules to identify trip purposes [17]. For example, the study in [3] presents a straightforward approach, which directly employs the activity type of the nearest POI as trip purpose.

Probability-based algorithms compute the probabilities of candidate trip purposes by using statistical models (e.g., Bayes' rules, topic model) with the revealed information [7, 14, 16, 31, 38]. For example, the study in [7] takes both the fine-grained spatial and temporal patterns of human behaviors into consideration, and adopts the Bayes' rules to model the probabilities of POIs being visited by the passenger. Compared with heuristic rules, probability-based algorithms are less dependent on the domain knowledge of researchers and are thus more transferable [30].

In recent years, machine learning algorithms are emerging in the prediction of trip purpose [8, 11, 15, 27, 28, 31]. On the basis of large datasets and intensive computation, they have been demonstrated to outperform the rule-based and probability-based algorithms [11, 28]. Since 2014, *Random Forest (RF)* [4] is widely adopted in the trip purpose prediction [11, 15, 29]. It consists of multiple decision trees, each of which predicts a trip purpose based on the given features, and the final result is determined by a voting mechanism. Besides, owing to the effectiveness in nonlinear regression, neural networks also show impressive performance in identifying trip purpose with complex input features [10, 28, 41]. For example, a three-layer *Artificial Neural Network (ANN)* with particle swarm optimization achieves 96.53% accuracy on the prediction of 6 candidates [41]. However, most of these models are based on the travel survey data. In addition to the privacy issue, it also might suffer from two limitations: 1) The training data is composed of daily travel logs from limited respondents, i.e., low data diversity; 2) Features might be inaccurately reported in the prompted recall surveys, i.e., unstable data quality.

There are also a few machine learning models that don't require the survey data. To name a few, topic model (i.e., LDA) is used to infer trip purposes with the cellular network and POI data [44], where trips and users are regarded as words and documents respectively. Since LDA is based on the "bag-of-word" exchangeability assumption, it might have the limitation of semantic loss during the computation. In [34], RF model identified by auto machine learning is used with the smartphone traces. However, the traces were collected from 19 respondents, so that the model's generalizability might be limited. In an unsupervised manner, autoencoder and a clustering algorithm are used to extract and cluster latent trip features from the GPS and POI data [8]. Then trip purposes are interpreted based on the semantics of cluster centers. Since this approach doesn't utilize ground truth, the correctness of clustering and interpretation is not guaranteed.

3 PRELIMINARY

3.1 Definitions

Definition 1 (Trajectory Data). Trajectory data is collected from vehicles in the scenarios of door-to-door ride services. Each trajectory consists of a sequence of GPS points on roads, and each GPS point contains the geographic position and time information about the vehicle, denoted by $l_i = (lng_i, lat_i, t_i)$.










Definition 2 (Trip's OD Pair). The trip in this study is represented by its origin-destination pair. An OD pair is a pair of GPS points (l_o, l_d) from a trajectory at the origin (i.e., pick-up location) and destination (i.e., drop-off location).

Definition 3 (Point of Interest). A POI refers to a place that is the very basic unit of taking human activities for people. POIs are usually represented by their positions and POI category information.

Definition 4 (POI Category). A POI category is the semantic label for a set of POIs, indicating the type of potential human activities at these POIs.

Table 1 shows the 9 primary POI categories provided by a Chinese LBSN called *Jiebang* [24]. Similar to other world-wide popular LBSNs such as *Foursquare* and *Gowalla*, people can also use *Jiebang* to track and share life moments with friends. In this table, each category is typically related to one kind of human activity. For

Table 1. Nine primary POI categories and their corresponding trip purposes.

k	Primary POI Categories	POI Icons	Human Activities/ Trip Purposes
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

example, the *Restaurant* corresponds to the “Dining” activity, while the *Apartment and Residence* corresponds to the “Homing” activity. Note that these human activities are served as the candidate trip purposes that we intend to predict in this study. (We introduce how to obtain the 9 primary POI categories in Appx. B.)

Definition 5 (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 the information about the user’s identity, the check-in time and the corresponding POI venue.

Generally, the number of check-ins during a time window could reveal the period popularity of a POI. Moreover, the check-in data is also able to demonstrate the geographic distribution and time-dependent heatmaps of different POIs in a city.

3.2 Problem Statement

The problem of predicting the passenger’s trip purpose can be viewed as identifying the most likely activity that the passenger takes after being dropped off. Specifically, the problem is formulated as:

Given:

- (1) Trip’s OD pair (l_o, l_d) : A pair of GPS points collected by the vehicle, with regard to the pick-up and drop-off locations respectively.
- (2) A set of POIs and their corresponding historical check-in records CI in the designated city.
- (3) A set of candidate activities \bar{A} in Tab. 1 (i.e., trip purposes).

Predict $\hat{p}(y = \bar{a} | l_o, l_d, CI)$, $\bar{a} \in \bar{A}$: The probability of a candidate activity \bar{a} being the actual activity y that the passenger would take, on the condition of (l_o, l_d, CI) .

Return the activity with the largest probability as the predicted purpose for this trip.

4 METHODOLOGY

4.1 Overview

The framework of our prediction model is illustrated in Fig. 1, which consists of three stages, namely *trip context augmentation*, *dual-attention graph embedding*, and *classification*.

- *Trip Context Augmentation*: When and where the passenger taking a trip are two foremost significant clues for trip purpose prediction, but the *raw* trajectory data still lacks semantics regarding passenger’s potential activity. In this stage, trajectory data and POI check-in data are aggregated and used to augment the semantic meaning of trip contexts, including *spatiotemporal (ST) context*, and *OD POI contexts*.

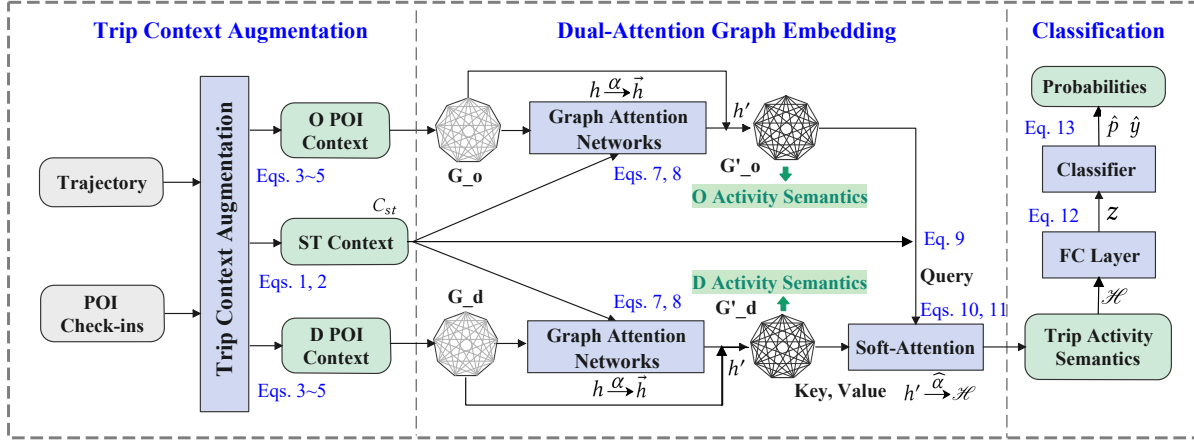


Fig. 1. Framework of our trip purpose prediction model.

- **Dual-Attention Graph Embedding:** The augmented trip contexts obtained in the first stage is expected to convey the primary activity semantics from different *partial* perspectives. Thus, in this stage, a dual-attention graph embedding is established to extract the comprehensive activity semantics of passenger's trip in the high-level feature space. Specifically, the augmented OD POI contexts are first converted into the graph structure, namely G_o, G_d . On top of that, two branches of GATs are employed to capture the neighboring activity semantics of each POI category in G_o and G_d , respectively. Note that the ST context is used as the complementary information. Then, the *soft-attention* mechanism is employed to aggregate three kinds of trip contexts (i.e., G'_o, G'_d and ST context), so as to extract the comprehensive activity semantics at the trip level. In this *soft-attention*, the activity semantics of G'_o and spatiotemporal context are served as the *query*, and the activity semantics of each POI category in G'_d is served as both the *key* and *value*. Such a manner is capable of modelling the different contributions of POI categories in G'_d for passenger's trip purpose on the condition of specific origin activity semantics and spatiotemporal context.
- **Classification:** In this stage, a full-connected layer is first employed to fuse the extracted activity semantics of the passenger's trip. Then *softmax* function is adopted as the classifier to output the probabilities of different candidate trip purposes. Besides, this stage also computes the loss of entire neural network for the back propagation process during the training phase.

We also mark all important symbols, equations in Fig. 1 for the easy check. To gain a better understanding of this trip purpose prediction procedure, we recommend readers to keep referring to this figure throughout this section.

4.2 Trip Context Augmentation

4.2.1 Spatiotemporal Context Augmentation. Human activities naturally present strong temporal regularity, such as working, homing. In this sense, the temporal context of passenger's trip is of paramount importance for understanding why people travel. Hence, for each trip, three kinds of temporal contexts are extracted from the GPS trajectory, including the type of day (i.e., workday or non-workday) and the hour time when this trip started and ended, and the travel time. In particular, the hour time value t is converted to the radian of a unit circle in the coordinates centered on $(0, 0)$, i.e., $[0, 24) \Rightarrow [0, 2\pi)$. Then the hour time is represented by the coordinate of a point in the unit circle based on the radian θ , as shown in Eq. 1. Such a representation could maintain the time

similarity between 00 : 00 and 23 : 00.

$$H(t) = (\cos \theta, \sin \theta), \theta = 2\pi \left(\frac{t}{24} \right) \quad (1)$$

In addition, the spherical distance between the origin and destination is computed and used as the spatiotemporal cost of this trip, together with the travel time. The underlying rationale is that people often travel with long time and distance for unusual activities like business. At last, the spatiotemporal context C_{st} of a trip (i.e., tr) is represented as Eq. 2.

$$C_{st}(tr) = [TYP(tr), H(t_o), H(t_d), t_d - t_o, l_d - l_o] \quad (2)$$

where $TYP(tr)$ and $H(t)$ obtain the corresponding day type and hour time. $t_d - t_o$ and $l_d - l_o$ refer to the travel time and distance, respectively.

4.2.2 OD POI Contexts Augmentation. The POI check-in data at origin and destination locations, conveys the real condition of 9 categories of human activities at reasonably fine temporal and spatial levels. Generally, such POI context indicates the land-use/functionality of a location, which has been recognized to be useful for understanding trip purpose [7, 16]. For instance, a passenger entering the residential area may imply a trip for the “Homing” purpose with high confidence. Thus, it is important to depict the *static* POI context at O/D location. However, as cities undergo rapid sprawl, the functionality of a city region is usually complex and mixed [42]. In real cases, different human activities would dominate the same region during different time periods, that is, the region functionality evolves with the time of a day. As a result, it is also crucial to consider the time-evolving functionality of O/D locations, i.e., *dynamic* POI context. Furthermore, for the problem of trip purpose prediction, POI context should be able to reveal which type of POI (activity) is more attractive to the passenger.

In light of the above perspectives, we selected the POI check-in data nearby the O/D location within a radius of r meters. According to the studies of land-use buffer for human trips [6], we set r to 250 meters after a few tests. In terms of the *dynamic* POI context, we employ the period popularity of POIs to depict the dynamic functionality of O/D location, which also reveals the attractiveness of different POIs explicitly. Specifically, based on the check-in data CI , we compute the total times that the k -th POI category had been checked during the given time period T in history, i.e. $|CI|_k^T$. Then, the *period popularity* of the k -th POI category is formulated as Eq. 3.

$$PP(k) = -\log_2 \left[1 - \left(\frac{|CI|_k^T}{\sum_{k \in K} |CI|_k^T} \right) \right] \quad (3)$$

where K denotes the number of all POI categories (i.e., $K = 9$). In order to imply the finished activity of the passenger, for the origin location, we set $T \in [t_o - 2, t_o]$ to 2h before the hour time of trip starting, since human activities commonly last less than 2 hours [8]. For the destination location, we set $T \in [t_d, t_d + 2]$ to 2h after the passenger gets off the vehicle.

In terms of the *static* POI context, we extract the *distance* and *uniqueness* of each POI category to characterize the POI distribution around trip's O/D location. As we all know, passengers always choose pick-up/drop-off points as close as possible to their departing/heading locations, so that those closer POIs should be assigned with more weights. Thus, for the k -th POI category, we compute the ratio of the minimum distance between POIs and the drop-off point l_d , as shown in Eq. 4.

$$Dis(k) = -\log_2 \left(\frac{\min(\text{distance}(\text{POIs}^k, l_d))}{r} \right) \quad (4)$$

Note that a POI category may be wrongly ranked more popular than the other just due to the fact that the number of the corresponding POIs is bigger than others. To alleviate this problem, we further employ the

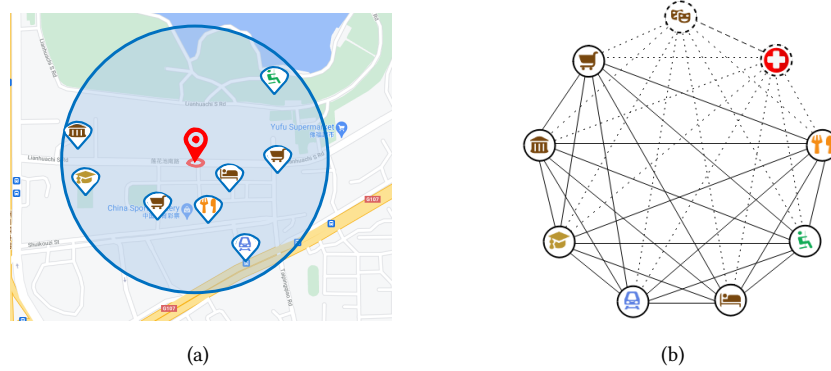


Fig. 2. Illustration of how to convert the POI context into the graph structure: a) The POI context near a pick-up/drop-off location (red landmark). The radius of the circle is 250 meters. b) The corresponding graph structure. 9 nodes represent 9 POI categories, and those nodes drawn with dotted lines refer to nonexistent POI categories within the marked circle.

uniqueness feature of the k -th POI category, as shown in Eq. 5.

$$Uniq(k) = -\log_2 \left(\frac{|POIs^k|}{\sum_{k \in K} |POIs^k|} \right) \quad (5)$$

In summary, for a origin/destination location, at most 9 categories of human activities are revealed based on the POI check-in data. Each of them is augmented with three discriminative features, namely, *period popularity*, *distance* and *uniqueness*.

4.3 Dual-Attention Graph Embedding for Trip's Activity Semantics Extraction

Here, we elaborate on the details of the dual-attention graph embedding that can extract the latent activity semantics underlying trip contexts. Generally, it consists of three stages, namely *graph construction*, *POI semantics extraction* and *trip semantics extraction*, detailed as follows.

4.3.1 Graph Construction. After the augmentation, each POI category is individually represented by its own features. As mentioned before, some human activities are usually associated with each other (e.g., “Recreation” and “Shopping”). Hence, it is also important to model the inherent correlations between different POI categories for activity semantics extraction.

Inspired by the idea that employs the graph representation learning to model the transition patterns of different driving states for driving behaviours understanding in [9], we convert the OD POI contexts into two graphs. As illustrated in Fig. 2 (a), there are 7 different categories of POIs within a 250-meter radius near the pick-up/drop-off point. This POI context is represented by the graph in Fig. 2 (b). The undirected completed graph is defined as $G = (V, E)$, where V is a set of nodes representing 9 POI categories (2 nodes drawn with dotted lines refer to the nonexistent POI categories within the marked circle) and E is a set of edges representing their potential correlations. In addition, each node contains its own augmented features $h \in \mathbb{R}^F$ (i.e., *period popularity*, *distance* and *uniqueness*), and $F = 3$ denotes the dimension of node features. As a result, the OD POI contexts can be represented by G_o and G_d , respectively.

4.3.2 GAT for POI Semantics Extraction. Based on the constructed graphs, we are able to extract and aggregate the twofold activity semantics of OD POI contexts, namely the augmented features and the neighboring semantics.

Graph Neural Networks (GNNs) are commonly adopted to extract high-level features from graph-structured data sources, such as social media. In this study, the OD POI contexts are converted into the graph structure, i.e., G_o and G_d . Nevertheless, G is an arbitrarily structured graph in the real world. Some nodes in G may not exist, since a location cannot always have all 9 categories of POIs nearby, as illustrated in Fig. 2. To solve the dilemma, we adopt the *GAT* which has great performance in inductive tasks [36]. *GAT* is able to model the non-identical correlations of neighboring nodes to the central node, and accordingly extract the high-level node representation. Specifically, it adopts the attention mechanism to learn attention coefficients between the central node u and its neighboring nodes N_u . Generally, the coefficient α_{uv} measures the correlation between u and a neighbor $v \in N_u$, which can be computed based on the *softmax* function as presented in Eq. 6.

$$\alpha_{uv} = \frac{\exp(g(\mathbf{att}^T [\mathbf{W}h_u \parallel \mathbf{W}h_v]))}{\sum_{n \in N_u} \exp(g(\mathbf{att}^T [\mathbf{W}h_u \parallel \mathbf{W}h_n]))} \quad (6)$$

where $\mathbf{W} \in \mathbb{R}^{F' \times F}$ is a shared weight matrix which linearly transforms the input node features h into higher-dimension features. Besides, \parallel and \cdot^T represent the concatenation and transposition operations. The attention mechanism is a single-layer feedforward neural network, parametrized by a weight vector $\mathbf{att} \in \mathbb{R}^{2F'}$, applying the activation function $g(\cdot)$, i.e., *LeakyReLU*.

We can find that *GAT* is capable of extracting the neighboring semantics of each POI category by modelling their mutual correlation with an attention mechanism. However, as shown in Eq. 6, for a central node u , its attention coefficients with different neighbors in N_u are computed with the same parameters, i.e., the coefficients are simply determined by their numerical values. It means that if different neighbors have the same value, their coefficients to u would also be the same. However, in the real world, different kinds of 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. Hence, the attention computations for neighboring nodes need to further consider their inherent differences. To that end, we modify the attention mechanism in the existing *GAT* to be *category-aware*, so as to deal with the POI context G . In addition, the correlation between different activities also demonstrates a time-dependent characteristic, so that the computation of attention coefficients should further account the time features \mathbb{T} (i.e., day type and hour time feature). Consequently, the Eq. 6 is rewritten as Eq. 7.

$$\alpha_{uv} = \frac{\exp(g(\mathbf{att}_{uv}^T \mathbf{W}h_u + \mathbf{att}_1^T \mathbf{W}h_v + \mathbf{att}_2^T \mathbb{T}))}{\sum_{n \in N_u} \exp(g(\mathbf{att}_{un}^T \mathbf{W}h_u + \mathbf{att}_1^T \mathbf{W}h_n + \mathbf{att}_2^T \mathbb{T}))} \quad (7)$$

where $\mathbf{att}_{uv} \in \mathbb{R}^{F'}$ is a *unique* weight matrix of the center node u towards a specific neighbor v , so that the modified *GAT* could learn category-aware correlations between u and different neighbors. Since there are K different POI categories and each of them has $K - 1$ neighbors, there are $K * (K - 1)$ *unique* matrices in total. Different from that, $\mathbf{att}_1 \in \mathbb{R}^{F'}$ and $\mathbf{att}_2 \in \mathbb{R}^{|TYP|+|H|}$ are *shared* attention weight matrices for different neighbors and time features. Then, attention coefficients are used to combine the neighbors’ features in a weighted sum manner, and the result is served as the neighboring semantics of u . In addition, we also adopt the multi-head mechanism to increase *GAT*’s expressive capability meanwhile stabilize the learning process. Specifically, a total of M independent attention mechanisms are used to extract neighboring features from different perspectives. Then those extracted neighboring features are concatenated and transformed into the final neighboring feature \vec{h}_u , as shown in Eq. 8.

$$\vec{h}_u = \mathbf{W}' \left(\left\| \sum_{m=1}^M \sigma \left(\sum_{v \in N_u} \alpha_{uv}^m \mathbf{W}^m h_v \right) \right\| \right) \quad (8)$$

where α_{uv}^m and W^m are the attention coefficient and linear transformation weight matrix of the m -th attention mechanism. 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. In this study, the graph attention networks contains two stacked multi-head GATs. Each of them is following the computation in Eq. 7 and Eq. 8.

By performing the GATs, each POI category in G_o and G_d could obtain its neighboring activity semantics. In particular, according to the computation in Eq. 7, the neighboring semantics of a POI category can be viewed as being derived from its own perspective. In this respect, for each POI category, we aggregate its augmented features and neighboring semantics (i.e., $h'_u = [h_u \| \tilde{h}_u]$, $h'_u \in \mathbb{R}^{F+F'}$). Finally, we obtain the POI context with twofold activity semantics at the origin and destination respectively (i.e., G'_o and G'_d).

4.3.3 Soft-Attention for Trip Semantics Extraction. In this stage, three kinds of activity semantics (from OD POI contexts and spatiotemporal context) are aggregated to derive the comprehensive semantics of passenger's trip. However, the aggregation is not straightforward. On one hand, among three trip contexts, the D POI context is more important in the trip purpose prediction, since destination is where a passenger takes the final activity. On the other hand, POIs are basic units of human activities, thus POI categories are with different contributions to the passenger's activity (i.e., trip purpose).

Soft-attention can be described as mapping a *query* and a set of *key-value* pairs to an output, where *query* and *keys* are from different domains [35]. The output is the weighted sum of *values*, where the weight (i.e., contribution) for each *value* is 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, *soft-attention* is adopted to extract the comprehensive activity semantics from three kinds of trip contexts meanwhile modelling their dependency on the trip purpose. The *query* is the combination of origin activity semantics G'_o and trip's spatiotemporal cost C_{st} . The *keys* are equal to *values*, which consist of the activity semantics of POI categories in the destination, i.e., $h'_u \in G'_d$. Specifically, we first employ a *Flatten* operation to convert G'_o into a 1-dimension vector, and concatenate it with C_{st} . Then, we use a full connected layer (parameterized by $\mathbf{W}^{fc1} \in \mathbb{R}^{F_{ost} \times (K|h'_u| + |C_{tr}|)}$ and \mathbf{b}^{fc1}) to fuse the features to serve as the *query* of *soft-attention*, as shown in Eq. 9.

$$h_{ost} = \tanh \left(\mathbf{W}^{fc1} \left[\text{Flatten}(G'_o) \| C_{st} \right] + \mathbf{b}^{fc1} \right) \quad (9)$$

A feed-forward network with a single hidden layer is employed as the compatibility function. Then, the coefficient $\hat{\alpha}_u$ of a POI category $u \in G'_d$ (i.e., *key*) can be obtained according to Eq. 10.

$$\hat{\alpha}_u = \frac{\exp(\tanh(\mathbf{att}_q^T h_{ost} + \mathbf{att}_k^T h'_u + \mathbf{b}))}{\sum_{s \in V} \exp(\tanh(\mathbf{att}_q^T h_{ost} + \mathbf{att}_k^T h'_s + \mathbf{b}))} \quad (10)$$

where $\mathbf{att}_q \in \mathbb{R}^{F_{ost}}$ and $\mathbf{att}_k \in \mathbb{R}^{|h'_u|}$ are learnable parameters for *query* and *key*, respectively. \mathbf{b} is the bias, and \tanh is a nonlinear activation function. h' refers to the node's features in G'_d . We also employ the multi-head attention to compute coefficients from multiple perspectives. At last, the coefficients are used to aggregate the activity semantics of different POI categories in G'_d , which is viewed as the comprehensive activity semantics \mathcal{H} of passenger's trip. The computation is shown in Eq. 11.

$$\mathcal{H} = \mathbf{W}'' \left(\left\| \sum_{m'=1}^{M'} \sigma \left(\sum_{u \in V} \hat{\alpha}_u^{m'} h'_u \right) \right\| \right) \quad (11)$$

where M' denotes the number of attention heads, and $\hat{\alpha}^{m'}$ denotes the learned coefficient in the m' -th attention. $\mathbf{W}'' \in \mathbb{R}^{|h'_u| \times M' |h'_u|}$ is a learnable parameter matrix that transforms the concatenated features into $|h'_u|$ dimensions.

4.4 Classification

The prediction of trip purpose is viewed as a classification task among several candidate trip purposes \bar{A} . In this study, a fully connected layer with *softmax* function is adopted as the classifier to output the probabilities of candidates. Specifically, on top of the extracted activity semantics \mathcal{H} , the fully connected layer with $|\bar{A}|$ neurons is first used to output the raw results z , as shown in Eq. 12.

$$z = \mathbf{W}^{fc2} \mathcal{H} + \mathbf{b}^{fc2} \quad (12)$$

where \mathbf{W}^{fc2} and \mathbf{b}^{fc2} are learnable parameters of the fully connected layer. Then, the probability \hat{p} of the i -th candidate activity \bar{a}_i being the purpose of a trip tr , can be obtained by performing the *softmax* function, as shown in Eq. 13. At last, the prediction result \hat{y} is the candidate activity with the highest probability.

$$\hat{p}(y = \bar{a}_i | tr) = \frac{\exp(z_i)}{\sum_{j=1}^{|\bar{A}|} \exp(z_j)}, \quad (z_i, z_j) \in z$$

$$\hat{y} = \arg \max_i \hat{p}(y = \bar{a}_i | tr)$$
(13)

The loss function of this network is based on the *cross-entropy*, which computes the distance between the predicted probability distribution and the actual probability distribution. The overall cost function is shown in Eq.14.

$$L = -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^{|\bar{A}|} y_i^{(j)} \log(\hat{p}_i^{(j)}) \quad (14)$$

where N denotes the number of samples. $y^{(j)}$ and $\hat{p}^{(j)}$ corresponds to the actual probability and predicted probability of the j -th candidate activity, respectively.

5 EVALUATION

5.1 Experimental Setup

5.1.1 Data Preparation. Our experiments are conducted based on two real-world datasets. In this study, the typical application scenario is taxi trip. However, people usually won't reveal their purpose information in taxis, so that there is no sufficient labeled taxi trips for our supervised learning method. Fortunately, Shenzhou UCar is also a door-to-door ride service and similar to taxis on many aspects [20], while it further possesses the information of passengers' trip purposes in the generated orders. Hence, the large-scale UCar data is used as the labeled dataset to evaluate our prediction model. Note that in all experiments, models' predictions only utilize the vehicle's GPS data (i.e., without any other information from UCar), so that the evaluation observations are able to be generalized to other door-to-door ride services like taxi trips.

Beijing UCar Trajectory Data. This data contains 780,494 vehicle trips collected by Shenzhou UCar in the city of Beijing, China, December 2015. Each record was generated when an anonymous and arbitrary passenger completed a trip with the RoD service. It contains the GPS information of pick-up&drop-off points on roads, and the description of a POI where this passenger actually heads for (e.g., *Beijing Restaurant* or *Beijing Tiantan Hospital*). Such a description intuitively reveals the activity type for this trip which is served as the passenger's trip purpose in this dataset (e.g., "Dining" or "Health"). The detailed mapping operations can be found in Appx. A.

Jiebang POI Check-in Data. This data was generated by over 11,080 users with *Jiebang* APP in Beijing from August 2011 to September 2012. It contains 511,133 POI check-ins, and each record contains an anonymous user ID, a check-in timestamp and the corresponding POI information (i.e., longitude&latitude, category and name of the POI).

There is a time misalignment issue between these two datasets. We argue that this issue reflects an objective challenge of urban computing: *it is usually difficult or even impossible to find perfect data sources with the exact time consistency in practice*. In this study, the POI check-in data is adopted to reveal the human activity semantics at different areas, while it is observed that the POIs and human activities in Beijing are relatively stable: 1) Most functional regions in Beijing are slightly changed per year [42]; 2) The overall spatial patterns of the “Restaurant” distribution are basically unchanged in two years [40]. Thus, using datasets with a 3-4-year separation in Beijing is relatively effective for our study. Moreover, our model focuses on the high-level ratios of different POI categories, which is relatively durable to the time misalignment issue [37].

Finally, we select 366,783 purpose-labelled trajectories in a square area around the Five-Ring of Beijing city, and divide them into the training, validation and test datasets at a ratio of 6 : 1 : 1.

5.1.2 Baseline Algorithms and Evaluation Metrics. Comparison experiments are conducted to evaluate the performance of our model, and several methods in existing studies are employed as baseline algorithms. Note that in this paper, the evaluation of these methods is based on the same data sources as ours (i.e., trajectory and POI check-in data).

- *Nearest*: A rule-based algorithm used in [3]. It simply sets the POI that is closest to the drop-off location as the ultimate destination of the passenger. Thus, the human activity related to that POI is served as the predicted trip purpose.
- *Bayes’s Rule*: A probability-based algorithm used in [16]. It considers a set of spatial and temporal rules to calculate the visiting probabilities of POIs near the drop-off point. Finally, the human activity related to the most likely POI is served as the predicted trip purpose.
- *Artificial Neural Network (ANN)*: A machine learning algorithm used in [41]. It is an artificial neural network with several hidden layers. The inputs are trip characteristics, including the day type and the land-use of trip’s end which is derived from the nearby POI categories (binary coding for each category). The outputs are a set of probabilities for candidate trip purposes.
- *Random Forest (RF)*: A machine learning algorithm used in [11]. The input variables include the nearby place characteristics (i.e., the percentages of different POI categories) and time characteristics (including day type and time period of a day). The outputs are also a set of probabilities for candidates.

Furthermore, we conduct an ablation study to evaluate the effectiveness of four important components in our model.

- *Ours-GATs*: Ablate the *GATs* component in the model to evaluate the effectiveness of considering the neighboring semantics of POI categories in OD POI contexts.
- *Ours-G’_o*: Ablate the POI context at the origin location to evaluate the effectiveness of considering the activity semantics before the passenger starts this trip.
- *Ours-C_{st}*: Ablate the spatiotemporal context when computing the attention coefficients in *GATs* and *soft-attention*, to evaluate its effectiveness in modelling the correlation between different POI categories.
- *Ours-S_Att*: Ablate the *soft-attention* component to evaluate its effectiveness in extracting the comprehensive trip activity semantics from three kinds of trip contexts.

To compare the performance of different algorithms, we adopt four commonly used metrics in the following groups of experiments, including *accuracy*, and macro-averaged *precision*, *recall*, *F₁-score*. Specifically, accuracy is a ratio of correctly predicted samples to the total samples. Besides, for the prediction results of *i*-th class, the precision and recall can be computed according to its false positive (FP) rate and true positive (TP) rate as shown in Eq. 15. *F₁-score* is the harmonic mean of precision and recall, and it is a more generic metric to the evaluation on an uneven class distribution. For our multi-class classification task, we further employ the *macro precision*, *macro recall* and *macro F₁-score* (i.e., the arithmetic mean of class-wise precision/recall/*F₁-score* as shown in

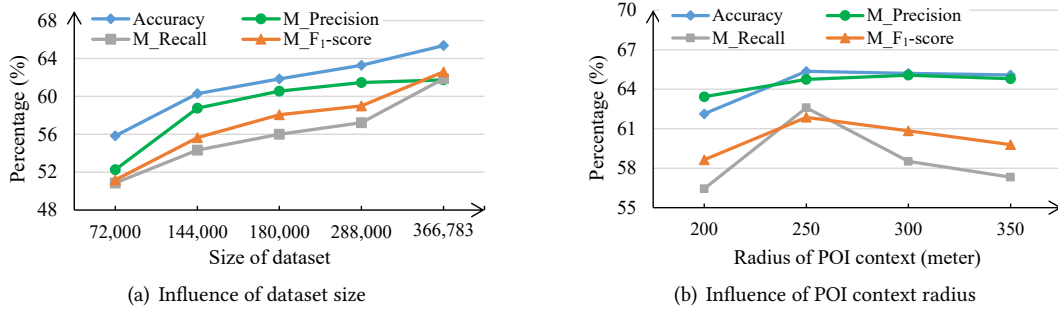


Fig. 3. Performance of our trip purpose prediction model when varying two parameters.

Eq. 16), thereby all classes can be treated equally to evaluate the overall performance of models.

$$Precision_i = \frac{TP_i}{TP_i + FP_i}, \quad Recall_i = \frac{TP_i}{TP_i + FN_i}, \quad F_1\text{-score}_i = \frac{2 * Precision_i * Recall_i}{Precision_i + Recall_i} \quad (15)$$

$$M_Precision = \frac{\sum_i^N Precision_i}{N}, \quad M_Recall = \frac{\sum_i^N Recall_i}{N}, \quad M_F_1\text{-score} = \frac{\sum_i^N F_1\text{-score}_i}{N} \quad (16)$$

5.1.3 Evaluation Environment and Parameter Settings. All the experiments are programmed using Python 3.7 with TensorFlow-2.0, and running on a PC with 4 NVIDIA GeForce RTX 2080 Ti GPU and 192 GB RAM.

For models in the following experiments, their hyperparameters are selected by comparing the performance of different groups of settings on the validation dataset with an early stopping method. For both of our model and ANN, we employ the Adam to optimize the loss function with a learning rate $l_r = 0.0001$. The batch size and L2 regularizer parameter are set to 64 and 0.0001, respectively. Additionally, for our model, the dimension of feature transformation F' in GATs is set to 50, and the number of heads for GATs and soft-attention (M, M') is set to 10 and 20, respectively. The dimension of fused origin POI context and spatiotemporal context F'_{ost} is set to 50. The ANN model consists of two hidden layers, in which the number of layer's neurons is set to 200. For the RF model, the number of decision trees is set to 1000, and the voting mechanism chooses the most popular prediction from all the decision trees.

5.2 Parameter Sensitivity Study

In this section, we investigate the impacts of two important parameters on our model's performance, namely the size of dataset and radius of POI context.

Most of existing studies have less than 100,000 samples in their datasets, while our dataset is much bigger which has over 300,000 samples. Here, we investigate whether a large-scale dataset is significant for our model's performance. Fig. 3 (a) shows the performance of our model when training on the increasing dataset. We can find that the performance improves a lot with the growth of dataset. Specifically, with regard to all the metrics, the model's performance achieves around 10% improvement when the dataset grows from 72,000 to 366,783. Thus, training our model on a large-scale dataset is beneficial for the performance.

The radius of POI context is used to select POI check-in data nearby the O/D location. We investigate how the radius affects our model by increasing this value from 200 meters to 350 meters. As shown in Fig. 3 (b), we can find the model achieves the best performance at a radius of 250 meters. In addition, when this value grows from 200 meters to 250 meters, the model's performance changes fast. It is because the passenger's activity place may

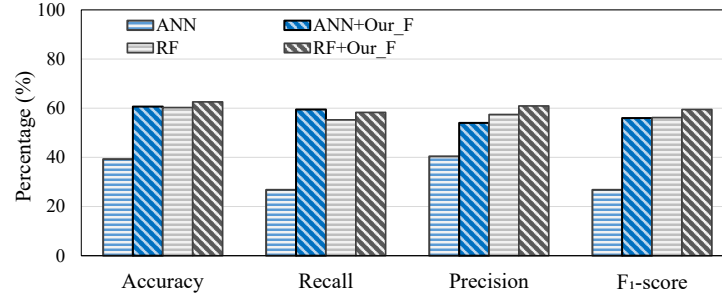


Fig. 4. Prediction results of ANN and RF models with the time and POI features in [41] and [11], and with our augmented temporal and POI features.

not be included in the small range of POI context. When the radius keeps growing, the impact would gradually decrease, since the POI context would include more noise information.

5.3 Effectiveness of Trip Context Augmentation

Trip context augmentation is used to augment the primary semantic meaning of passenger's trip. Specifically, *{day type, hour time, travel time and travel distance}* are employed to represent the trip's spatiotemporal context, and *{period popularity, distance, uniqueness}* are employed to represent the OD POI contexts. Those augmented trip contexts are served as the input features of our deep model. As mentioned before, baseline algorithms *RF* [11] and *ANN* [41] also adopt the temporal and POI context for trip purpose prediction, but with different feature engineering. Readers can refer to [11, 41] for more details. Thus, to highlight the effectiveness of our trip context augmentation, we further compare the performance of these two baseline algorithms (i.e., *ANN*, *RF*) with their original features and with our augmented features (i.e., *ANN+Our_F*, *RF+Our_F*), respectively.

Figure 4 shows that both algorithms can improve their performance by using our augmented time and POI features. Especially the prediction accuracy, macro recall and macro F_1 -score of *ANN* model have increased more than 20%. Such results demonstrate that our feature engineering (i.e., trip context augmentation) is more discriminative and effective in predicting trip purpose. Besides, the improvement of *RF* model's performance is relatively insignificant. It might be because this model's original features in [11] contain the percentage of POI categories information, which is quite similar to the *uniqueness* essentially. Hence, we can also conclude that *ANN* is more sensitive to the richness of input features, and *RF* model is competitive with *ANN* in predicting trip purpose.

5.4 Effectiveness of Dual-Attention Graph Embedding Model

To evaluate the effectiveness of our dual-attention graph embedding model, we compare its performance with four different kinds of baseline algorithms, namely *Nearest*, *Bayes' rule*, *ANN* and *RF*. Besides, in addition to the 9-class trip purposes in Tab. 1, we further evaluate the model's performance on fewer candidates. In reality, there is usually no need to predict all kinds of trip purposes for applications, and the neural network could show better performance when dealing with fewer candidates. In this paper, we take the in-car advertising for instance, and it is recognized that a few ad categories work well, namely *entertainment and sports*, *shopping*, *food and restaurants*. Thus we set the number of candidates to 4 (i.e., "Recreation, Shopping, Dining, Others"). Note that this number is not fixed, and developers can specify candidates according to their applications.

Table 2 presents the prediction results of different algorithms with the GPS trajectory and public POI check-ins. We can find that:

Table 2. Prediction results of different kinds of trip purpose prediction algorithms.

Comparison Algorithms	Accuracy (%)		M_Precision (%)		M_Recall (%)		M_F1-score (%)	
	9-class	4-class	9-class	4-class	9-class	4-class	9-class	4-class
Nearest	26.08	39.05	32.75	34.6	26.56	37.88	24.76	30.32
Bayes' rule	35.04	50.73	38.50	39.31	34.62	44.78	33.03	38.73
ANN	39.3	69.59	40.41	67.58	26.77	28.28	26.83	26.65
RF	60.26	77.57	57.41	65.34	55.27	60.05	56.2	62.31
Ours	65.63	79.76	61.74	70.1	62.59	60.97	61.86	64.41

Table 3. Confusion matrix analysis for our prediction model.

Trip Purposes	Predicted Results									Recall (%)	F ₁ (%)
	Recreation	Outdoors	Shopping	Dining	Education	Transportation	Homing	Health	Working		
Recreation	810	51	171	138	57	126	150	45	138	48.04	51.25
Outdoors	15	483	45	45	0	66	126	27	81	54.39	47.63
Shopping	81	60	4413	255	30	408	486	51	591	69.24	66.83
Dining	132	99	462	2901	54	441	531	102	498	55.57	59.8
Education	21	39	105	75	846	84	309	78	153	49.47	52.08
Transportation	180	168	486	399	186	5487	870	171	1083	60.76	65.77
Homing	90	123	486	333	210	405	5097	171	588	67.93	64.13
Health	18	21	36	30	21	63	102	1971	69	84.56	77.48
Working	126	96	630	306	135	576	723	141	7521	73.35	71.71
Precision (%)	54.99	42.37	64.59	64.73	54.97	71.67	60.72	71.49	70.15		

- **Our model outperforms baseline algorithms.** Our model shows a considerable improvement in predicting trip purpose. Especially on the 9-class trip purposes, it achieves a lead over 5% regarding the accuracy, macro recall and macro F₁-score.
- **Machine learning algorithms are better.** Not surprisingly, machine learning algorithms (i.e., *Ours*, *RF*, *ANN*) outperform the probability-based and rule-based algorithms (i.e., *Bayes' rule*, *Nearest*). It is because human activities are complicated and associated with many factors, while machine learning algorithms are more capable of dealing with such tasks in a data-driven manner.
- **The proposed dual-attention neural network is effective.** Both the *ANN* and our model are neural networks, but the *ANN* model performs much worse. It is because *ANN* simply aggregates all inputs in the latent space, while our model with two attention mechanisms, carefully models the inherent correlations of features in the latent space. Such results further demonstrate the significance of using “dual-attention” to extract activity semantics from trip contexts.
- **Our prediction model is applicable.** Our model could achieve 64.57% prediction accuracy on the 9-class trip purposes and 79.76% accuracy on the 4-class purposes. Such results show that our prediction model is generally applicable for real-life applications.

Table 3 presents the confusion matrix of our model on the test dataset. Each row shows the predicted results of a set of trips (belong to the same trip purpose). Each column shows the actual distribution of trips' labels which are predicted to be a given trip purpose. We can observe that the F₁-scores of “Health” and “Working” purposes are over 70%, while the “Recreation” and “Outdoors” are about 50%. It might be because the POI configurations near the “Health” and “Working” activities are usually simpler than “Recreation” and “Outdoors”. Otherwise,

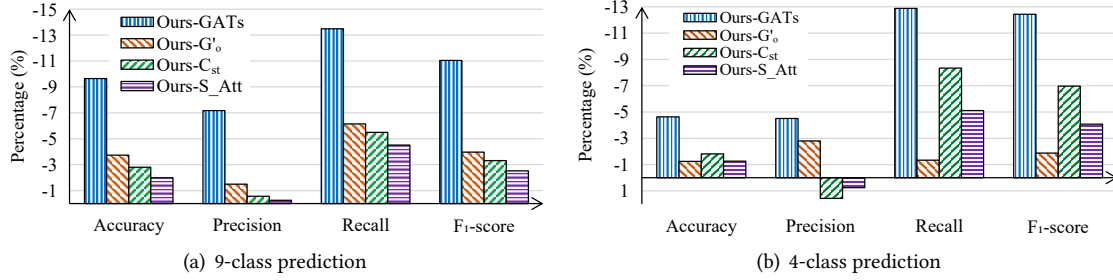


Fig. 5. Performance degradation of our trip purpose prediction model with different ablation settings.

some human activities are usually associated with each other at time and space, so that the model can not distinguish them very well. As we can see in the table, the “Recreation” purposes are more likely to be predicted as “Shopping”, since their corresponding POIs often appear together in streets. Similarly, many “Transportation” purposes are wrongly predicted as the “Working” purposes. We also observe the F₁ score of “Homing” is around 65%, which is not as good as our expectation. It might be because our model employs the check-in data from a social network to reveal human activities in the city, but relatively few users would leave a check-in record for the “Homing” activity. Consequently, the POI context for “Homing” purposes may be biased in some cases.

5.5 Ablation Study

The dual-attention graph embedding model is carefully designed to extract the activity semantics underlying the passenger’s trip for trip purpose prediction. Here, we conduct an ablation study to evaluate the effectiveness of different components and features, namely the *graph attention networks*, *soft-attention*, POI context at the origin location and trip’s spatiotemporal context.

The results are shown in Fig. 5. We can find that:

- **Graph attention networks play the most important role.** When ablating the *GATs* (i.e., *Ours-GATs*), the model’s performance degrades dramatically (e.g., the macro F₁-score is reduced by 10% on both the 9-class and 4-class prediction). In this study, *GATs* are employed to capture the twofold activity semantics of each POI category by modelling their neighboring correlations. Such results also demonstrate the effectiveness of converting POI context into the graph structure, and the significance of modelling the correlation between POI categories for the trip purpose prediction.
- **Considering the origin POI context is necessary.** The performance of *Ours-G'o* degrades on both two groups of prediction tasks. However, the degradation in Fig. 5 (b) is not as much as that in Fig. 5 (a), especially on the prediction recall. It indicates that the origin POI context is not very discriminative for 4-class trip purposes (i.e., “Recreation, Shopping, Dining, Others”). Such results might be because people’s origin locations for those human activities are relatively random compared with other activities. For example, people usually head for “Working” from residential areas, but for “Dining”, the origin could be their home, working place or any recreation facilities.
- **Spatiotemporal context helps improve the prediction recall.** As for the prediction results of *Ours-Cst*, the prediction macro recall is the most degraded among four metrics. Such results indicate more trip purposes can be correctly predicted by using the spatiotemporal context especially on the 4-class trip purposes (the macro recall is reduced over 8%). The spatiotemporal context includes the time information and travel cost. Thus, those time-dependent human activities would be easier to be correctly predicted, such as “Homing”, “Working” and “Dining”. On the contrary, activities like “Recreation” are usually with



Fig. 6. A use case of our trip purpose prediction model in the real-life scenario.

unfixed time period, and the spatiotemporal context might lead to the false prediction (i.e., the reversed precision).

- **Soft-attention is effective in aggregating trip contexts for activity semantics extraction.** In our model, the *soft-attention* is employed to aggregate three kinds of trip contexts to derive trip's comprehensive activity semantics. While in *Ours-S_Att*, different trip contexts are directly fused by full connected layers. As we can see, the overall performance of *Ours-S_Att* also degrades on both the 9-class and 4-class prediction, especially the macro recall shows a gap of 5%. It is because the *soft-attention* further models the dependency of different trip contexts on the passenger's trip purpose, and that is important for the prediction.

5.6 Case Study

Figure 6 illustrates a use case of our trip purpose prediction model in the real-life scenario. The taxi is equipped with an advertising system including a GPS device and our prediction model. It picked up a passenger on non-workday and traveled 4.9 kilometers in 14 minutes. During this trip, *the in-car system treats the passenger as a black box, and only senses the taxi's moving trajectory on roads*. When the taxi stops at the drop-off point at 3 PM, the trip purpose prediction model is triggered immediately. According to the time, travel cost and POI contexts, the prediction model outputs 4 probabilities of candidate trip purposes within 0.31 seconds. As shown in the figure, the "Shopping" purpose is with the largest probability. It indicates the passenger is most likely going to perform the shopping activity. Accordingly, the in-car advertising system presents this passenger some discount information (i.e., coupons for the nearby shopping mall) before he/she gets off the taxi. The whole response time is less than 0.5 seconds. It should be noted that the presented running times are tested on the desktop computer.

This case study demonstrates the model's feasibility in practice, also shows that our system protects the passenger's privacy from two aspects: 1) It has no connection with the passenger in the digital space, and does not require cooperative efforts; 2) It does not record or use any passenger's personal information, especially the identity, thus it cannot link the predicted trip purpose to a specific individual (i.e., no privacy exposure issue).

6 CONCLUSIONS AND FUTURE WORK

This paper presents a new deep model for the passenger's trip purpose prediction, aiming at offering a more *ubiquitous* and *applicable* approach in the scenarios of door-to-door ride services. To that end, the proposed model only utilizes the passenger's insensitive information for prediction. Specifically, the *vehicle's GPS trajectory on roads* and *public POI check-in data* are first aggregated to augment the semantics meaning of trip contexts (i.e., spatiotemporal context and OD POI contexts). Based on that, a dual-attention graph embedding network (i.e.,

graph attention networks and soft-attention) is established to extract the comprehensive activity semantics of passenger's trip for the trip purpose prediction. The model is trained on a large-scale labeled travel dataset in a supervised way. At last, extensive experiments demonstrate the model's great performance, and the case study shows its feasibility in the real deployment environment.

Generally, our trip purpose prediction is based on modelling the high-level human activity semantics, while the nature of human activities is similar between different cities [19]. Thus our approach is able to be generalized to other cities. Additionally, since the utilized data sources are relatively pervasive in door-to-door ride services, we believe this approach is also applicable among similar ride services like taxis and UCars. Nevertheless, our deep model is trained in a supervised manner, while many cities or application scenarios only have limited labeled data in reality. Thus the generalization may not be straightforward. For these situations, new models can be obtained by performing the fine-tuning operation with a pre-trained model, or employing semi-supervised learning and active learning techniques [32].

In the future, we plan to broaden and deepen this work in several directions. Firstly, since our prediction model requires many computations while the vehicle's compute capability is usually limited, we plan to design a new platform to deploy our model in real taxi fleets. Secondly, the uneven sampling rate issue in the LBSNs check-in data could have a negative effect on our POI semantics augmentation (i.e., popularity), thus we plan to explore potential solutions to alleviate its impact. Finally, we intend to examine the extension possibility of our trip purpose prediction algorithm in door-to-door services to public transportations such as bus and metro [2], and evaluate the performance.

ACKNOWLEDGMENTS

We thank the anonymous reviewers for their constructive feedback and comments. This work is supported by the National Natural Science Foundation of China (No. 62172066, 61872050, 62002135, 61960206008), the National Science Foundation for Distinguished Young Scholars of China (No. 61825204), the Beijing Outstanding Young Scientist Program (No. BJJWZYJH01201910003011), the Science and Technology Projects in Guangzhou (No. 202102021164). This work is sponsored by DiDi GAIA Research Collaboration Plan.

REFERENCES

- [1] Laura Alessandretti, Ulf Aslak, and Sune Lehmann. 2020. The scales of human mobility. *Nature* 587, 7834 (2020), 402–407.
- [2] Azalden Alsger, Ahmad Tavassoli, Mahmoud Mesbah, Luis Ferreira, and Mark Hickman. 2018. Public transport trip purpose inference using smart card fare data. *Transportation Research Part C: Emerging Technologies* 87 (2018), 123–137.
- [3] Wendy Bohte and Kees Maat. 2009. Deriving and validating trip purposes and travel modes for multi-day GPS-based travel surveys: A large-scale application in the Netherlands. *Transportation Research Part C: Emerging Technologies* 17, 3 (2009), 285–297.
- [4] Breiman. 2001. Random forests. *Machine Learning* 45 (2001), 5–32.
- [5] Pablo Samuel Castro, Daqing Zhang, Chao Chen, Shijian Li, and Gang Pan. 2013. From Taxi GPS Traces to Social and Community Dynamics: A Survey. *ACM Computing Surveys*. 46, 2, Article 17 (2013), 34 pages.
- [6] Cynthia Chen, Hongmian Gong, Catherine Lawson, and Evan Bialostozky. 2010. Evaluating the feasibility of a passive travel survey collection in a complex urban environment: Lessons learned from the New York City case study. *Transportation Research Part A: Policy and Practice* 44, 10 (2010), 830–840.
- [7] Chao Chen, Shuhai Jiao, Shu Zhang, Weichen Liu, Liang Feng, and Yasha Wang. 2018. TripImputor: Real-Time Imputing Taxi Trip Purpose Leveraging Multi-Sourced Urban Data. *IEEE Transactions on Intelligent Transportation Systems* 19, 10 (2018), 3292–3304.
- [8] Chao Chen, Chengwu Liao, Xuefeng Xie, Yasha Wang, and Junfeng Zhao. 2019. Trip2Vec: a deep embedding approach for clustering and profiling taxi trip purposes. *Personal and Ubiquitous Computing* 23, 1 (2019), 53–66.
- [9] Chao Chen, Qiang Liu, Xingchen Wang, Chengwu Liao, and Daqing Zhang. 2021. semi-Traj2Graph: Identifying Fine-grained Driving Style with GPS Trajectory Data via Multi-task Learning. *IEEE Transactions on Big Data* (2021), 1–1.
- [10] Yu Cui, Chuishi Meng, Qing He, and Jing Gao. 2018. Forecasting current and next trip purpose with social media data and Google Places. *Transportation Research Part C: Emerging Technologies* 97 (2018), 159–174.
- [11] A. Ermagun, Y. Fan, J. Wolfson, G. Adomavicius, and K. Das. 2017. Real-time trip purpose prediction using online location-based search and discovery services. *Transportation Research Part C Emerging Technologies* 77 (2017), 96–112.

- [12] Alireza Ermagun, Yingling Fan, Julian Wolfson, Gediminas Adomavicius, and Kirti Das. 2017. Real-time trip purpose prediction using online location-based search and discovery services. *Transportation Research Part C: Emerging Technologies* 77 (2017), 96–112.
- [13] Tao Feng and Harry J.P. Timmermans. 2015. Detecting activity type from GPS traces using spatial and temporal information. *European Journal of Transport & Infrastructure Research* 15, 4 (2015), 662–674.
- [14] Barbara Furletti, Paolo Cintia, Chiara Renso, and Laura Spinsanti. 2013. Inferring Human Activities from GPS Tracks. In *Proceedings of the 2nd ACM SIGKDD International Workshop on Urban Computing*. Article 5, 8 pages.
- [15] Lei Gong, Ryo Kanamori, and Toshiyuki Yamamoto. 2018. Data selection in machine learning for identifying trip purposes and travel modes from longitudinal GPS data collection lasting for seasons. *Travel Behaviour and Society* 11 (2018), 131–140.
- [16] Li Gong, Xi Liu, Lun Wu, and Yu Liu. 2016. Inferring trip purposes and uncovering travel patterns from taxi trajectory data. *Cartography and Geographic Information Science* 43, 2 (2016), 103–114.
- [17] Lei Gong, Takayuki Morikawa, Toshiyuki Yamamoto, and Hitomi Sato. 2014. Deriving Personal Trip Data from GPS Data: A Literature Review on the Existing Methodologies. *Procedia - Social and Behavioral Sciences* 138 (2014), 557–565.
- [18] Marta C Gonzalez, Cesar A Hidalgo, and Albert-Laszlo Barabasi. 2008. Understanding individual human mobility patterns. *Nature* 453, 7196 (2008), 779–782.
- [19] Sébastien Grauw, Stanislav Sobolevsky, Simon Moritz, István Gódor, and Carlo Ratti. 2015. Towards a comparative science of cities: Using mobile traffic records in New York, London, and Hong Kong. In *Computational Approaches for Urban Environments*. 363–387.
- [20] Suiming Guo, Chao Chen, Jingyuan Wang, Yaxiao Liu, Ke Xu, Zhiwen Yu, Daqing Zhang, and Dah Ming Chiu. 2019. Rod-revenue: Seeking strategies analysis and revenue prediction in ride-on-demand service using multi-source urban data. *IEEE Transactions on Mobile Computing* 19, 9 (2019), 2202–2220.
- [21] Moritz UG Kraemer, Chia-Hung Yang, Bernardo Gutierrez, Chieh-Hsi Wu, Brennan Klein, David M Pigott, Louis Du Plessis, Nuno R Faria, Ruoran Li, William P Hanage, et al. 2020. The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science* 368, 6490 (2020), 493–497.
- [22] Cory M. Krause and Lei Zhang. 2019. Short-term travel behavior prediction with GPS, land use, and point of interest data. *Transportation Research Part B: Methodological* 123 (2019), 349–361.
- [23] John Krumm and Dany Rouhana. 2013. Placer: Semantic Place Labels from Diary Data. In *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. 163–172.
- [24] Defu Lian, Yin Zhu, Xing Xie, and Enhong Chen. 2014. Analyzing location predictability on location-based social networks. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*. 102–113.
- [25] Tongtong Liu, Zheng Yang, Yi Zhao, Chenshu Wu, Zimu Zhou, and Yunhao Liu. 2018. Temporal understanding of human mobility: A multi-time scale analysis. *PLoS ONE* 13, 11 (2018).
- [26] Yijing, Lu and Lei. Zhang. 2015. Imputing trip purposes for long-distance travel. *Transportation* 42, 4 (2015), 581–595.
- [27] Suxing Lyu and Takahiko Kusakabe. 2021. Graph-aware Chained Trip Purpose Inference. In *2021 IEEE International Intelligent Transportation Systems Conference (ITSC)*. 3691–3697.
- [28] Chuishi Meng, Yu Cui, Qing He, Lu Su, and Jing Gao. 2017. Travel purpose inference with GPS trajectories, POIs, and geo-tagged social media data. In *2017 IEEE International Conference on Big Data (Big Data)*. 1319–1324.
- [29] Lara Montini, Nadine Rieser-Schüssler, Andreas Horni, and Kay W. Axhausen. 2014. Trip Purpose Identification from GPS Tracks. *Transportation Research Record* 2405, 1 (2014), 16–23.
- [30] Minh Hieu Nguyen, Jimmy Armoogum, Jean-Loup Madre, and Cédric Garcia. 2020. Reviewing trip purpose imputation in GPS-based travel surveys. *Journal of Traffic and Transportation Engineering* 7, 4 (2020), 395–412.
- [31] Marcelo G. Simas Oliveira, Peter Vovsha, Jean Wolf, and Michael Mitchell. 2014. Evaluation of Two Methods for Identifying Trip Purpose in GPS-Based Household Travel Surveys. *Transportation Research Record* 2405, 1 (2014), 33–41.
- [32] Burr Settles. 2009. Active learning literature survey. (2009).
- [33] Li Shen and Peter R. Stopher. 2013. A process for trip purpose imputation from Global Positioning System data. *Transportation Research Part C: Emerging Technologies* 36 (2013), 261–267.
- [34] Elton F de S Soares, Kate Revoredo, Fernanda Baião, Carlos A de MS Quintella, and Carlos Alberto V Campos. 2019. A combined solution for real-time travel mode detection and trip purpose prediction. *IEEE Transactions on Intelligent Transportation Systems* 20, 12 (2019), 4655–4664.
- [35] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser, and Illia Polosukhin. 2017. Attention is All you Need. In *Advances in Neural Information Processing Systems*. 5998–6008.
- [36] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. 2018. Graph Attention Networks. In *International Conference on Learning Representations*.
- [37] Leye Wang, Xu Geng, Xiaojuan Ma, Daqing Zhang, and Qiang Yang. 2019. Ridesharing car detection by transfer learning. *Artificial Intelligence* 273 (2019), 1–18.
- [38] Pengfei Wang, Guannan Liu, Yanjie Fu, Yuanchun Zhou, and Jianhui Li. 2017. Spotting Trip Purposes from Taxi Trajectories: A General Probabilistic Model. *ACM Transactions on Intelligent Systems and Technology* 9, 3, Article 29 (2017), 26 pages.

- [39] Jean Wolf, Randall Guensler, and William Bachman. 2001. Elimination of the Travel Diary: Experiment to Derive Trip Purpose from Global Positioning System Travel Data. *Transportation Research Record* 1768, 1 (2001), 125–134.
- [40] Mingbo Wu, Tao Pei, Wenlai Wang, Sihui Guo, Ci Song, Jie Chen, and Chenghu Zhou. 2021. Roles of locational factors in the rise and fall of restaurants: A case study of Beijing with POI data. *Cities* 113 (2021), 103185.
- [41] Guangnian Xiao, Zhicai Juan, and Chunqin Zhang. 2016. Detecting trip purposes from smartphone-based travel surveys with artificial neural networks and particle swarm optimization. *Transportation Research Part C: Emerging Technologies* 71 (2016), 447–463.
- [42] Jing Yuan, Yu Zheng, and Xing Xie. 2012. Discovering regions of different functions in a city using human mobility and POIs. In *Proceedings of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 186–194.
- [43] Kuan Zhang, Jianbing Ni, Kan Yang, Xiaohui Liang, Ju Ren, and Xuemin Sherman Shen. 2017. Security and Privacy in Smart City Applications: Challenges and Solutions. *IEEE Communications Magazine* 55, 1 (2017), 122–129.
- [44] Xueliang Zhao, Zhishuai Li, Yu Zhang, and Yisheng Lv. 2020. Discover Trip Purposes from Cellular Network Data with Topic Modelling. *IEEE Intelligent Transportation Systems Magazine* (2020).

A DATA LABELLING FOR UCAR TRIPS

As a ride-on-demand service, UCar further records the POI descriptions of passengers’ actual destinations (e.g., *Beijing Restaurant*) in orders. Such descriptions intuitively reveal the passengers’ trip purposes. Here, we elaborate on how to automatically label UCar trips with the corresponding trip purposes.

The process of data labelling consists of two *mapping* operations: 1) Mapping from the destination description text to the primary POI category; 2) Mapping from the primary POI category to the true trip purpose. The details are as follows:

- (1) The first mapping is automatically accomplished by using a pre-trained NLP model (i.e., ERNIE from PaddlePaddle, which achieved SOTA results in more than 40 typical NLP tasks. <https://pypi.org/project/paddle-ernie/>). Specifically, this NLP model is fine-tuned with 510,000 text samples from the JiePang dataset, since this dataset contains both the POI descriptions in Beijing and the corresponding POI categories. On top of that, this NLP model is used to predict the POI categories of destination descriptions in the UCar dataset. To evaluate this NLP model’s performance in the UCar dataset, we manually label the destination descriptions of 1000 randomly selected trips, then compare these labels with the model’s results. After 3 testing rounds, the average prediction accuracy of this NLP model achieves around 99.3%.
- (2) The second mapping is much easier. It can be simply fulfilled by using the illustrative mapping in Tab. 1 (i.e., POI categories and the corresponding trip purposes).

B POI CATEGORY MAPPING IN JIEPANG DATASET

The POI categories in JiePang dataset are hierarchical, namely primary POI categories and subcategories. Generally, there are 8 primary categories in this dataset, namely {“Recreation and Culture Facilities”, “Outdoors and Sightseeing Places”, “Shop and Service Facilities”, “Catering”, “School and Educational Facilities”, “Transportation Facilities”, “Professional Building and Residence”, “Others”}.

For the trip purpose prediction task, we follow some distinguished related works [7, 16], to adopt the primary POI categories to indicate different kinds of human activities. At the same time, we find the “Hospital” and “Residence” are considered as important primary POI categories [7, 16, 44]. However, in JiePang dataset, both the “Hospital” and “Residence” are the subcategories of “Professional Building and Residence”. Additionally, we note that in JiePang, the “Others” check-ins are usually generated when users check-in some private POIs, which are relatively uninformative, and such check-ins are very sparse in the dataset (around 1.4%).

With these in mind, in order to keep consistent with existing works, we map the original 8 primary POI categories to 9 categories. The detailed mapping operations are listed in the following.

- (1) For check-ins with the primary category of “**Professional Building and Residence**”, we divide them into three groups by examining their subcategories. The obtained three new categories are “Apartment and

Residence”, “Hospital and Clinic”, and “Office and Business Buildings”. Note that for each check-in, the tags of primary POI category and subcategory are directly presented in the data entry.

- (2) For the “**Others**” check-ins, we directly exclude these uninformative records in this study.
- (3) For check-ins with the rest **6** primary categories, we directly adopt their original category information.