Trip purpose imputation for taxi data

Author(s):
Li, Aoyong; Axhausen, Kay W.

Publication Date:
2018-05

Permanent Link:
https://doi.org/10.3929/ethz-b-000264721 →

Trip purpose imputation for taxi data

Aoyong Li, Institute for Transport Planning and Systems
Kay W. Axhausen, Institute for Transport Planning and Systems

Conference Paper STRC 2018

STRC | **18th Swiss Transport Research Conference**
Monte Verità / Ascona, May 16-18, 2018

Trip purpose imputation for taxi data

Aoyong Li
IVT
ETH Zurich
CH-8093 Zurich
T: +41 44 633 27 19
E: Aoyong.li@ivt.baug.ethz.ch

Kay W. Axhausen
IVT
ETH Zurich
CH-8093 Zurich
T: + 41 44 633 39 43
E: axhausen@ivt.baug.ethz.ch

May 2018

Abstract

The rapid development and application of sensor networks and information and communication technologies (ICT) make it possible to obtain a large amount of longitudinal movement data in high temporal and spatial accuracy. The movement data becomes an important alternative data source to monitor transport status, manage intelligent transportation, etc. However, the lack of semantic information, especially trip purpose, is the main limitation for these passively collected data. Given that human movement data is derived from people's needs for participating in activities, it is possible and necessary to enrich this human movement data with its trip purpose. A host of related studies mostly focuses on GPS track data, mobile data and smart card data. Nevertheless, taxi trajectory data, as an important part of movement data, receives little attention. Focusing on taxi trajectory data, the paper proposes a framework, which can infer the purpose of the passenger. The framework will introduce a point-polygon contextual data repository including semantic point and semantic polygon. To infer the trip purpose, the framework will consider the location and time of the drop-off point as well as the trajectory form. Moreover, Didi data and four other algorithms will be introduced to evaluate the framework.

Keywords

Trip purpose imputation; taxi trajectory data; trip purpose

1. Introduction

During these past years, human travel behavior analysis is becoming more and more important and complex than before since modern cities are undergoing rapid growth in most of the world (Chen *et al.*, 2015). Travel-related data is an important and valuable source for understanding travel behavior. These data can help urban planners and policy makers to address urban planning, management and operating issues (Jiang *et al.*, 2016). Traditionally, travel-related data was mainly collected manually by face-to-face interview or telephone interview and so on. These methods suffer from the drawbacks such as high cost, short reporting period, space coverage, small sample sizes and low frequencies.

The rapid development and application of sensor networks and information and communication technologies (ICT) make it possible to obtain large amount of time-stamped locational data of individuals. Such data contains a wealth of travel behavior information, such as when and where the passengers travel in the city in a high resolution and sometimes on which the routes they travel. For instance, taxi trajectory data record the physical coordinates (longitudes and latitudes) and the exact times that a passenger was picked up and dropped off. Therefore, how to use GPS-based data is a hot trend. However, the collected GPS data is raw. In particular, it lacks semantic information especially the transport mode or trip purpose. That is, we do not know the trip mode and the trip purpose of a passenger. This information is important for transport planning and urban computing. Furthermore, compared to enriching the raw data with trip mode, existing method on trip purpose imputation are still far more accurate. So, there is a dilemma that trajectory is captured well but activity information is poor. Therefore, the paper attempts to narrow the gap between the raw data and human trip purpose. In particular, we focus on analyzing taxi passengers' trip purposes.

Trip purpose imputation has been a long-standing research topic (Schuessler and Axhausen, 2009; Huang *et al.*, 2010; Furletti *et al.*, 2013; Parent *et al.*, 2013; Montini, 2016;) But previous studies have rarely addressed the following two issues: 1) Prior research mainly focus on predicting trip purpose at an aggregate level. Thus only smart urban services at the macro level can be used. However, the imputation of trip purpose at the individual level is necessary because it can support micro urban services such as recommendations services according to trip purpose; 2) Limited existing research focusses on taxi trip purpose imputation. Taxi data trip imputation are more difficult than personal wearable GPS trajectory because of two reasons. Firstly, taxi data cannot record personal information which is important in existing algorithms. Secondly, taxi trajectory only record a part of human trajectory.

To enable the taxi trip purpose imputation at the individual level, we need to address the following two challenges:

More information. The existing algorithm are mainly based on the personal socio-background information and spatiotemporal of drop-off point. But for taxi data, no personal background information can be recorded. So, further more information should utilized.

Lack of ground-truth information. The ground-truth for trip purpose is usually collected by the prompted recall, where only a small fraction of users can annotate their traces correctly.

With the research objectives and challenges discussed above, the main contributions of paper are:

- We introduce a two layer context data repository into the algorithms.
- We extract new rules based on the trajectory besides the spatiotemporal attributes of drop off point.
- We collected a Didi dataset to evaluate the algorithms. The dataset can help us evaluate the algorithm at the individual level.

2. Related work

2.1 Semantic Trajectory Enrichment

With the development and application of Global Positioning System (GPS) technologies and sensor networks, the passive collection of large-scale trajectory data with both locational information and timestamps becomes easy, both technically and economically. These data come from different sources: GPS tracking from GPS-enabled floating car, the call detail records from mobile phone, smart card data from Smart Card Automated Fare Collection (SC-AFC) systems and so on. These data vary in resolutions and formats because of different location recording approaches. For example, a GPS-enabled device records the physical location of the moving object, such as taxi data; a tower based device records the physical location of the tower, such as mobile phone data and WLAN hot spot data. Because of their advantages, these data prompt different research, such as human mobility (Gonzalez, 2008; Zhao, 2018) city structure (Liu *et al.*, 2014) and so on. However, the data generally lack an explicit trip purpose (Chen *et al.*, 2018)

The activity information is important in modelling people's behaviours because travel demands are derived from people's need for participating in activities (e.g. Axhausen, 1992) How to extract high-level semantics from raw data and use them to understand why people move has attracted researchers' attention. Several algorithms has been explored to infer travel purpose in terms of travel activities after the trip. These algorithms range from rule-based approaches to advanced machine learning based approaches (Feng and Timmermans, 2016) The rule based approaches in general rely on the position of the activity, time information and context data (point-of-interests, land use data) building different rules to infer trip purpose information. The rules are mostly based on spatiotemporal attributes of drop off point. For example, the nearest distance rules (Xie *et al.*, 2009) voting rules (Phithakkitnukoon *et al.*, 2010) Some rule-based methods combine rules with some existing models, such as the gravity model (Furletti *et al.*, 2013) One of the first examples (Wolf, 2001) considers the land use data and uses a set of rules to infer trip purpose. The machine learning algorithms used mainly include probabilistic models (Furletti *et al.*, 2013; Gong *et al.*, 2016) fuzzy logistic regression (Schuessler and Axhausen,

2009) random forest (Montini, 2016) and so on. To name an example, Gong (2016) infers taxi purpose based on Bayes' rules by considering the time and location of drop off point.

Activity information imputed from trajectory can be applied to GPS travel survey (Kim, 2015) recommendation (Chen *et al.*, 2018; Xiao *et al.*, 2012) city structure (Gong *et al.*, 2016; Wu *et al.* 2014) and so on. Kim and Pereira (2015) proposed a framework, which can recognize the users' activity of travellers when their movements are tracked by mobile sensors. In terms of recommendation, Xiao *et al*. (2012) first recognize human activity from GPS trajectory data to construct human's semantic activity traces. Then they measure the users' similarity of behaviour by semantic activity history instead of physical locations. Besides, Wu *et al.* (2014) extract human trajectory from location social network (LBS) data and human activity from texts. From the result, they derived the transition probability matrix between activities at different time.

2.2 Taxi data mining

Taxi data is an important type of trajectory data, with the advantages of high positioning accuracy, long time series and large scales. The information mined from taxi data can apply to transport and urban planning and so on, benefitting various parties, including taxi drivers, passengers and city planners.

In transportation area, the traffic administration can monitor and predict traffic congestion (Liu *et al.*, 2009; Shang *et al.* (2014) and use taxi trajectory data to infer the traffic volume, energy consumption and emissions of vehicles on the road network.

In terms of urban planning, many existing studies have taken trajectory data for spatial pattern finding (Liu *et al.* 2012) Recent studies also combine taxi trajectory with other data such as POI, check-in data, to infer building functions. Yuan *et al.* (2012) proposed a framework which can discovering regions of different functions in a city using both taxi data and POIs.

Some other applications are useful for individuals. By exploiting the taxi traces, some researchers try to build different recommender systems which can help passengers find a vacant taxicab easily and let the driver know where can they find the passenger close by (Ge *et al.*, 2010; Yuan *et al.* 2014)

3. Methods

In this work, we apply a rule-based model infer the taxi trip purpose. The existing methods infer trip purpose by applying a framework integrating the spatiotemporal information of the trip end point, ignoring the importance of the context data which provide the semantic information and the trajectory. In the first part, we put more value on these two elements. Firstly, we build a two layers context framework based on POI and AOI. Then, we build the rules considering the time, spatial location and trajectory formation. Based on these rules, we can infer the possibilities of each activities for the trip.

3.1 Basic concepts and problem statement

3.1.1 Basic concept

Road network: Road network is a graph $G(N, E)$ which contains a node set N and an edge set E.

Taxi drop-off point: A taxi drop-off point is the time-stamped point where passengers drops off. It can be denoted as (x_i, y_i, t_i).

Passenger taxi trajectory: The passenger taxi trajectory is the taxi trajectory with passenger, which is consist a consecutive of GPS sample points. The last sample point of passenger taxi trajectory is the taxi drop-off point.

Point of interest (POI): A point of interest (POI) is a useful and interesting place, denoted as a point. We denote POI as $(id_i, x_i, y_i, category_i, time_i)$. POI category is a semantic label for a place, indicating potential human activities information at the place. Time is the business hour for the POI, indicating whether the POI can be visited at the drop off time.

Area of interest (AOI): Area of interest (AOI) is an area which contains many different POI. It is denoted as $(id_i, boundary_i, enterPOIs_i, innerPOIs_i)$. Compared to POI which has no concept of shape and size, AOI is denoted as a polygon. The POIs for AOI is divided into two parts: enterPOIs denoted the entrances of the AOI; innerPOIs denote the POI which can provide an activity place.

Context data repository: Context data repository is the external data sources that provides semantic information.

3.1.2 Problem statement

Taxi trip purpose imputation problem can be seem as inferring the probabilities of taking different activities when getting off the taxi. It can be formulated as:

Given:

1) A raw passenger taxi trajectory, which consist of the GPS sample points.

2) POIs and AOIs in the city concerned.

3) A road network in the city concerned.

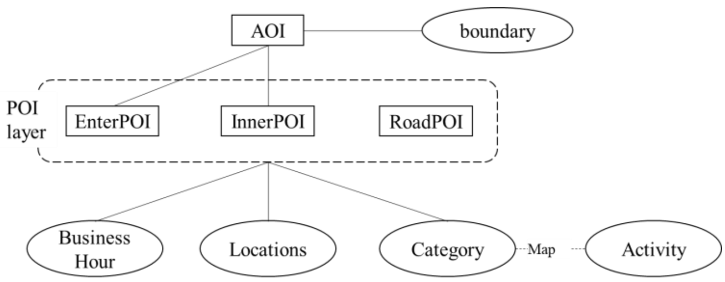
Question:

We define the question as, predicting the different activities the passengers will to do when they get off the taxi.

3.2 POI-AOI context repository

Existing research only adopts the landuse or POI in the context data repository. However, the POI information in a database is normally limited to an address and a coordinate pair. It is not easy to combine the trajectory with POI. On the other hand, the landuse is normally a mix of different activity possibilities. To overcome these two shortcoming, we combine the POI and AOI together into a two layers context repository framework. The AOI mainly refers to a building, housing estate, shopping mall, school. The framework is shown in Figure 1.

Figure 1: POI-AOI context repository framework



AOI is the first layer and POI is the second layer. POI have three types. EnterPOI means the entrance of AOI, such as the door of a building, gate of a school. InnerPOI denotes that the POI is in the AOI, such as a store in a shopping mall. RoadPOI denoted the POI which can be reached directly from the road. AOI is used to match the trajectory and POI layer provides the activity.

Because AOI is denoted as a polygon, another problem is the distance between the drop-off point and AOI. The AOI is only reached by an entrance. The distance between them is the smallest distance between the drop-off point and the nearest EnterPOI of the AOI, denoted as (1).

$$\mathrm{dis}(drop, AOI) = \min(disntace(Loc_{drop}, Loc_{EnterPOI})) \tag{1}$$

3.3 Rules building

Compared to trajectory of personal GPS trajectory, especially the GPS wearable device which can record the users full trajectory, it is more difficult to construct rules. It is mainly because two reasons: firstly, as public transport, taxi trajectory cannot record personal information which is important to recognize the travel purpose; secondly, taxi trajectory data cannot record the trajectory of the passengers after them leaving the taxi. It means we need to further consider other features to construct rules. The existing researches only consider the attributes of drop-off points. In this paper, we try to build rules from the passenger taxi trajectory besides the spatiotemporal rules for the drop-off points.

1.1.1 Time rule

Any POI has its business hours. For example, banks open at 7.30 a.m. and closes at 17.00 p.m. in general; supermarkets open at 8 a.m. and close at 8 p.m. When the passengers get out off the taxi, the utilities of different activities are different. It is denoted as equation (2)

$$p_{isInTime} = \begin{cases} 0 & \text{(drop off time in business hour)} \\ 1 & \text{(drop off time not in business hour)} \end{cases} \tag{2}$$

Besides, the attractiveness is different for passenger at different times. For example, restaurants are more attractive at meal time. This kind of time varying factor are defined as time dynamic function. To simplify, we apply the results of Huang *et al.* (2010)

1.1.2 Spatial rule

Taxi drivers usually drop off passengers as close as possible to their destinations. We use maximum walking distance to describe the region of passengers' movements, denoted as maxWalkDis. On the other side, the passenger prefer to select a near-by POI after leaving the taxi, exhibiting a distance decay effect, which can be formulated with a distance decay function (Wang, 2012) Combining the maximum walk distance and distance decay function, the spatial rule can be denoted as (3)

$$p_{spa} = \begin{cases} dis(POI, dropPoint)^{-\beta} & (\text{dis<=maxWalkDis}) \\ 0 & (\text{dis >maxWalkDis}) \end{cases} \quad (3)$$

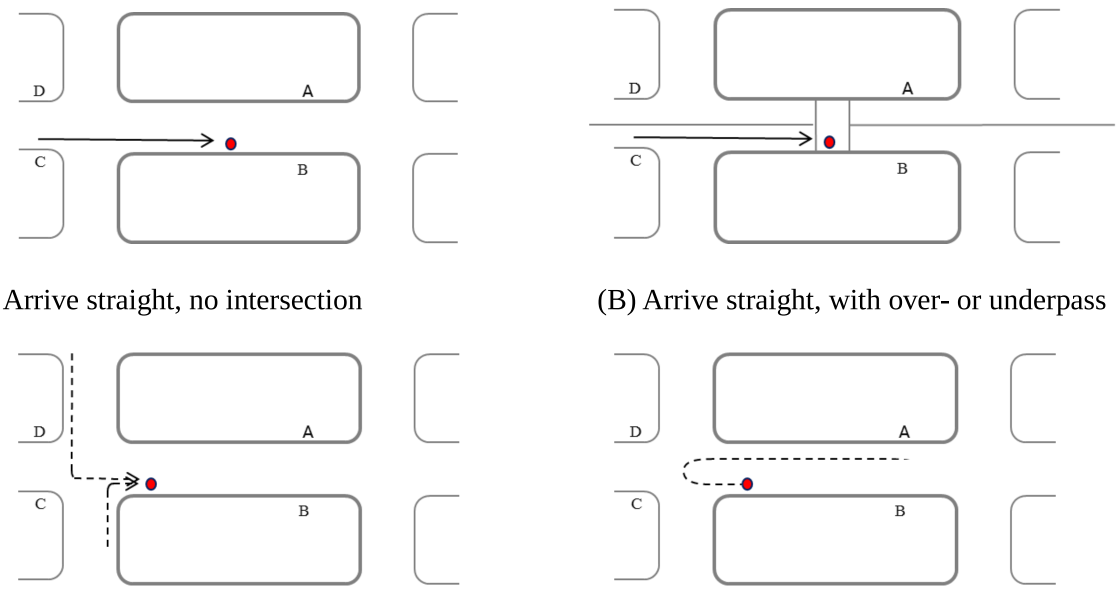
Another question is how to measure the distance between the drop-off point and the POI. In terms of the road POI, the distance the Euclidean distance between drop-off point and the road point. In terms of an inner POI, we assume that passengers would like to ignore the distance between the inner POI and the EnterPOI. So the distance between the inner POI and drop-off point is equal to the distance between the AOI EntryPOI and the drop-off point.

1.1.3 Trajectory form rule

Besides the distance, when getting off the taxi, the passengers have higher possibility to select the same side road although some roads are narrow enough to walk cross. For the main-wide road, selecting the opposite road side is normally expensive in terms of walking and waiting time. This situation will be changed, if there is an underpass or overpass. Based on the above observation, we provide the concept of trajectory form. The trajectory form refers to the morphological characteristics of the end of the trajectory, and the relative spatial relationship between end point and the intersection point. Here, trajectory form mainly refers to the direction of the end of the passenger trajectory.

The trajectory form can divided into the following four types (Figure 2)

Figure 2: The four types of trajectory form



(A) Arrive straight, no intersection

(B) Arrive straight, with over- or underpass

(C) Turn Left and Right

(D) Turn round

Arrive straight: Arrive straight means in the end of the passenger trajectory without no direction changing. It is not for the whole passenger trajectory. According to the existence of intersections, Arrive straight is divided into two parts, as Fig 2(A) and Fig 2(B) respectively. If there is no intersection, the passenger can only reach the POI on the same side, which is denoted as the B area. If a crossing exists, it means the passengers can go across another side easily

Turn Left and Turn Right: Turn Left and Turn Right are in the near area of intersection and only consider the end part of the passenger trajectory. Like the figure 2(C) if the taxi arrives at the destination after turning right, it means the passengers will not go to A, B and D area.

Turn round: Turn round means the taxi arrive at the destination after turning round. Like the figure 2(D) the taxi passes the A area before arriving at the B area. It means the destination of the passenger is in area B.

3.4 The algorithm

The algorithm includes two steps, candidate POI selection and activity allocation.

(A) Candidate POI selection

The main goal of this step is select the candidate POI where the passengers can go after dropping off the taxi. The candidate POI set is determined by the three rules, indicated by (4)

$$CandPOI = ruleTime(POI) \cap ruleSpa(POI) \cap ruleForm(POI) \tag{4}$$

Firstly, we apply the Trajectory form rule to select all the possible AOI and append all the inner POI into the CandPOI. Considering the data quality we only consider Arrive in straight and turn round. That is, if the taxi arrive straight with intersection, we append all the AOI on each side; if the taxi arrive straight with no crossing, we only append the AOI on the same side; if the taxi arrives after turning round, we only append the AOI on the same side.

Then, considering the spatial rule, i.e. the max walk distance. We select the max walk distance as half of the average distance of the road, i.e. 300 m. In terms of the time rule, it means all the selected POI should be open.

(B) Activity allocation.

In this step, we will assign all the candidate POI with a visiting possibility, and then calculate the final possibilities for all the possible activities. The possibility of visiting a POI is determined by the spatial rule, time rule and trajectory form, which is denoted as (5)

$$p_{POI} = p(time, loc, form) = p(time) * p(loc) * p(form) \tag{5}$$

After obtaining the visiting possibility, then the visiting possibilities for the activity is denoted as (6)

$$\mathrm{P}_{act} = \frac{(\sum_i p(POI))}{\sum_{act}(\sum_i p(POI))} \tag{6}$$

After calculating the visiting possibility for all the activities, we will select the activity with the highest visiting probabilities as the trip purpose.

4. Results

4.1 Data

Due to the lack of ground-truth data for the taxi trips, many studies evaluate the effeteness of their algorithms indirectly by comparing them to travel survey data at an aggregate level (Gong *et al.*, 2016; Chen *et al.*, 2018) But these results are not reasonable for two reasons: firstly, the general travel survey is for all the travel modes instead of for taxi data only; secondly, individual choices do not necessarily equal to the aggregate shares.

It is difficult to get the taxi's trajectory and the real passenger purpose at the same time. Thanks to Didi, which is a Chinese TNC, the vehicle driver can record the passengers' origin place and destination in text by their App, e.g. Didi Order. From this Didi Order, we can obtain the passengers' travel purpose. The data collected include Didi order data and trajectory data simultaneously. The sample size is 1862.

Besides the trajectory data, we also need POI, AOI and the road network.

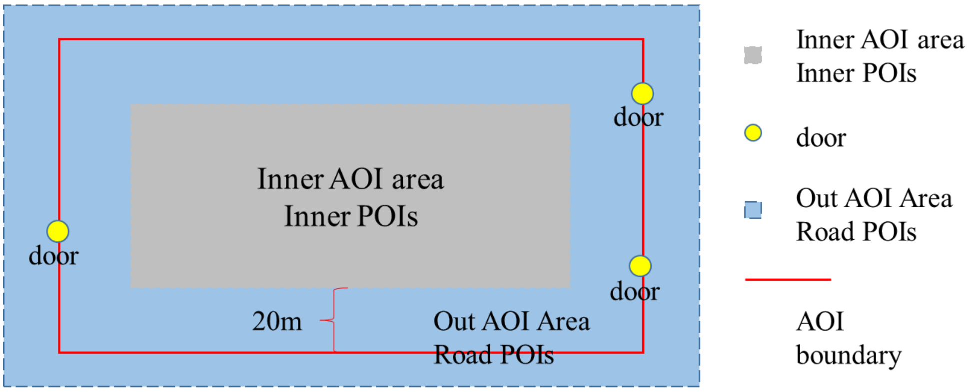
The POI and AOI are downloaded from AutoNavi Map API. These POIs are managed in two layers. The first layer has 23 categories and the second layer has 262 subcategories. Each POI has a unique id, name, address, latitude, longitude and POI category. According to their categories, we can map these POI to different trip purpose, which is shown in table 1.

Table 1: POI-purpose mapping table

Trip purpose	**POI Type**
Work-related	Office building, government and so on
In-Home	Residential subdivisions, hotel and so on
Transportation transfer	Airport, railway station, subway station and so on
Dining	Restaurant
Shopping	Shopping mall, supermarket, store
Recreation	Culture facilities; sport facilities; park and so on
Schooling	University, high school, primary school, technique school etc.
Errand	Beauty salon, laundry, ticket office and so on
Medical	Hospital

Besides these attributes of the POI, the AOI also has boundary information. After downloading the POI, we build the relationship between AOI and POI according to their relationship (Figure 3)

Figure 3: POI-AOI relationships



Firstly, we define a buffer for the boundary of AOI in the inner direction. The grey area is the inner AOI area and the POI in that area belongs to Inner POIs. The other blue area is the Out AOI area and the POI in that area belong to road POIs.

4.2 Base algorithms

To evaluate the algorithm, four algorithms are used as baseline algorithms. The existing taxi trip purpose algorithms includes four types, including voting method (Phithakkitnukoon *et al.*, 2010) the nearest distance (Xie *et al.*, 2009) only consider time and spatial constrains (Furlettid *et al.*, 2013) Bayes' formula (Gong *et al.*, 2016)

Table 2: Baseline algorithms

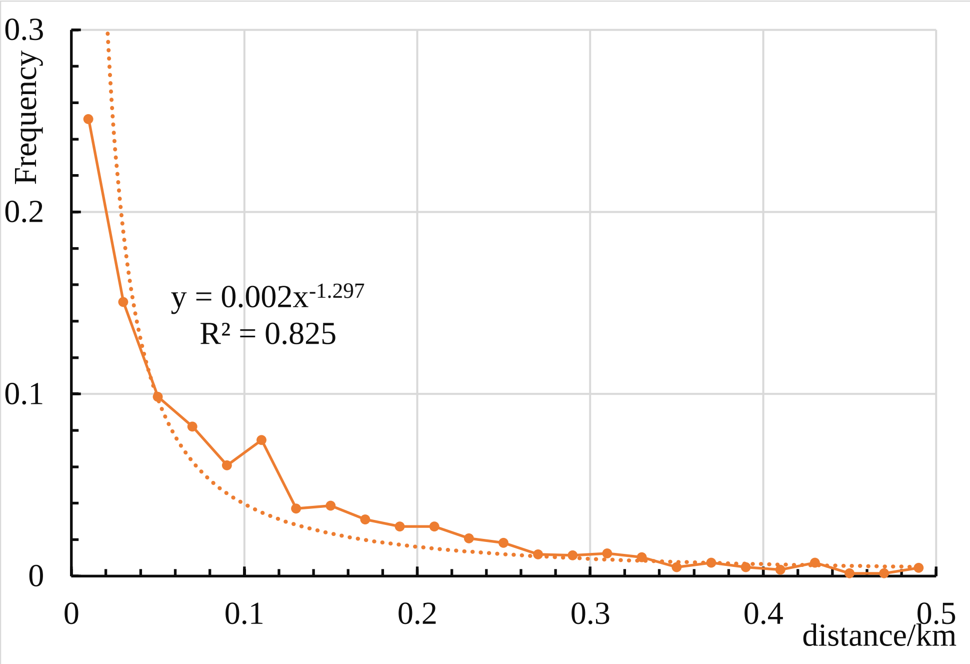
Index	Methods	Note
#1	Voting method	Select the activity with the most POIs
#2	Nearest distance method	Select the nearest POI's activity
#3	Spatial-temporal constrains	Only considering spatial-temporal constrains
#4	Bayes formula based	Based on Bayes formula, considering a number of factors
#5	Methods of this paper	POI-AOI context repository, spatal-temporal constraints and trajectory form

To evaluate our method, we compare these five methods with the aforementioned dataset. If different methods have the same factor, we adopt the same values. For example, #3, #4, #5 adopt the max walk distance simultaneously, we set all the max walk distance as 300m. For the information we cannot obtain, we adopt the value from the reference paper. For example, time varying factor is from Huang (2010) and we adopt it.

For the distance decay function in method #4 and #5, we adopt the function as equation (3) The exponent (β) is determined from the dataset. To get β , we firstly extract the real destination POI from Didi order and get the coordinate by Gaode. Then we calculate the distance between the drop off point and the real destination POI. The distance distribution is denoted as Figure 4. And the distance equation is:

$$\text{dis}Decay = dis^{-1.3} \quad \text{(dis/km)} \tag{7}$$

Figure 4: The distance decay model for walking after drop off taxi



4.3 Evaluation

We adopt accuracy as the evaluating indicator, which is measured by the ratio of the right prediction of and all the result (7)

$$precision = \frac{\sum_i^n Is(InfAct = TrueAct)}{n} \tag{8}$$

In this equation, InfAct is the predicted purpose. TrueAct is the real trip purpose. Is() will be 1 if the two values are equal. The result is illustrated as Table 3.

Table 3: Predicted result comparison

Index	**Accuracy (%)**
#1	37.00
#2	26.58
#3	36.98
#4	55.34
#5	67.95

From the result, we observe:

(1) On the whole, the proposed method can improve the accuracy of predicting the taxi trip purpose by introducing the trajectory form and POI-AOI context repository.

(2) If we compare the methods separately, we can find that different factors have different influence for the result. Method #1 and method #2 are the two basic methods. Method #1 only considers the distance. Method #2 select the activity which has the most POIs as the predicted activity. To some degree, method #2 is similar to predicting activity by the land use. The accuracy of #1 is higher than #2. It means people are more influenced by distance. Although method #2 introduces the gravity model to describe the visiting probability of different POI, the result is similar to method #1. According to the paper (Furlettid *et al.*, 2013) the square of the distance between POI and drop off point is the denominator. It means, the influence of distance is increased. So there is no difference between method #1 and method #3. Through Method #3, #4, #5, method #4, #5 apply power function distance decay model instead of gravity model. Method #4, #5 are better. We can assume that the power function distance models the decay better. Comparing method #4 and #5, although #5 does not use the time varying factor instead introducees the trajectory form, it is better than #4. One can say, the trajectory form is useful to improve the accuracy of imputing taxi trip purpose.

(3) Although #5 improve the accuracy of trip purpose imputation, it introduce additional factors. It means the we need higher quality data. It is a challenge. So we need to think how to find the most efficient factors and drop the less useful factors.

5. Discussion

In this paper, we present a rule-based algorithm for the taxi trip purpose imputation. Because the limited information for taxi trajectory data, we introduce the trajectory form to build rules besides the spatiotemporal attributes for the drop off point for the trajectory. In addition, we define a POI-AOI two layers contextual data repository which can bind the POI and AOI together. Finally, we evaluate the algorithms using the real-world datasets at the individual level instead of at the aggregate level. The results demonstrate the proposed taxi trip purpose imputation achieves a promising performance in accuracy.

However, some problems exist. Rule-based algorithms are seen as ad-hoc algorithms, which normally involve the specific data and maybe not useful for other datasets. (Feng and Timmermans, 2016) Without exception, our algorithms may have the same problems because we only evaluate the algorithm in one dataset. In addition, our algorithm introduce additional rules which increase the complexity and we cannot guarantee that all the rules are useful.

In the future, we plan to broaden this work in the following direction. First, we plan to apply more advanced machine learning problem which are more flexible in handling such complex problems. Second, evaluating the algorithm using Didi order data is a direction. However, the evaluation dataset is small. It is necessary to enrich the dataset.

6. References

Axhausen, K. W., and T. Gärling (1992) Activity-based approaches to travel analysis: conceptual frameworks, models, and research problems, *Transport Reviews*, **12** (4) 323-341.

Bohte, W., and K. 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) 285-297.

Chen, C., Jiao, S., Zhang, S., Liu, W., Feng, L., and Wang, Y. (2018) TripImputor: Real-Time Imputing Taxi Trip Purpose Leveraging Multi-Sourced Urban Data. *IEEE Transactions on Intelligent Transportation Systems*, 1-13.

Chen, C., Ma, J., Susilo, Y., Liu, Y., and Wang, M. (2016). The promises of big data and small data for travel behavior (aka human mobility) analysis. *Transportation Research Part C*: *Emerging Technologies*, **68**, 285-299.

Jiang, S., Ferreira, J., and Gonzalez, M. C. (2017). Activity-Based Human Mobility Patterns Inferred from Mobile Phone Data: A Case Study of Singapore. *IEEE Transactions on Big Data*, **3**(2), 208-219.

Kim, Y., Pereira, F. C., Zhao, F., Ghorpade, A., Zegras, P. C., and Ben-Akiva, M. (2015) Activity recognition for a smartphone and web based travel survey. *arXiv preprint arXiv:1502.03634*.

Feng, T., and Timmermans, H. J. P. (2016) Comparison of advanced imputation algorithms for detection of transportation mode and activity episode using GPS data. *Transportation Planning and Technology*, **39** (2) 180-194.

Furletti, B., Cintia, P., Renso, C., and Spinsanti, L. (2013) Inferring human activities from GPS tracks. *Proceedings of the 2nd ACM SIGKDD International Workshop on Urban Computing*.

Ge, Y., Xiong, H., Tuzhilin, A., Xiao, K., Gruteser, M., and Pazzani, M. (2010) An energy-efficient mobile recommender system. *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*.

Graaff, V. d., de By, R.A., van Keulen, M. and Flokstra, J. (2013) Point of interest to region of interest conversion. *Proceedings of the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. Orlando, Florida, ACM: 388-391.

Gonzalez, M. C., Hidalgo, C. A., & Barabasi, A. L. (2008) Understanding individual human mobility patterns. *Nature*, **453**(7196) 779-782.

Huang, L., Li, Q., & Yue, Y. (2010) Activity identification from GPS trajectories using spatial temporal POIs' attractiveness. *Proceedings of the 2nd ACM SIGSPATIAL International Workshop on Location Based Social Networks*, San Jose, California.

Montini, L. (2016). Extraction of transportation information from combined position and accelerometer tracks, ETH Zurich.

Liu, K., Deng, K., Ding, Z., Li, M., and Zhou, X. (2009) Moir/mt: Monitoring large-scale road network traffic in real-time. *Proceedings of the VLDB Endowment*, **2**(2) 1538-1541.

Liu, Y., Kang, C., Gao, S., Xiao, Y., & Tian, Y. (2012) Understanding intra-urban trip patterns from taxi trajectory data. *Journal of Geographical Systems*, **14**(4) 463-483.

Liu, Y., Sui, Z., Kang, C. and Gao, Y. (2014) Uncovering patterns of inter-urban trip and spatial interaction from social media check-in data. *PLoS One*, **9**(1) e86026.

Moiseeva, A., Jessurun, J., and Timmermans, H. (2010) Semiautomatic imputation of activity travel diaries: use of global positioning system traces, prompted recall, and context-sensitive learning algorithms. *Transportation Research Record: Journal of the Transportation Research Board* (2183) 60-68.

Parent, C., SPACCAPIETRA, S. and RENSO, C. (2013) Semantic trajectories modelling and analysis. *ACM Computing Surveys*, **45**(4) 1-32.

Phithakkitnukoon, S., Horanont, T., Di Lorenzo, G., Shibasaki, R., and Ratti, C. (2010) Activity-aware map: Identifying human daily activity pattern using mobile phone data. *International Workshop on Human Behavior Understanding*.

Schuessler, N. and K. Axhausen (2009) Processing Raw Data from Global Positioning Systems Without Additional Information. *Transportation Research Record: Journal of the Transportation Research Board* **2105**: 28-36.

Shang, J., Zheng, Y., Tong, W., Chang, E. and Yu, Y. (2014) Inferring gas consumption and pollution emission of vehicles throughout a city. *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*.

Xie, K., Deng, K. and Zhou, X. (2009) From trajectories to activities: a spatio-temporal join approach. *Proceedings of the 2009 International Workshop on Location Based Social Networks*, Seattle, Washington, 25-32.

Yuan, J., Zheng, Y. and Xie, X. (2012) Discovering regions of different functions in a city using human mobility and POIs. *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, Beijing, China.

Yuan, N. J., Zheng, Y., Zhang, L. and Xie, X. (2013) T-Finder: A Recommender System for Finding Passengers and Vacant Taxis. *IEEE Transactions on Knowledge and Data Engineering*, **25**(10) 2390-2403.

Wang, F. (2012) Measurement, Optimization, and Impact of Health Care Accessibility: A Methodological Review. *Ann Assoc Am Geogr*, **102**(5) 1104-1112.

Wolf, J., Guensler, R., & Bachman, W. (2001) Elimination of the travel diary: Experiment to derive trip purpose from global positioning system travel data. *Transportation Research Record: Journal of the Transportation Research Board*, (1768) 125-134.

Zhao, Z., Koutsopoulos, H. N., & Zhao, J. (2018) Individual mobility prediction using transit smart card data. *Transportation Research Part C: Emerging Technologies*, **89**, 19-34.