R4TLI
ISSN: 2513-2504

A TRIP PURPOSE INFERENCE FRAMEWORK USING SPATIAL CLUSTERING AND BAYESIAN PROBABILITY

Dineth Dhananjaya, Thillaiampalam Sivakumar
Department of Transport and Logistics Management, Faculty of Engineering, University of Moratuwa, Sri Lanka
d.dineth.dananjaya@gmail.com, tsivakumar@uom.lk

ABSTRACT - Taxis are one of the most widely used modes of transport among urban communities. The use of GPS devices in modern taxi vehicles has enabled the estimation of travel patterns through emitted and collected massive scale trip records. The only necessity that requires for this is a suitable trip purpose inference model as the GPS data are unable to provide the exact purpose of a trip but the neighborhood of travelers' destination. Thus, this study attempted to develop a trip purposes inference framework that can be used reliably in uncovering travel patterns. The proposed framework consists of three layers: (1) Trip purpose imputation for regular trips using spatial clustering, (2) Identifying the trips attracted to residential trips, and (3) Purpose inference using Bayesian probability. The model was tested using taxi trips data from a service provider operating in Colombo District, Sri Lanka, and compared that with the activity proportions data taken from a household travel survey. The results indicates that the proposed model is capable of providing plausible travel patterns through identified spatial dynamics and temporal patterns.

Keywords: Travel patterns, GPS data, POI data, taxi trips, spatial clustering, Bayesian probability

1. INTRODUCTION

Every single trip made by a passenger has a particular purpose or an activity to be fulfilled. In this context, taxis play a pivotal role as para transport mode among the available alternative transportation modes among urban communities for decades. For this reason, identifying the activity-based travel patterns of taxi users assist the city planners and transport planners in tasks such as traffic management, city planning, and investment decision makings. Modern taxi services are often embedded with either in-built or mobile GPS and the emitted data are stored by the service providers to uncover the knowledge behind their business. Although the temporal and spatial trip information can be extracted without much effort, identifying the trip purposes requires a well-developed accurate methodology and it has created the trip purpose inference problem.

Nguyen et.al, (2019) reviewed a vast scope in deriving trip information from GPS data and categorized the methods that had been used to trip purpose inference into three methods: rule-based, probabilistic, and machine learning-based. The rule-based models mainly incorporate a single parameter to directly assign a purpose for a trip. [1]. Eliminating the deterministic behavior in rule-based models, the probabilistic models attempt to give a probability for each possible purpose using several parameters such as attractiveness, the temporal impact of a trip, etc. [2]. The advantages of machine learning-based models come with the ability to utilize additional parameters such as origin context, and trip time for the inference [3]. The utilization of these methods for applications depends on factors such as transferability, data scale, number of purposes, and more importantly the land use context. However, the trip purpose inference problem has not been thoroughly addressed in developing countries despite the fact that the applicability of a trip purpose inference model extremely depends on the land use context. [4]. Hence, this study attempted to ameliorate an accurate model (base model) based on taxi data collected in Colombo District, Sri Lanka.

2. MATERIALS AND METHODS

2.1. Data description.

2.1.1. GPS data

Taxi trips originated from the Colombo District and attracted to "Dehiwala", "Sri Jayawardhanapura Kotte" and "Thimbirigasyaya" Divisional Secretariat (DS) divisions for a period of a week (2019/11/11 – 2019/11/17) was used for the assessment of the proposed framework. Dataset cleaning was performed based on assessing the trip time and linear distance reliability for the selected study area. The dataset provided GPS records of 107,617 taxi trips with features relating to the origin, and destination timestamped coordinates.

2.1.2. Point of Interest data

The main supplementary data source for trip purpose inference is the Point of Interest (POI) data as it provides insights into the nature of land use around the context of the destination at which the passenger may perform an activity. 42,418 POIs were collected for the selected study area by applying the Nearby search request of the Google Places API. The purpose for each POI was mapped manually referring to the provided category type for a POI by the API.

2.2 Three-layer trip purpose inference framework.

2.2.1. Layer1: Regularity based purpose imputation

Imputation of the trip purpose of commuters who regularly travel between identical locations is considered in the first layer. Density-Based Spatial Clustering of Applications with Noise (DBSCAN) is used to identify the spatial clusters of Pick-Up Points (PUP) and Drop Off Points (DOP) of passenger taxi trips and the defined time bins are utilized to add the temporal frames to select regular trips intended for a purpose. Two scenarios (1) round trips, (2) one-way trips are identified from this, and the purposes are imputed considering the selected time bins for selected purposes.

2.2.2. Layer 2: Residential trip imputation

The second layer attempts to identify the trips attracted to residential places as the POI data does not contain such data by referring to three assumptions. The first assumption is that residential places do not attract a higher number of trips per day which is captured from the outlier of GPS records from the DBSCAN algorithm. The second is that residential places do not locate close to the main road network of the study area, and this is referenced by using k-means clustering for the measured distance of DOPs to main roads from Open Street Routing Machine (OSRM) and removing the clusters with shorter distances intuitively. The third assumption is that if an attractive POI locates around the DOP that observation is also removed referring to the number of reviews as there is a higher probability that a passenger may visit that place.

2.2.3. Layer 3: Destination based trip purpose inference

Following the improvements made from the first two-layer, the final layer applied the selected base for the trip purpose inference for the rest of the taxi trips. Dhananjaya & Sivakumar (2021) , showed that the Bayes-theorem-based model proposed by Li Gong et al., (2016) provides more reliable results for the land use contexts in developing countries [5], [6]. Hence, that model is attached to the final layer of the proposed three-layer trip purpose inference framework in this study.

3. RESULTS AND DISCUSSION

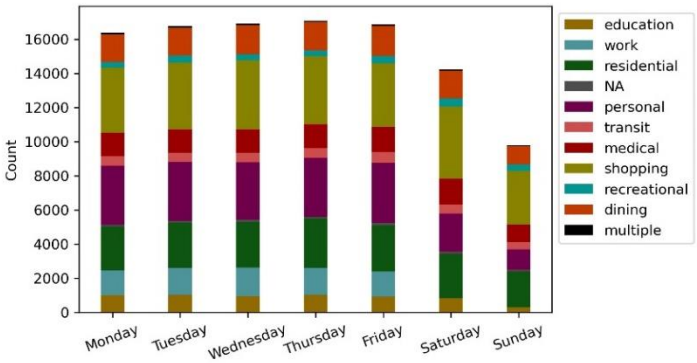

Figure 1. Daily Distribution of Trip Purpose Inference Results

The above Figure 1 depicts the achieved daily distribution of trip counts from the proposed model. It can be identified that shopping and personal are the most widely involved activities by taxi trips from the users in the selected study region and activities such as recreational and transit had received a lower attraction. The study will be continued to identify the travel patterns in terms of temporal distributions and spatial dynamics. It is expected to obtain insights relating to what times provide the peak and what are the places that attract a higher number of trips for different activity types

ACKNOWLEDGEMENT

This research was supported by the Accelerating Higher Education Expansion and Development (AHEAD) Operation of the Ministry of Higher Education funded by the World Bank

REFERENCES

1. K. Xie, K. Deng, and X. Zhou, “From trajectories to activities: A spatio-temporal join approach,” *GIS Proc. ACM Int. Symp. Adv. Geogr. Inf. Syst.*, pp. 25–32, 2009, doi: 10.1145/1629890.1629897.
2. B. Furletti, P. Cintia, C. Renso, and L. Spinsanti, “Inferring Human Activities from GPS Tracks,” 2013, doi: 10.1145/2505821.2505830.
3. C. Chen, C. Liao, X. Xie, Y. Wang, and J. Zhao, “Trip2Vec: a deep embedding approach for clustering and profiling taxi trip purposes,” *Pers. Ubiquitous Comput.*, vol. 23, no. 1, pp. 53–66, 2019, doi: 10.1007/s00779-018-1175-9.
4. M. H. Nguyen, J. Armoogum, J. L. Madre, and C. Garcia, “Reviewing trip purpose imputation in GPS-based travel surveys,” *J. Traffic Transp. Eng. (English Ed.*, vol. 7, no. 4, pp. 395–412, 2020, doi: 10.1016/j.jtte.2020.05.004.
5. D. D. Dhananjaya and T. Sivakumar, “Inferring the purposes of taxi trips using GPS and POI data considering the destination context,” in *2021 International Conference on Data Analytics for Business and Industry (ICDABI)*, 2021, pp. 295–300, doi: 10.1109/ICDABI53623.2021.9655943.
6. L. Gong, X. Liu, L. Wu, and Y. Liu, “Inferring trip purposes and uncovering travel patterns from taxi trajectory data,” *Cartogr. Geogr. Inf. Sci.*, vol. 43, no. 2, pp. 103–114, 2016, doi: 10.1080/15230406.2015.1014424.