

**IDENTIFYING THE TRAVEL PATTERNS OF
ON-DEMAND TAXI TRIPS
THROUGH INFERRED TRIP PURPOSES**

Durage Dineth Dhananjaya

218014X

Degree of Master of Science

Department of Transport and Logistics Management

University of Moratuwa

Sri Lanka

SEPTEMBER 2022

**IDENTIFYING THE TRAVEL PATTERNS OF
ON-DEMAND TAXI TRIPS
THROUGH INFERRED TRIP PURPOSES**

Durage Dineth Dhananjaya

218014X

Degree of Master of Science

Department of Transport and Logistics Management

University of Moratuwa

Sri Lanka

SEPTEMBER 2022

DECLARATION OF ORIGINALITY

I declare that this is my own work, and this thesis/dissertation does not incorporate without acknowledgement any material previously submitted for a Degree or Diploma in any other University or institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text.

Also, I hereby grant to University of Moratuwa the non-exclusive right to reproduce and distribute my thesis/dissertation, in whole or in part in print, electronic or other medium. I retain the right to use this content in whole or part in future works (such as articles or books).

Signature:

Date: 23/09/2022

UOM Verified Signature

STATEMENT OF THE SUPERVISOR

The above candidate has carried out research for the Degree of Master of Science under my supervision.

Name of the supervisor: Dr. T. Sivakumar

Signature of the Supervisor:

Date: 23/09/2022

UOM Verified Signature

ABSTRACT

Activity-based modeling has become the backbone behind transportation planning, and trip purposes that can be inferred from large GPS datasets of different travel means are paving the path to augmenting its accuracy. In this context, the trip purpose inference problem has emerged since the GPS is unable to capture the trip purposes explicitly. This 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. Hence, this study attempted to ameliorate an accurate model (base model) for a chosen study area in Colombo District, Sri Lanka using the on-demand taxi trips data.

Point of Interest (POI) data is often accompanied by the purpose inference models as it provides a complete insight into the land use around origin and destinations. In the pursuit of the main objective of the study, machine-learning-based text classification was tested to improve the number of informative POIs and its outcome indicated that the Support Vector Machine (SVM) classifier can be utilized effectively and efficiently. The designed trip purpose inference model referring to a base model is proposed as a three-layer trip purpose inference framework in which a method to impute purpose based on trip regularities and the method to identify residential trips was included as two layers before using the base model. The validity of the proposed model was evaluated with the assistance of household travel survey data and with respect to the purpose proportions and division level R-square. Furthermore, travel patterns of the on-demand taxi trips were assessed in terms of temporal regularity, trip lengths, and spatial dynamics. It is recommended to conduct further studies to assess the applicability of other parameters such as trip origin context and trip time with the assistance of unsupervised learning methods.

Keywords:

Travel behavior, trip purpose inference, human mobility, taxi trips, GPS data, POI data, semantic enrichment, developing countries.

ACKNOWLEDGEMENTS

I would like to convey my heartfelt gratitude to Dr. T. Sivakumar for providing me an opportunity to conduct this study under his supervision and for assisting me to drive this towards success from the very beginning. The guidance he has given me assisted to improve the quality of this work gradually and ultimately add a valuable context of knowledge for the community regarding the study area. I would like to extend my appreciation to Senior Professor Amal S. Kumarage for further helping me to identify the sections that required improvements in making this work more worthwhile for both academics and industry.

My warm thankfulness goes to Dr. Amal S. Perera for helping me in data collection and also providing me the additional knowledge from a computer science background which was highly useful during the data processing task. I would also like to thank all the members who worked on the research project “Data-Driven Transport Analytics” and the member of the “DataSearch” research center for the assistance and comments on the work. Furthermore, I would like to convey my sincere gratitude to Accelerating Higher Education Expansion and Development (AHEAD) Operation for financially assisting this study through the above-mentioned project.

Most importantly, none of this would be possible without the love and care of my family. I am grateful to my parents and my brother for strengthening me throughout this study period and helping me to start again with stamina during the critical periods which motivated me to complete this study at level best.

TABLE OF CONTENT

DECLARATION OF ORIGINALITY	i
STATEMENT OF THE SUPERVISOR	ii
ACKNOWLEDGEMENTS	iv
TABLE OF CONTENT	v
LIST OF FIGURES	viii
LIST OF TABLES	ix
LIST OF EQUATIONS	ix
LIST OF ABBREVIATIONS	x
1. INTRODUCTION	1
1.1 Background of the Study	1
1.2 The Trip Purpose Inference Problem	3
1.3 Research Gaps	4
1.4 Research Objectives	4
1.5 Significance of the Study	5
1.6 Chapter Summary	6
2. LITERATURE REVIEW	7
2.1 Overview of the Systematic Literature Reviews	7
2.2 Methods of Trip Purpose Inference	9
2.2.1 Rule-based models	9
2.2.2 Probabilistic models	9
2.2.3 Machine Learning (ML) based models	10
2.3 Usage of Different Data Types	12
2.3.1 Taxi GPS data	13
2.3.2 Point of Interest (POI) data	13
2.3.3 Other supplementary data	14
2.4 Trip Purpose Inference Model Validation	16
2.5 Travel Pattern Identification	16
2.6 Chapter Summary	17

3. METHODOLOGY	18
3.1 Selecting the Study Area.....	18
3.2 Data Collection and Preprocessing	18
3.2.1 Taxi GPS data	18
3.2.2 Point of Interest (POI) data	21
3.2.3 Route network data.....	24
3.2.4 Household travel survey data	25
3.3 Selecting the trip purpose inference models to compare	26
3.3.1 Selecting the candidate POIs for a DOP	26
3.3.2 Selected trip purpose inference models to compare.....	27
3.3.3 Crowdsourced reviews to measure the POIs’ attractiveness.	28
3.4.5 Assessing the temporal impact for trip purposes within the study area.....	29
3.4 Enhancing the POI Data for Trip Purpose Inference	30
3.4.1 Predicting the category of a POI.....	30
3.5 Identifying the Trips Attracted to Residential Places	33
3.6 Identifying the Purposes from Trip Regularity Analysis	34
3.6.1 Scenario identification from spatial regularities	35
3.6.2 Selecting the time bins for trip purposes.....	36
3.6.3 Purpose imputation of the round trips and onward trips.....	37
3.7 Designed Three-Layer Framework for Trip Purpose Inference.....	38
3.8 Validating the Results	40
3.9 Identifying the Travel Patterns.....	41
3.11 Chapter summary	41
4. RESULTS AND DISCUSSION	43
4.1 Selected Model Comparison Results	43
4.2 Enhanced POI data Results	44
4.2.1 Predicted POI categories	44
4.2.2 Identified entering locations of POIs	46
4.3 Purpose Imputation from the Three-Layer Framework	49
4.3.1 Regularity-based purpose imputation	49
4.3.2 Residential trip imputation.....	51
4.3.3 Destination-based trip purpose inference.....	53

4.3.4. Complete results.....	54
4.4 Model validation	54
4.5 Identified Travel Patterns.....	57
4.5.1 Temporal distributions	57
4.5.2 Trip length distribution	60
4.5.2 Spatial dynamics	64
4.6 Chapter Summary	68
5. CONCLUSION AND RECOMMENDATIONS	70
REFERENCE LIST	75
APPENDICES	79
Appendix 1: Questionnaire to measure the temporal impact for different trip purposes	79
Appendix 2: Determined discrete values to use in including the temporal impact to the bayes-theorem based model referring to the achieved results of the survey.....	81

LIST OF FIGURES

Figure 1: Proportions of Trip purpose inference studies by countries	8
Figure 2: Proportion of studies highlighted the POI importance	14
Figure 3: Study count for validation methods.....	16
Figure 4: Selected study area for the study	18
Figure 5: Taxi trip data preprocessing flow	19
Figure 7: Taxi trip travel time distribution	20
Figure 7: Taxi trip distribution over the period	20
Figure 8: Number of trips by each mode	20
Figure 9: Spatial distribution of taxi trips within the country.....	20
Figure 10: Spatial distribution of selected taxi trips	21
Figure 11: Sample API call.....	22
Figure 12: Grid-based POI data extraction approach.....	23
Figure 13: POI data spatial distribution	24
Figure 14: OSM Road network data (left: all roads, right: main roads)	25
Figure 15: Candidate POI selection	27
Figure 16: Attractiveness value estimation from number of reviews	29
Figure 17: POI entrance location identification method.....	32
Figure 18: Residential trips identification method.....	33
Figure 19: Trip distribution between commuters and non-commuters.....	34
Figure 20: Identified regular trip scenarios. (1) Round trip, (2) Onward trip, (3.1) Regular PUP, (3.2) Regular DOP	35
Figure 21: Time bin selection for regular trips	36
Figure 22: Steps of the designed purpose imputation method through trip regularities .	37
Figure 23: Designed three-layer framework for trip purpose inference	39
Figure 24: Total, Predicted vs labelled POIs for each purpose.....	44
Figure 25: Spatial distribution difference between original and improved POI data (left: original data, right: total data).....	45
Figure 26: POI selection for entrance identification.....	46
Figure 27: Sample entrance location identification	48
Figure 28: Distribution of distance between POIs and entrance locations	48
Figure 29: Daily distribution of imputed trip purposes from the layer 1	50
Figure 30: Daily distribution of clustered and individual trips in layer 2.....	51
Figure 31: Daily distribution of identified residential trips from layer 2.....	52
Figure 32: Daily distribution of trip purpose inference results from layer 3	53
Figure 33: Daily distribution of trip purpose inference from the proposed three-layer framework.....	54
Figure 34: DS division level shopping trip proportion comparison between proposed (right) and base model (left).....	56
Figure 35: DS division level educational trip proportion comparison between proposed (right) and base model (left).....	56
Figure 36: Temporal regularities of work, education and medical trips.....	57
Figure 37: Temporal regularities of the personal, shopping and residential trips.....	58

Figure 38: Trip length distributions for work, educational, medical and personal trips (in meters).....	60
Figure 39: Trip length distributions for personal, residential, dining and transit trips (in meters).....	61
Figure 40: Box plots for the trip length distributions for each purpose.....	62
Figure 41: Trip length distribution for recreational trips (in meters).....	62
Figure 42: Base layer of the spatial dynamics' visualization.....	64
Figure 43: Spatial dynamics of work trips.....	65
Figure 44: Spatial dynamics of educational trips.....	66
Figure 45: Spatial dynamics of medical trips.....	66
Figure 46: Spatial dynamics of personal trips.....	66
Figure 47: Spatial dynamics of shopping trips.....	67
Figure 48: Spatial dynamics of transit trips.....	67
Figure 49: Spatial dynamics of recreational trips.....	67
Figure 50: Spatial dynamics of dining trips.....	68
Figure 51: Spatial dynamics of recreational trips.....	68

LIST OF TABLES

Table 1: Differences between TS and HG-related studies.....	8
Table 2: Comparison of trip purpose inference methods.....	12
Table 3: Collecting methods, importance, and accuracy of data for trip purpose inference.....	15
Table 4: Taxi trip data features information.....	19
Table 5: POI categorization.....	25
Table 6: ML classifiers results for POI category prediction.....	31
Table 7: Descriptions for identified regular trip scenarios.....	35
Table 8: Descriptions for travel patterns.....	41
Table 9: Achieved purpose counts from the compared models.....	43
Table 11: Maximum distance and POI counts from distance-based clustering.....	47
Table 10: POI and maximum review counts from attractiveness-based clustering.....	47
Table 12: Purpose count for entrance location identified POIs.....	47

LIST OF EQUATIONS

Equation 1: Closest Assignment.....	27
Equation 2: Gravity model.....	27
Equation 3: Bayes-theorem based model.....	27
Equation 4: Explicit equation for bayes-theorem based model.....	28
Equation 5: Designed equation for attractiveness measure.....	29
Equation 6: F1 score to evaluate ML classifiers.....	31

LIST OF ABBREVIATIONS

AFC - Automatic Fare Collection (AFC)

API - Application Programming Interfaces (API)

AGPS - Assisted GPS (AGPS)

AIM - Activity Inference Model (AIM),

CDR - Caller Detail Records (CDR)

DBSCAN - Density-Based Spatial Clustering of Applications with Noise (DBSCAN)

DOP - Drop off Point (DOP)

DMR - Dirichlet-Multinomial Regression (DMR)

DS - Division Secretary (DS)

DT – Decision Tree

KDE - Kernel Density Estimation (KDE)

KNN – K Nearest Neighbors

GPS - Geographical Processing System (GPS)

HG - Human Geography (HG).

HDOP - Horizontal Dilution of Precision (HDOP)

LBS - Location Based Services (LBS)

LR – Logistics Regression

MNB – Multinomial Naïve Bayes

ML - Machine Learning (ML)

NLP - Natural Language Processing (NLP)

NSAT - Number of sync SATellites (NSAT)

OSM - Open Street Map OSM

OSRM - Open Street Routing Machine (OSRM)

POI - Point of Interest (POI)

PUP - Pick Up Point (PUP)

R - Walking Radius (R)

SML - Supervised Machine Learning (SML)

SVM – Support Vector Machine

TRID - Transport Research International Documentation (TRID).

TS - Transportation Science (TS)

UML - Unsupervised Machine Learning (UML)