

# **USE OF CROWDSOURCED TRAVEL TIME DATA IN TRAFFIC ENGINEERING APPLICATIONS**

Sakitha Pathmila Kumarage

(178042H)

Thesis submitted in partial fulfillment of the requirements for the degree  
Master of Science in Civil Engineering

Department of Civil Engineering

University of Moratuwa  
Sri Lanka

June 2018

## **Declaration**

I declare that this is my own work and this thesis 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).

S.P. Kumarage

Date 20/6/2018

The above candidate has carried out research for the Masters thesis under my supervision.

Dr. G.L.D.I. De Silva

Date 20/6/2018

## **Dedication**

This thesis is dedicated to  
Ven. Lankapura Saripuththa Thero,  
my parents Lalith Kumarage and Sandhya Kumarage,  
my sisters Senuri Kumarage and Osuri Kumarage  
and Wanuji Abewickrema.

## **Acknowledgement**

To my supervisor Dr. Dimantha De Silva for the constant encouragement and guidance at all stages of this research without whose invaluable commitment and dedication this research might not have been a reality. I would like to appreciate the valuable suggestions and guidance given by Prof. Saman Bandara, the senior professor at Transport Engineering Group, University of Moratuwa.

Special thank goes to the Senate Research Committee of University of Moratuwa for providing financial support under a short-term grant SRC/ST/2017/32 to facilitate a part of the research work. I like to thank Colombo Municipal Council for giving Infrared Traffic Logger (TIRTLE) to collect traffic data and Sri Lanka Police for granting permission to collect traffic flow data. Special thank goes to Google Inc for enabling Google Maps APIs and allowing free usage limits which made this study a success. I would like to acknowledge Director Postgraduate Studies, Dean Faculty of Graduate Studies, Head of Civil Engineering Department, Research Coordinator Department of Civil Engineering and all the academic staff members of the Department of Civil Engineering for providing the opportunity to complete this research.

I would like to extend my heartfelt gratitude for colleagues at Department of Civil engineering for supporting the research work. Special thank goes to Deshan Bandara, Chamith Dimantha, Chathura Vidanapathirana, Sasika Ranawaka, Gimhan Jayasiri, Milinda Amarasinghe for the support extended in probe vehicle data collection and license plate surveying. I would like to appreciate the support given by Ms. Gayani Rajapaksa to me in conducting the traffic flow estimation using machine learning principles. I would like to acknowledge the support given by Ms. Anuja Fernando, Lecturer Department of Transport and Logistic Studies in preparation of the thesis. I would like to mention the support given by non-academic staff at Department of Civil Engineering. Special thank goes to Ms. Melani Jayakody, the technical officer at Traffic Engineering Laboratory.

Finally, a special thank goes to my parents, my sisters and Wanuji for the constant support and motivation which encourage me to make this research a reality.

## Abstract

Transport planning and management are required to provide quality and reliable transport service. Collection of data required for this purpose has been always a challenge and obtaining reliable traffic information will ensure proper planning and management of transport activities efficiently. There are many methods followed in travel time data collection by incorporating both fixed detectors such as traffic sensors and moving detectors such as probe vehicles. Collection of travel time data under both methods requires significantly high investment and technical expertise.

With the development of intelligent transport systems economical ways of traffic data collection based on advanced detection principles were introduced. communication and detection methodologies have faced up with the advancement of crowdsourced data mining allowing more readily extractable information on transport and mobility. this research focuses on development of an economical method for obtaining crowdsourced travel time data. Scalability to larger networks, consistent data collection and data collection at multiple locations simultaneously and ensuring the reliability are issues which are addressed.

Travel time data obtained from Google Distance Matrix API which is a processed information released based on crowdsourced mobile phone data, is used in this study to identify use of crowdsource travel time data and transport planning activities. A cloud-based data acquisition platform was prepared for the data collection by accessing the Google Distance Matrix API. The travel time observed from Google Distance Matrix API was verified with the travel time information collected by using GPS enabled probe vehicles. The results indicate that there is a significant agreement between the travel time given by Google Distance Matrix API and actually observe data for both short distance and long-distance trips.

Several applications are illustrated to understand use of travel time information obtained by the Google Distance Matrix API. A traffic flow estimation model based on machine learning principles is proposed for urban roads, A bottleneck identification method based on spatio temporal analysis of travel time and space mean speed variation is illustrated to analyse corridor traffic. Further evaluating the traffic impact of implementation of bus priority lanes and evaluating the traffic impact of implementation of reversible lanes were discussed with respect to Colombo Metropolitan Area.

With the successful implementation of this research it was identified that use of travel time information given by Google Distance Matrix API is a reliable consistent and economical method of collecting travel time information and it is recommended that the public authorities and organisations responsible in managing city traffic use this tool to improve traffic management plans and transport policies

**Keywords:** Transport Planning, Crowdsourced data, Google Distance Matrix API, Traffic Analysis, Traffic Flow Estimation, Bottleneck Identification

## Table of Content

Dedication .....	II
Acknowledgement .....	III
Abstract .....	IV
Table of Content .....	V
List of Figures .....	IX
List of Tables .....	XII
List of Abbreviations .....	XIII
List of Appendices .....	XV
1 Introduction.....	1
1.1 Context.....	1
1.2 Research Problem .....	3
1.3 Research Objectives.....	4
1.4 Organization of the thesis .....	5
2 Preliminaries .....	8
2.1 Definitions .....	8
2.2 Uses of travel time data .....	9
2.2.1 Current methods of collecting travel time data .....	10
2.2.1.1 Test vehicle techniques .....	10
2.2.1.2 License plate matching techniques.....	11
2.2.1.3 ITS-based techniques .....	12
3 Use of crowdsourced hybrid positioning in transport planning.....	17
3.1 Introduction to crowdsourced data .....	17
3.2 Use of crowdsourced data in transport engineering .....	18
3.3 Positioning systems .....	23
3.3.1 GPS-based location identification.....	23
3.3.2 Cellular-tower trilateration.....	24
3.3.3 WIFI Positioning System .....	25

3.3.4	Hybrid positioning systems .....	26
3.4	Using crowdsourced data in hybrid positioning systems .....	27
3.5	How Google collects user location .....	28
3.6	Google privacy policy on anonymous data sharing.....	30
4	Methodology of developing a travel time data mining platform .....	32
4.1	Use of mobile probes in travel time data collection .....	32
4.2	Crowdsourcing methods of travel time data collection .....	33
4.3	Collecting data from Google traffic layers .....	35
4.4	Collecting User Location from Smartphone users.....	36
4.4.1	Location collect from Android phones.....	38
4.4.2	Location information from iOS mobile devices.....	39
4.4.3	Google Service Architecture of Location-Based Services.....	40
4.5	Using the Google Distance Matrix API.....	41
4.5.1	Input Parameters to API .....	41
4.5.1.1	Required parameters.....	41
4.5.1.2	Optional parameters .....	43
4.5.2	Results Given by API.....	45
4.5.3	Example: Calling the Distance Matrix API and obtaining results ...	46
4.6	Development of data collection server. ....	47
4.6.1	Development of PHP scripts to collect data.....	47
4.6.2	Scheduling the data collection.....	51
5	Verification of Google travel time data .....	53
5.1	Evaluation of the network infrastructure of Sri Lanka. ....	53
5.1.1	Mobile usage and subscription in Sri Lanka .....	53
5.1.2	Signal availability of Sri Lanka.....	57
5.2	Using probe vehicle techniques in verifying the Google travel time .....	62
5.2.1	Development of GPS device to collect travel time data.....	62
5.2.2	The methodology followed in collecting data.....	64
5.2.3	Analysis to verify data with Google travel time.....	65

5.2.3.1	Evaluation of short distance trips - Peak time traffic .....	66
5.2.3.2	Evaluation of short distance trips – Off-peak time traffic.....	72
5.2.3.3	Evaluation of long-distance trips.....	78
5.3	Verification of Google travel time data for different vehicle types.....	83
5.3.1	The methodology of license plate survey.....	84
5.3.2	Results obtained and analysis.....	86
6	Applications of travel time data obtained from Google Distance Matrix API ..	89
6.1	Traffic flow estimation urban roads-based on Google travel time data and machine learning principles. ....	89
6.1.1	The requirement of the study .....	89
6.1.2	Literature review .....	90
6.1.2.1	Use of Machine Learning in Flow Prediction .....	91
6.1.3	Methodology .....	92
6.1.3.1	Evaluation Using Machine Learning.....	93
6.1.4	Analysis .....	94
6.1.4.1	Results and Discussion.....	94
6.1.5	Conclusion and future work in traffic flow prediction.....	99
6.2	Identification of road bottlenecks along corridors using Google travel time .....	101
6.2.1	Objectives.....	101
6.2.2	Literature review .....	102
6.2.2.1	Definitions on road bottlenecks.....	102
6.2.2.2	Bottleneck Identification .....	103
6.2.2.3	Ranking and Reliability of Bottlenecks.....	105
6.2.3	Methodology .....	106
6.2.3.1	Bottleneck identification .....	107
6.2.3.2	Ranking of Bottlenecks .....	108
6.2.4	Analysis and evaluation .....	110

6.2.5	Conclusion and future work in bottleneck identification .....	114
6.3	Using Google travel time data for evaluation of transport projects .....	115
6.3.1	Implementation of bus priority lanes in CMA .....	115
6.3.1.1	Objectives .....	115
6.3.1.2	Methodology .....	117
6.3.1.3	Analysis and evaluation .....	117
6.3.2	Implementation of reversible lanes in CMA .....	122
6.3.2.1	Methodology .....	122
6.3.2.2	Analysis and evaluation .....	124
7	Discussion and Conclusion .....	128
7.1	Summary of main findings .....	128
7.2	Evaluation of the Data Collection Method .....	130
7.3	Evaluation of the Verification method .....	131
7.4	Evaluation of the Applications .....	132
7.5	Best practices to follow .....	135
7.6	Directions to future work.....	139
7.7	Concluding Remarks and recommendations .....	140
	Reference List .....	141
	Appendix A : List of Supported Languages of Google Distance Matrix API .....	148
	Appendix B : Arduino Code for retrieval of GPS Data .....	150
	Appendix C : Spatio temporal variation of space mean speed with the implementation of bus priority lanes .....	154
	Appendix D : Data set for verification of API travel time with GPS travel time – Short Distance .....	155
	Appendix E : Data set for verification of API travel time with GPS travel time – Long Distance (50km-260km).....	160
	Appendix F : Statistical evaluation of Google traveltime data and GPS traveltime data .....	162

## List of Figures

Figure 1 : Left: Microwave vehicle detectors, Right: Inductor loop vehicle detectors .....	12
Figure 2 : Location identification from GPS .....	23
Figure 3 :Reflection of GPS signals.....	23
Figure 4 : Cellular tower trilateration .....	24
Figure 5: Collecting traffic data from Google Maps .....	35
Figure 6 : Location information collected from users by Google.....	36
Figure 7: Map of Daily activity of a mobile phone user who use Google services ...	37
Figure 8 : Location enabling options in Android phones.....	38
Figure 9: iOS settings to change accessibility of Google Services.....	39
Figure 10 : Basic method of calling the Distance Matrix API.....	41
Figure 11 : Results obtained from calling the Distance Matrix API.....	47
Figure 12 : Data collection framework .....	48
Figure 13 : Flow chart of the PHP script .....	49
Figure 14 : Input Origin-Destination file .....	50
Figure 15 : Sample of data collected.....	51
Figure 16 : Cron job scheduler used to execute the PHP script.....	52
Figure 17: Cellular phone subscriptions 1997-2017 .....	54
Figure 18 : Mobile Broadband subscriptions from 2009-2017.....	55
Figure 19 : GSMA Connectivity indexes.....	56
Figure 20 : Network coverage of Mobile network providers in Sri Lanka.....	58
Figure 21: Network signal coverage map developed by OpenSignal.....	60
Figure 22:Network Coverage map of different network providers.....	61
Figure 23 : Design diagram of the Arduino-based GPS Device .....	63
Figure 24: GPS Device mounted in a probe vehicle.....	64
Figure 25 : Map of GPS locations use for travel time evaluation of short distance peak hour traffic condition .....	67
Figure 26 : Travel time Comparison of two methods, compared to segment distance – Peak traffic.....	69
Figure 27: Comparison of Google API travel times with Probe vehicle travel times - Short Distance-Peak traffic Conditions .....	70

Figure 28: Statistical analysis of Google travel time and GPS travel time for short distance trips in peak traffic conditions.....	71
Figure 28 : Map of GPS locations use for travel time evaluation of short distance off-peak time traffic condition.....	72
Figure 30: Map of GPS locations use for travel time evaluation of short distance - peak hour traffic condition .....	73
Figure 31: Travel time Comparison of two methods, compared to segment distance – Off - peak traffic .....	75
Figure 32: Comparison of Google API travel times with Probe vehicle travel times - Short Distance-Off-peak traffic Conditions .....	76
Figure 33: Statistical analysis of Google travel time and GPS travel time for short distance trips in off-peak traffic conditions.....	77
Figure 34 : Map of GPS locations use for travel time evaluation of long-distance trips .....	80
Figure 33:Travel time Comparison of two methods in Long Distance trips .....	81
Figure 36: Comparison of Google API travel times with Probe vehicle travel times - Long Distance.....	82
Figure 37: Statistical analysis of Google travel time and GPS travel time for long distance trips .....	82
Figure 38: Map of the road used for license plate matching survey .....	84
Figure 39: Identification method of license plates.....	85
Figure 40: Temporal variation of travel time for different vehicle types .....	86
Figure 41 : Variation of travel time for different vehicle types .....	87
Figure 42: Percent change of travel time from API travel time for different types of vehicles .....	88
Figure 43:Flowchart of the Methodology of data collection for traffic flow prediction .....	92
Figure 44: Training and Test Sample Distribution .....	94
Figure 45:K- Selection Error Graph .....	95
Figure 46:Predicted Vs Observed Traffic Flow for ANN .....	96
Figure 47: Predicted Vs Observed Traffic Flow for SVR .....	96
Figure 48:Predicted and Observed Flow Values Under KNN Regression .....	97
Figure 49: Distribution of Directional and Opposite Link Speeds with Predicted Flow .....	98

Figure 50: The Percentage Gain of Each Training Dataset .....	99
Figure 51: Spatiotemporal traffic state matrix STM; HCM-2010 Transport Research Board .....	107
Figure 52: Map of the segments used for bottleneck analysis .....	111
Figure 53: Spatio-temporal graph of bottleneck formation .....	112
Figure 54: Mode share by trip purpose - .....	116
Figure 55 : Bus Priority lane implementation Rajagiriya; Diyatha Uyana – Senanayke junction.....	117
Figure 56: Travel time variation before implementation of bus priority lanes.....	118
Figure 57: Travel time variation after implementing bus priority lanes .....	119
Figure 58: Speed variation before and after implementation of bus priority lanes..	119
Figure 59: Average speed variation with the implementation of Bus Priority Lanes .....	121
Figure 60: Map of the road segment in which reversible lanes were operated.....	123
Figure 61 : Travel time variation before implementation of reversible lanes.....	124
Figure 62 : Travel time variation towards Colombo after implementation of reversible lanes .....	125
Figure 63: Travel time variation towards Battaramulla after implementation of reversible lanes .....	125
Figure 64 : Travel time saving towards Colombo after implementation of reversible lanes.....	126
Figure 65: Travel time saving towards Battaramulla after implementation of reversible lanes .....	126
Figure 66: Defining the GPS coordinate of an origin or destination .....	136
Figure 67 : Marking origin and destination correctly on the map .....	137
Figure 68 : Origin and destination selection to avoid alternative route .....	138
Figure 69 : Selection of origin and destination to include junction delay .....	138

## List of Tables

Table 1 : Cost comparison of different sensor systems .....	13
Table 2 : Comparison of different probe vehicle techniques ref : The Federal Highway Administration US .....	16
Table 3 : Cost structure of the Distance Matrix API.....	43
Table 4 : Distance Matrix API call parameters.....	46
Table 5: Statistical analysis of travel time and space mean speed for different vehicle types.....	87
Table 6: KNN Input Attributes .....	93
Table 7. Evaluation of Regression Models .....	95
Table 8. Evaluation for Model Over fit .....	98
Table 9: Performance measure of bottleneck events .....	113
Table 10: Segments in which Bus priority lane analysis was carried out.....	117

## List of Abbreviations

Acronym	Definition
ANN	Artificial Neural Network
AoA	Angle of Arrival
API	Application Program Interface
AVI	Automatic Vehicle Identification
CMA	Colombo Metropolitan Area
CSV	Comma Separated Values
DGPS	Differential Global Positioning System
ETA	Estimated Time Of Arrival
FHWA	Federal Highway Administration
GPRS	General Packet Radio Service (GSM - UMTS)
GPS	Global Positioning System
GSM	Global System For Mobile Telecommunications
HCM2010	Highway Capacity Manual 2010
HTTP	Hypertext Transfer Protocol
IBM	International Business Machines Corporation
IEEE	Institute Of Electrical And Electronic Engineers
IOS	iOS operating systems
IP	Internet Protocol
ITS	Intelligent Transport Systems
JSON	JavaScript Object Notation
KNN	K- Nearest Neighbour
LOS	Level Of Service
MAC	Media Access Control (IEEE 802)
MAE	Maximum Absolute Error
OS	Operating System
PCs	Personal Computers
PHP	Hypertext Preprocessor
RAM	Random Access Memory

RFID	Radio-Frequency Identification
RMSE	Root Mean Squared Error
RSS	Residual Sum Of Squares
RTMS	Remote Traffic Microwave Sensor
SSID	Service Set Identifier
STM	Spatio-Temporal Marix
SVR	Support Vector Regression
TDoA	Time Delay of Arrival
ToA	Time of arrival
TTF	Time To First Fix
Wi-Fi	Wireless Fidelity
WLAN	Wireless Local Area Network
XML	Extensible Markup Language

## List of Appendices

Appendix A : List of Supported Languages of Google Distance Matrix API .....	148
Appendix B : Arduino Code for retrieval of GPS Data .....	150
Appendix C : Spatio temporal variation of space mean speed with the implementation of bus priority lanes .....	154
Appendix D : Data set for verification of API travel time with GPS travel time – Short Distance .....	155
Appendix E : Data set for verification of API travel time with GPS travel time – Long Distance (50km-260km) .....	160
Appendix F : Statistical evaluation of Google traveltime data and GPS traveltime data .....	162