DEVELOPING A TRIP DISTRIBUTION MODEL FOR IDENTIFIED MOBILITY GROUPS USING BIG DATA

Buddhi Ayesha Rathnayaka

188006L

Thesis/Dissertation submitted in partial fulfillment of the requirements for the Degree of Master of Science in Computer Science and Engineering

Department of Computer Science & Engineering

University of Moratuwa

Sri Lanka

February 2020

DECLARATION

I declare that this is my own work and this 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 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:

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

Name of the Supervisor: Dr. Charith Chitraranjan
Signature of the Supervisor: Date:

Name of the Supervisor: Dr. Amal Shehan Perera
Signature of the Supervisor: Date:

Name of the Supervisor: Prof. Amal S. Kumarage
Signature of the Supervisor: Date:

ACKNOWLEDGEMENT

First and foremost, I would like to thank my supervisors Dr. Amal Shehan Perera, Dr. Charith Chitraranjan and Prof. Amal S. Kumarage for the continuous support and guidance I received in every aspect while completing this research. I could not have imagined having a better set of advisors for my research. I am highly grateful for their insightful guidance and encouragement provided throughout my research work.

I would also like to thank my progress review committee, Dr. Lochandaka Ranatunge, and Dr. Dilum Bandara for their valuable insights and guidance. Their advice helped me to improve the state of my research work.

A very special gratitude goes to Sriganesh Lokanathan (Team leader, LIRNEasia), and I also thank Lasantha Fernando (Researchers - LIRNEasia). Further, I would like to extend my gratitude towards LIRNEasia organizations for all the support given during the study period.

This work was funded by a Senate Research Committee (SRC) Grant of the University of Moratuwa. Further, I would like to thank LK Domain Registry for the support given for the conference participation via travel grants.

I am sincerely thankful to all the staff members and the non-academic staff of the Department of Computer Science and Engineering. Finally, I would like to thank my fellow lab mates in Data Science and Engineering Research Hub of University of Moratuwa and all of my friends and family. Not just academic work, I was able to enjoy my time while involving different activities and outings. It helped me to keep my life balanced. I thank my parents who have given me a very fortunate life and always believing, trusting and supporting me.

ABSTRACT

The need for frequent transportation planning has become a key factor since people started becoming more mobile making urban traffic patterns more complex. The primary source for analysing such travel behavior is through manual surveys. These surveys are expensive, time consuming and often are outdated by the time the survey is completed for analysis. To overcome these issues, Mobile Network Big Data (MNBD) which concerns large data sets can be used over such traditional data collection processes. Call Detail Records (CDR) which is a subset of MNBD is readily available as most of the telecommunication service providers maintain CDR. Thus, analyzing CDR leads to an efficient identification of human behavior and location.

However, many researches on CDRs have been done focusing to identify travel patterns in order to understand human mobility behavior. Relatively high percentage of sparse data and other scenarios like the Load Sharing Effect (LSE) causes difficulties in identifying precise location of the user when using CDR data. Existing approaches for identifying precise user location patterns have certain constraints. Past researches utilizing CDRs have used primary approaches in recognizing load sharing effects and have given minimum consideration to the transmission power of the respective cell towers when localizing the users. Furthermore, these studies have neglected the differences in mobility behavior of different segment of users and taken the entire community of users as a single cluster.

In this research, a novel methodology to overcome these limitations is introduced for locating users from CDRs by dividing the users into distinct clusters for identifying the model parameters and through enhanced identification of load sharing effects by taking the transmission power into consideration. Further, this study contributes to the transport sector by identifying secondary activities from CDR data, without limiting to the primary activity recognition. This research uses approximately 4 billion CDR data points, voluntarily collected mobile data and manually collected travel survey data to find techniques to overcome the existing limitations and validate the results.

Proposed dynamic filtering algorithm for load shared records identification showed a significant improvement on accuracy over previous predefined speed based filtering methods. Further, we found that, IO-HMM outperforms standard HMM results on activity recognition.

Keywords: Travel Demand Modeling with Mobile Network Big Data, User Localization Based on CDRs, Activities identification from CDRs.

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

CA	Call Activity
CAC	Call Activity Count
CDR	Call Detail Records
CLS	Cordon Line Surveys
DSD	Divisional Secretariat Divisions
HBW	Home-based Work
HMM	Hidden Markov Model
HVS	Household Visit Survey
IO-HMM	Input/Output Hidden Markov Model
JICA	Japan International Cooperation Agency
LSE	Load Sharing Effect
LSR	Load Sharing Records
MNBD	Mobile Network Big Data
O-D	Origin-Destination
PCA	Principal Component Analysis
SLS	Screen Line Survey
SVM	Support Vector Machine
TGS	Trip Generation Survey
TSS	Travel Speed Survey
VLR	Visitor Location Register

TABLE OF CONTENTS