HÖGSKOLAN
DALARNA

Dalarna Doctoral Dissertations

A Multi-Dimensional Approach to Human Mobility and Transportation Mode Detection Using GPS Data

PARIA SADEGHIAN

Microdata Analysis
School of Information and Engineering
Dalarna University, Borlänge, Sweden
2024

Dissertation presented at Dalarna University to be publicly examined in room Clas Ohlson, Friday, 31 May 2024 at 13:00 for the Degree of Doctor of Philosophy. The examination will be conducted in English. Opponent: Professor Kevin Heaslip (University of Tennessee).

Abstract

Sadeghian, P. 2024. A Multi-Dimensional Approach to Human Mobility and Transportation Mode Detection Using GPS Data. *Dalarna Doctoral Dissertations* 33. Borlänge: Dalarna University. ISBN 978-91-88679-62-8.

GPS tracking data is an essential resource for analyzing human travel patterns and evaluating the effects on transportation systems. The primary challenge, however, is to accurately identify the modes of transportation within unlabeled GPS data. These approaches range from simple rule-based systems to advanced machine-learning techniques. This dissertation aims to bridge this gap by examining the critical features and techniques of these methods and proposing a novel approach for detecting transportation modes in GPS tracking data. To achieve this goal, a comprehensive understanding of individual journeys is crucial. Thus, this research adopts a microdata analytic approach, encompassing data collection, processing, analysis, and decision-making stages. Doing so contributes to advancing human mobility research and transportation mode detection.

Paper I undertook a systematic review of transport mode detection methodologies to fill the research gap, emphasizing the predominance of supervised learning algorithms and highlighting the need for further research to address the limitations of small datasets. Paper II introduced a stepwise methodology, integrating unsupervised learning, GIS, and supervised algorithms to detect transport modes while minimizing reliance on labelled data. The Random Forest algorithm emerged as a precise but time-intensive solution. Paper III showcased a novel approach to transport mode detection using deep learning models, outperforming traditional machine learning methods. This paper signals the potential of deep learning in the field and demonstrates the importance of raw GPS data in enhancing accuracy. Paper V addressed the challenge of predicting human mobility patterns under the Hidden Markov Model (HMM) framework, highlighting the applicability of HMMs to understanding and predicting complex mobility behaviour. This paper emphasized the need for GPS tracking data in developing advanced mobility models. Paper IV ventured into hybrid methodology by combining K-means clustering with the ANP-PSO algorithm to enhance transportation mode classification. This pioneering approach improved classification accuracy while reducing dependence on labelled datasets.

Collectively, these papers underscore the opportunities and limitations in human mobility research, offering insights into future directions for mitigating data quality issues and improving the accuracy of transportation mode detection. These innovative methodologies have practical implications for transportation planning, resource allocation, and intelligent transportation system development, ultimately shaping the future of transportation research and decision-making. Standardized data collection, processing, and labelling methods are crucial and need attention in future research. Future research can focus on developing such benchmarks and validation protocols to enhance the reliability and comparability of results.

Keywords: Transport mode detection, Machine learning, Statistical learning, Rule-based method, Data labelling

Paria Sadeghian, Microdata Analysis

ISBN 978-91-88679-62-8
urn:nbn:se:du-48308 (http://urn.kb.se/resolve?urn=urn:nbn:se:du-48308)

To my family, friends, and well-wishers.

Acknowledgements

Completing this PhD thesis has been a profound journey that would not have been possible without the guidance, support, and encouragement of many. I am immensely grateful to my main supervisor, Johan Håkansson, for his invaluable advice, unwavering support, and insightful feedback throughout this research process. His expertise and dedication have been fundamental to my academic development and work completion. I thank Mia Xiaoyun Zhao, my second supervisor, for her expert guidance and encouragement, which enriched my research. I'm also grateful to Mengjie Han, my third supervisor, for his invaluable advice and support. Special thanks to Kenneth Carling for his invaluable support during my PhD. Grateful to everyone at Dalarna University for creating a dynamic and supportive environment.

I am also deeply thankful to my friends in Borlänge, Serena, Mahesh, Nipunika, Sahar and Hamed, who have provided me with endless encouragement and motivation throughout this journey. To my parents and my sister, Parastoo and Olof, thank you for your unwavering love and encouragement and for always being there for me. Last but certainly not least, to my love, Arman. Your endless love, patience, and support have helped me through this journey and made all the difference in my life.

This thesis is not just a reflection of my work but a testament to the collective effort and support of each of you. Thank you all from the bottom of my heart.

List of papers

This thesis comprises the following papers, referred to in the text by their Roman numerals.

I. Sadeghian, P., Håkansson, J., & Zhao, X. (2021). Review and evaluation of methods in transport mode detection based on GPS tracking data. *Journal of Traffic and Transportation Engineering (English Edition)*, *8*(4), 467-482.
II. Sadeghian, P., Zhao, X., Golshan, A., & Håkansson, J. (2022). A stepwise methodology for transport mode detection in GPS tracking data. *Travel Behaviour and Society*, *26*, 159-167.
III. Sadeghian, P., Golshan, A., Zhao, M.X. et al. A deep semi-supervised machine learning algorithm for detecting transportation modes based on GPS tracking data. *Transportation (2024).*
IV. Sadeghian, P. Enhanced Clustering Approach for Transportation Mode Classification using GPS Data and Particle Swarm Optimization. *Journal of Traffic and Transportation Engineering (English Edition) (under review)*
V. Sadeghian, P., Han, M., Håkansson, &. J Zhao, X., Sadeghian, P., Han, M., Håkansson, J., & Zhao, M. X. (2024). Testing feasibility of using a hidden Markov model on predicting human mobility based on GPS tracking data. *Transportmetrica B: Transport Dynamics*, 12(1).

My contributions to the included papers were:

Paper I – Research design, data collection, data processing, analysis, writing, and manuscript revising.

Paper II – Research design, data processing, analysis, writing, and manuscript revising.

Paper III – Research design, data processing, analysis, writing, and manuscript revising.

Paper IV – Research design, data processing, analysis, writing, and manuscript revising.

Paper V – Research design, data processing, analysis, writing, and manuscript revising.

Note: Since the papers IV and V listed above are under review, no digital version is appended to this introduction.

Papers not included in the thesis:

VI. Johari, F., Peronato, G., Sadeghian, P., Zhao, X., & Widén, J. (2020). Urban building energy modeling: State of the art and future prospects. *Renewable and Sustainable Energy Reviews*, 128, 109902.

VII. Zhang, F., Saeed, N., & Sadeghian, P. (2023). Deep Learning in Fault Detection and Diagnosis of building HVAC Systems: A Systematic Review with Meta Analysis. Energy and AI, 100235.

VIII. Mahdavi Ardestani, S. F., Adibi, S., Golshan, A., & Sadeghian, P. (2023). Factors Influencing the Effectiveness of E-Learning in Healthcare: A Fuzzy ANP Study. In Healthcare (Vol. 11, No. 14, p. 2035). MDPI.

IX. Sadeghian, P., Mojarrad BB. Towards Sustainable Mobility: Gendered and Temporal Perspectives from Falun, Sweden. Applied Geography (under review).

Contents

Introduction

This PhD thesis consists of several papers that address the challenge of accurately estimating transportation and human mobility behavior from GPS tracking data. The introduction guides the reader through the natural evolution of the papers, how they are linked, what the aims of each paper are, and the methodology used to address the research gaps in environmental and socioeconomic goals related to transport management. The concluding section of the introduction provides a succinct summary of the main purpose of the thesis and the contributions to scientific literature.

Background

Historically, the necessity for micro-level mobility, defined as the small-scale movements and activities individuals engage in within their immediate environments to navigate and interact, was paramount for human survival, allowing for exploring immediate environments and procuring resources (Nei & Roychoudhury, 1993). As societies evolved, so too did the complexity of these movements. The development of tools and the advent of agriculture enhanced human ability to sustain and expand communities, thus increasing the need for intricate daily movements for hunting, farming, and, later, for trade (Archer & Archer, 2000; Massey, 2002). The Industrial Revolution marked a fundamental shift in the patterns of daily mobility. The separation of workspaces from living areas necessitated the mass migration of people to urban centers and introduced a new era of micro-level mobility characterized by the daily commute. From the industrial revolution and to present day also saw a transformation of household activities, with many tasks outsourced to external services, thereby further diversifying the reasons and routes for daily movements (Ho & Mulley, 2013; Z. Wang et al., 2018).

Technological advancements in transportation have expanded the modes and reach of daily travel, responding to and further stimulating the demand for micro-level mobility. This evolution has not only facilitated greater movement and access to services but has also led to a significant increase in the environmental footprint of daily human activities, manifesting in heightened carbon emissions, air pollution, and traffic congestion. The progression from survival-driven movements to the complex web of daily mobility seen today reflects the profound impact of technological, societal, and economic developments on human life. However, this increased mobility comes with challenges, necessitating carefully considering its environmental implications and pursuing sustainable practices. Therefore, this thesis pivots to an in-depth analysis of daily mobility at the micro-level, examining the intricate paths individuals navigate in their everyday movements.

Micro-level human mobility involves daily, short-distance trips that occur frequently and have a significant impact on individuals and the environment. Studies conducted in the EU and the US have revealed that transportation accounts for 15 and 25 percent of households' total expenditures, respectively (Barbosa et al. 2018; Kraemer et al. 2020; Yao, Zhang, and Chen 2023). Strikingly, transportation ranks as the second-largest expense category in human life, underscoring the crucial role of mobility in daily activities. Furthermore, transportation is the second largest contributor to greenhouse gas emissions, highlighting the urgent need to understand and manage human mobility behaviour to mitigate its impact on society and the environment. This underscores the importance of accurately assessing human mobility patterns for urban planners to make informed decisions about transportation management and its impact on society and the environment.

According to the International Energy Agency (Dhakal 2009), transportation is a leading contributor to greenhouse gas emissions, with an expected increase of nearly 50% by 2030 and over 80% by 2050, largely due to intra-urban car use. This highlights the need for a sustainable transport system, as outlined in the United Nations' 2015 goals. To achieve this, it is essential to understand better travel patterns and their environmental impact (Barbosa et al. 2018; Miao et al. 2022). Understanding how people move around in cities, whether they walk, cycle, drive, or take public transport, is an important precondition for making our urban areas better places to live (Xu et al. 2022; Sadeghian et al. 2024). When we know how people prefer to travel, we can design cities that reduce traffic jams and pollution and make it easier for everyone to get around (Dabiri and Heaslip 2018).

An essential part of the urban metabolism is the population mobility pattern in the cities (Khoroshevsky and Lerner 2017). Human mobility prediction, on the other hand, relies on accurate and reliable transport mode detection to generate meaningful insights and predictions (Jitta and Klami 2017a). One of the problems in city areas is the increasing of using car as a principal transportation mode (Sadeghian et al., 2021). To reduce the environmental impact of human mobility behavior in urban areas, it is necessary to capture and predict the human mobility behavior. By combining information about travel patterns with other data sources, such as weather, social media, or event calendars, human mobility prediction can provide valuable information for urban planners, transportation operators, and emergency responders.

A potential tool that can be used in variety of field from speech recognition to understanding human mobility behavior is called Hidden Markov Model (HMM) (Mathew, Raposo, and Martins 2012; Jitta and Klami 2017b). HMMs requires detailed input data, since human mobility behavior is highly complex (Alvarez-Garcia et al. 2010). The conventional methods of collecting travel behaviour data on individuals, such as travel surveys or travel diaries, have been dominant for a long time. However, these methods can be expensive, imprecise, and have limited sample sizes (Dabiri et al. 2020; McGowen and McNally 2007; Namdarpour et al. 2022).

Over the past few decades, alternative methods for collecting travel behaviour data, such as GPS tracking devices, have emerged. These devices have the potential to collect large amounts of data with high geographical precision, continuously and over a long time (Murakami and Wagner, 1999; B. Wang et al., 2017; Zito et al., 1995).

Collected human mobility data in cities may be somewhat inaccurate, mostly due to the interference of high buildings (McGowen and McNally 2007). However, the development of GPS devices used to collect data is improving rapidly and generated data has become more accurate in recording human geographical positions (Wu, Yang, and Jing 2016). Markos & Yu (2020) have demonstrated that in an urban setting, the GPS are accurate and reliable and can serve as tool for collecting detailed information about human mobility. This improvement in data accuracy has made GPS devices an attractive alternative to traditional travel surveys and diaries for collecting mobility data.

GPS devices can also be used with other sensors, such as accelerometers, gyroscopes, and magnetometers, to determine the mode of transportation used accurately (Yang, Yao, and Jin 2015). For instance, accelerometer data can be used to distinguish between walking, cycling, and motorized transport, while GPS data can provide location and speed information to further differentiate between different types of motorized transport, such as cars, buses, or trains (Z. Wang, He, and Leung 2018). Machine learning algorithms can also be used to analyze the data and automatically identify the transport mode (Dabiri et al. 2020). Accurately identifying transport modes is crucial for transportation and city planning, as it allows for developing strategies to promote sustainable transport modes and reduce emissions. However, only a few studies have been applied HMM to predict human mobility based on GPS tracking dataset. One limitation of these studies is that the dataset used is relatively small and based on a limited number of users (Alvarez-Garcia et al. 2010).

Several studies have also explored using machine learning algorithms for transport mode detection. For instance, Lin et al. (2013) used a decision tree algorithm to classify transport modes based on GPS data, achieving an accuracy of 85%. Similarly, Patterson et al. (2003) used a Bayesian network to identify transport modes with an accuracy of up to 80%. Deep learning approaches have also been employed for transport mode detection in recent years, with promising results. For example, Li et al. (2020) used a deep neural network to classify transport modes and achieved an accuracy of over 90%. The accuracy of transport mode detection using GPS tracking data varies depending on the method and dataset used. The use of advanced machine learning algorithms, GIS data, and user input can improve accuracy.

The lack of labelled data and the time-consuming labelling process are major challenges in developing accurate and efficient transport mode detection algorithms using GPS tracking data. However, several approaches can help address these challenges. One such approach is to use semi-supervised learning techniques, where a small amount of labelled data is combined with a larger amount of unlabeled data to improve the algorithm's accuracy (Wu, Yang, and Jing 2016). Another approach is to use transfer learning, where a pre-trained model on a different dataset is fine-tuned on a smaller labelled dataset for the specific task of transport mode detection (Prelipcean, Gidófalvi, and Susilo 2017). Moreover, crowdsourcing, and participatory sensing can be used to collect labelled data cost-effectively (Gong et al. 2014). Overall, these approaches can help improve the accuracy of transport mode detection algorithms using GPS tracking data, even with limited labelled data.

Aim and Research Questions

This thesis aims to develop classification and predictive mobility tools designed to enhance the efficacy of mobility modeling. Through a comprehensive approach that includes identifying essential features for transport mode detection, creating methodologies for processing GPS tracking data, and applying machine learning techniques, this research endeavors to provide robust solutions for the accurate analysis and prediction of human mobility patterns. By leveraging high-resolution GPS data, the thesis aims to bridge the gap between current transportation planning practices and the dynamic needs of urban mobility management. Figure 1. shows the research framework and contributions for the thesis.

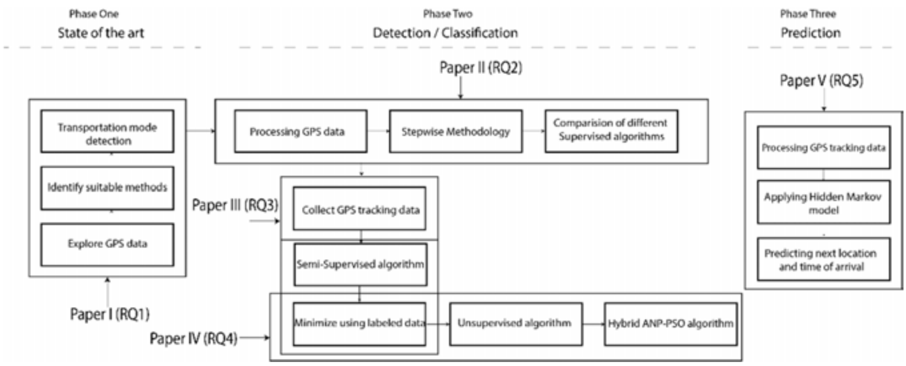

Figure 1. The Research framework and contributions for this thesis

Paper I, surveys various methods employed in the detection of transportation modes, offering insights and directions for future research. Phase two starts with Paper II which identifies transportation modes using a dataset that is fully labeled. In contrast, Paper III reduces the reliance on labeled data. Paper IV advances this approach by detecting transportation modes using a completely unlabeled dataset. Phase three, focuses on prediction of human mobility pattern with different scenarios.

The below research questions are formulated to achieve the aim of this thesis, which are addressed in the papers attached to the thesis.

1. What methodologies can be employed for identifying transportation modes via GPS tracking data, and how can the efficacy of diverse algorithms, encompassing both machine learning and statistical approaches, be evaluated?

2. How can a systematic methodology be developed to efficiently process entire GPS tracking datasets for the purpose of transport mode identification, while minimizing the need for extensive data labeling?

3. How can a novel semi-supervised deep learning algorithm be proposed for transportation mode detection that requires minimal labeled data, with the aim of improving accuracy and efficiency in classification?

4. How can transportation mode classification models be enhanced by reducing reliance on labeled data, and how can the employment of a combination of clustering and hybrid algorithms achieve high accuracy?

5. What approach would be involved in deploying a prediction-oriented algorithm under the HMM framework for the purpose of predicting human mobility patterns, and how could its effectiveness be measured across various scenarios?

This thesis identifies and addresses research gaps in the literature on micro-level human mobility analysis: the need for more accurate, efficient, and scalable methods to detect transport mode and to estimate and predict human mobility patterns using GPS tracking data. Despite advancements in GPS technology and machine learning, previous studies have often been limited by the precision of mobility data, the granularity of transport mode detection, and the scalability of data analysis methods. By integrating advanced machine learning algorithms with high-resolution GPS tracking data, this research offers novel approaches to overcome these challenges.

It provides empirical evidence on the effectiveness of these methods in various urban contexts, thereby filling a significant gap in the literature. The link between identifying transportation modes and predicting human mobility underscores the significance of comprehending how people move in urban environments. By improving our ability to detect transport modes and predict human mobility, we can create more efficient, sustainable, and resilient transportation systems that meet the needs of individuals and communities.

Synthesizing Insights Across Three Phases: A State of the Art, Classification, Detection, and Prediction in Five Key Papers

This section provides a summary of five papers. The five papers were divided into three phases, state of the art, classification and detection, and prediction. It briefly describes each paper's aim, methodology, results, and conclusion.

Phase One: State of the art (Paper I)

This paper aims to conduct a comprehensive review of the key features and techniques utilized in identifying transportation modes through GPS tracking data in order to answer the first research question. The paper examines the methods used in identifying transportation modes and compares and evaluates their effectiveness. The PRISMA methodology statement was utilized to review the subject matter systematically. Three primary categories were selected to identify relevant articles during the literature search. The search began by looking for relevant keywords and their synonyms.

Following the PRISMA statement, the paper used a systematic review methodology to gather information on transport mode detection using GPS tracking data. Three categories of keywords were used in the literature search to ensure that all relevant scientific work was included. The first category consisted of words related to transport mode detection and identification, such as travel mode detection and identifying travel patterns. The second category included keywords related to machine learning and statistical methods, such as supervised and unsupervised learning. The third category used broader terms, including GPS data processing methods, trip identification, and map matching, to ensure a comprehensive search.

In December 2019, a literature search was conducted, and the articles' metadata were saved. The eligibility criteria for inclusion in the review were that the articles must be written in English and published in a peer-reviewed scientific journal, conference proceeding, or book, and they must be relevant to the topic and focus on the subject. A total of 22 articles met the criteria and were included in the review. In Paper I, the methods and algorithms for detecting the transport mode were evaluated and discussed. The results indicated that most of the studies used supervised learning algorithms for transport mode detection.

According to the review of the articles, statistical methods, rule-based methods, and machine learning methods were found to be the dominant methods for detecting transport modes based on GPS tracking data. The paper mainly focused on machine learning methods, including supervised, semi-supervised, and unsupervised learning, and compared and evaluated the algorithms applied in transport mode detection. Semi-supervised learning was found to be a good candidate for providing accurate results with a small amount of labelled data, while unsupervised algorithms with a combination of deep learning could also be effective.

One common limitation in most reviewed articles was the use of small datasets. This often stems from the difficulty in collecting large GPS tracking datasets, attributed to challenges such as privacy concerns, resource limitations, and the specialized focus of some studies, despite the common assumption of easy access to large-scale data. Finally, further research is needed to improve the state-of-the-art method in transport mode detection.

Phase two: Detection and Classification (Paper II)

The research aimed to develop a systematic approach that can efficiently work with the entire GPS tracking dataset without requiring extensive labelling and allow the integration of various techniques to achieve optimal performance. The study aimed to create a methodology for identifying the mode of transportation using raw GPS tracking data to answer the second research question. The data collection phase involved selecting 21 individuals monitored for six months, from May to September 2017, using the Triplog mobile application ('Triplog' 2010). The application captured details such as date, time, latitude, longitude, and speed at a frequency of 30 seconds. The participants were required to carry their mobile phones throughout the day while on the move. The data was collected in Falun, a city in the central part of Sweden, and 370,000 positional records were collected from different parts of Sweden and stored in CSV format. The researchers utilized the Geographic Information System (GIS) of the public network and public transportation station to describe the transportation network for the study.

The four main phases of the methodological framework presented in this paper are depicted in Figure 2. The first phase involves the use of raw GPS data and GIS layers. The second phase, the pre-processing phase, aims to reduce noise in the data. The third phase is applying the stepwise methodology, and the last phase is detection of transportation modes. This is achieved by employing two primary techniques: filtering the data based on temporal features and addressing signal shortages. The paper provides a detailed explanation of these techniques.

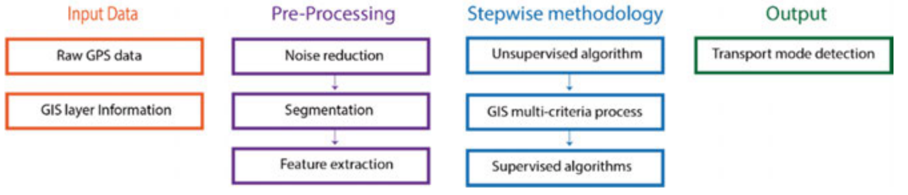

Figure 2. The framework of applied methodology

The participants' trips were divided into segments, and the relevant features for determining the transport modes were extracted. Additional features were needed to achieve a high level of accuracy in detecting the mode of transport. As a result, attributes such as average speed, average acceleration, maximum and minimum speed, acceleration during each segment, segment distance, direction, and duration were computed for each segment. The processed and clean GPS data was inputted into the stepwise methodology approach. The stepwise methodology approach for identifying the mode of transportation using an unlabeled GPS dataset. The proposed approach incorporates three methods: unsupervised learning algorithms, GIS multi-criteria process, and supervised learning algorithms. Using the same dataset, the study evaluated five common supervised learning algorithms, Decision Tree, Random Forest, K Nearest Neighbor, Naïve Bayes, and Multinomial Logistic Regression.

The paper found that the Random Forest algorithm performed better than the other supervised algorithms in detecting transport modes based on precision and F1 score. However, it takes a longer time to execute than the other algorithms. The proposed stepwise methodology was shown to be effective in reducing the amount of labelled datasets needed while achieving high accuracy in transport mode detection. The evaluation of supervised algorithms revealed that with only 20% of the labelled dataset, the Random Forest algorithm could achieve an accuracy of 99%.

According to the results, the stepwise methodology suggested in this study can effectively manage a large volume of unlabeled GPS data and accurately identify transport modes with only 10% of the data labelled. However, this study is limited because it only used one unsupervised learning algorithm and five supervised learning algorithms for testing. Additionally, the accuracy and efficiency levels of the methodology may vary depending on different situations and intended applications, as they depend on the amount of GPS tracking data used.

Phase two: Detection and Classification (Paper III)

The traditional methods of data collection for transportation research are time-consuming and error prone. GPS devices have become essential tools in understanding human mobility behavior, traffic demands, and the impacts of transportation infrastructure on people. While GPS tracking data can record spatial and temporal information on human movement during trips, it cannot indicate which transportation mode is used. Therefore, transportation mode detection studies have become an essential area of research, and several data processing methods and machine learning algorithms have been applied. The raw GPS tracking data were collected from September 2019 to September 2020 from 91 volunteers in Borlänge, a city in central Sweden (Figure 3). Each volunteer carried a GPS device, “Renkforce GPS logger,” for three weeks. The device recorded date, time, latitude, longitude, elevation, and speed every five seconds. Volunteers were requested to carry the device when they started the day and continue to the end of the day. More than five million positional recordings were recorded, and over 6% of the dataset was labelled with the correct transportation modes.

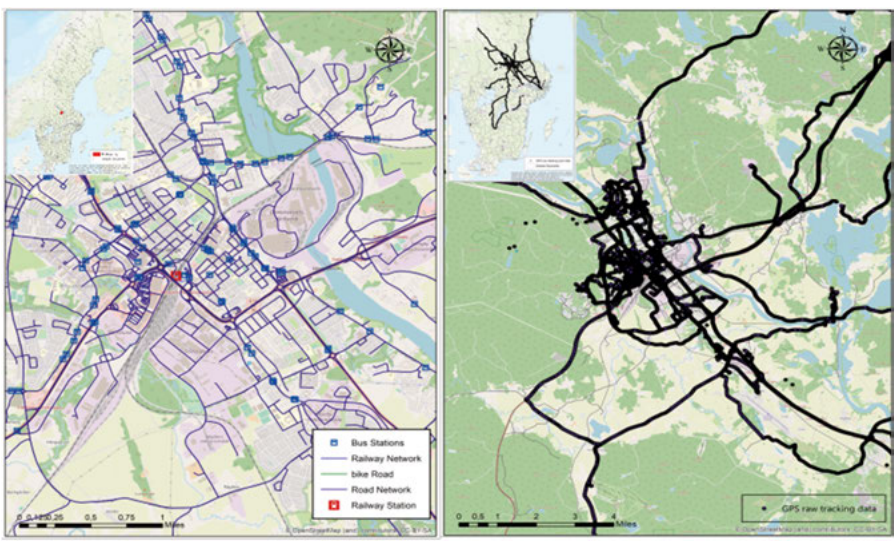

Figure 3. Input data location of Borlänge city in Sweden (a), the public transportation network in the study area (b), the raw GPS tracking data (c), and the GPS tracking data in the study area.

This paper proposes a novel semi-supervised deep learning algorithm for detecting transportation modes using minimal labelled data to answer third research question (Figure 4). The algorithm combines the Long Short-Term Memory (LSTM) Autoencoder and deep neural network architecture with three hidden layers. The proposed method can accept GPS trajectory with adjustable length and extract latent information to detect transportation modes. The accuracy of the proposed method is evaluated by applying it to different case studies, and an accuracy of 93.6% can be achieved, which outperforms similar studies.

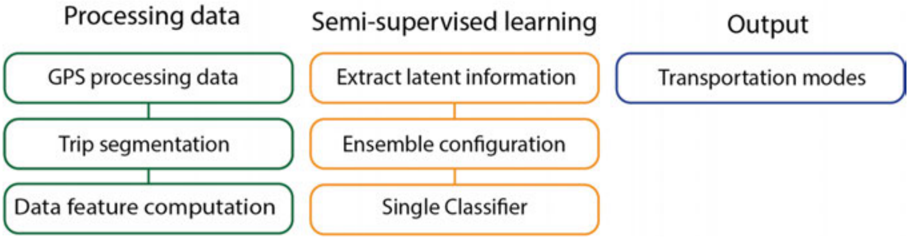

Figure 4. The framework of the proposed method

Most studies in detecting transportation modes using GPS tracking data have two steps: cleaning and processing the raw GPS tracking, removing the outliers, and using all features in an algorithm to detect transportation modes. Three main methodological approaches are commonly used in the second step: machine learning algorithms, statistical methods, and role-based algorithms. Fully supervised learning algorithms require the identification and formulation of features manually before applying a machine learning algorithm. Unsupervised learning algorithms used a fully unlabeled GPS tracking dataset to detect the correct transportation mode.

Using semi-supervised learning, only a small portion of raw GPS data needs to be labelled, which can improve efficiency and accuracy. A few studies have used the semi-supervised learning algorithm to detect the transportation modes, but they rely on a relatively old GPS tracking dataset, and the accuracy level of the proposed methods is not significant. The proposed method in this paper explores the advantages of applying semi-supervised learning methods and is applied to a large volume GPS tracking dataset.

The proposed method is a significant contribution to the field of transportation research, where a higher volume of data can be acquired with GPS tracking devices, and it can efficiently detect transportation modes with minimal labelled data. The method can be applied to a large volume GPS tracking dataset, improving the accuracy level of transportation mode detection.

Phase two: Detection and Classification (Paper IV)

In the context of contemporary urban planning and intelligent transportation systems, the accurate classification of transportation modes constitutes a pivotal facet of data-driven decision-making. This paper embarks on a journey to enhance existing models in the field, to reduce reliance on labelled data and foster a novel approach that maximizes the utilization of raw GPS data. A two-step process unfolds, commencing with applying the K-means clustering algorithm to delineate initial clusters representing distinct transportation modes to answer forth research question. Subsequently, the paper introduces the ANP-PSO hybrid algorithm, a novel amalgamation of multi-criteria decision-making and meta-heuristic optimization, to refine these clusters and achieve superior classification. Input variables encompass essential parameters such as average speed, mean speed, total distance, total time, average acceleration, mean bearing, maximum speed, and maximum acceleration. The results exhibit the successful formation of differentiated clusters via the K-means clustering algorithm, with the ANP-PSO hybrid algorithm emerging as the vanguard, elevating classification accuracy.

Moreover, the convergence trend underscores the algorithm's consistent refinement. This, however, serves as the precursor to the ANP-PSO hybrid algorithm's resounding impact. The novel hybrid algorithm enhances classification accuracy, a fact reflected in a convergence trend that points consistently upwards. In its final iteration, the algorithm attains an accuracy rate exceeding 88%, underscoring its potential to discern transportation modes based on GPS data accurately. Furthermore, a computational efficiency assessment reveals that the algorithm operates within reasonable timeframes, bolstering its applicability in real-time scenarios.

This research underscores the practical implications of the advanced model for transportation mode classification. The real-world applications of such a model are manifold, with the potential to benefit transportation planning, resource allocation, and the development of intelligent transportation systems. Accurate transportation mode detection, underpinned by raw GPS data and the ANP-PSO hybrid algorithm, offers insights to optimize transportation services, improve resource allocation efficiency, and boost overall transportation system performance. The advanced model also extends its utility to transportation research, facilitating the analysis of travel behaviour, assessment of transportation policies, and evaluation of alternative transportation modes.

In conclusion, this paper presents an innovative approach to transportation mode classification that minimizes reliance on labelled data and maximizes the use of raw GPS data. Through the strategic integration of the K-means clustering algorithm and the ANP-PSO hybrid method, the study enhances classification accuracy while offering practical applicability. The model is a testament to the possibilities of data-driven solutions and their potential to shape the future of transportation research and decision-making. This work invites the academic community and practitioners to explore and build upon these promising findings to advance the field of transportation mode detection further.

Phase three: Prediction (Paper V)

The mobility behaviour of human beings is predictable and not random, which has the potential to improve urban planning, location selection, and intelligent transportation services. Predicting human mobility patterns can also help to provide efficient and sustainable urban development strategies. However, modelling urban mobility patterns sometimes can be complex and stochastic because several factors, such as dynamic population, weather conditions The paper provides a valuable contribution to urban mobility behaviour prediction, and access to multiple transportation modes, influence mobility incentives. To answer the fifth research question, a prediction-oriented algorithm under the HMM framework has been proposed to predict the next location and time of arrival based on two schemes of human mobility. GPS-tracking data was used as input data in the models. The paper provides a valuable contribution to urban mobility behaviour prediction by proposing a novel algorithm under the framework of an HMM. The evaluation shows that the proposed algorithm is stable and consistent in predicting the next location of users based on their past trajectories.

HMMs are statistical tools applied in various fields, ranging from speech recognition to bioinformatics. While HMMs may not be as widely used for predicting human mobility patterns as they are in other areas, their flexibility and ability to model complex data sequences make them a potentially powerful tool for understanding and predicting human behaviour in various contexts. However, HMMs require detailed data input since the urban mobility behaviour of citizens is highly complex and volatile.

The paper discusses traditional approaches that rely on expensive manual surveys, which give a relatively coarse-grained description of a person's mobility. In addition, some of the commonly used data collection methods are based on stated preferences rather than revealed preferences of actual behaviour. Carrying a GPS tracker effectively records detailed outdoor positions with longitude, latitude, date, time, and speed. In combination with GIS layers, activities like going to work, shopping, leaving kids to kindergarten, and sports activities could also be identified. Although many studies have been conducted to identify human mobility patterns using GPS tracking data, a limited number of studies have implemented the data in an HMM.

The paper presents a method for predicting human mobility patterns by learning the destination choice of travelers based on large multi-day GPS tracking data in two schemes. In Scheme 1, the point of interest (POI) is a latent variable, and the considered time slot is an observable variable, while in Scheme 2, the representations are swapped. The evaluations of different schemes for the representation of latent and observable variables in HMMs are essential because the choice of representation scheme can impact the performance of an HMM. Various schemes can lead to different levels of accuracy, computational efficiency, and robustness to noise and uncertainty. By comparing the performance of various schemes under a range of conditions and scenarios, a better understanding of their strengths and weaknesses can be gained, and more informed decisions about which scheme to use in a particular application can be made. The paper provides a valuable contribution to urban mobility behaviour prediction, and the results indicate that the proposed method outperforms in both schemes compared to previous studies.

Contributions

This thesis contributes to research in multiple areas. A GPS tracking device is used to capture a large and recently collected set of human mobility behaviors. The performance of various machine learning models, from supervised to deep learning, was evaluated in detecting transportation modes (phase two). In addition to detecting transportation modes, this work also discusses understanding and predicting human mobility behaviors. Human mobility was predicted using the HMM model, based on a GPS tracking dataset (phase three). The proposed model designs a prediction-oriented algorithm based on HMM to predict the next destination and time of arrival.

Microdata Analysis

The contribution to the subject of Microdata Analysis (MDA) of this thesis is discussed in this section. In doing so, focus the discussion on the data collection, data processing, data analysis and decision making.

Data Collection

This thesis's core centers on utilizing high-resolution GPS tracking data to dissect daily human mobility patterns. The collection and analysis of this data, primarily engaged in from papers II through V, are pivotal to our research. Our methodology leveraging publicly available datasets but also, where necessary, generating new datasets tailored to explore specific aspects of human mobility. The unique contribution of this research lies in its approach to data collection and processing. In this study, we collected rather large and recent GPS tracking data from 91 individuals over a year, amounting to hundreds of gigabytes. Each participant carried a GPS device for three weeks to gather this data. The dataset is a mix of data labeled with correct transportation modes and unlabeled data. This extensive dataset helps us understand how people move and utilize various modes of transportation, which is essential for improving city transport plans and developing more eco-friendly travel options.

Data Processing

The collected raw GPS data undergoes significant assessment and transformation to prepare it for analysis. This involves processing raw GPS signals to extract meaningful information, such as speed, direction, and mode of transport. Through Papers II to V, the focus is on developing methodologies and algorithms that can accurately interpret these signals, transforming them into a format that can be analyzed to detect transport modes. This transformation process often involves filtering noise, interpolating missing data, and preparing the dataset to be used in different prediction models.

Data Storage

The thesis also focuses on effective and ethical data storage solutions for managing large GPS datasets. It emphasizes scalable, secure storage methods that respect privacy by anonymizing GPS data, ensuring it remains useful for analysis without compromising personal privacy. The chosen storage approach is ethically driven and ensures data integrity, making it easily accessible for analysis. This strategy, although not technologically innovative, is crucial for maintaining data quality and aligns with best data management practices, highlighting the thesis's commitment to ethical data handling.

Data Analysis

Each paper within this thesis employs various analytical methods. The first paper utilizes PRISMA methodologies to review and assess the gathered articles. The second paper introduces a stepwise methodology aimed at optimizing data labeling in transport mode detection studies. The third paper employs a novel semi-supervised algorithm for identifying transportation modes. The fifth paper applies the Hidden Markov Models (HMMs) to track and forecast human mobility, including the next destination and time of arrival. Meanwhile, Paper IV delves into a hybrid methodology aimed at boosting the accuracy of transport mode classification by implementing the ANP-PSO algorithm to manage unlabeled GPS datasets effectively.

Reports, and Decisions

This thesis demonstrates how the strategic handling of GPS data; from collection to analysis guides decisions towards sustainable transportation. It explores transport mode detection and human mobility, with notable studies introducing algorithms for predicting movements (Paper V) and improving transport mode classification (Papers II, III, IV). These efforts enhance urban transport planning, support efficient public networks, and advocate for eco-friendly mobility. The thesis sets benchmarks for algorithm evaluation, facilitating informed choices about their application. By linking detailed data analysis with actionable insights, it plays a role in advancing transportation systems that are both sustainable and user-centric, influencing decision-making in transportation planning and strategy development.

Conclusion

This thesis provided insights into prediction of transport mode detection and human mobility behaviour in using GPS tracking data. The collective contributions of this research illuminate a path from traditional machine learning frameworks toward more efficient deep learning methods, which are less dependent on extensive labelled GPS datasets and can be more adaptable with a small-labelled dataset.

A key contribution is the development of a novel methodology that integrates unsupervised learning, geographic information system (GIS) multi-criteria analysis, and supervised learning algorithms. This approach enhances the accuracy of transport mode detection and substantially reduces the need for labelled data, a common bottleneck in the field. Adopting deep learning models, such as Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks, marks another significant advancement, showcasing superior performance in capturing the nuances of spatial-temporal data compared to traditional machine learning methods.

The exploration of Hidden Markov Models (HMM) for predicting human mobility patterns introduces another predictive dimension to the thesis, extending its utility in urban planning and transportation management. This forward-looking perspective not only improves understanding of current mobility trends but also aids in anticipating future movements, enhancing the strategic planning of urban infrastructures. By providing more concrete examples, such as the potential for optimized route planning, improved traffic management, and enhanced public transport services, this discussion aims to clarify the substantial societal benefits that can be derived from the research findings.

Future Works

Based on the thesis suggestions for future research is the need for larger, more diverse datasets presents a call to action for the research community to expand data collection efforts, potentially through collaborative or crowdsourced initiatives. Exploring semi-supervised and unsupervised learning algorithms as viable alternatives to traditional supervised methods opens new avenues for reducing the reliance on labelled data, encouraging further innovation in algorithm development.

In conclusion, this thesis lays the groundwork for transformative changes in collecting, processing, and analyzing GPS tracking data for transport mode detection and predicting human mobility behaviour. By highlighting potential areas for future research, such as algorithm optimization, data scalability, and real-world application, it sets the stage for continued advancements in the field, driving towards more informed and effective urban planning.

References

Alvarez-Garcia, J A, J A Ortega, L Gonzalez-Abril, and F Velasco. 2010. 'Trip Destination Prediction Based on Past GPS Log Using a Hidden Markov Model'. *Expert Systems with Applications* 37 (12): 8166–71. https://doi.org/https://doi.org/10.1016/j.eswa.2010.05.070.

Barbosa, Hugo, Marc Barthelemy, Gourab Ghoshal, Charlotte R James, Maxime Lenormand, Thomas Louail, Ronaldo Menezes, José J Ramasco, Filippo Simini, and Marcello Tomasini. 2018. 'Human Mobility: Models and Applications'. *Physics Reports* 734: 1–74.

Dabiri, Sina, and Kevin Heaslip. 2018. 'Inferring Transportation Modes from GPS Trajectories Using a Convolutional Neural Network'. *Transportation Research Part C: Emerging Technologies* 86 (December 2017): 360–71. https://doi.org/10.1016/j.trc.2017.11.021.

Dabiri, Sina, Nikola Marković, Kevin Heaslip, and Chandan K. Reddy. 2020. 'A Deep Convolutional Neural Network Based Approach for Vehicle Classification Using Large-Scale GPS Trajectory Data'. *Transportation Research Part C: Emerging Technologies* 116 (May): 102644. https://doi.org/10.1016/j.trc.2020.102644.

Dhakal, Shobhakar. 2009. 'Urban Energy Use and Carbon Emissions from Cities in China and Policy Implications'. *Energy Policy* 37 (11): 4208–19.

Gong, Lei, Takayuki Morikawa, Toshiyuki Yamamoto, and Hitomi Sato. 2014. 'Deriving Personal Trip Data from GPS Data: A Literature Review on the Existing Methodologies'. *Procedia-Social and Behavioral Sciences* 138 (0): 557–65.

Jitta, Aditya, and Arto Klami. 2017a. 'Partially Hidden Markov Models for Privacy-Preserving Modeling of Indoor Trajectories'. *Neurocomputing* 266: 196–205. https://doi.org/https://doi.org/10.1016/j.neucom.2017.05.035.

———. 2017b. 'Partially Hidden Markov Models for Privacy-Preserving Modeling of Indoor Trajectories'. *Neurocomputing* 266: 196–205. https://doi.org/https://doi.org/10.1016/j.neucom.2017.05.035.

Khoroshevsky, Faina, and Boaz Lerner. 2017. 'Human Mobility-Pattern Discovery and next-Place Prediction from GPS Data'. In *Multimodal Pattern Recognition of Social Signals in Human-Computer-Interaction: 4th IAPR TC 9 Workshop, MPRSS 2016, Cancun, Mexico, December 4, 2016, Revised Selected Papers 4*, 24–35. Springer.

Kraemer, Moritz U G, Chia-Hung Yang, Bernardo Gutierrez, Chieh-Hsi Wu, Brennan Klein, David M Pigott, Open COVID-19 Data Working Group†, Louis Du Plessis, Nuno R Faria, and Ruoran Li. 2020. 'The Effect of Human Mobility and Control Measures on the COVID-19 Epidemic in China'. *Science* 368 (6490): 493–97.

Li, Linchao, Jiasong Zhu, Hailong Zhang, Huachun Tan, Bowen Du, and Bin Ran. 2020. 'Coupled Application of Generative Adversarial Networks and Conventional Neural Networks for Travel Mode Detection Using GPS Data'. *Transportation Research Part A: Policy and Practice* 136 (March): 282–92. https://doi.org/10.1016/j.tra.2020.04.005.

Lin, Miao, Wen-jing Hsu, and Zhuo Qi Lee. 2013. 'Detecting Modes of Transport from Unlabelled Positioning Sensor Data'. *Journal of Location Based Services* 7 (4): 272–90. https://doi.org/10.1080/17489725.2013.819128.

Markos, Christos, and James J.Q. Yu. 2020. 'Unsupervised Deep Learning for GPS-Based Transportation Mode Identification', 1–6. https://doi.org/10.1109/itsc45102.2020.9294673.

Mathew, Wesley, Ruben Raposo, and Bruno Martins. 2012. 'Predicting Future Locations with Hidden Markov Models'. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, 911–18.

McGowen, Patrick, and Michael McNally. 2007. 'Evaluating the Potential to Predict Activity Types from GPS and GIS Data'. In *Transportation Research Board 86th Annual Meeting, Washington*. Citeseer.

Miao, Qing, Min Li, Wenhui Lin, Zhigang Wang, Huiqin Shao, Junwei Xie, Nanfei Shu, and Yuanyuan Qiao. 2022. 'Human Mobility Prediction with Calibration for Noisy Trajectories'. *Electronics* 11 (20): 3362.

Murakami, Elaine, and David P Wagner. 1999. 'Can Using Global Positioning System (GPS) Improve Trip Reporting?' *Transportation Research Part c: Emerging Technologies* 7 (2–3): 149–65.

Namdarpour, Farnoosh, Mahmoud Mesbah, Amir H Gandomi, and Behrang Assemi. 2022. 'Using Genetic Programming on GPS Trajectories for Travel Mode Detection'. *IET Intelligent Transport Systems* 16 (1): 99–113.

Patterson, Donald J, Lin Liao, Dieter Fox, and Henry Kautz. 2003. 'Inferring High-Level Behavior from Low-Level Sensors'. In *UbiComp 2003: Ubiquitous Computing*, edited by Anind K Dey, Albrecht Schmidt, and Joseph F McCarthy, 73–89. Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/https://doi.org/10.1007/978-3-540-39653-6_6.

Prelipcean, Adrian C, Gyözö Gidófalvi, and Yusak O Susilo. 2017. 'Transportation Mode Detection–an in-Depth Review of Applicability and Reliability'. *Transport Reviews* 37 (4): 442–64.

Sadeghian, Paria, Arman Golshan, Mia Xiaoyun Zhao, and Johan Håkansson. 2024. 'A Deep Semi-Supervised Machine Learning Algorithm for Detecting Transportation Modes Based on GPS Tracking Data'. *Transportation*. https://doi.org/10.1007/s11116-024-10472-x.

Sadeghian, Paria, Johan Håkansson, and Xiaoyun Zhao. 2021. 'Review and Evaluation of Methods in Transport Mode Detection Based on GPS Tracking Data'. *Journal of Traffic and Transportation Engineering (English Edition)* 8 (4): 467–82. https://doi.org/https://doi.org/10.1016/j.jtte.2021.04.004.

'Triplog'. 2010.

Wang, Bao, Linjie Gao, and Zhicai Juan. 2017. 'Travel Mode Detection Using GPS Data and Socioeconomic Attributes Based on a Random Forest Classifier'. *IEEE Transactions on Intelligent Transportation Systems* 19 (5): 1547–58.

Wang, Zhenzhen, Sylvia Y He, and Yee Leung. 2018. 'Applying Mobile Phone Data to Travel Behaviour Research: A Literature Review'. *Travel Behaviour and Society* 11: 141–55. https://doi.org/https://doi.org/10.1016/j.tbs.2017.02.005.

Wu, Linlin, Biao Yang, and Peng Jing. 2016. 'Travel Mode Detection Based on GPS Raw Data Collected by Smartphones: A Systematic Review of the Existing Methodologies'. *Information* 7 (4): 67. https://doi.org/10.3390/info7040067.

Xu, Yang, Dan Zou, Sangwon Park, Qiuping Li, Suhong Zhou, and Xinyu Li. 2022. ‘Understanding the Movement Predictability of International Travelers Using a Nationwide Mobile Phone Dataset Collected in South Korea’. *Computers, Environment and Urban Systems* 92: 101753. https://doi.org/https://doi.org/10.1016/j.compenvurbsys.2021.101753.

Yang, Fei, Zhenxing Yao, and Peter J Jin. 2015. ‘Gps and Acceleration Data in Multimode Trip Data Recognition Based on Wavelet Transform Modulus Maximum Algorithm’. *Transportation Research Record* 2526 (1): 90–98. https://doi.org/https://doi.org/10.3141%2F2526-10.

Yao, Yuhao, Haoran Zhang, and Qi Chen. 2023. ‘Trip Segmentation and Mode Detection for Human Mobility Data’. In *Handbook of Mobility Data Mining*, 97–115. Elsevier.

Zito, Rocco, G D’este, and M A P Taylor. 1995. ‘Global Positioning Systems in the Time Domain: How Useful a Tool for Intelligent Vehicle-Highway Systems?’ *Transportation Research Part C: Emerging Technologies* 3 (4): 193–209.