FEDERAL UNIVERSITY OF PARÁ
INSTITUTE OF TECNOLOGY
GRADUATE PROGRAM IN ELECTRICAL ENGINEERING

MATHEUS MORAES DE BRITO

PERSONALIZED ROUTE SELECTION METHODS IN A URBAN COMPUTING SCENARIO

DM: 09/23

Belém

2023

MATHEUS MORAES DE BRITO

PERSONALIZED ROUTE SELECTION METHODS IN A URBAN COMPUTING SCENARIO

Master thesis submitted to the judging panel at the Federal University of Pará as part of the requirements for obtaining a Master's Degree in Electrical Engineering in the area of Applied Computing.

Advisor: Dr. Denis Lima do Rosário

Co-advisor: Dr. Eduardo Coelho Cerqueira

Belém

2023

Dados Internacionais de Catalogação na Publicação (CIP) de acordo com ISBD
Sistema de Bibliotecas da Universidade Federal do Pará
Gerada automaticamente pelo módulo Ficat, mediante os dados fornecidos pelo(a) autor(a)

B862p Brito, Matheus Moraes de.
PERSONALIZED ROUTE SELECTION METHODS IN A URBAN COMPUTING SCENARIO / Matheus Moraes de Brito. — 2023.
65 f. : il. color.

Orientador(a): Prof. Dr. Denis Lima do Rosário
Coorientador(a): Prof. Dr. Eduardo Coelho Cerqueira
Dissertação (Mestrado) - Universidade Federal do Pará, Instituto de Tecnologia, Programa de Pós-Graduação em Engenharia Elétrica, Belém, 2023.

1. Seleção de rotas. 2. Computação urbana. 3. Mobilidade urbana. I. Título.

CDD 004.072

UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

"PERSONALIZED ROUTE SELECTION METHODS IN A URBAN COMPUTING SCENARIO"

AUTOR: **MATHEUS MORAES DE BRITO**

DISSERTAÇÃO DE MESTRADO SUBMETIDA À BANCA EXAMINADORA APROVADA PELO COLEGIADO DO PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA, SENDO JULGADA ADEQUADA PARA A OBTENÇÃO DO GRAU DE MESTRE EM ENGENHARIA ELÉTRICA NA ÁREA DE COMPUTAÇÃO APLICADA.

APROVADA EM: 22/03/2023

BANCA EXAMINADORA:

Prof. Dr. Denis Lima do Rosário
(Orientador – PPGEE/UFPA)

Prof. Dr. Eduardo Coelho Cerqueira
(Avaliador Interno – PPGEE/UFPA)

Prof. Dr. Marcos César da Rocha Seruffo
(Avaliador Interno – PPGEE/UFPA)

Prof. Dr. Iago Lins de Medeiros
(Avaliador Externo – Pesquisador/UFPA)

VISTO:

Prof. Dr. Diego Lisboa Cardoso
(Coordenador do PPGEE/ITEC/UFPA)

I devote this work to my family, Daniel, Aucilene (in memoriam), and Ana Maria, who always believed in my potencial and supported me my whole life.

Dedico este trabalho à minha família, Daniel, Aucilene (in memoriam) e Ana Maria, que sempre acreditaram no meu potencial e me apoiaram durante toda a minha vida.

Acknowledgement

I dedicate my sincere thanks to:

– To Professor Denis Rosário for the guidance, support, and patience he offered me to achieve academic maturity.

– To Professor Eduardo Cerqueira, for the precious methodology knowledge and pieces of advice in the research field and life.

– To all my GERCOM colleagues, especially Camilo, Bruno, Iago, Lucas, and Hugo, who stood late at night with reviews and tips.

– To my friends and Adiele, who gave me hope and joy to face life's challenges.

– To the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPQ) for financed this work.

– To the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001.

– To the Fundação Amazônia de Amparo a Estudos e Pesquisas (FAPESPA) for financed this work under grant 054/2021.

"If everyone is moving forward together, then success takes care of itself"

Henry Ford

Resumo

Resumo da Dissertação apresentada à UFPA como parte dos requisitos necessários para obtenção do grau de Mestre em Engenharia Elétrica.

Context-aware Route Selection for Urban Computing

Orientador: Denis Lima do Rosário
Palavras-chave: Seleção de Rotas; Computação Urbana; Mobilidade Urbana

Com o crescimento populacional nas áreas urbanas, a infraestrutura urbana mais extensa enfrenta diversos problemas que afetam a saúde e a qualidade de vida da população. As mudanças na mobilidade urbana tornaram-se intensas, pois a revolução tecnológica mundial trouxe diversas ferramentas e métodos para prevenir tendências nocivas ao transporte urbano. Nesse contexto, as soluções de Internet das Coisas (IoT) realizam uma forma ubíqua de perceber a mobilidade da população e o contexto de mobilidade local como criminalidade, acidentes e qualidade do ar próximo à infraestrutura viária, complementando a mobilidade da cidade. Da mesma forma, as Redes Sociais baseadas em Localização (LBSN) dispõem de dados geolocalizados dos usuários, permitindo a identificação de padrões de mobilidade, fluxos de tráfego e recomendações de modais alternativos de transporte. Nesse sentido, novas soluções de mobilidade devem atender às questões da cidade em relação ao transporte público, criminalidade, fatores que influenciam o tráfego e comprometimento da qualidade do ar. Além disso, os métodos de seleção de rotas devem considerar características de conforto, tornando as viagens urbanas mais agradáveis. Esta dissertação de mestrado propõe e avalia duas abordagens de seleção de rotas conscientes da poluição, um método de rotas híbridas multimodais e um método de seleção de rotas personalizado multicritério, para melhoria do fluxo de mobilidade dos cidadãos urbanos. A solução multimodal híbrida supera o monomodal, oferecendo opções de viagens menos exorbitante e menos poluídas. Os perfis personalizados da solução multicritério superam a escolha de um único critério no mesmo contexto considerando todas as possibilidades de rota calculadas.

Abstract

Abstract of Dissertation presented to UFPA as a partial fulfillment of the requirements for the degree of Master in Electrical Engineering.

Context-aware Route Selection for Urban Computing

Advisor: Denis Lima do Rosário

Keywords: Route Selection; Urban Computing; Smart Mobility

With population growth in urban areas, the more extensive city infrastructure faces several problems affecting the population's health and quality of life. Urban mobility changes became intense since the worldwide technological revolution brought many tools and methods to prevent harmful tendencies regarding urban transportation. In this context, Internet of Things (IoT) solutions perform a ubiquitous way of sensing the population mobility and the local mobility context as criminality, accidents, and air quality near the road infrastructure, complementing the city mobility. Likewise, Location-based Social Networks (LBSN) dispose of users' geolocated data, allowing the identification of mobility patterns, traffic flows, and alternative modal transport recommendations. In this matter, novel mobility solutions must attend city issues regarding public transportation, criminality, traffic influencing factors, and air quality compromising. Also, route selection methods must consider comfort features, making more pleasant urban trips. This master's dissertation proposes and evaluates two pollution-aware route selection approaches, a multi-modal hybrid routes method and a multi-criteria personalized route selection method, for urban citizens' mobility flow improvement. The hybrid multi-modal solution surpasses the single-modal, offering less expensive and less polluted trip options. The multi-criteria solution personalized profiles outperform the single-criterion choice in the same context considering all calculated route possibilities.

Contents

List of abbreviations

AHP	Analytic Hierarchy Process
APE	Air Pollution Exposure
API	Application Programming Interface
CA	Fuel Consumption
CC	Energy Consumption
CI	Consistency Index
CO	Carbon Monoxide
CR	Consistency Ratio
CNN	Convolutional Neural Networks
EC	Emitted Carbon
ECO_2	Carbon Dioxide Emission
GHG	Greenhouse Gas
GIS	Geographic Information Systems
GRVAD	Geographic Routing method based on Velocity, Angle, and Density
HPV	Hired Private Vehicle
HRP	Healthier Route Planning
ICT	Information and Communication Technology
IPCC	Intergovernmental Panel on Climate Change
LAQN	London Air Quality Network
LBSN	Location Based Social Networks
MCDA	Multi-criteria Decision Analysis
MCDM	Multi-criteria Decision Making
NO_2	Nitrogen Dioxide
NO_x	Nitrogen Oxides
O_3	Ozone
OD	Origin-Destination
PDFKS	Percent Deviation From a Known Standard

PM	Particulate Matter
POI	Point-of-Interest
QC	Carbon Content
R2V	Road to Vector
RI	Random Index
SO_x	Sulfur Oxides
TJ	Tera-joule
UN	United Nations
VANET	Vehicular Ad Hoc Networks

List of Symbols

tEP	Fuel amount measurement to an equivalent oil ton
Fconv	Fuel calorific value
Fcorr	Calorific Value Correction Factor
GgC	Carbon Gigagram
R	Latitude-longitude pair path set
C	Path alternatives
S	London sensor set
$p[f]$	Average value from f
p	Selection Profile
f	Trip feature
$std[f]$	Known Standard

List of Figures

List of Tables

CHAPTER 1

Introduction

This chapter introduces the route selection need in the urban mobility context, and also how tools can provide a smart user mobility for either single or multi-modal. Furthermore, we outline this dissertation's main contributions and text organization.

1.1 Overview

Urban growth had exponential growth in all sectors since the migration began from rural areas, and citizens concentrated in large cities. The culture of productivity established the accelerated rhythm of the urban routine, where all citizens work in their designated position to advance society and quality of life [1]. The population demands benefits and facilities that urban areas can provide, which leads to greater demand for space occupied by housing and work infrastructure. Hence, vertical expansion becomes a strong trend with residential and business buildings [2].

The urban scenario verticalization benefits housing and businesses, but urban transport mobility needs to improve the infrastructure of obsolete urbanization projects. Numerous types of mobility are growing amidst the mobility-engineered highway system of the past. From transporting people to transporting goods, whether for work or a simple tourist trip, urban transport has become an essential sector in large cities that does not stop working; it only has a decrease in flow at night or on holidays [3].

Different types of transportation, such as walking, cycling, public transit, and private vehicles, have been used nowadays. These different transportation modes compose the urban mobility, where an efficient city mobility and transportation corroborate to economic growth, better social relations, and the population's quality of life [4]. This trend brings problems to the mobility fluidity, compromising productivity, with significant traffic jams that affect air quality. Such effects also decrease the quality of life, causing respiratory problems and stress for citizens. Hence, it is important to efficient use the different transportation modes based on user, city, and transportation context information.

In this context, the modernization and popularity of connected devices bring a plethora of data to be used in different areas, such as, urban mobility. For instance, many web-connected things interact among them and users, resulting in extensive data acquisition. In addition, the Internet of Things (IoT) represents systems and physical objects interconnected to the Internet for data exchange between heterogeneous devices without human intervention [5] [6]. Thus, IoT applied to urban mobility can present smart elements, such as traffic lights, parking lots, and flow management, which provide a vast amount of data for a experience-aware services [7] [8].

The Information and Communication Technologies (ICT) provides a set of services and applications, changing the urban mobility scenario [9]. The ICT advance brought mobile devices and internet access proliferation for the population. Traffic information, public transportation schedule, and ride-hailing services ease urban travel decisions. Beyond that, the connected devices permit a larger data-gathering, with the pervasive sensing advance in human behavior and transportation infrastructure.

1.2 Motivation and Challenges

The extensive urbanization process and consumer demand in many world scenarios cause a delay in urban mobility evolution and represent significant global challenges. Passenger and goods transportation services imply fleet growth, causing traffic congestion and poor air quality in city areas. However, urban mobility flow improvement relies on economic-ecological sustainability allied to drivers' satisfaction attainment [10]. Smart city services benefit urban life by reducing the population's exposure by redirecting to less concentrated vehicle pollutants areas.

The IoT technologies represent one of the key enabling subject to ubiquitous computing, referring to physical network objects with sensors to connect between them and exchange information [11]. The connected ecosystem combines data from physical and online world, for human behaviour identification. Thus, the information gathering provides knowledge to develop mobility, society, economy, and culture solutions [12] [13]. The information generation in urban context for users and infrastructure became a part of the urban context, specially in crisis' times.

The personalized experience-aware route selection method assists the decision-making on routes focusing on the user experience, considering multiple factors in the urban scenario, affecting comfort and health. Developers must consult and attend the Technology, policy, community, and environment sectors to improve city transportation and reach the smart mobility status [14]. Through the information technologies implementation, the data processing and exchange between sectors contribute to urban mobility method solutions, which address all quality of life and business-affecting factors when implemented in the city environment.

The traditional vehicular navigation systems dispose of much context information about traffic and hazards along the way, used to assist drivers [15]. Although, the typical navigation applications, ordinarily available in mobile devices and onboard car systems, do not consider the area criminality level and accidents' historical level. The contextual data scarcity may lead to potentially dangerous and unpleasant paths, risking user integrity [16]. Moreover, our method permits personalizing the route selection, allowing users to apply their distinct preferences, aiming at time-saving, eco-friendly, or less dangerous urban trips. Therefore, the Multi-criteria Decision Making (MCDM) method application helps the route selection problem, considering some of the most important criteria which affect trips in a simpler and accessible way. Further, The method deals with the route alternatives, which more can be created depending on travel distance and city context [17].

Besides that, the growth in the urban vehicular fleet motivates the alternative path suggestion, as the number of cars increase with the urban population growth. The inefficient transportation systems in greater city scenarios impact in larger congestion, longer travel time, and negative impacts to the environments [18]. In this context, Multi-modal transportation offers an alternative for the population, with economic relief and less commute time, combining different transportation modes and their features [19]. The bus trips may be cheaper but represent a bigger walking distance and waiting time, when Hired Private Vehicles (HPV) represent a time-saving way but are much more expensive. The hybrid multi-modal method with bus and HPV suggests balanced routes between economic and time-efficient [20].

The poor air quality in the city scenario affects the quality of life and threatens the environment. Vehicular emissions, industrial activity, and dust contribute to air pollution rise. The most common pollutants in urban areas are particulate matter (PM), nitrogen oxides (NO_x), sulfur oxides (SO_x), ozone(O_3), carbon monoxide (CO), and carbon dioxide (CO_2). The last represents 75% of global greenhouse gas (GHG) pollution, with the transportation sector as the primary source of emissions [21].

In this context, smart mobility solutions aim to reduce the population's exposure to an urban area with a high concentration of pollutants, preserving the quality of life of its citizens. For instance, users could rely on a service to select less polluted routes for smart mobility since air pollution impacts the inhabitants' health. Methods to identify pollution use active governmental sensors at specific points for collection and subsequent analysis, allowing the identification of low air quality areas. However, regarding the cost of implementation, there is a need to use passive sensing through existing services, such as emission calculations, if the open data portals do not share the air quality data for analysis [22].

1.3 Goals

This master thesis presents two route selection methods for urban areas as a solution for more comfortable, healthier, securer, and eco-friendly paths. We integrate a multi-modal routing method with a pollution calculation, combining public transportation with HPV for economical, efficient, and eco-friendly trips [9]. In addition, we present a personalized multi-criteria route selection with comfort, security, and air quality features to suggest better urban paths based on different user preferences [16].

Thus, the work objectives includes:

- Present the multi-modal route selection method with air pollution calculation.
- Compare the hybrid routes with single-modal routes in economic, trip-related and air quality features.
- Present the multi-criteria route selection method application.
- Compare the personalized profiles selection with greedy simpler preferences, selecting balanced routes for each user profile.

1.4 Contribution

This master thesis has the following main contributions:

- Route selection methods based on economic, comfort, security, and health features for urban mobility environment.
- Combination between urban transportation modes, proposing alternative routes for better time and cost management.
- Efficient local context open data use for route selection based on user profiles preferences, achieving faster, healthier, or more pleasant routes.
- Algorithm evaluation of the proposed selection methods compared to the state-of-art.

1.5 Text Organization

The remaining of this document is structured as follows:

- Chapter 2: presents the theoretical references about urban computing and their applications. Also, explains the AHP method for selection and the evaluation metrics for both proposed methods.

- Chapter 3: presents the state-of-art related works about multi-modal routing and multi-criteria route selection, highlighting their differences.
- Chapter 4: Details the multi-modal routing method, explaining the performance evaluation.
- Chapter 5: Describes the multi-criteria route selection and the user profiles evaluation.
- Chapter 6: Concludes this work.

CHAPTER 2

Basic Concepts

This chapter presents the main concepts and paradigms of urban computing and smart mobility. The geolocated data on social networks and open data repositories is discussed in the importance of urban mobility solutions for considering local contextual data and air quality. After that, the route selection for car navigation applications characteristics is described.

2.1 Urban Computing

A United Nations (UN) report shows that 55% of the global population lives in urban areas nowadays, and prospects warn the growth to 68% until 2050 [23]. In this matter, large urban centers need more infrastructure with expressive population growth. Transportation, housing, water, and energy supplies need solutions to avoid overload, compromising the constant rhythm. The urban computing concept mainly develop solutions for different urban areas, integrating technologies to assist and help the correct resource allocation, without restructuring the whole system [24].

Urban computing also uses diverse data types to understand, manage and improve urban environments. The IoT and Ubiquitous technologies deal with data collection, processing, analysis, and dissemination about their inhabitants and city authorities. Urban planners apply computer science, statistics, and data analysis through many big data sources, including social media, mobile devices, sensors, and others. In this way, they develop tools for proper management of the urban scenario [25].

Some solutions and methods focus on traffic improvement, natural disaster prediction, and mitigation or public safety enhancement. New urban services and applications emerge to assist people in better navigating around the city, easing the chaotic daily life, such as mobile mapping, public transportation, smart parking, and pedestrian navigation. These services dispose of much relevant urban information to aid the population in route planning [26].

Besides that, urban services also help understand and mitigate urban pollution since it is a great concern in many global cities. Urban pollution refers to air, water, and soil contamination around urban areas due to anthropological actions, such as transportation, industrial production, and energy generation [27]. Urban computing services help with environmental issues by integrating data collection about pollution levels, simulating urban solutions for further application, and enabling intelligent transportation systems for congestion mitigation and electric vehicle implementation [8].

The Location-Based Social Networks (LBSN) is a definitive solution applied to urban computing services. It generates one specific data type in every user interaction with social media and networks. Each interaction generates geographic and temporal records when users agree to share personal information. Urban developers leverage these interactions beyond social and POI promotions, such as Foursquare and Twitter.

This work uses the LBSN as leverage for human behavior identification. Our method identifies valid urban trips through social interactions and builds multi-modal routes for public transportation and Hired Private Vehicles. Social media data returns the most realistic expression of human mobility, especially in urban centers.

2.2 Analytic Hierarchy Process - AHP

MCDM methods introduce alternative decision processes based on mathematics and psychology. Many business, industry, and government applications use the decision methods, assisting when facing multiple criteria. Specifically, Thomas L. Saaty [28] created the Analytic Hierarchy Process (AHP) in the 1970s and kept refined, updated, and enhanced it with newer and intelligent methods such as the fuzzy AHP. The method aids decision-makers with many alternatives and deals with subjective judgments such as "I guess".

In this work, we use the AHP method to implement the route selection method considering the trip, health, and comfort-related factors. For the correct route selection among alternatives with different contextual information, we need to choose a decision-making method that is simple and robust, aiming at scalability for any urban environment application. The AHP method for route selection offers robustness, considering dense criteria and weight with simple mathematical calculations for selection.

The built hierarchy between elements and the weight definition to each alternative route tuple defines the path preference order. Saaty [28] describes the correct AHP use with decision matrix determination for every alternative relating to a criterion, defining the criteria normalized indexes. Nonetheless, the collected dataset for analysis normalizes its raw value indexes for each criterion, allowing the full use of the AHP method.

Saaty's scale [28] defines the element importance degree to another and allows the comparison matrix built, as shown in Equation 2.1. Specifically, M represent the decision matrix with all $f_{n,n}$ pairwise comparison. The matrix objectives are problem complexity level reduction and driver's profile preference definition, facilitating method application due to elevated criteria quantity in various problems.

In the pairwise comparison, the AHP method uses a verbal judgments scale ranging from "equal" to "extreme" (equal relevance, great relevance, greater relevance, huge relevance, and extreme relevance), referring to a criterion comparison importance to another for the problem solution reach. Numerical judgments represent every verbal judgment, being "equal" equivalent to 1 and "extreme" to 9 (1, 3, 5, 7, and 9) with the intermediate values (2, 4, 6, and 8).

$$M = (F_{i,j})_{n \times n} = \begin{pmatrix} f_{1,1} & f_{1,2} & \cdots & f_{1,n} \\ f_{2,1} & f_{2,2} & \cdots & f_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ f_{n,1} & f_{n,2} & \cdots & f_{n,n} \end{pmatrix} \tag{2.1}$$

A consistency validation for the matrix is an AHP method step for the correct matrix built. Equation 2.2 shows consistency ratio (CR) and consistency index (CI) fraction to obtain random index (RI), calculating decision-makers' judgments consistency.

$$CR = \frac{CI}{RI} \tag{2.2}$$

The maximum matrix auto-value (λ_{max}) must be equal to matrix dimension n for matrix consistency maintenance. The $n - 1$ value is used for logically deduced pairwise comparison. Therefore, the fraction between these elements obtains the CI, shown in Equation 2.3.

$$IC = \frac{\lambda_{max} - n}{n - 1} \tag{2.3}$$

The maximum eigenvalue indicates the judgment consistency measure, calculated through the judgment matrix (A) and the priority column vector (w) product, which splits the vector mean, as seen in Equation 2.4.

$$\lambda_{max} = \text{vector mean} \frac{Aw}{w} \tag{2.4}$$

The author also defines the RI as a constant value applied to defined decision matrices for the hierarchy analysis method. This paper uses the 1,41 RI value for using an eight elements matrix [28], as shown in table 1 . The formulas calculation obtains a valid CI for all profiles since it is less than 10%, as further shown in Chapter 5.

The AHP method sums the problem in 3 steps: state the objective, define the

Table 1: Average Random Index for AHP

Decision Matrix Elements	1	2	3	4	5	6	7	8
gray!10 **Random Index**	0.00	0.00	0.58	0.90	1.12	1.24	1.32	1.41

criteria, and pick the alternative from the weights. The following example explains the AHP practical approach didactically [29] [30].

Hypothetically, a man called Bob needs to buy a new car. However, Bob must evaluate the best car for him and his family. For him, the car must be stylish, reliable, and economical. Bob chooses the AHP tool to aid the decision-making. Firstly, Bob needs to perform the three initial steps.

He starts the AHP application by defining the objective: to buy a new car. Then, he defines the evaluation criteria regarding cars: style, reliability, and fuel economy. Finally, Bob defines the available alternatives: Honda, Chevrolet, Renault, and Ford. After the steps, the hierarchical tree arranges the information, as seen in Figure 1.

After declaring the objective, criteria, and alternatives, Bob needs to ponder how each criterion relates to another. Based on the decision-makers judgments, the AHP prioritizes one criterion over another. So Bob needs to relate the best criterion to select a new car. Due to the best car and option available, being beautiful, super reliable, and spending low fuel can not exist, Bob and every other decision-maker need help when dealing with distinct alternatives.

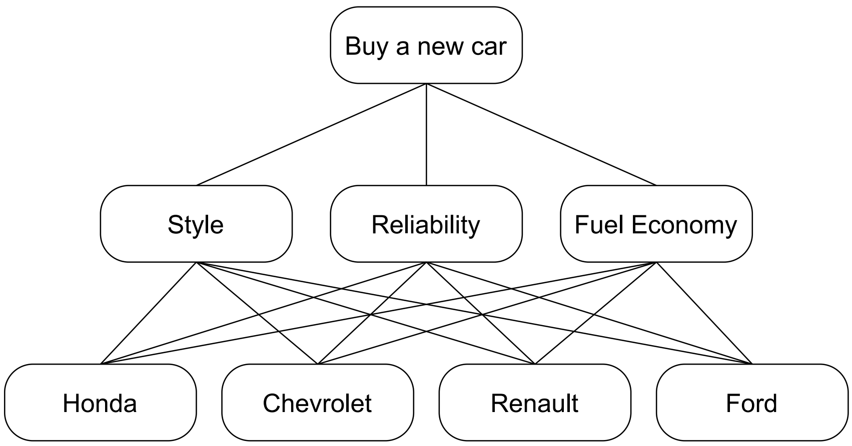


Figure 1: Buy a car problem AHP hierarchy

Then Bob starts relating the three criteria, which depend on predefined judgment values, as mentioned in Section 2.2 and shown in Table 2, proposed initially by Saaty [28]. The Scale of Relative Importance translates linguistic choices into numeric numbers and indicates the importance of a criterion to another. Our route selection method uses 1 to 10 for positive relationships and 1/1 to 1/5 for negative relationships.

For *e.g.*, Bob favors Style over Economy if the Style criterion is three times more important than the Fuel Economy when selecting a new car. Also, the inverse relationship

Intensity of Importance	Definition
gray!101	Equal Importance
3	Slight Importance
gray!105	Moderate Importance
7	Moderate Plus Importance
gray!109	Strong Importance

Table 2: Scale of Relative Importance for AHP

is valid because Economy is 1/3 more important than Style, consequently three times less important. We conclude that Bob moderately disfavors Economy over Style when selecting a new car.

The Judgment Matrix, shown in Equation 2.1, aggregates the relation between criteria. For Bob, Reliability is the most important criterion, followed by Style and Fuel Economy as the least important. Based on these judgments, the matrix in 2.5 shows Bob's preference in the pairwise comparison. The matrix contains the relative importance and the inverse relation assuring the matrix consistency. Bob must implement the idea of turning the Judgment Matrix into a priority ranking.

$$\begin{array}{l} \\ Style \\ Reliability \\ FuelEconomy \end{array} \begin{array}{c} \begin{array}{ccc} Style & Reliability & FuelEconomy \end{array} \\ \left(\begin{array}{ccc} 1 & 1/3 & 5 \\ 3 & 1 & 7 \\ 1/5 & 1/7 & 1 \end{array} \right) \end{array} \tag{2.5}$$

With the eigenvector solution, the AHP method turns the pairwise comparison matrix into the priority ranking, also known as the Vector of Priorities. The eigenvector is a non-zero vector set to represent a square matrix. One way to obtain the eigenvector is to apply a multiplication loop: by squaring the Judgment Matrix n times, summing and normalizing the rows, and then checking if the difference between two consecutive calculations is smaller than a predefined value or has no difference. The obtained criteria priority eigenvector is shown in Equation **??**

$$\text{Vector of Priorities} = \begin{array}{l} \\ Style \\ Reliability \\ FuelEconomy \end{array} \begin{array}{c} Priorities \\ \begin{bmatrix} 0.3196 \\ 0.5584 \\ 0.1220 \end{bmatrix} \end{array} \tag{2.6}$$

After defining the criteria ranking, the method needs the same procedure for the alternatives for the criteria evaluation. Bob judged all four options (Honda, Chevrolet, Renault, and Ford) under the Style (Equation 2.7) and Reliability (Equation 2.8) criteria. For Fuel economy, Bob evaluates the options by the miles per gallons consumption for the most economical car, as shown in Equation 2.9.

$$\text{Style} = \begin{array}{c|cccc|} & Honda & Chev & Renault & Ford \\ Honda & 1 & 1/4 & 4 & 1/6 \\ Chev & 4 & 1 & 4 & 1/4 \\ Renault & 1/4 & 1/4 & 1 & 1/5 \\ Ford & 6 & 4 & 5 & 1 \end{array} = \begin{array}{c|c|} & Priorities \\ Honda & 0.1160 \\ Chev & 0.2470 \\ Renault & 0.0600 \\ Ford & 0.5770 \end{array} \tag{2.7}$$

In terms of style (Equation 2.7) and reliability (2.8), the same judgments made in the criteria matrix, Bob made for the criteria relating to the options, obtaining three criteria eigenvector for the four car alternatives.

$$\text{Reli.} = \begin{array}{c|cccc|} & Honda & Chev & Renault & Ford \\ Honda & 1 & 2 & 5 & 1 \\ Chev & 1/2 & 1 & 3 & 2 \\ Renault & 1/5 & 1/3 & 1 & 1/4 \\ Ford & 1 & 1/2 & 4 & 1 \end{array} = \begin{array}{c|c|} & Priorities \\ Honda & 0.3790 \\ Chev & 0.2900 \\ Renault & 0.0740 \\ Ford & 0.2570 \end{array} \tag{2.8}$$

Regarding Fuel Economy, Bob used the miles per gallons consumption information in each alternative car, obtaining the eigenvector through the average calculation. The average normalization process indicates the eigenvector in a 0 to 1 scale, as seen in Equation 2.9, for the AHP next step.

$$\text{Fuel Economy} = \begin{array}{c|c|} Honda & 34/113 \\ Chevrolet & 27/113 \\ Renault & 24/113 \\ Ford & 28/113 \end{array} = \begin{array}{c|c|} Honda & 0.3010 \\ Chevrolet & 0.2390 \\ Renault & 0.2120 \\ Ford & 0.2480 \end{array} \tag{2.9}$$

When quantitative information is unavailable, verbal judgments achieve qualitative information about the alternatives. The preference judgments for style and reliability features are complex for different decision-makers' verbal preferences and represent qualitative information. Fuel Economy is a feature with numeric values and quantitative information, making it simpler to obtain the alternative ranking through its characteristics.

After completing all judgments matrix, a necessary step for the decision matrix before going further is to check the consistency with the CR, as seen in Equation 2.2. Equation 2.10 calculates the maximum eigenvector value by multiplying the matrix column sum and the priority eigenvector. Then, Equation 2.11 obtains the CI for priorities to be analyzed.

$$\lambda_{max} = \begin{bmatrix} 3.33 & 1.75 & 8 \end{bmatrix} . \begin{bmatrix} 0.3196 \\ 0.5584 \\ 0.1220 \end{bmatrix} = 3.0174 \tag{2.10}$$

$$CI = \frac{\lambda_{max} - n}{n - 1} = \frac{3.0174 - 3}{3 - 1} = 0.0087 \tag{2.11}$$

After that, Bob obtained the CR by dividing the CI for the RI defined in Table 1. In Bob's case, RI is 0.58 for three criteria elements. Finally, the CR is consistent if lower than 0.10%, otherwise will be a need to recheck the criteria relation. For the criteria priority, the CR was 1.5%, indicating consistency, as seen in Equation 2.12.

$$CR = \frac{CI}{RI} = \frac{0.0087}{0.58} = 0.015, \text{ where } \begin{cases} CR < 10\% \rightarrow \text{ adequate} \\ CR \geq 10\% \rightarrow \text{inadequate} \end{cases} \tag{2.12}$$

For the last step in the AHP method, Bob needs to multiply the priorities vector by the alternative ranking under each criterion, as shown in Equation 2.13. A single matrix aggregates the alternative ranking for each feature, built in four lines and three columns. Finally, after the multiplication, Bob gets the Ford alternative as the highest value above all alternatives under his preference, as shown in Equation 2.14.

$$\begin{array}{l} \\ Honda \\ Chevrolet \\ Renault \\ Ford \end{array} \begin{array}{c} \begin{array}{ccc} Style & Reliability & FuelEconomy \end{array} \\ \begin{bmatrix} 0.1160 & 0.3790 & 0.3010 \\ 0.2470 & 0.2900 & 0.2390 \\ 0.060 & 0.0740 & 0.2120 \\ 0.5770 & 0.2570 & 0.2480 \end{bmatrix} \end{array} . \begin{array}{l} \\ Style \\ Reliab. \\ FuelEcon. \end{array} \begin{array}{c} Priorities \\ \begin{bmatrix} 0.3196 \\ 0.5584 \\ 0.1220 \end{bmatrix} \end{array} \tag{2.13}$$

$$\begin{array}{l} \\ Honda \\ Chevrolet \\ Renault \\ Ford \end{array} \begin{array}{c} Criteria \\ \begin{pmatrix} 0.3060 \\ 0.2720 \\ 0.0940 \\ 0.3280 \end{pmatrix} \end{array} \tag{2.14}$$

With the example, AHP reveals a logical framework helping individuals and groups in complex decision analysis, showing the benefits between alternatives based on a set of criteria in a given problem in a structured and flexible way.

2.3 Evaluation Metrics

Evaluation metrics for urban mobility methods vary depending on the specific objective for the user's need. Navigation systems and ride or bike-sharing rely on much urban information to provide efficient transportation alternatives. The comparison between methods indicates a superior performance when correctly performed [31].

Travel time indicates the time users take from origin to destination. This metric determines the method's efficiency in time-saving, affecting the urban productivity rhythm that most users seek to mitigate. Our methods uses the travel time for indicating the time spent in route alternatives or different transportation modes [32].

Walking distance measures the need to walk to different transportation modes' stop points to urban destinations. We also can measure accessibility with the percentage of accessible destinations within a specific time limit or walking distance to enter the transportation mode, such as bus stops [33].

The modal shift can be related to the multi-modal routes method and measure the population's possibility to switch urban transportation modes as a alternative hybrid route. This method can avoid traffic congestion, improve air quality with fewer emission vehicles, and encourage collective transportation [34].

In terms of air quality, carbon emission metrics introduces the sustainability-aware urban mobility solutions. Measuring the carbon emission and air quality help determine the environmental impact of the actual transportation system for further changes [35] [36]. We assess the pollution measure in our proposed two method, with CO_2 emission calculation, as explained in Section 4.4 and through air quality sensors readings, as explained in Section 5.4.

Nature and Tourist Attractions can fulfill the user's need or happiness, which indicates comfort features. Geographic Information Systems (GIS) and navigation services contain geo-located information about green areas and Point-of-Interest (POI). In our personalizing method, we attend to the comfort and pleasant of users with customized preferences and weighted profiles, offering the best route alternative in a criteria hierarchy [37].

The cost-effectiveness metric relates the solution benefits to the cost it takes. Many solutions can offer a fast and comfortable alternative in transportation like HPV services, such as Uber, but much expensive. In contrast, public transportation can be a cheaper option, but need walk to the bus top and can be slow or uncomfortable [38].

The safety metric assess the accident risk or injuries regarding a mode of transportation or a mobility solution. Criminality can be considered a safety and security metric to be accounted. Many methods retrieve the safety level identification by historical data analysis, avoiding dangerous areas and alerting the local authorities for infrastructural changes [39].

For the multi-criteria route selection method, we choose the percent deviation from a known standard (PDFKS) to evaluate the personalized user profiles to greedy options, simulating the traditional navigation systems [16] [40]. Section 5.4 explains the PDFKS metric calculation and application in the route selection problem.

2.4 Chapter Conclusions

This Chapter provided the theoretical background and insights about the Urban Computing paradigm, the AHP method with an example, and the Evaluation Metrics for the proposed route selection methods. The presented concepts its necessary for this dissertation thesis understanding. In the following Chapter, we will approach the methods' related works in the state-of-the-art literature.

CHAPTER 3

Related Works

This chapter presents the leading works in the literature regarding multi-modal routing, multi-criteria route selection, and pollution-aware urban routes relating to this work's proposed approaches. In the state-of-art methods, route selection approaches consider the social networks data for human and mobility flows analysis, hybrid routes involving different transportation modes, and air quality measurement, but not the three elements in the same solution. Similarly, in the multi-criteria route selection literature, many works need to consider health, comfort, and security factors when offering urban routes without customizing the preference. We highlight the difference between state-of-art approaches to our methods and all related objectives attended.

3.1 Multi-modal Methods

This section presents the essential concepts about location-based social network sharing usage, urban multi-modal routing solutions, and urban mobility pollution analysis. All discussed concepts focused on specific solutions and were categorized into LBSN data usage, direct integration of multi-modal routing, mobility flow analysis based on social networks, and air pollution approach.

Ferreira et al. [41] investigated how LBSN check-ins could be used to study tourists' mobility behavior. The authors chose and evaluated social media-generated data, such as Twitter, Flickr, and Foursquare, for behavior analysis of tourists and city residents, showing their important locations and visiting time. The method combines data mining techniques with spatial analysis and natural language processing to extract useful information from social interactions.

The authors also used clustering techniques to group similar tweets, photos, and check-ins into semantic categories for popular activities and location identification, such as dining, nightlife, or sightseeing [41]. The chosen data treatment and modeling are appropriate for mobility identification and reveal insight into spatiotemporal and semantic aspects of tourists' movements. However, data collection in the used case can not explicitly address essential mobility elements.

Rodrigues et al. [42] presented a framework called SMAFramework, to integrate heterogeneous urban data sources and find their correlations. The work aims to help city planners with urban mobility data and data-driven solutions to facilitate citizen trip experiences around the city. Tool development was made for the framework users to manage, standardize, and integrate the data from different sources for information extraction.

For standardization, the authors use a Multi-Aspect Graph model, which is an extension of the traditional graph concept and allows the representation of time-varying constraints. Also, a part of the framework contains some developed tools to collect and analyze data from different urban sources, such as the Fuzzy Matcher, which evaluates the correlation between space and time with real data from New York taxis company [42]. The author's approach is relevant to data leveraging for urban applications; however, pollution issues were not raised.

Rodrigues et al. [43] proposed a hybrid multi-modal routing solution and an evaluation method for large-scale use, providing route impact analysis on mobility flow in the use case urban area. Multi-modal route generation includes walking, bus, subways, ferries, other modes, and Hired Private Vehicle (HPV), such as taxis, and application services, like Uber.

The authors' method identifies the most relevant city flows using an efficient clustering technique, creating personalized routes and avoiding traffic congestion. The work provides a comparison of the hybrid routes and traditional routes with only one transportation mode [43]. However, The authors focus on cost reduction without significantly impacting the user experience and trip time, not considering the pollution and air quality.

Kalajdjieski et al. [35] proposed a prediction air pollution method based on convolutional neural networks (CNN) using camera images. The authors preferred IoT infrastructure security and traffic light camera images instead of air pollution sensors, which are most useful in countries with fewer sensors installed. The authors developed the air pollution estimation approach based on cameras to facilitate real-time heat maps of air pollution creation and to track pollution sources.

The work evaluates four architectures that classify images and perform data augmentation for imbalanced datasets, with enhanced generative approaches to reach a better performance related to state-of-the-art [35]. Although, the approach had significant dependence on public infrastructure, which implies an occasional bottleneck in the system. Simplified and decentralized methods can offer system robustness.

Zou et al. [36] proposed a near real-time healthier route planning (HRP) method with an experimental online implementation service for air pollution exposure (APE) minimization in daily travels by the method reliability and feasibility validation. The study concentrates on the HRP method with steps, including the fine-scale air pollution concentration mapping, risk weight estimation of road segment exposure to air pollutants, dynamic updating mechanism of the Dijkstra algorithm in healthier route search, and development of the online service.

The method boosts healthier route planning through the fine-scale risk estimation achievement, providing an alternative to minimize exposure risk and protect human health in improved travel behaviors [36]. The author's objective was defined as a route planner and minor exposure to pollution without multi-modal integration and proper data mining for urban planning analysis and modification.

Table 3 shows the main related works comparison under four categories. Based on the related work, this work proposes the main features of LBSN data usage, multi-modal routing, mobility flow analysis integration, providing statistics to be used by users who opt for less polluting modes of transport and by urban transport managers who can optimize values to improve the quality of life.

Table 3: Multi-modal approach related works features comparison

Work	LBSN data usage	Multimodal routing	Mobility flow analysis	Air quality measure
gray!10 Ferreira et al. [41]	yes	no	no	no
Rodrigues et al. [43]	yes	yes	yes	no
gray!10 Rodrigues et al. [42]	yes	no	yes	no
Kalajdjieski et al. [35]	no	no	yes	yes
gray!10 Zou et al. [36]	no	no	no	yes
blue!10**Multi-modal Method**	yes	yes	yes	yes

3.2 Multi-criteria Selection

This section presents the main state-of-art works related to multi-criteria method, navigation systems, and trip influencer factors. The related work selection criteria contain health, well-being, comfort, and security factors on route selection based on driver preference.

Wu et al. [44] proposed a traffic-route selection optimization approach using symmetry/asymmetry contextual traffic data and multi-criteria decision analysis for urban-logistic routing recommendation. All potential routes from the delivery service employee's point of origin to his POI are collected and identified as candidate paths. The Multi-criteria Decision Analysis (MCDA) generates a ranking of candidate paths based on the contextual data evaluation.

The author aims to collect contextual data from the urban transportation database and Google Maps routing Application Programmin Interface (API) metadata, constructing a context-based social network. The work contains a comparison to the produced ranking by different MCDA methods [44]. The work presented an urban route selection for deliverymen used in urban logistics, needing to be more beneficial for citizen use. Besides, the criteria for selection are limited to average speed, congestion degree, distance, and worker personal interest.

Sarraf et al. [45] added an analytic choice method for the developed safer route planning application, comparing different multi-criteria methods outcomes for the same objective. The work develops a fully automated approach to acting as a transportation expert, analyzing road safety levels and user judgments, and selecting the most suitable path among alternative routes from a source to a destination.

Besides that, the authors present a comparative study of five MCDM methods and three approaches to derive fuzzy criteria weights [45]. Regardless, the proposed system analyzed historical and live monitoring, considering vehicle accidents in the analysis area, offering safer routes and urban infrastructure reports for future enhancement.

Kaivonen et al. [46] proposed real-time monitoring with data gathering on pollution through urban public transportation networks, covering the whole city area and addressing air quality issues as urban environment criteria. The authors evaluated data collection on mobile sensors compared to stationary air sensors, choosing an efficient way to map pollution in the urban environment.

The work proposes a bus routes identification method that can provide good town coverage, passing through highly polluted areas which require attention. Route planning is also studied to select bus routes that can acquire measurements from important locations. The method develops the solution via image analysis on a bus route map provided by the local bus company [46]. Although the advance in the state-of-the-art, the solution does not return a less polluted route alternative for user selection.

Zhang et al. [47] proposed a routing method for Vehicular *Ad Hoc* Network based on a fuzzy logic system. The method considers the relative speed, the angle between the node and neighbors, the connection angle between the destination node and its neighbors, and the node density of neighbors as the input of fuzzy logic, combining all these criteria into a node location prediction algorithm.

Thus, the work proposes the Geographic Routing method based on Velocity, Angle, and Density (GRVAD). Simulation results show that the proposed method outperforms the previous methods in many performance evaluations. Their proposal is an efficient routing approach for Vehicular Ad Hoc Networks (VANET) but does not consider contextual data for humanized mobility and only improves communication metrics between devices.

Hsun et al. [48] proposed a route recommendation method for taxi drivers to keep picking up passengers and receive a better profit while letting drivers successfully arrive at the reservation's location on time. The method considers real-time predictions and traffic network information, aiming for higher profit. The criteria for this approach rely on pick-up probability, drop-off distribution, road network, distance, and time factors.

The authors developed a novel framework that intelligently combines two prediction modules, traffic network information, and a search algorithm. Besides that, they design an attentive heuristic function search scheme and propose three indicators to evaluate the effectiveness in generating routes [48]. The authors did not consider health, comfort, and risk factors, only the usual navigation criteria compared to standard methods.

Solé et al. [40] proposed a method for feature measurement that composes a route, which affects driver security and pleasure on urban trips. The article evaluated different route selection methods with single or multi-criteria and some pre-defined profiles. The authors introduced two novel multi-criteria route selection methods; the first method is called Route to Vector (R2V) and translates the route features into vectors and finds the closest to the best vector.

The second, called Hierarchical with Variable Tolerance (HVT), follows a user-defined feature order to reach the best route. The authors build a novel dataset containing criminality, traffic, accidents, nature, tourist, attractions, and trajectory information data about 3170 routes from the city of London to apply their proposed method; the results indicated decreased driving risks without significant time-related penalties [40]. The authors intended the best route identification from the created dataset containing weights of different trip-influencing factors but did not consider air pollution as a criterion.

Table 4 shows the relation between previous works and this paper on different issues, such as the multi-criteria approach, various criteria in the selection, including air pollution, and providing the best route ranking based on defined user profiles preference. This paper presents contributions on each element integration, contrasting state-of-art approaches.

Table 4: Features addressed in related works

Work	**Multi-criteria Approach**	**Air pollution factor**	**Comfort and security factors**	**Routes ranking**	**User Profiles**
gray!10Wu et al. [44]	yes	no	no	yes	no
Sarraf et al.[45]	yes	no	yes	yes	no
gray!10Kaivonen et al. [46]	no	yes	no	no	no
Zhang et al.[47]	yes	no	no	no	no
gray!10 Hsieh et al. [48]	yes	no	no	yes	no
Solé et al. [40]	yes	no	yes	yes	yes
blue!10 **Multi-criteria Method**	yes	yes	yes	yes	yes

3.3 Chapter Conclusions

Based on the two methods' related works analysis, we perceive the need for integration on every service feature, providing a route selection considering some of the most important factors in a personalized way and integrating transportation modes to create hybrid economic and time-saving alternatives.

After analyzing the main issues related to multi-modal method addressed by the state-of-art, it can be highlighted the need to integrate methodologies for data collection and analysis, multi-modal routing from urban flows, and the calculation of greenhouse gas emission according to modal used. For the multi-criteria method, the literature review indicates the integration need for other factors in vehicle trip suggestion, using emerging technology to enhance the data acquisition step for route selection from each driver's necessities.

CHAPTER 4

Multi-modal Route Selection Method in a Urban Computing Scenario

As the first method in this master thesis, this Chapter introduces the multi-modal route selection method. The proposed method considers the pollution level calculation for each transport mode in urban routes built. The evaluation performance compares the hybrid multi-modal routes with the traditional single-modes in several trip-related methods and the GHG emission calculation.

4.1 Overview

Figure 2 depicts the air pollution-aware multi-modal urban route selection method overview. In the "Data Processing" part, we use the dataset acquisition through LBSN data collection to identify all valid urban trips. In the "Flow Identification" part, we apply the trip clustering process for the main flow identification, representing the average paths in-between all trip information. At least, in the "Method Evaluation" part, we analyze all trip-related features to compare the three route alternatives: performed by bus, only performed by HPV, and the hybrid alternatives.

Figure 3 presents the method workflow. This is important by smart cities services aim to reduce the exposure of the population to environments with a high concentration of pollutants are likely to have an impact on the quality of life of its citizens. Specifically, the multi-modal routing with pollution calculation methodology can generate new results for the analyzed emissions for each modal in certain areas where the user's location data are collected. The transportation mode distance traveled, and the fuel type calculation can return the CO_2 amount in every trip.

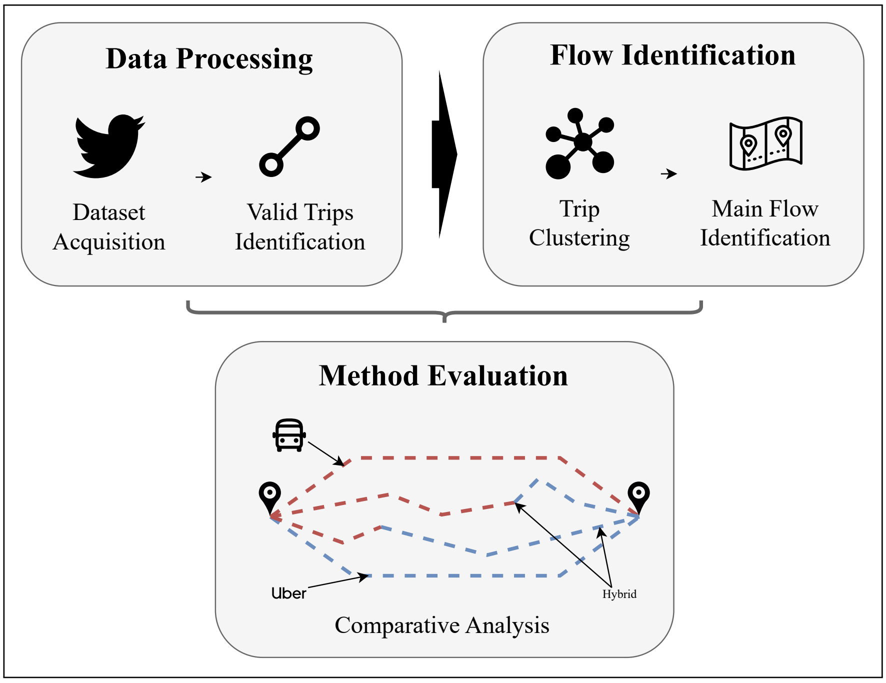


Figure 2: Multi-modal method overview

In this context, we consider two approaches for the air pollution-aware multi-modal urban route selection method, namely, hybrid multi-modal urban routes and emission approach. The "Record Linkage" step represents all LBSN data collection and processing, validating the user trips around the city. The "Clustering" step represents the data reduction method, identifying the average mobility flows of all valid trips. The "Routing" and "Visualization of Routes" steps represent the build routes between main flows and enable the routes to map through a mapping tool. At least, the "Comparative Analysis of Routes" step demonstrates the economic, trip, and pollution-related metrics comparison between the hybrid and single-modal approaches.

The emission calculation approach relates to the hybrid multi-modal through the gathered information obtained in the "Visualization of Routes" step and merges the pollution levels in the "Comparative Analysis of Routes" step. The "Fuel Data Gathering" step obtains the vehicle fuel local specification, obtaining the fuel mixture composition values. The "Fuel Data Filtration" step selects the specific fuel for each transport mode used for the analysis. After this process, the "Fuel and Routing Data Linkage" step links the fuel values to each route information gathered in the past process, obtaining the consumed fuel value. In conclusion, the "CO_2 Emission Determination" step returns for the "Comparative Analysis of Routes" step the emitted pollutant amount.

4.2 Data Acquiring and Mining

The geolocated data choice is arbitrary for the data-acquiring method, which is based on the cities' urban characteristics and user behavior. This methodology uses

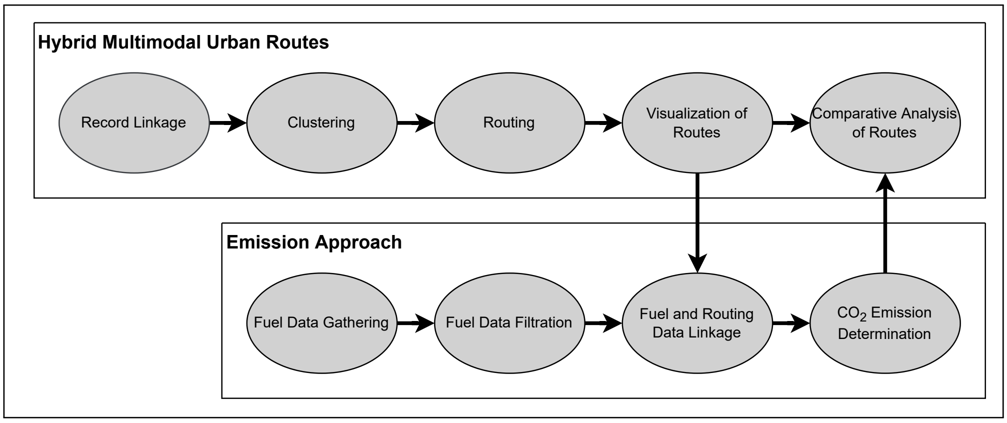


Figure 3: Emission calculation inserted to the multimodal service workflow

anonymous social media records with coordinates and timestamp information, for mobility flow retrieval and analysis. There is a need to filter valid user trips, allowing the identification of urban mobility flows, excluding the random and unlinked user records.

For the anonymous user data preparation for the analysis, the LBSN attend the location and time information requirements and have great popularity among the population, specifically in urban scenario. Rodrigues et al. [23] [1] use in their methodology two famous social networks which provide the necessary elements for the practical urban mobility analysis, *i.e.*, Foursquare and Twitter. We only used one city sector for the methodology analysis because of the records' limited dataset coverage area.

In our methodology, we use the dataset acquired by Rodrigues et al. [43] from the Twitter API Tool. The authors acquired all raw tweets data in the São Paulo metropolitan area, with minimal difference in time between interactions. The authors used data mining to identify valid trips by matching the trip distance with Origin-Destination (OD) pair, time variation verification, and matching speed variation.

4.3 Flow Identification

The flow pattern identification simplifies the mobility analysis, and in-depth characterizes its relationship with multi-modal transport. Rodrigues et al. [49] used mathematical calculation and visualization tools in the identification method for clustering all valid trips into mobility flows. The developed framework uses different clustering programming tools and assists the urban mobility analysis.

After the valid trip data mining, the next step is identifying the most frequent urban flow from all proper trips. From the main flows, we can identify the impacts of the transport modes in the most frequented urban routes without calculating for every route in the treatment phase. The grouping of (OD) pairs for each trip identifies the most relevant zones along the dataset, and we classify the flows into trending and secondary.

[1] https://github.com/diegopso/hybrid-urban-routing-tutorial-sbrc

The results from the technique execution indicate 12 mobility flows from the collected Twitter data, as shown in Figure 4. The flows concentrations reside in *"Jardim Itatinga"*, *"Olímpico em São Caetano do Sul"*, *"Jardim dos Perdizes"*, and *"Aclimação"*. Only seven flows include the central four districts, and the authors consider for the user experience analysis [23].

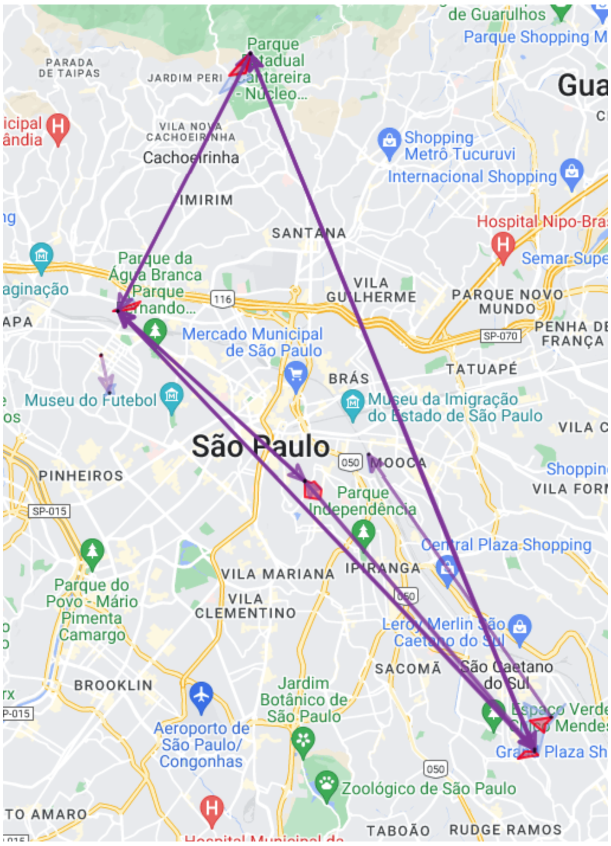


Figure 4: 12 main flows acquired by analysed Twitter Data

4.4 Emission Calculation

Regarding the emission calculation per transport mode, we implemented a standardized method for any road vehicles, used fuel, and applied it to the Brazilian urban context [50]. The inventory calculation method proposed by the Intergovernmental Panel on Climate Change (IPCC) defines gas emissions as harmful to the atmosphere and population health. The emission results obtain the CO_2 values from burning fuel with the proposed "top-down" approach for emission estimation performance. Therefore, we integrated these calculations into the proposed methodology, defining the correct emission values for multi-modal transportation.

The "top-down" approach determines the greenhouse gas emission estimation in three main equations. Equation 4.1 defines the energy consumption (CC) value, measured by tera-joule (TJ). The fuel consumption (CA) value indicates by liters, the physical unit conversion factor of the fuel amount measurement to an equivalent oil ton (tEP), based on the fuel higher calorific value (Fconv), the equivalent oil ton value is 45.2×10^{-3} TJ, and the upper-to-lower calorific value correction factor (Fcorr).

$$CC = CA \times 45,2 \times 10^{-3} \times Fcorr \tag{4.1}$$

The Algorithm 1 applies the emission calculation in the route determination method code, determining the CO_2 total emission value with the traveled distance value for each vehicle into possible routes. The emission calculation method is implemented through Python programming tools and guarantees the independent value functionality for different fuels, with conversion, correction, and emission modularity for any type. Table 5 indicates the Fconv of the physical fuel unit quantity measurement and the Fcorr, alternating between solid and liquid (0,9) and gaseous fuels (0,95).

Table 5: Conversion factor values.

Fuel Types	**Fconv values (tEP/m³)**
gray!10**Gasoline**	0.771
Anhydrous alcohol	0.520
gray!10**Hydrated alcohol**	0.496
Diesel	0.848
gray!10**Dry natural gas**	0.857

Algorithm 1 Energy consumption algorithm

```
procedure ENERGY(CA, Fconv, Fcorr)
    FconvValue ← ""
    FcorrValue ← ""
    FconvList ← FconvListValues
    FcorrList ← FcorrListValues
    for element ∈ FconvList do
        if element[0] == Fconv then
            FconvValue ← element[1]
    for element ∈ FcorrList do
        if element[0] == Fcorr then
            FcorrValue ← element[1]
    CC ← (CA ∗ FconvValue ∗ FcorrValue ∗ 45.2 ∗ 10^(−3)
    return CC
```

After obtaining the energy consumption value, the carbon content (QC) equation uses it for the local-specific fuel burn. The carbon content value expresses the value in carbon gigagram (GgC), after multiplying the energy consumption with the carbon emission factor and converting the gigagram to tons of carbon (10^{-3}), as shown in Equation 4.2.

$$QC = CC \times Femiss \times 10^{-3} \tag{4.2}$$

The Algorithm 2 calculates the emitted carbon (EC) amount by vehicle. Table 6 presents the carbon emission factor published by the IPCC [51], and we integrated the calculation into our method.

Table 6: Carbon emission factor values.

Fuel Types	Femiss values (tC/TJ)
gray!10**Gasoline**	18.9
Anhydrous alcohol	14.81
gray!10**Hydrated alcohol**	14.81
Diesel	20.2
gray!10**Dry natural gas**	15.3

Algorithm 2 Carbon emission algorithm

procedure CARBON($CC, Femiss$)
  $FemissValue \leftarrow$ ””
  $FemissList \leftarrow FemissListValues$
  for $element \in FemissListValues$ **do**
    if $element[0] == Femiss$ **then**
      $FemissValue \leftarrow element[1]$
  $EC \leftarrow (CC * FemissValue * 10^{(-3)})$
  return EC

At least, Equation 4.3 converts the carbon emission value, considering 44 tons of CO_2 correspond to 12 tons of carbon. We relate the emission calculation to a consumption per vehicle of 8 kilometers per liter in the HPV case and 2 for buses. The analyzed equations form an excellent methodology for obtaining greenhouse gas emissions from terrestrial transportation sources, calculating the fuel amount burned, the carbon content, and the corresponding emissions of CO_2 (ECO_2).

$$ECO_2 = EC \times 44/12 \tag{4.3}$$

The Algorithm 3 implements the emitted carbon to CO_2 converting function. The proposed method integrates the emission calculation approach into the hybrid multi-modal urban routes and has greater precision due to considering the global annual fuel consumption for the given analyzed environment, not the used vehicle specificity.

Algorithm 3 CO2 emission algorithm

procedure CO2(EC)
  $ECO2 \leftarrow ((EC * 44/12) * 10^{-6})$
  return $ECO2$

4.5 Routing and Visualization

This section defines the performed routes from the identified flows, alternating between transportation modes as possibilities for each route. The method uses the route alternatives with hybrid or single mode defining the user experience metrics. The method also adds the CO_2 emission calculation from the traveled distance for each vehicle.

For the routing process stage, we compute route possibilities alternating between urban modals in the analyzed city context with the SMAFramework [49] aid. In addition, the TomTom Routing API tools permit the congestion areas identification, suggesting alternative routes using transit and global positioning services. The Google Directions API computes the generated routes processed by the multi-modal approach, allowing its visualization.

In the multi-modal routes formation stage, we use three transportation modes: foot, bus, and hired vehicle. In the visual representation of each formed route for each mobility flow, red represents the bus, some green stretches define the walking, and blue represents the hired vehicle sections, calculated with the Uber price estimation, as defined in [23]. The author introduces two types of hybrid routes: “Hybrid 1” has a more expressive bus section along the route, and “Hybrid 2” with more Uber section. The walking sections along the dataset-obtained flows characterize the bus stop way performed by foot and alternates in the route beginning, middle, or end.

After processing and distributing the route among alternative options, we generated the geographical representation, which displays in a 2D map the determined sections, defined as “steps” for each transport mode suggested for the user to use. The developed algorithm initializes the map generation of the suggested route output. Each route section contains information such as origin, destination, travel time, overview polyline, and transportation mode. The Overview Polyline is a suggested route-coded representation as an approximated path to obtain the point sequence that forms the decoded route. Some colored drawn parts in the map define the mode to be taken by the user.

4.6 Evaluation

This section presents the obtained results on the data analysis methodology from the social networks data for the São Paulo urban transportation metrics. The generated route possibilities with transportation modes have different settings, allowing the comparative analysis for experience and economic metrics for urban users. The values correspond to the average value obtained for each analyzed flow [23].

The evaluation methodology uses the Python Matplotlib library for graph generating, which shows the comparative analysis between the average metrics for the main flows’ alternative routes. The graph representation indicates the analyzed urban environment status to each presented metric and enables the analytic selection based on the user preference by comparing the alternatives, choosing less expensive routes or with less duration.

After defining the multi-modal method resulting in graphs, we add the CO_2 emission calculation methodology to the resulting comparative analysis. We integrate the emission estimation graph into the user experience metrics approach, containing the pollution amount for each vehicle for the traveled distance along the generated route.

Four main factors represent the experience metrics along the identified seven main flows: the distance traveled, the transport mode estimated price, the elapsed waiting time for vehicle arrival, and the user's walking distance. In the route performance comparison, the walking distance integrates the total distance traveled in a trip, while an entire graph indicates the average between modes.

4.6.1 Previous Analysis

The multi-modal routing work [23] performed a comparative analysis for the generated routes considering the obtained main flows, which considers the walking and traveled distance, the estimated price for each transport mode, the travel time, and the user waiting time to the vehicle arrival. The metrics bar graphs explain the difference between the average for all routes in the main flows, with the difference between single modes and the "Hybrid 1" and "Hybrid 2" alternatives, as explained in Section 4.5.

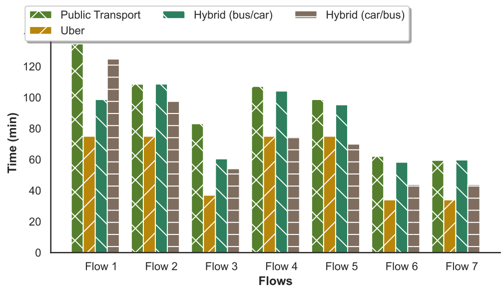
Figure 5 presents obtained values for the main flows, comparing metrics and evidencing the advantages between different transport modes chosen for analysis. The distance, time duration, cost, and waiting time metrics represent urban trip-related features, and the presented evaluations are essential for the hybrid method validation.

Regarding the estimated duration for each trip and type of route, Figure 5a indicates the more significant time taken to complete the trip using a bus since they have mandatory stops and can take longer and unnecessary routes for most users, contrasting to the routes performed by Uber. The decrease in time duration represents 30% to 50% of the total duration, and Hybrid routes have a negligible decrease compared to traditional types of routes. The graph depicts the estimated duration similarity of the hybrid routes to single-mode, differentiating in price.

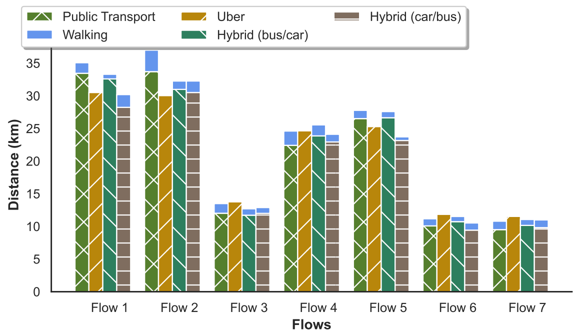
Figure 5b presents the distance traveled by different transport modes and flows; the difference is insignificant, representing a 1% to 10% variance due to the similar route construction and exact departure and arrival points. The graph's blue bar indicates the walking distance taken on the route, which is more frequent on bus routes because of the need to walk to its boarding point/stop. The distance graph with the walking course depicts the little distance walking on Uber and hybrid routes, with the cost advantage for the hybrid routes.

Figure 5c presents the disparity between the cost of each transportation mode, mainly showing a 70% to 80% decrease in spending on buses and Uber. The hybrid routes have a balanced price for use less Uber mode, as the most expensive transportation. Also, the hybrid offers less cost than a whole Uber route, more convenience, less waiting time, and a minor trip duration for users. The cost estimation graph observes the large disparity between the two single-mode routes and validates the hybrid alternative need.

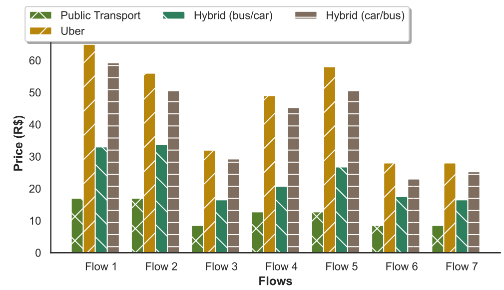
Figure 5d presents the user waiting time for boarding in the transport mode. The wait time is longer on routes with only buses, compared to routes performed only by Uber, increasing from 60% to 90%. Alternatively, hybrid routes decrease the waiting time by 5% to 20% when compared to the bus-only routes. The waiting time graph depicts the little time for Uber with a higher cost and the hybrid alternative to decrease the time.



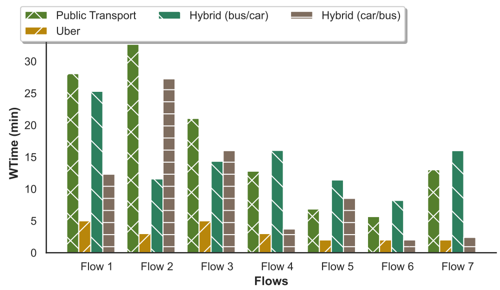
(a) Estimate trip duration.



(b) Travel Distance for each flow.



(c) Estimate price.



(d) Waiting time per flow.

Figure 5: Obtained Results for Different Route Selection Methods.

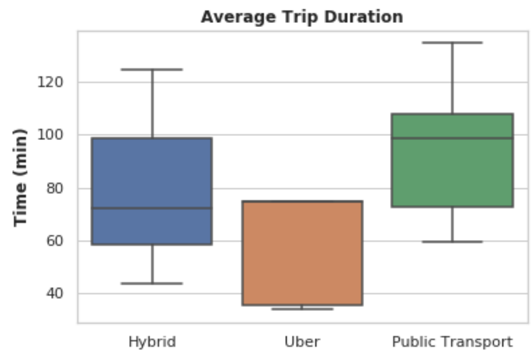
Figure 6 shows the average impact on the routes built from main flows. The graphical representations made explicit the comparison between the multi-modal approach with the traditional routes with the whole trip on a bus or Uber. Figure 6a shows the analyzed flows' average travel time, demonstrating the relative decrease in the hybrid route advantage compared to whole bus routes. There is less variation between the route duration, mainly performed by Uber, due to the duration consistency shown in the application.

Figure 6b shows the similar indicated evidence in Figure 5b. The average distance taken by the three route types is similar, despite the reduced variation occurrence in the Uber-performed routes, with greater endurance than whole bus routes.

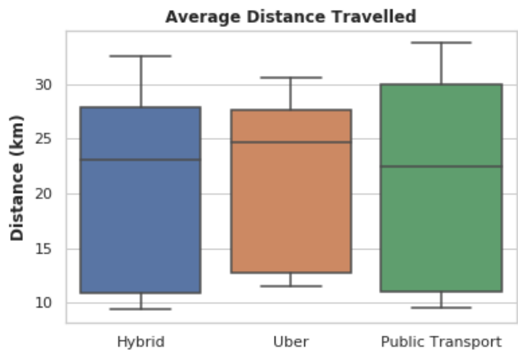
Figure 6c presents the average price among the transportation types in the routes. The analysis presents a disparity between bus and Uber cases, increasing about 75% of the trip cost. The alternative hybrid mode appears to increase only 50% for a user who wants to save economically.

Figure 6d demonstrates the waiting time disparity between trip types, with the worst value for bus compared to Uber routes, increasing 75% to 90% the time spent. The hybrid route type offers an alternative, balancing the waiting time between the two transport modes, as seen in (Figure 5d.

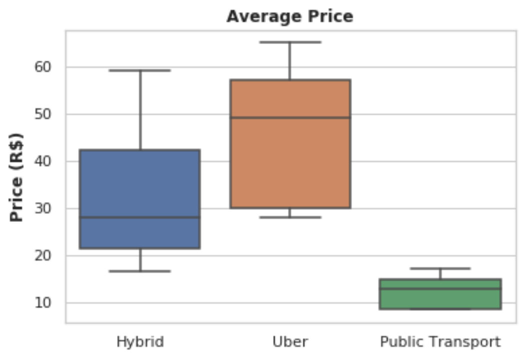
Figure 6e presents the average walking distance traveled separated from the total distance traveled, evidencing the disproportion between bus and Uber routes. The hybrid route brings feasibility in reducing the necessary distance for the user to walk to the boarding point and start the trip.



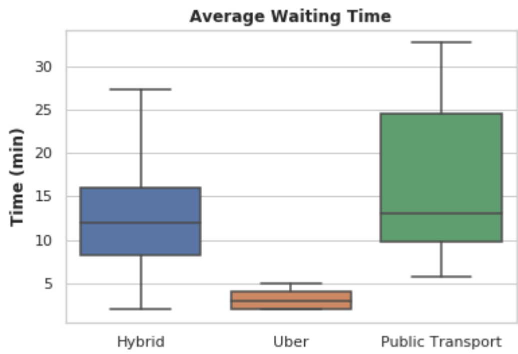
(a) Trip duration.



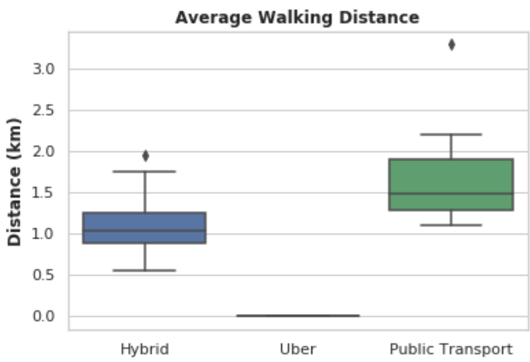
(b) Travel Distance.



(c) Price.



(d) Waiting time.



(e) Walking Distance.

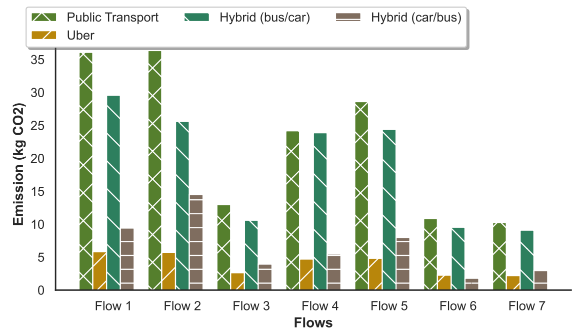
Figure 6: Average performance of routes metrics, considering flows 1 to 7.

4.6.2 Emission Analysis

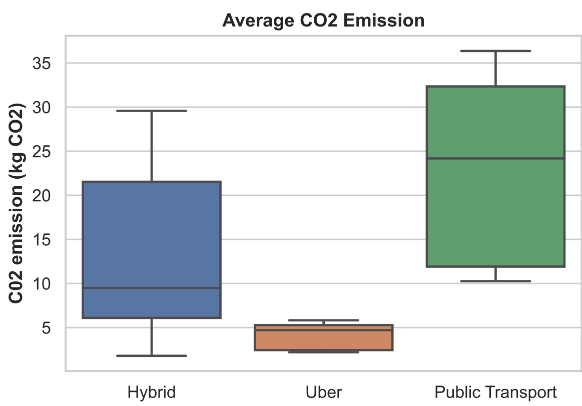
Figure 7a represents the obtained results for the emission calculation value in the seven main flows, acquired through the calculated traveled distance by the vehicle types used along the route and converted to the fuel amount consumed and integrated into the calculation. The comparative routes analysis indicates the pollution emitted by different transportation types in the main flows.

Bus-only routes perform a higher emission than all types due to the high fuel consumption, with about an 85% increase compared to Uber use and more significant emission adjustment in the hybrid alternative route. However, we need to emphasize the greater passenger transport capacity of the bus, justifying the higher emission by the mode when hired or private vehicles bring no more than five occupants.

Figure 7b presents the average emission values applied to all routes, indicating the highest pollutants concentration among all the generated routes transportation modes. The bus trip sections have a significant pollution occurrence, but it is necessary to consider again the occupants carried in the public transportation compared to HPV and private vehicles.



(a) CO_2 emission in defined routes.



(b) Average CO_2 emission in defined routes.

Figure 7: CO_2 emission calculation results

4.7 Chapter Conclusions

This chapter explained the multi-modal route selection in an urban computing scenario with the building dataset from LBSN interaction data. After the data analysis, the main mobility flows aggregate the user behavior to demonstrate the hybrid routes' effectiveness. The emission calculation algorithm adds the environmental feature to consider in urban routes since the air pollution concentration growth and health threats. The comparison of performance evaluation shows the trade-off between single-mode and hybrid routes when choosing a cheaper or faster alternative. The hybrid method introduces a viable way to overcome economic and time-related problems in traditional urban transportation.

CHAPTER 5

Multicriteria Route Selection Method in a Urban Computing Scenario

This Chapter describes how our method works, combining the multi-criteria decision-making method and the urban routes' contextual data for the personalized route selection. We also define the necessary steps for validating the AHP criteria preferences and introduce the personalized user profiles. Thus, we detail the method application specifying the implemented algorithm, highlighting its efficiency in the personalized profiles performance evaluation.

This Chapter presents a different application in the urban scenario, considering the described concepts in Chapter 4. The personalized multi-criteria provides an alternative selection for single-modal trips but considers more trip-related features.

5.1 Scenario Overview

Figure 8 presents the overview for a personalized experience-aware multi-criteria route selection scheme. The criteria definition step processes all contextual route-related data. The selection method step defines the AHP method criteria and alternatives validation with user profile weights description. Finally, the method evaluation steps evaluate the method application by comparing the custom profiles with greedy profiles.

The data acquisition phase in the criteria definition step consists in retrieving all contextual and physical data for the dataset build. The authors [52] collected open data from websites and geographic tools to complete the dataset. We added the pollution factor through the local air quality open database [53]. In the Data characterization phase, we insert all contextual and physical feature values to the route alternatives, updating the

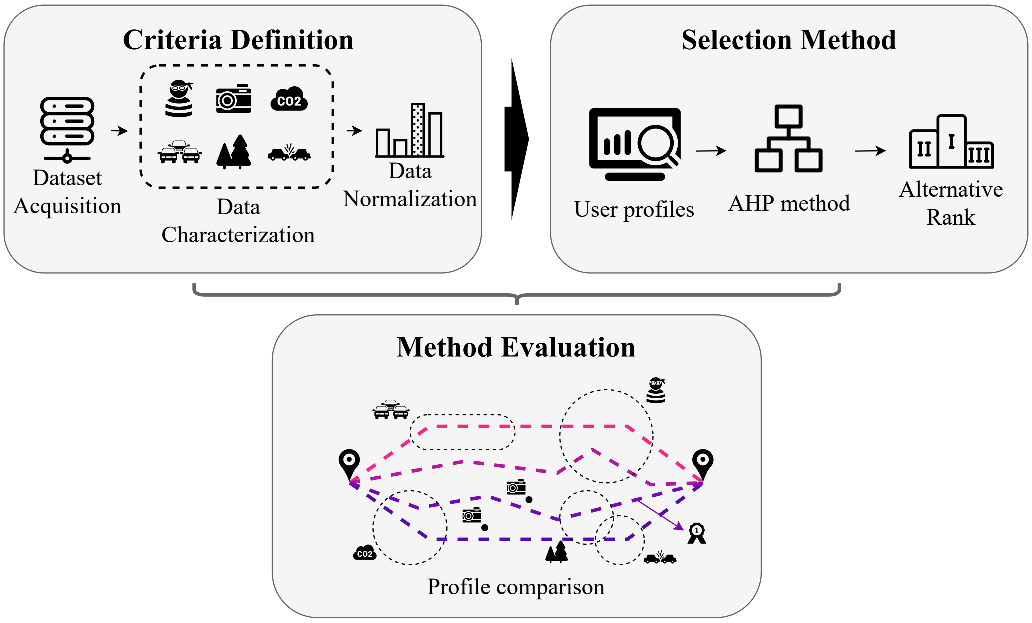


Figure 8: Methodology overview for route selection

London routes [1] dataset with the pollution level. The updated dataset contains eight criteria elements, described as follows:

- **Crime**: This criterion is related to the criminality level considering crime event history in determined areas. An open data United Kingdom police repository [54] containing all geolocated crime records are analyzed, and the average crime severity assigns the crime criterion value.
- **Accidents**: Defines a danger level to vehicle accidents near a determined route. The United Kingdom government's open data repository [54] provides geolocated accident records. The accident severity degree and the fatalities that occurred define the criterion value.
- **Nature**: Natural landscapes and "green" areas affect trip aesthetics. Parks, gardens, marinas, golf fields, nature reserves, lawns, meadows, and water define a pleasant trip and decrease driver stress. The Overpass API provides the natural occurrence through OpenStreetMap API [55], allowing the criterion value through the intersected area between nature polygons.
- **Attractions**: Defines the tourist attractions near the route traces. Overpass API [55] provides geolocated POI data. The attraction level indicates the POIs number in the region.
- **Duration**: Defines a traditional parameter for a vehicular navigation system affecting driver trip perception. Long trips may be a stressful experience and widely avoided. HERE API [56] provides the estimated duration for each route for alternative route tuple adding.

[1] https://ieee-dataport.org/open-access/crawdad-ufrjlondon-trajectories

- **Traffic**: Represents the most stress-related trip factor, implying in-route travel time. HERE API [56] provides the route vehicle density level. The traffic level is the comparison between duration with and without vehicle density.
- **Length**: Navigation system elementary factor provided by HERE Maps API [56], directly impacting vehicle combustion consumption and travel financial cost.
- **Pollution**: We added the pollution factor for the dataset with the London Air Quality Open Data [53], avoiding threatening population's health. The raw values the NO_2 concentration level near the route, with a $300m$ sensor tolerance.

The data normalization phase standardizes each criteria raw value from 0 to 1 for multi-criteria application on alternative selections. The Equation 5.1 calculates the normalized value for each criterion raw value (X_i) in the dataset, with X_{max} representing the maximum value and X_{min} the minimum value, which the lower occurrence indicates a better index for selection, such as crime and accidents occurrence, estimate duration, trip length, and pollution level. The Equation 5.2 normalizes the criterion raw value inversely for some criteria for which the higher occurrence indicates a better index for selection, such as natural areas, tourist attractions, and traffic ratio.

$$\widehat{X_i} = \frac{X_{max} - X_i}{X_{max} - X_{min}} \tag{5.1}$$

$$\widehat{X_i} = \frac{X_i - X_{min}}{X_{max} - X_{min}} \tag{5.2}$$

In the user profile phase of the selection method step, we define the pairwise preference comparison between features for the four profiles (Worker, Green, Safe, and Tourist). Afterward, the AHP method phase guarantees the matrix consistency for preference if it needs any correction. The alternative rank phase defines the best route selection as the product between the preference and alternative weights.

Finally, the method evaluation step analyses the best result for all routes for the user and greedy profiles under a profile comparison, corresponding to selection preference with higher priority on only one feature. These comparison objectives validate the user preference as a practical way to make a route choice.

5.2 Selection Method

In this section, we introduce the AHP methodology to obtain the selection weights for the proposed criteria. Figure 9 shows the route selection objective hierarchy after the method application, defining the objective, criteria, and alternatives. The MCDM method applied to the route selection context complements the traditional car navigation system factors with pleasant, health, and security factors. As initially arranged in the dataset, alternative route options for OD pair sets range from two to seven selection paths, as seen in the Figure 9 bottom line.

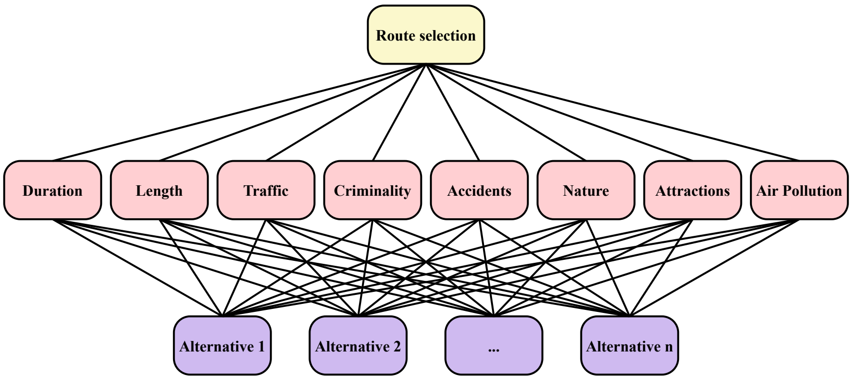


Figure 9: Hierarchy model for route selection representation

We define the four customized user preferences for the multi-criteria route selection method application: Worker, Green, Safe, and Tourist. In this way, we define four profile matrices to achieve the relative weights for further method application, as shown in Table 7. The result weights for Higher criteria weights indicate a higher preference, while smaller criteria indicate the opposite. The alternative evaluation process will use criteria weights for selection.

Table 7: Trip feature weights for different profiles

AHP Profile	Crime	Accident	Nature	Attraction	Duration	Traffic	Length	Pollution
gray!10 Worker	0.046	0.063	0.021	0.021	0.260	0.227	**0.328**	0.034
Green	0.085	0.040	0.280	0.087	0.031	0.055	0.059	**0.362**
gray!10 Safe	**0.369**	0.244	0.024	0.023	0.073	0.129	0.047	0.092
Tourist	0.164	0.101	0.117	**0.394**	0.058	0.018	0.011	0.044

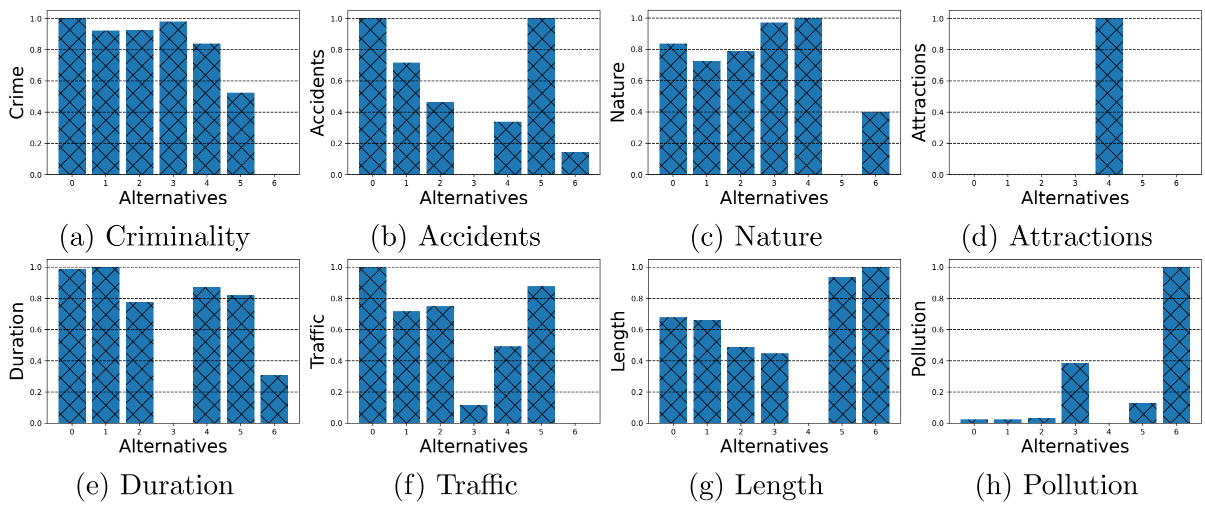
For instance, the Worker profile has a higher weight in the Length feature, followed by Duration and Traffic, aiming for faster trips [57]. The Green profile feature rank is Pollution and Nature for a bucolic and healthier trip [58]. Safe profile seeks a securer trip, prioritizing Crime occurrence and Accidents [59]. At last, the Tourist profile is for travelers and visitors, with higher weights on the Attraction feature [60]. For the research purpose, we consider only four profiles for demonstration; the method can work with any preference as long as the matrix is valid for the AHP method.

The route selection method defines the alternative paths' preference order using the relationship between the criteria index in each alternative tuple and the criteria weights. The used dataset contains a normalization acquired from contextual data information for each feature; we can apply the decision method directly since the normalized index represents a quantitative value, as defined in Section 2.2.

Figure 10 depicts the relationship between each criteria index and the seven route alternatives from the OD pair "setID 56", comparing the performance in every route alternative. We calculate the performance of each alternative by multiplying it with the defined criteria weight. Figure 10a presents the index scale for criminality in every route, with the route `0` achieving the value one and considered the securer; route `6` is the most dangerous path with value zero.

Figure 10b presents the accidents historical nearby the route alternative, with value one indicating a less chance of car accidents, indicated by routes `0` and `5`. Figure 10c and 10d depicts the value one for routes nearby green areas and tourist attractions, with route `4` the higher value. For attraction and nature, more occurrence means a good level justifying the reverse normalization.

Figures 10e, 10f, and 10g relate to Duration, Traffic, and Length, respectively; value one represents a minor level for those three and better performances among alternatives. Finally, Figure 10h indicates the pollution level for each alternative route, with value one in route `6` for routes with less pollution in the sensor readings.



(a) Criminality (b) Accidents (c) Nature (d) Attractions

(e) Duration (f) Traffic (g) Length (h) Pollution

Figure 10: Criteria index seven routes alternatives from "setID 56"

5.3 Method Application

Algorithm 4 computes the method application for matrix consistency calculation and the route evaluation for each OD pairs in order. `line 1 to line 4` declares the consistency index, the consistency ratio, and the F value for matrix weights . In `lines 5` to `14`, we applied the AHP method for attributing the preference weights. In `lines 15` to `17`, we define the variables for alternative evaluation. At least, in `lines 18` to `20`, we calculate the alternative route performance by multiplying every preference tuple value with each normalized feature value, returning the result array for further comparison, as seen in `line 21`.

For computational method implementation and flexible route selection, many

Algorithm 4 Decision matrix consistency and route evaluation

Require: M,C
$incRat \leftarrow 1.41$
$consistencyRatio \leftarrow 0.10$
$F \leftarrow shape(M, 1)$
$weights \leftarrow$ a list of zeros in the range of F
for $i \leftarrow 1$ to $length(F)$ **do**
    $weights[i] \leftarrow reduce(F(x, y) = x.y, M[i, :]^{(1/F)})$
$weights \leftarrow weights/$sum of all elements in weights
$\lambda_{Max} \leftarrow mean(sum(M.weights)/weights)$
$consInd \leftarrow (\lambda_{Max} - F)/(F - 1)$
$RC \leftarrow consInd/incRat$
if $RC > consistencyRatio$ **then**
    return $\emptyset$
$routesParams \leftarrow C[columns[parameters]]$
$resultsArray \leftarrow \emptyset$
$N \leftarrow length(routesParams)$
for $i \leftarrow 1$ to N **do**
    $resultsArray \leftarrow sum(multiply(routesParams[i], weights))$
return $resultsArray$

programming tools achieve geolocated data filtering and route alternative order definition goals. The method framework aims at data analysis of factors, including pollution, and inserting each alternative tuple. Each criteria-defined value distinguishes the best route and the alternative order for any OD pair.

The AHP method application, jointly with the route evaluation algorithm, as shown in the Algorithm 4, presents a time complexity, in the worst case, as O*(n+m)*, where m is the number of features to be evaluated in the profile, and n represents the number of alternative routes within an OD pair. We consider the presented complexity efficient due to its asymptotic value being limited by a polynomial, guaranteeing the algorithm's scalability in a future application.

5.4 Evaluation

In this section, we describe the dataset and introduce the mechanism of pollution level attribution for paths near air quality sensors. Furthermore, we define the statistical tool for profile comparison to establish our personalized approach efficiency.

5.4.1 Methodology

The London routes were designed for selection methods evaluation, containing different factors besides the standard time, length, and traffic. The criminality, accidents, nature, and attractions metrics consideration imply more pleasant and safe trips through the city. To consider the drivers' health and well-being, we introduce the air quality attribution to routes through sensor readings and add the pollution value to the dataset.

In this way, we consider the London public pollution data, which allows air quality level attribution for each alternative route with collected readings timestamp in the same dataset date. London Air Quality Network (LAQN) API provides pollution sensor readings, with sensors installed in and around London. Integrated sensors network has a real-time data collection of main pollution-related gaseous substances: O_3, nitrogen dioxide (NO_2), and inhalable particles with a diameter smaller than 10 and 2,5 micrometers. The API request retrieves the 2020 readings information, with a significant presence of NO_2. Pollution feature considers NO_2 level, with normalization for route selection method application.

We assume a set of routes in the dataset as a latitude-longitude pair path set $R = \{1, ..., n\}$, $R \in R^{n\times2}$. The methodology considers $C = \{R|R \in R^{nx2}\}$ as the path alternatives with standard departure and arrival OD pair, and then it is considered $R_i \in C_k$ e $R_j \in C_k$ if and only if for the same arrival and destination paths. With $S = \{1, ..., m\}$, $S \in R^{m\times2}$ the available London sensor set for determined pollution agent, so for each point $r_i \in R$ the pollution record for sensors $s \in S$, as shown in Algorithm 5.

The London Sensor Network API used for pollution factor attribution does not attend routes far from sensors tolerance in the built dataset. Similarly, other routes can contain readings from more than one sensor and consider an incoherent pollution level. For such cases, the proposed method excludes invalid routes for correct attribution and considers a tolerance `t = 300 m`.

Algorithm 5 assigns pollution levels to alternative route points for the given OD pair. The pollution record to each latitude-longitude pair attribute the pollution value, calculated from $pollution(s_j)$. After declaring C as the path set and S as the available sensor set, `line 1` defines an empty set for pollution associated with the determined path, filled with the pollution value from the nearest sensor. `lines 2` and `l3` receive path length and starts sensor distance as 0. In `lines 4` to `6`, the algorithm starts the for-loop iterating $routePollution$ to every $routesPollution$ (path alternatives). In `lines 7` to `10`, a second for-loop inside the first one is initiated, iterating associated points near the path. `lines 11` to `15` begin the third loop iterating existing pollution sensor data for $pointPollution$. All loops finish attributing the values in `lines 16` to `21`. Finally, in `line 22`, the algorithm returns the pollution value for each route alternative in the given set ID. The algorithm returns 142 OD pairs with pollution-normalized values attributed, allowing the correct method application.

The Haversine function, shown in Equation 5.3 and `line 13`, calculates the distance between a tracepoint and a sensor considering the earth curvature. State-of-art solutions use geographic coordinates handling with this equation for the appropriate distance obtaining, represented by OD pair.

Algorithm 5 Routes pollution attribution and exclusion

```
Require: C ≠ ∅, S ≠ ∅, t = 300
routesPollution = ∅
N ← length(C)
distance ← 0
for k ← 1 to N do
    routePollution = ∅
    minorDistanceRoute ← ∞
    M ← length(c_k)
    for i ← 1 to M do
        P ← length(S)
        pointPollution ← 0
        minorDistancePoint ← ∞
        for j ← 1 to P do
            distance ← haversine(r_j, s_j)
            if distance < minorDistancePoint then
                minorDistancePoint ← distance
                pointPollution ← pollution(s_j)
        routePollution ← pointPollution
        if minorDistancePoint < minorDistanceRoute then
            minorDistanceRoute ← minorDistancePoint
    if minorDistanceRoute > t then
        return ∅
    routesPollution ← routePollution
return routesPollution
```

$$D = 2arcsin\left[\sqrt{sin^2(\frac{r_1 - s_1}{2}) + cos(r_1)cos(s_1)sin^2(\frac{r_2 - s_2}{2})}\right] \tag{5.3}$$

We compare the results from the four user profiles proposed to each greedy profile. The eight greedy profiles choose the alternative tuple with maximum value for only one criterion of all, as a common method in commercial navigation systems considers only Length, Traffic, and Duration when choosing a car trip. We measure the relation between every profile and the maximum value for a route choice with a mathematical method.

We use the PDFKS [40] for comparing the profiles to greedy profiles. The Equation 5.4 calculates the PDFKS value, where $p[f]$ is the average value from feature f for each profile p, and $stf[f]$ is the best average value for feature f among all profiles results from the dataset values.

$$M_{pf} = \frac{p[f] - std[f]}{std[f]} \times 100\% \tag{5.4}$$

5.4.2 Results

Figure 11 shows the PDFKS matrix, where the M matrix rows represent the selection profiles (p), *i.e.,* four user profiles and eight greedy profiles, and the columns represent the trip features (f) for evaluation. The PDFKS metric value represents the best average value for each profile to the known standard ($std[f]$). For example, the first Safe profile index shows a 3.6% increase to the best average value for crime, validating the Safe profile route selection with less crime than the other profiles after the onlyCrimes. Also, Safe presented a 39.5% deviation from the best average for accident feature, representing its second priority for route choice. This value does not represent the more significant performance for the accident due to the higher priority for crime feature and the data arrangement. In a feature less related to the main priority, the Safe profile scored a -46% deviation for the attraction feature, which its weight has less impact on selection.

In contrast, Nature, Attraction, and Traffic ratio have negative percentage PDFKS values in the matrix because the method searches for the higher nature and attraction occurrence and a higher traffic ratio that indicates a less congested road, implying on a raw value less than the known standard, resulting negative percentage. A lower raw value indicates the best route selection for all other features containing positive PDFKS values; the better selection method is with features closest to 0%.

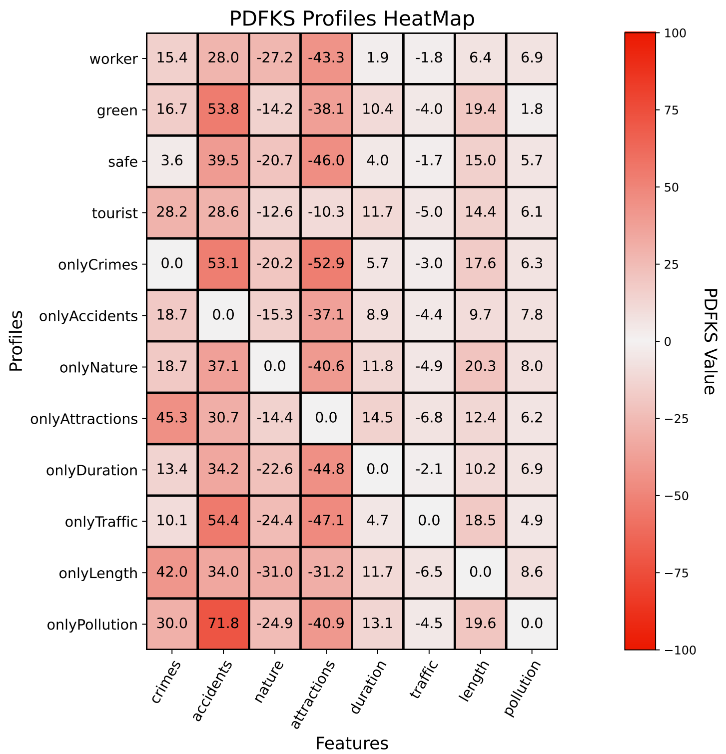


Figure 11: Methodology overview for route selection

We apply the absolute sum method for all 12 profiles, validating the user profiles, summing all elements without considering negative values. The lowest absolute sum of PDFKS for each profile represents the better selection method, considering all routes, as shown in Figure 12. We can note that greedy preferences have the known standard value (0%), indicating the best routes for a single feature, but tend to deviate more from all other features. Each proposed profile (Worker, Green, Safe, and Tourist) correlate to more than one feature, where we differentiate with colors the relationships and compare the resulting performance for all cases.

The Green profile has the closest value for the pollution standard and outranks the onlyPollution in other features and has the second best deviation from nature feature (-14.2%), resulting in a greener experience route. The Safe profiles outrank onlyCrimes (3.6%) and onlyTraffic (-1.7%), when the Crime feature is its higher priority, and higher traffic indicates slower paths and more dangerous routes, with the best deviation from crime. The worker profile overcomes its higher weighted features: onlyLength (6.4%), onlyDuration (1.9%), and onlyTraffic (-1.8%), surpassing the standard navigation systems in selection. The Tourist profile has better route selection than onlyAttraction, deviating from the standard attraction value with the best performance (-10.3%) and from nature feature (-12,6), with better performance for tourist users.

We note that all greedy profiles have the best performance for its features priorities but have a higher deviation in other features. The onlyAccidents have the lower absolute sum representing the best profile for selection in the evaluated environment. Otherwise, we note that the existence of an AHP profile with balanced weights for each criterion obtains better performance than single-criterion profiles, explaining the onlyAccidents higher performance. In other words, for diverse environments datasets, a specific profile with distributed weights, a priority for a few features, can outrank a greedy option.

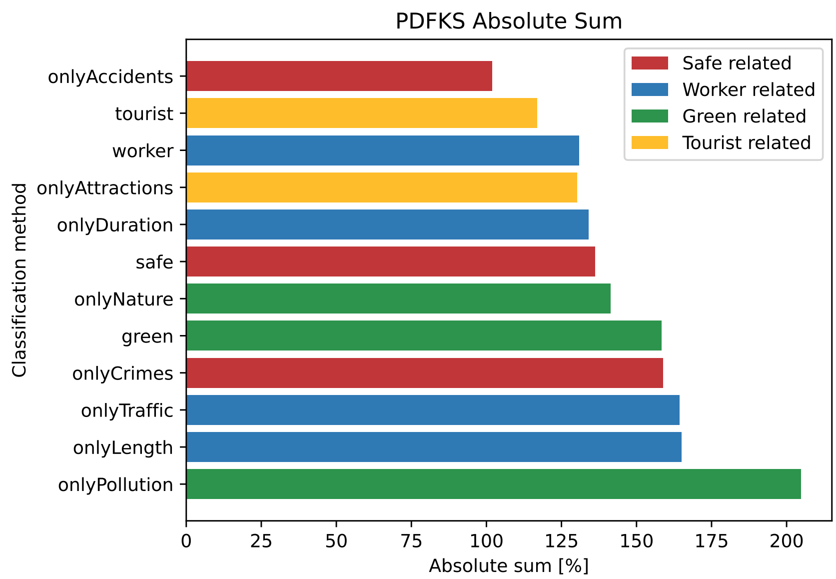


Figure 12: PDFKS absolute sum for each selection profile

In summary, our proposed user profile obtained a higher performance than its greedy opponents or was close to surpassing them. Tourist profile obtained the second (116,9%) best absolute sum and outranked the onlyAttractions greedy opponent by 13,4%, using the difference percentage. The Worker is the third (130,9%) best performance compared to greedy options, outranking them by 23,53% in the average difference percentage. Also, Safe and Green profiles surpassed their greedy opponents by 23,8% and 46,4%, respectively. This result indicates excellent usability for our method for considering all contextual data for selection than prioritizing only one criterion.

5.5 Chapter Conclusions

This chapter presented the personalized multi-modal route selection. The AHP method for selecting routes under the comfort, security, pleasant, and health features brings completeness to an urban navigation solution. The personalized user profiles ensure the selection based on customized preference, suggesting the most suitable path for a different need. The selection and pollution feature addition algorithms were explained as advanced compared to the state-of-the-art, adding a simpler method with bigger criteria. The performance evaluation depicts the customized profiles as an alternative to greedy options, like traditional navigation systems. The method shows the possibility of considering a large contextual data variety for the best route selection for any user preference.

The multi-modal method, seen in Chapter 4, considers five features for comparison between hybrid and single-modal, and the multi-criteria single-mode considers eight. Despite the difference between applications, both presented methods were developed for urban scenario applications and provide an efficient, personalized route selection with contextual information and pollution estimation.

CHAPTER 6

Conclusion

This Chapter presents the work conclusions. In Section 6.1, we present the main contributions of this master's thesis. Section 6.2 brings new implementations for the two route selection methods. Finally, Section 6.3 references the previous academic research supporting this master's thesis production.

6.1 Contributions

In this master thesis, We propose two route selection methods: a novel approach to multi-modal urban pollution-aware routing needs by using location-based social network collected data and an MCDM method with personalized routes for urban trip path choice considering all eight features. The multi-modal method offers less expensive, healthier trips for the population and data collection about carbon emissions for mobility planners to consider when turning urban scenarios dynamic and sustainable.

After acquiring and submitting the geo-located data from social networks, the mobility flows are identified with separated multi-modal used on routes. Each chosen modal CO_2 emission is calculated on the identified routes and evaluated on specific metrics, such as waiting time, walking distance, and price estimation. The proposed algorithm proves its efficiency and can be used by users of route applications, city authorities, and environmental studies on urban mobility, developing a better and more productive life quality for smart cities.

The multi-modal method considers all eight trip-related features, comparing the personalized profiles to greedy options for each feature, and we observe a result closest to the best value from all features' best average value compared. This method shows that traditional navigation systems can offer faster or healthier routes but can lead to dangerous or unpleasant paths. With the pollution factor addition, our method can prevent and alert the drivers and authorities to the polluted air threats, raising the quality of life. Furthermore, we developed the method with simple mathematical methods for easy applicability in navigation systems.

6.2 Future Works

For the multi-modal route selection method, a more efficient data acquiring method can be applied for a large set of geo-located data in less populated urban areas with pollution active sensing coverage in order of reliable emission sensing and mobility analysis.

Furthermore, a routing system can be built integrating various features and a multi-modal method for intelligent public transportation. The system can consider IoT-enable feature prediction for real-time route selection, integrating user devices into a more extensive urban computing solution.

6.3 Published Work

1. **BRITO, M.**; SANTOS, C.; OLIVEIRA, H.; CERQUEIRA, E.; ROSÁRIO, D. "Air Pollution Calculation for Location Based Social Networks Multimodal Routing Service", in *Anais do VI Workshop de Computação Urbana (CoUrb - 2022)*, June 2022. *Received Honorable Mention Award.*

2. **BRITO, M.**; MARTINS, B.; SANTOS, C.; MEDEIROS, I.; ARAÚJO, F.;SERUFFO, M.; OLIVEIRA, H.; CERQUEIRA, E.; ROSÁRIO, D. "Personalized Experience-aware Multi-criteria Route Selection for Smart Mobility", in *Anais do XLI Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos (SBRC - 2023)*, June 2023.

References

[1] J. Monstadt, "Urban and infrastructural rhythms and the politics of temporal alignment," *Journal of urban technology*, vol. 29, no. 1, pp. 69–77, 2022.

[2] A. Asadzadeh, T. Kötter, A. Fekete, M. Moghadas, M. Alizadeh, E. Zebardast, D. Weiss, M. Basirat, and G. Hutter, "Urbanization, migration, and the challenges of resilience thinking in urban planning: Insights from two contrasting planning systems in germany and iran," *Cities*, vol. 125, p. 103642, 2022.

[3] L. Bertolini, "From "streets for traffic" to "streets for people": can street experiments transform urban mobility?" *Transport reviews*, vol. 40, no. 6, pp. 734–753, 2020.

[4] K. C. Von Schönfeld and L. Bertolini, "Urban streets: Epitomes of planning challenges and opportunities at the interface of public space and mobility," *Cities*, vol. 68, pp. 48–55, 2017.

[5] D. C. Nguyen, M. Ding, P. N. Pathirana, A. Seneviratne, J. Li, D. Niyato, O. Dobre, and H. V. Poor, "6g internet of things: A comprehensive survey," *IEEE Internet of Things Journal*, vol. 9, no. 1, pp. 359–383, 2021.

[6] R. Mota, A. Riker, and D. Rosário, "Adjusting group communication in dense internet of things networks with heterogeneous energy sources," in *Anais do XI Simposio Brasileiro de Computacao Ubiqua e Pervasiva*. SBC, 2019.

[7] E. Lobato, L. Prazeres, I. Medeiros, F. Araújo, D. Rosário, E. Cerqueira, M. Tostes, U. Bezerra, W. Fonseca, and A. Antloga, "A monitoring system for electric vehicle charging stations: A prototype in the amazon," *Energies*, vol. 16, no. 1, p. 152, 2023.

[8] F. Araújo, S. Pinheiro, I. Medeiros, D. Rosário, E. Cerqueira, U. Bezerra, M. Tostes, and A. Antloga, "Plataforma para coleta de dados de eletropostos para veículos elétricos: Prototipação em bancada de testes real," in *Anais do XIV Simpósio Brasileiro de Computação Ubíqua e Pervasiva*. SBC, 2022, pp. 121–130.

[9] M. Brito, C. Santos, H. Oliveira, E. Cerqueira, and D. Rosário, "Air pollution calculation for location based social networks multimodal routing service," in *Anais do VI Workshop de Computação Urbana*. SBC, 2022, pp. 280–293.

[10] R. Savithramma, B. Ashwini, and R. Sumathi, "Smart mobility implementation in smart cities: a comprehensive review on state-of-art technologies," in *2022 4th International Conference on Smart Systems and Inventive Technology (ICSSIT)*. IEEE, 2022, pp. 10–17.

[11] R. M. Andrade, R. M. Carvalho, I. L. de Araújo, K. M. Oliveira, and M. E. Maia, "What changes from ubiquitous computing to internet of things in interaction evaluation?" in *Distributed, Ambient and Pervasive Interactions: 5th International Conference, DAPI 2017, Held as Part of HCI International 2017, Vancouver, BC, Canada, July 9–14, 2017, Proceedings 5*. Springer, 2017, pp. 3–21.

[12] S. Ribeiro-Navarrete, J. R. Saura, and D. Palacios-Marqués, "Towards a new era of mass data collection: Assessing pandemic surveillance technologies to preserve user privacy," *Technological Forecasting and Social Change*, vol. 167, p. 120681, 2021.

[13] M. Umair, M. A. Cheema, O. Cheema, H. Li, and H. Lu, "Impact of covid-19 on iot adoption in healthcare, smart homes, smart buildings, smart cities, transportation and industrial iot," *Sensors*, vol. 21, no. 11, p. 3838, 2021.

[14] J. Legaspi, S. V. Bhada, P. Mathisen, and J. DeWinter, "Smart city transportation: A multidisciplinary literature review," in *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2020, pp. 957–964.

[15] P. Bellini, P. Nesi, and G. Pantaleo, "Iot-enabled smart cities: A review of concepts, frameworks and key technologies," *Applied Sciences*, vol. 12, no. 3, p. 1607, 2022.

[16] M. Brito, B. Martins, C. Santos, I. Medeiros, F. Araujo, M. Seruffo, H. Oliveira, E. Cerqueira, and D. Rosário, "Personalized experience-aware multi-criteria route selection for smart mobility," in *Anais do XLI Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos*. SBC, 2023.

[17] G. Yannis, A. Kopsacheili, A. Dragomanovits, and V. Petraki, "State-of-the-art review on multi-criteria decision-making in the transport sector," *Journal of traffic and transportation engineering (English edition)*, vol. 7, no. 4, pp. 413–431, 2020.

[18] S. Penazzi, R. Accorsi, and R. Manzini, "Planning low carbon urban-rural ecosystems: An integrated transport land-use model," *Journal of cleaner production*, vol. 235, pp. 96–111, 2019.

[19] S. Lee, "Exploring associations between multimodality and built environment characteristics in the us," *Sustainability*, vol. 14, no. 11, p. 6629, 2022.

[20] X. Zhang, L. Han, H. Wei, X. Tan, W. Zhou, W. Li, and Y. Qian, "Linking urbanization and air quality together: A review and a perspective on the future sustainable urban development," *Journal of Cleaner Production*, p. 130988, 2022.

[21] J. Zawieska and J. Pieriegud, "Smart city as a tool for sustainable mobility and transport decarbonisation," *Transport Policy*, vol. 63, pp. 39–50, 2018.

[22] T. Sobral, T. Galvão, and J. Borges, "Visualization of urban mobility data from intelligent transportation systems," *Sensors*, vol. 19, no. 2, p. 332, 2019.

[23] D. Rodrigues, F. Santos, G. Rocha Filho, A. Akabane, R. Cabral, R. Immich, W. L. Junior, F. Cunha, D. Guidoni, T. Silva, D. Rosário, E. Cerqueira, A. Loureiro, and L. Villas, "Computação urbana da teoria à prática: Fundamentos, aplicações e desafios," in *Minicursos do XXXVII Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos*. SBC, May 2019, pp. 51–91.

[24] Y. Zheng, L. Capra, O. Wolfson, and H. Yang, "Urban computing: concepts, methodologies, and applications," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 5, no. 3, pp. 1–55, 2014.

[25] S. E. Bibri and J. Krogstie, "Smart sustainable cities of the future: An extensive interdisciplinary literature review," *Sustainable cities and society*, vol. 31, pp. 183–212, 2017.

[26] A. Baklanov, B. Cárdenas, T.-c. Lee, S. Leroyer, V. Masson, L. T. Molina, T. Müller, C. Ren, F. R. Vogel, and J. A. Voogt, "Integrated urban services: Experience from four cities on different continents," *Urban Climate*, vol. 32, p. 100610, 2020.

[27] S. Grimmond, V. Bouchet, L. T. Molina, A. Baklanov, J. Tan, K. H. Schlünzen, G. Mills, B. Golding, V. Masson, C. Ren *et al.*, "Integrated urban hydrometeorological, climate and environmental services: Concept, methodology and key messages," *Urban Climate*, vol. 33, p. 100623, 2020.

[28] T. L. Saaty, "How to make a decision: the analytic hierarchy process," *European journal of operational research*, vol. 48, no. 1, pp. 9–26, 1990.

[29] I. Medeiros, "Qoe and qos-aware handover for video transmission in heterogeneous vehicular networks," Master's thesis, Federal University of Pará, 2018.

[30] R. Haas and O. Meixner. An illustrated guide to the analyic hierarchy process. [Online]. Available: https://goo.gl/TA4b9g

[31] S. Porru, F. E. Misso, F. E. Pani, and C. Repetto, "Smart mobility and public transport: Opportunities and challenges in rural and urban areas," *Journal of traffic and transportation engineering (English edition)*, vol. 7, no. 1, pp. 88–97, 2020.

[32] R. Rothfeld, M. Fu, M. Balać, and C. Antoniou, "Potential urban air mobility travel time savings: An exploratory analysis of munich, paris, and san francisco," *Sustainability*, vol. 13, no. 4, p. 2217, 2021.

[33] P. Rode, G. Floater, N. Thomopoulos, J. Docherty, P. Schwinger, A. Mahendra, and W. Fang, "Accessibility in cities: transport and urban form," *Disrupting mobility: Impacts of sharing economy and innovative transportation on cities*, pp. 239–273, 2017.

[34] F. Canitez, P. Alpkokin, and S. T. Kiremitci, "Sustainable urban mobility in istanbul: Challenges and prospects," *Case studies on transport policy*, vol. 8, no. 4, pp. 1148–1157, 2020.

[35] J. Kalajdjieski, E. Zdravevski, R. Corizzo, P. Lameski, S. Kalajdziski, I. M. Pires, N. M. Garcia, and V. Trajkovik, "Air pollution prediction with multi-modal data and deep neural networks," *Remote Sensing*, vol. 12, no. 24, p. 4142, 2020.

[36] B. Zou, S. Li, Z. Zheng, B. F. Zhan, Z. Yang, and N. Wan, "Healthier routes planning: A new method and online implementation for minimizing air pollution exposure risk," *Computers, Environment and Urban Systems*, vol. 80, p. 101456, 2020.

[37] G. Inturri, N. Giuffrida, M. Le Pira, M. Fazio, and M. Ignaccolo, "Linking public transport user satisfaction with service accessibility for sustainable mobility planning," *ISPRS International Journal of Geo-Information*, vol. 10, no. 4, p. 235, 2021.

[38] J. M. Diez, M. E. Lopez-Lambas, H. Gonzalo, M. Rojo, and A. Garcia-Martinez, "Methodology for assessing the cost effectiveness of sustainable urban mobility plans (sumps). the case of the city of burgos," *Journal of Transport Geography*, vol. 68, pp. 22–30, 2018.

[39] S. Jat, R. S. Tomar, and M. S. P. Sharma, "Traffic analysis for accidents reduction in vanet's," in *2019 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE)*. IEEE, 2019, pp. 115–118.

[40] L. Solé, M. Sammarco, M. Detyniecki, and M. E. M. Campista, "Towards drivers' safety with multi-criteria car navigation systems," *Future Generation Computer Systems*, vol. 135, pp. 1–9, 2022.

[41] A. P. Ferreira, T. H. Silva, and A. A. Loureiro, "Uncovering spatiotemporal and semantic aspects of tourists mobility using social sensing," *Computer Communications*, vol. 160, pp. 240–252, 2020.

[42] D. O. Rodrigues, A. Boukerche, T. H. Silva, A. A. Loureiro, and L. A. Villas, "Combining taxi and social media data to explore urban mobility issues," *Computer Communications*, vol. 132, pp. 111–125, 2018.

[43] D. O. Rodrigues, J. T. Fernandes, M. Curado, and L. A. Villas, "Hybrid context-aware multimodal routing," in *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2018, pp. 2250–2255.

[44] M.-Y. Wu, C.-K. Ke, and S.-C. Lai, "Optimizing the routing of urban logistics by context-based social network and multi-criteria decision analysis," *Symmetry*, vol. 14, no. 9, p. 1811, 2022.

[45] R. Sarraf and M. P. McGuire, "Integration and comparison of multi-criteria decision making methods in safe route planner," *Expert Systems with Applications*, vol. 154, p. 113399, 2020.

[46] S. Kaivonen and E. C.-H. Ngai, "Real-time air pollution monitoring with sensors on city bus," *Digital Communications and Networks*, vol. 6, no. 1, pp. 23–30, 2020.

[47] D.-G. Zhang, J.-X. Wang, J. Zhang, T. Zhang, C. Yang, and K.-W. Jiang, "A new method of fuzzy multicriteria routing in vehicle ad hoc network," *IEEE Transactions on Computational Social Systems*, pp. 1–13, 2022.

[48] H.-P. Hsieh and F. Lin, "Recommending taxi routes with an advance reservation – a multi-criteria route planner," *International Journal of Urban Sciences*, vol. 26, no. 1, pp. 162–183, 2022.

[49] D. O. Rodrigues, A. Boukerche, T. H. Silva, A. A. Loureiro, and L. A. Villas, "Smaframework: Urban data integration framework for mobility analysis in smart cities," in *2017 20th ACM International Conference on Modelling, Analysis and Simulation of Wireless and Mobile Systems (MSWiM)*. ACM, 2017, p. 227–236.

[50] O. d. M. Álvares Jr. and R. R. A. Linke, "Metodologia simplificada de cálculo das emissões de gases do efeito estufa de frotas de veículos no brasil," *São Paulo: CETESB*, vol. 182, 2001.

[51] IPCC, "Ipcc guidelines for national greenhouse gas inventories," *Reference manual*, vol. 3, 1996.

[52] L. S. Rodrigues, M. Sammarco, M. Detyniecki, and M. E. M. Campista, "CRAWDAD dataset ufrj/london-trajectories (v. 2021-03-22)," Downloaded from https://crawdad.org/ufrj/london-trajectories/2021-03-22, Mar. 2021.

[53] F. J. Kelly and J. Kelly, "London air quality: a real world experiment in progress," *Biomarkers*, vol. 14, no. sup1, pp. 5–11, 2009.

[54] S. E. Bibri and J. Krogstie, "The emerging data–driven smart city and its innovative applied solutions for sustainability: The cases of london and barcelona," *Energy Informatics*, vol. 3, pp. 1–42, 2020.

[55] OpenStreetMap. (2017) Planet dump retrieved from https://planet.osm.org. [Online]. Available: https://www.openstreetmap.org

[56] HERE, "Build apps with here maps api and sdk platform access: Here developer," 2023. [Online]. Available: https://developer.here.com/

[57] J. B. Chiquetto, A. R. Leichsenring, F. N. Ribeiro, and W. C. Ribeiro, "Work, housing, and urban mobility in the megacity of são paulo, brazil," *Socio-Economic Planning Sciences*, vol. 81, p. 101184, 2022.

[58] M. Asghari, S. M. J. M. Al-e *et al.*, "Green vehicle routing problem: A state-of-the-art review," *International Journal of Production Economics*, vol. 231, p. 107899, 2021.

[59] S. Sohrabi, Y. Weng, S. Das, and S. G. Paal, "Safe route-finding: A review of literature and future directions," *Accident Analysis & Prevention*, vol. 177, p. 106816, 2022.

[60] J. Ruiz-Meza and J. R. Montoya-Torres, "Tourist trip design with heterogeneous preferences, transport mode selection and environmental considerations," *Annals of Operations Research*, vol. 305, no. 1-2, pp. 227–249, 2021.