Mobility Patterns through a Mobile App for Public Transport Users

José Gonçalo Simões Rodrigues

Abstract—Today’s transportation systems rely more and more on the ability to offer a quick and nonrestrictive experience to their users. To allow for such an experience the need to make available relevant information to be utilized to facilitate the interaction of user and operator, must to be observed. In order to facilitate this transaction of information, this work will shows in the form of an Android application, that it is possible to create a system capable of gathering information about the mobility patterns of a public transportation user with the intent of improving the decision making process both users and operators make when interacting. This solution will focus on offering a low cost system to operators that is easy to implement on their existing systems, and seamless in the form it gathers information from the users.

Index Terms—Mobility Patterns, Transportation System, Seamless, Transparent, Smartphone.

1 INTRODUCTION

With higher user expectations about the value of a reliable and effective Public Transportation System, a major necessity always exists to ensure the optimization of resources available to the operator of a Public Transportation System (PTS). This factor can prove to be the defining difference that helps users choose with what PTS they will engage with.

While the traditional methods that are employed to obtain information about resource use in a PTS often rely on first-hand accounts from the PTS workers, or from user feedback to inform the decision making of the PTS operator, such means of communication may prove not to have a sufficiently quick response time for the PTS operator to react in real-time to this information.

1.1 Motivation

As a general trend, the number of public transportation users has been increasing at a constant rate in the European Union, and is expected that this growth will not slow down in the near future [1]. With this increase in trips, PTSs are expected to handle the demanding requirements users impose on the offer of wide reaching and on-schedule transportation [2].

As such, in order to facilitate and empower the ability for a PTS operator to manage their resources based on reliable data, there exists a need for a solution that facilitates the exchange of information about how users interact with a PTS. By obtaining patterns that individual users experience on a PTS, it allows for the operator to better modulate their resources by making use of an informed decision making process and ultimately help to meet the requirements users expect.

1.2 Concept Overview

Mobility Patterns (MP) Are sets of data that directly represent the patterns of movements realized by individuals, generated on a given trip from a point A to a point B during an interval of time. Were an example can be the patterns a user makes on their daily journey to and from work.

PTS Network (PTSN) A PTSN is a network
consisting of various public transportation systems. These reflect a number of possible transportation formats, and possible trajectories a PTS user can realize. While most PTSN operators do in fact only employ one type of transportation, it should still be kept in consideration that users do not.

1.3 Objectives

The objective of this work was to create a simplified prototype built as an Android application, that is minimal in its impact on the preexisting infrastructure used by a PTSN operator, and the user, allowing for a more seamless ability to easily inspect their user’s mobility patterns.

2 STATE OF THE ART

For the principle of tracking users to improve the usability of a service, an operator’s ability to effectively and efficiently make use of mobility patterns is essential to improve any public transportation service as previously stated in Section 1.1 as such establishing a method capable of tracking users without an explicit action on their part to initiate the tracking is an important factor in this improvement.

2.1 Tracking Methods

Tracking Methods make use of technologies that enable the tracking of users in an environment, such as:

Global Positioning System (GPS): GPS is a tracking technology that makes use of a network of satellites (in an orbit at approximately 20,000 kilometers from the surface of the Earth) to deduce the location of a capable receiver [3].

Global System for Mobile Communications (GSM): GSM is a standard used to describe cellphone communication [4]. Some studies have tried to varying levels of success to utilize GSM as a form of tracking. Such a case was realized by utilizing signal towers as a form of tracking a user location via signal triangulation [5].

Wireless Fidelity (WiFi): WiFi is a technology utilized for wireless communication between two or more devices by means of a Wireless LAN working on the Electromagnetic (EM) spectrum. WiFi signal is transmitted using a WiFi - Access Point (WiFi-AP) that is also responsible for managing the network to all devices trying to, or already connected with it. Solutions based on WiFi are well studied such as:

- The use of fingerprint methodologies [6].
- The use of triangulation methodologies [7].
- The use of both triangulation and fingerprint methodologies [8].
- The use of other technologies to create hybrid systems for tracking users [9].

Computer Vision (CV): CV is the field that studies how a computer is capable of interpreting and extracting information from an image [10]. CV has been shown to be capable of tracking individuals by pairing a detected face in a photo/video to the location where the recording took place.

BlueTooth (BT): BT is a wireless technology standard utilized for medium to short range wireless communication between two devices [11]. BT has been frequently studied as a solution to the acquire the location of users [12]. Much as WiFi, BT location technologies derive into two possible solutions, fingerprinting and triangulation. It should be noted that both solutions of fingerprint and triangulation require the creation (in some form) of a network of BT signal sources of well-known fixed coordinates, as BT does not share the network layout made available by the use of WiFi-APs.

Dead Reckoning (DR): DR is an inertial tracking system. It utilizes the natural motion of its intended user to deduce direction of movement, speed, and distance traveled [13]. Such systems require the use of an accelerometer and a magnetometer, both widely spread on smartphones, of which
neither is responsible for a noticeable battery drain [14]. Tracking systems using DR have been created to track user activity [15], and even to track users’ positions relative to themselves (how much they have walked and in what direction) [16].

2.2 Tracking Auxiliaries

Are technologies that, while not capable of directly tracking a user, can help in the task. Examples of such tracking auxiliaries are:

**BlueTooth Beacon (BB):** BB are small devices capable of holding limited information, no more than a few bytes representing their own identification [17]. These beacons are designed to operate unattended for long periods of time ranging from one month to three years. Thanks to these characteristics, BB’s can complement tracking via BT, by operating as stationed signal emitters.

**Near-Field Communication (NFC):** NFC is a protocol that enables devices to communicate while at short ranges from each other, and is a continuation of Radio-Frequency Identification (RFID) [18]. NFC has been used on event-based tracking systems [19]. These systems require the user to explicitly “state” their location, which is done when the user realizes a specific tasks that utilize NFC (such as authenticating themselves on a bus) [20].

**Wearables:** Wearables are a form of mobile devices that are directly worn by their users (smartwatch, and smartband are two examples of such devices). While these devices are not capable of complex computation, they are usually paired with a smartphone to bolster their small range of sensors (where the most common sensors are an accelerometer and BT sensor, but not the only ones present among wearable devices).

**Crowd Sensing:** Crowd Sensing is the concept where by taking advantage of the mass dispersal of smartphones into the general population [21], applications can be employed, that will measure certain information criteria, such as average population density on a public zone, from a population of users rather than from a single user [22]. This use of “volunteer sensors” to effectively create an array of sensors that can generate a set of data that reflects the target criteria.

**Context Aware Detection:** Context Aware Detection is the concept of utilizing the various available sensors or meta-sensors available in a device to stipulate the current context the device and its user might be experiencing based on previously analyzed data from said sensors [23].

**Machine Learning (ML):** ML is the concept of creating programs capable of learning how to accomplish a task without the need to explicitly program the capability to deal with said task [24]. This concept has been utilized to identify patterns in large and complex data sets, in the effort of obtaining usable information, in a process named Data Mining [25]. Machine Learning can also be employed to analyze sets of data in an effort to understand complex situations, such as the location of mobile users [26].

2.3 Related Work

2.3.1 Mobility Agents: Guiding and tracking public transportation users [27]

Focuses on creating a system that allows for caregivers to manage and track possible transport options. Using a “city radar” that could map points-of-interest to a user, realized via a radar like display on the mobile device of the user, where the information to be displayed is acquired by crawling agents that seek for relevant information on the web regarding the general area the system is being used on. While among the first of its kind, it proved to be a successful implementation of how a user can have some improvement in their use of a PTSN, as the work allows for a less restrictive interaction between a previously constrained user and a PTSN operator, granting a greater
general level of usability to the users of the PTSN.

2.3.2 Learning and inferring transportation routines [28]
Focuses on the creation of a system capable of learning a given user Mobility Patterns (MP), in order to allow for improvements on the users day-to-day life.
This system makes use of the GPS signal to derive both location and velocity of a user (the later being determined considering the time and distance from between each sampling), to determine what mode of transportation a user is on (none, car, foot), and partitions the whole trip into segments that represent each mode of transportation.
To correctly deduce the criteria stated above, the system uses a learning mechanism, that will from a baseline, improve on its ability to accurately realize the MP of its user.
The prototype developed by the authors demonstrated that it is possible to obtain a MP using a smartphone, but the application was restricted due to the over-reliance in GPS, resulting in sometimes incorrect MP acquisition.

2.3.3 Exploring ticketing approaches using mobile technologies: QR Codes, NFC and BLE [29]
By taking into consideration the form in which a PTSN operator grants users access to the transports in their network, this work proposes that the current ticketing experience based on “token” authentication can be improved by the use of mobile technologies. The explored mobile technologies were Quick Response (QR) codes, NFC and BT.
Of particular interest are the characteristics shown when utilizing each technology:
- BT proved to be poor in determining distances, but showed no problem detecting “beacon zones”, thus allowing for a seamless experience to the user only affected by the rapidity in which the system detected the zone.
- NFC was the most reliable but still required an unobstructed sensor-to-reader path to obtain the best results.

This work shows how existing methods can be augmented or replaced with smartphone aware substitutes that allow for a more user friendly experience, with the added benefit of making possible the tracking of users in a PTSN.

2.3.4 Understanding individual and collective mobility patterns from smart card records: A case study in Shenzhen [30]
Proposes that by utilizing only preexisting means such as RFID smartcards, it is possible to attain the MP associated with use of a PTSN. While creating a solution restricted to a metro system (an environment where entry and exits are tightly controlled), this work exposes the readiness in which anonymous MPs can be extracted when employing proper Data Mining (DM) methods.

2.3.5 IoT Middleware for Precision Agriculture - Agricultural Labor Monitoring for Specialty Crops [31]
Proposes a system for tracking worker activity in specialized crops. The solution involves the extensive use of DR to acquire the worker activity and position, utilizing GPS only to compensate for the errors DR tracking accumulates over time.
This system proves that it is possible to build a system primarily reliant on DR to track user activity by matching sensor readings with ‘learned’ patterns from the DR sensors, where this tracking of activity is in effect creating a mobility pattern for worker activity on the crop.

2.3.6 Using Mobile Phones to Determine Transportation Modes [32]
Focuses on the idea of utilizing embedded sensors in a smartphone, such as an accelerometer and GPS sensor to deduce the current transportation mode of a user, therefore extracting information about how users behave in terms of their mobility. To accomplish this, DR means are utilized to understand the current context of mobility that an individual is experiencing, as well as fluctuations in GPS signal strength to understand if a user as entered/Exited a certain environment. However this solution suffered
greatly from battery drain derived from the necessity to constantly sample GPS signal.

2.3.7 My Mobile Assistant [33]
Focuses on a solution to track user mobility patterns during the daily commute to and from work. The work was realized with the objective of providing the user a form of tracking their own activities, in order to further improve the usability of a PTSN. While the prototype developed showed promise in capturing the MP of a user, the system still required user input on some contexts where generated readings proved to create ambiguous situations for the system, requiring manual intervention by the user.

2.3.8 Predicting the Location of Mobile Users: A Machine Learning Approach [26]
Focuses on analyzing the validity of applying ML mechanisms to the detection and prediction of the location of mobile users. Showing promising results in the ability to correctly determine the location of users, when using a large enough pool of information and the correct context (such as the temporal context, or activity context), enabling the various ML mechanisms can be further enhanced.

3 ENVIRONMENT
To ensure that the utilized architecture can accomplish the goals of this work, the typical environment and interaction a user has with a PTS must first be analyzed. As seen in Figure 1 the environments where the proposed solution will be employed are classified as either:

Outside Environment: An abstract environment that corresponds to all possible environments not associated with a transport network (such as a home, or work space).
Transport Environment: An environment that encapsulates the domain of the targeted PTSN.

This environment can be scoped as: the city of Lisbon, with all stops and vehicles maintained by a given Operator.

The Transport Environment can further be classified to have discrete locations that help restrain the scope that the proposed architecture must have:

Entry Point: These are the locations where the user enters a target transport. These locations can require the user to perform an action that validates their use of the service as legitimate, often by some form of ticketing practices.
Exit Point: These are the locations where the user exits a target transport. These locations can require the user to perform an action that validates the use of the service as having been legitimate.
In Transit: This is a general classification of the location a user is in while utilizing a transport, and describes the spatial dislocation of the user while making use of the PTSN. It can be further sub-classified as either being in motion or stopped (such as awaiting for traffic to move).

Users may in themselves have many states representing their current activity in a PTSN, in effect representing the context of the user in a PTSN, these contexts might be of greater relevance the more detail specific they are, however in an effort to keep some degree of simplicity, the developed work will focus only on the following contexts of user activity:

a) Walking - Considered to be when the user is in the Outside Environment or Transport Environment, moving on foot to a given destination.
b) Awaiting transport - Considered to be when a user is awaiting at a point of entry.
for a transport.
c) In target transport - When the user is considered to be in transit (and utilizing a transport of the targeted operator).
d) In transport - When the user is considered to be in some form of use of a vehicle while on the Outside Environment.
e) Other - Considered to be any possible activity either not targeted by the solution, while on the outside environment, or situations where proper context attribution is ambiguous.

4 USE CASE
As mentioned in Section [3] it is also vital to understand how a user will interact with the targeted PTSN, as can be exemplified in Figure 2 where the various User Contexts (UC) are visible:
- UC1: Represents any context where the user is interacting with the outside environment (either walking or using another transport, corresponding with points a), d) and e)).
- UC2: Represents the context where a user is awaiting for a transport in an entry point (such as awaiting for a bus at a bus station, corresponding with point b)).
- UC3: Represents the context where a user is utilizing a target transport, effectively going from a given point A to B (such as when the user is on a bus, corresponding with point c)).
- UC4: Represents the context where a user might pass near an entry/exit point but not enter or exit the transport, these might happen multiple times or not at all during a journey.
- UC5: Represents the context where a user will exit the transport, for some transportation systems this case might not be verified.

5 ARCHITECTURE
The proposed architecture for the mobile application follows a model that facilitates the collection of a user’s MP. It must be noted that the data acquired on the mobile application will be stored the device’s local database that allows further analysis, but it is not in the scope of this work to delve into how this is realized, or on how to create a data aware application of the collected MPs. As such the architecture is divided into various components (described below) that are in constant communication with each other, as detailed in Figure 3 were a general overview of the system is given.

5.1 Data Annotation
Data Annotation is realized by a service running constantly on the host device. To accomplish this, the system uses a conjunction of modules further denominated as Mobility Pattern Module (MPM). These modules will generate one or more values to be used by the MP Determination service in order to process how the user is currently traveling (bus, on foot, boat, train), further designated as User Activity Context (UAC).
This module utilizes the presence of free WiFi-AP offered by some PTSNs to help determine if the user is indeed inside the transport or merely near it.

5.1.2 Location Module
The coordinates obtained from GPS will be the main driving force in understanding the current location of a user, as well as their speed.

5.1.3 DR Module
DR by itself can detect when a user is walking, approximately of how far he has walked, and the general direction of movement based on a geomagnetic reference point. These features of DR are exploited by understanding that a UAC can generally be understood as (but not limited to) standing, walking, sitting inside a transport, or any other.

5.2 MP Determination
In order to minimize the required time of development, the use of Machine Learning (ML) means is applied to generate the decision mechanism whose output is the calculated UAC, and inputs are the various MPM outputs (as seen in Figure 4).

In order to accomplish this, a large data sample must be collected and adequately annotated, so that the employed ML mechanism can properly attempt to find relations in the data, in the form of a decision tree as seen in Figure 5.

5.3 MP Visualization
To help a user understand both his current and past MP, the system provides a screen dedicated to displaying this information over a geographic map. This displayed information will try to help the user on the effort of better understanding how he is utilizing a given PTSN.

5.4 Manager
The Manager serves as the orchestrator of the system, being the responsible entity for starting and terminating each MPM, as well as initializing the various components and their dependencies. It also serves as the entry and exit point to be utilized by any application that makes use of the developed system.

6 IMPLEMENTATION
For the implementation of each of the components as well as the prototype application Personal Assistant of a Transportation Environment Reflecting Natural-use (PATTERN), what follows is an overview of the implemented components and modules.

6.1 Data Gathering
In order to effectively collect data so that the MP determination module can be properly constructed by the chosen ML mechanism, each of the proposed MPM is provided with the
extra capability to collect data without feeding the Determination component of PATTERN (as seen in Figure 6). In this form, every MPM is capable of operating in:

**Data Collection Mode:** In this mode, every MPM will record its respective data in a local SQL-Lite database, while the application tags said data with reports from the user via the Data Processing Module, ensuring a true representation of the current UAC is captured and its MPM context, information that will in turn be fed to WeKa [34] to acquire the decision tree.

**Data Analysis Mode:** In this mode, only one MPM will record collected data into the local database (the location MPM), while other MPM will use a shared holder structure designated as the Accumulator (detailed in Section 7).

### 6.2 Mobility Pattern Module (MPM)

For each of the implemented MPM, the following outputs are generated:

**Activity Module:** The current user activity, represented as: Being on foot, being on a vehicle, being still, or unknown.

**Location Module:** The current location of the user (geographic coordinates), if the user is near a station, and if the user is inside the operating area of the target PTSN.

**WiFi Module:** If the user is currently near a target Vehicle equipped with at least one WiFi-AP.

**BT Module:** If the user is currently near a target vehicle or a target station with at least one BT beacon.

### 7 Accumulator

The accumulator functions with much of the same intent as the database previously described, by holding the outputs of the various MPMs. The use of the accumulator over the local database is to provide a unique access point where the Data Analysis task can query for information to feed into the decision tree, without the need to manage the acquisition of data directly from the MPM, or the access time to a database. The use of a volatile structure also ensures that the captured information is discarded when the application terminates.

### 8 Data Determination

To implement the Data Determination component of the prototype, an android asynchronous task is utilized to query the accumulator, and feed the available data to the decision tree, whose output is stored in the local database, as well as visualized by the Visualization Component, if one is currently active (as seen in Figure 7).

### 9 Data Visualization

In order to visualize current MP, this component utilizes the record of locations offered by
the Location MPM, as well as the output of the Data Determination component to create a line of locations color coded according to the detected UAC.

9.1 Challenges
During the development of this work, two major challenges were faced:

**Decision Tree**: the output generated by WeKa did not perform as expected during tests as the data set utilized was not sufficiently large to handle all corner cases generated by the various MPM outputs. However as the generated tree was a good template, this challenge was surpassed by simply modifying key aspects of the tree that did not make sense.

**GPS Signal Fluctuations**: Due to the natural interference present in the urban environment where the development took place, acquired coordinates from the GPS MPM sometimes had a large and sudden variance in location. To handle this problem a filter was employed that ignored sudden changes in location that would require unrealistic speed and acceleration by the user.

9.2 Tests
In order to accomplish the proposed targets of battery autonomy, accuracy and fidelity, the prototype was tested in normal use case scenarios (such as testing while a user is on a target bus, or on daily activities), and on a module basis (testing each module capability to generate results that reflect the real UAC of the user). In particular the tests took place in Lisbon utilizing both the intended operator (Carris) and various other transportation methods inside the same PTSN, with the prototype application running in a OnePlus 3T Android device.

As shown in Figure 8, where the slope of each plot indicate, the degree by which battery autonomy is affected, PATTERN should be run with location services set to a low power policy in order to reduce the impact on the battery of the host device.

Further more during tests, a record was kept of how often a new UAC was reported correctly and incorrectly during a journey, with the prototype being able to correctly indicate the correct UAC 93.42% of the time with a standard deviation of 0.077%.

These results allow for the affirmation that the prototype does indeed prove capable of correctly calculating the current UAC, as well as track the location of the user with a high degree of accuracy.

10 Conclusion
The developed prototype has proven to be able to have a high degree of accuracy in acquiring the current User Activity Contexts, while also observing a low cost in power when considering the location policy in use by the host device, proving that it is possible to create a system that does not require the addition of new infrastructure by the operator in order to effectively track user Mobility Pattern. However due to the lack of a larger data sample, the goal of developing a system capable of determining a correct Mobility Pattern with help from Machine Learning Techniques proved to be ineffective, while still showcasing the viability of applying Machine Learning to the solution as the incorrect output tree was still a good basis to work out the correct decision tree.

Acknowledgments
The author would like to complement for the participation and help in developing this work,
Ana Máximo, Miguel Guerra, Pedro Arvela, André Aparicio, as well as Engenheiro João Almeida for his input during development of the work, Professor Miguel Pardal for giving special insight into possible improvements to the initial concept of the work, and Professor Alberto Cunha for helping guide and develop this work.

REFERENCES


