Context-aware Vehicle Route Recommendation Platform: Exploring Open and Crowdsourced Data

Frances A. Santos[1,4], Diego O. Rodrigues[1], Thiago H. Silva[2],
Antonio A. F. Loureiro[3], Richard W. Pazzi[4] and Leandro A. Villas[1]
[1]Institute of Computing, University of Campinas, Campinas, SP – Brazil
[2]Department of Informatics, Federal University of Technology of Paraná, Curitiba, PR – Brazil
[3]Computer Science Department, Federal University of Minas Gerais, Belo Horizonte, MG – Brazil
[4]Faculty of Business and Information Technology, University of Ontario Institute of Technology, Oshawa, ON – Canada

***Abstract*—An increasing number of users have been adopting route recommendation systems, mostly motivated by the convenience that those systems bring to their traffic experiences. Usually, those systems observe the historical and current traffic conditions in order to evaluate and recommend the fastest routes. However, besides mobility aspects, more contextual information such as unplanned street events and neighborhood safety, are not taken into account in the recommendation process. With this in mind, we propose a platform to support context-aware route recommendation systems. The proposed platform aims to improve existing recommendation algorithms or enable the proposal of new ones. To assess it, we use datasets of routes suggested by Google Maps in the city of Curitiba, Brazil, official open data provided by the city and also data generated voluntarily by citizens in a participatory sensing fashion. Our results show the existence of an opportunity for route planners to provide personalized services to users, which is an important step towards the development of context-aware vehicular networks. Besides, these results illustrate how publicly available big data can be explored to improve context-aware route recommendations.**

I. Introduction

An increasing number of drivers have been adopting route recommendation systems. One reason for this trend is the ability to check possible routes for faster travel time taking into account traffic jam problems, which are getting worse in major cities [1]. In general, route planners aim to allow users to find the best route for traveling between an origin and a destination in the city. However, the concept of *best* route is very subjective; the best route for a certain user may not be the best for another user.

Vehicular route recommendation systems, such as Waze and Google Maps, commonly consider historical and current traffic conditions for recommending faster routes. However, other context information, usually harder to gather and interpret using traditional sensors, such as unexpected demonstrations and the safety level of areas, may not be taken into account in the recommendation. For example, recently a couple followed a route recommended by Waze was shot while entering a dangerous area in Rio de Janeiro, Brazil [2]. News like this one is not uncommon. Such episodes show the importance of considering more context data associated with routes beyond mobility-related context, helping users to receive personalized recommendation considering their preferences and needs.

A vehicular network provides communication between different vehicles or between vehicles and road-side units aiming to provide efficient and secure transportation depending on the driving context, i.e., any information that describes the driving situation [3]. With the advancement of information and communications technology, vehicular networks reached a new phase, being able to offer context-aware applications to drivers [4]. The idea of context-aware vehicular networks is to explore driving context to adapt and personalize various decisions. For instance, to help with traffic management, avoid collisions and offer personalized route recommendations [5].

In this paper, we propose a platform that aims to assist the process of recommending context-aware vehicular routes. The platform is composed of two key components: (i) identification of contextual areas, i.e., geographical areas of the city that stands out regarding a certain type of context; and (ii) identification of patterns of routes frequently used by people in the real world, instead of inferring better routes based on the shortest path. We discuss these components and present feasible approaches for their implementations. The platform allows the improvement of existing route recommendation algorithms and also enables the proposal of new ones.

We have collected a dataset of routes suggested by Google Maps in the city of Curitiba, Brazil, to identify frequent patterns on these routes in order to evaluate and illustrate an instance of the proposed platform. In addition, insecure areas were identified in Curitiba using official open data provided by the city and also data generated voluntarily by citizens in a participatory sensing fashion. We observe evidence that security issues are not being considered in route recommendation. With that, we verify exploring a real scenario that more alternative secure routes could be recommended without adding excessive delays. It is noteworthy observing that our results indicate an opportunity for route planners to provide personalized services to users, which is an important step towards the development of context-aware vehicular networks. Besides, we show how publicly available big data, including crowdsourced data and open data shared by cities, can be explored to improve context-aware route recommendations.

The remainder of this paper is organized as follows. Section

II presents the related work. Section III describes the proposed platform. Section IV shows the analyzed databases. Section V investigates the use of the platform in a real scenario. Finally, Section VI concludes the paper.

II. Related Work

A common goal of context-aware applications for smart cities is to improve the quality of life of its citizens. Among many examples of context-aware application, we can mention security [6], [7] and traffic planning/management [1], [8]. In the domain of a vehicular network, a context-aware application is able to extract contextual information about the city thanks to several solutions available, such as vehicle-to-sensor and vehicle-to-Internet [9]. This enables real-time traffic monitoring, traffic sign warnings and speed and distance estimation providing drivers the ability to make better decisions while driving.

Regarding traffic planning, a topic that has received a lot of attention from researchers is the recommendation of routes for vehicles or pedestrians [1]. To this end, several approaches are proposed to obtain the "best" route, such as [10], [11], where one of the main assumptions adopted is that the highest priority of the user is to save time on his/her trip. However, this is not always the case, making the problem of finding the best route very subjective, and in some cases going way beyond finding the fastest route. This is because the fastest route could have problems that might affect certain users, such as permeate insecure or heavily polluted areas, among other problems, which result in unpleasant experiences.

In this way, route recommendation systems can have additional goals according to the particular interests of users. In this direction, CrowdSafe [6], for example, is a mobile device system that allows users to report information about crimes they have suffered or witnessed. The crime information is used by the system to identify areas of insecurity in the city, allowing it to provide the recommendation of a more secure route, and also to generate crime statistics of the city. However, the mapping of insecurity on CrowdSafe does not differ among possible categories of crime (e.g., theft, burglary attempts, assault, homicide) and consequently, it may not reflect the security level required.

In the same line, the approach proposed by [12] considers people's perceptions of routes in the city for recommending more pleasant routes for pedestrians. In that approach, people shall actively participate in all stages of the recommendation system and, therefore, keeping users motivated is an essential and critical part of the approach [13]. In [7], the mobility of agents (e.g., vehicles) is modeled as a particle system, where information from different city contexts such as traffic, pollution, crime, and street events represent forces exerted on the particles. In that model, agents are particles that must move between origin-destination locations in a 2D space, which represents a geographical area of interest. The resultant of the attraction/repulsion forces must guide the particle across space. As the particles represent the individuals' movement through a free space, i.e., without barriers, the model does not consider the limitations of the street and avenues network existing in cities.

A context-aware route recommendation protocol is proposed in [4], for reducing or avoiding traffic jams. It also considers critical services in order to provide the desired path towards each user's destination. For this, vehicles and road-side units are responsible for gathering, processing and sharing the knowledge about their surrounding road segments, which enables to recommend routes aware of the traffic conditions. For the proper functioning of the proposed protocol, the authors assume the existence of a road-side unit at each road intersection, an assumption that may be considered unattractive from a financial point of view.

Our study differs from those discussed above since we propose a generic platform to assist the development of context-aware vehicular networks. The platform enables the identification of patterns of routes frequently used by people in the real world, instead of inferring better routes based on the shortest path. Likewise [7], the proposed platform allows the identification of contextual areas in urban data sources. However, our platform considers the aggregated analysis of urban data, from different publicly available sources, to minimize errors and to enrich the quality of the information obtained.

III. Platform to Assist Route Recommendation in Context-Aware Vehicular Networks

A. Overview

The wide availability of different data sources on various aspects of the city facilitates the process of understanding and solving problems faced in urban centers, and thus offers more sophisticated services focused on improving the lives of users in the city. In this sense, we have studied an approach to ease the development of context-aware vehicular routes recommender systems.

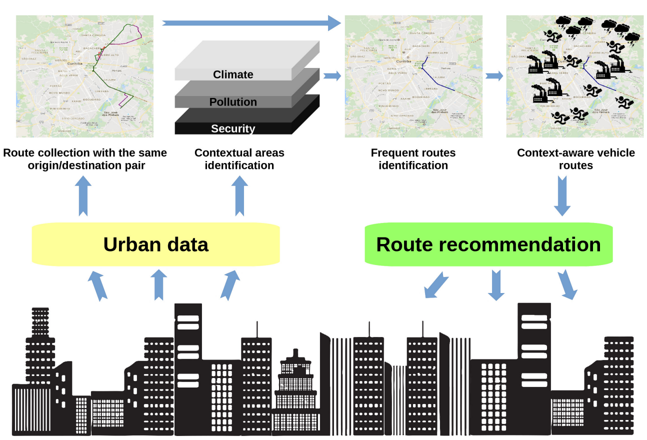


Fig. 1. Overview of the proposed platform.

In order to achieve this goal, the proposed platform (Figure 1) receives as input a set of routes and data about one or more aspects of the city, i.e., contextual information of interest. After collecting the data, which can be done in several manners [14], they are processed to obtain contextual areas.

For that, it is necessary to have a set containing a significant number of routes taken by users. Let $U = \{u_1, u_2, \dots, u_n\}$, be a set of users, $R = \{r_1, r_2, \dots, r_m\}$, a set of routes and $f_i \in [0, 1]$ is the frequency of some route $r_i \in R$. Each route $r_i \in R$ is traveled by an user $u_j \in U$, which can travel one or more routes. Given R, we have to identify the routes we desire to study their context, which depends on the intended recommendation to be performed. For that, we must answer the following questions:

- Will the route recommendation be made for a specific user u_j? If so, we shall filter all routes in R that can be used by u_j. The same idea goes for any subset of U. Otherwise, we consider all routes without applying any filter;
- What is a suitable frequency f_i for a given route r_i? The frequency value depends on the context under study. For example, driving eventually through routes that permeate areas with high levels of air pollution might be considered more acceptable than in insecure areas where the lives of people might be at serious risk. Thus, the frequency value for air pollution context should be higher than the security one, because traveling through polluted areas is more acceptable than through insecure ones. More details regarding this aspect are discussed in Section III-B.

Note that we can have multiple data sources about different urban aspects and, thus, we have the opportunity to enrich routes with diverse contexts.

In this sense, it is possible to consider different context simultaneously on various aspects of cities, such as security, pollution, and climate. In this way, enabling the recommendation of routes according to the requirements of users. In this study, we demonstrate an approach that considers contextual areas identified for a region of the city. In this case, areas of different contexts may be overlapped in the same geographical space. The proposed platform is generic enough to allow the use of other approaches, such as the one used in [15], [16], which proposes models to group different urban data sources. However, a cost-benefit assessment of each approach is beyond the scope of this study.

Our focus is the identification of contextual areas in distinct urban data sources, $\mathcal{F} = \{F_1, F_2, \dots, F_n\}$, where each $F_i \in \mathcal{F}$ is a dataset about an urban aspect. These data are comprised of a pair: geographical location d and the time t of its observation. For example, a dataset about crime has the address (d) and time (t) of the crime occurrence. With the $\mathcal{F}$ set, our goal is to identify contextual areas that better represent the coverage area a of data in the F_i element, instead of considering only the specific point represented by d.

The aforementioned coverage depends on the target context under study. For example, let F_i be a dataset about crime and F_j a dataset about air pollution, it is reasonable to consider a smaller size a associated with F_i than the size associated with F_j. The occurrence of a crime in d does not only indicate that the point d may be insecure but at least few blocks surrounding d. Whereas the high level of pollution sensed in d may indicate that a larger area is affected, for instance, a neighborhood sized area. Section III-C is dedicated to discuss a strategy to identify contextual areas.

Figure 1 also illustrates the possibility to generate context-aware routes by combining frequent routes and one or more contextual areas. Particularly, routes aware of a certain context could be used by route recommendation algorithms. For instance, users can explicitly request routes to avoid insecure areas, such as areas with high level of crime or routes that may compromise their security in adverse weather conditions. Another possibility is to offer an automatic recommendation service, i.e., without the user's explicit involvement. To that end, it is necessary to analyze typical routes used by a given user to be able to provide information that help him/her to make better decisions.

B. Frequent Routes Identification

Drivers can navigate existing city routes on their own, or use some route recommendation system to help with their daily journeys. In both cases, there may be routes that are regularly used, since traffic routine is not unusual, as noted by Karnadi et al. [17]. Finding regular patterns is particularly interesting to the proposed platform because it enables the study and recommendation of context-aware routes.

Let a given user u_j who moves daily from his/her home to work and vice versa, and let R be a set of routes. We can filter R to obtain a set of routes, $R_i \subseteq R$, used by u_j to travel that itinerary (i.e., home-work). Let r_k be the most frequent route in R_i where $f, 0 \leq f \leq 1$, is the frequency that r_k occurs in R_i. Depending on the users context of interest and the value of f, an alternative route could be recommended for u_j. For instance, suppose that u_j is interested in avoiding areas with high level of air pollution, r_k has frequency $f \geq 0.85$ (hypothetically high bound) and r_k penetrates several polluted areas (identified using the approach discussed in Section III-C).

Hence, the recommendation system is able to provide for u_j a new route $r_y \in R$ to avoid polluted areas, even if r_y is longer than r_k, and, consequently, it increases travel time. In case u_j has no interest in avoiding polluted areas, even so the recommendation system could still alert the user to the risk of serious diseases contraction. However, whether $f \leq 0.15$ (hypothetical low bound) and r_k is significantly lower than r_y, then u_j might not be favorable to take r_y. Similarly, we can consider routes of all users or any subset of U. If we consider a set of routes R used by all users, we can find a subset of routes $R_p \subseteq R$, where $\forall r_x \in R_p$, r_x is a route often frequented by all users. In this way, if we consider the contextual areas that r_x penetrates, it is possible to provide recommendations as discussed above.

To identify the most frequent routes, we represent each route by a connected digraph $D(V, A)$, where a vertex $v_i \in V$ is a specific point of a route (e.g., a street corner) and an arc $a_{i,j} \in A$ represents a given street between vertices v_i and

v_j. Digraph D has at least two vertices, where one represents the origin (source vertex) and the other one represents the destination (sink vertex) of a route, which are the only ones that have indegree and outdegree 0 in D, respectively. If R contains n routes, then we denote by $\mathcal{D} = \{D_1, D_2, \ldots, D_n\}$ the collection of n digraphs D that represent the n routes $r \in R$. The collection $\mathcal{D}$ is used to identify the collection of maximal subsets $\mathcal{Q} = \{Q_1, Q_2, \ldots, Q_m\}$, where each $Q_i \in \mathcal{Q}, 1 \leq i \leq m$, is a maximal subset that occurs with frequency f in $\mathcal{D}$, and Q_i represents a path, i.e., an entire route or a smaller part, which is common to f routes in R.

To find every maximal subset Q_i through isomorphism test is known to be $\mathcal{NP}$–hard problem [18]. A possible way to get around this problem is through the algorithm called *graph-based substructure pattern mining* (gSpan), which minimizes this problem with its adopted strategy [19]. gSpan expects as initial parameters a collection of graphs $\mathcal{D}$ and a minimum value of frequency f that a given subset must match with other subsets. After determined those parameters, the algorithm returns a list of subsets $\mathcal{Q}$.

C. Contextual Areas Identification

In this section, we discuss more details about another key component of our platform: contextual area identification. Urban data, when analyzed individually, may not reflect precisely the current status of a specific context. This is because data may be outdated, incorrect, or the granularity is insufficient to represent a contextual area.

To illustrate a problem related to the location of the event, we discuss what is observed for crime data. Official data about a crime may not guarantee the crime location precision, by geolocation or complete address. Typically, crime location corresponds to a city spot, such as a street name and district, which is the case of open data provided by the city of Curitiba, Brazil (available in https://goo.gl/RJv7P6). Alternative sources of crime data, such as www.ondefuiroubado.com.br, enables the citizens to voluntarily report crime occurrences with the help of geolocation. However, users may incorrectly inform the location where they have witnessed a crime, since they may, for instance, not remember the exact location of the occurrence when reporting the crime. Therefore, perform aggregated analysis of urban data is important to reduce errors and increase the quality of information.

To this end, a potential approach for contextual area identification is clustering, which is an unsupervised organization of a dataset in groups according to some similarity measure [20]. Thus, when clustering a dataset F_i, m disjoint subsets $C_i^1, C_i^2, \ldots, C_i^m, m \geq 0$, are identified, called clusters. Data within a cluster has higher similarity among themselves than data in different clusters [20]. Here, the similarity measure adopted is geographic distance obtained by the Haversine formula [21].

Among the data clustering techniques, the algorithm Density-Based Spatial Clustering of Application with Noise (DBSCAN [22]) is especially relevant, because it can create clusters with different formats and sizes, and, it also achieves satisfactory results even in the presence of noises. For this reason, DBSCAN is adopted in this study. DBSCAN is based on neighborhood density to identify clusters, where points present in clusters with low density are called *outliers*. Each cluster C consists of a set of points, whose minimum amount of points is η and they are in a neighborhood with radius ε. The parameters η and ε shall be provided to the input of DBSCAN algorithm. Let $p_i \in C$ be a point, which may be a *core* point, whether p_i is at maximum distance ε from at least $\eta - 1$ points. Otherwise, p_i is a *border* point of a cluster with a core point $p_j \in C$, satisfying the first condition. Outliers are not core, nor border points and are discarded by the algorithm. This characteristic is essential to identify and discard noises.

Note that the need to adjust the η and ε parameters of DBSCAN is essential when dealing with multiple contexts. Taking two different contexts, it is likely they have distinct properties. For instance, suppose F_i represents crime context and F_j represents climate context. Thus, a cluster C_k obtained from F_i represents an insecure area in the city, whereas C_y obtained from F_j represents weather conditions from an area inside the city. Probably, the radius ε' of a cluster C_y will be defined greater than radius ε'' of cluster C_k, because an insecure area might be typically set with a granularity of a district, whereas weather conditions might be referring to an area that covers the entire city.

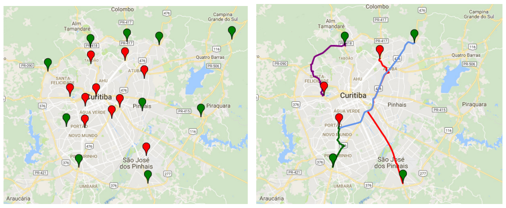
In addition, the cluster's "validity" can also be different among them according to the context. Indeed, data about the weather are valuable during a short period of time (e.g., hours or days) and, therefore, they should be updated regularly. In contrast, data about crimes are valid for a longer period (e.g., weeks or months), because an insecure area is not likely to become secure in a few days. Therefore, for each cluster C should be defined a time window $\mathcal{T} = [t_{\min}, t_{\max}]$, where any point $p \in C$ represents a location d at time instant t, such that $t_{\min} \leq t \leq t_{\max}$.

IV. Datasets Studied

In this section, we describe the datasets used in our study. Section IV-A presents the set of routes suggested by Google Maps. Section IV-B presents the sets of unofficial and official data about crime that have occurred in Curitiba, Brazil.

A. Routes Suggested by Google Maps

A route recommendation system is a powerful tool used by people for helping them navigate within a city or between cities. Usually, people inform their origin, which may be obtained automatically, a destination, and the navigation mode (e.g., drive, walking, or public transport), and then the system suggests a route based on the provided information. Google Maps and Waze are two well-known route recommendation systems, which are commonly used by people due to their ability to suggest "best" routes in a particular traffic condition. Such systems consider information about historical and real-time conditions of traffic to estimate the best route at a particular time.



(a) Origin (green) and destination (red) points. (b) Examples of suggested routes.

Fig. 2. Routes suggested by Google Maps in Curitiba.

In this work, we consider the city of Curitiba, Brazil, as our scenario to find patterns on route recommendations from Google Maps. For that, we defined two distinct sets of addresses to represent origin and destination points, where each set has ten distinct points, considering the movement from users departing from all regions towards downtown. Figure 2(a) shows a map of Curitiba, with origin and destination points displayed in green and red, respectively. With both sets defined, we built an origin/destination matrix denoted by $M_{10\times10}$, which is used to request routes recommendation of each origin to all destinations from Google Maps Directions API[1], exploring the navigation mode "Drive". Each iteration obtains 100 suggested routes, which contain meta-data and traffic conditions along the route.

In total, we conducted 420 iterations during 60 days, with 7 hours apart between consecutive iterations. Thus, the dataset of routes of Curitiba has 42,000 routes, collected at different periods of the day (i.e., morning, afternoon, evening and night), containing detailed instructions to follow each route, the estimated time of arrival, and total distance. Figure 2(b) shows four examples of routes of our collected dataset, where routes are displayed in blue, red, green and purple.

B. *Crime Reports*

Regarding public security context, we collected two crime datasets in Curitiba: one from reports made by users of the system www.ondefuiroubado.com.br (data obtained in a crowdsourcing fashion), and another provided by the City Guard (official open data). In the following, we present a brief description of the datasets.

1) Crowdsourced Data: The website www.ondefuiroubado.com.br enables users to contribute voluntarily with data about crimes suffered/witnessed by them. In this way, the system is able to map insecure regions of the cities helping users to protect themselves. Crimes that have occurred in Brazilian cities within the past four months from the current date are available on the website. The main attributes regarding each crime report are: address (street, district, city, state, country, and zip code), geolocation (latitude and longitude), text of the crime description, date and time of occurrence, crime nature, estimated loss, and whether the user reported the crime for the police. During the collection period, from July to October 2016, users reported 331 crimes in Curitiba, where 36.8% were robbery cases and 26.5% were theft cases. From all occurrences, around 39% of them were not recorded officially. There are several reasons which could help to explain this fact. One of them is the convenience provided by a Web system compared to an official channel, where, in Brazil, you have to go in person to perform the report in some specific places. This is a clear hint on the relevance of collective collaboration of users to enhance knowledge concerning the city security.

2) Official Data: The official dataset contains crime occurrences registered by Curitiba's City Guard since 2009, and it is updated monthly, where the last update considered in this work occurred in November 1st, 2016. The dataset has 18,235 crime records in the metropolitan region of Curitiba in 2016. Some of the main attributes regarding each crime report are: address of the occurrence, crime nature (e.g., robbery), subcategory for the crime (e.g., car robbery), and date and time of occurrence. The official dataset has crime records of several types that happened for a long period, however, we consider only a subset of them that happened at the same period of the unofficial (crowdsourced) dataset, and refers to a type of crime that exists in the unofficial dataset. This filter resulted in 1,313 official crime occurrences.

V. Experimental Analysis

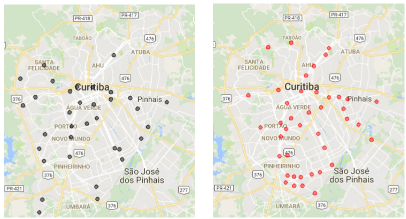
As proof of concept of our proposed platform, in this section we consider the datasets of Section IV to analyze the recommended routes regarding security.

A. *Insecure Areas Identification*

To identify insecure areas in Curitiba, which are the contextual area considered, we apply the proposed methodology in our platform (discussed in Section III-C). We represent every crime occurrence that occurred in a location d at instant t by o. This means that for the same locality d, we can have several occurrences o. In this way, a cluster $C = \{o_1, o_2, \ldots, o_n\}$ contains n crime occurrences and a coverage a corresponding to an insecure area in Curitiba. The size of a can vary according to the density of the points in a specific area, ranging from some blocks to one or more neighborhood-size area. We consider a a circular area, whose center of a is a centroid (any *core* point) of C and the radius of a is used by the clustering algorithm (DBSCAN) to identify C, i.e., ε is the radius of a. By doing that, we tend to minimize eventual inaccurate information provided by users regarding the occurrence of crimes.

We define the parameters $\varepsilon = 250\,$meters and $\eta = 5$. We experimented with other values, but the results with those parameters represent more accurate insecure areas in the studied city. In practice, those parameters may change according to the data available and studied areas. Additionally, we consider a time window of four months ($\mathcal{T} = [$*July/2016, October/2016*$]$), which is the same validity period used by www.ondefuiroubado.com.br. Then, datasets about crimes are separated according to the degree of severity

[1]More details about Google Maps Directions API can be found in: https://goo.gl/kiXllw.



(a) Moderate crimes (b) Serious crimes

Fig. 3. Areas threatened by moderate and serious crimes in Curitiba.

of the crime. Two classes were created to differentiate them: (i) moderate crimes (e.g., theft and car burglary) involving a total of 611 occurrences; and (ii) serious crimes (e.g., firearm assault and kidnapping) with 1033 occurrences recorded.

Figure 3 shows Curitiba's insecure areas, where clusters of moderate crimes are displayed in black areas in Figure 3(a), and cluster of serious crimes are displayed in red areas in Figure 3(b). Overall, we identified 75 insecure areas, where 34 of them refer to moderate crime clusters and 41 to serious crime clusters, and there are overlaps in some cases.

B. Security of Most Frequent Routes

In this section, we identify paths (i.e., maximal subgraphs) of whole routes that are regularly recommended by Google Maps, where we desire to verify whether these paths intersect insecure areas. For this, we consider the collected routes R_{ij} with the same origin/destination pair $i, j : 1 \leq i, j \leq 10$, as described in Section IV-A. Next, we apply the approach described in Section III-B to identify frequent routes. We define the frequency $f = 0.85$, which means that Q_{ij} is contained in at least 85% of the routes R_{ij}. Thus, if Q_{ij} penetrates insecure areas in Curitiba, then users who request a route with origin i and destination j, are likely to pass through insecure areas.

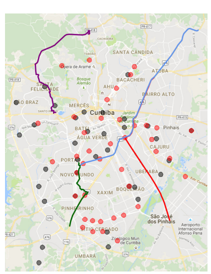


Fig. 4. Frequent path contained in routes recommended by Google Maps, which permeate insecure areas in Curitiba.

Figure 4 shows Curitiba's map and some illustrative examples of frequent paths (for $f = 0.85$), which are contained in routes shown in Figure 2(b) and available in R_{ij}. In Figure 4, we can also observe insecure areas identified with our approach. As we can observe, all paths displayed in Figure 4 intersect one or more insecure areas. It is worth mentioning that not all suggested routes pass through insecure areas.

C. Alternative routes

Our results suggest that Google Maps are not considering security aspects in the recommendation process. Even so, it is one of the most popular navigation systems due to its capacity to provide faster routes, taking account mainly the mobility aspects. In that sense, our platform offers an additional possibility to the recommendation systems like Google Maps to explore context information about various aspects of cities, in order to recommend the best routes to, potentially, improve the user experience.

Regarding security, the best route should be the one that avoids insecure areas as much as possible, without significantly increasing the distance and travel time. Thus, for every suggested route, for instance by Google Maps, there are at least two possible cases. The route does not permeate insecure areas, in this case, it must be attributed to the user without further modification. Otherwise, the route intercepts insecure areas, and under this circumstance, the context-aware recommendation system should provide an alternative route. Moreover, some users may be willing to take the risk of following a route that permeates insecure areas, where the system should leave them to decide the security levels of their routes.

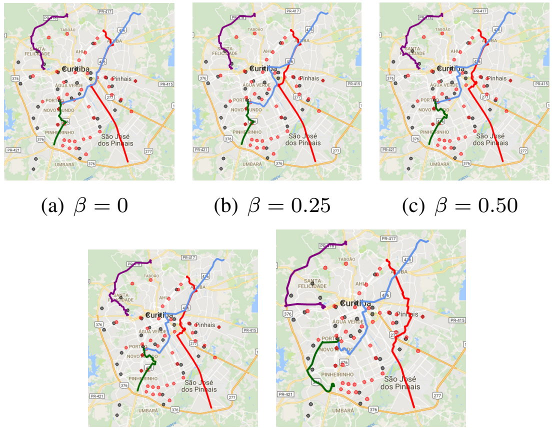
Figure 5 shows alternative routes with distinct security levels $\beta, 0 \leq \beta \leq 1$. All of them were calculated using Google Maps API considering the same departure time, where our system calculated waypoints to avoid insecure areas according to β. The alternative routes have the same origin/destination points as shown in Figure 2(b), and they are denoted by r_1 (purple route), r_2 (green route), r_3 (blue route), and r_4 (red route). When security context is not taken into account in the route recommendation, i.e. $\beta = 0$, we have the case discussed in Figure 5(a). By increasing $\beta, \beta \geq 0.25$, the routes suggested are presented in Figures 5(b)–(e). As we can see, the suggested routes become different of frequent paths found, especially when $\beta = 1$, because they are avoiding undesired contextual areas.

TABLE I

SECURITY LEVEL, TOTAL DISTANCE AND TRAVEL TIME.

	Total distances (km)				Travel time (min)			
β	r_1	r_2	r_3	r_4	r_1	r_2	r_3	r_4
0.00	15.18	8.46	21.83	25.40	23.15	14.91	36.08	34.43
0.25	15.18	8.46	25.45	25.40	23.15	14.91	47.46	34.43
0.50	16.98	10.28	24.60	26.48	24.78	20.08	48.23	38.93
0.75	16.98	10.28	25.52	26.48	24.78	20.08	50.31	38.93
1.00	23.10	14.09	27.60	25.44	21.78	22.03	53.81	47.80

Taking into account the metrics (*i*) total distance and (*ii*) travel time, we perform a comparison among routes r_1, r_2, r_3, r_4, considering the security level β in range $[0, 0.25, 0.50, 0.75, 1]$. Table I shows results for (*i*) and (*ii*) metrics. As we can see, safer routes do not always result in larger routes or significant delays in travel time. For instance, route r_3 is smaller with $\beta = 0.50$ than $\beta = 0.25$ and might demand almost the same travel time. Similarly, route r_4 is smaller with $\beta = 1$ than $\beta = 0.50$ and $\beta = 0.75$, and practically the same with $\beta = 0.25$ and $\beta = 0$, however, it might demand on average 11 minutes more to be traveled.



(a) $\beta = 0$ (b) $\beta = 0.25$ (c) $\beta = 0.50$

(d) $\beta = 0.75$ (e) $\beta = 1$

Fig. 5. Examples of routes in Curitiba according to different β.

In this particular case, choosing $\beta = 0.75$ only adds a delay of 4 minutes. Another example is considering route r_1 in the safest scenario, i.e., $\beta = 1$, because despite being the longest route it is also the fastest one. With our examples, we want to illustrate that current route recommendation mechanisms could be upgraded to meet the requirement of security, and any other context of interest, tasks that could be modeled and performed in a simpler fashion with our platform.

VI. Conclusion

In this study, we explored the use of heterogeneous urban data sources along with routes that are frequently used by people in their daily journeys, to provide key mechanisms towards the development of context-aware vehicular networks. To this end, we have proposed a platform to enrich the vehicular route recommendation process. By exploring an example based on a real scenario considering security aspects, we found evidence that current route recommendation systems could be improved even further, e.g., avoiding recommendation of frequent routes through insecure areas within a city. Our platform is generic enough to handle the same principle for other types of context individually or simultaneously, i.e., it offers flexibility to handle different context homogeneously. Moreover, our platform helps to ease some aspects of the task of handling large amounts of data, being helpful to improve existing route recommendation algorithms, or the proposal of new ones. As future work, we intend to evolve our platform to handle not only spatial data but also other types of large-scale heterogeneous data sources, such as preference data expressed in Foursquare-like applications and free-text messages shared in Twitter-like microblogging, in order to enable the identification of non-trivial information and their association with specific areas.

Acknowledgment

The authors would like to thank the grants Sao Paulo Research Foundation (FAPESP 2015/07538-1), CAPES (PDSE 88881.132016/2016-01), CNPq (grant 401802/2016-7) and CNPq-URBCOMP project (grant 403260/2016-7) by the financial support.

References

[1] H. Wang, G. Li, H. Hu, S. Chen, B. Shen, H. Wu, W.-S. Li, and K.-L. Tan, "R3: A real-time route recommendation system," in *Proc. of the ACM VLDB'14*. ACM, 2014, pp. 1549–1552.

[2] D. Phillips, *How directions on the Waze app led to death in Brazils favelas*, Washington Post, October 2015. [Online]. Available: https://goo.gl/QxHdKv

[3] S. Fuchs, S. Rass, B. Lamprecht, and K. Kyamakya, "Context-awareness and collaborative driving for intelligent vehicles and smart roads," in *Proc. of the Workshop on ITS for an Ubiquitous ROADS*, 2007, pp. 1–6.

[4] M. B. Younes and A. Boukerche, "A performance evaluation of a context-aware path recommendation protocol for vehicular ad-hoc networks," in *Proc. of the IEEE GLOBECOM'13*. IEEE, 2013, pp. 516–521.

[5] H. Hartenstein and K. Laberteaux, *VANET vehicular applications and inter-networking technologies*. John Wiley & Sons, 2009, vol. 1.

[6] S. Shah, F. Bao, C.-T. Lu, and I.-R. Chen, "Crowdsafe: Crowd sourcing of crime incidents and safe routing on mobile devices," in *Proc. of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. ACM, 2011, pp. 521–524.

[7] M. De Domenico, A. Lima, M. C. González, and A. Arenas, "Personalized routing for multitudes in smart cities," *EPJ Data Science*, vol. 4, no. 1, 2015.

[8] Y. Wang, J. Jiang, and T. Mu, "Context-aware and energy-driven route optimization for fully electric vehicles via crowdsourcing," *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 3, pp. 1331–1345, 2013.

[9] N. Lu, N. Cheng, N. Zhang, X. Shen, and J. W. Mark, "Connected vehicles: Solutions and challenges," *IEEE Internet of Things Journal*, vol. 1, no. 4, pp. 289–299, 2014.

[10] A. W. ter Mors, C. Witteveen, J. Zutt, and F. A. Kuipers, "Context-aware route planning," in *Proc. of the MATES 2010*. Springer, 2010, pp. 138–149.

[11] R. Bader, E. Neufeld, W. Woerndl, and V. Prinz, "Context-aware poi recommendations in an automotive scenario using multi-criteria decision making methods," in *Proc. of the CaRR'11*. ACM, 2011, pp. 23–30.

[12] D. Quercia, R. Schifanella, and L. M. Aiello, "The shortest path to happiness: Recommending beautiful, quiet, and happy routes in the city," in *Proc. of the 25th ACM Conference on Hypertext and Social Media*. ACM, 2014, pp. 116–125.

[13] F. A. Santos, T. H. Silva, T. Braun, A. A. F. Loureiro, and L. A. Villas, "Towards a sustainable people-centric sensing," in *Proc. of the IEEE ICC'17*. IEEE, 2017, pp. 1–6.

[14] T. Silva, C. Celes, J. Neto, V. Mota, F. Cunha, A. Ferreira, A. Ribeiro, P. Vaz de Melo, J. Almeida, and A. Loureiro, "Users in the urban sensing process: Challenges and research opportunities," in *Pervasive Computing: Next Generation Platforms for Intelligent Data Collection*. Academic Press, 2016, pp. 45–95.

[15] T. H. Silva, P. Vaz de Melo, J. Almeida, A. Viana, J. Salles, and A. Loureiro, "Participatory sensor networks as sensing layers," in *Proc. of the IEEE SocialCom'14*. IEEE, 2014.

[16] S. M. Oteafy, "A framework for heterogeneous sensing in big sensed data," in *Proc. of the IEEE GLOBECOM'16*. IEEE, 2016, pp. 1–6.

[17] F. K. Karnadi, Z. H. Mo, and K. c. Lan, "Rapid generation of realistic mobility models for vanet," in *Proc. of the IEEE WCNC'17*. IEEE, 2007, pp. 2506–2511.

[18] S. A. Cook, "The complexity of theorem-proving procedures," in *Proc. of the ACM Symposium on Theory of Computing*. ACM, 1971, pp. 151–158.

[19] X. Yan and J. Han, "gspan: Graph-based substructure pattern mining," in *Proc. of the IEEE ICDM'02*. IEEE, 2002, pp. 721–724.

[20] A. K. Jain, M. N. Murty, and P. J. Flynn, "Data clustering: A review," *ACM Comput. Surv.*, vol. 31, no. 3, pp. 264–323, 1999.

[21] R. W. Sinnott, "Virtues of the Haversine," *Sky and Telescope*, vol. 68, no. 2, pp. 159+, 1984.

[22] M. Ester, H.-P. Kriegel, J. Sander, X. Xu *et al.*, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *Proc. of the KDD*, vol. 96, no. 34, 1996, pp. 226–231.