Hybrid Context-Aware Multimodal Routing

Diego O. Rodrigues[1], João T. Fernandes[2], Marilia Curado[2], and Leandro A. Villas[1]
[1]*Institute of Computing, University of Campinas, Campinas, Brazil*
[2]*Department of Informatics Engineering, University of Coimbra, Coimbra, Portugal*
Email: diego@lrc.ic.unicamp.br, {joaonf, marilia}@dei.uc.pt and leandro@ic.unicamp.br

***Abstract*—Selecting the appropriate route in urban multimodal transport networks may require information from several sources, such as user preferences and contextual information. Such information could be used to choose different transport modes for a route, to avoid the use of hired private vehicles (HPV), like taxis, in certain segments of the trip with poor traffic conditions, for example. Therefore, this paper proposes the use of hybrid Hired Private Vehicles and Transit (HPV-transit) to enable a more personalized urban routing. User's impressions over trips' cost and duration are taken into account to select better hybrid routes. A taxi dataset was used to create graphs that map the main mobility flows in New York City. To create these graphs we propose a novel flow-based clustering technique which identifies trending mobility flows based on spatio-temporal datasets containing departures and arrivals. These flows are used to evaluate the performance of the hybrid HPV-transit trips with metrics such as cost, duration, and user experience. We compare the proposed solution with traditional transit or HPV only trips. It is possible to conclude that the adoption of HPV-transit trips can bring benefits in terms of costs and durations for urban mobility. The results obtained contribute with the quantification and comparison of the gain using hybrid routes.**

1. Introduction

Due to the fast development of urban areas, cities are facing a series of issues, many of them related to urban mobility, such as traffic congestion, and public transportation waiting times, and others. Therefore, cities are making efforts to improve their services by, for example, understanding people's behavior. This is accomplished by collecting data about city aspects [1], analyzing contextual urban data [2], [3], and building a knowledge base to aid governments [4]. The results found in these studies are critical for research purposes because they allow researchers to understand and simulate different situations, which can lead to changes in the city and its services.

During the past years, Google improved their software Google Maps[1] to provide a variety of functionalities, such as providing real-time traffic information or suggestion of urban routes selected by different criteria (regarding time or transporting mode). However, the route suggestion only focuses on specific transportation modes selected by users (e.g., car, foot or transit). Public transportation modes could be used to avoid poor traffic conditions while riding Hired Private Vehicles (HPV), such as Uber[2]. Such routes are being called in this study "hybrid HPV-transit" routes.

1. https://www.google.pt/maps/
2. http://www.uber.com/

We propose a method to improve route selection, by considering traffic information and using different transportation modes to produce a hybrid transit-HPV route. We identify the most relevant flows of a city using a novel flow-based clustering technique proposed in this paper. Then these flows are used to analyze the impact of hybrid routes (i.e., HPV and Transit) on avoiding traffic congestion and creating personalized routes. The results are aligned with common sense about using a combination of HPV and transit routes, although we provide an enhanced understanding of this knowledge by quantifying it and enabling a comparison with traditional routes using only transit or only HPV. The next sections are organized as follows: Section 2 describes different techniques related to our work regarding multimodal routing, clustering, and modeling; Section 3 presents our proposed method to select hybrid HPV-transit routes and how to evaluate it; Section 4 shows the results for the evaluation performed; Finally, Section 5 presents the conclusions.

2. Background

The task of selecting best routes in urban multimodal transport networks was explored for different fields of study, and several techniques were built to compute, select and evaluate multimodal routes. In this section, we briefly describe some of these techniques.

2.1. Multimodal Routing

Urban routing systems were developed to compute routes based on different information (such as user's input or traffic conditions). P. Campigotto et. al. [5] developed an algorithm that takes into account user's preferences. FAVOUR (FAVourite ROUte Recommendation) uses a Bayesian strategy to understand and estimate the best routes that the user prefers. The Bayesian strategy was used to understand the user's behavior and constant imprecision during decisions. D. Bucher et. al. [6] proposed a preset rule-based heuristic model defined as a set of rules for different transport modes, such as walk, bus, car and train.

Song et.al. [2] proposed a hierarchical routing algorithm to compute routes using real-time traffic data. The authors evaluated their solution by generating randomly 150 trips to be performed in Jinan, China. While the authors obtained good results in these trips, it is not clear how their solution would behave with real-world cases. Zhou et.al. [3] also proposed an algorithm to compute safe urban routes. The paper identified danger indexes for streets based on traffic accident data in New York City. The focus of their paper is to use safety contextual information to aid planning routes to address certain safety requirements (e.g., planning school bus routes). In our study, we identify and use the main mobility flows in New York City to explore a different contextual information about route selection and analyze traffic data to investigate positive impacts on using hybrid routes to avoid traffic congestions.

2.2. Clustering

Exploring different urban data allows the identification of patterns and trends in human behavior that could be explored to improve city services. To identify such patterns, one commonly used technique is clustering, which allows merging data samples according to similar characteristics. One clustering technique used to analyze similarity of spatial distribution is the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) [7]. This method allows discovering clusters with arbitrary shape and is suitable to find high-density regions. To use DBSCAN, two parameters need to be defined: (1) the spatial neighborhood in terms of radius – called *eps* – and (2) a threshold density of samples in the neighborhood to identify a cluster – called *min-samples*. R. Campello et. al. [8] presented an improvement to DBSCAN called Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN). The authors added a hierarchical tree structure that will find a suitable radius threshold for every cluster (removing the necessity of informing it). DBSCAN and HDBSCAN are well-adopted clustering algorithms to identify regions in position datasets. HDBSCAN is used in the present study to aid the identification of mobility flows in New York City.

H. Hamedmoghadam-Rafati et. al. [4] applied a coarse-graining method to analyze New York City taxi trips. The authors first filter the data to remove possible outliers and trips with wrong coordinates. Then, they split the data in a grid with 13161 zones, to analyze each zone and understand the traffic flow within the city. M. Momtazpour and N. Ramakrishnan [9] used a similar approach to divide taxi pickups and drop-offs datasets in a grid. The authors also used 6-hour time slots to divide the 24 hours day period to study the taxi behavior in a highly populated city. In our work, instead of using a grid system, we applied HDBSCAN to identify the main regions. This algorithm was chosen due to its efficacy clustering large spatial datasets with noise.

2.3. Modeling

People can be affected by certain attributes of a route (e.g. time walking, waiting), and this can help improving route selection for different user's profiles. For instance, M. Wardman [10] studied how users perceived time in different transportation means. The author presented different values for several time variables (such as walk, waiting). The idea was to model and understand the perception of time according to different time variables, by exploring several surveys performed in England. P. Abrantes and M. Wardman [11] went further and improved this research, by adding a better division of different travel moments (such as congestion). P. Ryus et. al. [12] and N. Schaap et. al. [13] analyzed how a user decides to access a certain transportation mode (such as bus, subway). The authors concluded that it was possible to identify the different predisposition curves of the users, according to the traveling distance to access the service.

T. Arentze and E. Molin [14] presented a model where the users had to choose two options for a specific journey. At the end, the authors traced socio-demographics characteristics from the sample to understand the differences between different classes, ages, education levels, among others. Achieving precise values is a complex task, especially considering human subjectivity when making decisions. The results provided by these studies are used in our proposal to obtain more realistic results.

3. Proposal

The present paper aims to evaluate the use of hybrid multimodal routes producing more personalized routes and the impacts created by these routes. These multimodal routes may include transport modes such as walk, bus, subways, ferries, among other modes, and also consider HPV, such as taxis and similar services, as Uber. Hybrid routes may use HPV for town trips, but may also apply transit routing to prevent the higher costs of this mode due to congestion time. We study the hypothesis of cost reduction with low impact on user experience and trip time. To check our hypothesis, we use data from real trips occurred in New York City. This data is presented in Section 3.1. The identification of trip trends to reduce the dimensionality of the dataset is developed using clustering techniques described in Section 3.2 and the framework to obtain and evaluate the impacts of the hybrid trips is described in Section 3.3.

3.1. Data Characterization

The New York City has a program to make the data created by its agencies available to the public. This project is called NYC Open Data[3] and has published datasets related to wealth, mobility, safety and other urban aspects. The current work explores the Yellow Taxis dataset, containing data about pickups and drop-offs of trips in the city yellow cabs. This dataset describes variables, such as start and end locations, duration, passenger count, and cost. To perform the analysis proposed in this paper, we took data from March/2016, which contains 11.618.824 valid taxi trips. The trips from 2016 were chosen mainly because of two reasons: (1) the datasets before June/2016 represent the start and end location of the trip with GPS positions, rather than zone IDs used nowadays – the GPS positions allow a better analysis since the zones used to represent the data currently are too large which causes imprecisions, and (2) the taxi industry, and consequently the amount and representativity of trips, are more significant in previous years due to a reduced influence of the popularization of HPV smartphone applications. We also remove invalid data from the dataset by establishing a 3D bounding box (i.e., geo-coordinates and

3. https://opendata.cityofnewyork.us/

time boundaries) which encapsulates the city in the month of March/2016. Also, some of the trips were wrongly recorded (e.g., end timestamp before start timestamp). To deal with such samples we adopted speed thresholds, where trips using taxis, with an average speed faster than 100Km/h or slower than 5Km/h, were also considered invalid. Even after applying the filter for valid trips we still have a significant amount of data, which needed to be reduced in order to proceed with the analysis. This data reduction is described in Section 3.2.

3.2. Data Reduction

To allow the analysis of the hybrid routes, we had to reduce the amount of data, otherwise, it would take much time and resources. Thus, the focus of this section is to describe the effort made to reduce the dataset magnitude. Our objective was to identify the main flows in the city, this way allowing the analysis of the impact of hybrid routing on these flows. The approach to identifying the main taxi trip flows in New York consists of four steps:

i **Hourly Split:** divide the data by hours and also between weekends and weekdays;
ii **Functional Region Identification:** cluster the start and end point of the trips to identify the main regions where they take place;
iii **Flow Accounting:** count the flows between the identified regions;
iv **Flow Classification:** classify the flows between main flows and secondaries.

The first step is to split the dataset into several hour-based smaller ones. We observed that an hourly division of the data could capture the main changes on trip frequencies, as shown in Figure 1. It is possible to visualize major changes in the number of trips and passengers traveling in different hours of the day and, also, the difference in the behavior between weekdays and weekends. The observation of this trend guided us to divide the original dataset according to the time when trips started and then perform the identification of the regions in the smaller datasets. This division also helps to match hour-specific traffic conditions with flows. For instance, Figure 1 shows some peaks of taxi usage, which were classified in: (1) Low Peak: weekdays from 8h to 16h, (2) High Peak: weekdays from 18h to 23h, and (3) weekends from 11h to 1h – passing through 0h.

After splitting the dataset, the next step is to cluster the origins and destinations of the trips to identify interest regions where the flow concentrates. The use of the HDBSCAN algorithm [8] is desirable because this algorithm analyses the density of a distribution of points and groups them. HDBSCAN is an improved generalization of the well-known DBSCAN clustering algorithm; these algorithms are used to cluster large spatial databases with the presence of noise, therefore HDBSCAN was selected to our experiment. The clustering was performed individually for every tuple <event, hour, day-type>; where the event can be either origin or destination, the hour is the hour when that event happened, and the day-type is the classification between weekday or weekend. By clustering the data in this way we obtained 96 sets of clusters (i.e., 2 events × 24 hours × 2 day-types). Then, the clusters for the events of origin and destination for the same hour and day-type were used to count the flows – for every origin cluster, we computed the number of trips going to each of the destination clusters. After filtering non-trending areas, the number of flows was still significant, therefore further downsampling is needed.

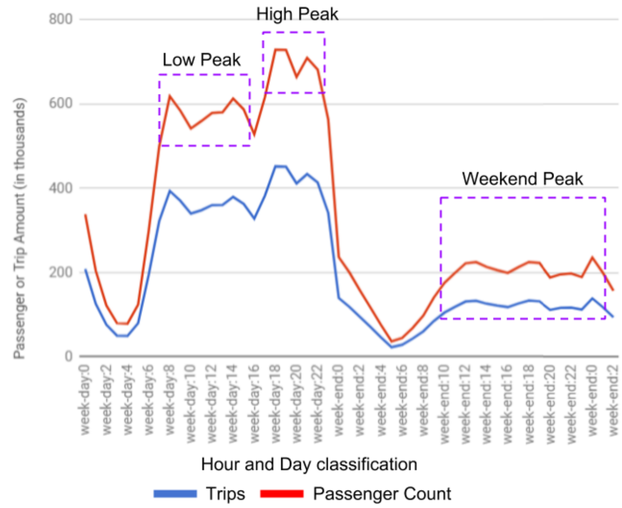

Figure 1. Distribution of trips and passengers per hour in weekdays and weekends.

To perform this, we classified the flows between trending and secondary by the number of trips in the flow. It is important to notice that the importance of the flow is bound to the hour it happened, i.e., a trending flow on weekend may not be considered trending if it happened on a weekday. Hence, the classification of the flows was made based on the hour and day-type when it happened. To classify them, for every pair <hour, day-type>, we sorted the flows according to the number of trips forming a long-tail curve (i.e., exponential decay like curve), in a descending order. After this, the resulting curve is analyzed using the following equation:

$$\kappa = \frac{|x'y'' - y'x''|}{(x'^2 + y'^2)^{3/2}} \tag{1}$$

The curve of sorted trips per flow must be parametrized in two functions x and y – where x is a sequenced discrete index for the flows and y is the number of trips in each flow. Using the first and second order derivatives of the parametric functions, it is possible to evaluate the curvature equation shown in Equation 1. The point of transition, where the short-head of the curve gives place to the long-tail, is the global maximum of κ, which can be found using its derivative (i.e., $\kappa' = 0$). Finally, all flows to the left of the transition point are considered to be trending, and all flows to the right are taken as secondary flows. The developed method was inspired on the approach suggested to evaluate the best EPS value in the DBSCAN algorithm [15] – this EPS works as a boundary for neighborhood searches in the DBSCAN. Using this methodology it was possible to select the main flows in the city. The results of data reduction are shown and discussed in Section 4. Once the amount of data is reduced, it is possible to start the viability analysis of the hybrid routes; the description of the analysis process is presented in Section 3.3.

3.3. Framework of Analysis

Mobility in urban centers is affected by several situational aspects, such as traffic conditions, weather, city events and also some user-centric aspects, such as preferences and perceptions. Hence, creating a framework that allows the integration of these different aspects which may influence the way people move within the city is necessary. In the present paper, we integrate transit and HPV routing with contextual information. Precisely, we use traffic conditions data and urban transportation user's models to evaluate the use of transit transport modes to avoid traffic congestions in the city. The framework presented in this section is used to obtain and evaluate the impact of these hybrid routes with different metrics.

In order to evaluate the actual impact of using hybrid HPV-transit routes, we need to analyze the real trip behavior of the city. Thus, the first step in the methodology is to find a position dataset that can represent this behavior. Once the data was obtained, the following step consists in downsampling it and creating a flow graph containing the main trip flows observed – as described in Section 3.2. Every flow will then be used to create an initial route, which is a way a vehicle can use to navigate from the origin of the flow to its destination. This initial route is used to match traffic conditions data and to identify segments with poor traffic conditions. From these segments we extract the exchange point candidates, i.e., points where the user may change transportation mode in order to avoid congestion. The knowledge of these points is used to produce the hybrid routes by replacing congested segments that would be traversed in a hired vehicle by a transit option. Once all the hybrid and non-hybrid routes are known we use different models to evaluate the impact of the routes in time, cost, and user experience. All the routing steps in our work were made using the Google Directions API[4]. The traffic data was collected from the Here API[5].

After matching traffic and routing data, it is possible to extract the exchange points based on the congested segments of the routes. These exchange points are used to identify where users can change to other transportation modes, thus we build hybrid options for every exchange point to be evaluated. After obtaining the route options, the evaluation takes place. We analyze the cost of the trip, the real time spent and, using a mathematical model, the user-perceived time. The model to measure users perception was built based on different studies that performed surveys with passengers and evaluate what were the impressions about the trips they made. For instance, the work from Wardman [10] establishes weights for the time spent in each phase (e.g., access, egress, headway) of the trip; these weights are designed using as a basis the in-vehicle time-lapse perspective of the users. Also, the book Transit Capacity and Quality of Service [12] presents curves describing the willingness of users to walk to access transit stations based on the distance to the station. The pseudo-code representing the algorithm to compute the hybrid HPV-transit routes is shown in Algorithm 1.

Algorithm 1 Compute available hybrid HPV-transit options for a given origin and destination.

```
 1: procedure get_hybrid_route(orig, dest)
 2:     drive_way ← get_driving_way(orig, dest)
 3:     trans_start_candidates ← newList()
 4:     trans_end_candidates ← newList()
 5:     options ← newList()
 6:     for each (index, step) in drive_way.steps do
 7:         if step.lenght > 500 then
 8:             frags ← split_step(step, 500)
 9:             splice(drive_way.steps, index, 1, frags)
10:             continue
11:         end if
12:         traffic = get_traffic(step.orig, step.dest)
13:         if is_congested(traffic) then
14:             append(trans_start_candidates, step.orig)
15:             append(trans_end_candidates, step.dest)
16:         end if
17:     end for
18:     append(trans_start_candidates, orig)
19:     append(trans_end_candidates, dest)
20:     for each ts in trans_start_candidates do
21:         for each te in trans_end_candidates do
22:             opts ← get_options(orig, ts, te, dest)
23:             concat(options, opts)
24:         end for
25:     end for
26:     return options
27: end procedure
```

Algorithm 1 starts by initializing some variables. The $drive\_way$ is the driving route from the origin to the destination used as a base to match the traffic data with the path that an HPV would take. This driving way is created using the Google Directions API. After we have three lists, the $trans\_start\_candidates$ and $trans\_end\_candidates$ shelters the candidate's positions to start and end, respectively, the transit steps in the hybrid route. Finally, the $options$ list contains all the hybrid route options to be evaluated. The first loop on Line 6 walks through all the steps in the $drive\_way$. These steps are the commands given to drivers to reach their destinations. Each of these steps has an origin, an end, and length properties. We want to get the traffic conditions for each step, thus, Lines 12-16 get these data (from HERE Traffic API) and check if there is a congestion in the segment. If congestion is found, the origin of the step becomes a transit start candidate, and its destination becomes a transit end candidate. Since some of these steps may have a large length, we established a boundary of 500m for step length; therefore, Lines 7-11 split steps with greater length in smaller fragments, up to 500m length. These fragments replace the split step in the $drive\_way.steps$ list.

After the identification of transit starts and ends candidates, the trip's origin and destination are also added as transit start and end candidates. Lines 20-25 then combine the candidates for start and end transit mode creating hybrid route options. The method $get\_options$ access Google API

4. https://developers.google.com/maps/documentation/directions/
5. https://developer.here.com/documentation/maps/topics/traffic.html

to consult possible transit routes between the candidates and then uses Uber API to estimate the values of their services, this way crating a set of combinations given the reference transition points. All these options are concatenated (i.e., bulk appended) to the options list (on Line 23). Notice that by start, or end with transit we also consider the option of starting and ending with walking steps. Once the route options are computed, it is possible to evaluate their impact using the proposed metrics described in Section 3.3.

The evaluation of the route options was made by comparing the values of real trip duration, perceived trip duration, trip cost and also effective cost relations derived from these three variables. The time variable is evaluated using the Google Directions API time estimations, while the perceived time is computed using the time spent in each step of the route (e.g., access to transit, wait for HPV, headway) and the weighted perception of the users [10], [12]. The trip cost for transit trips was estimated using the current prices of buses and subways in New York City, while the HPV prices were estimated using the Uber Estimate API. The effective cost relations were obtained in two different ways: (1) using the real duration and (2) the perceived duration normalized values. This effective cost is the product of trip duration and its cost. Also we defined a user profile σ as a monetary value users are willing to gain by waiting more to reach their destination, for example a profile $\sigma = 2$ means that for every extra minute of duration of a trip the user should save USD 2.00. These profiles are used to evaluate the influence of users preferences in the personalized routes. Section 4 shows the results obtained using the framework.

4. Results and Discussions

We studied the possibility of combining transit transport with HPV to reduce the cost of the hired trips due to poor traffic conditions. The results obtained in this study are divided into two parts: firstly, we identify the main urban mobility flows in New York City; secondly, these main flows are analyzed to check the impact of hybrid HPV-transit trips; Finally we study some specific cases to verify the impact of user's preferences in the route selection. The general impact of the usage of hybrid routes is described in Section 4.1, latter Section 4.2 discusses the influence of adding user's views to perform route selection.

4.1. Overall Usage of Hybrid Routing

As described in Section 3, the main flows were identified using clustering techniques, which allowed us to create a flow graph for every pair <hour, day-type>. Figure 2 presents some examples of mobility flow graphs created from the yellow taxi trips. The red areas are the trip origins and the blue areas are the destinations. The arrows connecting the regions represent the flow between those regions, and its opacity represents the weight (i.e., normalized amount of trips) of that flow. According to the hour of the day, different graphs were obtained. It is possible to identify some trending shapes, like, for instance, Figure 2-A shows a graph containing a ring-shaped hub, where most of the flows are. Also, in Figure 2-B there are flows coming from many regions and converging to one; the opposite situation, where flows diverge to peripheral zones was also seen in later hours, such as 8-9 P.M. on weekdays. Finally, Figure 2-C shows an example where the methodology applied did not have a good performance in identifying zones, thus leading to meaningless flows that were not considered in the analysis.

The analysis of the mobility flow graphs allowed the creation of insights on which are the main regions of the city and how citizens move between them. Information as this can help, for instance, in the planning of bus lines. We use these mobility graphs to study the impact of hybrid HPV-transit routes. Figure 3 shows the impact of the proposed hybrid routes in different traffic scenarios.

The traffic scenarios in Figure 3 were extracted from the data presented in Figure 1. These scenarios are the weekend peak, from 11h to 23h and also 0h and 1h, and the low peak, from 8h to 16h on weekdays; and high peak from 18h to 23h also on weekdays. We established this division to analyze the impact due to different traffic conditions in these scenarios, which may have a different influence on the impact created by the usage of hybrid HPV-transit routes. We studied five metrics of the trips: cost, duration, perceived duration, effective cost and effective cost perceived, variables described in Section 3. Figure 3 compares each metric for every studied transport mode. All the variables were normalized to be shown in the graph, the bars represent a weighted average obtained for the metrics in each trip (i.e., flow) using as weight the number of trips in that flow.

Since the metrics presented in Figure 3 are cost-based, the lower values mean the best options. For specific flows, we observed situations where the HYBRID and TRANSIT mode had smaller durations than the HPV mode. In some cases, the HYBRID approach had higher costs than the HPV mode. Overall, the HPV trips are the fastest ones, but also the most expensive. However, traffic conditions and the cost raising for trips in these scenarios leads the values of effective cost of this mode to raise. In general, the variation of the metrics for HYBRID routes are smaller than in HPV, it just do not happen for the cost metric. Comparing traffic scenarios leads towards a blind comparison of a variety of trip lengths. Thus, we also compared the modes according to trip lengths to observe what is the behavior of the hybrid HPV-transit routes. The results for different trip lengths are shown in Figure 4.

In Figure 4, the objective was to identify patterns on the usage of hybrid HPV-transit routes in trips with similar length. This lead to a more stable variation of costs and durations than comparing trips in the same traffic scenarios. Figure 4 shows that the impact of the cost for hybrid routes reduces, in comparison to HPV only routes, as the trip length increases. This is reflected in the reduction of the average effective cost and effective cost perceived for hybrid routes, while the HPV mode is increasing. This trend of reduction is kept even for longer trips, that were not shown. This behavior is observed because by adding transit modes in small trips we also add waiting times that are more significant for short duration trips.

As the data shows, using hybrid HPV-transit routes may reduce the cost of trips and have low impact on duration

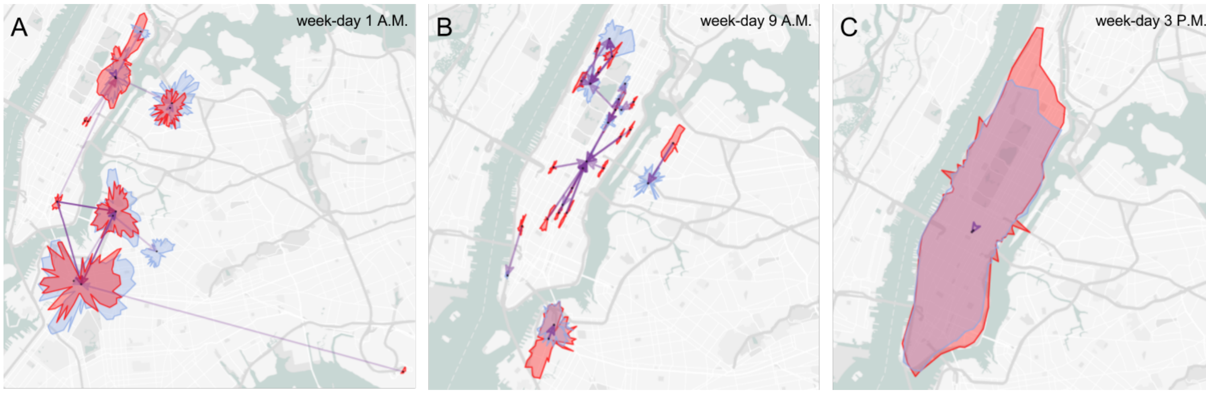

Figure 2. Graphs representing important flows between New York City regions.

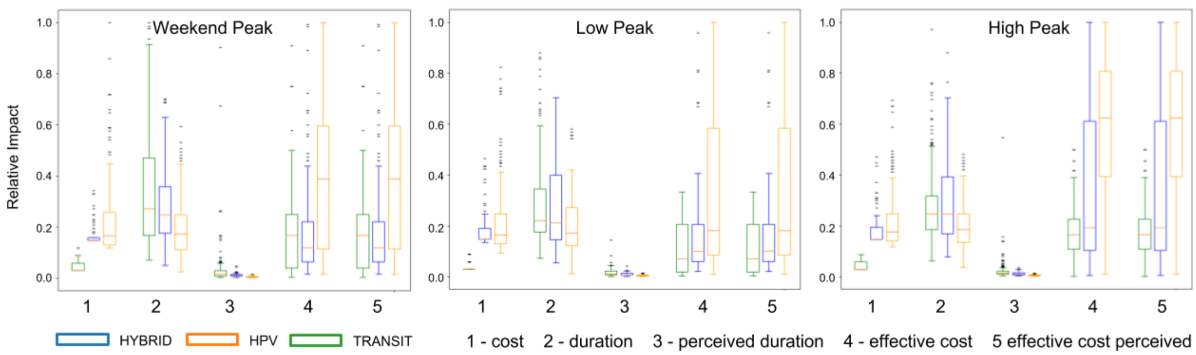

Figure 3. Impact of using different routing options for different traffic conditions. HPV represents the trips using hired private vehicles, estimated using Uber API, TRANSIT represents the traditional public urban transport modes (e.g., bus, subway), and HYBRID represents our proposal of routes using both HPV and TRANSIT.

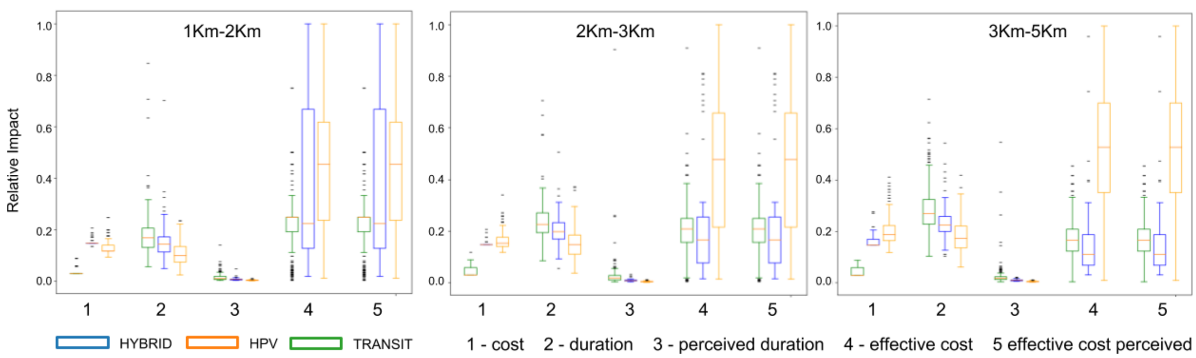

Figure 4. Impact of using different routing options for different trip lengths. HPV represents the trips using hired private vehicles, estimated using Uber API, TRANSIT represents the traditional public urban transport modes (e.g., bus, subway), and HYBRID represents our proposal of routes using both HPV and TRANSIT.

compared to HPV only trips. Also, the average hybrid trip is faster than the transit only trip. Furthermore, when comparing the user experience (i.e., perceived duration) in the trip, the HPV-transit routes obtained values comparable with the HPV only routes, while the transit only trips were as high as twice the perceived duration of the hybrid routes.

4.2. Route Selection Using User's Profile

To evaluate the impact of user's preferences in the route section we choose two trips from the identified flows. The first trip (**T1**) is the route where the gains in cost by using hybrid trips were the greatest when compared to HPV; the second trip (**T2**) is the longest flow identified, which consists in a trip from the center of Manhattan to the JFK airport. Table 1 shows the cost, duration and congestion time of these trips for each of the considered routing modes. To evaluate the influence of the user preference we consider user profiles containing a monetary value for minute of time.

TABLE 1. COMPARISON OF TWO TRIPS SELECTED FROM THE MAIN FLOWS OF THE DATASET.

	Category	Cost	Duration	Congestion Time
T1	HYBRID-1	USD 17.50	53 min	5 min
	HYBRID-2	USD 63.50	43 min	5 min
	TRANSIT	USD 2.50	53 min	0 min
	HPV	USD 85.00	33 min	13 min
T2	HYBRID	USD 8.00	60 min	1 min
	TRANSIT	USD 5.00	72 min	0 min
	HPV	USD 66.00	45 min	2 min

Trip **T1** is a 8Km distance from origin to destination and happened at 15h on a weekend, **T2** has about 25Km and happened at 13h on a week-day. In **T1** it was possible to reduce the amount of time spent in congestions from 13 to 5 min using the hybrid routes. Even with the saving in congestion, due to the usage of the public transport, the full duration of the trip has increased from 33 to 53 minutes in HYBRID-1 and for 43 in HYBRID-2 (the fastest transit only trip has). Thus, a user profile where $\sigma \geq 3.40$ would be willing to have more 20 minutes in his trip to save USD 67.00. In this scenario it would be even better to take the TRANSIT option, since the duration is the same as the HYBRID-1. Another relevant option would be HYBRID-2, even more restrict profiles where $\sigma = 2.20$ would take it and it saves 10 minutes compared to the TRANSIT option while still saving USD 22.00 regarding the HPV option.

In trip **T2** it was possible to save USD 38.00 by adding a 15 minutes delay, which indicates that a profile $\sigma > 2.53$ would take the HYBRID option instead of the HPV. In this case we can observe also a reasonable economy of time when comparing the HYBRID and the TRANSIT options. We could also observe that in 5% of the main flows (i.e., 50 out of 998) the cost and duration of the HPV-transit trips were lower than HPV only. However, such situations cannot be observed in the average's graph in Section 4.1 due to the low frequency. In these flows the average savings were USD 9.83 (first quartile: 3.50, third quartile 14.50), and 4 minutes (first quartile: 1.5, third quartile 5.6). If the usage of such routes was made easier, the trip experience of users in public transit could be improved while also reducing cost and duration for some trips.

5. Conclusion

The present study proposes the use of hybrid Hired Private Vehicles and Transit (HPV-transit) urban routes to produce personalized urban routes. We evaluate the impact of our proposal in terms of reduction of cost and duration of trips, and also congestion avoidance. Five metrics are used to compare the traditional HPV or transit only routing with the proposed hybrid HPV-transit routing. We identify some situations where the HPV-transit routing outperforms HPV only routes, either in trip duration and cost. However, these situations have a low frequency which reduces their relevance in the overall scenario. Even with the observed values for trip durations lower in HPV only trips than the HPV-transit, there is a reduction in the average trip costs with small impact on their durations. Our main contributions are the proposition of a clustering technique to identify trending mobility flows and the measurement of the gain obtained by using hybrid HPV-transit to avoid traffic congestions.

Acknowledgments

This work was carried out in the scope of the MobiWise project: from mobile sensing to mobility advising (P2020 SAICTPAC/0011/2015), cofinanced by COMPETE 2020, Portugal 2020-POCI, European Regional Development Fund of European Union, and the Portuguese Foundation for Science and Technology.

References

[1] J. Liu, Y. Li, M. Chen, W. Dong, and D. Jin, "Software-defined internet of things for smart urban sensing," *IEEE Communications Magazine*, vol. 53, no. 9, pp. 55–63, September 2015.

[2] Q. Song, D. Li, and X. Li, "Traffic prediction based route planning in urban road networks," in *2017 Chinese Automation Congress (CAC)*, 2017, pp. 5854–5858.

[3] E. Zhou, S. Mao, and M. Li, "Investigating street accident characteristics and optimal safe route recommendation: A case study of new york city," in *2017 25th International Conference on Geoinformatics*, Aug 2017, pp. 1–7.

[4] H. Hamedmoghadam-Rafati, I. Steponavice, M. Ramezani, and M. Saberi, "A Complex Network Analysis of Macroscopic Structure of Taxi Trips," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 9432–9437, 2017.

[5] P. Campigotto, C. Rudloff, M. Leodolter, and D. Bauer, "Personalized and Situation-Aware Multimodal Route Recommendations: The FAVOUR Algorithm," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 1, pp. 92–102, 2017.

[6] G. Gartner, *Progress in Location-Based Services 2016*, 2017.

[7] M. Daszykowski and B. Walczak, "Density-Based Clustering Methods," *Comprehensive Chemometrics*, vol. 2, pp. 635–654, 2010.

[8] R. J. G. B. Campello, D. Moulavi, A. Zimek, and J. Sander, "Hierarchical Density Estimates for Data Clustering, Visualization, and Outlier Detection," *ACM Transactions on Knowledge Discovery from Data*, vol. 10, no. 1, pp. 1–51, 2015.

[9] M. Momtazpour and N. Ramakrishnan, "Characterizing Taxi Flows in New York City," *UrbComp*, p. 9, 2015.

[10] M. Wardman, "A review of British evidence on time and service quality valuations," *Transportation Research Part E: Logistics and Transportation Review*, vol. 37, no. 2-3, pp. 107–128, 2001.

[11] P. A. L. Abrantes and M. R. Wardman, "Meta-analysis of UK values of travel time: An update," *Transportation Research Part A: Policy and Practice*, vol. 45, no. 1, pp. 1–17, 2011.

[12] r. E. Peteauthor = Ryus, Paul and Danaher, Alan and Walker, Mark and Nichols, Foster and Carter, Bill and Ellis, Elizabeth and Cherrington, LInda and Bruzzone, Anthony, *Transit Capacity and Quality of Service Manual, Third Edition*, 3rd ed., 2013, ch. 4, pp. 4–18.

[13] N. Schaap, L. Harms, M. Kansen, and H. Wust, *Cycling and Walking: the grease in our mobility chain*. Kennisinstituut voor Mobiliteitsbeleid / Netherlands Institute for Transport Policy Analysis, 2016.

[14] T. A. Arentze and E. J. Molin, "Travelers' preferences in multimodal networks: Design and results of a comprehensive series of choice experiments," *Transportation Research Part A: Policy and Practice*, vol. 58, pp. 15–28, 2013.

[15] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A Density-based Algorithm for Discovering Clusters a Density-based Algorithm for Discovering Clusters in Large Spatial Databases with Noise," in *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, ser. KDD'96. AAAI Press, 1996, pp. 226–231.