Future Generation Computer Systems 135 (2022) 1–9

Contents lists available at ScienceDirect

Future Generation Computer Systems

journal homepage: www.elsevier.com/locate/fgcs

Towards drivers' safety with multi-criteria car navigation systems

Leonardo Solé [a,*], Matteo Sammarco [b], Marcin Detyniecki [b], Miguel Elias M. Campista [a]

[a] Universidade Federal do Rio de Janeiro (UFRJ), Brazil
[b] Axa, Paris, France

ARTICLE INFO

Article history:
Received 31 July 2021
Received in revised form 9 December 2021
Accepted 19 April 2022
Available online 30 April 2022

Keywords:
Car navigation systems
Alternative routes
Multi-criteria classification

ABSTRACT

Usual car navigation systems are configured to propose either the shortest or the fastest path between any origin–destination pair, neglecting the particularities of the territory. Some roads are impracticable when raining, some others are to avoid at night for the scarce lighting, or less safe for the presence of criminality and high accident ratio. On the other hand, longer paths can be safer and more pleasant as they pass through less noisy zones, with the presence of beautiful landscapes. In this paper we analyze the faults in current car navigation systems, especially quantifying the trade-off between safety and traveling time or path length. We propose two multi-criteria route planning methods, HVT (Hierarchical with Variable Tolerance) and R2V (Route to Vector), suggesting the best path to drivers also considering safety or multiple drivers' specific needs. A dataset of 3,170 paths from 600 origin/destination pairs within London is created and shared to the research community. With this dataset, we show that selecting routes with reduced driving risks is indeed possible with a marginal increase in travel times.

1. Introduction

The navigation system market, including automotive applications, is expected to register a CAGR of 11.3% over the forecast period 2020–2025 [1]. In particular, car navigation systems contain digital maps, with information about the neighboring areas, and use route planning algorithms to give directions to drivers. These route planning algorithms tend to focus on finding the fastest or the shortest path between origin–destination (OD) pairs [2].

The downside of the single-criterion approach, however, is the chance to lead drivers through unpleasant or even dangerous routes without significant time-related gains. Other characteristics may have an impact on drivers' overall satisfaction when traveling as well. The contact with nature has been found to have a positive effect on mental health and wellness, even if just visual, as is the case for drivers [3]. On the other hand, traffic may increase stress levels and even mortality due to noise and air pollution [4]. Besides health-related problems, drivers also experience anger issues on road, that influences the drivers' attitude, with an increase in aggressive driving, risky driving, driving errors, and accidents [5]. Choosing alternative routes, therefore, can have an enormous impact on the drivers' safety [6–9].

Fig. 1 shows an example of a set of routes for a given OD within the London area. The map includes an overlay heatmap representing the criminality rates and a second overlay indicating locations of vehicle accidents with relative severity. Route 1, represented as a bold yellow path, is the fastest route and it is recommended by most car navigation systems. It passes through an area with high criminality rate though. On the other hand, Route 2, represented by a red line, goes through an area with less criminality and more nature, and could be preferred over the first route, especially if the duration of both routes is similar.

This paper aims to provide methods to measure different route characteristics that may influence the driver's safety and pleasantness perception. We evaluate different route selection methods with either a single or a multiple criteria approach, with the intent to find the best path from a set of previously calculated routes. We introduce two novel multi-criteria route selection methods. The first method, called Route to Vector (R2V), translates route features into vectors and finds the one closest to the best vector. The second, called Hierarchical with Variable Tolerance (HVT), follows a user-defined feature order to reach the best route.

To evaluate our approaches, we have beforehand built a novel dataset containing criminality, traffic, accidents, nature, tourist attractions, and trajectory information data about 3170 routes from the city of London [10]. Safety-related parameters are individually evaluated with the proposed methods to emphasize safety gains. The obtained results show that our multi-criteria approaches decrease driving risks without significant time-related penalties. Also, we confirm that relying just on duration or path length may lead to dangerous and unpleasant trajectories.

* Corresponding author.
E-mail addresses: sole@gta.ufrj.br (L. Solé), matteo.sammarco@axa.com (M. Sammarco), marcin.detyniecki@axa.com (M. Detyniecki), miguel@gta.ufrj.br (M.E.M. Campista).

https://doi.org/10.1016/j.future.2022.04.019
0167-739X/

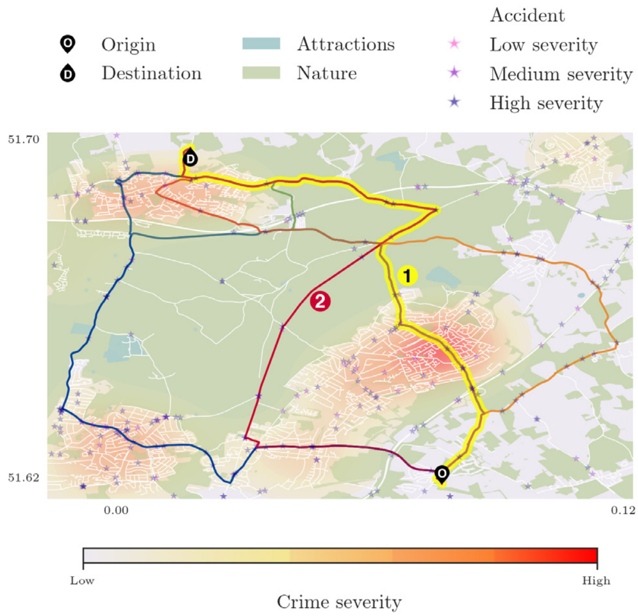

Fig. 1. Example of a set of routes for a given origin–destination (OD) pair within the London area. The underlying heatmap represents the criminality rates, while the stars indicate positions of previous vehicle accidents. Route 1 is the fastest route recommended by most car navigation systems and it goes directly through a higher criminality area. Route 2, instead, is slightly slower, but it passes through an area with less criminality and more nature.

Throughout this paper, the names trajectory, route and path are used interchangeably.

The remainder of this paper is structured as follows. Section 2 provides an overview of car navigation systems, while in Section 3 we examine previous works on route classifications. Next, in Section 4, we describe our collected dataset containing a set of route parameters that can influence the driver's satisfaction. Then, we propose two route selection methods, the Route to Vector and the Hierarchical with Variable Tolerance, based on multiple criteria in Section 5. We evaluate the performance of both proposed methods to assess route classification and drivers' security improvements in Section 6. Finally, Section 7 concludes this paper and draw future directions.

2. Car navigation systems overview

Car navigation systems are in-car or smartphone based systems implemented to aid drivers in planning a route for a given OD pair. Upon receiving the OD pair translated into coordinates of a digital map, these systems compute and advise a set of possible routes, usually considering duration as the preferential single criterion. Car navigation systems can also have additional features such as real-time alerts about traffic events. However, this functionality can be viewed as an add-on available thanks to integrated hardware or software. We divide car navigation systems according to the presence of real-time, predictive, multi-modal, or multi-criteria planning.

- **Real-time planning**: This feature contrasts with car navigation systems which exclusively compute routes based on coordinates of a digital map. In fact, they can rerun the route planning algorithm if the car moves away from the planned route. Early car navigation systems relied solely on GPS coordinates and a digital map to compute routes. This strategy does not consider the route dynamics imposed by accidents or any other event that could affect the set of routes previously selected. To handle real-time planning, car navigation systems must have access to a real-time data source through the Internet or even through the collaboration of other vehicles, such as in crowdsourcing approaches [11]. Google Maps and Waze are examples of car navigation systems that rely on the participation of users, either passively or not, to keep fresh information regarding traffic conditions.
- **Predictive planning**: Traffic events follow certain known patterns, for instance, traffic jams can happen on a certain road in rush hours but not at different times along the day. Hence, based on time and also on an increasing trend in the number of vehicles, the system could be able to anticipate a probable traffic jam and advice an alternative path beforehand [2]. This strategy relies on a previously collected history of events and thus need additional infrastructure such as that provided by cloud facilities.
- **Multi-modal planning**: Public transport may consist of several different modes of transport, for instance, bus, train, subway, or ferry. This introduces the need for multi-modal route planning, where all different modes available are considered. This type of planning may also include other transport modes that are not considered public, such as carpooling [12]. As an example, ROSE combines pedestrian navigation and public transport to minimize walks and waiting time through an A*-like algorithm [13].
- **Multi-criteria planning**: The best route is a very subjective definition and can vary depending on evaluated features and personal preferences. Although some navigation systems take into account multiple criteria, the best path decision may still be based on a single criterion, for instance, evaluating different features to find the fastest route. This contrasts with the subjective nature added by including users into the loop. Even though finding the best path based on multi-criteria algorithms is known to culminate in NP-complete problems, it is important to draw heuristics to at least enlarge the concept of best-path finding to include not only quantitative but also qualitative metrics, which is our main contribution in this paper.

3. Related work on multi-criteria planning

In this section, we present an overview of recent navigation systems that provide different route planning methods, taking into account the notion of multi-criteria car navigation systems. These systems are complex and do not exclude real-time, predictive, and multi-modal planning. The multiple criteria approach may be implemented in different ways. We present different methods with three different objectives: the first and second ones focus on tourist needs, the third one is directed to cyclists, and the last one aims to avoid traffic congestion.

Huang et al. propose a route planning framework, called Multi-Task Deep Travel Route Planning, addressed to tourists [14]. It goes through three stages: a feature extraction stage, which gathers data from points of interest, travel routes, and user's preferences; a learning model stage, which uses a deep learning model that considers features collected to uncover the probability for the best next point of interest; and as the last stage, a route generator, that provides a travel route based on the user's preferences. For a given OD pair, the system finds the best route for sightseeing in according to the user's preferences.

Quercia et al. suggest, in their work, routes that are not only short but also emotionally pleasant, i.e., places considered beautiful, quiet, and happy [15]. At first, a certain number of short OD paths are identified through the Eppstein's algorithm. Then, the path with the highest scores for those three features, relying on crowdsourcing data and user generated content posted on online social networks, is considered as the best one.

Derek et al. [16] introduce a bicycle route planning system with multiple criteria analysis. The system includes both leisure and safety-related features, such as road segment length, road type, slope grade, distance to the emergency unit, and distance to the drinking water source, and calculates the difficulty of the road segments accordingly. It then proposes three different routes, one for beginners and two others for more advanced riders. This is an example of a non-subjective multi-criteria approach since it analyzes different features, that uses a well defined single criterion decision metric, the route difficulty level. Another example of a multi-criteria selection system with single criteria decision was proposed by Xu et al. to select alternative paths to avoid traffic congestion [17]. The model divides road segments by road grade and has an index of traffic performance for each entry. It uses a deep learning approach with traffic features such as traffic volume, speed limit, distance, traffic lights, and weather features to find the best alternative path to avoid heavy traffic.

Although there are many car navigation systems with different route planning techniques, there is a gap when it comes to a more subjective and personal selection. Huang et al., Quercia et al., and Xu et al. propose multi-criteria approaches with a focus on tourism, emotional reaction from drivers, and traffic congestion, respectively. We, instead, propose two methods with configurable preferences, following the drivers' wish. Unlike Derek et al. the system is mainly proposed to drivers and not to bikers. In addition, the proposed methods are simple and can easily quantify the multiple criteria employed for route computation using point count, weighted point count, intersection area, and scalars. The proposed multi-criteria methods are presented in more details in Section 5.

The main contributions of the proposed methods are customization and simplicity. Instead of focusing on a specific type of user or parameter, we give the user the ability to choose the most important characteristics the best route should have. In addition, the proposed methods rely on simple computation algorithms, which can easily quantify the multiple parameters involved.

4. Dataset for multi-criteria planning

To compare our two route selection proposals with the baseline single-criterion approach, we built a dataset composed by several alternative paths for the same OD pairs within the city of London [10]. The dataset contains 3170 paths from 600 random OD pairs. Each OD pair creates a route set that contains a varying number of partially or totally disjoint routes generated using the HERE Maps API [18], along with duration estimations. The HERE Maps API does not return paths exceeding the fastest one in 20%, individually considering duration or path length. This is an internal API parameter that users cannot tweak, further determining the number of paths in each set. The city of London was chosen as scenario due to the availability of crimes and accidents open data repositories [19,20].

Each route is described by a n-tuple $\mathbf{f}$, where n is the number of features and each feature has a numerical value assigned f, which will be normalized as $\hat{f}$. In particular, we consider seven features:

1. **Duration:** estimated time to reach the destination.
2. **Length:** kilometers to drive to reach the destination.
3. **Traffic:** ratio between the duration without traffic and the current duration estimation.
4. **Crime:** level of danger along a route due to previous acts of criminality.
5. **Accident:** level of danger along a route due to previous road accidents.

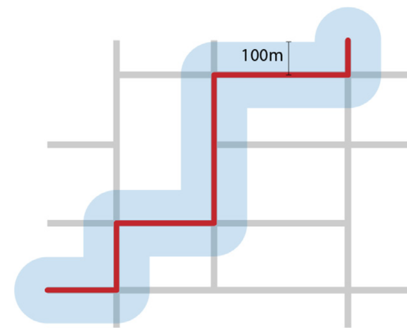

Fig. 2. Example of a Route Region ($\mathcal{R}_r$).

6. **Nature:** presence of natural surroundings along the path.
7. **Attractions:** presence of touristic attractions along the path.

The features described in this section are route characteristics that may influence the driver's choice. While duration, length, and traffic are pragmatic path features to consider, crime and accident are serious concerns for drivers, especially when driving in unknown places. The presence of nature and attractions is, instead, a subjective preference. Without loss of generality, other features might be added if available such as noise, air pollution, and road conditions.

For crime, nature and attractions, a visible region around the route, referred to as Route Region ($\mathcal{R}_r$), is estimated taking into account the route segments and a margin of 100 m for the left and right side, as shown in Fig. 2. In addition, the value assignment strategy considers four different feature types: point count for attractions; weighted point count for crime and accidents; intersection area for nature; and scalars for duration, length, and traffic. Among the selected features, nature and attractions cannot be used as a single criterion since they are not upper bounded: the more nature or attractions we have, the better the route. Hence, the selected route would be longer and longer. In these cases, we need an additional feature to previously determine a set of route candidates and, consequently, an upper bound for nature and attractions.

The departure times chosen for each set of routes was a random time within each of the five time slots previously defined: {0 am–7 am, 7 am–10 am, 10 am–16 pm, 16 pm–19 pm, 19 am–24 pm}. These five sets were chosen as they have different vehicular traffic profiles. The first (0 am–7 am) and the last one (19 am–24 pm) have low traffic, the second (7 am–10 am) and the fourth one (16 pm–19 pm) capture rush hours, and the third one (10 am–16 pm) has midday traffic. The value used for the time dependent features duration and traffic was the mean value found for each path.

Following, we describe in detail how to compute each feature, taking into account all other routes from the dataset.

Duration: Route duration is an important parameter commonly used in most car navigation systems that undoubtedly impacts drivers' perception of the trip. Since driving can be a stressful experience and time is a scarce resource for many people, the sooner one arrives to destination, the better. Also, reducing the exposure time on the road reduces the overall driving risk. Route duration estimation is acquired directly from HERE maps which takes into account the starting time [18]. *Hence, the duration feature used is the direct value obtained from our dataset.*

Length: Route length is another important practical parameter that is also commonly used in many car navigation systems. This parameter has a considerable impact on fuel consumption and

Table 1
Crime severity weights per category. These weights are computed as an average of all crimes related to a given category.

Category	Weight
Anti-social behavior	3
Public order	10
Shoplifting	13
Bicycle theft	16
Other theft	33
Possession of weapons	75
Theft from the person	86
Vehicle crime	124
Drugs	250
Other crime	256
Burglary	438
Criminal damage arson	439
Robbery	746
Violent crime	1098

consequently on the expenses of the trip. HERE maps returns the exact length of each route [18]. *The length feature used is the direct value obtained from our dataset.*

Traffic: Traffic is a known cause of stress among drivers and has a major impact on traveling time. Rather than scoring this feature by the time length or amount of cars on each route, the traffic value is calculated as the ratio between the duration of the trip in a scenario without any traffic and the route duration at the time the data is collected, as returned by the HERE Maps API [18]. *The traffic is then computed as the ratio between the no-traffic route duration (d_n) and the route duration considering traffic (d_t).*

$$T = \frac{d_n}{d_t} \tag{1}$$

This assigns the maximum value of 1 to routes without traffic.

Crime: Choosing a route passing through areas with high criminality rates may result in an additional level of stress and risk. To assign crime scores to the routes within the dataset, we employ the UK government open data repository. In fact, the UK Police Data repository [20] gathers information about crimes that occurred in the city of London in the year 2018.

Such crimes are geolocalized with latitude and longitude coordinates and labeled with a category. To consider the severity of each crime, categories are weighted based on the UK Office for National Statistics Crime Severity Score [21], whose weights vary from 3 to 7979, with an average of 526. Since the categories listed in the UK crime dataset have subcategories, the weights used for our calculations are set as the average weights of all crimes related to the same category. The categories considered in our dataset are listed in Table 1, each one with its related average weight value. *The crime value is computed as the weighted sum of all crimes found inside the route area R_a.*

The crime value C is computed as the sum of all crime severity found inside each Route Region $\mathcal{R}_r$, as seen in Eq. (2), where c_k is a point with the coordinates of the crime with severity c_k^s, and $\mathcal{S}(c_k)$ is a set with all accidents that are found inside $\mathcal{R}_r$.

$$C = \sum_{\mathcal{S}(c_k)} c_k^s \text{ , where } \mathcal{S}(c_k) = \{c_k \mid c_k \in \mathcal{R}_r\} \tag{2}$$

Accident: One of the most stressful and dangerous events that can happen during a trip is to be involved in an accident. Hence, data providing paths having less overall accidents are acquired through a repository maintained by the UK government: the UK Government Open Data [19]. This data repository gathers information of accidents that happened in the city of London in the year 2018, with geographical coordinates for each accident.

Similarly to crime definition, we denote a point a_k with the coordinates of the accident with a_k^c casualties and severity $a_k^s \in [1, 3]$. *Thus, the overall value of accidents for each route A_c is calculated as the sum of all severity and casualties found in it.*

As the geolocalization for the accidents is not always accurate, to determine if an accident happened along the considered route r inside $\mathcal{R}_r$ or in a street nearby, the minimum distance from the accident location to the route r ($d(r, a_k)$) is compared with the minimum distance to all streets around ($d(\bullet, a_k)$). Thus, an accident is only considered if the distance to the route is less than or equal to the distance to every other street around. as seen in Eq. (3). We define the set formed by these accidents as $\mathcal{S}(a_k)$. The idea of such an approximation is to compensate for possible errors that may occur from different coordinate sources.

$$A_c = \sum_{\mathcal{S}(a_k)} (a_k^s + a_k^c) \text{ , where } \mathcal{S}(a_k) = \{a_k \mid d(r, a_k) \leq d(\bullet, a_k)\} \tag{3}$$

Nature: One parameter that affects the aesthetics of a trip is the presence of natural and green sights. These natural sights are defined as: parks, gardens, marinas, golf courses, natural reserves, grass, greenfields, meadows and water. A route with more natural surroundings can reduce drivers' overall stress level, resulting in a more pleasant journey. To evaluate the amount of nature in each route, we employ the Overpass API [22], that provides geolocalization information from the OpenStreetMap database, with the following keywords:

```
"leisure"~"park|garden|marina|
golf_course|nature_reserve"
"landuse"~"grass|greenfield|meadow"
"water"
"natural"
```

The data returned is used to create polygons on the map that are intersected with each Route Region $\mathcal{R}_r$. *The nature feature N is calculated as the total area of all intersections between nature polygons ($\mathcal{N}_k$) and the Route Region $\mathcal{R}_r$.* The set of all $\mathcal{N}_k$ that intersects $\mathcal{R}_r$ is defined as $\mathcal{S}(\mathcal{N}_k)$, as observed in Eq. (4). We assume that different nature polygons do not overlap.

$$N = \sum_{\mathcal{S}(\mathcal{N}_k)} \|\mathcal{N}_k \cap \mathcal{R}_r\| \text{ , where } \mathcal{S}(\mathcal{N}_k) = \{\mathcal{N}_k \mid \{\mathcal{N}_k \cap \mathcal{R}_r\} \neq \emptyset\} \tag{4}$$

Attractions: Another feature that contributes to a more pleasant route is the number of tourist attractions along the route. We retrieve this feature from the Overpass API [22] with the following query:

```
"tourism"~"[^information|hostel|
guest_house|hotel|apartment|yes]"
```

The attractions data is used to create points (t_k) on the map. *Hence, the attraction value A_t is calculated as the number of attractions found inside each Route Region $\mathcal{R}_r$.* Let $\mathcal{S}(t_k)$ be the set of all attractions inside $\mathcal{R}_r$. Thus, Eq. (5) computes the cardinality of $\mathcal{S}(t_k)$ to find A_t as follows:.

$$A_t = |\mathcal{S}(t_k)| \text{ , where } \mathcal{S}(t_k) = \{t_k \mid t_k \in \mathcal{R}_r\} \tag{5}$$

5. Multi-criteria classification methods

The proposed multi-criteria methods follow a workflow composed of: (*i*) per route feature calculation, (*ii*) feature normalization, and (*iii*) routes ranking. The input is a set of routes R that can be randomly selected or previously calculated considering any preliminary metric of interest. Hence, each route R_i is described by a n-tuple $\mathbf{f}$ considering features of interest. We can select any possible subset computed using any combination of features available. In this paper, we use path duration to select the input set of routes since this is the most common approach used by car navigation systems and we aim to avoid much longer paths with respect to the shortest one. After selecting the input set, we compute and normalize all features in the n-tuple $\hat{\mathbf{f}}$.

A normalization procedure is required to allow a multi criteria classification of routes, since features described in $\mathbf{f}$ may have different value ranges. Note that this normalization is used only for the sake of computing the best route, the analysis (Sections 6 and 7) will be done with the non-normalized values. Given the ith route feature f_i, its normalized value is computed as follows:

$$\hat{f}_i = \frac{f_{max} - f_i}{f_{max} - f_{min}}, \tag{6}$$

where f_{max} and f_{min} are the maximum and minimum values, respectively, for feature f found in R. It is worth noting that, for some features, lower values represent better choices: e.g., lower duration or lower length values indicate fastest and shortest paths, respectively. For other features, instead, the opposite is true, like the number of attractions or the ratio of green sights present along the route. In this case, the normalization is inverted, as shown in Eq. (7):

$$\hat{f}_i = \frac{f_i - f_{min}}{f_{max} - f_{min}} \tag{7}$$

Eventually, $\hat{f}_i \in [0 - 1]$, where 0 and 1 denote the worst and the best value, respectively. Moreover, this normalization ensures that, for each feature, at least one route with value 0 and another with value 1 exist. For our chosen features, duration, length, crime, and accidents, we use the normalization shown in Eq. (6); whereas traffic, nature, and attractions use the process described in Eq. (7).

After normalization, all metrics have the same weight for a multi-criteria approach. Hence, the last step is to apply one of our proposed methods, called Hierarchical with Variable Tolerance (HVT) and Route to Vector (R2V) methods, as explained next.

5.1. Hierarchical with variable tolerance (HVT)

The Hierarchical with Variable Tolerance algorithm is presented as a flowchart in Fig. 3. It assumes that features are sorted following drivers' preferences. Each feature filters the best routes according to its individual metric and outputs them to the next feature in the pipeline. The algorithm runs until there is only one route left. After a round of feature analysis, if the final output reveals more than one route, the entire process restarts. We call this method hierarchical because the sooner the feature appears in the pipeline, the more important it gets.

The number of routes filtered from one feature to another must converge to a single one. Hence, we use a tolerance `tol` value previously defined. This value is used to define the subset of routes, which will be filtered out by the next feature. This subset is composed of the best route according to the current feature under analysis and all the others satisfying the same normalized feature tolerance. Note that after each feature filtering step, if only one route is left, the HVT algorithm stops. Thus, the tolerance value assigns more or less importance to the earlier features in the pipeline. If the tolerance is very rigorous, then the algorithm tends to stop earlier. On the other hand, if after a filtering round for all the features there is still more than one route left, the pipeline restarts with the remaining routes using an updated tolerance value equal to $\texttt{tol} - \epsilon$. In particular, $\epsilon = \frac{\texttt{tol}}{n}$, where $n \in \mathbb{N}$, is the algorithm iteration counter. Intuitively, we gradually reduce the tolerance to avoid early terminations without going through all features at least once. If there is no conclusion after going through the entire pipeline, this means that we can reduce the tolerance and check again.

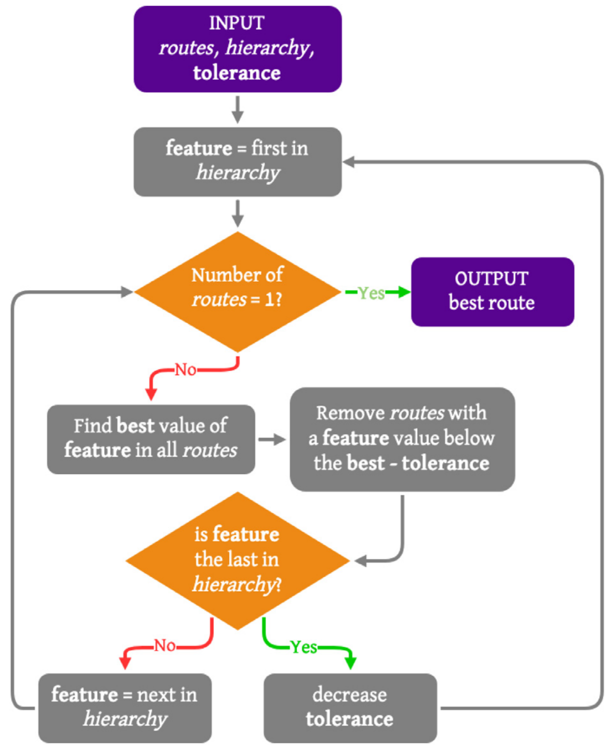

Fig. 3. HVT algorithm flowchart. The set of eligible routes are filtered out as each feature in the pipeline is individually analyzed. The sequence of features follows the drivers' preferences and the procedure ends when a single best route is found.

5.2. Route to vector (R2V)

The second route selection method considers each route as an n-dimensional vector, where n is the number of features available in the n-tuple. We call this the Route to Vector (R2V) method. Compared to the HVT method, R2V considers all the features at the same time and not as in a pipeline sorted according to drivers' preferences. With the R2V approach, the larger the vector magnitude, the better the route. This is because the better the route in a given feature, the higher the normalized value. The magnitude of a vector is computed according to its Euclidean distance from the origin coordinates at the n-dimensional space, considering the n-tuple of normalized features. Even though our definition assumes all features have the same importance, these can be weighted according to the drivers' will.

6. Multi-criteria models evaluation

In this section, we analyze the feature space beforehand, then we evaluate the proposed methods for multi-criteria route selection with respect to single criterion approaches. Finally, we highlight security gains and aspects that could be emphasized using our multi-criteria contributions.

The values presented in this section were gathered from three different sources. HERE Maps API provided route data from the random OD coordinate pairs, such as path coordinates, trip duration with and without traffic, and length. We use path coordinates to calculate the other features. Nature and attractions data were fetched from the Overpass API and calculated using, respectively, the intersection area and the point count methods. Crime and accident data were gathered from the UK Government database and were calculated using weighted point count. Crime computation used points inside the route region, whereas accidents used points closer to the path.

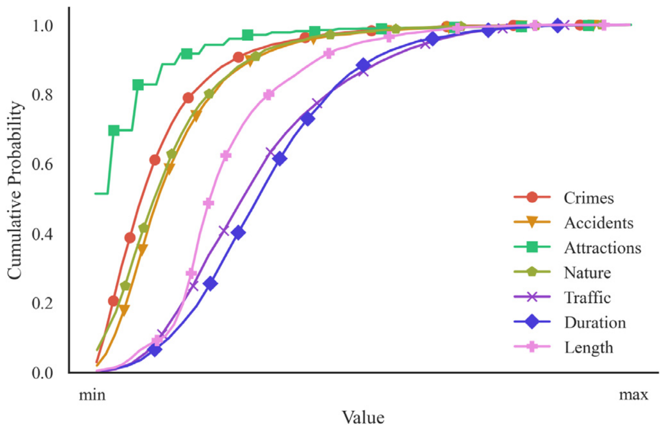

Fig. 4. Features cumulative distribution.

	Crime	Duration	Traffic	Nature	Attractions	Length
Duration	0.43					
Traffic	0.38	0.43				
Nature	0.13	-0.00	-0.01			
Attractions	-0.14	-0.10	-0.07	0.08		
Length	-0.08	0.27	-0.06	-0.29	-0.08	
Accident	0.20	0.22	0.11	0.05	-0.01	0.33

Fig. 5. Pearson correlation among features.

6.1. Feature distribution

Fig. 4 shows the cumulative distribution function for each feature. In order to be able to compare the distributions from different features, we decided to show the distribution according to the minimum and the maximum value found in the dataset. Around 50% of the routes collected had no tourist attractions whatsoever. We can also see that crimes, accidents and nature follow a very similar distribution, and that traffic is more sparsely distributed.

We also compute and show the Pearson correlation among features from the 600 OD sets in Fig. 5. The highest correlation magnitude is between duration and traffic, showing that traffic impacts duration; and duration and crime, showing that slower paths tend to be more dangerous, because faster streets have less pedestrians. Crime is also correlated with traffic, probably for the same reason.

Some other notable correlations regard length and nature, showing that paths with more nature between OD in London usually tale extra-urban roads making the path longer; length and accidents, because accidents are computed throughout the whole path, longer paths can go through more accidents; and length and duration, since longer paths tend to be more time consuming.

6.2. Route selection methods comparison

This section compares the performance of R2V and HVT with respect to single criterion approaches. The initial tolerance used

	Duration	Traffic	Length	Attractions	Nature	Crimes	Accidents
R2V	3%	13%	12%	-51%	-19%	14%	25%
R2V (safety)	8%	25%	11%	-56%	-22%	12%	13%
HVT (worker)	1%	11%	14%	-53%	-27%	16%	48%
Fastest (HERE)	0%	15%	13%	-55%	-28%	18%	47%
Fastest (Google)	2%	23%	10%	-58%	-29%	21%	39%
Least traffic	6%	0%	23%	-54%	-25%	19%	57%
Shortest	10%	37%	0%	-52%	-40%	50%	31%
HVT (tourist)	16%	37%	22%	-12%	-13%	50%	46%
Most attractions	20%	45%	19%	0%	-15%	63%	40%
Most natural	12%	27%	27%	-51%	0%	23%	48%
HVT (safety)	6%	18%	17%	-59%	-22%	3%	42%
Least crime	7%	20%	22%	-61%	-23%	0%	57%
Least accident	10%	28%	8%	-51%	-27%	27%	0%

Fig. 6. Percent deviation from a known standard in relation to the best classification for each feature.

for the HVT method is 0.2, with a decrement step of 0.01. We consider three different examples of features order having various driving scenarios in mind:

- **HVT(safety).** Features are in the following order: 1. crime, 2. accident, 3. duration, 4. traffic, 5. nature, 6. attractions, 7. length. This is the order preferred by drivers who want to minimize safety risks when driving in unsafe or unknown places.
- **HVT(tourist).** The following order privileges paths for well-being and tourism, regardless time constraints: 1. attractions, 2. nature, 3. crime, 4. traffic, 5. duration, 6. accident, 7. length.
- **HVT(worker).** This order takes into account time constraints above all: 1. duration, 2. traffic, 3. crime, 4. nature, 5. attractions, 6. length, 7. accident. It is intended for users who drive back and forth to work every day and they want primarily to avoid traffic to arrive home as fast as possible.

The Fastest (HERE) row shows the route selected by HERE, while the Fastest (Google) row represents the route selected by Google Maps, but with the duration calculated by the HERE Maps API.

We show the comparison between the proposed methods and the single-criterion ones in Fig. 6. The rows of the matrix M represent the different route selection and classification methods (m), while the columns represent the features (f) to be evaluated. As metric, we use the Percent Deviation From a Known Standard (PDFKS), with the best average value for each feature as the known standard ($std[f]$), due to the different value ranges of the features. The PDFKS is calculated as follows:

$$M_{mf} = \frac{m[f] - std[f]}{std[f]} \times 100\%, \tag{8}$$

where m is the classification method, $m[f]$ is the average value from feature f for m, and $std[f]$ is the best average value for feature f among all classification methods. For instance, the first cell shows that the R2V method finds routes that have, on average, a

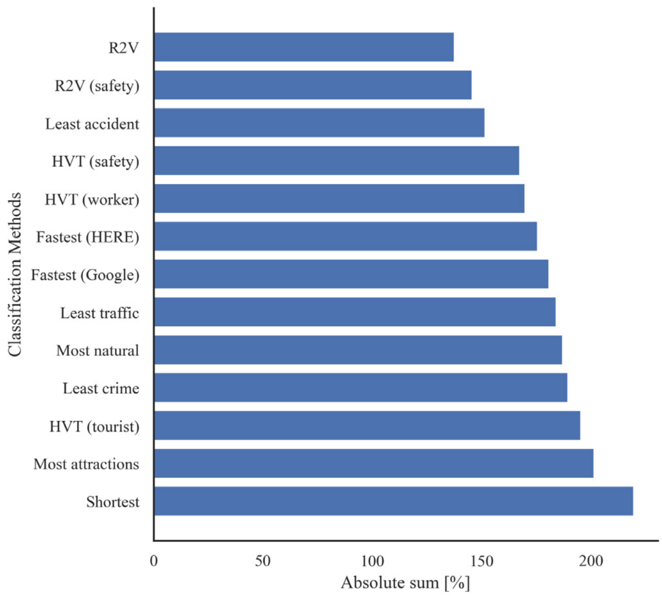

Fig. 7. Absolute PDFKS sum of all classification methods.

3% increase in duration when compared to the Fastest route (as we compare to the duration column).

It is important to note that some features are better with low values, such as duration, crime, traffic, accidents and length; and others are better with high values, such as nature and tourist attractions. This results in some columns (Nature and Attractions) having negative percentages. Thus, for the PDFKS, percentage values close to 0% represent better route choices, and a low overall absolute sum of all percentages represent better classification methods (considering that all features are equally important), since we are only interested in the magnitude of the values. This sum is represented in Fig. 7.

We can note that the Shortest path method finds the best path if we consider duration and length only, with a sum of 10% of both features. On the other hand, this method comes out with one of the most dangerous routes in relation to criminality (50%) and the one with least nature (−40%). The Fastest (HERE) path is one of the best single criteria methods, with an overall sum of 176%, behind the Least accident method (152%). The routes selected by this method tends to have a high accidents value (47%), low attractions (−55%), and low nature (−28%) nearby. Compared to the other methods, the Fastest (HERE) path also performs well on traffic (15%).

The HVT (worker) method, with duration as the first evaluated feature, shows improvements for crime, nature, attractions and a more significant improvement in traffic when compared with the fastest path, which performs slightly better in duration, accidents and length. The HVT (safety) method presents a similar performance to the least crime method, but with improved results in all other features. The HVT (tourist), instead, gets the second highest average value for attractions nature, but lacks good performance for the other metrics. All HVT methods have a better overall performance when compared with the first feature from its hierarchy.

Broadly speaking, the HVT methods tend to select paths similar to the ones selected by the first evaluated feature of its order, but with more personalized and better overall results. For instance, the HVT (tourist) goes through more nature than the most attractions method, the HVT (safety) has fewer accidents than the least crime method, and the HVT (worker) has less traffic than the fastest method.

The R2V performs better, in general, than the fastest one which remains the most common route selection method used by navigation apps. In fact, besides duration, all the other features present better values, and the difference in travel time is just of 3%. This approach has thus the best overall performance, with an overall sum of 138%, among all the compared methods. It presents the third best values for accidents and traffic.

Concerning travel times, our proposed methods R2V and HVT (worker) have very low additional duration values (3 and 1%, respectively) compared with the duration time found in the Fastest(HERE) method.

6.3. Safety improvements

In this Section, we evaluate safety related gains for seven route selection methods: Fastest (HERE), Fastest (Google), Least crime, Least accidents, R2V, R2V safety, HVT worker, and HVT safety. We rely on four metrics: crime severity, number of vehicle related crimes (events falling within the vehicle crime category of Table 1), accident severity, and number of accident casualties. Each metric is computed as the sum of such events occurring along the route. It is worth noting that the Fastest ones, the Least crime, and the Least accidents methods rely on a single criterion.

Fig. 8 shows, for each route selection method, the PDFKS with respect to the best value for each metric. In particular, the least crime method presents the best result for vehicle crime and criminality metrics, while the least accidents approach has the best result for accident casualties and accident severity metrics. As an example, routes selected by the R2V method present 8% and 12% more vehicle crimes and criminality acts, respectively, compared with routes selected by the least crime method. On the other hand, the least crime method selects routes with around 50% more of accident severity and 60% more accident casualties than the least accidents method.

We have also evaluated an additional R2V route selection, called R2V (safety), which relies on customized weights for the n-tuple features describing the route to reduce safety risks. We compute using the following weights: 10 for crime and accidents, 2 for duration and traffic, and 1 for all the remaining metrics. Although not excelling in a specific metric, all the multi-criteria methods proposed have a better performance over the four safety metrics considered than the fastest method. Moreover, the least crime method has the worst accident casualties metric, and the least accident method shows also high crime metrics. These results reinforce our intuition that a multi-criteria selection method reduces the overall safety risk.

The safety improvements after customizing the proposed methods are noticeable, especially for the R2V method. The R2V method has more than 10% reduction in the casualties metric, further improved by reductions in the other metrics. The hierarchy shift from the worker order to the safety order also shows an improvement, with emphasis on criminality, with reduction of also more than 10%. This proves that multi-criteria methods is flexible enough to be tuned and hence produce results taking into account drivers' preferences.

7. Conclusion

In this paper, we focus on route selection improvements considering alternative paths beyond the fastest one. In particular, we stress that the fastest (or even the shortest) path proposed by car navigation systems may present a higher level of safety risks, while slightly slower (or longer), well-selected paths, could reduce such risks. Thus, we propose and analyze two multi-criteria route selection methods, HVT (Hierarchical with Variable Tolerance) and R2V (Route to Vector), and four techniques: point

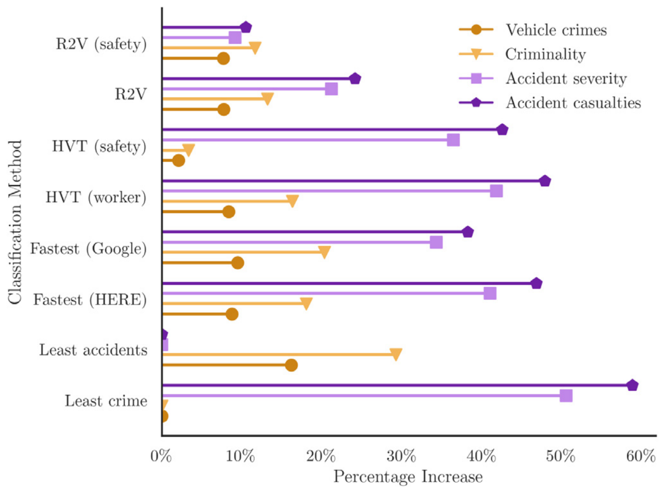

Fig. 8. Safety PDFKS comparison.

count, weighted point count, intersection area and scalars, to measure the different features.

To demonstrate our claim, we introduce safety and pleasant features for paths beyond length and duration. The techniques used to gather information about these features were: point count, for attractions; weighted point count, for crime and accidents; intersection area, for nature; and scalars, for duration, length and traffic.

The R2V method describes routes as n-tuples of features and ranks them according to their magnitude. It chooses better routes than the traditional fastest approach, without significant increase in duration. The customization of this method, through the use of weighted features, can give more personalized results according to the users' preferences. The HVT method directly ranks paths upon drivers' preferences. This method presents just a little decrease on its first evaluated feature when compared to the single criterion approach on that specific feature. In addition, it presents much better results regarding the other features.

We built a dataset to evaluate our classification methods and quantify the improvements in route features. These improvements are indicators that relying on a single criterion neglects other important parameters, and that multi-criteria approaches should be pursued for a better driving experience.

In the future, other possible features can be introduced, such as air pollution, noise and road conditions. Also, a personalized machine learning can be used in the classification methods to set the weights and hierarchy.

CRediT authorship contribution statement

Leonardo Solé: Software, Investigation, Data curation, Writing - original draft, Visualization. **Matteo Sammarco:** Conceptualization, Validation, Writing - review & editing, Project administration. **Marcin Detyniecki:** Resources, Funding acquisition. **Miguel Elias M. Campista:** Methodology, Formal analysis, Writing - review & editing, Supervision, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This paper was partially supported by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001. It was also supported by CNPq, Brazil, FAPERJ, Brazil Grants E-26/211.144/2019 and E-26/202.689/2018, and FAPESP, Brazil Grant 15/24494-8.

References

[1] Mordor Intelligence, Navigation system market - growth, trends, and forecasts (2020 - 2025), 2019, URL https://www.mordorintelligence.com/industry-reports/navigation-system-market.
[2] Z. Yang, M. Zhu, A dynamic prediction model of real-time link travel time based on traffic big data, in: 2019 International Conference on Intelligent Transportation, Big Data Smart City (ICITBS), 2019, pp. 330–333, http://dx.doi.org/10.1109/ICITBS.2019.00087.
[3] M. Kondo, J. Fluehr, T. McKeon, C. Branas, Urban green space and its impact on human health, Int. J. Environ. Res. Public Health 15 (3) (2018) 445, http://dx.doi.org/10.3390/ijerph15030445.
[4] A. Tobías, A. Recio, J. Díaz, C. Linares, Health impact assessment of traffic noise in Madrid (Spain), Environ. Res. 137 (2015) 136–140, http://dx.doi.org/10.1016/j.envres.2014.12.011, URL http://www.sciencedirect.com/science/article/pii/S0013935114004629.
[5] T. Zhang, A.H. Chan, The association between driving anger and driving outcomes: A meta-analysis of evidence from the past twenty years, Accid. Anal. Prev. 90 (2016) 50–62, http://dx.doi.org/10.1016/j.aap.2016.02.009, URL http://www.sciencedirect.com/science/article/pii/S0001457516300422.
[6] K.L. Wolf, S.T. Lam, J.K. McKeen, G.R. Richardson, M. van den Bosch, A.C. Bardekjian, Urban trees and human health: A scoping review, Int. J. Environ. Res. Public Health 17 (12) (2020) 4371, http://dx.doi.org/10.3390/ijerph17124371.
[7] B. Grinde, G. Patil, Biophilia: Does visual contact with nature impact on health and well-being? Int. J. Environ. Res. Public Health 6 (9) (2009) 2332–2343, http://dx.doi.org/10.3390/ijerph6092332.
[8] G.C. Gee, D.T. Takeuchi, Traffic stress, vehicular burden and well-being: A multilevel analysis, Soc. Sci. Med. 59 (2) (2004) 405–414, http://dx.doi.org/10.1016/j.socscimed.2003.10.027, URL http://www.sciencedirect.com/science/article/pii/S0277953603005677.
[9] S. Bachoo, A. Bhagwanjee, K. Govender, The influence of anger, impulsivity, sensation seeking and driver attitudes on risky driving behaviour among post-graduate university students in Durban, South Africa, Accid. Anal. Prev. 55 (2013) 67–76, http://dx.doi.org/10.1016/j.aap.2013.02.021, URL http://www.sciencedirect.com/science/article/pii/S0001457513000626.
[10] L.S. Rodrigues, M. Sammarco, M. Detyniecki, M.E.M. Campista, CRAWDAD dataset ufrj/london-trajectories (v. 2021-03-22), 2021, http://dx.doi.org/10.15783/ejej-de94, Downloaded from https://crawdad.org/ufrj/london-trajectories/2021-03-22.
[11] C.H. Carlos, D.S. Medeiros, M.E.M. Campista, Towards participatory sensing of regions of interest with adaptive sampling rate, Veh. Commun. 25 (2020) http://dx.doi.org/10.1016/j.vehcom.2020.100254.
[12] H. Huang, D. Bucher, J. Kissling, R. Weibel, R. Martin, Multimodal route planning with public transport and carpooling, IEEE Trans. Intell. Transp. Syst. PP (2018) 1–13, http://dx.doi.org/10.1109/TITS.2018.2876570.
[13] B. Ludwig, B. Zenker, J. Schrader, Recommendation of personalized routes with public transport connections, in: D. Tavangarian, T. Kirste, D. Timmermann, U. Lucke, D. Versick (Eds.), Intelligent Interactive Assistance and Mobile Multimedia Computing, Springer Berlin Heidelberg, Berlin, Heidelberg, 2009, pp. 97–107.
[14] F. Huang, J. Xu, J. Weng, Multi-task travel route planning with a flexible deep learning framework, IEEE Trans. Intell. Transp. Syst. (2020) 1–12, http://dx.doi.org/10.1109/tits.2020.2987645.
[15] D. Quercia, R. Schifanella, L.M. Aiello, The shortest path to happiness: Recommending beautiful, quiet, and happy routes in the city, in: Proceedings of the 25th ACM Conference on Hypertext and Social Media, in: HT '14, Association for Computing Machinery, New York, NY, USA, 2014, pp. 116–125, http://dx.doi.org/10.1145/2631775.2631799.
[16] J. Derek, M. Sikora, Bicycle route planning using multiple criteria GIS analysis, in: 2019 27th International Conference on Software, Telecommunications and Computer Networks, SoftCOM 2019, 2019, http://dx.doi.org/10.23919/SOFTCOM.2019.8903800.
[17] J. Xu, Y. Zhang, C. Xing, An effective selection method for vehicle alternative route under traffic congestion, in: International Conference on Communication Technology Proceedings, ICCT, Vol. 2019-Octob, IEEE, 2019, pp. 494–499, http://dx.doi.org/10.1109/ICCT.2018.8600055.
[18] HERE Technologies, HERE maps API, 2020, [Online; accessed 20-July-2020]. URL https://developer.here.com.
[19] UK Government, Road traffic statistics, 2020, [Online; accessed 20-July-2020]. URL https://roadtraffic.dft.gov.uk.
[20] UK Police, UK police open data, 2020, [Online; accessed 20-July-2020]. URL https://data.police.uk.

[21] M. Bangs, Research outputs: developing a Crime Severity Score for England and Wales using data on crimes recorded by the police, 2016, [Online; accessed 20-July-2020]. URL https://www.ons.gov.uk/releases/researchoutputsdevelopingacrimeseverityscoreforenglandandwalesusingdataoncrimesrecordedbythepolice.

[22] OpenStreetMap Wiki, Overpass API — OpenStreetMap Wiki, 2020, [Online; accessed 20-July-2020]. URL https://wiki.openstreetmap.org/w/index.php?title=Overpass_API&oldid=2015229.

Leonardo Solé Rodrigues received his Social Communications degree from the Superior School of Marketing and Propaganda (ESPM), Brazil, in 2011, and is currently pursuing his Electronic and Computer Engineering degree from the Federal University of Rio de Janeiro (UFRJ). He is part of the Group of Teleinformatics and Automation Group (GTA) from UFRJ, working with computer networking and machine learning.

Matteo Sammarco received his the B.Sc. and M.Sc. degrees in Computer Engineering from Università degli Studi di Napoli Federico II, Italy, in 2008 and 2010, respectively. In 2014 he received the Ph.D. in Computer Science from Sorbonne Université, France. He has worked as researcher at Télécom ParisTech and Laboratory of Information, Networking and Communication Sciences, both in France. Currently, he is research data scientist at Axa Group Operations, mainly working on machine learning applied to Safety, Smart Mobility and IoT.

Marcin Detyniecki studied mathematics, physics and computer science at the Sorbonne Université Paris. In 2000 he obtained his Ph.D. in Artificial Intelligence from the same university. Between 2001 and 2014, he was a research scientist of the French National Center for Scientific Research (CNRS). He has been researcher at the University of California at Berkeley and at Carnegie Mellon University (CMU) and visiting researcher at the University of Florence and at British Telecom Research labs.

He is currently Group Chief Data Scientist and Global Head of Research at AXA Group Operations. His work focuses on Machine Leaning, Artificial Intelligence, Computational Intelligence, Multimedia Retrieval, Fair and Transparent AI.

Miguel Elias M. Campista received a D.Sc. in Electrical Engineering from the Federal University of Rio de Janeiro (UFRJ), in 2008. He is an associate professor in the Electronic and Computer Engineering Department at Poli/UFRJ and a full professor in the Electrical Engineering Program (PEE) at COPPE/UFRJ. Prof. Campista is an Affiliate Member of the Brazilian Academy of Science (ABC), is on the board of directors of the Brazilian Computer Networking Lab, is an Editor of the Annals of Telecommunications Journal, and an IEEE senior member. His research interests are on data and network science.