Understanding Cycling Trip Purpose and Route Choice Using GPS Traces and Open Data

SURAJ NAIR*, University of Maryland, USA
KIRAN JAVKAR*, University of Maryland, USA
JIAHUI WU, University of Maryland, USA
VANESSA FRIAS-MARTINEZ, University of Maryland, USA

Many mobile applications such as Strava or Mapmyride allow cyclists to collect detailed GPS traces of their trips for health or route sharing purposes. However, cycling GPS traces also have a lot of potential from an urban planning perspective. In this paper, we focus on two important issues to characterize urban cyclist behavior: trip purpose and route choice. Cycling trip purpose has been typically analyzed using survey data. Here, we present a method to automatically infer the purpose of a cycling trip using cyclists' personal data, GPS traces and a variety of built-in and social environment features extracted from open datasets characterizing the streets cycled. We evaluate the proposed method using GPS traces from over 7,000 cycling routes in the city of Philadelphia and report F1 scores of up to 86% when four trip purposes are considered. On the other hand, we also present a novel statistical method to identify the role that certain variables characterizing the built-in and social environment play in the selection of a specific cycling route. Our results show that cyclists in Philadelphia tend to favor routes with green areas, safety and centrality.

CCS Concepts: • **Human-centered computing → Empirical studies in ubiquitous and mobile computing**; • **Computing methodologies → Supervised learning by classification**;

Additional Key Words and Phrases: trip purpose classification, statistical analysis of route choice, spatio-temporal cycling traces

ACM Reference Format:
Suraj Nair, Kiran Javkar, Jiahui Wu, and Vanessa Frias-Martinez. 2019. Understanding Cycling Trip Purpose and Route Choice Using GPS Traces and Open Data. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 3, 1, Article 20 (March 2019), 26 pages. https://doi.org/10.1145/3314407

1 INTRODUCTION

There exists a plethora of mobile applications that allow cyclists to collect detailed GPS traces of their trips such as Strava, Mapmyride, Cyclemeter or MyTracks [22, 44, 49, 65]. Cyclists typically collect such traces either for health purposes, to track the amount of exercise done by an individual; or for informational purposes, to share cycling experiences with other fellow cyclists willing to explore new routes. However, cycling GPS traces also

*Both authors contributed equally to the paper.

Authors' addresses: Suraj Nair, Department of Computer Science, University of Maryland, College Park, Maryland, USA, srnair@cs.umd.edu; Kiran Javkar, Department of Computer Science, University of Maryland, College Park, Maryland, USA, kjavkar@cs.umd.edu; Jiahui Wu, College of Information Studies, University of Maryland, College Park, Maryland, USA, jeffwu@umd.edu; Vanessa Frias-Martinez, College of Information Studies, University of Maryland, College Park, Maryland, USA, vfrias@umd.edu.

2474-9567/2019/3-ART20 $15.00
https://doi.org/10.1145/3314407

have a lot of potential from an urban planning perspective so as to help cities better understand the ways in which cyclists use the streets on a daily basis. Furthermore, given the large growth in the number of citizens cycling for commuting and non-commuting purposes, understanding cycling behavior becomes highly critical for cities willing to improve the urban infrastructure so as to maximize cyclist satisfaction [31, 43, 47, 72].

In this paper, we focus on two important issues characterizing urban cyclist behavior: trip purpose and route choice. Trip purpose refers to the classification of cycling trips by its main objective *e.g.,* commuting, exercise or shopping, among others. Understanding trip purpose allows decision makers to identify different urban areas as major hubs for specific purposes and to develop appropriate infrastructure accordingly. For example, if an urban area mostly sees commuting trips, urban planners could focus on providing long-term, secure bicycle parking options to commuters; while if another region is mostly visited by cyclists exercising, it might be relevant to provide dedicated bike lanes, rather than lanes shared with traffic, so as to allow citizens to relax and enjoy their exercise routines. Urban planners have typically analyzed cycling trip purposes through expensive surveys which, as a result, are only run every certain number of years. Although GPS traces have been used to automatically infer car trip purpose [24], there are no solutions for cycling trip purpose inference, which can be much more challenging due to the less constrained nature of these trips as cyclists enjoy much more freedom than vehicles in their route choices. In fact, cyclists can often times choose to cycle through bike-only park trails instead of roads; they can also use contra-flow lanes, which allow cyclists to travel both with and against the traffic flow [56]; they can decide to walk their bikes on the sidewalk or across bridges while moving against the traffic flow [10]; or even cycle in the wrong direction, which is not legal, but happens with certain frequency [9]. This plethora of route choices makes the cycling trip purpose inference much more challenging. In this paper, we explore the applicability of existing car trip purpose prediction techniques to cycling, explore its limitations, and propose a novel approach that improves existing solutions by combining XGBoost with features characterizing the built-in and social environment where the trip takes place, cyclists' personal data and trip information extracted from the GPS traces collected via cycling tracking tools. Since the main objective is to provide affordable inference methods that cities and urban planners can run with high frequency, we propose to use contextual information about the environment exclusively extracted from open datasets such as open data city portals or Open Street Maps.

Route choice, on the other hand, refers to understanding the role that certain variables characterizing the built-in and social environment play in the selection of a specific cycling route. For example, cyclists might generally prefer streets with bike lanes or no slope; or might express a preference for streets in green areas when they are exercising. Understanding route choice alone, or in combination with trip purpose would help urban planners in the identification of potential improvements to the urban bicycle infrastructure. Route choice has been largely studied by researchers in transportation analysis [6]. The common approach is to generate the *choice set* of routes that cyclists can possibly take, and to compare these against the actual routes cycled so as to identify the route features that are favored by cyclists. Since it is not trivial to generate all possible routes, specially given the fact that cyclists enjoy much more freedom than drivers in their choices, choice set generation algorithms typically select the subsets of all the possible routes that make more sense, using different types of cyclist behavioral assumptions and hypotheses. For example, some scholars create the choice set assuming that cyclists always prefer the shortest or fastest routes [5], while others accept that cyclists will always choose routes with the lowest traffic [3]. Although partially correct, none of the behavioral assumptions represent all the decisions behind route choice in cyclists. In fact, related work has shown that different route choice set generation algorithms directly affect the behavioral findings observed, reaching sometimes contradictory conclusions [3].

To address this issue, we propose a novel method that uses Google's cycling directions to generate the route choice set; followed by statistical methods to analyze the similarities and differences between the actual GPS-recorded routes by cyclists and Google's suggested routes. Our main hypothesis is based on the fact that Google's cycling directions uses a mash-up of data including publicly available bike maps as well as user-generated routes from cyclists worldwide, which implicitly incorporate information about cyclists route choices *i.e.,* there are no assumptions made as of the types of routes chosen, but rather, these represent actual routes taken by cyclists with certain frequency [26, 33]. The statistical methods we propose will use built-in and social environment features extracted from open datasets, rather than from proprietary data as traditionally done in the field of transportation analysis, making route choice analyses more accessible to cities with low resources.

This paper advances the state of the art along several directions:

- The design of an automatic method to infer cycling trip purpose using cyclists' personal data, cyclist-generated GPS data and contextual information about the environment extracted, exclusively, from open datasets.
- The design of a method to generate a route choice set without any cyclist behavioral assumption, and a statistical method to analyze route choice using open datasets.
- The evaluation of these methods using a open dataset with over 7,000 GPS traces from cyclists in Philadelphia collected using the Cycle Philly mobile application [19].

The rest of the paper is organized as follows. Section 2 covers related work. Sections 3 and 4 present a description and evaluation of the methods for trip purpose inference and route choice; Section 5 describes how the proposed methods could be embedded within a decision support tool for decision makers; and, finally, sections 6 and 7 cover the discussion and main conclusions.

2 RELATED WORK

2.1 Trip Purpose Inference

There exists a lot of work focused on activity inference based on geolocated traces *i.e.,* to identify whether a person is walking, cycling or running using traces collected via smartphones (GPS) [17, 25, 36]; or to extract travel/transportation mode *i.e.,* whether a person is traveling by bus, train, car or bicycle using smartphone GPS data or cell phone CDR data [60, 69, 74]. However, work on trip purpose inference is mostly focused on car trips, not cycling. One of the seminal papers, by Wolf et al. [71], focused on trip purpose inference using GPS data collected from cars. The authors inferred trip purpose by manually building a database of land uses, places and trip purposes, and by manually identifying the exact characteristics of the destination location (address and land use) and assigning a trip purpose to it. Later work, explored more automatic inference approaches using GPS data collected from cars, and typically combined with information extracted from travel logs [32]. Generally, most inferences use two sets of features: spatio-temporal information of the destination, or origin, or both [15, 27, 41, 51, 62, 71]; or a combination of spatio-temporal information of destination and/or origin with personal information about the driver [24, 28, 29, 40, 46, 48]. The spatio-temporal information of the destination or the origin typically includes day and time of the trip, type of land use or POIs (points of interest) at the origin or destination, duration of the trip, or distance traveled; while personal information refers to demographic and socio-economic characteristics of the driver. On the other hand, three types of inference approaches have been used in the literature:rule-based methods that match the GPS information with a series of predefined heuristic rules to infer trip purpose [11, 54, 64, 71]; probabilistic methods that estimate the probability for each trip purpose and select as final the one with the highest value [2, 15]; or machine learning approaches that use classification, regression trees or discriminant analysis to predict the purpose of the trip [24, 46]. Nevertheless, to the best of our knowledge, there are no papers that focus on trip purpose inference for cycling, which can be much more challenging due to the less constrained nature of these trips as cyclists enjoy much more freedom than cars in their route choices: bike-only trails through urban parks, contra-flow lanes, walking the bike on sidewalks or bridges, or cycling against traffic. In this paper, we explore the applicability of existing car trip purpose prediction techniques to infer cycling trip purpose, explore its limitations, and extend the state of the art by developing an inference method to determine the purpose of a cycling trip given cyclists' personal data, a set of GPS traces collected via mobile applications and characteristics of the built-in and social environment - available via open datasets - throughout the cycling trip.Since we envision city halls and urban planners using these GPS traces, and to avoid privacy breaches, our approach works with trips whose origin and destination points have been obfuscated to the closest intersection. The inference method we propose will work with obfuscated traces rather than the fully disclosed origin-destination traces.

2.2 Route Choice

The generation of the choice set of cycling routes has been implemented using multiple approaches. The most straight forward approach is the use of K-shortest paths to minimize the generalized path costs [67]. However these algorithms assume implicitly cyclist awareness of all the link attributes, which is highly unrealistic. Ben-Akiva *et al.* [4] proposed an approach that generated possible paths based on different optimal criteria that cyclists might pursue including shortest route, shortest travel time or lowest slope. Nevertheless, this approach assumes that travelers might have different objective functions, which has not been proved in the literature either. Simulation methods based on Monte Carlo approaches have also been proposed to take into account that travelers might erroneously perceive link attributes [12, 14]. Although all these assumptions and hypothesis are correct, they are never complete, since that is precisely the specific route choice problem we are trying to solve.

On the other hand, the comparison between cycled routes and the choice set has been carried out in the literature using a variety of discrete choice analysis methods including C-Logit [13], Path Size Logit [5], PCL model [18], CNL [68], GNL [70] or multinomial logit path (MLP) [23], all of which focus on the identification of specific built-in or social environment features that are statistically significant from a route choice perspective. All these choice analysis methods assume the existence of distinct, separable and mutually exclusive alternatives, condition that does not hold when using routes from Google's cycling directions which often times overlap with each other for certain subsets of streets. Additionally, they model cycling choices as individual decisions, without taking into account that cycling decisions might be influenced by what other cyclists are doing on the road and by their route choices as well. Although corrective methods have been proposed [59], eliminating just the overlapping routes would heavily limit the amount of options available since Google's cycling directions outputs up to three routes. To overcome all these limitations, we propose an approach that combines (a) Google's cycling directions to extract a route choice set that incorporates a complete overview of different cyclist behaviors, choices and influences, without the need to make any behavioral assumptions; and (b) a reproducible and simple statistical approach to analyze the role that the environment features play in cyclists' route choice, when it is known that routes might heavily overlap with each other.

3 TRIP PURPOSE INFERENCE

In this section, we design and evaluate a method to infer the purpose of a cycling trip using cyclists' personal data, features extracted from the trip's GPS traces and characteristics of the built-in and social environment visited during the trip. We assume that the GPS traces are collected using one of the multiple existing cycling tracking tools that ask cyclists to provide personal information upon installation (such as cycling experience or demographic data) and that allow them to collect cycling traces from their trips [19, 61]. For evaluation purposes, we will use a dataset that additionally contains trip purpose information for each set of GPS traces, labeled as commuting, errand, exercise, school, shopping or social, among others. However, we aim to design a trip purpose inference method that will only require cyclists to share their GPS traces without the need to provide any trip purpose labels. Additionally, cycling tracking tools typically allow cyclists to define certain privacy settings, including the creation of privacy zones which obfuscate the exact origin and destination points. In this paper, we focus on the design of a privacy-preserving trip purpose inference method which will not use the exact origin and destination locations, but rather their obfuscated representations defined as the closest road intersection to the actual location. Although this approach makes trip inference more challenging, it provides a balanced solution between the collection of private cyclist data and its use by city planners for urban planning purposes.

3.1 Baseline

Previous work has already shown that the day and time of the trip as well as the type of land use at the destination can statistically differentiate with high significance different car-based trip purposes [15, 41, 51, 71, 73].Thus, we define a basic baseline (*BS*) that exclusively uses spatio-temporal information of the cycling trip including day, hour and month of the trip, its duration and the distance traveled, together with the points of interest (POI) at the obfuscated destination point to infer trip purpose. This information can be easily retrieved from the GPS traces collected, except for the POI, which we extract from Open Street Maps using the following 10 categories: residential, schools and universities, outdoors and recreation, retail shops, art, professional service, food, nightlife spots, cycling facilities and transportation facilities [34]. To retrieve the POI at the obfuscated destination points, which are defined as the closest road intersection to the actual location, we compute the number of POI per category present across all the streets involved in that intersection. On the other hand, other approaches that infer car trip purpose have successfully combined spatio-temporal origin and destination information with personal driver data [24, 40, 46, 48]. As a result, we will consider a second baseline that combines *BS* with personal data from the cyclist ($BS + P$).

Nevertheless, the two baselines described do not take into account information about the types of streets cycled during the trip. Given the route choice freedom that cyclists enjoy, we hypothesize that these features could be highly informative of the trip purpose.In fact, depending on the cyclist and/or the purpose of the trip, cyclists might choose routes with different built-in or social environment features. For example, an experienced cyclist might prefer a direct route without bike lanes when commuting to work, while a reluctant cyclist might favor secondary roads with bike lanes. Thus, we aim to design a trip purpose inference method that uses the spatio-temporal features and cyclists' personal data described in the baselines as well as built-in and social environment features characterizing each street cycled, as predictors of the purpose of the trip. To achieve this goal, we propose four consecutive steps as shown in Figure 1 (top). First, we retrieve the streets associated to the collected GPS traces; second, we retrieve the built-in and social environment features that characterize each street cycled; third, we design a novel method that uses personal cyclist data, spatio-temporal information of the trip and street features to infer trip purpose; and fourth, we evaluate the method. Next, we describe each step in detail.

3.2 Retrieving Street Segments

Streets can be long and might traverse areas with largely diverse built-in and social environment features. To carry out a more granular characterization of the streets cycled, we propose to work with street segments instead, defined as the road between any two street intersections. Retrieving the street segments associated to the GPS traces of a cycling trip is not straightforward since GPS sensors have errors, and more so in urban environments where when surrounded by tall buildings the GPS might lose signal or record a location quite far away from the actual visited location. As a result, we retrieve the list of street segments cycled using Mapbox's Map Matching API, which snaps fuzzy, inaccurate GPS traces to actual segments in the road network [42]. Internally, Mapbox uses the map-matching algorithm by Newson and Krumm, based on Hidden Markov Models (HMM) that find the most likely street segment in the network that is represented by the collected GPS location [50].

3.3 Street Segment Features

Once each cycling route is represented as a set of street segments, we characterize each segment with a set of built-in and social environment features that we hypothesize could be predictive of the trip purpose. Since our

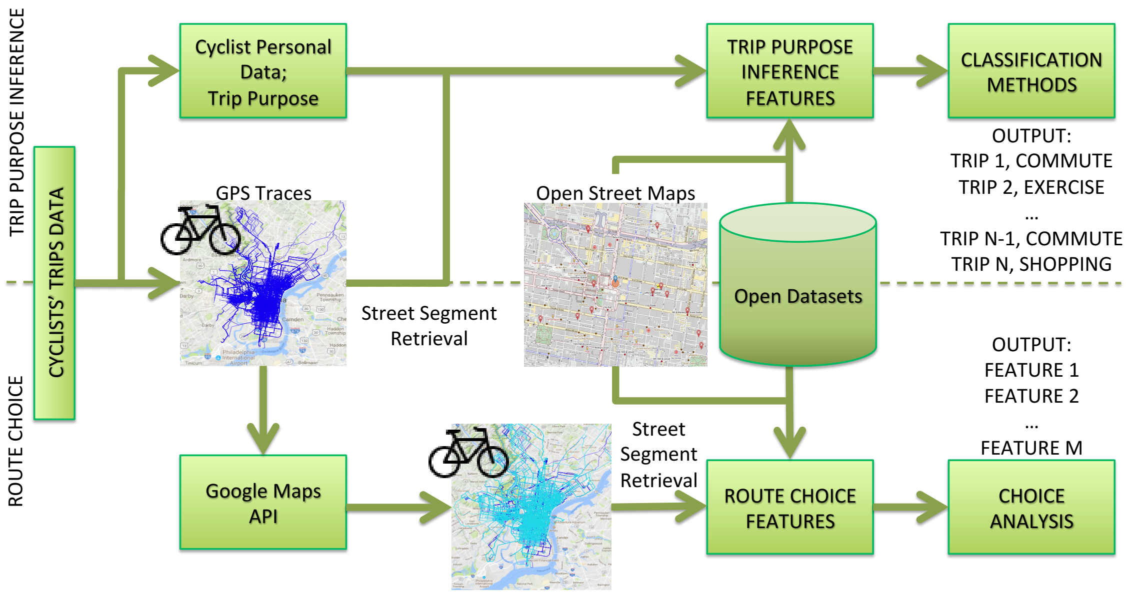
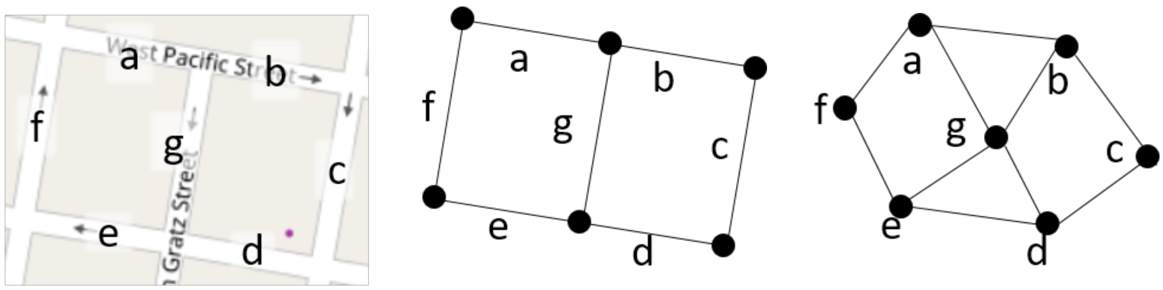


Fig. 1. Main components of the Trip Purpose Inference (top half figure) and Route Choice methods (bottom half figure) proposed. Trip purpose is inferred using cyclist personal data and social and built-in environment features of the route. Route choice is evaluated with Google Maps suggested routes and the same social and built-in features.

objective is to design affordable and accessible tools for city planners, all the proposed features can be typically retrieved from open data portals or from Open Street Maps. While over 2600 cities worldwide offer access to open data portals with city information [52], Open Street Maps is available for over 4 million small- to mid-sized cities [34], which guarantees the possibility to replicate these studies across other cities and countries.

We propose to use the following built-in environment features: (a) **road network** features of each street segment including the average slope as well as the minimum and maximum segment's slope (to account for cases where the segment might have multiple peaks), computed using elevation data from several segment points extracted using USGS free point query service; these features might reflect how cyclists favor or avoid routes with slopes depending on their trip purpose [66]; (b) **graph-based characterization** of the street segment in terms of centrality measures that quantify the importance of the segment in the overall road network *i.e.,* whether it is a central segment that is typically cycled through to go between any two points in the city, or more of an outlier segment, which could reflect preference of direct or indirect routes depending on the purpose of the trip. We use the SNAP package [38] and the road network extracted from Open Street Maps to evaluate various centrality measures such as degree, betweenness or page rank, among others, considering the road network of the city both as an undirected and directed graph (taking into account the direction of the traffic flow). Additionally, we will evaluate both primal and dual road network approaches that consider either each segment as an edge and each intersection as a node, or vice versa (Figure 2 shows an example) [57, 58]; (c) **presence of bike facilities** in the street segment, extracted using Open Street Maps as well as open datasets from cities that provide the shapefiles for the bike lanes. These features might be favored by certain types of trip purposes *e.g.,* cyclists exercising favor



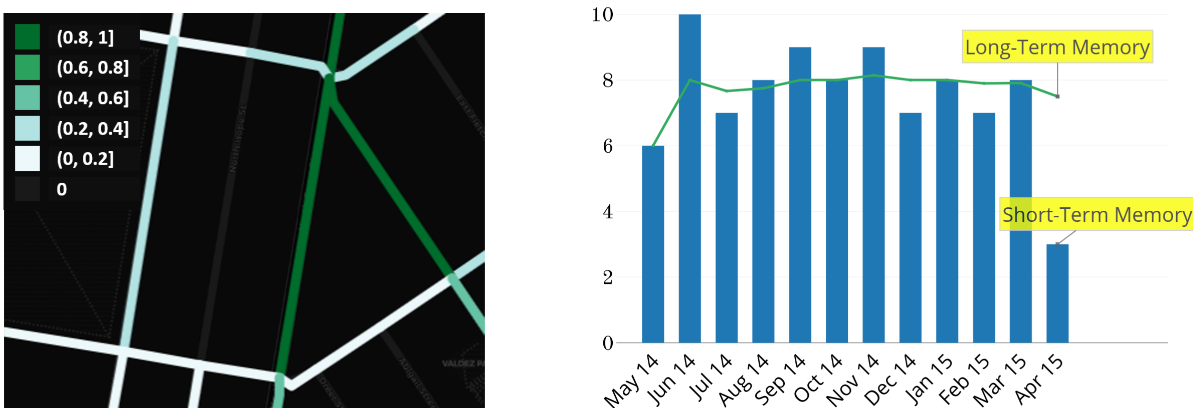
(a) Road Network (b) Primal Graph (c) Dual Graph

Fig. 2. Transformation of the road network defined by two blocks in Philadelphia into its undirected primal and dual graphs. The road network consists of street segments *a to g*. Given the graphs, graph-based features that characterize the centrality of the street segments are computed using the SNAP package.

the safety provided by separated bike lanes; and (d) **presence of different types of POI** in the street segment, extracted from Open Street Maps, which could be indicative of trip purpose.

On the other hand, we also work with social environment features that characterize human interactions with the built-in environment over time. These features, typically provided by cities through their open data portals, are time-stamped and geolocated. Specifically, we propose to work with four types of features: (e) **crime statistics** which could play a role in route selection if cyclists consider a neighborhood to be more or less safe; (f) **crash statistics** per street segment, to characterize the cycling safety of a given street. Types of crashes reported in open datasets typically include *collision with fixed car* or *hit and run*, among others; (g) **parking violations** per street segment, under the assumption that the volume of violations might shape the perception of safety or convenience for a given street segment. These violations are classified into different types such as *parked car obstructing sidewalk* or *parked car obstructing driveway*, among others; and finally, (h) **311 requests**, 311 is a non-emergency service that people can use in many cities to make complaints or report problems like road damage; in this paper, we focus on requestsrelated to the built-in environment of a street segment such as volumes of curb or pothole repairs, which could be indicative of cycling safety and potentially play a role in trip purpose. Figure 3(a) shows the normalized distribution of all 311 requests in a small area of Philadelphia in 2015-16. As shown, 311 requests are not uniformly distributed, with certain street segments having higher volumes than others.

To assess the role that the social features (crime, crash, parking and 311) might play in a cyclist's choice of streets and on the inference of trip purpose, we explore two approaches: (i) short-term memory (*SM*) , which assumes that only the most recent events will shape a cyclist's perception of street safety or route convenience, and (ii) long-term memory (*LM*), which considers that all events from the past might play a role in that perception. The former is computed as the average for the month when the trip took place, while the latter is computed using the monthly averages across all past available data up to the day when the trip took place (see Figure 3(b) for an example). These two approaches to feature calculation will also provide insights into the role that single or time-series representations of the social features play in the trip purpose inference. The evaluation in section 3.5 will describe further details about short- and long-term memory features applied to a specific dataset.



(a) Counts of all 311 requests per segment in 2015-2016, scaled to range of [0, 1]). Darker green represents larger volumes of 311 requests.

(b) Short-term memory (SM) and Long-term memory representations for the crime statistics feature. For a trip undertaken in April 15, SM considers only the crime count for that month for a given segment while LM considers all the previous months by taking an average of the crime counts.

Fig. 3. Social environment features per street segment. Features are time-stamped and geolocated per street segment, with different densities across the road network. Each feature is represented using two approaches: short- and long-term memory.

3.4 Inference Method

Formally, given a set of m trips with cyclist-collected GPS data and their trip purpose labels, we first characterize each trip as $C_i = \{SP, S, P\}$, where SP represents the spatio-temporal features of the trip (time, day, month, duration, distance of the trip, and POI at the obfuscated destination point); $S = \{s_1, ..., s_n\}$ represents the set of segments cycled in that trip; and P the set of personal features that characterize the individual who cycled that trip. Each segment s_i visited during the trip is characterized by its built-in and social environment features $F = \{f_1, ..., f_p\}$. Bear in mind that the number of segments n can be different across trips *i.e.,* one cyclist might cycle through 30 segments while other can cover a much larger number, depending on the length of the trip and on the cyclist's speed. With this representation, we frame the trip purpose inference as a classification problem where the set of trips $\vec{C}$ and their trip purpose labels $\vec{L}$ are used to train and test various trip classification methods $\vec{L} = M(\vec{C})$ and to pinpoint to the most relevant features for the identification of a given trip purpose.

We evaluate classification methods that have already been used for car trip purpose inference (although with fewer or different features) including logistic regression, SVM or Random Forest (RF) [24] using scikit-learn [53] and propose a novel method, based on XGBoost [16], that handles better large feature vectors like the ones we have in this setting. Logistic regression, RF and XGBoost require the number of features for each training and testing sample to be the same *i.e.,* equal feature size. However, since trips can have different number of segments, the total number of features per sample will be $n * p$ where n might vary. To overcome this issue, we re-define each trip C_i as a set of N segments where N is the total number of street segments in the city under study. Thus, a trip that goes through q different segments, will be represented as a vector of N street segments where only q segments have non-zero values for the built-in and social environment features, while all others vector elements are zero. However, with this set up it would be impossible to differentiate cycled segments whose features have all zero values from segments that have not been cycled. To disentangle this situation, we apply a smoothing

Table 1. Personal information collected by Cycle Philly App, including demographic: age, gender, ethnicity or income; and cycling experience information: cycling frequency, type of cyclist and cyclist history. Numbers in brackets indicate the total number of cyclists belonging to each category.

Personal Features	**Subtypes**
Age	'18 - 24' (28), '25 - 34' (121), '35 - 44' (55), '45 - 54' (19), '55 - 65' (7), '>65' (4)
Gender	'Male' (162), 'Female' (71)
Ethnicity	'White' (192), 'Asian' (9), 'Hispanic/Mexican/Latino' (3), 'African American' (11), 'Multi-racial' (4), 'Other' (6)
Income	'< \$20,000' (17), '\$20,000 – \$39,999' (30), '\$40,000 – \$59,999' (44), '\$60,000 – \$74,999' (20), '\$75,000 – \$99,999' (33), '> \$100,000' (53)
Cycling frequency	'Less than once a month' (22), 'Several times per month' (53), 'Several times per week' (116), 'Daily' (35)
Cyclist Type	'Interested, but concerned' (5), 'Comfortable, but cautious' (71), 'Enthused & Confident' (119), 'Strong & Fearless' (42)
Cyclist History	'Just trying it out/Just started' (13), 'One year or less' (11), 'Several years' (85), 'Since childhood' (128)

parameter $\alpha = 0.5$ to all the zero-valued features from segments that have been cycled, and leave the zero values for the non-cycled segments. All values are normalized using MaxAbsScaler which scales each feature by its maximum absolute value, preserving sparsity among the values.

On the other hand, SVM requires equal feature size when using traditional linear or nonlinear kernels (polynomial, GRB); but can handle different feature sizes when used with kernels adapted for time series analysis. In this paper, we evaluate linear- and GRB-based SVM as well as SVMs with a Global Alignment kernel (GAK), that computes the similarity of any two sequential distributions [8, 21]. Unlike Dynamic Time Warping (DTW) kernels, GAKs are faster to compute and do not require any type of correction on the Gram matrices to define positive definite kernels, which are required to compare time-series data [63]. Additionally, the SVM-GAK approach will also respect the order in which the segments were visited, which could potentially play a role in the identification of the trip purpose. Finally, we also evaluate the impact that class imbalance and spatial autocorrelation might have on the accuracy of the classification methods.

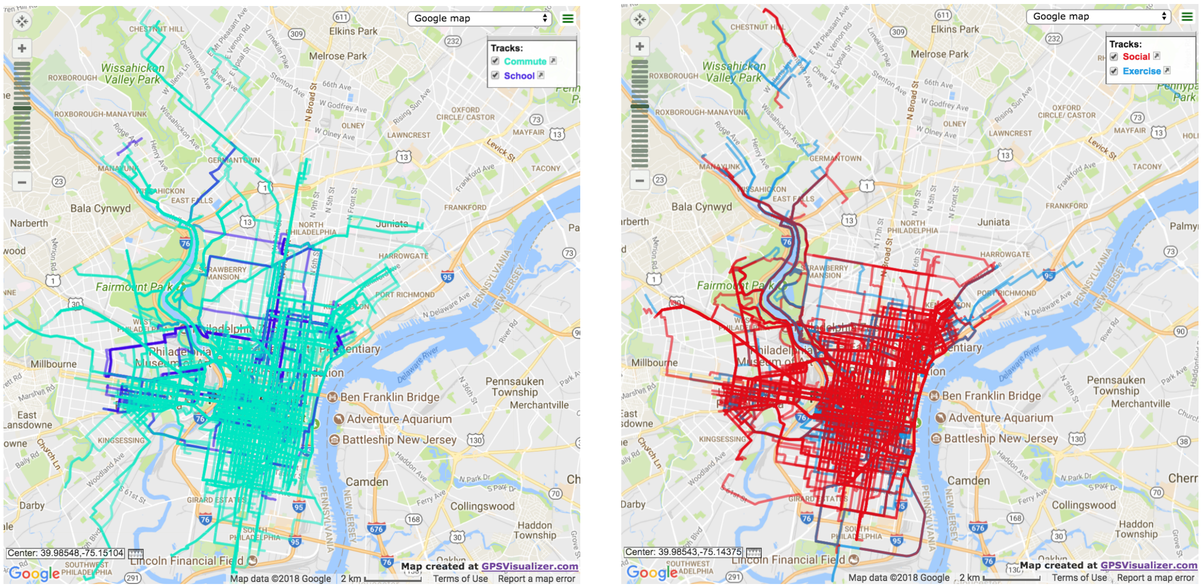
3.5 Evaluation

3.5.1 Datasets. We evaluate the proposed trip purpose classification methods using cyclist data from the City of Philadelphia. We focus on data collected by cyclists using the Cycle Philly App, an application developed by the city to allow cyclists record their cycling trips using their personal smartphones [19]. The mobile application was promoted by the city, as well as by regional authorities, encouraging cyclists to share their data for urban planning purposes. After the data collection period ended, the city gave open data access online [19].

Upon signing up, the Cycle Philly App asked cyclists to provide seven variables characterizing their demographic and cycling experience information. Specifically, cyclists were asked for their age, gender, ethnicity, income, cycling frequency, cyclist type, and cyclist history (see Table 1 for details). Additionally, every time cyclists rode their bicycles, they were asked to provide their trip purpose by choosing between commute, social, exercise, errand, work-related, shopping, school or other.

The data collection was started manually by the cyclists meaning that if they did not want a specific personal trip to be recorded, they had the choice to do so. The cycling GPS traces were collected until the end of the

[1]Images generated using GPS Visualizer.



(a) GPS traces for trips labeled as Commute (in cyan) and School (in dark blue)

(b) GPS traces for trips labeled as Social (in red) and Exercise (in blue)

Fig. 4. Example of cycling trips collected with Cycle Philly App.[1] Labels were provided by cyclists and trips were collected as GPS traces.

trip, and both origin and destination points were obfuscated to the closest intersection for privacy preserving purposes. The complete dataset was collected over a period of two years, from May 2014 to April 2016 and included $|\vec{C}| = 7,367$ trips made by 255 cyclists distributed as: 4,472 commuting trips, 967 social trips, 470 errand trips, 414 work-related trips, 346 exercise trips, 296 trips labeled as *other*, 226 shopping trips, and 176 school trips. After computing the street segments associated to each trip, the average number of segments per trip was 46. Figure 4 shows a map of Philadelphia with the GPS traces collected from four different types of trip purposes: commute, school, social and exercise.

The GPS traces were used to compute the six spatio-temporal features for each trip; while the built-in and social environment features to characterize the street segments were extracted from several datasets in Philadelphia's OpenDataPhilly website [55]; as well as from Open Street Maps. The built-in environment features per segment included three road network features (slope, minimum and maximum values); 23 graph-based features (different segment importance and centrality measures with primal and dual approaches); 10 POI features; and one bike lane feature. For each of the four social environment features, we collected all the events available during the same time range as the trip collection period *i.e.,* between May 2014 and April 2016. We computed monthly averages per social feature using both the short- and long-term memory approaches. The final dataset used to train and test the trip purpose classification methods consists of 7,367 trips with their purpose labels, where each trip is characterized by all spatio-temporal ($|SP| = 6$), personal ($|P| = 7$) and built-in and social environment features ($|F| = 41$). We will give open access to this final dataset for other researchers willing to continue this line of work.

Table 2. Trip purpose classification for eight trip purposes: commuting, social, exercise, errand, work-related, shopping, school or other. Results are reported as averaged micro-F1/Macro-F1 scores over ten random runs for each method and feature set. Nonlinear-SVM was run with RBF and $C = 100$ and $Gamma = 10$. XGBoost was run with $ntree = 500$ trees and a depth of $max\_depth = 6$. Car trip inference results reported in the literature are provided in the last line of the Table to show that the scores obtained using our cycling datasets are not as good, thus motivating the need for a new trip purpose inference approach.

METHOD	BS	BS+P	BS+SEG[SM]	BS+SEG[LM]	BS+SEG[SM]+P	BS+SEG[LM]+P
LogisticR	0.59/0.20	0.58/0.26	0.64/0.42	0.59/0.28	0.64/0.42	0.58/0.30
LinearSVM	0.58/0.20	0.58/0.25	0.64/0.42	0.59/0.26	0.64/0.42	0.59/0.29
NonLinearSVM	0.61/0.30	0.65/0.36	0.62/0.13	0.67/0.40	0.62/0.13	0.68/0.41
SVM (GAK)	N/A	N/A	0.47/0.12	0.47/0.12	N/A	N/A
RandomForest	0.66/0.19	0.66/0.20	0.61/0.11	0.65/0.21	0.61/0.11	0.66/0.22
XGBoost	**0.66/0.40**	**0.71/0.42**	0.71/0.43	**0.72/0.44**	0.72/0.45	**0.73/0.47**
XGBoost + Unders.	0.41/0.28	0.52/0.32	0.41/0.31	0.47/0.34	0.45/0.34	0.48/0.36
XGBoost + Overs.	0.68/0.40	0.69/0.46	0.69/0.46	0.72/0.48	0.70/0.48	**0.73/0.50**
Car Trip Inference	[0.70-0.90]	[0.72-0.82]	N/A	N/A	N/A	N/A

3.5.2 Classification Results. Table 2 shows the F1 scores for different combinations of methods and spatio-temporal, street segment and personal features. To account for the effect of the imbalanced nature of our dataset (larger proportions of commuting trips than any other type), we report both micro- and macro-F1 scores (m-F1 and M-F1). Significantly lower micro scores when compared to macro values, reflect high misclassification among the most common labels, with labels with lower numbers of samples being correctly classified. On the other hand, macro scores significantly lower than micro scores are associated to poor classification rates among labels with lower numbers of samples, with common labels being correctly classified.

For each method, the dataset is divided into training (70%) and testing (30%) sets randomly selected ten different times, and average F1 scores across all runs are reported. We evaluate the methods described in the previous section for the following sets of features: (a) each trip is exclusively characterized by its spatio-temporal features (*BS*), (b) each trip is characterized by a combination of spatio-temporal features and personal data about the cyclist (*BS+P*), (c) each trip is characterized by a combination of spatio-temporal and segment features computed using both the short- and the long-term memory approaches (*SEG[SM/LM]*), and (d) each trip is characterized by a combination of spatio-temporal, segment features computed using both short- and long-term memory approaches, and personal data about the cyclist. Results for the SVM (GAK) method are only reported when using segment features since it only works when the training data can be specified as a time series, which is not the case for the spatio-temporal or personal information. For comparison purposes, the last row of the Table contains the range of accuracy values reported in the literature [15, 24, 40, 41, 46, 48, 51, 71] for different combinations of methods (SVM, Logistic, Random Forest) and sets of features to infer trip purpose using car-based GPS data.

The first important observation is that the features and methods that are currently used for car trip purpose inference (first two columns in the Table) do not work as well for cyclist-collected GPS data. We can observe that the results achieved using BS and $BS + P$ features extracted from car-generated GPS data provide accuracies between 0.7 and 0.9; while the same features extracted from cyclist-collected GPS data to infer trip purpose provide much lower scores with maximum values (after transforming F1 scores to accuracies) between 0.66 for BS and 0.71 for $BS + P$. Thus, as proposed, we set out to explore results when the inference methods are enhanced using street segment features from the cycling trip. As the table shows, adding segment features to the baselines improved the F1 scores to $m\text{-}F1 = 0.72$, $M\text{-}F1 = 0.44$ (third column); and this result was also enhanced by adding the personal data of the cyclist, which boosted F1 scores to $m\text{-}F1 = 0.73$, $M\text{-}F1 = 0.47$ (fourth column). The

SVM (GKA) method showed really poor scores compared to all other approaches. We hypothesize that these low scores are probably due to the fact that only street segment features are considered, and thus, not incorporating the spatio-temporal and personal features heavily impacts the final scores. Given the imbalanced nature of our sample, it is also important to compare these results against a simple majority classifier that considers all labels as the one with the largest presence in the dataset (commute in our case). Such classifier resulted in much worse F1 scores, $m\text{-}F1 = 0.61$, $M\text{-}F1 = 0.09$, which confirmed that XGBoost with spatio-temporal, segment and cyclists' personal features had the highest predictive power.

These results indicate that incorporating personal data and information about the built-in and social environment visited during the trip helps in improving trip purpose inference when only spatio-temporal features are considered by 7%; which might in turn indicate that different types of cyclists choose different types of built-in and social environment features (consciously or unconsciously) depending on the purpose of their trips, since the personal data and segment features help in differentiating across trip purposes. It is also important to highlight that the best results were obtained with the long-term memory option (LM) *i.e.,* considering the evolution of the various built-in and social environment features over all past months has a stronger predictive power than simply considering their values during the month when the trip took place; which could signal that the mental map that cyclists form about their environment is constructed over months of events, rather than just a few weeks surrounding the trip.

Focusing on the best method, XGBoost with spatio-temporal data, long-term memory segment data and personal data (with ntree=500 and max_depth=6), we next explore the most relevant features in the identification of trip purpose. Since XGBoost is trained using an ensemble of decision trees, the technique can automatically provide the importance of each feature during training (see Figure 5). The top most relevant features in identifying different types of trip purpose were: trip duration, hour at which the trip took place, slope, number of green areas (as POI), 311 cycling-related complaints, trip distance, network centrality of the segments, crime or presence of cycling facilities. As the list shows, trip purposes can be identified, in large part, through basic spatio-temporal features of the trip, through POI visited, through a few social environment features including crime rates and the quality of the segments, both in terms of maintenance and road network connectivity, and through some personal data such as cycling experience. Crash, parking violations or road type data did not appear among the top features, which indicate that these features are not as relevant in identifying trip purpose *i.e.,* their values are more homogeneous across trip purposes. Section 4 will delve more into the identification of the specific types of features that are important for each trip purpose.

3.5.3 Class Imbalance. The differences between F1 macro and micro scores (26% for the best approach) show that poor classification rates happened more among labels with lower numbers of samples, while common labels were being correctly classified. Exploring the confusion matrix for the best approach (XGBoost), we observe that the vast majority of misclassified samples were being wrongly classified either as social or commute, the two majority classes (see Figure 6). This could be due to the imbalanced nature of the dataset, or to the fact that certain labels from the original dataset, such as shopping or errands, might be difficult to differentiate among themselves or from others. To assess both hypotheses, we explore two approaches and report their F1 scores: (1) over/undersampling and (2) reducing the labels to a smaller, concise set. For the first approach, undersampling reduces the number of samples of each class to the smallest value, and repeats the process several times to account for selection biases; while oversampling creates synthetic samples, via k-nearest neighbors, for all classes until they reach the number of samples for the majority class. We used the imbalanced-learn toolbox [37] to implement both methods and the resulting F1 scores show that while undersampling provided worse F1-scores, oversampling slightly improved the classification rates for the minority classes (macro-F1 increased 3%, see Table 2 last rows).

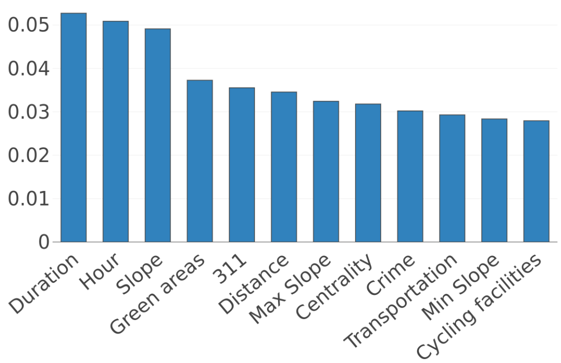


Fig. 5. Top most relevant features ordered by importance with XGBoost. We can observe that basic spatio-temporal features such as duration of the cycling trip or hour at which the trip took place are among the most determinant of trip purpose.

Fig. 6. Confusion Matrix for XGBoost with eight classes. Darker colors represent best trip purpose classification values. Most of the classification errors wrongly assign as trip purpose either commute or social as can be observed in those two columns and rows.

Table 3. Trip purpose classification for four trip purposes: commuting, exercise, school and social. Results are reported as averaged micro-F1/Macro-F1 scores over ten random runs for each method and feature set.

METHOD / FEATURES	BS	BS+SEG[SM]	BS+SEG[LM]	BS+SEG[SM]+P	BS+SEG[LM]+P
LogisticR	0.77/0.35	0.81/0.62	0.79/0.45	0.81/0.62	0.79/0.49
LinearSVM	0.78/0.34	0.81/0.63	0.79/0.43	0.81/0.63	0.79/0.48
NonLinearSVM	0.78/0.49	0.80/0.28	0.82/0.58	0.80/0.27	0.83/0.56
SVM(GAK)	N/A	0.69/0.22	0.69/0.23	N/A	N/A
RandomForest	0.80/0.31	0.80/0.26	0.81/0.33	0.80/0.26	0.81/0.33
XGBoost	**0.83/0.58**	**0.85/0.66**	**0.85/0.66**	**0.85/0.67**	**0.85/0.67**

For the second approach, reducing the set of trip purpose labels to a smaller number of distinct purposes, we use the confusion matrix from the best XGBoost solution to merge the classes that have a majority of misclassified samples as another label into that label. As a result, errands, shopping and work-related trips are merged into commute. The final set of labels: commute, exercise, social and school, although reduced, is still highly useful from an urban planning perspective. Table 3 shows the F1 scores for this new setting. We observe that the F1 scores improve across methods by 12% in the best-case scenario, which is XGBoost, with $m\text{-}F1 = 0.85$, $M\text{-}F1 = 0.67$.

As can be seen, both macro and micro scores improved, indicating that the classification accuracy improved for both the majority and minority labels. Similarly to the experiment were all trip purposes were considered, the best F1 scores were obtained when the trips were represented not only with spatio-temporal trip features, but also with specific segment features as well as personal information about the cyclists. However, in this case, no difference in the F1 scores was observed between using a short-term or a long-term representation of the built-in and social environment features. This might be due to the fact that the baseline with only four trip purposes is already quite good ($m\text{-}F1 = 0.83$, $M\text{-}F1 = 0.58$), and as a result, the segment features improve slightly the F1 scores, without being able to capture the impact of the long or short term approaches. This is confirmed by looking at feature importance for XGBoost, which shows that the social features, which are modeled in terms of short- and long-term memory, are of lower importance than general spatio-temporal and personal features.

3.5.4 Spatial Autocorrelation. Many of the built-in and social environment features we use might be spatially autocorrelated *i.e.,* their value could be very similar to, or completely different from, those in the neighboring segments. For example, if the number of crimes in a street segment is high, the number of crimes in segments nearby might also be high. Here, we explore the use of spatial autocorrelations as yet another feature to predict cycling trip purpose. First, we identify all autocorrelated features and enhance each segment's feature vector with as many elements as autocorrelated features have been detected; with each element representing the average value for that feature across all of the neighboring segments. Next, all the new segment representations are used to re-run the predictive algorithms for all different combinations of built-in and social environment features. The spatial autocorrelation is analyzed using Moran's I statistic with a spatial weights matrix that defines neighboring segments using a weighting strategy based on the distance between the mid-points of any pair of segments, with a distance band (cutoff point after which feature values are ignored) of $150m$, since it gave the best prediction results. Features are deemed autocorrelated when the p-values for the Moran's I test were $p < 0.05$. In our dataset, 129 out of 131 features have been identified as being positively spatially autocorrelated with I values in the range of $0.02 < I < 1$, indicating that built-in and social environment features tend to cluster, rather than being dispersed or randomly distributed.

Figure 7 shows the results for the best approach identified for four and eight purpose classes: XGBoost with oversampling using baseline, segment (long-term memory) and personal features. As the table shows, the F1 scores improved for the four-class setting and remained the same for eight classes. Using four classes improved the micro-F1 score by 1% and the macro-F1 by 3%, showing that including spatial information into the prediction helped in better identifying trip purpose.

3.5.5 Classification By Cyclist or Demographic Type. The previous experiments have explored trip purpose classification for all types of cyclists, independently of their demographic and cycling expertise characteristics. In this section, we evaluate trip purpose classifiers exclusively built for individuals with specific demographic features or types of cycling expertise. The objective of this experiment is to identify whether trip purpose prediction works better for certain types of individuals, based on their demographic characteristics or cycling expertise. For example, through this analysis we will be able to evaluate whether trip purpose prediction for *fearless* cyclists is more or less accurate than that for *cautious* cyclists; whether trip purpose prediction works better for certain age groups; or whether trip purpose is more predictable for females than for males.

To carry out this experiment, we first take from the pool of 7,367 trips only those whose cyclists have provided personal demographic information, which accounts for 51% of the total. We then divide the trips by cyclist type, age range or gender, and repeat the trip purpose inference experiments for each individual variable across all of its types. To guarantee statistical significance of our inference experiments, we only report results for those features, and their types, that have at least 10 samples per trip purpose. Figure 8 shows the micro-F1 scores computed using XGBoost and oversampling averaged over ten runs and four trip purpose labels.

We can observe that as cyclists report less confidence in their cycling skills from *strong* to *comfortable*, the micro-F1 scores decrease 9% showing that trip purpose is most easily predictable for strong cyclists. This result could reflect that highly confident cyclists tend to be more repetitive in their routes, and thus more predictive; while cautious cyclists might be more entropic in their behaviors and thus less predictable. A similar result is observed across age groups, with 25-34 and 35-44 being the groups with the most predictable trip purposes, possibly reflecting a more homogeneous selection of types of routes than younger 18-34 or older 45-54 age groups with F1 scores 14% lower than the best age group. Finally, the smallest differences across F1 scores were observed for gender, with F1 scores 1% higher for males than for females, which indicate that both genders are almost equally predictable in terms of trip purpose. Overall, this experiment shows that if demographic features or

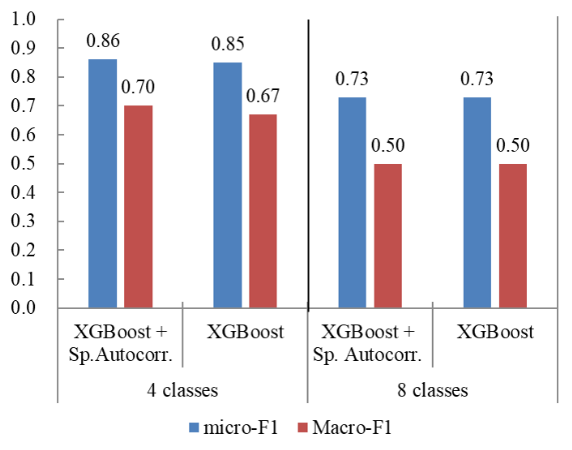


Fig. 7. micro-F1/Macro-F1 scores with and without spatial autocorrelation information for four and eight trip purposes.

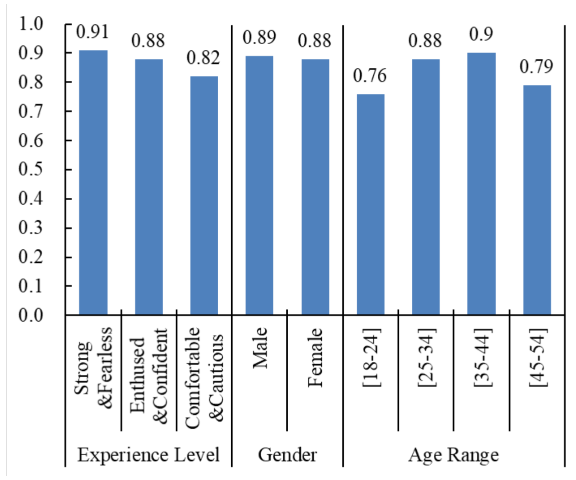


Fig. 8. micro-F1 scores for specialized trip purpose classifiers built for cyclist experience levels *strong*, *enthused* and *comfortable*; gender male and female; and age ranges 18-24, 25-34, 35-44 and 45-54.

cycling experience were to be collected for the majority of cyclists in a city, trip purpose prediction could be improved for *strong* and *enthused* cyclists and for age groups 25-34 and 35-44; while for the other features, a general classifier would do a better work.

4 ROUTE CHOICE

In this section, we propose and evaluate a novel method to analyze route choice *i.e.,* to understand the specific built-in and social environment features that cyclists appear to favor when choosing a cycling route. The method has two main components: the *route collection* component, that collects the route choice set *i.e.,* the set of routes that the cyclists could have taken to go from the trip origin to a destination, and characterizes each route by their built-in and social environment features; and the *choice analysis* component, that compares the cyclists' chosen routes against the route choice set to identify the set of built-in and social environment features that cyclists appear to favor with statistical significance (see Figure 1, bottom, for details). As stated in the introduction, our method improves current approaches by using crowdsourced route choice sets provided by Google cycling directions which incorporate route choices that represent different mental models and that take into account the influence that others might have on the selection of a specific route.

4.1 Route Collection

The route collection component uses Google's maps cycling directions and open datasets with information about the built-in and social environment features of the street segments. Given a set of cyclist-collected trips characterized by their GPS traces, we first retrieve the route choice set from Google's cycling directions API. For a cyclist-collected trip (C_i) between an origin and a destination, Google provides a maximum of three suggested routes between the same origin and destination points ($G_i^{1,..,3}$). As explained in Section 3.4, each cyclist-collected and Google-suggested route is transformed into a list of street segments, and each segment is then characterized with a set of built-in and social environment features *i.e.,* $C_i = \{s_1, ..., s_n\}$ and $G_i^j = \{s'_1, ..., s'_m\}$ $j \in [1-3]$ where s_i is a segment in the trip characterized by its features $s_i = \{f_1, ..., f_l\}$ and where cyclist-collected and

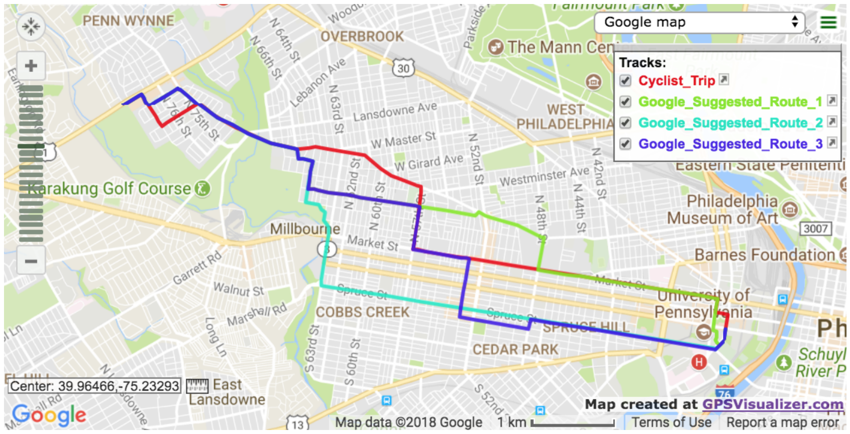


Fig. 9. Example of a cyclist-collected route (in red) together with the three Google-suggested routes. We can observe that certain parts of the cyclist route overlap with the suggested ones, while others are different. The focus of this section is to understand the potential reasons behind cyclists' route choices.[2]

Google-suggested routes can have different number of segments ($m = n$ or $m \neq n$). Figure 9 shows an example with a cyclist-collected trip, and with the route choice set retrieved from Google.

4.2 Choice Analysis

The choice analysis component focuses on the identification of the differences between the cyclist-collected routes and the other Google-suggested routes that cyclists could have chosen. These differences, which will be characterized using the built-in and social environment features of the cycled routes, will throw some light into the potential reasons that cyclists might have to favor one route versus another.

To achieve this objective, we propose a two-step method. Figure 10 shows the main components for each step. In the first step, we aim to identify the pairs of cyclist-collected and Google-suggested routes that are statistically significantly different based on the distributions of their segments' features. Since each route can have a different number of segments and each segment is characterized by a set of p features, we propose to run a multivariate, two-sided Kolmogorov-Smirnov test (KS) [39]. For each cyclist-collected trip C_i we will run a KS test with each of its possible Google-suggested routes *i.e.,* $KS(C_i, G_i^j)$ $j \in$ [1-3]. These non-parametric tests will output the pairs of trips for which the built-in and social environment features of the street segments cycled are statistically significantly different *i.e.,* cyclists appear to favor certain features versus others available in the suggested Google routes. Due to the large number of tests we run, it is highly possible that some of the significant results will be false positives. To control for this, we apply the Benjamini-Hochberg procedure with a false discovery rate of 0.05 [7].

In the second step, we focus on the identification of the specific features that make the pairs of trips identified in step one statistically different, and that might be playing a role in the selection of a route. For that purpose, we run, for each built-in and social environment feature f_l a multivariate, two-sided Mann-Whitney (MW) test [35] between all cyclist-collected and Google-suggested routes that have been identified as statistically significant different by the KS test in step one *i.e.,* $MW(C_i(f_l), G_i(f_l))$ where $i = 1, ..., q$ is the number of significantly different pairs extracted in step one and f_l is one specific built-in or social segment feature. These non-parametric tests will identify whether the medians of each feature are significantly similar or different across the segments of the cyclist-collected and Google-suggested routes. To evaluate in depth the magnitude and direction of the differences

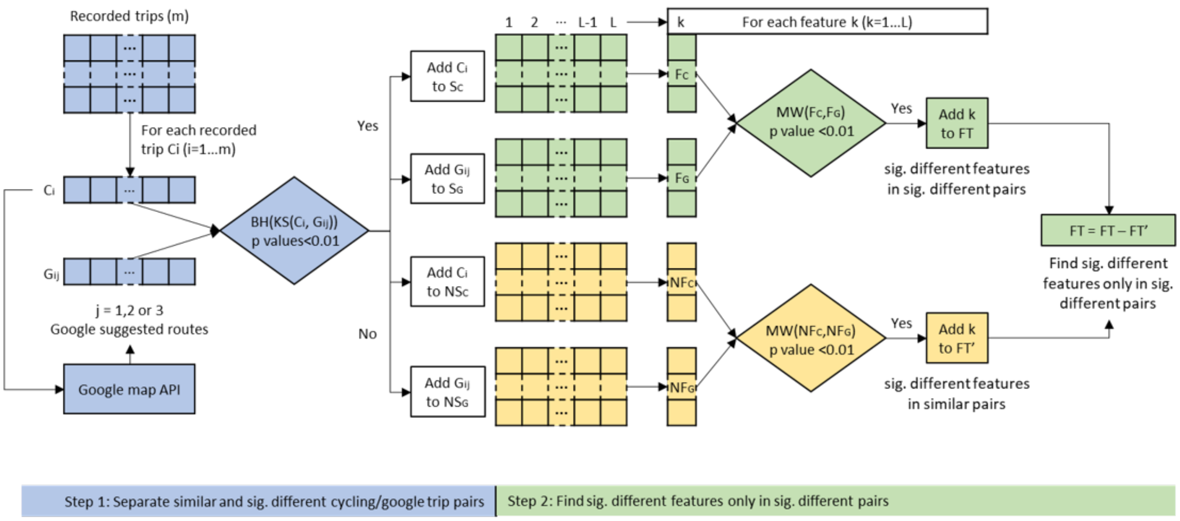


Fig. 10. Flowchart of the proposed two-step statistical method to analyze Route Choice.

found, averages and standard deviations per feature across both cyclist collected and Google-suggested routes are also computed.

Some of the features identified in step two as significantly different across cyclist-collected and Google-suggested trips could also appear in pairs of routes that were not deemed significantly different by the KS test in step one *i.e.,* certain pairs of routes could have features that are statistically different, but not enough features to make the two trips statistically different. To identify these features, we run step two over the set of cyclist-collected and Google-suggested trips that were not identified as being statistically significantly different by the KS test in step one; and eliminate them from the list of significant features extracted in the first run of step two. We use this final set of features to explain the types of route choices made by cyclists in the urban environment under study. It is important to clarify that our statistical analyses will be affected by any existing bias in the open and crowdsourced datasets *i.e.,* certain social features, such as 311 requests, might be more present in affluent areas and as a result our analyses could fail to reveal findings in less affluent areas.

4.3 Evaluation

To evaluate the proposed approach, we run three experiments. The first experiment focuses on the identification of the street segment features that might affect route choice in general, while the second experiment focuses on the identification of such features for each specific trip purpose: commute, exercise, social and school. Finally, we use the built-in and social environmental features identified to explore the predictability of a route being selected by a cyclist.

4.3.1 Dataset. We run the experiments using the $|\vec{C}| = 7,367$ cyclist-collected trips from the city of Philadelphia described in section 3.5; and characterize the segments in each trip using the set of 41 built-in and social environment features presented in section 3.3. The analyses are both done when the social environment features (crime, crash, parking and 311) are computed using the long-term memory approach *i.e.,* all events from May 2014 onwards are counted as potentially influencing a cyclist choice; and the short-memory approach *i.e.,* only

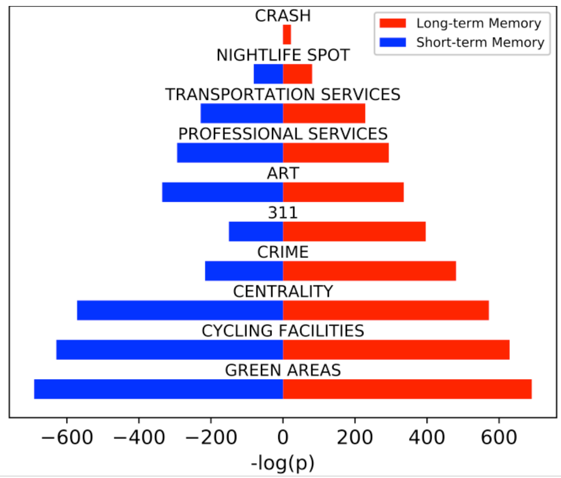
[2]Image generated using GPS Visualizer.

the events from the month when the trip takes place are considered as potential features that could influence route choice.

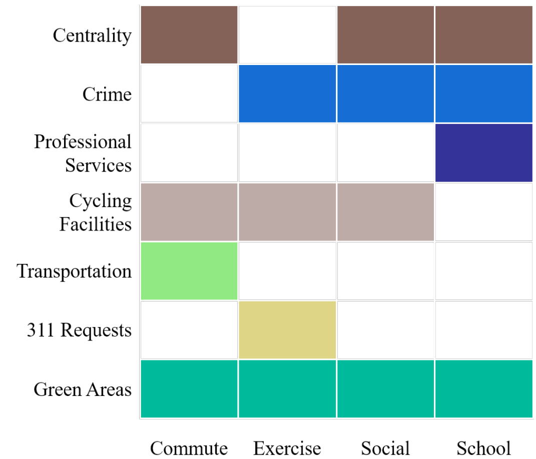
4.3.2 Analysis across Trip Purposes. Figure 11(a) shows the features identified as playing a significant role in route choice at $p < 0.01$ (for both KS and MW tests). For clarity purposes, we rank the features by their $-log(p)$ where p is the p-value for the statistical test, and show results for both the short- and long-term memory approaches. Recall that the smaller the p-value (the larger its $-log(p)$), the more significant the feature is *i.e.,* we can assure with higher certainty that the feature does play a role when cyclists decide which route to take. The first observation is that the top three most important features are common for both long- and short-term memory: presence of green areas, presence of cycling facilities and road centrality, specifically eigenvector centrality computed when street segments are considered as nodes. Our analysis indicates that when cyclists in Philadelphiachoose a route, they first look for agreeability, safety and convenience. In fact, cyclists appear to highly favor routes that are agreeable, defined as the presence of green areas and parks on the route; routes that are safe, with safety defined by the presence of cycling facilities such as various types of protected (or not) bike lanes; and by the convenience of the route, represented by the centrality of the streets *i.e.,* how connected the route is to other routes. Previous work has found that the topological centrality of streets is significantly correlated to retail commerce vitality, popularity and human way-finding [20]. Our findings for the City of Philadelphiacould indicate that the mental model cyclists have of their cities is heavily influenced by major, commercial and popular roads. In fact, the list of statistically significant features (although with smaller p-values) also includes some POIs such as professional service or nightlife spots, typically present in major central roads.

Our analysis also identifies that the crime rates of the cycled routes play a significant role in route choice, with cyclists avoiding areas with high crime in Philadelphia. Although this feature is important in both long- and short-term memory analyses, the p-value is much lower in the longer-term possibly indicating that the mental association of crime to parts of the city takes a long time to form. Similar results have been reported in other mobility studies that look at the impact of crime in, for example, the choice of public transportation routes [1]. Figure 11(a) also shows that the quality of the road, measured via 311 reports about cycling conditions in Philadelphia, also plays a role in route choice, and that that role is more significant in the long-term approach. As with crime rates, this indicates that forming a mental map of road quality is not based on single events but rather on an accumulation of knowledge and route familiarity over a long period of time. A detailed analysis by type of specific 311 reports showed that the most significant issues affecting cyclist route choice and forcing to look for alternative routes are the presence of double parked cars, the presence of loading zones, illegal dumping, street light outage or the presence of construction sites. Interestingly, crash rates, although significant, were identified as having much less importance in cyclists' route choice than all the other previously discussed features, and was only identified as significant for the long-term approach. Another interesting feature is the presence of transportation services (*e.g.,* metro or bus stops), which is identified also among the top most significant features in cyclists' route choice for Philadelphia. This factor could be revealing multi-modal transportation patterns where cyclists might combine bicycle with metro or bus, and as a result, having these types of services on the routes they follow becomes an important variable for them.

4.3.3 Analysis by Trip Purpose. Figure 11(b) shows the top four most statistically significant features ($p < 0.01$) for each of the following four trip purposes: commute, exercise, social, and school. We only report results for the features computed using the long-term approach, although the short-term revealed similar findings. We observe that road centrality is a feature that cyclists take into account for commute, social and school purposes, but not when the trip purpose is exercise. Cyclists in Philadelphia appear to favor routes that go through central, well communicated street segments that will potentially drive them faster to their final locations. However, and logically, this feature is not found significant when the trip purpose is exercise, where cyclists are not so interested in rapidly moving between locations but rather on enjoying the route. Our results also show that cyclists in



(a) Statistically highly significant segment features for route choice across all trip purposes in Philadelphia. Results for both short- and long-term memory are shown. Larger $-log(p)$ values are associated to higher statistical significance for any given feature.



(b) Top statistically significant segment features for each of the four trip purposes using long-term memory. Each row has four elements, one per trip purpose: commute, exercise, social and school. A given element is colored if there is a statistically significant relationship between segment feature and trip purpose.

Fig. 11. Statistically significant social and built-in environment segment features for route choice.

Philadelphia favor routes that go through regions with low crime rates except for the case of commuting when cyclists appear to be giving more importance to reaching a destination quickly (centrality) than to the safety of the route. Finally, cyclists in Philadelphia also appear to choose routes that have cycling facilities available when the trip purpose is commute, exercise or social, but not for school purposes, which could reveal the fact that when going to school, speed (measured via road centrality) is given more importance than cycling safety.

4.3.4 Predicting Cycled Routes. We have already shown that certain built-in and social environment features appear to be favored by cyclists. In this section, we evaluate if we can use these features to predict whether a route will be cycled by a cyclist or not in Philadelphia. As proposed, this new experiment provides an analysis that is complementary to the two-step statistical significance test, by predicting whether a route - characterized by its built-in and social environmental features - might be taken or not by a cyclist. Such predictive model could potentially help urban planners in making informed design decisions about new routes as well as on the analysis of cycling route likeability.

We frame the prediction as a binary classification problem (is the route taken by a cyclist or not), and focus on regression methods to easily identify the most predictive features via coefficient and significance analysis. We evaluate the following two regression methods: Logistic and Ridge (with a binomial function). Additionally, to assess the role that feature selection might have on the prediction accuracy, we evaluate these methods considering different sets of built-in and social environmental features as independent variables. Specifically, we consider the following three sets of independent features: (i) all built-in and social environmental features (for both long- and short-term memory), (ii) all built-in and social environmental features identified as statistically significant in the two-step statistical significance test presented in Section 4.2 (for both long- and short-term memory) and (c) similar to (b), but with an effect size correction for both long- and short-term memory, using the Z-score ($|Z\text{-score}| > 1.96$) [30].

Table 4. Route selection prediction results using two methods: Logistic and Ridge Regression and six different subsets of built-in and social environment features.

METHODS	Logistic Regression			Ridge Regression		
Features	AIC	BIC	R-sq	AIC	BIC	R-sq
1- All features (long-term memory)	62.509	371.488	0.6243	63.444	372.421	0.0583
2- All features (short-term memory)	62.519	371.499	0.6262	63.442	372.421	0.0575
3- Significant features by p-value (long-term memory)	49.136	294.980	0.6386	50.089	295.933	0.0416
4- Significant features by p-value (short-term memory)	47.128	285.522	0.6371	48.086	286.481	0.0402
5- Significant features (p-value and effect size) (long)	43.108	266.603	0.6335	44.067	267.561	0.0308
6- Significant features (p-value and effect size) (short)	41.102	257.147	0.6325	42.064	258.109	0.0294

All models resulting from the combination of the two methods and the six feature sets are evaluated using AIC, BIC and the pseudo R-squared values. It is important to clarify that since our models are binary classifiers, the R-squared we use is a pseudo-R-squared value (McFadden 's) [45]. Table 4 shows the main results. We can observe that the Logistic-based models outperform the Ridge models when compared for the same subsets of features *i.e.,* lower AIC and BIC, and higher R-squared values. Furthermore, the prediction models generally work better when only the statistically significant features identified in Section 4.3.2 are used as predictors, with slightly better results for short-term memory features (see lines 4 and 6 versus lines 3 and 5). We can also observe that as we incorporate more features into the models i.e., consider all features as input to the regression, the AIC and BIC values increase a lot, while R-squared values decrease. This result can be explained by the fact that AIC penalizes complex sets of features, which in our case is represented by all features being considered (lines 1 and 2 in the Table). Additionally, we can also argue that inputting only the statistically significant features from the previous section, we are performing a feature selection process that appears to positively affect the accuracy of the inference method (AIC and BIC values are lower, R-squared values are higher). To select the best model, we look both into AIC/BIC values as well as into the pseudo-R-squared values. Logistic-based models show much higher R-squared values than Ridge-based models with values between 0.6190 and 0.7175 as opposed to 0.02 and 0.05, respectively. As a result, it would be advisable to select any Logistic-based model between lines 3 and 6 in the Table, with a slight preference for model 3 given its slightly higher R-squared value.

Finally, looking into the significant independent features of the best Logistic regression and its coefficients, we observe that both centrality measures and cycling facilities are considered as the most influential features in the predictive model, with coefficients between $|1.1|$ and $|0.08|$ (at $p < 0.01$); while green areas are also significant but with a much smaller coefficient ($|0.004|$ with $p < 0.01$). These features represent a subset of the features identified in Figure 11(a) and with a slight order change in significance. However, both results are coherent since the two-step analysis focuses on identifying significance while the Logistic regression model focuses on identifying significance from a prediction point of view, which is always more restrictive.

5 DECISION SUPPORT TOOL FOR DECISION MAKERS

In this paper, we propose to use GPS data from cycling experiences to infer trip purpose and analyze route choice. Ultimately, we aim to help decision makers evaluate what built-in and social environmental changes could be implemented, and in which locations, so as to attract more cycling activity. In this section, we explain in detail the decision support tool that the methods proposed in this paper provide, and discuss a couple of use cases.

[2] Map visualization using Leaflet and Carto tiles.

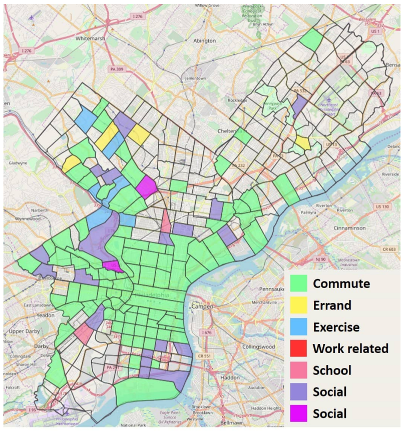


Fig. 12. Majority Origin by Purpose.

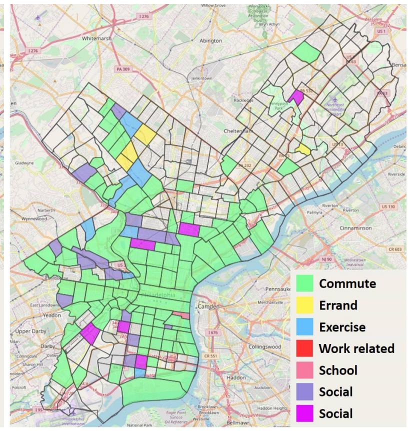


Fig. 13. Majority Destination by Purpose.

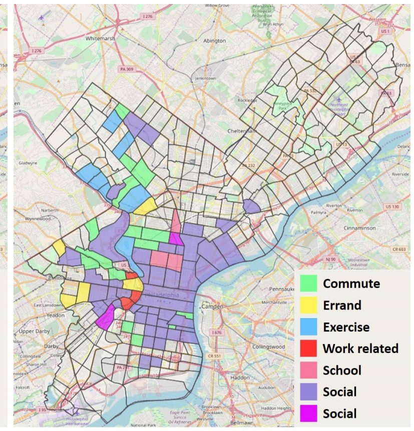


Fig. 14. Top Second Destination.

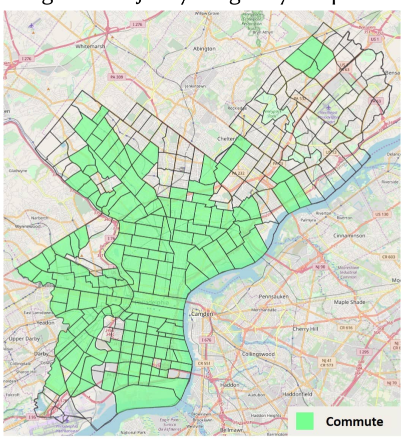


Fig. 15. Along-the-route Commuting.

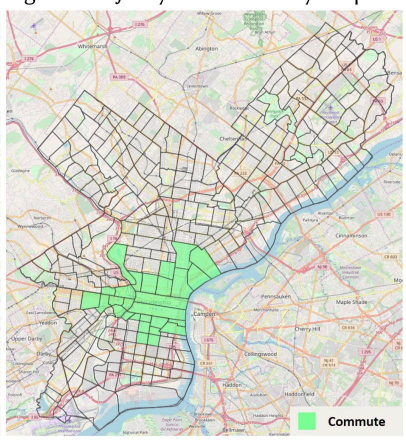


Fig. 16. ALT Commuting with Threshold.

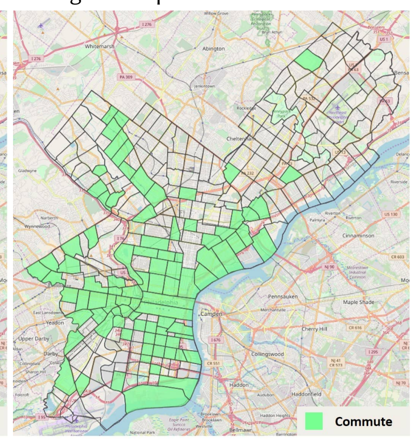


Fig. 17. Destination Commuting.

We bootstrap the decision support tool with the GPS traces and trip labels used in this paper. These traces are used to build two tools: the trip inference method and the route choice model. With these tools in place, cyclists in Philadelphia can be encouraged to continue to collect their GPS traces, but without the need to label trip purpose every time they cycle. The assumption is that it is easy for cyclists to collect GPS traces (just press a button in the mobile application), but it is more tedious to ask them to label their trips. As urban planners access newly collected GPS cycling traces, they can use the trip purpose inference method as a black box to label each trip collected with a trip purpose. As we have shown in Section 3.5, inference methods can achieve F-1 scores of up to 0.86, giving high confidence to the assigned labels; if necessary, the tool can be customized to accept or reject inferences based on a desired minimum accuracy. The inferences can be visualized as aggregated inferred trip purpose maps that divide cities into small geographical areas characterized by the trip purpose(s) in that region, as shown in Figures 12-17. With our methods, urban planners can have access to three different types of maps: aggregated distribution of trip purposes by origin, by destination or along the route, meaning that the maps associate to each region the trip purpose(s) of cycling trips that start in the region, that end in the region or that pass through that region, respectively. Figures 12, 13 and 14 show examples of such maps, where each geographical area is color-coded with the trip purpose that represents the majority (more than 50%) of origin or destination trips for that region, or the second majority trip purpose destination, respectively. Lack of color implies that there are no recorded trips for a given trip purpose for that geographical region *i.e.,* cyclists appear not to start in, finish at, or pass through that geographical region for any trip purpose. Interestingly, these maps show that the most cyclist active areas in Philadelphia are in the south-east, as reported in [9].

Once trip purpose labels are assigned to each individual trip, these can be used as input to the route choice analysis, to assess the built-in and social environmental features favored by cyclists for each type of trip purpose, and to rank them by importance, as shown in Section 4. Combining trip purpose maps with the route choice findings, urban planners can propose policy or infrastructure improvements for a given region that is considered mostly of a specific type of trip purpose. For example, we have shown that cyclists in Philadelphia highly favor routes with small numbers of 311 reports (such as double-parking), possibly related to cycling safety and speed; or that cyclists in Philadelphia highly favor cycling facilities, including bike parking (see Figure 11(a)). These are statistically significant findings that characterize cyclist preferences in Philadelphia, however, such preferences are not necessarily present throughout the cycling infrastructure. Thus, urban planners willing to improve cycling conditions for, for example, commuters, could use an along-the-route trip purpose map, as shown in Figure 15, to identify all the geographical regions where the majority of the passing trips are commuting, and suggest to put in place more parking enforcement officers in those areas, in the hopes of decreasing the number of double-parked cars, improving cycling safety and speed and, as a result, increasing the number of cycling commuting trips. One could argue that placing more parking enforcement officers in all the regions shown in Figure 15 is not doable due to potential lack of funding. Given the type of data we collect, the decision support tool allows urban planners to put minimum thresholds in the number of trips per region, thus identifying the most popular cycling areas where parking enforcement changes would have more impact. As Figure 16 shows, the along-the-route (ALT) commuting regions heavily diminish when a minimum number of trips per month is required (in the Figure, that number is approximated to at least the average number of trips across all regions with cycling traffic). Thus, such map could be used, instead of Figure 15, to assess locations where assigning more parking enforcement officers would have the largest impact in improving cycling conditions. Continuing with the commuting example, urban planners could use the destination trip purpose map (Figure 17) to identify all the regions where the majority (or a minimum threshold) of the commuting trips have a given location as its destination, and propose to build bike parking facilities in those regions, which could potentially encourage more cyclists to commute to work.

Overall, these examples show that using the proposed decision support tool, urban planners could identify policy and infrastructure changes that, if implemented, would have a positive impact for the cyclist community in Philadelphia. But even more interesting, the decision support tool could also be used to assess the proposed policy and infrastructure changes. In fact, by collecting GPS traces from cyclists prior- and post-implementation of changes, the support tool would allow urban planners to assess the impact that the new policies or infrastructures have in the cycling behaviors, with the final aim of encouraging cycling in the city.

6 DISCUSSION

In this paper, we have presented novel methods to infer cycling trip purpose and to analyze cyclists' route choices, and we have evaluated them for the City of Philadelphia. Due to its cost, traditional, survey-based approaches to understanding cycling behaviors highly limit the frequency at which cycling information can be collected. The methods presented here are solely based on open and crowdsourced data that cities and cyclists are already collecting for other purposes. While cities worldwide are embracing the use of open data repositories [52], cyclists are using a plethora of mobile applications to collect their cycling GPS traces. As a result, the methods presented in this paper have two important implications: (1) cycling behaviors, and its changes over time, can be modeled and analyzed with higher frequency, and (2) cycling behaviors can now be studied at lower costs, since the datasets used by the proposed methods are already being collected for other purposes.

Our paper also offers a set of important insights for cities and cyclist advocacy groups willing to replicate these automatic approaches in other cities. We have shown that cycling trip purpose inference methods work best when not only the spatio-temporal data of the trip is considered, but rather when that data is enhanced with personal information about the cyclist, as well as social and built-in environment features. This result highlights the importance of (a) cities having access to a plethora of social and built-in environment features and (b) cyclists willing to share their obfuscated traces together with some personal information.

Our analyses for the City of Philadelphia have also shown that the mental map that cyclists form of their environment requires processing months of experiences and events, and are not formed in short periods of time. As a result, cities and cyclist organizations willing to use GPS traces collected by cyclists should make sure that the data collection period is long enough to be able to fully assess the role that different social and built-in environment features play in how cyclists perceive and choose the streets they go through. The inference by cyclist and demographic type has also shown the presence of certain biases in the classification, with worse classification rates for more unexperienced, younger and older cyclists. This result might point to the need to collect more GPS traces for these populations, as well as for longer periods of time, since their behaviors might be more entropic and thus harder to model in terms of trip purpose or route choice.

The route choice analyses discussed for the city of Philadelphia have confirmed many behavioral findings previously reported in survey-based studies. This has important implications for urban planners working on both cyclist infrastructure and the general layout of cities, since it provides a novel automatic method to analyze cyclist behavior and route choice at large scale, only using open and crowdsourced data. Finally, the decision support tool described with use cases for the City of Philadelphia could potentially aid urban planners in other cities to identify specific policy or infrastructure changes that would improve cycling experiences in their regions.

7 CONCLUSIONS

In this paper, we have presented novel methods to analyze two important issues characterizing urban cyclist behavior: trip purpose and route choice. We have proposed the use of spatio-temporal and personal cyclists' information as well as features characterizing the routes' street segments built-in and social environment to predict the purpose of a cycling trip. Our results using over 7,000 trips and their corresponding GPS traces for the City of Philadelphia show that the best approach yielded an F1-score of 86% using a combination of XGBoost and oversampling to classify across four trip purposes. We have also presented a statistical method to understand the role that various built-in and social environment features play in the way cyclists choose a route. We have shown that cyclists in Philadelphia tend to favor routes with green areas, cycling facilities and road centrality, which reflect general preference for healthier parts of the city, safer experiences through for example, protected bike lanes, and faster routes through central street segments. However, trip purpose might alter the relevance given to such features. In fact, our analyses reveal that commuting trips in Philadelphia favor faster routes while exercise trips focus on enjoyable routes; and that social or school trips give a lot of importance to finding routes through areas with low crime. Finally, we have described a decision support tool that shows that the methods presented are highly useful for urban planners willing to improve cycling experiences in their cities; and accessible to many cities since they are based on open and crowdsourced data.

REFERENCES

[1] Bruce Appleyard and Christopher Ferrell. 2015. The Meaning of Mean Streets for Sustainable & Active Travel: Crime,(Pedestrian/Bike) Casualties and Mode Choice. *Journal of Transport & Health* 2, 2 (2015), S59–S60.

[2] Kay W Axhausen, Stefan Schönfelder, J Wolf, M Oliveira, and U Samaga. 80. weeks of gps-traces: approaches to enriching the trip information (2003). (80).

[3] Shlomo Bekhor, Tomer Toledo, and Joseph N Prashker. 2008. Effects of choice set size and route choice models on path-based traffic assignment. *Transportmetrica* 4, 2 (2008), 117–133.
[4] Moshe Ben-Akiva, MJ Bergman, Andrew J Daly, and Rohit Ramaswamy. 1984. Modeling inter-urban route choice behaviour. In *Symposium on Transportation and Traffic Theory*. VNU Science Press Utrecht, The Netherlands, 299–330.
[5] Moshe Ben-Akiva and Michel Bierlaire. 1999. Discrete choice methods and their applications to short term travel decisions. *Handbook of transportation science* 23 (1999), 5–33.
[6] Moshe E Ben-Akiva, M Scott Ramming, and Shlomo Bekhor. 2004. Route choice models. In *Human Behaviour and Traffic Networks*. Springer, 23–45.
[7] Yoav Benjamini and Yosef Hochberg. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the royal statistical society. Series B (Methodological)* (1995), 289–300.
[8] Donald J Berndt and James Clifford. 1994. Using dynamic time warping to find patterns in time series. In *Proceedings of the 3rd International Conference on Knowledge Discovery and Data Mining*. AAAI Press, 359–370.
[9] Bicycle_Coalition. 2017. Bicycle Statistics for Philadelphia. http://www.arcgis.com/apps/Cascade/index.html?appid=6b42ffc54fb941df90cd0be7bd41aac3.
[10] Bicycle_Coalition. 2018. Bikes on Transit Bridges. https://bicyclecoalition.org/ride-your-bike/bikes-on-transit-bridges/.
[11] Wendy Bohte and Kees Maat. 2009. Deriving and validating trip purposes and travel modes for multi-day GPS-based travel surveys: A large-scale application in the Netherlands. *Transportation Research Part C: Emerging Technologies* 17, 3 (2009), 285–297.
[12] Piet HL Bovy and Stella Fiorenzo-Catalano. 2007. Stochastic route choice set generation: behavioral and probabilistic foundations. *Transportmetrica* 3, 3 (2007), 173–189.
[13] Ennio Cascetta, Agostino Nuzzolo, Francesco Russo, and Antonino Vitetta. 1996. A modified logit route choice model overcoming path overlapping problems. Specification and some calibration results for interurban networks. In *Internaional symposium on transportation and traffic theory*. 697–711.
[14] SF Catalano and N Van der Zijpp. 2001. A forecasting model for inland navigation based on route enumeration. In *Proceedings of the AET European Transport Conference*.
[15] Cynthia Chen, Hongmian Gong, Catherine Lawson, and Evan Bialostozky. 2010. Evaluating the feasibility of a passive travel survey collection in a complex urban environment: Lessons learned from the New York City case study. *Transportation Research Part A: Policy and Practice* 44, 10 (2010), 830–840.
[16] Tianqi Chen and Carlos Guestrin. 2016. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*. ACM, 785–794.
[17] Tanzeem Choudhury, Sunny Consolvo, Beverly Harrison, Jeffrey Hightower, Anthony LaMarca, Louis LeGrand, Ali Rahimi, Adam Rea, G Bordello, Bruce Hemingway, et al. 2008. The mobile sensing platform: An embedded activity recognition system. *IEEE Pervasive Computing* 7, 2 (2008).
[18] Chaushie Chu. 1989. A paired combinatorial logit model for travel demand analysis. In *Transport Policy, Management & Technology Towards 2001: Selected Proceedings of the Fifth World Conference on Transport Research*, Vol. 4.
[19] CodeForPhilly. 2017. Code for Philadelphia. http://www.cyclephilly.org/.
[20] Paolo Crucitti, Vito Latora, and Sergio Porta. 2006. Centrality in networks of urban streets. *Chaos: an interdisciplinary journal of nonlinear science* 16, 1 (2006).
[21] Marco Cuturi. 2011. Fast Global Alignment Kernels. In *Proceedings of the 28th International Conference on International Conference on Machine Learning (ICML'11)*.
[22] Cyclemeter. 2017. Cyclemeter. https://abvio.com/cyclemeter.
[23] Carlos F Daganzo and Yosef Sheffi. 1977. On stochastic models of traffic assignment. *Transportation science* 11, 3 (1977), 253–274.
[24] Zhongwei Deng and Minhe Ji. 2010. Deriving rules for trip purpose identification from GPS travel survey data and land use data: A machine learning approach. In *Traffic and Transportation Studies 2010*. 768–777.
[25] Tamara Denning, Adrienne Andrew, Rohit Chaudhri, Carl Hartung, Jonathan Lester, Gaetano Borriello, and Glen Duncan. 2009. BALANCE: towards a usable pervasive wellness application with accurate activity inference. In *Proceedings of the 10th workshop on Mobile Computing Systems and Applications*. ACM, 5.
[26] Richard Devine. 2017. Google Slips walking and cycling tracking card in google now. https://www.androidcentral.com/google-slips-walking-and-cycling-tracking-card-google-now.
[27] Vanessa Frias-Martinez, Cristina Soguero, and Enrique Frias-Martinez. 2012. Estimation of urban commuting patterns using cellphone network data. In *Proceedings of the ACM SIGKDD International Workshop on Urban Computing*. ACM, 9–16.
[28] Vanessa Frias-Martinez, Jesus Virseda, and Enrique Frias-Martinez. 2010. Socio-economic levels and human mobility. In *International Conference on Information and Communication Technologies for Development, Qual Meets Quant Workshop*.
[29] Vanessa Frias-Martinez, Jesus Virseda-Jerez, and Enrique Frias-Martinez. 2012. On the relation between socio-economic status and physical mobility. *Information Technology for Development* 18, 2 (2012), 91–106.

[30] Catherine O Fritz, Peter E Morris, and Jennifer J Richler. 2012. Effect size estimates: current use, calculations, and interpretation. *Journal of experimental psychology: General* 141, 1 (2012), 2.
[31] Alyson L Geller. 2003. Smart growth: a prescription for livable cities. *American journal of public health* 93, 9 (2003), 1410–1415.
[32] Lei Gong, Takayuki Morikawa, Toshiyuki Yamamoto, and Hitomi Sato. 2014. Deriving personal trip data from GPS data: a literature review on the existing methodologies. *Procedia-Social and Behavioral Sciences* 138 (2014), 557–565.
[33] Google. 2016. System, apparatus and procedure for providing route suggestions. https://patents.google.com/patent/US9255807B2/en.
[34] Mordechai Haklay and Patrick Weber. 2008. Openstreetmap: User-generated street maps. *IEEE Pervasive Computing* 7, 4 (2008), 12–18.
[35] Jin-Whan Jung and Gary G Koch. 1999. Multivariate non-parametric methods for Mann–Whitney statistics to analyse cross-over studies with two treatment sequences. *Statistics in medicine* 18, 8 (1999), 989–1017.
[36] Nicholas D Lane, Ye Xu, Hong Lu, Shaohan Hu, Tanzeem Choudhury, Andrew T Campbell, and Feng Zhao. 2011. Enabling large-scale human activity inference on smartphones using community similarity networks (csn). In *Proceedings of the 13th international conference on Ubiquitous computing*. ACM, 355–364.
[37] Guillaume Lemaître, Fernando Nogueira, and Christos K. Aridas. 2017. Imbalanced-learn: A Python Toolbox to Tackle the Curse of Imbalanced Datasets in Machine Learning. *Journal of Machine Learning Research* 18, 17 (2017), 1–5.
[38] Jure Leskovec and Rok Sosič. 2016. SNAP: A General-Purpose Network Analysis and Graph-Mining Library. *ACM TIST* 8, 1 (2016).
[39] James D Loudin and Hannu E Miettinen. 2003. A multivariate method for comparing n-dimensional distributions. In *Proceedings of the Conference on Statistical Problems in Particle Physics, Astrophysics and Cosmology (PHYSTAT)*. 207–210.
[40] Yijing Lu and Lei Zhang. 2015. Imputing trip purposes for long-distance travel. *Transportation* 42, 4 (2015), 581–595.
[41] Yijing Lu, Shanjiang Zhu, and Lei Zhang. 2012. A machine learning approach to trip purpose imputation in GPS-based travel surveys. In *4th Conference on Innovations in Travel Modeling, Tampa, Fla.*
[42] Mapbox. 2017. Map Matching API. https://www.mapbox.com.
[43] Jeff Mapes. 2009. *Pedaling revolution: How cyclists are changing American cities*. Oregon State University Press.
[44] MapMyRide. 2017. Map My Ride. www.mapmyride.com.
[45] Daniel McFadden. 1974. The measurement of urban travel demand. *Journal of public economics* 3, 4 (1974), 303–328.
[46] Patrick Tracy McGowen and Michael G McNally. 2007. *Evaluating the potential to predict activity types from GPS and GIS data*. Technical Report.
[47] Peter Midgley. 2011. Bicycle-sharing schemes: enhancing sustainable mobility in urban areas. *United Nations, Department of Economic and Social Affairs* (2011), 1–12.
[48] Lara Montini, Nadine Rieser-Schüssler, Andreas Horni, and Kay W Axhausen. 2014. Trip purpose identification from GPS tracks. *Transportation Research Record* 2405, 1 (2014), 16–23.
[49] MyTracks. 2017. My Tracks. https://play.google.com/store/apps/details?id=com.zihua.android.mytracks&hl=en.
[50] Paul Newson and John Krumm. 2009. Hidden Markov Map Matching Through Noise and Sparseness. In *Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. 336–343.
[51] Marcelo Oliveira, Peter Vovsha, Jean Wolf, and Michael Mitchell. 2014. Evaluation of two methods for identifying trip purpose in GPS-based household travel surveys. *Transportation Research Record: Journal of the Transportation Research Board* 2405 (2014), 33–41.
[52] OpenDataSoft. 2018. https://www.opendatasoft.com/a-comprehensive-list-of-all-open-data-portals-around-the-world/. [accessed May-2-2018].
[53] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12 (2011), 2825–2830.
[54] Francisco Pereira, Carlos Carrion, Fang Zhao, Caitlin D Cottrill, Chris Zegras, and Moshe Ben-Akiva. 2013. The future mobility survey: Overview and preliminary evaluation. In *Proceedings of the Eastern Asia Society for Transportation Studies*, Vol. 9.
[55] PhiladelphiaCityHall. 2017. Open Datasets for the City of Philadelphia. https://www.opendataphilly.org/.
[56] PhiladelphiaStreets. 2018. Contraflow Bicycle Lanes. https://www.philadelphiastreets.com/survey-and-design-bureau/bicycle-network/chop-contraflow-bicycle-lane/.
[57] Sergio Porta, Paolo Crucitti, and Vito Latora. 2006. The network analysis of urban streets: a dual approach. *Physica A: Statistical Mechanics and its Applications* 369, 2 (2006), 853–866.
[58] Sergio Porta, Paolo Crucitti, and Vito Latora. 2006. The network analysis of urban streets: a primal approach. *Environment and Planning B: planning and design* 33, 5 (2006), 705–725.
[59] Carlo Giacomo Prato. 2009. Route choice modeling: past, present and future research directions. *Journal of choice modelling* 2, 1 (2009), 65–100.
[60] Sasank Reddy, Min Mun, Jeff Burke, Deborah Estrin, Mark Hansen, and Mani Srivastava. 2010. Using mobile phones to determine transportation modes. *ACM Transactions on Sensor Networks (TOSN)* 6, 2 (2010), 13.
[61] RideReport. 2017. Ride Report Oregon. https://ride.report.

[62] Alberto Rubio, Vanessa Frias-Martinez, Enrique Frias-Martinez, and Nuria Oliver. 2010. Human Mobility in Advanced and Developing Economies: A Comparative Analysis.. In *AAAI Spring Symposium: Artificial Intelligence for Development*.

[63] Hiroshi Shimodaira, Ken-ichi Noma, Mitsuru Nakai, and Shigeki Sagayama. 2002. Dynamic time-alignment kernel in support vector machine. In *Advances in neural information processing systems*. 921–928.

[64] Peter Stopher, Camden FitzGerald, and Jun Zhang. 2008. Search for a global positioning system device to measure person travel. *Transportation Research Part C: Emerging Technologies* 16, 3 (2008), 350–369.

[65] Strava. 2017. Strava. www.strava.com.

[66] USGS. 2017. The National Map: Elevation Point Query Service. https://nationalmap.gov/epqs/.

[67] NJ Van der Zijpp and S Fiorenzo Catalano. 2005. Path enumeration by finding the constrained K-shortest paths. *Transportation Research Part B: Methodological* 39, 6 (2005), 545–563.

[68] Peter Vovsha. 1997. *The cross-nested logit model: application to mode choice in the Tel-Aviv metropolitan area.* Transportation Research Board.

[69] Huayong Wang, Francesco Calabrese, Giusy Di Lorenzo, and Carlo Ratti. 2010. Transportation mode inference from anonymized and aggregated mobile phone call detail records. In *Intelligent Transportation Systems (ITSC), 2010 13th International IEEE Conference on*. IEEE, 318–323.

[70] Chieh-Hua Wen and Frank S Koppelman. 2001. The generalized nested logit model. *Transportation Research Part B: Methodological* 35, 7 (2001), 627–641.

[71] Jean Wolf, Randall Guensler, and William Bachman. 2001. Elimination of the travel diary: Experiment to derive trip purpose from global positioning system travel data. *Transportation Research Record: Journal of the Transportation Research Board* 1768 (2001), 125–134.

[72] Jiahui Wu, Lingzi Hong, and Vanessa Frias-Martinez. 2018. Predicting Perceived Cycling Safety Levels Using Open and Crowdsourced Data. *IEEE International Conference on Big Data* (2018).

[73] Mingsheng Wu and Vanessa Frias-Martinez. 2015. Crowdsourcing Biking Times. In *Adjunct Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers*.

[74] Yu Zheng, Like Liu, Longhao Wang, and Xing Xie. 2008. Learning transportation mode from raw gps data for geographic applications on the web. In *Proceedings of the 17th international conference on World Wide Web*. ACM, 247–256.

Received May 2018; revised November 2018; accepted January 2019