What's Your Value of Travel Time?
Collecting Traveler-Centered Mobility Data via Crowdsourcing

Cristian Consonni,[1] **Silvia Basile,**[1] **Matteo Manca,**[1] **Ludovico Boratto,**[1] **André Freitas,**[2]
Tatiana Kovacikova,[3] **Ghadir Pourhashem,**[3] **Yannick Cornet**[3]

[1]Data Science and Big Data Analytics, EURECAT - Centre Tecnològic de Catalunya - Barcelona (Spain),
[2]TIS.pt - Consultores em Transportes Inovao e Sistemas, S.A. - Lisbon (Portugal),
[3]Dept. of International Research Projects - ERAdiate+, University of ilina - ilina (Slovakia)
cristian.consonni@acm.org, silvia.bas95@gmail.com, matteomanca@gmail.com, ludovico.boratto@acm.org,
andre.freitas@tis.pt, {tatiana.kovacikova,ghadir.pourhashem,yannick.cornet}@uniza.sk

Abstract

Mobility and transport, by their nature, involve crowds and require the coordination of multiple stakeholders - such as policy-makers, planners, transport operators, and the travelers themselves. However, traditional approaches have been focused on time savings, proposing to users solutions that include the shortest or fastest paths. We argue that this approach towards travel time value is not centered on a traveler's perspective. To date, very few works have mined data from crowds of travelers to test the efficacy and efficiency of novel mobility paradigms. In this paper, we build upon a different paradigm of *"worthwhile time"* in which travelers can use their travel time for other activities; we present a new dataset, which contains data about travelers and their journeys, collected from a dedicated mobile application. Each trip contains multi-faceted information: from the transport mode, through its evaluation, to the positive and negative experience factors. To showcase this new dataset's potential, we also present a use case, which compares corresponding trip legs with different transport modes, studying experience factors that negatively impact users using cycling and public transport as alternatives to cars. We conclude by discussing other application domains and research opportunities enabled by the dataset.

Introduction

The extraction of actionable knowledge from user mobility has been a central perspective mainly for mobility stakeholders - such as policy-makers, planners, and transport operators - who, in turn, used this knowledge to adapt the services to the end users. This process, known as behavioral-data mining, aims at extracting patterns from the travelers' behavior, in order to better characterize them (Manca, Boratto, and Carta 2018; Boratto et al. 2018). Under this paradigm, end-users have mostly been passive actors.

In the last 20 years, the integration between information and communication technologies (ICT) and transport have radically changed mobility: from the mere availability of information about transport such as bus lines or time tables, we now have new forms of offerings in terms of on-demand services and even whole new business models for transport,

where new services offer the opportunity to share rides or even private cars. New Web platforms, such as journey planners (Sourlas and Nathanail 2019), have transformed users into *active actors* in providing their mobility preferences. Users can sort trips based on different types of preferences (e.g., length, duration, emissions) and decide to complete a trip by combining services of different operators. Hence, novel mobility paradigms emerged, in which users' *value of travel time* (VTT) is defined as the combination of different factors (e.g., a trip by bike could be longer than taking other means of transport, but might valuable from multiple perspectives, such as emissions, costs, and fitness for the user) (Devarasetty, Burris, and Douglass Shaw 2012).

However, other aspects of transportation, such as the reliance upon private cars have not changed. We argue that one of the main opportunities offered by the integration between ICT and transport lies in the possibility of gather data from a vast audience of travelers about both their travel preferences and wishes and their actual day-to-day transport usage. In this sense, crowdsourcing, intended as the process of collecting data contributed by vast amounts of people, mainly via the Internet, can become a valuable means towards the accomplishment of this goal.

While novel paradigms related to value of travel time have been introduced (Karadimce, Lugano, and Cornet 2018; Kováčiková, Lugano, and Pourhashem 2018; Lugano et al. 2019), existing studies capture user behavior in mobility from a single perspective. When moving from academic research to industrial applications, a one-to-one relationship between value of travel time and shortest path exists, so that existing services - such as Google Maps and Waze - mainly rank trips by shortest path. We believe that this might be due - at least from the research side - to the lack of knowledge on *what is actually valuable for the users in their mobility choices*. While several existing datasets capture mobility by considering a single perspective (e.g., one transport mode, or only trip coordinates), *no dataset capturing both explicit preferences in terms of value for the users when making their mobility choices and implicit information coming from the trip (e.g., coordinates) exist*. This, in turn, has been reflected in industrial applications, which do not capture explicit feedback on additional factors. Hence, value of travel

time remains a concept trapped inside the mobility community, which cannot be fully exploited by the Web community, to create actionable knowledge for Web platforms. We believe that **crowdsourcing can be a powerful tool to directly collect feedback on the travelers' value of travel time and their mobility choices.**

To overcome these issues, in this paper we present the *MoTiV* dataset, a collection of traveler-centered information about trips and the value of travel time behind traveler's choices.

To show how the extraction of knowledge on users' value of travel time can concretely impact transport stakeholders, we also present a use case that characterizes what are the negative factors associated to a trip, when this is performed by means of public transport or by cycling, w.r.t. the same trip performed by car.

This dataset was collected in the context of the Horizon 2020 *MoTiV* (Mobility and Time Value) project, whose goal was to provide novel definitions of value of travel time (VTT) (Malichová et al. 2020).

Related Work

In this section, we present datasets related to the one presented in this paper, and conclude by highlighting the difference between our dataset and the existing ones.

User Behavior Mining

Data coming from online platforms was previously used to mine user behavior, to consider aspects such as their willingness to pay for services (Zografos, Androutsopoulos, and Apospori 2012) and challenges in behavior change (Schrammel et al. 2015). Manca et al. (Manca et al. 2017) presented a survey on mobility patterns considering social media data. González, R., and Barabási (González, R., and Barabási 2008), instead, consider mobile phone data, while Calabrese et al. (Calabrese et al. 2013) extract mobility patterns from urban sensing data. Goulias (Goulias 2018) surveyed the existing travel behavior models. Data analysis can produce insights that serve as input for other purposes, such as the improvement of transport services by considering user needs (Sierpiński and Staniek 2017), the promotion of changes of the user habits (Schrammel et al. 2015), and the improvement of journey planners and transport portals (Esztergár-Kiss 2016; Vargas, Weffers, and da Rocha 2011). Other studies go beyond data analysis, e.g., to extract topic models from geo-location data (Hasan and Ukkusuri 2014), to forecast the evolution of preferences over time thanks to a Markov model (Zarwi, Vij, and Walker 2017), or to provide a personalized journey planning (Jakob et al. 2014). As previously mentioned, the knowledge coming from the user analytics can also be used as a form of actionable knowledge, e.g., to improve transport service according to the user needs (Sierpiński and Staniek 2017), to promote changes in the user habits (Schrammel et al. 2015) such as the adoption of greener and healthier solutions (Gabrielli et al. 2014), or improve the usability and services provided by journey planners and transport portals (Esztergár-Kiss 2016; Vargas, Weffers, and da Rocha 2011).

Crowdsourced Datasets

Social and trip data Microsoft GeoLife is a social network that also allows users to share their experience, both through GPS data and with pictures. The publicly available GPS trajectory dataset was collected by 182 users, for over two years[1]. No explicit info about the means of transport is available, and no evaluation of the trips is offered.

Check-in data Check-in data, coming from platforms such as Twitter, Foursquare, or Gowalla, is usually used to consider user preferences related to their mobility. Indeed, knowing where the users go and with which frequency, allows to characterize users' mobility and their preferences (especially if check-ins can be paired with reviews). Examples of datasets belonging to this class can be found here [2]. While this class of datasets is widely employed in various personalization algorithms, the concept of a trip is entirely missing, thus losing all the information about the means of transport, or the relevant factors for the users.

Trip-only data Another class of datasets collect trip information. A selection of 250 datasets available for research purposes can be found here[3]. Trip datasets usually do not associate the trips to a user, thus not allowing a characterization of the individual user mobility. In addition, they are usually associated to a single transport of mode (e.g., taxis).

Contextualizing our Contributions

As this analysis has presented, mobility datasets usually *capture a single perspective* and usually the full user-mobility picture as lost. The aim of our data collection is to capture multiple perspectives behind user mobility. This is done by:

- having trips with different transport modes (something not available in classic trip datasets);
- collecting trip coordinates (this is usually not available in check-in datasets);
- collecting reviews and evaluations of each trips;
- working at a lower granularity, by not only considering the trip as a whole, but also the activities that compose it; This allows to consider the different experience factors in user mobility, at different stages of the trip.

Background and Crowdsourcing Process

Worthwhile time. *Worthwhile time is a central concept for our data collection. While the traditional view is to consider travel time as something to minimize, we consider travel time as an opportunity, i. e. time that that*

---

[1]https://www.microsoft.com/en-us/research/project/geolife-building-social-networks-using-human-location-history/#!downloads

[2]https://sites.google.com/site/yangdingqi/home/foursquare-dataset

[3]https://data.world/datasets/trips

can be characterized by other activities. We have collected data through a dedicated mobile app, called Woorti[4]*, to ground the definition of worthwhile time into multiple dimensions.*

A key challenge of collecting a European-wide dataset was engaging users across multiple countries for an extended time. In fact, from a traveler's point-of-view, the data collection translated to using the smartphone app actively for at least 14 days. We framed the process as a challenge between user: data-collection campaigns (DCCs) would group users together and reward the most active ones in each campaign. To facilitate the process, we have partnered with thematic organizations across Europe - such as cycling associations - and we have appointed DCC managers as liaison with the community, whose tasks were recruiting participants, promoting and monitoring the data collection process. The application supported both Android and iOS devices and was available in 11 languages. Data collection targeted 10 European countries: Belgium, Croatia, Finland, France, Italy, Norway, Portugal, Slovakia, Spain, and Switzerland. Overall, the data considered in this paper cover a period of 8 months, from May 1st, 2019 to December, 13rd 2019.

The use of the app consists of three main phases:

1. *Onboarding*: upon installing the application and registering a new account, the user is introduced to the functionalities of the app. During this process, the user enters their travel preferences as well as some basic demographic information.

2. *Trip recording*: the user can start a new trip and the app automatically collects data in background.

3. *Trip validation*: when a trip is finished, the user can review the data, validate it and insert other data regarding the trip (trip purpose, mood, etc). When validating a trip, the user must choose one leg of the trip as the reviewed leg.

Within the application, a user can access their data and visualize and edit their profile information and trips. Furthermore, the application features a dashboard that presents to the user statistics related to their validated trips.

The value of travel time is analyzed from a travelers perspective, assuming that time and cost savings are not always the main criteria influencing route and mode choice. Depending on the traveler's transport attitude and context, other criteria such as environmental impact, comfort, or even weather conditions may influence the perceived value of a trip. In particular, we adopt the perspective that travel time can be worthwhile, i. e. it can be allocated for activities that the user finds useful, enjoyable, or productive. For this reason, when validating a trip, the user is asked about which activities they have performed during the trip, which factors in their have influenced the trip positively or negatively, and which was the trip purpose.

_____
[4]The name of the app is a play on the words "worth it" referring to worthwhile travel time.

By collecting this data, we are able to shifts perspective from considering travel time as spent - or, worse, wasted - to time that can be characterized by other activities. Furthermore, this characterization is not limited by defining time as productive or unproductive time, because it is not necessarily related to its evaluation in terms of cost. Worthwhile time is independent of what can be monetized.

The *MoTiV* dataset and the analyses presented here contain only validated trips. During data collections, local campaign managers stressed to users that validating a trip meant sharing it with the project, i. e. validated trips were uploaded to a central server. This process preserves users' privacy since they were always in control of what data they are willing to share with the project. Furthermore, the data collection process was overseen and approved by an independently appointed ethics advisor and the ethics committee at the University of Žilina, *MoTiV*'s coordinator.

Dataset Description

Dataset and soruce code. *The dataset can be downloaded at:*
https://zenodo.org/record/4027465.
The code used for pre-processing the raw data and performing the case study is available at:
https://github.com/MoTiV-project/data-analysis.

The dataset is comprised of 13 tables saved as comma-separated value (.csv) files:

1. user_details.csv
2. user_generic_worthwhileness_values.csv
3. user_specific_worthwhileness_values.csv
4. mots.csv
5. trips.csv
6. legs.csv
7. legs_coordinates.csv
8. activities.csv
9. experience_factors.csv
10. purposes.csv
11. weather_legs.csv
12. weather_raw.csv
13. worthwhileness_elements_from_trips.csv

User Details The table user_details.csv (1) contains the data collected during the onboarding phase. Only the fields user_id, registration_timestamp, gender, and age are required.

- user_id: the user identifier, a string of 29 characters.
- registration_timestamp: the trip start date, formatted as %Y-%m-%d %H:%M:%S.%f.

- `gender`: the gender of the user. Three values are possible: `Other`, `Male`, `Female`.

- `age`: the user's age in a range. Possible values are: `16-19`, `20-24`, `25-29`, `30-39`, `40-49`, `50-64`, `65-74`, `75+`.

- `language`: the language used within the app expressed as a 3-letter ISO 639-2/B code.

- `city`: the user's city of residence.

- `country`: the user's country of residence expressed in ISO 3166-1 alpha-3 code.

- `education_level`: the user's education, possible values are: `Basic (up to 10th grade)`, `High School (12th grade)`, and `University`.

- `marital_status`: the user's marital status, possible values are: `Single`, `Registered relationship`, `Married`, `Divorced`, `Widowed`.

- `number_of_people`: the number of people in the household, it can be a number from 1 to 4 or 5+.

- `labour_status`: the users' labour status. Possible values are: `Student`, `Employed full-time`, `Employed part-time`, `Pensioner`, `Unemployed`.

- `years_of_residence`: the number of years of residence in the household. Possible values are: `Less than 1`, `1 to 5`, `More than 5`.

User's Generic and Specific Worthwhileness Elements
Users' preferences and experiences are characterized along the three dimensions of fitness, enjoyment, and productivity, defined as follows (in parenthesis, we report a description of each dimension, visualized by users during data collection):

- fitness measures how much the user values the fact that when traveling they can exercise (When you walk, cycle, or even run on your travels, you are getting exercise and keeping in shape);

- enjoyment is related to how the travel can be used for fun or relaxing activities (Relaxing or having fun: taking time to listen to music, rest or meditate; engaging in social media; observing the surroundings);

- productivity captures how much the user values the possibility of using travel time to complete some tasks, either personal or work-related. (Using travel time to get things done, not only for work or study, but also personal things like managing home or family stuff). It is further diving in two aspects: *Paid work* and *Personal tasks*.

User preferences and experiences are encoded in two main sets of values, called worthwhileness[5] values:

- *generic worthwhileness values:* they are a triplet of values (F, E, P) for *fitness*, *enjoyment*, and *productivity*, respectively. They measure how much the user values these dimension in general when traveling;

- *specific worthwhileness values:* they are triplets of values (F, E, P) that the user is asked to assign for each specific mode of transport chosen in the onboarding phase. The transport modes that the user selects during the onboarding phase are called *preferred transport modes*. Specific worthwhileness values are the measure of how much the user values *fitness*, *enjoyment*, and *productivity* when using that particular transport mode.

During the onboarding phase, the user is asked to provide both the generic and specific worthwhileness values on a scale from 1 to 100. When evaluating trips, the user is asked to provide an evaluation for each dimension of *fitness*, *enjoyment*, and *productivity* using a scale from low to high (`low`, `medium`, `high`). This difference in data collection depended on the design of the app, whose description is out of scope for this paper. For consistency with the evaluation values, we scale the onboarding values to the same three classes: `low`, for values in $[0 - 33]$; `medium`, $[34 - 66]$; and `high`, $[67 - 100]$.

Generic worthwhileness values are stored in the table `user_generic_worthwhileness_values.csv` (2), its columns are:

- `userid`, the user identifier;

- `fit`, the value for fitness $[0 - 100]$;

- `prod`, the value for productivity $[0 - 100]$;

- `enjoy`, the value for enjoyment $[0 - 100]$;

Specific worthwhileness elements are related to specific mode of transports, chosen by the user during the *Onboarding* phase. These are referred as the traveler's favorite modes of transport. Specific worthwhileness values are stored in the table `user_specific_worthwhileness_values.csv` (3). Its columns are:

- `userid`, the user identifier;

- `motid`, the mode of transport identifier;

- `fit`, the value for fitness $[0 - 100]$;

- `prod`, the value for productivity $[0 - 100]$;

- `enjoy`, the value for enjoyment $[0 - 100]$;

The mapping between mode of transport ids and text is contained in `mots.csv` (4).

Trips Info Trip data are contained in the table `trips.csv` (5), whose fields are reported in Table 1. A trip is a collection of *legs* and *waiting events*: the former are parts of a journey where the app has detected some movement, while waiting events are intervals of time where the app did not detect any significant displacement. We processed trip data to merge legs that were erroneously split by the app, furthermore we performed outlier detection to eliminate the trips that were in top and bottom percentiles in terms of length and duration.

---

[5]Although this diction of the word is less widespread than the more common variant "*worthiness*," it is used throughout the project, so we keep it for consistency with the project itself.

[5]The app offers mode detection algorithms to infer the type of mode of transport used; however during the trip validation ad evaluation the user was able to override the modes suggested.

Field	Description and admissible values
tripid	Trip identifier, a string in the format #{n1}:{n2} where n1 and n2 are two numbers.
userid	User identifier.
start_date	Trip start date, formatted as '%Y-%m-%d %H:%M:%S.%f'.
end_date	Trip end date, formatted as '%Y-%m-%d %H:%M:%S.%f'.
average_speed	Average speed during trip in km/h.
max_speed	Max speed during trip.
distance	Leg distance in meters.
duration	Leg duration in seconds.
mood_rating	Evaluation of trip mood on a scale from 1 to 5.
did_you_have_to_arrive	Answer to the question *"Did you have to arrive on time"*, a boolean value.
how_often	Answer to the question *"How often do you make this trip?"*, possible values are integers from 0 to 3. −1 means that the question was not answered.
use_trip_more_for	Answer to the question *"How often do you make this trip?"*, possible values are integers from 0 to 3. −1 means that the question was not answered.
manual_start	Answer to the question *"Has the trip recording been started manually?"*, a boolean value.
manual_end	Answer to the question *"Has the trip recording been ended manually?"*, a boolean value.
validation_date	Trip validation date, formatted as '%Y-%m-%d %H:%M:%S.%f'.
os	Operating system of the phone, can be iOS or Android, a string.
os_version	Operating system version, a string.
model	Model of the phone, a string.

Table 1: Trip information stored in trips.csv.

Field	Description and admissible values
legid	Leg identifier, a string in the format #{n1}:{n2} where n1 and n2 are two numbers.
class	Leg type, either Leg or WaitingEvent.
userid	User identifier.
tripid	Trip identifier.
motid	Mode of transport identifier, an integer number.
start_date	Leg start date, formatted as %Y-%m-%d %H:%M:%S.%f.
end_date	Leg end date, formatted as %Y-%m-%d %H:%M:%S.%f.
true_distance	Leg distance in meters.
leg_distance	Leg distance in meters.
leg_duration	Leg duration in seconds.
worthwhileness_rating	Worthwhileness rating, −1 if not set.
transport_category	Transport category.
campaign	Data collection campaign.
weekday	Day of the week when the leg was performed.
weekday_class	Classification of the day, either Working_day or Weekend.

Table 2: Leg information stored in legs.csv.

Legs Info Leg data are contained in the table legs.csv (6), whose fields are reported in Table 2.

Legs Coordinates Leg coordinates are contained in the table legs_coordinates.csv (7), whose fields are reported in Table 3. Coordinates are anonymized depending on the fact that the point is located in a urban, suburban or rural area. To classify the points we used the "functional urban areas by country" classification provided by the Organisation for Economic Co-operation and Development (OECD).[6]

Points were anonymised using the following criterion, applied both to latitude and longitude:

- for urban areas, points are rounded to the third decimal place;

_____
[6]Functional urban areas by country, https://www.oecd.org/regional/regional-statistics/functional-urban-areas.htm

- for sub-urban areas, the third decimal place is rounded to the nearest 0.5;

- for rural areas, the second decimal place is rounded to the nearest 0.5;

Activities Data about activities performed during trips are contained in table activities.csv (8). A user can select multiple activities during a trip, but we do not record their duration. Columns in activities.csv are:

- legid, leg identifier

- activity, activity name, possible values are Accompanying, Browsing, Cycling, Driving, Eating, Listening, PersonalCare, ReadingDevice, ReadingPaper, Relaxing, Talking, Thinking, Walking, Watching, and Other

Field	Description and admissible values
leg_id	The identifier of the leg in the format #n1:n2, with n1 and n2 being two integer numbers.
start_lat	a real number with 3 decimals. The latitude of the starting point of the leg.
start_lon	a real number with 3 decimals. The longitude of the starting point of the leg.
end_lat	a real number with 3 decimals. The latitude of the ending point of the leg.
end_lon	a real number with 3 decimals. The longitude of the ending point of the leg.
start_name	The inferred name of the city where the starting point of the leg is located. This field is obtained in the preprocessing by integrating the leg coordinates with OECD data.
start_country	The ISO 3166 Alpha-3 code of the inferred country where the starting point of the leg is located. This field is obtained in the preprocessing by integrating the leg coordinates with OECD data1b.
start_class	Possible values are: urban, sub-urban, or rural. The classification of the starting point of the leg. This field is obtained in the preprocessing by integrating the leg coordinates with OECD data1c.
end_name	The inferred name of the city where the ending point of the leg is located. This field is obtained in the preprocessing by integrating the leg coordinates with OECD data.
end_country	The ISO 3166 Alpha-3 code of the inferred country where the ending point of the leg is located. This field is obtained in the preprocessing by integrating the leg coordinates with OECD data1b.
end_class	Possible values are: urban, sub-urban, or rural. The classification of the ending point of the leg. This field is obtained in the preprocessing by integrating the leg coordinates with OECD data1c.

Table 3: Legs coordinates stored in `legs_coordinates.csv`

Experience Factors Table `experience_factors.csv` (9) contains information about experience factors affecting trips, its columns are:

- `legid`, leg identifier;

- `factor`, experience factor name. There are 48 possible values, for context we list here a sample: `Air_Quality`, `Cleanliness`, `Crowdedness_Seating`, `Internet_Connectivity`, `Privacy`, `Reliability_Of_Travel_Time`, `Todays_Weather`, `Toilets`, `Vehicle_Quality`, `Vehicle_Ride_Smoothness`;

- `type`, factor categorization;

- `minus`, a boolean value, the factor was rated as negative;

- `plus`, a boolean value, the factor was rated as positive.

Trip purposes The table `purposes.csv` (10) contains information about trips purposes, its columns are:

- `tripid`, the trip identifier;

- `purpose`, the trip purpose, possible values are `Business_Trip`, `Everyday_Shopping`, `Home`, `Leisure_Hobby`, `Personal_Tasks_Errands`, `Pick_Up_Drop_Off`, `Work`, and `Other`.

Weather data Weather data has been collected through the API provided by OpenWeatherMap.[7] The API was queried regularly for a set of 66 cities of interest in the scope of the project. Weather information was collected for the times of 09:00, 12:00 and 18:00 for each day from July 8th, 2019 to December 18th, 2019. The dataset contains two tables related to weather data: `weather_raw.csv` (11) and `weather_legs.csv` (12).

The table `weather_raw.csv` contains the data parsed has obtained from the OpenWeather API, the documentation of each field is available on the OpenWeatherMap website.[8] The table `weather_legs.csv` contains the association between trips legs and the corresponding weather for the time and place, the available fields are presented in Table 4.

For the computation of weather scenarios, different sources of data were combined. Specifically, we computed the apparent temperature, which is equivalent of the temperature perceived by humans, based on effects of air temperature, relative humidity and wind speed. The formula for apparent temperature introduced by Robert Steadman in 1984 (Steadman 1984) was used, which takes into consideration four environmental factors: wind, temperature, humidity and radiation from the sun.

Worthwhileness elements from trips Table `worthwhileness_elements_from_trips.csv` (13) contains the evaluation of each trip along the worthwhileness elements dimensions on scale from 0 (low) to 2 (high). The fields of the table are:

- `tripid`, the trip identifier;

- `legid`. the leg identifier;

- `worthwhileness_element`, name of the worthwhileness element. Possible values are `Enjoyment`, `Fitness`, `Paid_work`, `Personal_tasks` and `Unknown`;

- `value`, the value of the worthwhileness element, possible values are: 0, 1, 2 and −1;

Records where the worthwhileness element and the value is −1 are the same.

A User's Story

In the following section, we present a brief user story of a user using the dedicated app to describe the data collected in context. We will call our user Luigi.

---

[7]https://openweathermap.org/, OpenWeatherMap is an online service owned by OpenWeather Ltd.

[8]OpenWeatherMap historical weather API guide, https://openweathermap.org/history.

Field	Description and admissible values
weatherid	Weather identifier, a string of 24 characters.
legid	Leg identifier.
request_date	Weather request timestamp, formatted as '%Y-%m-%d %H:%M:%S.%f'.
centroid_x	Longitude of the centroid of the leg.
centroid_y	Latitude of the centroid of the leg.
country	Country the leg is located in.
weather_scenario	Classification of weather scenarios.
apparent_temperature	Apparent temperature, computed using Steadman's equation.
net_radiation	Net radiation received by the terrain in the location of the centroid.
temperature_category	Categorization based on apparent temperature.
temperature_description	Description of the temperature based on apparent temperature.
cloud_category	Categorization based on cloud cover.
cloud_main	Description based on cloud cover.
precipitation_category	Categorization based on precipitation level.
precipitation_main	Description based on precipitation level.
wind_beaufort_number	Categorization using Beaufort's number, based on wind speed.
wind_category	Categorization based on wind speed.
wind_description	Description based on wind speed.

Table 4: Weather info stored in `weather_legs.csv`.

One day, Luigi sees an advertisement of a new app just released: *Woorti*. *Woorti* keeps track of personal trips along with all the aspects that influenced the travel, it seems interesting! So he decides to download it. He registers to the app and inserts some demographic information (**user_details**). After that, he is asked how much he values three aspects of life while he travels: enjoyment, how much fun he has; productivity, if he is able to work or do personal tasks; and fitness, if he is able to exercise and stay healthy; both in general terms (**user_generic_worthwhileness_values**) and with respect to his favorite transport modes (**user_specific_worthwhileness_values**).

Luigi wants to try the app while going to work (**purposes**). He starts from his apartment, walks until the bus stop where he waits for 5 minutes the bus. He gets on the bus for 20 minutes and in the meantime he reads a book and use the smartphone to check something on internet (**activities**). He easily finds a seat and enjoys his time on the bus since it is not crowded and he likes to watch the landscape outside (**experience_factors**). The bus stop is just few minutes from his office and since it's a beautiful and sunny day (**weather**) he decides to walk until there. Before starting his working day, he stops at the bar to take a coffee and spends some minutes filling the questionnaire on the app after his trip. He validates the trip checking that the start and end points were correct (**legs_coordinates**): he did in total four legs during his trip (walking, waiting event, bus, walking) (**legs**). He decides to review his leg on the bus and he completes all the questions, giving also a rate for the whole trip and for the leg he reviewed (**trips**), included a score for the enjoyment, productivity and fitness (**worthwhileness_elements_from_trips**).

Case Study

We find users that have traveled a given route multiple times using both cars and alternative modes of transport and look at which are the factors that have impacted negatively the travel experience when using bikes or public transport.

> **Experience factors impacting negatively use of public transport, cycling and walking versus private cars** *We leverage the experience factor evaluations to compare trips done by cars versus trips conducted with other modes of transport: public transport, cycling and other emerging micromobility, and walking. With this use case we want to answer the question: "What are the negative experience factors of cyclists and users of public transport for the same trip legs performed by car?".*

In this way, we want to get some insights on which are the experience factors that are hindering the use of modes of transports alternative to cars. This analysis is performed along the following steps:

1. we select trip legs performed by car and those users that have performed at least one such trips;

2. we restrict this set to the users that have chosen at least one preferred transport mode within the transport categories: *cycling and emerging micromobility*, *public transport (short distance)*, and *public transport (long distance)*. For the sake of the study we call these the alternative transport category

3. we identify users that have performed similar trips using different transport modes; specifically, we select users that performed the same journey using a car - in the *private motorized* category - or using a bike or public transport.

4. For the same user, we look at negative experience factors for trip legs performed using modes of transports in the alternative categories.

For the definition of similar trips we match trips using their starting and ending points: When using different modes

of transport can lead to taking different paths, but for our analysis we are interested in the fact that a user needs to travel from a given pair of locations. To identify points that are close in space we adopt the following procedure, illustrated in Figure 1: each point is transformed to a curved square, where each side is an arc of length 0.004 degrees. If the squares representing two points intersect, we consider them a matching pair. To simplify the computational complexity of the matching process we only compare trip legs in the same country. To estimate the distance between these two point we use an approximate conversion between the precision of decimal degrees in the EPSG:7030/WGS 84,[9] 0.001 degrees correspond $78.71\,\mathrm{m}$ so at most two points are matching if they are within a maximum distance of $445\,\mathrm{m}$:

$$D = \sqrt{2} \cdot 0.004\,\mathrm{deg} \cdot 78.71\,\mathrm{km/deg} \approx 445\,\mathrm{m}$$

Table 5 presents some statistics related to each step of the process as presented above. For brevity, we will refer to trip legs performed with a mode of transport in the category *private motorized* as legs performed "by car" and trips in the categories *cycling and emerging micromobility*, *public transport (short distance)*, *public transport (long distance)* as legs performed "by alternative modes of transport."

Tables 6 and 7 present the results of our case study: Table 6 contains the overall top-ten negative experience factors for all trip legs in the categories *cycling and emerging micromobility*, *public transport (short distance)*, and *public transport (short distance)*. Table 7 presents the top negative experience factors only for the trip legs select by our case study for the same transport categories.

When cycling we find two main areas of concern: safety (availability of bicycle paths, safety from other cars, visibility and traffic signals) and quality (noise level, air quality). Road path directness has a somewhat a more important role when cycling is used as an alternative to traveling by car w.r.t. the general negative experience factors. Weather is ranked among the top-10 negative experience factors in

---

[9]http://epsg.io/7030-ellipsoid, taken as the E/W at 45 degrees N/S.

both bases. For short-distance public transport the main obstacles are lack of privacy and crowdedness in many forms (including noise level and air quality), while reliability of travel time do not appear in the top-10.

Research Opportunities

Our dataset can support numerous research initiatives, with applications that can provide benefits both to the end users and to transport stakeholders, such as transport operators.

Cost-benefit Analyses The approach showed in this study challenges conventional cost-benefit narratives and paradigms on the value of travel time, which is the current bedrock of how policy decision in the realm of transport are made. We envision that the "worthwhile time" approach can be further developed to become a viable alternative that takes more into account the travelers' perspective, especially when addressing significant infrastructure investments.

User profiling and clustering Crossing user mobility with their experience factors can be directly used to profile and cluster the users (Basile et al. 2020). The identification of users with similar behavior and a similar value of travel time might directly impact the shaping of journey planners, with solutions that target that cluster being presented first (e.g., if a user belongs to a cluster associated with low emissions, a sorting by emissions might be the default options).

Recommender systems Current systems associate mobility to Point-of-Interest recommendation (Liu et al. 2017). While in this domain collaborative filtering can be enriched with geographical information, our dataset can offer much richer notions of peer users. With the new prospective provided by our analysis, a peer user is not only someone who visited PoI similarly to another, but can be someone who gives values to the same experience factors. In addition, our dataset can enable novel forms of recommendation, based on the previous observations, such as the suggestion of activities to perform given a type of trip (e.g., reading in trains).
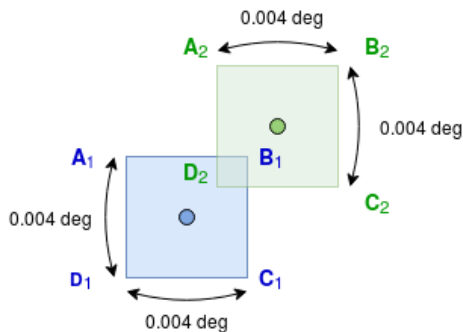

Figure 1: Procedure to identify close points in space: each point is transformed to a curved square, where each side is an arc - with sides of length 0.004 degrees. If the squares of two points intersect, we consider them a matching pair.

Step	Step description	Trips	Legs	Users
1	All trips, legs, users	64,098	158,897	3,269
2	Users that have performed at least one trip leg by car	21,764	62,227	2,083
3	Users that have selected at least one preferred alternative transport mode	51,973	132,259	1,771
4	Users that have performed at least one trip leg by car and by alternative transport modes	20,032	27,921	1,376

Table 5: User statistics considered by our case study.

Cycling and micromobility		Public transport (short distance)		Public transport (long distance)	
Factor	#	Factor	#	Factor	#
Cars other vehicles	1,751	Privacy	500	Internet connectivity	53
Air Quality	1,277	Crowdedness seating	499	Privacy	49
Road path availability and safety	1,219	Other people	463	Seating quality personal space	44
Noise level	1,097	Seating quality personal space	445	Reliability of travel time	43
Road path quality	969	Noise level	437	Noise level	36
Traffic signals crossings	854	Internet connectivity	418	Other people	35
Crowding congestion	837	Charging opportunity	406	Vehicle ride smoothness	34
Today's weather	591	Air quality	349	Today's weather	34
Simplicity difficulty of the route	417	Scenery	309	Crowdedness seating	33
Facilities shower lockers	371	Reliability of travel time	290	Food drink available	31

Table 6: Overall top-ten negative experience factors for all trip legs in the categories cycling and emerging micromobility, public transport (short distance), and public transport (short distance).

Cycling and micromobility		Public transport (short distance)		Public transport (long distance)	
Factor	#	Factor	#	Factor	#
Road path availability and safety	315	Privacy	133	Today's weather	2
Road path quality	307	Other people	111	Reliability of travel time	2
Cars other vehicles	267	Air quality	103	Vehicle ride smoothness	1
Air quality	173	Noise level	100	Ability to do what I wanted	1
Noise level	132	Crowdedness seating	99	Privacy	1
Road path directness	131	Seating quality personal space	99	Other people	1
Today's weather	123	Today's weather	84	Cleanliness	1
Traffic signals crossings	122	Internet connectivity	77		
Crowding congestion	111	Charging opportunity	76		
Lighting visibility	88	Scenery	67		

Table 7: Top negative experience factors for the trip legs selected by our case study.

Ad targeting Most platforms nowadays survive thanks to advertisements, which are usually personalized based on the experience of the users in a platform (Saia et al. 2016). It is clear that in the context of value of travel time, ad targeting should go beyond this, to consider the factors that positively or negatively impact user mobility, and tailor ads around them. These ads would be much more effective, thus benefiting the platform, and by presenting the users with "complementary" services, based on their preferences.

Conclusions and Future Work

Extracting actionable knowledge from user behavior in traveling can help knowing them more and providing them with better services (e.g., personalized rankings). For this reason, it is important not only to monitor the users, but to understand them and the choices they make, according to the value of their travel time. Plus, trips are complex entities, made up of several legs, which might disclose information about user behavior at a finer granularity.

To accomplish the goal of knowing users, their travel preferences and their value of travel time, we presented a mobility dataset collected via an app and we performed an analysis of factors impacting negatively the usage of cycling and public transport. Our dataset collects information about user mobility, capturing both raw information about trips and their legs (e.g., coordinates, time, and weather), plus information about the worthwhileness of a trip leg.

To assess the impact that our dataset can have in the real-world, we presented also a use-case to analyze what are experience factors that can negatively impacting the use of public transport, by comparing cycling and walking versus private cars. As our use case has shown, our dataset can provide valuable information to transport operators and service providers, so that their services can be tailored on the needs of their users. Our deliverable also presents opportunities in other domains, such as personalized recommendation or customer clustering. In addition, we plan to use the dataset to take action in concrete real-world scenarios, such as the design of cost-benefit analyses, to see how to improve infrastructures and services, thanks to the needs and preferences of the users.

Acknowledgements

lection managers for their invaluable support in the coordination of the activities in each country.

References

Basile, S.; Consonni, C.; Manca, M.; and Boratto, L. 2020. Matching User Preferences and Behavior for Mobility. In *ACM HT 2020*.

Boratto, L.; Carta, S.; Kaltenbrunner, A.; and Manca, M. 2018. Behavioral-Data Mining in Information Systems and the Big Data Era. *Inf. Syst. Frontiers* 20(6): 1153–1156. doi:10.1007/s10796-018-9884-1. URL https://doi.org/10.1007/s10796-018-9884-1.

Calabrese, F.; Diao, M.; Di Lorenzo, G.; Ferreira, J.; and Ratti, C. 2013. Understanding individual mobility patterns from urban sensing data: A mobile phone trace example. *Transportation research part C: emerging technologies* 26: 301–313.

Devarasetty, P. C.; Burris, M.; and Douglass Shaw, W. 2012. The value of travel time and reliability-evidence from a stated preference survey and actual usage. *Transportation Research Part A: Policy and Practice* 46(8): 1227–1240. doi:10.1016/j.tra.2012.05.002. URL https://ideas.repec.org/a/eee/transa/v46y2012i8p1227-1240.html.

Esztergár-Kiss, D. 2016. *Optimization of multimodal travel chains*. Ph.D. thesis.

Gabrielli, S.; Forbes, P.; Jylhä, A.; Wells, S.; Sirén, M.; Hemminki, S.; Nurmi, P.; Maimone, R.; Masthoff, J.; and Jacucci, G. 2014. Design challenges in motivating change for sustainable urban mobility. *Computers in Human Behavior* 41: 416–423.

González, M. C.; R., C. A. H.; and Barabási, A. 2008. Understanding individual human mobility patterns. *Nature* 453(7196): 779.

Goulias, K. G. 2018. Travel behavior models. In *Handbook of Behavioral and Cognitive Geography*, chapter 5, 56–73. USA: Edward Elgar Publishing Limited.

Hasan, S.; and Ukkusuri, S. V. 2014. Urban activity pattern classification using topic models from online geo-location data. *Transportation Research Part C: Emerging Technologies* 44: 363 – 381. ISSN 0968-090X.

Jakob, M.; Hrncir, J.; Oliva, L.; Ronzano, F.; Zilecky, P.; and Finnegan, J. 2014. Personalized Fully Multimodal Journey Planner. ECAI'14, 1225–1226. ISBN 978-1-61499-418-3.

Karadimce, A.; Lugano, G.; and Cornet, Y. 2018. Modelling the Quality of User-perceived Travel Experience. In *ACT 2018*, 142–148.

Kováčiková, T.; Lugano, G.; and Pourhashem, G. 2018. From Travel Time and Cost Savings to Value of Mobility. In *Reliability and Statistics in Transportation and Communication*, 35–43. ISBN 978-3-319-74454-4.

Liu, Y.; Pham, T.; Cong, G.; and Yuan, Q. 2017. An Experimental Evaluation of Point-of-interest Recommendation in Location-based Social Networks. *PVLDB* 10(10): 1010–1021.

Lugano, G.; Kurillova, Z.; Hudák, M.; and Pourhashem, G. 2019. Beyond Travel Time Savings: Conceptualizing and Modelling the Individual Value Proposition of Mobility. In *Data Analytics: Paving the Way to Sustainable Urban Mobility*, 714–721. ISBN 978-3-030-02305-8.

Malichová, E.; Pourhashem, G.; Kováiková, T.; and Hudák, M. 2020. Users' Perception of Value of Travel Time and Value of Ridesharing Impacts on Europeans' Ridesharing Participation Intention: A Case Study Based on MoTiV European-Wide Mobility and Behavioral Pattern . *Sustainability* 12(10): 1–19. URL https://ideas.repec.org/a/gam/jsusta/v12y2020i10p4118-d359553.html.

Manca, M.; Boratto, L.; and Carta, S. 2018. Behavioral data mining to produce novel and serendipitous friend recommendations in a social bookmarking system. *Inf. Syst. Frontiers* 20(4): 825–839. doi:10.1007/s10796-015-9600-3. URL https://doi.org/10.1007/s10796-015-9600-3.

Manca, M.; Boratto, L.; Roman, V. M.; i Gallissà, O. M.; and Kaltenbrunner, A. 2017. Using social media to characterize urban mobility patterns: State-of-the-art survey and case-study. *Online Social Networks and Media* 1: 56–69.

Saia, R.; Boratto, L.; Carta, S.; and Fenu, G. 2016. Binary sieves: Toward a semantic approach to user segmentation for behavioral targeting. *Future Gener. Comput. Syst.* 64: 186–197. doi:10.1016/j.future.2016.04.006. URL https://doi.org/10.1016/j.future.2016.04.006.

Schrammel, J.; Prost, S.; Mattheiss, E. E.; Bothos, E.; and Tscheligi, M. 2015. Using Individual and Collaborative Challenges in Behavior Change Support Systems: Findings from a Two-Month Field Trial of a Trip Planner Application. volume 9072 of *Lecture Notes in Computer Science*, 160–171. Springer.

Sierpiński, G.; and Staniek, M. 2017. Heuristic approach in a multimodal travel planner to support local authorities in urban traffic management. *Transportation Research Procedia* 27: 640 – 647. ISSN 2352-1465.

Sourlas, D.; and Nathanail, E. 2019. A Thorough Review and Analysis of Journey Planners. In *Data Analytics: Paving the Way to Sustainable Urban Mobility*, 690–697. ISBN 978-3-030-02305-8.

Steadman, R. G. 1984. A universal scale of apparent temperature. *Journal of Climate and Applied Meteorology* 23(12): 1674–1687.

Vargas, A.; Weffers, H.; and da Rocha, H. V. 2011. Analyzing user interaction logs to evaluate the usability of Web applications. In *2011 3rd Symposium on Web Society*, 61–67. ISSN 2158-6993.

Zarwi, F. E.; Vij, A.; and Walker, J. 2017. Modeling and Forecasting the Evolution of Preferences over Time: A Hidden Markov Model of Travel Behavior. *arXiv preprint arXiv:1707.09133* .

Zografos, K. G.; Androutsopoulos, K. N.; and Apospori, E. 2012. User Acceptance and Willingness to Pay for the Use of Multimodal Trip Planning Systems. *Procedia - Social and Behavioral Sciences* 48: 2405 – 2414. ISSN 1877-0428.