An online monitoring tool for local infrastructure decision makers using smartphone-generated bicycle data

W. Beng, M. Jean, A. Chasse

Department of Control, Signal and System
IFP Énergies nouvelles
1-4 Avenue du Bois Préau, 92852 Rueil-Malmaison
e-mail: {william.beng, maxime.jean, alexandre.chasse}@ifpen.fr

ABSTRACT

Navigation applications for bicycle have been widely adopted by commuters and sportsmen and women alike. This generates a new wealth of data that can potentially provide new insights about bicycle usage. The aim of our work is to process these data and create an interactive tool that will help decision makers spot risky or unpractical roads. Because automatic detection is less accurate and less expressive, it is to be stressed that the main goal is to provide a monitoring tool that detects problematic portions of road so that they can be studied more thoroughly by infrastructure specialists. Steps of the workflow include GPS track pre-processing and mapmatching to the infrastructure, the derivation of flow indicators, waiting times, delays, dangerous braking and road roughness. Results include the extraction of braking and delays from 1 year worth of Geovelo data and the comparison with survey data from French Cyclists association Fédération des usagers de la bicyclette. In this survey participants were asked to indicate on a map danger zones or portions of roads where significant delays were encountered. Traffic volumes were extracted and compared to fixed automatic counting stations. An indicator for road roughness was extracted and mapped as well. This resulted in the creation of a web-based application that has been made available for two groups of local infrastructure decision makers to test and benchmark.

Keywords: GPS, smartphone, navigation, traffic monitoring, congestion, safety, road roughness, data visualization, infrastructure planning.

1 INTRODUCTION

Nowadays, a good proportion of cyclists are equipped with smartphones and use bicycle navigation systems that have the ability to record a complete cycling route. From these sensor rich data, inconveniences felt during trips that are captured in conventional surveys could potentially be extracted by the derivation of indicators in a cheaper and faster manner. They could offer a wider coverage in terms of number of surveyed people and kilometres travelled, and could help reveal discomfort usually not expressed in surveys. The goal of this paper is to present the workflow implemented in order to process raw GPS traces generated from a navigation system into an online monitoring tool usable by decision-making staff. Extensive literature review of big data, crowdsourced data and passive data methods is performed in Lee et al. (2019). Growing interest is shown by public initiatives, like the Civitas initiative co-financed by the European Union. Advances on such mobility data analysis are also made by the private sector, exemplified by the works of Strava, Uber Movement or Bike Citizens.

The platform described in this paper is very much part of the works of this research community. While focusing on related orientations, the platform aims at collecting quality data on major French metropolitan areas, in order to feed dedicated algorithms, from simple aggregations to advanced traffic flow models.

2 METHOD

Bike trips are generated by the Geovelo application. Each of these trips contains time-series of GPS locations, GPS speeds and accelerometer measurements which can be joined with cyclists information. Steps of the workflow include GPS track pre-processing, cleaning and mapmatching to the infrastructure, the derivation of indicators such as traffic intensity, speed, waiting times, delays, dangerous braking and road roughness.

2.1 Traces processing

Basic filtering was done directly in PostGIS and consists in filtering out duplicates and out of date-range traces as well as low quality signals. Non-bike trips are also detected using geometric properties of the GPS traces and filtered out. The next step consists in mapmatching the GPS trace to the infrastructure using the HERE® API. Although this API provides a detailed description of the mapmatched GPS traces, it had to be completed for the needs of this study by a computed curvilinear abscissa giving the total travelled length as a function of time. This has been implemented by cautious interpolation procedures. The HERE® matching algorithm shows its limitations in the trip for sharp turns, areas of particular interest for safety considerations. The simple projection operation performed by HERE® generally violates an expected deceleration of the cyclists before each corner. Observing that the time integral of the GPS speed (Doppler measurement) and the previous curvilinear abscissa previously presented are two estimations of the same time series, a data fusion optimization offers a third estimation taking full advantage of the strength of these two different sources of information. A constrained minimization of an appropriate quadratic cost function is a handy formulation of this problem and was performed using CVXPY from Diamond and Boyd (2016). It imposes the monotony of the estimated curvilinear abscissa, while ensuring a good compatibility regarding the two input time series. Furthermore, a robust estimation of the cyclist speed is given directly by the filtering of the GPS speed measurements. The estimated speed is smooth enough to be differentiated by simple differences, offering a valuation of the cyclist acceleration.

2.2 Indicator computation

Indicators computed as aggregations of individual contributions produce crowd sourced insights of the influence of the infrastructure on the cyclists behaviours ready for interpretation. Traffic intensity is assessed as the number of trips passing by each routelink. Collected GPS speeds can also provide some precious information by using aggregates like average speeds or histogram of speeds on each routelink, or by analysing time-series inside individual trips. For example, average speed and average moving speed per routelink are computed as the mean value of the estimated speed over the entire routelink length and averaged over the entire community. Analysing user speed along the whole trajectory can bring insights about waiting times and dangerous braking. Waiting times are derived by detecting time ranges where speed falls below a certain threshold. Acceleration profile is derived from speed profile and the minimum acceleration between the last peak in the speed profile and the stopping phase can be considered as the braking intensity. The location of the braking event is then retrieved and the result can finally be displayed on a map as point clouds or as a heatmap (see Figure 3). A road roughness indicator has been computed on each routelink. It relies on an isotropic indicator of vibrations computed over the smartphone accelerometer signals and on a community-scale normalization procedure ensuring the automated reconciliation of the data sent by each member, see Jean et al. (2019).

2.3 Data processing pipeline

As the number of tasks comprising the workflow grew, manual execution became increasingly error prone and time consuming. To alleviate these difficulties Apache Airflow was used as a data pipeline management tool. It offers keys automation advantages such as task scheduling, logging, retry on failure, database connection management, and overall a better visualization of the whole pipeline. For faster response times during serving, data is split in multiple tables each corresponding to a major metropolitan area and further divided into monthly partitions. See Figure 1 for the number of trips for each month of collected data.

2.4 Data visualization using Kepler.gl

The next step consists in visualizing the computed indicators. An interactive interface was built using a modified version of the open-source tool Kepler.gl, a powerful tool capable of handling large-scale datasets. Using dedicated controls, the end user can choose an arbitrary date range as well as the days of week and time of day of interest. Custom data layer can also be added.

3 RESULTS

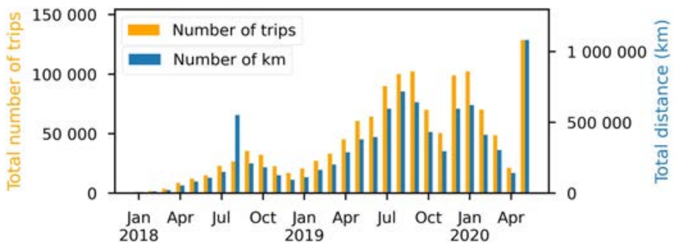

Figure 1. General statistics about the Geovelo database from January 2019 to May 2020.

3.1 The Geovelo dataset

General statistics about the database are shown in Figure 1 (left). A growth trend due to popularity growth of the navigation app can be seen from 2018 to 2019. Seasonality is also captured. Cycling trends due to important events such as the 2019-20 French pension reform strike from December 2019 to January 2020 as well as the lockdown due to COVID-19 during March and April 2020 and the subsequent easing are well captured in the dataset.

Figure 2. Traffic intensity and mean speed in Paris, France during May 2020 (speed from low in red to high in green, traffic intensity as a function of thickness).

4 RESULTS AND DISCUSSION

4.1 Comparison of intense braking event detection with manual survey data

Figure 3. Automatic discomfort detection using smartphone data (heat map in blue shows concentration of braking events weighted by the absolute value of deceleration during braking phase) and manual signalling from FUB (red dots) in downtown Lyon, France.

The comparison of Figure 3 shows that the proposed method is able to confirm major problematic locations for cyclists. This shows how crowd-sourced sensor data can help add weight to survey data. In the same spirit, mapping waiting times after braking can reveal congested sections of the network.

4.2 Traffic volumes comparison with fixed automatic counters

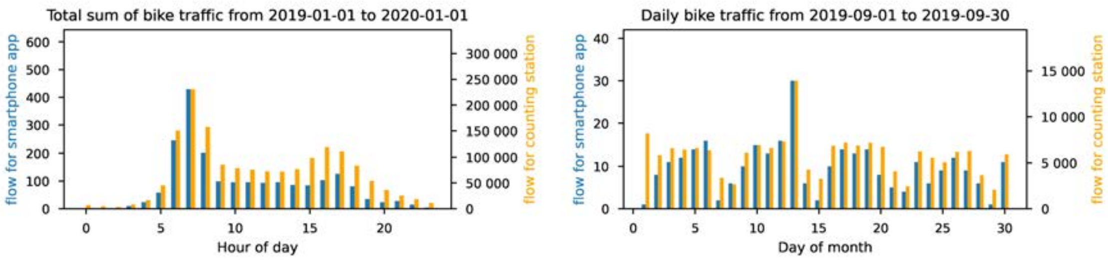

Figure 4. Total hourly sum of bike traffic during year 2019 (left) and daily traffic during September 2019 (right) for the Geovelo dataset (blue) and the Paris dataset (orange) at one fixed counting station

The city of Paris has made publicly available on its OpenData platform hourly bike count from more than 80 automatic counter locations. Comparisons are shown in Figure 4. Further comparison with other counting stations showed good correlation but is out of scope for this paper. A future publication will address the issue of fusing these two sources of information.

4.3 Representativeness of data

As seen in Figure 3, the flow generated by cyclists using the app is only a tiny fraction of the total flow measured by fixed counting stations. The difference in ratio for weekends and week days also suggests that the population of app users does not represent the entire cycling population. In fact, commuters might be overrepresented in the Geovelo community at the expense of sportsmen and women, and other specific cyclist categories. Unfortunately, no registration is required to use the app and even for the registered users, no demographic data is collected. Thus results should be interpreted with care. Data fusion, currently under progress, with external sources of information like fixed counting sensors, will help in assessing the representation bias to build adjustment methods.

Nonetheless, indicators representing physical quantities such as road roughness computed using vibrations, waiting times or braking intensity computed using GPS speeds, could require less post-processing and could be more readily used.

4.4 Mapmatching

Map-matching the GPS-signals to the infrastructure can be risky and difficult, especially if separate cycling infrastructures are present alongside a roadway. Due to GPS uncertainty, it is impossible in most cases to know if the cyclist used the roadway or the cycle path/cycle lane/cycle track that was available. Per routelink, only the global impact of the infrastructure of all the lanes on the behaviours can be observed, with no means to break it down between each one of them. However, the efficacy of particular infrastructure can be assessed over routelinks seen as indivisible entities.

4.5 Further work

A model, assuming cyclists select their route by optimizing a given cost function, will be tuned over the metropolitan area to reflect observed behaviours. At a routelink scale, systematic discrepancies between predicted and actual cyclist paths are expected to indicate actively avoided roads and crossings.

As it is in very early stage of development, this tool is for now aimed mainly towards maintenance management. Infrastructure planning would require more powerful simulation tools such as MATSim.

5 CONCLUSION

Leveraging crowd sourced data collection capabilities offered by smartphones, an online monitoring tool has been created. It was made available to two districts bordering the west of Paris (Suresnes and Puteaux) who showed strong interest in the tool and used it to monitor bike usage on temporary bicycle tracks during and after the COVID-2019 lockdown. The detection of intense braking events as well as road roughness, speed distributions and traffic flow will be cross-referenced with other survey data so that decision makers can decide on where to best send their agents. As the database grows in time, the area covered by the observations will expand and the precision of computed estimators is expected to improve, progressing towards the initial goal of the platform. Other features such as origin-destination flow visualization will soon be included. Future development will tackle data fusion from various sources especially fixed counting stations.

6 REFERENCES

Civitas, cleaner and better transport for cities, 2020 https://civitas.eu/

S. Diamond and S. Boyd, “CVXPY: A Python-embedded modeling language for convex optimization,” Journal of Machine Learning Research, vol. 17, no. 83, pp. 1–5, 2016.

Fédération des usagers de la bicyclette (FUB) (2019). Opendata Parlons-Vélo

M. Jean, A. Chasse and W. Beng, "Road Roughness Crowd-Sensing with Smartphone Apps," 2019 IEEE Intelligent Transportation Systems Conference (ITSC), Auckland, New Zealand, 2019, pp. 1079-1084

K. Lee and I.N. Sener, “Emerging data for pedestrian and bicycle monitoring: sources and applications”, Trans. Res. Interdiscip. Perspectives. (2020)

Open Data de la Ville de Paris (2019). Comptage vélo - Données compteurs