URBANITE Mobility Data Analysis Tools

Ignacio (Iñaki) Olabarrieta†
TECNALIA, Basque Research and Technology Alliance (BRTA),
P. Tecnológico Bizkaia, Ed. 700, 48160 Derio, Spain
ignacio.olabarrieta@tecnalia.com

Ibai Laña
TECNALIA, Basque Research and Technology Alliance (BRTA),
P. Tecnológico Bizkaia, Ed. 700, 48160 Derio, Spain
ibai.lana@tecnalia.com

Urrotz Larrañaga
Bilboko Udala ITS Engineer
Ernesto Erkoreka Plaza, 12, 48007 Bilbo, Bizkaia
ularranaga@bilbao.eus

Sergio Campos
TECNALIA, Basque Research and Technology Alliance (BRTA),
P. Tecnológico Bizkaia, Ed. 700, 48160 Derio, Spain
sergio.campos@tecnalia.com

Raquel Gil
Bilboko Udala Mobility Management Deputy Director,
Ernesto Erkoreka Plaza, 12, 48007 Bilbo, Bizkaia
raquel.gil@bilbao.eus

Shabnam Farahmand
Forum Virium Helsinki, Unioninkatu 24, 00130 Helsinki, Finland
shabnam.farahmand@forumvirium.fi

ABSTRACT

The decision-making process in the policy making should rely on data driven evidence, in most of the cases the raw data needs to be processed to transform it into actionable information. For this purpose, several tools have been developed within the URBANITE project to transform urban mobility data into usable information. Specifically: (1) traffic prediction models based on historical data, (2) Origin-Destination (OD) matrix estimation models and (3) a methodology to analyse the locations visited in several trajectories.

KEYWORDS

Traffic prediction, Origin-Destination Matrix Computation, Data Analysis, Artificial Intelligence.

1 INTRODUCTION

URBANITE project goal is to provide tools for the decision-making in the urban transformation field using disruptive technologies and a participatory approach. These tools should aid the process of taking decisions guiding it on data evidence. The main features of the URBANITE architecture include:

- **Modularity**, i.e., each component provides specific functionalities and exposes clear interfaces,
- **Adaptability** to heterogeneous city and region contexts and ICT maturity levels, from complete implementation to complementary add-on components.
- **Interoperability**, i.e., vertical, and horizontal interoperability among modules and with existing systems.

And using the European standards as much as possible.

The main elements that URBANITE offers are the following:

† Corresponding Author

Information Society 2021: 24th international multiconference, *4–8 October 2021, Ljubljana, Slovenia*

Social Policy Lab – an environment to promote digital co-creation with methodologies and methods to support the communication among public servants, private companies, and citizens. The aim of the Social Policy Lab is to develop joint ideas and to produce co-creation guidance for policies. **Data Management Platform** – to provide automatic support to the whole data processing chain and its life cycle, starting with the collection process all the way up to the use of the data. **Decision-Making Support System** – powerful tools which combine multiple data sources with advanced algorithms, a simulation engine, a recommendation, and visualisation system. These tools include predefined analysis pipelines to be used by non-technical users, intuitive and understandable visualisations, and setups to perform simulations of new mobility policies and situations that allow their evaluation. URBANITE is implemented in four different use cases: Amsterdam, Bilbao, Helsinki, and Messina.

The analysis tools that are presented in this communication belong to the Decision-Making Support System. More concretely they belong to the set of algorithms designed to obtain information from the historical data stored in the URBANITE Data Management Platform. The results obtained from these algorithms can be used to understand better what is the state of the mobility at a given time, or, alternatively, they can be used as input for simulations of new policies.

Among all the tools within the Decision-Making Support System three components are explained in this communication, namely: traffic prediction, OD matrix estimation, and trajectory location analysis. These components are discussed in the following sections, sections 2-4. This communication ends with some concluding remarks in section 5.

2 TRAFFIC PREDICTION

Road traffic forecasting has been a topic of study since the sixties [1] when time series analysis methods where mainly used [2][3][4]. In the last two decades, heuristic machine learning methods [5] started being used allowing to find more complex relations within the traffic data. Nowadays traffic prediction component has become one key tool for any ITS system. The component developed within the URBANITE project can forecast what is the traffic flow that a sensor within the city would measure for a given set of features.

The set of features at the time of this communication include the day of the week and the time of the day but other ones are in the process of being incorporated. Some of these features include if the day of the forecast is a bank or school holiday, weather features (precipitation and temperature), the arrival of ferries to Helsinki port or sport events (soccer games) in Bilbao (using the method developed in [6]). Note that the approach within URBANITE is not to consider previous measurements as features since the data is not available in real time. Therefore, this approach can be considered as long-term prediction because the predicting horizon is only limited by the accessibility of external features (i.e., access to a weather prediction for example).

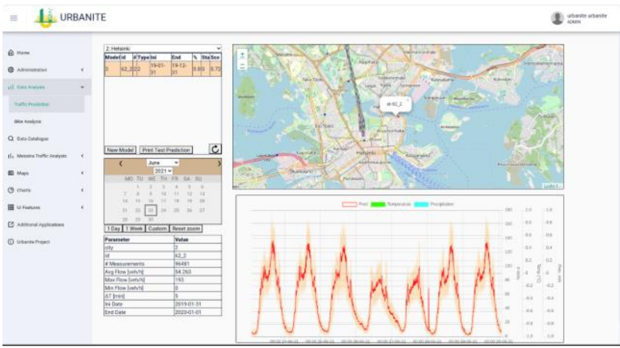

Figure 1: Integrated tool to perform traffic prediction showing the Helsinki use case.

The web portal to the integrated tool can be seen in Figure 1, the tool allows to train a new model, to perform a prediction and to visualize the results.

The process of performing the training implies that the user needs to choose the following:

- The regression model type, two options are available: random forest [7] and distribution inference [8] (only for features with discrete values).
- The number of features to consider: 1. considers only the day of the week, 2. also considers the time and so on.
- The time resolution, typically either 5 or 15 minutes. This is the aggregation period on which the individual counts of vehicles moving over the sensor are combined to produce a time series.
- The traffic sensor, this is chosen by selecting the available sensors within a map.
- The period of the training data, the period can be chosen from the available data within Data Management Platform, being able to change this period allows for instance to avoid choosing the anomalous period due to the restrictions due to COVID-19. In addition, a percentage of the training data can be reserved to test the goodness of the model, this percentage can also be specified.

Once a model is trained this can be used to perform a prediction, there are different ways to perform this, one way is to use the URBANITE web visualization tool to choose a given date and perform the prediction for either the following 24 hours or the following 7 days. Alternatively, specific set of features can be feed to the model using the REST Web service in JSON format to obtain a result at a given instant of time.

An example of the result can be seen in Figure 2 where in addition to the prediction (red line) the confidence interval is shown (orange band). The details of how to compute the confidence interval are explained in [9]. In the Figure the result of the prediction for a week is shown, where the peaks for the different days are clearly visible, including the difference in the pattern due to the weekend (fourth and fifth peaks in the series).

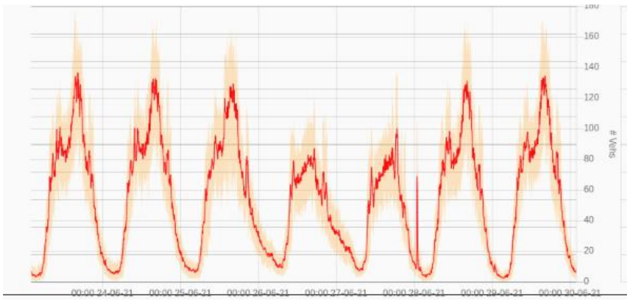

Figure 2: Detailed of the visualization for the result of the traffic flow prediction for 7 days including the confidence interval.

3 OD MATRIX ESTIMATION

The OD Matrix estimation works in a similar way than the prediction module. In this case we use data from bike rental city service, specifically we consider the origins and the destinations of each one of the rentals. These are both temporally and spatially aggregated by providing the time resolution (the same way as for the traffic prediction) and by providing a set of geographic areas where to aggregate the origins and the ends of each rental. These areas can be specified either via a GEOJSON or by specifying a set of points, the URBANITE web can be used to obtain the Voronoi areas [10] associated with those points and use those to perform the spatial aggregation.

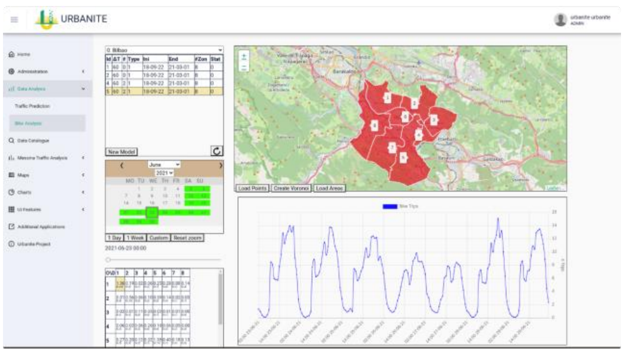

Figure 3: Integrated tool to perform OD matrix estimation for the Bilbao use case.

Training a model to perform OD matrix estimation implies choosing a regression model type, the number of features (in this case 0 can be chosen, which implies the use of only spatial information), the time resolution, and the period of training data.

The result of the estimation, at a specific instant of time, consists of a square matrix of size NxN where N is the number of different areas considered for the spatial aggregation. The web tool within URBANITE allows to compute and visualize these estimations for all the instants within a period (typically a day or a full week). In Figure 3 a detail of the web tool is shown where it can be seen the result at a given instant in the form of a matrix (lower left hand side) and the time evolution of one of the matrix components for a whole week (lower right hand side).

It is worth mentioning that this process to estimate OD matrixes, by means of the use of regression algorithms, have the capability to generalize the values measured obtaining results even in regions of the feature space where no values have been obtained yet.

4 TRAJECTORY LOCATION ANALYSIS

Finally, the last component that we explain in this communication consists in a tool able to analyze not only the origin and the destination of trajectories but also what happens in between. More specifically, and to fix ideas, we can think this tool's goal to be obtaining the points more popular to visit in a trajectory. The processing consists in two different phases: the cleaning phase, and the aggregation phase.

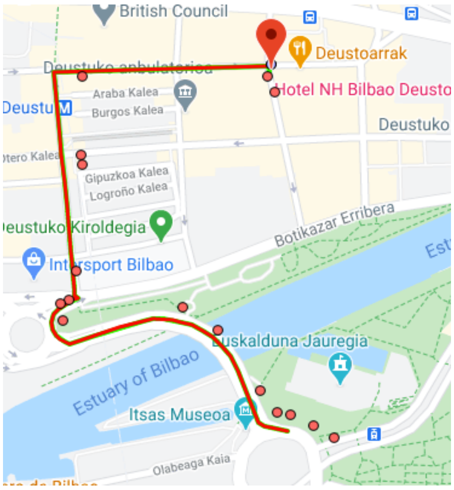

Figure 4: Result of the cleaning phase for a set of GPS points obtained from a single bike city rental in Bilbao.

The cleaning phase is a crucial phase when processing GPS data obtained from affordable, not very accurate sensors or in areas with tall buildings (urban environment) where the multi-path of the satellite signal can increase the noise of the measurements. The purpose of this phase is to align the obtained measurements with the navigational road network, i.e., the possible allowed positions for the vehicles. In URBANITE, Hidden Markov models [11] are used in this phase. Moreover, this process provides an additional result, which consists in the most likely points between measurements.

The second phase, the aggregation phase, compares the points obtained in the cleaning phase for all the trajectories. Probably the simplest of these aggregations is to compute the number of times a location is visited independently of the trajectory it belongs. The result of this process applied to the trajectories of the bike city service in Bilbao is shown in Figure 5.

Other types of aggregations can also be performed, as for example most likely points to be visited depending of the day of the week and the time of the day, the most popular chain of consecutive points visited, the longest route accomplished, etc.…

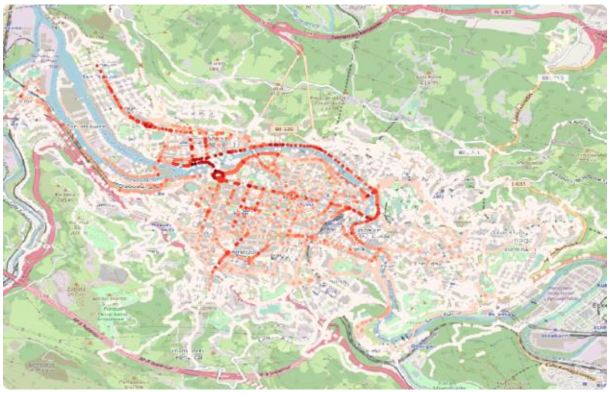

Figure 5: The most popular points corresponding to trajectory locations are labelled with darker color.

5 CONCLUSIONS

In this paper we have introduced three components developed within the URBANITE project to convert data into information. The first component is designed to obtain a prediction of the typical traffic flow at a particular sensor location given a set of features, the second component aims to produce OD matrixes from bicycle data and finally the last component consists in a methodology to analyse trajectory locations. These results have been achieved during funding project from the European Union's Horizon 2020 research and innovation programme under grant agreement #870338.

REFERENCES

[1] I. Laña, J. Del Ser, M. Velez, and E. I. Vlahogianni, (2018) *Road traffic fore-casting: Recent advances and new challenges*, IEEE Intelligent Transportation Systems Magazine, vol. 10, no. 2, pp. 93–109.

[2] M. S. Ahmed and A. R. Cook, (1979) *Analysis of freeway traffic time-series data by using Box-Jenkins techniques*, Transportation Research Record, no. 722, pp. 1-9.

[3] M. Levin and Y.-D. Tsao. (1980) *On forecasting freeway occupancies and volumes (abridgment)*, Transportation Research Record, no. 773.

[4] C. Moorthy and B. Ratcliffe, (1988) *Short term traffic forecasting using timeseries methods*, Transportation planning and technology, vol. 12, no.1, pp. 45–56.

[5] E. I. Vlahogianni, M. G. Karlaftis, and J. C. Golias (2007), *Spatio-temporal short-term urban traffic volume forecasting using genetically optimized modular networks*, Computer-Aided Civil and Infrastructure Engineering, vol. 22, no. 5, pp. 317–325.

[6] I. Olabarrieta, I. Laña. (2020). *Effect of Soccer Games on Traffic, Study Case: Madrid*. 1-5. 10.1109/ITSC45102.2020.9294749.

[7] L. Breiman (2001). *Random Forests*. Machine Learning. 45 (1): 5–32. doi:10.1023/A:1010933404324.

[8] G. Upton, I. Cook, (2008) *Oxford Dictionary of Statistics*, OUP. ISBN 978-0-19-954145-4.

[9] I. Laña, I. Olabarrieta, J. Del Ser. (2021) Output Actionability of Traffic Forecasting Models: Measuring and Understanding Uncertainty of Forecasts to provide Confidence Levels (in preparation).

[10] G. Voronoi, (1908a). *Nouvelles applications des paramètres continus à la théorie des formes quadratiques*. Premier mémoire. Sur quelques propriétés des formes quadratiques positives parfaites" (PDF). Journal

für die Reine und Angewandte Mathematik. 1908 (133): 97–178. doi:10.1515/crll.1908.133.97. S2CID 116775758.

[11] P. Newson, J. Krumm. Nov. (2009*). Hidden Markov Map Matching Through Noise and Sparseness*. 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (ACM SIGSPATIAL GIS 2009), November 4-6, Seattle, WA