Vehicle mobility data analysis and Individual Mobility Networks for crash prediction

Mirco Nanni
ISTI-CNR
Pisa, Italy
mirco.nanni@isti.cnr.it

Leonardo Longhi
Sistematica S.p.A.
Rome, Italy
leonardo.longhi@grupposistematica.it

***Abstract* — This paper describes an ongoing work on the analysis of mobility data of individual users aimed to predict the user's risk of experiencing a crash in the near future. The suggested approach looks at descriptive features of the user's driving behavior, both in terms of driving style and general mobility demand. The latter point is addressed through a Individual Mobility Network model. Preliminary promising results and next challenges are finally discussed.**

Keywords—Mobility, Crash risk, Driving behavior.

I. Introduction

Driving safety is one major issue in modern urban setting, due to the presence of distressing traffic situations and frequent accident events involving vehicles. While some factors have been clearly identified as potential causes of accidents (e.g. long drives, lack of proper visibility, bad weather conditions), identifying risky situations and predicting crashes still remains an open area for research.

In this work we address the problem from an individual mobility data perspective, analyzing the mobility of users based on GPS traces of private vehicles, and trying to identify features of the mobility and driving behavior of the user that are correlated with accidents. Differently from most existing approaches, that examine the problem either from a geographical and temporal perspective (i.e. identifying the areas and the hours of the day where the risk of accidents is higher) or from a real-time one (i.e. identify the instantaneous conditions that are likely to lead to accidents), our proposal consists in analyzing the mobility of users on the medium and long term, extracting overall behaviors and regularities that can be linked to higher rates of accidents.

In the following we introduce a basic tool for summarizing the mobility of an individual, namely Individual Mobility Networks, and then discuss how the user's driving can be characterized. Afterward, we point out existing issues in applying predictive models in different geographical areas. Finally, we show some preliminary results and conclude the paper with final remarks and future works.

II. Individual Mobility Networks

Individual Mobility Networks are a concise graph representations of the mobility history of individuals. From raw GPS traces the trajectories of a single mobility user are reconstructed and processed to infer the relevant locations that the user visited (the nodes of IMNs) and aggregate the trips between two locations (the edges of IMNs). Nodes and edges are enriched with several statistics of the associated trips, such as temporal distributions and distances. Figure 1 shows a pictorial representation of a IMN.

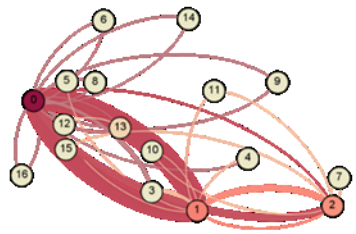

Fig. 1. Sample Individual Mobility Network. Nodes represent locations, edges (direction is clock-wise) represent flows between locations.

Users with different mobility needs and habits show different IMNs, as it can be seen from Figure 2, which depicts several examples extracted from different cities in Tuscany, which reveals a high variability of mobility behaviours both within and across cities. That suggests that characterizing IMNs can provide a distinctive description of users.

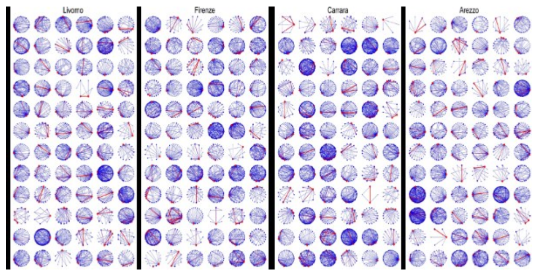

Fig. 2. IMNs from several individuals over four cities in Tuscany. Thicker and red nodes/edges represent frequent ones.

III. Crash Prediction

Crash risk can be defined as the probability of accidents, which are (in statistical terms) rare events. That, together with the lack of a clear set of predictive indicators to adopt, make the risk prediction a difficult task.

The solution proposed and under development in this work follows a standard machine learning approach: first we analyze the trajectories of each user to extract descriptive features, and then a predictive model is trained on a labelled dataset (i.e. a dataset containing users that had accidents and users that did not, with a label distinguishing them). The key point is to define an effective set of features, able to capture those aspects that are correlated with the crash risk.

Here we consider two types of features: those describing basic driving behaviours of the user (speeds, frequency of accelerations, cornerings, etc.), and IMNs ones, describing how driving features are distributed on the IMNs.

A. Movement features

Through the analysis of the travels performed by the user, we extract the following categories of features:

- Travel features: they include statistics about the length and duration of the trips, average speed, also split into periods of the day or of the week.
- Events features: they include the frequency and intensity of driving events, i.e. accelerations and decelerations, also divided in temporal intervals

Where available, additional information about the vehicle (brand or model) is considered.

B. IMNs features

When the single trips of a user are organized in a IMN, we are able to capture all the semantic categories that IMNs allow to infer (frequent routes, central vs. peripheral locations w.r.t. the IMN). A first set of features includes network statistics of the IMN structure, such as number of locations and links, network clustering coefficients, network modularity index, etc. Moreover, all the movement statistics mentioned in the previous section can be transferred to the corresponding portion of the IMN, thus enabling to extract more semantic aggregates, e.g. frequency of accelerations on routinary trips.

Finally, IMNs can be enriched with information about the types of environment the user usually traverses: road categories, weather conditions, expected traffic density during driving time, etc., which can in turn be used for inferring even more specific aggregates.

IV. Geographical Adaptivity of Models

Most analyses and models extracted from mobility data are highly dependent on the characteristics of the territory under study. In particular, it is known that mobility models extracted in one region might not work well in other ones, thus raising an issue of transferring models across different areas. At the same time, the analyses considered in this work requires the availability of massive, labelled mobility data that is difficult to have in all areas of interest, therefore producing a model out of a data-rich area and making it usable in other areas would be of extreme value.

In this direction, the ongoing work of this research tries to address the problem by characterizing different areas based on a wide variety of indicators, with the aim of better assessing the similarity of different geographical areas. The idea is that models are more easily transferrable between similar areas, and it might be possible to devise mechanisms to adapt models across areas with different characteristics. Preliminary experiments on a simpler traffic prediction problem show that this approach, carried out considering the two families of features described above, can help identifying areas across which the model transfer has better chances of working.

V. Experimental setting and preliminary results

The approaches described above have been preliminary tested over a dataset of GPS vehicle traces that include acceleration events and a number of actual crashes. The results commented here focus on users of the area of Rome (Italy), while ongoing experiments are generalizing them also to Tuscany and the area of London (see Figure 3).

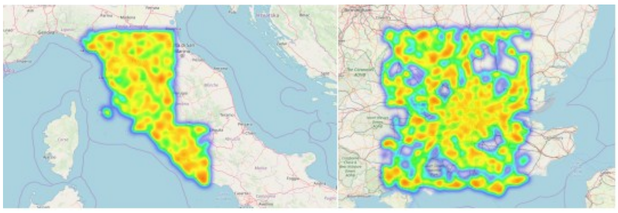

Fig. 3. Area of experimentation. Preliminary results are focused on Rome area (lower part of left figure)

For each user the data over 4 months was used to extract all the movement features mentioned in Section III.A (ongoing work also includes IMNs features). In addition, each user is labeled according to the presence of accidents during the next month. The resulting dataset has been used as input for various traditional machine learning models, including Random Forests, Support Vector Machines and Neural Networks. RFs yielded the best and more stable results, shown in Figure 4. The table also divides performances over different subsets of features (traj = travel only, evnt = events only, evnt = both, all = include also additional information about brand and model).

features	f1-score	precision	recall	test_accuracy	train_accuracy
all	0.659024	0.709972	0.655093	0.712644	0.815271
evnt	0.641626	0.734188	0.663095	0.672414	0.757389
traj	0.636443	0.723219	0.655688	0.669540	0.748768
trev	0.603573	0.651994	0.609672	0.658046	0.795567

Fig. 4. Performances over different subsets of features

The results show that all feature types bring some improvement. Also, we remark that the problem is imbalanced (around 1 crash every 5 users in this sample dataset), therefore a high recall is as valuable as a high accuracy.

VI. Conclusions

This ongoing work tackles a challenging task from an alternative, individual-based angle, aiming to improve predictability of crashes. Yet, predicting risk is only the first step to prevention. The future lines of research of this work will aim to extract prescriptive rules out of predictions, which is a difficult task because best performing predictors are usually black-boxes, inherently non human-readable.

Acknowledgment

This work is funded by the EU H2020 program under G.A. 780754, “Track & Know”.

References

[1] Salvatore Rinzivillo, Lorenzo Gabrielli, Mirco Nanni, Luca Pappalardo, Fosca Giannotti and Dino Pedreschi. The Purpose of Motion: Learning Activities from Individual Mobility Networks. In 2014 Int. Conf. on Data Science and Advanced Analytics (DSAA14), 2014.

[2] Sinno Jialin Pan and Qiang Yang. A Survey on Transfer Learning. IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, 2009.

[3] Riccardo Guidotti, Mirco Nanni, Salvatore Rinzivillo, Dino Pedreschi and Fosca Giannotti. Never Drive Alone: Boosting carpooling with network analysis. In Information Systems journal (IS), Elsevier, 2017

[4] M. E. J. Newman, Modularity and community structure in networks, in PNAS, vol. 103, nº 23, 2006, pp. 8577–8582

[5] Gravity models: A tool for migration analysis. IZA World of Labor 2016: 239