Exploring Real Mobility Data with M-Atlas

R. Trasarti, S. Rinzivillo, F. Pinelli, M. Nanni, A. Monreale,
C. Renso, D. Pedreschi, and F. Giannotti

Pisa KDD Laboratory, ISTI - CNR, Italy

1 Introduction

Research on moving-object data analysis has been recently fostered by the widespread diffusion of new techniques and systems for monitoring, collecting and storing location aware data, generated by a wealth of technological infrastructures, such as GPS positioning and wireless networks. These have made available massive repositories of spatio-temporal data recording human mobile activities, that call for suitable analytical methods, capable of enabling the development of innovative, location-aware applications [3]. The M-Atlas is the evolution of the system presented in [5] allows to handle the whole knowledge discovery process from mobility data. The analysis capabilities of M-Atlas system have been applied onto a massive real life GPS dataset, obtained from 17,000 vehicles with on-board GPS receivers under a specific car insurance contract, tracked during one week of ordinary mobile activity in the urban area of the city of Milan; the dataset contains more than 2 million observations leading to a set of more than 200,000 trajectories (see Fig.1).

2 The M-Atlas System

A system able to master the complexity of the knowledge discovery process over mobility data needs to support at least four functionalities: (i) trajectory data need to be created, stored and queried through spatio temporal primitives; (ii) trajectory models and patterns representing collective behavior have to be extracted using trajectory mining algorithms; (iii) such patterns and models have to be represented and stored in order to be re-used or combined; (iv) new mining algorithms may be added. The M-Atlas system allows the user to combine all these aspects through an innovative Data Mining Query Language (DMQL). This language can be used to express the whole knowledge discovery process as a sequence of queries to be submitted to the system. The GUI interface gives the user the possibility to use pre-defined analysis (i.e. O/D Matrix) or to use the console to write down his/her own DMQL queries.

In the next sections we will give a short example of the capabilities of the system on the real dataset of Milan described above.

3 Understanding Mobility

To grasp a general vision on the dataset we performed a series of statistical analysis on the dataset (the charts representing the results are shown in Figure 2):

J.L. Balcázar et al. (Eds.): ECML PKDD 2010, Part III, LNAI 6323, pp. 624–627, 2010.

The Movements Distribution analysis estimates the active movements in each hour of the week. From the figure it is evident the drop in the movements during the night hours, clearly separating the different days, hence the seven groups represent the seven days from Sunday to Saturday. The shapes of the days are very similar especially during the week-days: each day contains two peeks, one in the morning and one in the late afternoon. This analysis validates the dataset comparing it with a survey obtained from the Mobility agency of Milan in 2005.

The Cumulative Lengths Distribution represents the cumulative number of trajectories having the same length. From this analysis it is clear that in the city there are many short movements and few long movements obtaining a distribution which follows a power law.

The Density of Length over Speed represents in cold colors a low density and the warm colors an high density. In this way we can see the variance of lengths for each speed value, looking at the *existence* of points and the common value looking for the warmer ones. For Example the trajectories with speed around 20 km/h have length from 1 to 190 km with a very high density zone between 1 and 30 km.

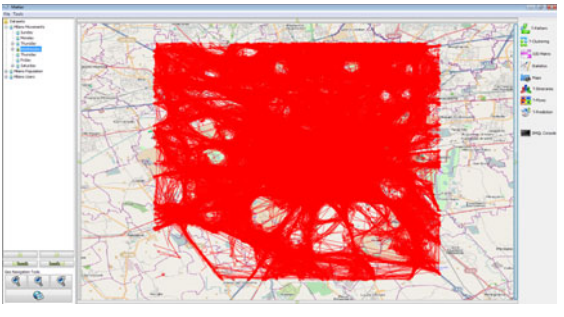

Fig. 1. The Trajectory Dataset of Milan

In the next section we show some analytical tools used in combination with data mining algorithms in order to answer to a specific request of the mobility manager.

4 Discovering Mobility Behaviors

The M-Atlas system integrates a set of analytical and data mining tools such as the construction of Origins-Destinations Matrix, the construction of georeferenced density maps according to different measures, extracting of T-Patterns [2], T-Clustering [6], T-Itineraries [1] and T-Prediction [4]. In this section we present an example of these tools as basic bricks to build a discovery process. The presented analysis answers to a specific requirements of the mobility manager: *How the people leave the city center during the day?*.

Computation of O/D Matrix: The common exits of the city has been discovered building an O/D Matrix between the zones of the city represented in a big scale by a 3×3

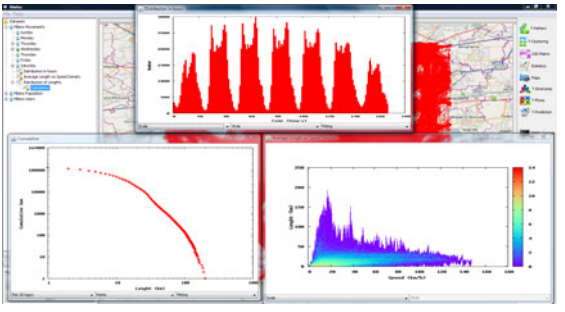

Fig. 2. Statistical Analysis of the dataset: Movements Distribution (top), Cumulative Lengths Distribution (left) and Density of Length over Speed (right)

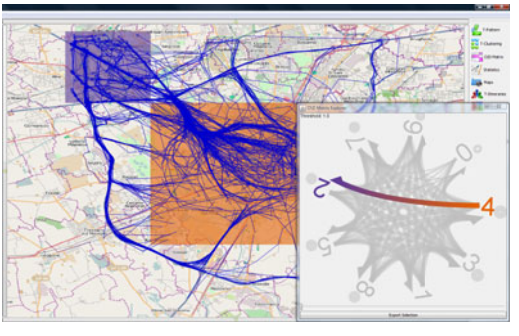

Fig. 3. The O/D Matrix between zones of the city

matrix. Each zone in the border corresponds to an exit and the zone in the middle corresponds to the city center. In Fig.3 in the bottom right corner we can see a conceptual representation of the flows between the zones (in gray) and a particular flow from the zone n.4 (the city center, in orange) to the zone n.2 (the north-west exit, in violet) highlighted as the *most dense flow starting from the city center* (according to the size of the arrow). Hence the set of trajectories which is part of the selected flow can be extracted (see Fig.3) which can used for further analysis.

T-Clustering: Given the trajectories leaving the city center toward the north-west exit, we want to group together similar trajectories in order to discover common behaviors. Therefore a data mining tools called *T-Clustering* is used. M-Atlas provides different methods to describe the similarity between two trajectories, one of them is *route similarity* which perform a temporal alignment and then compares spatially the two trajectories. The result of this computation is shown in Fig.4 where each cluster is represented by a different color. There are three major clusters: (i) the *red* one is the group of people which leaves the very center of the city via the north-west city gate and then goes straight to the north-west exit, (ii) the *purple* one represents the people leaving the center from a peripheral part of the city center and (iii) the *blue* one is the group of people which prefers to avoid the crowd and opts for a

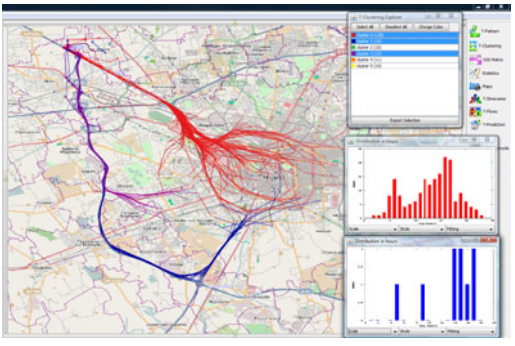

Fig. 4. The result of T-Clustering tools and the distribution of movements during the day for the red and blue clusters

longer way but probably faster. This last hypothesis is supported also by the analysis of the temporal distribution of the two cluster (red and blue) which highlight the simultaneity of the peaks of the movements distribution in the afternoon as shown in Fig.4 (right).

The same analysis can be applied also for the other exit directions, presenting to the mobility manager a complete view of behaviors.

5 Conclusion

The presented process of analysis is just a sketch of the power of the M-Atlas system which allows the analyst to combine tools in order to build his/her own discovery knowledge process in an iterative and interactive way.

Acknowledgments. The authors wish to acknowledge the Mobility Agency of Milano and Octotelematics S.p.A. for providing the anonymized dataset of private cars with on-board GPS receivers.

References

1. Benkert, M., Gudmundsson, J., Hbner, F., Wolle, T.: Reporting flock patterns. Computational Geometry 41, 111–125 (2008)
2. Giannotti, F., Nanni, M., Pinelli, F., Pedreschi, D.: Trajectory pattern mining. In: KDD, pp. 330–339 (2007)
3. Giannotti, F., Pedreschi, D. (eds.): Mobility, Data Mining and Privacy - Geographic Knowledge Discovery. Springer, Heidelberg (2008)
4. Monreale, A., Pinelli, F., Trasarti, R., Giannotti, F.: Wherenext: a location predictor on trajectory pattern mining. In: Proc. ACM KDD, pp. 637–646 (2009)
5. Trasarti, R., Giannotti, F., Nanni, M., Pedreschi, D., Renso, C.: A Query Language for Mobility Data Mining. International Journal of Data Warehousing and Mining (2010) (print on)
6. Andrienko, G., Andrienko, N., Rinzivillo, S., Nanni, M., Pedreschi, D.: A Visual Analytics Toolkit for Cluster-Based Classification of Mobility Data. In: SSTD, pp. 432–435 (2009)