Traffic Jams Detection Using Flock Mining

Rebecca Ong, Fabio Pinelli, Roberto Trasarti, Mirco Nanni,
Chiara Renso, Salvatore Rinzivillo, and Fosca Giannotti

Pisa KDD Laboratory,
ISTI - CNR, Italy

1 Introduction

The widespread use of GPS devices on cars enables the collection of time-dependent positions of vehicles and, hence, of their movements on the road network. It is possible to analyze such huge collection of data to look for critical situation on the traffic flow. The offline analysis of traffic congestions represents a challenging task for urban mobility managers. This kind of analysis can be used by the traffic planner to predict future areas of traffic congestions, or to improve the accessibility to specific attraction points in a city. Many traffic systems adopt *ad-hoc* sensors like cameras, induction loops, magnetic sensors to monitor the status of the traffic flows: these systems are very expensive for installation and maintenance, and they are restricted to the local monitoring of the road arcs where they are installed. On the contrary, the use of GPS data to check the traffic conditions requires low installation costs (a part for the installation on the vehicle) and it enables to virtually monitoring the entire road network.

In this demo we present an innovative tool that exploits the data collected from GPS-enabled cars to detect the occurrences of traffic jams on the road network. The detection of potential traffic jams is based on the discovery of slowly moving *flock patterns*, i.e. a set of objects slowly moving together for a minimum amount of time [2,1,5]. The tool has been integrated in the M-Atlas system [4,7] exploiting the implementation of the T-Flock algorithm provided by the system. To the best of our knowledge this is the first system that uses GPS data, combined with flock mining, to detect traffic congestions. Most of the approaches available in the literature for traffic analysis are based on aggregation of spatial or temporal data focusing on predefined areas [3]. It is important to point out that this tool does not provide real time analysis, but instead it allows the analysis of the historical data. We will sketch here a case study on a dataset of around 40,000 GPS-tracked cars in the surrounding of Pisa in Italy. The demo will highlight the tight integration of the spatio-temporal and data mining tools of the M-Atlas system and the graphical user interface that assists the DM analyst in driving his/her analysis.

2 Problem Definition

We define a *traffic jam* as a group of vehicles moving close and slowly for a minimum period. Following the definition given in [6], a *T-Flock* is formally described as:

Definition 1 (T-Flock). *Given a set of n trajectories where each trajectory consists of τ line segments a flock in a time interval $[t_i, t_j]$, where $j - i + 1 \geqslant k$ consists of at least m entities such that for every discrete time step t_l, $i \leqslant l \leqslant j$, there is a disk of radius r that contains all the m entities.*

According to Definition 1, a T-Flock represents a generic movement of different entities moving together. For the objectives of our analysis, we enrich the T-Flock definition with a speed threshold, ensuring that the discovered patterns do not move more quickly w.r.t. the given threshold.

Definition 2 (T-Flock Jam). *Given a speed threshold s, a T-Flock T is a T-Flock Jam if its speed is lower than s.*

We propose a new method for traffic jam detection based on the combination of the *(1)* data mining query language and T-Flock algorithm implementation provided by M-Atlas, and *(2)* a graphical user interface to refine and interpret the results. Initially, the M-Atlas system is used to extract the T-Flock patterns from a dataset. The extracted patterns are rendered on the screen by means of visual metaphors to highlight relevant properties of the flocks, like support, velocity, duration. The visual interface allows also the definition and the combination of constraints on the patterns, in order to select the subset of T-Flocks patterns that satisfy the T-Flock Jam definition. The integration of the methodology within the M-Atlas system allows the transparent integration of the analytical process with the other tools provided by the system.

3 The Case Study

We briefly sketch here an analytical case study performed on a set of GPS-tracked vehicles in the Pisa surroundings.

The Dataset. The Pisa dataset contains a set of timestamped points of the form (*id*, *lat*, *long*, *t*) from around 40,000 cars, which represents 2% of registered cars in the coastal areas of Tuscany. These points were tracked using GPS receivers with a sampling rate of $\approx 30s$ and a positioning system error of 10-20m in normal conditions over a period 5 weeks. As deeply discussed in [7], this sample of data indicates strong evidence to the validity and coherence of GPS data and its representativeness power. We have concentrated on the points observed during a one week period from June 14, 2010 to June 20, 2010, consiting of $\approx 4,000$ cars per day. To avoid the extraction of stopping T-Flocks (e.g. groups of cars parked in the same area), the traces of the vehicles have been preprocessed by cutting each trajectory in smaller trips: a stop longer than two hours determines the end of the current trip and the start of a new one.

Flock Discovery. According with the T-Flock definition, the traces of the vehicles have been divided into line segments with a time interval of $30s$. The T-Flock patterns are extracted considering a disk radius of $200m$, at least 6 *line segments* –i.e. vehicles moving together for at least 180 seconds– and a minimum flock support of at least 3 vehicles. The discovered flocks on the June 14 data are shown in Figure 1.

Traffic jam detection. To distinguish the traffic jams from all the flocking vehicles, the analyst can apply a set of constraints based on the T-Flock attributes: *duration*, *length*,
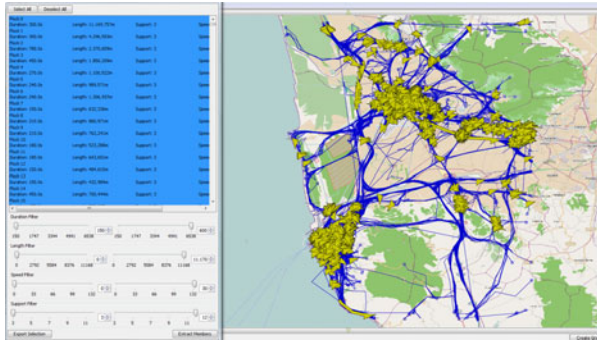
Fig. 1. The complete set of flocks extracted in the Pisa and surrounding area

speed, and *support*. In particular, the flock *speed* constraint can be used to select all the patterns within the given threshold. Figure 2 shows the whole set of flocks discovered, and how it is possible, using the interactive constrains, filter the flocks taking only the slow ones.

Dynamic Speed Constraint. A different approach to detect traffic jams from the set of discovered flocks consists of computing the ration of the speed of a flock with the average speed of all the cars passing through the same area. Actual traffic jams can be identified through their low speed ratio, which means that the involved cars have a considerably lower speed compared to the average speed of passing cars in the considered area at a specific period of time. Figure 3 shows how previously detected traffic jams in Pisa urban area can further be classified with respect to the dynamic speed constraint. It is important to notice that this new constraint is useful to detect events during the day which do not follow the typical mobility behavior in that area. The system is also open to consider different ways to compute the typical speed, which can be for example the using of a map matching algorithm between the flocks and the road network.
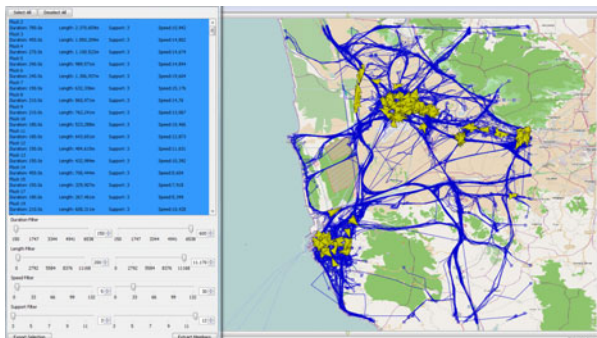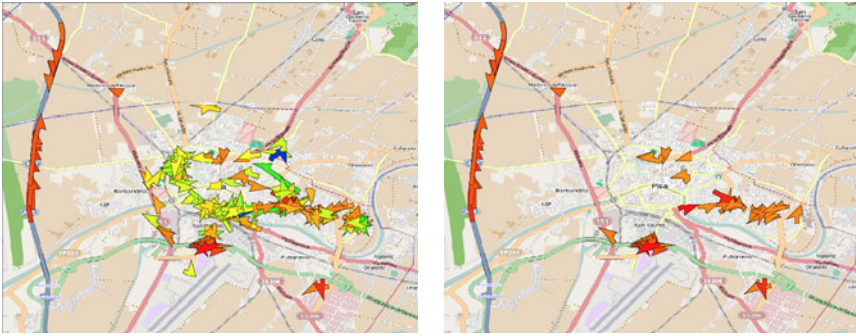

Fig. 2. Detecting traffic jams candidates using the static speed constraint: greater then 5km/h and less than 30 km/h

Fig. 3. Applying the dynamic speed constraint on selected candidates highlights how this constraint distinguishes between real traffic jams (red/yellow gradient) and usual congested areas (blue/green gradient). (Left) All the selected flocks in Pisa urban area. without applying any threshold. (Right) Applying a threshold corresponding to a decrement of 60% of average speed.

4   Additional Analyses

During the demo, the case study presented in the previous section will be extended with further analyses: (i) we will show how supplementary information (e.g. points of interest, city gates, meteorological conditions, etc.) can be exploited to add semantics to the models, and (ii) how the traffic jams can be temporally characterized.

References

1. Gudmundsson, J., van Kreveld, M.J.: Computing longest duration flocks in trajectory data. In: GIS (2006)
2. Laube, P., Imfeld, S., Weibel, R.: Discovering relative motion patterns in groups of moving point objects. International Journal of Geographical Information Science 19(6), 639–668 (2005)
3. Gennadu, A., Natalia, A.: A general framework for using aggregation in visual exploration of movement data. The Cartographic Journal 47(1), 22–40 (2010)
4. Trasarti, R., Rinzivillo, S., Pinelli, F., Nanni, M., Monreale, A., Renso, C., Pedreschi, D., Giannotti, F.: Exploring Real Mobility Data with M-Atlas. ECML/PKDD (3), 624–627 (2010)
5. Wachowicz, M., Ong, R., Renso, C., Nanni, M.: Discovering Moving Flock Patterns among Pedestrians through Spatio-Temporal Coherence. International Journal of Geographical Information Science (in press)
6. Benkert, M., Gudmundsson, J., Hübner, F., Wolle, T.: Reporting Flock Patterns. In: Azar, Y., Erlebach, T. (eds.) ESA 2006. LNCS, vol. 4168, pp. 660–671. Springer, Heidelberg (2006)
7. Trasarti, R., Giannotti, F., Nanni, M., Pedreschi, D., Renso, C.: A query language for mobility data mining. International Journal of Data Warehousing and Mining (IJDWM) 7(1), 24–45 (2011)