Building Visual Summaries of Clusters of Trajectories

Gennady Andrienko, Natalia Andrienko

Fraunhofer Institute IAIS (Intelligent Analysis and Information Systems), Sankt Augustin, Germany

ABSTRACT

A big problem in analysis of movement data (trajectories of moving agents) is how to visualize clusters of trajectories so as to have an overview of all clusters and to be able to compare different clusters. Clusters of trajectories are not disjoint in space but intersect and overlap. Hence, they need to be shown separately in multiple small displays. Such displays can be legible only if the clusters are presented in a schematized form. We suggest a method for building schematic representations of clusters of trajectories. The method transforms trajectories into aggregate flows between areas. The areas are automatically defined on the basis of the spatial distribution of the characteristic points from the trajectories. We use numeric measures of the abstraction quality and techniques for controlling the quality.

1 INTRODUCTION

In our current research, we develop approaches to visual exploration of large amounts of movement data, such as results of tracking animals, people, or vehicles. Traditional visualization techniques, particularly, animated map and space-time cube, can effectively visualize a small number of geometrically simple trajectories made during a relatively short time period. However, with increasing number of moving entities, length of the time period, and/or geometric complexity of the trajectories, these displays soon become illegible. A typical approach to visualizing large sets of trajectories is density maps [4][8]. While these may be suitable for certain analysis tasks, they do not fully capture the essence of movement as change of spatial position.

Another approach is grouping of trajectories by similarity and spatial proximity using clustering techniques [1][2][7] and consideration of the clusters instead of the individual trajectories. A question is how to visualize a resulting group (cluster) of similar trajectories so as to give a clear idea about the commonalities between the trajectories and at the same time about the degree of internal variance in the cluster.

Furthermore, there is a need of simultaneous visualization of multiple clusters of trajectories. The main problem is that trajectories are not disjoint in space; they intersect and overlap. As a consequence, summarized representations of groups of trajectories will also intersect and overlap if put in the same display. Hence, it would be appropriate to use a "small multiples" view, i.e. multiple juxtaposed maps or other displays each representing a single cluster. Since each of these displays has to be small, the clusters need to be represented in a highly schematic way such that only the principal features of each cluster are visible (but these features must be very easy to grasp). Such a representation may be called graphical model [3][5].

Our approach to building spatial graphical models of clusters of trajectories is based on the idea of summarizing trajectories into aggregate moves between appropriately defined areas [1]. An aggregate move between two areas summarizes a set of fragments of trajectories starting in the first area and ending in the second area. An aggregate move is represented on a map by an arrow with the thickness proportional to the number of trajectory fragments summarized in the move.

The main problem here is the generation of appropriate areas. We have devised a special method for partitioning the territory based on extracting significant points from the trajectories. The resulting abstraction conveys the principal characteristics of the movement. The level of the abstraction can be controlled through the parameters of the method. We use local and global numeric measures of the quality of the generalization and interactive and automated techniques for improving the quality in selected parts of the territory where this is deemed necessary.

2 THE METHOD

In brief, we suggest a method that extracts specific points from the trajectories (starts, ends, turns and stops), groups them into spatial clusters, and uses the centers of the clusters as generating points for Voronoi tessellation of the territory. The resulting cells are used for aggregating movement data and building flow maps. The degree of the generalization depends on the sizes of the Voronoi cells which, in turn, depend on the spatial extents of the point clusters. The desired spatial extent (radius) is a parameter of the method. The algorithms will be presented in the poster.

We illustrate the method using the dataset collected by GPS-tracking of 17,241 cars in Milan (Italy) during one week. The records include car identifier, time stamp (date and time of the day), and geographical coordinates. The data have been kindly provided by the Municipality of Milan. For clustering, we apply the density-based clustering algorithm OPTICS [6]. Our implementation allows the use of different distance functions assessing the similarity between trajectories [2][7]. The algorithm requires the data to be loaded in RAM, which is impossible for the whole dataset. Therefore, we use a subset of the data consisting of about 4200 trajectories.

The following examples demonstrate the application of the generalization method to clusters of trajectories obtained with the use of three different distance functions. We use special flow symbols in the form of half-arrows indicating the direction of the movement. The widths of the half-arrows are proportional to the number of the summarized trajectory segments. Fig.1 presents examples of clusters of trajectories having close end positions, i.e. common destinations. We clearly see the destination place of each cluster, the major routes leading to it, the relative frequencies of using the routes, and small deviations from them.

Fig.2 shows clusters of trajectories having close starts and end positions. We see the common origins and destinations of the trips in each cluster. We also see that people typically use the same routes although deviations also occur.

Fig. 3 shows clusters of trajectories having similar routes. In each cluster, we see the common part of the trajectories as a sequence of thick arrows (main flow). We also see that some trajectories start before the main flow and some continue farther. We can estimate the relative amounts of these trajectories with respect to the cluster size. We also see branching in some clusters.

http://geoanalytics.net/and andrienko@geoanalytics.net

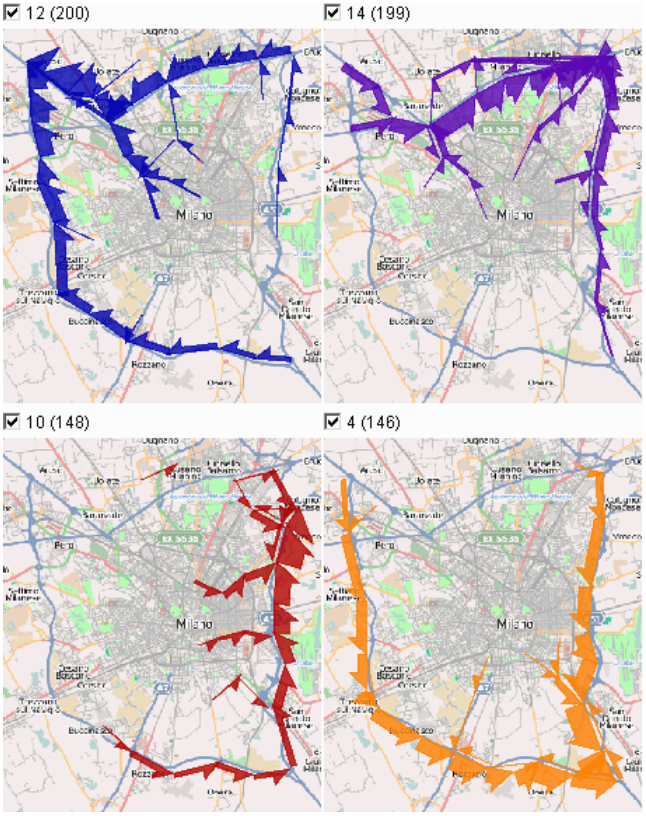

Figure 1. Examples of clusters of trajectories with close ends.

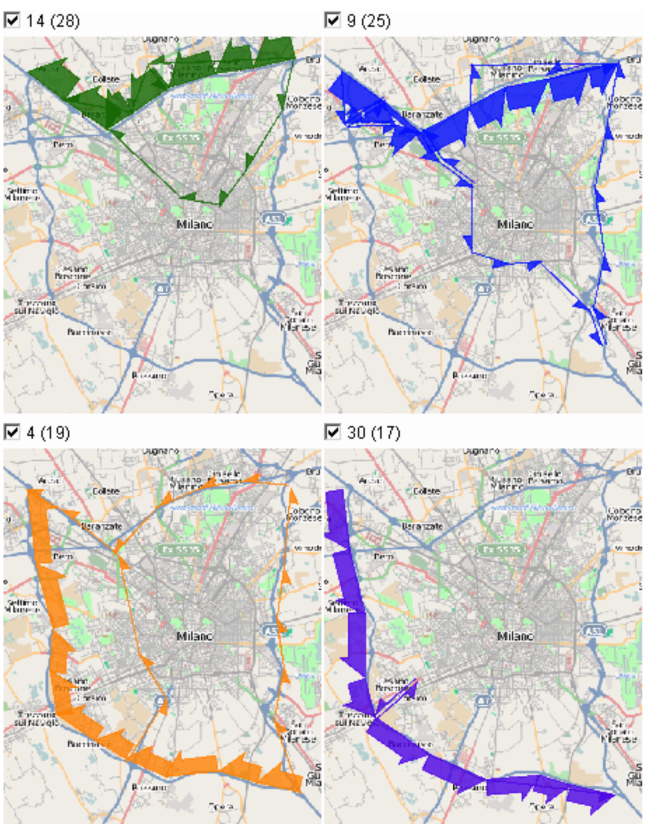

Figure 2. Examples of clusters of trajectories with close starts and ends.

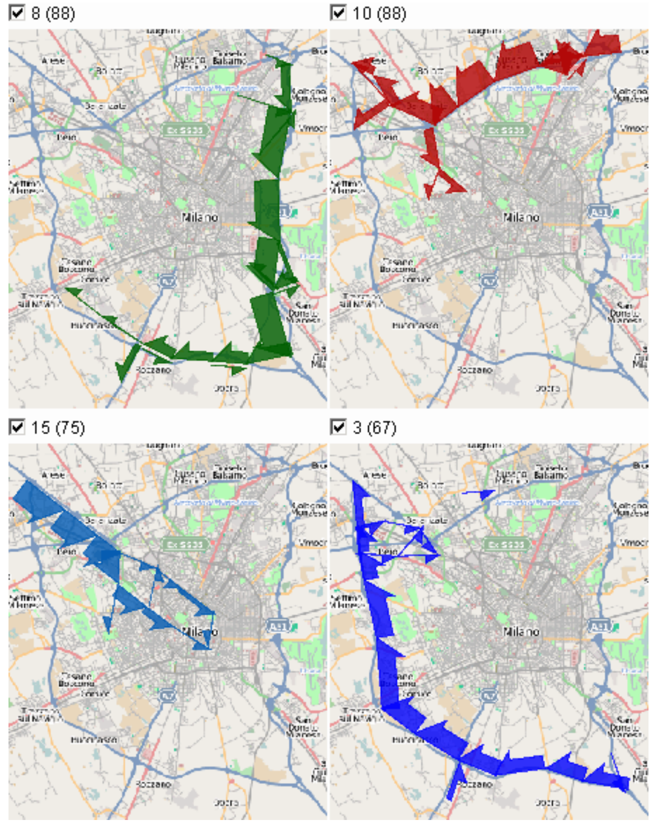

Figure 3. Examples of clusters of trajectories with similar routes.

The examples show that the spatial graphical models convey well the principal spatial and quantitative characteristics of the trajectories in each cluster. In the poster we shall introduce the numeric measures of the abstraction quality. Further research work will be done on building spatio-temporal graphical models of clusters of trajectories with similar temporal properties.

REFERENCES

[1] Andrienko, G., & Andrienko, N. 2008. Spatio-temporal aggregation for visual analysis of movements. In Proceedings VAST 2008, IEEE Computer Society Press, 51-58

[2] Andrienko, G., Andrienko, N., Rinzivillo, S., Nanni, M., Pedreschi, D., and Giannotti, F. 2009. Interactive Visual Clustering of Large Collections of Trajectories. In Proceedings VAST 2009, in press

[3] Cheylan, J.-P., Libourel, T., & Mende, C. 1997. Graphical modelling for geographic explanation. Spatial Information Theory: A Theoretical Basis for GIS (LNCS Vol. 1329), Springer, 473-483

[4] Dykes, J. A. & Mountain, D. M. 2003. Seeking structure in records of spatio-temporal behavior: visualization issues, efforts and applications, *Computational Statistics and Data Analysis*, 43 (Data Visualization II Special Edition), 581-603.

[5] van Elzakker, C.P.J.M. 2004. The use of maps in the exploration of geographic data. ITC Dissertation 116, Utrechtse Geografische Studies 326, University of Utrecht, Utrecht

[6] Ester, M., Kriegel, H.-P., Sander, J., and Xu, X. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proc. ACM KDD* 1996, 226-231

[7] Rinzivillo, S., Pedreschi, D., Nanni, M., Giannotti, F., Andrienko, N., & Andrienko, G. 2008. Visually–driven analysis of movement data by progressive clustering. Information Visualization, 7(3/4), 225-239

[8] Willems, N., van de Wetering, H., van Wijk, J.J. 2009. Visualization of vessel movements. Computer Graphics Forum, 28(3), 959-966