

Hermoupolis: A Trajectory Generator for Simulating Generalized Mobility Patterns

Nikos Pelekis¹, Christos Ntrigkogias², Panagiotis Tampakis²,
Stylios Sideridis², and Yannis Theodoridis²

¹ Dept. of Statistics and Insurance Science, Univ. of Piraeus, Greece
npelekis@unipi.gr

² Dept. of Informatics, Univ. of Piraeus, Greece
xdrigog@gmail.com, {ptampak,ytheod}@unipi.gr, siderste@yahoo.gr

Abstract. During the last decade, the domain of mobility data mining has emerged providing many effective methods for the discovery of intuitive patterns representing collective behavior of trajectories of moving objects. Although a few real-world trajectory datasets have been made available recently, these are not sufficient for experimentally evaluating the various proposals, therefore, researchers look to synthetic trajectory generators. This case is problematic because, on the one hand, real datasets are usually small, which compromises scalability experiments, and, on the other hand, synthetic dataset generators have not been designed to produce mobility pattern driven trajectories. Motivated by this observation, we present *Hermoupolis*, an effective generator of synthetic trajectories of moving objects that has the main objective that the resulting datasets support various types of mobility patterns (clusters, flocks, convoys, etc.), as such producing datasets with available ground truth information.

Keywords: Mobility Data Mining, Trajectory Patterns, Synthetic Generators.

1 Introduction

The explosion of mobile devices and positioning technologies has now made possible and easier the collection of trajectory data of moving objects. The rapid growth of these technologies has increased the interest for data analysis upon trajectory datasets. As such, the field of mobility data mining has already many success stories to narrate, as these are described by works that identify various types of patterns, including, among others, clusters of entire trajectories [11] or of sub-trajectories [9][14], moving clusters [7], flocks [8][5], sequential trajectory patterns [3], convoys [6], swarms [10], and top- k representative trajectory samples [12].

The effectiveness of most of the afore-mentioned methods has been evaluated with the use of small datasets w.r.t. potential sizes of real-world datasets, which however, are not available, usually due to privacy issues. Even when datasets are available, the ground truth for such kind of patterns is absent, thus researchers have to evaluate their proposals with general-purpose validation metrics (e.g. intra vs. inter cluster distance). On the other hand, utilizing synthetic generators is a typical approach for

researchers since it can support scalability experiments, however, synthetic datasets cannot guarantee the cardinality (or even the existence) of patterns within the synthetic population. For instance, experimentation of density-based clustering algorithms, like T-OPTICS [11] and TRACCLUS [9], may be biased if the distribution of the data under experimentation does not include a sufficient number of density-connected groups of objects; similarly for other mobility patterns. We argue that the effectiveness, efficiency and scalability experiments should not be applied independently and over different datasets. Like efficiency and scalability, effectiveness should be tested in very large datasets; like effectiveness, efficiency and scalability should be tested in datasets that include patterns of varying, known cardinality. Only this way, experimental results are interpretable and useful.

To meet the above-described requirements, we present *Hermoupolis*¹, a *pattern-aware synthetic trajectory generator*, which produces annotated trajectories of moving objects following given mobility patterns. These mobility patterns imply the different profiles of movement that we want to reproduce, covering many of the examples cited earlier. On the other hand, related work includes data generators simulating either movement in free space, including GSTD [15], CENTRE [4] and C4C [11], or network-constrained movement, including Brinkhoff [1] and BerlinMOD [2]. Although the above generators present very interesting features, they cannot be considered pattern-aware, in the sense that we described earlier.

In what follows, Section 2 presents the details about *Hermoupolis* methodology and Section 3 describes demo specifications.

2 Generating Pattern-Aware Synthetic Trajectories

Hermoupolis takes as input a set of *Generalized Mobility Patterns (GMP)* along with a road network and a set of *Points of Interest (PoI)*, and generates a set of network-constrained trajectories conforming to the requirements posed by *GMP*.

More formally, a *GMP* is a time ordered sequence of pairs $\langle AMP, c \rangle$, where c is the cardinality of the trajectories that will be simulated and *AMP* is an *Atomic Mobility Pattern*. In turn, a *AMP* is a triplet $\langle MBB, MFV, AFV \rangle$, where *MBB* is a *Minimum Bounding Box* that approximates the spatio-temporal space where the motion of simulated trajectories takes place, *MFV* is a *Movement Feature Vector* containing parameters that affect movement (distribution of speed, duration, agility, etc.) and *AFV* is an *Annotation Feature Vector* that contains textual information that is used to annotate the simulated recordings. Having in mind the recent advances in *semantic trajectory* modeling [13], *AFV* includes meta-information about a Stop at a *PoI*, where the user performed e.g. a leisure activity, or a Move (or trip) between two *PoI*, in-between which was performed e.g. by foot for a fitness activity.

In order to perform the trajectory generation task, *Hermoupolis* simulates all concurrent *AMP* from the set of *GMP*, while at the same time it interprets the sequence of each *GMP* and appropriately simulates Stops and Moves. Stop could be either stillness at or jerky movement around a *PoI*; Move could be a navigation from one Stop to another.

The simulation of the trips between Stops is actually performed by the well-known Brinkhoff data generator [1], which has been appropriately adapted in order to

¹ Hermoupolis, polis (=town) of Hermes, is the capital of Cyclades prefecture in Greece.

constrain movements according to the spatio-temporal restrictions set by *MBB* and *MFV*. Moreover, road networks and PoI databases that are public available at OpenStreetMap open source repository can be automatically loaded in *Hermoupolis*.

3 Demo Specifications

Hermoupolis design allows for the simulation of an extensive range of mobility patterns. For instance, Fig. 1(a) illustrates the generation of a set of trajectories representing the mobility behaviour of six different profiles of people (each depicted with a different colour) throughout a weekday. Although non-visible, each of these profiles corresponds to a *GMP* and is accompanied by a respective activity pattern (e.g. the green profile reflects young, single, working men following Home – Campus – Leisure - Home pattern). Obviously, such a dataset is appropriate for evaluating clustering techniques, such as [11] and [9].

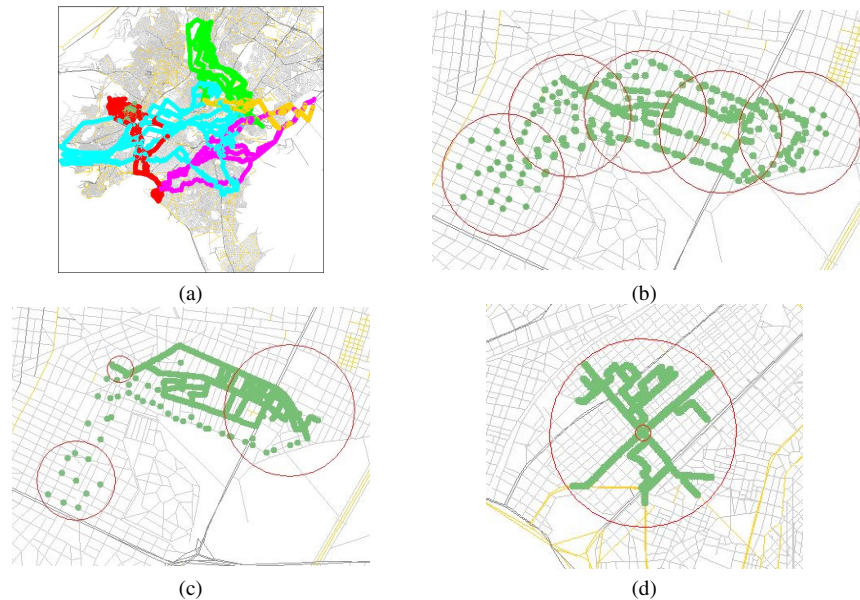


Fig. 1. Hermoupolis in action

Moreover, as a food for thought, in Fig. 1(b-d), we illustrate three simple different *GMP* aiming to simulate mobility patterns already available in the literature. In particular, Fig. 1(b) illustrates a *GMP* consisting of 9 *AMP* (5 Stops, for which we visualize their bounding circles, and 4 Moves). As the respective *MBB* of the 5 Stops are overlapping, one expects that some kind of flock or convoy pattern be hidden in the resulted trajectories. Then, in Fig. 1(c), we depict a *GMP*, composed of 3 Stops with varying spatial extent (indicated by the circles' radiuses). The first Move (between the first and second Stop) has large speed and small agility, while the opposite is true for the second Move (between the second and third Stop). We argue that such a simulation is appropriate for evaluating both sub-trajectory clustering algorithms, as well as moving object clusters like convoys, swarms, even patterns that

can capture the thickness of the mobility patterns. Finally, in Fig. 1(d), we exhibit a *GMP* that simulates objects starting from a wide Stop (the large circle) routed towards a short Stop (the small circle) that lies inside the first! Intuitively, this is a simple way of simulating that several people from a region converge to a “meeting place”.

Throughout the demonstration², users will be able to interact with *Hermoupolis* and generate the volumes of synthetic trajectories they wish by simulating various mobility patterns, such as the ones discussed above.

Acknowledgements. This work has been supported by the European Union’s FP7-DATASIM (grant 270833) and IRSES-SEEK (grant 295179) projects.

References

1. Brinkhoff, T.: Generating network-based moving objects. *GeoInformatica* 6(2), 153–180 (2002)
2. Duntgen, C., Behr, T., Guting, R.H.: BerlinMOD: a benchmark for moving object databases. *The VLDB Journal* 18(6), 34 (2008)
3. Giannotti, F., Nanni, M., Pedreschi, D., Pinelli, F.: Trajectory Pattern Mining. In: Proc. of SIGKDD (2007)
4. Giannotti, F., Mazzoni, A., Puntoni, S., Renso, C.: Synthetic generation of cellular network positioning data. In: Proc. of ACM GIS, pp. 12–20 (2005)
5. Gudmundsson, J., Kreveld, M.J., Speckmann, B.: Efficient detection of patterns in 2d trajectories of moving points. *GeoInformatica* 11(2), 195–215 (2007)
6. Jeung, H., Yiu, M.L., Zhou, X., Jensen, C.S., Shen, H.T.: Discovery of Convoys in Trajectory Databases. In: Proc. of VLDB (2008)
7. Kalnis, P., Mamoulis, N., Bakiras, S.: On discovering moving clusters in spatio-temporal data. In: Medeiros, C.B., Egenhofer, M., Bertino, E. (eds.) SSTD 2005. LNCS, vol. 3633, pp. 364–381. Springer, Heidelberg (2005)
8. Laube, P., Imfeld, S., Weibel, R.: Discovering relative motion patterns in groups of moving point objects. *IJGIS* 19(6), 639–668 (2005)
9. Lee, J.G., Han, J., Whang, K.Y.: Trajectory clustering: A partition-and-group framework. In: Proc. of SIGMOD, pp. 593–604 (2007)
10. Li, Z., Ding, B., Han, J., Kays, R.: Swarm: mining relaxed temporal moving object clusters. In: Proc. of PVLDB, vol. 3(1-2), pp. 723–734 (2010)
11. Nanni, M., Pedreschi, D.: Time-focused clustering of trajectories of moving objects. *JGIS* 27(3) (2006)
12. Panagiotakis, C., Pelekis, N., Kopanakis, I., Ramasso, E., Theodoridis, Y.: Segmentation and Sampling of Moving Object Trajectories based on Representativeness. In: TKDE (2011)
13. Parent, C., Spaccapietra, S., Renso, C., Andrienko, G., Andrienko, N., Bogorny, V., Damiani, M.L., Gkoulalas, D.A., Macedo, J.A., Pelekis, N., Theodoridis, Y., Yan, Z.: Semantic trajectories modeling and analysis. *ACM Computing Surveys* 35(4) (2013)
14. Pelekis, N., Kopanakis, I., Kotsifakos, E., Frentzos, E., Theodoridis, Y.: Clustering Uncertain Trajectories. *KAIS* 28(1), 117–147 (2011)
15. Theodoridis, Y., Silva, J.R.O., Nascimento, M.A.: On the generation of spatiotemporal datasets. In: Proc. of SSD (1999)

² A preview of *Hermoupolis* including screenshots from the GUI, and an accompanying video is available at: <http://infolab.cs.unipi.gr/pubs/pkdd2013/>.