MOVE project (COST Action IC0903)
STSM COST-STSM-IC0903-12660:
Scientific Report

STSM Applicant: *Corrado Loglisci, Department of Computer Science, University of Bari, Italy, 70125(IT) , loglisci@di.uniba.it*
Host: *Yannis Manolopoulos, Department of Informatics, Aristotle University of Thessaloniki, Thessaloniki(EL), manolopo@csd.auth.gr*
STSM Topic: *Mining group patterns in moving objects*
Period: *2013-06-29 00:00:00 to 2013-07-29 00:00:00*

1 Motivation and Purpose of the STSM

The tremendous advances of the positioning technologies, such as telemetry, GPS equipment and smart mobile phones, have enabled tracking of any type of moving objects and collection of the spatio-temporal data into growing repositories for a variety of domains. For instance, in location-based social networks, people travel in the real world and leave their location history in a form of trajectories. These trajectories do not only connect locations in the physical world but also bridge the gap between people and locations [8]. The portable GPS devices allow the recording in traffic management systems of the locations in the road networks of the vehicles where they are installed on [7]. Such information often include data on the human mobility. Thanks of the telemetry systems also the animal mobility can be tracked. Indeed, zoologists are investigating the impacts of the varying levels of urbanization on the migration, distribution, and habitat use of animals [6]. In these types of applications, one can be interested in the discovery of groups of objects which move together or similarly: for instance, in car pooling it could be useful to determine people with the same route to share the car.

In the literature, we can find an extensive research which investigates this kind of problems with techniques for mining groups of moving objects such as flock [1], convoy [3], and swarm[5], which differentiate from each other on the spatio-temporal properties which characterize the resulting groups. In particular, a flock contains at least m objects moving in the same direction within an circular region with a user-defined radius. Variants of the flock include also a notion of time-interval (with minimum duration defined by the user) according to which in each time-stamp of the interval a disc containing m objects can be identified. However, the fixed circular shape could be a drawback and can lead to miss groups with arbitrary forms. The introduction of the notion of density avoids this disadvantage and allows of discovering groups, named as convoys, which have not limitations on the shape and size. A convoy is defined as a cluster of objects and identified through a density-based clustering technique which checks for the condition of density-connectedness among the objects for all the time-stamps of a time-interval [3]. A more general type of the group is represented by the swarm which, differently from the flock and convoy, is not required to hold in all the time-stamps of a time-interval but it can occur more sporadically and anyway for at least k time-stamps not necessarily consecutive.

However, flocks, convoys and swarms allow to represent common and similar trajectories which members of those groups can exhibit but do not capture two relevant and interrelated aspects of the movements, namely *i)* the interactions among the objects of a group and *ii)* the dynamic behaviour of the objects in their movements. Interactions reflect the possible relationships in which the objects can be involved in space-time, they can provide a more complete description of the groups by explaining even the cause of their formation. Interactions can evolve because the objects can move near each other and then move away. Indeed, moving objects intrinsically are dynamic, their motion is not necessarily linear and it can be influenced by the properties or needs of each object and even by the interactions with other objects. For instance, in social studies, people expect to discover groups in which individuals start from different locations, they come near more and more until to join together in proximity of a point of interest, they remain there for a time and then go away from each other. These individuals may be members of a group even without having followed common trajectories. In that kind of problems, a group can turn out to be interesting not only when its members are spatially close and move similarly but also when they are far apart, they have different movements but common or similar interactions. In this mission, we investigated and designed a computational solution for discovering this kind of groups of moving objects which we refer as *communities* given the particular focus on the aspect of the interactions.

2 Description of the work carried out during the STSM

The contribution of Corrado Loglisci can organized in two main phases. Before than the visit, he studied the related literature concerning the techniques for the discovery of group patterns, such as those listed in the section 1, and algorithms for trajectory grouping, such as the clustering algorithms. During the visit, the work followed what planned in the STSM application.

First, he investigated the difference between the concept of *trajectory* and the concept of *movement*: the trajectory can intended as a geometrical abstraction of the movement path over time, while the movement models the dynamic behaviour of the moving objects conveying the physical notion of motion and can give complementary insight into the movement behaviour of objects, identifying patterns and their causes [2].

Following the idea on the interactions, he formalized a model to represent the trajectory data based on a set of movement parameters. These were defined to describe pairs of trajectories produced by pairs of objects involved in the interactions. This set of parameters can be modelled as a vector of geo-spatial features with categorical and numerical values.

Then, he designed and implemented an ad-hoc algorithm able to discover communities and provide a spatio-temporal characterization in terms of the feature vector before defined. The algorithm does not rely the clustering decision strictly on a distance/similarity notion and adopts a two-stepped procedure: *i)* arrangement of the vectors in an hierarchical structure and *ii)* exploration of this structure in order to identify groups of objects which share geo-spatial features across time. The grouping step can involve all possible pairs of the objects.

The implemented technique was empirically evaluated on a real world dataset and some preliminary experiments have been performed. The dataset concerns the trajectory of the animal during the migration[1]. Also, he identified some competitor techniques and other real world trajectories datasets[2].

Finally, as planned, he investigated the aspects of the current *centralized* implemented version which can be upgraded to develop a *parallel* version which resorts to the Map-Reduce framework. In particular, following the structure of the algorithm illustrated above, the trajectory data are partitioned on the hosts of the framework, each partition inputs the first step of the algorithm of clustering on each host. The produced partial results are finally "reduced" to perform the second step of the algorithm and obtain the final aggregated groups.

3 Description of the obtained results

The results obtained in this visit are of different form: the beneficiary researcher and the host institution share them via usual web-based file sharing platforms. The first deliverable is the implementation of the algorithm as a prototypical stand-alone system (encoded in Java) interfacing a relational database (MySQL DBMS).

The experiments were performed to test the behaviour of the algorithm with respect to the influence of the input parameters: we collected responses which we consider as scientific results.

Another result consists of the accessibility to the software tools implementing two competitor algorithms (among which TRACLUS [4]) and the representation of real world trajectories and the same datasets used by the competitors in the formalism of relational database. Software routines for converting data from a representation to the relational formalism were developed too.

The study of the MapReduce framework gave the possibility of acquiring expertise on the *i)* approaches to scale up a centralized algorithm to massive trajectories, and *ii)* modalities to develop on the Hadoop platform a parallel version of an existing implementation of a centralized algorithm.

4 Future collaboration with host institution

Thanks to this visit, a very promising collaboration has started between the host laboratory DELAB in Aristotle University of Thessaloniki and the home

[1] http://www.fs.fed.us/pnw/starkey/mapsdata.shtml
[2] http://research.microsoft.com/en-us/downloads/b16d359d-d164-469e-9fd4-daa38f2b2e13/

institution (Department of Computer Science of the University of Bari) on the basis of shared research interests regarding the definition of the techniques of spatio-temporal data mining.

First, we plan to complete the work on the parallel version for the Map-Reduce framework. Then, the distributed version will be experimented on the same datasets used for the centralized version and compared with it.

Second, another future work is the improvement of the effectiveness of the initial algorithm. Indeed, grouping all possible pairs of objects can imply a great computational cost which can be mitigated with a preliminary clustering of pairs suitable to form the communities.

Third, another future work will concern the upgrade of existing mobility data mining algorithms in the Hadoop framework. The existing algorithms will constitute the contribution of the home institution while the approaches for the parallel computation will constitute the contribution of the host institution.

5 Foreseen Publications

The first submission (currently in progress) is for a conference paper on Databases Systems and Knowledge Discovery topics. It concerns the implementation of the centralized algorithm and the comparison with the competitors. Preliminary experiments will be extended and performed on other datasets.

The second submission (which could be for a journal paper as well) will concern the implementation of the parallel version of the algorithm and the comparison with the centralized one.

References

1. M. Benkert, J. Gudmundsson, F. Hübner, and T. Wolle. Reporting flock patterns. *Comput. Geom.*, 41(3):111–125, 2008.
2. S. Dodge, P. Laube, and R. Weibel. Movement similarity assessment using symbolic representation of trajectories. *International Journal of Geographical Information Science*, 26(9):1563–1588, 2012.
3. H. Jeung, M. L. Yiu, X. Zhou, C. S. Jensen, and H. T. Shen. Discovery of convoys in trajectory databases. *PVLDB*, 1(1):1068–1080, 2008.
4. J.-G. Lee, J. Han, and K.-Y. Whang. Trajectory clustering: a partition-and-group framework. In *Proceedings of the 2007 ACM SIGMOD international conference on Management of data*, pages 593–604. ACM, 2007.
5. Z. Li, B. Ding, J. Han, and R. Kays. Swarm: Mining relaxed temporal moving object clusters. *PVLDB*, 3(1):723–734, 2010.
6. Z. Li, J. Han, M. Ji, L. A. Tang, Y. Yu, B. Ding, J.-G. Lee, and R. Kays. Movemine: Mining moving object data for discovery of animal movement patterns. *ACM TIST*, 2(4):37, 2011.
7. J. Yuan, Y. Zheng, C. Zhang, W. Xie, X. Xie, G. Sun, and Y. Huang. T-drive: driving directions based on taxi trajectories. In D. Agrawal, P. Zhang, A. El Abbadi, and M. F. Mokbel, editors, *GIS*, pages 99–108. ACM, 2010.
8. Y. Zheng, X. Xie, and W.-Y. Ma. Geolife: A collaborative social networking service among user, location and trajectory. *IEEE Data Eng. Bull.*, 33(2):32–39, 2010.