Project Proposal

Andres Calderon

October 13, 2016

Recently increase use of location-aware devices (such as GPS, Smart-phones and RFID tags) has allowed the collection of a vast amount of data with a spatial component linked to them. Different studies have focused in analyzing and mining this kind of collections [Leung, 2010][Miller and Han, 2001]. In this area, trajectory datasets have emerged as an interesting field where diverse kind of patterns can be identified [Zheng and Zhou, 2011][Vieira and Tsotras, 2013]. For instance, authors have proposed techniques to discover motion spatial patterns such as moving clusters[Kalnis et al., 2005], convoys[Jeung et al., 2008] and flocks [Benkert et al., 2006][Gudmundsson and van Kreveld, 2006]. In particular, [Vieira et al., 2009] and [Turdukulov et al., 2014] propose two novel algorithms to find moving flock patterns in very large spatio-temporal datasets.

A flock pattern is defined as a group of entities which move together for a defined lapse of time [Benkert et al., 2006]. Applications to this kind of patterns are diverse and range from surveillance to integrated transport systems. For example, [Turdukulov et al., 2014] explore the finding of this class of patterns to discover similarities between tropical cyclone paths. Also, [Calderon Romero, 2011] finds moving flock patterns in iceberg trajectories to understand their movement behavior and how they related to changes in ocean's currents.

The algorithms proposed by [Vieira et al., 2009] and [Turdukulov et al., 2014] share the same initial strategy to detect flock patterns. In that, first they find clusters of points which could be close enough to initiate a flock for each time interval. This is a costly operation due to the large number of points and intervals to be analyzed. The technique uses a grid-based index and a stencil (see figure 1) to speed up the process but the complexity is still high.

One alternative that I would like to explore is the use of distributed parallel platforms such as Hadoop and Spark. Recently, systems as Simba [Xie et al., 2016] or GeoSpark [Yu et al., 2016] added indexing and spatial operation capabilities to the Spark environment allowing distributed spatial analysis in-memory.

As a result of this project, I expect to implement a parallel version of the first part of the algorithm proposed in [Vieira et al., 2009] using Simba and then deploy a set of experiments comparing its performance with the sequential version.

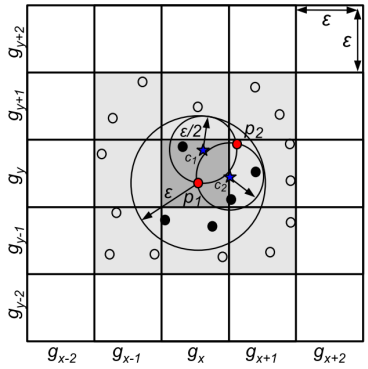

Figure 1: Grid-based index used in [Vieira et al., 2009].

References

Marc Benkert, Joachim Gudmundsson, Florian Hübner, and Thomas Wolle. Reporting Flock Patterns. In Yossi Azar and Thomas Erlebach, editors, *Algorithms – ESA 2006*, number 4168 in Lecture Notes in Computer Science, pages 660–671. Springer Berlin Heidelberg, September 2006. ISBN 978-3-540-38875-3 978-3-540-38876-0. URL `http://link.springer.com/chapter/10.1007/11841036_59`. DOI: 10.1007/11841036_59.

Andres Oswaldo Calderon Romero. Mining moving flock patterns in large spatio-temporal datasets using a frequent pattern mining approach. Master's thesis, University of Twente, 2011. URL `https://www.itc.nl/library/papers_2011/msc/gem/calderon.pdf`.

Joachim Gudmundsson and Marc van Kreveld. Computing Longest Duration Flocks in Trajectory Data. In *Proceedings of the 14th Annual ACM International Symposium on Advances in Geographic Information Systems*, GIS '06, pages 35–42, New York, NY, USA, 2006. ACM. ISBN 1-59593-529-0. doi: 10.1145/1183471.1183479. URL `http://doi.acm.org/10.1145/1183471.1183479`.

Hoyoung Jeung, Man Lung Yiu, Xiaofang Zhou, Christian S. Jensen, and Heng Tao Shen. Discovery of Convoys in Trajectory Databases. *Proc. VLDB Endow.*, 1(1): 1068–1080, August 2008. ISSN 2150-8097. doi: 10.14778/1453856.1453971. URL `http://dx.doi.org/10.14778/1453856.1453971`.

Panos Kalnis, Nikos Mamoulis, and Spiridon Bakiras. On Discovering Moving Clusters in Spatio-temporal Data. In Claudia Bauzer Medeiros, Max J. Egenhofer, and Elisa Bertino, editors, *Advances in Spatial and Temporal Databases*, number 3633 in Lecture Notes in Computer Science, pages 364–381. Springer Berlin Heidelberg, August 2005. ISBN 978-3-540-28127-6 978-3-540-31904-7. URL `http://link.springer.com/chapter/10.1007/11535331_21`. DOI: 10.1007/11535331_21.

Yee Leung. *Knowledge Discovery in Spatial Data*. Springer Science & Business Media, March 2010. ISBN 978-3-642-02664-5.

Harvey J. Miller and Jiawei Han. *Geographic Data Mining and Knowledge Discovery*. Taylor & Francis, Inc., Bristol, PA, USA, 2001. ISBN 0415233690.

Ulanbek Turdukulov, Andres Oswaldo Calderon Romero, Otto Huisman, and Vasilios Retsios. Visual mining of moving flock patterns in large spatio-temporal data sets using a frequent pattern approach. *International Journal of Geographical Information Science*, 28(10):2013–2029, October 2014. ISSN 1365-8816. doi: 10.1080/13658816.2014.889834. URL `http://dx.doi.org/10.1080/13658816.2014.889834`.

Marcos R. Vieira and Vassilis Tsotras. *Spatio-Temporal Databases: Complex Motion Pattern Queries*. Springer Science & Business Media, October 2013. ISBN 978-3-319-02408-0.

Marcos R. Vieira, Petko Bakalov, and Vassilis J. Tsotras. On-line Discovery of Flock Patterns in Spatio-temporal Data. In *Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, GIS '09, pages 286–295, New York, NY, USA, 2009. ACM. ISBN 978-1-60558-649-6. doi: 10.1145/1653771.1653812. URL `http://doi.acm.org/10.1145/1653771.1653812`.

Dong Xie, Feifei Li, Bin Yao, Gefei Li, Liang Zhou, and Minyi Guo. Simba: Efficient in-memory spatial analytics. In *Proceedings of the 2016 International Conference on Management of Data*, SIGMOD '16, pages 1071–1085, New York, NY, USA, 2016. ACM. ISBN 978-1-4503-3531-7. doi: 10.1145/2882903.2915237. URL `http://doi.acm.org/10.1145/2882903.2915237`.

J. Yu, J. Wu, and M. Sarwat. A demonstration of geospark: A cluster computing framework for processing big spatial data. In *2016 IEEE 32nd International Conference on Data Engineering (ICDE)*, pages 1410–1413, May 2016. doi: 10.1109/ICDE.2016.7498357.

Yu Zheng and Xiaofang Zhou. *Computing with Spatial Trajectories*. Springer Science & Business Media, October 2011. ISBN 978-1-4614-1629-6.