

INVITED ARTICLE

Inferring movement patterns from geometric similarity

Maike Buchin

Department of Mathematics, Ruhr University Bochum, Germany

Carola Wenk

Department of Computer Science, Tulane University, USA

Received: October 17, 2020; accepted: November 4, 2020

Abstract: Spatial movement data nowadays is becoming ubiquitously available, including data of animals, vehicles and people. This data allows us to analyze the underlying movement. In particular, it allows us to infer movement patterns, such as recurring places and routes. Many methods to do so rely on the notion of similarity of places or routes. Here we briefly survey how research on this has developed in the past 15 years and outline challenges for future work.

Keywords: Movement data, movement patterns, similarity measures, popular places, popular routes

1 Introduction

Due to technological advances more and more precise movement data is being collected. Movement data often comes in the form of tracking data using GPS, but can also be recorded in other ways. This data is collected and analysed in different application domains, such as ecology, sports, and urban planning. Also in every day life it is becoming more common to track ones movement, e.g., with a smart phone or watch. Such data allows to gain insights into the underlying movement, in particular to infer *patterns of movement*. A pattern generally refers to a recurring theme or aspect. In the context of movement, the most common patterns are a place repeatedly visited or a route repeatedly traveled. Other patterns for instance depend on the interaction of two or more entities, such as a follower-leader pattern or joint movement [16, 17, 19, 24, 35].

Extracting movement patterns is a challenge that has been addressed by researchers from different disciplines in recent years. In particular, this has been addressed in application domains, such as movement ecology, in geographical information science, and in computer science. Two main questions arise for these patterns: how to effectively model them and how to efficiently detect them. Many patterns of movement, such as frequently visited places or routes, are based on geometric similarity of places or routes. That is, a frequently visited place or route is one that is similar to a place or route taken by many in the observed data. Hence these patterns can be determined by quantifying and detecting the underlying similarity.

Several different measures of similarity have been proposed for places, routes and other aspects of movement, or in terms of geometric objects: for points, curves and other objects. How similarity is defined determines the type of movement pattern that is detected. For instance, it makes a difference whether/how the similarity of places or routes take into account the temporal component or the geographic context of the movement. But also, how similarity is defined determines how efficiently movement patterns based on it can be computed.

In the past 15 years, computational movement analysis [18, 23], which more broadly studies ways of analysing movement data, has been an active research area. However many challenges still remain, including how to handle uncertainty, how to include context or data from multiple sources, and very large amounts of data [13–15, 25, 26, 29].

2 Similarity measures

Many similarity measures have been proposed for different geometric objects. For point sets, a popular measure is the Hausdorff distance, which matches each point to its closest neighbor in the other set. Another measure is the Earth movers distance, which finds a global matching of both point sets. Whereas the first can be efficiently computed (in log-linear time), for the latter only approximation algorithms (with fairly high running times) are known. Beyond that several other measures exist [5, 32].

For curves, also many different distance measures exist [6, 32]. These differ, among other things, how they treat the time component of a curve resulting from a movement path. A simple measure is again the Hausdorff distance applied to the underlying point set of the curve. However, this ignores the course of the curves, which in contrast is taken into account by the Fréchet distance. Another popular similarity measure is the equal-time distance, i.e., comparing distances at equal time stamps. Of these three, the latter can be computed most efficiently (in linear time), the first nearly as efficiently (log-linear time), while the Fréchet distance is less efficient to compute (log-quadratic time). Furthermore, more distance measures have been considered, for instance variants of the measures above and of edit distance and dynamic time warping [30].

For geometric graphs, measuring the similarity becomes even more complex. In particular, the question arises, how to take into account both the geometry and topology of the graph in the comparison. Hence many different approaches to graph similarity exist [2]. Measures that solely consider topology, such as graph isomorphisms have little meaning for geometric graphs. Ideally, both the geometry and topology should be taken into account. Two measures that do this are a geometric edit distance for graphs [12] and graph similarity based on graph mappings [4].

3 Movement patterns

In recent years, many patterns of movement have been considered [16,17,24,35]. Here we focus on those that inherently build on geometric similarity of the places visited.

Perhaps the most common movement pattern are recurring or popular places. This pattern has applications for many types of movement data, such as detecting home places and hot spots. A popular place can be modeled as a place often visited, more specifically, this can be modeled as a cluster of similar trajectory points, or by a disc or rectangle often visited by the given trajectories. Note that this pattern differs from the notion of a stay point in a trajectory, i.e., a location and time span, where a one entity remained in one place for some time. Detecting and removing stay points is often used in preprocessing of trajectory data [28]. For detecting hot spots in trajectory data several methods exist [21,27]. And for the more general problem of clustering arbitrary point sets a multitude of methods exists [33].

A second common movement pattern are recurring routes. These occur for instance as commuting or migration routes, or as parts of a road network. A natural model of recurring routes in a set of trajectories is as a set of similar subtrajectories. Here in particular, the computational complexity (i.e., how efficiently these can be computed) depends heavily on the similarity measure used and whether one seeks for similar subtrajectories or whole trajectories. For instance, the Fréchet distance is a good model for subtrajectory similarity, but costly to compute. And similar subtrajectories are harder to find, as one also needs to find start and endpoints of the subtrajectories. Hence several different approaches exist, which apply in different scenarios [34].

For recurring routes, several related problems occur: determine a single or a set of recurring routes, which cover all or only part of the data. Also finding a good (i.e., simple but realistic) representation for such a recurring route is another problem in itself [3,9].

Oftentimes, one does not want to find a single recurring route, or a set of such routes, but rather the entire underlying travel network. More specifically, this could be the street network used by cars in a city, or the migration network of one (or more) animal species. This can be modeled for instance by a graph that represents all the given trajectories. In recent years, much research has focused on the so-called *map construction problem* where the goal is to reconstruct the underlying road network. For this several different methods have been proposed [1,2]. In map construction, the aim is to construct the geometry and topology of the underlying network. Given the large amount and uncertainty of data, this quickly becomes a hard problem. Furthermore, for smaller inputs, e.g., groups of animals travelling together, one may want to extract further information, e.g., who was travelling with whom and when [10,11].

Another important type of movement pattern includes those that are defined based on the interaction of individuals. This includes the follower-leader and the joint movement pattern. The follower-leader pattern can be modeled by following similar routes with a slight delay [8,22], or that the follower always is behind (and sees) its leader [7]. A number of approaches have been proposed for detecting joint movement of several entities. In particular, the flock pattern [20,31], which models joint movement as a group of entities moving within a disk for some time, has received considerable attention.

4 Challenges

There are several challenges that apply to movement data in general, these include data storage, availability of algorithms, anonymisation of data, and linking data of different sources. Although many methods for detecting movement patterns have been proposed, many of these are not yet efficient, effective or reliable enough to deal with today's large and varying data. Specifically, the following challenges remain:

- **Efficiency:** Can we devise more efficient algorithms for very large data sets? Current algorithms for computing movement patterns have relatively high runtimes or may even be NP-hard. Different algorithmic approaches may include sampling, parallelization, or approximations.
- **Modeling and semantics:** Can we include more semantics in the modeling and analysis of patterns? Movement patterns could be enriched with more information on the type of movement, motivation for movement and/or its geographical context. To do so, we could rely not only on the movement data itself, but also on other data types such as social data.
- **Data sources:** How can other or multiple data sources be used to infer movement patterns? Geospatial tweets or check-in data for instance provide movement data at a much coarser level. Also, data providers often only make aggregate tracking data available due to privacy or other concerns. How can such aggregate data be used to infer movement patterns? And are there ways to aggregate data so that movement patterns can still be inferred?
- **Behavioral patterns and ethics:** How can we model and infer behavioral movement patterns? Behavioral patterns for both humans and animals typically involve interacting with each other or with the geographic context, and hence are more complex to model. While detecting behavioral patterns has important applications for example in emergency management in urban areas, detecting such patterns in human data raises ethical issues that need to be addressed.

Solving these problems in particular requires close interdisciplinary collaborations between computer scientists and domain scientists from various disciplines including ecology, geography, urban studies, and social and behavioral sciences.

Acknowledgments

C. Wenk was partially supported by National Science Foundation grant CCF-1637576.

References

- [1] AHMED, M., KARAGIORGOU, S., PFOSE, D., AND WENK, C. A comparison and evaluation of map construction algorithms using vehicle tracking data. *GeoInformatica* 19, 3 (2014), 601–632. doi:10.1007/s10707-014-0222-6.
- [2] AHMED, M., KARAGIORGOU, S., PFOSE, D., AND WENK, C. *Map Construction Algorithms*. Springer, 2015. doi:10.2307/2265559.

- [3] AHN, H.-K., ALT, H., BUCHIN, M., OH, E., SCHARF, L., AND WENK, C. Middle curves based on discrete Fréchet distance. *Computational Geometry: Theory and Applications* 89, 101621 (2020).
- [4] AKITAYA, H. A., BUCHIN, M., KILGUS, B., SIJZEN, S., AND WENK, C. Distance measures for embedded graphs. In *30th International Symposium on Algorithms and Computation, ISAAC (2019)*, vol. 149 of *LIPICs*, Schloss Dagstuhl - Leibniz-Zentrum für Informatik, pp. 55:1–55:15.
- [5] ALT, H., AND GUIBAS, L. Discrete geometric shapes: Matching, interpolation, and approximation: A survey. *Handbook of Computational Geometry* (03 1997).
- [6] ALT, H., KNAUER, C., AND WENK, C. Comparison of distance measures for planar curves. *Algorithmica* 38 (2003), 45–58. doi:10.1007/s00453-003-1042-5.
- [7] ANDERSSON, M., GUDMUNDSSON, J., LAUBE, P., AND WOLLE, T. Reporting leaders and followers among trajectories of moving point objects. *GeoInformatica* 12, 4 (2008), 497–528. doi:10.1007/s10707-007-0037-9.
- [8] BUCHIN, K., BUCHIN, M., AND GUDMUNDSSON, J. Detecting single file movement. In *Proceedings of the 16th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems* (2008), GIS '08. doi:10.1145/1463434.1463476.
- [9] BUCHIN, K., BUCHIN, M., VAN KREVELD, M., LÖFFLER, M., SILVEIRA, R. I., WENK, C., AND WIRATMA, L. Median trajectories. *Algorithmica* 66, 3 (2013), 595–614.
- [10] BUCHIN, K., BUCHIN, M., VAN KREVELD, M., SPECKMANN, B., AND STAALS, F. Trajectory grouping structure. In *Algorithms and Data Structures* (2013), Springer, pp. 219–230.
- [11] BUCHIN, M., KILGUS, B., AND KÖLZSCH, A. Group diagrams for representing trajectories. *International Journal of Geographical Information Science* (2019), 1–33. doi:10.1080/13658816.2019.1684498.
- [12] CHEONG, O., GUDMUNDSSON, J., KIM, H.-S., SCHYMURA, D., AND STEHN, F. Measuring the similarity of geometric graphs. In *Experimental Algorithms* (Berlin, Heidelberg, 2009), Springer, pp. 101–112.
- [13] DEMSAR, U., BUCHIN, K., CAGNACCI, F., SAFI, K., SPECKMANN, B., VAN DE WEGHE, N., WEISKOPE, D., AND WEIBEL, R. Analysis and visualisation of movement: an interdisciplinary review. *Movement Ecology* 3 (2015). doi:10.1186/s40462-015-0032-y.
- [14] DODGE, S., GAO, S., TOMKO, M., AND WEIBEL, R. Progress in computational movement analysis – towards movement data science. *International Journal of Geographical Information Science* (2020). doi:10.1080/13658816.2020.1784425.
- [15] DODGE, S., WEIBEL, R., AHEARN, S., BUCHIN, M., AND MILLER, J. Analysis of movement data. *International Journal of Geographical Information Science* 30 (2016), 1–10. doi:10.1080/13658816.2015.1132424.

- [16] DODGE, S., WEIBEL, R., AND LAUTENSCHÜTZ, A.-K. Towards a taxonomy of movement patterns. *Information Visualization* 7, 3-4 (2008), 240–252. doi:10.1057/PALGRAVE.IVS.9500182.
- [17] GIANNOTTI, F., NANNI, M., PINELLI, F., AND PEDRESCHI, D. Trajectory pattern mining. In *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (2007), Association for Computing Machinery, p. 330–339. doi:10.1145/1281192.1281230.
- [18] GUDMUNDSSON, J., LAUBE, P., AND WOLLE, T. *Computational Movement Analysis*. Springer, 2012, pp. 423–438. doi:10.1007/978-3-540-72680-7_22.
- [19] GUDMUNDSSON, J., LAUBE, P., AND WOLLE, T. Movement patterns in spatio-temporal data. In *Encyclopedia of GIS*, S. Shekhar, H. Xiong, and X. Zhou, Eds. Springer, 2017, pp. 1362–1370.
- [20] GUDMUNDSSON, J., AND VAN KREVELD, M. Computing longest duration flocks in trajectory data. In *Proceedings of the 14th Annual ACM International Symposium on Advances in Geographic Information Systems* (2006), pp. 35–42. doi:10.1145/1183471.1183479.
- [21] GUDMUNDSSON, J., VAN KREVELD, M., AND STAALS, F. Algorithms for hotspot computation on trajectory data. In *Proceedings of the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems* (2013), pp. 134–143. doi:10.1145/2525314.2525359.
- [22] KONZACK, M., MCKETTERICK, T., OPHELDERS, T., BUCHIN, M., GIUGGIOLI, L., LONG, J., NELSON, T., WESTENBERG, M. A., AND BUCHIN, K. Visual analytics of delays and interaction in movement data. *International Journal of Geographical Information Science* 31, 2 (2017), 320–345. doi:10.1080/13658816.2016.1199806.
- [23] LAUBE, P. *Computational Movement Analysis*. Springer, 2014. doi:10.1007/978-3-319-10268-9.
- [24] LAUBE, P., IMFELD, S., AND WEIBEL, R. Discovering relative motion patterns in groups of moving point objects. *International Journal of Geographical Information Science* 19, 6 (2005), 639–668. doi:10.1080/13658810500105572.
- [25] LONG, J. A., WEIBEL, R., DODGE, S., AND LAUBE, P. Moving ahead with computational movement analysis. *International Journal of Geographical Information Science* 32, 7 (2018), 1275–1281. doi:10.1080/13658816.2018.1442974.
- [26] MAZIMPAKA, J. D., AND TIMPF, S. Trajectory data mining: A review of methods and applications. *Journal of Spatial Information Science* 13 (2016), 61–99.
- [27] NIKITOPOULOS, P., PARASKEVOPOULOS, A., DOULKERIDIS, C., PELEKIS, N., AND THEODORIDIS, Y. Hot spot analysis over big trajectory data. *2018 IEEE International Conference on Big Data* (2018), 761–770.
- [28] PARENT, C., SPACCAPIETRA, S., RENSO, C., ANDRIENKO, G., ANDRIENKO, N., BOGORNY, V., DAMIANI, M. L., GKOUALALAS-DIVANIS, A., MACEDO, J., PELEKIS, N., THEODORIDIS, Y., AND YAN, Z. Semantic trajectories modeling and analysis. *ACM Comput. Surv.* 45, 4 (2013). doi:10.1145/2501654.2501656.

- [29] PURVES, R., LAUBE, P., BUCHIN, M., AND SPECKMANN, B. Moving beyond the point: An agenda for research in movement analysis with real data. *Computers, Environment and Urban Systems* 47 (2014). doi:10.1016/j.compenvurbsys.2014.06.003.
- [30] TOOHEY, K., AND DUCKHAM, M. Trajectory similarity measures. *SIGSPATIAL Special* 7, 1 (2015), 43–50. doi:10.1145/2782759.2782767.
- [31] TURDUKULOV, U., ROMERO, A. O. C., HUISMAN, O., AND RETSIOS, V. 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 (2014), 2013–2029. doi:10.1080/13658816.2014.889834.
- [32] VELTKAMP, R. Shape matching: Similarity measures and algorithms. In *International Conference on Shape Modeling and Applications* (06 2001), pp. 188–197. doi:10.1109/SMA.2001.923389.
- [33] XU, D., AND TIAN, Y. A comprehensive survey of clustering algorithms. *Annals of Data Science* 2 (2015). doi:10.1007/s40745-015-0040-1.
- [34] YUAN, G., SUN, P., ZHAO, J., LI, D., AND WANG, C. A review of moving object trajectory clustering algorithms. *Artificial Intelligence Review* 47 (2016). doi:10.1007/s10462-016-9477-7.
- [35] ZHENG, Y. Trajectory data mining: An overview. *ACM Transactions on Intelligent Systems and Technology* 6, 3 (2015). doi:10.1145/2743025.