Predicting trajectories with probabilistic time geography and massive unconstrained movement data

A. Graser[1,2], J. Schmidt[1], P. Widhalm[1]

[1]AIT Austrian Institute of Technology GmbH, Giefinggasse 2, 1210 Vienna, Austria
[2]University of Salzburg, Kapitelgasse 4-6, 5020 Salzburg, Austria
Email: anita.graser@ait.ac.at, johanna.schmidt@ait.ac.at, peter.widhalm@ait.ac.at

1. Introduction

While existing work in time geography focused primarily on representation and semantics (Hägerstrand, 1970; Dodge et al., 2016), current movement research in GIScience aims to extend these approaches towards information extraction and predictive modelling. So far, methods for extracting useful information from increasingly massive movement data are lagging behind the technology for generating data (Long & Nelson, 2013). Therefore, frameworks and predictive models for movement data are an important research avenue towards understanding, simulating, and predicting complex movement (Long & Nelson, 2013, Dodge et al., 2016).

Classical time geography does not account for object kinetics and other physical constraints and thus resulting space-time cones and prisms overestimate movement opportunities as shown by Kuijpers et al. (2011). Kinetic time geography (Kuijpers et al., 2011) provides a more realistic representation of movement opportunities and constraints. Extending this approach, Long et al. (2014) integrate kinetic and probabilistic time geography (Winter, 2009) which accounts for movement probabilities. Other current work in probabilistic time geography (Winter & Yin, 2011; Song & Miller, 2014; Tang et al., 2016; Song et al., 2016; Kuijpers et al., 2017) models visit probabilities between observed locations, but is not used to extrapolate future locations. The integration of probabilistic time geography into prediction models is a promising option to account for uncertainty.

Commonly used trajectory prediction approaches are simple linear prediction with constant speed and direction, or approaches taking into account acceleration and rate of turn. Another common approach are Hidden Markov models (HMMs) which are used to predict both animal and vehicle movement (Langrock et al., 2012; Mathew et al., 2012; Michelot et al., 2016; Ye et al., 2016). Particularly in research on vehicle movement, kinematic movement models that are specific to certain moving objects (for example, a specific ship type) are built from kinematic equations whose development requires information that is often not readily available. Instead, taking advantage of historical massive movement data, novel data-driven approaches that emulate kinematic models are being developed (Zorbas et al., 2015).

In this paper we introduce and compare the results of two novel data-driven trajectory prediction approaches computing future locations of moving objects from massive historical movement data:

1. **Statistical Model Learning** provides predictions based on learned location-dependent direction and velocity distributions, combining massive movement

data models with the field-based perspective of time geography (Miller & Bridwell, 2009)
2. **Similar trajectory search** provides predictions solely based on identifying similar trajectories in historical massive movement datasets, corresponding with an object-based perspective of time geography

The two prediction approaches are suitable for settings where objects move freely and are not constrained by a transport network. We demonstrate the applicability of our approaches for cargo vessel movements recorded by Automatic Identification System (AIS) data in the port of Gothenburg, Sweden. This AIS vessel tracking data is produced at irregular time intervals and therefore not suitable as input for HMMs as described by Langrock et al. (2012) and Michelot et al. (2016). Furthermore, while cargo vessels tend to follow shipping networks, extracting a usable vessel movement graph from AIS is still an open research issue (Dobrkovic et al., 2016). Network-based prism approaches are therefore not considered in this work because issues in the data-derived movement graph would negatively impact prediction results. Our approaches do not require the computation of a movement graph and therefore avoid this potential source of errors.

2. Data-driven trajectory prediction

The statistical model learning approach uses historical data to learn which movement directions and speeds were commonly observed at a certain location, represented as prototypes (Graser & Widhalm, 2018). This corresponds to the field-based perspective of time geography (Miller & Bridwell, 2009). To predict future locations, we follow the random walk approach introduced by Winter & Yin (2011) to deal with probabilistic time geography and applied by Long et al. (2014) to incorporate kinetic-based time geography. While Long et al. (2014) use skew normal distribution to model kinetic-based movement probabilities, we employ a data-driven approach where the distribution is learned from historical data. Based on the last observation of a moving object's location (including movement direction and velocity), we determine the best fitting learned model prototype. Then we draw from this prototype's heading and velocity distribution and project the position one step into the future. This is repeated for as many steps as necessary to reach the desired prediction time which corresponds to the time budget concept in time geography. The fact that we need to identify the best fitting prototype at each step makes the prediction computationally expensive.

The similar trajectory search approach (inspired by Wijaya & Nakamura (2013)) does not require learning a statistical model. Applying the object-based perspective of time geography, we could derive moving objects' kinematic properties such as maximum velocity, acceleration, and rate of turn from historic data and compute a corresponding kinetic space-time cone, but this would ignore local movement constraints such as speed limits, traffic separation schemes, or coast lines. Therefore, in the absence of detailed information about movement constraints, historical trajectories provide an approximation or sample of the potential paths. To be able to deal with the large amounts of historical data required for this approach, moving object data is stored in a spatially enabled big data cluster. To predict future locations, we first search the historical data for moving objects that were moving along a similar path. (While different distance metrics are conceivable to determine dissimilarity, the following demonstration uses GeoMesa's RouteSearchProcess to find movement data records that are heading along a given route using a maximum heading and distance tolerance.)

Then we determine the locations of those moving objects after the defined prediction time frame or available time budget.

We demonstrate the applicability of our data-driven prediction approaches using vessel movement data provided by the Danish Maritime Authority that also covers sections of the Swedish coast. The training dataset covers the month of June 2016 and consists of 360 million location records. Figure 1 shows the learned movement prototypes of the data-driven model (Graser & Widhalm, 2018) for cargo vessels at the entrance of the port of Gothenburg, Sweden. The spatial resolution is 680 meters, and each 680x680m cell can contain up to five prototypes. The learned prototypes represent movements along main vessel routes (with narrow distributions of movement direction and relatively high speeds, indicated by narrow dark blue triangles), as well as movement towards and stops at quays and in anchoring areas (indicated by wider light blue or white triangles).

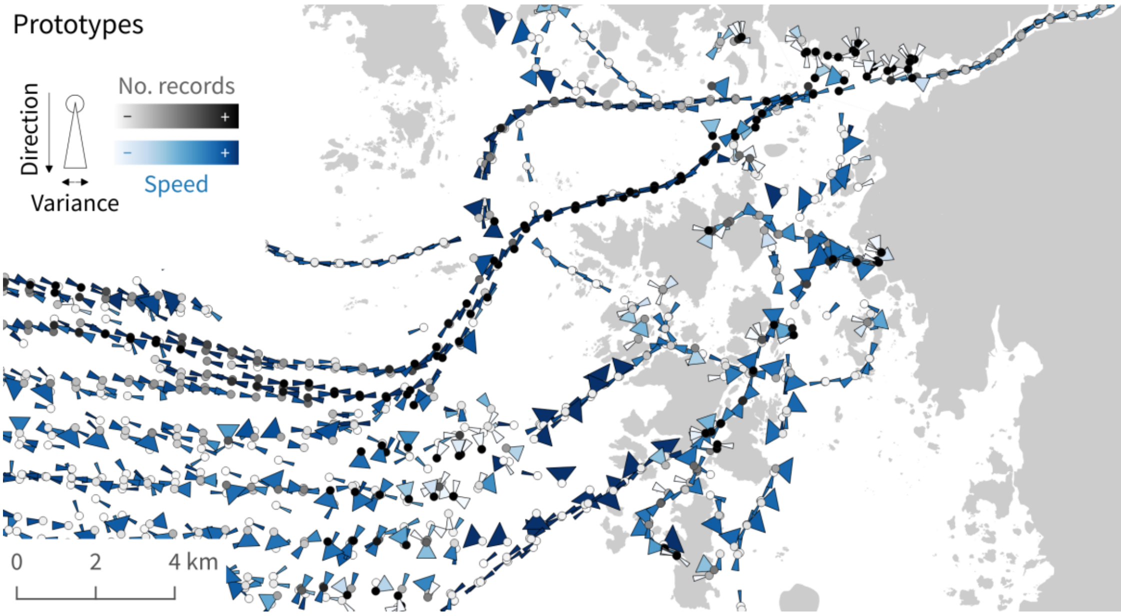

Figure 1. Learned prototypes in Gothenburg: speed (triangle colour), mean direction (triangle orientation), variability of direction values (triangle width), and number of records (circle colour). Background map data: OSM.

Figure 2 shows prediction examples for a vessel moving towards Gothenburg based on its observed position, direction, and speed indicated by the black arrow in the bottom left corner of the figures. Predictions are provided for 5, 10, and 15 minutes into the future. The statistical model learning predictions (blue dots in the top row) include 100 random walk iterations. The similar trajectory search results (red dots in the bottom row) include future locations of 87 similar trajectories.

Both approaches agree on the overall direction and general speed of future movements. The random walk methods are more clustered than similar trajectory search results. Figure 2 also illustrates some shortcomings of the prediction methods. For example, the statistical model learning predictions include locations on islands, since this approach is not aware of constraints such as coastlines. In contrast, the similar trajectory search approach profits from the implicit knowledge about where cargo vessels have previously appeared, thus successfully avoiding locations on land.

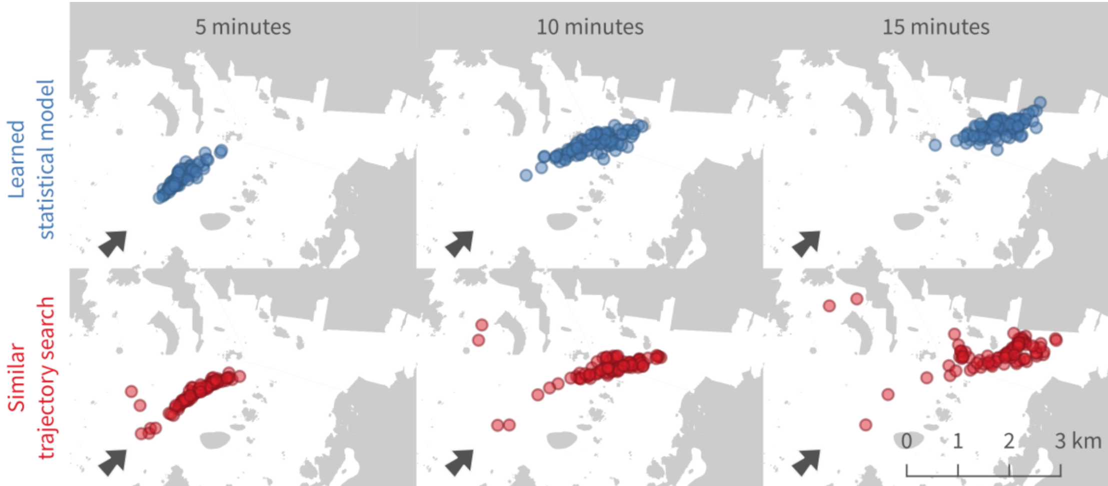

Figure 2. Potential future locations after 5/10/15 minutes based on the learned statistical model (top row, blue dots) and similar trajectory search (bottom row, red dots). Start position and heading depicted by the black arrow.

For our evaluation, we extracted one sample segment from each trajectory of the 14 cargo vessels that approached Gothenburg on July 1st 2016 with an average speed of 13kn (24km/h). The vessels have an average width of 24m and length of 169m. The statistical model was trained using data for June 2016 and the similar trajectory search uses June 2016 data as well. The linear prediction uses reported speed and course over ground. The statistical model learning results represent 5 random walk iterations per prediction with a step size of 20s. Each prediction step was logged and the predictions for 5, 10, and 15 minutes are presented here. They are therefore dependent on each other. The similar trajectory search uses the first 10 identified similar trajectories per prediction (heading tolerance: 10°, buffer size: 15m). Figure 3 shows the 15 minutes results for two incoming (red and blue) and two outgoing (orange and green) cargo vessels.

Table 1 presents a first quantitative evaluation of the data-driven prediction models as well as a comparison with commonly used linear prediction. These results show that data-driven predictions outperform linear prediction once prediction time exceeds 5min. Similar trajectory search results in smaller average location prediction errors but with higher standard deviation than the statistical approach. However, similar trajectory search's better prediction accuracy comes at higher data storage costs compared to statistical model learning.

Table 1. Average location prediction error (in meters) and standard deviation (in brackets).

Minutes	Linear prediction	Learned statistical model	Similar trajectory search
5	520 (±426)	582 (±404)	436 (±309)
10	1247 (±950)	1029 (±565)	919 (±823)
15	1923 (±1063)	1522 (±969)	1344 (±1383)

Figure 3. Predicted future locations after 15 minutes for two incoming (red and blue) and two outgoing (orange and green) cargo vessel trajectories.

3. Conclusions and future work

Our results show that both data-driven approaches provide promising results for predicting future locations of moving objects in settings that are unconstrained by a network. The geographic spread of potential future locations provides a visual impression of the uncertainty of the prediction. To compute these results, we used empirically defined parameters values (model resolution, random walk iterations, step size, heading tolerance, buffer size). For future work, an evaluation of parameter space could help identify how different parameters influence the results, and which parameter sets work best. Further next steps include methodological and performance improvements to both data-driven models as well as a comparison to kinematic prediction models. Furthermore, we aim to provide specific location predictions including quantitative indicators of the corresponding uncertainty.

Acknowledgements

This work was supported by the Austrian Federal Ministry for Transport, Innovation and Technology (BMVIT) within the programme "IKT der Zukunft" under Grant 861258 (project MARNG). The authors would like to thank the project partners Frequentis, indicate.digital.design.vision, and TeleConsult Austria.

References

Dobrkovic A, Iacob ME and Van Hillegersberg J, 2016, Maritime Pattern Extraction from AIS Data Using a Genetic Algorithm. In: IEEE, *International Conference on Data Science and Advanced Analytics (DSAA),* 642-651.

Dodge S, Weibel R, Ahearn SC, Buchin N and Miller JA, 2016, Analysis of movement data. *International Journal of Geographical Information Science*, 30(5), 825-834.

Graser A and Widhalm P, 2018, Modelling massive AIS streams with quad trees and Gaussian Mixtures. In: *Geospatial Technologies for All: short papers, posters* *and poster abstracts of the 21th AGILE Conference on Geographic Information Science*. Lund University 12-15 June 2018, Lund, Sweden.

Hägerstrand T, 1970, What about people in regional science? *Papers of the Regional Science Association*. 24 (1): 6–21.

Kuijpers B, Miller HJ and Othman W, 2011, Kinetic space-time prisms. In: 19th *ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. New York, 162–170.

Kuijpers B, Miller HJ and Othman W, 2017, Kinetic prisms: incorporating acceleration limits into space-time prisms. *International Journal of Geographic Information Science*, 31, 2164-2194.

Langrock R, King R, Matthiopoulos J, Thomas L, Fortin D and Morales JM, 2012, Flexible and practical modeling of animal telemetry data: hidden Markov models and extensions. *Ecology*, 93(11), 2336-2342.

Long JA and Nelson TA, 2013, A review of quantitative methods for movement data. *International Journal of Geographical Information Science*, 27(2), 292-318.

Long JA, Nelson TA and Nathoo FS, 2014, Toward a kinetic-based probabilistic time geography. *International Journal of Geographical Information Science*, 28(5), 855-874.

Mathew W, Raposo R, and Martins B, 2012, Predicting future locations with hidden Markov models. In *Proceedings of the 2012 ACM conference on ubiquitous computing* (pp. 911-918). ACM.

Michelot T, Langrock R and Patterson TA, 2016, moveHMM: An R package for the statistical modelling of animal movement data using hidden Markov models. *Methods in Ecology and Evolution*, 7(11), 1308-1315.

Miller HJ and Bridwell SA, 2009, A field-based theory for time geography. *Annals of the Association of American Geographers*, 99(1), 49-75.

Song Y and Miller HJ, 2014, Simulating visit probability distributions within planar space-time prisms. *International Journal of Geographical Information Science*, 28(1), 104-125.

Song Y, Miller HJ, Zhou X and Proffitt D, 2016, Modeling Visit Probabilities within Network-Time Prisms Using Markov Techniques. *Geographical Analysis*, 48(1), 18-42.

Tang J, Song Y, Miller HJ and Zhou X, 2016, Estimating the most likely space-time paths, dwell times and path uncertainties from vehicle trajectory data: A time geographic method. *Transportation Research C,* 66, 176-194.

Wijaya WM and Nakamura Y, 2013, Predicting ship behavior navigating through heavily trafficked fairways by analyzing AIS data on Apache HBase. In: IEEE, *First International Symposium on Computing and Networking (CANDAR),* Matsuyama, Japan, 220-226.

Winter S, 2009, Towards a probabilistic time geography. In: *17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. New York. 528–531.

Winter S and Yin ZC, 2011, The elements of probabilistic time geography. *GeoInformatica*, 15(3), 417-434.

Ye N, Zhang Y, Wang R and Malekian R, 2016, Vehicle trajectory prediction based on Hidden Markov Model. *KSII Transactions on Internet & Information Systems*, 10(7).

Zorbas N, Zissis D, Tserpes K and Anagnostopoulos D, 2015, Predicting object trajectories from high-speed streaming data. In: IEEE, *Trustcom/BigDataSE/ISPA*, Helsinki, Finland, vol. 2, 229-234.