[AM-06-090] Computational Movement Analysis

Abstract

Computational Movement Analysis (CMA) develops and applies analytical computational tools aiming at a better understanding of movement data. CMA copes with the rapidly growing data streams capturing the mobility of people, animals, and things roaming geographic spaces. CMA studies how movement can be represented, modeled, and analyzed in GIS&T. The CMA toolbox includes a wide variety of approaches, ranging from database research, over computational geometry to data mining and visual analytics.

Keywords: conceptual models for movement and movement spaces, movement mining, movement patterns, moving object databases (MOD), time geography, tracking data, trajectory analysis

Author & citation

Laube, P. (2017). Computational Movement Analytics. The Geographic Information Science & Technology Body of Knowledge (2nd Quarter 2017 Edition), John P. Wilson (ed.), DOI: 10.22224/gistbok/2017.2.6

Explanation

1. Definitions
2. Why study movement in GIS&T?
3. Conceptual models of movement and movement spaces
4. Movement data
5. Trajectory operations
6. Context
7. Movement patterns
8. Visual analytics of movement

1. **Definitions**

Computational Movement Analysis (CMA): the interdisciplinary research field studying the development and application of computational techniques for capturing, processing, managing, structuring, and ultimately analyzing data describing movement phenomena, both in geographic and abstract spaces, aiming for a better understanding of the processes governing that movement (Laube, 2014, p. 4)

After Laube (2014, p. 5), CMA investigates the scientific fundamentals related to:

- the specific **characteristics** and peculiarities of the geographic phenomenon of movement and the spatio-temporal data describing it, including data quality (uncertainty, accuracy), scale issues, and spatio-temporal autocorrelation,
- the peculiarities of established and emerging integrated spatial systems serving as direct or indirect **tracking systems** capturing raw or enriched movement data,
- capturing, (pre-)processing, integrating, storing, managing, and querying the rapidly

growing **data** streams describing movement phenomena,
- the **conceptual models** for moving objects and movement processes, and the spaces embedding that movement, the data structures implementing these models, and the implications of models and structures on the CMA process,
- the development and evaluation of **analysis techniques and operations** structuring low-level movement data and deriving high-level process knowledge from that data. This draws on methods from spatio-temporal analysis, geography, computational geometry, scientific visualization, data mining and KDD, and statistics,
- the characteristics and **semantics** of the wide range of current **applications** of CMA, and the assessment of the potential of prospective applications areas, and
- the **societal issues**, including ethics and privacy, as well as issues around user-generated and open data.

2. **Why study movement in GIS&T?**

Understanding how and why people, animals, goods, and more generally things move about in space-time is one of the fundamental questions in geography. Early work on movement analysis in GIS&T was very much data and technology driven, mainly triggered by technological progress in GPS tracking technology (Laube, 2015). Whereas GPS data of a dozen caribou offered previously unseen tracking data two decades ago, today we are not that far away from tracking populations of entire countries via mobile phone data. Movement data is inherently spatio-temporal and hence allows adding true dynamics to the otherwise predominantly static GIS&T environment. Major achievements in CMA have emerged from the field of movement ecology, where biologists strive for a better understanding of animal movement. Other important application areas of CMA include transportation research, surveillance and security, marketing and consumer behavior, and increasingly also sports analytics.

3. **Conceptual models of movement and movement spaces**

Movement always happens in a space enabling and possibly constraining that movement. Be it animals that move through their habitats, cars driving along streets, people commuting through public transit, or shoppers browsing a mall – the characteristics of the spaces containing the movement have a big influence on the movement itself. From a computer science perspective, both the movement and its embedding space must be modeled using conceptual models such that movement analysis becomes possible in a computational environment (Laube, 2014).

Within any GIS&T environment, the classic entity- and field-based conceptual models for representing space lend themselves as basic movement spaces. These two are then complemented with network spaces for street networks and public transit. Often, such 2D-spaces are extended to include a third temporal dimension (x, y and t for time) making reference to Hägerstrand’s time geography. The model for the movement space then rules how the movement itself is modeled, and subsequently what analysis techniques are possible: Whereas, for example, a 2D Euclidean field-space produces trajectories as sequences of x, y and t-tuples, a transportation network rather produces trajectories in the form of sequences of visited network edges or nodes (Andrienko et al., 2008; Laube, 2017).

Movement in the space-time cube, by contrast, can be modelled as space-time paths or lifelines or then space-time prisms representing the potential movement given some constraints (e.g., maximal speed). These three different forms of movement illustrate, how subsequently different algorithmic approaches for analyzing the movement are required. Trajectories can be tackled with time-enabled approaches for line analysis (e.g. polyline similarity measures). Space-time prisms analysis requires rather complex 3D volume intersections. Finally, node sequence data can be analyzed using sequence mining algorithms, akin to association rule mining.

4. **Movement data**

In any CMA project, a substantial fraction of time will be spent on preprocessing, cleaning, and structuring the raw movement data. Given the multitude of systems able to track movement, movement data comes in a plethora of different forms, each with its peculiarities worth studying when aiming for efficient and effective processing in GIS&T and supporting environments.

An increasingly popular characterization of different tracking perspectives is borrowed from physics, the distinction between the Lagrangian and Eulerian observation perspective. The Lagrangian perspective considers changes in a moving object's position, where the moving object produces a sequence of position fixes as it moves across space (e.g. GPS tracking). The Eulerian perspective, by contrast, describes movement as changes in position of moving objects relative to known, fixed locations in space. These check-points can be GSM antennas, traffic gantries, or smart-card readers at train stations. Whereas Lagrangian tracking data is mostly limited to smaller samples, it typically offers finer spatio-temporal granularities and a higher reliability. Eulerian tracking systems, however, promise access to much larger samples, but for the price of coarser spatio-temporal granularity and often much less control over the sampling regime. The two perspectives also correspond to different conceptual models of movement and movement spaces, again resulting in different analytical approaches for the resulting movement data.

Obviously, movement data is spatial, so it comes with all the known issues of spatial data, including questions about reference systems, data quality issues (inaccuracy, uncertainty), scale issues, and autocorrelation. Importantly, movement data is also temporal, still challenging rather static GIS&T with the inclusion of temporal reference systems, temporal relations, temporal granularity or sampling issues, and questions about temporal autocorrelation. Tracking data quality is clearly an issue for CMA in GIS&T, especially when inaccurate positional data is used for computing descriptive parameters such as speed, turning angle or sinuosity. Equally important when deriving such movement attributes is scale, or here the spatio-temporal sampling granularity (Laube & Purves, 2011).

5. **Trajectory operations**

This reference article uses the simple notion of a trajectory as a "time-stamped sequence of visited locations representing a moving object's trace in space-time." This may hence include time-enriched polylines, 3D life lines, network edges sequences, and sequences of check-ins and –outs of public transit smart card system. Irrespective of its peculiar form, a first set of GIS&T operations focuses on the trajectory entities alone; that is shape or arrangement characteristics of the trajectories not yet embedding them in their geographic context (see Section 6, below. Whereas some trajectory operations can be performed in GIS environments, the temporal dimension often requires the use of additional data processing tools (e.g. R packages).

A first crucial task is enriching trajectories with descriptive variables such as speed, step length, acceleration, turning angle, or sinuosity. A second key trajectory operation is segmentation, decomposing the trajectory into self-similar segments (Buchin et al., 2011). Very often a first segmentation will separate moves from stops, that is periods when the object is not moving at all. This may be important with tracking systems that keep recording positions whilst the object is static, potentially producing pseudo-movement resulting from inaccurate measurements rather than actual movement.

Comparing trajectories is a further important trajectory operation. Comparing trajectories is important for grouping or clustering objects that express similar or dissimilar movement. This is another good example of the importance of the modeling process in CMA, as the different conceptual models used for movement will require rather different trajectory similarity measures (Toohey & Duckham, 2015). The catalogue of proposed approaches reaches from geometry-based line similarity measures to time series analysis to edit distance concepts borrowed from string analysis and genetics (Dodge et al., 2012). Once a suitable similarity metric has been found, conventional clustering techniques allow grouping of objects according to their movement properties.

6. **Context**

The GIS&T core strength of integrating spatial variables through a spatial reference system (overlay) is the basis for contextualizing trajectories. The key to a better understanding of movement processes very often lies in putting the trajectories in context to the underlying geography, hence studying relations between the trajectories and the enabling and constraining geographic context (Nathan et al 2008). In the simplest case this may mean adding a land-use category to every observed position of a moving entity. More complex forms of such semantic enrichment of trajectories may involve buffer operations or topological operations. Dodge et al. (2013) present a framework for annotating animal movement data with environmental variables (e.g. weather data) in the widely used Movebank system.

Another challenge for CMA comes in the form of additional sensor data further characterizing mobility behavior (Shamoun-Baranes et al., 2012, Williams et al., 2014). The most prominent such sensor is the accelerometer, nowadays-standard equipment even on mobile phones. Also many GPS trackers are complemented with accelerometers, adding further data dimensions to the mobility monitoring exercise. Combined with speed readings accelerometer data may help qualifying different behaviors when monitoring animals or detecting travel modes when monitoring commuters. Further sensors allowing such semantic enrichment of trajectories include sensors tracking physiological parameters such as heart rate or body temperature, acoustic sensors, or light sensors.

7. **Movement patterns**

The search for salient movement patterns clearly is one of the key promoters of CMA within GIS&T. Movement patterns can refer to frequently used routes by commuters, spatially-explicit manifestations of animal behavior such as hunting excursions of feral animals, or arrangement and coordination patterns relating sets of mobile objects such as flocking, convoys, or leadership (Andrienko et al., 2008; Long & Nelson, 2013).

The GIS&T community has been very successful in adopting a range of tools from related disciplines for this challenging task. First of all, data mining algorithms were deemed to be a natural fit, with borrowings from spatio-temporal association rule mining known from market basket analysis or time series analysis from stock exchange analysis. Then, movement data and its processing also led to significant developments in the database community, with moving object database (MOD) specifically being tailored for managing movement data and querying patterns. Furthermore, the computer science discipline of computational geometry produced a wide range of movement pattern approaches focusing mostly on the geometric aspects of arrangements of points and lines representing movement. Although the GIS&T community has produced some attempts aiming for a generic categorization of movement patterns, the diverse nature of the involved application areas keeps challenging such noble efforts (Dodge, Weibel, & Lautenschütz, 2008).

Detecting movement patterns in essence means structuring data: Be it finding spatio-temporal hot-spots of space-use, clusters of objects expressing similar movement or coordination events in groups of moving objects. The next knowledge discovery step of attaching meaning and process understanding to those same patterns, is much harder. Galton (2005) highlights this difficult analysis task as “bridging the semantic gap.”

8. **Visual analytics of movement**

Visualization and visual analytics is another GIS&T field that has enthusiastically adopted the challenge of CMA. With its natural affinity to large noisy spatio-temporal datasets, information visualization and exploratory data analysis are a natural fit for movement data (Andrienko & Andrienko, 2012; Demšar et al., 2015). The oft-cited ideal match between the human ability to perceive patterns and trends with the computer’s power to process and present large datasets also works well for CMA – adhering to Shneiderman’s mantra “overview first, zoom and filter, then details-on-demand.” Linked views and animation are especially prominent elements in the many visualization settings put forward for movement data. Typical views include maps, 3D space-time spaces, and parameter spaces typically depicting time series of derived movement parameters such as speed or acceleration. Also, the interactive visualization of previously mined movement patterns in map views – hence the combination of data mining and visualization promoted as visual analytics – works well for exploring large streams of movement data.

References

Andrienko, N., & Andrienko, G. (2012). Visual analytics of movement: An overview of methods, tools and procedures. Information Visualization. 12(1):3-24.

Andrienko, N., Andrienko, G., Pelekis, N., & Spaccapietra, S. (2008). Basic concepts of movement data. In Mobility, data mining and privacy (pp. 15-38). Springer Berlin Heidelberg.

Buchin, K., Buchin, M., Van Kreveld, M., & Luo, J. (2011). Finding long and similar parts of trajectories. Computational Geometry, 44(9), 465-476.

Demšar, U., Buchin, K., Cagnacci, F., Safi, K., Speckmann, B., Van de Weghe, N., ... & Weibel, R. (2015). Analysis and visualisation of movement: an interdisciplinary review. Movement Ecology, 3(1), 1-24.

Dodge, S., Bohrer, G., Weinzierl, R., Davidson, S. C., Kays, R., Douglas, D., ... & Wikelski, M. (2013). The environmental-data automated track annotation (Env-DATA) system: linking animal tracks with environmental data. Movement Ecology, 1(1), 3. DOI: 10.1186/2051-3933-1-3

Dodge, S., Laube, P., & Weibel, R. (2012). Movement similarity assessment using symbolic representation of trajectories. International Journal of Geographical Information Science, 26(9), 1563-1588.

Dodge, S., Weibel, R., & Lautenschütz, A. K. (2008). Towards a Taxonomy of Movement Patterns. Information Visualization, 7(3-4), 240-252.

Galton, A. (2005). Dynamic collectives and their collective dynamics. In International Conference on Spatial Information Theory (pp. 300-315). Springer Berlin Heidelberg.

Laube, P. (2014). Computational Movement Analysis. Springer Verlag, Berlin.

Laube, P. (2015). The low hanging fruit is gone: achievements and challenges of computational movement analysis. SIGSPATIAL Special, 7(1), 3-10.

Laube, P. (2018). Representation, Trajectories. In: D. Richardson, N. Castree, M. Goodchild, W. Liu, A. Kobayashi, & R. Marston (Eds.) The International Encyclopedia of Geography: People, the Earth, Environment, and Technology. Wiley/AAG, Hoboken, NJ.

Laube, P., & Purves, R. S. (2011). How fast is a cow? Cross-Scale Analysis of Movement Data. Transactions in GIS, 15(3), 401-418.

Long, J. A., & Nelson, T. A. (2013). A review of quantitative methods for movement data. International Journal of Geographical Information Science, 27(2), 292-318.

Nathan, R., Getz, W. M., Revilla, E., Holyoak, M., Kadmon, R., Saltz, D., & Smouse, P. E. (2008). A movement ecology paradigm for unifying organismal movement research. Proceedings of the National Academy of Sciences, 105(49), 19052-19059.

Shamoun-Baranes, J., Bom, R., van Loon, E. E., Ens, B. J., Oosterbeek, K., & Bouten, W. (2012). From sensor data to animal behaviour: an oystercatcher example. PloS one, 7(5), e37997.

[Toohey, K., & Duckham, M. (2015). Trajectory similarity measures. SIGSPATIAL Special, 7(1), 43-50.]

[Williams, T. M., Wolfe, L., Davis, T., Kendall, T., Richter, B., Wang, Y., ... & Wilmers, C. C. (2014). Instantaneous energetics of puma kills reveal advantage of felid sneak attacks. Science, 346(6205), 81-85.]