Künstl Intell (2012) 26:215–221
DOI 10.1007/s13218-012-0215-2

ÜBERSICHTSBEITRAG

Spatiotemporal Modeling and Analysis—Introduction and Overview

Christine Körner · Michael May · Stefan Wrobel

Received: 8 June 2012 / Accepted: 12 June 2012 / Published online: 26 June 2012

Abstract Over the past five to seven years the analysis of trajectory data has established itself as an independent research discipline within the area of data mining. In this article we provide an overview on data characteristics, state-of-the-art preprocessing and analysis methods of trajectory data. We conclude the article with a collection of challenges that arise due to the increasing variety of spatiotemporal data sources and which have to be solved for the application of spatiotemporal analysis methods in practice.

Keywords Spatiotemporal data · Mobility mining

1 Introduction

Nearly every object, event or phenomenon can be referenced in space and time. The analysis of spatiotemporal data is therefore of interest in many industry and service sectors, e.g. traffic management, agriculture, disaster management, location planning or location based services. Consequently, the demand for analysis and modeling techniques has increased in parallel with the volume of collected data over the past years. Spatiotemporal data has many different forms and includes data such as georeferenced time series, remote-sensing images or moving object trajectories. In this overview we will focus on trajectory data, i.e. data containing the movement history of mobile objects. The analysis of trajectory data is an interdisciplinary research field and involves communities from geographic information science (GIScience), database technology, data mining, visual analytics, sensor networks, distributed systems, transportation science as well as privacy. In the data mining community, the analysis of trajectory data is often referred to as *trajectory* or *mobility mining*.

Mobility mining is still a young research discipline and has emerged from the field of spatial data mining about five to seven years ago. Its analysis techniques differ from classic data mining techniques because of data characteristics and the complexity of the data. One principal difference between spatial and non-spatial data is *autocorrelation*, which means that objects with similar characteristics are typically clustered in space. This characteristic contrasts with the often made assumption of independent, identically distributed data samples in classic data mining and results in inaccurate or inconsistent models if autocorrelation is ignored [45]. In addition, spatial characteristics do not necessarily spread evenly in space. For example, natural borders such as rivers and mountains influence the variation of spatial phenomena. For movement behavior the street network has an even higher significance because it canalizes mobility. Besides these spatial dependencies, movement behavior is influenced by physiological and social circumstances. For example, the maximum walking speed and available means of transportation restrict the daily range of operation. The habit of returning to a home base for sleep and fixed work schedules are further constraints that impose a strong regularity on human movement behavior. A first classification of such constraints has been given by Hägerstrand [20] in the area of time geography in 1970.

C. Körner (✉) · M. May · S. Wrobel
Fraunhofer IAIS, Sankt Augustin, Germany
e-mail: christine.koerner@iais.fraunhofer.de

M. May
e-mail: michael.may@iais.fraunhofer.de

S. Wrobel
e-mail: stefan.wrobel@iais.fraunhofer.de

S. Wrobel
University of Bonn, Bonn, Germany

For the development of algorithms and analysis methods for spatiotemporal data this means that it is only a first step to adapt existing data mining algorithms to a new data structure. The second, more challenging step is to incorporate expert and background knowledge into the discovery process. Especially the analysis of trajectory data requires background information because movement is seldom an end in itself. Instead, it is the means to achieve some goal at a given place. The quantitative and qualitative growth of georeferenced data has facilitated the incorporation of background knowledge over the past years. However, it has also given rise to a number of future research challenges.

In the following we will give an introduction to knowledge discovery from trajectory data. Section 2 introduces trajectory data sources and data structures. Section 3 describes preprocessing and feature extraction methods and Section 4 provides an overview on analysis tasks and trajectory data mining methods. We conclude the paper with a section on research challenges of mobility mining, which are decisive for its application in practice.

2 Trajectory Data Sources and Data Structures

The wide availability of positioning technology as well as advances in trajectory data analysis have opened new ways for the collection of movement information. While past mobility studies recorded personal movement data primarily through an active involvement of the user using questionaries or interviews, technologies such as the Global Positioning System (GPS), Global System for Mobile Communications (GSM), Bluetooth or Radio Frequency Identification (RFID) allow the passive collection of movement information.

Depending on the applied positioning technology, trajectory data differs in its spatial and temporal resolution. For example, GPS records have a very high spatial resolution while their temporal resolution typically depends on the battery power of the employed devices. Positions may be recorded as often as every second or in intervals of a few minutes or even hours (e.g. when monitoring wild animals). GSM data, in contrast, possess a coarse spatial resolution, which ranges between a few hundred meters and several kilometers depending on the structure of the cellular network. Typically, GSM data is provided in form of call detail records (CDR), which are used for billing purposes. In this type of data, the radio cell of a user is only known during a call activity. Thus, the position records are of sporadic nature. The advantage of GSM technology in comparison to GPS lies in its high penetration within the population, which allows to form large, long-term data samples.

Several data models exist in order to store trajectory data and to analyze it in a trajectory database or a trajectory data warehouse. All models have to decide between a discrete and a continuous temporal representation of movement. As positioning techniques are bound to record position data at discrete points in time, it is natural to choose a discrete representation of trajectories, too. Typically such a trajectory in 2-dimensional geographic coordinate space is represented simply as a sequence of tuples

$$\big((x_1, y_1, t_1), \ldots, (x_n, y_n, t_n)\big)$$

with (x_i, y_i) denoting coordinates in a given geographic reference system and t_i denoting points in time with $t_1 < \cdots < t_n$ $\forall i \in 1..n$. A more advanced model is the moving objects data model [16, 19]. The model was designed to express continuous changes of objects over time, both in their position and extent. It is thus able to model moving point objects as well as moving regions. The model has been implemented in the database engine SECONDO [19]. A second database engine, called HERMES, has been designed specifically for the support of trajectory data analysis [39]. It provides a rich set of operations and distance functions and possesses extensions for privacy preserving data analysis.

3 Trajectory Preprocessing and Feature Extraction

Preprocessing and feature extraction require a significant amount of attention and time in spatiotemporal data analysis because of noisy position records and time consuming spatiotemporal operations. In spatial and spatiotemporal data mining two approaches of feature extraction are possible. It can be done either previous to the mining step or it can be incorporated into it. The former approach is advantageous if many different mining algorithms are tested on the data. In addition, it allows to apply non-spatial data mining methods. The latter approach allows to exploit spatial features dynamically during the discovery process. It has the advantage that only those regions of the search space need to be explored that are likely to contain interesting hypotheses.

In the database literature trajectory preprocessing is also referred to as *trajectory construction*. It comprises the steps data cleaning, data compression and data segmentation [51]. During data cleaning measurement noise and outliers are removed. Data compression reduces the amount of trajectory sample points because frequent position records easily lead to large volumes of data. A recent evaluation of compression techniques is given in [34]. Data segmentation is the division of a trajectory into meaningful sub-trajectories, which are required for the extraction of stops and moves and subsequently for trajectory annotation. Different strategies for trajectory segmentation exist, including the detection of stops [31, 46], the detection of sequences with homogeneous movement characteristics [10] or the detection of representative sub-trajectories in a trajectory database [37].

For trajectory data analysis characteristics of a single trajectory as well as the relationship between two or more trajectories are important. Andrienko et al. [5] group features of a single trajectory into two basic categories: *moment-related* and *overall* characteristics. Moment-related characteristics can be extracted for each point in time whereas overall characteristics rely on a trajectory interval. Examples of moment-related characteristics of a trajectory are a moment's spatial and temporal reference, direction or speed of movement. Examples of overall characteristics of a trajectory are the trajectory's geometry, length and duration or its minimum, average and maximum speed [5].

In order to specify the relationship between two or more trajectories, distance functions and topological relations can be used. Pelekis et al. [38] make a basic distinction between distance functions relying on spatiotemporal characteristics or on spatial characteristics only. In the case of spatiotemporal characteristics a small distance is assumed if mobile entities follow similar routes concurrently whereas in the case of spatial characteristics only the similarity of routes is decisive. In addition, derived characteristics of a trajectory such as speed and direction can be considered in the distance function. Concrete examples of distance functions are given in [4, 38, 40].

In order to define topological relations between trajectories again their relation in space and in time can be considered. The predominant formal model to describe topological relations between two spatial objects has been developed by Egenhofer [15] and is called the 9-intersection model. For time intervals Allen [1] has defined seven temporal relations. In order to express spatiotemporal topological relations, both relations can be combined [12, 13].

4 Analysis Tasks and Methods

In this section we give an introduction to the most prominent analysis tasks and data mining algorithms for mobile entities, namely clustering, pattern analysis, location prediction and trajectory annotation. A comprehensive overview on the topic can be found in [36].

4.1 Clustering

The clustering of trajectories, i.e. the segmentation of trajectories into groups with similar movement characteristics and determination of group representatives, generally takes place on a set of trajectory sections rather than on the lifelong trajectories of entities. State-of-the-art clustering techniques for trajectories rely on traditional clustering algorithms and put their main effort into the definition of meaningful similarity (distance) functions as described in Section 3. Nanni and Pedreschi [35] recommend the usage of density-based algorithms for trajectory clustering as, for example, the OPTICS algorithm. A common approach for the clustering of trajectory data is also the stepwise, visually aided application of clustering algorithms. The gradual refinement of clusters has the advantage that it breaks down complexity with respect to comprehensibility as well as to computational resources [6, 40].

4.2 Pattern Analysis

Trajectory patterns describe interesting behaviors of groups of moving objects. Hereby, two tasks are considered in the literature: the detection of frequent movement patterns and the detection of pattern occurrences. In the first case the goal is to identify the pattern itself, for example, a frequent movement from location A to location B to location C. In the second case the goal is to identify when and where a specific pattern occurs and which entities participate in it, for example, the convergence of a group of entities to some location. In the following we will discuss both data mining tasks in more detail.

Mining frequent trajectory patterns is the task of extracting (parts of) routes that are frequently followed by the objects of interest. Frequent trajectory patterns can hereby be defined using spatial or spatiotemporal characteristics of the trajectories. In the first case only the sequence of the visited locations is considered. In the second case the transition times between the locations are also important. In order to detect frequent spatiotemporal patterns, Giannotti et al. defined the concept of *temporally annotated sequences* (TAS) [17] and later generalize it to *trajectory patterns* (also called T-patterns) [18]. A TAS is a sequence of items along with a sequence of transition times (i.e. the temporal annotations) between the items. The items hereby represent geographic locations. In T-patterns the items are substituted with pairs of coordinates in two-dimensional geographic coordinate space. In order to specify the containment of a T-pattern in a trajectory the authors define a neighborhood function. One challenge of trajectory pattern mining is the handling of continuous geographic coordinate space. Clearly, two persons that travel along a street will not yield trajectories with the same coordinates. Therefore trajectories are typically discretized previous to pattern mining [18, 23].

The detection of pattern occurrences naturally requires a specification of the pattern to be detected. The most commonly used patterns for this task are *group patterns*. Group patterns refer to objects that conform to a specified collective behavior and may involve derived information concerning the whole group of objects (e.g. average speed). Intuitively, a group is formed by a number of objects that stay close in space for a meaningful period of time. Algorithms that detect such groups of objects are provided by Wang et al. [50] for regularly spaced trajectories and by Hwang et al. [22] for irregularly spaced trajectories assuming linear movement. In addition to the general definition of spatiotemporal closeness, a group can be specified by some characteristic internal structure. For example, a group could be headed by some individual who anticipates the group motion. This pattern is called *leadership* and was introduced by [26] under the general concept of *relative motion* (REMO). Other basic spatiotemporal group patterns of REMO are *flock*, *convergence* and *divergence*. Algorithms for the efficient computation of REMO patterns and their extensions are provided, for example, in [3, 8, 49].

4.3 Location Prediction

The prediction of future locations of moving objects has been of interest in mainly two research areas, namely moving object database systems and wireless communication networks. Moving object databases employ location prediction, for example, in range queries [41, 48] or nearest neighbor queries [7, 47] over the future positions of moving objects. In wireless communication networks location prediction serves mainly resource allocation and cell paging, which requires to anticipate the motion of users for the near future. Several algorithms have been investigated to accomplish this task using neural networks [30] or Gauss-Markov models [28]. More recently, further methods have been developed for the prediction of the most likely route [25] or of destinations [27, 32, 53]. The underlying assumption hereby is that people follow daily or weekly routines and have only a few frequently visited locations, which is one of the main characteristics of human movement behavior.

4.4 Trajectory Annotation

Trajectory annotation means to lift a trajectory from its representation in physical space to a semantic space. A semantic trajectory has the advantage that it contains information about *why* and *how* people move. Trajectories can be annotated with different types of semantic information, among them are stop locations and activities [2, 29, 54] or means of transportation [43]. An automated framework for trajectory annotation has recently been proposed by Yan et al. [51]. In addition, the work in [52] is a first step toward online segmentation and annotation of trajectories.

5 Challenges in Spatiotemporal Data Analysis

The analysis of spatiotemporal data has to face a number of challenges in the next years. Besides known issues related to privacy and missing data, the tremendous increase in geographic and time referenced data sets challenges the development of new algorithms. On the one hand, big data analysis requires scalable and distributed data mining algorithms as well as incremental algorithms because more and more data sets are becoming available in near real-time. On the other hand, the variety of data sources and the complexity of spatiotemporal data makes it a major research task to work on data harmonization and integration.

From our applied projects with the German and Swiss outdoor advertising industry, we have experienced that missing data is a serious problem in GPS mobility surveys. Test persons easily forget to carry the device or to charge its battery, which leads to a significant amount of missing measurement days. Therefore, the removal of test persons with missing data is not an option. If missing data is ignored, i.e. missing measurement periods are treated as immobility, mobility quantities derived from the data will be underestimated. Therefore, analysis and modeling techniques have to be advanced that are either able to impute missing values or to handle missing values methodically. Special care has thereby to be taken if the absence of data relates to the mobility behavior of a person because it induces a bias into the data sample [24].

Privacy is a second, serious challenge for spatiotemporal data analysis because movement trajectories are inherently connected to a person's places of interest, including home and work location as well as attended medical or religious services. It is therefore not sufficient to simply remove personal identifiers from trajectory data [9]. Some approaches already exist that apply the concept of *k-anonymity* [33, 42] or of *differential privacy* [11, 14]. However, it remains a challenge to ensure privacy with respect to the large amount of available background information that may be linked to the data in question. Further, the intensity of anonymization always needs to be balanced against the remaining data utility.

Finally, big data is certainly a challenge in the spatiotemporal domain because nearly any event or process can be referenced in space and time. Next to the wide availability of monitoring technologies, the realization of local authorities and business companies about the value of their data collections (e.g. traffic counts, call detail records) and their quest to make use of it have contributed to the growth of data. This means, on the one hand, to master many structurally and semantically different data sets. On the other hand, this requires to develop scalable methods that can handle the amounts of data in the requested time. The diversity among data sets can mainly be attributed to characteristics as their representativeness, sampling rate and resolution, applying each to the spatial, temporal or population dimension of a data set. In the following we will take a closer look into the various dimensions and problems that can arise, which we have collected from our experience in practice over the past years.

Spatial dimension. The spatial representativeness of a data set can be biased if measurements are taken at specific locations only. For example, traffic counts are typically taken at major roads while the majority of streets are minor roads. Furthermore, data sets may be available only for specific regions or cities. As mobile behavior (e.g. average daily travel distance or use of public transportation) depends on local structures (e.g. size of city), an extrapolation of mobility characteristics to other regions is not straightforward. In addition, the density of measurements influences the choice of analysis methods and evaluations. Clearly, dense GPS data will allow to estimate travel distance and even requires to remove points within stop locations for distance estimation while CDR data will underestimate this quantity [44]. Finally, the spatial resolution of data records can vary between a few meters (GPS), a few tens or hundreds of meters (Bluetooth, GSM) or even kilometers (GSM). In order to use data with a coarse spatial resolution on a detailed geographic level of scale, appropriate disaggregation methods have to be developed as e.g. in [21].

Temporal dimension. The temporal representativeness of trajectory data is tightly coupled to the sampling mechanism. For example, positions can be recorded in regular time intervals (with high or low frequency) or in dependence of certain events as, for example, call activities in the case CDR data. In the latter case representativeness in time as well as space is not guaranteed. In addition, the observation period is important because traffic differs over the day, between weekdays and weekends as well as between working periods and vacation times. Two data sets taken in different time periods may therefore require a scaling factor in order to be comparable. Scaling factors about the temporal distribution of traffic are also required to adapt the temporal resolution of data sets, for example, to convert daily traffic counts to the level of hours.

Population dimension. The sample size of spatiotemporal data sources has immensely gained by the availability of big data. However, population representativeness is a major issue because data sets are primarily collected for other purposes. Data sets collected through companies are typically limited to the range of customers, and over-represent specific sociodemographic groups in the population depending on the business strategy. The correction of such a bias is often aggravated by the fact that sociodemographic information cannot be shared due to privacy restrictions. A similar problem is encountered in Bluetooth data, because Bluetooth enabled devices such as smart phones are not evenly distributed in the population. However, as long as data representativity cannot be ensured, the value of mined patterns is questionable. When analyzing data of a regional population (e.g. the trajectory data is collected from inhabitants of a given city) it is furthermore important to consider effects of externals as, for example, commuters or tourists. The resolution of trajectory data sets with respect to the population is typically on the level of individuals. However, changing identifiers over time due to privacy reasons can complicate analysis as well as the usage of multiple tracking devices per person (e.g. mobile phones).

In summary, the availability of diverse movement data sources makes it an important task to develop intelligent analysis methods not only for the specific task at hand but also for the identification and handling of mechanisms that disturb data representativeness. Only then will it be possible to benefit from the availability of various data sources and to remedy the weakness (e.g. in sampling rate or granularity) of one data set with the strength of another.

References

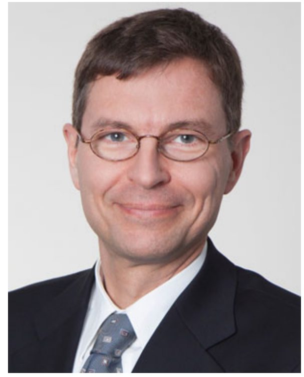
1. Allen JF (1984) Towards a general theory of action and time. Artif Intell 23(2):123–154
2. Alvares LO, Bogorny V, Kuijpers B, de Macedo JAF, Moelans B, Vaisman A (2007) A model for enriching trajectories with semantic geographical information. In: Proc of the 15th annual ACM international symposium on advances in geographic information systems (GIS'07). ACM, New York, pp 1–8
3. Andersson M, Gudmundsson J, Laube P, Wolle T (2008) Reporting leaders and followers among trajectories of moving point objects. Geoinformatica 12(4):497–528
4. Andrienko G, Andrienko N, Wrobel S (2007) Visual analytics tools for analysis of movement data. SIGKDD Explor Newsl 9(2):38–46
5. Andrienko N, Andrienko G, Pelekis N, Spaccapietra S (2008) Basic concepts of movement data. In: Giannotti F, Pedreschi D (eds) Mobility, data mining and privacy. Springer, Berlin, Chap 1
6. Andrienko G, Andrienko N, Rinzivillo S, Nanni M, Pedreschi D, Giannotti F (2009) Interactive visual clustering of large collections of trajectories. In: Proc of the IEEE symposium on visual analytics science and technology (VAST'09). IEEE, New York, pp 3–10
7. Benetis R, Jensen CS, Karciauskas G, Saltenis S (2006) Nearest and reverse nearest neighbor queries for moving objects. VLDB J 15(3):229–249
8. Benkert M, Gudmundsson J, Hübner F, Wolle T (2008) Reporting flock patterns. Comput Geom 41(3):111–125
9. Bonchi F, Lakshmanan LV, Wang H (2011) Trajectory anonymity in publishing personal mobility data. SIGKDD Explor Newsl 13(1):30–42
10. Buchin M, Driemel A, van Kreveld M, Sacristán V (2010) An algorithmic framework for segmenting trajectories based on spatio-temporal criteria. In: Proc of the 18th SIGSPATIAL international conference on advances in geographic information systems (ACM GIS'10). ACM, New York, pp 202–211
11. Chen R, Fung BCM, Desai BC (2011) Differentially private trajectory data publication. CoRR abs/1112.2020
12. Claramunt C, Jiang B (2000) A representation of relationships in temporal spaces. In: Innovations in GIS VII: geocomputation. Taylor & Francis, London, pp 41–53
13. Claramunt C, Jiang B (2001) An integrated representation of spatial and temporal relationships between evolving regions. Geogr Syst 3:411–428
14. Dwork C (2006) Differential privacy. In: Proc of the 33rd international colloquium on automata, languages and programming (ICALP'06). Lecture notes in computer science. Springer, Berlin, pp 1–12
15. Egenhofer MJ (1991) Reasoning about binary topological relations. In: Günther O, Schek HJ (eds) Proc of the 2nd international symposium on advances in spatial databases (SSD). Springer, Berlin, pp 143–160

16. Forlizzi L, Güting RH, Nardelli E, Schneider M (2000) A data model and data structures for moving objects databases. In: Proc of the 2000 ACM SIGMOD international conference on management of data (SIGMOD'00). ACM, New York, pp 319–330
17. Giannotti F, Nanni M, Pedreschi D (2006) Efficient mining of temporally annotated sequences. In: Proc of the 6th SIAM international conference on data mining (SDM'06). SIAM, Philadelphia, pp 346–357
18. Giannotti F, Nanni M, Pinelli F, Pedreschi D (2007) Trajectory pattern mining. In: Proc of the 13th ACM SIGKDD international conference on knowledge discovery and data mining (KDD'07). ACM, New York, pp 330–339
19. Güting RH, Schneider M (2005) Moving objects databases. Morgan Kaufmann, San Mateo
20. Hägerstrand T (1970) What about people in regional science? Pap Reg Sci Assoc 24:7–21
21. Hecker D, Körner C, Stange H, Schulz D, May M (2011) Modeling micro-movement variability in mobility studies. In: Geertman S, Reinhardt W, Toppen F (eds) Advancing geoinformation science for a changing world. Lecture notes in geoinformation and cartography. Springer, Berlin, pp 121–140
22. Hwang SY, Liu YH, Chiu JK, Lim EP (2005) Mining mobile group patterns: a trajectory-based approach. In: Proc of the 9th Pacific-Asia conference on advances in knowledge discovery and data mining (PAKDD'05). Lecture notes in computer science, vol 3518. Springer, Berlin, pp 713–718
23. Kang J, Yong HS (2010) Mining spatio-temporal patterns in trajectory data. J Inf Process Syst 6(4):521–536
24. Körner C (2012) Modeling visit potential of geographic locations based on mobility data. PhD thesis, University of Bonn
25. Laasonen K (2005) Clustering and prediction of mobile user routes from cellular data. In: Proc of 9th European conference on principles and practice of knowledge discovery in databases (PKDD'05). Springer, Berlin, pp 569–576
26. Laube P, Imfeld S (2002) Analyzing relative motion within groups of trackable moving point objects. In: Proc of the 2nd international conference on geographic information science (GIScience'02). Springer, London, pp 132–144
27. Lei PR, Shen TJ, Peng WC, Su IJ (2011) Exploring spatial-temporal trajectory model for location prediction. In: Proceedings of the 2011 IEEE 12th international conference on mobile data management (MDM'11). IEEE Computer Society, Washington, pp 58–67
28. Liang B, Haas ZJ (2003) Predictive distance-based mobility management for multidimensional PCS networks. IEEE/ACM Trans Netw 11(5):718–732
29. Liao L, Fox D, Kautz H (2007) Extracting places and activities from GPS traces using hierarchical conditional random fields. Int J Robot Res 26(1):119–134
30. Liou SC, Huang YM (2005) Trajectory predictions in mobile networks. Int J Inf Technol 11(11):109–122
31. Marketos G, Frentzos E, Ntoutsi I, Pelekis N, Raffaetà A, Theodoridis Y (2008) Building real-world trajectory warehouses. In: Proc of the seventh ACM international workshop on data engineering for wireless and mobile access (MobiDE'08). ACM, New York, pp 8–15
32. Monreale A, Pinelli F, Trasarti R, Giannotti F (2009) WhereNext: a location predictor on trajectory pattern mining. In: Proc of the 15th ACM SIGKDD international conference on knowledge discovery and data mining (KDD'09). ACM, New York, pp 637–646
33. Monreale A, Andrienko G, Andrienko N, Giannotti F, Pedreschi D, Rinzivillo S, Wrobel S (2010) Movement data anonymity through generalization. Trans Data Priv 3(2):91–121
34. Muckell J, Hwang JH, Lawson CT, Ravi SS (2010) Algorithms for compressing GPS trajectory data: an empirical evaluation. In: Proc of the 18th SIGSPATIAL international conference on advances in geographic information systems (ACM GIS'10). ACM, New York, pp 402–405
35. Nanni M, Pedreschi D (2006) Time-focused clustering of trajectories of moving objects. J Intell Inf Syst 27(3):267–289
36. Nanni M, Kuijpers B, Körner C, May M, Pedreschi D (2008) Spatiotemporal data mining. In: Giannotti F, Pedreschi D (eds) Mobility, data mining and privacy. Springer, Berlin, Chap 10
37. Panagiotakis C, Pelekis N, Kopanakis I, Ramasso E, Theodoridis Y (2011) Segmentation and sampling of moving object trajectories based on representativeness. IEEE Trans Knowl Data Eng 24:1328–1343
38. Pelekis N, Andrienko G, Andrienko N, Kopanakis I, Marketos G, Theodoridis Y (2011) Visually exploring movement data via similarity-based analysis. J Intel Inf Syst 1–49
39. Pelekis N, Frentzos E, Giatrakos N, Theodoridis Y (2011) HERMES: a trajectory db engine for mobility-centric applications. Int J Knowl-Based Organ, in press
40. Rinzivillo S, Pedreschi D, Nanni M, Giannotti F, Andrienko N, Andrienko G (2008) Visually driven analysis of movement data by progressive clustering. Inf Vis 7(3):225–239
41. Saltenis S, Jensen CS, Leutenegger ST, Lopez MA (2000) Indexing the positions of continuously moving objects. In: Proc of the 2000 ACM SIGMOD international conference on management of data (SIGMOD'00). ACM, New York, pp 331–342
42. Samarati P (2001) Protecting respondents' identities in microdata release. IEEE Trans Knowl Data Eng 13:1010–1027
43. Schuessler N, Axhausen KW (2009) Processing raw data from global positioning systems without additional information. Transp Res Rec 2105:28–36
44. Schulz D, Bothe S, Körner C (2012) Human mobility from GSM data—a valid alternative to GPS? In: Proc of the mobile data challenge workshop
45. Shekhar S, Raju VR, Celik M (2009) Spatial and spatio-temporal data mining: recent advances. In: Kargupta H, Han J, Yu P, Motwani R, Kumar V (eds) Next generation of data mining. Chapman & Hall/CRC, London, Chap 26
46. Stopher PR (2009) Collecting and processing data from mobile technologies. In: Transport survey methods—keeping up with a changed world. Emerald Group Publishing Limited, Bingley, Chap 21
47. Tao Y, Papadias D (2002) Time-parameterized queries in spatio-temporal databases. In: Proc of the 2002 ACM SIGMOD international conference on management of data (SIGMOD'02). ACM, New York, pp 334–345
48. Tao Y, Papadias D, Sun J (2003) The TPR*-tree: an optimized spatio-temporal access method for predictive queries. In: Proc of 29th international conference on very large data bases (VLDB'03). Morgan Kaufmann, San Mateo, pp 790–801
49. Wachowicz M, Ong R, Renso C, Nanni M (2011) Finding moving flock patterns among pedestrians through collective coherence. Int J Geogr Inf Sci 25(11):1849–1864
50. Wang Y, Lim EP, Hwang SY (2003) On mining group patterns of mobile users. In: Proc of the 14th international conference on database and expert systems applications (DEXA'03). Springer, Berlin, pp 287–296
51. Yan Z, Chakraborty D, Parent C, Spaccapietra S, Aberer K (2011) SeMiTri: a framework for semantic annotation of heterogeneous trajectories. In: Proc of the 14th international conference on extending database technology (EDBT'11). ACM, New York, pp 259–270
52. Yan Z, Giatrakos N, Katsikaros V, Pelekis N, Theodoridis Y (2011) SeTraStream: semantic-aware trajectory construction over streaming movement data. In: Proc of the 12th international symposium on advances in spatial and temporal databases (SSTD'11), pp 367–385

53. Ying JJC, Lee WC, Weng TC, Tseng VS (2011) Semantic trajectory mining for location prediction. In: Proc of the 19th ACM SIGSPATIAL international conference on advances in geographic information systems (ACM GIS'11). ACM, New York, pp 34–43
54. Zhou C, Frankowski D, Ludford P, Shekhar S, Terveen L (2007) Discovering personally meaningful places: an interactive clustering approach. ACM Trans Inf Syst 25(3)

Christine Körner is research scientist at the Fraunhofer Institute for Intelligent Analysis and Information Systems (IAIS). She studied Business Informatics at the University of Leipzig and Jönköping International Business School, Sweden. She received her Ph.D. degree in Computer Science from the University of Bonn in 2011. Her interest lies in the analysis of spatiotemporal data and mobility mining.

Michael May is heading the Knowledge Discovery department at the Fraunhofer Institute for Intelligent Analysis and Information Systems (IAIS) since 1999. His main research interest is in Spatial Data Mining. He studied Philosophy and Computer Science at the University of Hamburg and obtained his Ph.D. 1997 at the DFG doctoral program in Cognitive Science.



Stefan Wrobel is director of the Fraunhofer Institute for Intelligent Analysis and Information Systems (IAIS) and Professor of Computer Science at the University of Bonn. He studied Computer Science in Bonn and Atlanta, GA, USA (M.S. degree, Georgia Institute of Technology), receiving his doctorate from the University of Dortmund. He is a founding member of the International Machine Learning Society and an elected member of its board. Professor Wrobel is speaker of the German Computer Society's special interest group on Knowledge Discovery and Machine Learning.