

Raumzeitliches Data Mining

Patrick LAUBE

Winterthurerstrasse 193, CH-8057 Zürich, Schweiz

E-Mail: patrick.laube@geo.uzh.ch

Abstract: Data Mining schürft Wissen aus grossen Datenbeständen. Data Mining sucht computergestützt und algorithmisch nach Mustern, Regeln und Beziehungen, die allgemein gültig, neu aber nicht trivial, sowie nützlich und verständlich sind. Raumzeitliches Data Mining ist eine Spezialisierung von Data Mining und sucht entsprechend nach Mustern in Objekten, Ereignissen und Prozessen in Raumzeit. Dieser Artikel gibt eine Einführung in Data Mining und diskutiert typische Muster und Beziehungen, welche durch raumzeitliches Data Mining in Geodatenbeständen gefunden werden können.

1 Data Mining – Aus Daten Wissen schaffen

Während die Geoinformatik in einer Zeit der Datenarmut und geringer Rechenleistung entstand, ist die Disziplin heute in einer datenreichen Gegenwart mit fast unbeschränkter Rechenleistung angekommen (Miller und Han 2009). Im Zuge von Google Earth und Openstreetmap werden die Schnittstellen zwischen den noch vor kurzem proprietär behüteten Geoinformatik-Systemen immer durchlässiger und erlauben zunehmend den plattform-übergreifenden Zugriff auf immer umfangreichere Geodatenbestände – „ad-hoc“ und „on the fly“. Stetig wachsende, zunehmend heterogene und wenig strukturierte Geodatenbestände verlangen nach Alternativen zur klassischen Raumanalyse in Geographischen Informationssystemen (GIS). Raumzeitliches Data Mining ist eine Alternative und bietet das ideale Werkzeug zum rechnergestützten Wühlen im stetig wachsenden Geodatenberg.

Data Mining „schürft“ Wissen aus grossen Datenbeständen. Der Begriff bezeichnet den zentralen Analyseschritt im übergeordneten Prozess des *Knowledge Discovery in Databases* (KDD), welcher den Inhalt grosser Datenbanken zu Information und letztlich nützlichem Wissen destillieren soll (Hand, Manilla und Smyth 2001, Fayyad, Piatetsky-Shapiro und Smyth 1996). Während Data Mining im engeren Sinne die eigentliche Analyse bezeichnet, schliesst KDD im weiteren Sinne auch die Datenvorverarbeitung (Auswahl,

Integration, Normalisierung, Filterung) und die Nachverarbeitung (Interpretation und Evaluation) mit ein (Abbildung 1). Die Abgrenzung zwischen diesen beiden Begriffen ist allerdings unscharf, oft werden sie sogar synonym verwendet.

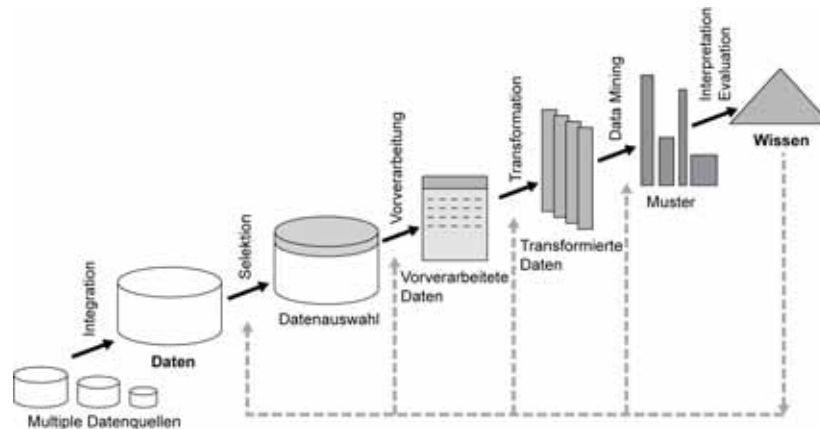


Abbildung 1: Data Mining als zentraler Schritt von *Knowledge Discovery in Databases*, KDD (Quelle: Fayyad *et al.* 1996, verändert).

Data Mining bezeichnet die Analyse grosser Datenmengen, mit dem Ziel unerwartete Beziehungen zu finden, und die Daten für den Datennutzer auf nützliche und verständliche Art zusammenzufassen (Hand *et al.* 2001). Zu den typischen Data Mining Verfahren zählen Clusteranalyse, Klassifikation, Regression und Mustererkennung. Resultate des Data Mining Prozesses sind Muster, Regeln und Beziehungen, die allgemein gültig, neu aber nicht trivial, sowie nützlich und verständlich sein sollen (Runkler 2010). Nur mit Hilfe von Anwendungsexperten kann entschieden werden, welche Resultate diese Kriterien erfüllen und damit sinnvolle Ergebnisse des iterativen Data Mining Prozesses sind.

Data Mining kommt immer dort zum Einsatz, wo aus grossen Datenmengen in zentralisierten oder verteilten Datenbeständen Mehrwert geschaffen werden soll, z.B. im Marketing (Kundenprofile und Warenkorbanalyse im Handel und im Bankenwesen), in der Bioinformatik (Genetik, Molekularbiologie), im Gesundheitswesen (Patientendaten, Versicherungen), sowie bei vielen web-Anwendungen (Suchmaschinen, Textdaten und strukturierte Daten).

Im Falle von raumzeitlichem Data Mining ist die Anwendung geographischer Natur. *Raumzeitliches Data Mining* ist eine Spezialisierung von Data Mining, die nach interessanten Mustern in Objekten, Ereignissen und Prozessen in Raumzeit sucht. Typische Anwendungsfelder für raumzeitliches Data Mining umfassen die Bereiche Umweltmonitoring und Klimaforschung, Staatsschutz, Transport- und Verkehrsplanung, oder aber Fernerkundung und die aufstrebende Earth System Science (Pitman 2005).

2 Raumzeitliche Eigenheiten

Räumliche und raumzeitliche Daten unterscheiden sich von herkömmlichen Datenbeständen und haben daher folgerichtig zur Entstehung der räumlichen Informationswissenschaft und zur Entwicklung geographischer Informationssysteme und räumlicher Datenbanken geführt. „Spatial is special“ — folgende Eigenheiten zeichnen räumliche, respektive geographische Information aus (Anselin 1989, Goodchild 2001):

- räumliche Abhängigkeit, räumliche Autokorrelation ist charakteristisch für geographische Information (was nahe beieinander liegt, zeigt oft ähnliche Eigenschaften),
- der geographische Raum ist heterogen (ein kleiner Raumausschnitt ist selten repräsentativ für den gesamten Untersuchungsraum),
- geographische Attribute und Modelle sind skalenabhängig,
- geographische Information ist unsicher und unscharf,
- geographische Phänomene sind dynamisch.

Was aber bedeuten diese Eigenheiten räumlicher Information für die Entwicklung raumzeitlicher Data Mining Ansätze (Miller und Han 2009)?

Geographische Bezugssysteme. Wohl sind auch Datensätze im Marketing oder in der Bioinformatik multidimensional. Aber bei geographischen Daten sind die drei räumlichen Dimensionen (und die vierte zeitliche) nicht unabhängig. Die Repräsentation raumzeitlicher Information erfordert daher spezielle topologische und geometrische Bezugssysteme, und diese wiederum bestimmen, welche geographischen Attribute gemessen und was für Muster daraus abgeleitet werden können.

Räumliche Abhängigkeit und räumliche Heterogenität als Chance. Diese beiden Eigenheiten geographischer Information hebeln einige herkömmliche statistische Verfahren aus. Diese beruhen auf der Grundannahme der Unabhängigkeit von Stichproben, was gerade bei räumlichen Stichproben

selten gegeben ist. Raumzeitliches Data Mining greift hier an und macht sich diese beiden Eigenheiten raumzeitlicher Daten zu Nutze. In einigen Verfahren geht es eben genau darum, diejenigen Datenpunkte zu finden, die aus der Reihe tanzen, respektive andere Eigenschaften als ihre Nachbarn aufweisen.

Geographische Objekte und Beziehungen sind komplex. Während sich viele nicht-räumliche Objekte und Beziehungen problemlos zu Punkten in multidimensionalen Datenräumen reduzieren lassen, steckt das Wesen räumlicher Objekte und ihrer Beziehungen gerade in ihrer räumlichen Ausdehnung, ihrer Form oder ihren Nachbarschaftsbeziehungen. Raumzeitliche Data Mining Ansätze nützen solche Beziehungen aus und stützen sich gezielt auf Distanzbeziehungen, Ähnlichkeit oder Vernetzung zwischen Geoobjekten.

3 Raumzeitliche Muster, Regeln und Beziehungen suchen und finden

Nachdem der vorherige Abschnitt veranschaulicht wurde, weshalb es durchaus spezifisch raumzeitliche Data Mining Ansätze braucht, illustriert der folgende Abschnitt, wie die Eigenheiten geographischer Phänomene zur Suche raumzeitlicher Muster, Regeln und Beziehungen ausgenutzt werden können.

3.1 Räumliche Ausreisser (Spatial Outliers)

Ausreisser sind Beobachtungen, die nicht ins Gesamtbild eines Datensatzes passen. Falls Ihnen schon mal auf einer Auslandsreise nach dem ersten Einsatz ihrer Kreditkarte eben diese gesperrt wurde, sind Sie wohl Opfer eines Ausreisser-Algorithmus ihrer Bank geworden. *Räumliche Ausreisser* sind verortete Objekte, deren nicht-räumliche Attribute sich signifikant von den Attributen ihrer räumlichen Nachbarn unterscheiden. Ein Neubau in einem Viertel voller Altbauhäuser ist z.B. ein räumlicher Ausreisser (Shekhar *et al.* 2003b).

Oft können räumliche Ausreisser bereits in einfachen Scatterplots identifiziert werden, wenn z.B. die Attributwerte von Geopunkten gegen ihre X- oder Y-Koordinate aufgetragen werden. Es gibt aber auch statistische Tests, die spezifisch die Attributwerte eines Geopunktes mit denen seiner Nachbarn vergleichen (Shekhar, Lu und Zhang 2003a, Ng 2001).

3.2 Räumliche Assoziationsanalyse (Spatial Association Rules)

Die Assoziationsanalyse wird am besten mit der Warenkorbanalyse veranschaulicht. Marketingabteilungen grosser Firmen sind z.B. daran interessiert, Kundestämme abzugrenzen, indem sie Kunden respektive ihre Einkäufe suchen, die sich durch bestimmte Warenkombinationen auszeichnen. Ein solche Assoziationsregel wäre zum Beispiel: „In Warenkörben mit Bier finden sich oft auch Kartoffelchips.“ Sind Biertrinker einmal identifiziert, können sich gezielt für Kartoffelchips beworben werden.

Ganz ähnlich werden räumliche *Assoziationsregeln* formuliert. In einer Immobilien-Datenbank könnte man zum Beispiel nach Häusern in bevorzugter Lage suchen: „Liegt ein Haus nahe am Strand, ist es teuer.“ Ein wesentlicher Unterschied zum Warenkorb, der als binärer Zustand Bier und Chips enthält oder eben nicht, kann die räumliche Relation „nahe“ auf vielfältige Weise definiert werden und ist oft nicht als solche explizit in der räumlichen Datenbank gespeichert. Weiter müssen auch die Masse für die Güte einer Regel, gegeben als *Support* (Anzahl Warenkörbe mit Bier und Chips / alle Warenkörbe) und *Konfidenz* (Anzahl Warenkörbe mit Bier und Chips / alle Warenkörbe nur mit Bier), für den räumlichen Fall eigen definiert werden (Laube, de Berg und van Kreveld 2008).

3.3 Kollokation (Co-location)

Kollokations-Muster beschreiben die räumliche Verteilung von Objekten im Raum. Symbiotische Lebensformen bieten hier ein anschauliches ökologisches Beispiel. Bestimmte Ameisen und Blattläuse leben in Symbiose, entsprechend befinden sich Ameisennester oft in räumlicher Nähe zu Blattlauskolonien. *Kollokations-Muster* beschreiben Subsets räumlicher Objekte, deren Instanzen sich oft in räumlicher Nähe befinden (Shekhar *et al.* 2003b). Ein weiteres Beispiel bietet die räumliche Kollokation von Waldbrandherden und Raststätten.

Einerseits werden räumliche Korrelationsmasse aus der räumlichen Statistik verwendet um Kollokations-Muster zu finden (O'Sullivan und Unwin 2003). Andere Ansätze betrachten die räumliche Anordnung von Objekten als Transaktionen und verwenden sodann Methoden ähnlich der räumlichen Assoziationsanalyse (Koperski und Han 1995).

3.4 Raumzeitliche Clusteranalyse

Das Auffinden von Klumpungen gleicher oder ähnlicher Objekte in Raumzeit ist vielleicht die naheliegendste raumzeitliche Data Mining Technik. Raumzeitliche Cluster beschreiben zum Beispiel den Herd einer aufkeimenden Pandemie oder aber auffällige „hot spots“ in der räumlichen Verteilung von Delikten in der Verbrechensbekämpfung. Raumzeitliches Clustering gruppiert Objekte, die ähnliche Eigenschaften haben und sich damit von Objekten in anderen Clustern abgrenzen lassen. Raumzeitliches Clustering basiert damit auf der Grundannahme der räumlichen Abhängigkeit (Abschnitt 2): Objekte die räumlich beieinander liegen, sind oft ähnlich. Clusteranalyse kann dazu dienen, Objekte in Raumzeit zu klassifizieren und ist deshalb oft ein erster Schritt in der Entwicklung von Erklärungshypothesen (O'Sullivan und Unwin 2003).

In der räumlichen Statistik wird zum Testen auf räumliches Clustering typischerweise die Nullhypothese der zufälligen Punktverteilung als Vergleichsbasis herangezogen. Das Punktmuster selbst kann dann entweder bezüglich seiner Dichteverteilung oder bezüglich der Punktabstände untersucht werden (O'Sullivan und Unwin 2003). Dichtemasse beschreiben die absolute Lage der Punkte und ihre Variation bezüglich der Anzahl Punkte pro Raumeinheit (*Effekte erster Ordnung*). Häufig verwendete Dichtemasse sind Quadratanzahlmethoden und die Berechnung von Dichteoberflächen (z.B. Kernel Density Estimation, siehe Abbildung 2A). Relative Angaben zu den Abständen der Punkte untereinander widerspiegeln mögliche Interaktionen zwischen den Punktobjekten (*Effekte zweiter Ordnung*). Zum Beispiel die Methode des nächsten Nachbarn (engl. *Nearest Neighbor Method*) wurde entwickelt, um eine Unterscheidung von Punktverteilungen in gleichmässige, zufällige und geklumpete Verteilungsmuster vornehmen zu können.

3.5 Bewegungsmuster

Rasanter technologischer Fortschritt in den Bereichen mobile Kommunikation und Lokalisation führen zu immer umfangreicheren Datenbeständen zu dynamischen Geoprozessen. So erlauben zum Beispiel GPS-bestückte smart phones und Autos mit Navigationssystemen die fast lückenlose Erfassung der raumzeitlichen Bewegung von Individuen im Alltag. Die so entstehenden Datenbestände eignen sich besonders dazu, urbane Pendlerströme oder das Massenverhalten bei Grossanlässen systematisch zu untersuchen. Bewegungsdaten wachsen schnell ins Unermessliche und sind wegen ihrer raumzeitlichen Dichte und unumgänglichen Überlappungen schlecht

visualisierbar. Raumzeitliches Data Mining bietet sich hier als Werkzeug, um interessante Bewegungs- und Verhaltensmuster zu finden und damit dynamische Geoprozesse besser verstehen zu können.

Auch für das Mining von Bewegungsmustern prägen das räumliche Referenzsystem und die Modellierung des Bewegungsraums die Form und Struktur der möglichen Muster. Abbildung 2 illustriert typische Bewegungsräume für Menschen, Tiere oder Fahrzeuge und Waren: Den freien Bewegungsraum (A), topologische Räume in Gebäuden (B), den Raumzeitwürfel (C), Rasterräume (D), Bewegungsräume aus Mobilfunkzellen (E), und Verkehrsnetzwerke (F). Die raumzeitlichen Primitiven der Bewegungsanalyse bilden sog. *Trajektorien*. Polylinien zwischen zeitgestempelten und verorteten Beobachtungspunkten.

Einzelobjekte und Gruppen können Bewegungsmuster ausbilden. So zeigt zum Beispiel Objekt f_1 in Abbildung 2F ein Abbiegen nach rechts, gefolgt von einer langen Geraden. Ein zweites Beispiel wäre die Sequenz besuchter Mobilfunkzellen, etwa [I>F>G] bei Objekt e_2 in Abbildung 2E (Du Mouza und Rigaux 2005). Bewegungsmuster von Objektgruppen beschreiben kollektives Verhalten und können so auf mögliche Beziehungen zwischen Objekten hindeuten. Abbildung 2A veranschaulicht die Muster *Gänsemarsch* (Objekte folgen einander in einer Reihe, a_3 , Buchin, Buchin und Gudmundsson 2008) und *Herde* (Objekte bewegen sich zusammen für eine bestimmte Dauer, a_4 , Laube, van Kreveld und Imfeld 2004). Die Bewegungsmusteranalyse versucht derartige Muster zunächst konzeptionell zu beschreiben, dann zu formalisieren und schliesslich Algorithmen zu entwickeln, die derlei Muster zuverlässig und effizient erkennen (Laube 2009). Da sich viele Bewegungsmuster durch mit Zeit versehene geometrische Primitiven ausdrücken lassen, entwickeln sich viele Ansätze zum Mining von Bewegungsmustern an der Schnittstelle zwischen Geographie und theoretischer Informatik, resp. deren Unterdisziplin *Algorithmische Geometrie* (Gudmundsson, Laube und Wolle 2008, Gudmundsson, van Kreveld und Speckmann 2007). Die Definition von Distanzmassen zwischen Trajektorien zur Analyse deren Ähnlichkeit ist ein weiteres Beispiel für diese fruchtbare interdisziplinäre Zusammenarbeit.

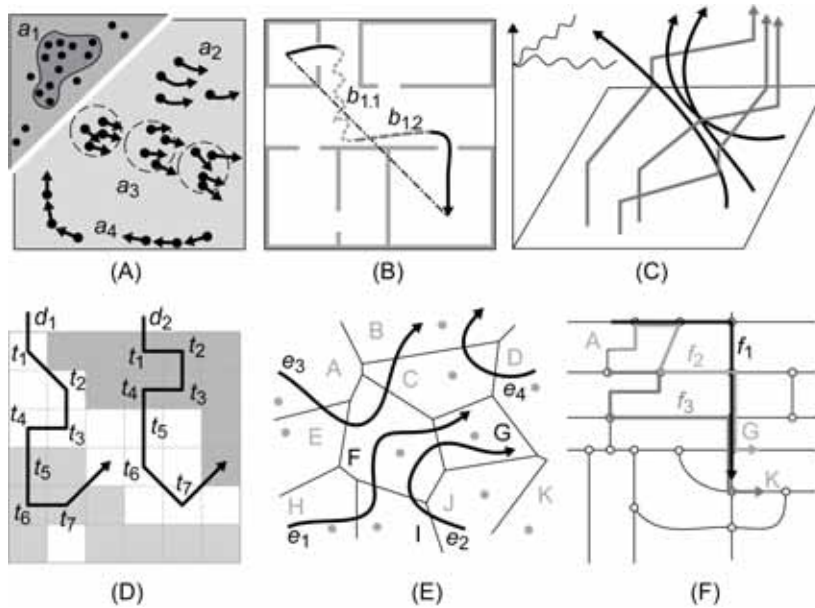


Abbildung 2: Bewegungsräume und -muster. (A) Freier Raum mit home range (a_1), Anführer (a_2), Herde (a_3), und Gänsemarsch (a_4); (B) Indoor-Raum mit gewundenem ($b_{1,1}$) und geradem ($b_{1,2}$) Bewegungsabschnitt; (C) Raumzeitwürfel mit Trajektorienbündel; (D) Rasterraum mit zwei ähnlichen Trajektorien; (E) Bewegung von Zelle zu Zelle; (F) Verkehrsnetz, drei unterschiedliche Wege von A nach K (Quelle: Laube 2009, verändert).

3.6 Ähnlichkeitsanalyse

Wie in Abschnitt 2 erläutert, zeichnen sich räumliche und raumzeitliche Primitiven dadurch aus, dass sie wesentlich komplexere Formen und Beziehungen bilden können als herkömmliche Datenobjekte. Es ist daher naheliegend, Muster und Beziehungen zu suchen, die sich auf die Ähnlichkeit von geographischen Objekten konzentrieren. Welche Parzellen sind ähnlich geformt? Welche Bewegungs-Trajektorien sind ähnlich und zeigen damit ähnliches Pendlerverhalten?

Die Definition von Distanzmassen zwischen Trajektorien und darauf aufbauende Ähnlichkeitsmasse für Trajektorien hat sich in jüngster Zeit zu

einem eigenen Forschungszweig entwickelt. Während viele Ansätze zunächst die Ähnlichkeit der rein räumlichen Polylinie der Trajektorie untersuchen (z.B. mit der Fréchet Distanz), gilt es v.a. auch die zeitlichen Aspekte der Bewegung zu berücksichtigen (Pelekis et al. 2007, Vlachos, Kollios und Gunopulos 2002). Andere Ansätze zielen darauf ab, Trajektorien in Unterabschnitte ähnlicher Eigenschaften zu segmentieren (Buchin *et al.* 2010). Dies kann einerseits ein Vorverarbeitungs-Schritt sein, kann aber auch als Data Mining im engeren Sinne betrachtet werden, versucht eine Segmentation doch nichts anderes als Struktur in den Daten zu finden. Letztlich dienen Ähnlichkeitsmasse dazu, Trajektorien zu gruppieren und dadurch ein besseres Verständnis der Bewegung zugrunde liegender Prozesse zu erlangen.

4 Schlussbemerkungen und Ausblick

Data Mining ist ein kraftvolles Instrument zur Bewältigung der stetig wachsenden räumlichen Datenflut. Da wo der menschliche Betrachter alleine überfordert ist, können Algorithmen helfen interessante und bisher unbeachtete Zusammenhänge zwischen Objekten in Raumzeit aufzudecken. Durch die zunehmende Vernetzung vormals getrennter Datenbestände entstehen neue Informationsquellen, die es erlauben nach Mustern und Beziehungen zu suchen, die bisher unerkannt bleiben mussten. Diese Entwicklung birgt allerdings auch Gefahren. Im Windschatten von Google Earth und räumlichen Apps kann derzeit eine Goldgräberstimmung in einer Vielzahl von Anwendungsfeldern beobachtet werden. Oft zählt mehr was technisch möglich ist, als was semantisch sinnvoll wäre. Die Integration heterogener Datenquellen birgt die Gefahr, dass wesentliches Kontextwissen und Hintergrundinformation zu den einzelnen Datensätzen auf der Strecke bleiben und somit Daten in unzulässiger Weise verwendet werden und ungültige Schlüsse gezogen werden. Die Geoinformatik hat hier die Chance aus dem Erfahrungsschatz der Disziplin zu schöpfen und zur Entwicklung räumlich und geographisch informierter Data Mining Ansätze beizutragen.

Im Sinne eines Ausblickes soll hier noch auf zwei aktuelle Entwicklungen hingewiesen werden. Während data mining vornehmlich auf bereits in Datenbanken strukturiert abgelegte Datenbestände angewendet wird, verlangen auch stetig wachsende Bestände *unstrukturierter* Daten, wie sie etwa Textdokumente im Internet darstellen, nach Methoden zur Wissensgenerierung. *Information Retrieval* (IR) und die räumliche Unterdisziplin *Geographic Information Retrieval* (GIR) versucht in grossen Dokumentbeständen

inhaltsbezogene Abfragen unter Zuhilfenahme von Relevanzabschätzungen zu ermöglichen (Purves und Jones 2006). Wer in einer Suchmaschine mit einem unvollständigen Suchbegriff etwas sucht, macht nichts anderes als Information Retrieval. Ein räumliches Beispiel wäre die Suche nach „Hotels südlich von Zürich“ in Internet, was neben der Verortung von Dokumenten auch die Berücksichtigung von topologischen Beziehungen erfordert.

Die zweite Entwicklung betrifft einen sich abzeichnenden Wandel der IT-Infrastruktur. Konventionelle GIS und räumliche Datenbanken basieren darauf, dass kraftvolle Rechner grosse Datenmengen integrieren und zentralisiert verarbeiten. Derzeit kann allerdings beobachtet werden, dass räumliche Anwendungen sich zunehmend aus den Desktop-Computern weniger Spezialisten verabschieden und in die mobilen Geräte von Laien einwandern. Im Zuge dieses *pervasive computing* Trends verändern sich auch die Rahmenbedingungen für räumliche Datenprozessierung. Es kann nicht immer nur davon ausgegangen werden, dass räumliche Informationen zentral verarbeitet werden können. Es stellt sich vermehrt auch die Frage, ob bereits lokal in verteilten System interessante Muster durch verteilte Data Mining Ansätze gefunden werden können. Als Beispiel sei hier wiederum auf die Bewegungsanalyse verwiesen. In nicht allzu ferner Zukunft wird es so sein, dass Autos miteinander kommunizieren und nur durch verteilte Kontrollsysteme die Ausbildung von Stau verhindert unterdrückt wird (Laube, Duckham und Wolle 2008).

5 Literatur

- Anselin, L. 1989. What is special about spatial data. Alternative perspectives on spatial data analysis. In *Technical Report*. Santa Barbara, Calif.: National Center for Geographic Information and Analysis.
- Buchin, K., M. Buchin & J. Gudmundsson. 2008. Detecting single file movement. In *Proceedings of the 16th ACM SIGSPATIAL international conference on Advances in geographic information systems*. Irvine, California: ACM.
- Buchin, M., A. Driemel, M. van Kreveld & V. Sacristan. 2010. An Algorithmic Framework for Segmenting Trajectories based on Spatio-Temporal Criteria In *18th International Conference on Advances in Geographic Information Systems (ACM SIGSPATIAL GIS 2010)*. San Jose, California: ACM.
- Du Mouza, C. & P. Rigaux (2005) Mobility patterns. *Geoinformatica*, 9, 297-319.
- Fayyad, U., G. Piatetsky-Shapiro & P. Smyth (1996) From Data Mining to Knowledge Discovery in Databases. *AI Magazine*, 17, 37-54.
- Goodchild, M. F. 2001. A Geographer Looks at Spatial Information Theory. In *Spatial Information Theory*, 1-13. Springer Berlin / Heidelberg.

- Gudmundsson, J., P. Laube & T. Wolle. 2008. Movement Patterns in Spatio-temporal Data. In *Encyclopedia of GIS*, 726-732. Springer US.
- Gudmundsson, J., M. van Kreveld & B. Speckmann (2007) Efficient Detection of Patterns in 2D Trajectories of Moving Points. *Geoinformatica*, 11, 195-215.
- Hand, D. J., H. Manilla & P. Smyth. 2001. *Principles of Data Mining*. Cambridge, MA: MIT Press.
- Koperski, K. & J. Han. 1995. *Discovery of Spatial Association Rules in Geographic Information Databases*. Springer-Verlag.
- Laube, P. 2009. Progress in Movement Pattern Analysis. In *Behaviour Monitoring and Interpretation - BMI - Smart Environments*, eds. B. Gottfried & H. Aghajan, 43-71. CRC Press.
- Laube, P., M. d. Berg & M. v. Kreveld. 2008. Spatial Support and Spatial Confidence for Spatial Association Rules. In *Headway in Spatial Data Handling*, eds. A. Ruas & C. Gold, 575-593. Springer Berlin Heidelberg.
- Laube, P., M. Duckham & T. Wolle. 2008. Decentralized Movement Pattern Detection amongst Mobile Geosensor Nodes. In *GIScience 2008*, eds. T. J. Cova, K. Beard, M. F. Goodchild & A. U. Frank, 199-216. Springer.
- Laube, P., M. van Kreveld & S. Imfeld. 2004. Finding REMO - Detecting Relative Motion Patterns in Geospatial Lifelines. In *Developments in Spatial Data Handling*, ed. P. F. Fisher, 201-214. Berlin Heidelberg, DE: Springer.
- Miller, H. & J. Han. 2009. *Geographic Data Mining and Knowledge Discovery*, Second Edition. Boca Raton, FL: CRC Press.
- Ng, R. T. 2001. Detecting outliers from large datasets. In *Geographic data mining and knowledge discovery*, eds. H. J. Miller & J. Han, 218-235. London, UK: Taylor & Francis.
- O'Sullivan, D. & D. J. Unwin. 2003. *Geographic Information Analysis*. Hoboken, NJ: John Wiley and Sons.
- Pelekis, N., I. Kopanakis, G. Marketos, I. Ntoutsi, G. Andrienko & Y. Theodoridis. 2007. Similarity Search in Trajectory Databases. In *Temporal Representation and Reasoning, 14th International Symposium on*, 129-140.
- Pitman, A. J. (2005) On the role of Geography in Earth System Science. *Geoforum*, 36, 137-148.
- Purves, R. & C. Jones (2006) Geographic Information Retrieval (GIR). *Computers, Environment and Urban Systems*, 30, 375-377.
- Runkler, T. A. 2010. *Data Mining, Methoden und Algorithmen intelligenter Datenanalyse*. Wiesbaden, Germany: Vieweg+Teubner.
- Shekhar, S., C. T. Lu & P. S. Zhang (2003a) A unified approach to detecting spatial outliers. *Geoinformatica*, 7, 139-166.
- Shekhar, S., P. Zhang, Y. Huang & R. R. Vatsavai. 2003b. Trends in Spatial Data Mining. In *Data Mining: Next Generation Challenges and Future Directions*, eds. H. Kargupta, A. Joshi, K. Sivakumar & Y. Yesha. MIT/AAAI Press.
- Vlachos, M., G. Kollios & D. Gunopulos. 2002. Discovering Similar Multidimensional Trajectories. In *18th International Conference on Data Engineering (ICDE'02)*.