

Neue Erkenntnisse zum Mobilitätsverhalten dank Data Mining?

**De nouvelles découvertes sur le comportement de
mobilité par Data Mining?**

**New findings on the mobility behavior through
Data Mining?**

Zusammenfassung d, f, e

**büro widmer ag
Paul Widmer
Thomas Buhl**

**Institut für Datenanalyse und Prozessdesign (idp), zhaw
Andreas Ruckstuhl
Markus Dettling
Sina Rüeger**

**Forschungsauftrag SVI 2004/014 auf Antrag der Schweizerischen
Vereinigung der Verkehrsingenieure und Verkehrsexperten (SVI)**

Mai 2011

ZUSAMMENFASSUNG

Unter Data Mining versteht man im engeren Sinn das systematische (in der Regel automatisierte oder halbautomatisierte) Entdecken und Extrahieren von vorher unbekanntem statistischen Informationszusammenhängen aus grossen Datenmengen. Im deutschen Sprachgebrauch steht Data Mining oft für den ganzen Analyse-Prozess, der auch die Vorbereitung der Daten sowie die Bewertung der Resultate umfasst. Data Mining wird in verschiedenen Bereichen erfolgreich eingesetzt. Anwendungsbeispiele aus schweizerischen Verkehrsplanungen fehlen aber bisher weitgehend. Ziel der Forschungsarbeit war es, den praktisch tätigen Verkehrsplaner mit dem Prozess und den Methoden von Data Mining vertraut zu machen und die Möglichkeiten von Data Mining als Hilfsmittel in der Verkehrsplanung auszuloten.

Data Mining wird als iterativer, lernender Prozess dargestellt, in welchem die Phasen von der Fragestellung über das Sammeln und Aufbereiten der Daten, die Modellierung und die Auswertung der Ergebnisse bis zu deren Umsetzung in die Praxis mehrfach durchlaufen werden. In dieser Arbeit wird dieser Prozess genauer beschrieben und ein Überblick über eine Auswahl von Methoden, die in der Modellierung verwendet werden, gegeben.

Beispiele aus der Literatur illustrieren das breite Anwendungsspektrum von Data Mining in der Verkehrsplanung (z.B. Verkehrserzeugung, Autobesitz, Verkehrsmittel- und Routenwahl oder Klassifikation von Mobilitätsmustern). Bei den beschriebenen Beispielen handelt es sich um Forschungsarbeiten. Deren Ergebnisse haben noch kaum breiten Eingang in die Praxis gefunden.

An Fallbeispielen wird demonstriert, wie Data Mining in der Praxis angewendet werden kann. Als Datensätze werden der Mikrozensus Verkehr 2005 und Raumstrukturdaten des Bundesamtes für Statistik verwendet. Die Fallbeispiele behandeln die Analyse der Häufigkeit von Wegeketten, die Vorhersage der Anzahl Wegeketten pro Person und Tag, die Klassifikation nach Mobilitätstypen sowie die Vorhersage des Mobilitätstyps einer Person aufgrund sozio-demographischer Merkmale und Raumstrukturinformationen zu den Wohn- und Zielorten.

Aus der grossen Vielfalt von Software-Lösungen für Data Mining wird eine Auswahl proprietärer und frei verfügbarer Pakete, welche für den Einsatz in der Verkehrsplanung als grundsätzlich geeignet beurteilt werden, grob und ohne Wertung beschrieben.

Die Studie kommt zum Schluss, dass Data Mining in der Verkehrsplanung sicher nutzbringend anwendbar ist, dass aber nicht – wie vielleicht erhofft – automatisch auf alle Fragen gute Antworten erwartet oder ohne Zutun des Anwenders aus vorhandenen Datensätzen neue Erkenntnisse gewonnen werden können.

Empfehlenswerte Einsatzgebiete für Data Mining in der Verkehrsplanung sind beispielsweise:

- Klassifikationen, z.B. des Mobilitätsverhaltens
- Visualisierung komplexer mehrdimensionaler Datensätze zum raschen Erkennen von Mustern resp. Clustern
- Rasche und automatische Erkennung der (aus statistischer Sicht) wichtigsten Prädiktoren des Mobilitätsverhaltens
- Analyse der Entscheidungsprozesse, z.B. bei der Verkehrsteilnahme.

Zusammenhänge, die mit Data Mining Methoden extrahiert werden, sind grundsätzlich datengetrieben und müssen keine Kausalitäten widerspiegeln. Deshalb wird empfohlen, aus Kausalitätsüberlegungen abgeleitete Modelle weiterhin mit statistischen Methoden an die Daten anzupassen. Konventionelle Modellansätze und Data Mining sollen als sich ergänzende und gegenseitig unterstützende Methoden eingesetzt werden.

Um Data Mining zukünftig auch in der Verkehrsplanung nutzbringend einsetzen zu können, bedarf es keiner weiteren Forschung. Vielmehr sind möglichst viele praktische Anwendungen erwünscht, mit denen Verkehrsplaner und Data Mining Experten in interdisziplinärer Zusammenarbeit Erfahrungen sammeln und weitergeben können.

RÉSUMÉ

Le "Data Mining" (aussi appelé exploration ou fouille de données, extraction de connaissances) est un procédé utilisé avec succès dans différents domaines afin de découvrir des types caractérisés à partir de très nombreuses données. Des exemples d'application provenant de planifications des transports en Suisse manquent jusqu'à présent largement. Le but de la recherche était de familiariser le planificateur en transports dans sa pratique avec le procédé et les méthodes de "Data Mining" et de sonder les possibilités du "Data Mining" comme aide à la planification des transports.

Le "Data Mining" se présente comme un procédé d'apprentissage itératif dans lequel les phases sont parcourues plusieurs fois, du questionnement à la récolte et la préparation des données, de la modélisation et du dépouillement des résultats à leur mise en œuvre pratique. Pour donner une vue d'ensemble, une sélection des méthodes souvent appliquées est décrite en distinguant entre celles de l'apprentissage "surveillé" et "non surveillé".

Des exemples issus de la bibliographie illustrent le large éventail d'applications du "Data Mining" dans la planification des transports (par exemple génération de trafic, possession d'auto, choix du moyen de transport et de l'itinéraire ou classification de types de mobilité). Les exemples décrits proviennent de travaux de recherche. Leurs résultats n'ont pas encore vraiment trouvé leur place dans la pratique.

A l'aide d'exemples caractéristiques, il est démontré comment le "Data Mining" peut être mis en œuvre dans la pratique. Les données utilisées proviennent du micro-recensement sur le comportement en matière de transports de 2005 et des données structurelles territoriales de l'Office fédéral du développement territorial. Les exemples typiques traitent de l'analyse de la fréquence des chaînes de déplacement, de la prévision du nombre de chaînes de déplacement par personne et jour, de la classification selon les types de mobilité ainsi que des prévisions du type de mobilité d'une personne sur la base de ses caractéristiques sociodémographiques et des informations structurelles territoriales des lieux de domicile et de destination.

Parmi la grande variété de solutions logicielles pour le "Data Mining", un choix d'ensembles propriétaires et disponibles gratuitement, qui sont en principe jugés appropriés pour une utilisation dans la planification des transports, est présenté sommairement et sans évaluation.

L'étude arrive à la conclusion que le "Data Mining" est certainement applicable utilement dans la planification des transports mais qu'il ne faut pas s'attendre – comme peut-être espéré – à de bonnes réponses ou que de nouvelles connaissances puissent émerger des données existantes sans intervention de l'utilisateur. Les domaines recommandés de recours au "Data Mining" dans la planification des transports sont par exemple:

- Classifier, par exemple les comportements en matière de mobilité
- Visualiser des données complexes et pluridimensionnelles pour rapidement reconnaître des types respectivement des groupes
- Reconnaître rapidement et automatiquement les variables déterminantes importantes (du point de vue statistique) du comportement en matière de mobilité
- Analyser les procédures de décision, par exemple en matière de déplacements.

Comme ni les règles de classification ni les méthodes pour reconnaître automatiquement les variables déterminantes statistiquement significatives ne tiennent compte des causalités entre ces dernières et les variables visées, les méthodes de "Data Mining" ne peuvent pas remplacer les méthodes conventionnelles pour l'estimation de modèles. Il est bien plutôt recommandé de recourir à des approches conventionnelles de modélisation et au "Data Mining" en les considérant comme méthodes se complétant et se soutenant réci-proquement.

Afin de pouvoir recourir utilement au "Data Mining" à l'avenir aussi dans la planification des transports, il n'est pas nécessaire d'entreprendre de nouvelles recherches. Sont bien plutôt souhaitées des applications pratiques les plus nombreuses possibles par lesquelles les planificateurs en transports et les experts en "Data Mining" pourront rassembler des expériences en collaboration interdisciplinaire et les transmettre plus loin.

SUMMARY

Data mining as a process for revealing yet unknown patterns from large amounts of data has been employed successfully in various fields. Application examples from Swiss transport planning however are largely missing so far. The aim of this research paper is to familiarize the practicing transportation planner with both process and methods of data mining and to explore the possibilities of data mining as an aid in transport planning.

Data mining is commonly perceived as an iterative learning process, in which all actions such as problem formulation, data collection, data preparation, modeling and evaluation are run through several times and finally completed by the deployment step. In this work, this process is described in some detail and an overview of a selection of methods used in the modeling is given.

A literature review illustrates the wide range of applications data mining is used for in transport planning. These include traffic generation, car ownership, mode and route choice or classification of mobility patterns. Because they mostly result from research activities, their outcomes have not yet found wide acceptance in practice.

With case studies we demonstrate how data mining can be successfully applied in a transportation planner's daily routine. The "Micro Census on Travel Behavior 2005" and the data on spatial structure of the Federal Office for Spatial Development (ARE) are used. The case studies encompass a frequency analysis of trip chaining, predicting the number of trip chains per person per day, the segmentation of people into different mobility behavior types and predicting these for a person based on their socio-demographic characteristics and spatial structure information of home and destination locations.

We continue with a review of a selection of data mining software, including both proprietary and freely available tools. All the reviewed packages are basically regarded as suitable for the use in transport planning but are not evaluated further.

As a conclusion of our research we believe that data mining can be beneficial to and yield a substantial contribution in quantitative transportation research. However, it does not – as might be hoped for by non-professionals – automatically generate sensible answers to some vague questions floating around. It still needs a clear aim and the input of an expert user to extract new insights from existing data sets.

Recommended application areas for data mining in transportation planning include:

- Classification, e.g. of the mobility behavior
- Visualization of complex multidimensional data sets for a simple identification of patterns and clusters, respectively.
- Quick and automated detection of (from a statistical point of view) most important predictors of mobility behavior
- Analysis of decision processes, e.g. in the context of traffic participation.

It is also important to keep in mind that any association that is extracted with data mining methods is inherently data driven and might not reflect causalities. We therefore recommend continuing fitting models that arise from causality considerations by traditional statistical methods. Conventional modeling approaches and data mining may be used complementary or as mutually supportive methods.

To exploit the potential of data mining in transportation planning, we think that no further research is needed. But rather a variety of practical application cases should be carried out, where transportation engineers and data mining experts can collect and share experiences in such interdisciplinary collaborations.