

Thilo Siegle

Fahrerverhalten und Ermüdung, Data Mining in Fahrerermüdungsverhalten

## Einführung

Mit dieser Arbeit wird der Versuch unternommen, die relativ junge Technik Data Mining (Grob & Bensberg, 1999; SPSS Inc., 1999) auf eine Fragestellung in einer psychologischen Domäne anzuwenden. Dies geschieht exemplarisch anhand des arbeitspsychologischen Themas der Müdigkeitsdetektion von Autofahrern. Anwendungsbasis sind Daten eines Fahrsimulatorversuchs der Daimler Chrysler AG, Abteilung RBP-BA, zur Müdigkeitsdetektion im April 2002.

Dabei erhebt diese Arbeit weder den Anspruch, ein Überblicksartikel zum Stand der Forschung im Bereich Fahrermüdigkeit zu sein, noch eine Einführung in Data Mining zu ersetzen. Konzepte und Erkenntnisse aus beiden Bereichen, auf denen die Untersuchung aufbaut, werden aber in Kapitel 1 eingeführt.

In Kapitel 2 werden der Fahrsimulatorversuch, die gewonnenen Rohdaten und deren Aufbereitung sowie die eingesetzten Modellierungsmethoden vorgestellt. Kapitel 3 gibt einen Überblick über die Ergebnisse der Untersuchung. Dem Prozess Data Mining ist dabei geschuldet, dass diese beiden Kapitel in ihrer Abfolge nicht stringent der Chronologie der Untersuchung entsprechen, nach der man zwischen den Kapiteln mehrfach springen müsste. Um die Darstellungsform einer empirischen psychologischen Diplomarbeit einzuhalten, ist die Aufteilung in Methoden und Ergebnisse trotzdem aufrechtzuerhalten.

Kapitel 4 gibt eine kritische Würdigung der Ergebnisse und zeigt Möglichkeiten der Optimierung sowie Chancen, die sich mit einer eingesetzten Technik des Data Mining in der untersuchten Domäne Fahrerermüdung bieten.

## Ermüdung und Autofahren

Ermüdung beim Autofahren und das damit verbundene Phänomen des Sekundenschlafs sind ein seit langem und intensiv bearbeitetes Forschungsgebiet (Brown, 1994). Ursachen, Wirkmechanismen, Erkennungsmöglichkeiten und Gegenmaßnahmen sind Thema von Forschungsprojekten (AWAKE, 2003) und Forschungsverbunden (AMCoNet, 2003), in denen Arbeitswissenschaftler, Ingenieure, Psychologen und Mediziner interdisziplinär nach Lösungen mit einer wesentlichen Gefährdung des Straßenverkehrs suchen.

In offiziellen Unfallstatistiken werden lediglich etwa 2 % der Unfälle der Ermüdung zugeschrieben (Brown, 1994, S. 309; Kircher, Uddman & Dandin, 2002, S.12). Weil Ermüdung jedoch keine eindeutigen Spuren hinterlässt wie beispielsweise der Blutalkoholgehalt bei Unfällen unter Alkoholeinfluss und nur aus charakteristischen Begleitumständen geschlossen werden kann, dürften diese Zahlen die wahre Prävalenz deutlich unterschätzen. Brown(1994, S.309) zitiert in seinem Überblicksartikel zur Fahrerermüdung Studien, die bei eingehender Untersuchung bei mindestens 10 % der Unfälle, an denen nur ein Fahrzeug beteiligt ist, wird dieser Anteil sogar auf mindestens 25 % geschätzt. Durch Ermüdung verursachte Unfälle treten typischerweise bei hoher Geschwindigkeit auf, der Fahrer ist dabei eher nicht versucht, den Unfall zu vermeiden (NCSDR/NHTSA Expert Panel on Driver Fatigue and Sleepiness, 1998,S.11). Vermutlich aus diesem Grund ist der Anteil ermüdungsbedingter Unfälle an schweren Unfällen höher (a.a.O.,S.9).

Felder, auf denen Forschung zum Thema betrieben wird, sind neben der Analyse müdigkeitsbedingter Unfälle unter anderem die Erkennung und Vorhersage von Ermüdung und die mögliche Umsetzung dieser Erkenntnisse in Fahrerassistenzsystemen oder die Gestaltung von Schichtplänen bei Berufskraftfahrern (AMCoNet,2003, Struktur).

Die vorliegende Arbeit ist im Bereich der Erkennung von Ermüdung angesiedelt.

Die in der Forschung auch bearbeitete Fragestellung, auf welche Art ein System zur Ermüdungserkennung auf den Fahrer im Fall erkannter Müdigkeit einwirken sollte (Bekiaris, 1999; Fairbanks, Fahey & Wierwille, 1995, Part II), wird hier nicht erörtert. Diese Frage setzt ein zuverlässiges System der Erkennung voraus, welches mit dem jetzigen Stand der Forschung trotz intensiver Bemühungen noch nicht gegeben ist (siehe 1.2).

Die Bedeutung, die einem solchen hypothetischen System beigemessen wird, zeigt sich auch am Interesse außerhalb des wissenschaftlichen Fachpublikums (Feller, 2003; Wüst,2002).

Im November letzten Jahres startete die National Sleep Foundation in den USA eine Website, die sich ausschließlich mit dem Thema Ermüdung und Autofahren an die Öffentlichkeit wendet, und versucht, diese für das Phänomen zu sensibilisieren (National Sleep Foundation , 2003a, 2003b).

## Data Mining

Data Mining ist ein Forschungsbereich, der seine Ursprünge im Feld der Datenbankanwendungen in der Informatik hat.

Berry und Linoff (1997, zitiert nach Grob & Bensberg, 1999, S. 3):“(...) sehen Data Mining als Exploration und Analyse großer Datenmengen mit automatischen oder semiautomatischen Werkzeugen, um bedeutungsvolle Muster und Regeln aufzufinden .“.

Knowledge Discovery in Databasis (KDD) bezieht sich auf den Prozess, in dem Verfahren des Data Mining zum Einsatz kommen (Fayyad et al., 1996, zitiert nach Nakhaeizadeh, Reinartz & Wirth, 1998, S. 4). Dieser später näher beschriebene Prozess ist die entscheidende Gemeinsamkeit der verschiedenen Ziele des Data Mining und der nach Herkunft und Funktionsweise sehr unterschiedlichen Verfahren.

Mit zunehmender Rechenkapazität erweiterten sich Möglichkeiten, auch mit rechenintensiven Verfahren nach Mustern und Gesetzmäßigkeiten in Datenbanken zu suchen. In den 1990er Jahren hat Data Mining so eine schnelle Entwicklung durchlaufen, vom ersten interdisziplinären Workshop „Knowledge Discovery in Databases (KDD)“ auf der Joint Conference of Artificial Intelligence 1989 in Detroit bis zur Herausgabe der Zeitschrift „Data Mining and Knowledge Discovery“ 1997.

## Data Mining in der Psychologie

In der Psychologie ist Data Mining ein noch wenig eingesetztes und untersuchtes Konzept. Eine Suche in der Datenbank PsycINFO im Mai 2002 mit dem Suchbegriff „Data Mining“ ergab 17 Treffer. Abgesehen von den rein methodischen Fundstellen zeigen aber bereits diese wenigen Treffer mögliche Ansatzpunkte für Data Mining in der Psychologie:

Die Anwendung von Verfahren des Data Mining ist die naheliegendste Einsatzmöglichkeit. Der gewinnbringende Einsatz erfordert Anwendungsfelder, in denen Datenerhebungsmethoden eingesetzt werden, die Daten in entsprechend großer Menge möglichst automatisch erzeugen. Im psychologischen Kontext ist das nicht naturgemäß gegeben. Zwei Varianten großer Datensätze sind aber auch hier anzutreffen: Untersuchungen mit (a) sehr vielen Beobachtungen pro Versuchsperson und mit (b) sehr vielen Versuchspersonen pro Beobachtung. Fall (a) ist in der Biopsychologie anzutreffen ( 6 Treffer in Datenbanksuche aus der Psychopharmakologie und der Neuropsychologie), Fall (b) im Marketing (3 Treffer). Die gefundenen Anwendungsbeispiele sind dabei interdisziplinärer Natur und könnten auch in der Medizin oder den Wirtschaftswissenschaften angestoßen worden sein.

Untersuchung von Anforderungen im Data-Mining –Prozess an Auswertende und Entwicklung von Verfahren, die Aufgaben beim Data Mining bezüglich ihres psychologischen Anforderungsprofils optimieren. Mit 2 Treffern zu kognitiven Aufgaben im Data-Mining Prozess ist dieses Feld noch seltener untersucht als die Anwendung von Verfahren.

Optimierung der Darstellung von Ergebnissen aus dem Data-Mining-Prozess. Ebenfalls nur zwei Publikationen bearbeiten die Unterstützung der Vermittlung von Ergebnissen im Data Mining durch adäquate Visualisierung.

Fahrsimulatorexperimente sind daher sehr datenintensiv. In einer darauf aufbauenden Arbeit bieten sich somit Verfahren des Data Mining (1.) zur Anwendung auf einem Datensatz mit sehr vielen Beobachtungen pro Versuchsperson (a) an. Diese Domäne, sei es Mensch-Maschine-Interaktion oder Arbeitspsychologie, findet sich in PsycINFO noch nicht im Zusammenhang mit Data Mining.