

Vahlens Großes Marketinglexikon

Herausgegeben von

Prof. Dr. Hermann Diller

2., völlig überarbeitete und erweiterte Auflage

Verlag C.H. Beck München
Verlag Franz Vahlen München

VERLAG
VAHLEN
MÜNCHEN
www.vahlen.de

Die Deutsche Bibliothek – CIP-Einheitsaufnahme

Vahlens Großes Marketinglexikon / hrsg. von Hermann Diller. –
2., völlig überarb. und erw. Aufl. – München : Beck; München : Vahlen, 2001
ISBN 3-8006-2689-6

ISBN 3 8006 2689 6

© 2001 Verlag Franz Vahlen GmbH, Wilhelmstraße 9, 80801 München

Satz: Fotosatz H. Buck, 84036 Kumhausen

Druck: Druckerei C.H. Beck

(Adresse wie Verlag)

Bindung: Großbuchbinderei Monheim

Umschlaggestaltung: Bruno Schachtner, Dachau

Gedruckt auf säurefreiem, alterungsbeständigem Papier
(hergestellt aus chlorfrei gebleichtem Zellstoff)

Data Mining (DM)

umfasst die automatische Identifikation nicht trivialer, aber dennoch interpretierbarer Zusammenhänge zwischen zwei oder mehr Variablen in großen Datenmengen und stellt einen zentralen Schritt im „*Knowledge Discovery in Databases*“-Prozess (kurz: KDD-Prozess) dar. Dabei steht die explorative Analyse der Daten im Vordergrund. Die Abgrenzung zur traditionellen

→ Multivariatenanalyse, in der zum Teil identische Methoden (z.B. Clusteranalyse) eingesetzt werden, ergibt sich aus der Perspektive, die dem jeweiligen Forschungsansatz zugrunde liegt. Traditionellerweise wird an die Datenanalyse – vereinfacht ausgedrückt – die auf substanziiell begründete Hypothesen gestützte Aufgabe gestellt „Zeige, was im vorliegenden Kontext interessant ist!“ (Top down-Ansatz). Die Aufgabe an das Data Mining kann hingegen durch die Maxime „Finde heraus, was interessant ist!“ umschrieben werden (Bottom up-Ansatz). Der zuletzt genannte Ansatz bietet zwar aufgrund des theorieleeren anmutenden „Buddelns in der Datenmine“ einerseits durchaus Ansatzpunkte für Kritik, er ermöglicht jedoch andererseits durch die explorative Vorgehensweise die Aufdeckung bisher nicht beachteter bzw. bekannter Zusammenhänge. Darüber hinaus eignet er sich gerade auch für die Auswertung von Datenbeständen mit sehr vielen Variablen, bei denen die Formulierung expliziter Hypothesen in Bezug auf die möglichen Variableninteraktionen komplexitätsbedingt aussichtslos erscheint.

Wesentliches Ziel des DM ist somit die Identifikation von *Mustern* in den vorliegenden Daten, um auf dieser Grundlage dann beispielsweise zukünftige Werte zu prognostizieren. Dazu sind Aufgaben der → Klassifikation (s.a. → AID) bzw. Segmentierung (→ Clusteranalyse) sowie der → Assoziationsanalyse zu bewältigen. Ein typisches Beispiel für eine entsprechende Klassifikationsaufgabe ist die Partitionierung potenzieller Kundenkreise entsprechend der Lukrativität unterschiedlicher Verkaufsförderungsmaßnahmen. Weitere in der Literatur dokumentierte DM-Anwendungen entstammen u.a. der Kreditwirtschaft (z.B. die Entscheidungsunterstützung bei Bonitätsfragen im Rahmen der Kreditvergabe oder die frühzeitige Aufdeckung von Kreditkartenmissbräuchen) sowie dem → Direktmarketing (z.B. die Selektion von Kundenadressen für gezielte Mailing-Aktionen). Assoziationsanalysen können z.B. die Untersuchung der Affinität der Käufer eines bestimmten Produktes zum Kauf anderer Produkte zum Gegenstand haben. Die damit korrespondierende Warenkorbanalyse kann sogar als Paradebeispiel eines Anwendungsgebietes von DM im Marketingkontext gesehen werden. Als fast schon „klassisches“ Szenarium wird in der Literatur der gemeinsame Kauf von

Windeln und Bier durch Väter, die beim Einkauf von Babyutensilien offensichtlich auch sich selbst etwas Gutes tun wollen, angeführt. Nichttriviale Verbundbeziehungen dieser Art dürften auf dem Wege der statistischen Prüfung a priori formulierter Forschungshypothesen nur schwer zu identifizieren sein. Der im Rahmen eines entsprechenden KDD-Prozesses gefundene Zusammenhang erfüllt auch die eingangs genannte Forderung nach einer sinnvollen Interpretierbarkeit.

Wird eine Bezugsvariable im DM extern vorgegeben, so bezeichnet man den resultierenden KDD-Prozess als „gelenkt“, anderenfalls als „ungelenkt“. Um die Aufgaben im Rahmen von Klassifikation und Assoziationsanalyse zu bewältigen, werden unterschiedliche DM-Ansätze, wie z.B. Assoziationsregeln, → Clusteranalysen, → Diskriminanzanalysen, Entscheidungsäume und → neuronale Netze eingesetzt. Eine umfassende Systematisierung von Verfahren und korrespondierenden Anwendungen erscheint derzeit allerdings noch nicht sinnvoll, da laufend neue Vorschläge zur Diskussion gestellt werden. In vielen Fällen ist überdies der kombinierte Einsatz mehrerer Verfahren angebracht. Da bei der skizzierten Vorgehensweise Scheinkorrelationen im Vorhinein nicht gänzlich auszuschließen sind, empfiehlt es sich, aus einer großen Datenbasis mehrere disjunkte Stichproben zu ziehen, um die gewonnenen DM-Resultate zu validieren oder gegebenenfalls auch zu falsifizieren. Neben „klassischen“ Datengrundlagen (wie z.B. POS-Scannerdaten) können DM-Prozesse natürlich auch auf den mehrdimensionalen Datengrundlagen eines betrieblichen → Data Warehouse gestartet werden. Einschlägige Software hierfür ist durch kommerzielle Anbieter verfügbar, beispielsweise in Form des IBM-*Intelligent Miner* und des SAS-*Enterprise Miner*.
R.D./R.Wa.

Literatur: Fayyad, U.M.; Piatetsky-Shapiro, G.; Smyth, P.: From Data Mining to Knowledge Discovery: An Overview, in: Fayyad, U.M.; Piatetsky-Shapiro, G.; Smyth, P.; Uthurusamy, R. (Eds.): *Advances in Knowledge Discovery and Data Mining*, Cambridge 1996. Gaul, W.; Schader, M.: Data Mining: A New Label for an Old Problem?, in: Gaul, W.; Schader, M. (Hrsg.): *Mathematische Methoden der Wirtschaftswissenschaften*, Heidelberg 1999. Decker, R.; Wagner, R.; Knauff, N.: Data Warehouse – Ein Instrument für Marktforschung und Management?, in: *Marktforschung & Management*, Jg. 42 (1998), Nr. 5, S. 172–176.