

Data Mining im Controlling

– Methoden, Anwendungsfelder und Entwicklungsperspektiven –

Dr. Peter Gentsch

Sonderheft **Controlling Anwendungssysteme**
der Zeitschrift für Controlling und Management (ZfCM)

Autoreninformation:

Dr. Peter Gentsch ist Director CRM/ Analytics der *Business Intelligence Group* (Frankfurt, Berlin), einem führenden Dienstleister im Bereich Business Intelligence/ Data Mining. Zudem hat er Lehraufträge an der Universität München, Aalen und Berlin.

Data Mining stellt durch die Möglichkeit, automatisiert große Datenbeständen zu durchforsten und zu analysieren, ein leistungsstarkes Instrument für das Controlling dar. Das Unterstützungspotential liegt dabei insbesondere im Vertriebs-Controlling und in der Kundenperspektive der Balanced Scorecard. Durch das Aufzeigen von Ursache-Wirkungsbeziehungen, Vertriebs- und Kundenwertpotenzialen, Abwanderungswahrscheinlichkeiten sowie Cross- und Up-Selling-Potenzialen kann Data Mining einen substantiellen Beitrag sowohl für das operative als auch für das strategischen Controlling leisten. Die automatisierte Analyse von Texten durch das Text Mining sowie die Analyse von Internet-Daten durch das Web Mining stellen darüber hinaus innovative Ansätze insbesondere für das Controlling von Internettransaktionen dar.

Data Mining – Motivation und Methoden

„Our goal is to challenge the data to ask questions, rather than asking questions to the data “ (Keim/ Kriegel/ Seidel 1994, S. 302 ff.)

Das einführende Zitat bringt die grundlegende Intention des Data Mining zum Ausdruck. Im Sinne eines aktiven Analyseparadigmas sollen rechnergestützte Systeme den Anwender auf interessante, möglicherweise überraschende Muster und Strukturen, die sich in den Geschäftsdaten befinden, hinweisen. Damit sind also keine SQL- oder OLAP-Standardabfragen des Standard-Reportings und des Briefing Books gemeint. Treu nach dem Motto „Häufig sind die wichtigsten Dinge die, von denen wir gar nicht wissen, dass wir sie nicht wissen“ geht es beim Data Mining insbesondere um die Entdeckung früher Signale, die aufgrund des Geschäftsalltages und fehlender Frühwarnsysteme untergehen.

Die in der Literatur am häufigsten genutzte bzw. als Basis für ähnliche Definitionen fungierende Begriffsbestimmung für Data Mining stammt von Fayyad: „Data Mining is the non-trivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data“. (Fayyad/Piatetsky-Shapiro/Smyth 1996, S. 6).

Implizit liegt der zitierten Definition des Data Mining die Annahme strukturierter Daten zugrunde. Die Definition lässt sich jedoch prinzipiell auch auf unstrukturierte Daten wie z.B. Texte übertragen. Der Prozeß der Wissensentdeckung in unstrukturierten Datenbeständen wird in der Literatur unter dem Begriff des Text Mining diskutiert. Während Data Mining Muster und Strukturen in strukturierten Datenbeständen sucht, versucht Text Mining Muster und Strukturen in unstrukturierten Datenbeständen zu entdecken. Interessanterweise werden beide Konzepte – trotz starker Analogien bei den Methoden und den wissenschaftlichen Bezügen – in der Literatur weitgehend getrennt voneinander behandelt. In dem vorliegenden Artikel werden zusätzlich noch die Controlling-Potenziale des Web Mining untersucht. Die Datengrundlage hierbei sind strukturierte und unstrukturierte Daten aus dem Intra-, Extra- oder Internet.

Die Mustererkennung ist gemäss der zugrundeliegenden Definition nicht Ergebnis einer einzelnen Funktion, sondern eines Prozesses von der Selektion und Aufbereitung von Daten über das Generieren interessanter Beziehungsmuster

(eigentliches Mining) bis hin zur Ergebnis-Repräsentation und –Interpretation (Abbildung 1).

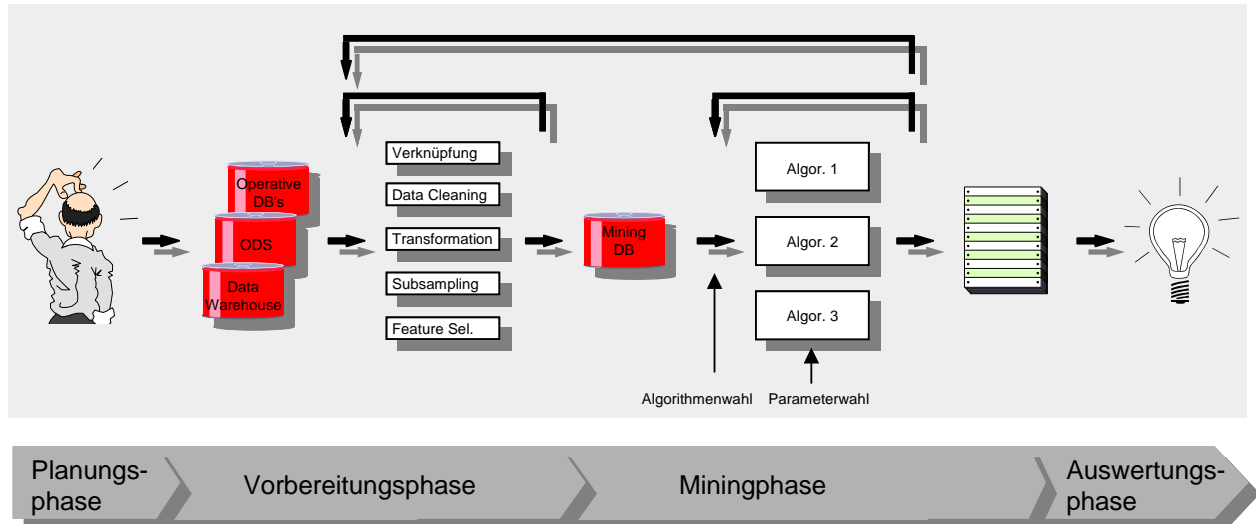


Abbildung 1: Data Mining als Prozess

Die einzelnen Phasen können wie folgt skizziert werden:

Planungsphase – In dieser Phase werden die konkrete Zielsetzung, die zu erwartenden Ergebnisse sowie die grobe Vorgehensweise des Data Mining festgelegt. Data Mining darf kein Selbstzweck sein, sondern muss problemorientiert auf die Bedürfnisse der Fachabteilung ausgerichtet sein.

Vorbereitungsphase – Die Durchführung von Data-Mining-Projekten erfordert, je nach der unternehmensspezifischen Datensituation (Anzahl der verschiedenen Datenbanksysteme, Verteilungsgrad der Daten, Anzahl und Art der Daten, Existenz eines gepflegten Datenmodells, Existenz eines Data Warehouse) eine unterschiedlich intensive Aufbereitung der Datengrundlage. In der Vorbereitungsphase werden die zu untersuchenden Datenbestände (bei sehr großen Datenmengen repräsentative Stichproben) selektiert, anschließend aus den verschiedenen Quellsystemen zusammengeführt und referenzielle und semantische Fehler bereinigt. Ebenso erfolgen hier je nach Erfordernissen der angewandten Methoden des Data Mining Reskalierungen und Datentransformationen. Ergebnis ist eine Mining Base, die als aufgabenbezogene,

aufbereitete und konsolidierte Datenbasis den Ausgangspunkt für die weiteren Analyseschritte darstellt.

Mining-Phase – In dieser Phase erfolgt die Anwendung des jeweiligen Data-Mining-Algorithmus auf den vorbereiteten Datenbestand und damit die eigentliche Suche nach interessanten Mustern. Dieser Schritt ist Kern des Prozesses und wird als das eigentliche Data Mining bezeichnet. Innerhalb der Mining-Phase erfolgt die Spezifikation des Algorithmus, dessen Parametrisierung sowie abschließend die eigentliche Anwendung des parametrisierten Algorithmus auf die vorbereiteten Daten.

Auswertungsphase – Gegenstand dieser Phase ist die inhaltliche Interpretation sowie die Aufbereitung der Ergebnisse. Die in der Mining-Phase generierten Ergebnisse werden auf statistische und sachlogische Validität und Adäquanz geprüft. Gegebenenfalls müssen die Modellparameter aufgrund der Resultate modifiziert und ein erneuter Durchlauf des gesamten Prozesses durchgeführt werden.

Damit ist der Data Mining-Prozess durch das Zusammenspiel von Mensch und Maschine gekennzeichnet: Während nur der Mensch die sachgerechte und aufgabenfokussierte Datenaufbereitung sowie die Interpretation der Ergebnisse vornehmen kann, kann der Rechner mit Hilfe intelligenter Algorithmen schnell große Datenbestände auf Auffälligkeiten hin untersuchen. Damit wird deutlich, dass Planung und Ausführung des gesamten Prozesses nur in bestimmten Grenzen automatisierbar sind.

In der betriebswirtschaftlichen Praxis und insbesondere im Controlling existieren unterschiedliche Frage- und Aufgabenstellungen, die durch die vier Kern-Methoden des Data Mining „Klassifikation“, „Segmentierung“, „Prognose“ und „Assoziation“ bearbeitet werden können. Für die Methoden können unterschiedliche Verfahren und Algorithmen eingesetzt werden (vgl. Grothe/Gentsch 2000, S. 179 ff.).

Im Folgenden werden die vier Kern-Methoden kurz skizziert:

Assoziationsanalyse: Assoziationsanalysen verfolgen im Rahmen des Data Mining die Zielsetzung, eigenständig Assoziationsregeln zu formulieren, die häufig auftretende - in den Datenbeständen versteckte - Regeln oder Muster beschreiben. Eine typische Anwendung der Assoziationsanalyse ist die Warenkorbanalyse. So lassen sich beispielsweise auf Basis von Kassentransaktionsdaten komplementäre Beziehungen zwischen einzelnen Artikeln identifizieren. Diese komplementären Artikelbeziehungen charakterisieren das Einkaufsverhalten der Konsumenten.

Klassifikation: Eine typische Fragestellung der Klassifikation beschäftigt sich damit, wie bestimmte Kunden charakteristisch beschrieben werden können, d.h. wie entsprechende Kundenprofile erstellt werden können. Auf den Bankensektor bezogen, kann exemplarisch die Beurteilung des Kreditrisikos neuer Bankkunden anhand ihres Profils genannt werden.

Prognose: Die Data Mining-Aufgabe „Prognose“ ist sehr ähnlich zur Klassifikation. Die Prognose erweitert die Klassifikation um eine zeitliche Komponente. Während es bei der Klassifikation nur darum geht, die heutige Klassenzugehörigkeit oder die heutige Ausprägung einer im Zeitablauf veränderlichen Zielgröße zu bestimmen, so geht es bei der Prognose darum, deren zukünftige Klassenzugehörigkeit oder zukünftige Ausprägung zu ermitteln.

Segmentierung: Die Segmentierung als Aufgabe des Data Mining kann den Anwender dergestalt unterstützen, daß sie „datengetrieben“ große Datenbestände, in kleinere, homogene und betriebswirtschaftlich zweckmäßige Teilmengen unterteilt. Die in den Clustern (Teilmengen) zusammengefaßten Datensätze teilen eine bestimmte Anzahl interessierender Eigenschaften.

Je nach Fragestellung kommen damit unterschiedliche Methoden des Data Mining in Betracht. Die Auswahl der für die jeweilige Fragestellung geeigneten Methode ist einer der zentralen Erfolgsfaktoren bei der Durchführung eines Data Mining-Projektes.

Im Folgenden wird basierend auf diesem Grundverständnis das Unterstützungs- und Verbesserungspotential des Data Mining zum einen für die klassischen Controllingfeldern wie der Bilanz & GuV, des Vertriebs-Controllings und der

Balanced Scorecard aufgezeigt. Zum anderen werden auch die neueren Herausforderungen für das strategische Controlling sowie das Controlling von Internetaktivitäten diskutiert.

Data Mining im Controlling

Zu den Prämissen, die den Einsatz von Data Mining erfolgreich versprechen lassen, gehört das Vorhandensein großer digitaler Datenbestände, die aufgrund ihrer internen Struktur ein gewisse Komplexität aufweisen, die es verhindert, Muster und Strukturen mit dem bloßen Auge zu erfassen. Die Prämisse grosser Datenmengen ist zumindest für das Controlling großer Unternehmen gewährleistet. Sofern erscheint der Einsatz für das Data Mining im Controlling zunächst grundsätzlich möglich. Abbildung 2 illustriert das grundsätzliche Anwendungspotenzial von Data Mining im Controlling (vgl. Küppers 1998).

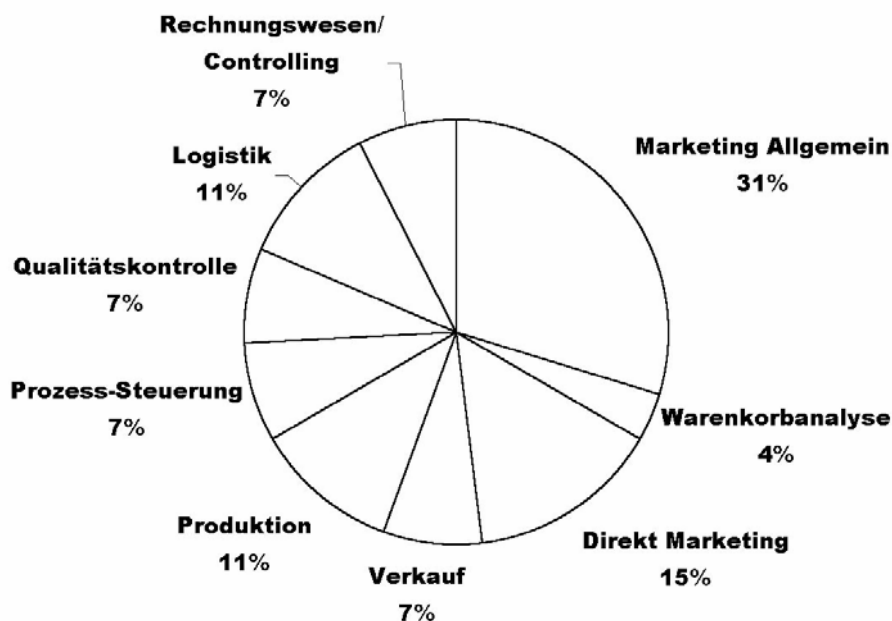


Abbildung 2: Anwendungsgebiete des Data Mining nach Funktionen

Für eine differenzierte Aussage ist jedoch eine genaue Unterscheidung des Controlling-Begriffs notwendig. Betrachtet man im klassischen Controlling die Datenbasis und Strukturen von GuV und Bilanz sind jedoch die zugrundeliegenden Strukturen i.d.R. übersichtlich und bekannt. Für die zentralen Fragestellungen des klassischen Controlling sind die entsprechenden Abfragen

und Hypothesen bekannt und „Schätze“, die überraschend durch Data Mining geborgen werden könnten, damit eher unwahrscheinlich. Anders sieht dies für das Vertriebs-Controlling und für das Controlling von Kundenbeziehungen aus (z.B. die Bewertung von Kundenzufriedenheit und Kundenbindung der Kunden-Perspektive der Balanced Scorecard).

Betrachtet man beispielsweise ein mittelständisches Unternehmen mit 14 Produkthauptgruppen, 4 Erzeugnisbereichen, 126 Kundengruppen, 12 Branchen 4 Erzeugnisbereiche, 11 Herstellern, 5 Vertriebsbereichen, 3 Vertretergruppen sowie 2.080 Sorten, der seine Daten- und Lösungsraum in einem – im Controlling zunehmend Einzug erhaltenden – OLAP-System abbildet, so ergeben sich hier bereits 2.384 Einstiegspunkte für entsprechende OLAP-Abfragen. Bedenkt man nun zusätzlich die kombinatorischen Möglichkeiten, die einzelnen Datenpunkte des OLAP-Systems mit einander zu verknüpfen, wird schnell die Komplexität eines entsprechenden Vertriebscontrolling insgesamt deutlich. So ist es durchaus denkbar, dass einem Standard-Reporting ein Umsatzrückgang eines Produktes einer Region, der durch ein ganz bestimmtes Zusammenspiel von Erzeugnisbereich, Vertretergruppe, Kundengruppe sowie zeitlichem und regionalen Bezug zu erklären ist, verborgen bleibt. In dem komplexen Geflecht potentiell interessanter mehr-elementigen Datenbezüge, können erkenntnisbringende Ursache-Wirkungs-Beziehungen aufgrund fehlender Transparenz untergehen.

Neben dem Vertriebs-Controlling weist die **Kundenperspektive** der Balanced Scorecard das grösste Potenzial für das Data Mining auf. Gegenstand der Kundenperspektive ist die Identifizierung der Kunden- und Marktsegmente, in denen die Unternehmung konkurrenzfähig sein will. Innerhalb der Kundenperspektive wird die Unternehmensstrategie durch die Entwicklung von spezifischen strategischen Zielen auf die als strategische bedeutsam identifizierten Kunden- und Marktsegmente übertragen. Die Marktsegmentierung ist daher der Ausgangspunkt für weitere Überlegungen innerhalb der Kundenperspektive. Die Segmentierung des Kundenstamms kann im Rahmen des unüberwachten Lernens durch den Einsatz der Data Mining-Methode der Segmentierung erfolgen. Das Ergebnis ist die Segmentierung der Kunden in Gruppen, die in sich möglichst homogen und untereinander möglichst heterogen sind. Zur Segmentierung sollte eine möglichst differenzierte Datenbasis gewählt

werden, die neben Transaktions- und Vertragsdaten insbesondere auch soziodemographische, mikrogeografische und psychographische Daten besitzen sollte. Der zweite Schritt konzentriert sich auf die Beschreibung der identifizierten Segmente. Die ermittelten Kundensegmente dienen nun als Ausgangspunkt für weitere Analysen. So können z.B. durch Assoziationsanalysen Regeln gebildet werden, die das Kaufverhalten einer Kundengruppe widerspiegeln. Sequenzanalysen erlauben darüber hinaus die Entdeckung von zeitlichen Kaufmustern. Insgesamt stellen die ermittelten Kundensegmente mit der spezifischen Segmentbeschreibungen die Basis für die Ermittlung der strategischen Ziele sowie die Kern-Kennzahlen der Kundenperspektive. Hierfür ist die von Kaplan und Norton vorgeschlagene Vorgehensweise einsetzbar (vgl. Kaplan/ Norton 1997, S. 66 ff).

Neben den einfachen Kennzahlen der Kundenperspektive wie Kundenanzahl oder Anzahl der Reklamationen/ Garantiefälle sind im CRM-Controlling insbesondere auch Phänomene wie Kundenbindung und –zufriedenheit zu messen und zu steuern. Die Operationalisierung kann durch so genannte Kundenwert-Modelle erfolgen.

Betrachtet man komplexere Kundenwert-Modelle, wie sie gerade gegenwärtig im Rahmen des CRM-Controlling diskutiert werden, wird schnell deutlich, dass diese nicht ohne entsprechendes Data Mining zu realisieren sind. Sollen aussagekräftige Kundenwerte ermittelt werden, die insbesondere auch das Kundenpotenzial reflektieren, reichen die klassische DB-Modelle nicht mehr aus. Es werden komplexere Kundenwertkonstrukte benötigt, die neben retrospektiven insbesondere auch prospektive Elemente wie Abwanderungswahrscheinlichkeit, Schadenswahrscheinlichkeit, Cross- und Up-Selling-Potenziale oder das Weiterempfehlungspotential des Kunden berücksichtigen: Klassifikationsverfahren ermitteln typische Multiplikatoren, Assoziationsverfahren zeigen die Cross-Selling-Potenziale auf, durch Prognoseverfahren werden Abwanderungs- und Schadenswahrscheinlichkeiten ermittelt.

Auch innerhalb der **Prozessperspektive** lässt sich das Data Mining sinnvoll einsetzen. So lassen sich Produktionsengpässe und Qualitätsprobleme von Fertigungsprozessen prognostizieren, besonders zuverlässige Lieferanten profilieren oder die Ursachen für Maschinen-Ausfallzeiten analysieren und erklären.

Auch die Perspektive „**Lernen und Innovation**“ bietet Anwendungspotential für das Data Mining. Diese Perspektive sichert die Grundlage für die Erreichung der Ziele der anderen drei Perspektiven, indem sie die dafür notwendige Basis und Infrastruktur schafft. Die wesentlichen Treiber zur Erreichung einer adäquaten Infrastruktur leiten sich aus den Personalpotentialen, der technologischen Infrastruktur und dem Arbeitsklima ab.

Einsatzpotentiale für das Data Mining innerhalb dieser Perspektive bietet z.B. die Analyse des Informations- und Kommunikationsverhaltens der Mitarbeiter im Hinblick auf die Nutzung der unternehmensinternen Informationssysteme. Basis für das Data Mining bzw. Web Mining stellen die automatisch erstellten Protokolldateien der Intranet-Applikationen und -Servern dar. Auf Basis von Nutzungssegmenten, Navigationspfaden oder Suchprofilen lassen sich konkrete Initiativen und Maßnahmen ableiten, um das Informationsverhalten der Mitarbeiter zu optimieren, eine Neuplatzierung und Gruppierung von Informationsinhalten vorzunehmen oder die Schwerpunkte der Informationsversorgung anzupassen.

Auch innerhalb der übergeordneten **Finanzperspektive** lassen sich Data Mining-Methoden anwenden. Da jede Balanced Scorecard-Kennzahl Teil einer Kette von Ursache und Wirkung ist, fließen in der Finanzperspektive die Ergebnisse aller Perspektiven zusammen. Die Kennzahlen der Finanzperspektive sind daher ein massgeblicher Indikator für Veränderungen in den zugrundeliegenden Perspektiven.

Ansatzpunkte für das Data Mining ergeben sich in der Finanzperspektive vor allem in der Ursachenanalyse von signifikanten Abweichungen der Kennzahlen von Normwerten oder Planwerten.

Darüber hinaus lässt sich das Data Mining auch **Perspektiven-übergreifend** anwenden. Die perspektivenübergreifende Verbindung der Balanced Scorecard-Kennzahlen durch eine Kausal-Kette von Ursache-Wirkungsbeziehungen beschreibt die Unternehmensstrategie in ihrer Wirkungsrichtung. Die Identifikation der Ursache-Wirkungsbeziehung zwischen den Kennzahlen ist daher einer der erfolgskritischen Punkte bei der Implementierung der Balanced Scorecard. Kaplan und Norton schlagen als Methode zur Identifikation dieser Kausalbeziehungen die Generierung von Hypothesen innerhalb von Managementteams vor (vgl. Kaplan/Norton 1997, S. 246). Die Verifikation der so ermittelten Hypothesen kann durch Korrelationsanalysen erfolgen. Neben der Ermittlung der Kennzahlenbeziehung

lässt sich so auch die Beziehungsstärke der Kennzahlen zueinander quantifizieren.

Diese Verfahren erlauben jedoch keine Ableitung der Reaktionszeiten zwischen den Veränderungen der Kennzahlen. Auf Basis der Data Mining-Methode der Sequenzanalyse können jedoch diese Reaktionszeiten gemessen werden. Untersucht wird dabei der zeitliche Verlauf eines Kennzahlpaares und die Identifizierung von einander ähnlichen Teilsequenzen.

Insbesondere vor dem Hintergrund, dass ein Grossteil der Unternehmensdaten in unstrukturierter Form vorliegt, müssen neben den Einsatzmöglichkeiten des Data Mining auch die Unterstützungs- und Verbesserungspotentiale des Text Mining für das Controlling untersucht werden. Während Data Mining i.d.R. mehr auf interne Datenbestände ausgerichtet ist, sind beim Text Mining auch insbesondere externe Informationsquellen von Interesse. Die zunehmende Verfügbarkeit von Wirtschaftsinformationen im Internet macht dieses Medium als externe Mining Base für das Text Mining interessant. In den kommerziellen und freien Internet-Seiten finden sich Informationen zur Technologie- und Wettbewerbssituation, zu Produkten, Dienstleistungen, Vorankündigungen oder Unternehmensverflechtungen. Im Gegensatz zu den kommerziellen besitzen freie Seiten i.d.R. keine thematisch zusammengefassten Informationen zur Technologie- und Wettbewerbssituation, sondern meistens Einzel-Informationen, die erst noch entsprechend verdichtet werden müssen. Durch Text Mining lassen sich automatisiert Technologie- und Markt-Cluster bilden (sowie die Verbindungen dieser Cluster und ihre Veränderung im Zeitablauf deutlich machen), so dass Unternehmen schnell Überblick über neue oder bereits bestehende Märkte, Produkte und Konkurrenten gewinnen können.

Das Text Mining kann in diesem Verständnis zum Aufbau eines qualitativen Controlling-Radars genutzt werden, das frühe und „leise“ Indikatoren erkennt und verfügbar macht. Insbesondere im Rahmen des sogenannten Competitive Intelligence werden für wissensbasierte Wettbewerbs- und Marktanalysen immer stärker Daten- und Wissensbestände relevant, die aufgrund ihres geringen Formalisierungsgrades nicht bzw. nur unzureichend von den traditionellen Datenbanksystemen erfasst werden.

Data Mining für das Internet-Controlling

*„Beyond merely creating customer profiles and performance metrics, clickstream analysis and customer intelligence solutions **will become critical** in determining how to allocate costly sales/support channels to the most profitable customer segments.“* (Einschätzung der META Group, 2002)

Die wachsende Bedeutung des Internets und der zunehmende kommerzielle Einsatz von Web Sites geht einher mit erhöhten Investitionen in diesem Bereich. Insbesondere in jüngerer Zeit versuchen Unternehmen effiziente Methoden zu finden, um den Erfolg ihrer Web Site zu messen. Dies macht den Einsatz einer Erfolgskontrolle von Web Sites notwendig.

Das Internet bietet dem Controller den entscheidenden Vorteil Interaktionsdaten (Informations-, Kommunikations- und Transaktionsdaten) in Echtzeit zu erfassen und den Kunden bei seinen Handlungen genau zu beobachten. Auf der einen Seite entstehen bei Internet-basierten Interaktionen damit neue Möglichkeiten, automatisiert umfangreiche und aussagekräftige Daten für die Kennzahlenbildung zur Verfügung zu stellen. Auf der anderen Seite müssen für das Web Controlling relevante Metriken zur Erfolgsmessung von Internet-Präsenzen definiert werden.

Bei den meisten Unternehmen fehlt derzeit ein umfassendes, bedarfsorientiertes und betriebswirtschaftlich ausgerichtetes Internet-Controlling, das für die systematische und zielgruppen-orientierte Gestaltung und kontinuierliche Optimierung des Internet notwendig wäre. Die gegenwärtige Auswertung der Logfiles wird im überwiegenden Maße zur technischen Überwachung genutzt. Im Rahmen dieser Analysen werden lediglich allgemeine Besucherzahlen ausgewertet, welche aber keine Analysen des Nutzungsverhaltens und der Benutzerakzeptanz zulassen. So ist zwar i.d.R. bekannt, wieviele unterschiedliche IP-Adressen das Internet genutzt haben und wie viele Page Impressions dabei erzeugt wurden, es kann aber hinsichtlich der Erreichung der Marketing- und Kommunikationsziele keine Aussage getroffen werden. Daraus ergibt sich, dass Maßnahmen zur Verbesserung und zur Gestaltung der Inhalte des Internet-Präsenz nicht im ausreichenden Maße ergriffen werden können.

Dies ist neben den rudimentären Auswertungstools (Webalyzer und WebSuccess sind Logfile-Analyse-Tools, die in der betrieblichen Praxis häufig für das Internet-

Controlling eingesetzt werden) insbesondere in der fehlenden Datenbasis für ein aussagekräftiges Controlling begründet. Auf der Basis der derzeitigen Reports lässt sich i.d.R. nicht fundiert erkennen, welche Content-Bereiche verbessert werden sollten, ob eine geringe Content-Nutzung inhaltliche, technische oder navigationsbedingte (Usability) Ursachen hat oder ob bestimmte Usertypen besondere Navigationspfade und Suchabfragen nutzen.

Neben dem Controlling des Content-basierten Internet ist insbesondere für die Internet-Applikationen wie z.B. „Buchungssysteme“, „Bestellsysteme“ oder „eLearning“-Anwendungen ein differenziertes Nutzungsbild wichtig. Nur so lassen sich diese Applikationen betriebswirtschaftlich operationalisieren und optimieren.

Beim Internet-Controlling sollte nicht vom „prinzipiell Messbaren“ ausgegangen werden. Das **richtige Kennzahlensystem** kann nur von den **Internet-Zielen** des jeweiligen Unternehmens abgeleitet werden. Dies gilt insbesondere auch für die Gewinnung bzw. Verdichtung von Kennzahlen durch das Data Mining.

Die Definition der Kennzahlen kann des Weiteren nicht ohne das technisch Machbare (bzw. aufwandsmäßig vertretbare) Daten-Tracking erfolgen. Bei den verschiedenen Datenerfassungsverfahren (Tracking-Verfahren: Prinzipiell stehen zur Datenerfassung folgende Verfahren zur Verfügung: Applikation Monitoring, Logfile-Analyse, Zählpixel-Verfahren, Cookies, Revers-Proxy, URL-Rewriting. Ausführlich hierzu: Gentsch et al. 2001), welche die notwendige Datenbasis für die jeweiligen Kennzahlen sicherstellen, sind damit insbesondere die **informationstechnologischen Rahmenbedingung** bzw. auch **Restriktionen** zu berücksichtigen. Insgesamt gilt abzuwägen, welche Kennzahlen mit welchem Aufwand der korrespondierenden Datenerfassungstechnologie erhoben werden sollen. Generell gilt, je umfangreicher und detaillierter das Daten-Tracking ist, umso umfassendere Kennzahlen lassen sich abbilden. Insbesondere erfordern die auf Basis des Data bzw. Web Mining ermittelten Kennzahlen eine besondere Datentiefe. Der optimale Mix aus Kennzahlensystem und Tracking-Aufwand lässt sich nur anhand des konkreten Zielsystems des Unternehmens klären.

Für den zielorientierten und strukturierten Aufbau des Internet-Kennzahlensystems empfiehlt sich eine Einteilung und sukzessive Entwicklung gemäss der Kennzahlenebenen **Basis-Kennzahlen**, **erweiterte Kennzahlen** und **Mining-basierte Kennzahlen**. Abbildung 3 zeigt für die einzelnen Kennzahlen-Ebenen exemplarisch einige Kennzahlen.

| Kennzahlenebene | Aussagefähigkeit/ Pre- & Post-Processing | Kennzahl | Definition |
|-------------------|--|---|--|
| Basics | | Hits | $H = S$ Zeilen Logfile |
| | | Page Impressions | $PI = \sum H \mid H = \text{html-Dokument oder ähnliches}$ |
| Advanced | | Stickiness ("Klebrigkeit") | $ST = F * D * TSR \rightarrow$ $ST = SL / \text{Besucher}_{\text{unique}}$ |
| | | Focus | $F = \text{durchschnittlich Besuchte Anzahl an Seiten pro Bereich} / \text{Anzahl der Seiten in diesem Bereich}$ |
| Data Mining-based | | Trail, Kundencluster (<i>Trampelpfade in Intranet-Bereichen. Welchen Weg zum einem bestimmten Content-Angebot gehen welche Usertypen?)</i>) | <i>Zeitliche Assoziationen zwischen einzelnen Intranet-Seiten bzw. -Bereichen (Sequenz-Analysen auf Basis von Data Mining)</i> |

Abbildung 3: Kennzahlenebenen für das Internet mit Beispielen

Die auf Basis des Data Mining ermittelten Kennzahlen (Ebene 3) weisen die grösste Informations- bzw. Erklärungstiefe auf. So geben diese Kennzahlen darüber Aufschluss, welche Kundengruppen welche Präferenzen haben, welche Navigationspfade signifikant zur Konversion führen, ob es Trampelpfade bestimmter Kundentypen gibt oder welche Seiten und Produkte häufig zusammen angeschaut bzw. gekauft werden. Analog zur Balanced Scorecard kann hier von der Entwicklung der *Web Scorecard* als Management-Instrument gesprochen werden. Das diese Form des Internet-Contolling wesentlich stärker die betriebswirtschaftliche Aspekte wie Profitabilität einer Internet-Seite adressiert, als dies die herkömmlich technisch-orientierten Logfile-Analysen tun, soll durch das folgende Praxisbeispiel eines großen deutschen E-Shops illustriert werden. Ziel des zugrunde liegenden Projektes war es, das Kaufverhalten der Kunden in dem E-Shop transparent zu machen: Welcher Kundentypus kauft welche Waren? Welche Waren werden zusammengekauft? Welche Waren werden in einem

bestimmten Zeitmuster nachgefragt? Welche Seiten besuchen Kunden typischer Weise vor und nach einem Kauf?

Abbildung 4 zeigt zunächst exemplarisch, wie auf Basis des unüberwachten Clusterings differenziertes Wissen über die User und Kunden aufgebaut werden kann. So gibt es im betrachteten Shop-Beispiel zum einen „*Business-Cluster*“, der 11,24% des gesamten Kundenbestandes ausmacht. Kunden dieses Segments geben im Vergleich zur Grundgesamtheit signifikant mehr im Shop aus, kommunizieren weniger, haben überdurchschnittlich viele Seitenabrufe und sind älter als der Durchschnitt. Im Gegensatz zu diesem Männer-dominierten Kundensegment gibt es ein auffällig großes Segment vornehmlich jüngerer weiblicher Kundschaft, die sich weniger auf Kaufseiten sondern vielmehr auf unterhaltsamen Content-Seiten befindet und demnach auch einen geringen Umsatz erbringt („*Fun-Cluster*“). Insgesamt wurden sechs aussagekräftige Kundensegmente gefunden, die als Basis für die Definition kundenorientierter Kennzahlen und Ziele dienen.

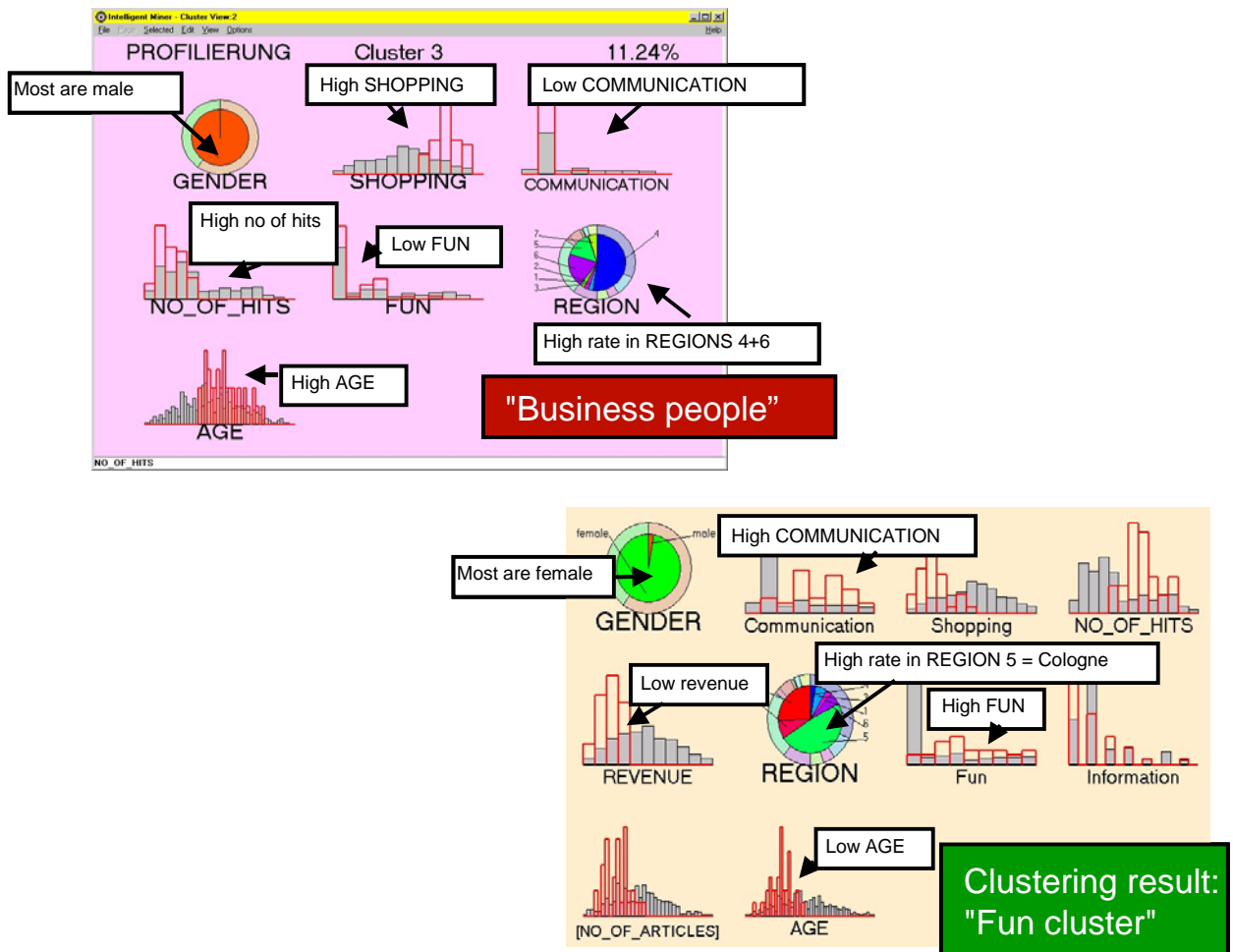


Abbildung 4: Kundensegmente, die auf Basis von Data Mining entwickelt wurden

So lässt sich nun auf Basis dieser Kundentypologie weiteres Controlling-relevantes Wissen entwickeln. Im Folgenden soll exemplarisch die Internet-Profitabilitäts-relevante Frage beantwortet werden, welche Seiten Kunden typischer Weise vor und nach einem Kauf besuchen? Auf Basis der automatisch fortgeschriebenen Navigationsdaten wurde mit Hilfe von Sequenzanalysen das zeitliche Verhalten von Kaufprozessen analysiert. Abbildung 5 zeigt als beispielhaftes Ergebnis signifikante Navigationspfade innerhalb des Online-Shops. Es handelt sich dabei um eine sequentielle Folge von Besuchen (Click Streams) innerhalb des Shops, die besonders auffällig sind. Durch die Analysen wurden so die assoziativen Folgen mit „buy.html“, die also einen konkreten Kauf mit einschließen, transparent (vgl. Schommer/Müller 2001).

| Support | Itemsets |
|---------|---|
| 5.172 | [womens-fashion.html] [buy.html] [postcards.html] [chat.html] |
| 5.172 | [buy.html] [womens-fashion.html] [chat.html] [travel.html] |
| | [buy.html] [womens-fashion.html] [chat.html] [mens-fashion.html] |
| | [buy.html] |

Abbildung 5: Auffällige Navigationspfade in einem Online-Shop

Interessant ist nun die Interpretation der gefundenen Assoziationen: so interessieren sich 5.172% der User zunächst für verschiedene Artikel in Bereich „womens-fashion.html“, tätigen dann die eigentliche Kauftransaktion (buy.html), verschicken darauf hin eine virtuelle Postkarte (postcard.html) und beginnen dann zu chatten (chat.html). Diese Resultate sprechen dafür, dass die Shop-Besucher zunächst einkaufen und dann ihrem Kommunikationsbedürfnis durch das Versenden von Postkarten und dem Chatten in der Online-Community nachgehen.

Weitere 5.172% der Shop-Besucher kaufen zunächst ein (buy.html), besuchen dann den Modebereich (womens-fashion.html) und steuern nach dem Chatten (chat.html) den Reisebereich (travel.html) bzw. den Männermodebereich (mens-fashion.html) an. Die Kaufabsicht scheint bei diesen Shop-Besuchern im Mittelpunkt ihres Besuches zu stehen. Die informativen Bereiche werden erst nach getätigtem Kauf aufgesucht.

Auf Basis dieses differenzierten Wissens der Kundenperspektive wurden nun verschiedene Maßnahmen zur gezielteren Kundenansprache und Verkaufsförderung diskutiert. Eine Möglichkeit besteht darin, das Angebote je nach Kundentypus zu personalisieren oder entsprechende Restrukturierungen des Shops vorzunehmen.

Durch das Internet-basierte CRM-Controlling können insbesondere auch Aussagen über die Profitabilität von Usern und Kunden gemacht werden. Sie können z.B. Kundencluster in einem E-Shop deutlich machen, dass es häufig Mädchen und junge Frauen im Alter von 15-25 Jahren aus den Großräumen Hamburg und Mannheim sind, die zwar häufig den Shop besuchen und auch dann und wann mit geringem Volumen kaufen, aber hauptsächlich E-Cards abrufen, chatten und sich Reiseseiten ansehen und damit noch nicht zu den profitablen Kunden zu zählen sind. Eine Möglichkeit dieses generierte Wissen umzusetzen, könnte in dem Angebot von Low-Budget-Reisen auf den Chat- und Ecardseiten bestehen, das junges weibliches Publikum adressiert.

Data Mining im Controlling – eine kritische Einschätzung

Der vorliegende Artikel hat insbesondere die Anwendungspotentiale des Data Mining für das Controlling aufgezeigt. Der erfolgreiche Einsatz des Data Mining in der betrieblichen Praxis ist jedoch nicht unumstritten. Im Sinne einer kritischen Auseinandersetzung mit dem Thema wird folgend eine kurze Darstellung der geäußerten Kritikpunkte mit einer entsprechenden Bewertung bzw. auch „Entkräftung“ vorgenommen.

Zum einen kommt die Kritik aus der klassischen Statistik, die den hypothesenfreien Ansatz des Data Mining problematisiert. Während bei der konventionellen Statistik die Analyse-Fragestellung vollständig modelliert wird (abhängige, unabhängige Variablen, Signifikanzniveau, Verteilung), kann Data Mining (insbesondere die Methoden des unüberwachten Lernens: Ausführlich hierzu Grothe/Gentsch 2000, S. 179 ff.) als ein Prozeß der Modellbildung verstanden werden, d.h. die abgeleiteten Informationen stellen sich als Modell hinter den zugrundeliegenden Daten dar. Dies ist insbesondere dann notwendig, wenn noch kein Modellverständnis vorliegt. Der Fall tritt besonders dann auf, wenn eine unübersichtlich große Anzahl von potentiell erklärungskräftigen Einflussgrößen existiert und keine klaren Ursache-Wirkungs-Beziehungen bekannt sind.

Data Mining stellt damit einen zur klassischen Statistik komplementären Ansatz dar. Der Anspruch des Data Mining, Modell-bildend zu sein, muss jedoch auch relativiert werden. Werden die Ergebnisse auf Basis nicht korrekter Datenmodelle durchgeführt, können diese statistisch korrekt sein, inhaltlich jedoch zu falschen

Rückschlüssen führen. Es kann im Rahmen eines zugrunde liegenden Datenmodells weitgehend ungerichtet und automatisiert vorher nicht bekannte Strukturen und Muster in den – dem Datenmodell entsprechenden – Daten entdecken. Data Mining hilft damit, die extensionale Sicht der Daten besser zu verstehen. Die intensionale Sicht in Form des Datenmodells (z.B. Entity-Relationship-Modell, Star-Schema oder OLAP-Modell) muß jedoch bereits existieren. Die intensionale Ebene entspricht der Typenebene eines Datenmodells: z.B. Produkte werden von Maschinen gefertigt und deren Inputfaktoren von Lieferanten zur Verfügung gestellt. Beim Data Mining geht es um die Mustererkennung auf der Ausprägungsebene (extensionale Ebene: Ausprägung der Informationsobjekten samt ihrer Attribute): z.B. Welche Rohstoffe führen bei welchem Fertigungsverfahren auf welcher Maschine bei welcher Rahmenbindung mit welcher Wahrscheinlichkeit zu einer signifikant hohen Ausschussquote? Ebenso muss ein Verständnis über die potentiellen Einflussgrößen und Störgrößen auf die zu untersuchende Zielgröße vorhanden sein. Damit muss bereits ein grobes Erklärungsmodell für die Data Mining-Analyse vorliegen. Ebenso ist der Anspruch der ungerichteten, daten-getriebenen Analyse zu relativieren. Durch die Selektion der relevanten Datenbestände und Vorgabe der Zielvariable erfolgt eine benutzer-getriebene Spezifikation bzw. Einschränkung des potentiellen Lösungsraumes.

Ein anderer Kritikpunkt richtet sich auf die Problematik der rechnerbasierten Strategieableitung. In diesem Zusammenhang kann nur ausdrücklich darauf hingewiesen werden, dass die Umsetzung bzw. Kapitalisierung entsprechender Data Mining-Erkenntnisse nur nach einer fundierten kritischen Überprüfung auf Basis des jeweiligen Unternehmens- und Domänenwissens erfolgen darf. So liess sich zum Beispiel in der Vergangenheit ein signifikanter Zusammenhang zwischen dem Rückgang an Storchen und dem Rückgang der Geburtenrate von Babys feststellen. Das Beispiel macht deutlich, dass eine kritische inhaltliche Reflektion von Korrelation und Kausalität notwendig ist.

Eine anderer Problembereich bezieht sich auf die Attribute „novel“ und „potentially useful“ der Data Mining-Definition. Neben interessanten Mustern werden im Rahmen des Data Mining i.d.R. auch verschiedene triviale oder redundante Regeln als "ungewollte Nebenprodukte" generiert. Sind beispielsweise die Erlöse und variablen Kosten eines Produktes bekannt, ist der Deckungsbeitrag des

Produktes eine redundante Größe. Wenn nun eine Regel zwischen einem dieser redundanten Attribute und der Zielgröße entdeckt wird, generieren Data Mining-Tools i.d.R. gleichzeitig eine weitere Regel, die sich auf die Beziehung der Zielgröße mit dem anderen redundanten Attribut bezieht. Diese weitere Regel bringt jedoch aufgrund ihrer Redundanz keinen zusätzlichen Informationsgewinn und erschwert es, in der Vielzahl generierter Regeln die wirklich neuen zu entdecken. Bezieht sich beispielsweise eine Regel auf sämtliche Artikel einer Artikelgruppe, so reicht es vom Informationswert her aus, wenn die Regel für die Artikelgruppe formuliert ist und nicht zusätzlich redundant auf die einzelnen Artikel bezogen wird.

Neben den redundanten Regeln und Mustern werden zudem i.d.R. auch triviale Ergebnisse generiert. So ist beispielsweise die „Entdeckung“ trivial, dass bei der Attributsausprägung „Konto geschlossen“ das Attribut „Umsatz“ zu einem auffällig hohen Prozentsatz die Ausprägung „sehr gering (bzw. 0)“ aufweist. Um diese Trivialitäten möglichst früh herauszufiltern, besteht die Möglichkeit, sich nur auf Veränderungen in Datenstrukturen zu konzentrieren. Da die trivialen Regeln i.d.R. zeitlich stabil sind, können sie auf diese Weise unterdrückt werden. Eine andere Möglichkeit besteht darin, das Data Mining auf einen „Normaldatensatz“ aufzusetzen, der keine wirklich interessanten Regeln enthält. Die in diesem Datensatz entdeckten Muster bilden die „Normalität“ ab und können daher bei zukünftigen Data Mining-Einsätzen schon im Ansatz unterdrückt werden (vgl. Bissantz/Hagedorn 1996)

Für ein Data Mining-Tool kann es demnach keine inhaltlich signifikanten, sondern grundsätzlich nur statistisch signifikante Muster geben. Auch wenn die redundanten und trivialen Regeln das Ergebnis anscheinend unnötig aufblähen, zeigen diese doch mit ihrer statistischen Signifikanz, daß die Tools automatisiert statistisch „richtig“ arbeiten können und tragen damit insgesamt zur Ergebnisvalidierung bei. Zudem sind solche Ergebnisse auch aus Sicht der Sicherung und Erhöhung der Datenqualität interessant. In einem durchgeführten Projekt wurden z.B. signifikante Soll-Ist-Abweichungen entdeckt, die sich nach genauerem Nachfragen als Datenschmutz erklären ließen: So wurde häufig von den Mitarbeitern der Soll-Wert nicht sauber in dem System eingegeben und gepflegt. Das System hat diese fehlenden Werte *default-mässig* auf „Null“ gesetzt, so dass durch das Data Mining für diese Fälle signifikante Soll-Ist-Abweichungen

entdeckt wurden. Auch wenn solche Ergebnisse nicht inhaltlich interessant sind, helfen sie doch Schwachstellen bei der – für das Data Mining so wichtigen – Phase der Dateneingabe und -pflege zu entdecken: „It all begins with data“.

Fazit und Ausblick

Der vorliegende Beitrag hat gezeigt, dass Data Mining in ausgewählten Bereichen des Controlling ein deutliches Unterstützungs- und Verbesserungspotential besitzt. Neben dem Einsatz im Vertriebs-Controlling liegt das Potential insbesondere in der Unterstützung der Konzepte der Balanced Scorecard respektive der Web Scorecard. Data Mining-Methoden können hier unter anderem im Prozeß der Zielfindung und im Prozess der Identifizierung von Maßnahmen zur Zielerreichung helfen. Des Weiteren können Ergebnisse von Data Mining-Analysen dazu verwendet werden, um Rückschlüsse über die Effektivität einzelner Balanced Scorecard-Initiativen hinsichtlich ihres Erfolgspotentials abzuleiten. Um Data Mining erfolgreich im Controlling einzuführen, müssen die Fallstricke und Restriktionen des Data Mining kritische Berücksichtigung finden. Insgesamt schränken diese jedoch nicht das prinzipielle Potential des Data Mining als rationalitätssicherndes und komplexitätsreduzierendes Instrument für das Controlling ein.

Im Gegensatz zur traditionellen, durch den Benutzer gesteuerten Daten- und Dokumentenanalyse, wird beim Data und Text Mining daten- bzw. dokumentengetrieben nach interessanten Mustern und Strukturen gesucht. Der Vorteil dieser automatisierten, weitgehend ungerichteten Analyse besteht darin, dass der Rechner schnell große Mengen von Daten und Dokumenten unvoreingenommen verarbeiten kann. Damit steigt die Wahrscheinlichkeit, auch unerwartete, überraschende Signale wahrzunehmen, die bei einer konventionellen, manuellen Analyse möglicherweise a priori ausgeschlossen worden wären. Durch den relativ hohen Automatisierungsgrad des Data und Text Mining können zudem zeitliche und personelle Einsparungspotentiale entstehen.

Insgesamt muss der insbesondere durch die populärwissenschaftliche Literatur geprägte Mythos, dass der Computer ohne Interaktion des Menschen völlig autonom Wissen generieren kann, sowohl für das Data Mining wie auch für das Text Mining klar entkräftet werden. Dass der Faktor Mensch bei dieser Art der

Wissensentwicklung eine tragende Rolle spielt, wird durch die Tatsache deutlich, dass nur er die Interpretation und Plausibilitätsprüfung übernehmen kann. Es bedarf der menschlichen Erfahrung und Intelligenz, um letztlich das wirklich relevante Wissen aus digitalen Datenbeständen heraus zu entwickeln. Der vollständige Transformationsprozess, der aus Daten handlungsorientiertes Wissen entstehen lässt, ist ohne die fachspezifische und intellektuelle Interpretation des Menschen nicht möglich.

Im Sinne integrierter Business Intelligence-Lösungen werden Softwareanbieter zunehmend eine verbesserte informationstechnologische Integration von Datenbank/ Data Warehouse, OLAP-System und Data Mining-Applikation anbieten (Abbildung 6).

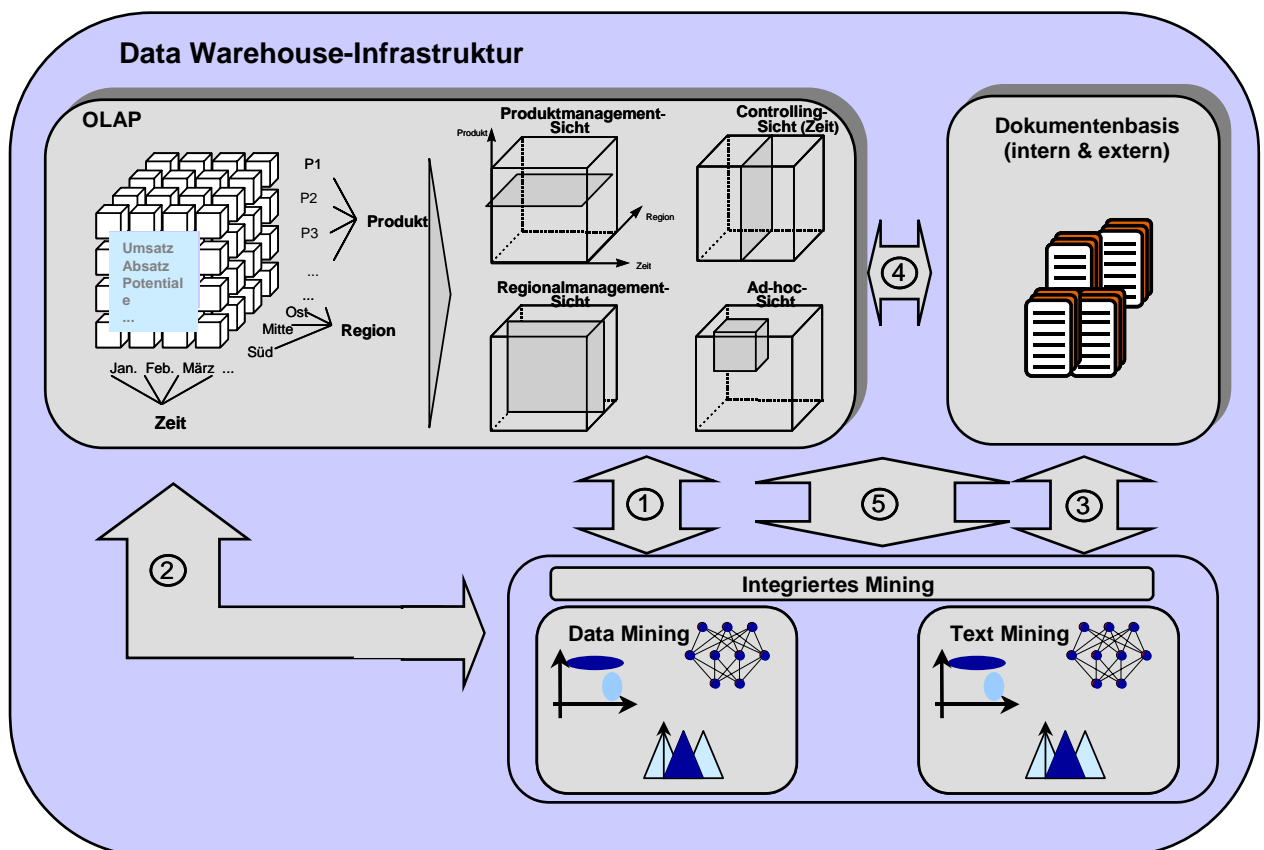


Abbildung 6: Zusammenspiel von OLAP und DWH

Lösungsräume, die mit entsprechenden „Drill Down-“ und „Slice & Dice“-Operationen erzeugt wurden, können direkt in das Data Mining gespielt werden, um hier interessante, bisher durch das OLAP-System nicht erkannte Muster und Strukturen zu entdecken (1). Umgekehrt können Mining-Ergebnisse, zur einer genaueren Erklärung und Fundierung direkt in das OLAP-System gespielt werden

(2). Neben dem typischen Mining auf flachen Strukturen wird nun zunehmend das mehrdimensionale Mining möglich. So entsteht ein erkenntnisfördernder und rationalitätssichernder Closed-Loop-Prozess aus hypothesen-gestützten und hypothesen-freien Analysen. Diese Entwicklung wird beispielsweise durch den SQL-Server von Microsoft oder von MIS ALEA zusammen mit dem DeltaMiner von Bissantz & Company vorangetrieben.

Analog zum Data Mining-Loop lässt sich auch ein Text Mining-Loop skizzieren (3). Für das Controlling relevante und interessante Beziehungen zwischen Geschäftsobjekten eines Unternehmens findet man jedoch teilweise erst dann, wenn man eine integrierte Analyse von sowohl strukturierten als auch unstrukturierten Daten durchführt (4). So befinden sich beispielsweise strukturierte Kundeninformationen wie Name, Anschrift, Telefonnummer und Anzahl der Kundengespräche in der relationalen Marketingdatenbank. Korrespondenz, Aktennotizen und Verträge sind jedoch in der dokumentenorientierten Datenbank auf Basis von Lotus Notes abgelegt. Auftragspositionen, aktuelle Umsatzzahlen und Auftragsbestätigungen befinden sich als strukturierte Informationen wiederum in einer relationalen Datenbank, die sich in der Buchhaltung befindet. Die neuesten Nachrichten und Börsennotierungen des Kunden kommen online über Internet und liegen damit als HTML-Dokumente vor. Gerade zwischen diesen unterschiedlich strukturierten Informationen können signifikante Relationen verborgen sein. Möglicherweise existieren interessante Beziehungen zwischen der Korrespondenz, welche die Beschwerden der Kunden zum Ausdruck bringt und bestimmten Sachbearbeitern oder Produkten. Durch die Verbindung strukturierter Kundeninformationen wie Alter, Geschlecht oder Anzahl der Kinder mit textuellen Informationen wie Beschwerden über Produkte oder Serviceleistungen, Feedbackschreiben oder Verbesserungsvorschlägen kann wichtiges Wissen für die Produktentwicklung und das Marketing entwickelt werden. Integriertes Mining (5) hilft dem Unternehmen, mehr über seine Kunden zu lernen, seine Beschwerden besser zu verstehen sowie die Wahrnehmung des eigenen Unternehmens durch Kunden und Konkurrenten zu erfahren. Ein ganzheitliches Mining ist nur durch die integrative Perspektive aller Datentypen und -medien möglich. Daher sind Hersteller bemüht, dem Anwender beide Mining-Perspektiven in einer integrierten Business Intelligence-Lösung anzubieten. Die integrierten

Mining-Tools befinden sich noch in einem relativ frühen Entwicklungsstadium, werden aber zunehmend an Marktreife und Bedeutung gewinnen.

Mit zunehmenden internen und externen Dokumentenvolumina erhöht sich das Potential des Text Mining. Damit steigt nicht nur die Notwendigkeit, den Computer bei der Analyse von Dokumenten unterstützend zu nutzen, sondern auch die Möglichkeiten, diese Technologie im Unternehmen sinnvoll einzusetzen. Die zunehmende Verbreitung des XML-Standards wird die Semantik der unstrukturierten Daten erhöhen und damit die Anwendungsmöglichkeiten und „Entdeckungspotentiale“ der Text Mining-Tools im Inter- oder Intranet deutlich vergrößern.

Die steigende Bedeutung des Themas Data Mining für die unternehmerische Praxis ist ganz entscheidend auf die zunehmende Verbesserung der Leistungsfähigkeit und Bedienerfreundlichkeit entsprechender Data Mining-Softwarewerkzeuge zurückzuführen. Die rasante technologische Entwicklung der Datenbank- und Rechnersysteme hat den Data Mining-Tools in Bezug auf die wirtschaftliche Anwendbarkeit in der unternehmerischen Praxis eine neue Qualität verliehen. Durch die stärkere informationstechnologische Verankerung der Controlling-Anwendung mit Data Mining-Systemen auf der Grundlage von Data Warehouse-Infrastrukturen ergeben sich Synergien, die positive Auswirkungen auf die Befähigung zur qualitativen und quantitativen Analyse haben, die eine moderne, wettbewerbsfähige Controlling-Applikation erfüllen sollte (vgl. Gentsch et al. 2000).

Ein weiteres strategisches Anwendungsfeld für das Data Mining im Controlling werden zukünftig stärker auch Frühwarnsysteme sein, die auf Basis von Prognoseverfahren das Verhalten von Stakeholdern bis zu einem bestimmten Grad besser kalkulierbar machen. Einen besonders innovativen Ansatz stellt hier das Predictive IR dar, das Ursache-Wirkungs-Beziehungen bezogen auf das Anlageverhalten von Aktionären prognostiziert (ausführlich hierzu Gehra/Niedermaier/Hess 2003).

„Wir können jemandem Pardon gewähren, der besiegt wurde, aber niemanden, der sich hat überraschen lassen“

Friedrich der Große

Literaturhinweise

Bissantz, N./Hagedorn, J.: Data Mining im Controlling, Erlangen 1996.

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.(Hrsg.): Advances in Knowledge Discovery and Data Mining, Menlo Park 1996, S. 1-34.

Gehra, B./Niedermaier, O./Hess, T.: Fallstudie Predictive Investor Relations - Früherkennung durch Nutzung operativer Datenbestände bei der Computershare Ltd., Zeitschrift für Controlling und Management (ZfCM), 47. Jg. (2003), 2. Sonderheft.

Gentsch, P./Mandzak, P./Roth, M.: Data Mining: 12 Tools im Vergleich, Würzburg 2000.

Gentsch, P./Mandzak, P./Roth, M.: Web-Personalisierung und Web-Mining für eCRM: 12 Tools im Vergleich, Würzburg 2001.

Grothe M./Gentsch P.: Business Intelligence „Aus Informationen Wettbewerbsvorteile gewinnen, Berlin 2000.

Kaplan, R. S./Norton, D. P.: Balanced Scorecard - Strategien erfolgreich umsetzen, Stuttgart 1997.

Keim, D. A./Kriegel, H.-P./Seidel, T.: Supporting Data Mining of Large Databases by Visual Feedback Queries, in: IEEEEEE 1994, S. 298–313.

Küppers, B.: Data Mining in der Praxis, Berlin 1998.

Schommer, C./Müller, U.: Data Mining im E-Commerce – ein Fallbeispiel zur erweiterten Logfileanalyse, HMD 04/2001.