

Warenkorbanalysen mit Data Mining

Stephan Müller, Institut für Agrarwirtschaft, ETH Zürich

Bei zunehmender Preiskonkurrenz im Detailhandel wird das Marketing zum kritischen Erfolgsfaktor, denn sämtliche Aktivitäten sind noch verstärkter auf die Bedürfnisse der Kunden auszurichten. Innerhalb der Marketingforschung spielt der einzelne Kaufakt eine wichtige Rolle und macht eine ganzheitliche Analyse des Warenkorbes notwendig. Fortschritte in der Informations-technologie ermöglichen es, dass riesige Datenmengen gespeichert werden, die einen wertvollen Pool an „Geschäftserfahrungen“ darstellen. Dieses Wissenspotential lässt sich mit konventionellen statistischen Ansätzen nur bedingt verfügbar machen. In diesem Kontext bezeichnet Data Mining einen Analyseprozess zur Gewinnung handlungsrelevanter Informationen aus grossen Datenbanken.

Ziel der Studie, die als Diplomarbeit an der ETH Zürich eingereicht wurde, ist es, einen möglichen Data Mining-Prozess für die Auswertung von Verkaufsdaten im Detailhandel aufzuzeigen. Dabei stehen die Effizienz des Prozessmodells und die Abklärung der fachlichen Bedürfnisse im Fokus der Untersuchung. Anhand einer Warenkorbanalyse wird der Data Mining-Prozess in Anlehnung an Hippner systematisch bearbeitet. Das Kernstück des Prozesses bildet die Assoziationsanalyse, welche handlungsrelevante Beziehungen zwischen verschiedenen Artikeln eines Sortiments aufdeckt.

Probleme bei der Auswertung und Interpretation der gefundenen Assoziationsregeln erschweren die Umsetzung in marketingrelevante Empfehlungen. Trotzdem stellte sich der Data Mining-Prozess als äusserst effizient heraus.

Schlüsselwörter: Marketingforschung, Data Mining, Warenkorbanalyse, Assoziationsanalyse

1. Einleitung

1.1 Ausgangslage und Problemstellung

Im Detailhandel verschärft der zunehmende Preis- und Innovationsdruck den Wettbewerb. Als Folge rückt der Kunde stärker ins Zentrum der unternehmerischen Tätigkeit. Kundenorientiertes Handeln bedingt ein effizientes Marketing, dessen Kernaufgabe die Erfüllung von Kundenbedürfnissen darstellt.

In diesem Kontext ist die Aufgabe der Marketingforschung die Gewinnung und Bereitstellung von Informationen über das Kundenverhalten. Um wirksame Marketingmassnahmen ergreifen zu können, ist das detaillierte Verständnis des Kaufakts¹ und die damit verbundene Zusammensetzung des Warenkorbes unumgänglich.

Nebst Abgrenzungsproblemen ist allen Warenkorbanalysen gemeinsam, dass die Zusammensetzung des Kaufakts, im Sinne einer ganzheitlichen Analyse, nur teilweise möglich ist. Eine Alternative zu den konventionellen statistischen Verfahren stellt die Auswertung des Warenkorbes mit Hilfe des elektronischen Kassabons dar. In diesem Zusammenhang eröffnet der Einsatz von Data Mining² neue Möglichkeiten in der Marketingforschung.

1.2 Forschungsziele und methodisches Vorgehen

- Die Arbeit zeigt die Möglichkeiten zur data mining-basierten Auswertung des Warenkorbes mit Hilfe des elektronischen Kassabons anhand eines Fallbeispiels. Dabei sollen die folgenden Zielsetzungen erreicht werden:
- Am Beispiel der Software „Clementine 8.5“ von SPSS soll aufgezeigt werden, wie die Umsetzung einer Warenkorbanalyse zu realisieren wäre.

¹ Unter Kaufakt wird der vollständige Einkaufsprozess, bis zur Bezahlung des Warenkorbes an der Kasse, verstanden.

² Prozessorientierte Definition von Data Mining: „Data Mining bezeichnet den Prozess, der durch die Anwendung von Methoden auf einen Datenbestand Muster identifiziert“ (GROB ET AL. 1999).

- Die Evaluation des verwendeten Prozessmodells und der Analyse-methode dient zur Abklärung von fachlichen Bedürfnissen im Detailhandel bei der systematischen Durchführung von Warenkorbanalysen.
- Der theoretische Teil der Arbeit besteht aus einer Situationsanalyse des Schweizerischen Detailhandels. Dabei wird die Bedeutung der Marketingforschung für die Gestaltung unternehmerischer Aktivitäten herauskristallisiert. Im Anschluss wird das Data Mining-Prozess-modell nach HIPNER ET AL. (2001) beschrieben, und die einzelnen Prozessstufen werden systematisch durchlaufen.
- Der empirische Teil besteht aus der Durchführung einer Assoziationsanalyse im Rahmen der Umsetzung des Prozess-modells. Zum Schluss werden die angewandte Methode und die generierten Resultate einer, auf die Methode fokussierten, Diskussion unterzogen.

2. Situation des Schweizer Detailhandels

Das dynamische Umfeld der europäischen Märkte beeinflusst zunehmend die Wettbewerbsbedingungen der Detailhändler in der Schweiz. Französische und deutsche Unternehmen versuchen durch Expansion grössenbedingte Wettbewerbsvorteile zu nutzen. In der Schweiz ist der Lebensmitteldetailhandel durch die Vormachtstellung einiger weniger Unternehmungen gekennzeichnet. Der hohe inländische Agrarprotektionismus wirkt sich strukturerhaltend aus (BRÜHLMANN 2004). Zudem erschweren die gute vertikale Integration in die geschützten vor- und nachgelagerten Stufen und die umfassenden Kenntnisse der Marktverhältnisse den Zutritt für europäische Unternehmungen, können diesen aber nicht verhindern. Durch den Markteintritt europäischer Discounter sehen sich die Schweizer Detailhändler mit einem verschärften Preiswettbewerb konfrontiert, wodurch der Konsument noch stärker ins Zentrum der handelsbetrieblichen Vorgänge rückt.

In diesem Kontext gewinnt das Marketing, welches alle unternehmerischen Tätigkeiten auf die Kundenbedürfnisse ausrichtet, zusätzlich an Bedeutung (SEILER 2001). Zur Erfüllung dieser Aufgabe liefert die Marketingforschung Informationen zuhanden der strategischen Marketingplanung und bereitet, basierend auf einem vertieften Verständnis des Konsumenten und dessen Verhalten, unternehmerische Entscheidungen vor.

Die mögliche Nutzung von Scannerdaten stellt heute einen Schwerpunkt in der quantitativen Marketingforschung im Handel dar. Beim Point-of-Sale (POS) Scanning erfolgt gleichzeitig mit dem Kassiervorgang die automatische Aufzeichnung eines elektronischen Kassabons, welcher die Zusammensetzung des Warenkorbes und damit implizit das Einkaufsverhalten jedes einzelnen Konsumenten festhält. Das POS-Scanning dient als Grundlage für datengestützte Entscheidungen. In jüngster Zeit sind besonders Data Mining-Ansätze zur Auswertung grosser Mengen von Scanningdaten in den Mittelpunkt des wissenschaftlichen und praktischen Interesses gerückt.

3. Data Mining im Marketing

3.1 Das Data Mining-Konzept

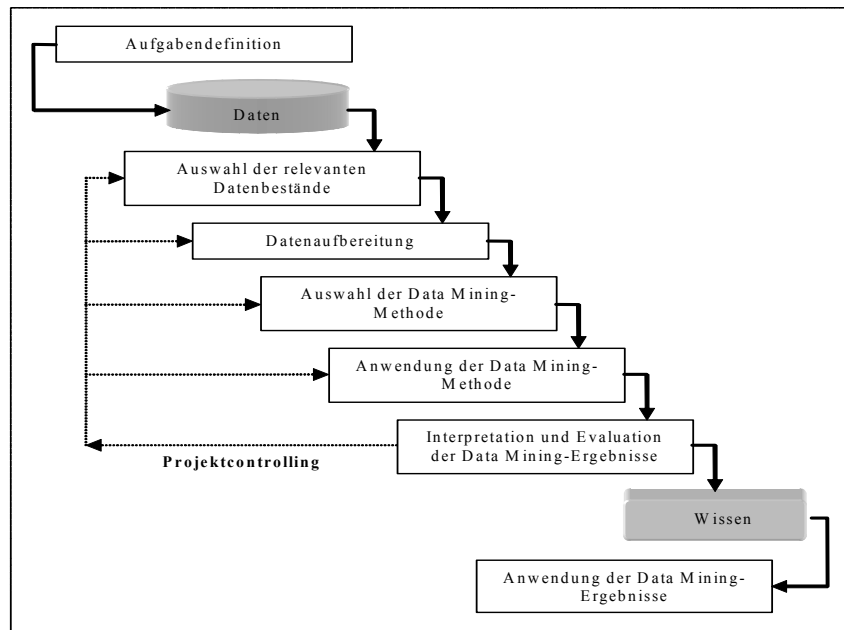
Der Einsatz betriebswirtschaftlicher Systeme zur Unterstützung operativer Geschäftsprozesse konfrontiert Unternehmungen zunehmend mit dem Phänomen der Datenflut. Trotz explodierendem Datenangebot stehen dem Management nur wenige zusätzliche Informationen zur Verfügung. Erst die wachsende Segmentierung der Märkte in immer kleinere, differenziertere Segmente und der steigende Wettbewerbsdruck, führten zur Erkenntnis, dass die traditionellen Geschäftsprozesse im Marketing den Ansprüchen des Marktes nicht mehr gewachsen sind (BERRY ET AL. 1997). Rasche und radikale Wechsel der strategischen Position erfordern ein Handeln ausserhalb des gewohnten Erfahrungsbereichs (HIPPER ET AL. 2001). Damit die in den operativen Informationsfeldern versteckten Geschäftserfahrungen für das Management verfügbar werden, braucht es Support-Systeme, die über die folgenden drei Grundelemente verfügen:

- 1. Informationsbeschaffung:** Die in den operativen Systemen entstandenen Daten werden zusammengeführt.
- 2. Informationsbereitstellung:** Diese gespeicherten Informationen müssen vom Management abgerufen werden können.
- 3. Informationsauswertung und -präsentation:** Innerhalb dieses Systems gewährt Data Mining die Analyse und Interpretation der Daten, sowie die Präsentation der Ergebnisse.

Der Data Mining-Prozess ist in einen übergeordneten Managementzyklus von Planung, Massnahmen und Wirkungsanalyse eingebunden. Letztere liefert die Grundlage für eine verbesserte Planung künftiger Marketingaktionen und ist Kernaufgabe des Data Minings.

3.2 Das Prozessmodell im Data Mining

Data Mining umfasst den Prozess der Gewinnung handlungs-relevanter Informationen aus grossen Datenbanken und deren Bereitstellung für betriebswirtschaftliche Entscheidungen (CABBENA ET AL. 1998). Im Zentrum dieses Prozesses stehen verschiedene Data Mining-Methoden (siehe dazu Abbildung 1). Allen ist gemeinsam, dass sie ohne zielorientierte Datenauf- bzw. Ergebnismachbearbeitung keine oder nur unbrauchbare Informationen generieren.



Quelle: In Anlehnung an HIPPER ET AL. 2001

Abb. 1: Der Data Mining-Prozess.

Der Prozessablauf lässt sich in die folgenden Phasen unterteilen:

- 1. Aufgabendefinition:** Die erste Prozessphase beinhaltet die exakte Definition der betriebswirtschaftlichen Problemstellung.
- 2. Auswahl der relevanten Datenbestände:** Hauptbestandteil der Phase zwei ist die Evaluation geeigneter Datenbestände.
- 3. Datenaufbereitung:** Die Aufgabe der dritten Phase besteht in der Transformation der Daten in eine für die Analyse geeignete Form.
- 4. Auswahl von Data Mining-Methoden:** Die vierte Phase beinhaltet sowohl die Bestimmung geeigneter Data Mining-Methoden³ als auch die Auswahl einer Data Mining-Software.
- 5. Anwendung der Data Mining-Methode:** In der fünften Phase erfolgt der Methodeneinsatz zur Identifikation interessanter Informationen.
- 6. Interpretation und Evaluation der Data Mining-Ergebnisse:** Die sechste Prozessphase soll die generierten Resultate hinsichtlich ihrer handlungsrelevanten Bedeutung bewerten.
- 7. Anwendung der Data Mining-Ergebnisse:** Die Aufgabe der siebten Phase besteht in der Integration der Resultate in die Führungsentscheidungen im Marketing.

Im Zentrum des Data Mining-Prozesses stehen die verschiedenen Data Mining-Methoden. Hier soll auf die für Warenkorbanalysen üblichste, die Assoziationsanalyse, genauer eingegangen werden.

3.3 Die Assoziationsanalyse

Die Assoziationsanalyse hat zum Ziel, strukturelle Zusammenhänge, sogenannte Assoziationsregeln, in Datenbanken zu identifizieren. Gegeben sei eine Anzahl von Warenkörben und darin enthalten eine Menge von Artikeln. Das Ziel ist es, diejenigen Artikel zu ermitteln, die das Auftreten anderer Artikel innerhalb eines Warenkorbes implizieren. Assoziationsregeln sind entsprechend immer Wenn-Dann-Bedingungen. Zum Beispiel: „WENN ein Kunde Pizzateig kauft, DANN kauft er in 10 % der Fälle auch Mozzarella.“

³ Unter einer „Data Mining-Methode“ wird eine generelle Beschreibung der Vorgehensweise verstanden und beinhaltet damit mehr als nur die Auswahl einer geeigneten Analyse-methode.

Um die „Interessantheit“ einer aufgedeckten Assoziationsregel zu bestimmen, werden drei Masse verwendet: Support, Konfidenz und Lift. Diese drei „Interessantheitsmasse“⁴ basieren auf Häufigkeitszählungen. Bezogen auf das genannte Beispiel, beantwortet der Support die Frage, „In wie vielen Warenkörben kommt sowohl Pizzateig, als auch Mozzarella vor?“ Die Konfidenz einer Regel gibt an, mit welcher Wahrscheinlichkeit ein Kunde Mozzarella kauft, wenn er bereits Pizzateig gekauft hat. Ein Lift von 2 bedeutet, dass Kunden, die Pizzateig gekauft haben, doppelt so oft auch Mozzarella kaufen, wie Kunden, die ebenfalls Mozzarella kaufen, aber keinen Pizzateig.

Es gilt weiter zu beachten, dass Assoziationsregeln, die nur für wenige Warenkörbe gelten, häufig von keinem grossen Interesse sind. Um die wertvollen von den trivialen Regeln zu trennen, werden minimale Werte für den Support, Lift und die Konfidenz definiert.

4. Warenkorbanalysen mit Data Mining

4.1 Einleitung und Problemstellung der Warenkorbanalyse

Für eine konsumorientierte Marketingpolitik gewinnen Warenkorbanalysen an Bedeutung (BRÜHLMANN 2004). Die Warenkorbanalyse versucht mit Hilfe eines datenbasierten Ansatzes Erklärungen zum Käuferverhalten zu machen.

Dabei steht der spezifische Warenkorb, als Resultat eines Kaufakts, im Zentrum des Interesses. Werden im Rahmen eines solchen Kaufakts an einem Ort, zu einem bestimmten Zeitpunkt mehrere Artikel gleichzeitig erworben, dann spricht man vom einem Kaufverbund, der somit das direkt beobachtbare Kundenverhalten beschreibt. Die Assoziationsanalyse hat zum Ziel, solche Kaufverbunde innerhalb eines Sortiments ausfindig zu machen.

Für die Messung von Kaufverbunden ist nominales Skalenniveau der Daten genügend. Es ist ausreichend zu wissen, ob ein Artikel in einem

⁴ Interessantheitsmasse bezeichnen Filtersysteme, welche die gefundenen Assoziationsregeln in interessante und uninteressante aufteilen können.

Warenkorb gar nicht oder mindestens einmal vorkommt. Oder anders formuliert, es reicht zwischen Kauf und Nichtkauf zu unterscheiden.

4.2 Umsetzung der Warenkorbanalyse

Die folgenden Ausführungen stellen aus Datenschutzgründen nur einen Auszug aus der Fallstudie in der Diplomarbeit dar.

Aufgabendefinition der Warenkorbanalyse

Folgende betriebswirtschaftliche Fragestellung soll bei der Umsetzung des Data Mining-Prozesses beantwortet werden:

- Welche Beziehungen bestehen zwischen im Verbund gekauften Artikeln innerhalb des betrachteten Sortiments?

Der Fokus liegt auf einem hypothesengenerierenden Forschungsansatz, da die Theoriedefizite im vorliegenden Forschungsfeld einen hypothesenprüfenden Ansatz nur in beschränktem Umfang ermöglichen.

Auswahl des Datenbestandes

Aus dem Internetgeschäft eines Schweizer Detailhändlers wurden Verkaufsdaten des Jahres 2004 ausgewählt.

Die analysierten Kassabons werden mit Hilfe von neun verschiedenen Variablen beschrieben. Für die Assoziationsanalyse genügen die Kassabonnummer und die zugehörigen Artikelbezeichnungen.

Aufbereitung des ausgewählten Datenbestandes

Die Analyse von Verbundbeziehungen kann auf verschiedenen Hierarchieebenen, zum Beispiel Artikel- oder Warengruppen⁵, erfolgen. Im vorliegenden Fallbeispiel wurde die Assoziationsanalyse auf der Ebene der Warengruppe durchgeführt. Zu diesem Zweck wurde jeder Artikelnummer die entsprechende Warengruppe zugeordnet.

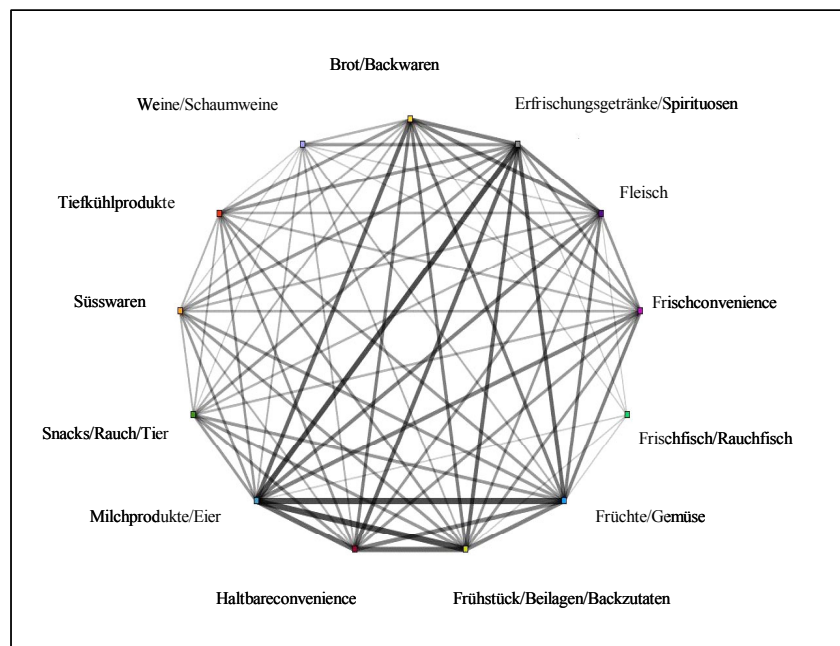
⁵ Man unterscheidet zwischen: Hauptwarengruppen, Warengruppen, Unterwarengruppen, Unterunterwarengruppen und den einzelnen Artikel.

Auswahl und Anwendung der Data Mining-Methode

Zur Umsetzung des Data Mining-Prozesses wurde die Software „Clementine 8.5“ von SPSS verwendet. Der vollständige Analyse-prozess kann auf einer graphischen Benutzeroberfläche dargestellt und verwaltet werden. Dadurch steigt die Nachvollziehbarkeit der Analyse und der einzelnen Auswertungsschritte beträchtlich.

Darstellung ausgesuchter Data Mining-Ergebnisse

Die Verbundbeziehungen werden durch ein Netzdiagramm sehr übersichtlich dargestellt. Daraus können erste Hinweise auf Beziehungen innerhalb eines Sortiments abgeleitet werden. Abbildung 2 zeigt ein solches Netzdiagramm. Je ausgeprägter die Verbindungslinien zwischen zwei Elementen, desto häufiger sind diese gemeinsam auf einem Kassabon anzutreffen. Hier werden Intensitäten zwischen Hauptwarengruppen abgebildet.



Quelle: Eigene Darstellung

Abb. 2: Print Screen, Clementine 8.5, verändert.

Die Interpretation des Webknotens zeigt die hohe Verbundwirkung von Milchprodukten/Eiern und Früchten/Gemüse. Diese beiden Warengruppen gehören offensichtlich zu den grundlegenden Elementen eines Warenkorbes. Im Weiteren besteht eine intensive Beziehung zwischen Milchprodukten/Eiern, Backen/Backwaren und Frühstück/ Beilagen/ Backzutaten. Diese Kombinationen werden sowohl bei der Realisierung eines Backrezepts, als auch bei der Zusammenstellung des Frühstück-angebots miteinander gekauft. Auffallend ist die hohe Bedeutung der Warengruppe Erfrischungsgetränke/Spirituosen. Es könnte sein, dass der Internetshop als Lieferkanal für schwere Artikel wie zum Beispiel Mineralwasser genutzt wird.

Das Auffinden der Assoziationsregeln erfolgt durch die Anwendung des Apriori Algorithmus auf den Datensatz. Die in Tabelle 1 aufgeführten Regeln zeigen die stärksten Beziehungen zwischen den einzelnen Warenuntergruppen.

Tabelle 1: Resultate, selektierte Assoziationsregeln

WENN_Bedingung	DANN_Bedingung	Anzahl	Anzahl %	Konfidenz %	Support %	Lift
Senf	Mayonnaise	316	6.87	47.78	3.28	5.58
BiscmitSchoko	BischohneSchoko	233	5.06	45.06	2.28	5.39
Essig	Öl	341	7.41	52.49	3.89	4.48
GlacéDessertbeilag	TK_Gemüse	244	5.30	43.44	2.30	4.38
Weissmehl	Zucker	244	5.30	45.49	2.41	4.29
Salzgebäck/Gebäck	Chips	391	8.49	38.62	3.28	3.29
ÜbrigeSalate	Blattsalate	788	17.12	56.47	9.67	2.62
Früchte_Beerenobst	ExotFrüchte/Gemüse	391	8.49	35.81	3.04	2.54
Früchte_Beerenobst	ÜbrigeSalate	391	8.49	43.22	3.67	2.52
Pilze	Zwiebelprodukte	377	8.19	33.16	2.72	2.48
Fischkonserven	Gemüsekonserven	286	6.21	43.71	2.72	2.47
ÜbrigeSalate	Tomaten	788	17.12	61.93	10.60	2.44
Pilze	ÜbrigeSalate	377	8.19	41.64	3.41	2.43
Früchte_Beerenobst	Tomaten	391	8.49	61.38	5.21	2.42
Pilze	ExotFrüchte/Gemüse	377	8.19	33.95	2.78	2.41
Pilze	Tomaten	377	8.19	61.01	5.00	2.41
Pilze	Gemüse	377	8.19	85.68	7.02	2.34
ExotFrüchte/Gemüse	Früchte_Agrumen	649	14.10	51.16	7.21	2.31
Reis	Teigwaren	483	10.49	49.48	5.19	2.30
Fruchtkonserven	Gemüsekonserven	328	7.13	40.55	2.89	2.29
Pilze	Blattsalate	377	8.19	49.34	4.04	2.29
Pastasaucen	Teigwaren	798	17.34	48.50	8.41	2.26
ExotFrüchte/Gemüse	Früchte_Bananen	649	14.10	63.17	8.91	2.24
Pilze	VerarbeiteteSalate	377	8.19	51.72	4.24	2.23
Blattsalate	Tomaten	993	21.57	56.19	12.12	2.22

Quelle: Eigene Darstellung

Betrachtet man die Resultate der Assoziationsanalyse in Tabelle 1, dann fällt auf, dass viele Vermutungen bezüglich Artikelkombinationen bestätigt werden. Zum Beispiel besteht ein starker Kaufverbund zwischen Senf und Mayonnaise. Senf wird gesamthaft 316 Mal gekauft, bzw. kann in 6,87 % aller Warenkörbe festgestellt werden. In 47,78 % der Fälle gelangt zusätzlich zum Senf auch gleich noch Mayonnaise in den Einkaufswagen, so dass die Kombination WENN Senf, DANN Mayonnaise in 3,28 % aller Warenkörbe vorkommt. Auch Artikel, die zu ähnlichen Warengruppen gehören, wie Salzgebäck/Gebäck und Chips, Reis und Teigwaren, Fisch- und Gemüsekonserven, gelangen zusammen in denselben Warenkorb. Erstaunlich ist die grosse Häufigkeit, mit der die Untergruppe „Pilze“ in der WENN_Bedingung vorkommt. Pilze sind offensichtlich ein Indikator für das Vorhandensein von Früchten oder Gemüse im Warenkorb.

5. Diskussion der Methode

Eine Herausforderung beim Arbeiten mit Data Mining besteht im fehlenden standardisierten Prozessschema. Diese Tatsache ist eine Folge der Komplexität anfallender Problemstellungen. Der verwendete Ansatz von Hippner stellte sich als äusserst effizient heraus. Das Prozessdesign ermöglicht dem Methoden-Laien einen leichten Einstieg in die Thematik. Gezielt wird der Anwender auf Schwachstellen der Vorgehensweise hingewiesen, so dass der Prozessablauf frühzeitig optimiert werden kann. Trotzdem verlangt das Analysedesign eines Data Mining-Prozesses gründliche Kenntnisse, sowohl in der Datenverwaltung, als auch bei der Datenanalyse und dem jeweiligen Anwendungsfeld - hier dem Detailhandel.

Der Ansatz der Assoziationsanalyse brachte bei der Umsetzung des Fallbeispiels Vorteile mit sich. Bei hypothesenprüfenden Verfahren fallen ca. 80 % des zeitlichen Aufwandes und der Prozesskosten auf die Datenvor- bzw. Datenaufbereitung. Die Durchführung einer Assoziationsanalyse verlangt lediglich nominales Skalenniveau, und normalerweise können die Daten bereits mit geeignetem Skalenniveau bezogen werden, was die Datenaufbereitung erleichtert. Im Weiteren kann durch das Zusammenfassen von Artikeln zu Produktgruppen die Rechenzeit erheblich reduziert werden. Dieses Vorgehen verringert jedoch die Aussagekraft der Resultate. Zur Festlegung der optimalen Hierarchiestufen finden sich leider keine Hinweise in der Literatur. Allgemein stellte sich

das Generieren von Assoziationsregeln jedoch als erstaunlich einfach heraus.

Die Software „Clementine 8.5“ von SPSS überzeugte durch ihre einfache Benutzeroberfläche und übersichtliche Darstellung des gesamten Analyseprozesses. Leider machten es fehlende Instrumente zur Ergebnisauswertung praktisch unmöglich, die interessanten, von den naheliegenden Regeln zu trennen.

6. Schlussfolgerungen

6.1 Der Data Mining-Prozess

Bei der Umsetzung des Fallbeispiels ist das Potential des Data Mining-Prozesses deutlich geworden. Diese Schlussfolgerung bestätigt eine Studie der Universität Eichstätt-Ingoldstadt, welche belegt, dass 87 % der Unternehmen, die Data Mining einsetzen, mit der Profitabilität dieser Projekte zufrieden sind (LEBER 2002). Diese Tatsache, verbunden mit der Erkenntnis, dass sich alle 18 Monate die Informationsmenge auf der Welt verdoppelt, führt zum Schluss, dass die Bedeutung von Data Mining in der Datenanalyse zunehmen wird. Noch wirksamere Data Mining-Methoden sind dringend notwendig. Mit deren Entwicklung werden die heute noch vorhandenen Theoriedefizite schnell kompensiert sein. Trotzdem sind im Rahmen des Data Minings wichtige Fragen ohne Antwort geblieben.

Die Assoziationsanalyse generiert eine kaum überblickbare Anzahl an Regeln. Der effektive Aussagegehalt dieser Regeln wird jedoch erst durch die sorgfältige Interpretation durch den Anwender manifest. Leider fehlt für die Ergebnisinterpretation ein bewährtes Vorgehen, um diese umfangreiche Informationsmenge angemessen auszuwerten. Die verwendete Software bietet zwar eine Reihe von Interessantheitsmassen an, diese tragen jedoch nur wenig zur Ergebnisauswertung bei. Es wäre wünschenswert, mit mehreren Masszahlen ganze Modelle von Bewertungskriterien zu konstruieren, die jeweils auf die spezifischen Fragestellungen angepasst werden könnten.

Eine weitere Problematik zeigte sich in der Tatsache, dass verschiedene Algorithmen unterschiedliche Assoziationsregeln generieren. Folglich können je nach verwendeter Software bzw. verwendetem Algorith-

mus, gänzlich andere Empfehlungen an die Marketingabteilung resultieren. Optimalerweise werden die Daten mit verschiedenen mathematischen Verfahren ausgewertet und die Resultate abgeglichen.

Die Assoziationsregel, dass Pastasauce mit Teigwaren gekauft wird, verdeutlicht eine weitere Problematik. Das erwähnte Beispiel kann völlig gegensätzlich interpretiert werden. Soll zur Nutzung des „Cross Sellings-Effekts“ neben dem Gestell mit Pastasauce eines mit Teigwaren eingerichtet werden? Oder sollen die Pastasaucen und die Teigwaren, da diese Artikel so oder so zusammen gekauft werden, möglichst weit auseinander platziert werden, damit der Kunde auf dem Weg dazwischen möglichst viele Impulskäufe tätigt? Die Assoziationsanalyse also ist im heutigen Entwicklungsstadium noch nicht reif für ein routinemäßiges Controlling von absatzpolitischen Massnahmen, wie z.B. Promotionen oder Ladengestaltung. Ohne zusätzliche Datenanalysen können solche Fragen kaum beantwortet werden.

Aus betriebswirtschaftlicher Sicht stellt sich die Frage nach der monetären Bewertung von Data Mining-Analysen. Erst wenn nämlich aufgezeigt werden kann, dass sich standardisierte Data Mining-Analysen direkt auf den Unternehmenserfolg auswirken, wird deren Notwendigkeit mess- und kommunizierbar.

6.2 Anwendungsempfehlungen

Durch eine Anpassung oder Erweiterung der Datengrundlage mit zusätzlichen Informationen kann die Aussagekraft der Resultate erhöht werden.

- Die einfachste Verfeinerung besteht in der weiteren Zerlegung von Warenkorbklassen in ausgewählte Hierarchieebenen oder Klasseneinteilungen.
- Einführung virtueller Items. Jeder Warenkorb könnte mit einer Dummy-Variablen für „Aktion“ ergänzt werden. Assoziationsregeln, die mit der Bezeichnung „Aktion“ versehen sind, könnten gezielter untersucht werden.
- Durch den Einbezug des Einkaufszeitpunktes und die Zuordnung des Kassabons zum jeweiligen Kunden, könnte dessen Verhalten über die Zeit abgebildet werden.

- Es ist jedoch zu überprüfen, ob die Erweiterung der Datenbasis hinsichtlich des Zielerreichungsgrades und der Aussagekraft der Ergebnisse sinnvoll ist.
- Abschliessend ist darauf hinzuweisen, dass nur anonymisierte Kundendaten das tatsächliche Einkaufsverhalten abbilden. Der Einsatz von Kundenkarten zum Beispiel, ermöglicht zwar die Erfassung von Kaufdaten individueller Kunden, diese sind jedoch nur repräsentativ, wenn der Kunde für jeden einzelnen Einkauf auch die Kundenkarte benutzt (REUTTERER 1998). Zudem sind Kundenbindungsprogramme in der Regel auf die Erhöhung der Kauffrequenz ausgerichtet, womit Kundenkartenbesitzer eher zu den „Vielkäufern“ gezählt werden dürften.

Aus Informatiksicht weisen Data Mining-Systeme, die Datenvolumina im Terabyte-Bereich verarbeiten sollen, immer noch Effizienzprobleme auf. Dieser Herausforderung ist mit einer Weiterentwicklung vorhandener und der Konzipierung neuer Algorithmen zu begegnen.

Neben effizienteren Algorithmen stellt sich die Frage nach der Datenqualität. Leider besteht bis heute keine Möglichkeit, die Daten von den operativen Vorkaufsystemen, dem Point-of-Sale, in hinreichend hoher Qualität zu den am „Frontend“ ansetzenden Data Mining-Methoden schicken zu können. Diese Feststellung darf nicht über die Tatsache hinwegtäuschen, dass die Herausforderungen des Data Minings bei der Interpretation und Anwendung der Resultate liegt.

6.3 Fazit

Durch eine stärkere Berücksichtigung der Data Mining-Ansätze in der Marketingforschung könnten langfristig weitreichende Wettbewerbsvorteile erzielt werden. Die Leistung eines Detailhändlers ist sein Produktesortiment. Es wird immer ein Verbund von Artikeln verkauft, welcher die Bedürfnisse der Konsumenten möglichst optimal abdecken soll. Mit den bisherigen statistischen Methoden werden die verkauften Produkte jedoch isoliert betrachtet. Leider sind die Tools zur Messung solcher komplexer Wechselwirkungen noch nicht vorhanden oder erst in der Entwicklungsphase. Trotzdem kann Data Mining erste Antworten über Beziehungen zwischen im Verbund gekauften Artikeln geben.

7. Literatur

BERRY M.J.A., LINOFF G., 1997. Data Mining techniques for marketing, sales and customer support, Wiley corp., New York.

BRÜHLMANN R., 2004. Analyse des Potentials von Kassabon-Analysen im Detailhandel für eine effektive und effiziente Promotionsplanung, Diplomarbeit, ETH Zürich.

CABBENA P., HADJIINIAN P., STADLER R., VERHEES J., ZANASI A., 1998. Discovering data mining – from concept to implementation, Upper Saddle River.

GROB H.L., BENSBERG F., 1999. Das Data Mining-Konzept, URL: <http://www.wi.uni-muenster.de/aw/download/publikationen/CGC8.pdf>
Zugriff: 28.01.05.

HIPPNER H., KÜSTERS U., MEYER M., WILDE K., 2001. Handbuch - Data Mining im Marketing, Knowledge Discovery in Marketing Databases, Vieweg Verlag, Braunschweig/Wiesbaden.

LEBER M., 2002. Data Mining: Jeder zweite Grossbetrieb wertet Kundendaten systematisch aus, Katholische Universität Eichstätt-Ingoldstadt.

MÜLLER S., 2004. Abklärung der fachlichen Bedürfnisse im Detailhandel zur Einführung standardisierter Data Mining-Analysen, Diplomarbeit Nr. 5070, Institut für Agrarwirtschaft, ETH Zürich.

REUTTERER T., 1998. Die entscheidungsrelevante Aufbereitung von POS Scanning-Daten als Herausforderung für die Handelsforschung und –praxis, WU-Jahrestagung, Service Fachverlag, S. 213-224, Wien.

SEILER A., 2001. Marketing, BWL in der Praxis IV, 6. überarbeitete Auflage, Orell Füssli Verlag, Zürich.

Summary

Market Basket Analysis with Data Mining

With increasing price-competition in retail business, marketing becomes the critical success factor. All activities have to be oriented towards customer needs. Within marketing research, the individual act of purchase plays a decisive role and necessitates a holistic analysis of the market basket. Moreover, progress in information technology has enabled us to store huge amounts of data, including a valuable pool of business experiences. It is not possible to make this potential knowledge completely available with conventional statistical approaches. This is where Data Mining provides an ideal approach. It is an analysis process for extracting information from large databases.

The aim of the study, submitted as diploma thesis at the ETH Zurich, is to show a possible Data Mining Process for analyzing sales data in retail business. Thereby the analysis is focused on the efficiency of the process model and the identification of professional needs. According to Hippner, the Data Mining Process is applied systematically on the basis of a Market Basket Analysis. The Association Analysis, which detects relevant correlations between different products of an assortment, is the core of the process. Problems of interpreting identified association rules make a transformation into marketing recommendations very difficult. Nevertheless, the Data Mining Process turned out to be very efficient.

Keywords: Marketing Research, Data Mining, Market Basket Analysis, Association Analysis

Anschrift des Verfassers:

Stephan Müller
Institut für Agrarwirtschaft
Sonneggstrasse 33
8092 Zürich

stephan.mueller@iaw.agrl.ethz.ch