

# Empirischer Vergleich alternativer Ansätze zur Verbundanalyse im Marketing

Reinhold Decker

Lehrstuhl für BWL und Marketing, Universität Bielefeld,  
33501 Bielefeld  
E-Mail: rdecker@wiwi.uni-bielefeld.de

## Abstract

Der vorliegende Beitrag beschäftigt sich mit alternativen Ansätzen zur Analyse des Kaufverbundes. Kern der Arbeit ist die empirische Überprüfung der Eignung der in SAS® implementierten Ansätze für die praktische Entscheidungsunterstützung im Einzelhandelsmarketing auf Basis realer POS-Scannerdaten. Im Einzelnen richten sich die Betrachtungen auf das seit vielen Jahren etablierte Konzept der Assoziationskoeffizienten (inklusive bedingter Wahrscheinlichkeiten), auf die aus der aktuellen Data Mining-Diskussion bekannten Assoziationsregeln und auf einen neuen Ansatz auf Basis von Self-organizing MapsText

## 1. Verbundproblematik und Entscheidungsunterstützung

Kaufverbund bezeichnet ein aufgrund bereits realisierter Kaufakte entstandenes und – bei entsprechender Datenverfügbarkeit – auch identifizierbares Beziehungsgeflecht zwischen einzelnen Teilen eines Warensortiments. Die Relevanz dieser Thematik für das Marketing resultiert aus ihrem besonderen Stellenwert für die im Einzelhandel alltäglich zu treffenden Sortiments- und Platzierungsentscheidungen, ihrer Affinität zur Preisfestsetzung und der Tatsache, dass ein gezieltes Category Management ohne fundierte Verbundkenntnisse nicht wirklich sinnvoll zu betreiben ist. Die hieraus abzuleitenden Anforderungen an die Verbundanalyse bzw. -modellierung umfassen zum einen die grundsätzliche Fähigkeit zur Aufdeckung der „relevanten“ Sortimentsbeziehungen und zum anderen die Möglichkeit der geeigneten Dokumentation bzw. Visualisierung derselben als Grundlage für eine praxisnahe Entscheidungsvorbereitung. Ein zentrales und bis dato noch nicht zur Gänze gelöstes Problem besteht allerdings im zuverlässigen Nachweis entsprechender Beziehungen. Um diesem Anspruch gerecht werden zu können, bedarf es zunächst einmal geeigneter Datengrundlagen. Zum gegenwärtigen Zeitpunkt sind das in erster Linie die heute schon fast in allen größeren Einzelhandelsgeschäften quasi als Nebenprodukt des Kassiervorganges anfallenden Point-of-Sales (POS)-Scannerdaten (im folgenden kurz als Scannerdaten bezeichnet).

Aber auch das beste Datenmaterial ist für die Entscheidungsunterstützung im Hinblick auf die oben genannten Aspekte ohne geeignete Auswertungswerkzeuge nahezu wertlos. Insoweit erscheint es auch nicht überraschend, dass sich die Marketingforschung in der Vergangenheit schwerpunktmäßig auf die Modifikation bzw. Erweiterung des verfügbaren Instrumentariums konzentriert hat. Dennoch besteht in der einschlägigen Literatur bis heute weitgehend Uneinigkeit bzw. Ungewissheit im Hinblick auf die vielzitierte „Best Practice“. Umfassende empirische Gegenüberstellungen unter Einbeziehung der neueren Ansätze sind bis dato noch so gut wie nicht verfügbar. Erwähnenswert scheinen in diesem Zusammenhang allenfalls die entsprechenden Ausführungen bei *Schmalen et al. (1996)*.

In jüngster Vergangenheit ist die quantitative Verbundforschung durch verschiedene Arbeiten zum Data Mining und zum Knowledge Discovery in Databases, die den Kaufverbund gerne als Paradebeispiel für eine entsprechende praktische Problemstellung heranziehen, erneut in den Mittelpunkt des Interesses gerückt. Tabelle 1 gibt einen ersten Eindruck von der Breite des zwischenzeitlich verfügbaren Instrumentenspektrums.

Setzt man sich intensiver mit der einschlägigen Literatur zur Verbundanalyse auseinander, so erkennt man schnell, dass das Gros der bisher erschienenen Publikationen aus dem deutschsprachigen Raum stammt. Autoren wie *Böcker (1975, 1978)* und *Merkle (1981)*, die sich primär mit dem Konzept der Verbundkoeffizienten beschäftigten, sowie *Hruschka (1991)* und *Lukanowicz, Buchta (1992)*, deren einschlägige Forschungsbemühungen sich auf entsprechende Anwendungen des multivariaten Logitmodells konzentrierten, aber auch *Agrawal, Imielinski, Swami (1993)*, deren Augenmerk auf der Nutzung von Assoziationsregeln lag, haben durch ihre Arbeiten maßgeblich zum heutigen Stand der Entwicklung beigetragen.

	Sechziger/Siebziger Jahre	Achtziger Jahre	Neunziger Jahre
Modellierung der Verbundbeziehung:	Kreuzelastizitäten Lineare Programmierung Modell der bedingten Wahrscheinlichkeiten Assoziationskoeffizienten Korrelationsmodelle Lineare Regression	Polynomial-Lag-Regression Logistische Regression	Multivariate Logitmodelle Neuronale Netze Binomialverteilungsmodell Assoziationsregeln
Visualisierung der Verbundbeziehung:	Multidimensionale Skalierung Clusteranalyse Faktorenanalyse		Self-organizing maps Hypergraphen

Tabelle 1: *Instrumente der quantitativen Verbundanalyse*

## 2. Verwendete Datengrundlage

Um sich bei der Gegenüberstellung der Ansätze nicht auf theoretische Aspekte beschränken zu müssen, wurde für die nachfolgenden empirischen Betrachtungen ein eigens aufbereiteter Datensatz herangezogen. Hierbei handelt es sich um reale Scannerdaten aus dem Konsumgüterbereich, die sich wie in Tabelle 2 dargestellt charakterisieren lassen.

Lf. Nr.	Warengruppe (WG)	WG-Nr.	Lf. Nr.	Warengruppe (WG)	WG-Nr.
1	Haarshampoos	6	13	Gebissreiniger	222
2	Haarspülungen	9	14	Gebissshaftmittel	225
3	Haarkuren	12	15	Sonnenschutzmittel	240
4	Tampons	30	16	After Sun Lotions	241
5	Damenbinden	33	17	Nassrasiermittel	326
6	Katzennahrung	69	18	Klingen für Nassrasierer	329
7	Belohnungen (für Hund/Katze)	70	19	Schlankheitskost	356
8	Säfte (für Kleinkinder)	183	20	Fitnesskost	362
9	Desserts (für Kleinkinder)	186	21	Hustenbonbons	364
10	Gemüse (für Kleinkinder)	189	22	Kaugummi	366
11	Junior Kost	195	23	Herz- und Nerventonicum	368
12	Kindermenüs	198	24	Augenmakeup	562
			25	Lippenpflege	563

Tabelle 2: *Scannerdaten aus dem bundesdeutschen Einzelhandel*

Die in insgesamt mehr als 7.500 Warenkörben respektive Einkäufen beobachteten (und vom Verfasser gezielt ausgewählten) 25 Warengruppen lassen individuelle Kaufverbände wahrscheinlich erscheinen. So ist z.B. zu vermuten, dass die Nahrungsmittel für Kleinkinder, d.h. die Warengruppen 183 bis 189, aufgrund ihrer Zielgruppenspezifität und ihres mutmaßlich komplementären Charakters durchaus häufiger gemeinsam in einem Warenkorb auftreten werden. Als Käufer dürften hier wohl vorwiegend jüngere Konsumenten und Konsumentinnen in Frage kommen. In analoger Weise sprechen die Gebisspflegemittel, d.h. die Warengruppen 222 und 225, wohl eher eine ältere Klientel an. Eine ausgeprägte

Verbundbeziehung zu dem zuvor genannten Kindernahrungsmittelsortiment scheint hier eher unwahrscheinlich. Die Warengruppen 364 (Hustenbonbons) und 366 (Kaugummi) schließlich stellen typische Mitnahmeprodukte dar, die an der Kasse häufig spontan in den Warenkorb wandern. Eine starke Verbundbeziehung zu den anderen, in der Tabelle aufgelisteten Warengruppen ist hier wohl nicht zu vermuten.

### 3. Vergleichende Darstellung alternativer Analyseansätze

Bevor wir uns dem ersten der drei zu betrachtenden Messkonzepte zuwenden, sollen zunächst noch einige wichtige Bezeichner eingeführt werden. Es sei  $I = \{i_1, K, i_k, K, i_n\}$  eine Menge von Items (im vorliegenden Fall Warengruppen), die auf Verbundbeziehungen hin untersucht werden sollen. Die Menge der betrachteten Transaktionen respektive Warenkörbe  $T_j \subset I$  ( $j=1, K, m$ ) wird mit  $W$  und ein damit korrespondierender Transaktionsvektor mit  $\mathbf{t}_j = (t_{j1}, K, t_{jk}, K, t_{jn})$  bezeichnet, wobei gilt:

$$t_{jk} = \begin{cases} 1, & \text{falls } i_k \in T_j \\ 0, & \text{sonst} \end{cases} \quad \forall j, k$$

Ergänzend sei noch angemerkt, dass die Menge  $I$  nur jene Items enthält, die in mindestens einer der betrachteten Transaktionen vorkommen und dass auch Transaktionen mit nur einem Item Berücksichtigung finden.

#### 3.1 Verbundmessung auf Basis von Assoziationskoeffizienten

Die seit ca. 30 Jahren bekannte Verbundmessung mittels Assoziationskoeffizienten stellt wahrscheinlich das in der Vergangenheit am intensivsten (wissenschaftlich) diskutierte Konzept dieser Art dar. Ausgangspunkt ist hierbei die Annahme symmetrischer Verbundbeziehungen<sup>1</sup>.

Als Basis für die Modellierung der Verbundbeziehungen wird hier die sogenannte Frequenzmatrix

$$F = (f_{kl})_{k,l=1,K,n} \quad \text{mit} \quad f_{kl} = \begin{cases} \sum_{j=1}^m t_{jk} \cdot t_{jl}, & \text{falls } k \neq l \\ 0, & \text{sonst} \end{cases}$$

herangezogen, mit der die Häufigkeit des gemeinsamen Auftretens zweier unterschiedlicher Items in den Transaktionen erfasst wird<sup>2</sup>. Aus der Frequenzmatrix können paarweise Vierfeldertafeln (sogenannte Plasmoden) generiert werden, die ihrerseits die Grundlage für die sich anschließende Gruppenbildung darstellen. Die im Folgenden zugrunde gelegte Struktur dieser Tafeln wird durch Abbildung 1 veranschaulicht.

<sup>1</sup> Eine ausführliche Diskussion der Symmetrieproblematik in der Verbundanalyse findet sich bereits bei Böcker (1978).

<sup>2</sup> Der neueren Literatur folgend beschränken sich die weiteren Betrachtungen auf die ungewichtete Einzählung.

		Item $i_k$		
		„1“	„0“	
Item $i_l$	„1“	$a$	$b$	$a + b$
	„0“	$c$	$d$	$c + d$
		$a + c$	$b + d$	$a + b + c + d$

**Legende:**

Feld  $a$ : Anzahl der Transaktionen, die Item  $i_k$  und Item  $i_l$  enthalten

Feld  $b$ : ... Item  $i_l$ , aber nicht Item  $i_k$  enthalten

Feld  $c$ : ... Item  $i_k$ , aber nicht Item  $i_l$  enthalten

Feld  $d$ : ... weder Item  $i_k$  noch Item  $i_l$  enthalten

Abbildung 1: Allgemeiner Aufbau einer Vierfeldertafel

Entscheidend für die hieraus generierbare Verbundaussage ist allerdings die Wahl des die paarweisen Verbundintensitäten zum Ausdruck bringenden Verbundmaßes. Der für die weiteren Berechnungen herangezogene Tanimoto-Koeffizient

$$v_{kl} = \frac{a}{a+b+c} \quad \text{mit} \quad 0 \leq v_{kl} \leq 1 \quad \forall k, l$$

gilt als eines der für die vorliegende Problemstellung am besten geeigneten Maße<sup>3</sup>. Die in der resultierenden Ähnlichkeitsmatrix verborgenen Verbundbeziehungen können sodann mittels hierarchischer Clusteranalyse sichtbar gemacht werden.

Die bei diesem Ansatz unterstellte Symmetrie der Verbundbeziehungen muss allerdings keineswegs immer angemessen sein. Eine naheliegende Möglichkeit der expliziten Modellierung asymmetrischen Verbundes stellt die ebenfalls bereits in den siebziger Jahren in die Diskussion eingebrachte Verwendung bedingter Wahrscheinlichkeiten dar.

Die Wahrscheinlichkeit eines Kaufes von Item  $i_k$  in Abhängigkeit von einem Kauf des Items  $i_l$  lässt sich direkt aus der bereits erwähnten Vierfeldermatrix bestimmen. Es gilt nämlich

$$P(i_k | i_l) = \frac{a}{a+b} \quad \text{mit} \quad 0 \leq P(i_k | i_l) \leq 1 \quad \text{und} \quad P(i_l | i_k) = \frac{a}{a+c} \quad \text{mit} \quad 0 \leq P(i_l | i_k) \leq 1$$

Aus dem analog zu erhaltenden Erwartungswert für Feld  $a$ <sup>4</sup>:

$$\begin{aligned} E(a) &= P(i_l) \cdot P(i_k) \cdot (a+b+c+d) \\ &= \frac{a+b}{a+b+c+d} \cdot \frac{a+c}{a+b+c+d} \cdot (a+b+c+d) = \frac{(a+b)(a+c)}{a+b+c+d} \end{aligned}$$

lassen sich Aussagen über die Art einer möglicherweise bestehenden Verbundbeziehung ableiten. Für komplementären Verbund muss gelten:

$$a > E(a) \Rightarrow ad > bc \Leftrightarrow \frac{ad}{bc} > 1 \quad (\text{„odds ratio“}) \quad \Leftrightarrow d > \frac{bc}{a}$$

während im Falle von  $a \approx E(a)$  von einem zufälligen Mitnahmeeffekt auszugehen ist. Der Tatbestand der zufälligen Mitnahme kann dabei, Böcker (1978) folgend, als die stochastische Unabhängigkeit zweier Ereignisse respektive zweier Kaufhandlungen verstanden werden. Man erkennt den sich hieraus ergebenden Skaleneffekt „irrelevanter“ Sortimentsteile. Aufgrund der in aller Regel starken Dominanz des Feldes  $d$ , d.h. der Anzahl an beobachteten Nichtkäufen des gerade betrachteten Item-Paares, scheidet im

<sup>3</sup> Vgl. hierzu auch Bordemann (1985, S. 262).

<sup>4</sup> Vgl. hierzu auch Schmalen et al. (1996, S. 216 f.).

vorliegenden Fall der nahe liegende Chi-Quadrat-Test auf Unabhängigkeit als Instrument zur Quantifizierung bestehender Verbundbeziehungen aus.

Wendet man die skizzierte Vorgehensweise auf die eingangs vorgestellten Scannerdaten an, so kommt man, wie Abbildung 2 zu entnehmen ist, je nach gewähltem Clusteralgorithmus, zu auf den ersten Blick recht unterschiedlich anmutenden Dendrogrammen. Die in den Graphiken verwendeten Nummerierungen entsprechen denen aus obiger Datenbeschreibung. Bei genauerem Hinsehen erkennt man allerdings die nahezu völlige Übereinstimmung der berechneten Gruppierungen. Aufgrund der ausgeprägteren Hierarchiestruktur ist die Ward-Lösung allerdings einer inhaltlichen Interpretation leichter zugänglich. So erkennt man dort z.B. sehr schön das Kindernahrungsmittelcluster (laufende Nr. 8 bis 12) im unteren Bereich des Dendrogramms. Ebenfalls sehr gut zu erkennen ist das Senioren-Zahnpflegecluster (laufende Nr. 13 und 14), das in durchaus nachvollziehbarer Weise auch noch die Warengruppe Herz- und Nerventonicum (laufende Nr. 23) enthält. Insgesamt liegt eine sehr gute Reproduktion der a priori vermuteten Verbundbeziehungen innerhalb der jeweiligen Gruppen vor.

Dennoch verbleiben einige nicht zu vernachlässigende Probleme. So ermöglicht das skizzierte Verfahren z.B. noch nicht die Herausrechnung zufälliger Mitnahmeeffekte<sup>5</sup>. Darüber hinaus hat natürlich auch die Wahl des Verbundkoeffizienten – und die einschlägige Literatur geizt hier nicht gerade mit Vorschlägen – einen direkten Einfluss auf die resultierende Verbundaussage. Des Weiteren können stark unterschiedlich besetzte Warengruppen u.U. zu Ergebnisverzerrungen führen. Gleichzeitig ist die Vorgehensweise aber intuitiv plausibel und einsichtig. Alles in allem stellt das Konzept der Assoziationskoeffizienten in Verbindung mit der hierarchischen Clusteranalyse ein durchaus brauchbares, in der hier dargestellten traditionellen Form allerdings noch keinesfalls „optimales“ Instrument zur Verbundmodellierung dar.

---

<sup>5</sup> Ein möglicher Ansatz zur Identifikation zufälliger Mitnahmeeffekte auf Basis des Tanimoto-Koeffizienten wird in einer vom Verfasser gemeinsam mit H. Schimmelpfennig zur Veröffentlichung eingereichten Arbeit dargestellt.

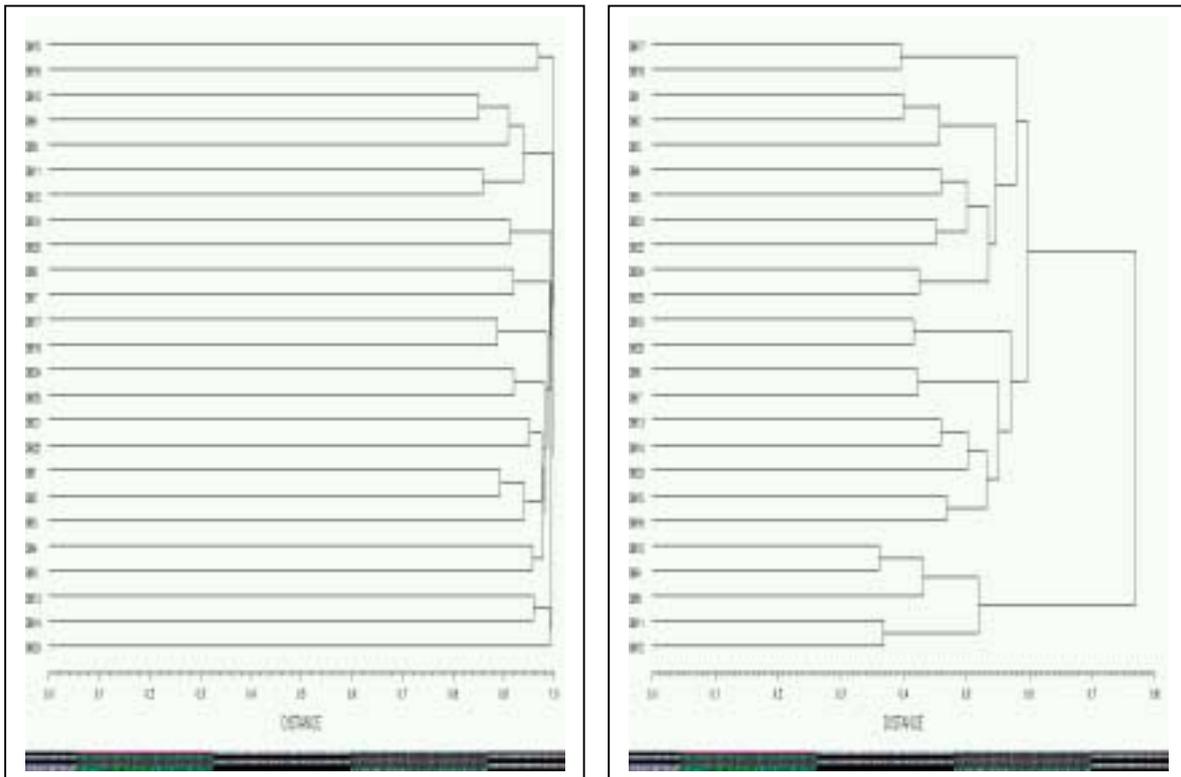


Abbildung 2: Dendrogramme einer Average Linkage- (links) und einer Ward-Lösung (rechts) für die vorliegenden Scannerdaten

### 3.2 Verbundmessung mittels Assoziationsregeln

Eine wahre Renaissance hat die Verbundforschung Mitte der neunziger Jahre im Zusammenhang mit anwendungsorientierten Publikationen zum Data Mining mittels Assoziationsregeln erfahren. Bei diesem Ansatz bedarf es keiner expliziten Annahme über die Symmetrieeigenschaften der betrachteten Verbundbeziehungen. Mit  $A, B \subset I$  können in bekannter Weise der Support<sup>6</sup>

$$\text{sup}(A \rightarrow B) = \frac{|\{T \in W \mid (A \cup B) \subseteq T\}|}{|W|}$$

und die Konfidenz

$$\text{conf}(A \rightarrow B) = \frac{|\{T \in W \mid (A \cup B) \subseteq T\}|}{|\{T \in W \mid A \subseteq T\}|}$$

einer Regel  $A \rightarrow B$  angegeben werden. Eine Verbundbeziehung zwischen  $A$  und  $B$  wird genau dann als möglich erachtet, wenn beide Größen vorzugebende Mindestniveaus überschreiten. Allgemein gilt:

$$\text{conf}(A \rightarrow B) \geq \text{sup}(A \rightarrow B)$$

<sup>6</sup> Vgl. hierzu auch *Agrawal, Srikant (1994)* und *Bollinger (1996)*. Der Übersichtlichkeit halber wird im Folgenden auf eine Indizierung von  $T$  verzichtet.

Wie sich leicht zeigen lässt, handelt es sich beim Support um ein symmetrisches, bei der Konfidenz hingegen um ein asymmetrisches Verbundmaß, was letzteres im Verbundkontext auch interessanter erscheinen lässt. Es gilt nämlich:

$$\text{sup}(A \rightarrow B) = \text{sup}(B \rightarrow A)$$

aber

$$\text{conf}(A \rightarrow B) \neq \text{conf}(B \rightarrow A), \text{ falls } |\{T \in W \mid A \subseteq T\}| \neq |\{T \in W \mid B \subseteq T\}|$$

Ein besonderes Problem bei der Anwendung des Konzepts der Assoziationsregeln stellt die Bestimmung der unter Verbundgesichtspunkten tatsächlich „interessanten“ Regeln dar. Erste Anhaltspunkte in dieser Richtung liefern verschiedene in der Literatur vorgestellte Maße. Der sogenannte Lift

$$\text{lift}(A \rightarrow B) = \frac{\text{conf}(A \rightarrow B)}{\text{sup}(B)}$$

beispielsweise gibt an, um wie viel häufiger bzw. seltener der Regelkopf  $B$  unter der Voraussetzung des Eintretens von  $A$  eintritt als dies in der Grundgesamtheit der Fall ist. Von einer Verbundbeziehung wird ausgegangen, wenn der Lift größer als 1 ist. Die sogenannte Conviction

$$\text{conv}(A \rightarrow B) = \frac{\text{sup}(A) \cdot \text{sup}(\neg B)}{\text{sup}(A \rightarrow \neg B)}$$

liefert darüber hinaus Hinweise auf die Richtung der zwischen  $A$  und  $B$  bestehenden Abhängigkeit, wobei der Support von  $\neg B$  den Anteil der Transaktionen an allen Transaktionen angibt, die die Item-Menge  $B$  nicht enthalten. Die statistische Signifikanz einer Regel schließlich kann mit einem Chi-Quadrat-Test auf Basis der Nullhypothese „ $H_0$ :  $A$  und  $B$  sind voneinander unabhängig.“ und der Teststatistik

$$\chi^2(A \rightarrow B) = |W| \cdot \sum_{u=1}^2 \sum_{t=1}^2 \frac{(\text{sup}(A_t \rightarrow B_u) - \text{sup}(A_t) \cdot \text{sup}(B_u))^2}{\text{sup}(A_t) \cdot \text{sup}(B_u)}$$

mit  $A_1 = A, A_2 = \neg A, B_1 = B$  und  $B_2 = \neg B$  bei einem Freiheitsgrad überprüft werden.

Setzt man den Mindest-Support – der Struktur des vorliegenden Datensatzes entsprechend – mit einem Wert von 0,5 % und die Mindest-Konfidenz mit 10 % an, so können auf dieser Basis z.B. mit Hilfe des SAS Enterprise Miner® eine Reihe interessanter Regeln generiert werden. Die gefundenen und in Tabelle 3 wiedergegebenen Regeln waren größtenteils aufgrund der vorausgegangenen Betrachtungen auch zu erwarten und bestätigen die Leistungsfähigkeit dieses Datenauswertungskonzeptes im Verbundkontext. Die in der Tabelle aufgeführte Regelmengende zeigt keine allzu ausgeprägte Asymmetrie in den einzelnen Verbundbeziehungen, was zumindest teilweise die guten Resultate der Verbundmessung auf Basis von Tanimoto-Koeffizient und Clusteranalyse zu erklären scheint. Aber auch die aufgetretenen Asymmetrien, z.B. die Warengruppen 6 (Haarshampoos) und 12 (Haarkuren) betreffend sind unter Kaufverhaltensgesichtspunkten nachvollziehbar. Mit Blick auf die angegebenen Lift-Werte lassen sich einige „besonders“ interessante Konstellationen identifizieren. Regeln, wie etwa die für die relativ hochpreisige Warengruppenkombination 241 und 240, erweisen sich trotz eines vergleichsweise geringen Supports als sehr interessant. Offensichtlich entscheiden sich die Käufer von After Sun Lotions (Warengruppe 241) auch regelmäßig für den Kauf von Sonnenschutzmitteln (Warengruppe 240). Das Umgekehrte scheint jedoch nicht der Fall zu sein, was aus Anwendersicht auch durchaus plausibel ist. Hier liegt also ganz offensichtlich eine gewisse Verbundasymmetrie vor, was hingegen aufgrund des vergleichsweise geringen Lift-Wertes bei der zuvor genannten Kombination 12 und 6 nicht mit letzter Sicherheit gesagt werden kann. Leider werden aber auch einige vom koeffizientenbasierten Ansatz mehr oder weniger deutlich aufgedeckte Beziehungen im vorliegenden Fall nicht zutage gefördert. Ein offensichtliches Beispiel hierfür

sind die bereits zitierten Gebisspflegemittel aus den Warengruppen 222 und 225. Bemerkenswert ist schließlich noch die Tatsache, dass bei den vorgegebenen Mindestniveaus für Support und Konfidenz keine Regeln mit mehr als zwei Elementen generiert werden.

Regel	conf(.)	lift(.)	Regel	conf(.)	lift(.)
6 → 9	12,01	2,38	186 → 198	11,54	5,48
9 → 6	47,66	2,38	198 → 186	31,88	5,48
12 → 6	24,50	1,22	12 → 9	11,39	2,25
329 → 326	16,96	3,67	9 → 12	11,98	2,25
326 → 329	24,50	3,67	195 → 198	28,69	13,63
186 → 189	18,10	8,04	198 → 195	21,88	13,63
189 → 186	46,78	8,04	189 → 183	19,88	5,10
195 → 186	57,38	9,86	362 → 356	11,07	7,26
186 → 195	15,84	9,86	183 → 189	11,49	5,10
186 → 183	15,38	3,95	356 → 362	29,31	7,26
183 → 186	22,97	3,95	195 → 183	15,57	4,00
563 → 562	17,85	2,65	195 → 189	11,48	5,10
562 → 563	12,33	2,65	241 → 9	11,76	2,33
69 → 70	11,55	3,17	241 → 30	11,76	1,92
70 → 69	20,94	3,17	241 → 240	11,76	18,63
			241 → 329	11,76	1,76

**Tabelle 3:** Regeln und deren Wertigkeit (Support  $\geq 0,5$  % und Konfidenz  $\geq 10$  %)

Senkt man den Mindest-Support indes auf 0,2 % und erhöht man gleichzeitig die Mindest-Konfidenz auf 20 % (letzteres dient hier nur der Begrenzung der generierten Regelmenge), so erhält man ein deutlich gewandeltes Bild. Die in Tabelle 4 dargestellte Lösung zeigt nun eine doch recht ausgeprägte Asymmetrie in einzelnen Verbundbeziehungen. Auffällig ist, dass die Warengruppe 186 (Desserts für Kleinkinder) bei den zweielementigen Regeln ausschließlich im Regelkopf auftaucht. Offensichtlich sind die anderen Kindernahrungsmittel im vorliegenden Fall Auslöser für den Kauf von Kinderdesserts. Das Umgekehrte scheint hingegen nicht der Fall zu sein. Ein Blick auf die Regeln mit drei Warengruppen legt sogar die Vermutung nahe, dass sich die in Abbildung 2 so gut erkennbare Gruppierung der Kindernahrungsmittel (respektive die Warengruppen 183 bis 198) sehr wahrscheinlich im Wesentlichen aufgrund der Beziehungen zur Warengruppe 186 gebildet hat.

Regel	conf(.)	lift(.)	Regel	conf(.)	lift(.)
9 → 6	47,66	2,38	9, 12 → 6	63,04	3,14
12 → 6	24,75	1,22	6, 12 → 9	29,00	5,72
326 → 329	24,50	3,67	195, 198 → 186	57,14	9,80
189 → 186	46,78	8,04	186, 198 → 195	39,22	24,36
195 → 186	57,38	9,86	186, 195 → 198	28,57	13,53
183 → 186	22,97	3,95	183, 189 → 186	47,06	8,07
70 → 69	20,94	3,17	183, 186 → 189	23,53	10,43
198 → 186	31,88	5,48	186, 189 → 183	20,00	5,12
195 → 198	28,69	13,63			
198 → 195	21,88	13,63			
356 → 362	29,31	7,26			

**Tabelle 4:** Regeln und deren Wertigkeit (Support  $\geq 0,2$  % und Konfidenz  $\geq 20$  %)

Abschließend kann festgehalten werden, dass sich das Konzept der Assoziationsregeln – neben der Möglichkeit der Berechnung von Verbundbeziehungen zwischen mehr als zwei Items (auch unterschiedlicher Taxonomieebenen) – vor allem durch seine Anwendbarkeit auf

große Datenmengen und die Differenzierbarkeit zwischen symmetrischem und asymmetrischem Verbund auszeichnet. Darüber hinaus ermöglicht es die implizite, d.h. über die Interessantheitsmaße gesteuerte Isolation zufälliger Mitnahmeeffekte. Als problematisch erweist sich hingegen – wie auch die Gegenüberstellung der Inhalte von Tabelle 3 und 4 zeigt – die zum Teil recht subjektive Festlegung der Mindestniveaus für Support und Konfidenz. Alles in allem hat man es aber mit einem für die exploratorische Analyse „kleiner bis mittlerer“ Sortimente zweifellos sehr nützlichen Instrument zu tun, das die Entscheidungsbasis des Einzelhandelsmanagements zukünftig deutlich zu verbessern vermag.

### 3.3 Verbundmodellierung mittels Self-organizing Maps

Abschließend soll noch kurz auf einen neuen Ansatz zur Modellierung von Verbundbeziehungen mittels sogenannter Self-organizing Maps (Kohonen-Netze), im folgenden mit SOM abgekürzt, eingegangen werden. Ausgangspunkt hierfür ist u.a. die Annahme, dass Verbundbeziehungen in ähnlichen „Warenkorbmustern“ zum Ausdruck kommen und nur auf der Multi-Item-Ebene sinnvoll messbar sind. Die grundsätzliche Idee dieser Form der Modellierung von Verbundbeziehungen bringt Abbildung 3 zum Ausdruck.

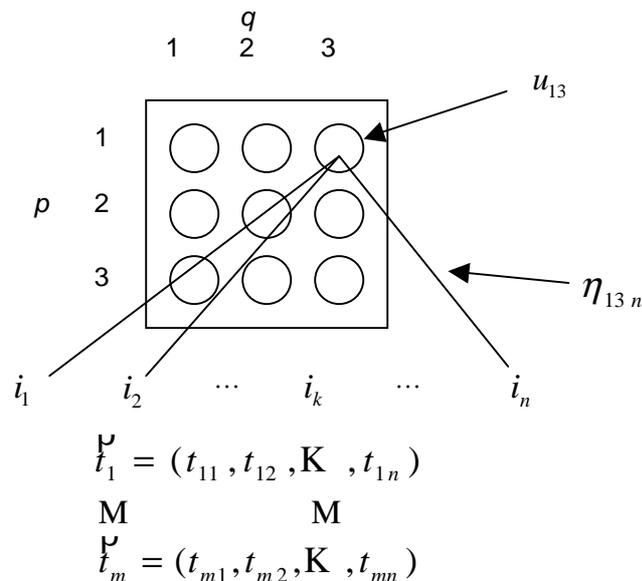


Abbildung 3: Struktureller Aufbau eines 3x3-SOM-Modells

Die Modellparameter  $\eta_{pqk}$  der einzelnen SOM-Units werden dabei so bestimmt, dass sie die beobachteten Transaktionen respektive Warenkörbe bei gegebener Unit-Zahl – im vorliegenden Fall wurde exemplarisch von einem 3x3-Layer ausgegangen – bestmöglich repräsentieren. D.h. jede Gruppe von sich ähnelnden Warenkörben wird nach Möglichkeit durch ein Tupel von Parametern (man spricht in diesem Zusammenhang auch von einem „Prototypen“) repräsentiert. Die simultane Minimierung der die Clusterbildung ( $\rightarrow$  Bestimmung des optimalen Prototypen-Systems  $\{\mathbf{h}_{c_p c_q}^p\}$ ) und die topologische Anordnung der Cluster ( $\rightarrow$  Bestimmung der optimalen Anpassung) betreffenden Zielfunktionen<sup>7</sup>

$$\text{Min} \left( \sum_{j=1}^m d(\mathbf{t}_j^p, \mathbf{h}_{c_p c_q}^p) = \sum_{j=1}^m \min_{p,q} \{d(\mathbf{t}_j^p, \mathbf{h}_{pq}^p)\} \right)!$$

mit

<sup>7</sup> Zur generellen Methodik von SOM-Ansätzen siehe auch Reutterer (1997).

$$d(t_j^p, \eta_{pq}^p) = \|t_j^p - \eta_{pq}^p\|^2 = \sum_{k=1}^n (t_{jk}^p - \eta_{pqk}^p)^2$$

und

$$\text{Min} \left( \sum_{c_p, c_q} d(\eta_{c_p c_q}^p, \eta_{\tilde{c}_p \tilde{c}_q}^p) = \sum_{c_p, c_q} \sum_{\tilde{c}_p, \tilde{c}_q} \sum_{k=1}^n (\eta_{c_p c_q k}^p - \eta_{\tilde{c}_p \tilde{c}_q k}^p)^2 \right)$$

mit

$$d(\eta_{c_p c_q}^p, \eta_{\tilde{c}_p \tilde{c}_q}^p) = \sum_{\tilde{c}_p, \tilde{c}_q} \|\eta_{c_p c_q}^p - \eta_{\tilde{c}_p \tilde{c}_q}^p\|^2 \quad (\text{Nachbarschaftsdistanz für Unit } u_{\tilde{c}_p \tilde{c}_q} \in N_{c_p c_q})$$

ermöglicht eine intuitiv plausible Interpretation der resultierenden Karten unter Verbundgesichtspunkten. Hierbei kommen den einzelnen Bezeichnern die folgenden Bedeutungen zu: Das Tupel  $u_{11}, \mathbf{K}, u_{pq}, \mathbf{K}, u_{PQ}$  kennzeichnet die SOM-Units einer  $P \times Q$ -Repräsentation. Der Vektor  $\eta_{pq}^p = (\eta_{pq1}, \eta_{pq2}, \mathbf{K}, \eta_{pqn})'$  repräsentiert entsprechend die Kantengewichte für Unit  $u_{pq}$ , während  $N_{c_p c_q}$  für die Nachbarschaftsmenge von Winning-Unit  $u_{c_p c_q}$  steht.

Zwei Items stehen demnach in einer gegenseitigen Verbundbeziehung, wenn die betreffenden Kantengewichte einer Unit einen Wert „nahe“ 1 aufweisen. In der vorliegenden Untersuchung wurde die Untergrenze für die Kantengewichte exemplarisch auf 0,8 festgelegt. Darüber hinaus können aber auch Items benachbarter Units (mit entsprechend großen Parameterwerten) als miteinander im Verbund stehend interpretiert werden, da aufgrund der für die topologische Struktur formulierten Optimierungsaufgabe zu erwarten ist, dass ähnliche Prototypen durch benachbarte Units repräsentiert werden. Als Maß für die Güte der erzielten Repräsentation kann die Heterogenität

$$\text{het} \left\{ \eta_{c_p c_q}^p \right\} = \frac{1}{m} \sum_{j=1}^m d(t_j^p, \eta_{c_p c_q}^p)$$

des Prototypen-Systems und die Einfachheit

$$\text{simpl} \left\{ \eta_{c_p c_q}^p \right\} = \sum_{c_p, c_q} d(\eta_{c_p c_q}^p, \eta_{\tilde{c}_p \tilde{c}_q}^p)$$

der Topologie herangezogen werden. Beide Kenngrößen sollten möglichst klein sein.

Wendet man den skizzierten SOM-Ansatz auf die Scannerdaten an, so liefert dies interessante Einblicke in die bestehenden Verbundbeziehungen. Die größte Schwierigkeit stellt hierbei allerdings die geeignete Festlegung der relevanten Steuergrößen<sup>8</sup> und insbesondere die der Unit-Anzahl dar. Eine hinsichtlich beider Gütemaße (Heterogenität und Einfachheit) zufriedenstellende und in Abbildung 4 dargestellte 8x8-Lösung liefert auch eine unter inhaltlichen Gesichtspunkten einsichtige Verbundstruktur. Dies lässt sich z.B. wieder an der geblockten Lage der Kindernahrungsmittel (Warengruppe 183 bis 198) ebenso festmachen wie an der Gruppierung der Tiernahrungsmittel (Warengruppe 69 und 70) oder den Produkten für die Rasur (Warengruppe 326 und 329). Als sehr plausibel kann auch die vergleichsweise breite „Streuung“ der Mitnahmeprodukte aus den Warengruppen 364 (Hustenbonbons) und 366 (Kaugummi) bezeichnet werden. Der traditionelle Clusteransatz in der hier vorgestellten Form liefert in dieser Hinsicht hingegen keine tieferen Erkenntnisse. Die näherungsweise Zweiteilung der Zuordnung der Warengruppe 562 (Augenmakeup) zum einen, in Verbindung mit Warengruppe 563 (Lippenpflege), zu den Haarpflegemitteln und zum anderen zu den Kindernahrungsmitteln ist schließlich ebenso nachvollziehbar wie die wohl verwendungsbedingte Trennung der Hygienewarengruppen 30 und 33. Letzteres scheint im übrigen ein gutes Beispiel für eine substitutive Beziehung zwischen zwei Warengruppen zu sein.

<sup>8</sup> Siehe hierzu auch die entsprechenden Angaben in Abbildung 4.

Der SOM-Ansatz stellt somit gewissermaßen eine interpretationserleichternde Erweiterung des traditionellen Clusteransatzes auf den mehrdimensionalen Fall dar. Die Identifikation zufälliger Mitnahmeeffekte scheint zumindest implizit möglich. Leider stehen der inhaltlichen Plausibilität der SOM-Resultate unter praktischen Gesichtspunkten einige noch nicht zur Gänze gelöste methodische Probleme gegenüber. Hierzu zählen z.B. die Schwierigkeiten im Zusammenhang mit der statistischen Überprüfung der erzielten Resultate. Ein Aspekt, dem gerade unter dem Blickwinkel einer möglichst zuverlässigen Entscheidungsunterstützung in Zukunft auf jeden Fall besondere Aufmerksamkeit zu schenken ist. Alles in allem stellt der SOM-Ansatz aber eine zweifellos interessante und ausbaufähige Alternative für die exploratorische Verbundmodellierung dar.

a/b	1	2	3	4	5	6	7	8
1	366	33 366	33	6 33	6	6	6 326	6 12
2	366 562		33 362	33 364	6 364	6 562	6 222	9 12
3	562	362 562	362	362 368	6 366	6 563	9 562	6 9
4	195	198	356		368	563	9 563	9
5	186 195	186	183	240	222	225	30 366	9 366
6	186 364	189	183 189	12	241	30 33	30	6 30
7	364 562	364 366	69 366	12 69	70	326	30 326	6 329
8	364	69 364	69	69	69 70	326	326 329	329

**Eckdaten der Optimierung:**

Anzahl Iterationen: 1.000.000    Anteil zuordenbarer Warengruppen: 100 %  
 Startwert der Lernrate: 0,7    Heterogenität: 0,656  
 Nachbarschaftsparameter: 3    Einfachheit: 1,861

**Abbildung 4:** Verbundrepräsentation mittels 8x8-Karte

#### 4. Fazit und Ausblick

Die Verbundanalyse war in den letzten Jahrzehnten gekennzeichnet von einem steten Wandel des eingesetzten Instrumentariums. Fokussiert man hierbei auf den Stellenwert der einzelnen Ansätze für die praktische Entscheidungsunterstützung im Einzelhandelsmarketing, so kann zumindest in der vorliegenden Untersuchung eine in jeder Hinsicht eindeutige Überlegenheit eines der betrachteten Instrumente nicht festgestellt werden. Der Stringenz klassifizierender Ansätze steht die inhaltliche Aussagekraft kausaler Modelle gegenüber. Für die Zukunft bleibt anzumerken, dass nach Ansicht des Verfassers vor allem ein dringender Bedarf an objektiven Qualitätskriterien für die Verbundmodellierung besteht. Darüber hinaus stellt die zuverlässige und möglichst vollständige Isolation zufälliger Mitnahmeeffekte aufgrund ihrer besonderen Relevanz für die Entscheidungsunterstützung auch weiterhin eine zentrale Herausforderung an die quantitative Verbundforschung dar.

#### Literatur

**Agrawal, R.; T. Imielinski; A. Swami (1993):** Mining Association Rules between Sets of Items in Large Databases, in: Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data, Washington.

**Agrawal, R.; R. Srikant (1994):** Fast Algorithms for Mining Association Rules, in: Proceedings of the 20th International Conference on Very Large Databases, Santiago, 487 – 499.

**Böcker, F. (1975):** Die Analyse des Kaufverbunds – Ein Ansatz zur bedarfsorientierten Warentypologie, Zeitschrift für betriebswirtschaftliche Forschung, Jg. 27, Heft 5, 290 – 306.

- Böcker, F. (1978):** Die Bestimmung der Kaufverbundenheit von Produkten, Berlin.
- Bollinger, T. (1996):** Assoziationsregeln – Analyse eines Data Mining Verfahrens, Informatik-Spektrum, Jg. 19, Heft 5, 257 – 261.
- Bordemann, H.-G. (1985):** Analyse von Verbundbeziehungen zwischen Sortimentsteilen im Einzelhandel, Duisburg.
- Hruschka, H. (1991):** Bestimmung der Kaufverbundenheit mit Hilfe eines probabilistischen Messmodells, Zeitschrift für betriebswirtschaftliche Forschung, Jg. 43, Heft 5, 418 – 434.
- Lukanowicz, M.; C. Buchta (1992):** Auswirkungen von absatzpolitischen Maßnahmen auf den Sortimentsverbund, Forschungsbericht Nr. 296, Institut für höhere Studien, Wien.
- Merkle, E. (1981):** Die Erfassung und Nutzung von Informationen über den Sortimentsverbund in Handelsbetrieben, Berlin.
- Reutterer, T. (1997):** Analyse von Wettbewerbsstrukturen mit neuronalen Netzen: Ein Ansatz zur Kundensegmentierung auf Basis von Haushaltspaneldaten, Wien.
- Schmalen, H.; H. Pechtl; W. Schweitzer (1996):** Sonderangebotspolitik im Lebensmittel-Einzelhandel, Stuttgart.