

**Analyse von Marktsegmenten mit Hilfe
konnexionistischer Modelle**

Harald HRUSCHKA
Martin NATTER

Forschungsbericht/
Research Memorandum No. 293

März 1992

Die in diesem Forschungsbericht getroffenen Aussagen liegen im Verantwortungsbereich des Autors/der Autorin (der Autoren/Autorinnen) und sollen daher nicht als Aussagen des Instituts für Höhere Studien wiedergegeben werden. Nachdruck nur auszugsweise und mit genauer Quellenangabe gestattet.

All contributions are to be regarded as preliminary and should not be quoted without consent of the respective author(s). All contributions are personal and any opinions expressed should never be regarded as opinion of the Institute for Advanced Studies.

This series contains investigations by the members of the Institute's staff, visiting professors, and others working in collaboration with our departments.

Zusammenfassung

Der Erfolg von Maßnahmen der Marktsegmentierung hängt u.a. vom Einsatz geeigneter Datenanalysetechniken ab. Eine Alternative zu bekannten Verfahren wie der linearen Diskriminanzanalyse oder der logistischen Regression stellen konnexionistische Modelle (künstliche neurale Netzwerke) dar. Sog. Backpropagation-Netzwerke erzielen deutlich höhere Klassifikationsraten als logistische Regressionsmodelle in einem Anwendungsfall der a-priori Segmentierung. Dies dürfte auf die Fähigkeit konnexionistischer Modelle zur Aufdeckung nichtlinearer Beziehungen zwischen Deskriptoren und Segmentzugehörigkeiten sowie von Interaktionseffekten zwischen Deskriptoren zurückzuführen sein. Besondere Aufmerksamkeit wird der Spezifikation konnexionistischer Modelle und der Interpretation der mit deren Hilfe gewonnenen Ergebnisse gewidmet.

Abstract

Success of market segmentation depends on the use of appropriate data analysis techniques. Connectionist models (artificial neural networks) constitute an alternative to well-known methods such as linear discriminant analysis or logistic regression. So-called backpropagation networks attain higher classification rates than logistic regression models in a pilot-study of a-priori segmentation. This may be caused by the capability of connectionist models for discovering nonlinear relationships between descriptors and segment memberships as well as interaction effects between descriptors. Special attention is drawn to specification of connectionist models and interpretation of results obtained from estimating their parameters.

1 Einleitung

Eine Aufgabe der Marktsegmentierung besteht in der Bestimmung von Regeln der Zuordnung von (potentiellen) Abnehmern zu Segmenten auf Grundlage der Eigenschaften von (potentiellen) Abnehmern mit bekannter Segmentzugehörigkeit. Als weitere Aufgabe ergibt sich dann die Bestimmung der Implikationen derartiger Regeln für weitere Marketing-Entscheidungen, etwa mit Hinblick auf Werbebotschaft, Produktgestaltung, Mediaselektion (3). Der Erfolg der Marktsegmentierung hängt somit vom Einsatz geeigneter Datenanalysetechniken ab.

Bei der oben angeführten Bestimmung von Regeln der Zuordnung von Abnehmern zu Segmenten handelt es sich um ein Klassifikationsproblem (14). Als wohl am häufigsten im Bereich der Marktsegmentierung eingesetzte Klassifikationstechnik läßt sich die multiple lineare Diskriminanzanalyse nennen. Zunehmende Verbreitung findet auch die Methode der logistischen Regression.

Eine Alternative zu diesen bekannten statistischen Verfahren stellen konnexionistische Modelle bzw. künstliche neurale Netzwerke dar, die vor allem in der Informatikliteratur verstärkte Beachtung erfahren haben (12,39,42). In jüngster Zeit wendet sich das betriebswirtschaftliche Schrifttum der Thematik von Einsatzmöglichkeiten (18,25,43) bzw. Anwendungserfahrungen (5,8,30,19) konnexionistischer Modelle zu.

Zwischen bestimmten konnexionistischen Modellen und verbreiteten statistischen Verfahren bestehen enge Beziehungen (18). Die lineare Diskriminanzanalyse etwa bildet einen Spezialfall eines konnexionistischen Modells. Der größere Allgemeinheitsgrad bestimmter konnexionistischer Modelle legt daher den hier präsentierten Einsatz als Klassifikationstechnik für Probleme der

a-priori Segmentierung nahe.

Eine Pionieranwendung eines konnexionistischen Modells zur Erklärung von Marktsegmenten stammt von Mazanec (28). Mazanec verwendet die sog. Counter-Propagation als Schätzverfahren zur Bestimmung der Gewichte eines konnexionistischen Modells, das er mit der multiplen linearen Diskriminanzanalyse vergleicht. Dieses konnexionistische Modell basiert nicht auf den ursprünglichen Beobachtungen, sondern verarbeitet 16 Gruppen von Auskunftspersonen, die mit Hilfe eines Clusteranalysealgorithmus aufgrund der Prädiktoren und der Zugehörigkeit zu Marktsegmenten gebildet werden. Pro Gruppe fungieren die Häufigkeiten der Werte einzelner Prädiktoren als Inputvariablen, die Häufigkeiten der entsprechenden Zugehörigkeiten zu Segmenten als Outputvariablen.

2 Marktsegmentierung

Ein Marktsegment entspricht einem homogenen Teilmarkt eines heterogenen Gesamtmarktes. Die Nachfrager eines Marktsegments reagieren in ähnlicher Weise auf den Einsatz von Marketing-Instrumenten.

Im Zusammenhang mit Problemen der Marktsegmentierung hat sich in der wissenschaftlichen Marketing-Literatur eine verhaltensorientierte Forschungsrichtung konstituiert, die Fragestellungen des Aufdeckens und der Erklärung von Marktsegmenten betont (44,10). Als weiteres Spezifikum dieser Richtung läßt sich die Betrachtung hypothetischer, d.h. nicht direkt beobachtbarer, Konstrukte (wie z.B. Einstellungen, Markentreue, soziale Schicht usw.) herausstellen, die in Beziehung zu mehreren beobachteten Variablen stehen.

Die Bildung von Segmenten erfolgt aufgrund der Ausprägungen von Kriteri-

umsvariablen, wobei i.d.R. beobachtbare Kaufverhaltensgrößen oder Präferenzen wegen ihrer Relevanz für die Marketing-Ziele der Anbieter im Vordergrund stehen. Außerdem dienen neben Präferenzen auch weitere psychografische Variablen (wie Bekanntheit, Wahrnehmungen, Einstellungen usw.) als Segmentierungskriterien. Bei der hier behandelten a-priori Segmentierung sind die Segmente (meist durch Ausprägungen nur einer Kriteriumsvariablen) vorgegeben. A-priori Segmentierungen basieren häufig auf der Annahme unterschiedlichen Verhaltens von Käufern und Nicht-Käufern, Intensiv- und Nicht-Intensivverwendern u.ä.

Die Erklärung der Zugehörigkeit zu Marktsegmenten geht von sog. Deskriptorvariablen aus, die sich auf (potentielle) Abnehmer beziehen und (hypothetisch) sowohl mit Kriteriumsvariablen als auch mit Marketing-Zielen in Zusammenhang stehen.

Bei den Deskriptorvariablen lassen sich als Hauptgruppen psychografische und soziodemografische Variablen unterscheiden. Psychografische Variablen können sich ganz generell auf eine Auskunftsperson oder auf deren Verhältnis zu einem Produktfeld bzw. zu Marken beziehen. Zur ersten Kategorie zählen Variablen wie Persönlichkeit oder allgemeiner Lebensstil, deren Einfluß auf das Konsumverhalten oft gering ausfällt. Dies unterscheidet derartige Variablen etwa von Einstellungen von Auskunftspersonen, die eine Marke oder Produktgruppe zum Gegenstand haben.

Ein Verzicht auf soziodemografische Variablen und eine Beschränkung auf psychografische Variablen bei der Untersuchung von Marktsegmenten erscheint beim gegenwärtigen Stand der Forschung zumindest voreilig. Die Bedeutung soziodemografischer Variablen folgt aus dem Weiterbestehen von Unterschieden hinsichtlich Einkommen, Ausbildung, Beruf usw (35,45). So ergibt eine neuere Untersuchung i.d.R. höhere Korrelationen von soziode-

mografischen Variablen mit Verhaltensgrößen wie Produktbesitz bzw. Mediennutzung gegenüber zwei unterschiedlichen Batterien von Werte- bzw. Lebensstilitems (29). In den meisten Fällen dürfte für die Analyse von Marktsegmenten die Kombination von soziodemografischen und psychografischen Variablen von Vorteil sein.

Hinter den durch soziodemografische Variablen identifizierten Unterschieden können freilich divergente Zielvorstellungen, Wertsysteme oder Verhaltensweisen von Abnehmern stehen (9,15). In diesem Fall fungieren entsprechende soziodemografische Variablen als Ersatzindikatoren. So lassen sich häufig subkulturelle Unterschiede des Konsumentenverhaltens etwa in Bezug auf Ortsgrößen (z.B. Großstädte, Kleinstädte usw.) oder Regionen feststellen.

Aus einfachen soziodemografischen Variablen lassen sich Konstrukte ableiten. Ein Beispiel für ein derartiges Konstrukt stellen soziale Schichten dar, d.s. Aggregate von Personen, deren Mitglieder ungefähr den gleichen sozialen Status besitzen (24). Unter Status wird die Wertschätzung verstanden, die einem Menschen im sozialen System entgegengebracht wird. Zu den wichtigsten Schichtungskriterien zählen Beruf, Ausbildung und Einkommen. Im Gegensatz zur Forderung, soziale Schichtungen in Abhängigkeit von Segmentierungskriterien zu bestimmen, bedient sich die Marktforschungspraxis meist einfacher Indizes mit fixen Punktwerten für diverse Ausprägungen der einfachen Variablen. Interessanterweise erweist sich für eine Reihe von Gütern des täglichen Bedarfs die soziale Schichtung einfachen Deskriptoren wie dem Einkommen als überlegen (34).

3 Konnexionistische Modelle

Konnexionistische Modelle bestehen aus Einheiten, die über Verbindungen, deren Stärke durch Gewichte angegeben wird, interagieren (12). Als Synonym für konnexionistische Modelle findet man den Ausdruck künstliche neuronale Netzwerke. Die Einheiten werden dann als (künstliche) Neuronen bezeichnet. Hier sollen lediglich die später eingesetzten Backpropagation-Netzwerke mit verborgenen Einheiten erörtert werden. Umfassendere Einführungen in konnexionistische Modelle findet der interessierte Leser in (32,40) oder (17,27). Mit dem Einsatz potentiell geeigneter konnexionistischer Modelle zur Datenanalyse im Marketing befaßt sich (18) .

Bei konnexionistischen Modellen wirkt eine Menge von Einheiten mit Zustandswerten x_1, x_2, \dots, x_n auf eine andere Einheit k ein. Jeder Zustandswert wird mit dem ihm zugeordneten Gewicht w_1, w_2, \dots, w_n multipliziert. Positive Gewichte bezeichnen exzitatorische, negative Gewichte inhibitorische Verbindungen der jeweiligen Einheiten. Die Summe der derartig gewichteten Zustandswerte ergibt dann das Potential z_k der Einheit k :

$$z_k = \sum_{i=1}^n w_i x_i \quad (1)$$

Die Verarbeitung des Potentials z_k durch eine Transferfunktion ergibt schließlich den Zustand der Einheit, der diskrete (oft binäre) oder kontinuierliche Werte annehmen kann. Die hier erörterten Modelle verwenden als Transferfunktion folgende logistische Funktion:

$$y_k = \frac{1}{1 + \exp^{-z_k}} \quad (2)$$

Der Zustand y_k dieser logistischen Funktion liegt im Intervall $[0, 1]$. Für die infinitesimale Änderung des Zustands in bezug auf das Potential gilt:

$$\frac{\delta y_k}{\delta z_k} = y_k(1 - y_k) \quad (3)$$

Die logistische Transferfunktion führt zu geringen Änderungen bei Werten des Zustands nahe Null oder Eins, zu großen Änderungen bei Werten des Zustands nahe 0.5.

Die logistische Transferfunktion kann auch einen konstanten Term aufweisen, dessen Zustand auf Eins fixiert ist. Ein positiver (negativer) Wert des konstanten Terms verschiebt die Transferfunktion nach rechts (links), d.h. bewirkt höhere (niedrigere) Zustandswerte. Daher läßt sich der mit minus Eins multiplizierte konstante Term als Schwellenwert für die Inputs einer Einheit interpretieren.

Schichten (Layers) zerlegen die Einheiten in disjunkte Teilmengen, wobei bei den meisten Modellen Verbindungen nur zwischen Einheiten unmittelbar aufeinanderfolgender Schichten vorliegen. Die hier betrachteten Modelle bestehen aus drei Schichten, die sich jeweils aus Inputeinheiten (Inputvariablen), verborgenen Einheiten und Outputeinheiten (Outputvariablen) zusammensetzen. Bei Inputeinheiten erfolgt keine Summierung, sie geben lediglich Werte an die Gewichte ihrer Verbindung zu verborgenen Einheiten weiter. Verborgene Einheiten unterscheiden sich von Input- und Outputeinheiten durch Abgeschlossenheit gegenüber der Außenwelt. Bei Backpropagation-Modellen beginnt die Berechnung der Zustände mit der ersten Schicht verborgener Einheiten, sie setzt dann mit der unmittelbar folgenden Schicht fort, um mit der Schicht der Outputvariablen zu enden.

Konnexionistische Modelle mit verborgenen Schichten und linearen Trans-

ferfunktionen lassen sich in äquivalente lineare Modelle ohne verborgene Schichten transformieren. Lineare konnexionistische Modelle können den Inputraum nur durch Hyperebenen in Inputregionen aufteilen. Selbst relativ einfache Probleme, wie das exklusive Oder, sind auf diese Weise nicht lösbar. Diese Beschränkung gilt selbstverständlich auch für analoge lineare statistische Modelle (wie lineare Regressionsanalyse oder Diskriminanzanalyse), die in der Marketing-Forschung häufig verwendet werden.

Die verborgenen Einheiten konnexionistischer Modelle erinnern an die latenten Variablen von linearen Strukturgleichungsmodellen des LISREL- oder PLS-Typs (22,2). Folgt man einer weiten Definition einer latenten Variablen, so kann man zwei Formen unterscheiden (21). In der ersten Form stellt eine latente Variable die Ursache einer oder mehrerer direkt beobachteter Variablen dar, deren empirischer Gehalt nur indirekt durch Überprüfung der Zusammenhänge zwischen beobachtbaren Variablen geschätzt werden kann. Eine enge Definition einer latenten Variablen beschränkt sich im übrigen auf diese erste Form (4). In der zweiten Form ist eine latente Variable die Folge einer oder mehrerer beobachtbarer Variablen. Falls alle (auch die verborgenen) Einheiten eines konnexionistischen Modells lineare Transferfunktionen besitzen, entspricht dessen algebraische Struktur der eines linearen Strukturgleichungsmodells mit latenten Variablen. Nachfolgend wird die Bezeichnung latente Variable synonym zum Begriff verborgene Einheit verwendet.

Konnexionistische Modelle mit latenten Variablen und einer nichtlinearen Transferfunktion (wie der hier verwendeten logistischen Funktion) sind mächtiger als lineare Modelle, da sie konvexe Regionen des Inputraumes bilden können. Backpropagation-Modelle mit einer Schicht verborgener Einheiten besitzen außerdem die Fähigkeit, jede kontinuierliche Funktion zu approximieren (6,11).

Zur Schätzung der Gewichte der in diesem Beitrag erörterten konnexionistischen Modelle dient das Backpropagation-Verfahren (33). Die schrittweise Anpassung der Gewichte von Verbindungen beginnt dabei mit den Outputeinheiten. Die Fehler zwischen tatsächlichen und berechneten Outputs werden dann Schicht für Schicht bis zurück zu den Inputeinheiten fortgepflanzt (propagated).

Als zu minimierende Zielfunktion geht man meist vom Fehlermaß E aus, das der Hälfte der Summe der quadratischen Differenzen zwischen tatsächlichen t_{kp} und berechneten Outputwerten y_{kp} über alle p Beobachtungen entspricht. In diesem Fall ist Backpropagation eine Kleinstquadratprozedur.

$$E = \frac{1}{2} \sum_p (t_{kp} - y_{kp})^2 \quad (4)$$

Vor der Schätzung werden die Gewichte mit kleinen Zufallszahlen (hier im Intervall $[-0.1, +0.1]$) initialisiert. Der Schätzalgorithmus läuft dann einige Iterationen lang, die jede eine Vorwärts- und eine Rückwärtsphase aufweisen. Während der Vorwärtsphase werden Outputwerte Schicht für Schicht, beginnend mit den Inputeinheiten, berechnet.

Der Index p für Beobachtungswerte entfällt im weiteren. Während der Rückwärtsphase werden die Gewichte w_{ij} der Verbindungen wie folgt verändert:

$$\Delta w_{ij} = -\eta \frac{\delta E}{\delta w_{ij}} \quad (5)$$

Dabei ist η eine Lernkonstante im Intervall $[0.1, 1.0]$.

Dieser Ausdruck zeigt deutlich, daß Backpropagation ein Gradientenverfahren ist. Mit Hilfe der Kettenregel ermittelte partielle Ableitungen des

Fehlermaes E nach Gewichten fhren zur sog. verallgemeinerten Delta-Regel. Die Ausfhrungen beschrnken sich auf die entsprechenden Hauptergebnisse.

Zunchst wird der Schtzvorgang fr Gewichte zwischen verborgenen und Outputeinheiten beschrieben. Der Fehler zwischen tatschlichen und berechneten Werten $t_k - y_k$ der Outputeinheit k multipliziert mit der Ableitung der logistischen Funktion ergibt δ_k :

$$\delta_k = y_k(1 - y_k)(t_k - y_k) \quad (6)$$

Das neue Gewicht fr die Verbindung der verborgenen Einheit j mit der Outputeinheit k ergibt sich durch Addition von $\eta\delta_k$ multipliziert mit dem Zustand y_j von j zu dessen alten Wert (n ist der Iterationszhler):

$$w_{jk}^{n+1} = w_{jk}^n + \eta\delta_k y_j \quad (7)$$

Der Schtzvorgang fr Gewichte, die Inputeinheiten mit verborgenen Einheiten verbinden, mu ohne externe Outputwerte auskommen. Der Faktor $\delta E/\delta y_j$ kann nicht direkt, sondern nur ausgehend von anderen, bekannten Werten bestimmt werden.

Der bei den Outputeinheiten berechnete Fehler pflanzt sich durch Summieren der Produkte $\delta_k w_{jk}$ ber alle mit dem verborgenen Neuron j verbundenen Outputeinheiten auf die Einheit j fort. δ_j entspricht schlielich dieser Summe multipliziert mit der Ableitung der logistischen Funktion fr die verborgene Einheit j .

$$\delta_j = y_j(1 - y_j)\sum_k \delta_k w_{jk} \quad (8)$$

Neue Gewichte erhält man durch Addition von $\eta\delta_j$ multipliziert mit dem Input x_i zu den Werten aus der vorangegangenen Iteration.

$$w_{ij}^{n+1} = w_{ij}^n + \eta\delta_j x_i \quad (9)$$

Eine Zerstörung bereits bestimmter Zusammenhänge findet bei diesem Verfahren nicht statt, falls die Gewichte erst nach Kumulierung der Änderungen über alle Beobachtungswerte geändert werden. Anschließend beginnt eine neue Iteration mit der Vorwärtsphase, solange bis ein Stopkriterium erfüllt wird.

Der hier verwendete Backpropagation-Algorithmus zeichnet sich durch einige Erweiterungen aus, die zu einer deutlichen Beschleunigung des Schätzvorgangs führen. Zunächst wird ein Verzögerungsterm α hinzugefügt (16). Für die Änderung eines Gewichts gilt dann:

$$\Delta w_{ij}^{n+1} = -\eta \frac{\delta E}{\delta w_{ij}} + \alpha \Delta w_{ij}^n \quad (10)$$

Im Unterschied zum grundlegenden Backpropagation-Verfahren werden außerdem sowohl der Lernparameter η als auch der Verzögerungsterm α nicht konstant gehalten, sondern durch gleichverteilte Zufallszahlen im Intervall $[0, 0.1]$ bzw. $[0, 0.9]$ variiert.

Beim Einsatz konnexionistischer Modelle ist die Frage der Modellspezifikation zu lösen, d.h. insbesondere der Bestimmung von Art und Anzahl latenter Variablen sowie der Verbindungen zwischen den betrachteten Variablen eines Modells. White (41) schlägt vor, zu diesem Zweck in der Statistik bewährte Kreuzvalidierungsmethoden einzusetzen (37). Kreuzvalidierte Klassifikatoren ignorieren bei der Klassifikation einer Beobachtung die mit dieser Beobach-

tung verbundene Information und liefern auf diesem Weg ein Maß für die Klassifikationsgüte außerhalb der betrachteten Stichprobe.

Die kreuzvalidierte Klassifikationsrate entspricht dem Prozentsatz richtiger Klassifikationen der p Beobachtungswerte durch den Klassifikator, der lediglich durch Verarbeitung der restlichen $p-1$ Beobachtungen gewonnen wurde. Als weitere Gütemaße seien die mittlere und die maximale Klassifikationsrate genannt. Die mittlere Klassifikationsrate entspricht dem Mittelwert der Prozentsätze richtiger Klassifikationen dieser auf jeweils $p-1$ Beobachtungen zurückgehenden p Klassifikatoren für alle p Beobachtungswerte. Die maximale Klassifikationsrate ist der Prozentsatz richtiger Klassifikationen eines Klassifikators, der auf allen p Beobachtungswerten basiert. Klarerweise ist die kreuzvalidierte Klassifikationsrate kleiner als die mittlere Klassifikationsrate, letztere wiederum liegt unter der maximalen Klassifikationsrate, da sich hier Klassifikator und Klassifikation auf dieselbe Stichprobe beziehen.

Bei der in Abschnitt 6 erörterten empirischen Untersuchung dient die kreuzvalidierte Klassifikationsrate zur Bestimmung und Auswahl geeigneter Klassifikationsmodelle.

4 Klassifikation mittels Backpropagation

Bei Klassifikationsproblemen nehmen die jeweils einer Klasse entsprechenden Outputvariablen eines konnexionistischen Modells Werte im Einheitsintervall (d.h. zwischen Null und Eins) an. Für die Erklärung einer a-priori Segmentierung ergibt sich dann je eine Outputvariable pro Segment, wobei ein Wert von Eins (Null) für Zugehörigkeit (Nicht-Zugehörigkeit) der Auskunftsperson zum jeweiligen Segment steht. Als Inputvariablen dienen die zu untersuchenden Deskriptoren zur Erklärung der a-priori Segmentierung.

Entsprechende lineare konnexionistische Modelle ohne verborgene Einheiten sind nur eine andere Darstellungsform der wohlbekannten multiplen linearen Diskriminanzanalyse, wobei allerdings i.d.R. andere (rekursive) Schätzverfahren (Perzeptron- bzw. Widrow-Hoff-Algorithmus) zur Bestimmung der Gewichte (Diskriminanzkoeffizienten) Anwendung finden (7).

Erfolgt die Bestimmung der Gewichte eines konnexionischen Modells durch Minimierung eines Kleinstquadratkriteriums, so stellen die auf diese Weise geschätzten Outputgrößen Approximationen an a-posteriori Wahrscheinlichkeiten der Klassenzugehörigkeiten dar (36). Unabhängig von der verwendeten Transferfunktion approximieren derartige konnexionistische Modelle die optimale Bayes'sche Diskriminanzfunktion (31). Die mit Hilfe eines solchen konnexionistischen Modells für bestimmte Inputs ermittelten Outputwerte sind Schätzer der bedingten Verteilung der Klassenzugehörigkeiten. Werden diese Werte normiert, so erhält man die bedingte Verteilung der geschätzten Zufallsvariablen der Zugehörigkeiten, die als Bewertung des Vertrauens in die Klassifikation interpretiert werden kann. Die im letzten Abschnitt erörterten Approximationseigenschaften konnexionistischer Modelle mit nichtlinearen Aktivierungsfunktionen und einer Schicht von latenten Variablen bzw. verborgenen Einheiten lassen relativ genaue Näherungen der Klassenzugehörigkeiten erwarten (38).

5 Klassifikation mittels logistischer Regression

Das Ziel der logistischen Regression besteht ebenfalls in einer guten Annäherung an die Bayes'sche Diskriminanzfunktion. Die logistische Regression basiert auf der Annahme, daß die logarithmierten bedingten Wettchancen

auf die Zugehörigkeit zu einer Klasse (einem Segment) g gegenüber der Zugehörigkeit zur Klasse G eine lineare Funktion der Prädiktoren (Deskriptorvariablen) darstellen (1). Die Wettchance entspricht dem Quotienten aus der Wahrscheinlichkeit der Zugehörigkeit zur Klasse g (L_g) und der Wahrscheinlichkeit der Zugehörigkeit zur Klasse G (L_G). Bei G Klassen (Segmenten) gilt:

$$\frac{\ln L_g}{\ln L_G} = b_0^g + \sum_i b_i^g x_i \quad \text{mit } g = 1, \dots, G - 1 \quad (11)$$

b_0^g konstanter Term für Gruppe g

b_i^g Koeffizient der Inputvariablen x_i für Gruppe g

Die Schätzung der Koeffizienten der logistischen Regression erfolgt i.d.R. mit Hilfe einer Maximum-Likelihood-Methode. Nach der Bestimmung der Koeffizienten erfordert die Klassifikation einer Auskunftsperson nur die Berechnung der Werte folgender linearer Diskriminanzfunktionen:

$$z_g = b_0^g + \sum_i b_i^g x_i + \ln(p_g/p_G) \quad \text{mit } g = 1, \dots, G - 1 \quad (12)$$

p_g relative Häufigkeit der Gruppe g

Die Zuordnung erfolgt zu jener Gruppe, für die die Diskriminanzfunktion z_g für $g = 1, \dots, G$ maximal ist (wobei $z_G = 0$ gilt).

Die logistische Regression dient als Vergleichsbasis für die konnexionistischen Modelle, da bei der Marktsegmentierung meist nominale oder ordinale Prädiktoren überwiegen. Im Gegensatz zur logistischen Regression liefert die häufiger verwendete multiple lineare Diskriminanzanalyse bei nicht normalverteilten Prädiktoren inkonsistente Schätzungen (26). Ein weiterer Man-

gel gegenüber der logistischen Regression ist die Gefahr erheblicher Fehlklassifikationen bei nicht-metrischen Prädiktoren (13).

6 Empirische Untersuchung

6.1 Problemstellung und Daten

Die zu analysierende a-priori Segmentierung ergibt sich durch die Kriteriumsvariable Erstpräferenz einer Zufallsstichprobe österreichischer Hausfrauen in Bezug auf fünf Marken (A, B, C, D, E) der Produktgruppe Haushaltsreiniger. Ein Segment entspricht dann jenen Auskunftspersonen, die alle dieselbe Marke als am stärksten bevorzugte Marke angeben. Als diese Erstpräferenzsegmente möglicherweise erklärende Deskriptoren werden sowohl psychografische Variablen (Items, die die Einstellung zur Hausarbeit bzw. zu Haushaltsreinigern messen) als auch soziodemografische Variablen (wie Einkommen, Schulbildung, Ausmaß der Berufstätigkeit usw.) postuliert (vgl. dazu Tabelle 1). Sowohl in den logistischen als auch in den konnexionistischen Modellen werden die Deskriptoren durch binäre Dummyvariablen repräsentiert.

Wegen der ungleichmäßigen Verteilung auf die betrachteten 5 Segmente werden für die Klassifikation die ursprünglich 435 Auskunftspersonen auf 215 reduziert. Pro Segment steht dann dieselbe Anzahl (nämlich 43) zur Verfügung. Für diese Vorgangsweise spricht die Tatsache, daß (ungefähr) gleich große Beobachtungszahlen je Klasse zu besseren Klassifikationsergebnissen führen (20).

Table 1: Verwendete Deskriptoren

Einstellung zur Hausarbeit generell	8 Items
Einstellung zu Reinigern	4 Items
Alter	20-29 30-39 40-49 50-59
Haushaltsgröße	1 bis mindestens 5 Personen
Anzahl Kinder	0 bis mindestens 3
Schulbildung	Volks- und Hauptschule Fachausbildung Matura oder Universität
Berufstätigkeit	Ganztägig Teilzeit keine
Zweitwohnung	ja/nein
Ortsgröße	bis 2000 bis 5000 bis 50000 über 50000
Anzahl der Einkommensbezieher des Haushalts	1 oder mindestens 2
Einkommensklasse des Haushalts	4 Kategorien

6.2 Ergebnisse der logistischen Regressionsmodelle

Tabelle 2 gibt an, welche Deskriptorvariablen im jeweiligen logistischen Regressionsmodell enthalten sind. Berücksichtigt werden nur jene Deskriptoren, deren Kontingenzkoeffizient mit der beobachteten Segmentzugehörigkeit größer als 0.18 ist. Die jeweils 4 Gleichungen weisen bei allen Modellen einen konstanten Term auf.

Table 2: Spezifikation der Logistischen Modelle

Deskriptor	Modellkennung				
	L1	L2	L3	L4	L5
Einstellung 2	x	x	x		
Einstellung 12	x	x	x		
Einstellung 14	x	x	x	x	x
Haushaltsgröße	x	x	x		
Schulbildung	x				
Berufstätigkeit	x	x	x	x	
Ortsgröße	x	x	x		
Einkommensbezieher	x	x	x		
Einkommen	x				

Tabelle 3 zeigt, daß logistische Modelle mit steigender Anzahl von Deskriptoren zu einem höheren Prozentsatz korrekter Klassifikationen führen. Allerdings sind die meisten Koeffizienten der Modelle L1, L2 und L3 nicht signifikant von Null verschieden. Das Modell L4 enthält jene zwei Deskriptoren (Einstellung 14, Berufstätigkeit) des Modells L3, für die die Nullhypothese, daß deren Koeffizienten in allen vier Gleichungen gleich Null sind, aufgrund der Ergebnisse geeigneter näherungsweise Chi-Quadrat-verteilter Likelihood-

Table 3: Klassifikationsgüte der logistischen Modelle

Modell	Parameter	maximale Klassifikationsrate	Prediction Criterion
L1	76	42.3	1.21
L2	60	36.7	1.12
L3	52	32.6	1.10
L4	16	30.7	0.80
L5	8	25.6	0.80

Ratio-Tests mit einer Irrtumswahrscheinlichkeit von 0.05 verworfen werden kann.

Daneben wird das Prediction Criterion als Gütemaß, das auch die unterschiedliche Anzahl von Parametern verschiedener Modelle bewertet, angegeben. Es wurde von Amemiya ausgehend vom mittlerem quadratischen Fehler entwickelt und bestraft eine höhere Parameteranzahl stärker als das bei der multiplen linearen Regressionsanalyse verwendete korrigierte Bestimmtheitsmaß. Für die Anzahl von Fehlklassifikationen f , Beobachtungen p und Parametern l berechnet es sich wie folgt (23):

$$\frac{f}{p-l} \left(1 + \frac{l}{p}\right) \quad (13)$$

Das Prediction Criterion erreicht den optimalen Wert von Null, falls alle Beobachtungen der richtigen Klasse zugeordnet werden. Es steigt mit der Anzahl von Fehlklassifikationen und dem Quotienten aus Parameteranzahl und Anzahl der Beobachtungen.

Das Prediction Criterion nimmt für die hier geschätzten logistischen Modelle

L1, L2 und L3 relativ ungünstige (hohe) Werte an und sinkt mit abnehmender Zahl von Deskriptoren bzw. Koeffizienten. Dies würde für Modelle mit einer geringen Anzahl von Deskriptorvariablen sprechen. Jedoch klassifiziert das Logitmodell L5 mit nur einem Deskriptor (dem Einstellungsitem 14) alle Auskunftspersonen alternativ in lediglich zwei der fünf a-priori Segmente (Erstpräferenz für Marke A bzw. Marke E). Eine Zuordnung zu den anderen drei Segmenten findet fälschlicherweise nicht statt. Darüberhinaus liegt der Prozentsatz richtiger Klassifikationen insbesondere für Modell L5 nur unwesentlich über jenem Wert, der bei rein zufälliger Klassifikation zu erwarten wäre (dieser beträgt bei 5 Klassen 20 Prozent). Die kreuzvalidierten Klassifikationsraten der Modelle L4 und L5 liegen auch nur um die 20 Prozent. Insgesamt gesehen schließen diese unbefriedigenden Klassifikationsergebnisse für logistische Modelle die Sinnhaftigkeit einer näheren Interpretation oder gar einer Ableitung von Schlußfolgerungen für Marketing-Entscheidungen aus.

6.3 Ergebnisse der konnexionistischen Modelle

Die Parameteranzahl der untersuchten konnexionistischen Modelle liegt nicht über jener der zuvor behandelten logistischen Modelle. Im Gegensatz zu vielen Anwendungen konnexionistischer Modelle bestehen keine vollen Verbindungen zwischen Inputvariablen (Deskriptoren) und latenten Variablen. Jeder Deskriptor steht mit nur einer latenten Variablen in Beziehung. Dieser vergleichsweise restriktive Aufbau reduziert die Anzahl der Parameter und erleichtert die Interpretation der geschätzten konnexionistischen Modelle. Die latenten Variablen sind mit den fünf Segmentzugehörigkeiten voll verbunden. Die Schätzung der Gewichte erfolgt mit Hilfe des in Abschnitt 3 dargestellten Backpropagation-Algorithmus.

Die gebildeten latenten Variablen entsprechen Gruppen verwandter Deskrip-

toren und lassen sich als psychografische bzw. soziodemografische hypothetische Konstrukte interpretieren (vgl. Tabelle 4).

Table 4: Latente Variablen

Latente Variable	Deskriptoren
1 Einstellung zur Hausarbeit	5 Items
2 Einstellung zu Reinigern	4 Items
3 Haushaltgröße	Dummyvariablen für 5 Kategorien
4 Ortsgröße	Dummyvariablen für 4 Kategorien
5 Soziale Schichtung	Dummyvariablen für Schulbildung, Berufstätigkeit, Einkommensbezieher, Einkommensklasse

Modell K1 umfaßt alle in Tabelle 4 angegebenen Deskriptoren und latenten Variablen. Sowohl die latenten Variablen als auch die Segmentzugehörigkeiten besitzen konstante Terme. In Tabelle 5 findet man einige Meßgrößen der Klassifikationsgüte dieses Modells. Der Vergleich der Werte für das Prediction Criterion ebenso wie für die maximale und die kreuzvalidierte Klassifikationsrate zeigt die Überlegenheit gegenüber den zuvor behandelten logistischen Modellen.

Tabelle 6 enthält kreuzvalidierte Klassifikationsraten für diverse Modelle, die sich vom Modell K1 durch Entfernen jeweils eines Deskriptors oder einer Gruppe von Deskriptoren unterscheiden. Dabei fällt auf, daß die kreuzvalidierte Klassifikationsrate von ursprünglich 34.1 Prozent bei Entfernung bestimmter Einstellungsitems bzw. aller konstanter Terme (sowohl der latenten Variablen als auch der Segmentzugehörigkeiten) nicht oder nur wenig abnimmt. Daher wird Modell K1 mit zwei weiteren Modellen (K2 und K3)

näher verglichen. Modell K2 wird aus K1 durch Entfernung der Einstellungsitems 4, 11 und 2 sowie der konstanten Terme für die Segmentzugehörigkeiten gebildet. Modell K3 verzichtet darüberhinaus auch auf konstante Terme für die latenten Variablen.

Table 5: Klassifikationsgüte ausgewählter konnexionistischer Modelle

Modell	Parameter	Klassifikations- rate			Prediction Criterion
		Maximum	Mittel	Kreuzvalidiert	
K1	58	47.4	42.1	34.1	0.91
K2	50	44.2	39.0	34.4	0.90
K3	45	40.5	39.2	33.6	0.91

Die folgende Interpretation beschränkt sich auf Modell K2 wegen dessen höherer kreuzvalidierter Klassifikationsrate. Der erste Interpretationsschritt hat die Beziehung zwischen Deskriptoren und latenten Variablen zum Gegenstand. Zu diesem Zweck wird bestimmt, welche Werte die jeweilige latente Variable für diverse Kombinationen der Ausprägungen der mit ihr in Zusammenhang stehenden Deskriptoren annimmt.

Die latente Variable 1 kann man als positive Einstellung zur Hausarbeit bezeichnen. Sie basiert auf folgenden Items:

- Ich schätze es, wenn mir meine Familie bei der Hausarbeit hilft;
- Wenn man nicht auf unbedingte Sauberkeit im Haushalt achtet, können leicht Krankheiten übertragen werden;
- Ich empfinde Hausarbeit als ausgesprochen lästig.

Table 6: Kreuzvalidierte Klassifikationsraten

bei Entfernung von	Klassifikationsrate
Einstellung 4	33.7
Einstellung 6	32.6
Einstellung 10	32.0
Einstellung 11	37.2
Einstellung 12	28.2
Einstellung 2	35.3
Einstellung 13	28.2
Einstellung 14	25.9
Einstellung 15	30.0
Einstellung zur Hausarbeit (Items 4-12)	28.2
Einstellung zu Reinigern (Items 2-15)	30.4
Haushaltsgröße	26.8
Ortsgröße	28.2
Schulbildung	30.5
Beruf	31.7
Einkommensbezieher	32.9
Einkommensklasse	29.4
allen konstanten Termen	33.7

Die latente Variable 1 nimmt den Wert Eins an, falls den ersten beiden Items zugestimmt wird und dem dritten Item nicht zugestimmt wird, bzw. den Wert Null, falls die Auskunftsperson genau umgekehrt antwortet.

Die latente Variable 2 kann man als negative Einstellung zu Haushaltsreinigern auffassen. Die entsprechenden Items lauten:

- Die meisten Reinigungsmittel sind zu scharf;
- Für verschiedene Reinigungsarbeiten im Haushalt braucht man mehrere spezielle Putzmittel;
- Ich probiere gerne neue Putzmittel aus.

Die Werte dieser latenten Variablen liegen zwischen 0.683 (erstes Item ja, zweites und drittes Item nein) und Null (erstes Item nein, zweites und drittes Item ja).

Die Gewichte der Verbindungen der durch vier binäre Dummyvariablen gemessenen fünf Haushaltsgrößen mit der latente Variable 3 reproduzieren eine nichtlineare Funktion, wobei sich für Haushalte mit einer oder mindestens vier Personen Werte um Eins, sonst Werte um 0.90 für die latente Variable 3 ergeben.

Die mit drei binären Dummyvariablen für vier Ortsgrößen in Verbindung stehende latente Variable 4 nimmt bei der kleinsten Ortsgröße einen Wert von 0.76, sonst Werte zwischen 0.94 und 0.99 an.

Die latente Variable 5 repräsentiert eine kontinuierliche soziale Schichtung der Auskunftspersonen. Tabelle 7 gibt den Zusammenhang zwischen den Werten dieser latenten Variablen und Ausprägungen der Indikatoren Ausbildung, Berufstätigkeit, Einkommensbezieher und Einkommensklasse an. Die latente

Variable 5 nimmt bei Auskunftspersonen mit Matura ausschließlich mittlere und niedrige Werte an, mittlere Werte bei (teilweiser) Berufstätigkeit der Auskunftsperson und geringerem Einkommen je Einkommensbezieher. Bei Berufsschulbildung nimmt die latente Variable 5 mit steigendem Einkommen je Einkommensbezieher bzw. bei Teilzeitbeschäftigung geringere Werte an. Ähnlich verhält es sich bei Hauptschulbildung, wobei niedrige Werte aber auch bei Teilzeitbeschäftigung, einem Einkommensbezieher und niedrigem Einkommen auftreten.

Der zweite Interpretationsschritt betrifft die Beziehungen zwischen latenten Variablen und Segmentzugehörigkeiten bzw. der Erstpräferenz (diese entspricht dem maximalen Zugehörigkeitswert). Zu diesem Zweck werden die Werte der fünf Outputvariablen für diverse Kombinationen von Ausprägungen der fünf latenten Variablen ermittelt (vgl. Tabellen 8,9).

Personen mit Erstpräferenz für Marke A lassen sich durch positive Einstellung sowohl zur Hausarbeit als auch zu Reinigern sowie durch Wohnsitz in einem größerem Ort charakterisieren.

Beim Segment mit einer Erstpräferenz für Marke B nimmt die latente Variable 5 meist den Wert Null an. Dies trifft auf Personen mit Matura zu, die einerseits teilzeitbeschäftigt sind, über ein mittleres Einkommen bei einem Einkommensbezieher (ein hohes Einkommen bei mehreren Einkommensbeziehern) verfügen oder nicht berufstätig sind. Weiters trifft dies auf Personen mit Berufschulabschluß zu, die teilzeitbeschäftigt sind bei hohem Einkommen und einem Einkommensbezieher im Haushalt. Darunter fallen schließlich auch Personen mit Hauptschulabschluß, die teilzeitbeschäftigt oder nicht berufstätig sind, ein hohes Einkommen bei einem Einkommensbezieher aufweisen oder teilzeitbeschäftigt bei niedrigem Einkommen und einem Einkommensbezieher sind.

Table 7: Zusammenhang der latenten Variablen 5 (soziale Schichtung) mit Indikatoren

Ausbildung	Berufstätigkeit	Einkommens- bezieher	Einkommen	Latente Variable
Matura	Ganztätig	ein	niedrig	0.37 bis 0.47
Matura	Teilzeit	mehrere	mittel	0.37 bis 0.47
Matura	Teilzeit	ein	mittel	0.0 bis 0.17
Matura	Teilzeit	mehrere	hoch	0.0 bis 0.17
Matura	keine	ein		0.0 bis 0.17
Berufsschule	Ganztätig	mehrere		1.0
Berufsschule	Ganztätig	ein	mittel	1.0
Berufsschule	Teilzeit	mehrere	niedrig	1.0
Berufsschule	Teilzeit	ein	mittel	0.37 bis 0.47
Berufsschule	Teilzeit	ein	hoch	0.0 bis 0.17
Hauptschule	keine	mehrere	niedrig	1.0
Hauptschule	Ganztätig	mehrere	mittel	1.0
			hoch	
Hauptschule	Teilzeit	mehrere	hoch	0.37 bis 0.47
Hauptschule	keine	ein	mittel	0.37 bis 0.47
Hauptschule	Teilzeit	ein	hoch	0.0 bis 0.17
	keine			
Hauptschule	Teilzeit	ein	niedrig	0.0 bis 0.17

Table 8: Beziehungen zwischen latenten Variablen und Segmentzugehörigkeiten

Latente Variable					Segment					Erstpräferenz
1	2	3	4	5	A	B	C	D	E	
H	N	H	H	H	0.36	0.02	0.29	0.11	0.22	
H	N	N	H	H	0.36	0.01	0.32	0.09	0.22	A
H	N	N	H	N	0.29	0.31	0.27	0.00	0.13	A
H	H	H	H	N	0.06	0.42	0.21	0.00	0.31	B
N	H	H	H	N	0.02	0.72	0.16	0.00	0.10	B
H	H	N	H	N	0.07	0.42	0.14	0.00	0.37	B
N	H	N	H	N	0.02	0.61	0.23	0.00	0.14	B
H	N	H	H	N	0.23	0.50	0.16	0.00	0.11	B
N	N	H	H	N	0.06	0.77	0.15	0.00	0.02	B
H	H	H	N	N	0.05	0.58	0.09	0.00	0.28	B
H	N	H	N	N	0.29	0.43	0.11	0.00	0.17	B
H	N	N	N	N	0.26	0.49	0.13	0.00	0.12	B
N	N	N	H	N	0.09	0.64	0.23	0.00	0.04	B
N	H	H	H	H	0.03	0.36	0.26	0.12	0.23	B
N	N	H	N	H	0.07	0.74	0.06	0.08	0.05	B
N	H	N	H	H	0.04	0.10	0.33	0.23	0.30	C

H, N Maximal- bzw. Minimalwert der latenten Variablen

Table 9: Beziehungen zwischen latenten Variablen und Segmentzugehörigkeiten (Fortsetzung)

Latente Variable					Segment					Erstpräferenz
A	B	C	D	E	1	2	3	4	5	
0.27	0.07	0.09	0.39	0.18	H	N	H	N	H	D
0.28	0.04	0.11	0.36	0.21	H	N	N	N	H	D
0.02	0.19	0.12	0.47	0.20	N	H	N	N	H	D
0.08	0.08	0.19	0.60	0.05	N	N	H	H	H	D
0.11	0.04	0.27	0.50	0.08	N	N	N	H	H	D
0.03	0.18	0.12	0.47	0.20	N	N	N	N	H	D
0.10	0.06	0.25	0.03	0.56	H	H	H	H	H	E
0.09	0.12	0.13	0.12	0.54	H	H	H	N	H	E
0.11	0.02	0.33	0.02	0.52	H	H	N	H	H	E
0.10	0.08	0.16	0.11	0.55	H	H	N	N	H	E
0.09	0.27	0.17	0.02	0.45	H	H	N	N	N	E

H, N Maximal- bzw. Minimalwert der latenten Variablen

Marke C wird als Erstpräferenz vor allem von Personen genannt, die eine negative Einstellung zur Hausarbeit und zu Reinigern aufweisen, in größeren Gemeinden in einem Haushalt mit zwei bis drei Personen wohnen. Die latente Variable 5 (soziale Schichtung) nimmt meist den Wert Eins an. Dieser Wert ergibt sich bei Personen mit Berufsschulbildung, die ganztätig berufstätig sind und in deren Haushalt mehrere Einkommensbezieher oder ein Einkommensbezieher bei mittlerem Haushaltseinkommen wohnen sowie bei Personen mit Berufsschulbildung, die teilzeitbeschäftigt sind und über ein geringes Einkommen bei mehreren Einkommensbeziehern verfügen. Außerdem trifft dies auf Personen mit Hauptschulabschluß und mehreren Einkommensbeziehern im Haushalt zu, die entweder nicht berufstätig sind und ein niedriges Einkommen aufweisen oder ganztätig berufstätig sind und über ein mittleres oder hohes Einkommen verfügen.

Für Personen mit einer Erstpräferenz für Marke D gilt mit Hinblick auf die soziale Schichtung dasselbe wie für Personen mit einer Erstpräferenz für Marke C. Daneben verfügen sie über eine negative Einstellung zur Hausarbeit oder eine positive Einstellung zu Reinigern.

Als typisch für Personen mit einer Erstpräferenz für Marke E erweist sich eine positive Einstellung zur Hausarbeit, aber eine negative Einstellung zu Reinigern.

Der unterschiedliche Einfluß der betrachteten Deskriptoren erlaubt die Ableitung von Implikationen für Marketingmaßnahmen, die sich an die fünf a-priori Segmente wenden. Dabei sind für die Mediaselektion v.a. soziale Schichtung und Ortsgrößen, für die Gestaltung von Werbebotschaften soziale Schichtung und Einstellungen, für die Produktpolitik Einstellungen zu Reinigern relevant. Generell dürfte im vorliegenden Fall die Anpassung des Angebots an die Einstellungen als absatzpolitische Strategie überlegen sein (24).

Eine Änderung der hier betrachteten, sich nicht auf einzelne Marken beziehenden Einstellungen dürfte demgegenüber weniger aussichtsreich sein.

7 Abschluß

Der Erfolg der Marktsegmentierung hängt u.a. vom Einsatz geeigneter Datenanalysetechniken ab. Konnexionistische Modelle weisen im Vergleich zu üblichen Techniken wie Diskriminanzanalyse oder logistischer Regression einige Vorteile auf. Vor allem erlauben sie die Aufdeckung nichtlinearer Beziehungen zwischen Deskriptoren und Segmentzugehörigkeiten und von Interaktionseffekten zwischen Deskriptoren, wobei bei einem geeigneten Aufbau des konnexionistischen Modells die Anzahl der Parameter ähnlich groß ist wie bei den verbreiteten Verfahren. Berücksichtigt man Interaktionseffekte, kann die Parameteranzahl sogar geringer ausfallen. Durch Gruppierung von Deskriptoren zu latenten Variablen, die hypothetischen Konstrukten entsprechen, ergeben sich konnexionistische Modelle, die sich durch geringe Anzahl an Parametern und zufriedenstellende Interpretierbarkeit der Modellergebnisse auszeichnen.

Die lineare Diskriminanzanalyse stellt einen Sonderfall eines konnexionistischen Modells dar. Daher sollte die Klassifikationsgüte bei Segmentierungsproblemen zumindest gleich, gegebenenfalls besser als jene der bekannteren Methoden sein. Im vorliegenden Anwendungsfall erlaubt erst die ausreichend höhere Klassifikationsgüte eines konnexionistischen Modells, die Zugehörigkeit von Abnehmern zu a-priori Segmenten zu erklären und Implikationen für Marketing-Maßnahmen abzuleiten. Die logistische Regression dagegen führt zu keinen verwertbaren Ergebnissen. Insgesamt gesehen könnten konnexionistische Modelle zur Lösung von Klassifikationsproblemen eine sinnvolle Er-

weiterung des Methodenvorrats der Marketing-Forschung darstellen.

References

- [1] Anderson, J.A. (1982): Logistic Discrimination, in: Krishnaiah, P.R./Kanal, L.N. (eds.): Handbook of Statistics, Vol. 2, 169–191.
- [2] Bagozzi, R.P. (1980): Causal Models in Marketing.
- [3] Bagozzi, R.P. (1986): Principles of Marketing Management.
- [4] Bentler, P. (1980): Multivariate Analysis with Latent Variables: Causal Modeling, in: Annual Review of Psychology, 419-456.
- [5] Bischoff, R./Bleile, C./ Graalfs, J. (1991): Der Einsatz Neuronaler Netze zur betriebswirtschaftlichen Kennzahlenanalyse, in: Wirtschaftsinformatik, 375–385.
- [6] Cybenko, G. (1989): Continuous Value Neural Networks with Two Hidden Layers are Sufficient, in: Mathematics of Control, Signal and Systems, 303–314.
- [7] Duda, R.O./Hart, P.E. (1973): Pattern Classification and Scene Analysis.
- [8] Erxleben, K./Koch, H. (1991): Früherkennung von Unternehmenskrisen. Ein Vergleich von Neuronalen Netzen und Diskriminanzanalyse. Arbeitspapier 4, Abteilung Wirtschaftsinformatik, Universität Erlangen-Nürnberg.
- [9] Frank, R.E./Massy, W.F./Wind, Y. (1972): Market Segmentation.
- [10] Freter, H. (1983): Marktsegmentierung.

- [11] Funahashi, K. (1989): On the Approximate Realization of Continuous Mappings by Neural Networks, in: *Neural Networks*, 183–192.
- [12] Gallant, S.I. (1988): Connectionist Expert Systems, in: *Communications of the ACM*, 152–169.
- [13] Goldstein, M./Dillon, W.R. (1978): *Discrete Discriminant Analysis*.
- [14] Hand, D.J. (1981): *Discrimination and Classification*.
- [15] Hansen, F. (1972): *Consumer Choice Behavior. A Cognitive Theory*.
- [16] Hertz, J./Krogh, A./Palmer, R.G. (1991): *Introduction to the Theory of Neural Computation*.
- [17] Hinton, G.E. (1990): Connectionist Learning Procedures, in: Kodratoff, Y./Michalski, R. (eds.), *Machine Learning. An Artificial Intelligence Approach. Volume III*, 555–610.
- [18] Hruschka, H. (1991): Einsatz künstlicher neuronaler Netzwerke zur Datenanalyse im Marketing, in: *Marketing-ZFP*, 217–225.
- [19] Hruschka, H. (1992): Determining Market Response Functions by Neural Network Modeling. A Comparison to Econometric Techniques, erscheint in: *European Journal of Operational Research*.
- [20] Jain, A.K./ Chandrasekaran, B. (1982): Dimensionality and Sample Size Considerations in Pattern Recognition Practice, in: Krishnaiah, P.R./Kanal, L.N. (eds.): *Handbook of Statistics, Vol. 2*, 835–855.
- [21] James, L.R./Mulaik, S.A./Brett, J.M.(1987): *Causal Analysis. Assumptions, Models and Data*. 4th Printing.
- [22] Jöreskog, K.G./Wold, H. (1982; eds.): *Systems under Indirect Observation. Part I*.

- [23] Judge, G.E./ Griffiths, W.E./Hill, R.C./Lee, T.C. (1980): The Theory and Practice of Econometrics.
- [24] Kroeber-Riel, W. (1984): Konsumentenverhalten. 3. Auflage.
- [25] Kurbel, K./Pietsch, W. (1991): Eine Beurteilung konnektionistischer Modelle auf der Grundlage ausgewählter Anwendungsprobleme und Vorschläge zur Erweiterung, in: Wirtschaftsinformatik, 355–364.
- [26] Maddala, G.S. (1983): Limited-Dependent and Qualitative Variables in Econometrics.
- [27] Másson, E./Wang, Y.-J. (1990), Introduction to Computation and Learning in Artificial Neural Networks, in: European Journal of Operational Research 47, 1–28.
- [28] Mazanec, J.A. (1992): Market Segmentation with a Neural Network Model. Preliminary Findings, erscheint in: Journal of Travel and Tourism Marketing.
- [29] Novak, Th.P./MacEvoy, B. (1990): On Comparing Alternative Segmentation Schemes. The List of Values (LOV) and Values and Life Styles (VALS), in: Journal of Consumer Research, 105–109.
- [30] Rehkugler, H./Poddig, Th. (1991): Künstliche Neuronale Netze in der Finanzanalyse. Eine neue Ära der Kursprognosen, in: Wirtschaftsinformatik, 365–374.
- [31] Ruck, D./Rogers, S.K./Kabrisky, M./Oxley, M.E./Suter, B.W.(1990): The Multilayer Perceptron as an Approximation to a Bayes Optimal Discriminant Function, in: IEEE Transactions on Neural Networks, 296–298.

- [32] Rumelhart, D.E./Hinton, G.E./McClelland, J.L. (1986): A General Framework for Parallel Distributed Processing, in: Rumelhart, D.E./McClelland, J.L. (eds.), *Parallel Distributed Processing. Explorations in the Microstructure of Cognition*, Volume 1, 45–76.
- [33] Rumelhart, D.E./Hinton, G.E./Williams, R.J. (1986): Learning Internal Representations by Error Propagation, in: Rumelhart, D.E./McClelland, J.L. (eds.), *Parallel Distributed Processing. Explorations in the Microstructure of Cognition*, Volume 1, 318–362.
- [34] Schaninger, Ch.M. (1981): Social Class versus Income Revisited. An Empirical Investigation, in: *Journal of Marketing Research*, 192–208.
- [35] Sheth, J.N (1977): Demographics in Consumer Behavior. In: *Journal of Business Research*, 129–138.
- [36] Shoemaker, P.A. (1991): A Note on Least-Squares Learning Procedures and Classification by Neural Network Models, in: *IEEE Transactions on Neural Networks*, 158– 160.
- [37] Stone, M. (1974): Cross-Validatory Choice and Assessment of Statistical Predictions. In: *Journal of the Royal Statistical Society, Series B*, 111–133.
- [38] Wan, E.A. (1990): Neural Network Classification. A Bayesian Interpretation, in: *IEEE Transactions on Neural Networks*, 303–305.
- [39] Wasserman, P.D. (1988): Neural Networks, Part 2, in: *IEEE Expert*, 10–15.
- [40] Wasserman, P.D. (1989): *Neural Computing. Theory and Practice*.

- [41] White, H. (1989): Connectionist Nonparametric Regression. Multilayer Feedforward Networks Can Learn Arbitrary Mappings. Working Paper, Department of Economics, University of California San Diego.
- [42] Widrow, B./Winter, R. (1988): Neural Nets for Adaptive Filtering and Adaptive Pattern Recognition, in: IEEE Computer, 25–39.
- [43] Wilbert, R. (1991): Kreditwürdigkeitsanalyse im Konsumentenkreditgeschäft auf der Basis Neuronaler Netze, in: Zeitschrift für Betriebswirtschaft, 1377–1393.
- [44] Wind, Y. (1978): Issues and Advances in Segmentation Research. In: Journal of Marketing Research, 317–337.
- [45] Wiswede, G. (1983): Marktsoziologie. In: Irle, M. (Hrsg.): Marktpsychologie. 1. Halbband, 151–224.