

planung & analyse

Zeitschrift für Marktforschung und Marketing
Eine Marke der dfv Mediengruppe

www.planung-analyse.de

1/2015 D11700F

Interview

mit GfK-CEO
Matthias Hartmann

Special

Information und
Telekommunikation

Kongress

rheingold spricht
Kunden Mut zu

Schwerpunkt

Von der Idee bis ins Regal

Statistik KOMPAKT

Clusteranalyse

Die Segmentierung von Nachfragern anhand ihrer Präferenzen bzw. von Produkten anhand ihrer Eigenschaften oder die Bestimmung von Konsumententypologien mittels psychografischer Merkmale zählen zu den klassischen Anwendungsgebieten der Clusteranalyse. Daneben eröffnen die verschiedenen Varianten viele weitere Anwendungsmöglichkeiten.

Homogenität und Heterogenität von Clustern

Ziel einer Clusteranalyse ist es zumeist, eine Menge von Objekten wie Personen oder Produkten so in Gruppen, Klassen bzw. Clustern zusammenzufassen, dass

- ▶ diese in sich homogen sind, das heißt die Objekte eines Clusters bezüglich der betrachteten Variablen zueinander ähnlich sind, und
- ▶ diese zueinander heterogen sind, das heißt die Objekte unterschiedlicher Cluster hinsichtlich der herangezogenen Variablen möglichst verschieden sind.

Alle Cluster bilden zusammen eine Klassifikation der Menge von Objekten. Diese heißt exhaustiv, wenn alle Objekte mindestens einem Cluster zugeordnet werden.

Kriterien zur Einteilung der Varianten der Clusteranalyse

Die verschiedenen Varianten zum Auffinden von Clustern lassen sich anhand folgender Kriterien charakterisieren:

▶ **Definition der Ähnlichkeit: Ausprägungen oder Zusammenhänge**
Objekte sind entweder zueinander ähnlich, wenn sie hinsichtlich der betrachteten Variablen ähnliche Ausprägungen aufweisen, oder wenn für sie ähnliche Zusammenhänge zwischen den Variablen gelten. Produkte können beispielsweise ähnlich sein, wenn ihre Preise gleich hoch sind, oder wenn die Entwicklung der Preise im Zeitablauf gleich war.

▶ **Deterministische oder probabilistische Zuordnung**
Bei einer deterministischen Zuordnung ist zu unterscheiden, ob ein Objekt eindeutig jeweils nur einem Cluster zugewiesen ist (nicht-überlappende Zuordnung bzw. Partition) oder ob es auch zu mehreren Clustern zählen kann (überlappende Zuordnung). Dagegen weist eine probabilistische Zuordnung jedem Objekt Wahrscheinlichkeiten zu, mit denen es zu den einzelnen Clustern gehört. Innerhalb einer Warenkorbanalyse ist es beispielsweise ein Ziel, Cluster aus Produkten oder Warengruppen zu identi-

fizieren, die häufig zusammen gekauft werden. Bei einer nicht-überlappenden Zuordnung besteht somit im Gegensatz zur überlappenden oder probabilistischen Zuordnung die Gefahr, dass der Kaufverbund eines Produkts zu den Produkten eines anderen Clusters übersehen wird, da es diesem nicht ebenfalls angehören kann.

▶ Einmodale oder zweimodale Verfahren

Einmodale Verfahren fassen entweder – wie zuvor beschrieben – Objekte oder – wie in einer Faktorenanalyse – Variablen zu Gruppen zusammen. Mit zweimodalen Verfahren werden Objekte und Variablen gleichzeitig klassifiziert. Ziel ist es, Cluster zu finden, in denen neben Objekten, die untereinander ähnlich sind, auch Variablen enthalten sind, die zu den Objekten in enger Beziehung stehen. Zweimodale Clusteranalysen eignen sich beispielsweise zur Werbewirkungskontrolle: In einem Test ist eine Reihe von Markennamen den entsprechenden Anzeigen zuzuweisen, ohne dass die Marke dort zu sehen ist. Die richtigen und falschen „Zuordnungen“ sind dann Basis einer Analyse, in der die Marken die Objekte und die Anzeigen die Variablen darstellen. Befinden sich Marke und die korrekte Anzeige in demselben Cluster, ist dies ein Indikator für die Eigenständigkeit der Anzeige.

Beispiel zur Illustration der Kriterien

Die Abbildung zeigt ein kleines Datenbeispiel mit fünf Objekten, die durch zwei Merkmale M1 und M2 charakterisiert sind. Für dieses Beispiel werden nachfolgend mögliche Ergebnisse von Varianten der Clusteranalyse vorgestellt, die in kommenden Beiträgen dieser Reihe erläutert werden.

Eine nicht exhaustive Klassifikation nur der Objekte O1 bis O4 anhand der Ausprägungen der Variablen M1 und M2, die aus zwei Clustern besteht und dem Ziel der Clusteranalyse gerecht wird, ist {O1,O2} und {O3,O4}. Beide Cluster sind in sich homogen, aber zueinander heterogen. Mithilfe der Mittelwerte der beiden Merkmale ließen sich die Cluster genauer beschreiben. Werden statt der Ausprägungen die Zusammenhänge der Variablen zugrunde gelegt, ergeben sich die Cluster {O1,O3} und {O2,O4}. Der rechte Teil der Abbildung verdeutlicht die gleich verlaufenden Profile der jeweils zu einem Cluster gehörenden Objekte.

Eine zweimodale Clusteranalyse anhand der Ausprägungen der Merkmale führt zu einer Zuordnung beider Variablen M1 und M2 zu dem Cluster {O3,O4}. Damit wird deutlich, dass (die Objekte) dieses Cluster(s) durch hohe Ausprägungen beider Variablen gekennzeichnet ist (sind). Ein Rückgriff zum Beispiel auf die Mittelwerte zur Charakterisierung der Cluster erübrigt sich damit.

Wird eine exhaustive Klassifikation mit zwei Clustern anhand der Merkmalsausprägungen gefordert, die auch Objekt O5 mit einschließt, wird dieses von einmodalen Verfahren bei einer deterministischen, nicht-überlappenden Zuordnung letztlich zufällig, aber eindeutig entweder dem Cluster {O1,O2} oder dem Cluster {O3,O4} zugeordnet. Ein probabilistisches Verfahren vermeidet eine „harte“ Zuordnung. Da O5 genau zwischen den Clustern liegt, weist dieses Verfahren O5 den beiden Klassen jeweils mit einer Wahrscheinlichkeit von 50 Prozent zu.

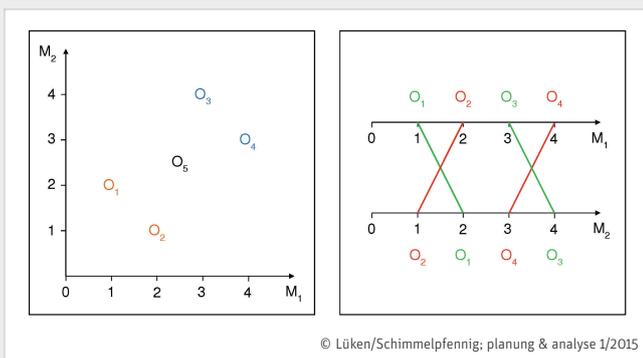
Johannes Lüken und **Prof. Dr. Heiko Schimmelpfennig**, Experten für Multivariate Analysen bei IfaD, Institut für angewandte Datenanalyse.

In Ausgabe 2/2015: *Hierarchische Clusteranalyse*

▶ Literatur

Bacher, J.; Pöge, A.; Wenzig, K.: Clusteranalyse. 3. Auflage, 2010, München, S. 15-33.

Schwaiger, M.: Wirkungskontrolle kommunikationspolitischer Maßnahmen, in: Reinecke, S.; Tomczak, T. (Hrsg.): Handbuch Marketingcontrolling, 2. Auflage, Wiesbaden, 2006, S. 521-548.



© Lüken/Schimmelpfennig; planung & analyse 1/2015

Abbildung: Beispiel mit zwei Merkmalen.