

**Vergleich von drei Verfahren der Clusteranalyse:
der HMEANS/KMEANS-Algorithmus
das Verfahren mittels unscharfer Partitionen
sowie die Klassifikation nach WARD**

J. FUNKE¹ und B. KOPP²

Zusammenfassung, Summary, Résumé

Drei Verfahren zur Aufdeckung von Subgruppenstrukturen in einem Datensatz unbekannter Struktur werden miteinander verglichen: ein modifizierter HMEANS/KMEANS-Algorithmus, die Methode unscharfer Partitionen von BOCK sowie das Verfahren von WARD. An einem Datensatz mit neun Persönlichkeitsvariablen und 135 Alkoholiker-Vpn werden die unterschiedlichen Analyseergebnisse demonstriert sowie die Zusammenhangsstrukturen der verschiedenen Cluster aufgezeigt. Die Beurteilung der Resultate führt zu keiner eindeutigen Präferenz eines der vorgestellten Verfahren.

A comparison of three methods of cluster analysis = the Hmeans/Kmeans-algorithm, the fuzzy set partition and the procedure of WARD

Three methods for the identification of sub-group-structures in a set of data with unknown structure are compared with: a modified Hmeans/Kmeans-algorithm, the fuzzy set partition developed by Bock and the procedure of Ward. Using the personality data of 135 alcoholics the results of the above mentioned methods are demonstrated. Thereby the relationships and structures of cluster centroids are shown. Evaluating the results does not lead to a clear preference for one or another of the three procedures of cluster analysis.

Comparison de trois méthodes d'analyse des structures de sous-groupes: l'algorithme HMEANS/ KMEANS, la méthode de «partition» imprecise et la classification d'après WARD

On compare trois méthodes de repérage des structures de sous-groupes pour un éventail de données portant sur une structure non connue: un algorithme HMEANS/KMEANS, la méthode de «partition» non précise de BOCK ainsi que la méthode de WARD. A l'aide d'un choix de données à 9 variables de personnalité et portant sur 135

- 1 Dipl.-Psych. Joachim Funke, Fachbereich I-Psychologie, Universität Trier, Schneidershof, 5500 Trier.
- 2 Dipl.-Math. Bruno Kopp, Gruner und Jahr AG und Co. Zentrale Datenverarbeitung, Alsterufer 1, 2000 Hamburg 36.

sujets alcooliques, on démontrera les différences entre les résultats des analyses, de même que les relations structurales entre les divers «clusters». Les résultats une fois analysés, on ne peut opter de façon décisive pour l'une ou l'autre des méthodes proposées.
(S. Schnettler)

1. Problemstellung

Die in den letzten Jahren in das Interessensfeld einer breiteren Öffentlichkeit gerückte Problematik differenzierter Indikationsstellung im klinischen Bereich hat dazu geführt, daß mathematische Algorithmen zur Aufteilung von Personengruppen in Substichproben mit spezifischen Profilen verstärkt Verwendung finden. Dieser Sachverhalt macht es notwendig, die Brauchbarkeit der verschiedenen Analyseverfahren näher zu bestimmen. Die vorliegende Arbeit greift drei Modelle der Clusteranalyse auf und wendet sie an einem Datensatz exemplarisch an.

2. Datensatz

Der zugrundegelegte Datensatz stammt aus einer Untersuchung von FUNKE, KLEIN und SCHELLER (1981). Insgesamt 135 unausgelesene Patienten einer Fachklinik für Suchtkranke (102 Männer und 33 Frauen) wurden dort mit einer Kurzform des „Freiburger Persönlichkeitsinventars FPI“ (FPI-K; FAHRENBERG, SELG und HAMPEL 1973²) untersucht. Die Rohwerte der Personen auf den 12 Skalen des FPI-K bilden das Ausgangsmaterial für eine hierarchische Clusteranalyse von WARD (1963), deren Ergebnisse zur Auswahl einer gut interpretierbaren Vier-Cluster-Lösung führte. Die praktische Bedeutsamkeit dieser Clusterbildung wird von SCHELLER und KLEIN (1981) bestätigt: die Autoren finden für die genannte Vier-Cluster-Lösung stark schwankende Erfolgsquoten in einer Ein-Jahres-Katamnese; als Konsequenz werden differentielle Therapiestrategien für die Clustermitglieder vorgeschlagen.

Für den vorliegenden Vergleich beschränken wir uns auf die ersten neun Skalen des FPI-K, da die Zusatzskalen 10–12 zum Teil durch Doppelauswertung der Items gebildet werden und somit rechnerische Abhängigkeiten enthalten. Die Rohdaten wurden vor der weiteren Verarbeitung standardisiert.

3. Beschreibung der Algorithmen

Bei den in dieser Studie benutzten Programmen handelt es sich um die PL/I-Typologieprogramme AUTOKLAS bzw. FUZCLU der Zentralen Datenverarbeitung von Gruner und Jahr AG & Co., Hamburg. AUTOKLAS ist eine Kombination der Algorithmen HMEANS und KMEANS (vgl. BOCK 1974, SPÄTH 1977, STEINHAUSEN und LANGER 1977, VOGEL 1975). Das Programm FUZCLU arbeitet auf der Grundlage unscharfer Partitionen (BOCK 1979). Da wir nicht unterstellen können, daß diese Algorithmen allgemein bekannt sind, wollen wir diese Methoden zuvor kurz erläutern.

Das Verfahren von WARD (1963) ist im Unterschied zu den beiden bisher genannten weiter verbreitet. In VELDMAN (1967), SPÄTH (1975), ROLLET und BARTRAM (1976) sowie in STEINHAUSEN und LANGER (1977) findet der Leser neben der Beschreibung des Verfahrens auch zugehörige Programme. VON EYE (1977) hat diesen Algorithmus bereits einem ersten Methodenvergleich unterzogen; SPÄTH (1975) sowie STEINHAUSEN und LANGER (1977) vergleichen den WARDschen mit weiteren divisiv-agglomerativen Abstandbestimmungen.

3.1. HMEANS

Dieses Verfahren läßt sich besonders einfach implementieren und ist entsprechend weit verbreitet. Da es besonders schnell ist, läßt es sich auf sehr große Datenbestände anwenden. Die Grundidee ist denkbar einfach: Man beginnt mit einer mehr oder weniger zufälligen Anfangspartition in, sagen wir m Klassen, berechnet deren m Zentroide und ordnet die Objekte neu zu nach Maßgabe ihrer Entfernung zu den Schwerpunkten (Minimaldistanzpartition). Danach werden die Schwerpunkte neu berechnet und das Verfahren wird iterativ fortgesetzt bis sich keine wesentliche Änderung mehr ergibt. Die Praxis zeigt, daß sich in der Regel nach 5 bis 10 Iterationen eine stabile Partition einstellt. Beim HMEANS-Verfahren werden im wesentlichen nur Randpunkte umgeordnet. Deshalb ist es sinnvoll, die erhaltene Partition durch den besseren Algorithmus KMEANS zu verfeinern.

3.2. KMEANS

Während beim HMEANS-Verfahren die Schwerpunkte jeweils nach Zuordnung *aller* Objekte neu berechnet werden, werden hier die Zentroide u. U. bereits nach der Zuordnung eines einzigen Objekts angepaßt. Der Algorithmus ist dementsprechend rechenintensiv; wir empfehlen eine Kombination der beiden Methoden: erst HMEANS, danach KMEANS.

Dies ist im Programm AUTOKLAS verwirklicht worden. Ein relativ neues Typologie-Verfahren liefert die Clusteranalyse mittels unscharfer Partitionen (Fuzzy Set Partitions).

3.3. Unscharfe Partitionen

Bei den beiden oben beschriebenen Algorithmen wurde jedes Objekt *genau* einer der m Klassen zugeordnet. Dabei ist es gleichgültig, ob das Objekt nahe beim Zentroid oder weiter entfernt davon liegt. Wie alle Objekte mit dem gleichen Gewicht in die Schwerpunktsberechnung eingehen, kann dies als eine Schwachstelle dieser Methoden angesehen werden. Die Clusteranalyse über unscharfe Partitionen vermeidet diesen Effekt dadurch, daß für jedes Objekt i Gewichte g_{ik} der Zugehörigkeit zu den m Klassen $k = 1, \dots, m$ interiert werden:

$$\sum_{i=1}^m g_{ik} = 1, \quad g_{ik} \geq 0.$$

Man könnte g_{ik} als eine Art „Klassenzugehörigkeitswahrscheinlichkeit“ ansehen. Bei der Ermittlung der Zentroide z_k wird eine Potenz r (z. B. $r = 2$) der g_{ik} als Gewicht verwendet:

$$z_k = \frac{\sum_{i=1}^N g_{ik}^r x_i}{\sum_{i=1}^N g_{ik}^r};$$

dabei ist N die Gesamtanzahl der Objekte und x_i der Vektor der Ausprägungen des i -ten Objekts.

Das Verfahren beginnt wieder mit einer mehr oder weniger zufälligen Anfangspartition der g_{ik} und verbessert diese nach der folgenden Vorschrift:

$$g_{ik} = \frac{d_i}{x_i - z_k}^{r-1}$$

Dabei gilt:

$$d_i = \sum_{j=1}^m |x_i - z_j|^{r-1}.$$

Aus den \bar{g}_{ik} ergeben sich neue Schwerpunkte, und das Verfahren wird iteriert.

Die Praxis zeigt, daß im allgemeinen Konvergenz erst nach einer großen Anzahl Iterationen (ca. 50) eintritt. Entsprechend hoch ist auch der Rechenaufwand.

3.4. Das Verfahren von WARD

WARDs (1963) Verfahren gehört zur Gruppe der hierarchischen, agglomerativen Clusteranalysen, bei denen ausgehend von der feinsten Partition (alle n Elemente bilden je ein eigenes Cluster) schrittweise jeweils die Elemente bzw. Cluster fusioniert werden, die den geringsten Zuwachs zu folgendem Varianzkriterium liefern:

$$\frac{n_p n_q}{N} |x_{g_p} - x_{g_q}|^2;$$

dabei gehen n_p und n_q die Elemente der zu fusionierenden Cluster p und q an, die x_g verweisen auf den jeweiligen Gruppenschwerpunkt. Fusioniert werden also die beiden Cluster, bei denen dieses „Heterogenitätsmaß“ am geringsten ansteigt. Zur weiteren Einordnung dieses Verfahrens in den Kontext hierarchischer Clusteranalysen siehe KOPP (1978).

4. Ergebnisse

Ergänzend zu den Ergebnissen der früheren Arbeit von FUNKE, KLEIN und SCHELLER (1981) wurden mit den Programmen AUTOKLAS und FUZCLU Partitionen in vier Gruppen gesucht. Tabelle 1 zeigt die Ergebnisse der drei Analyseverfahren im Überblick.

Neben den Zentroiden der jeweils vier Cluster wird hier die Clustergröße sowie die Anzahl abstinenter Clustermitglieder aufgeführt. Diese Information dient dazu, die praktische Bedeutsamkeit der Gruppierungen einschätzen zu helfen; gefragt wird, ob eines der Verfahren zu einer höheren Trennschärfe in bezug auf das – für die Klassifizierung irrelevante – Kriterium Abstinenz führt. Bei fehlenden statistischen Argumenten für die Bevorzugung oder Zurückweisung einzelner Lösungen kann dadurch zusätzliche Entscheidungshilfe geboten werden.

Tabelle 1

Zentroide der Vier-Cluster-Lösungen auf den neun FPI-K-Skalen, getrennt für drei Klassifikationsverfahren, sowie clusterspezifische Abstinenzquoten nach Ein-Jahres-Katamnese

FPI-K-Skala	AUTOKLAS-Cluster				FUZCLU-Cluster				WARD-Cluster				
	1	2	3	4	1	2	3	4	1	2	3	4	
1 Nervosität	3.44	3.84	1.83	5.31 ⁺	2.83	4.83	2.29	3.67	2.86	4.97 ⁺	1.69	4.12	
2 Aggressivität	0.65	4.29 ⁺	1.98	2.22	1.92	2.77	1.85	1.94	3.33 ⁺	2.97	1.64	1.06	
3 Depressivität	4.00	5.25 ⁺	2.77	6.55 ⁺	3.90	5.79 ⁺	3.19	4.72	4.57	6.16 ⁺	2.88	4.26	
4 Erregbarkeit	3.20	6.38 ⁺	2.40	5.00	3.54	5.26	2.87	4.12	5.67 ⁺	5.73 ⁺	2.00	3.44	
5 Geselligkeit	1.92 ⁻	2.57	4.81	2.79	3.53	2.31	4.05	2.68	4.24	2.05	4.57	1.88 ⁻	
6 Gelassenheit	2.77	4.41	5.09	1.46 ⁻	3.89	2.70	4.48	3.15	4.76	2.40	4.79	2.59	
7 Dominanzstreben	0.94	4.44 ⁺	2.01	4.23 ⁺	2.28	4.00	2.03	2.51	4.43 ⁺	4.40 ⁺	1.59	1.03	
8 Gehemtheit	5.39 ⁺	4.23	1.50	5.07	3.31	5.07	2.52	4.47	2.09	5.59 ⁺	1.88	5.53 ⁺	
9 Offenheit	3.21	5.26	3.79	5.44	3.99	5.15	3.90	4.25	4.38	5.62 ⁺	4.05	3.09	
Clustergröße	34	27	44	30	(15)	(42)	(46)	(32)	21	37	43	34	
dauernd abstinent Cluster mitglieder	N	21	11	23	13	4	17	28	19	11	14	20	23
	in %	62.8	40.8	52.3	43.3	26.7	40.5	60.9	59.4	52.4	37.8	46.5	67.6

Anmerkung: Die über dem dritten oder unter dem ersten Quartil der Eichstichprobe (N = 630) liegenden Meßwerte sind mit „+“ bzw. „-“ markiert.

Eine erste Inspektion der Daten aus Tabelle 1 läßt folgende Schlüsse zu: sowohl bei AUTOKLAS als auch bei WARD finden sich mehr Zentroidwerte, die unter dem ersten bzw. über dem dritten Quartil der Eichstichprobe liegen, als dies bei FUZCLU der Fall ist (10 und 11 Fälle respektive 2. Diese überrascht insofern, als FUZCLU hinsichtlich der Objektzahl pro Cluster einen größeren Range aufweist als die beiden anderen Verfahren und daher extremere Gruppeneinteilungen erwarten läßt; allerdings muß hinzugefügt werden, daß bei diesem Algorithmus keine eindeutige Clusterzugehörigkeit angegeben werden kann, sondern nur entsprechende Zugehörigkeitswahrscheinlichkeiten. Aus diesem Grund wurden die Fallzahlen in Tabelle 1 eingeklammert: sie geben die Clustergröße an, wenn man die jeweils höchste Wahrscheinlichkeit verwendet. Die Zentroide werden dadurch nicht beeinträchtigt.

Die aufgeführten Vier-Cluster-Lösungen der drei verschiedenen Verfahren sollen nicht näher interpretiert werden (zur Interpretation der WARD-Typen siehe FUNKE, KLEIN und SCHELLER 1981); stattdessen interessiert die Ähnlichkeit der drei Analyseresultate. Rangkorrelationen r_s zwischen den zwölf verschiedenen Clustern gibt Tabelle 2 an.

Tabelle 2

Rangkorrelationen r_s zwischen den Clusterzentroiden der verschiedenen Verfahren (Koeffizienten zusammengehöriger Cluster sind hervorgehoben)

	AUTOKLAS (K)				FUZCLU (F)				WARD (W)			
	1	2	3	4	1	2	3	4	1	2	3	4
A	1	-										
	2	07 ^a	-									
	3	-27	23	-								
	4	77	32	-25	-							
F	1	47	52	67	43	-						
	2	67	68	-25	83	45	-					
	3	20	08	85	-03	82	-10	-				
	4	93	35	-05	80	68	80	33	-			
W	1	-20	73	65	-12	52	23	43	07	-		
	2	67	68	-25	83	45	99	-10	80	23	-	
	3	18	05	83	-07	80	-12	98	32	38	-12	-
	4	97	05	-32	68	40	67	15	90	-18	67	17

a führende Nullen sowie Dezimalpunkte wurden ausgelassen

Zwei Aspekte des Zusammenhangsgefüges werden erkennbar: zum einen die Interkorrelation der vier Cluster innerhalb eines jeden Verfahrens, zum anderen die Korrelation der Cluster, die aus verschiedenen Algorithmen resultieren. Hinsichtlich des ersten Aspekts erwartet man möglichst geringe Zusammenhänge – wünschenswert sind Clusterlösungen mit möglichst unterschiedlichen Zentroiden –, jedoch zeigen sich in allen drei Verfahren substantielle Clusterähnlichkeiten. Die höchsten Werte (.77 bei AUTOKLAS, .82 bei FUZCLU und .67 bei WARD) lassen erkennen, daß die jeweiligen Verfahren keine orthogonalen Cluster definieren; FUZCLU liefert von allen drei hier vorgestellten Verfahren Cluster mit den größten „Intra“-Interkorrelation, während AUTOKLAS und WARD deutlich besser differenzieren.

Hinsichtlich des zweiten Aspekts – Clusterähnlichkeiten über die verschiedenen Verfahren hinweg – lassen sich in einem Fall, nämlich beim Vergleich von AUTOKLAS mit WARD, klare Zusammenhangsstrukturen aufweisen (d. h. es ist eine eindeutige Zuordnung der acht Cluster möglich), während dies beim Vergleich AUTOKLAS-FUZCLU sowie FUZCLU-WARD nicht so gut gelingt (erkennbar an der Strichlierung der

vorgeschlagenen Zuordnungen). Damit offenbart sich die Sonderstellung der „unscharfen Partitionen“: die Clusterstruktur, die dieses Verfahren ermittelt, entspricht nicht so deutlich der Struktur, die die beiden anderen Algorithmen aufdecken.

5. Diskussion

Die Ergebnisse des hier vorgestellten Methodenvergleichs mögen auf den ersten Blick enttäuschen: keines der drei Verfahren zur Gruppierung eines Datensatzes mit unbekannter Substruktur fällt als besonders geeignet oder nicht geeignet auf. Im Unterschied zu Plasmodienstudien (vgl. BAUMANN 1971), in denen die Datenstruktur von vornherein bekannt ist, läßt unsere Untersuchung keine klare Aussage über die Güte eines der Algorithmen zu.

Tendenziell scheinen die HMEANS/KMEANS-Algorithmus sowie das Verfahren von WARD eher einander ähnliche Befunde zu liefern bei gleichzeitig besserer Unterscheidbarkeit der Cluster innerhalb der Verfahren. Zieht man allerdings das Kriterium der praktischen Brauchbarkeit heran (in unserem Beispiel die differentiellen Rückfallquoten innerhalb der einzelnen Cluster), weist FUZCLU mit einem Range von 34.2 Prozentpunkten (Abstand des erfolgreichsten Clusters vom schlechtesten), gefolgt von WARD mit 29.8 sowie AUTOKLAS mit 18.5 das beste Ergebnis auf.

Abschließend bleibt festzustellen: die Analyseresultate lassen keine klare Präferenz zugunsten eines der vorgeführten Algorithmen erkennen. Es zeigt sich, daß der modifizierte HMEANS/KMEANS-Algorithmus sowie das Verfahren nach WARD (1963) sehr ähnliche Cluster liefern. Die Methode unscharfer Partitionen von BOCK (1979) liegt an der Spitze, wenn es um die Bildung möglichst trennscharfer Extremgruppen hinsichtlich eines nicht klassifikationsrelevanten Außenkriteriums geht.

Literatur

- Baumann, U.: Psychologische Taxometrie. Bern: Huber, 1971.
 Bock, H. H.: Automatische Klassifikation. Göttingen: Vandenhoeck & Ruprecht, 1974.
 Bock, H. H.: Clusteranalyse mit unscharfen Partitionen. In: Bock, H. H. (Ed.). Klassifikation und Erkenntnis III. Frankfurt: INDEKS-Verlag, 1979.
 Fahrenberg, J., Selg, H. und Hampel, R.: Das Freiburger Persönlichkeitsinventar FPI. Handanweisung. Göttingen: 2. Auflage, Hogrefe, 1973.
 Funke, J., Klein, M. und Scheller, R.: Zur Klassifikation von Alkoholikern durch Persönlichkeitsmerkmale. Psychologische Beiträge 23, 1981, 146–158.

- Kopp, B.: Hierarchical classification III: Averagelinkage, median, WARD, flexible strategy. Biometrical Journal 20, 1978, 703–711.
 Rollet, B. und Bartram, M. (Ed.): Einführung in die hierarchische Clusteranalyse für Psychologen, Pädagogen und Soziologen. Stuttgart: Klett, 1976.
 Scheller, R. und Klein, M.: Persönlichkeitspsychologische Determinanten des Therapieerfolgs bei Alkoholabhängigen. Zeitschrift für Differentielle und Diagnostische Psychologie (im Druck), 1981.
 Späth, H.: Cluster-Analyse-Algorithmen zur Objektklassifizierung und Datenreduktion. München: Oldenbourg, 1975.
 Steinhausen, D. und Langer, K.: Clusteranalyse. Einführung in Methoden und Verfahren der automatischen Klassifikation. Berlin: de Gruyter, 1977.
 Veldman, D. J.: FORTRAN programming for the behavioral sciences, New York: Rinehart & Winston, 1967.
 Vogel, F.: Probleme und Verfahren der numerischen Klassifikation. Göttingen: Vandenhoeck & Ruprecht, 1975.
 Von Eye, A.: Zum Vergleich zwischen der hierarchischen Clusteranalyse nach WARD und MACS, einer mehrdimensionalen, automatischen Clustersuchstrategie. Psychologische Beiträge, 19, 1977, 201–217.
 Ward, J. H.: Hierarchical grouping to optimize an objective function. Journal of the American Statistical Association 58, 1963, 236–244.