

Geleitwort

Zweimodale Clusteranalysen erfreuen sich in verschiedenen Anwendungsbereichen - darunter ist auch die Betriebswirtschaftslehre zu nennen - wachsender Beliebtheit. Dadurch, dass die zweimodalen Algorithmen z.B. Objekte mit den ihnen in besonderem Maße zugesprochenen Merkmalen in einem Cluster simultan abbilden, eliminieren sie die klassische Schwäche einmodaler Klassifikationsverfahren: Da Objekte und stark assoziierte Merkmale im selben Cluster vereinigt sind, erspart sich der Anwender den zur Interpretation eines Clusters in der einmodalen Variante erforderlichen mühsamen Rückgriff auf die Ausgangsdaten. Doch trotz beachtlicher relativer Aufmerksamkeitsgewinne muss leider festgehalten werden, dass die absolute Verbreitung zweimodaler Klassifikationsverfahren sehr schwach ist. Die Gründe hierfür sind zu hinterfragen.

Wird schon im einmodalen Bereich bemängelt, dass die Eingriffsmöglichkeiten des Datenanalytikers - z.B. bei der Festlegung der Distanzmaße und der Wahl des Fusionsalgorithmus - so umfangreich sind, dass sich nahezu beliebige Ergebnisse erzeugen lassen, so galt dieser Einwand bis dato in noch stärkerem Maße für die zweimodalen Varianten, denn im Gegensatz zur einmodalen Klassifikation existierten hier nur wenige Gütemaße, und selbst diese waren bisher nicht in nennenswertem Umfang untersucht worden.

Hier setzt die vorliegende Arbeit an: Rix stellt die Algorithmen zweimodaler hierarchischer Clusteranalyse und die verfügbaren Gütemaße vor. Dabei werden Fehler in der Originalliteratur aufgedeckt, Parallelen zwischen bestehenden Algorithmen herausgearbeitet und eigenständige Ideen zu Gütemaßen und Algorithmen ausgearbeitet. Kernstück der wissenschaftlichen Arbeit ist eine umfangreiche Simulationsstudie zur Evaluierung der Verfahren und Gütemaße. Ausgehend von der Beschreibung idealer Klassifikationen entwickelt der Autor zunächst sechs zu variierende Kriterien: den Wertebereich und die Varianz der

Ausprägungen, die Anzahl der untersuchten Objekte und Merkmale, sowie die Anzahl der Cluster und die Abstände zwischen den Gruppen. In der Folge führt er dann Störfaktoren ein, die in realen Anwendungen auftreten können. Dabei handelt es sich um die Clustergrößenstreuung, das Auftreten reiner Objekt- und/oder Merkmalscluster, solitäre Objekte bzw. Merkmale, und um fehlerhafte Daten (z.B. durch falsche Codierung).

Die aufwändige Erzeugung der Simulationsdaten ist ebenso ausführlich und damit verständlich beschrieben wie die Struktur des gesamten Simulationsablaufs selbst. Rix führt je nach Problemstellung zwischen 1.000 und 10.000 Simulationsläufe durch und vergleicht die erzeugten Klassifikationsergebnisse mit der wahren, bei der Ausgangdatenerzeugung vorgegebenen Klassenstruktur. Das hierzu benötigte goodness-of-fit Kriterium ist der Adjusted Rand Index, mit dessen Hilfe nicht nur ermittelt werden kann, welches Verfahren die am nächsten an der wahren Klassenstruktur liegende Klassifikation erzeugt, sondern auch, welches inneres Gütemaß am ehesten in der Lage ist, das beste Klassifikationsergebnis zu identifizieren. Beschränkungen und Variierungsmöglichkeiten für die Simulation sind so hervorragend gesetzt, dass der schwierige Trade-off zwischen Realitätsnähe und Analysierbarkeit der Ergebnisse als sehr gelungen zu bezeichnen ist. Der Verfasser zeigt, dass mit der quadrierten Zentroidabweichung (SCD) ein dominierendes inneres Gütemaß existiert und kann mit diesem Wissen nun ein zweimodales Austauschverfahren entwickeln. Er plädiert für die Berechnung zahlreicher Klassifikationsvarianten und die Identifikation des jeweils besten Ergebnisses mit Hilfe der Kennzahl SCD. Damit präsentiert Raimund Rix der scientific community im Bereich der explorativen Datenanalyse wesentliche Erkenntnisse und liefert dem Anwender eine wissenschaftlich fundierte Empfehlung zur Methodenwahl. Auf diese Weise leistet er auch der Verbreitung der Methoden in der Praxis Vorschub, denn eine möglichst objektive Bewertbarkeit der Ergebnisse ist eine der wichtigsten Voraussetzungen für die praktische Anwendung.

Ich wünsche dem Buch eine gute Aufnahme in den einschlägigen Kreisen und würde mich freuen, wenn die hier präsentierten Ergebnisse den Anstoß für weitere Forschungsvorhaben in diesem Bereich geben könnten.

Manfred Schwaiger