Zur Objektivität explorativer Bildstrukturerkennung mittels Clustering. Fehlerquellen und deren Vermeidung.

Ulrich Möller[1,2], Marc Ligges[2], Petra Georgiewa[2], Carolin Grünling[2], Bernhard Blanz[2] und Herbert Witte[1]

[1]Institut für Medizinische Statistik, Informatik und Dokumentation
[2]Klinik für Kinder- und Jugendpsychiatrie
Friedrich-Schiller-Universität Jena, 07740 Jena
Email: Ulrich.Moeller@med.uni-jena.de

Zusammenfassung. Ein wichtiger Bestandteil der (partitionierenden) Clusteranalyse ist die Validierung der in Frage kommenden Lösungen. Am Beispiel von drei Clusteralgorithmen und drei Validitätsfunktionen wurde u.a. für fMRI-Daten festgestellt, dass anhand der Validitätsfunktion die optimale Partition nur dann zuverlässig bestimmt werden kann, wenn zuvor die Zielfunktion der Clusteranalyse ausreichend optimiert wurde. Dazu wurde ein Konzept mit Versuchsserien und einer statistischen Evaluierung verwendet. Dies ermöglicht eine beliebig genaue Optimierung und liefert dem Anwender Hinweise, wann die Optimierung als ausreichend angesehen werden kann.

1 Einleitung

Bei der Interpretation medizinischer Bilddaten steht man oft vor dem Problem, dass a priori nicht genau bekannt ist, wie sich die gesuchten Merkmale in den Daten abbilden. Gründe dafür sind z.B. die Neuheit bzw. die technische oder/und methodische Komplexität des Verfahrens sowie die Variabilität des Untersuchungsgegenstands. Eine wichtige Rolle spielt daher die große Gruppe der explorativen Datenanalysemethoden, darunter die Clusteranalyse. Gemäß den Studien der letzten Jahre bieten sich besonders die partitionierenden Clusterverfahren an, um Strukturen anatomischer sowie funktioneller Bilder unüberwacht zu erkennen (vgl. [1] sowie zitierte Arbeiten in [2]). Dabei geht man nach einer zweistufigen Prozedur vor [3]. Zuerst wird mit einem Clusteralgorithmus die Partitionierung separat für festgelegte Clusteranzahlen entsprechend einer Zielfunktion optimiert. Danach gewinnt man aus einer Validitätsfunktion Hinweise darauf, bei welcher Clusteranzahl die Struktur der Daten am besten wiedergegeben wurde. Für eine adäquate Interpretation der Bilddaten ist es unverzichtbar, dass die Validitätsfunktion die tatsächliche bzw. optimale Clusteranzahl anzeigt, damit die ‚richtige' Partition zur Interpretation herangezogen wird.

Ein Scheitern der Prozedur bei der Erkennung vorgegebener Cluster wurde bislang auf Mängel der verwendeten Validitätsfunktion zurückgeführt und begründete die Suche nach neuen, geeigneteren Funktionen, z.B. [4, 5]. Im vorliegenden Beitrag wird demonstriert, dass insbesondere eine ungenügend optimierte Clusterbildung den Grund für falsche bzw. inkonsistente Validierungsergebnisse liefert, während man durch sorgfältige Optimierung ein korrektes Resultate erhält.

2 Methoden

Die erste Analysestufe wurde nacheinander mit drei Clusterverfahren realisiert: den bekannten Algorithmen c-means (CM) und fuzzy c-means (FCM) [1] sowie dem Algorithmus zur zufälligen Suche unter Clusterzentroiden (ZSZ, vgl. [2]). CM und FCM besitzen unterschiedliche Zielfunktionen. Beide sind Verfahren zur lokalen Optimierung. ZSZ besitzt dieselbe Zielfunktion wie CM, ist aber ein Verfahren zur globalen Optimierung. Für die zweite Analysestufe wurden drei Validitätsfunktionen verwendet [4-6], so dass die Ergebnisse jedes Clusterverfahrens mit mindestens zwei Funktionen validiert werden konnte. Eine geeignete Validitätsfunktion zeigt bei der wahren bzw. optimalen Clusteranzahl einen charakteristischen Peak (Extremum).

Die Ergebnisse der Verfahren CM und FCM sind abhängig von (zufällig gewählten) Startwerten; das Ergebnis des Verfahrens ZSZ kann vom zufälligen Verlauf des Suchprozesses abhängen. Es empfiehlt sich daher, jedes Verfahren mehrfach ablaufen zu lassen, das beste Ergebnis nach Möglichkeit zu reproduzieren und als Clusterlösung für die erste Stufe zu betrachten (vgl. [2, 6]). In der hier vorgestellten Untersuchung wurden die besten Partitionen aus A = 1, 3, 5, 10, 20 und 50 Algorithmusdurchläufen ermittelt, um die Abhängigkeit des Validierungsergebnisses vom Grad der Optimierung zu charakterisieren. Zur Untersuchung dieser Abhängigkeit wurde mit jedem Wert von A jede der 6 möglichen Analysevarianten (3 Clusterverfahren mit je zwei Validitätsfunktionen) 100 mal wiederholt, wobei die Startwerte, und bei ZSZ der Suchpfad, in jeder Analyse variierten.

3 Ergebnisse

Die Studie umfasste Analysen von über 40 Datenbeispielen, gegliedert in i) zweidimensionale Simulationsdaten, ii) multidimensionale simulierte Cluster mit typischen Merkmalen einer fMRI-Messung, und iii) Daten aus fMRI-Messungen. Die Studie ist an anderer Stelle ausführlicher dargestellt [7]. Die folgenden Ergebnisse (Abb. 1) besitzen exemplarischen Charakter. In Beispiel 1 wurden 9 zweidimensionale, normalverteilte Cluster simuliert (Abb. 2). In Beispiel 2 wurden 2618 80-dimensionale fMRI-Pixelzeitreihen analysiert, welche bei einer der möglichen Frequenzen einen herausragenden Leistungsanteil aufwiesen. Während der Aquisition der fMRI-Bilder 11-20, 31-40, 51-60 und 71-80 löste der Proband Aufgaben zur Sprachverarbeitung.

Abb. 1 (links) zeigt, dass bei geringem Optimierungsaufwand ($A = 1$) keine eindeutige Clustervalidierung möglich war. Mit wachsendem Optimierungsaufwand pro Analyse bildeten die Validitätsfunktionen aber eine eindeutige Kurvenstruktur aus (Abb. 1, rechts). Der kleinste Wert von A mit einem eindeutigen Validierungsergebnis kennzeichnet den Optimierungsaufwand, der mit dem jeweiligen Algorithmus im Mittel erforderlich war, um mit einer statistischen Sicherheit die optimale Clusteranzahl zu bestimmen. (Liefert eine von 100 Analysen eine Kurve, deren Extremum von dem der übrigen Kurven abweicht, dann beträgt die statistische Sicherheit 99%). Besitzen die Kurven ein klar hervortretendes Extremum bei der Clusteranzahl C^*, dann wird die Struktur der Daten durch eine Partition aus C^* Clustern am besten beschrieben. Dies trifft in drei Fällen zu (Abb. 1, rechts). Lediglich die Funktion VK im Beispiel 2 erreichte das globale Minimum für die Clusteranzahl 6 nur in ca. 60% der

Analysen, wobei der Optimierungsaufwand (A = 5) noch relativ gering war. Die Extrema der Kurven in Beispiel 1 weisen korrekt auf die Existenz von 9 Clustern hin (vgl. die Partition für C^* = 9 in Abb. 2 links). Die Extrema der Kurven in Beispiel 2 zeigen an, dass die Struktur der fMRI-Daten am besten durch 6 Cluster beschrieben wird (vgl. Teilergebnis in Abb. 3). Diese Cluster ließen sich gut in Zusammenhang mit den Hypothesen und dem Paradigma zum fMRI-Experiment interpretieren. Zudem wurden die Grundaussagen der Clusteranalyse mittels statistischer parameterischer Modellierung [8] bestätigt. Die Studie zeigte, dass die Funktion VK für gut strukturierte Daten eine eindeutige Validierung ermöglichen kann (vgl. Bsp. 1). Jedoch lieferte VK speziell mit fMRI-Daten weniger klare Ergebnisse als z.B. die Funktion VG. Simulationen zeigten, dass bei VG der Hinweis auf die optimale Clusteranzahl oft ein lokales Maximum ist, welches nach einem initialen Abfall der Kurve auftritt. (Die Analyse der Werte für die Zielfunktion in Abhängigkeit von der Clusteranzahl hilft, eine Fehlinterpretation des initialen Peaks zu vermeiden.)

Zwei Resultate belegen, welche Auswirkungen mangelnde Optimierung der Zielfunktion auf die Interpretation der Daten haben kann. Abb. 2 (rechts) zeigt, für Beispiel 1, die beste 4-Clusterpartition, welche – bei schlechter Optimierung für andere Clusterzahlen – als Ergebnis aus einer Validitätsfunktion abgeleitet wurde.

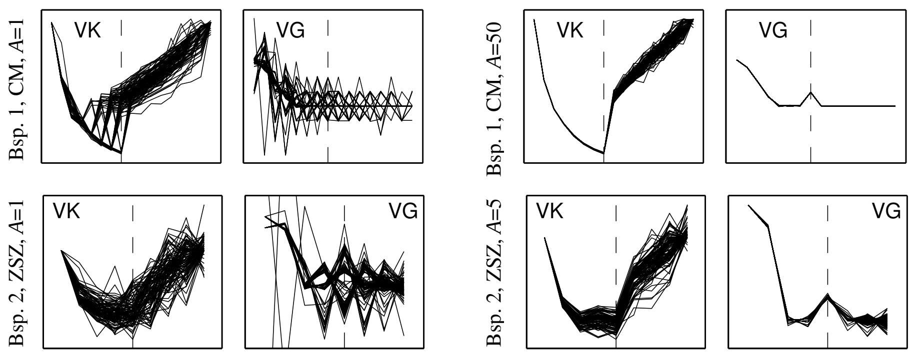

Abb. 1 Validitätsfunktionen VK nach [4] und VG nach [6] für 100 Clusteranalysen. Zur Algorithmuskonfiguration (CM, ZSZ und A) siehe den Abschnitt ‚Methoden'. VK und VG sind für die Clusterzahlen 2 bis 18 (Bsp. 1) bzw. 2 bis 10 (Bsp. 2) abgebildet. Die gestrichelten Linien markieren die Clusterzahlen 9 (Bsp. 1) bzw. 6 (Bsp. 2).

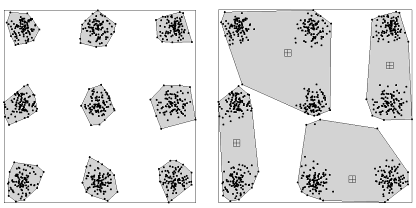

Abb. 2 Beispiel 1, 9-Clusterpartition, (tatsächliche Struktur) und 4-Clusterpartition

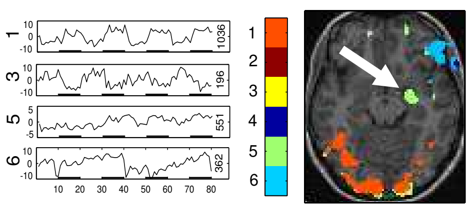

Abb. 3 Beispiel 2, Teil der 6-Clusterpartition mit einem von 46 anatomischen Schichtbildern

Abb. 3 enthält, für Beispiel 2, die Zentroide von 4 Clustern der besten 6-Clusterlösung und einen Teil der anatomischen Repräsentation der Cluster. Die Region aus Cluster 5, auf die der Pfeil zeigt, wurde durch die Sprachverarbeitungsprozesse aktiviert. In mehreren 3- und 4-Clusterpartitionen, die – bei schlechter Optimierung – als Ergebnis aus einer Validitätsfunktion abgeleitet wurden, wurde diese Region einem Cluster zugeordnet, dessen zeitliche Dynamik der von Cluster 3 in Abb. 3 entsprach. Damit würde diese Region als ein Hirnareal fehlinterpretiert, welches während der Kontrollbedingung aktiver ist als während der stimulierten Sprachverarbeitung.

Ähnliche Ergebnisse wurden auch mit den hier nicht durch Beispiele vertretenen Verfahrensvarianten erzielt, darunter den Methoden zum sog. fuzzy clustering [1, 5].

4 Diskussion

Es wurde gezeigt, dass die notwendige Validierung von Clusterergebnissen (Bestimmung der optimalen Clusteranzahl) eine ausreichende Optimierung der Partitionen voraussetzt. Letztere hängt von den Daten und der gewählten Verfahrensvariante ab. Dieser Befund wurde für mehrere der derzeit aktuellen Verfahrensvarianten (Clusteralgorithmen, Zielfunktionen, Validitätsfunktionen) bestätigt. Erzielt wurde der Befund mittels eines allgemein formulierten Gerüsts für eine mehrstufige Clusteranalyse, welches eine statistischen Evaluierung einschließt. Dies ermöglicht eine beliebig genaue Optimierung und liefert dem Anwender Hinweise, wann die Optimierung als ausreichend angesehen werden kann (siehe [7] für detailliertere Informationen).

Die Ergebnisse der Studie haben Bedeutung für die Anwendbarkeit des Clusteransatzes. Ihre Relevanz ergibt sich daraus, dass Clusteranalysen mit Algorithmen zur lokalen Optimierung weit verbreitet sind, wobei üblicherweise nicht evaluiert wird, ob der erreichte Grad der Optimierung eine sichere Validierung erlaubt.

Danksagung: Die Studie wurde gefördert durch die DFG (Projekt BL 435-3-1).

5 Literatur

1. Bezdek JC, Hall LO, Clarke LP: Review of MR image segmentation techniques using pattern recognition. Med. Phys. 20(4): 1033—1048, 1993.
2. Möller U, Ligges M, Grünling C et al.: Pitfalls in the clustering of neuroimage data and improvements by global optimization strategies. NeuroImage 14: 206—218, 2001.
3. Rezaee MR, Lelieveldt BPF, Reiber JHC: A new cluster validity index for the fuzzy c-mean. Pattern Recognition Lett 19: 237—246, 1998.
4. Kim DJ, Park YW, Park DJ: A novel validity index for determination of the optimal number of clusters. IEICE Trans Inf & Syst 84-D: 281—285, 2001.
5. Fadili MJ, Ruan S, Bloyet D et al.: On the number of clusters and the fuzziness index for unsupervised FCA application to BOLD fMRI time series. Med Image Anal 5: 55—67, 2001.
6. Goutte C, Toft P, Rostrup E et al.: On clustering fMRI time series. Neuroimage 9: 298—310, 1999.
7. Möller U, Ligges M, Georgiewa P et al.: How to avoid spurious cluster validation? A methodological investigation on simulated and fMRI data. NeuroImage, in Revision
8. SPM: Statistical Parametric Mapping. http:// www.fil.ion.ucl.ac.uk/spm