

Rezension von „*A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise*“ (Ester, et al. 1996).

## Aufbau

Der Aufbau der Abhandlung beginnt mit einer kurzen Zusammenfassung, der Einführung ins Thema in der die Problemstellung klar definiert wird, Vorstellung von anderen Clustering Algorithmen gefolgt von der DBSCAN Vorstellung die zugleich die Lösung zum zuvor beschriebenen Problem darstellt. Nachdem der Algorithmus im Detail vorgestellt wurde wird die Leistung verglichen. Im Vergleich, der anhand einer Referenzdatenbank mit realen geowissenschaftlichen Daten erstellt wird, mit anderen Lösungen wird der bis dahin bekannteste Algorithmus CLARANS herangezogen. Im letzten Abschnitt rekapitulieren die Autoren nochmals kurz den Algorithmus und geben einen Ausblick auf weitere Verbesserungen die sie geplant haben. Somit sind alle Punkte die in einer Abhandlung vorhanden sein sollten enthalten.

## Inhalt

Zu Beginn wird auf diverse Definitionen eingegangen, die die Herangehensweise bei der Findung des Algorithmus näherbringen sollen. Zugleich sind diese Definitionen fundamental für das Verständnis von DBSCAN selbst. Es ist – verglichen mit anderen Abhandlungen – einfach den Erläuterungen zu folgen. Schade ist, dass für mathematische Formeln kein geeignetes Format verwendet wurde. So benötigt der Leser teilweise mehr als einen Blick um herauszufinden ob es sich nun um eine Formel oder um ordinären Fließtext handelt. Die gegebenen grafischen Beispiele untermalen den Inhalt gut und helfen dem Leser so die Definitionen schneller zu erfassen und zu verstehen.

Den Definitionen folgt der eigentliche Algorithmus. Die Erläuterungen werden von Pseudocode begleitet. Die Lesbarkeit des Codes ist rückblickend schwer zu beurteilen, da vor nun doch fast 20 Jahren anders programmiert wurde. Persönlich halte ich die (C-typische) kurzschreibweise bei Variablen irreführend, sprechende Namen würden hier dem Leser sehr helfen. So müssen die Autoren im Fließtext teils selbst erläutern für was ein Variablenname steht und verwenden dort dann geeignete sprechende Namen (z. B. „*The `CLId` (clusterId) of points...*“).

In der Leistungsevaluierung wird mit dem CLARANS-Algorithmus verglichen. Der Vergleich zeigt einfach und

klar auf, dass DBSCAN ein viel besseres Verfahren zur Gruppenfindung ist als CLARANS. Die Ergebnisse sind durchweg korrekt, jedoch ist die Wahl der Daten so gewählt worden, dass DBSCAN durchweg sehr gute Ergebnisse liefert. Eine andere Datenwahl führt jedoch zu den gleichen Messergebnissen, einfach weil CLARANS wirklich schlechter ist als DBSCAN.

## Zusammenfassung

DBSCAN stellte damals ein völlig neues Verfahren zur Gruppenbildung in Spatial Databases (Geodatenbanken) dar und kommt somit der Lösung der Probleme von solchen Daten einen enormen Schritt näher als alle bis dahin bekannten Algorithmen. Heute stellt DBSCAN eine eigene Familie von Algorithmen dar, mit sehr vielen Abkömmlingen wie OPTICS, DeLi-Clu oder EnDBSCAN.

Die Problemstellung die DBSCAN zu lösen versucht ist bis heute ein großes Feld der Wissenschaft. In der Geowissenschaft gibt es unzählige Aufgabenstellungen in denen das Datenaufkommen unglaublich groß ist. Einen Verfahren zur einfachen Gruppenbildung ohne genaue Vorkenntnisse der Domäne selbst ist somit unerlässlich, jedoch auch sehr schwierig zu erarbeiten.

Alles in Allem halte ich die Abhandlung für sehr gut geschrieben, einfach zu verstehen und mit nur kleinen Kritikpunkten die vornehmlich die Formatierung von Formeln und Pseudocode betreffen. Jedoch vermisste ich eine klarere Ausarbeitung der Schwächen des Algorithmus. Es kommt der Irrglaube auf, dass DBSCAN eigentlich keine großen Schwächen besitzt, was nachweislich nicht der Fall ist.

## Literaturverzeichnis

Ester, Martin, Hans-Peter Kriegel, Jörg Sander, und

Xiaowei Xu. „A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise.“ *2nd International Conference on Knowledge Discovery and Data Mining (KDD-96)*. München: Institute for Computer Science, University of Munich, 1996. 6.

Wikipedia. *Cluster analysis*. 9. März 2013.

[http://en.wikipedia.org/wiki/Cluster\\_analysis](http://en.wikipedia.org/wiki/Cluster_analysis).