

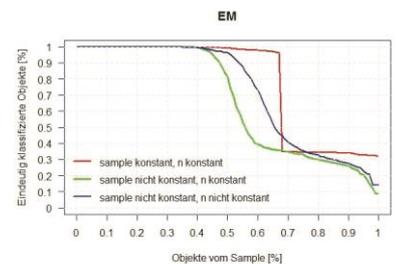
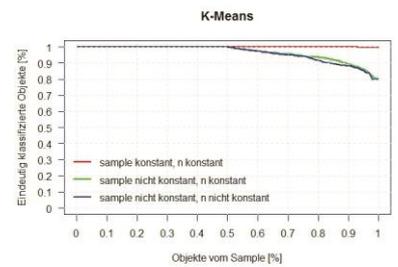
Vergleich des Stabilitätsverhaltens von zwei Clustering-Algorithmen auf einem Computertomografie-Datensatz

Aus verschiedenen Datenmengen sollen nützliche Erkenntnisse über gewisse Mustererkennung oder Gruppenbildungen sogenannte "Cluster" gezogen werden. Mittels deterministischen und stochastischen Algorithmen können solche Clusteranalysen durchgeführt werden. Da aber bei mehreren solcher Clusteranalysen auf einem Datensatz die Ergebnisse unterschiedlich sein können, zeigen die angewendeten Clusteralgorithmen kein stabiles Verhalten. In dieser Bachelorarbeit wurde ein Datensatz mit Computertomografieuntersuchungen an Patienten zur Analyse verwendet. Dieser Datensatz wurde mit einem deterministischen (K-Means) und einem stochastischen (EM: Expectation-Maximization für Normalverteilungen) Clusteralgorithmus analysiert. Dabei sollten die Algorithmen so optimiert werden, dass eine Clusteranalyse ein zuverlässiges Resultat liefert. Anschliessend sollte das Stabilitätsverhalten der Clusterbildungen mit unterschiedlichen Stichproben mittels einer Stabilitätsanalyse überprüft werden. Für eine Stabilitätsanalyse wurden gewisse Vorbereitungen getroffen. Die Clusteranzahl wurde mittels Methode von Ward bestimmt. Die Startwerte wurden so optimiert, dass die globalen Extrema gefunden werden konnten. Beim EM wurde überprüft, ob die Normalverteilungsannahme gilt. Dank der Optimierungen und Vorbereitungen konnten beim K-Means ein gutes Stabilitätsverhalten erreicht werden. Die Objekte in allen Stichproben wurden mit einer Wahrscheinlichkeit von 80% eindeutig demselben Cluster zugeordnet. Zu bemerken ist, dass bei 90% der Objekte schon eine Wahrscheinlichkeit von 90% erreicht wird. Die Abweichung der Zuteilung könnte eine Folge der Ausreisser im Datensatz sein. Der EM liefert kein befriedigendes Resultat für das Stabilitätsverhalten, da eine Verletzung der Normalverteilung vorliegt. Durch ein erweitertes Modell kann das Stabilitätsverhalten verbessert werden.



Diplomierende
Akilan Muthukirushnasamy
Benjamin Stocker

Dozierende
Christoph Zaugg
Elisabeth Dumont



Vergleich der Stabilitätsanalyse mit verschiedenen Stichproben