



Metadata-aware Clustering in R

Clustering-Verfahren wie selbstorganisierende Karten sind mächtige und in der modernen Informatik häufig genutzte Verfahren unüberwachten Lernens. Dank ihrer Fähigkeit aus Beispielen zu lernen ermöglichen sie eine Vielzahl innovativer Anwendungen. In der Abteilung Technische Informatik wird der Einsatz unüberwachter Lernverfahren unter anderem in Brain-Computer-Interfaces und zur biomedizinischen Datenanalyse erforscht.

Werden mehrere alternative Clusterings auf einem gegebenen Datensatz gefunden, ist es erforderlich, diese in weiteren Verarbeitungsschritten eindeutig und effizient unterscheiden zu können. Liegen für die Eingabedaten modell-basierte Metadaten vor, kann eine solche Unterscheidbarkeit durch Übertragung dieser Metadaten für jedes gefundene Clustering erreicht werden. In *R* und anderen Tools verfügbare Clustering-Packages erlauben jedoch derzeit keine automatisierte Erzeugung solcher Clustering-Metadaten aus Eingabedaten.

Im Rahmen der Bachelorarbeit soll ein neues *R*-Package entworfen und implementiert werden, das die automatisierte Generierung von modell-basierten Clustering-Metadaten ermöglicht und dazu eine metadaten-nutzende Repräsentation eines gegebenen Clusterings realisiert. Gleichzeitig soll eine einfache und modulare Integration bzw. Anbindung bestehender Clustering-Packages in das neue *R*-Package erreicht werden.

Die schriftliche Ausarbeitung kann auf Deutsch oder Englisch verfasst werden.

Anforderungen:

- Programmiererfahrung in der Programmiersprache *R*

Ansprechpartner:

Dr. Thomas Schmid
Paulinum, Raum P533
schmid@informatik.uni-leipzig.de