

Checkliste Lernziele:

- Wie lassen sich Sequenzdaten in R aufbereiten?
- Welche Möglichkeiten zur Visualisierung und Beschreibung von Sequenzen gibt es?
- Wie lassen sich Abstände nach dem Optimal Matching Verfahren berechnen?

Aufgaben:

1. Installieren und laden Sie das Paket `TraMineR`. Laden Sie den Datensatz `mvad`.
2. Bereiten Sie den in der ersten Aufgabe geladenen Datensatz so auf, dass Sie diesen mit den Befehlen des `TraMineR` Pakets auswerten können. Verwenden Sie hierbei Wertecodes und -labels.
3. Erstellen Sie eine Grafik, die die Sequenzen der Datenzeilen 11 bis 20 darstellt.
4. Erstellen Sie mit den Befehlen `seqfplot` und `seqdplot` Grafiken, wobei die Daten über die Variable `gcse5eq` in zwei Gruppen eingeteilt werden sollen. Interpretieren Sie die Ergebnisse!
5. Berechnen Sie für alle Beobachtungen Abstände nach dem Optimal Matching Verfahren. *indel*-Kosten sollen auf den Wert 1 gesetzt werden, alle Austauschkosten auf den Wert 2.
6. Berechnen Sie mit der Abstandsmatrix aus der letzten Aufgabe eine hierarchische Klassifikation. Verwenden Sie hierzu einen SAHN-Algorithmus Ihrer Wahl. Erstellen Sie anschließend ein Dendrogramm und diskutieren Sie die Ergebnisse (wieviele Cluster sollte man wählen?).
7. Erstellen Sie ausgehend von Ihrem Ergebnis in der letzten Aufgabe eine Gruppierungsvariable (`cutree`) und verwenden Sie diese, um Sequenzen nach Gruppen auszuwerten. Wie lassen sich die Gruppen interpretieren?

8. Führen Sie Aufgaben 6 und 7 erneut aus, allerdings unter Verwendung eines anderen SAHN-Algorithmus. Wie unterscheiden sich die Ergebnisse verglichen mit denen aus Aufgabe 6 und 7?