

Analyse et visualisation des séquences catégorielles

Nous utilisons dans ce travail la library TraMineR et arulesSequences conçues pour fouiller et visualiser les séquences d'états et le jeu de données mvad vu en cours.

Exercice 1: Création des séquences d'états

1. Charger TraMineR et mvad en utilisant les fonctions `library("TraMineR")` et `data("mvad")`.
2. Explorer le jeu de données mvad en affichant les premières lignes ainsi que les noms des colonnes, en calculant le nombre de lignes et de colonnes, et en affichant un résumé statistique.
3. Créer un objet séquences d'états avec les colonnes 17 à 86 en utilisant la fonction `seqdef()`.
4. Afficher les états mensuels de l'année 1994 des 10 premiers individus en format STS et SPS.

Exercice 2: Visualisation des séquences

1. Tracer le plot individuel (i-plot) des 20 premières séquences en utilisant la fonction `seqiplot()`.
2. Représenter les séquences en les regroupant selon l'obtention de diplômes à la fin de la scolarité obligatoire (gcse5eq), puis les trier par confession.
3. Donner les 5 séquences les plus fréquentes et tracer les avec le f-plot en utilisant respectivement les fonctions `seqtab()` et `seqfplot()`.
4. Visualiser les distributions transversales des états par groupe selon l'obtention de diplômes à la fin de la scolarité obligatoire avec la fonction `seqdplot()`.

Exercice 3: Calcul de l'entropie et de la turbulence

1. Calculer et comparer les entropies des deux premières séquences avec la fonction `seqient()`.
2. Calculer et comparer les turbulences des deux premières séquences avec la fonction `seqST()`.
3. Tracer la fréquence des turbulences pour toutes les séquences de mvad sous format d'histogramme.
4. Tracer une courbe pour comparer l'entropie et la turbulence de toutes les séquences

Exercice 4: Similarité des séquences

1. Calculer la distance entre les deux premières séquences en utilisant les méthodes vues en cours (LCP, LCS, ...)

Exercice 5: Extraction de motifs séquentiels

1. Appliquer l'algorithme `cSPADE` pour découvrir les motifs séquentiels fréquents dans les données mbad en utilisant la bibliothèque `arulesSequences` et en transformant la base de données dans un format adéquat.