

TraMineR: une librairie R pour l'analyse de données séquentielles

Alexis Gabadinho*, Nicolas S. Müller*, Gilbert Ritschard*, Matthias Studer*

*Département d'économétrie et Laboratoire de démographie, Université de Genève
{alexis.gabadinho,nicolas.muller,gilbert.ritschard,matthias.studer}@unige.ch
<http://mephisto.unige.ch/traminer/>

TraMineR est une librairie pour l'environnement statistique R destinée à la fouille, la description et la visualisation de séquences d'états ou d'événements, et plus généralement de données séquentielles discrètes. La librairie est librement disponible sur le CRAN <http://cran.r-project.org>. Elle est principalement conçue pour l'analyse de données biographiques longitudinales dans le domaine des sciences sociales, telles que des trajectoires professionnelles ou familiales. Cependant, les méthodes sont applicables à tout type de données séquentielles catégorielles.

La librairie TraMineR offre un ensemble unique de fonctionnalités dont de nombreuses originales. Elle s'intègre dans R, un logiciel libre, complet et multiplateforme offrant, par ailleurs, de nombreuses fonctions statistiques et graphiques avec lesquelles TraMineR peut interagir. Les fonctions de TraMineR permettent notamment de :

- manipuler et transformer différents formats de données longitudinales ;
- visualiser des séquences à l'aide de plusieurs types de graphiques ;
- déterminer le centrotype et mesurer la diversité d'un ensemble de séquences ;
- mesurer des caractéristiques longitudinales de séquences individuelles (longueur, durée par état, entropie, turbulence) ;
- déterminer des caractéristiques de l'ensemble des séquences (entropie par âge, taux de transitions, durées moyennes dans chaque état, fréquences) ;
- calculer la distance entre séquences à l'aide de plusieurs métriques (distance d'édition, plus longue sous-séquence commune, plus long préfixe commun) ;
- analyser l'hétérogénéité des séquences (méthodes du type analyse de variance et arbre d'induction) ;
- extraire les sous-séquences typiques d'événements ;
- identifier les sous-séquences les plus discriminantes.

La librairie est accompagnée d'un guide de l'utilisateur détaillé et illustré par de nombreux exemples que l'on trouve sur <http://mephisto.unige.ch>. TraMineR a été développé dans le cadre d'un projet de recherche sur la fouille d'histoires d'événements financé par le Fonds national suisse pour recherche scientifique.

