

## **Traiter des données de tracking avec R navigation sur un site web, suivi d'enquête web ou téléphonique**

**Anne GAYET**

Directrice Datamining  
A.I.D.  
4 rue Henri le Sidaner  
78000 Versailles  
agayet@aid.fr

**Mots clefs** : séquences d'événements, règles d'association, motifs séquentiels, co-clustering, trajectoires, données de tracking, webmining. Packages arules, arules sequences, TraMineR.

Les suites d'évènements que constituent les données de tracking nécessitent des méthodologies particulières pour être traitées. Ces données sont composées d'états ordonnés dans le temps, éventuellement associés à des durées. Par exemple :

- pour le suivi d'enquête téléphonique : les états sont les différentes tentatives d'appel,
- pour le suivi d'enquête web : les états sont les écrans affichant une ou plusieurs questions, les écrans sont horodatés lors de leur affichage,
- en webmining : les états sont les pages vues lors d'une visite sur un site web et sont horodatées.

Pour synthétiser et visualiser ces types de données, les méthodes traditionnelles d'analyse de données multidimensionnelles peuvent certes être utilisées : création d'indicateurs agrégés comme le nombre d'états, la durée totale, etc ... puis utilisation des méthodes d'analyse factorielle et / ou typologie. Toutefois elles ne prennent pas en compte la dimension temporelle des données, ni le besoin de traiter conjointement les individus (internauts, enquêtés) et les nombreux états suivis (pages vues, questions et items afférents). Des méthodes plus spécifiques le permettent : règles d'association, motifs séquentiels, clustering de séquences. Nous verrons comment mettre en œuvre ces méthodes avec les packages Arules, AruleSequences, et TraMineR.

Une autre façon de synthétiser ce type de données est d'utiliser le co-clustering (ou biclustering) sur la matrice des fréquences individus x nombre de passages par chaque état.

Par ailleurs, la visualisation des données détaillées d'une part, des synthèses obtenues d'autre part, nécessite des modes de représentation très différentes des visualisations traditionnelles. Nous verrons comment visualiser les associations ou motifs séquentiels en utilisant des graphes de liens dans lesquels les nœuds sont des états et les liens sont les fréquences de passage entre les états.

## Références

- [1] Nicolas S. Müller, Matthias Studer, Alexis Gabadinho, Gilbert Ritschard (2010). Analyse de séquences d'événements avec TraMineR. EGC 2010.
- [2] Alexandre Pollien (FORS), Dominique Joye (ISS), Michèle Ernst Stähli (FORS), Marlène Sapin (FORS) - Université de Lausanne (2012). Répondants et non-répondants dans les enquêtes, analyse des séquences de contact. 7ème colloque francophone sur les sondages, Rennes.
- [3] Alexis Gabadinho, Gilbert Ritschard, Matthias Studer and Nicolas S. Muller (2011) Department of Econometrics and Laboratory of Demography - University of Geneva. Mining sequence data in R with the TraMineR package: A user's guide <http://mephisto.unige.ch/traminer/>
- [4] Buchta, C., Hahsler, M. (2010). "arulesSequences: Mining frequent sequences". R package. <http://CRAN.R-project.org/package=arulesSequences>