

Analyse de séquences d'événements avec TraMineR

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

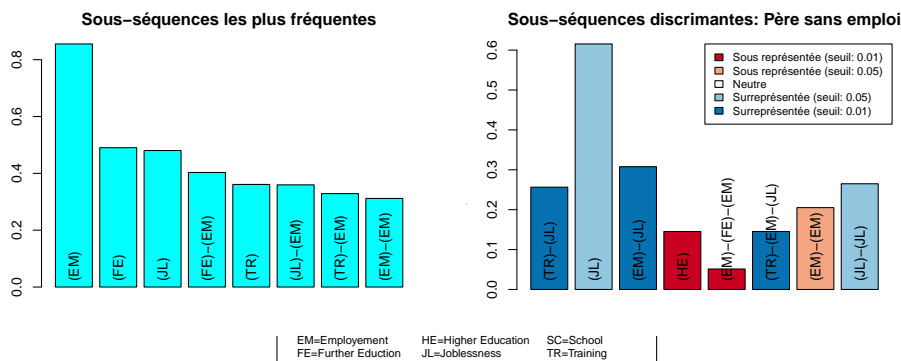
*Institut d'études démographiques et des parcours de vie, Université de Genève
{nicolas.muller, matthias.studer, alexis.gabadinho, gilbert.ritschard}@unige.ch
<http://mephisto.unige.ch/traminer/>

TraMineR et séquences d'événements

Les méthodes de fouille de motifs séquentiels fréquents ont donné lieu, depuis le travail fondateur de Agrawal et Srikant (1995), à un grand nombre de travaux de recherche. La librairie TraMineR pour l'environnement statistique R cherche à rendre ces méthodes accessibles aux chercheurs de divers domaines désirant analyser ou visualiser des ensembles de séquences d'événements. Il est à noter que TraMineR n'est pas adapté à l'étude d'une séquence unique d'événements. La librairie est disponible sur le CRAN (Comprehensive R Archive Network) et possède une aide en ligne ainsi qu'un manuel de l'utilisateur détaillé (Gabadinho et al., 2009).

TraMineR offre les fonctionnalités pour séquences d'événements suivantes :

- l'importation et la conversion de plusieurs formats de données et leur transformation en séquences d'événements,
- l'extraction des motifs séquentiels fréquents, en prenant en compte des contraintes temporelles telles que le temps maximum entre deux événements ou la taille de la fenêtre,
- le calcul de mesure de dissimilarité entre séquences d'événements afin de procéder, par exemple, à une classification non-supervisée des séquences (Studer et al., 2010)
- l'identification des motifs séquentiels fréquents les plus discriminants par rapport à une autre variable d'intérêt en se basant sur des tests du χ^2 de Pearson (Studer et al., 2010),
- la visualisation des motifs séquentiels discriminant à l'aide des résidus,



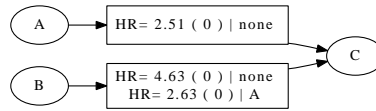


FIG. 1 – Règles d'association basées sur les modèles de durée

- l'extraction de règles d'association issues de la fouille de motifs séquentiels fréquents,
- et l'extraction de règles d'association séquentielles statistiquement significatives à l'aide d'une méthode basée sur des modèles de durée (Müller et al., 2010) ainsi que leur représentation graphique (Figure 1).

TraMineR ne se limite pas à l'analyse de séquence d'événements, mais permet également d'analyser des séquences d'états et des matrices de dissimilarités. Ainsi, cette librairie a déjà été utilisée en sciences sociales (Demont et Heuveline, 2008; Massoni et al., 2009) et dans d'autres domaines tels que la biologie ou l'analyse de comportement interface-utilisateur en informatique (Reilly et al., 2009).

TraMineR étant une librairie R, les résultats produits peuvent être analysés à l'aide de toute méthode statistique disponible dans R. Écrit en R et en C++, le code source de TraMineR est accessible en licence GPL et donc modifiable par tous. La librairie TraMineR est suivie de près par ses développeurs, maintenue et mise à jour régulièrement.

Références

- Agrawal, R. et R. Srikant (1995). Mining sequential patterns. In P. S. Yu et A. L. P. Chen (Eds.), *Proceedings of the International Conference on Data Engineering (ICDE), Taipei, Taiwan*, pp. 487–499. IEEE Computer Society.
- Demont, F. et P. Heuveline (2008). Diversity and change in cambodian households, 1998–2006. *Journal of Population Research* 25(3), 287 – 313.
- Gabadinho, A., G. Ritschard, M. Studer, et N. S. Müller (2009). Mining sequence data in R with TraMineR: A user's guide. Technical report, Department of Econometrics and Laboratory of Demography, University of Geneva, Geneva.
- Massoni, S., M. Olteanu, et P. Rousset (2009). Career-path analysis using optimal matching and self-organizing maps. In *Advances in Self-Organizing Maps: 7th International Workshop, WSOM 2009, St. Augustine, FL, USA, June 8-10, 2009*, Volume 5629 of *Lecture Notes in Computer Science*, pp. 154–162. Berlin: Springer.
- Müller, N. S., M. Studer, G. Ritschard, et A. Gabadinho (2010). Extraction de règles d'association séquentielle à l'aide de modèles de durée. *Revue des nouvelles technologies de l'information RNTI*. (in press).
- Reilly, D. F., K. M. Inkpen, et C. R. Watters (2009). Getting the picture : Examining how feedback and layout impact mobile device interaction with maps on physical media. *Wearable Computers, IEEE International Symposium*, 55–62.
- Studer, M., N. S. Müller, G. Ritschard, et A. Gabadinho (2010). Classifier, discriminer et visualiser des séquences d'événements. *Revue des nouvelles technologies de l'information RNTI*. (in press).