

Classer, discriminer et visualiser des séquences d'événements

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

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

Résumé. Cet article ¹ présente un ensemble d'outils destiné à analyser des séquences d'événements en sciences sociales et à visualiser les résultats obtenus. Nous commençons par formaliser la notion de séquence d'événements avant de définir une mesure de dissimilarité entre ces séquences afin de construire des typologies et de tester les liens entre ces séquences et d'autres variables d'intérêts. Initialement définie par Moen (2000), cette mesure se base sur la notion de distance d'édition entre séquences et permet d'identifier les différences d'ordonnement et de temporalité des événements. Nous proposons une extension de celle-ci afin de pouvoir prendre en compte la simultanéité des événements ainsi qu'une méthode de normalisation qui garantit le respect de l'inégalité triangulaire. Dans un deuxième temps, nous présentons un ensemble d'outils destinés à interpréter les résultats. Nous proposons ainsi deux méthodes de visualisation d'un ensemble de séquences et nous introduisons la notion de sous-séquence discriminante qui permet d'identifier les différences d'ordonnement des événements les plus significatives entre groupes. L'ensemble des outils présentés est disponible au sein de la librairie R TraMineR.

1 Introduction

L'analyse de séquences en sciences sociales se centre essentiellement sur deux types de questions. Premièrement, dans une optique descriptive, on cherche à construire une typologie des séquences, ce que l'on réalise généralement à l'aide d'une classification automatique. Deuxièmement, on s'intéresse à comprendre les liens entre ces séquences et d'autres variables d'intérêt afin de mettre en relation ces séquences avec leur contexte. Il existe déjà plusieurs méthodes pour répondre à ces questions dans le cas de séquences d'états. Dans cet article, nous développons un ensemble de méthodes destinées à analyser des séquences d'événements. Toutes ces méthodes sont disponibles au sein de la librairie R TraMineR (Gabadinho et al., 2009).

La méthode généralement utilisée en sciences sociales consiste à recoder les séquences d'événements sous forme de séquence d'états puis d'utiliser une mesure de dissimilarité dérivée des méthodes d'alignement de séquences (Abbott, 1990). Cette méthode a deux désavantages. Premièrement, les méthodes d'alignement de séquences ne prennent qu'indirectement en

1. Etude réalisée avec le soutien financier du Fonds national suisse (FNS), subside 100015-122230.