Position wise group-typical states

Matthias Studer

matthias.studer@unige.ch

"Position wise group-typical states" aims at looking at the relationship between a state sequences and a covariate (Studer, 2015). The aims of this brief introduction is to show how it can be interpreted and used with R using a short example: how the school-to-work trajectories in Northern Ireland are linked to the qualifications gained by the end of compulsory education (variable `gcse5eq`). If you use this method, please cite the paper Studer (2015). Before going further, let us build the state sequence object and print frequency table of variable `gcseq5eq`. In this variable, "yes" means "good results."

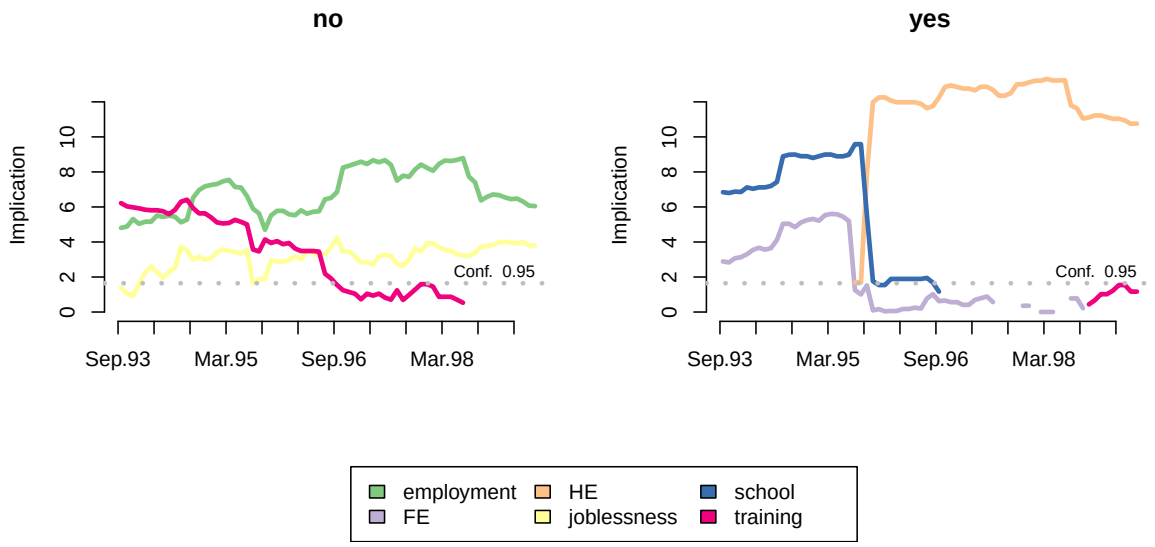
```
library(TraMineR)
## Load our example data set
data(mvad)
## Build the state sequence object
mvad.seq <- seqdef(mvad, 17:86, xtstep = 6)
```

This analyis is ran using the command `seqimplic` in the `TraMineRextras` package. We must specify the state sequence object and the covariate as follows:

```
library(TraMineRextras)
mvad.si <- seqimplic(mvad.seq, group=mvad$gcse5eq)
```

This object can then be plotted to visualize the results as follows (`lwd` is used to increase the line width):

```
plot(mvad.si, lwd=3)
```



This Figure offers a visualization of the differences between two or more groups of trajectories. It presents at each time points the typical states of a subpopulation (here according to qualifications at the end of compulsory schooling) using implicative statistics. The implicative statistic assesses the statistical relevance of a rule of the form "A implies B." Please read the short article Studer (2015) for a more detailed presentation or the appendix below.

The "no" subplot of the above figure presents at each time point t, the relevance of the rule "Not having qualification implies being in state A at time t". The horizontal dashed lines present the confidence

thresholds. A rule is considered statistically significant at the 5% if it exceeds the 95% confidence horizontal line. For instance, here, being in the group "Not having qualification" implies being employed (also in the joblessness state) and this rule is significant for the whole trajectory. Interestingly, some rules are only significant for a given period. Having a qualification implies being in School or Further Education in the beginning of the trajectories (but not later) and being in Higher Education afterward.

"Position wise group-typical states" allows identifying the most typical states of a group at a given position. It can be used to study the association between covariate and sequences, but it is also very useful to compare and interpret a cluster analysis. By using the "cluster" variable as the grouping variable, one can easily identify the typical states at each time position of each cluster.

References

Matthias Studer (2015), Comment: On the Use of Globally Interdependent Multiple Sequence Analysis, *Sociological Methodology* 45, DOI: 10.1177/0081175015588095

Gras, Régis and Kuntz, Pascale. (2008), An overview of the Statistical Implicative Analysis (SIA) development, in Gras, R., Suzuki, E., Guillet, F. and Spagnolo, F. (eds), *Statistical Implicative Analysis: Theory and application*, Series Studies in Computational Intelligence, Vol 127, Berlin: Springer-Verlag, pp 11-40.

Ritschard, G. (2005). De l'usage de la statistique implicative dans les arbres de classification. In Gras, R., Spagnolo, F., and David, J., editors, Actes des *Troisièmes Rencontres Internationale ASI Analyse Statistique Implicative*, volume Secondo supplemento al N.15 of Quaderni di Ricerca in Didattica, pages 305-314. Università a degli Studi di Palermo, Palermo.

Appendix: Implicative statistic

The implicative statistic framework aims to assess the statistical relevance of a rule of the form "A implies B" by measuring the gap between the expected and observed numbers of counter examples (Gras et al., 2008). The rule is considered to be strongly implicative if we observe much less counter examples than expected under the independence assumption. This gap and its significance are computed using adjusted residuals of a contingency table with continuity correction as proposed by Ritschard (2005). In order to improve the readability of the graphs, we use here the opposite of the implicative statistic, which is highly negative for significant rules. The statistic $I(A \rightarrow B)$ measuring the relevance of the rule "A implies B" reads as follows:

$$I(A \rightarrow B) = -\frac{n_{\bar{B}A} + 0.5 - n^e_{\bar{B}A}}{\sqrt{n^e_{\bar{B}A}(n_{B.}/n)(1 - n_{.A}/n)}}$$

Where $n_{\bar{B}A}$ is the observed number of counter-examples, $n^e_{\bar{B}A}$ the expected number of counter-examples under the independence assumption, $n_{B.}$ the number of times that B is observed, $n_{.A}$ the number of times that A is observed and n the total number of cases.