How many are too many? Methods to enumerate underlying trajectories with Mixture Hidden Markov Models and Sequence Analysis

Mauricio Garnier-Villarreal   Dimitris Pavlopoulos

Vrije Universiteit Amsterdam

Typological analysis of longitudinal data, i.e. classifying sequences of data from different individuals for a particular phenomenon, is becoming increasingly popular in social sciences. When we have longitudinal data, a main focus is to describe the longitudinal pattern, like the slope in growth curve and the transition matrix in Markov models. Conclusions about these patterns might be incorrect if there is group heterogeneity, meaning that subgroups of subjects present different longitudinal patterns (trajectories). Here we focus on methods that can detect this heterogeneity when the phenomena of interest is categorical, like Mixture Hidden Markov Models (MHMM) and Sequence Analysis (SA). The issue of class enumeration is not new, but most of the research has been done in cross-sectional models, and when looking at longitudinal models some focus has been done in with growth mixture models, where the phenomena is continuous in nature. With a simulation study, we evaluate methods to select the correct number of trajectories. For SA, we will compare the measures of the quality of a partition described by Studer (2013), and for MHMM we will compare the information criteria commonly used for class enumeration such as AIC, BIC (Masyn, 2013), and tests like BLRT (McNeish & Harring, 2017). We test these methods across the conditions of number of true trajectories, level of overlap, trajectory types, and unbalance samples. With those conditions we intend to evaluate the methods across realistic data conditions. We will present our simulation results, and recommendations for applied researchers in how to detect group heterogeneity in longitudinal patterns. This way we are able to properly describe longitudinal pattern heterogeneity.

References

Studer, Matthias (2013). WeightedCluster Library Manual: A practical guide to creating typologies of trajectories in the social sciences with R. LIVES Working Papers, 24. DOI: 10.12682/lives.2296-1658.2013.24.

Masyn, K. E. (2013). Latent Class Analysis and Finite Mixture Modeling. In P. E. Nathan & T. D. Little (Eds.), The Oxford Handbook of Quantitative Methods: Vol. Volume 2: Statistical Analysis (p. 63). Oxford University Press.

McNeish, D., & Harring, J. R. (2017). The Effect of Model Misspecification on Growth Mixture Model Class Enumeration. Journal of Classification, 34(2), 223–248. https://doi.org/10.1007/s00357-017-9233-y