Utilizing the multilevel hidden Markov model in social and behavioral data: the R CRAN package mHMMbayes and empirically based guidelines on sample size requirements

Emmeke Aarts & Sebastian Mildiner-Moraga

Utrecht University, Faculty of Social and Behavioural Sciences,

Department of Methodology and Statistics

Keywords: Hidden Markov model, Multilevel modeling, Bayesian estimation, Intense longitudinal data

The multilevel (also known as mixed or random effects) hidden Markov model - a generalization of the hidden Markov model (HMM) - is a promising vehicle to investigate latent dynamics over time in social and behavioral processes in intense longitudinal data. The multilevel HMM is tailored to accommodate data of multiple individuals simultaneously, allowing for heterogeneity in the model parameters (transition probability matrix and conditional distribution), while estimating one overall HMM. Hence, the multilevel framework facilitates the study of individual-specific trajectories and the study of individual differences.

An open-source implementation of the multilevel hidden Markov model is provided by the R CRAN package mHMMbayes. The model can be fitted on multivariate data with a categorical or normal (i.e., Gaussian) distribution, and include individual level covariates (allowing for e.g., group comparisons on model parameters). Parameters are estimated using Bayesian estimation utilizing the forward-backward recursion within a hybrid Metropolis within Gibbs sampler. The package also includes various visualization options, a function to simulate data, and a function to obtain the most likely hidden state sequence for each individual using the Viterbi algorithm.

In addition, we provide guidelines on sample size requirements - currently still lacking for typical social and behavioral data in combination with the multilevel HMM. The guidelines are based on extensive simulation studies and are driven by the complexity of the data and the study objectives of the practitioners.