Gaining Insight into Determinants of Physical Activity using Bayesian Network Learning
Simone C.M.W Tummers, Arjen Hommersom, Lilian Lechner, Catherine Bolman, and Roger Bemelmans
Open University of the Netherlands and Zuyd Hogeschool

Case study background
Interventions aim to influence behavioural determinants (factors determining a certain behaviour) in order to change (health-related-)behaviour of participants. In this paper, Bayesian network learning is applied to data from the Active Plus [3] intervention (aiming at influencing physical activity behaviour among older adults).

Bayesian network model learning
- Bayesian network (BN) [2] = a probabilistic model represented as directed acyclic graph
- Structure learning procedure:
  - Tabu search algorithm
  - Greedy search algorithm, avoiding local optima (by random restarts, option to select a slightly worse network in next iteration and tabulist)
  - Model selection criterion: Bayesian Information Criterion

Dealing with missing data
- Major problem in this case study
  - More than a fourth of the data is missing (in variables measuring intervention effects)
  - 360 complete records out of 1976
- Evaluated 2 methods to handle this problem:
  - Mean imputation
  - Structural Expectation Maximization (EM) algorithm [1]: combines BN model learning with the estimation of missingness (based on model parameters)
- Results from 10-fold cross-validation:
  - The structural EM outperforms mean imputation in this case study

Bayesian network model for intervention data
- Learnt a linear Gaussian temporal Bayesian network model
  - Applying described learning procedure with SEM algorithm for missingness
  - Bootstrapping applied to verify edge stability
    - Stable edges (in black, figure 3): appear in both the learnt model and in most models for bootstrap samples
  - Observations resulting model after bootstrapping:
    - Most edges are stable; quite some are not
    - Complete overview of complexity with which determinants are correlated and determine physical activity
  - Observations in highlighted submodel:
    - Previous result verified by the network: intention mediates intervention effect on physical activity
    - No influence found of intervention on self-efficacy (previous result); explained by path via several other determinants (whole model)

Conclusions
- Bayesian network model applied to new field
- The model reveals the complex dependence structure between physical activity and its determinants
  - Previous findings confirmed
  - Advantage compared to previous analyses: insight in complex mediation paths
- Missing data problem:
  - Magnitude in this case study shown
  - Verified that SEM algorithm outperforms mean imputation to handle it

Future research
- Dive more into structure found
- Consider general model over more studies (note: some unstable edges)
- Consider multiple Imputation to handle missing data

References

Contact Information: Simone.tummers@ou.nl +31 45 576 2874
Acknowledgement of funding: This research was funded by ZonMw.

Table 1. Derived variables included in analysis and number of missing values per variable

Table 2. Results from cross-validation analysis

Table 3. Results from cross-validation analysis

Scenario

Algorithm 1: Structural EM algorithm, given (θ_M, s)
For n = 0, 1,..., until convergence or predefined maximum number of iterations reached do
    Compute M_θ^N, where P_i(θ, s) = \text{maximize} \text{log-likelihood} \text{over } θ
    Maximize s: apply structure learning to determine M_θ, using data w/θ
    return M_θ
end if

Figure 1. Timeline and moments of measurement in Active Plus [3], intervention

Figure 2. Pseudocode of structural Expectation Maximization algorithm

Figure 3. Submodel of the averaged Bayesian network, including stable edges