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
    - In class of score-based algorithms
    - 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
- Parameter learning procedure: maximum likelihood estimation

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 missings
    (based on model parameters)
- Results from 10-fod 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 missings
- 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 network 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)

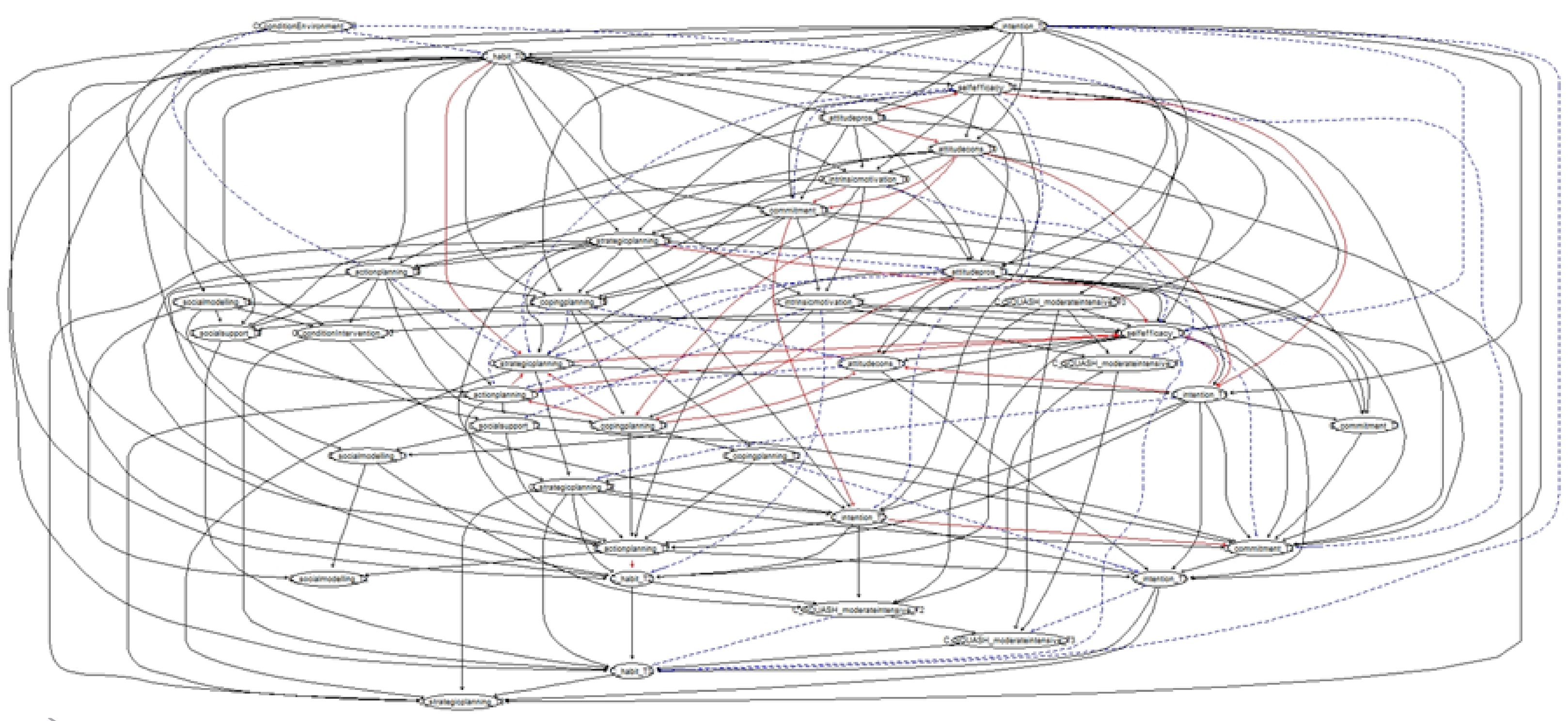

Figure 3. Averaged model, including false positives (blue) and false negatives (red) from the model learnt for original data set

Concept	Timeslot	Number of missing values (out of 1976)
Condition: intervention	T0	8
Condition: environment	T0	8
SQUASH outcome measure	T0	3
	T1	518
	T2	565
	T3	628
Self-efficacy	T0	229
	T1	638
Attitude(-pros)	T0	149
	T1	587
Attitude(-cons)	T0	167
	T1	597
Intrinsic motivation	T0	325
	T1	690
Intention	T0	141
	T1	571
	T2	654
	T3	748
Commitment	T0	31
	T1	531
	T2	573
Strategic planning	T0	156
	T1	601
	T2	652
	T3	661
Action planning	T0	182
	T1	604
	T2	686
Coping planning	T0	192
	T1	621
	T2	668
Habit	T0	136
	T2	633
	T3	662
Social modelling	T0	532
	T1	915
	T2	952
Social support	T0	68
	T1	561

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

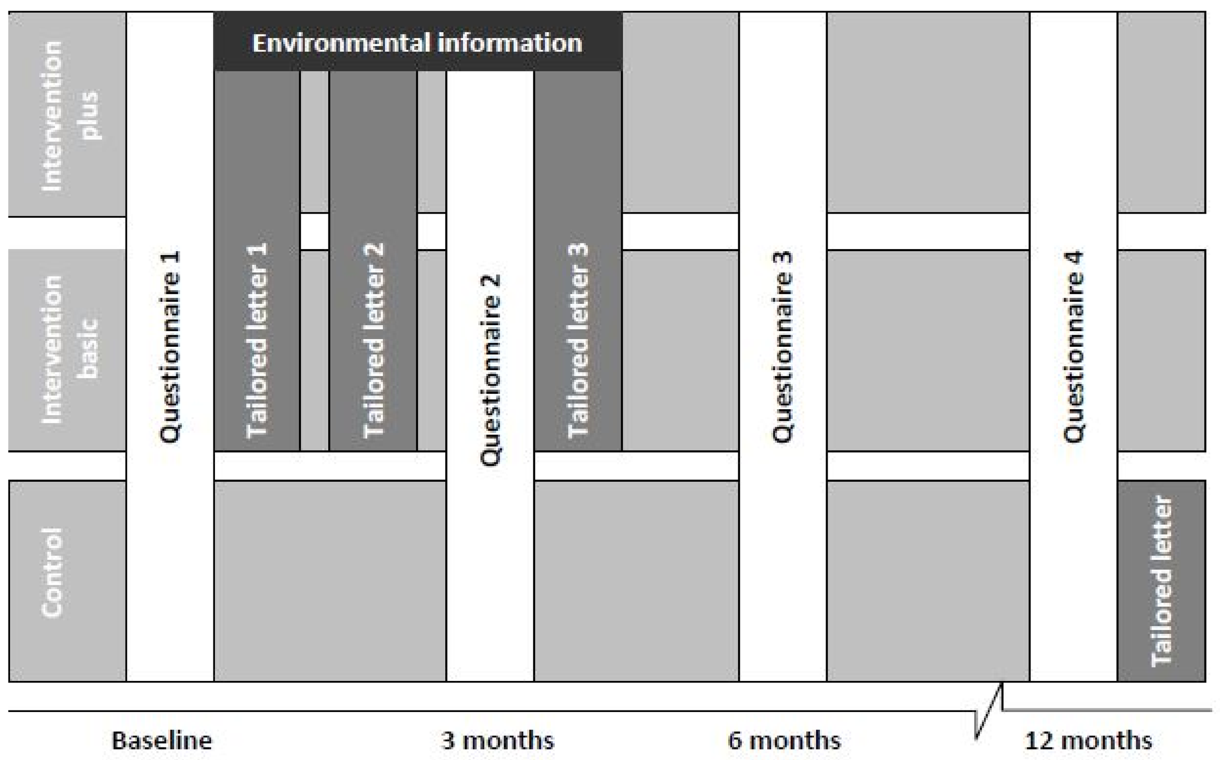

Figure 1. Timeline and moments of measurement in Active Plus [3] intervention

Algorithm 1 Structural EM algorithm, given $(M_0, \mathbf{o})$:

for $n = 0, 1, \ldots$ until convergence or predefined maximum number of iterations reached **do**
  Compute Θ^{M_n} using a parameter learning algorithm.
  Expectation-step:
  compute $\mathbf{h}^* = \arg\max_{\mathbf{h}} \mathbb{P}(\mathbf{h} \mid \mathbf{o}, M_n)$
  Maximization-step: apply structure learning to determine M_n using data $\mathbf{h}^* \cup \mathbf{o}$
  if $M_n = M_{n+1}$ or if stopping criterion is met **then**
    return M_n
  end if
end for

Figure 2. Pseudocode of structural Expectation Maximization algorithm

Handling missing values	Mean log-likelihood	95% Confidence Interval
Mean imputation	-4779	[-4832;-4726]
SEM algorithm	-4127	[-4183;-4071]

Table 2. Results from cross-validation analysis

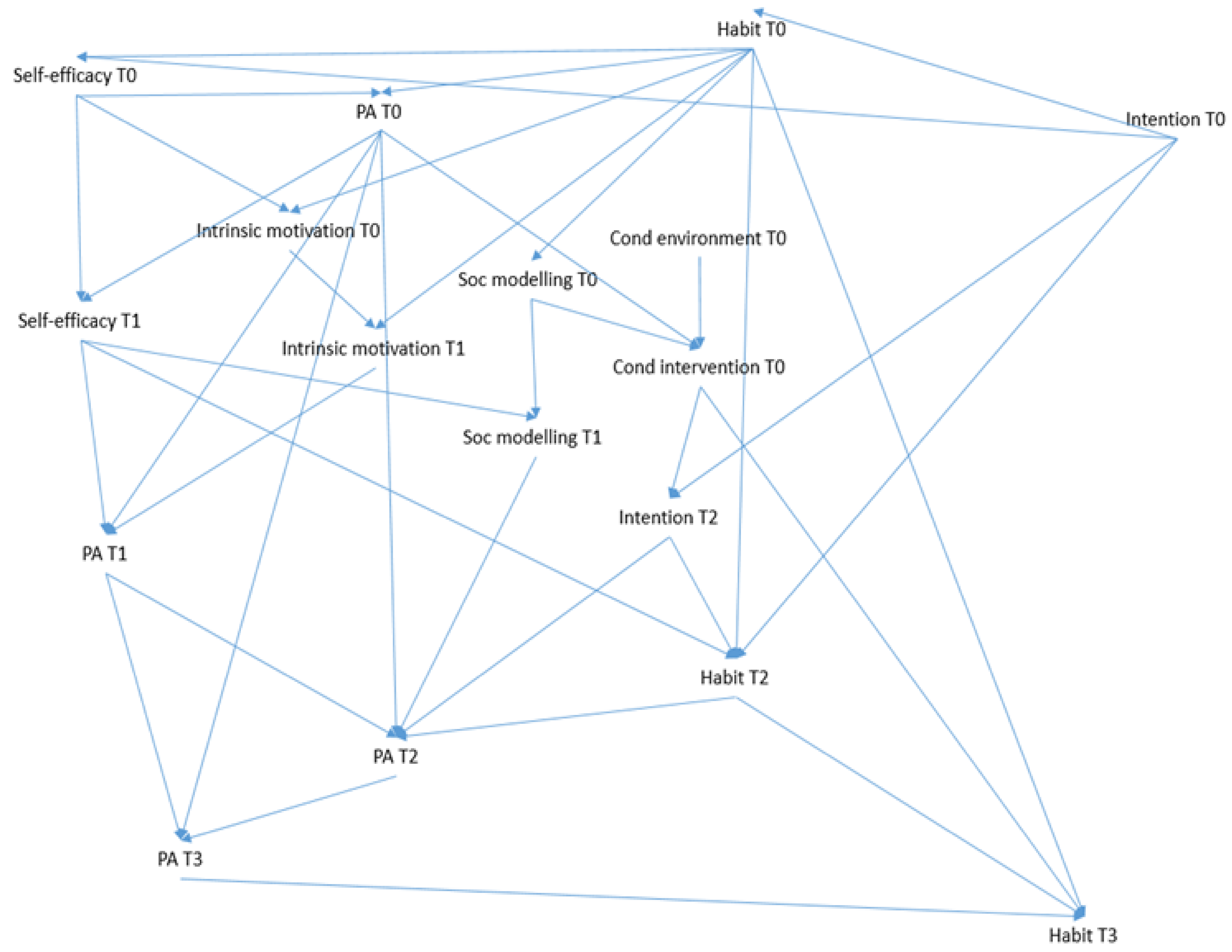

Figure 4: Submodel of the averaged Bayesian network, including stable edges

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

1. Friedman, N.: The Bayesian structural EM algorithm. In: Proceedings of the Four-teenth conference on Uncertainty in artificial intelligence. pp. 129–138 (1998)

2. Pearl, J.: Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann, San Mateo, CA (1988)

3. van Stralen, M.M., Kok, G., de Vries, H., Mudde, A.N., Bolman, C., Lechner, L.: The active plus protocol: systematic development of two tailored physical activity interventions for older adults. BMC Public Health 8 (2008)

Contact information: *simone.tummers@ou.nl, +31 45 576 2874*

Acknowledgement of funding: This research was funded by ZonMw.

Open Universiteit
www.ou.nl