Monotonicity in Bayesian Networks for Computerized Adaptive Testing

Martin Plajner, Jirka Vomlel

Faculty of Nuclear Sciences and Physical Engineering,
Czech Technical University
The Institute of Information Theory and Automation,
Czech Academy of Sciences

10.7.2017 - 14.7.2017
Lugano
Switzerland

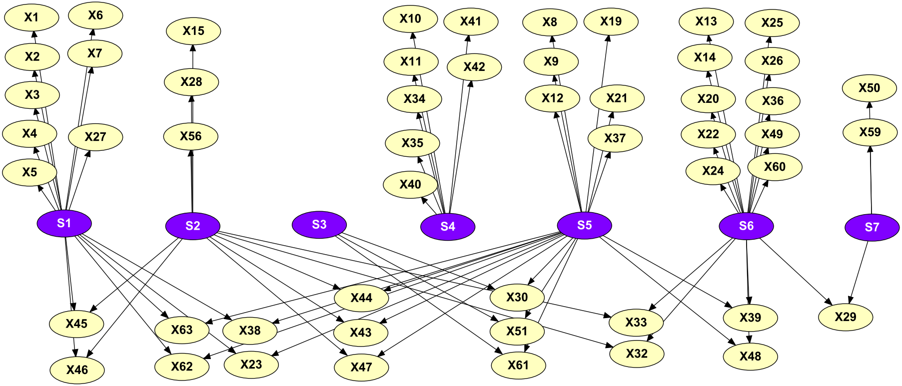
Adaptive Testing

- Selection of questions' subsets.
- Shorter test versions.
- Individual sets of questions.
- Improved precision and understanding of student's skills.
- Students are modeled by a student model. BN in our case.

Adaptive Testing Procedure

- Select a next question.
- Ask the question.
- Update the model.
- Estimate student's skills/answers.

Expert Network Model



Single Parent Monotonicity

With a natural ordering of states of a skill variable S_j

$$s_{j,1} \prec \ldots \prec s_{j,m_j} \quad ,$$

the monotonic effect on its child question variable X_i is

$$s_{j,k} \preceq s_{j,l} \quad \Rightarrow \quad P(X_i = 1 | S_j = s_{j,k}, \boldsymbol{s}) \; \leq \; P(X_i = 1 | S_j = s_{j,l}, \boldsymbol{s}) \quad .$$

Model Monotonicity

With multiple parents of a question X_i and their states configurations $(\boldsymbol{s}^i, \boldsymbol{r}^i)$,
we create a partial ordering of these configurations based on their effect on the child X_i

$$\boldsymbol{s}^i \quad \preceq_i \quad \boldsymbol{r}^i \quad .$$

Then the monotonicity condition is

$$\boldsymbol{s}^i \preceq_i \boldsymbol{r}^i \quad \Rightarrow \quad P(X_i = 1 | \boldsymbol{S}^i = \boldsymbol{s}^i) \ \leq \ P(X_i = 1 | \boldsymbol{S}^i = \boldsymbol{r}^i) \quad .$$

Reasons to Use Monotonicity

- Sensible requirement in many applications.
- Experts acceptance.
- Additional information $\rightarrow$ easier/more precise learning.

Learning Parameters under Monotonicity

- Monotonicity:
  van der Gaag, L., Bodlaender, H. L., and Feelders, A. J. (2004). Monotonicity in Bayesian networks. UAI2004
- Gradient learning method (motivation method):
  Altendorf, E. E., Restificar, A. C., and Dietterich, T. G. (2005). Learning from Sparse Data by Exploiting Monotonicity Constraints. UAI2005
- Isotonic regression EM (comparison method):
  Masegosa, A. R., Feelders, A. J., and van der Gaag, L. (2016). Learning from in- complete data in Bayesian networks with qualitative influences. IJAR

Learning Parameters under Monotonicity

Given the model parameters $\boldsymbol{\mu} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_n, \boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_m)$,

$$\theta_{i,\boldsymbol{s}^i} = P(X_i = 0 | \boldsymbol{S}^i = \boldsymbol{s}^i), \; \boldsymbol{\theta}_i = (\theta_{i,\boldsymbol{s}^i})_{\boldsymbol{s}^i \in Val(\boldsymbol{S}^i)} \;,$$

we use the model's log likelihood

$$LL(\boldsymbol{\mu}) \;,$$

which we penalize:

$$\begin{aligned} p(\theta_{i,\boldsymbol{s}^i}, \theta_{i,\boldsymbol{r}^i}) &= exp(c \cdot (\theta_{i,\boldsymbol{r}^i} - \theta_{i,\boldsymbol{s}^i})) \\ LL'(\boldsymbol{\mu}, c) &= LL(\boldsymbol{\mu}) - \sum_{i \in \boldsymbol{N}} \sum_{\boldsymbol{s}^i \preceq_i \boldsymbol{r}^i} p(\theta_{i,\boldsymbol{s}^i}, \theta_{i,\boldsymbol{r}^i}) \;. \end{aligned}$$

Learning Parameters under Monotonicity

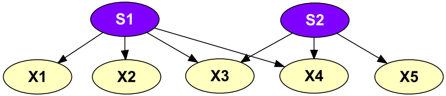
Penalized log likelihood

$$LL'(\boldsymbol{\mu}, c) = LL(\boldsymbol{\mu}) - \sum_{i \in \boldsymbol{N}} \sum_{\boldsymbol{s}^i \preceq_i \boldsymbol{r}^i} p(\theta_{i,\boldsymbol{s}^i}, \theta_{i,\boldsymbol{r}^i})$$

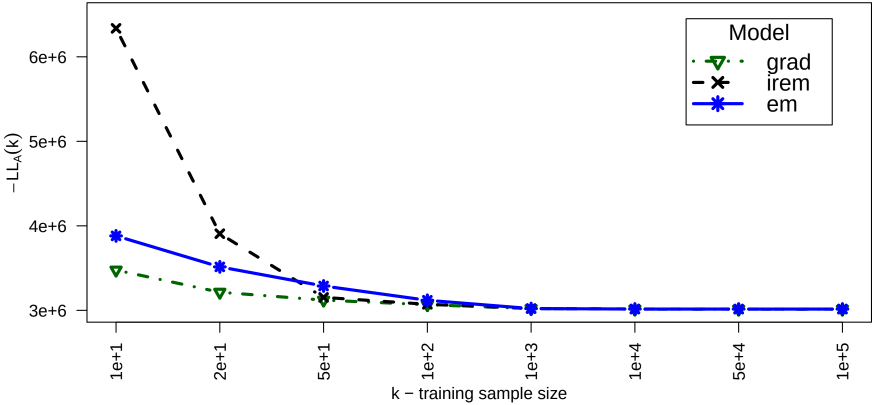
is optimized using gradient methods.

Experiments

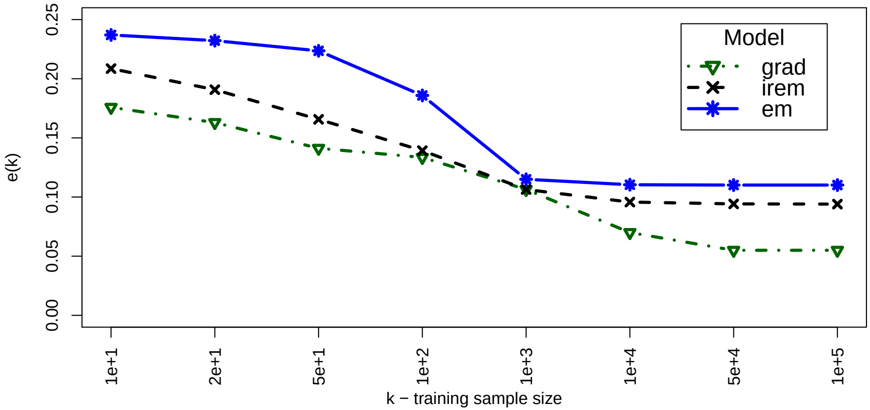
- Experimental evaluation with
  - Empirical data set - Math test, expert model, 281 cases
  - Synthetic data set - 100 000 cases



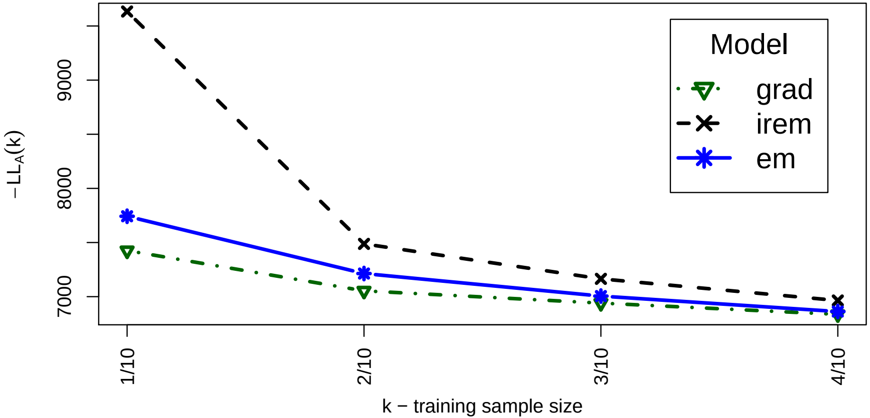
Results - Synthetic Model Log Likelihood



Results - Synthetic Model Parameters



Results - Empirical Model Log Likelihood



Conclusions

- Gradient method for monotonic parameters learning with hidden variables.
- Provides good results for small training sets.
- Comparable results with other methods for larger training sets.
- Generalization for less specific network structure is required.