Student Skill Models in Adaptive Testing

Martin Plajner, Jirka Vomlel

The Institute of Information Theory and Automation
Czech Academy of Sciences

6.9.2016 - 9.9.2016
Lugano
Switzerland

Outline

- Super-quick introduction
- Unifying cat models under the common hood
- Exploiting monotonicity property

Adaptive Testing

- Selection of questions' subsets
- Shorter test versions
- Individual sets of questions
- Our research is based on non-adaptive paper test

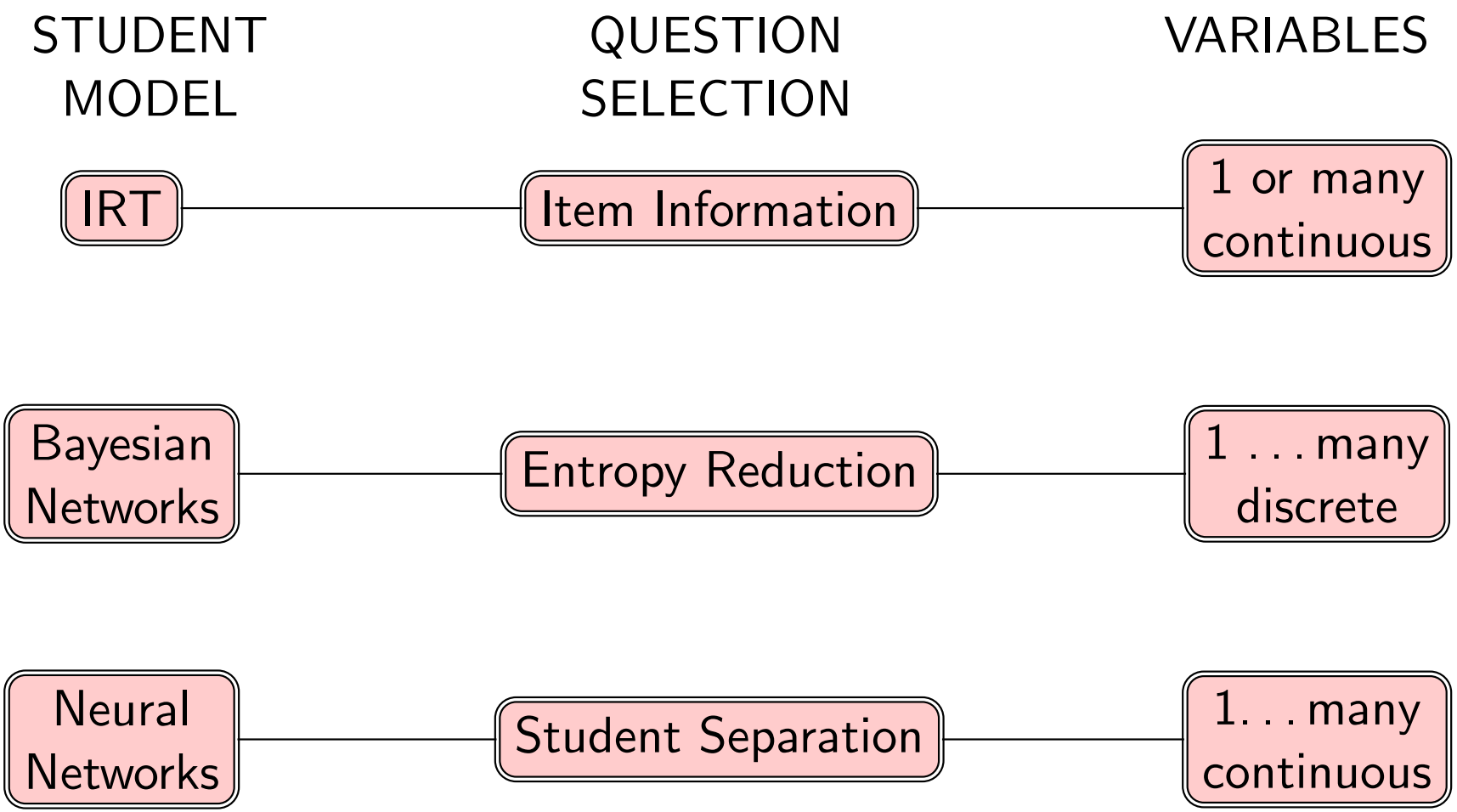
Adaptive Testing Procedure

- Select the next question
- Ask the question
- Update the model
- (optional) Estimate subsequent answers

Key Components

- Student model
- Question selection method
- Model scoring method

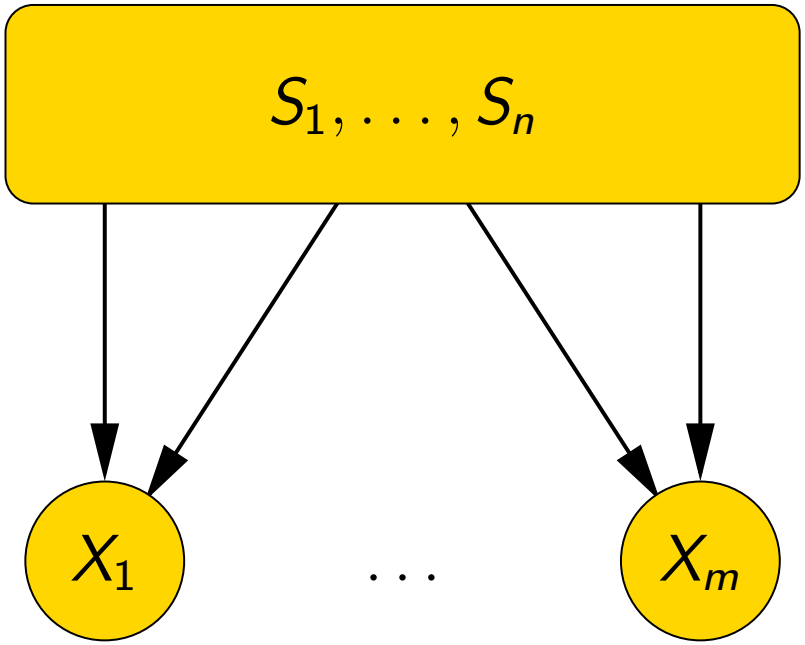
Adaptive Testing Models



Is it truly separated?

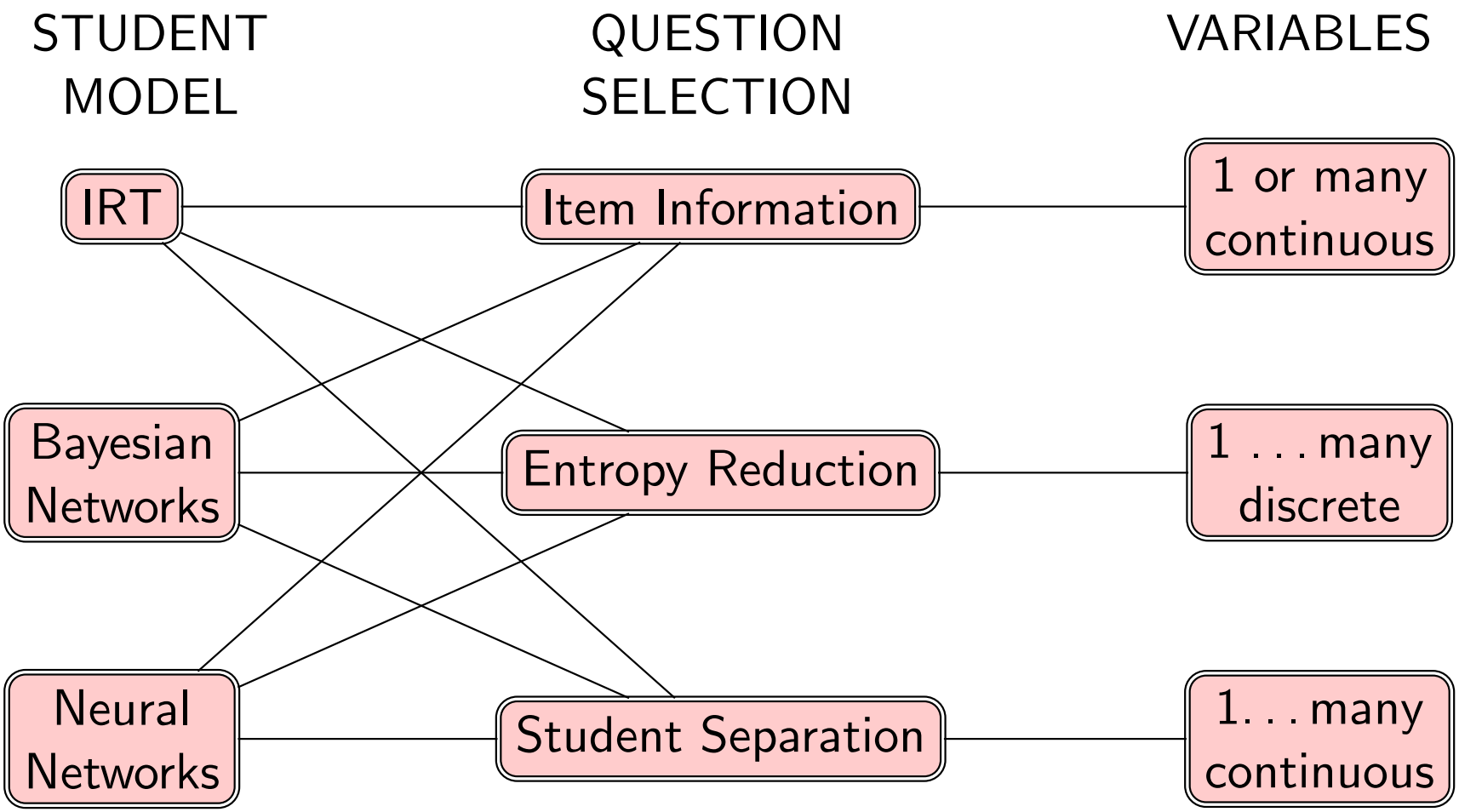
Unifying Generic Model

S_i – student skills, X_i – evidence about the student (answers)



$P(X_i = x_i | \boldsymbol{S}_{pa(i)})$ – specific model dependent
Associated question selection method

Adaptive Testing Models – Unified



Framework code

- The working version of source code (R) for the framework is available at the author's web page
- Including data sample
- We are working towards the full R package with examples

Model Monotonicity

With an ordering of states $s \preceq s'$ of a student skill S
while asking the question X resulting in the answer x we want

$$\begin{aligned} s \preceq s' &\rightarrow P(X = x|S = s) \leq P(X = x|S = s') \text{ , or} \\ s \preceq s' &\rightarrow P(X = x|S = s) \geq P(X = x|S = s') \end{aligned}$$

to hold.

Monotonicity Property in Individual Models

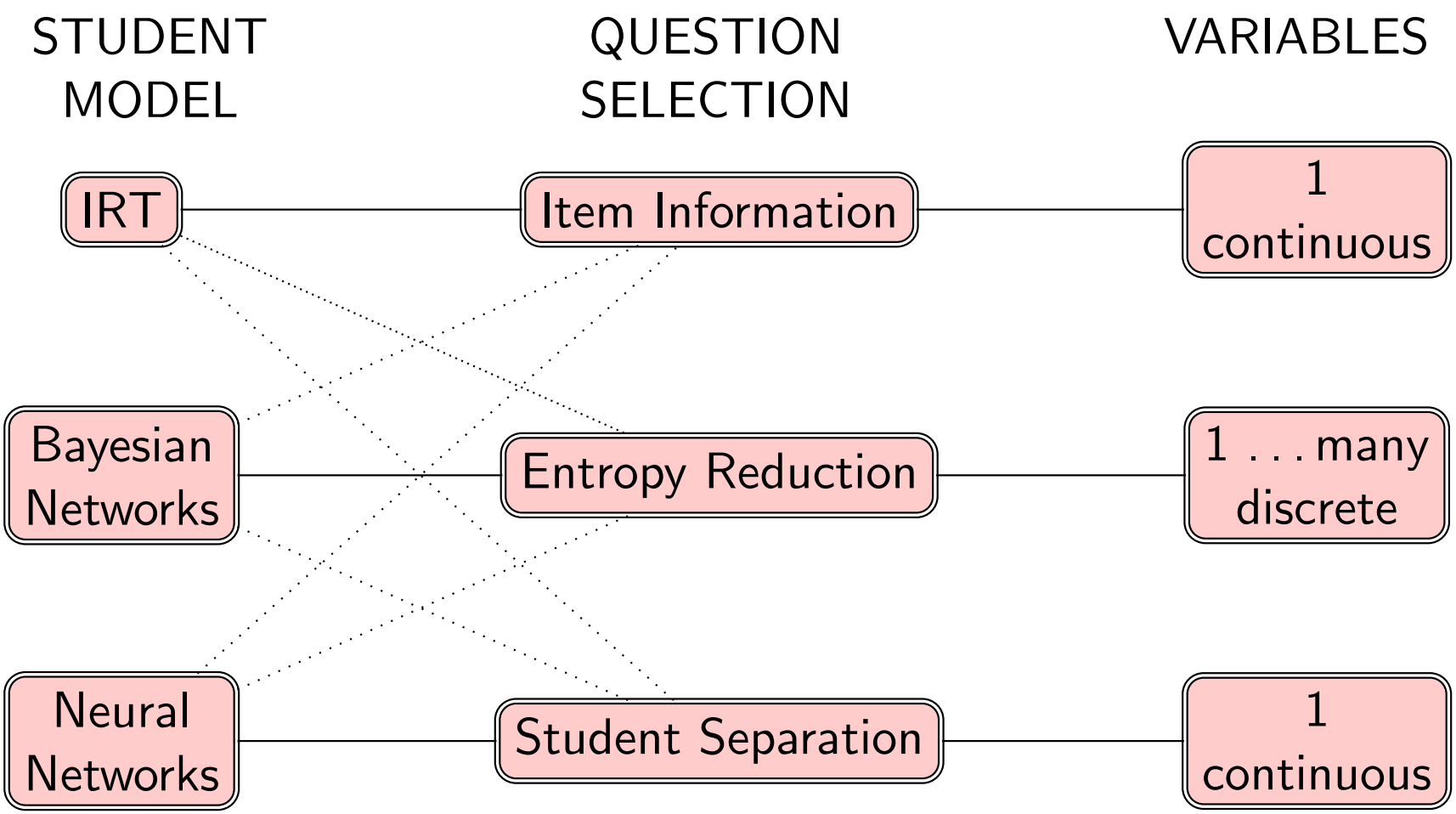
- IRT – yes (in a natural way)
- Bayesian Network – possible to ensure
- Neural Network – hard to tell (experts ?)

Monotonicity in BNs

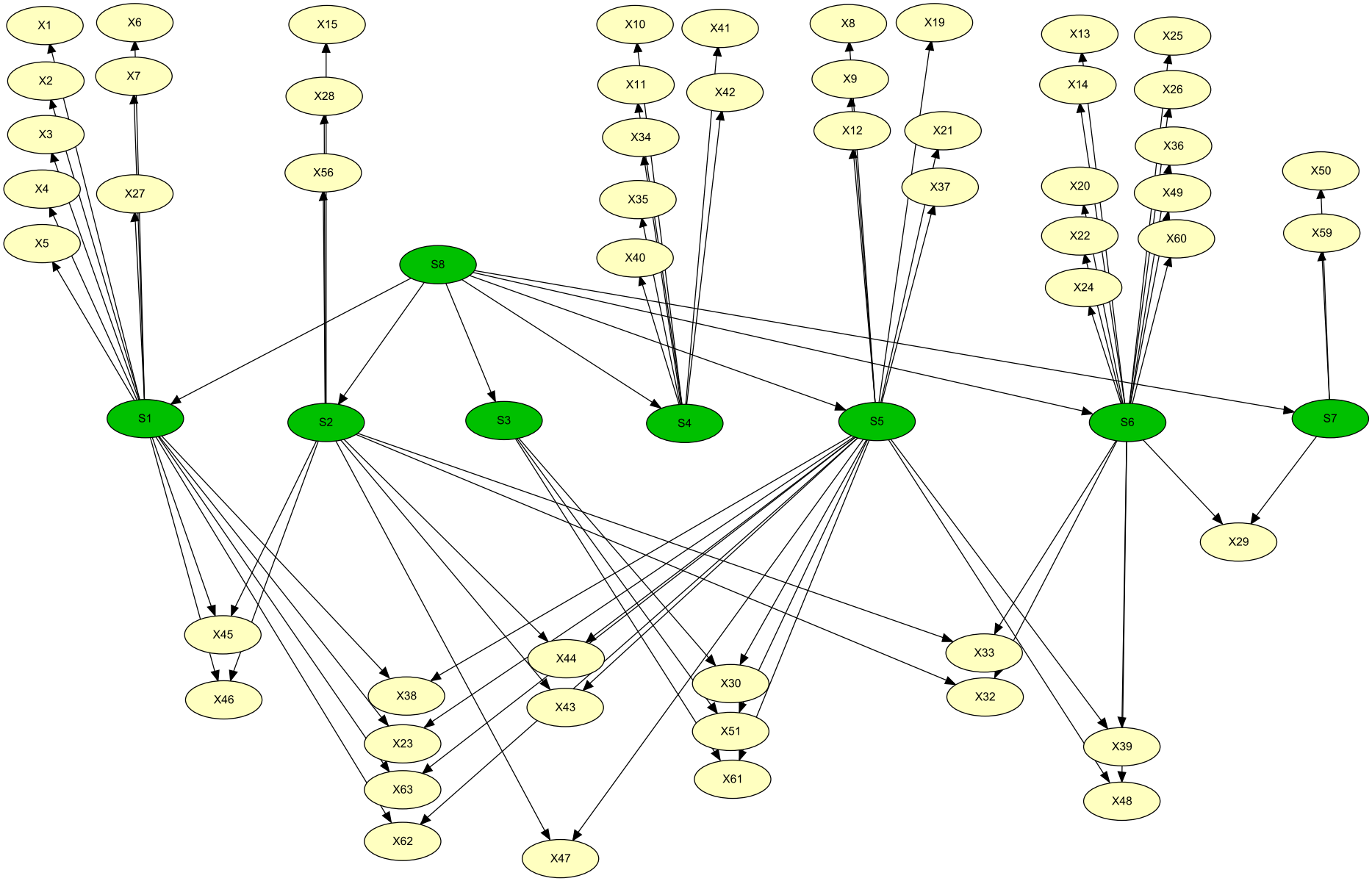
- The CPT of BN ($P(X_i = x_i|\boldsymbol{S}_{pa(i)})$) has to be in a special form.
- E.g, created from binomial family model (glm model with the logit link function) with parameters $\boldsymbol{\alpha_i}, \boldsymbol{\beta_i}$
-

$$P(X_i = 1|\mathcal{S}_{pa(i)} = s_{pa(i)}) \quad = \quad \frac{\exp(\boldsymbol{\alpha}_i + \boldsymbol{\beta}_i^T s_{pa(i)})}{1 + \exp(\boldsymbol{\alpha}_i + \boldsymbol{\beta}_i^T s_{pa(i)})} \quad .$$

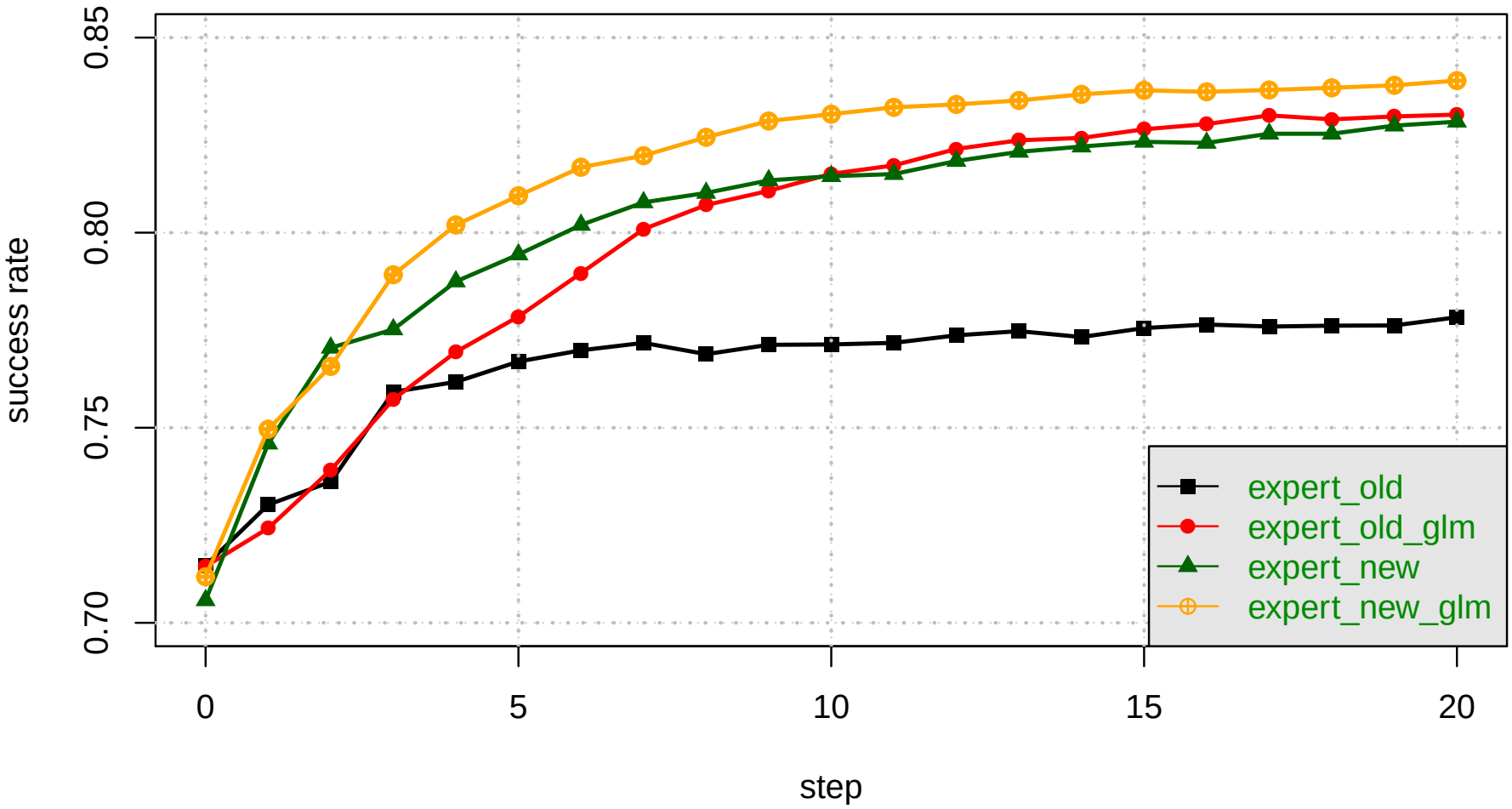
Models For Experiments



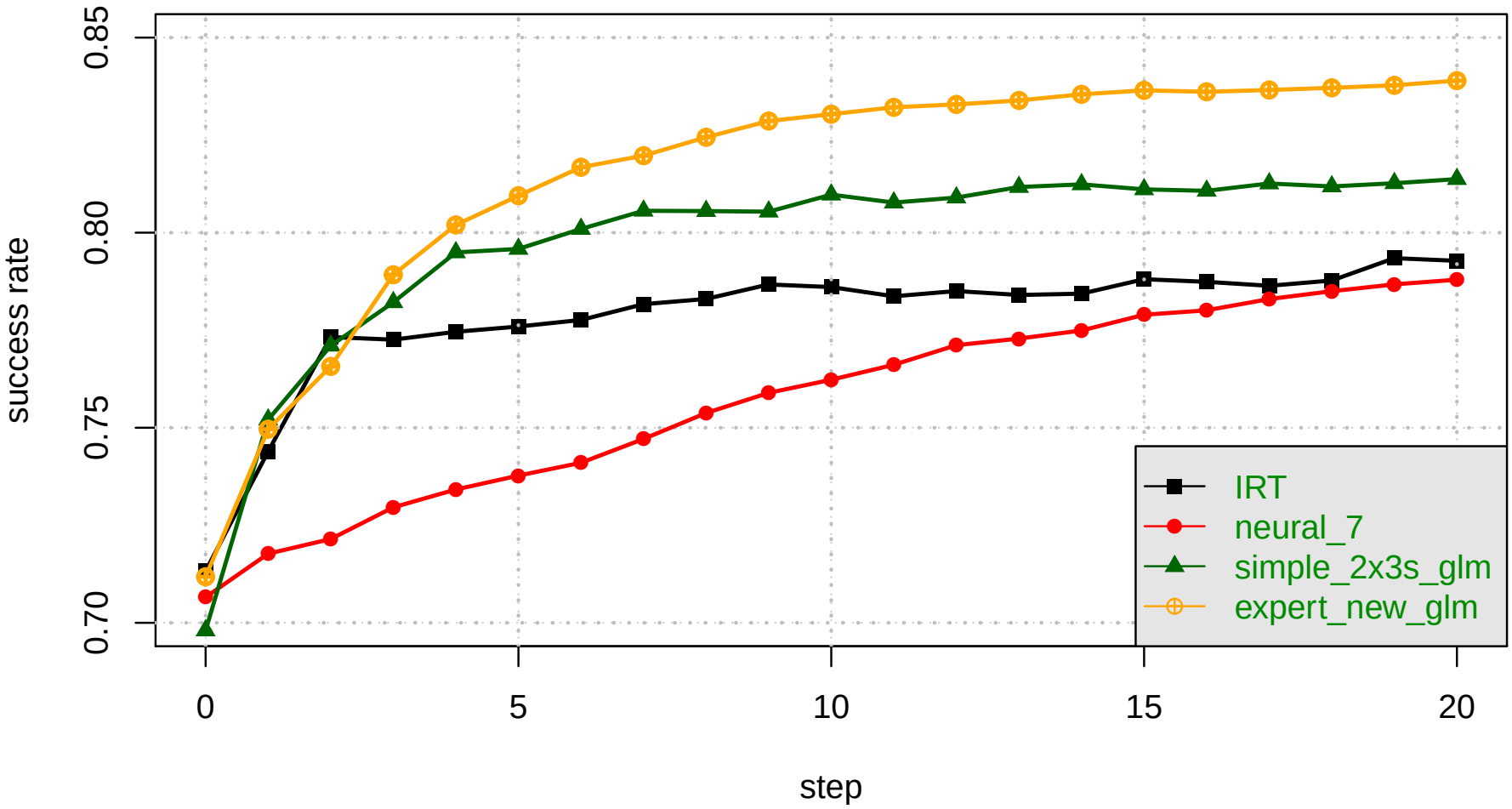
Expert network model



Results


success rate
0.70
0.75
0.80
0.85
0
5
10
15
20
step
expert_old
expert_old_glm
expert_new
expert_new_glm

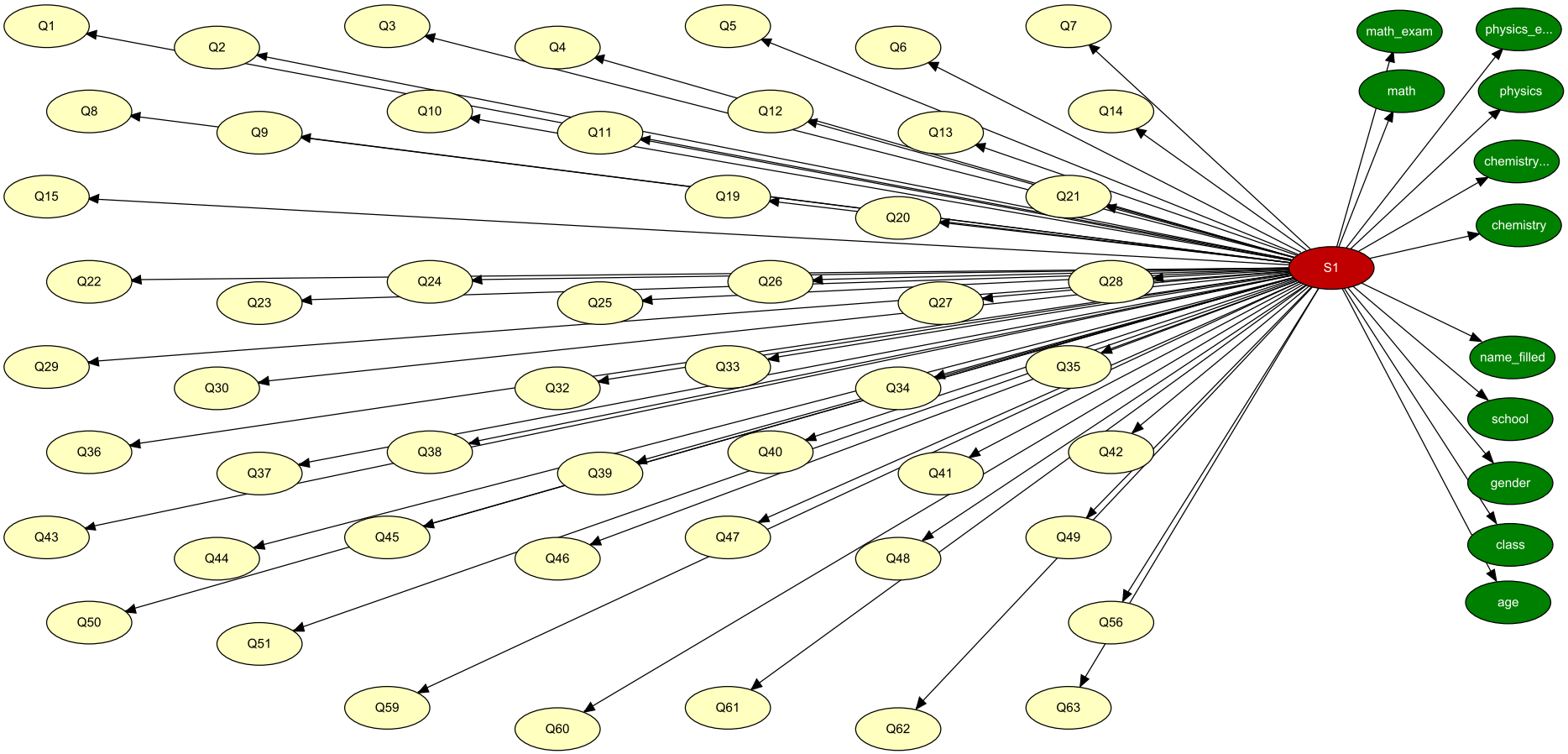
Results


success rate
0.70
0.75
0.80
0.85
0
5
10
15
20
step
IRT
neural_7
simple_2x3s_glm
expert_new_glm

Conclusions

- Establishment of an unifying generic CAT model
- Instantiation of this model by 3 specific models
- Empirical prove of the possibility of using these models
- Empirical verification of usefulness of the monotonicity property

Network model with one skill node and additional information



Question Selection Criteria

- The criterion C_j provides a value for question X_i based on the evidence e
- We select the next question by maximizing this criterion

$$X^*(e) = \arg\max_{X_i} C_j(X_i, e)$$

Item Information

For the continuous variables $\mathcal{S}$
and links to questions given by $p_i(X_i = 1|e)$:

$$C_1(X_i, e) = I(X_i, e) = \frac{(p_i'(X_i = 1|e))^2}{p_i(X_i = 1|e)(1 - p_i(X_i = 1|e))},$$

where p_i' is the derivation of p_i

Entropy Reduction

Cumulative Shannon entropy

$$H(e) = \sum_{k=1}^{n} \sum_{\ell=1}^{i_n} -P(S_k = s_{k,\ell}|e) \cdot \log P(S_k = s_{k,\ell}|e).$$

Expected entropy

$$EH(X_i, e) \quad = \quad \sum_{j=1}^{p} P(X_i = x_j|e) \cdot H(e_{i,j}) \quad .$$

Criterion

$$C_2(X_i, e) = IG(X_i, e) \quad = \quad H(e) - EH(X_i, e)$$

Question Selection Criteria

With $s_j|e_{i,j}$ as the predicted value of skill S_1 given extended evidence $e_{i,j} = e \cup \{X_i = x_j\}$
and $(\bar{s}|e_{i,j})$ its mean value,
the variance of S_1 given evidence $e_{i,j}$:

$$C_3(X_i, e) \quad = \quad \sum_{j=1}^{p} ((\bar{s}|e_{i,j}) - (s_j|e_{i,j}))^2 \cdot P(X_i = x_j|e) \quad .$$

Model evaluation

- $10 \cdot 10$ cross-validation
- For every (281) test repeat:
  - Select a question
  - Ask the selected question
  - Inserted answer as evidence into the BN
  - Distribute and collect evidence
  - Estimate subsequent answers

Model evaluation

- In every step compute $P(X_i|e)$ for all unanswered questions
- Extract $x_i^* = \arg\max_{x_l} P(X_i = x_l|e)$
- Compare it to the measured value x_i'

Success rate

$$\mathrm{SR}_s^t = \frac{\sum_{X_i \in \mathcal{X}_{s+1}} I(x_i^* = x_i')}{|\mathcal{X}_{s+1}|}, \text{ where}$$

$$I(expr) = \begin{cases} 1 & \text{if } expr \text{ is true} \\ 0 & \text{otherwise} \end{cases}$$

Data collection

- Paper test of mathematical knowledge (domain of functions)
- 4 grammar schools
- 281 test subjects (students)

Test description

- 29 questions (mathematical problems) - 53 sub questions

Example question

Decide which of the following functions

$$\begin{aligned} f(x) &= x^2 - 2x - 8 \\ g(x) &= -x^2 + 2x + 8 \end{aligned}$$

is decreasing in the interval $(-\infty, -1]$.

- Rated with correct / incorrect and points 0 - 4
- Total maximum of 120 points

Test results, correlations

Table: Average test scores of the four grammar schools.

GS1	GS2	GS3	GS4	Total
42.76	46.68	46.35	43.65	44.53

Table: Correlation of the grades and the test total score.

Mathematics	Physics	Chemistry
-0.60	-0.42	-0.41