

Model Selection For Adaptive Testing

Martin Plajner, Jirka Vomlel

The Institute of Information Theory and Automation
Czech Academy of Sciences

16.7.2015

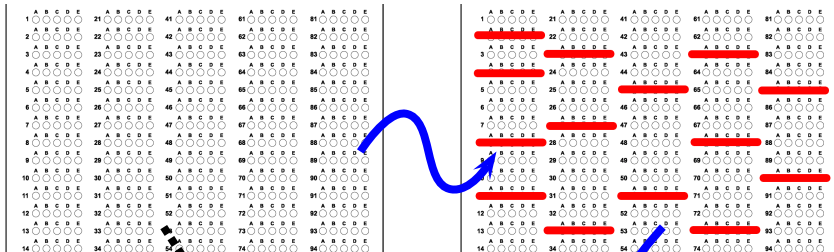
Bayesian Applications Workshop

UAI

Amsterdam

Adaptive testing

- Selecting subset of questions
- Shorter test versions
- Individual sets of questions



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

- 29 questions (mathematical problems) - 53 sub questions

Example question

Decide which of the following functions

$$f(x) = x^2 - 2x - 8$$

$$g(x) = -x^2 + 2x + 8$$

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

- Rated correct(0) / incorrect(1)
or graded 0 - 4
- Total maximum of 120 points

Test results, correlations

Table: Average test scores of the four grammar schools.

GS1	GS2	GS3	GS4	
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

Each model consist of

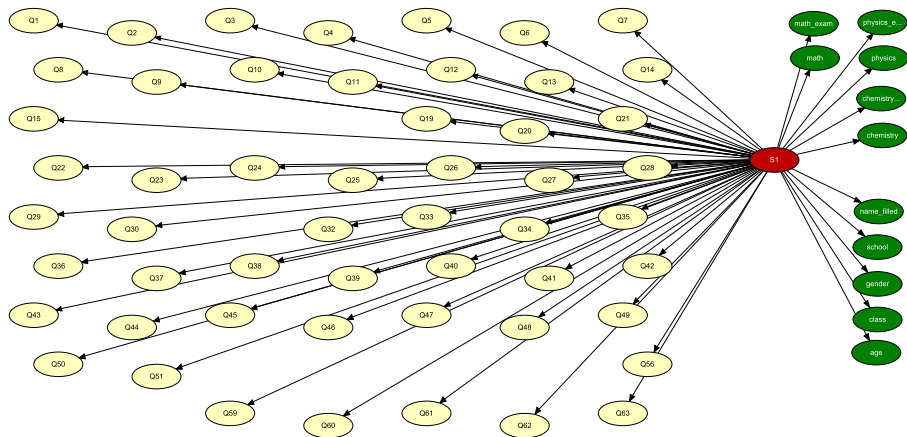
- A set of n variables $\{S_1, \dots, S_n\}$ - skills.
 - S will denote the multi variable (S_1, \dots, S_n) taking states $s = (s_1, \dots, s_n)$.
- A set of m variables $\{X_1, \dots, X_m\}$ - questions (mathematical problems)
 - X will denote the multi variable (X_1, \dots, X_m) taking states $x = (x_1, \dots, x_m)$.
- A set of arcs between variables

Options summary

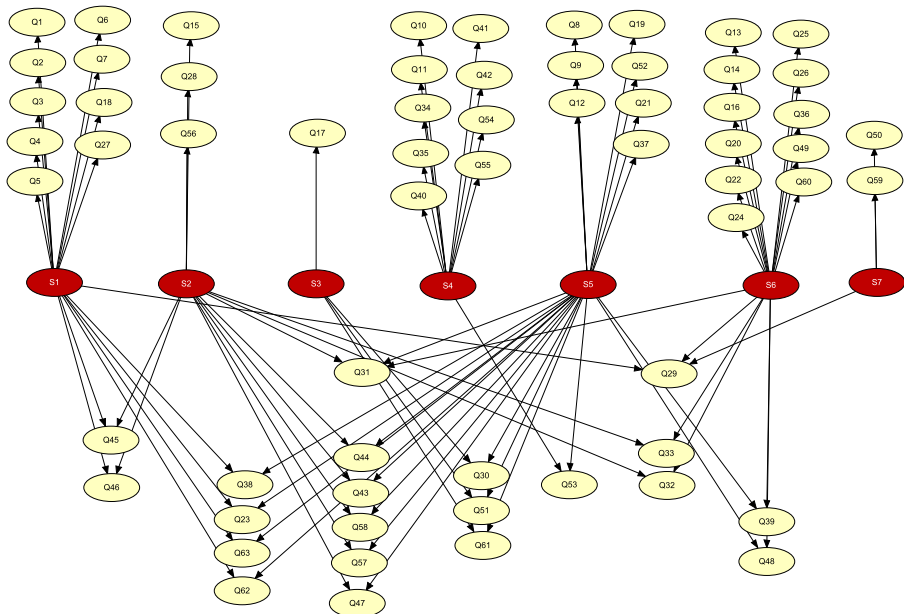
Question grading scales	Boolean	numeric
Number of skill nodes	1	7
Skill nodes states	2	3
Additional information	NO	YES

- Bayesian network models
 - 7 models for Boolean scale
 - 7 models for numeric scale

Network model with one skill node and additional information



Expert network model



- Select a next question
- Ask the question
- Update the network

Next question selection

- Selection based on information gain
i.e. reduction of expected entropy
- Cumulative Shannon entropy

$$H(e) = \sum_{i=1}^n \sum_{s_i} -P(S_i = s_i|e) \cdot \log P(S_i = s_i|e)$$

Expected entropy for the selection of X'

$$EH(X', e) = \sum_{j=1}^p P(X' = x'_j | e) \cdot H(e \cup \{X' = x'_j\})$$

Question X^* minimizing the expected entropy

$$X^* = \arg \min_{X' \in \mathcal{X}_s} EH(X', e)$$

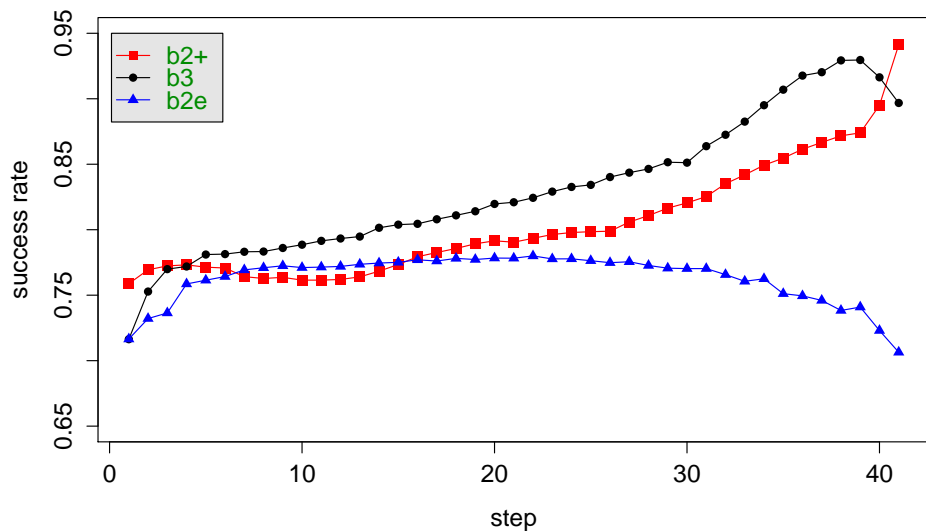
- 10 fold cross-validation
- For every (281) test repeat:
 - Select a question
 - Ask the selected question
 - Inserted answer as evidence into the BN
 - Distribute and collect evidence
 - Estimate subsequent answers

- We measure the success rate of a test
 - ⇒ By comparing the estimated value to the observed value

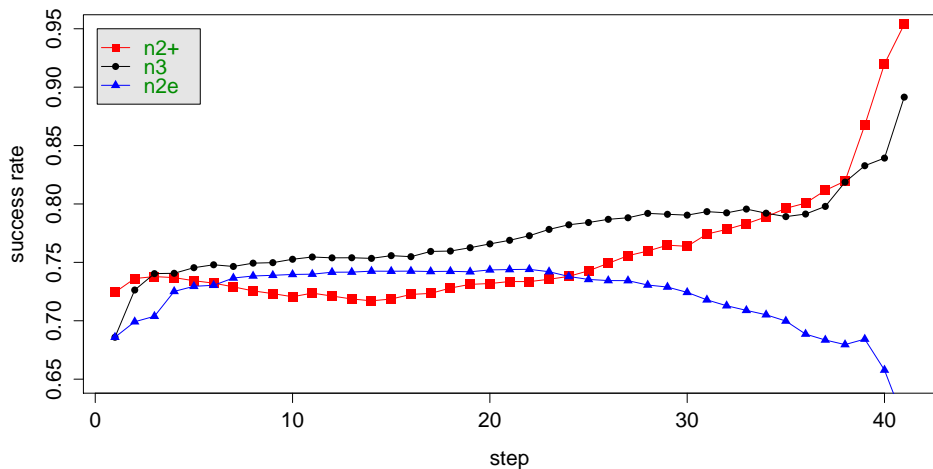
Success rate

$$SR_s^t = \frac{\sum_{x_i \in \mathcal{X}_{s+1}} I(x_i^* = x_i')}{|\mathcal{X}_{s+1}|}, \text{ where}$$
$$I(expr) = \begin{cases} 1 & \text{if } expr \text{ is true} \\ 0 & \text{otherwise} \end{cases}$$

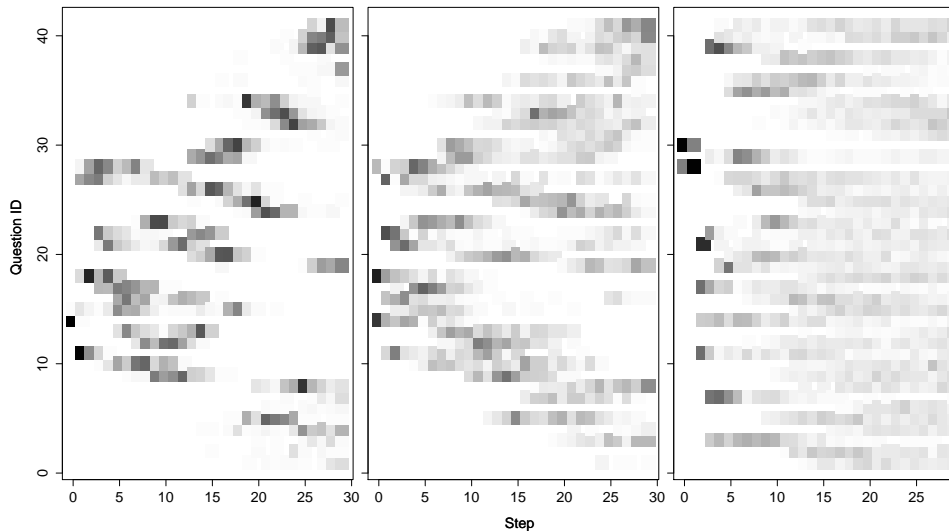
Boolean graded question answers



Numeric graded question answers



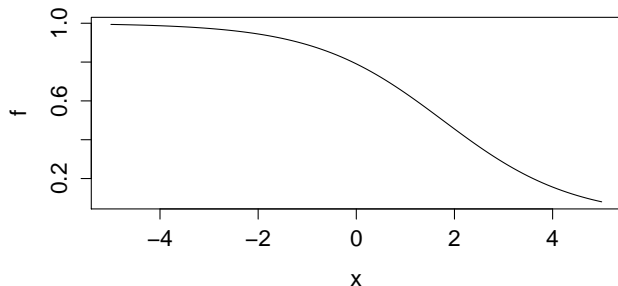
Question selection order



- Proposed models provide reasonable results
- 3-stated skill variables provide better results
- Expert model does not score as good as expected
- Additional information is not as important

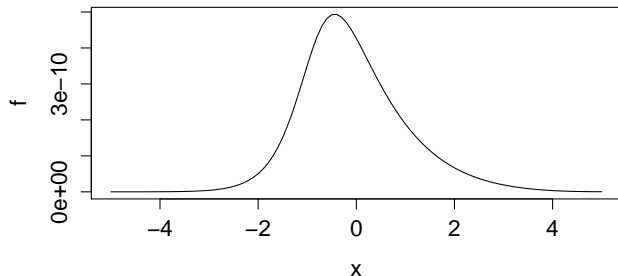
Future research

- Explore more options of parameters selection
- Comparison with other model types
 - standard Rash, IRT, other methods (neural network)
- Practical issues (fairness, constraints, . . .)



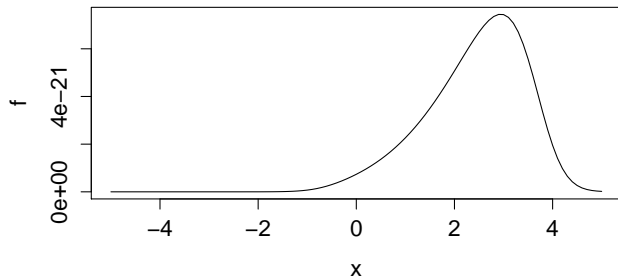
Future research

- Explore more options of parameters selection
- Comparison with other model types
 - standard Rash, IRT, other methods (neural network)
- Practical issues (fairness, constraints, . . .)



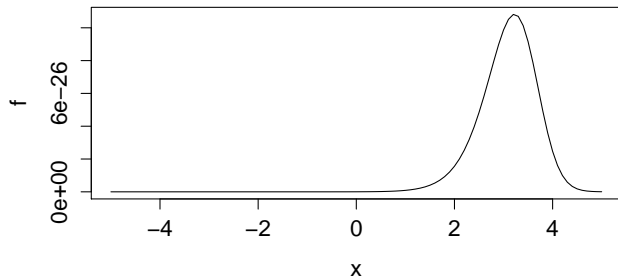
Future research

- Explore more options of parameters selection
- Comparison with other model types
 - standard Rash, IRT, other methods (neural network)
- Practical issues (fairness, constraints, . . .)



Future research

- Explore more options of parameters selection
- Comparison with other model types
 - standard Rash, IRT, other methods (neural network)
- Practical issues (fairness, constraints, . . .)



Future research

- Explore more options of parameters selection
- Comparison with other model types
 - standard Rash, IRT, other methods (neural network)
- Practical issues (fairness, constraints, . . .)

