Reliable Knowledge-Based Adaptive Tests by Credal Networks

Francesca Mangili, Claudio Bonesana, and Alessandro Antonucci

Istituto "Dalle Molle" di Studi sull'Intelligenza Artificiale
Lugano (Switzerland)

http://ipg.idsia.ch/

ISIPTA 17 & ECSQARU 2017, Lugano

Introduction

- Intelligent tutoring systems (computerized adaptive testing - CAT)
- Probabilistic graphical model
- Expert knowledge elicitation

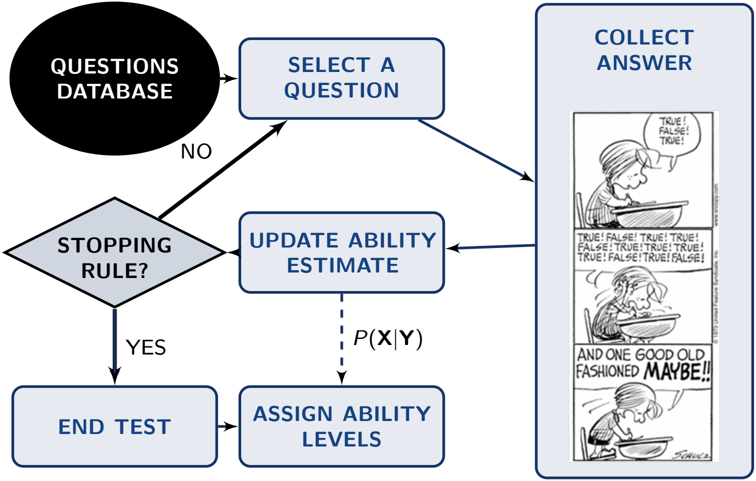
Adaptive Testing

Adaptive test? Test selecting the sequence of questions on the basis of the test taker ability.

Expected benefits

- reduced assessment time;
- increased accuracy;
- challenged/not discouraged test takers;
- improved security.

CAT procedure



CAT requirements

Elements necessary to develop a CAT:

- Test taker knowledge model (skills and answers model)
- Question selection rule
- Stopping rule

Test taker knowledge model

- Test taker skills $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ *Hidden*
- Answers to questions $\mathbf{Y} = (Y_1, \ldots, Y_m)$ *Observable*

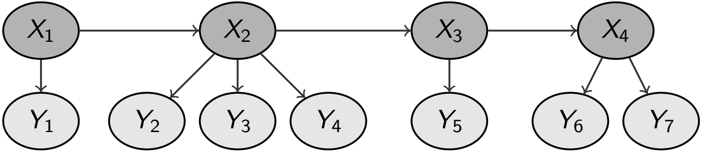


Figure : A directed graph for CAT.

Evaluation:

- Condition the model on the observed answers $\mathbf{y}$.
- Compute the marginal probabilities for each skill $P(X_i|\mathbf{y})$
- Assign each skill its most probable value.

Adaptive approach

- *Goal*: gather information about the student level $\mathbf{X}$
- *Selection criteria:* information gain

$$IG(Y_{next}) = H(\mathbf{X}|\mathbf{y}) - H(\mathbf{X}|\mathbf{y}, Y_{next} = y)$$

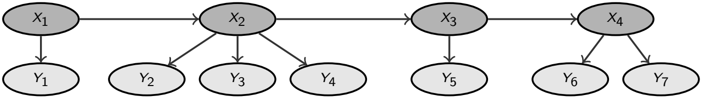
y is not observed yet $\rightarrow$ we work with the conditional entropy
$$H(\mathbf{X}|\mathbf{y}, Y_{next}) := \sum_{y \in \{0,1\}} H(\mathbf{X}|\mathbf{y}, Y_{next} = y) P(Y_{next} = y|\mathbf{y})$$

- *Question selection rule:* the question giving the maximum expected information gain is selected at each step

$$\tilde{Y}_{next} := \arg \max_{Y_{next} \in \mathbf{Y}} [H(\mathbf{X}|\mathbf{y}) - H(\mathbf{X}|\mathbf{y}, Y_{next})]$$

- *Stopping rule:* the test ends when entropy falls below a predefined threshold

Elicitation



- Experts should elicit:
  - $P(X_{i+1}|X_i), \ \forall i = 1, 2, \ldots$
  - $P(Y_j|\mathrm{X}_{\mathrm{Y_j}}), \ \forall j = 1, 2, \ldots.$
- Interval-valued probabilistic elicitation
  - can simplify the elicitation of the model

Judgement	*impossible*	*improbable*	*uncertain*	*fifty-fifty*	*expected*	*probable*
$P(Y_i\|X_{Y_i})$	17.5-20%	22.5-25%	30-35%	60-65%	75-80%	95-97.5%

  - higher realism in the modeling of the expert knowledge, which is typically qualitative.

Credal networks

Replacing conditional PMF in the BN above with sets of probabilities we obtain credal networks.

Main issues:

- numerical inferences will be interval-valued, thus making debatable both the evaluation criterion an the information measure to adopt
- inference tasks in CNs typically belongs to higher complexity classes than their Bayesian counterparts ((updating is NP^{PP}-hard).

Evaluation method

- Posterior probabilities of each skill are set-valued.
  They can be characterized by lower and upper bounds, say $\underline{P}(X_i|\mathbf{y})$ and $\overline{P}(X_i|\mathbf{y})$ for each $X_i \in \mathbf{X}$.
- *Maximality* criterion is used to assign levels:
  - levels which are less probable than another level for all the elements of the credal set are rejected;
  - less conservative than interval dominance;
  - can be easily reduced to standard CNs updating by auxiliary nodes;
  - multiple non-rejected levels induce a situation of indecision.

Question selection criteria

- To use the information gain as decision criteria, computation of entropies should be extended to credal sets.
- Cautious approach: take upper entropy $\overline{H}(\mathbf{X})$.
- We need maximum values of conditional entropies to compute
  - the joint entropy $H(\mathbf{X})$ and its posterior value;
  - the conditional entropies involved in question selection, i.e., $H(\mathbf{X}|\mathbf{y}, Y_{next})$
- Problem: computation of upper conditional entropies require the solution of a non-linear non-convex optimization.

Question selection based on upper entropy

We introduce a number of approximations:

- we separately consider the entropy of each skill
- we approximate $\overline{H}(X_i|Y',\mathbf{y})$ by taking the maximum over $P(1|\mathbf{y})$ of

$$\overline{H}(X_i|\mathbf{y},1)P(1|\mathbf{y})+\overline{H}(X_i|\mathbf{y},0)[1-P(1|\mathbf{y})]$$

- we approximate $\overline{H}(X_i|\mathbf{y},y')$ by taking the maximum of $H(X_i|\mathbf{y})$ over the credal set induced by the probability intervals $[\underline{P}(x_i|\mathbf{y}),\overline{P}(x_i|\mathbf{y})]$ This optimization is carried out by [Abellan and Moral, 2003].
- we select the question $\tilde{Y}'$ leading to the maximum information gain:

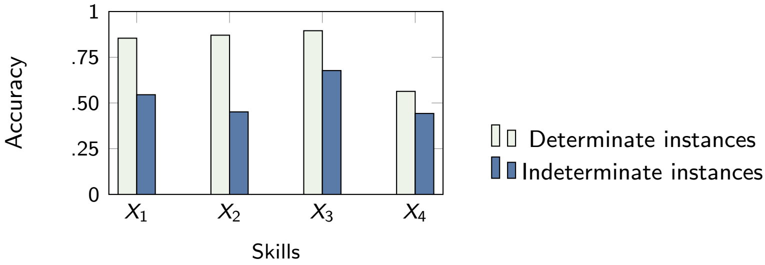
$$\tilde{Y}_{next} := \arg\max_{Y_{next}\in\mathbf{Y}}\left[\overline{\overline{H}}(X_{Y_{next}}|\mathbf{y})-\overline{\overline{H}}(X_{Y_{next}}|\mathbf{y},Y_{next})\right] \tag{1}$$

Application

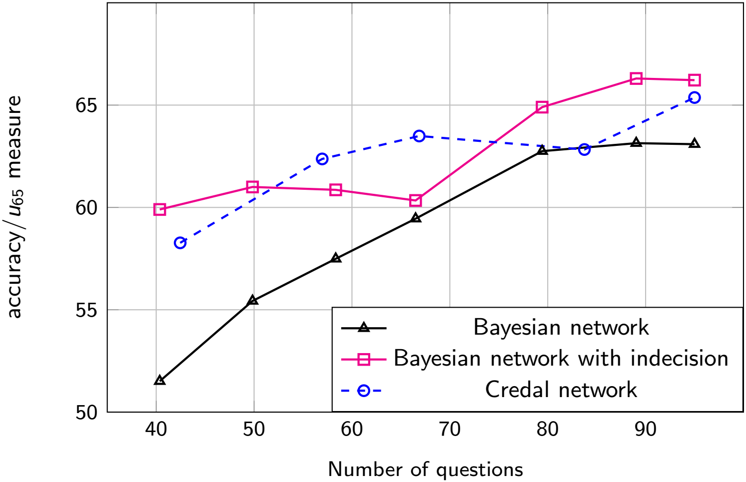
- Dataset: answers from 451 students to 95 question about German language.
- Each skill is assumed to have four ability levels: $X_i \in \{A1, A2, B1, B2\}$.
- Each question is influenced by a single background skill identified by the teachers.
- Three approaches have been simulated.
  - **Credal network model**
  - **Bayesian network model** with PMF obtained as centers of mass of the credal sets
  - **Bayesian network model with indecision**

Non adaptive test results

Algorithm	Average	X_1	X_2	X_3	X_4
BN (acc)	63.09%	67.56%	60.85%	75.84%	48.10%
CN (u_{65})	65.37%	67.71%	66.67%	70.33%	56.76%



Adaptive test results



Conclusion

- A procedure for adaptive testing built solely on expert knowledge has been proposed based on credal networks.
- Results are promising as the credal approach
  - simplifies the model elicitation
  - recognizes when a sharp decision about the student level should not be made
  - achieves an accuracy comparable to that of an indecisive Bayesian approach maximizing the expected u65 measure.
- However, the indecision of the CN test remains rather large.