Efficient Propagation for Computerized Adaptive Testing

Jiří Vomlel
jirka@cs.auc.dk

What is CAT?

What is CAT?	Computerized Adaptive Testing
Objective?	An optimal test for each examinee
How?	(1) examinee's knowledge level is estimated
	(2) questions appropriate for the level are selected.

Almond and Mislevy (1999) proposed to use graphical models.

- one student model (relations between skills, abilities, etc.)
- several evidence models, one for each task or question.

Goal: keep the actually used model small,

evidence models are connected to the student model only if there is an evidence in the model.

CAT for basic operations with fractions

$$
\begin{array}{llcl}
T_1: & \left(\frac{3}{4} \cdot \frac{5}{6}\right) - \frac{1}{8} & = & \frac{15}{24} - \frac{1}{8} = \frac{5}{8} - \frac{1}{8} = \frac{4}{8} = \frac{1}{2} \\
T_2: & \frac{1}{6} + \frac{1}{12} & = & \frac{2}{12} + \frac{1}{12} = \frac{3}{12} = \frac{1}{4} \\
T_3: & \frac{1}{4} \cdot 1\frac{1}{2} & = & \frac{1}{4} \cdot \frac{3}{2} = \frac{3}{8} \\
T_4: & \left(\frac{1}{2} \cdot \frac{1}{2}\right) \cdot \left(\frac{1}{3} + \frac{1}{3}\right) & = & \frac{1}{4} \cdot \frac{2}{3} = \frac{2}{12} = \frac{1}{6} \ .
\end{array}
$$

Elementary and operational skills

Label	Description	Example
CP	Comparison (common numerator or denominator)	$\frac{1}{2} > \frac{1}{3}$, $\frac{2}{3} > \frac{1}{3}$
AD	Addition (comm. denom.)	$\frac{1}{7} + \frac{2}{7} = \frac{1+2}{7} = \frac{3}{7}$
SB	Subtract. (comm. denom.)	$\frac{2}{5} - \frac{1}{5} = \frac{2-1}{5} = \frac{1}{5}$
MT	Multiplication	$\frac{1}{2} \cdot \frac{3}{5} = \frac{3}{10}$
CD	Common denominator	$\left(\frac{1}{2}, \frac{2}{3}\right) = \left(\frac{3}{6}, \frac{4}{6}\right)$
CL	Cancelling out	$\frac{4}{6} = \frac{2 \cdot 2}{2 \cdot 3} = \frac{2}{3}$
CIM	Conv. to mixed numbers	$\frac{7}{2} = \frac{3 \cdot 2 + 1}{2} = 3\frac{1}{2}$
CMI	Conv. to improper fractions	$3\frac{1}{2} = \frac{3 \cdot 2 + 1}{2} = \frac{7}{2}$

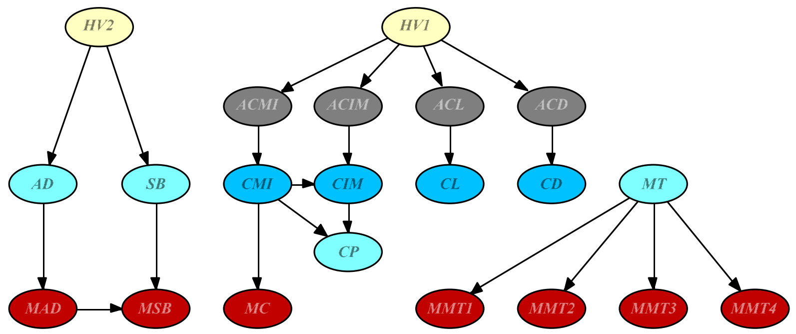
MISCONCEPTIONS

Label	Description	Occurrence
MAD	$\frac{a}{b} + \frac{c}{d} = \frac{a+c}{b+d}$	14.8%
MSB	$\frac{a}{b} - \frac{c}{d} = \frac{a-c}{b-d}$	9.4%
MMT1	$\frac{a}{b} \cdot \frac{c}{b} = \frac{a \cdot c}{b}$	14.1%
MMT2	$\frac{a}{b} \cdot \frac{c}{b} = \frac{a+c}{b \cdot b}$	8.1%
MMT3	$\frac{a}{b} \cdot \frac{c}{d} = \frac{a \cdot d}{b \cdot c}$	15.4%
MMT4	$\frac{a}{b} \cdot \frac{c}{d} = \frac{a \cdot c}{b+d}$	8.1%
MC	$a\frac{b}{c} = \frac{a \cdot b}{c}$	4.0%

CREATION OF A STUDENT MODEL

- what skills will be tested (KDE group),
- preparation of paper tests (KDE group),
- paper tests given to students at Brønderslev high school (Kirsten Bangsø Jensen), 149 students did the test.
- analysis of results, finding misconceptions, summarizing results into a data file (KDE group),
- learning model (PC-algorithm for structure, EM-algorithm for parameters),
- trying to explain some relations between skills and misconceptions using hidden variables,
- a new model learning phase with hidden variables included, certain edges required to be part of the learned model.

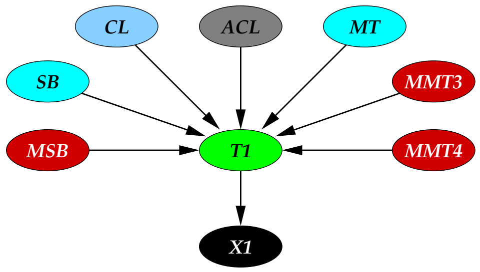
STUDENT MODEL



EVIDENCE MODEL FOR TASK T_1

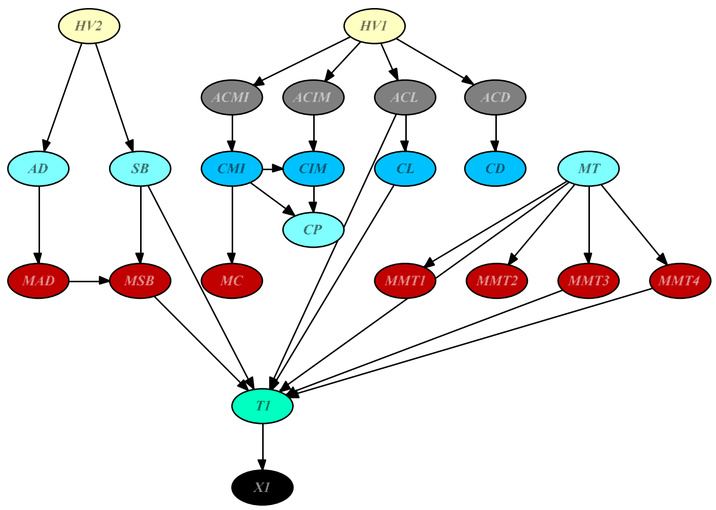
$$\left(\frac{3}{4} \cdot \frac{5}{6}\right) - \frac{1}{8} = \frac{15}{24} - \frac{1}{8} = \frac{5}{8} - \frac{1}{8} = \frac{4}{8} = \frac{1}{2}$$

$$T_1 \quad \Leftrightarrow \quad MT \;\&\; CL \;\&\; ACL \;\&\; SB \;\&\; \neg MMT3 \;\&\; \neg MMT4 \;\&\; \neg MSB$$

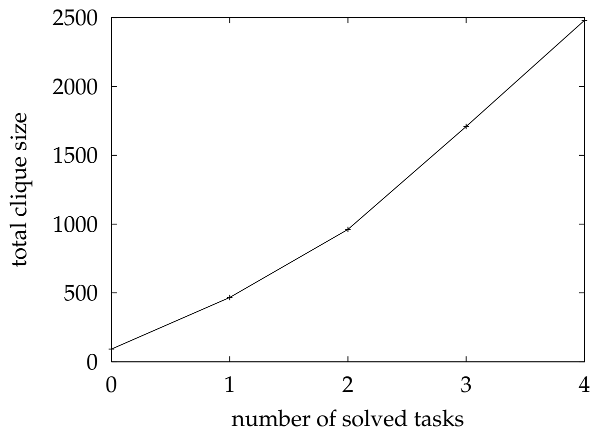


guessing	$P(X_1 \mid \neg T_1)$
mistakes	$P(\neg X_1 \mid T_1)$

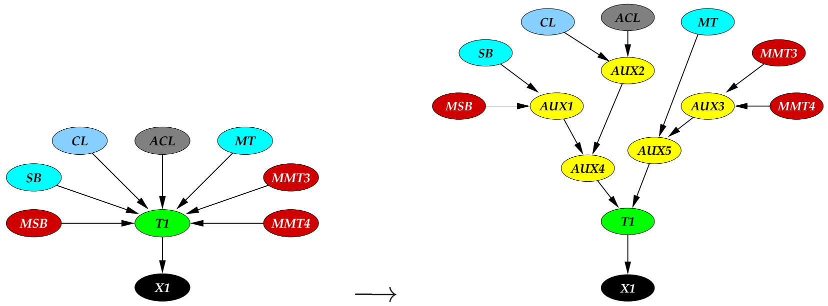
Student model + evidence model of T_1



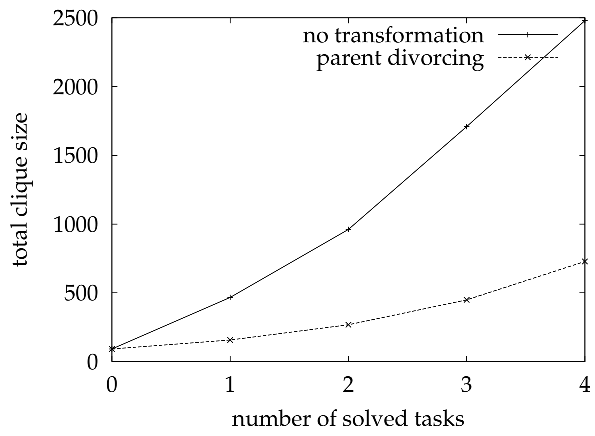
Total clique size

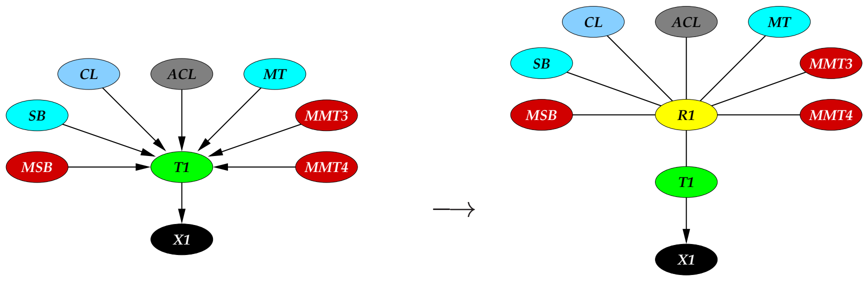


Hierarchical evidence model (=parent divorcing)



Total clique size





$\psi(X_1, \ldots, X_n)$

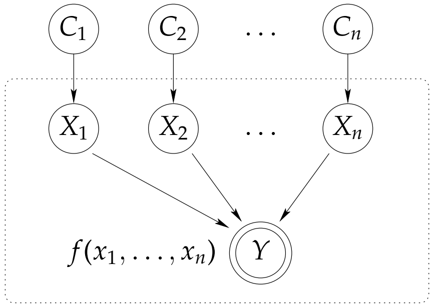
$$\sum_R \prod_{i=1}^{n} \varphi_i(X_i, R)$$

Lemma 1 *Belief updating performed in a probabilistic model with a potential replaced by a factorized model and in the original probabilistic model provide equivalent marginals.*

Y is functionally dependent on $X_1, \ldots, X_n$ if

$$\psi(y, x_1, \ldots, x_n) \quad = \quad \begin{cases} 1 & \text{if } y = f(x_1, \ldots, x_n) \\ 0 & \text{otherwise.} \end{cases}$$

Example - a model with independence of causal influence



Let $X_i \ = \ \{0, +1, +2\}, \ \ i = 1, 2 \quad y = f(x_1, x_2) \ = \ \max\{x_1, x_2\}$.

We want to get a factorization

$$P(Y|X_1, X_2) = \sum_R h(Y, R) \cdot g_1(X_1, R) \cdot g_2(X_2, R) \ .$$

	0			+1			+2		
	0	+1	+2	0	+1	+2	0	+1	+2
0	1								
+1		1		1	1				
+2			1			1	1	1	1

=

$\sum_{r_1, r_2, r_3}($

	r_1	r_2	r_3
0	1		
+1	-1	1	
+2		-1	1

$\times$

	r_1	r_2	r_3
0	1	1	1
+1		1	1
+2			1

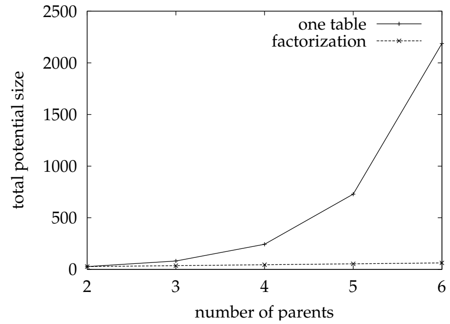
$\times$

	r_1	r_2	r_3
0	1	1	1
+1		1	1
+2			1

$)$

• more compact representation of a probability potential

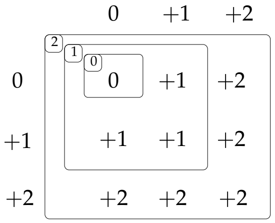
Noisy-MAX with n parents with three states: 3^{n+1} versus $(n+1) \cdot 3^2$.



Even more compact representation:
using probability trees instead of flat tables.

• smaller cliques in a junction tree

Factorization of MAX: different problem formulation — 17



Partition by values of Y:

$$\begin{aligned} \mathcal{Y}_0 &= \{ (0,0) \} \\ \mathcal{Y}_1 &= \{ (0,+1),(+1,0),(+1,+1) \} \\ \mathcal{Y}_2 &= \{ (0,+2),(+1,+2),(+2,0),(+2,+1)(+2,+2) \} \end{aligned}$$

Rectangular subspaces:

$$\begin{aligned} \mathcal{B}_0 &= \{ (0,0) \} \\ \mathcal{B}_1 &= \{ (0,0),(0,+1),(+1,0),(+1,+1) \} \\ \mathcal{B}_2 &= X_1 \times X_2 \end{aligned}$$

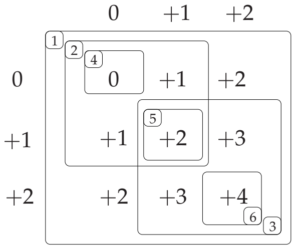
$$\mathcal{Y}_0 = \mathcal{B}_0, \qquad \mathcal{Y}_1 = \mathcal{B}_1 \setminus \mathcal{B}_0, \qquad \text{and } \mathcal{Y}_2 = \mathcal{B}_2 \setminus \mathcal{B}_1.$$

Proper difference and disjunctive union — 18

Let $\mathcal{A}, \mathcal{B} \subseteq \mathcal{X}$.

- If $\mathcal{A} \supseteq \mathcal{B}$ then proper difference of $\mathcal{A}$ and $\mathcal{B}$ is defined as $\mathcal{A} \ominus \mathcal{B} = \{\mathbf{x} \in \mathcal{A} \wedge \mathbf{x} \notin \mathcal{B}\}$.
- If $\mathcal{A} \cap \mathcal{B} = \emptyset$ then disjunctive union of $\mathcal{A}$ and $\mathcal{B}$ is defined as $\mathcal{A} \oplus \mathcal{B} = \{\mathbf{x} \in \mathcal{A} \vee \mathbf{x} \in \mathcal{B}\}$.

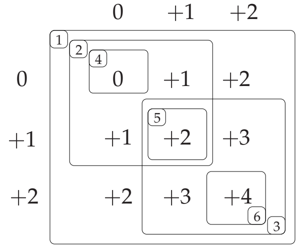
ADD: $Y = X_1 + X_2$



Minimal base of rectangles (MBR) — 19

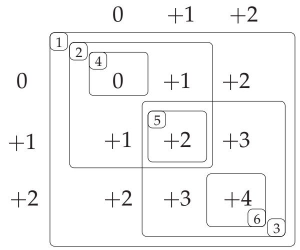
For a given partition $\mathcal{Y}_1, \ldots, \mathcal{Y}_q$ of $\mathcal{X} = \times_{i=1}^{n} \mathcal{X}_i$ find a set $\{\mathcal{B}_1, \mathcal{B}_2, \ldots, \mathcal{B}_k\}$ of minimal cardinality such that:

- for $j = 1, \ldots, k$: $\mathcal{B}_j = \times_{i=1}^{n} \mathcal{D}_i$, $\emptyset \neq \mathcal{D}_i \subseteq \mathcal{X}_i$, i.e. $\mathcal{B}_j$ is a rectangle,
- each element $\mathcal{Y}_\ell, \ell = 1, 2, \ldots q$ of the partition can be generated from base $\{\mathcal{B}_1, \mathcal{B}_2, \ldots, \mathcal{B}_k\}$ using operations $\ominus$ and $\oplus$.



$$\begin{aligned} \mathcal{Y}_1 &= \mathcal{B}_4 \\ \mathcal{Y}_2 &= \mathcal{B}_2 \ominus \mathcal{B}_4 \ominus \mathcal{B}_5 \\ \mathcal{Y}_3 &= \mathcal{B}_1 \ominus (\mathcal{B}_2 \ominus \mathcal{B}_5) \ominus (\mathcal{B}_3 \ominus \mathcal{B}_5) \\ \mathcal{Y}_4 &= \mathcal{B}_3 \ominus \mathcal{B}_5 \ominus \mathcal{B}_6 \\ \mathcal{Y}_5 &= \mathcal{B}_6 \end{aligned}$$

Correspondence of MBR and factorization — 20



$$\begin{aligned} \mathcal{Y}_1 &= \mathcal{B}_4 \\ \mathcal{Y}_2 &= \mathcal{B}_2 \ominus \mathcal{B}_4 \ominus \mathcal{B}_5 \\ \mathcal{Y}_3 &= \mathcal{B}_1 \ominus (\mathcal{B}_2 \ominus \mathcal{B}_5) \ominus (\mathcal{B}_3 \ominus \mathcal{B}_5) \\ \mathcal{Y}_4 &= \mathcal{B}_3 \ominus \mathcal{B}_5 \ominus \mathcal{B}_6 \\ \mathcal{Y}_5 &= \mathcal{B}_6 \ , \end{aligned}$$

$h(Y, R)$

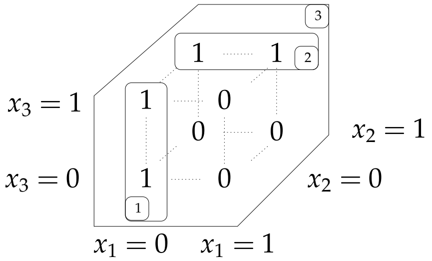
	$\mathcal{B}_1$	$\mathcal{B}_2$	$\mathcal{B}_3$	$\mathcal{B}_4$	$\mathcal{B}_5$	$\mathcal{B}_6$
y_1	0	0	0	+1	0	0
y_2	0	+1	0	−1	−1	0
y_3	+1	−1	−1	0	+2	0
y_4	0	0	+1	0	−1	−1
y_5	0	0	0	0	0	+1

$g_1(X_1, R)$, $g_2(X_2, R)$

	$\mathcal{B}_1$	$\mathcal{B}_2$	$\mathcal{B}_3$	$\mathcal{B}_4$	$\mathcal{B}_5$	$\mathcal{B}_6$
x_1	1	1	0	1	0	0
x_2	1	1	1	0	1	0
x_3	1	0	1	0	0	1

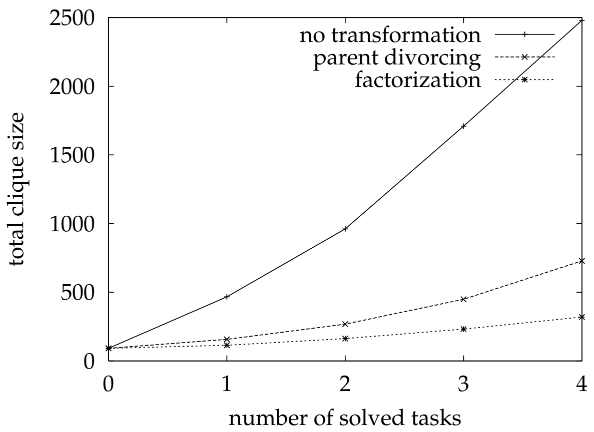
A Boolean function 21

$$
\begin{aligned}
Y &= (X_1 \vee X_2) \Rightarrow (X_2 \wedge X_3) \\
&= (\neg X_1 \wedge \neg X_2) \vee (X_2 \wedge X_3)
\end{aligned}
$$



$$
\begin{aligned}
\mathcal{Y}_0 &= \mathcal{B}_3 \ominus (\mathcal{B}_2 \oplus \mathcal{B}_1) \\
\mathcal{Y}_1 &= \mathcal{B}_2 \oplus \mathcal{B}_1
\end{aligned}
$$

Total clique size 22



Conclusions 23

Comparisons with Zhang and Poole (1996):

- no need for a special combination operator,
- no restriction on elimination ordering, and
- applicability to any model containing functional dependence (not only independence of causal influence).

Future work:

- solution method of the Minimal Base of Rectangles,
- exploring similiarity with problems in multiparty communication complexity, Kushilevitz and Nisan (1997),
- generalization to context specific independence (?),
- greedy vs. optimal approach to selecting questions (i.e. a generalization of troubleshooting).

References

Almond, R. G. and R. J. Mislevy (1999). Graphical models and computerized adaptive testing. *Applied Psychological Measurement 23*(3), 223–237.

Būtėnas, L., A. Brilingaitė, A. Čivilis, X. Yin, and N. Zokaitė (2001). Computerized adaptive test based on Bayesian network for basic operations with fractions. KDE1 student project report, Aalborg University, Denmark. `http://www.cs.auc.dk/library`.

Kushilevitz, E. and N. Nisan (1997). *Communication Complexity*. Cambridge University Press.

Takikawa, M. and B. D'Ambrosio (1999). Multiplicative factorization of noisy-max. In K. B. Laskey and H. Prade (Eds.), *Proc.of the Fifteenth Conf. on Uncertainty in AI*, pp. 622–630.

Zhang, N. L. and D. Poole (1996). Exploiting causal independence in Bayesian network inference. *Journal of Artificial Intelligence Research 5*, 301–328.