Two applications of Bayesian networks

Jiří Vomlel

Laboratory for Intelligent Systems
University of Economics, Prague

Institute of Information Theory and Automation
Academy of Sciences of the Czech Republic

This presentation is available from
http://www.utia.cas.cz/vomlel/

Contents:

- Bayesian networks and the tasks solved with them
- Building “good” strategies using Bayesian network models
- Application 1: Adaptive testing
- Application 2: Decision-theoretic troubleshooting

Bayesian network

- a directed acyclic graph $G = (V, E)$
- each node $i \in V$ corresponds one random variable X_i with a finite set $\mathbb{X}_i$ of mutually exclusive states
- $pa(i)$ denotes the set of parents of node i in graph G
- to each node $i \in V$ corresponds a conditional probability table $P(X_i \mid (X_j)_{j \in pa(i)})$
- the DAG implies conditional independence relations between $(X_i)_{i \in V}$
- d-separation (Pearl, 1986) can be used to read the CI relations from the DAG

Using the chain rule we have that:

$$P((X_i)_{i \in V}) \quad = \quad \prod_{i \in V} P(X_i \mid X_{i-1}, \ldots, X_1)$$

Assume an ordering of $X_i, i \in V$ such that if $j \in pa(i)$ then $j < i$.
From the DAG we can read conditional independence relations

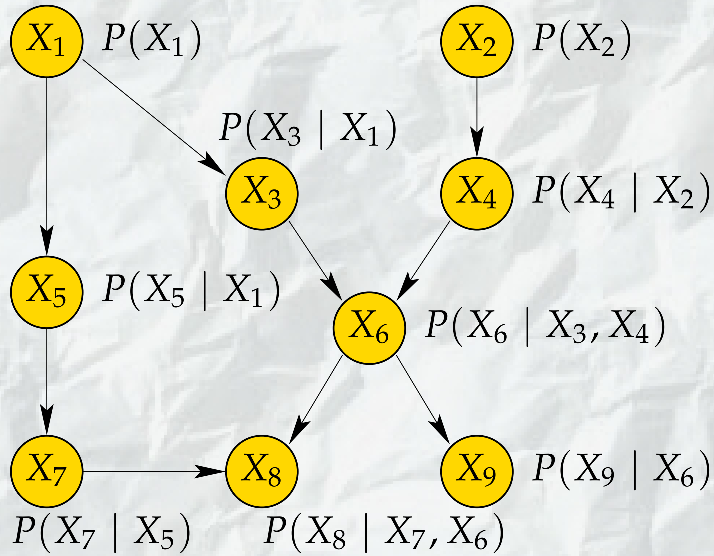
$$X_i \perp\!\!\!\perp X_k \mid (X_j)_{j \in pa(i)} \qquad \text{for } i \in V \text{ and } k < i \text{ and } k \notin pa(i)$$

Using the conditional independence relations from the DAG we get

$$P((X_i)_{i \in V}) \quad = \quad \prod_{i \in V} P(X_i \mid (X_j)_{j \in pa(i)}) \ .$$

It is the joint probability distribution represented by the Bayesian network.

Example:



$$
\begin{aligned}
P(X_1, \ldots, X_9) &= \\
&= P(X_9|X_8, \ldots, X_1) \cdot P(X_8|X_7, \ldots, X_1) \cdot \ldots \cdot P(X_2|X_1) \cdot P(X_1) \\
&= P(X_9|X_6) \cdot P(X_8|X_7, X_6) \cdot P(X_7|X_5) \cdot P(X_6|X_4, X_3) \\
&\quad \cdot P(X_5|X_1) \cdot P(X_4|X_2) \cdot P(X_3|X_1) \cdot P(X_2) \cdot P(X_1)
\end{aligned}
$$

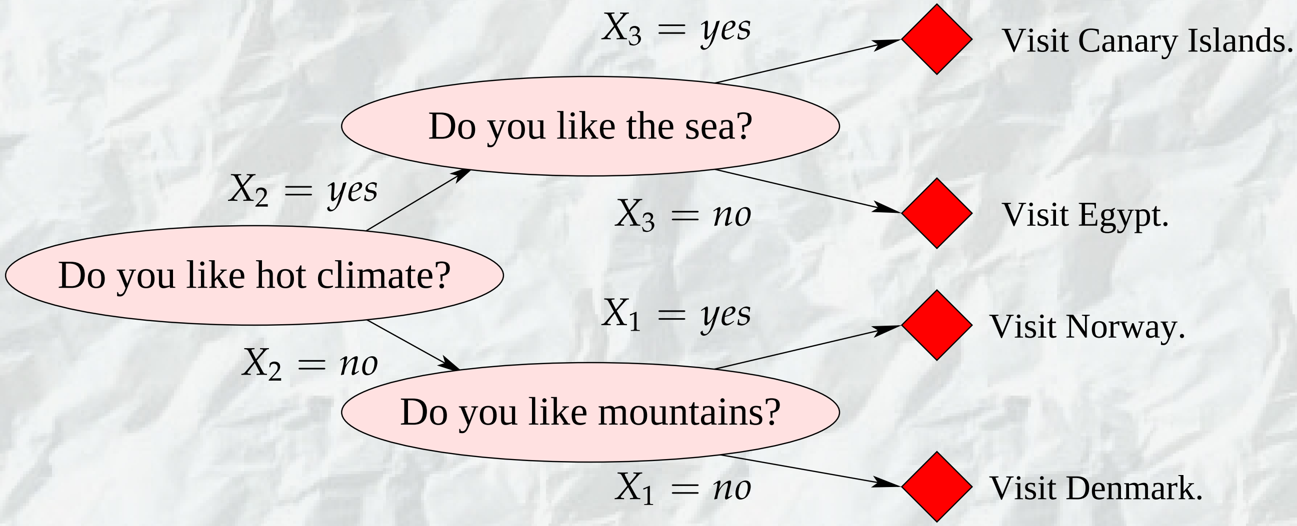
Typical use of Bayesian networks

- to model and explain a domain.
- to update beliefs about states of certain variables when some other variables were observed, i.e., computing conditional probability distributions, e.g., $P(X_{23}|X_{17} = yes, X_{54} = no)$.
- to find most probable configurations of variables
- to support decision making under uncertainty
- to find good strategies for solving tasks in a domain with uncertainty.

`Example 1: asia.net`

`Example 2: mildew4.net`

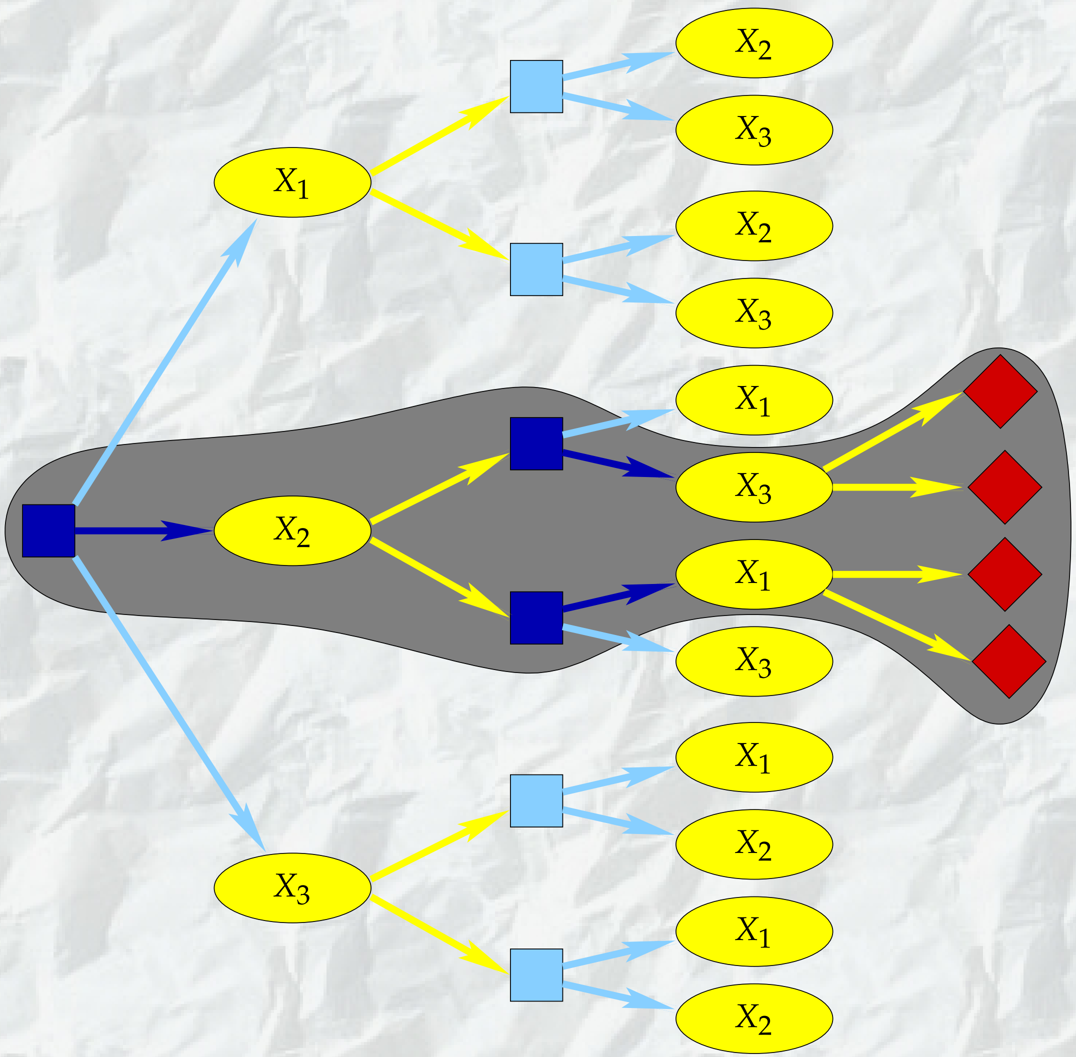
Strategy: “Holiday advisor”



For all nodes n of a strategy $\mathbf{s}$ we have defined:

- evidence $\mathbf{e}_n$, i.e. outcomes of steps performed to get to node n,
- probability $P(\mathbf{e}_n)$ of getting to node n, and
- utility $f(\mathbf{e}_n)$ being a real number.

Let $\mathcal{L}(\mathbf{s})$ be the set of terminal nodes of strategy $\mathbf{s}$.
Expected utility of strategy is $E_f(\mathbf{s}) = \sum_{\ell \in \mathcal{L}(\mathbf{s})} P(\mathbf{e}_\ell) \cdot f(\mathbf{e}_\ell)$.



Strategy $\mathbf{s}^\star$ is **optimal** iff it maximizes its expected utility.

Strategy $\mathbf{s}$ is **myopically optimal** iff each step of strategy $\mathbf{s}$ is selected so that it maximizes expected utility after the selected step is performed (*one step look ahead*).

Application 1: Adaptive test of basic operations with fractions

Examples of tasks:

$$T_1: \quad \left(\frac{3}{4} \cdot \frac{5}{6}\right) - \frac{1}{8} = \frac{15}{24} - \frac{1}{8} = \frac{5}{8} - \frac{1}{8} = \frac{4}{8} = \frac{1}{2}$$

$$T_2: \quad \frac{1}{6} + \frac{1}{12} = \frac{2}{12} + \frac{1}{12} = \frac{3}{12} = \frac{1}{4}$$

$$T_3: \quad \frac{1}{4} \cdot 1\frac{1}{2} = \frac{1}{4} \cdot \frac{3}{2} = \frac{3}{8}$$

$$T_4: \quad \left(\frac{1}{2} \cdot \frac{1}{2}\right) \cdot \left(\frac{1}{3} + \frac{1}{3}\right) = \frac{1}{4} \cdot \frac{2}{3} = \frac{2}{12} = \frac{1}{6} \; .$$

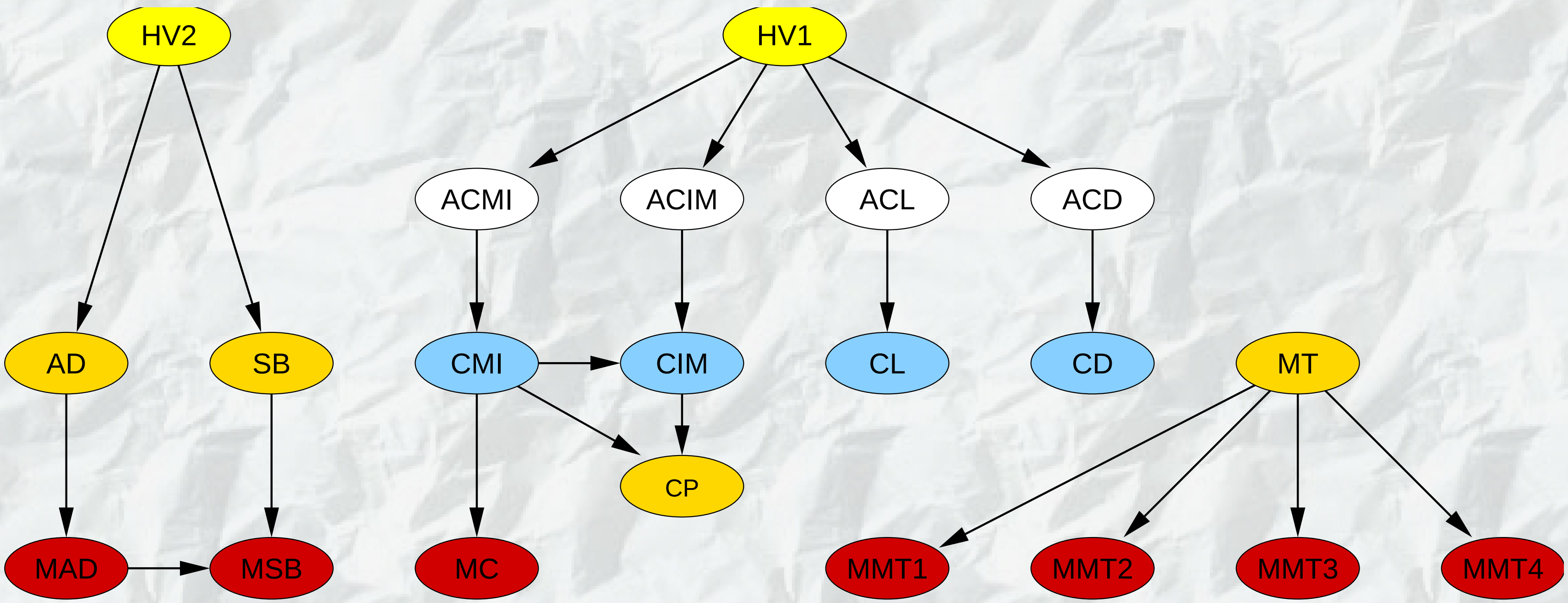
Elementary and operational skills

CP	Comparison (common numerator or denominator)	$\frac{1}{2} > \frac{1}{3}, \quad \frac{2}{3} > \frac{1}{3}$
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 improp. 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%

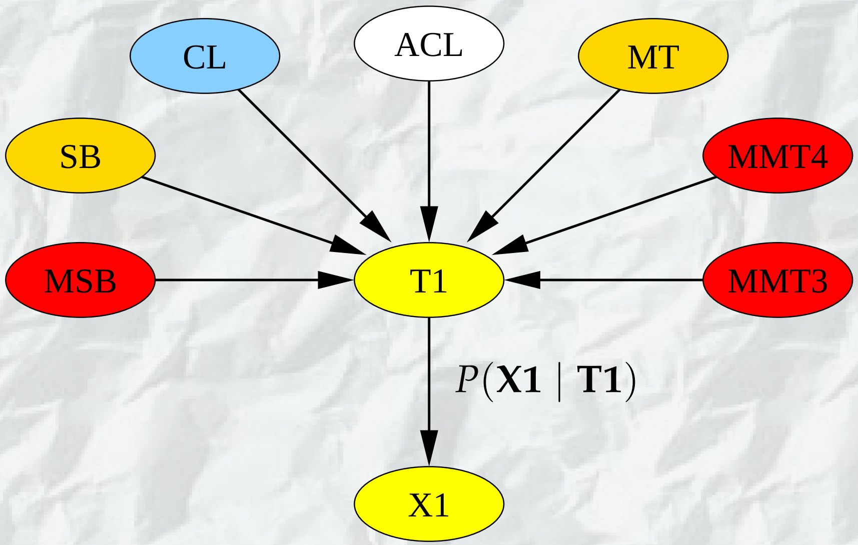
Student model



Evidence model for task $T1$

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

$$T1 \quad \Leftrightarrow \quad MT \;\&\; CL \;\&\; ACL \;\&\; SB \;\&\; \neg MMT3 \;\&\; \neg MMT4 \;\&\; \neg MSB$$

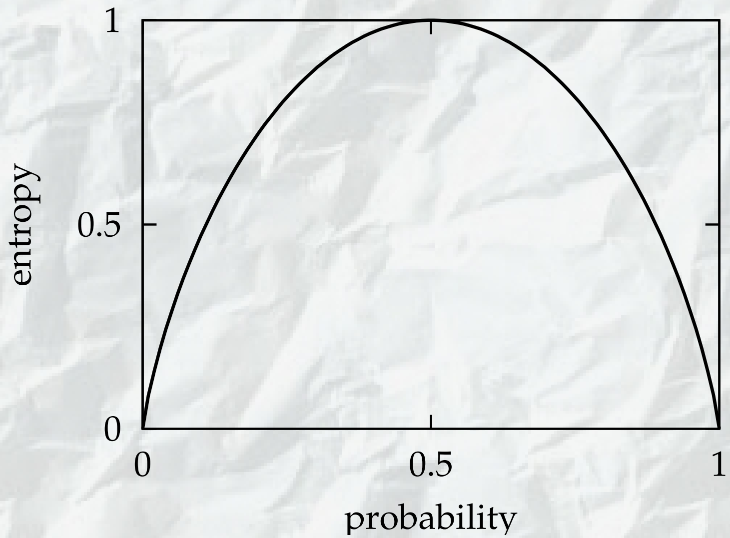


Hugin: `model-hv-2.net`

Using information gain as the utility function

"The lower the entropy of a probability distribution the more we know."

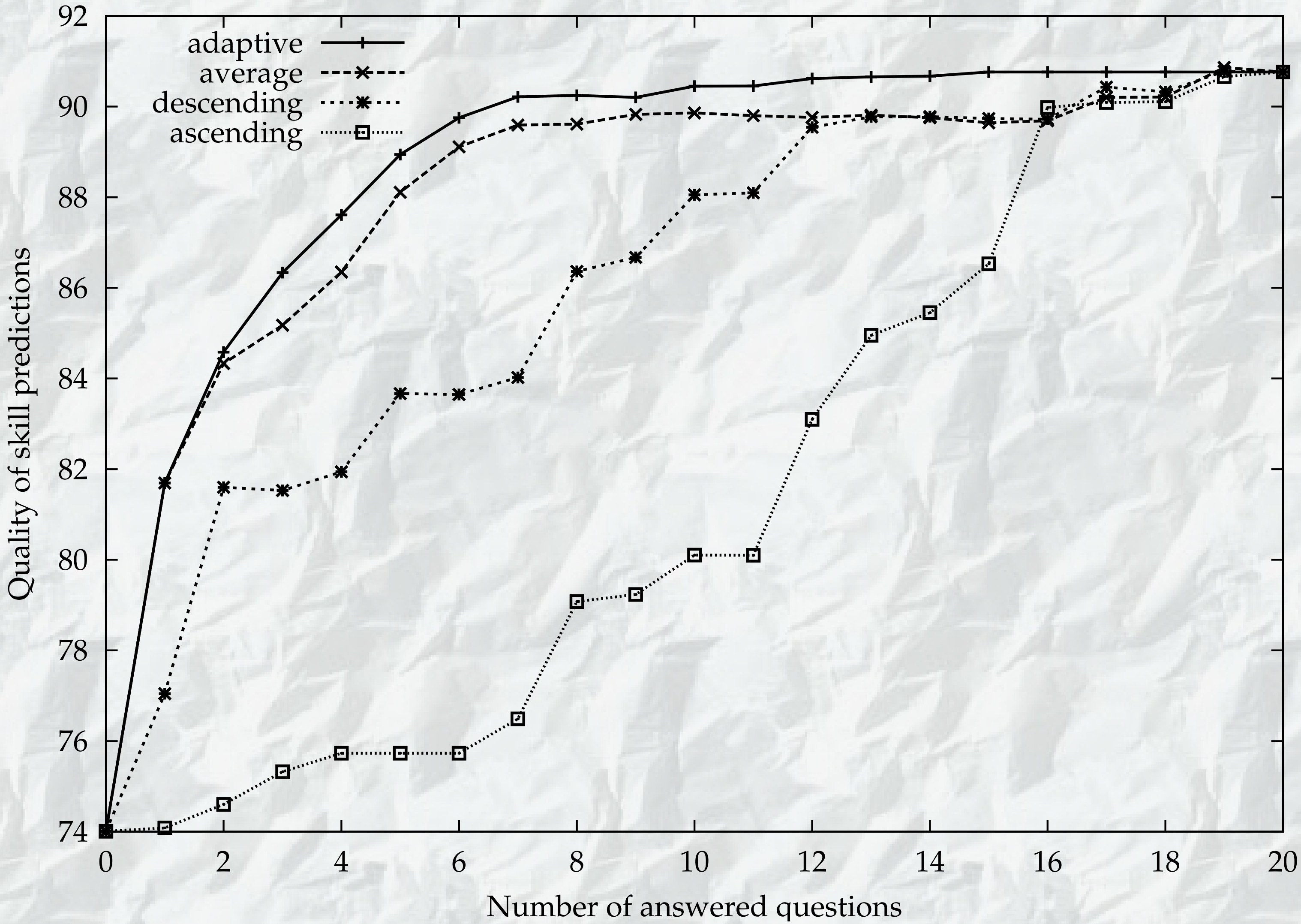
$$H\left(P(\mathbf{X})\right) = -\sum_{\mathbf{x}} P(\mathbf{X} = \mathbf{x}) \cdot \log P(\mathbf{X} = \mathbf{x})$$

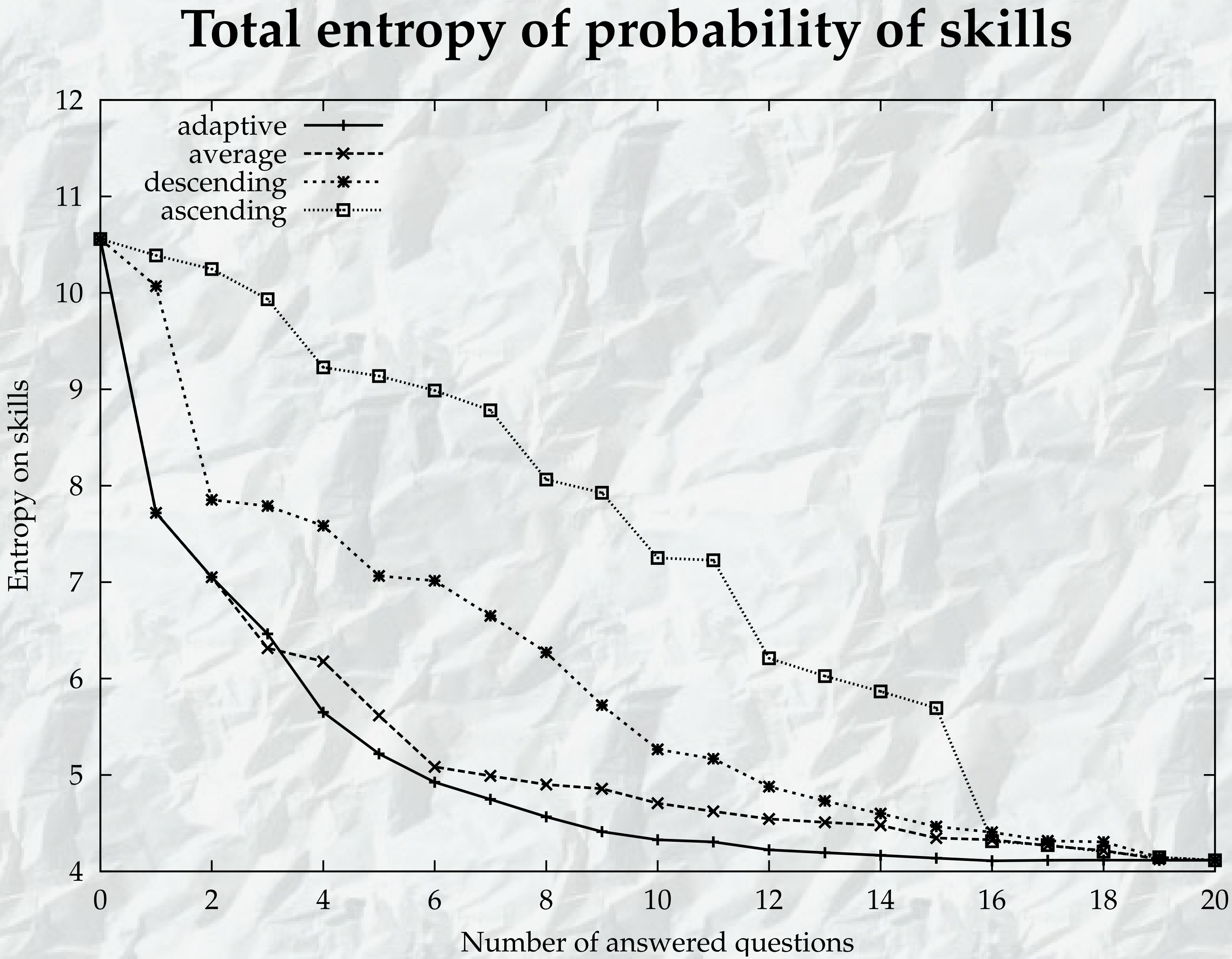


Information gain in a node n of a strategy

$$IG(\mathbf{e}_n) = H(P(\mathbf{S})) - H(P(\mathbf{S} \mid \mathbf{e}_n))$$

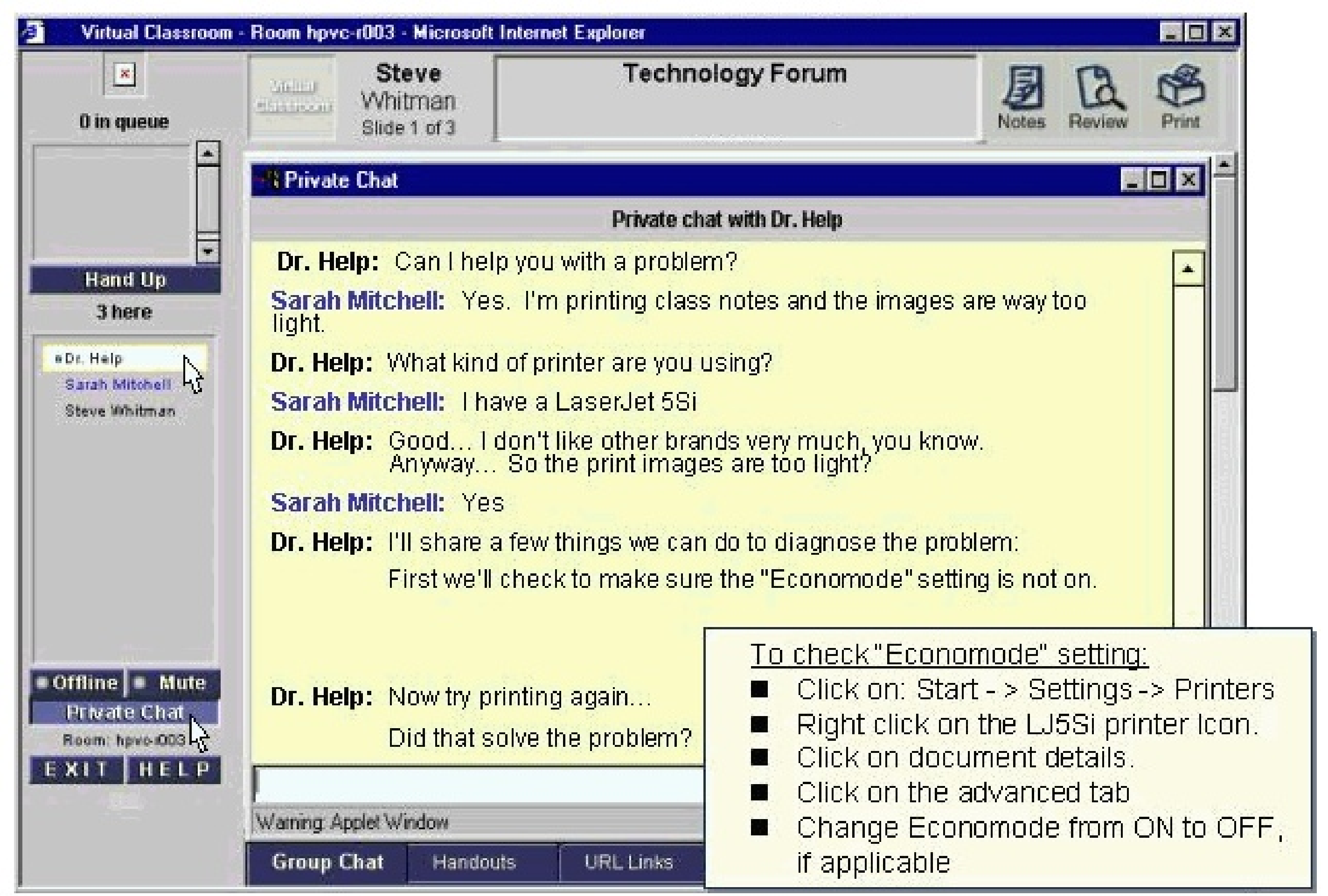
Skill Prediction Quality



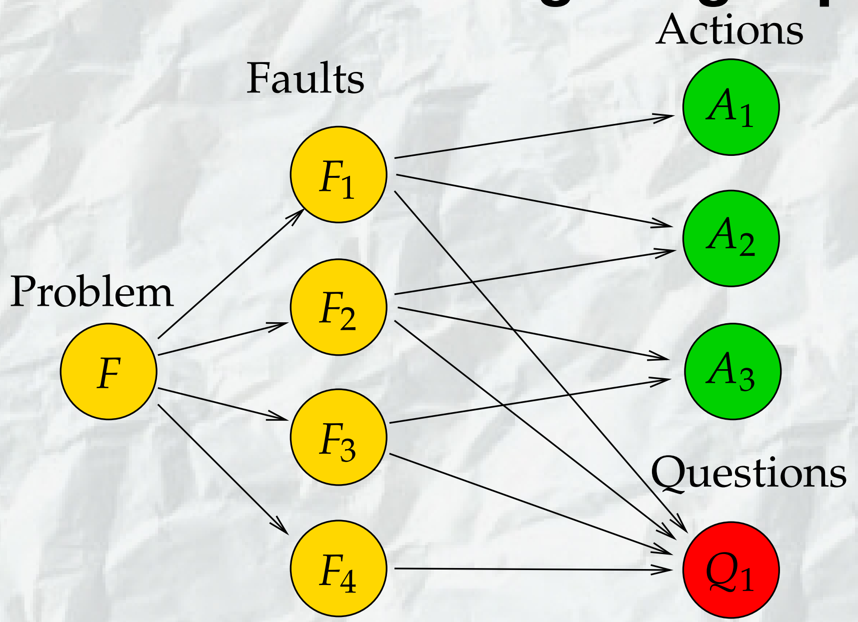


Total entropy of probability of skills
adaptive
average
descending
ascending
Entropy on skills
Number of answered questions
12
11
10
9
8
7
6
5
4
0
2
4
6
8
10
12
14
16
18
20

Application 2: Troubleshooting

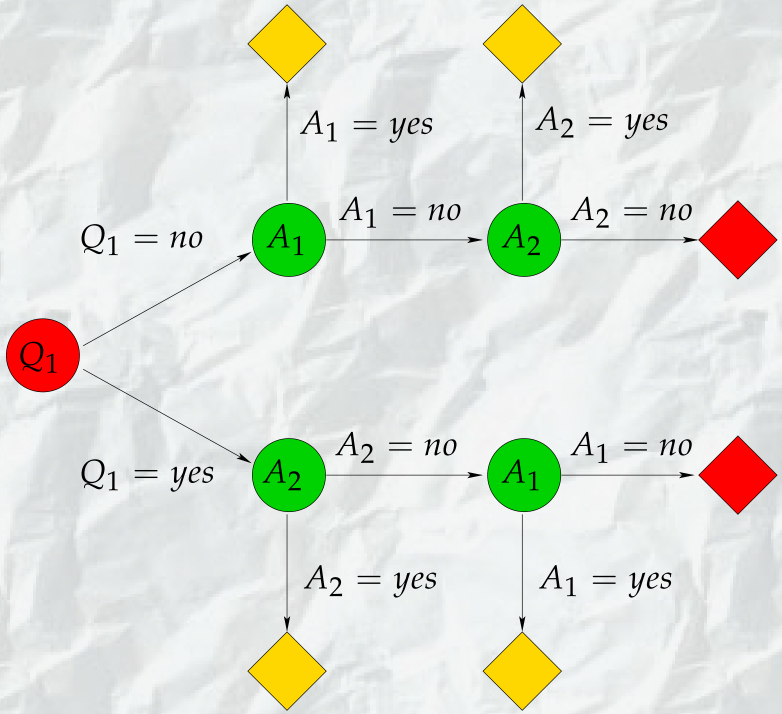


Application 2: Troubleshooting - Light print problem



- Problems: F_1 *Distribution problem*, F_2 *Defective toner*, F_3 *Corrupted dataflow*, and F_4 *Wrong driver setting*.
- Actions: A_1 *Remove, shake and reseat toner*, A_2 *Try another toner*, and A_3 *Cycle power*.
- Questions: Q_1 *Is the configuration page printed light?*

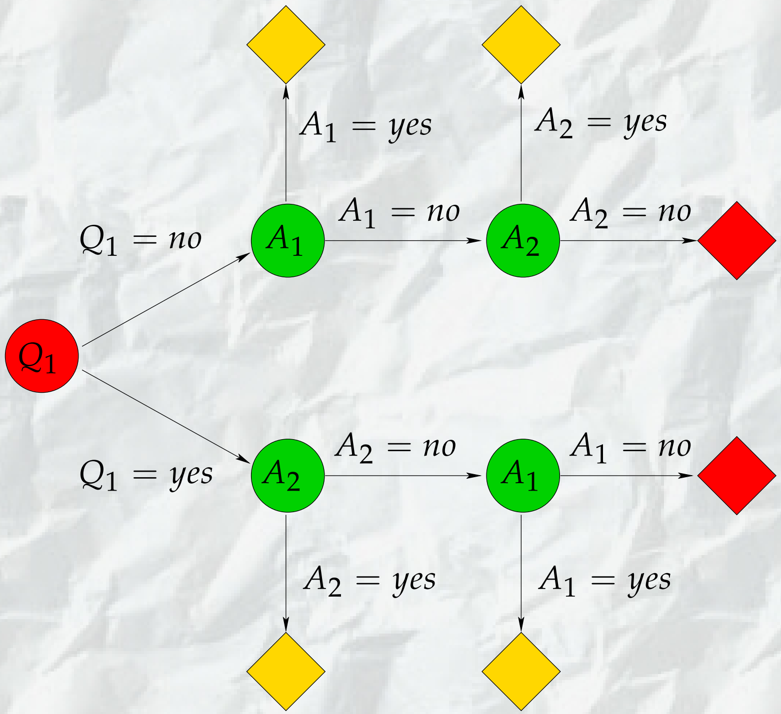
Troubleshooting strategy



The task is to find a strategy $\mathbf{s} \in \mathcal{S}$ minimising expected cost of repair

$$E_{CR}(\mathbf{s}) \quad = \quad \sum_{\ell \in \mathcal{L}(\mathbf{s})} P(\mathbf{e}_\ell) \cdot \left(\, t(\mathbf{e}_\ell) + c(\mathbf{e}_\ell) \, \right) \quad .$$

Expected cost of repair for a given strategy



$$
\begin{aligned}
E_{CR}(\mathbf{s}) \quad = \\
& P(Q_1 = no, A_1 = yes) \cdot (c_{Q_1} + c_{A_1}) \\
& +P(Q_1 = no, A_1 = no, A_2 = yes) \cdot (c_{Q_1} + c_{A_1} + c_{A_2}) \\
& +P(Q_1 = no, A_1 = no, A_2 = no) \cdot (c_{Q_1} + c_{A_1} + c_{A_2} + c_{CS}) \\
& +P(Q_1 = yes, A_2 = yes) \cdot (c_{Q_1} + c_{A_2}) \\
& +P(Q_1 = yes, A_2 = no, A_1 = yes) \cdot (c_{Q_1} + c_{A_2} + c_{A_1}) \\
& +P(Q_1 = yes, A_2 = no, A_1 = no) \cdot (c_{Q_1} + c_{A_2} + c_{A_1} + c_{CS})
\end{aligned}
$$

Demo: `headache_problem`

Commercial applications of Bayesian networks in educational testing and troubleshooting

- Hugin Expert A/S.
  software product: Hugin - a Bayesian network tool.
  http://www.hugin.com/
- Educational Testing Service (ETS)
  the world's largest private educational testing organization
  Research unit doing research on adaptive tests using Bayesian networks: http://www.ets.org/research/
- SACSO Project
  Systems for Automatic Customer Support Operations
  - research project of Hewlett Packard and Aalborg University.
  The troubleshooter offered as DezisionWorks by Dezide Ltd.
  http://www.dezide.com/