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 as a model for reasoning with uncertainty
- Building probabilistic models
- Building “good” strategies using the models
- Application 1: Adaptive testing
- Application 2: Decision-theoretic troubleshooting

Independence

If two discrete random variables are independent, the probability of the joint occurrence of values of two variables is equal to the product of the probabilities individually:

$$P(X = x, Y = y) \quad = \quad P(X = x) \cdot P(Y = y).$$

Also,

$$P(X = x | Y = y) = P(X = x)$$

- learning the value of Y does not influence your belief about X.

Example: `two_coins.net`

Conditional independence

If two variables are conditionally independent, the conditional probability of the joint occurrence given the value of another variable is equal to the product of the conditional probabilities:

$$P(X = x, Y = y|Z = z) \quad = \quad P(X = x|Z = z) \cdot P(Y = y|Z = z) \; .$$

- Also, learning the value of Z may influence your belief about X and about Y,
- but if you know the value of Z, learning the value of Y does not influence your belief about X.

$$P(X = x|Y = y, Z = z) \quad = \quad P(X = x|Z = z) \; .$$

Example: `two_biased_coins.net`

Pearl on Conditional independence (Pearl, 1988, p. 44)

- Conditional independence is not a grace of nature for which we must wait passively, but rather a psychological necessity which we satisfy actively by organizing our knowledge in a specific way.

- An important tool in such organization is the identification of intermediate variables that induce conditional independence among observables; if they are not in our vocabulary, we create them. *In medical diagnosis when some symptoms directly influence one another, the medical profession invents a name for that interaction (e.g., "syndrome", "complication," "pathological state") and treats it as a new auxiliary variable that induces conditional independence;*

- dependency between any two interacting systems is fully attributed to the dependencies of each on the auxiliary variable.

Building up complex networks

- Relationships among many variables are modeled in terms of important relationships among smaller subsets of variables.

Example: Wet grass on Holmes' lawn can be caused either by rain or by his sprinkler.

$$
\begin{aligned}
& P(Holmes, Watson, Rain, Sprinkler) \\
= \quad & P(Holm|Wat, Rn, Sprnk) \cdot P(Wat|Rn, Sprnk) \cdot P(Rn|Sprnk) \cdot P(Sprnk) \\
= \quad & P(Holm|Rn, Sprnk) \cdot P(Wat|Rn) \cdot P(Rn) \cdot P(Sprnk)
\end{aligned}
$$

Example: `wet_grass.net`

Building up complex Bayesian networks

- Acyclic directed graphs (DAGs):
- Nodes correspond to variables
- Directed edges represent explicit dependence relationships
- No edges means no explicit dependence, although there can be dependence through relationships with other variables.

`Example: asia.net`

Building Bayesian network models

three basic approaches

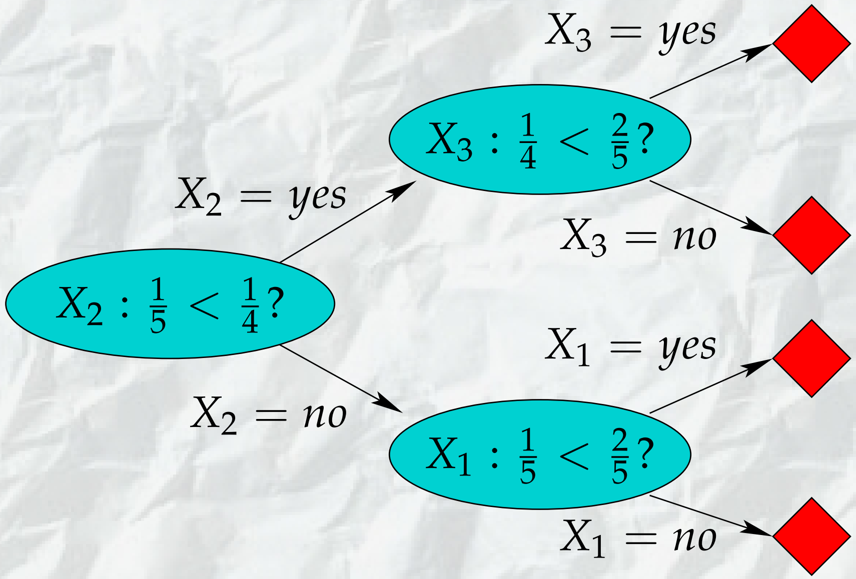
- Discussions with domain experts: expert knowledge is used to get the structure and parameters of the model
- A dataset of records is collected and a machine learning method is used to to construct a model and estimate its parameters.
- A combination of previous two: e.g. experts helps with the stucture, data are used to estimate parameters.

Typical tasks solved using Bayesian networks

Bayesian networks are used:

- 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 of a strategy



X_3 is more difficult question than X_2 which is more difficult than X_1.

Building strategies using the models

For all terminal nodes $\ell \in \mathcal{L}(\mathbf{s})$ of a strategy $\mathbf{s}$ we have defined:

- steps that were performed to get to that node together with their outcomes. It is called collected evidence $\mathbf{e}_\ell$.
- Using the probabilistic model of the domain we can compute probability of getting to that terminal node $P(\mathbf{e}_\ell)$.

During the process of collecting evidence $\mathbf{e}$ we update the probability of getting to a terminal node, which corresponds to conditional probability $P(\mathbf{e}_\ell \mid \mathbf{e})$, where $\mathbf{e}$ is evidence collected as far.

Building strategies using the models

For all terminal nodes $\ell \in \mathcal{L}(\mathbf{s})$ of a strategy $\mathbf{s}$ we have also defined:

- an evaluation function $f : \cup_{\mathbf{s} \in \mathcal{S}} \mathcal{L}(\mathbf{s}) \mapsto \mathbb{R}$.

For each strategy we can compute:

- expected value of the strategy:

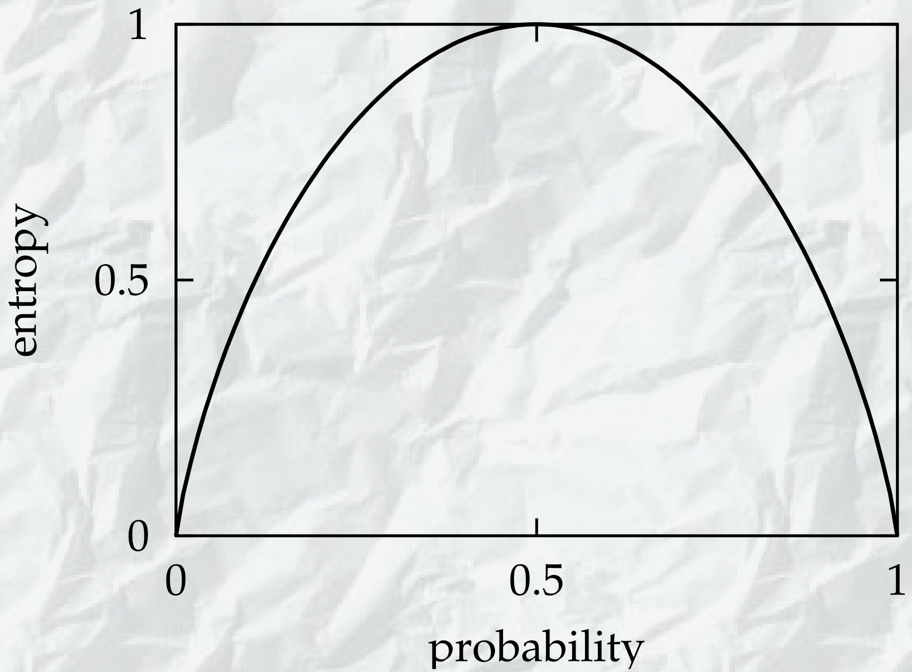
$$E_f(\mathbf{s}) = \sum_{\ell \in \mathcal{L}(\mathbf{s})} P(\mathbf{e}_\ell) \cdot f(\mathbf{e}_\ell)$$

The goal:

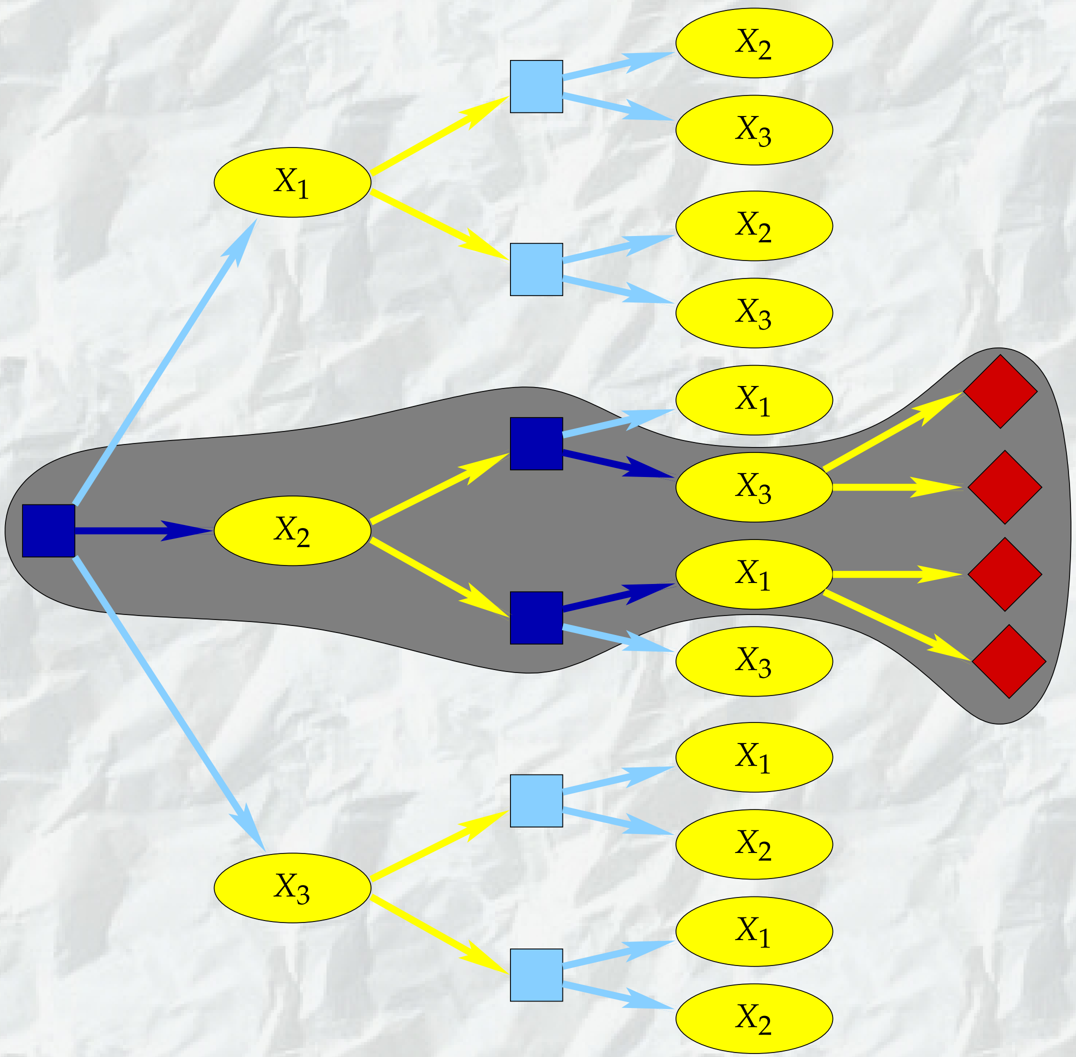
- find a strategy that maximizes (minimizes) its expected value

Using entropy as an information measure

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



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

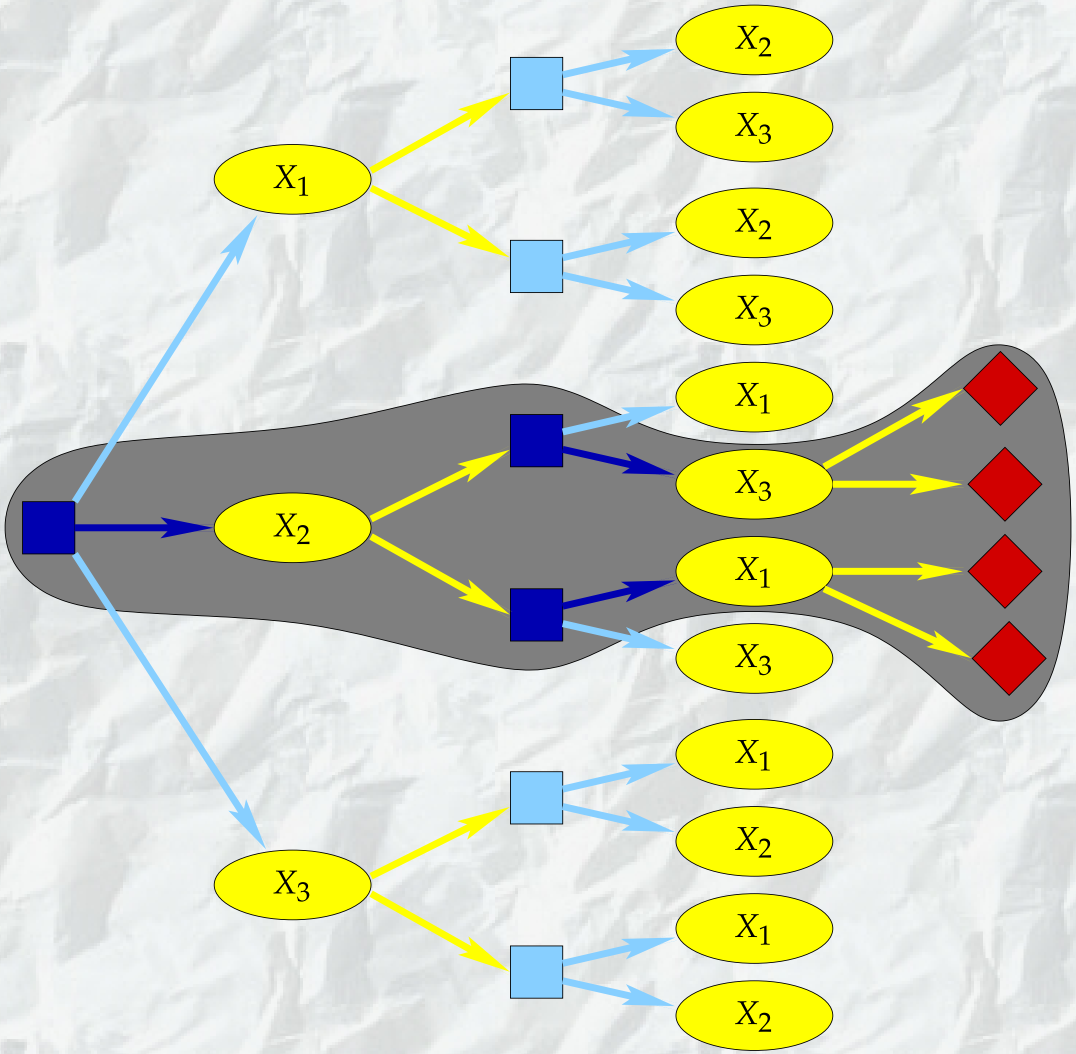


Entropy in node n

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

Expected entropy at the end of test $\mathbf{t}$

$$E_H(\mathbf{t}) = \sum_{\ell \in \mathcal{L}(\mathbf{t})} P(\mathbf{e}_\ell) \cdot H(\mathbf{e}_\ell)$$



$\mathcal{T}$ ... the set of all possible tests

A test $\mathbf{t}^\star$ is **optimal** iff

$$\mathbf{t}^\star = \arg\min_{\mathbf{t}\in\mathcal{T}} E_H(\mathbf{t}) \ .$$

A test $\mathbf{t}$ is **myopically optimal** iff each question $X^\star$ of $\mathbf{t}$ minimizes the expected value of entropy after the question is answered:

$$X^\star = \arg\min_{X\in\mathcal{X}} E_H(\mathbf{t}_{\downarrow X}) \ ,$$

i.e. it works as if the test finished after the selected question $X^\star$.

Application 1: Adaptive test of basic operations with fractions

Examples of tasks:

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

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

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

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

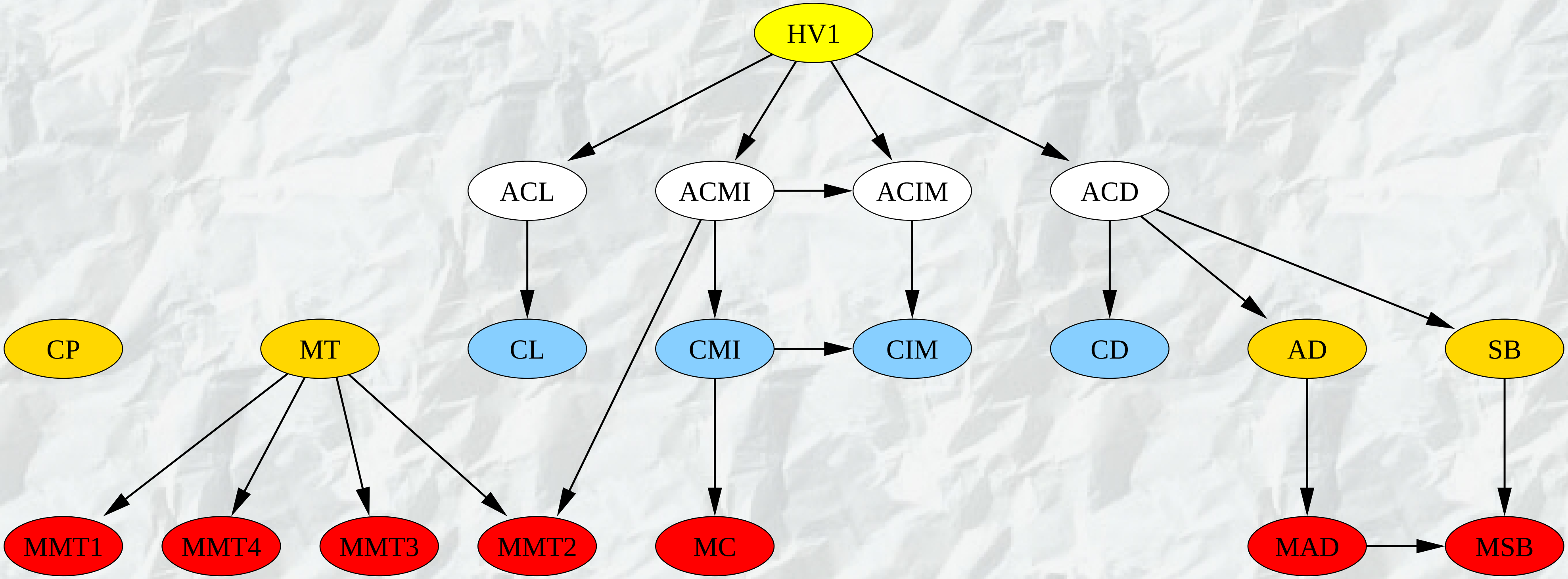
Elementary and operational skills

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%

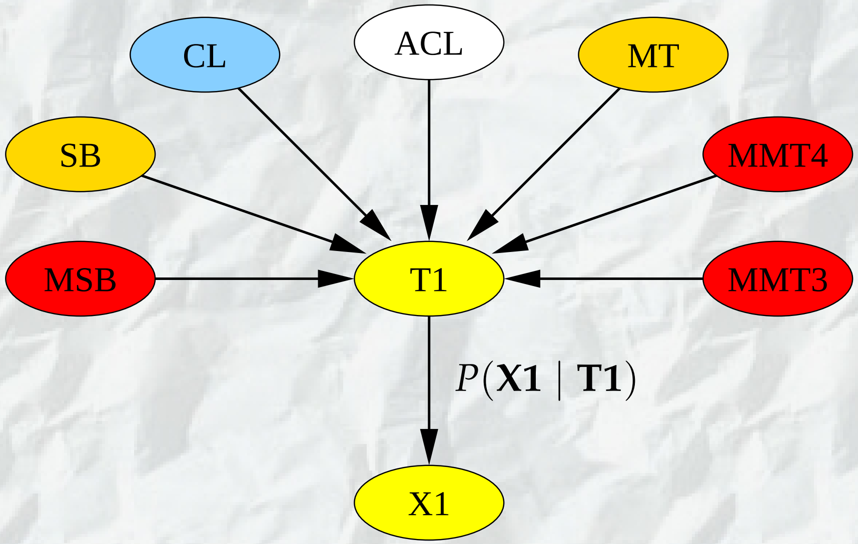
Student model



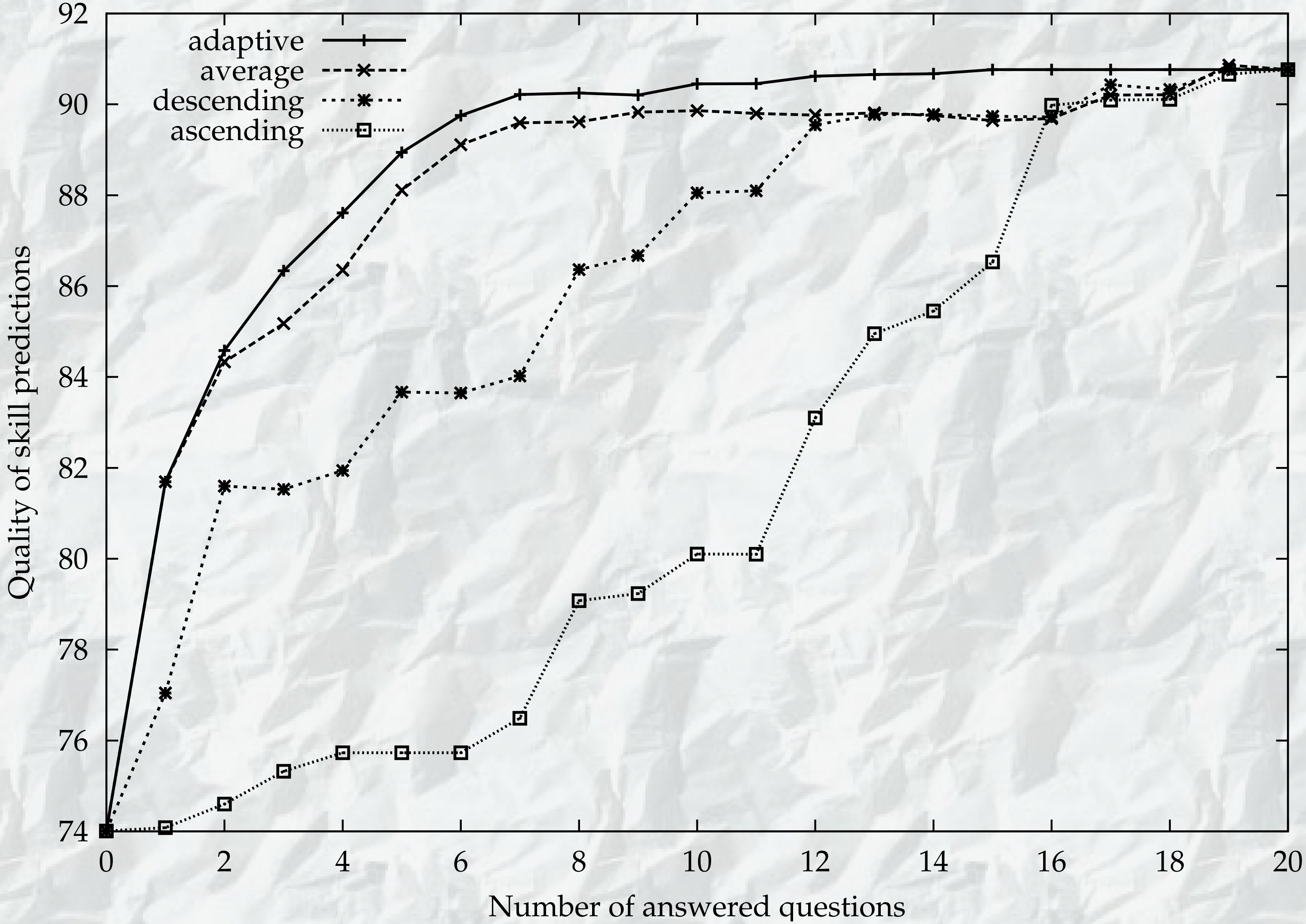
Evidence model for task $T1$

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

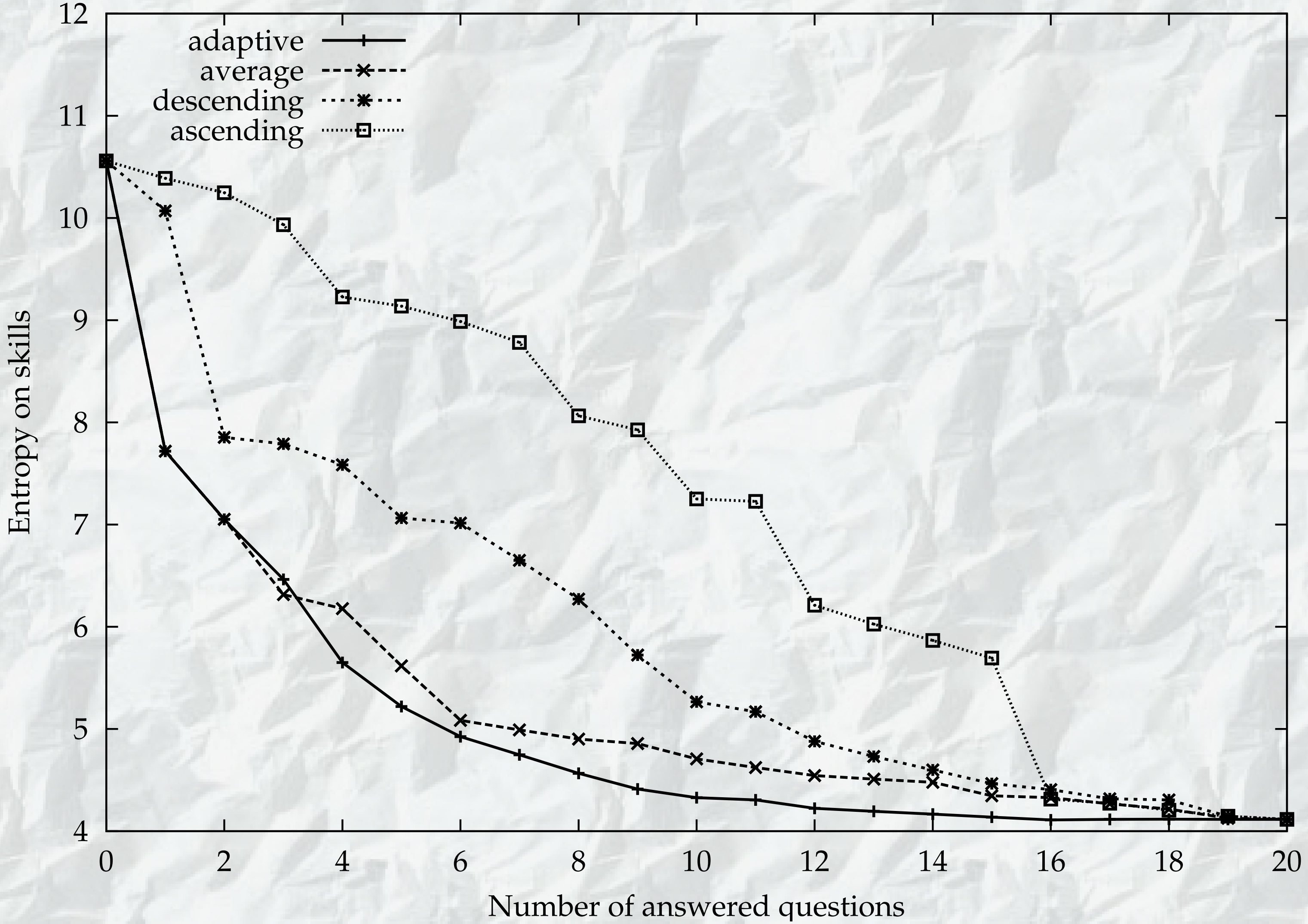
$$T1 \quad \Leftrightarrow \quad MT \,\&\, CL \,\&\, ACL \,\&\, SB \,\&\, \neg MMT3 \,\&\, \neg MMT4 \,\&\, \neg MSB$$



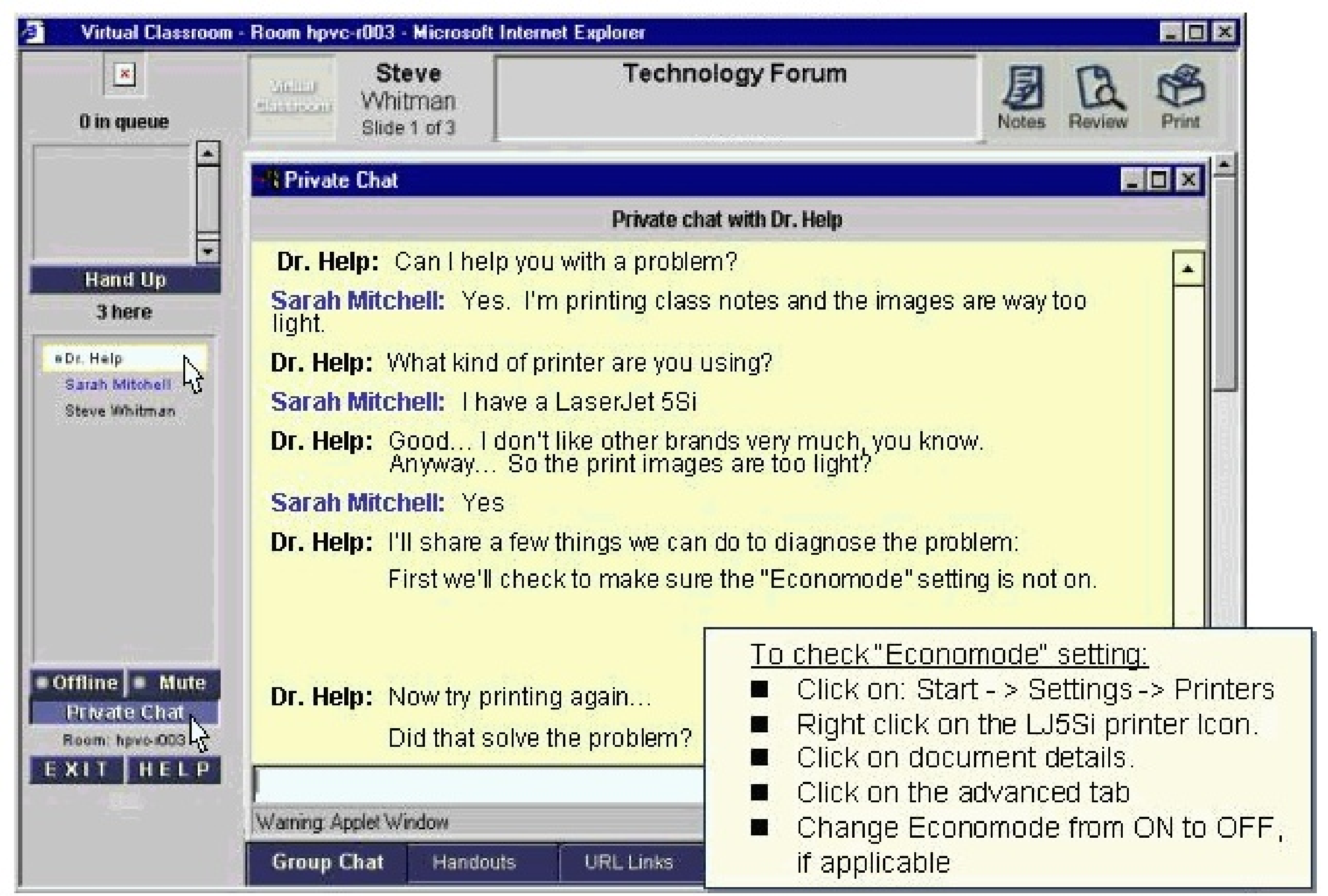
Skill Prediction Quality



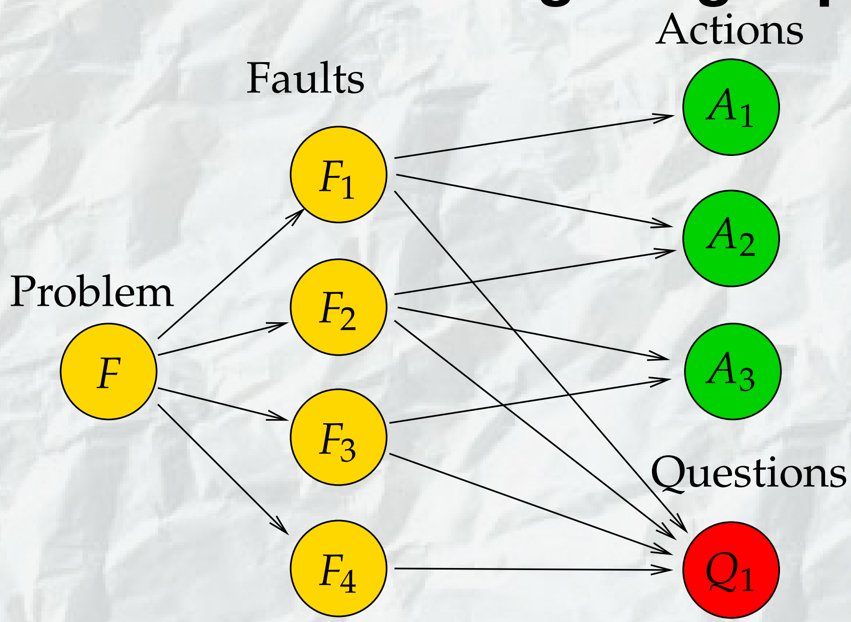
Total entropy of probability of skills



Application 2: Troubleshooting

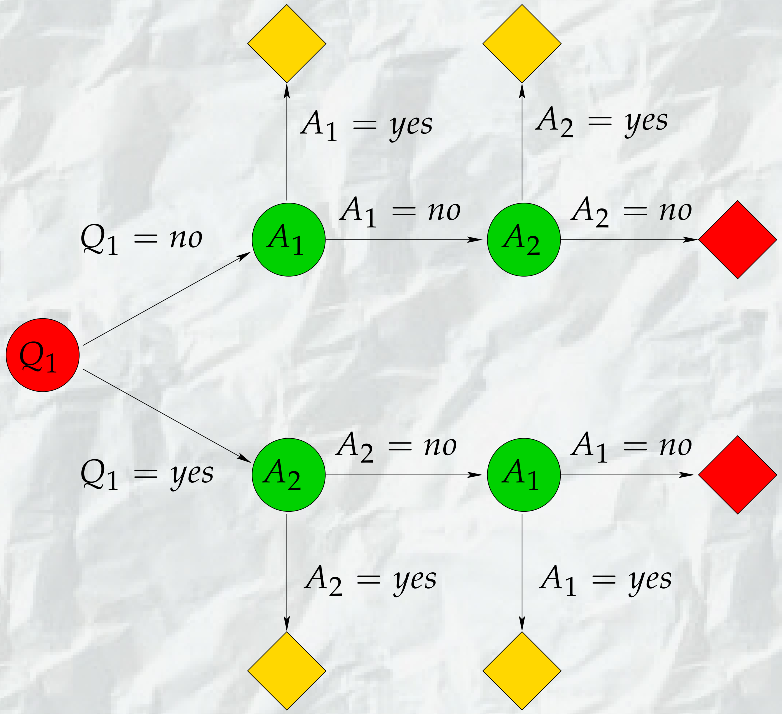


Application 2: Troubleshooting - Light print problem



- 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?*

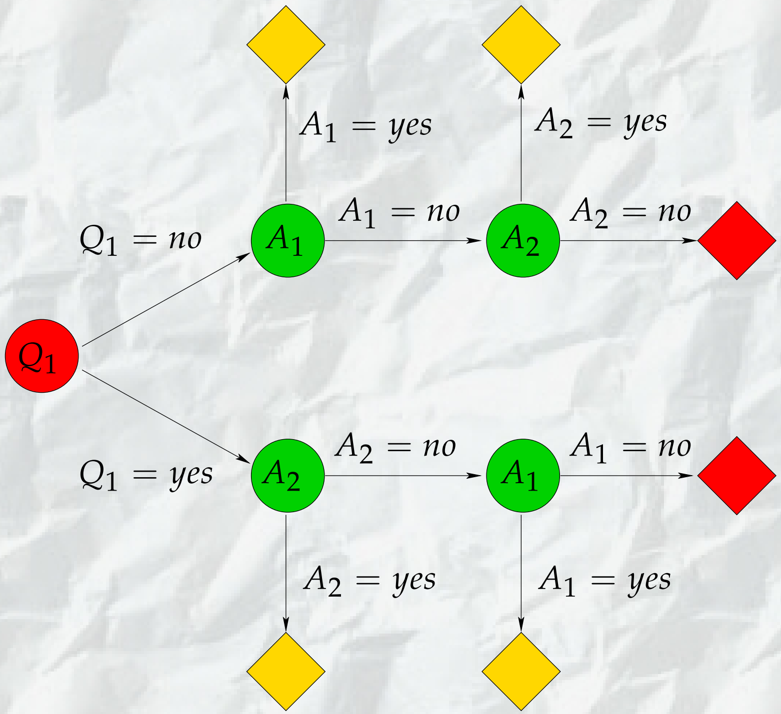
Troubleshooting strategy



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

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

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: `light_print_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/

...and it is time to end.