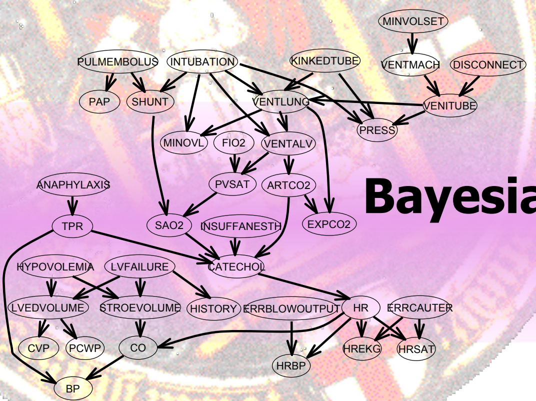
Mainly based on F. V. Jensen, „Bayesian Networks and Decision Graphs“, Springer-Verlag New York, 2001.

Advanced I WS 06/07

Bayesian Networks - Intro -



Wolfram Burgard, Luc De Raedt, Kristian Kersting, Bernhard Nebel

Albert-Ludwigs University Freiburg, Germany

Advanced I WS 06/07

Why bother with uncertainty?

Uncertainty appears in many tasks

- Partial knowledge of the state of the world
- Noisy observations
- Phenomena that are not covered by our models
- Inherent stochasticity

$$\forall \texttt{Patient}: \ \texttt{symptoms}(\texttt{P}, \texttt{toothache}) \Rightarrow \texttt{disease}(\texttt{P}, \texttt{cavity})$$

Advanced I WS 06/07

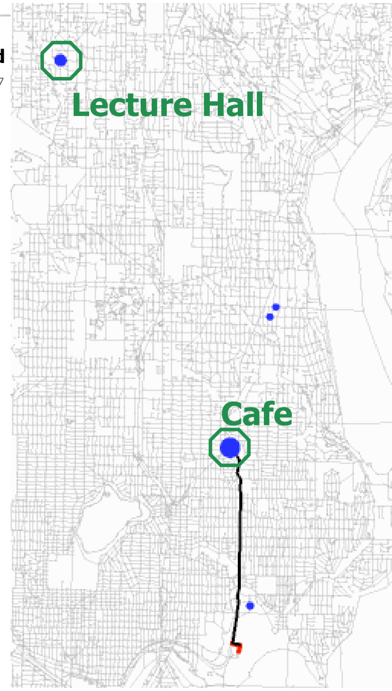
Recommendation Systems

Your friends attended this lecture already and liked it. Therefore, we would like to recommend it to you !

Advanced I WS 06/07

Activity Recognition

[Fox et al. IJCAI03]

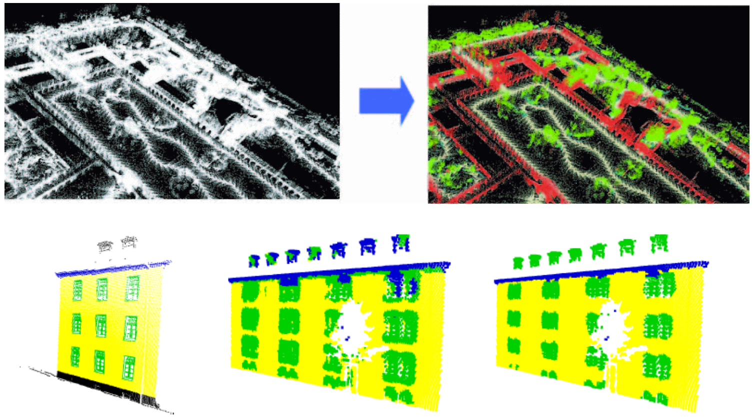


Will you go to the AdvancedAI lecture or will you visit some friends in a cafe?

3D Scan Data Segmentation

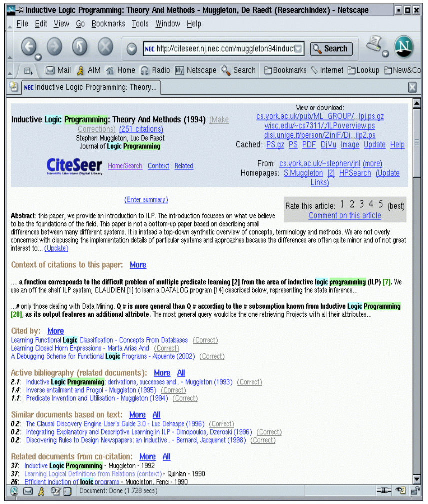
[Anguelov et al. CVPR05, Triebel et al. ICRA06]

How do you recognize the lecture hall?



(a) Training data (b) Bayes classifier (c) AMN + adaptive reduction

Duplicate Identification

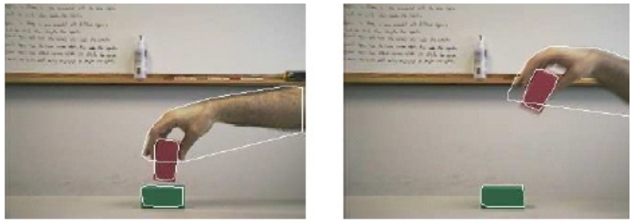


- L. D. Raedt
- L. de Raedt
- Luc De Raedt

- Wolfram Burgard
- W. Burgold
- Wolfram Burgold

Video event recognition

[Fern JAIR02,IJCAI05]



Frame 14 Frame 21

- What is going on?
- Is the red block on top of the green one?
- …

How do we deal with uncertainty?

- Implicit:
  - **Ignore** what you are uncertain if you can
  - Build procedures that are robust to uncertainty
- Explicit:
  - **Build a model** of the world that describes uncertainty about its state, dynamics, and observations
  - Reason about the effects of actions given the model

Graphical models = explicit, model-based

Probability

- A **well-founded** framework for uncertainty
- Clear semantics: **joint prob. distribution**
- Provides **principled answers** for:
  - Combining evidence
  - Predictive & Diagnostic reasoning
  - Incorporation of new evidence
- **Intuitive** (at some level) to human experts
- Can automatically be **estimated from data**

Joint Probability Distribution

- „truth table" of set $X_1, \ldots, X_n$ of random variables

X_1	X_2	X_3	X_1, X_2, X_3
true	1	green	0.001
true	1	blue	0.021
true	2	green	0.134
true	2	blue	0.042
...	...	...	...
false	2	blue	0.2

- Any probability we are interested in can be computed from it

Representing Prob. Distributions

- Probability distribution = probability for each combination of values of these attributes

Hospital patients described by
- Background: age, gender, history of diseases, …
- Symptoms: fever, blood pressure, headache, …
- Diseases: pneumonia, heart attack, …

- Naïve representations (such as tables) run into troubles
  - 20 attributes require more than $2^{20} \geq 10^6$ parameters
  - Real applications usually involve hundreds of attributes

Bayesian Networks - Key Idea

Exploit regularities !!!

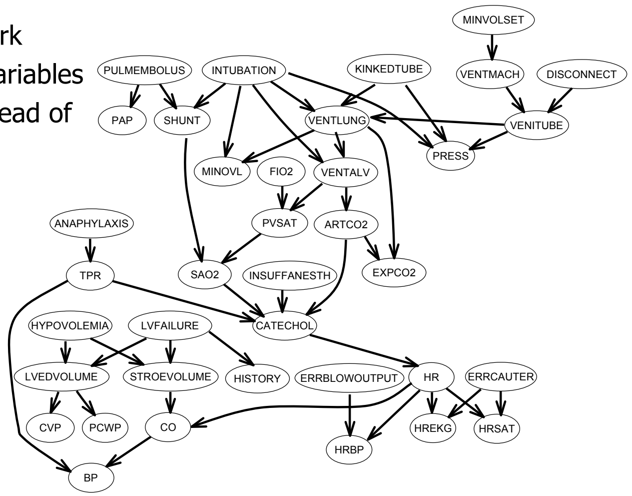
- Bayesian networks
  - utilize **conditional independence**
  - **Graphical Representation** of conditional independence respectively "causal" dependencies

A Bayesian Network

The "ICU alarm" network

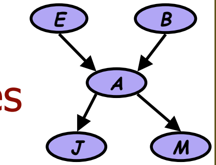
- 37 binary random variables
- 509 parameters instead of

$2^{37} = 10^{12}$



Bayesian Networks

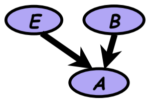
1. Finite, acyclic graph
2. Nodes: (discrete) random variables
3. Edges: direct influences
4. Associated with each node: a table representing a conditional probability distribution (CPD), quantifying the effect the parents have on the node



$$P(X_1, \dots, X_n) = \prod_{i=1}^{n} \boxed{P(X_i | \mathrm{pa}(X_i))}$$

Associated CPDs

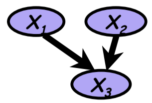
- naive representation
  - tables
- other representations
  - decision trees
  - rules
  - neural networks
  - support vector machines
  - ...



E	B	P(A | E,B)	
e	b	.9	.1
e	b	.7	.3
e	b	.8	.2
e	b	.99	.01

Bayesian Networks

$P(X_1)$
(0.2, 0.8)

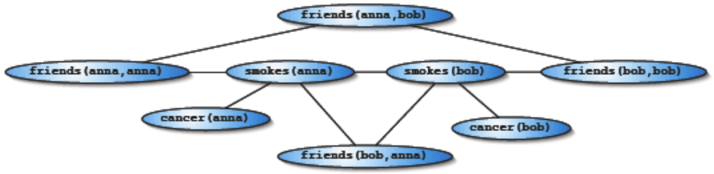


$P(X_2)$
(0.6, 0.4)

X_1	X_2	$P(X_3 \| X_2, X_1)$
true	1	(0.2,0.8)
true	2	(0.5,0.5)
false	1	(0.23,0.77)
false	2	(0.53,0.47)

$$\begin{aligned} P(X_1 = true, X_2 = 1, X_3 = false) &= P(X_1 = true) \cdot P(X_2 = 1) \\ &\quad \cdot P(X_3 = false | X_2 = 1, X_1 = true) \\ &= 0.2 \cdot 0.6 \cdot 0.8 = 0.96 \end{aligned}$$

Markov Networks

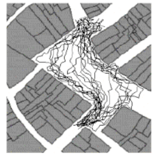


- Undirected Graphs $X_1, \ldots, X_n$
- Nodes = random variables
- Cliques = potentials (~ local jpd) ϕ_k

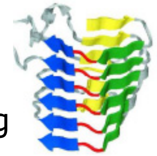
$$P(X = x) = \frac{1}{Z} \prod_k \phi_k(x_{\{k\}})$$

Fielded Applications

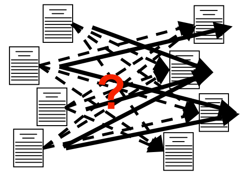
- Expert systems
  - Medical diagnosis (Mammography)
  - Fault diagnosis (jet-engines, Windows 98)



- Monitoring
  - Space shuttle engines (Vista project)
  - Freeway traffic, Activity Recognition

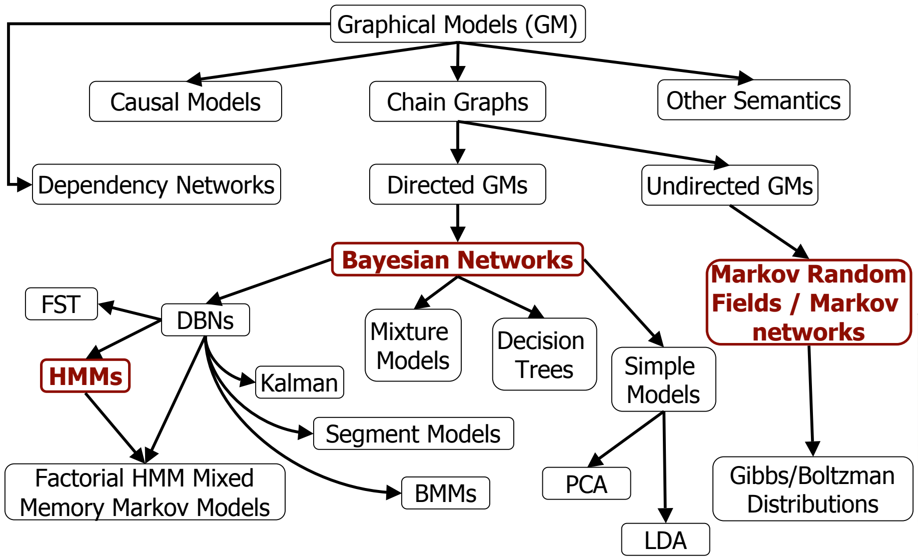


- Sequence analysis and classification
  - Speech recognition (Translation, Paraphrasing
  - Biological sequences (DNA, Proteins, RNA, ..)
- Information access
  - Collaborative filtering
  - Information retrieval & extraction



… among others

Graphical Models



Outline

- Introduction
- Reminder: Probability theory
- Basics of Bayesian Networks
- Modeling Bayesian networks
- Inference
- Excourse: Markov Networks
- Learning Bayesian networks
- Relational Models