Bayesian Networks: Theory and Applications

Prof. Dr. Rudolf Kruse
Computational Intelligence Group Magdeburg

Research: Data Sciences, Intelligent Systems

Software Tools: NEFCLASS, Information Miner,…

Transfers: Industrial Projects (BT, SAP, Siemens, Volkswagen, …), Spin Offs
### Property planning - Volkswagen

<table>
<thead>
<tr>
<th>Property family</th>
<th>Car body</th>
<th>Motor</th>
<th>Radio</th>
<th>Doors</th>
<th>Seat cover</th>
<th>Makeup mirror</th>
</tr>
</thead>
<tbody>
<tr>
<td>Property</td>
<td>Hatchback</td>
<td>2.8 L 150 kW Otto</td>
<td>Type alpha</td>
<td>4</td>
<td>Leather, Type L3</td>
<td>yes</td>
</tr>
</tbody>
</table>

### Complexity
- About 200 variables
- Typically 4 to 8, but up to 150 possible instances per variable
- More than $2^{200}$ possible combinations available
Knowledge about the Planning

Rules
- 10000 Technical Rules for Item Combinations, e.g.

\[ \text{IF } \text{Motor} = m_4 \ \text{AND} \ \text{Heating} = h_1 \]

\[ \text{THEN} \ \text{Generator} \in \{g_3, g_4, g_5\} \]

- Often 6-dimensional, sometimes more than 10 dimensions

- 500000 marketing oriented rules (with uncertainty)

- The rules are often changing

Data
- Specification of millions of built cars
Planning Tasks

**Calculation of part demands**
Compute the installation rate of a given item combination

**Simulation**
Analyze customers' preferences with respect to those persons who use a navigation system in a VW Polo

**Marketing and Sales stipulation**
Installation rate of Navigation system increase from 20% to 30%

**Capacity Restrictions**
Maximum availability of seat coverings in leather is 5000
How to manage the **uncertain** information about planning?

**Data**
- Vehicle orders
- Specifications of built vehicles (sample)

**Knowledge**
- Installation rules
- Combinability of properties

**Context**
- Model group
- Planning horizon

**Forecast**
- forecast/ set plan data
  - (frequencies, required quantities, capacities (restrictions), production plans, open purchase order quantities ...)

**Planning**

---

Warsaw, 25.09.2014
How to handle **uncertain** Information in AI Systems?

**Method 1 : Rule Based Approaches**

**Example Mycin (1975)** – Expert System for diagnosing bacterial infections

500 (uncertain) decision rules as knowledge base

If

1) the gram stain of the organism is gramneg, and
2) the morphology of the organism is rod, and
3) the aerobicity of the organism is anaerobic

then

there is suggestive evidence (0.6) that the identity of the organism is bacteroides
Certainty Factor $cf$: $-1 \leq cf \leq 1$

**Rules**

- $A \rightarrow B [0.80]$
- $C \rightarrow D [0.50]$
- $B \land D \rightarrow E [0.90]$
- $E \lor F \rightarrow G [0.25]$
- $H \rightarrow G [0.30]$

**Facts**

- $A [1.00]$
- $C [0.50]$
- $F [0.80]$
- $H [0.90]$

**Tree**
Rules for conjunction disjunction, serial und parallel combination combination, e.g.
$CF(B \land D) = \min\{CF(A), CF(B)\}$, $CF(D) = CF(C) \times CF(C \rightarrow D)$, $CF(G) = 0.27 + 0.2 - 0.27 \cdot 0.2 = 0.416$

Drawback of this method:
- No modelling of (in) dependences (CF ist truth functional !).
- Reasoning in one direction
- Manual Design of rules and certainty factors

Opinion at that time:
- Probabilistic solutions are too complex for big systems
The CF rule combination scheme is inconsistent in general.

Example: $\text{CF}(A) = 0.9$, $\text{CF}(D) = ?$

$\text{CF}(D) = 0.9 + 0.9 - 0.9 \cdot 0.9 = 0.99$

$\text{CF}(D) = 0.9$

Certainty factor is increased just because (the same) evidence is transferred over two different (parallel) paths!
Example Fuzzy Control

Automatic Gear Box for VW Beetle (1995), in series line

Seven Fuzzy rules evaluate the sportiveness of driver 12 times per second

The rules were found by fuzzy clustering.
How to find consistent rules in big applications?

Method 2  Data Driven Approaches: Big Data, Machine Learning, Deep Learning

Human Brain

Approx. 86 000 000 000 Neurons
1000-10000 inputs per neuron

Nature Inspired Models

Artificial Neuronal Networks
Learning Back-Propagation
Deep Learning
Lots of data, high performance computing, clever algorithms are needed

Typical Problems with pure data driven approaches: explainability, robustness, safety
Method 3  Hybrid Systems (Data + Knowledge)

Vision: A hybrid system accesses many sources and data, merges this information, filters and evaluates this information from the respective context, interacts with other systems and users, learns from information, uses knowledge about causalities, dependencies, associations, hits these base decisions, etc.
Our hybrid modeling approach for property planning

Installation Rates as Subjective Probabilities:
High dimensional finite probability space

Handling Complexity by using Decomposition:
Instead of one high dimensional global model use
several connected local low dimensional models

Implementation with Graph Based Probabilistic Models:
Relational, Bayes, and Markov Networks
How to find a suitable decomposition of a probability space?

A, B, C (Random) Variables  
A quality of ingredients  
B cook’s skill  
C meal quality

If C is not known, A and B are independent.

If C is known, then A and B become (conditionally) dependent given C.
If nothing is known about the restaurant success or meal quality or both, the cook’s skills and quality of the ingredients are unrelated, that is, independent.

However, if we observe that the restaurant has no success, we can infer that the meal quality might be bad.

If we further learn that the ingredients quality is high, we will conclude that the cook’s skills must be low, thus rendering both variables dependent.

Decomposition: \[ P(A,B,C,D) = P(A)P(B)P(C|A,B)P(D|C) \]
Separation in undirected graphs (u-Separation)

$Z$ u-separates $X$ from $Y$ if every path from $X$ to $Y$ is blocked by a node in $Z$.

In directed graphs the decomposition is more tricky.
Separation in directed Graphs (d-separation)

Z d-separates X from Y if all paths (pathes in reverse to arrows are allowed) from X to Y are blocked by a node in Z.

A node A is blocking a path, if its edge directions along the path - are of type 1 and $A \in Z$, or

- are of type 2 und neither A nor one of its descendants is in Z.

serial, head-to-tail
serial, head-to-tail
diverging, tail-to-tail

Typ 1

Typ 2

converging, head-to-head
A **Bayesian network** is a directed acyclic graphs (DAG) of random variables $X_v$ with the property

$$p(x) = \prod_{v \in V} p(x_v | x_{\text{pa}(v)})$$

$p(X)$ denotes the joint distribution of the random variables, $\text{pa}(v)$ is the set of parents of $v$ (i.e. those vertices pointing directly to $v$ via a single edge).

**Example**

$$P(X_1, \ldots, X_6) = P(X_6 | X_5) \cdot P(X_5 | X_2, X_3) \cdot P(X_4 | X_2) \cdot P(X_3 | X_1) \cdot P(X_2 | X_1) \cdot P(X_1)$$

**Theorem** If $Z$ d-separates $X$ und $Y$, then $X$ and $Y$ are conditionally independent with respect to $Z$. 
P(A,B,C,D) = P(A) \times P(B) \times P(C \mid A,B) \times P(D \mid C)
Modelling with Bayesian Networks

A is a possible cause for B and also an explanation for C. Both in turn could explain that D is true. E may also be associated with D.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Values</th>
<th>DAG</th>
<th>Probabilities</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>a1, a2</td>
<td><img src="image" alt="Diagram" /></td>
<td>$P(e_1 \mid c_1) = 0.8$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$P(e_1 \mid c_2) = 0.6$</td>
</tr>
<tr>
<td>B</td>
<td>b1, b2</td>
<td></td>
<td>$P(d_1 \mid b_1, c_1) = 0.8$</td>
</tr>
<tr>
<td>C</td>
<td>c1, c2</td>
<td></td>
<td>$P(d_1 \mid b_1, c_2) = 0.8$</td>
</tr>
<tr>
<td>D</td>
<td>d1, d2</td>
<td></td>
<td>$P(d_1 \mid b_2, c_1) = 0.8$</td>
</tr>
<tr>
<td>E</td>
<td>e1, e2</td>
<td></td>
<td>$P(d_1 \mid b_2, c_2) = 0.05$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$P(b_1 \mid a_1) = 0.8$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$P(b_1 \mid a_2) = 0.2$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$P(c_1 \mid a_1) = 0.2$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$P(c_1 \mid a_2) = 0.05$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$P(a_1) = 0.2$</td>
</tr>
</tbody>
</table>
Reasoning with Bayesian Networks

E is observed. The „Belief“ in D changes.

\[
P(A, B, C, D, E) = P(A) P(B | A) P(C | A) P(D | BC) P(E | C)
\]

\[
P(d_1) = 0.3200, \quad P(d_1 | e_1) = 0.3325
\]
There are many tools for handling BN‘s.
Free Download of the state of the art BN-System Hugin Lite 8.9
How to find efficient algorithms for large Bayesian Networks?

**Solution**: Global information can be shared locally by every entity. New knowledge is distributed by message passing.

Example: How many people are we?
How many people are we?

Forward propagation

backward propagation
**Problem:** How to avoid cycles in message pathing (as in Mycin)

**Solution:** Decomposition of the Probability space

- A singly connected structure is obtained by triangulating the graph and then forming a tree of maximal cliques, the so-called **join tree**.
- For evidence propagation a join tree is enhanced by so-called **separators** on the edges, which are intersection of the connected nodes → **junction tree**.
**Consistent Decomposition**

- Distributions only over the cliques (conditional independencies)
- Validity of all combinations registered and available

**Fast Propagation in Join Tree**

- Propagation by traversing the net twice (collect & distribute)

**Efficient planning**

- Planning requirements as conditional distributions
- Calculation of frequencies for arbitrary property combinations (focusing analyses, simulations) in real time
Example: Network for VW Bora

186 variables, 174 cliques
- Assistant System for Handling Estimates for Installation Rates
- Answers to the Questions in real time (seconds)
- Different Model Groups and Different Planning Intervals
- This daily use: 5000 networks (planning scenarios) handled by 350 planners worldwide
- Realization by a CI Group spin off, Leader Jörg Gebhardt
# Example for one planning week

<table>
<thead>
<tr>
<th>Data volume (cumulative per week)</th>
<th>Across all 166 Model groups</th>
<th>Single Model group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Markov Nets</td>
<td>75.787</td>
<td>1.054</td>
</tr>
<tr>
<td>Total size (compressed)</td>
<td>79.9 GB</td>
<td>3.8 GB</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>… for one planning interval</th>
<th>Model group 1</th>
<th>Model group 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Planning requirement</td>
<td>4.424</td>
<td>1.299</td>
</tr>
<tr>
<td>Number of variables</td>
<td>203</td>
<td>204</td>
</tr>
<tr>
<td>Number of cliques</td>
<td>174</td>
<td>156</td>
</tr>
<tr>
<td>Network size (compressed, RAM: ~Factor 10)</td>
<td>8.5 MB</td>
<td>17.9 MB</td>
</tr>
<tr>
<td>Largest clique (only positive probabilities)</td>
<td>130.806 tupels 9 variables</td>
<td>1.489.515 tupels 14 variables</td>
</tr>
</tbody>
</table>
How to learn new Planning Models?

Context: Mode group, Planning interval

History
context dependant samples of customer chosen vehicle specifications

Technique
Context dependent rules of Combinability of properties in planning interval

Estimate a-priori distribution of frequencies

Quantitative Learning
Markov net with structure of the relational net

Structural Learning
Transformation to a relational Net with hyper tree structure

Fusion

Planning Model
fused consistent Markov Net for Property planning
How to revise a Planning Model?

Prior Probability Distribution

New dimensional Conditional Probabilities: Marketing Stipulations, Capacity Restrictions

**REVISION**

*Principle of minimal Change*

Posterior Probability Distribution including *specified and inferred* changes.

Information-theoretically closest to the prior distribution
Revision Operator

- Iterative Proportional Fitting (Biproportional Fitting, RAS Algorithm, Matrix Scaling)

- Algorithms for Adapting the Marginal Distributions

- Stepwise Modification of the Probability Distribution

- Process converge for Non-Contradicting Revision Statements

- Approx. 7000 Revision Assignments per Week and Model
Inconsistencies

Complex Structure

Many Changes

Inconsistencies

Outer Inconsistencies
Revision assignments inconsistent with zero-values in prior distribution

Inner Inconsistencies
Revision assignments are inconsistent, independent of prior distribution
Inconsistencies: Examples

It is not easy to configure revision statements without creating inconsistencies!
How to explain the user how to change his desired revision statements?
Planning Operation: Revision

**Historical Data**
Context dependent sample of customer chosen vehicle specifications

Estimate a-priori distribution of frequencies

**Set plan data**
(forecast/frequencies, required quantities, capacities, restrictions, production plans, open purchase order quantities …)

**Planning Model**
Consistent Markov-Net for Planning (Model group, Planning Interval)

**Planning**

**Inconsistency Management**
Decision Support

**Revision**
Principle of Minimal Change
What comes next?

Associations (“observing”) Bayes, belief, conditional probability,…
What does a symptom say about a disease? How would observing X change my belief in Y?

Intervention (“doing”) Do-Operator, model revision,…
If I take aspirin, will my headache be cured? What would Y be if I did X?

Causality („understanding“) Counterfactuals,…
Was it X that caused Y? What if X hadn't occurred? What if I had acted differently? Did Aspirin Stop My Headache? What if I hadn't smoked in the last 2 years?

Reference
Judea Pearl and Dana Mackenzie, The Book of Why: The New Science of Cause and Effect, 2018