Evidenzpropagation in Bayes-Netzen und Markov-Netzen

Thomas Thüm

20. Juni 2006

Übersicht

1. Einleitung
   - Motivation
   - Einordnung der Begriffe
2. Evidenzpropagation
   - Schlussfolgerungsnetze
   - Ein Beispiel für eine Zerlegung
   - Schlussfolgern mit Projektionen
   - Zerlegungen mit Wahrscheinlichkeiten
3. Markov-Netze und Bayes-Netze
   - Bedingte Unabhängigkeit
   - Markov-Netz
   - Bayes-Netz
4. Zusammenfassung
   - Ausblick und Zusammenfassung
   - Literaturverzeichnis

Motivation

In vielen Fällen ist unser Wissen:

- unvollständig
  "Der Zeuge hat das Kennzeichen nicht gesehen."
- unscharf, unpräzise
  "Das Auto war grün."
- unsicher
  „Ich bin mir nicht sicher, ob es wirklich ein Audi war."

Wir müssen trotzdem agieren!

- Schließen bei Unschärfe $\Rightarrow$ Fuzzy-Logik
- Schließen unter Unsicherheit
  $\Rightarrow$ Schließen über Eintretenswahrscheinlichkeiten
  $\Rightarrow$ Schlussfolgerungsnetze

Einordnung der Begriffe

Andrej Markov (1856-1922)

probabilistische
Schlussfolgerungsnetze
(auch graphische Modelle)

Thomas Bayes (1702-1761)

Markow-Netze
(auch Markov Networks)

Bayes-Netze
(auch Bayessches-Netze
oder Bayesian Networks)

Mehrdimensionaler Raum

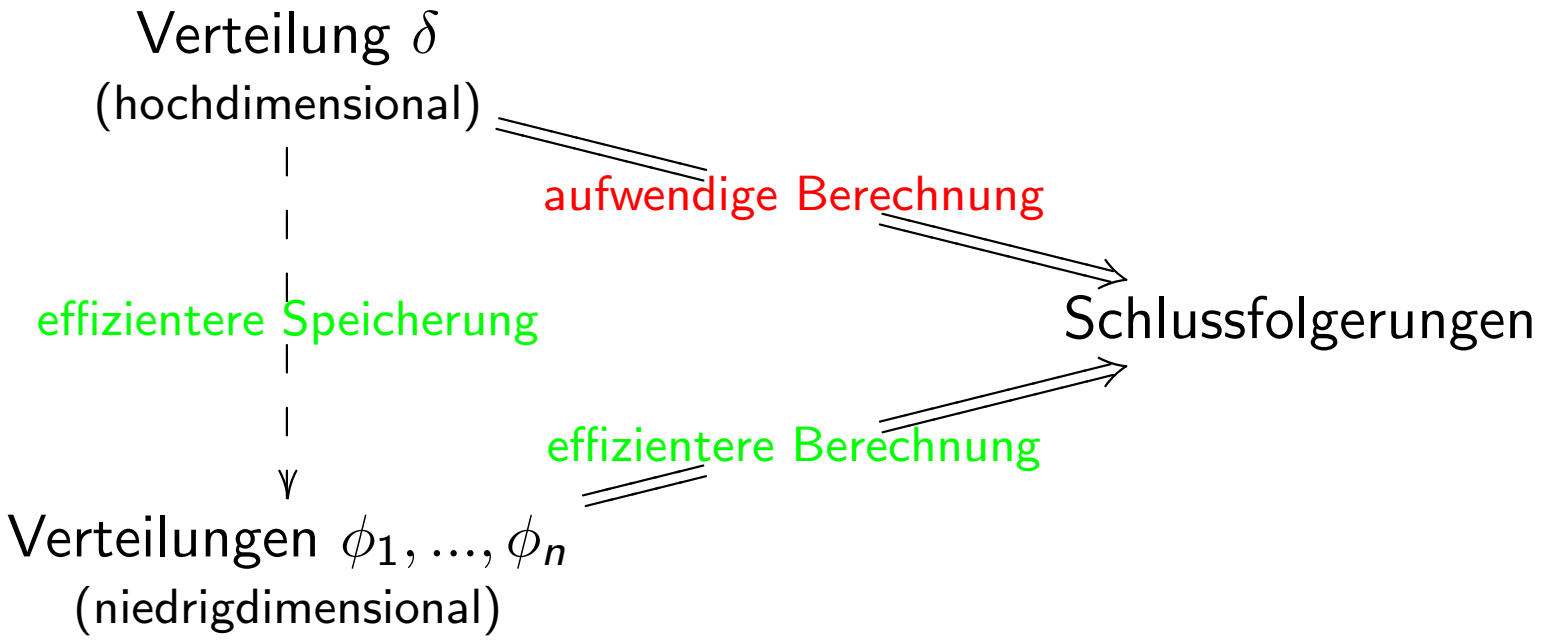
- Modellierung eines Weltausschnitts durch einen mehrdimensionalen Raum
  $\Rightarrow$ Beispiel: alle deutschen Autos
- Dimension $\Leftrightarrow$ Attribut
  $\Rightarrow$ Hersteller, Farbe, ...
- Attribute: erschöpfend, sich gegenseitig ausschließend
  $\Rightarrow$ jedes Auto hat genau einen Hersteller
- Fahrzeug $\Leftrightarrow$ Punkt im mehrdimensionalen Raum
  - mehrere Fahrzeuge $\Leftrightarrow$ ein Punkt
    $\Rightarrow$ alle Attribute sind gleich
  - ein Punkt $\Leftrightarrow$ kein Fahrzeug
    $\Rightarrow$ Fahrzeug mit den Attributen wurde nicht verkauft

Verteilung im mehrdimensionalen Raum

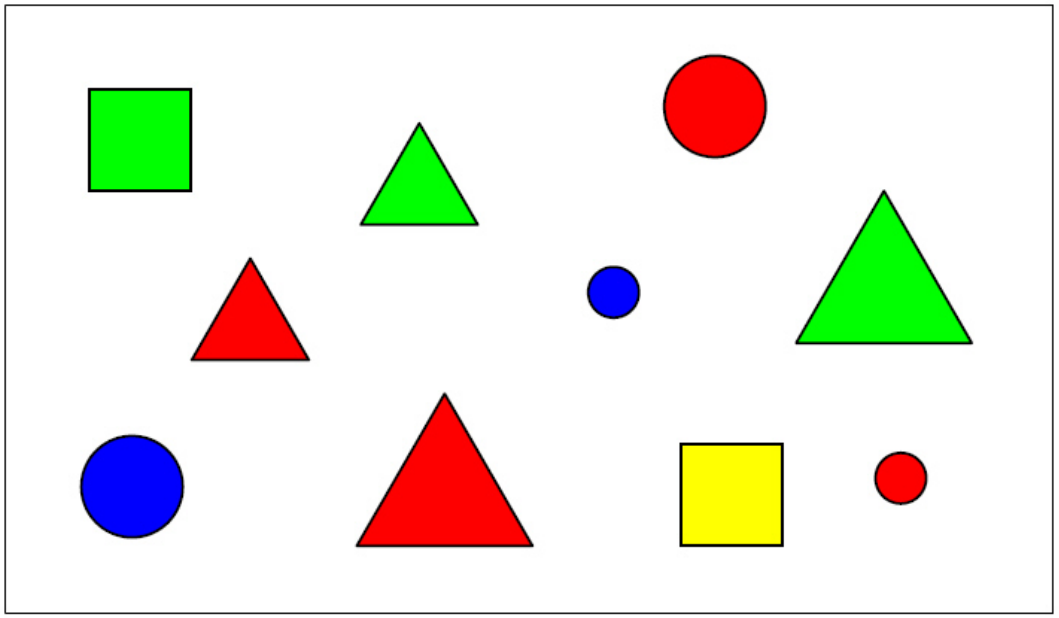
- wir betrachten nur Wahrscheinlichkeitsverteilungen
- ordnet jedem Punkt eine Zahl aus dem Intervall $[0, 1]$ zu
- gibt die Wahrscheinlichkeit an, dass der entsprechende Zustand vorliegt
- Daten werden geschätzt oder gehen aus statistischen Analysen hervor
- z.B. relative Häufigkeit mit der Fahrzeuge verkauft wurden

Zerlegung der Verteilung

Eine Verteilung δ auf einem Raum R kann in Verteilungen $\phi_1, ..., \phi_n$ auf niedrigdimensionalen Unterräumen von R (näherungsweise) zerlegt und wieder rekonstruiert werden.



Ein einfaches Beispiel

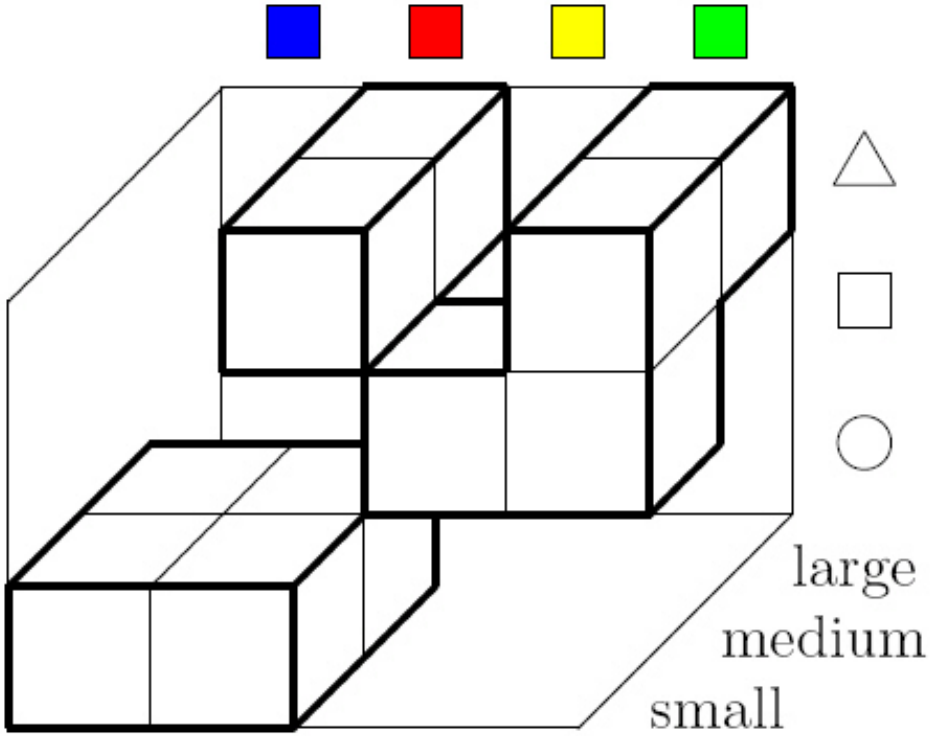


color	shape	size
■	○	small
■	○	medium
■	○	small
■	○	medium
■	△	medium
■	△	large
■	□	medium
■	□	medium
■	△	medium
■	△	large

- 10 einfache geometrische Objekte, 3 Attribute

Die Relation im Folgerungsraum

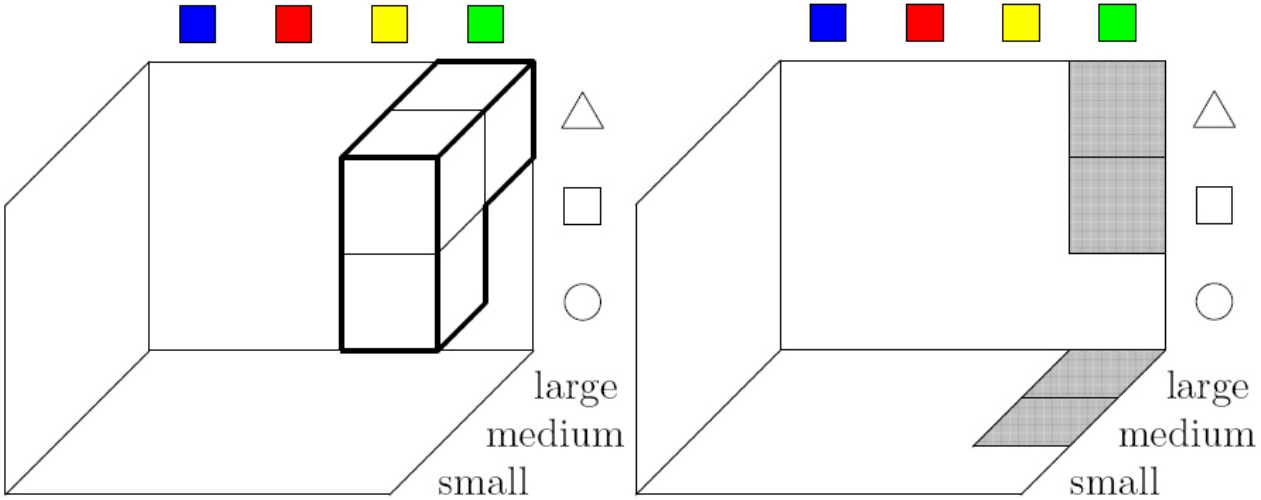
color	shape	size
■	○	small
■	○	medium
■	○	small
■	○	medium
■	△	medium
■	△	large
■	□	medium
■	□	medium
■	△	medium
■	△	large



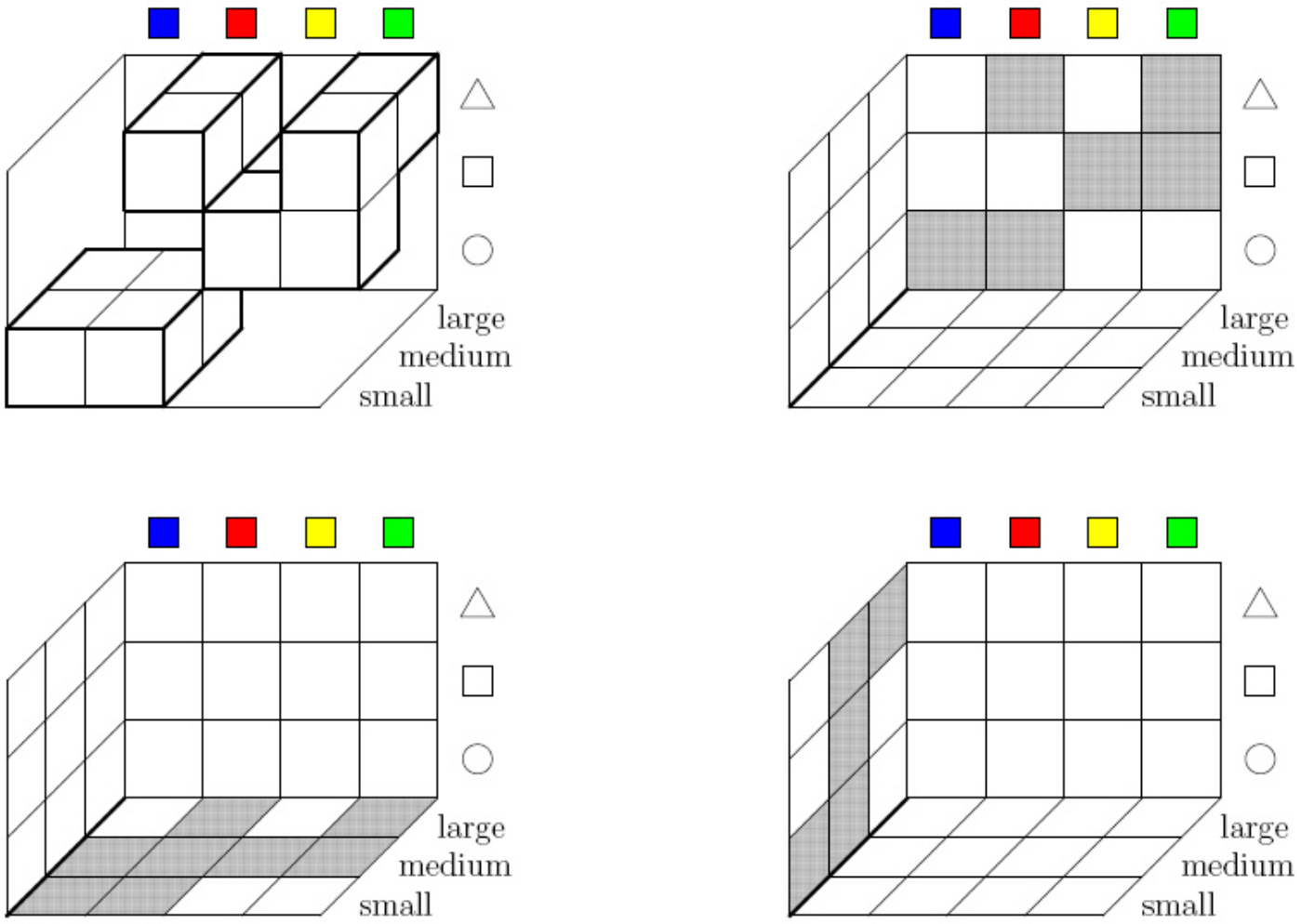
- Jeder Kubus repräsentiert ein Tupel.

Schlussfolgern

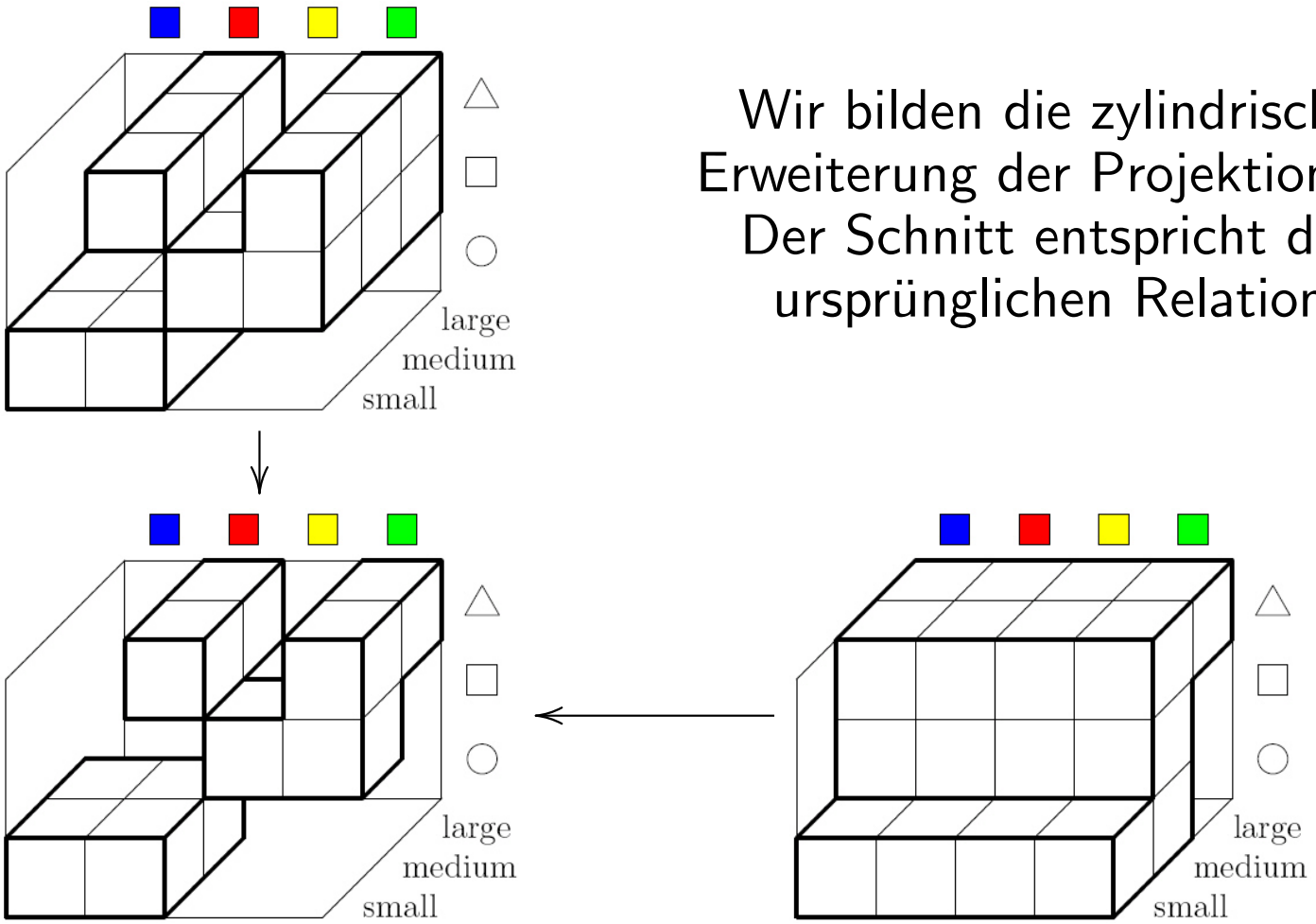
- Nehmen wir an, wir wissen das Objekt sei grün. Diese Information reduziert den Raum der möglichen Attributkombinationen.
- Aus diesem Wissen folgt, dass das Objekt ein Dreieck oder Quadrat sein muss und nicht klein sein kann.



A-priori-Wissen und Projektionen



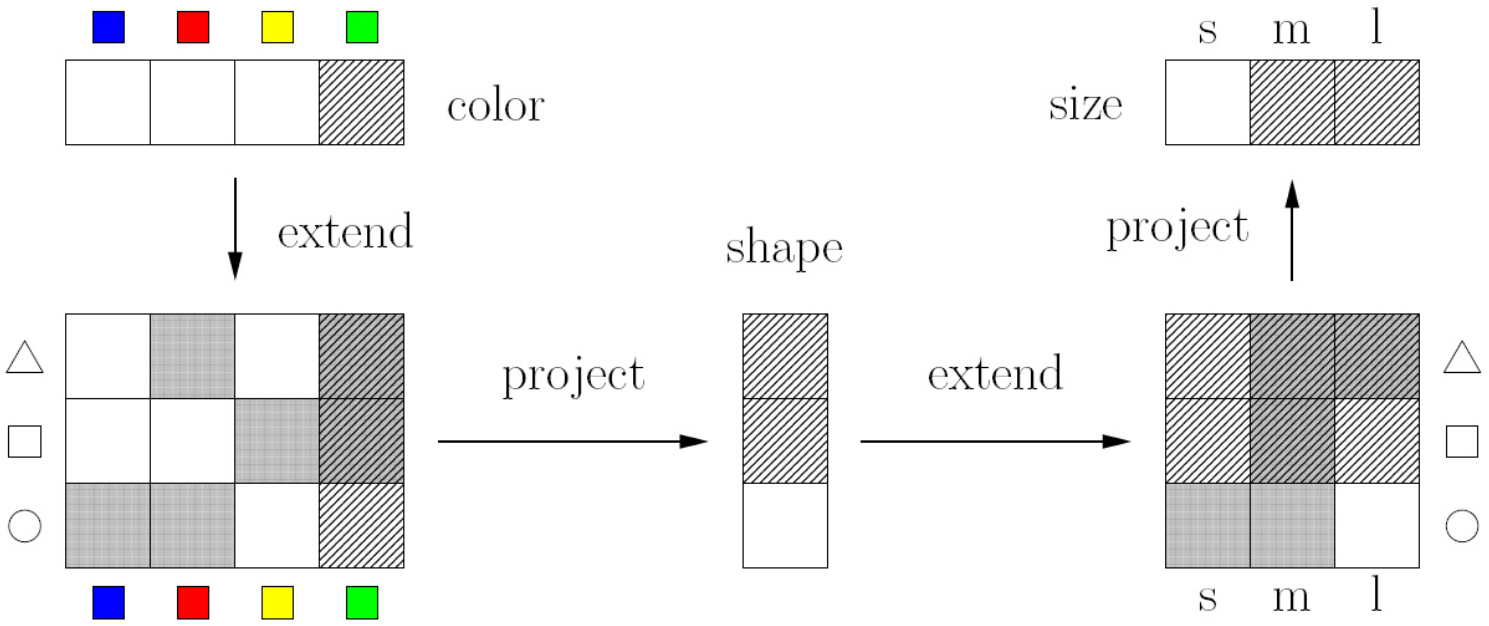
Zylindrische Erweiterung und ihr Durchschnitt



Wir bilden die zylindrische Erweiterung der Projektionen. Der Schnitt entspricht der ursprünglichen Relation

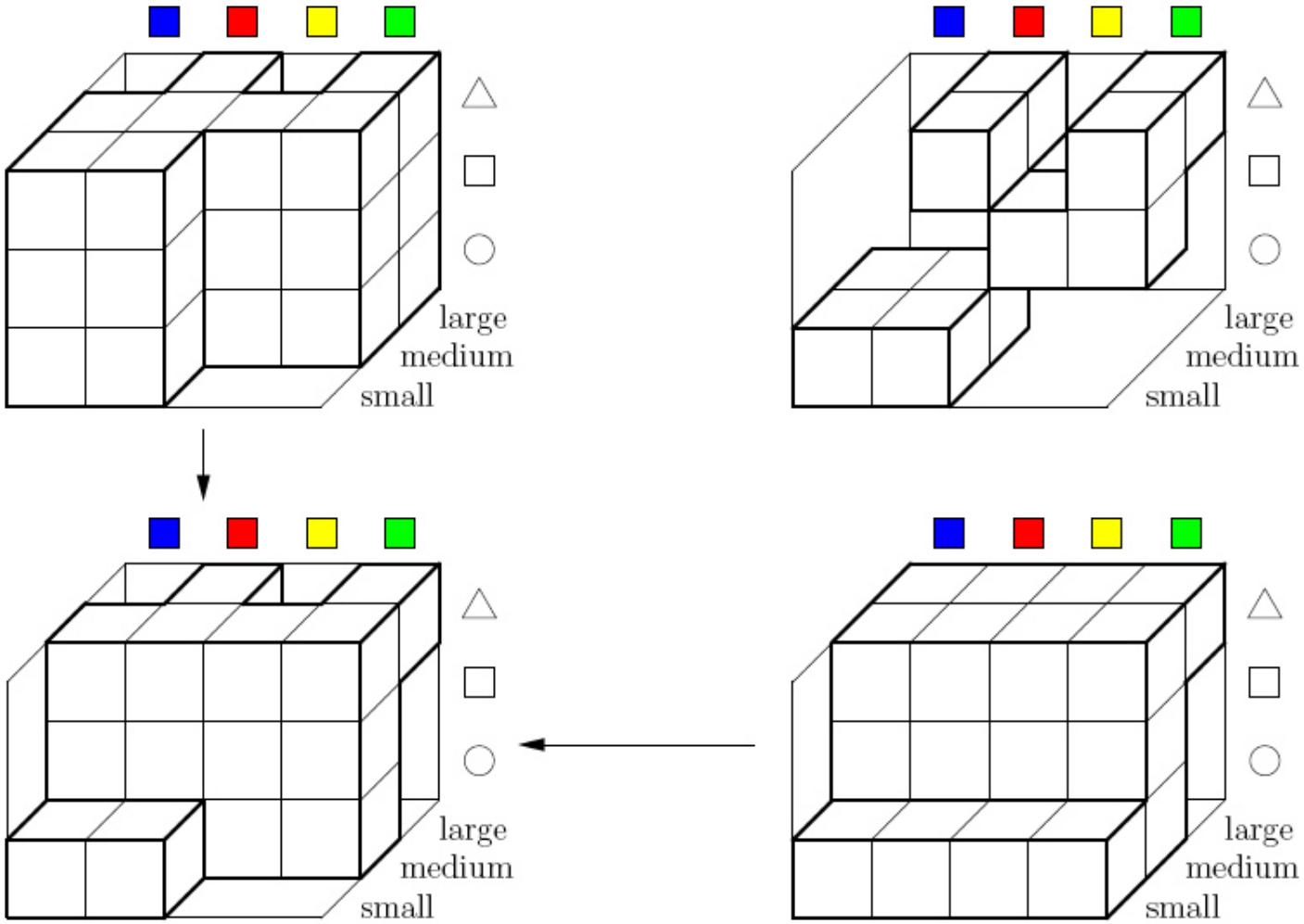
Schlussfolgern mit Projektionen

Das gleiche Ergebnis kann mit den Projektionen berechnet werden ohne das der 3-dimensionale Raum rekonstruiert werden muss:

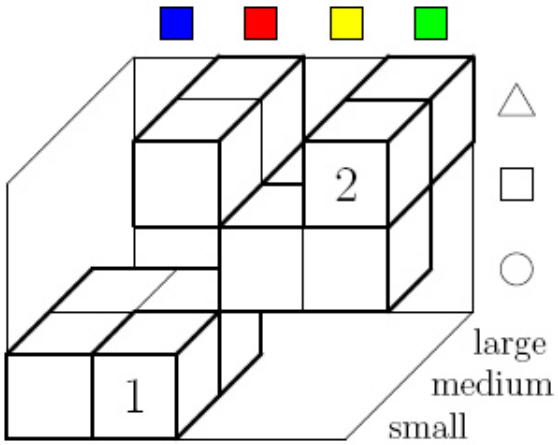


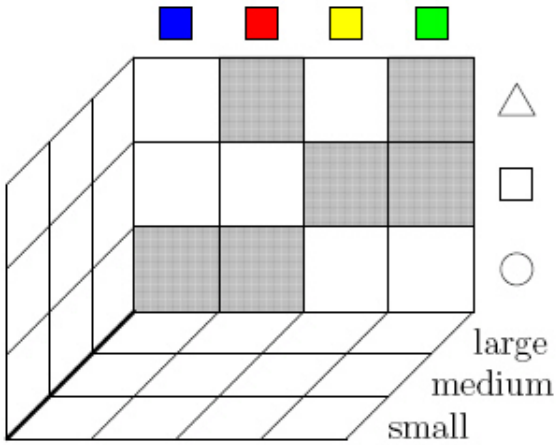
Das induziert folgendes Netzwerk:

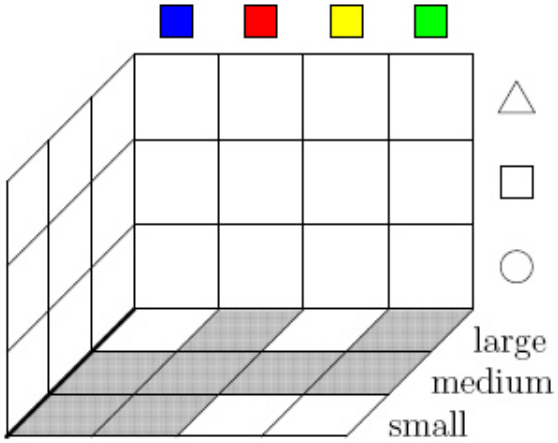
Andere Projektionen

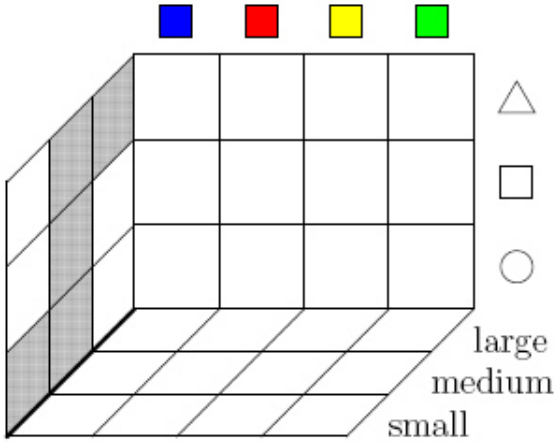


Zerlegung immer möglich?









Eine Wahrscheinlichkeitsverteilung

all numbers in parts per 1000

220	330	170	280

				△	400
20	90	10	80	□	240
2	1	20	17	○	360
28	24	5	3		

large

300

18	81	9	72
8	4	80	68
56	48	10	6

medium

460

△	2	9	1	8
□	2	1	20	17
○	84	72	15	9

small

240

△	40	180	20	160
□	12	6	120	102
○	168	144	30	18

	s	m	l
△	20	180	200
□	40	160	40
○	180	120	60

large	50	115	35	100
medium	82	133	99	146
small	88	82	36	34

Schlussfolgern - Bedingte Wahrscheinlichkeiten berechnen

0	0	0	1000

0	0	0	286	△	572
0	0	0	61	□	364
0	0	0	11	○	64

large

358

0	0	0	257
0	0	0	242
0	0	0	21

medium

520

△	0	0	0	29
□	0	0	0	61
○	0	0	0	32

small

122

△	0	0	0	572
□	0	0	0	364
○	0	0	0	64

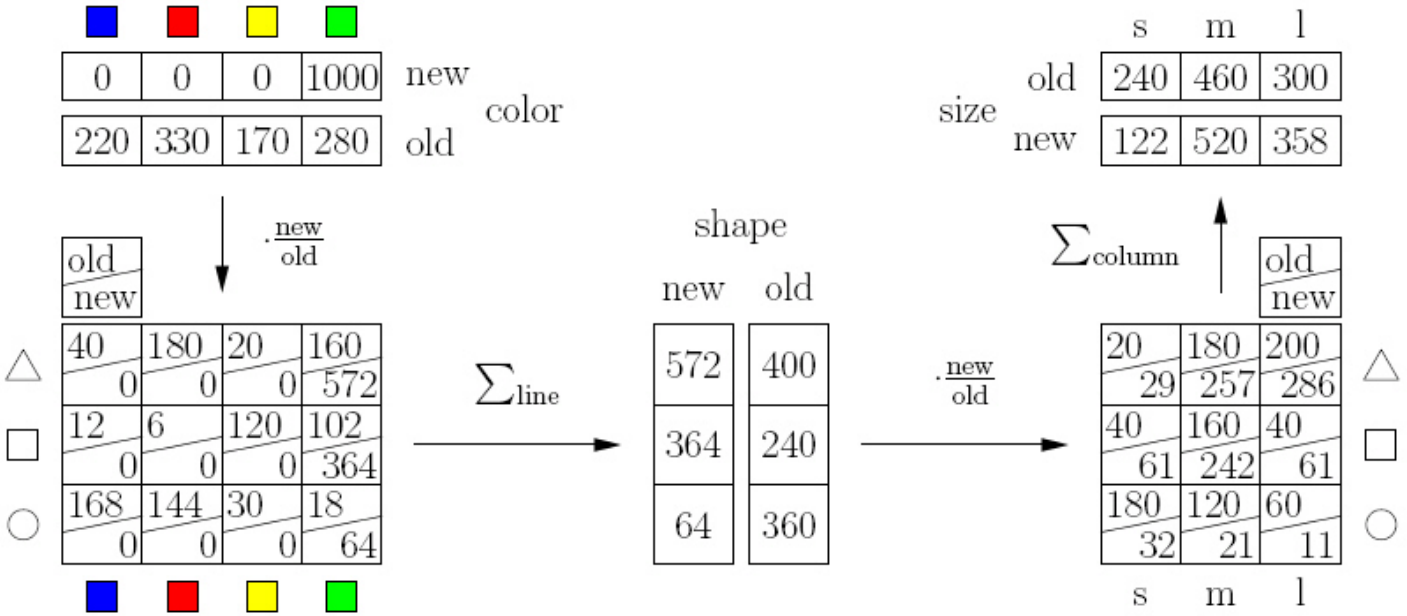
all numbers in parts per 1000

	s	m	l
△	29	257	286
□	61	242	61
○	32	21	11

large	0	0	0	358
medium	0	0	0	531
small	0	0	0	111

Schlussfolgern mit Projektionen

Das gleiche Ergebnis kann mit den Projektionen berechnet werden ohne das der 3-dimensionale Raum rekonstruiert werden muss:



Das induziert folgendes Netzwerk:

Bedingte Unabhängigkeit

Warum funktioniert die Evidenzpropagation im Beispiel?

- **graphentheoretisch:**
  weil der einzige Pfad zwischen A und C durch B geblockt wird
- **wahrscheinlichkeitstheoretisch:**
  weil A *bedingt unabhängig* von C gegeben das Attribut B ist
- **formal:**
  $P(a|c \wedge b) = P(a|b)$ wobei b die Evidenz ist
- es gilt auch die Umkehrung:
  $P(c|a \wedge b) = P(c|b)$
- vgl. bedingte Wahrscheinlichkeit:
  $P(x|y) = P(x)$ gdw. x unabhängig von y ist

Markov-Netz

Zerlegung eines ungerichteten Graphen

Sei δ die Wahrscheinlichkeitsverteilung über einer Menge von Attributen $U = \{A_1, ..., A_r\}$ und (U, E) der Graph G

1. maximale Cliquen $N = \{M_1, ..., M_s\}$ ermitteln
2. Funktionen ϕ_i suchen, sodass

$$\delta(\bigwedge_{A_i \in U} A_i = a_i) = \prod_{M_i \in N} \phi_i(\bigwedge_{A_i \in M_i} A_i = a_i)$$

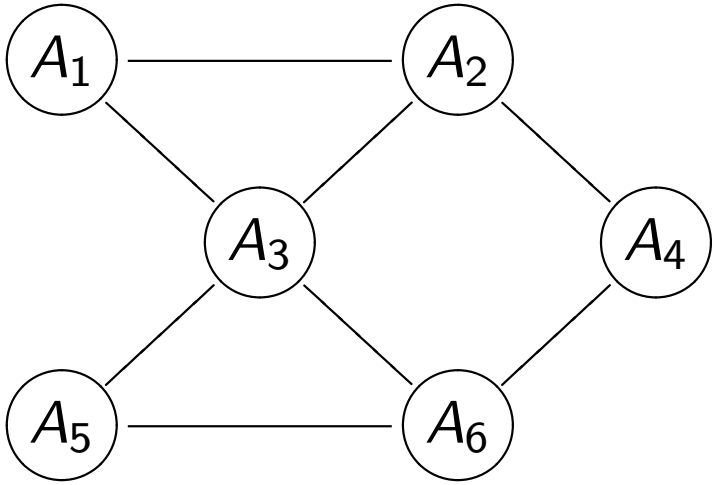
wobei $\forall a_i \in dom(A_i)$ und $dom(A)$ der Wertebereich des Attributes A ist

- δ heißt *zerlegbar* (faktorisierbar) genau dann, wenn die Funktionen ϕ_i existieren
- G heißt *bedingter Unabhängigkeitsgraph* für δ

G und ϕ_i bilden ein *Markov-Netz*

Markov-Netz

Beispielgraph:



1. maximale Cliquen: $M_1 = \{A_1, A_2, A_3\}$, $M_2 = \{A_3, A_5, A_6\}$, $M_3 = \{A_2, A_4\}$, $M_4 = \{A_4, A_6\}$
2. Zerlegung:

$$\delta(A_1 = a_1 \wedge \ldots \wedge A_6 = a_6) = \phi_1(A_1 = a_1 \wedge A_2 = a_2 \wedge A_3 = a_3) \cdot \phi_2(\ldots) \cdot \phi_3(\ldots) \cdot \phi_4(\ldots)$$

Bayes-Netz

Zerlegung eines gerichteten azyklischen Graphen (DAG)

Sei δ die Wahrscheinlichkeitsverteilung über einer Menge von Attributen $U = \{A_1, ..., A_r\}$ und $(U, \vec{E})$ der Graph $\vec{G}$

- Wir betrachten nun die folgende Gleichung

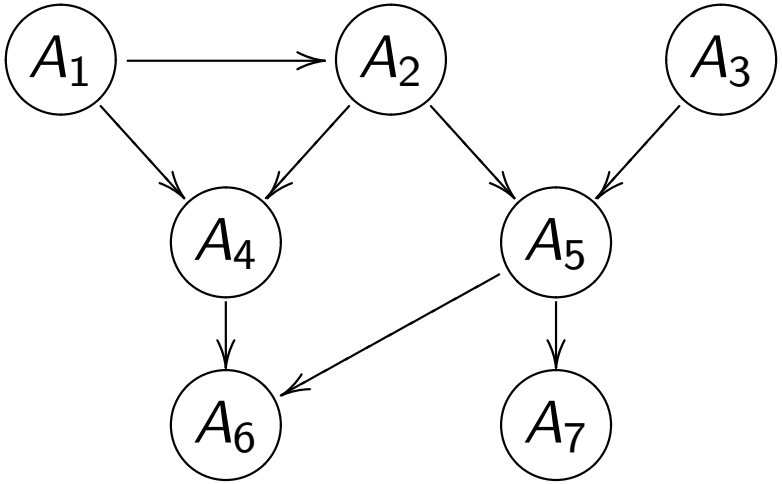
$$\delta(\bigwedge_{A_i \in U} A_i = a_i) = \prod_{A_i \in U} P(A_i = a_i | \bigwedge_{A_j \in parents(A_i)} A_j = a_j)$$

  wobei $\forall a_i \in dom(A_i)$ und $parents(B) = \{C | (C, B) \in \vec{E}\}$ gilt
- vgl. Produktsatz: $P(x \wedge y) = P(x|y)\dot{P}(y)$
- δ heißt *zerlegbar* (faktorisierbar) genau dann, wenn die obige Gleichung erfüllt ist
- $\vec{G}$ heißt *bedingter Unabhängigkeitsgraph* für δ

$\vec{G}$ und die bedingten Wahrscheinlichkeitsverteilungen bilden ein *Bayes-Netz*

Bayes-Netz

Beispielgraph:



$$\delta(A_1 = a_1 \wedge A_2 = a_2 \wedge \ldots \wedge A_6 = a_6) =$$
$$P(A_1 = a_1) \cdot P(A_2 = a_2 | A_1 = a_1) \cdot P(A_3 = a_3)$$
$$\cdot P(A_4 = a_4 | A_1 = a_1 \wedge A_2 = a_2) \cdot P(A_5 = a_5 | A_2 = a_2 \wedge A_3 = a_3)$$
$$\cdot P(A_6 = a_6 | A_4 = a_4 \wedge A_5 = a_5) \cdot P(A_7 = a_7 | A_5 = a_5)$$

Ausblick

Anwendungen für Schlussfolgerungsnetze:

- **Medizinische Diagnose**
  Wissen über Abhängigkeiten zwischen Symptomen
  $\Rightarrow$ Krankheiten identifizieren
- **Wahrscheinlichkeitstheoretische Wettervorhersage**
  Langjährige, statistische Daten über Niederschlag in verschiedenen Orten + aktuelle Daten von einem Ort
  $\Rightarrow$ Niederschläge in übrigen Orten

Arten der Evidenzpropagation:

- **exakt:** numerische Berechnung
- **aproximativ:** wenn kein exaktes Verfahren bekannt ist oder zu hohen Aufwand hat
- **symbolisch:** es wird eine Funktion von Symbolen (Variablen) ermittelt

Zusammenfassung

- Schlussfolgerungsnetze
  $\Rightarrow$ Schließen unter Unsicherheit
- Unsicherheit wird meist durch Wahrscheinlichkeiten modelliert
- praktische Anwendungen
  $\Rightarrow$ sehr große, hochdimensionale Verteilungen
  $\Rightarrow$ sind nicht zum Schlussfolgern geeignet (zu aufwendig)
  $\Rightarrow$ daher: mehrere kleine, niedrigdimensionale Verteilungen
- Evidenzpropagation transportiert Wissen zwischen niedrigdimensionalen Verteilungen
- ein Markov-Netz ist ein ungerichteter Graph und ein Bayes-Netz ein gerichteter azyklischer Graph

Literaturverzeichnis

C. Borgelt, H. Timm, and R. Kruse.
Graphical Models - Methods for Data Analysis and Mining.
2002.

Cofino, Cano, Sordo, and Gutierrez.
Bayesian Networks for Probabilistic Weather Prediction.
Technical report, 15th European Conference on Artificial Intelligence, 2002.

E. Castillo, J. M. Gutierrez, and A. S. Hadi.
Expert Systems and Probabilistic Network Models.
1997.

Werner Dilger.
Unsicherheit und Probabilistische Schlussfolgerungssysteme.
Technical report, Fakultät für Informatik, Technische Universität Chemnitz, April 2006.