
12. Verarbeitung unsicheren Wissens

Prof. Dr. Rudolf Kruse

University of Magdeburg

Faculty of Computer Science

Magdeburg, Germany

rudolf.kruse@cs.uni-magdeburg.de



12. Verarbeitung unsicheren Wissens

Inhalt:

- Motivation
- Grundlagen der Wahrscheinlichkeitstheorie
- Probabilistische Inferenz

Motivation

In vielen Fällen ist unser Wissen über die Welt

- **unvollständig: Wissensumfang ist nicht hinreichend**
 - (z.B.: „Ich weiß nicht, wie der Bus an Feiertagen fährt, da ich nur werktags fahre!“)
- **unscharf: Wissen ist nicht präzise genug**
 - (z.B.: „Der Bus fährt ungefähr zu jeder vollen Stunde!“)
- **unsicher: Wissen ist unzuverlässig**
 - (z.B.: „Der Bus fährt wahrscheinlich jede volle Stunde.“)

Wir müssen trotzdem agieren!

- **Schließen unter Unsicherheit**
- **Schließen über Eintretenswahrscheinlichkeiten**
- **... und Kosten/Nutzen**

Beispiel

- Ziel: Um 9:15 Uhr in der Uni sein, um an einer Vorlesung teilzunehmen.
- Es gibt mehrere Pläne, um das Ziel zu erreichen:
 - P_1 : Um 8:00 aufstehen, um 8:55 Haus verlassen, um 9:00 den Bus ...
 - P_2 : Um 7:30 aufstehen, um 8:25 Haus verlassen, um 8:30 den Bus ...
 - ...
- Alle Pläne sind *korrekt*, aber
 - sie implizieren verschiedene *Kosten* und verschiedene *Wahrscheinlichkeiten* das Ziel *tatsächlich* zu erreichen.
 - P_2 wäre der Plan der Wahl, da Vorlesungen wichtig ist, und die Erfolgsrate bei P_1 nur bei ca. 85 - 95% liegt.

Unsicherheiten bei Regeln (1)

Beispiel: Es sei ein einfaches Diagnose-Expertensystem für Zahnärzte gegeben, das z.B. folgende Regel beinhaltet:

$$\forall p [\text{Symptom}(p, \text{toothache}) \Rightarrow \text{Disease}(p, \text{cavity})]$$

■ Diese Regel ist inkorrekt! Besser wäre:

$$\forall p [\text{Symptom}(p, \text{toothache}) \Rightarrow \\ \text{Disease}(p, \text{cavity}) \vee \text{Disease}(p, \text{gum_disease}) \vee \dots]$$

... aber wenn wir gar nicht alle Ursachen kennen?

Vielleicht besser die *kausale* Regel?

$$\forall p [\text{Disease}(p, \text{cavity}) \Rightarrow \text{Symptom}(p, \text{toothache})]$$

■ Auch nicht korrekt!

Unsicherheiten bei Regeln (2)

Probleme mit der Logik:

- **Wir können nicht immer alle möglichen Ursachen aufzählen, und selbst wenn . . .**
- **Wir kennen nicht die Gesetzmäßigkeiten (in der Medizin), und selbst wenn . . .**
- **Es bleibt die Unsicherheit über den Patienten bestehen:**
 - **Karies und Zahnschmerzen können zufällig gleichzeitig auftreten.**
 - **Wurden alle Untersuchungen durchgeführt - und wenn ja: richtig?**
 - **Hat der Patient alle Fragen beantwortet - und wenn ja: angemessen?**
- **Ohne perfektes Wissen keine korrekten logischen Regeln!**

Unsicherheit bei Fakten

Beispiel: Nehmen wir an, wir wollten die Lokalisation eines Roboters durch ortsfeste Landmarken unterstützen. Aus dem Vorhandensein von Landmarken können wir auf den Raum schließen.

Problem: Sensoren sind ungenau!

- Aus der Tatsache, dass eine Landmarke erkannt wurde, kann man nicht mit Sicherheit schließen, dass der Roboter sich in dem entsprechenden Raum befindet.
- Analoges gilt, falls eine Landmarke nicht wahrgenommen wird.
- Es wird lediglich die Wahrscheinlichkeit erhöht oder erniedrigt.

Grade der Überzeugung

- Wir (oder andere Agenten) sind von Regeln und Fakten nur bis zu einem gewissen Grad überzeugt.
- Eine Möglichkeit, Grad der Überzeugung auszudrücken, besteht in der Benutzung von Wahrscheinlichkeiten.
- Der Agent ist von der Sensorinformation zu 0,9 überzeugt heißt dann: der Agent glaubt in 9 von 10 Fällen an die Richtigkeit der Information.
- Wahrscheinlichkeiten fassen die „Unsicherheit“ bedingt durch Unwissen zusammen.
- Wahrscheinlichkeiten sind nicht mit *Unschärfe (Vagheit)* zu verwechseln (→ Fuzzy Logik).
- Das Prädikat *groß* ist unscharf. Die Aussage „*Das könnte Peters Uhr sein.*“ ist unsicher.

Rationale Entscheidungen unter Unsicherheit

- Wir haben verschiedene *Aktionen* oder *Pläne* zur Auswahl
- Diese können mit verschiedenen *Wahrscheinlichkeiten* zu verschiedenen Ergebnissen führen.
- Die *Aktionen* verursachen verschiedene – ggf. subjektive – Kosten
- Die *Ergebnisse* haben verschiedenen – ggf. subjektiven – Nutzen
- Rational wäre es, die Aktion zu wählen, die den größten zu erwartenden Gesamtnutzen **hat!**

Entscheidungstheorie = Nutzentheorie + Wahrscheinlichkeitstheorie

Entscheidungstheoretischer Agent

function **DT-AGENT**(*percept*) returns **an action**

static: **a set probabilistic beliefs about the state of the world**

calculate updated probabilities for current state based on

available evidence including current percept and previous action

calculate outcome probabilities for actions,

given action descriptions and probabilities of current states

select *action* with highest expected utility

given probabilities of outcomes and utility information

return *action*

Entscheidungstheorie: Ein Agent ist rational genau dann, wenn er die Aktion wählt, die den größten zu erwartenden Nutzen gemittelt über alle möglichen Ergebnisse von Aktionen hat.

Unbedingte Wahrscheinlichkeiten (1)

$P(A)$ bezeichnet die unbedingte oder a priori Wahrscheinlichkeit, dass A eintreten wird im Fall, dass *keine* zusätzliche Information verfügbar ist, z.B.

$$P(Cavity) = 0.1$$

$Cavity$ ist eine Proposition (Aussage).

A priori Wahrscheinlichkeiten gewinnt man durch statistische Analyse oder aus allgemeinen Regeln.

Unbedingte Wahrscheinlichkeiten (2)

- Im Allgemeinen kann eine Zufallsvariable **nicht** nur die Werte *wahr* und *falsch*, sondern mehrere Werte annehmen:

$$P(\textit{Weather} = \textit{Sunny}) = 0.7$$

$$P(\textit{Weather} = \textit{Rain}) = 0.2$$

$$P(\textit{Weather} = \textit{Cloudy}) = 0.08$$

$$P(\textit{Weather} = \textit{Snow}) = 0.02$$

$$P(\textit{Headache} = \textit{TRUE}) = 0.1$$

- Propositionen können auch Gleichungen über Zufallsvariablen enthalten.
- Logische Konnektoren können zur Bildung von Propositionen verwendet werden, z.B.

$$P(\textit{Cavity} \wedge \neg \textit{Insured}) = 0.06$$

Unbedingte Wahrscheinlichkeiten (3)

$P(X)$ bezeichnet den Vektor der Wahrscheinlichkeiten für den (geordneten) Wertebereich der Zufallsvariable X :

$$P(\text{Headache}) = \langle 0.1, 0.9 \rangle$$

$$P(\text{Weather}) = \langle 0.7, 0.2, 0.08, 0.02 \rangle$$

definieren die Wahrscheinlichkeitsverteilung der Zufallsvariablen *Headache* und *Weather*.

$P(\text{Headache}, \text{Weather})$ ist eine 4 x 2 Tabelle von Wahrscheinlichkeiten aller Kombinationen der Werte einer Zufallsvariablen.

	Headache = TRUE	Headache = FALSE
Weather = Sunny	$P(W = \text{Sunny} \wedge \text{Headache})$	$P(W = \text{Sunny} \wedge \neg \text{Headache})$
Weather = Rain		
Weather = Cloudy		
Weather = Snow		

Bedingte Wahrscheinlichkeiten (1)

Neue Information kann die Wahrscheinlichkeit eines Ereignisses ändern.

Beispiel: Die Wahrscheinlichkeit von Zahnlöchern erhöht sich, wenn man weiß, dass der Patient Zahnschmerzen hat.

Liegt Zusatzinformation vor, darf nicht mehr mit a priori Wahrscheinlichkeiten gerechnet werden!

$P(A | B)$ bezeichnet die bedingte oder a posteriori Wahrscheinlichkeit von A , sofern die *alleinige* Beobachtung (die *Evidenz*) B gegeben ist:

$$P(Cavity | Toothache) = 0.8$$

$P(X | Y)$ ist die Tabelle aller bedingter Wahrscheinlichkeiten über alle Werte von X und Y .

Bedingte Wahrscheinlichkeiten (2)

$P(\text{Weather} \mid \text{Headache})$ ist eine 4 x 2 Tabelle von bedingten Wahrscheinlichkeiten aller Kombinationen der Werte einer Zufallsvariablen.

	Headache = TRUE	Headache = FALSE
Weather = Sunny	$P(W = \text{Sunny} \mid \text{Headache})$	$P(W = \text{Sunny} \mid \neg \text{Headache})$
Weather = Rain		
Weather = Cloudy		
Weather = Snow		

Bedingte Wahrscheinlichkeiten ergeben sich aus unbedingten Wahrscheinlichkeiten per Definition
(für $P(B) > 0$):

$$P(A \mid B) = \frac{P(A \wedge B)}{P(B)}$$

Bedingte Wahrscheinlichkeiten (3)

$\mathbf{P}(X, Y) = \mathbf{P}(X \mid Y)\mathbf{P}(Y)$ entspricht einem Gleichungssystem:

$$P(W = \textit{Sunny} \wedge \textit{Headache}) = P(W = \textit{Sunny} \mid \textit{Headache})P(\textit{Headache})$$

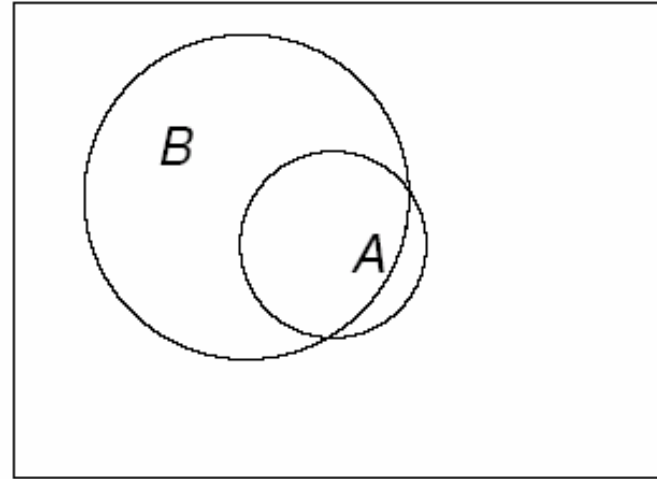
$$P(W = \textit{Rain} \wedge \textit{Headache}) = P(W = \textit{Rain} \mid \textit{Headache})P(\textit{Headache})$$

$$\vdots = \vdots$$

$$P(W = \textit{Snow} \wedge \neg \textit{Headache}) = P(W = \textit{Snow} \mid \neg \textit{Headache})P(\neg \textit{Headache})$$

Bedingte Wahrscheinlichkeiten (4)

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$



- **Produktregel:** $P(A \cap B) = P(A \mid B)P(B)$
- **Analog:** $P(A \cap B) = P(B \mid A)P(A)$
- **A und B heißen unabhängig voneinander, falls**
 $A) = P(B)$.

$$P(A \mid B) = P(A) \text{ und } P(B \mid$$

Dann und nur dann gilt $P(A \cap B) = P(A)P(B)$!

Interpretation bedingter Wahrscheinlichkeiten

Achtung: häufiger Interpretationsfehler:

*„ $P(A|B) = 0.8$ bedeutet, dass $P(A) = 0.8$ ist,
falls B wahr ist.“*

Diese Aussage ist aus mindestens zwei Gründen falsch!

- 1. $P(A)$ ist immer die a priori Wahrscheinlichkeit, nicht die a posteriori Wahrscheinlichkeit, gegeben eine Evidenz (B wahr).**
- 2. $P(A|B) = 0.8$ ist nur dann anwendbar, falls keine andere Evidenz außer B vorhanden ist! Wenn C bekannt ist, dann muss $P(A|B \wedge C)$ berechnet oder geschätzt werden. Im Allgemeinen gilt $P(A|B \wedge C) \neq P(A|B)$; z.B. könnte $C \rightarrow A$ gelten (C impliziert A).**

Axiomatische Wahrscheinlichkeitstheorie

Eine Funktion P von aussagenlogischen Formeln in die Menge $[0, 1]$ ist ein Wahrscheinlichkeitsmaß, falls für alle Aussagen A, B gilt:

1. $0 \leq P(A) \leq 1$

2. $P(True) = 1$

3. $P(False) = 0$

4. $P(A \vee B) = P(A) + P(B) - P(A \wedge B)$

Alle anderen Eigenschaften lassen sich aus diesen Axiomen ableiten, z.B.

$P(\neg A) = 1 - P(A)$ folgt aus

$P(A \vee \neg A) = 1$ und

$P(A \wedge \neg A) = 0.$

Wieso sind die Axiome sinnvoll?

- Wenn P eine *objektiv* beobachtbare Wahrscheinlichkeit bezeichnet, machen die Axiome natürlich Sinn.
- Aber wieso sollte ein Agent diese Axiome beachten, wenn er den *Grad seiner Überzeugung* modelliert?
- *Objektive vs. subjektive* Wahrscheinlichkeiten
- Die Axiome schränken die Menge der Überzeugungen ein, die ein Agent aufrechterhalten kann.
- Eines der überzeugendsten Argumente, warum subjektive Überzeugungen die Axiome respektieren sollten, wurde 1931 von de Finetti gegeben. Es basiert auf dem Zusammenhang zwischen Aktionen und dem Grad der Überzeugung:
Sind die Überzeugungen widersprüchlich, dann wird der Agent auf lange Sicht in seiner Umwelt scheitern!

Verbundwahrscheinlichkeit

- Wahrscheinlichkeit, die ein Agent jeder Aussage in der Domäne zuordnet.
- Ein atomares Ereignis ist eine Zuweisung von Werten an alle Zufallsvariablen $X_1, \dots, X_n$ (= vollständige Spezifikation eines Zustands).

Beispiel: Seien X, Y boolesche Variablen. Dann gibt es die folgenden 4 atomaren Ereignisse:

$$X \wedge Y, \quad X \wedge \neg Y, \quad \neg X \wedge Y, \quad \neg X \wedge \neg Y$$

- Die Verbundwahrscheinlichkeitsverteilung $P(X_1, \dots, X_n)$ weist jedem *atomaren Ereignis* eine Wahrscheinlichkeit zu:

	Toothache	$\neg$ Toothache
Cavity	0.04	0.06
$\neg$ Cavity	0.01	0.89

- Da alle atomaren Ereignisse disjunkt sind, ist die Summe über alle Felder 1 (Disjunktion der Ereignisse). Die Konjunktion ist notwendigerweise falsch.

Rechnen mit der Verbundwahrscheinlichkeit

Alle interessanten Wahrscheinlichkeiten lassen sich aus der Verbundwahrscheinlichkeit errechnen, indem wir sie als Disjunktion von atomaren Ereignissen formulieren.

Beispiel:

$$\begin{aligned} P(Cavity \vee Toothache) &= P(Cavity \wedge Toothache) \\ &+ P(\neg Cavity \wedge Toothache) \\ &+ P(Cavity \wedge \neg Toothache) \end{aligned}$$

Unbedingte Wahrscheinlichkeiten erhält man durch Aufsummieren von Zeile oder Spalte (Marginalisierung):

$$P(Cavity) = P(Cavity \wedge Toothache) + P(Cavity \wedge \neg Toothache)$$

Womit wieder bedingte Wahrscheinlichkeiten ableitbar sind:

$$P(Cavity \mid Toothache) = \frac{P(Cavity \wedge Toothache)}{P(Toothache)} = \frac{0.04}{0.04 + 0.01} = 0.80$$

Probleme mit der Verbundwahrscheinlichkeit

- Aus der Verbundwahrscheinlichkeit lassen sich alle Wahrscheinlichkeiten einfach ermitteln.
- Allerdings umfasst die Verbundwahrscheinlichkeit k^n Werte, wenn es n Zufallsvariablen mit k Werten gibt.
- Schwierig darzustellen
- Schwierig zu ermitteln

Fragen:

1. Gibt es eine dichtere Darstellung von Verbundwahrscheinlichkeiten?
 2. Gibt es eine effiziente Methode, diese Darstellung zu verarbeiten?
-
- Allgemein nicht, aber in vielen Fällen geht es. Moderne Systeme arbeiten direkt mit bedingten Wahrscheinlichkeiten (Diagnose-Kausalität) und machen Annahmen über die Unabhängigkeit von Variablen, um Rechnungen zu vereinfachen.

Die Bayessche Regel

Wir wissen (Produktregel):

$$P(A \wedge B) = P(A | B)P(B) \quad \text{und} \quad P(A \wedge B) = P(B | A)P(A)$$

Durch Gleichsetzen der rechten Seiten folgt:

$$P(A | B)P(B) = P(B | A)P(A)$$

$$\Rightarrow \boxed{P(A | B) = \frac{P(B|A)P(A)}{P(B)}}$$

Für mehrwertige Variablen (Menge von Gleichungen):

$$P(Y | X) = \frac{P(X | Y)P(Y)}{P(X)}$$

Verallgemeinerung (bzgl. Hintergrundevidenzen):

$$P(Y | X, E) = \frac{P(X | Y, E)P(Y | E)}{P(X | E)}$$

Anwendung der Bayesschen Regel

$$P(\textit{Toothache} \mid \textit{Cavity}) = 0.4$$

$$P(\textit{Cavity}) = 0.1$$

$$P(\textit{Toothache}) = 0.05$$

$$P(\textit{Cavity} \mid \textit{Toothache}) = \frac{0.4 \times 0.1}{0.05} = 0.8$$

Warum nicht gleich $P(\textit{Cavity} \mid \textit{Toothache})$ schätzen?

- Kausales Wissen wie $P(\textit{Toothache} \mid \textit{Cavity})$ ist i.A. robuster als diagnostisches Wissen $P(\textit{Cavity} \mid \textit{Toothache})$:
- Die Kausalität $P(\textit{Toothache} \mid \textit{Cavity})$ ist unabhängig von den a priori Wahrscheinlichkeiten $P(\textit{Toothache})$ und $P(\textit{Cavity})$.
- Nähme $P(\textit{Cavity})$ bei einer Karies-Epidemie zu, so bliebe die Kausalität $P(\textit{Toothache} \mid \textit{Cavity})$ unverändert, während sich $P(\textit{Toothache})$ proportional mit $P(\textit{Toothache} \mid \textit{Cavity})$ änderte.
- Ein Arzt, der $P(\textit{Cavity} \mid \textit{Toothache})$ geschätzt hätte, wüsste keine Regel zum Update von $P(\textit{Cavity} \mid \textit{Toothache})$. Ein Arzt, der die Bayesschen Regel benutzt hätte, wüsste um das proportionale Verhalten zwischen $P(\textit{Toothache} \mid \textit{Cavity})$ und $\textit{Cavity}$.

Relative Wahrscheinlichkeit

Annahme: Wir wollen auch die Wahrscheinlichkeit der Diagnose *GumDisease* betrachten.

$$P(\text{Toothache} \mid \text{GumDisease}) = 0.7$$

$$P(\text{GumDisease}) = 0.02$$

Welche Diagnose ist wahrscheinlicher?

Wenn nur die relative Wahrscheinlichkeit der Ursachen interessiert, muss die $P(T)$ nicht geschätzt werden:

$$P(C \mid T) = \frac{P(T \mid C)P(C)}{P(T)} \quad \text{oder} \quad P(G \mid T) = \frac{P(T \mid G)P(G)}{P(T)}$$

■ **Relative Wahrscheinlichkeiten können für einige Entscheidungen hinreichend sein!**

$$\begin{aligned} \frac{P(C \mid T)}{P(G \mid T)} &= \frac{P(T \mid C)P(C)}{P(T)} \times \frac{P(T)}{P(T \mid G)P(G)} = \frac{P(T \mid C)P(C)}{P(T \mid G)P(G)} \\ &= \frac{0.4 \times 0.1}{0.7 \times 0.02} = 28.57 \end{aligned}$$

Normierung (1)

Wenn wir die absolute Wahrscheinlichkeit von $P(C | T)$ bestimmen wollen und $P(T)$ nicht kennen, können wir auch eine *vollständige* Fallanalyse durchführen (z.B. für C und $\neg C$) und den Zusammenhang $P(C | T) + P(\neg C | T) = 1$ (hier boolesche Variable) ausnutzen:

$$P(C | T) = \frac{P(T | C)P(C)}{P(T)}$$

$$P(\neg C | T) = \frac{P(T | \neg C)P(\neg C)}{P(T)}$$

$$P(C | T) + P(\neg C | T) = \frac{P(T | C)P(C)}{P(T)} + \frac{P(T | \neg C)P(\neg C)}{P(T)}$$

$$P(T) = P(T | C)P(C) + P(T | \neg C)P(\neg C)$$

Normierung (2)

Durch Einsetzen in die oberste Gleichung:

$$P(C | T) = \frac{P(T | C)P(C)}{P(T | C)P(C) + P(T | \neg C)P(\neg C)}$$

Für mehrwertige Zufallsvariablen:

$$P(Y | X) = \alpha P(X | Y)P(Y)$$

wobei α eine Normierungskonstante ist, welche die Werte in $P(Y | X)$ zu 1 aufsummieren lässt, z.B.

$$\alpha(1, 1, 3) = \left(\frac{1}{5}, \frac{1}{5}, \frac{3}{5}\right).$$

Multiple Evidenzen (1)

Der Patient bejaht die Frage nach Zahlschmerzen. Aus dieser ersten Evidenz schließt der Zahnarzt:

$$P(\text{Cavity} \mid \text{Toothache}) = 0.8$$

Der Zahnarzt ertastet bei der Untersuchung mit einem Haken eine Bruchstelle im Zahn. Aus dieser zweiten Evidenz schließt der Zahnarzt:

$$P(\text{Cavity} \mid \text{Catch}) = 0.95$$

Beide Schlüsse könnten z.B mit der Bayesschen Regel berechnet worden sein. Aber was bringt die kombinierte Evidenz?

Mit der Bayesschen Regel könnte er weiter ermitteln:

$$P(\text{Cav} \mid \text{Tooth} \wedge \text{Catch}) = \frac{P(\text{Tooth} \wedge \text{Catch} \mid \text{Cav}) \times P(\text{Cav})}{P(\text{Tooth} \wedge \text{Catch})}$$

Multiple Evidenzen (2)

Problem: Er braucht $P(\textit{Tooth} \wedge \textit{Catch} \mid \textit{Cav})$, d.h. Diagnosewissen für alle Kombinationen von Symptomen im allgemeinen Fall.

Besser ist es, Evidenzen mit Hilfe der Evidenzenregel schrittweise aufzunehmen.

$$P(Y \mid X, E) = \frac{P(X \mid Y, E)P(Y \mid E)}{P(X \mid E)}$$

Mit einer bestimmten a priori Wahrscheinlichkeit hat der Patient ein Loch: $P(\textit{Cav})$. Er berichtet von Zahnschmerzen (Bayessche Regel):

$$P(\textit{Cav} \mid \textit{Tooth}) = P(\textit{Cav}) \times \frac{P(\textit{Tooth} \mid \textit{Cav})}{P(\textit{Tooth})}$$

Multiple Evidenzen (3)

$$\boxed{P(Cav \mid Tooth)} = P(Cav) \times \frac{P(Tooth \mid Cav)}{P(Tooth)} \quad (1)$$

Die Untersuchung ergibt *Catch*, also

$$P(Cav \mid Catch, Tooth) = \frac{P(Catch \mid Cav, Tooth)}{P(Catch \mid Tooth)} \times \boxed{P(Cav \mid Tooth)} \quad (2)$$

(1) in (2) einsetzen ergibt

$$P(Cav \mid Catch, Tooth) = P(Cav) \times \frac{P(Tooth \mid Cav)}{P(Tooth)} \times \frac{\boxed{P(Catch \mid Cav, Tooth)}}{P(Catch \mid Tooth)}$$

Multiple Evidenzen (4)

Annahme bedingter Unabhängigkeit von *Toothache* und *Catch* gegeben *Cavity*
(vereinfachtes Diagnosewissen):

$$\boxed{P(\text{Catch} \mid \text{Cav}, \text{Tooth})} = P(\text{Catch} \mid \text{Cav})$$
$$P(\text{Tooth} \mid \text{Cav}, \text{Catch}) = P(\text{Tooth} \mid \text{Cav})$$

$$P(\text{Cav} \mid \text{Catch}, \text{Tooth}) = P(\text{Cav}) \times \frac{P(\text{Tooth} \mid \text{Cav})}{P(\text{Tooth})} \times \boxed{\frac{P(\text{Catch} \mid \text{Cav})}{P(\text{Catch} \mid \text{Tooth})}}$$

Multiple Evidenzen (5)

$$P(Cav \mid Catch, Tooth) = P(Cav) \times \frac{P(Tooth \mid Cav)}{P(Tooth)} \times \frac{P(Catch \mid Cav)}{P(Catch \mid Tooth)}$$

Immer noch scheint die Betrachtung von Paaren (Tripeln etc.) von Symptomen wie $P(Catch \mid Tooth)$ nötig! Aber mit Produktregel folgt aus den Nennern der Brüche:

$$P(Tooth) \times P(Catch \mid Tooth) = P(Tooth \wedge Catch)$$

Dies führt zu:

$$P(Cav \mid Catch, Tooth) = P(Cav) \times \frac{P(Tooth \mid Cav) \times P(Catch \mid Cav)}{P(Catch \wedge Tooth)}$$

Für die Bestimmung von $P(Cav \mid Catch, Tooth)$ ist $P(Tooth \wedge Catch)$ ein Normierungsfaktor und eliminierbar, sofern $P(Catch \mid \neg Cav)$ und $P(Tooth \mid \neg Cav)$ bekannt sind.

Multiple Evidenzen (6)

Somit:

$$P(\text{Tooth} \wedge \text{Catch}) = P(\text{Tooth} \wedge \text{Catch} \mid \text{Cav}) P(\text{Cav}) + P(\text{Tooth} \wedge \text{Catch} \mid \neg \text{Cav}) P(\neg \text{Cav})$$

Mit bedingter Unabhängigkeit:

$$\begin{aligned} P(\text{Tooth} \wedge \text{Catch}) &= P(\text{Tooth} \mid \text{Cav}) P(\text{Catch} \mid \text{Cav}) P(\text{Cav}) \\ &+ P(\text{Tooth} \mid \neg \text{Cav}) P(\text{Catch} \mid \neg \text{Cav}) P(\neg \text{Cav}) \end{aligned}$$

Einzusetzen in:

$$P(\text{Cav} \mid \text{Catch}, \text{Tooth}) = P(\text{Cav}) \times \frac{P(\text{Tooth} \mid \text{Cav}) \times P(\text{Catch} \mid \text{Cav})}{\boxed{P(\text{Catch} \wedge \text{Tooth})}}$$

Also Ableitung möglich aus unbedingten Wahrscheinlichkeiten und einfachen kausalen Regeln!

Zusammenfassung Multiple Evidenzen

- Mehrfache Evidenzen sind berechenbar durch Reduktion auf
 - a priori Wahrscheinlichkeiten und
 - (kausale) bedingte Wahrscheinlichkeiten für eine Evidenz
 - unter Annahme der bedingten Unabhängigkeit.

Allgemeine Kombinationsregel:

$$P(Z \mid X, Y) = \alpha P(Z)P(X \mid Z)P(Y \mid Z)$$

für X und Y bedingt unabhängig bei gegebenem Z und mit Normierungskonstante α .

Zusammenfassung

- Unsicherheit ist unvermeidbar in komplexen und dynamischen Welten, in denen **Agenten zur Ignoranz gezwungen sind.**
- Wahrscheinlichkeiten formulieren die **Unfähigkeit eines Agenten, eine definitive Entscheidung zu fällen.** Sie drücken den Grad seiner Überzeugung aus.
- Bedingte und unbedingte Wahrscheinlichkeiten können über Aussagen formuliert werden.
- Verletzt ein Agent die wahrscheinlichkeitstheoretischen Axiome, so wird er unter bestimmten Umständen irrationales Verhalten zeigen.
- Die Bayessche Regel ermöglicht es, unbekannte Wahrscheinlichkeiten aus bekannten Wahrscheinlichkeiten zu berechnen.
- Multiple Evidenzen können bei bedingter Unabhängigkeit effektiv in die Berechnung einbezogen werden.

Beispiel: VW Planungssystem

- etwa 200 item families (Variablen, Merkmale)
- von 2 bis 50 items pro item family
(Werte pro Variable)



- mehr als 2^{200} mögliche Fahrzeug Spezifikationen
- Wahl der gültigen Spezifikationen ist beschränkt durch **REGEL SYSTEME**
(10.000 technische Regeln, noch mehr marketing- und produktions-orientierte Regeln)

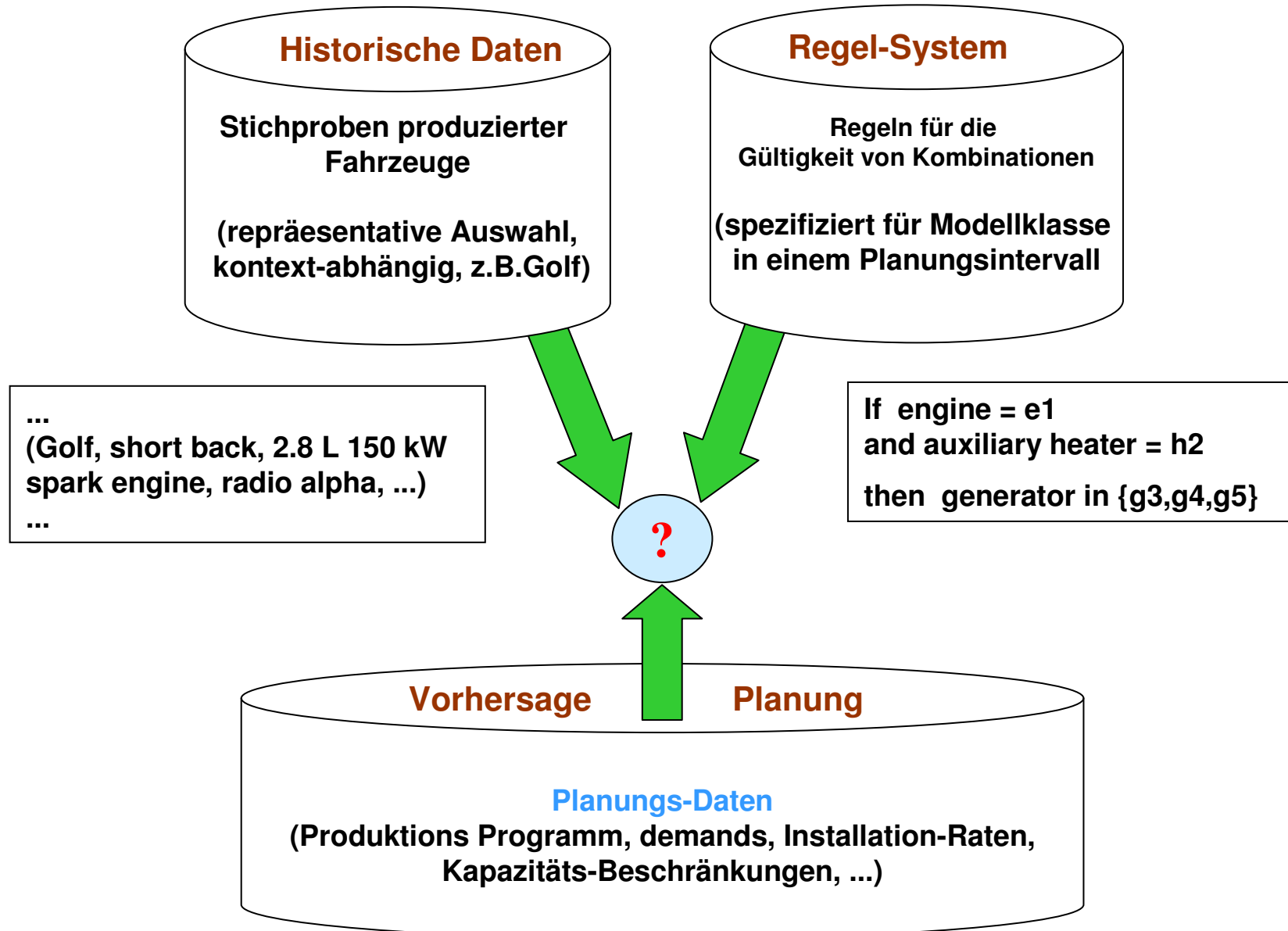
Beispiel

if $engine = e_1$ then $transmission = t_3$

if $engine = e_4$ and $auxiliary\ heater = h_1$

then $generator \in \{g_3, g_4, g_5\}$

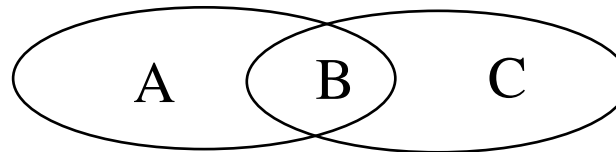
Das Problem aus methodischer Sicht



Probabilistische Graphische Modelle

Ein Probabilistisches Graphisches Model besteht aus einer **Zerlegung** sowie **lokalen Wahrscheinlichkeitsräumen**

Zerlegung: Hypergraph auf Variablenmenges



Lokale Modelle: Marginal Verteilungen von A,B
und B,C , die „zueinander passen“

Bayessche Netze

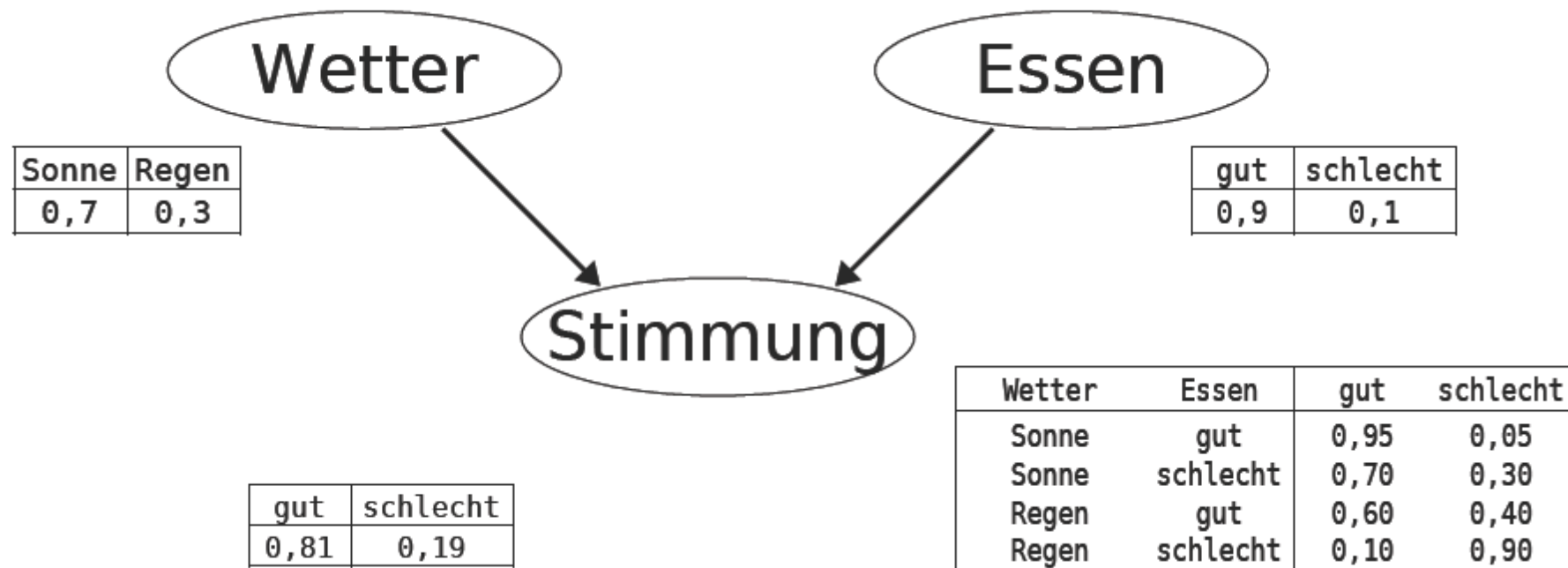
- Graphische Modelle
- Formalismus zur Darstellung kausaler Abhängigkeiten
- Möglichkeit, einen hochdimensionalen Raum durch mehrere niederdimensionale Räume zu repräsentieren
- Grundlage: Bayes Theorem

Wahrscheinlichkeit des Ereignisses B, falls Ereignis A bekannt ist

$$P(B | A) = \frac{P(A | B) P(B)}{P(A)}$$

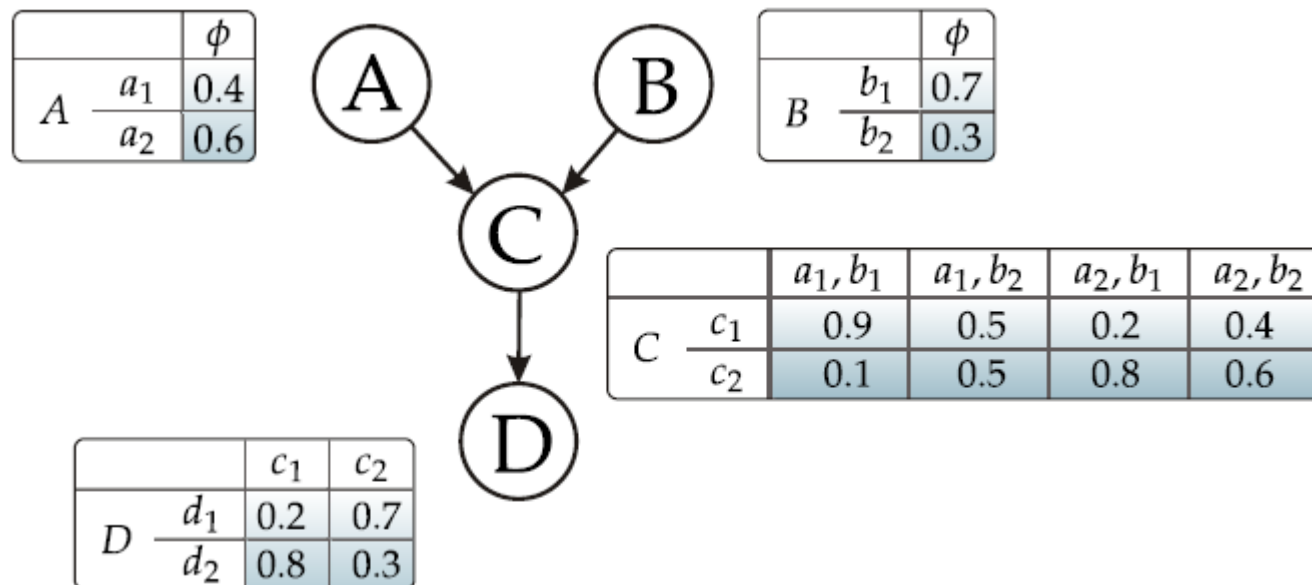
Beispiel eines simplen Bayes Netzes

- Knoten --> Variablen
- Kanten --> Abhängigkeiten der Variablen



Bayessche Netze

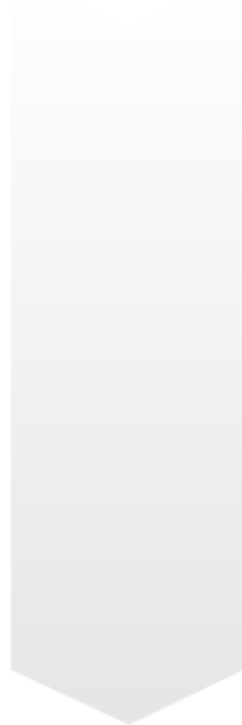
- Gerichteter, azyklischer Graph
- Knoten bedeuten Attribute
- Kanten symbolisieren (prob.) Abhängigkeiten
- Beinhalten **qualitative** und **quantitative** Information



Bayes Netze

Top-Down-Ansatz

Expertenwissen



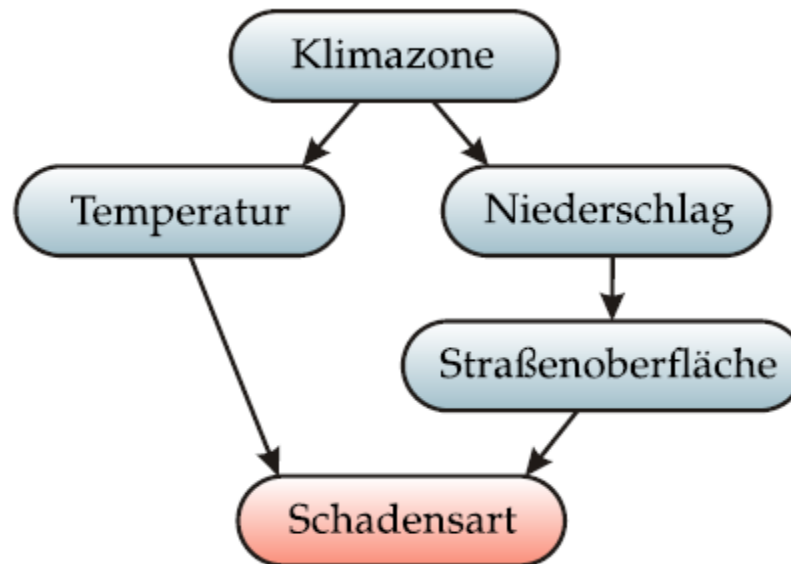
Daten

Bottom-Up-Ansatz

Expertenwissen

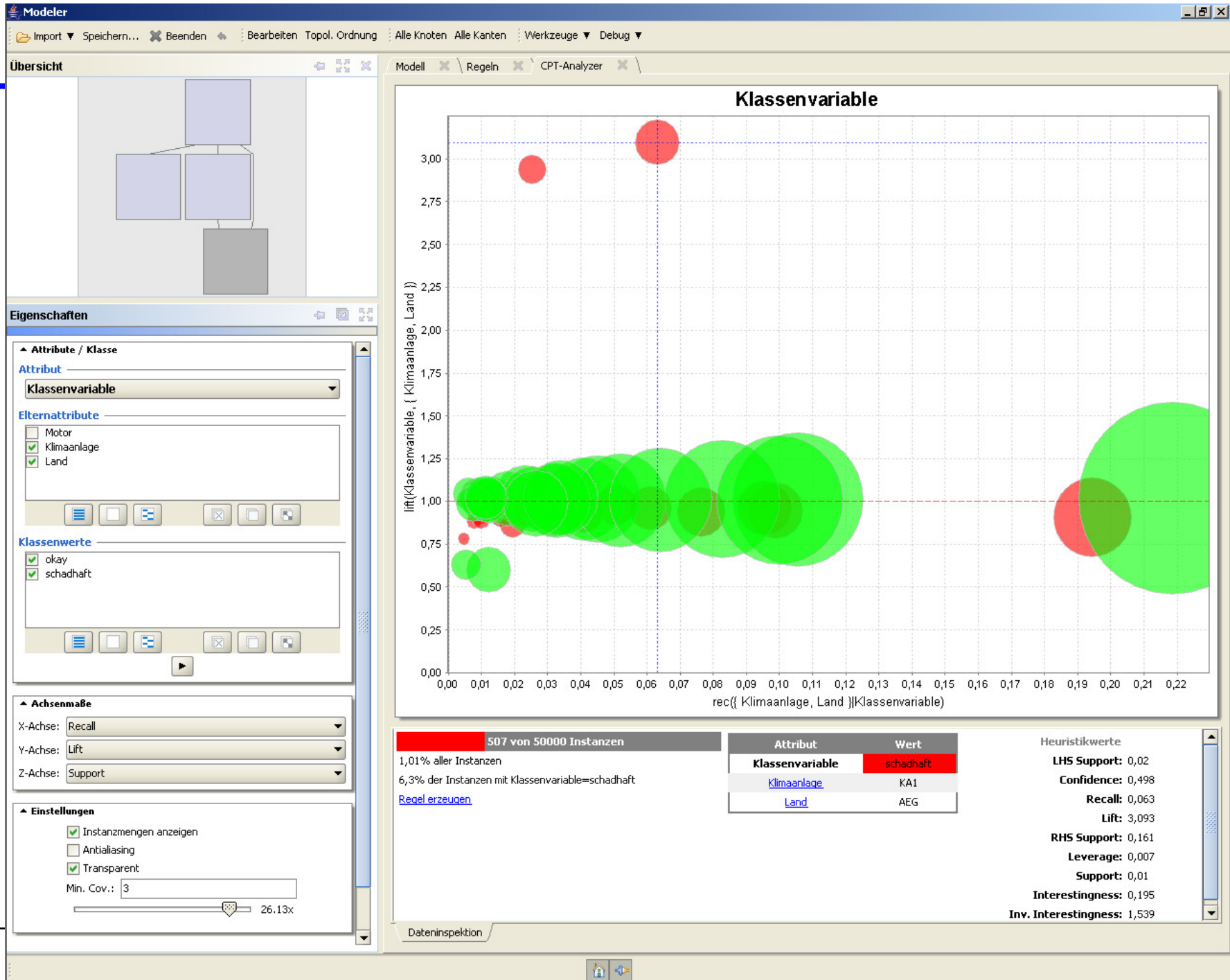


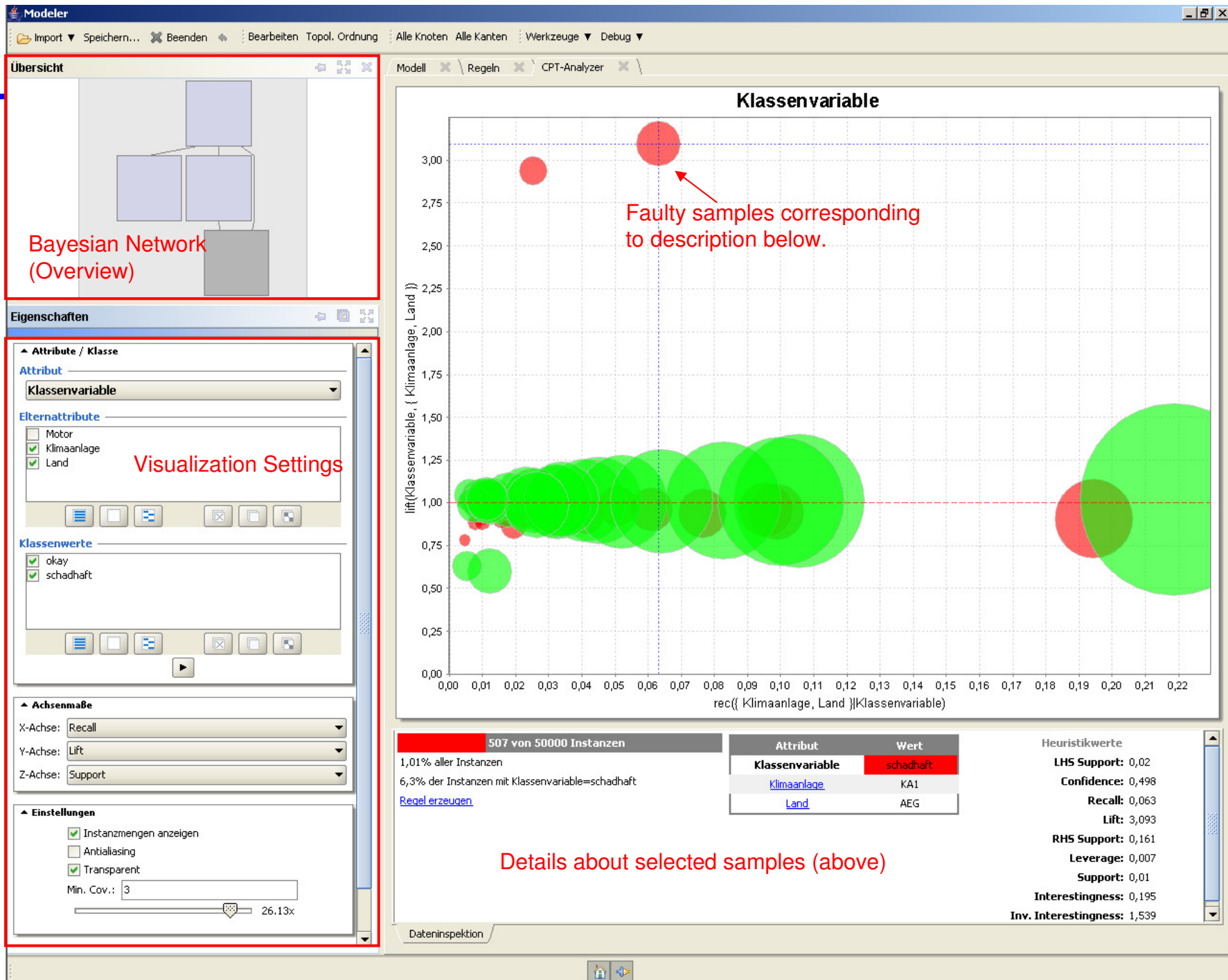
Daten



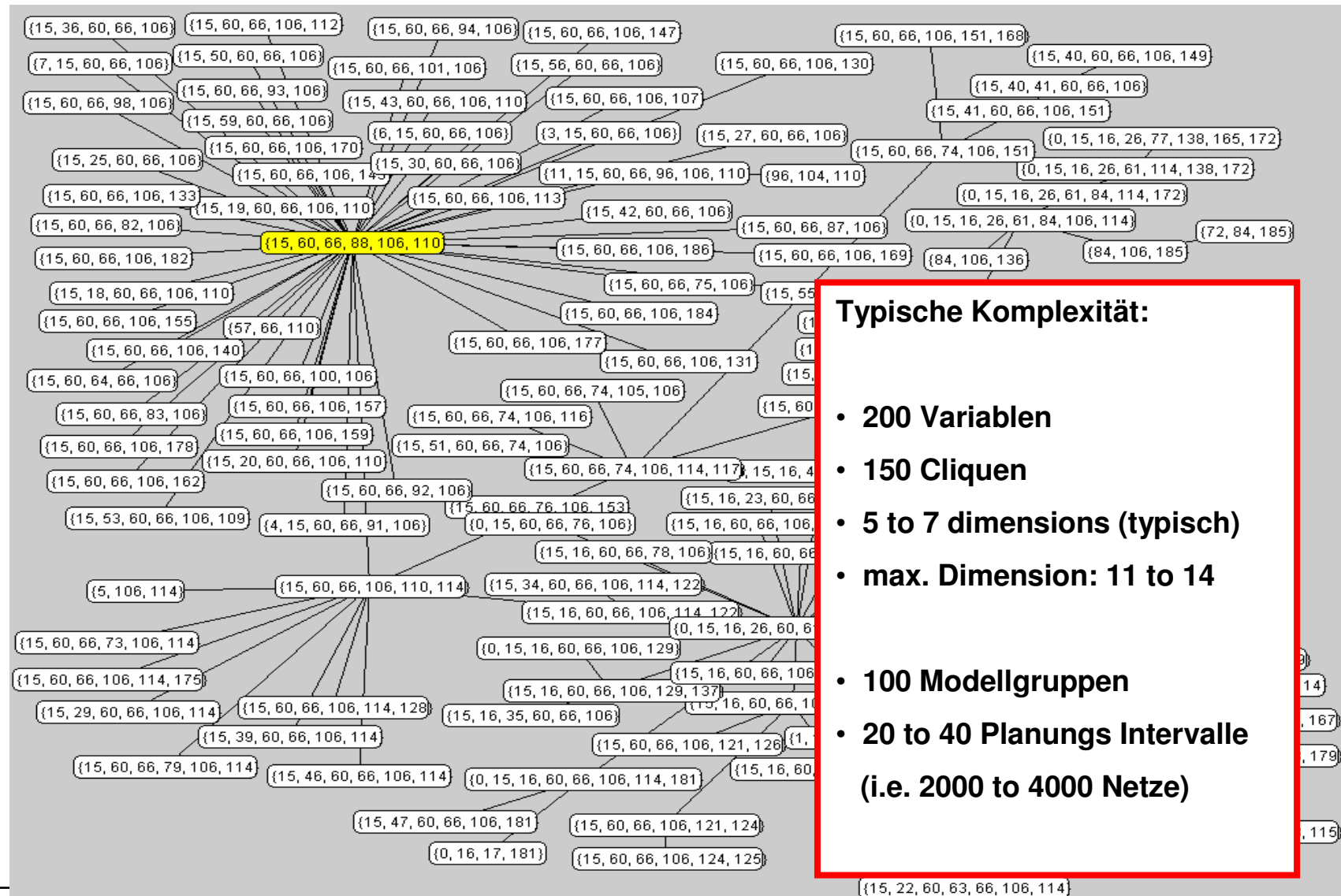
Erlernen von Bayes Netzen aus Daten

- Erstellung eines Bayesschen Netzes ist von Hand sehr aufwändig
- viele Zusammenhänge müssen erkannt werden
- häufig keine Experten mit ausreichendem Wissen verfügbar
- Daten oft in großen Mengen vorhanden
- Lernverfahren ermöglichen Entwicklung Bayesscher Netzwerke aus vorhandenen Daten
- Arten des Lernens:
 - Parameterlernen:
 - ♦ Netzstruktur bekannt, lediglich Parameter der Wahrscheinlichkeitsverteilungen müssen geschätzt werden
 - Strukturlernen:
 - ♦ keine Netzstruktur vorgegeben
 - ♦ Abhängigkeiten aus den Daten bestimmen
- Erweiterungen auf Lernverfahren für Hyperbaum Strukturen





Planungs-Modelle



Planungs-Operation: Konditionieren

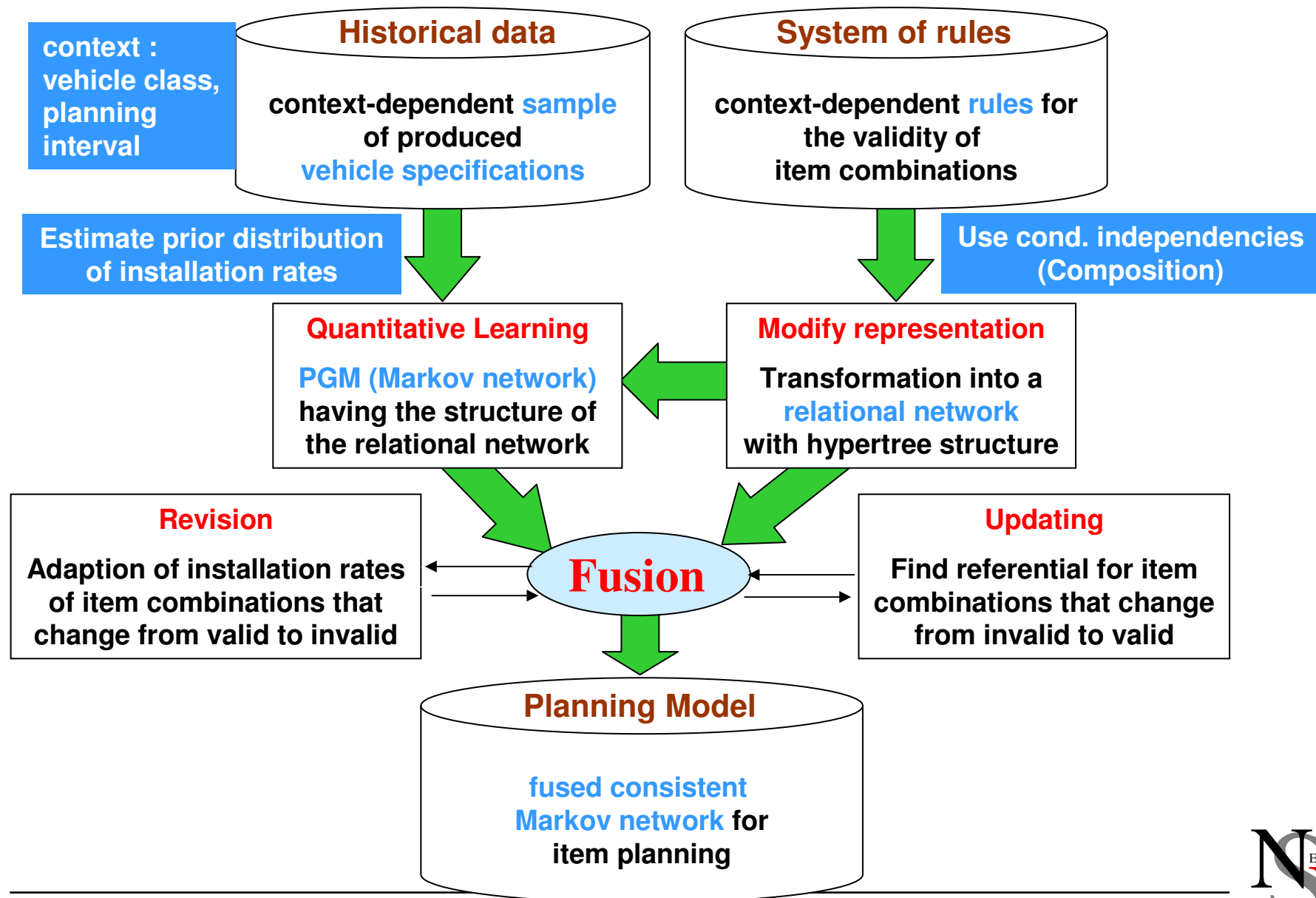
- Eingabe: Eigenschaftskombination
(Menge von Variablenbelegungen)

- Operation: Berechnen der bedingten Netz-Verteilung und der Wahrscheinlichkeit der gegebenen Eigenschaftskombination (propagation).

- Beispiele: Berechnung des Teilebedarfs
Berechnung der Installationsrate der Merkmalskombination
 (m_3, t_4, g_5)

Simulation
Analysiere Präferenzen aller Kunden die ein *Navigationssystem* in einem *VW Polo* gekauft haben.

Planning Model based on Belief Change



Planungsoperation : Updating

- Eingabe: Menge der Eingabetupel, die von ungültig auf gültig gesetzt werden sollen
- Operation: Kopieren der Abhängigkeitsstruktur (cross-product ratios) der Referenz Kombination auf die Eingabe Kombination und Initialisierung mit $\mathcal{E}$ -probabilities.
- Anwendung: Technische Modifikationen
Die Kombination von engine e_1 and transmission t_3 ändert sich von ungültig zu gültig, die quantitativen Abhängigkeiten von (e_2, t_3) werden übernommen.

Planning Operation: Revision

- Input Data: Family of marginal / conditional probability distributions

- Operation: Calculate Markov network with same structure that satisfies all input distributions and is conform to the principle of minimal change.

- Application: Marketing stipulations
Installation probability of item *air condition* increases by 10 % in case of *Golf all-wheel drive in France*.

Logistic restrictions
The *maximum availability* of engine e_1 in week 32/06 is 1.000 .

Benutzerschnittstelle

Name:	Golf - No. 02/07/05 - 17		v
Vehicle class:	Golf	v	Market: Germany v
	Planning interval: 36/05		v
Revision scheme:	Engines	Revision context:	Short back v Comfort v
		Context scheme:	Body v Equipment v

Partitioning:	Installation rates (%)		Restriction
	estimated	assigned	
Group of 1,8L spark engines v	5,79	9,00	≤ 500
Diesel engine X1 (single item) v	2,13	3,00	
Diesel engine X2 (single item) v	21,07	[18,20]	
Rest	71,01		

Software Entwicklung

- Client-Server System
 - Server on 6-8 Machines (16 GB each)
 - 4-Processor AMD Opteron system
 - Terabyte storage device
 - Betriebssystem: Linux
-
- 15 Entwickler Programmiersprache: JAVA
 - WebSphere Application Developer, Eclipse
 - DB-System: Oracle
 - Einführung weltweit: jetzt

