OTTO VON GUERICKE UNIVERSITÄT MAGDEBURG

INF FAKULTÄT FÜR INFORMATIK

Intelligente Systeme

Bayes-Netze

Prof. Dr. R. Kruse C. Braune C. Moewes

{kruse,cbraune,cmoewes}@iws.cs.uni-magdeburg.de

Institut für Wissens- und Sprachverarbeitung

Fakultät für Informatik

Otto-von-Guericke Universität Magdeburg

Übersicht

1. **Motivation**
2. Grundlagen der Wahrscheinlichkeitstheorie
3. Probabilistische Inferenz
4. Bayes'sche Netze

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!",

oder 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 Uni sein, um an Vorlesung teilzunehmen

mehrere **Pläne**, um Ziel zu erreichen:

- P_1: 8:00 Uhr aufstehen, 8:55 Uhr Haus verlassen, 9:00 Uhr Bus...
- P_2: 7:30 Uhr aufstehen, 8:25 Uhr Haus verlassen, 8:30 Uhr Bus...
- ...

Alle Pläne sind *korrekt*, aber

- sie implizieren verschiedene **Kosten** und verschiedene **Wahrscheinlichkeiten** Ziel *tatsächlich* zu erreichen
- P_2 am besten, denn Vorlesung ist wichtig, und Erfolgsrate bei P_1 liegt nur bei ca. 85–95%

Grade der Überzeugung

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

Rationale Entscheidungen unter Unsicherheit

- Verschiedene *Aktionen* oder *Pläne* zur Auswahl
- Können mit verschiedenen *Wahrscheinlichkeiten* zu verschiedenen Ergebnissen führen
- *Aktionen* verursachen verschiedene – ggf. subjektive – **Kosten**
- *Ergebnisse* haben verschiedenen – ggf. subjektiven – **Nutzen**
- Rational: wähle Aktion, die größten **zu erwartenden Gesamtnutzen** hat

Entscheidungstheorie = Nutzentheorie + Wahrscheinlichkeitstheorie

Entscheidungstheoretischer Agent

Algorithmus 1 DT-Agent

Eingabe: a *percept*

Ausgabe: an *action*

1: static: a set of probabilistic beliefs about the state of the world
2: calculate updated probabilities for current state based on available evidence including current percept and previous action
3: calculate outcome probabilities for actions, given action descriptions and probabilities of current states
4: select *action* with highest expected utility, given probabilities of outcomes and utility information
5: **return** *action*

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

Übersicht

1. Motivation

2. **Grundlagen der Wahrscheinlichkeitstheorie**

3. Probabilistische Inferenz

4. Bayes'sche Netze

Unbedingte Wahrscheinlichkeiten I

$P(A)$ bezeichnet **unbedingte** oder **a priori** Wahrscheinlichkeit, dass Ereignis A eintritt, falls *keine* zusätzliche Information verfügbar, wobei A eine Menge ist

man kann Wahrscheinlichkeiten auch für Aussagen nutzen:

$$P(\text{Karies}) = 0.1$$

Karies ist eine Proposition (Aussage)

Wahrscheinlichkeiten werden durch statistische Analysen oder aus allgemeinen Regeln gewonnen

Axiomatische Wahrscheinlichkeitstheorie

Funktion $P : 2^\Omega \to [0, 1]$ ist (im endlichen Fall) ein Wahrscheinlichkeitsmaß, falls für alle A, B gilt:

$$0 \le P(A) \le 1,$$
$$P(\Omega) = 1,$$
$$P(\emptyset) = 0,$$
$$P(A \vee B) = P(A) + P(B) - P(A \wedge B).$$

Andere 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 objektiv beobachtbare Wahrscheinlichkeiten beschreibt (z.B. Würfelexperiment), dann machen die Axiome natürlich Sinn

Aber wieso sollte ein Agent Axiome beachten, wenn er den *Grad seiner (subjektiven) Überzeugung* modelliert?

Axiome schränken Menge der Überzeugungen ein, die Agent aufrechterhalten kann

Eines der überzeugendsten Argumente, warum subjektive Überzeugungen den Axiome genügen sollten, stammt von de Finetti (1931):
Verhält sich ein Agent nicht rational, so hat er auf lange Sicht in seiner Umwelt Nachteile.

Unbedingte Wahrscheinlichkeiten I

Eine Zufallsvariable kann nicht nur die Werte wahr und falsch annehmen, sondern mehrere Werte

Zufallsvariable Wetter mit $P(\text{Wetter} = \text{sonnig})$ für $P(\{\omega \mid \text{Wetter}(\omega) = \text{sonnig})$

$$P(\text{Wetter} = \text{sonnig}) = 0.7$$
$$P(\text{Wetter} = \text{Regen}) = 0.2$$
$$P(\text{Wetter} = \text{bewölkt}) = 0.08$$
$$P(\text{Wetter} = \text{Schnee}) = 0.02$$
$$P(\text{Schmerz} = \text{wahr}) = 0.1$$

Logische Verknüpfungen von Propositionen sind möglich, z.B.

$$P(\text{Karies} \wedge \neg\text{Schmerz}) = 0.06$$

entspricht $P(\{\omega \mid \text{Karies}(\omega) = \text{wahr} \wedge \text{Schmerz}(\omega) = \text{falsch})$

Unbedingte Wahrscheinlichkeiten II

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

$$\mathbf{P}(\text{Schmerz}) = \langle 0.1, 0.9 \rangle$$
$$\mathbf{P}(\text{Wetter}) = \langle 0.7, 0.2, 0.08, 0.02 \rangle$$

Vektoren definieren Wahrscheinlichkeitsverteilung der Zufallsvariablen *Schmerz* und *Wetter*

$P(\text{Schmerz}, \text{Wetter})$ ist (4×2)-Tabelle von Wahrscheinlichkeiten aller Kombinationen der Werte einer Zufallsvariablen

	Schmerz = wahr	Schmerz = falsch
Wetter = sonnig	$P(W = \text{sonnig} \land S)$	$P(W = \text{sonnig} \land \neg S)$
Wetter = Regen		
Wetter = bewölkt		
Wetter = Schnee		

Verbundwahrscheinlichkeit

atomares Ereignis: Zuweisung von Werten an alle Zufallsvariablen $X_1, \ldots, X_n$ (= vollständige Spezifikation eines Zustands)

Beispiel: Seien X, Y boolesche Variablen, dann gibt es folgende 4 atomaren Ereignisse:

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

Verbundwahrscheinlichkeitsverteilung $P(X_1, \ldots, X_n)$ weist jedem atomaren Ereignis eine Wahrscheinlichkeit zu:

	Schmerz	¬Schmerz
Karies	0.04	0.06
¬Karies	0.01	0.89

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

Bedingte Wahrscheinlichkeiten I

Neue Information kann Wahrscheinlichkeit eines Ereignisses ändern

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

Bei Zusatzinformation: nicht mehr mit a priori Wahrscheinlichkeiten rechnen!

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

$$P(\text{Karies} \mid \text{Schmerz}) = 0.8$$

$P(X \mid Y)$: Tabelle aller bedingten Wahrscheinlichkeiten über alle Werte von X und Y

Bedingte Wahrscheinlichkeiten II

$P(\text{Wetter} \mid \text{Schmerz})$: (4×2)-Tabelle von bedingten Wahrscheinlichkeiten aller Kombinationen der Werte einer Zufallsvariablen

	Schmerz = wahr	Schmerz = falsch
Wetter = sonnig	$P(W = \text{sonnig} \mid S)$	$P(W = \text{sonnig} \mid \neg S)$
Wetter = Regen		
Wetter = bewölkt		
Wetter = Schnee		

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

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

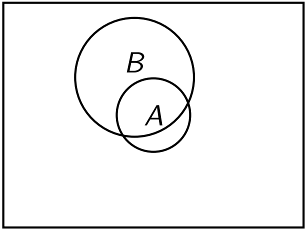
Bedingte Wahrscheinlichkeiten III

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

$$
\begin{aligned}
P(W = \text{sonnig} \land S) &= P(W = \text{sonnig} \mid S)P(S) \\
P(W = \text{Regen} \land S) &= P(W = \text{Regen} \mid S)P(S) \\
\vdots &= \vdots \\
P(W = \text{Schnee} \land \neg S) &= P(W = \text{Schnee} \mid \neg S)P(\neg S)
\end{aligned}
$$

Bedingte Wahrscheinlichkeiten IV

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



Produktregel: $P(A \wedge B) = P(A \mid B)P(B)$
Analog: $P(A \wedge B) = P(B \mid A)P(A)$
A und B heißen **unabhängig** voneinander, falls $P(A \mid B) = P(A)$ und $P(B \mid A) = P(B)$
A und B unabhängig $\iff P(A \wedge B) = P(A)P(B)$

Interpretation bedingter Wahrscheinlichkeiten

Die Interpretation:

„$P(A \mid B) = 0.8$ bedeutet, dass $P(A) = 0.8$ ist
falls B wahr ist“

ist aus mindestens zwei Gründen falsch!

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

Rechnen mit der Verbundwahrscheinlichkeit

Alle interessanten Wahrscheinlichkeiten sind aus Verbundwahrscheinlichkeit errechenbar

Beispiel:
$$P(\text{Karies} \lor \text{Schmerz}) = P(\text{Karies} \land \text{Schmerz}) + P(\neg\text{Karies} \land \text{Schmerz}) + P(\text{Karies} \land \neg\text{Schmerz})$$

Marginalisierung durch Aufsummieren von Zeile oder Spalte:

$$P(\text{Karies}) = P(\text{Karies} \land \text{Schmerz}) + P(\text{Karies} \land \neg\text{Schmerz})$$

Bedingte Wahrscheinlichkeiten ebenfalls ableitbar:

$$P(\text{Karies} \mid \text{Schmerz}) = \frac{P(\text{Karies} \land \text{Schmerz})}{P(\text{Schmerz})} = \frac{0.04}{0.04 + 0.01} = 0.80$$

Die Bayes'sche Regel

Es gilt

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

Durch Gleichsetzen der rechten Seiten folgt:

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

Für mehrwertige Variablen erhält man mehrere Gleichungen:

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

Verallgemeinerung (bzgl. Hintergrundevidenzen):

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

Anwendung der Bayes'schen Regel

$$P(\text{Schmerz} \mid \text{Karies}) = 0.4$$
$$P(\text{Karies}) = 0.1$$
$$P(\text{Schmerz}) = 0.05$$
$$P(\text{Karies} \mid \text{Schmerz}) = \frac{0.4 \cdot 0.1}{0.05} = 0.8$$

Warum nicht gleich $P(\text{Karies} \mid \text{Schmerz})$ schätzen?

Kausales Wissen wie $P(\text{Schmerz} \mid \text{Karies})$ ist i.A. robuster als diagnostisches Wissen $P(\text{Karies} \mid \text{Schmerz})$:

Kausalität $P(\text{Schmerz} \mid \text{Karies})$ ist unabhängig von a priori Wahrscheinlichkeiten $P(\text{Schmerz})$ und $P(\text{Karies})$

Nähme $P(\text{Karies})$ bei Karies-Epidemie zu, so bliebe Kausalität $P(\text{Schmerz} \mid \text{Karies})$ unverändert, während sich $P(\text{Schmerz})$ proportional mit $P(\text{Schmerz})$ und $P(\text{Karies})$ änderte

Ein Arzt, der $P(\text{Karies} \mid \text{Schmerz})$ geschätzt hätte, wüsste keine Regel zum Update von $P(\text{Karies} \mid \text{Schmerz})$

Ein Arzt, der Bayes'sche Regel benutzt hätte, wüsste um proportionales Verhalten zwischen $P(\text{Schmerz} \mid \text{Karies})$ und $P(\text{Karies})$

Übersicht

Multiple Evidenzen I

Patient bejaht Frage nach Zahnschmerzen.
Aus dieser ersten Evidenz schließt Zahnarzt:

$$P(\text{Karies} \mid \text{Schmerz}) = 0.8$$

Zahnarzt ertastet bei Untersuchung mit Haken Bruchstelle im Zahn.
Aus dieser zweiten Evidenz schließt er:

$$P(\text{Karies} \mid \text{Bruch}) = 0.95$$

Beide Schlüsse könnten z.B. mit Bayes'scher Regel berechnet werden, aber was bringt kombinierte Evidenz?
Mit Bayes'scher Regel könnte er weiter ermitteln:

$$P(\text{Karies} \mid \text{Schmerz}{\wedge}\text{Bruch}) = \frac{P(\text{Schmerz} \wedge \text{Bruch} \mid \text{Karies}) \cdot P(\text{Karies})}{P(\text{Schmerz} \wedge \text{Bruch})}$$

Multiple Evidenzen II

Problem: Er braucht $P(\text{Schmerz} \wedge \text{Bruch} \mid \text{Karies})$, d.h. Diagnosewissen für alle Kombinationen von Symptomen

Besser: Evidenzen mit Hilfe der Evidenzenregel schrittweise aufnehmen

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

Mit bestimmter a priori Wahrscheinlichkeit hat Patient Loch: $P(\text{Karies})$
Er berichtet von Zahnschmerzen (Bayes'sche Regel):

$$P(\text{Karies} \mid \text{Schmerz}) = P(\text{Karies}) \cdot \frac{P(\text{Schmerz} \mid \text{Karies})}{P(\text{Schmerz})}$$

Multiple Evidenzen III

$$P(\text{Karies} \mid \text{Schmerz}) = P(\text{Karies}) \cdot \frac{P(\text{Schmerz} \mid \text{Karies})}{P(\text{Schmerz})} \tag{1}$$

Untersuchung ergibt *Bruch*, also

$$P(\text{Karies} \mid \text{Bruch}, \text{Schmerz}) = \frac{P(\text{Bruch} \mid \text{Karies}, \text{Schmerz})}{P(\text{Bruch} \mid \text{Schmerz})} \cdot P(\text{Karies} \mid \text{Schmerz}) \tag{2}$$

(1) in (2) einsetzen ergibt

$$P(\text{Karies} \mid \text{Bruch}, \text{Schmerz}) = P(\text{Karies}) \cdot \frac{P(\text{Schmerz} \mid \text{Karies})}{P(\text{Schmerz})} \cdot \frac{P(\text{Bruch} \mid \text{Karies}, \text{Schmerz})}{P(\text{Bruch} \mid \text{Schmerz})}$$

Multiple Evidenzen IV

Annahme bedingter Unabhängigkeit von *Schmerz* und *Bruch* gegeben *Karies* (vereinfachtes Diagnosewissen):

$$P(\text{Bruch} \mid \text{Karies}, \text{Schmerz}) = P(\text{Bruch} \mid \text{Karies})$$
$$P(\text{Schmerz} \mid \text{Karies,Schmerz}) = P(\text{Schmerz} \mid \text{Karies})$$

$$P(\text{Karies} \mid \text{Bruch}, \text{Schmerz}) = P(\text{Karies}) \cdot \frac{P(\text{Schmerz} \mid \text{Karies})}{P(\text{Schmerz})} \cdot \frac{P(\text{Bruch} \mid \text{Karies})}{P(\text{Bruch} \mid \text{Schmerz})}$$

Multiple Evidenzen V

$$P(\text{Karies} \mid \text{Bruch}, \text{Schmerz}) = P(\text{Karies}) \cdot \frac{P(\text{Schmerz} \mid \text{Karies})}{P(\text{Schmerz})} \cdot \frac{P(\text{Bruch} \mid \text{Karies})}{P(\text{Bruch} \mid \text{Schmerz})}$$

Betrachtung von Paaren (Tripeln etc.) von Symptomen wie $P(\text{Bruch} \mid \text{Schmerz})$ ist notwendig, jedoch mit Produktregel folgt:

$$P(\text{Schmerz}) \cdot P(\text{Bruch} \mid \text{Schmerz}) = P(\text{Bruch} \wedge \text{Schmerz})$$

Man erhält:

$$P(\text{Karies} \mid \text{Bruch}, \text{Schmerz}) = P(\text{Karies}) \cdot \frac{P(\text{Schmerz} \mid \text{Karies}) \cdot P(\text{Bruch} \mid \text{Karies})}{P(\text{Bruch} \wedge \text{Schmerz})}$$

für Bestimmung von $P(\text{Karies} \mid \text{Bruch}, \text{Schmerz})$ ist $P(\text{Bruch} \wedge \text{Schmerz})$ Normierungsfaktor und eliminierbar, sofern $P(\text{Bruch} \mid \neg\text{Karies})$ und $P(\text{Schmerz} \mid \neg\text{Karies})$ bekannt

Multiple Evidenzen VI

Somit gilt

$$\begin{aligned} P(\text{Bruch} \wedge \text{Schmerz}) = {} & P(\text{Bruch} \wedge \text{Schmerz} \mid \text{Karies})P(\text{Karies}) \\ & + P(\text{Bruch} \wedge \text{Schmerz} \mid \neg\text{Karies})P(\neg\text{Karies}) \end{aligned}$$

und wegen der bedingten Unabhängigkeit folgt

$$\begin{aligned} P(\text{Bruch} \wedge \text{Schmerz}) = {} & P(\text{Schmerz} \mid \text{Karies})P(\text{Bruch} \mid \text{Karies})P(\text{Karies}) \\ & + P(\text{Schmerz} \mid \neg\text{Karies})P(\text{Bruch} \mid \neg\text{Karies})P(\neg\text{Karies}) \end{aligned}$$

und

$$\begin{aligned} P(\text{Karies} \mid \text{Bruch}, \text{Schmerz}) = {} & P(\text{Karies}) \\ & \cdot \frac{P(\text{Schmerz} \mid \text{Karies}) \cdot P(\text{Bruch} \mid \text{Karies})}{P(\text{Bruch} \wedge \text{Schmerz})} \end{aligned}$$

Multiple Evidenzen VII

Mehrfache Evidenzen sind berechenbar durch Reduktion auf

a priori Wahrscheinlichkeiten und

(kausale) bedingte Wahrscheinlichkeiten für Evidenz.

Zusätzlich werden gewisse bedingte Unabhängigkeiten vorausgesetzt.

Allgemeine Kombinationsregel:

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

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

Beispiel I

Expertenwissen:
„Metastatischer Krebs ist eine mögliche Ursache von Hirntumor, und ist auch eine Erklärung für erhöhte Kalziumwerte im Blut. Beide Phänomene zusammen können erklären, dass ein Patient ins Koma fällt. Schwere Kopfschmerzen sind auch möglicherweise verbunden mit einem Hirntumor.“

Spezieller Fall:
“Der Patient hat schwere Kopfschmerzen.”

Anfrage:
Wird der Patient ins Koma fallen?

Wahl des Zustandsraums

Ω ist endliche Menge von erschöpfenden, sich gegenseitig ausschließenden Werten

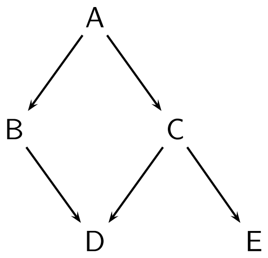
Beispiel:
Explizite Spezifikation der Zufallsvariablen und ihrer Werte

Zufallsvariable		**Mögliche Werte**
A	metastatischer Krebs	$a, \neg a$
B	erhöhter Calciumwert im Blut	$b, \neg b$
C	Hirntumor	$c, \neg c$
D	Koma	$d, \neg d$
E	schwere Kopfschmerzen	$e, \neg e$

Wahrscheinlichkeit für geg. Abhängigkeiten I

$$
\begin{aligned}
&P(A = a, B = b, C = c, D = d, E = e)\\
&= P(E = e \mid A = a, B = b, C = c, D = d)\\
&\cdot P(D = d \mid A = a, B = b, C = c)\\
&\cdot P(C = c \mid A = a, B = b)\\
&\cdot P(B = b \mid A = a) \cdot P(A = a)
\end{aligned}
$$

$$
\begin{aligned}
&= P(E = e \mid C = c)\\
&\cdot P(D = d \mid B = b, C = c)\\
&\cdot P(C = c \mid A = a)\\
&\cdot P(B = b \mid A = a)\\
&\cdot P(A = a)
\end{aligned}
$$



Was muss vom Experten spezifiziert werden?

a priori Wahrscheinlichkeiten der Knoten ohne Eltern

bedingte Wahrscheinlichkeiten eines Knotens bezogen auf Menge von elementaren Knoten (Markov-Feld)

Wahrscheinlichkeit für geg. Abhängigkeiten II

$P(e \mid c) = 0.8$ $P(e \mid \neg c) = 0.6$	Kopfschmerzen sind häufig, aber viel häufiger, falls Tumor vorhanden ist
$P(d \mid b, c) = 0.8$ $P(d \mid b, \neg c) = 0.8$ $P(d \mid \neg b, c) = 0.8$ $P(d \mid \neg b, \neg c) = 0.05$	Koma ist selten, aber häufig, falls einer der beiden Gründe vorliegt
$P(b \mid a) = 0.8$ $P(b \mid \neg a) = 0.2$	erhöhtes Kalzium ist ungewöhnlich, aber eine häufige Konsequenz von Metastasen
$P(c \mid a) = 0.2$ $P(c \mid \neg a) = 0.05$	Hirntumor ist selten, aber häufige Konsequenz von Metastasen
$P(a) = 0.2$	Vorkommen von metastatischem Krebs in relevanten Studien

11 Werte (anstatt 31) spezifiziert, andere berechenbar

Wahrscheinlichkeiten anhand von Fallstudien, Lehrbuchinformationen und massiven Tests

Lösung des Problems

Gegeben:	Wahrscheinlichkeit P
Evidenz:	Patient hat starke Kopfschmerzen
Frage:	Wie groß ist die Wahrscheinlichkeit, dass der Patient ins Koma fallen wird?

Evidenz:	$E = e$
Bedingte Wahrscheinlichkeit:	$P(\bullet \mid E = e)$
Frage:	$D = d?$
Randwahrscheinlichkeit:	$P(D = d \mid E = e)$

Analyse dieses speziellen Falles

„Der Patient hat starke Kopfschmerzen"

Idee: Beschreibung der bedingten Wahrscheinlichkeiten $P(\bullet \mid E = e)$

$$P(A = a', B = b', C = c', D = d', E = e' \mid E = e)$$

$$= \begin{cases} \frac{P(A=a',B=b',C=c',D=d',E=e')}{\sum_{a \in A} \sum_{b \in B} \sum_{c \in C} \sum_{d \in D} P(a,b,c,d,e)} & \text{falls } e' = e \\ 0 & \text{falls } e' = \neg e \end{cases}$$

Auswertung der Daten des Patienten

Wird der Patient ins Koma fallen?

Berechne Randwahrscheinlichkeiten $P(\{d\} \mid \{e\})$

$$P(\{d\} \mid \{e\}) = \sum_{a' \in A} \sum_{b' \in B} \sum_{c' \in C} \sum_{e' \in E} P(a', b', c', d, e' \mid e)$$

Probleme:

- Komplexität der Berechnungen der bedingten Randwahrscheinlichkeiten (100 Zufallsvariablen mit 3 Merkmalen: 3^{99} Summationen von kleinen Werten)
- Wie bestimmt man die Wahrscheinlichkeiten?
- Lösung: Bayes-Netze!

Übersicht

1. Motivation

2. Grundlagen der Wahrscheinlichkeitstheorie

3. Probabilistische Inferenz

4. **Bayes'sche Netze**

Bayes'sche Netze

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

Belief Netz

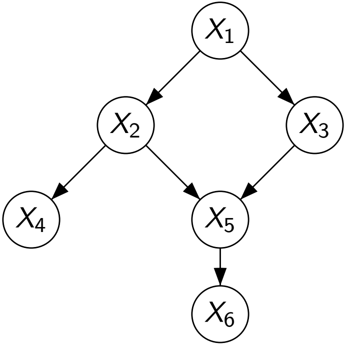
Ein *Belief Netz* (V, E, P) besteht aus einer Menge $V = \{X_1, \ldots, X_n\}$ von Zufallsvariablen und einer Menge E von gerichteten Kanten zwischen diesen Variablen.

Jede Variable hat eine endliche Menge von sich gegenseitig ausschließenden und in ihrer Vereinigung erschöpfenen Zuständen.

Die Variablen formen in Verbindung mit den Kanten einen azyklischen, gerichteten Graphen.

Jeder Variable X_i mit Elternknoten $B_1, \ldots, B_m$ wird eine Tabelle $P(X_i \mid B_1, \ldots, B_m)$ zugewiesen.

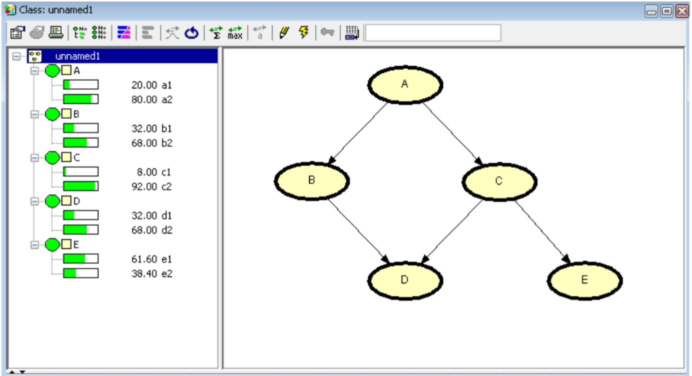
Probabilistiche Graphische Netze



Im Allgemeinen (nach Kettenregel):

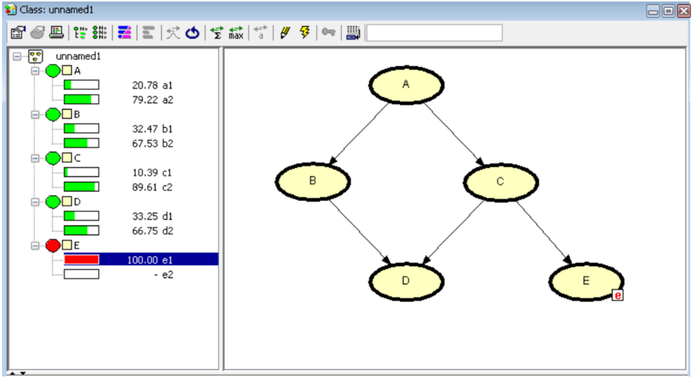
$$
\begin{aligned}
P(X_1, \ldots, X_6) = {} & P(X_6 \mid X_5, \ldots, X_1) \cdot \\
& P(X_5 \mid X_4, \ldots, X_1) \cdot \\
& P(X_4 \mid X_3, X_2, X_1) \cdot \\
& P(X_3 \mid X_2, X_1) \cdot \\
& P(X_2 \mid X_1) \cdot \\
& P(X_1)
\end{aligned}
$$

Beispiel 1 (Fortsetzung)



Randwahrscheinlichkeiten im HUGIN-Tool

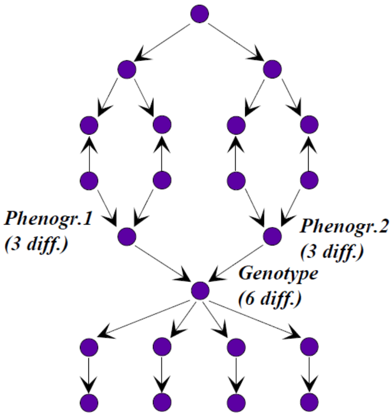
Beispiel 1 (Fortsetzung)



bedingte Randwahrscheinlichkeiten mit Evidenz $E = e_1$

Beispiel: Genotyp-Bestimmung des Jersey-Rinds

22 Variablen, $6 \cdot 10^{13}$ Zustände, 324 Parameter



graphisches Modell

Knoten $\rightarrow$ Zustandsvariablen

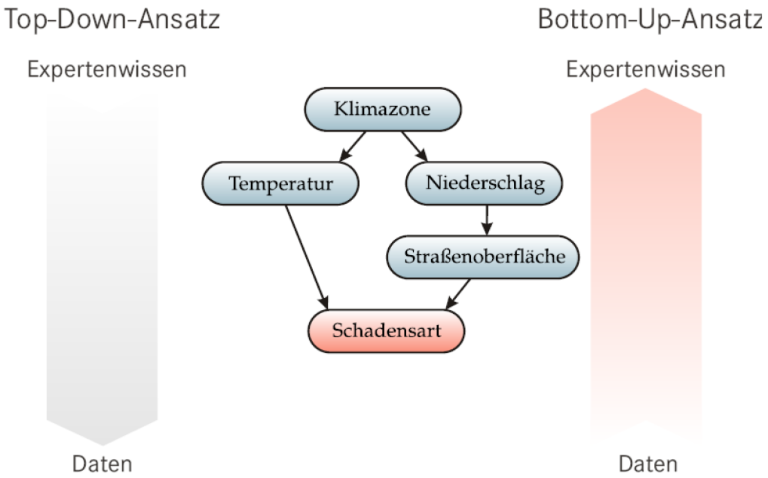
Kanten $\rightarrow$ Bedingte Abhängigkeiten

Zerlegung $\rightarrow$

$$P(X_1, \ldots, X_{22}) = \prod_{i=1}^{22} P(X_i \mid \text{Eltern}(X_i))$$

Diagnose $\rightarrow$ $P(\bullet \mid \text{Evidenz})$

Bayes'sche Netze



Beispiel IV: Planungssystem für VW

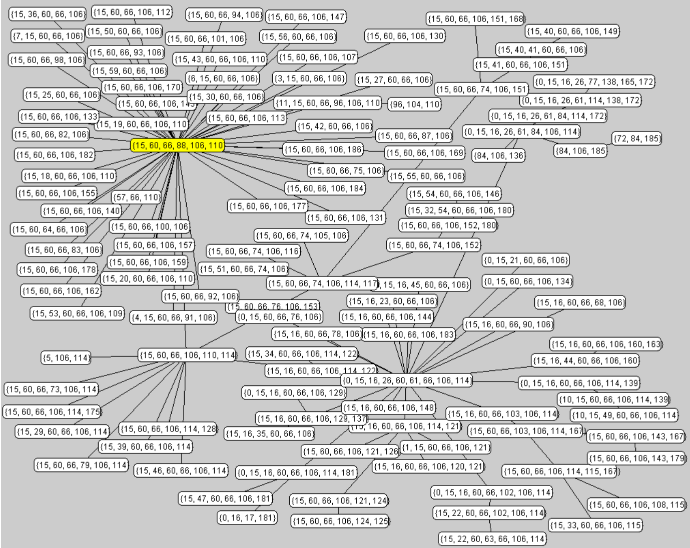
- etwa **200 item families** (Variablen, Merkmale)
- von **2 bis 50 items pro item family** (Werte pro Variable)
- mehr als 2^{200} mögliche Fahrzeug-Spezifikationen
- Wahl der gültigen Spezifikationen ist beschränkt durch **Regelsysteme** (**10.000 technische Regeln**, noch mehr marketing- und produktionsorientierte Regeln)

Beispiel:

$$\textbf{if } \mathit{engine} = e_1 \textbf{ then } \mathit{transmission} = t_3$$

$$\textbf{if } \mathit{engine} = e_4 \textbf{ and } \mathit{auxiliary\ heater} = h_1$$
$$\textbf{then } \mathit{generator} \in \{g_3, g_4, g_5\}$$

Reales Bayes'sches Netz



Zusammenfassung

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