Intelligente Systeme

Einführung

Prof. Dr. Rudolf Kruse   Georg Ruß
Christian Moewes

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

Arbeitsgruppe Computational Intelligence
Institut für Wissens- und Sprachverarbeitung
Fakultät für Informatik
Otto-von-Guericke Universität Magdeburg

Unsicherheit

Gliederung der Vorlesung

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!"
- 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 die Vorlesung wichtig ist, und die Erfolgsrate bei P_1 nur bei ca. 85–95% liegt.

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 ($\Rightarrow$ 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

Algorithm 1 DT-Agent

Input: a *percept*

Output: 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: 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.

Gliederung der Vorlesung

1. Motivation

2. Grundlagen der Wahrscheinlichkeitstheorie

3. Probabilistische Inferenz

4. Bayes'sche Netze

Unbedingte Wahrscheinlichkeiten I

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

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

- *Cavity* is eine Proposition (Aussage).
- A priori Wahrscheinlichkeiten gewinnt man durch statistische Analyse oder aus allgemeinen Regeln.

Unbedingte Wahrscheinlichkeiten II

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

$$\begin{aligned} P(\text{Weather} = \text{Sunny}) &= 0.7 \\ P(\text{Weather} = \text{Rain}) &= 0.2 \\ P(\text{Weather} = \text{Cloudy}) &= 0.08 \\ P(\text{Weather} = \text{Snow}) &= 0.02 \\ P(\text{Headache} = \text{TRUE}) &= 0.1 \end{aligned}$$

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

$$P(\text{Cavity} \wedge \neg\text{Injured}) = 0.06$$

Unbedingte Wahrscheinlichkeiten III

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

$$\boldsymbol{P}(\text{Headache}) = \langle 0.1, 0.9 \rangle$$
$$\boldsymbol{P}(\text{Weather}) = \langle 0.7, 0.2, 0.08, 0.02 \rangle$$

- diese Vektoren definieren die Wahrscheinlichkeitsverteilung der Zufallsvariablen *Headache* und *Weather*.
- $P(\text{Headache}, \text{Weather})$ ist eine (4×2)-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 I

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

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

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

Bedingte Wahrscheinlichkeiten II

- $P(\text{Weather} \mid \text{Headache})$ ist eine (4×2)-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 $P(B) > 0$):

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

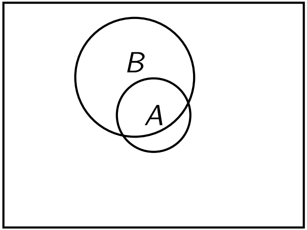
Bedingte Wahrscheinlichkeiten III

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

$$
\begin{aligned}
P(W = \text{Sunny} \wedge \text{Headache}) &= P(W = \text{Sunny} \mid \text{Headache})P(\text{Headache}) \\
P(W = \text{Rain} \wedge \text{Headache}) &= P(W = \text{Rain} \mid \text{Headache})P(\text{Headache}) \\
\vdots &= \vdots \\
P(W = \text{Snow} \wedge \neg\text{Headache}) &= P(W = \text{Snow} \mid \neg\text{Headache})P(\neg\text{Headach}
\end{aligned}
$$

Bedingte Wahrscheinlichkeiten III

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



- 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 $P(B \mid A) = P(B)$.
- Dann und nur dann gilt $P(A \wedge B) = P(A)P(B)$!

Interpretation bedingter Wahrscheinlichkeiten

Achtung: häufiger Interpretationsfehler:

„$P(A \mid B) = 0.8$ 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. $P(A \mid B) = 0.8$ 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. Im Allgemeinen gilt $P(A \mid B \wedge C) \neq P(A \mid B)$; 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. $P(\text{True}) = 1$,
3. $P(\text{False}) = 0$,
4. $P(A \vee B) = P(A) + P(B) - P(A \wedge B)$.

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

$$
\begin{aligned}
&P(\neg A) = 1 - P(A) \text{ folgt aus} \\
&P(A \vee \neg A) = 1 \text{ und} \\
&P(A \wedge \neg A) = 0.
\end{aligned}
$$

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

	Toothache	¬Toothache
Cavity	0.04	0.06
¬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:
$$\begin{aligned} P(\text{Cavity} \lor \text{Tootache}) &= P(\text{Cavity} \land \text{Tootache}) \\ &\quad + P(\neg\text{Cavity} \land \text{Tootache}) \\ &\quad + P(\text{Cavity} \land \neg\text{Tootache}) \end{aligned}$$

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

$$P(\text{Cavity}) = P(\text{Cavity} \land \text{Tootache}) + P(\text{Cavity} \land \neg\text{Tootache})$$

Womit wieder bedingte Wahrscheinlichkeiten ableitbar sind:

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

Die Bayes'sche Regel

Wir wissen (Produktregel):

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

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 (Menge von Gleichungen):

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

Verallgemeinerung (bzgl. Hintergrundevidenzen):

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

Anwendung der Bayes'schen Regel

$$P(\text{Tootache} \mid \text{Cavity}) = 0.4$$
$$P(\text{Cavity}) = 0.1$$
$$P(\text{Tootache}) = 0.05$$
$$P(\text{Cavity} \mid \text{Toothache}) = \frac{0.4 \cdot 0.1}{0.05} = 0.8$$

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

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

Relative Wahrscheinlichkeit

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

$$\begin{aligned} P(\text{Tootache} \mid \text{GumDisease}) &= 0.7 \\ P(\text{GumDisease}) &= 0.02 \end{aligned}$$

Welche Diagnose ist wahrscheinlicher?
Wenn nur die **relative Wahrscheinlichkeit** der Ursachen interessiert, muss die $P(T)$ nicht geschätzt werden:

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

Relative Wahrscheinlichkeiten können für einige Entscheidungen hinreichend sein!

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

Normierung I

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

$$
\begin{aligned}
P(C \mid T) &= \frac{P(T \mid C)P(C)}{P(T)} \\
P(\neg C \mid T) &= \frac{P(T \mid \neg C)P(\neg C)}{P(T)} \\
P(C \mid T) + P(\neg C \mid T) &= \frac{P(T \mid C)P(C)}{P(T)} + \frac{P(T \mid \neg C)P(\neg C)}{P(T)} \\
P(T) &= P(T \mid C)P(C) + P(T \mid \neg C)P(\neg C)
\end{aligned}
$$

Normierung II

Durch Einsetzen in die oberste Gleichung:

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

Für mehrwertige Zufallsvariablen:

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

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

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

Gliederung der Vorlesung

1. Motivation

2. Grundlagen der Wahrscheinlichkeitstheorie

3. **Probabilistische Inferenz**

4. Bayes'sche Netze

Multiple Evidenzen I

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 Bayes'schen Regel berechnet worden sein. Aber was bringt die kombinierte Evidenz?

Mit der Bayes'schen Regel könnte er weiter ermitteln:

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

Multiple Evidenzen II

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

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

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

Mit einer bestimmten a priori Wahrscheinlichkeit hat der Patientein Loch: $P(\text{Cav})$. Er berichtet von Zahnschmerzen (Bayes'sche Regel):

$$P(\text{Cav} \mid \text{Tooth}) = P(\text{Cav}) \cdot \frac{P(\text{Tooth} \mid \text{Cav})}{P(\text{Tooth})}$$

Multiple Evidenzen III

$$P(\text{Cav} \mid \text{Tooth}) = P(\text{Cav}) \cdot \frac{P(\text{Tooth} \mid \text{Cav})}{P(\text{Tooth})} \tag{1}$$

Die Untersuchung ergibt *Catch*, also

$$P(\text{Cav} \mid \text{Catch}, \text{Tooth}) = \frac{P(\text{Catch} \mid \text{Cav}, \text{Tooth})}{P(\text{Catch} \mid \text{Tooth})} \cdot P(\text{Cav} \mid \text{Tooth}) \tag{2}$$

(1) in (2) einsetzen ergibt

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

Multiple Evidenzen IV

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

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

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

Multiple Evidenzen V

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

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

$$P(\text{Tooth}) \cdot P(\text{Catch} \mid \text{Tooth}) = P(\text{Catch} \wedge \text{Tooth})$$

Dies führt zu:

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

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

Multiple Evidenzen VI

Somit:

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

Mit bedingter Unabhängigkeit:

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

Einsetzen in:

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

Also Ableitung möglich aus unbedingten Wahrscheinlichkeiten und einfachen kausalen Regeln!

Multiple Evidenzen VII

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

Allgemeine Kombinationsregel:

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

fur X und Y bedingt unabhangig 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 I

- **Expertenwissen:**
„Metastatischer Krebs ist eine mögliche Ursache von Hirntumor, und ist auch eine Erklärung für erhöhte Kalziumwerte im Blut. Wiederum beider Phänomene 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 eine endliche Menge von erschöpfenden, sich gegenseitig ausschließenden Werten.

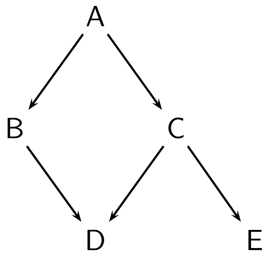
Beispiel:
Explizite Spezifikation der Attribute und ihrer Werte.

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

Wahrscheinlichkeit für gegebene Abhänigkeiten 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?

- Apriori-Wahrscheinlichkeiten der Knoten ohne Eltern
- Bedingte Wahrscheinlichkeiten eines Knotens bezogen auf Menge von elemenaren Knoten (Markov-Feld)

Wahrscheinlichkeit für gegebene Abhänigkeiten 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 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
$P(a) = 0.2$	Vorkommnis von metastatischem Krebs in relevanten Studien

- 11 Werte (anstatt 31) spezifiziert, anderen berechenbar
- Wahrscheinlichkeiten werden anhand von Fallstudien, Lehrbuchinformationen und massiven Tests gewonnen.

Mathematische 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 $P(\bullet \mid \{e\})$

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

Wird der Patient ins Koma fallen?

- **Idee:**
  Berechne Randwahrscheinlichkeiten $P(\{d\} \mid \{e\})$

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

- **Probleme:**
  - Komplexität der Berechnungen der bedingten Randwahrscheinlichkeiten (100 Attribute mit 3 Merkmalen: 3^{99} Summationen von kleinen Werten)
  - Können (oder müssen) wir erwarten, dass Nutzer mit Wahrscheinlichkeiten im Produktraum zurechtkommen.

Gliederung der Vorlesung

1. Motivation

2. Grundlagen der Wahrscheinlichkeitstheorie

3. Probabilistische Inferenz

4. Bayes'sche Netze

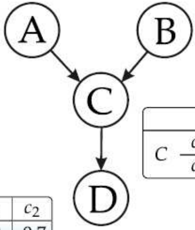
Bayes'sche Netze

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

Wahrscheinlichkeit des Ereignisses B, falls Ereignis A bekannt ist

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

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



		ϕ
A	a_1	0.4
	a_2	0.6

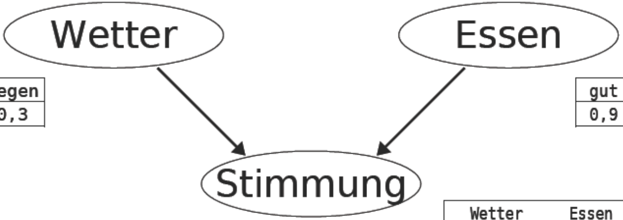
		ϕ
B	b_1	0.7
	b_2	0.3

		a_1, b_1	a_1, b_2	a_2, b_1	a_2, b_2
C	c_1	0.9	0.5	0.2	0.4
	c_2	0.1	0.5	0.8	0.6

		c_1	c_2
D	d_1	0.2	0.7
	d_2	0.8	0.3

Beispiel II

- Knoten $\longrightarrow$ Attribute
- Kanten $\longrightarrow$ Abhängigkeiten der Variablen



Sonne	Regen
0,7	0,3

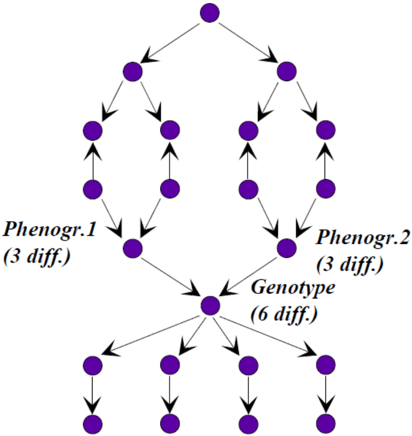
gut	schlecht
0,9	0,1

gut	schlecht
0,81	0,19

Wetter	Essen	gut	schlecht
Sonne	gut	0,95	0,05
Sonne	schlecht	0,70	0,30
Regen	gut	0,60	0,40
Regen	schlecht	0,10	0,90

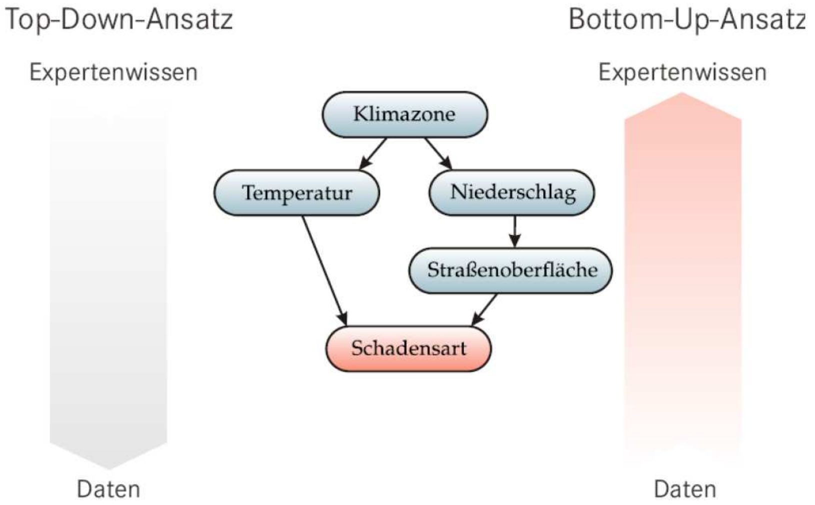
Beispiel III: Genotyp-Bestimmung des Jersey-Rinds

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



- Graphisches Modell
- Knoten $\rightarrow$ Zustandsvariablen
- Kanten $\rightarrow$ Bedingte Abhängigkeiten
- Zerlegung $\rightarrow P(X_1, K, X_{22}) = \prod_{i=1}^{22} P(X_i \mid \text{Eltern}(X_i))$
- Diagnose $\rightarrow P(\bullet \mid \text{Evidenz})$

Bayes'sche Netze

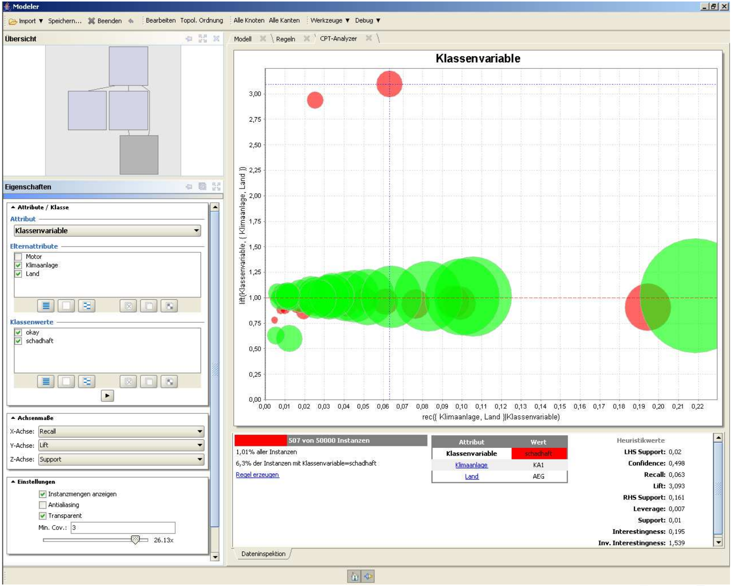


Erlernen von Bayes'schen Netzen aus Daten

- Erstellung eines BayesschenNetzes 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 Bayes'scher 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

Visual Data Mining



Beispiel IV: Planungssystem für VW

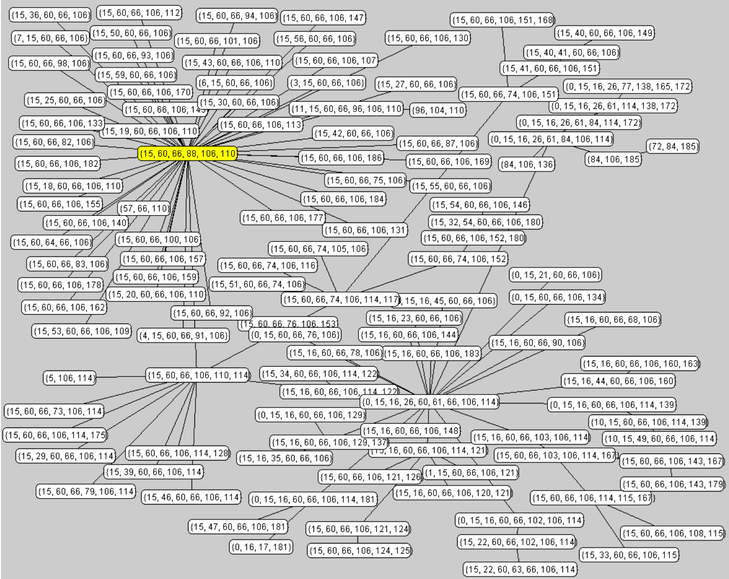
- etwa **200 item families** (Variablen, Merkmale)
- von **2 bis 50 items pro item family** (Werte pro Variable)
- mehr als 2^{200} mögliche Fahrzeug-Spezifikationen
- Wahl der gültigen Spezifikationen ist beschränkt durch **Regelsysteme** (**10.000 technische Regeln**, noch mehr marketing- und produktionsorientierte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

Beispiel:

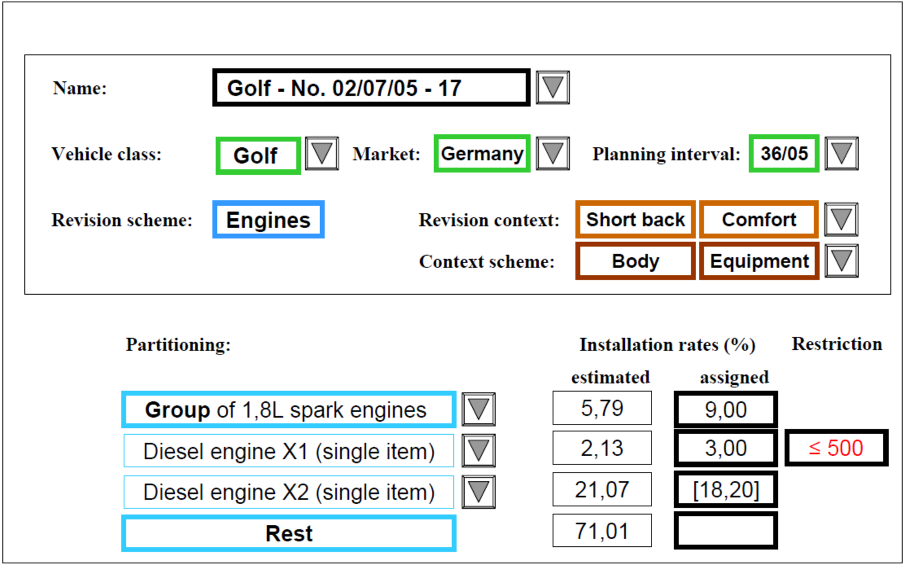
$$\textbf{if } \textit{engine} = e_1 \textbf{ then } \textit{transmission} = t_3$$

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

Reales Bayes'sches Netz



Einfache Benutzerschnittstelle



Software-Entwicklung

- Client-Server-System
- Server mit 6–8 Maschinen (jeder 16 GB)
- AMD Opteron Quad-Prozessor
- Terabyte-Speichersystem
- Betriebssystem: Linux

- 15 Entwickler
- Programmiersprache: Java
- WebSphere Application Developer, Eclipse
- DB-System: Oracle
- weltweit bereits eingeführt