

1 Umgang mit Unsicherheit und Probabilistisches Schließen

Dieses Kapitel behandelt die Grundlagen der Unsicherheit in der Künstlichen Intelligenz, die Wahrscheinlichkeitstheorie als Werkzeug zur Modellierung von Glaubensgraden (Degrees of Belief) sowie Bayesian Networks zur kompakten Repräsentation von Wissen und Inferenzmethoden.

1.1 Einführung in Unsicherheit

In der klassischen Logik gehen Agenten davon aus, dass Aussagen entweder wahr oder falsch sind und Aktionen deterministische Ergebnisse haben. Die reale Welt ist jedoch oft unsicher.

Gründe für Unsicherheit

- **Partielle Beobachtbarkeit:** Der Agent hat keinen Zugriff auf den vollständigen Zustand der Welt (z.B. unbekannte Verkehrslage).
- **Rauschende Sensoren:** Messwerte können fehlerhaft sein.
- **Unsicherheit in Aktionsergebnissen:** Aktionen gelingen nicht immer wie geplant (z.B. Reifenpanne).
- **Komplexität:** Es ist unmöglich, alle Eventualitäten (“Qualification Problem”) in logischen Regeln zu modellieren (z.B. “Ich komme pünktlich an, außer...”).

Anstatt strikter logischer Wahrheit modellieren wir **Wahrscheinlichkeiten** als Maß für den **Glaubensgrad** (Degree of Belief) eines Agenten. Eine Wahrscheinlichkeit von 0.1 für einen Stau bedeutet nicht, dass die Straße zu 10% verstopft ist, sondern dass der Agent in 10% der Fälle an das Vorliegen eines Staus glaubt.

1.2 Grundlagen der Wahrscheinlichkeitstheorie

1.2.1 Wahrscheinlichkeitsraum und Axiome

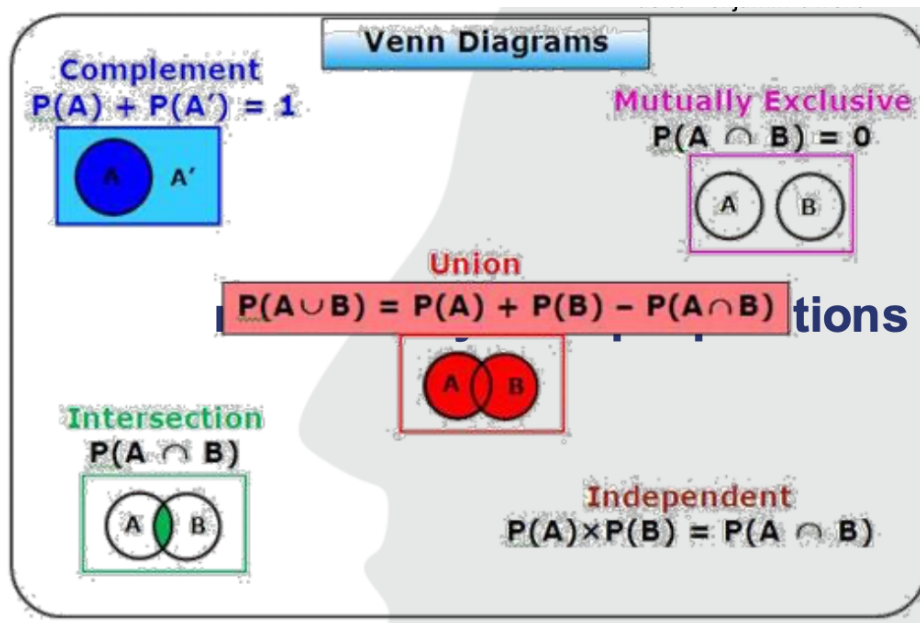
Ein Wahrscheinlichkeitsmodell besteht aus einem **Stichprobenraum** Ω (Menge aller möglichen Welten/Ergebnisse). Ein Ereignis A ist eine Teilmenge von Ω .

Kolmogorov-Axiome

Diese Axiome beschränken die Menge rationaler Glaubenssätze:

1. Alle Wahrscheinlichkeiten liegen zwischen 0 und 1: $0 \leq P(A) \leq 1$.
2. Wahre Aussagen haben Wahrscheinlichkeit 1, falsche 0: $P(\text{true}) = 1, P(\text{false}) = 0$.
3. Die Wahrscheinlichkeit einer Disjunktion (Vereinigung):

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



Ein Verstoß gegen diese Axiome führt zu irrationalem Verhalten. Dies wird durch das **Dutch Book Theorem** bewiesen: Wenn ein Agent Wahrscheinlichkeiten akzeptiert, die den Axiomen widersprechen, kann ein Gegner (Bookmaker) eine Wette so konstruieren, dass der Agent garantiert Geld verliert, unabhängig vom Ausgang des Ereignisses.

1.2.2 Zufallsvariablen

Anstatt mit atomaren Ereignissen zu arbeiten, nutzen wir **Zufallsvariablen** (Random Variables, RVs), um Komplexität zu reduzieren.

- **Diskret**: Endliche Menge von Werten (z.B. *Wetter* $\in \{\text{sonnig, regen, schnee}\}$). Werte müssen disjunkt und erschöpfend sein. **Kontinuierlich**: Unendlicher Wertebereich (z.B. Temperatur).

Eine Proposition (Aussage) wie *Wetter = regen* entspricht der Menge aller atomaren Ereignisse, in denen diese Aussage wahr ist.

1.3 Verteilungen und Rechenregeln

1.3.1 Joint Distribution (Verbundwahrscheinlichkeit)

Die **Joint Distribution** $P(X_1, \dots, X_n)$ weist jeder möglichen Kombination von Werten aller Zufallsvariablen eine Wahrscheinlichkeit zu.

$$P(x, y) = P(X = x \wedge Y = y)$$

Die Tabelle der Joint Distribution wächst exponentiell mit der Anzahl der Variablen ($O(2^n)$ bei binären Variablen), erlaubt aber die Beantwortung jeder beliebigen Wahrscheinlichkeitsanfrage.

1.3.2 Marginalisierung und Bedingte Wahrscheinlichkeit

Marginalisierung

Man kann Wahrscheinlichkeiten von Teilmengen von Variablen berechnen, indem man über die nicht interessierenden Variablen summiert ("Summing Out"):

$$P(Y) = \sum_{x \in X} P(x, Y)$$

Bedingte Wahrscheinlichkeit

Die Wahrscheinlichkeit von X , gegeben Evidenz Y :

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

Daraus folgt die **Produktregel**:

$$P(X, Y) = P(X|Y)P(Y) = P(Y|X)P(X)$$

Die **Kettenregel** (Chain Rule) erlaubt die Zerlegung einer Joint Distribution in bedingte Wahrscheinlichkeiten:

$$P(X_1, \dots, X_n) = \prod_{i=1}^n P(X_i | X_1, \dots, X_{i-1})$$

1.3.3 Satz von Bayes

Der Satz von Bayes ist fundamental für das Lernen aus Beobachtungen. Er erlaubt die Inversion von bedingten Wahrscheinlichkeiten.

$$\underbrace{P(H|E)}_{\text{Posterior}} = \frac{\underbrace{P(E|H)}_{\text{Likelihood}} \cdot \underbrace{P(H)}_{\text{Prior}}}{\underbrace{P(E)}_{\text{Evidenz/Marginalisierung}}}$$

Erweitert mit Marginalisierung im Nenner:

$$P(H|E) = \frac{P(E|H)P(H)}{P(E|H)P(H) + P(E|\neg H)P(\neg H)}$$

- **Prior**: Wahrscheinlichkeit der Hypothese vor Evidenz.
- **Likelihood**: Wahrscheinlichkeit der Evidenz, angenommen die Hypothese ist wahr.
- **Posterior**: Neue Wahrscheinlichkeit der Hypothese nach Beobachtung der Evidenz.

1.4 Bayesian Networks

Da die vollständige Joint Distribution ($O(2^n)$) speicher- und rechenintensiv ist, nutzen wir Unabhängigkeiten zur kompakten Darstellung.

1.4.1 Unabhängigkeit

Zwei Variablen X und Y sind **unabhängig**, wenn $P(X|Y) = P(X)$ bzw. $P(X, Y) = P(X)P(Y)$. Viel häufiger ist die **bedingte Unabhängigkeit**: X und Y sind unabhängig gegeben Z , wenn $P(X|Y, Z) = P(X|Z)$.

1.4.2 Definition und Semantik

Ein **Bayesian Network (BN)** ist ein gerichteter azyklischer Graph (DAG), bestehend aus:

- **Knoten**: Zufallsvariablen.
- **Kanten**: Direkte Abhängigkeiten ($X \rightarrow Y$ bedeutet X beeinflusst Y).
- **CPTs (Conditional Probability Tables)**: Jeder Knoten X_i hat eine Wahrscheinlichkeitsverteilung $P(X_i | Pa(X_i))$, wobei $Pa(X_i)$ die Elternknoten sind.

BN Semantik

Ein BN definiert die Joint Distribution als Produkt der lokalen bedingten Wahrscheinlichkeiten:

$$P(X_1, \dots, X_n) = \prod_{i=1}^n P(X_i | Pa(X_i))$$

Lokale Markov-Annahme: Ein Knoten ist bedingt unabhängig von all seinen Nicht-Nachfahren (Non-Descendants), gegeben seine Eltern.

1.5 Inferenz in Bayesian Networks

Inferenz beantwortet Anfragen der Form $P(Query|Evidence)$.

1.5.1 Exakte Inferenz: Variable Elimination

Variable Elimination vermeidet die Berechnung der vollen Joint Distribution, indem Summen “nach innen gezogen” werden. Der Prozess eliminiert Variablen, die weder Query noch Evidenz sind, nacheinander.

Algorithmus:

1. Schreibe die Joint Distribution als Produkt der Faktoren (CPTs).
2. Wähle eine Eliminationsreihenfolge für die versteckten Variablen.
3. Für jede zu eliminierende Variable Z :
 - Sammle alle Faktoren, die Z enthalten.
 - Multipliziere diese Faktoren zu einem neuen temporären Faktor.
 - Summiere Z aus diesem Faktor heraus (Marginalisierung).
 - Ersetze die alten Faktoren durch den neuen Faktor.
4. Normalisiere das Ergebnis am Ende.

Die Operationen sind:

- **Faktor-Produkt:** $f_3(A, B, C) = f_1(A, B) \cdot f_2(B, C)$.
- **Summing Out:** $f_{new}(B) = \sum_a f(a, B)$.

1.5.2 Komplexität

Exakte Inferenz in Bayesian Networks ist **NP-hard** (bewiesen durch Reduktion auf 3-SAT). In einfach verbundenen Netzen (Polytrees) ist die Komplexität linear, im schlimmsten Fall jedoch exponentiell.

1.6 Approximative Inferenz: Sampling

Da exakte Inferenz oft zu aufwendig ist, nutzen wir stochastische Verfahren (Monte Carlo Methoden), die gegen die wahre Wahrscheinlichkeit konvergieren.

1.6.1 Direct Sampling (ohne Evidenz)

Man zieht Stichproben in topologischer Reihenfolge (von Eltern zu Kindern) entsprechend der CPTs.

$$P(X) \approx \frac{N(X)}{N_{total}}$$

1.6.2 Rejection Sampling

Um $P(X|e)$ zu berechnen: Generiere Samples wie oben. Verwirf (reject) alle Samples, die der Evidenz e widersprechen. Zähle die verbleibenden Samples. Problem: Bei unwahrscheinlicher Evidenz werden fast alle Samples verworfen (ineffizient).

1.6.3 Markov Chain Monte Carlo (MCMC)

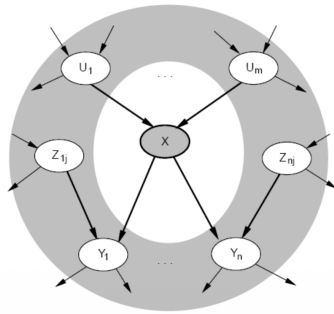
MCMC-Verfahren wie **Gibbs Sampling** wandern durch den Zustandsraum, anstatt Samples unabhängig zu generieren.

Markov Blanket

Der **Markov Blanket** einer Variable X besteht aus:

- Eltern von X
- Kindern von X
- Anderen Eltern der Kinder von X

X ist bedingt unabhängig von allen anderen Knoten im Netzwerk, gegeben seinen Markov Blanket.



$$P(X \mid U_1, \dots, U_m, Y_1, \dots, Y_n, Z_{1j}, \dots, Z_{nj}) = P(X \mid \text{allvariables})$$

1. Initialisiere alle Variablen mit zufälligen Werten (Evidenz fixieren).
2. Wiederhole für viele Schritte:
 - Wähle eine Nicht-Evidenz-Variable X_i .
 - Ziehe einen neuen Wert für X_i aus der Verteilung $P(X_i | \text{Markov Blanket}(X_i))$.
3. Schätze die Wahrscheinlichkeit basierend auf der Häufigkeit der Zustände in den gesammelten Samples.

Dies erzeugt eine Markov-Kette, deren stationäre Verteilung der gesuchten Posterior-Verteilung entspricht (unter der Bedingung, dass die Kette irreduzibel und ergodisch ist).