

**Technische Universität
München**

Fakultät für Informatik

Forschungs- und Lehrinheit Informatik VI

Bayesnetze

Seminar Kognitive Robotik (WS12)

Markus Kaiser

Betreuer: Alexander Perzylo

Leitung: Prof. Alois Knoll

Abgabetermin: 10. Februar 2013

Zusammenfassung

Bayesnetze repräsentieren mehrdimensionale Zufallsvariablen platzsparend aber exakt. Diese Arbeit führt die notwendigen Begriffe der Wahrscheinlichkeitstheorie ein und definiert die bedingte Unabhängigkeit anhand derer mit Hilfe der d-Separation die Semantik der Bayesnetze erarbeitet wird. Schließlich wird ein einfacher Algorithmus zur exakten Inferenz vorgestellt, die im Allgemeinen NP-vollständig ist, auf Polytrees jedoch in Linearzeit ausgeführt werden kann.

Inhaltsverzeichnis

1 Unsicheres Wissen	3
2 Bayesnetze	6
2.1 Aufbau	6
2.2 Semantik	7
3 Exakte Inferenz in Bayesnetzen	11

1 Unsicheres Wissen

Wissen ist selten sicher und exakt. Soll beispielsweise ein Arzt eine Diagnose stellen, so reicht perfekt logisches Schließen allein nicht aus. Jeder Patient hat eine Vorgeschichte, die unvollständig sein kann und Symptome, die vage sein können. Tests lassen mehrere alternative Schlüsse zu und Schulwissen ist fehlerbehaftet. Der Arzt muss diese Unsicherheiten abwägen und die plausibelste Diagnose und erfolgsversprechendste Therapie wählen. Soll dieses Vorgehen in computergestützten System modelliert werden, bietet sich ein Ansatz gestützt auf die Wahrscheinlichkeitstheorie an. Inhalt dieser Arbeit ist es, mit Bayesnetzen ein Verfahren vorzustellen, das es erlaubt unsichere Problemomänen platzeffizient aber vollständig zu repräsentieren.

Die Wahrscheinlichkeitstheorie definiert einen *Wahrscheinlichkeitsraum* $P = (\Omega, \mathcal{A}, \Pr)$ als ein Tupel bestehend aus einer Menge von eindeutigen Elementarereignissen Ω , *Ereignissen* \mathcal{A} und einer *Wahrscheinlichkeitsfunktion* $\Pr : \mathcal{A} \mapsto [0, 1]$. \mathcal{A} bildet eine σ -Algebra über Ω , erlaubt es also mehrere Elementarereignisse zu Gruppen zusammenzufassen. \Pr ordnet jedem Ereignis eine Wahrscheinlichkeit zwischen 0 und 1 zu, wobei $\Pr[\Omega] = 1$ und für zwei disjunkte Mengen A und B gilt $\Pr[A \cup B] = \Pr[A] + \Pr[B]$.

Im Folgenden werden die für Bayesnetze wesentlichen Sätze und Definitionen kurz und ohne Beweis eingeführt, diese können zum Beispiel [Gö03] entnommen werden.

Oft soll die Wahrscheinlichkeit für ein Ereignis A bestimmt werden, wenn schon bekannt ist, dass andere Ereignisse B , sogenannte *Evidenzen*, eingetreten sind. Diese Wahrscheinlichkeit nennt man *bedingte Wahrscheinlichkeit*.

Definition 1 (bedingte Wahrscheinlichkeit)

Seien A und B Ereignisse mit $\Pr[B] > 0$. Dann ist

$$\Pr[A \mid B] = \frac{\Pr[A \cap B]}{\Pr[B]}$$

die bedingte Wahrscheinlichkeit von A unter der Bedingung B .

Diese Definition erlaubt die Einführung des Begriffs der stochastischen *Unabhängigkeit*. Zwei Ereignisse sind dann unabhängig, wenn Wissen über das eine Ereignis die Wahrscheinlichkeit des anderen Ereignisses nicht beeinflusst.

Definition 2 (Unabhängigkeit)

Sind A und B mögliche Ereignisse, dann ist A von B unabhängig wenn

$$\Pr[A \mid B] = \Pr[A]$$

Unter Verwendung der bedingten Wahrscheinlichkeit lässt sich zeigen, dass diese Definition äquivalent ist zu:

Lemma 3

A ist unabhängig von B genau dann wenn

$$\begin{aligned}\Pr[A \cap B] &= \Pr[A] \cdot \Pr[B] \\ &= \Pr[A \mid B] \cdot \Pr[B]\end{aligned}$$

Mit diesen Begriffen können drei wichtige Sätze der Wahrscheinlichkeitstheorie hergeleitet werden:

Satz 4 (Produktsatz)

Für unabhängige Ereignisse A_i gilt

$$\Pr \left[\bigcap_{i=1}^n A_i \right] = \prod_{i=1}^n \Pr \left[A_i \mid \bigcap_{k=1}^{i-1} A_k \right]$$

Die Wahrscheinlichkeit, dass mehrere unabhängige Ereignisse gleichzeitig eintreten lässt sich also darstellen als eine mit Produkten verknüpfte Kette von bedingten Wahrscheinlichkeiten, die immer kürzer wird.

Man nennt eine Menge von Ereignissen eine *vollständige Ereignisdisjunktion* wenn kein Paar von Ereignissen gleichzeitig eintreten kann und gemeinsam alle Elementarereignisse abgedeckt werden. Für eine solche Disjunktion kann die Wahrscheinlichkeit eines Ereignisses ausgedrückt werden, indem alle Alternativen addiert werden.

Satz 5 (Satz über die totale Wahrscheinlichkeit)

Bilden die Ereignisse A_1, \dots, A_n eine vollständige Ereignisdisjunktion, dann gilt für die Wahrscheinlichkeit eines Ereignisses B

$$\Pr[B] = \sum_{i=1}^n \Pr[B \mid A_i] \cdot \Pr[A_i]$$

Der Satz von Bayes stellt eine Verbindung her zwischen der Bedingung eines Ereignisses A unter einem anderen Ereignis B und der Wahrscheinlichkeit, dass beide Ereignisse gleichzeitig eintreten. Der Satz illustriert die Interpretation, dass eine bedingte Wahrscheinlichkeit eine Renormalisierung von bereits bekannten Wahrscheinlichkeiten darstellt, indem der Ereignisraum eingeschränkt wird.

Satz 6 (Satz von Bayes)

Für zwei Ereignisse A und B mit $\Pr[B] > 0$ gilt

$$\begin{aligned}\Pr[A | B] &= \frac{\Pr[B | A] \cdot \Pr[A]}{\Pr[B]} \\ &= \frac{\Pr[A, B]}{\Pr[B]}\end{aligned}$$

Zur Vereinfachung der Notation verwendet man $\Pr[A, B]$ statt $\Pr[A \cap B]$, um auszudrücken, dass sowohl A als auch B zutreffen sollen. Werden mehrere Zufallsvariablen betrachtet, so definiert man die *gemeinsame Verteilung* als die Funktion, die einer beliebigen aber vollständigen Kombination von Evidenzen ihre Wahrscheinlichkeit zuordnet. Für n binäre Zufallsvariablen hat diese 2^n Einträge.

Lemma 7

Für unabhängige Zufallsvariablen A_1, \dots, A_n gilt

$$\Pr[A_1 = a_1, \dots, A_n = a_n] = \Pr[A_1 = a_1] \cdot \dots \cdot \Pr[A_n = a_n]$$

Beweis

Folgt direkt aus der Definition der Unabhängigkeit.

Für einfache Problemstellungen führt dieses Lemma direkt zu einer kompakten Darstellung der gemeinsamen Verteilung.

Beispiel 1 (Nachmittagsplanung)

Ein Student kann an einem Nachmittag Lebensmittel kaufen, Sport machen oder lernen. Er hat genügend Zeit für alle drei Tätigkeiten. Am Mittag entscheidet er für jede mögliche Tätigkeit mit Hilfe eines Würfelwurfs, ob er sie ausüben möchte.

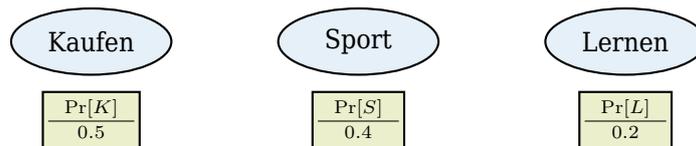


Abbildung 1: Graphische Repräsentation der Nachmittagsplanung.

Das Ergebnis jedes Würfelwurfes ist von allen vorherigen Würfeln unabhängig, es kann Lemma 7 angewendet werden. Für die Wahrscheinlichkeit, Sport zu machen, zu lernen aber nicht einkaufen zu gehen gilt also

$$\begin{aligned}\Pr[K, S, \neg L] &= \Pr[K] \cdot \Pr[S] \cdot \Pr[\neg L] \\ &= 0.5 \cdot 0.4 \cdot 0.8 = 0.16\end{aligned}$$

Zwar existieren exponentiell viele Kombinationen, diese müssen jedoch nicht alle explizit gespeichert werden, da sie sich aus linear vielen Informationen - den Einzelverteilungen - errechnen lassen.

K	S	L	Pr	K	S	L	Pr
F	F	F	0.24	T	F	F	0.24
F	F	T	0.06	T	F	T	0.06
F	T	F	0.16	T	T	F	0.16
F	T	T	0.04	T	T	T	0.04

Abbildung 2: Gemeinsame Verteilung der Nachmittagsplanung.

2 Bayesnetze

Der Ansatz, mehrdimensionale Zufallsvariablen auf Basis der Einzelverteilungen darzustellen, wird mit Hilfe von Bayesnetzen so erweitert, dass auch Systeme mit Abhängigkeiten repräsentiert werden können.

2.1 Aufbau

Bayesnetze stellen Abhängigkeiten in Form von gerichteten Graphen dar. Dabei sind Zufallsvariablen genau dann mit einer Kante verbunden, wenn sie sich direkt beeinflussen. Diese Darstellung führt zu einer vollständigen und in vielen Fällen sehr kompakten Darstellung von beliebigen Zufallsvariablen. [RN09] definiert Bayesnetze anhand von drei Regeln.

Definition 8 (Bayesnetz)

Ein Bayesnetz ist ein gerichteter azyklischer Graph, ein DAG.

1. Jeder Knoten repräsentiert eine diskrete oder kontinuierliche Zufallsvariable.
2. Existiert eine Kante vom Knoten X zum Knoten Y , dann heißt X Elternknoten von Y . Es existieren keine Kreise.
3. Jeder Knoten X_i besitzt eine bedingte Wahrscheinlichkeitsverteilung $\Pr[X_i \mid \text{Parents}(X_i)]$.

Im Folgenden soll gezeigt werden, dass diese Darstellung ausreichend ist, Einträge in der gemeinsamen Verteilung zu ermitteln und damit die Zufallsvariable vollständig zu beschreiben.

Beispiel 2 (Erweiterte Nachmittagsplanung)

Der Student möchte das Haus nicht verlassen müssen, wenn es regnet. Deshalb halbiert er bei Regen die Wahrscheinlichkeit für Aktivitäten außer Haus. Wenn er weder Sport macht noch lernt bekommt er ein schlechtes Gewissen. Nach seinem Würfelergebnis möchte er wissen, wie groß die Wahrscheinlichkeit ist, am Abend schlecht gelaunt zu sein.

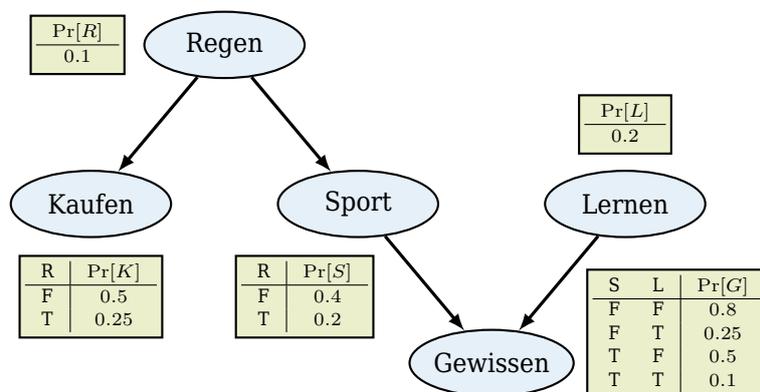


Abbildung 3: Ein Bayesnetz zur erweiterten Nachmittagsplanung. Jede Zufallsvariable wird durch einen Knoten repräsentiert, Kanten beschreiben direkte Abhängigkeiten. Es sind sowohl die Topologie des Netzes als auch die bedingten Wahrscheinlichkeitsverteilungen abgebildet.

Abbildung 3 zeigt ein Bayesnetz, das die erweiterte Nachmittagsplanung modelliert. Ob der Student am Abend ein schlechtes Gewissen hat ist allein davon abhängig, ob er Sport gemacht hat und gelernt hat. Zwar beeinflusst der Regen das schlechte Gewissen, jedoch nur indirekt, da er die Wahrscheinlichkeit für sportliche Aktivitäten verringert. Sport und Lernen haben keine Verbindung, da die Entscheidung für beide Tätigkeiten unabhängig voneinander getroffen wird.

Die bedingten Verteilungen sind in Form von Tabellen dargestellt. Dabei wird jeder möglichen Kombination von Zuständen der Elternknoten eine Wahrscheinlichkeit zugewiesen. Im Gegensatz zur kompletten gemeinsamen Verteilung werden hier nur lokale Zusammenhänge repräsentiert, die im Allgemeinen einfacher und intuitiver zu bestimmen sind. Im gegebenen Netz sind alle Variablen binär, daher wird jeweils nur die Wahrscheinlichkeit p für ein positives Ergebnis angegeben, da die Wahrscheinlichkeit für ein negatives Ergebnis $1 - p$ betragen muss. So müssen im gegebenen Netz die Topologie und 10 Werte gespeichert werden. Dies genügt, um alle $2^5 = 32$ Einträge der gemeinsamen Verteilung zu errechnen.

2.2 Semantik

Lemma 7 erlaubt die Konstruktion von Einträgen in der gemeinsamen Verteilung in sehr kompakter Weise, jedoch nur für unabhängige Zufallsgrößen. Dieses Lemma soll für Abhängigkeiten verallgemeinert werden, indem diese möglichst lokal dargestellt werden. Im Beispiel der erweiterten Nachmittagsplanung kann zwar offensichtlich nicht von Unabhängigkeit aller Variablen ausgegangen werden, es sind aber immernoch

Unabhängigkeiten vorhanden. Wissen darum, ob der Student am Nachmittag gelernt hat, ändert nichts an den Wahrscheinlichkeiten für sportliche Aktivität. Ist bekannt, ob der Student einkaufen war, lässt jedoch Rückschlüsse auf den Sport zu, da beides vom Regen beeinflusst wird. Dieser Rückschluss ist jedoch indirekt: Ist bereits bekannt ob es geregnet hat oder nicht, so beeinflussen sich beide Variablen nicht.

Diese Beobachtung führt zur Definition der bedingten Unabhängigkeit, eine Verallgemeinerung der im vorherigen Teil eingeführten Unabhängigkeit.

Definition 9 (Bedingte Unabhängigkeit)

Zwei Ereignisse A und B sind bedingt unabhängig unter einem Ereignis C wenn gilt

$$\Pr[A, B | C] = \Pr[A | C] \cdot \Pr[B | C]$$

Um eine kompakte Form zur Berechnung von Einträgen in der gemeinsamen Verteilung zu erhalten, können bedingte Unabhängigkeiten in Bayesnetzen allgemein bestimmt werden. Dieser Begriff fällt zusammen mit einer Definition aus der Graphentheorie, der sogenannten *d-Separation*.

Definition 10 (d-Separation)

Zwei Knotenmengen X und Y sind in einem DAG G durch Z d-separiert $\langle X | Z | Y \rangle_G$ wenn es keinen (ungerichteten) Pfad zwischen X und Y gibt für den gilt

1. Jeder Knoten an dem Kanten zusammenlaufen ist selbst in Z oder hat einen Nachfahren in Z
2. Jeder andere Knoten ist nicht in Z

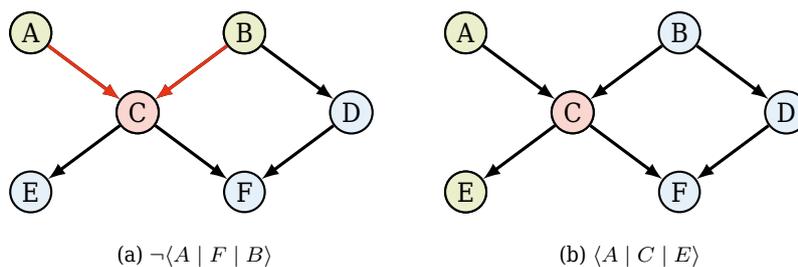


Abbildung 4: Zwei Beispiele zur d-Separation. (a) ist keine gültige d-Separation, da F Nachfahre von C ist. (b) ist gültig, da C in der separierenden Menge ist und F als einziger Knoten an dem Kanten zusammenlaufen nicht.

Intuitiv durchbricht eine d-Separation die Kette von möglichen Rückschlüssen in einem Bayesnetz durch das Hinzufügen von Evidenzen. [Gö03] zeigt, dass diese Intuition tatsächlich stichhaltig ist.

Satz 11 (Bedingte Unabhängigkeiten in Bayesnetzen)

Zufallsvariablen X und Y sind in einem Bayesnetz bedingt unabhängig unter E wenn $\langle X \mid E \mid Y \rangle$, E also X und Y d-separiert.

Speziell ist jeder Knoten bedingt unabhängig von

1. allen nicht-Nachfahren gegeben seine Eltern.
2. allen anderen Knoten gegeben seine Eltern, Kinder und Eltern der Kinder.

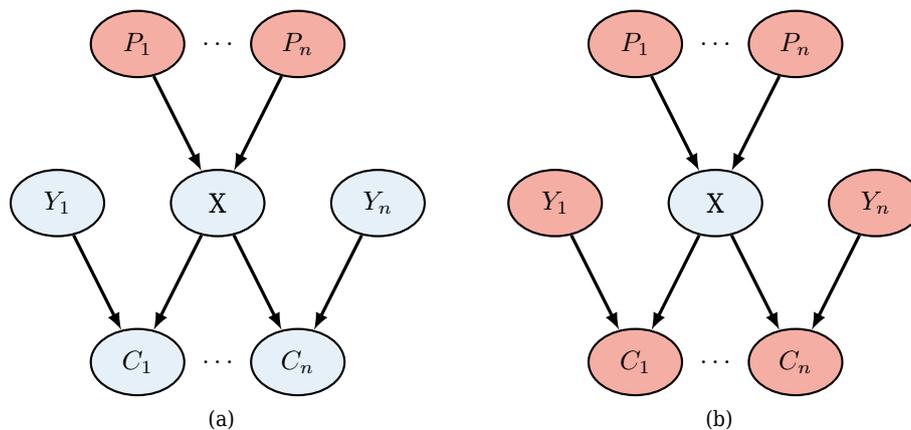


Abbildung 5: Bedingte Unabhängigkeit in Bayesnetzen. (a) Ein Knoten X ist bedingt unabhängig von allen nicht-Nachfahren gegeben seine Eltern (rot markiert). (b) Ein Knoten X ist bedingt unabhängig von allen anderen Knoten gegeben seine Markov-Einbettung.

Während die zweite Regel bei der Konstruktion von Inferenzalgorithmen zur Anwendung kommt, kann die erste Regel direkt verwendet werden, um Einträge in der gemeinsamen Verteilung zu errechnen. Anhand des Beispiels zur erweiterten Nachmittagsplanung soll errechnet werden, wie groß die Wahrscheinlichkeit ist, dass der Student bei schönem Wetter ein schlechtes Gewissen bekommt, obwohl er Sport gemacht hat und einkaufen war, jedoch nicht gelernt hat. Einsetzen in den Produktsatz ergibt

$$\begin{aligned} \Pr[G, S, \neg L, K, \neg R] &= \Pr[G \mid S, \neg L, K, \neg R] \cdot \Pr[S \mid \neg L, K, \neg R] \\ &\quad \cdot \Pr[\neg L \mid K, \neg R] \cdot \Pr[K \mid \neg R] \cdot \Pr[\neg R] \end{aligned}$$

Während für den Produktsatz prinzipiell eine beliebige Reihenfolge gewählt werden kann, erleichtert es die Rechnung, eine umgekehrte topologische Sortierung zu wählen. Dann ist keine Bedingung Kindknoten der bedingten Variable und der erste Satz zur Unabhängigkeit in Bayesnetzen kann direkt angewendet werden. Dadurch fallen alle Bedingungen weg, die keine direkten Elternknoten sind, man erhält

$$\begin{aligned}
\Pr[G, S, \neg L, K, \neg R] &= \Pr[G \mid S, \neg L] \cdot \Pr[S \mid \neg R] \\
&\quad \cdot \Pr[\neg L] \cdot \Pr[K \mid \neg R] \cdot \Pr[\neg R] \\
&= 0.5 \cdot 0.8 \cdot 0.4 \cdot 0.5 \cdot 0.9 = 0.072
\end{aligned}$$

Da die bedingte Wahrscheinlichkeitsverteilung gerade so definiert ist, Wahrscheinlichkeiten in Abhängigkeit der Elternknoten anzugeben, sind alle verbleibenden Wahrscheinlichkeiten bekannt und der Eintrag der gemeinsamen Verteilung kann errechnet werden.

Satz 12 (Gemeinsame Verteilung)

Für einen Eintrag $\Pr[x_1, \dots, x_n]$ in der gemeinsamen Verteilung eines Bayesnetzes gilt

$$\Pr[x_1, \dots, x_n] = \prod_{i=1}^n \Pr[x_i \mid \text{Parents}(X_i)]$$

Beweis

Die am Beispiel vorgeführte Konstruktion verwendet nur Satz 11 und die Existenz einer topologischen Sortierung, die für jeden DAG gegeben ist. Sie lässt sich also auf jedes Bayesnetz übertragen. Ein exakterer Beweis wird in [Gö03] geführt.

Durch die Konstruktion der gemeinsamen Verteilung wurde gezeigt, dass Bayesnetze eine mehrdimensionale Zufallsvariable komplett und exakt beschreiben, wenn alle Abhängigkeiten repräsentiert sind. Zwar können auf diese Weise prinzipiell beliebige Verteilungen modelliert werden, bei der Konstruktion von performanten Netzen ist jedoch eine Abwägung zwischen Exaktheit und Effizienz notwendig.

Die Gemeinsamen Verteilung einer Zufallsgröße mit n binären Dimensionen hat 2^n Einträge. Um zu vermeiden, dass auch im Bayesnetz exponentiell viele Einträge existieren, muss lokale Struktur geschaffen werden. Wird davon ausgegangen, dass es eine Konstante k gibt sodass alle Knoten maximal k Eltern haben, so enthalten die bedingten Wahrscheinlichkeitsverteilungen gemeinsam $n2^k$ Einträge, wachsen also nur noch linear mit der Größe des Netzwerks. Dies kann erreicht werden, indem einerseits gute Heuristiken beim Ordnen der Netzknoten angewendet werden (zum Beispiel eine Orientierung der Eltern-Kind-Beziehung an Ursache-Wirkung-Zusammenhängen) und andererseits geringe oder unwichtige Abhängigkeiten ignoriert werden.

Um kurze Laufzeiten zu erhalten, muss auch darauf geachtet werden, die Knotenverteilungen effizient zu repräsentieren. Während in dieser Arbeit nur binäre Verteilungen behandelt werden, erklärt [RN09] wie auch komplizierte kontinuierliche Verteilungen mit kleinen Netzen behandelt werden können, indem kanonische Verteilungen wie die Normalverteilung geschickt kombiniert werden.

3 Exakte Inferenz in Bayesnetzen

Beim Umgang mit mehrdimensionalen Zufallsvariablen ist nicht nur das Ermitteln von Einträgen in der gemeinsamen Verteilung interessant. Das Ermitteln von A-posteriori-Wahrscheinlichkeiten stellt eine zweite wichtige Anwendung dar, wobei bestimmte Evidenzen bekannt sind und die Wahrscheinlichkeitsverteilungen für weitere Zufallsvariablen bestimmt werden sollen.

Definition 13 (Inferenz)

Für ein Netz mit Knotenmenge $K = \{X\} \cup E \cup Y$ soll die Wahrscheinlichkeit ermittelt werden, dass eine Variable X einen Wert annimmt, unter der Bedingung, dass für die Evidenzvariablen E die Werte e beobachtet werden und die versteckten Variablen Y unbekannt sind.

$$\Pr[X = x \mid e] = \Pr[X = x \mid E_1 = e_1, \dots, E_n = e_n] = ?$$

Im weiteren Verlauf soll mit der Inferenz durch Aufzählen ein einfacher Inferenzalgorithmus vorgestellt werden. Hierzu wird das Beispiel der erweiterten Nachmittagsplanung in abgewandelter Form betrachtet.

Beispiel 3 (Bekämpfung des schlechten Gewissens)

Ein schlechtes Gewissen macht den Studenten unglücklich, er will es bekämpfen. Dazu lenkt er sich mit einem Film ab oder isst zur Aufmunterung Apfelringe. Zur Vereinfachung der Rechnung wird der Einfluss des Regens nicht mehr beachtet.

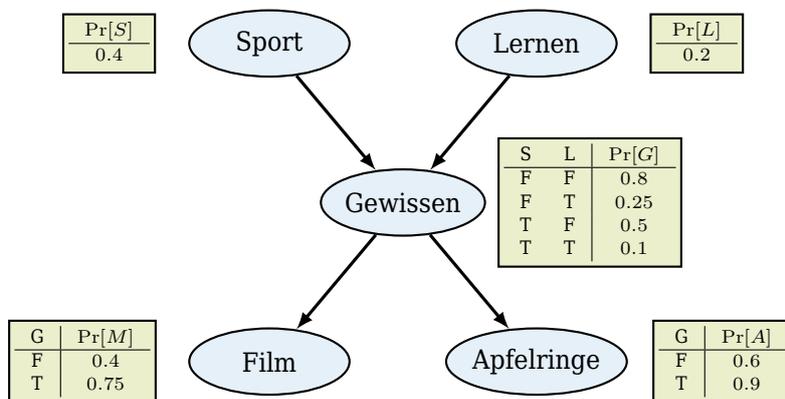


Abbildung 6: Bayesnetz zur Bekämpfung des schlechten Gewissens.

Es soll die Wahrscheinlichkeit errechnet werden, dass der Student ohne vorher Sport gemacht zu haben einen Film ansieht und dabei Apfelringe isst, also $\Pr[\neg S \mid M, A]$. Im Gegensatz zur gemeinsamen Verteilung sind hier nicht alle Variablen bestimmt, L und G sind versteckte Variablen. Außerdem existieren Bedingungen, die nicht unbedingt direkt mit Satz 11 eliminiert

werden können. Der Satz von Bayes ermöglicht es durch mehrmalige Anwendung, die Bedingung zu entfernen:

$$\begin{aligned}\Pr[\neg S \mid M, A] &= \frac{\Pr[\neg S, M \mid A]}{\Pr[M]} \\ &= \frac{\Pr[\neg S, M, A]}{\Pr[M] \cdot \Pr[A]} \\ &= \alpha \cdot \Pr[\neg S, M, A]\end{aligned}$$

Die Bestimmung von $\Pr[M]$ und $\Pr[A]$ ist zwar auch ein Inferenzproblem, kann jedoch vermieden werden, indem die Tatsache ausgenutzt wird, dass die zu errechnende Wahrscheinlichkeit und ihr Gegenereignis sich zu 1 addieren müssen. Beide Wahrscheinlichkeiten erfahren die gleiche Normalisierung durch den Satz von Bayes, weswegen diese im Faktor α zusammengefasst wird. Dieser Faktor wird eliminiert, indem zusätzlich auch das Gegenereignis inferiert und als letzter Schritt die Normalisierung manuell durchgeführt wird.

Die resultierende Wahrscheinlichkeit enthält keine Bedingungen mehr, ist jedoch im Allgemeinen noch kein Eintrag in der gemeinsamen Verteilung, da die versteckten Variablen nicht festgelegt sind. Um diese hinzuzufügen wird der Satz der totalen Wahrscheinlichkeit angewandt.

$$\begin{aligned}\Pr[\neg S \mid M, A] &= \alpha \cdot \Pr[\neg S, M, A] \\ &= \alpha \cdot \sum_g \Pr[\neg S, M, A, g] \\ &= \alpha \cdot \sum_g \sum_l \Pr[\neg S, M, A, g, l] \\ &\approx 0.585\end{aligned}$$

Diese Darstellung erlaubt Inferenz, da nur über Einträge der gemeinsamen Verteilung summiert wird, benötigt jedoch exponentielle Laufzeit. Für ein Netz mit n binären Zufallsvariablen muss im schlechtesten Fall über fast alle Variablen summiert werden, zum Beispiel indem die Wahrscheinlichkeit $\Pr[M]$ ausgerechnet wird. Jede Summe verdoppelt die Anzahl an zu berechnenden Einträgen in der gemeinsamen Verteilung und jeder dieser Einträge wird in linearer Zeit berechnet. Es ergibt sich eine Gesamtkomplexität von $\mathcal{O}(n \cdot 2^n)$.

Satz 14 (Inferenz durch Aufzählen)

Für ein Netz mit Knotenmenge $K = \{X\} \cup E \cup Y$ kann die Wahrscheinlichkeit $\Pr[X \mid e]$ berechnet werden durch

$$\Pr[X \mid e] = \alpha \cdot \sum_y \Pr[X, e, y]$$

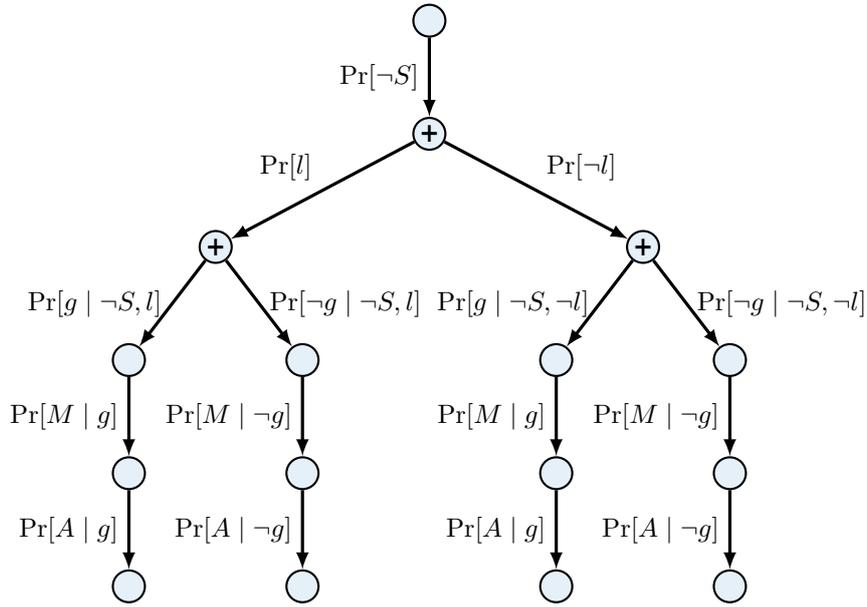


Abbildung 7: Berechnungsbaum erzeugt durch die verbesserte Inferenz durch Aufzählen. Unmarkierte Knoten stellen eine Multiplikation mit dem darunterliegenden Teilbaum dar, mit + markierte Knoten repräsentieren Summen.

Der Ansatz kann jedoch verbessert werden, indem die Unabhängigkeiten in Bayesnetzen verwendet werden, um die Gesamtwahrscheinlichkeit direkt aus den bedingten Verteilungen zu berechnen, ohne den Umweg über die gemeinsame Verteilung gehen zu müssen.

$$\begin{aligned}
 \Pr[\neg S \mid M, A] &= \alpha \cdot \sum_g \sum_l \Pr[\neg S, M, A, g, l] \\
 &= \alpha \cdot \sum_g \sum_l \Pr[\neg S] \cdot \Pr[l] \cdot \Pr[g \mid \neg S, l] \cdot \Pr[M \mid g] \cdot \Pr[A \mid g] \\
 &= \alpha \cdot \Pr[\neg S] \cdot \sum_l \Pr[l] \cdot \sum_g \Pr[g \mid \neg S, l] \cdot \Pr[M \mid g] \cdot \Pr[A \mid g]
 \end{aligned}$$

Eine Intuition zur Komplexität dieses verbesserten Algorithmus liefert Abbildung 7, die veranschaulicht, dass im schlechtesten Fall immer noch exponentiell viele Berechnungen durchgeführt werden müssen. Wird in einem Netz mit n binären Variablen wiederum die Wahrscheinlichkeit einer einzigen Variable errechnet, so existieren im resultierenden Baum $n-1$ Schichten aus Plus-Knoten. Diese haben immer zwei Kinder, es liegt also ein Baum vor, in dem $n-1$ vollbesetzte Binärbaumschichten existieren. Jeder Knoten in diesem Baum stellt eine Rechenoperation dar, es müssen also $\mathcal{O}(2^n)$ Rechenoperationen durchgeführt werden.

Im Allgemeinen lässt sich diese Komplexität wahrscheinlich auch nicht verbessern, da das SAT-Problem auf exakte Inferenz in Bayesnetzen reduzierbar ist. Diese wird erreicht, indem in den Knoten aussagenlogische Operatoren (\wedge , \vee , \neg) repräsentiert werden. Exakte Inferenz auf Bayesnetzen ist NP-schwer ([RN09]).

Es gibt jedoch mehrere Ansätze, Inferenz auf Bayesnetzen effizienter durchzuführen. Abbildung 7 zeigt, dass bei der Inferenz durch Aufzählen die selben Teilbäume mehrmals berechnet werden, im Beispiel die Pfade für M und A . Diese Redundanz kann mit einem Ansatz der dynamischen Programmierung eliminiert werden, in dem die einzelnen bedingten Wahrscheinlichkeitsverteilungen bei Bedarf zusammengefasst werden. Dieser Algorithmus mit Elimination von Variablen ([RN09]) hat zwar im schlechtesten Fall auch exponentielle Laufzeit, läuft jedoch linear auf Polytrees, also Netzen, die keine ungerichteten Kreise enthalten.

Eine Alternative stellen approximative Algorithmen dar, bei denen zwar kein exaktes Ergebnis erreicht werden kann, die jedoch auch auf großen oder topologisch ungünstigen Netzen einsetzbar sind. Der direkte Ansatz, bei dem sukzessive anhand der bedingten Verteilungen zufällig Belegungen erzeugt und mitgezählt werden lässt sich dabei verbessern, indem Schätzvariablen oder Markovketten eingesetzt werden, um schneller konvergierende Verfahren zu erhalten.

Literatur

- [Gö03] Günther Görz, editor. *Handbuch der künstlichen Intelligenz*. Oldenbourg, München [u.a.], 4th edition, 2003.
- [RN09] S. J. Russell and P. Norvig. *Artificial Intelligence: A Modern Approach*. Prentice Hall, 3rd edition, 2009.