FernUniversität
Gesamthochschule in Hagen

Fachbereich Wirtschaftswissenschaft

Probabilistische Wissensrepräsentation mittels Bayes-Netzen

Seminar:	Virtuelles Seminar des Lehrstuhls für Betriebswirtschaftslehre, insbesondere Operations Research von Professor Dr. Wilhelm Rödder
Studienjahr:	2001/2002
Matr.-Nr.:	4xxxxxx
Vorname Name:	Dr.-Ing. Ingo D. Rullhusen
Anschrift:	xxxxxxxxxxxxxxxxxxx, 28xxx Bremen
Abgabedatum:	2. Dezember 2001

Inhaltsverzeichnis

Abbildungsverzeichnis

1 Einführung

Die probabilistische Wissensrepräsentation, wie sie in modernen Expertensystemen zum Einsatz gelangt, bildet das Wissen über eine eng spezifizierte Domäne in einer Wahrscheinlichkeitsverteilung ab. Hierbei sollen die Zufallsvariablen mit ihren Ausprägungen alle Zustände des betrachteten Systems darstellen können. Bei komplexen Wissensdomänen gelangt man somit zu hochdimensionalen Verteilungsfunktionen, deren Erfassung und Handhabung Schwierigkeiten bereiten. Derzeit gibt es vor allem zwei Lösungsansätze für diese Problematik: zum einen das Entropieprinzip, welches in der Expertensystem-Shell SPIRIT[1] eingesetzt wird, und zum anderen das Bayes-Netz, welches beispielsweise in HUGIN[2] zum Einsatz gelangt.

Zur effektiven Erfassung des Wissens ist es wichtig, mit möglichst wenig Angaben, die hinreichend zuverlässig bestimmbar sind, das gesamte Systemverhalten der Wissensdomäne zu charakterisieren. Beim Entropieprinzip geht man zunächst von einer Gleichverteilung aus, da diese keinerlei Informationen enthält. Diese Startverteilung unterwirft man anschließend den Regeln, die in Form bedingter Wahrscheinlichkeiten gegeben sind. Diese Wahrscheinlichkeiten erhält man aus den Expertenaussagen, indem man eine Aussage wie „wenn das Ereignis A vorliegt, ergibt sich daraus das Ereignis B mit einer Plausibilität von p“ in die mathematische Form $P(B|A) = p$ überführt. Nach dem Entropieprinzip wird nun eine neue Wahrscheinlichkeitsverteilung gesucht, die diese bedingten Wahrscheinlichkeiten erfüllt und eine minimale relative Entropie zur alten Verteilung besitzt. Auf diese Weise wird das Erzeugen von fälschlicher Artefaktinformation minimiert, so daß die neue Verteilung nach Möglichkeit zusätzlich zum bisherigen Wissen nur die in dieser Art neu gelernten Regeln enthält.

Bei der Anwendung von Bayes-Netzen muß man das Expertenwissen vorweg formal nach den Abhängigkeiten der einzelnen Systemgrößen voneinander strukturieren. Der aus diesen Abhängigkeiten bzw. Unabhängigkeiten resultierende azyklische gerichtete Un-

[1] Siehe http://www.xspirit.de.
[2] Siehe http://www.hugin.com.

abhängigkeitsgraph zeigt die bedingten Wahrscheinlichkeiten auf, die zur Definition der Gesamtverteilung gemäß der Kettenregel der Wahrscheinlichkeit benötigt werden. Diese bedingten Wahrscheinlichkeiten sind dann von einem Experten zu erfragen. Läßt sich dieser gerichtete Unabhängigkeitsgraph eineindeutig bezüglich der Indizierung der Knoten den bedingten Unabhängigkeiten der Verteilung zuordnen, so bezeichnet man diesen als *Bayes-Netz*.

Die direkte Darstellung der Gesamtverteilung kann bei vielen Zufallsvariablen recht speicheraufwendig werden. Deshalb greift man auf eine Faktorisierung der Form $P(x_1;\ldots;x_n) = \prod_i \psi_i(\ldots;x_k;\ldots)$ zurück, wobei die Dimensionen der Faktorfunktionen ψ_i möglichst klein sein sollen. Beim Entropieansatz ist eine solche Faktorisierung auf der Grundlage der Regeln erst zu bestimmen, wohingegen der Bayes-Ansatz mit den bedingten Wahrscheinlichkeiten über die Kettenregel bereits explizit eine solche Faktorisierung darstellt. Die Effizienz der Faktorisierung ergibt sich dabei aus einer geschickten Wahl der Wurzel des Bayes-Netzes.

In dieser Arbeit soll im folgenden der Bayes-Ansatz in Form des Bayes-Netzes näher betrachtet werden. Nach einer Erläuterung der wahrscheinlichkeitstheoretischen Grundlagen ohne Beachtung der mathematischen Strenge wird das Bayes-Netz definiert. Anhand eines Beispiels wird dann dessen Aufstellung und Nutzung zur Entscheidungsfindung demonstriert.

2 Grundlagen der bedingten Wahrscheinlichkeit

Die Wahrscheinlichkeit ist eigentlich streng axiomatisch über die Kolmogorowschen Axiome definiert.[3] Hierbei stellt die Wahrscheinlichkeit ein Häufigkeitsmaß über eine unendliche Anzahl von Realisierungen der stochastischen Grundgesamtheit dar. Wenn im folgenden von der Wahrscheinlichkeit die Rede ist, handelt es sich häufig streng genommen um ein Plausibilitätsmaß für eine einzelne Realisierung, wie es durch Schätzung aus der Erfahrung gewonnen

[3] Siehe bspw. Bronstein/Semendjajew (1989).

wird. Da aber hierauf die gleichen Rechenregeln der Wahrscheinlichkeit angewandt werden, wird diese strenge Unterscheidung desweiteren nicht weiter herausgestellt.

Mit $P(A)$ sei die Wahrscheinlichkeit des Ereignisses A bezeichnet. Bedeutet das Ereignis A, daß die diskrete Zufallsvariable X_i die Ausprägung x_i annimmt, so schreibt man auch einfach $P(X_i = x_i)$ oder kürzer $P(x_i)$. Die Verbundwahrscheinlichkeit der Ereignisse A und B sei mit $P(A \cap B)$ bezeichnet und steht für die Wahrscheinlichkeit des gleichzeitigen Zutreffens der beiden Ereignisse. Ist nun $A = \{X_i = x_i\}$ und $B = \{X_j = x_j\}$ so verwendet man auch die Schreibweise $P(X_i = x_i, X_j = x_j)$ oder kürzer $P(x_i; x_j)$.

2.1 Grundregeln

Die *bedingte Wahrscheinlichkeit* $P(A|B)$ („P von A gegeben B“) wird über

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \qquad \text{mit} \qquad P(B) \neq 0 \tag{1}$$

definiert. Sie stellt also den Quotienten aus der Verbundwahrscheinlichkeit und der Wahrscheinlichkeit der Bedingung dar. Löst man diese Definition auf, so erhält man mit

$$P(A \cap B) = P(A|B)P(B) \tag{2}$$

das *Produktgesetz der Wahrscheinlichkeit.* Die rekursive Anwendung des Produktgesetzes auf die Verbundwahrscheinlichkeit $P(A_1 \cap \cdots \cap A_n)$ ergibt mit

$$\begin{aligned} P(A_1 \cap \cdots \cap A_n) = \\ P(A_n|A_{n-1} \cap \cdots \cap A_1) \cdot P(A_{n-1}|A_{n-2} \cap \cdots \cap A_1) \cdot \ \cdots \\ \cdot P(A_2|A_1) \cdot P(A_1) \end{aligned} \tag{3}$$

die *Kettenregel der Wahrscheinlichkeit.* Diese ist zugleich die formale Grundlage der Bayes-Netze. In dieser Form wäre das zugeordnete Bayes-Netz ein vollständiger gerichteter azyklischer Graph mit der Wurzel A_1, was später näher ausgeführt wird.

Für die *disjunkten* Ereignisse A und B gilt gemäß den Axiomen der Wahrscheinlichkeit

$$P(A \cup B) = P(A) + P(B).$$

Für *beliebige* Ereignisse wird hieraus das *verallgemeinerte Summengesetz*

$$P(A \cup B) = P(A) + P(B) - P(A \cap B), \tag{4}$$

da die beiden Ereignissen gemeinsame Schnittmenge bereits in beiden Einzelwahrscheinlichkeiten enthalten ist und somit noch einmal subtrahiert werden muß. Über das Summengesetz können Aussagen mit inklusivem Oder mithin auf einfache Verbundwahrscheinlichkeiten zurückgeführt werden. Auch komplexere Anfragen lassen sich also auf eine Berechnung von Verbundwahrscheinlichkeiten zurückführen. Dies ist deshalb von besonderer Bedeutung, weil ein Bayes-Netz zunächst einmal nur die globale Verbundwahrscheinlichkeit repräsentiert und die Berechnung der einzelnen Randverteilungen erleichtert. Aus diesen kann dann der Wahrscheinlichkeitswert einer konkreten Aussage über die Gleichungen (1) oder (4) berechnet werden.

Beachtet man, daß gemäß dem Produktgesetz (2)

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

gilt, so erhält man durch einfaches Umstellen das *Bayes-Theorem*

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}. \tag{5}$$

Dieses ist bei der Ermittlung der bedingten Wahrscheinlichkeiten von besonderer Bedeutung: Falls die im Bayes-Netz geforderte bedingte Wahrscheinlichkeit $P(A|B)$ nicht verfügbar ist, so kann man diese mit dem Bayes-Theorem bspw. aus der bedingten Wahrscheinlichkeit $P(B|A)$ ermitteln, die mit statistischem Material untermauert sein könnte. Es bietet sich auch die Möglichkeit, einfache Schätzungen der gewöhnlichen Wahrscheinlichkeiten $P(A)$ und $P(B)$ einfließen zu lassen.

2.2 Bedingte Unabhängigkeit

Von zentraler Bedeutung für die Effizienz von Bayes-Netzen ist der Begriff der *bedingten Unabhängigkeit.* Man nennt zwei Ereignisse A und B bedingt unabhängig bezüglich C, wenn gilt:

$$P(A|B \cap C) = P(A|C) \qquad \text{mit} \qquad P(B \cap C) \neq 0. \tag{6}$$

Hierfür schreibt man auch kurz $A \perp B|C$. Ferner gilt dann auch $B \perp A|C$, sofern $P(A \cap C) \neq 0$ ist. Sofern C das sichere Ereignis ist, folgt hieraus die gewöhnliche statistische Unabhängigkeit als Spezialfall. Die besondere Bedeutung dieser Eigenschaft für die Bayes-Netze ergibt sich aus der Kettenregel (3): Die Anzahl der Bedingungen in den einzelnen bedingten Wahrscheinlichkeiten kann mit Hilfe der Eigenschaft (6) reduziert werden. Im Falle von Zufallsvariablen resultiert daraus eine Verringerung der Dimension.

Diese Definition läßt sich unmittelbar von allgemeinen Ereignissen auf Zufallsvariable übertragen, indem man als Ereignis alle Ausprägungen einer Familie von Zufallsvariablen ansetzt. Seien in diesem Sinne X_A, X_B und X_C paarweise disjunkte Familien von Zufallsvariablen der Grundgesamtheit, so nennt man die Variablen in X_A und X_B bezüglich denen in X_C bedingt unabhängig, sofern

$$P(x_A|x_B; x_C) = P(x_A|x_C) \text{ mit } P(x_B; x_C) \neq 0 \quad \forall x_A, x_B, x_C \tag{7}$$

gilt, wobei bspw. mit x_A ein konkretes Tupel von Ausprägungen aller in der Familie X_A enthaltenen Zufallsvariablen bezeichnet wird. Die bedingte Unabhängigkeit muß mithin für alle Ausprägungen gegeben sein. Entsprechend schreibt man auch hier $X_A \perp X_B|X_C$.

Das bemerkenswerte an der bedingten Unabhängigkeit ist nun, daß die Unabhängigkeit quasi durch die Instanziierung der Bedingung bewirkt wird. Sofern die Bedingung x_C nicht gegeben ist, können die Zufallsvariablen x_A und x_B durchaus abhängig sein, wobei deren Abhängigkeit gerade durch die x_C vermittelt wird.

Zur Illustration sei ein einfaches Erntemodell erwähnt: Dieses Modell besitze die drei Zufallsvariablen

$$
\begin{aligned}
\text{Wetter} &= \{\text{gut}, \text{mittel}, \text{schlecht}\},\\
\text{Weizenertrag} &= \{\text{gut}, \text{mittel}, \text{schlecht}\},\\
\text{Gerstenertrag} &= \{\text{gut}, \text{mittel}, \text{schlecht}\}.
\end{aligned}
$$

Nun läßt sich sagen, daß gutes Wetter sowohl eine gute Weizen- als auch eine gute Gerstenernte bewirkt, vorbehaltlich stochastischer Schwankungen. Eine gute Weizenernte läßt über das Bayes-Theorem auf gutes Wetter und somit auch auf eine gute Gerstenernte schließen, sofern das Wetter unbekannt ist. Die Ernteerträge sind also statistisch abhängig. Bei gegebenen Wetterverhältnissen liefert der Ausgang der Weizenernte jedoch in diesem Modell keine zusätzliche Information über den Ausgang der Gerstenernte. Die Ernteerträge sind bei gegebenen Wetterverhältnissen also unabhängig voneinander. Insgesamt gilt somit

$$\text{Weizenertrag} \perp \text{Gerstenertrag} | \text{Wetter}.$$

3 Das Bayes-Netz

Die zuvor erläuterte bedingte Unabhängigkeit stellt das wesentliche Strukturmerkmal der Bayes-Netze dar. Diese stellen eine besondere Form von gerichteten azyklischen Unabhängigkeitsgraphen dar. Einen solchen Graphen erhält man, indem man jeder Zufallsvariablen genau einen Knoten aus einer vollständig geordneten Knotenmenge zuordnet und alle möglichen gerichteten Kanten bzw. Pfeile zwischen den zugeordneten Knoten in Ordnungsrichtung einträgt. Aus diesem vollständigen azyklischen gerichteten Graphen dürfen dann solche Pfeile entfernt werden, denen eine bedingte Unabhängigkeit zwischen den beiden Pfeilknoten bezüglich aller Vorgängerknoten dieser entspricht. Für das Erntemodell ist dies in der Abbildung 1 verdeutlicht.

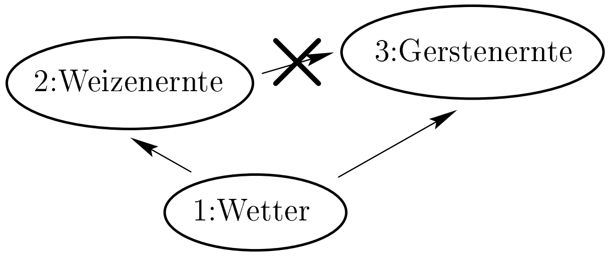

Abbildung 1: Unabhängigkeitsgraph des einfachen Erntemodells; durch Streichen eines Pfeils gemäß der bedingten Unabhängigkeit zwischen Weizenernte und Gerstenernte bezüglich Wetter entsteht ein Bayes-Netz.

3.1 Definition

Ein Bayes-Netz ergibt sich aus einem gerichteten Unabhängigkeitsgraphen dadurch, daß aus diesem alle maximal möglichen Pfeile gestrichen werden. Hieraus resultiert dann, daß die Zuordnung zwischen der Verteilung und dem Bayes-Netz bei gegebener Knotenordnung eindeutig ist.

Dies läßt sich in der folgenden Definition formalisieren:[4] Es sei $G = (V, E)$ ein gerichteter azyklischer Graph, dessen Knoten V der Pfeilrichtungen von E folgend vollständig geordnet sind und es sei P eine gemeinsame Verteilung der Familie $(X_v)_{v \in V}$. Dann heißt G ein *gerichteter Unabhängigkeitsgraph* für P, wenn gilt:

$$(v, w) \notin E \Rightarrow X_v \perp X_w | X_{\mathrm{an}(\{v,w\})}.$$

G heißt ein *Bayes-Netz* für P, wenn sogar die Äquivalenz gilt:

$$(v, w) \notin E \Leftrightarrow X_v \perp X_w | X_{\mathrm{an}(\{v,w\})}.$$

Hierbei wird mit $\mathrm{an}(\{v, w\})$ die Menge aller Vorgänger (engl. antecedents) von v und w bezeichnet.

Aus dieser Definition folgt nun, daß jeder Pfeil im Bayes-Netz eine vorhandene bedingte Abhängigkeit bezüglich aller Vorgänger darstellt. Somit erhält man aus dem Bayes-Netz Angaben zu den bedingten Wahrscheinlichkeiten, die zur Anwendung der Kettenregel der Wahrscheinlichkeit (3) notwendig sind, um die gesamte Verbundwahrscheinlichkeit darzustellen. Für jede Knotenvariable

[4] Siehe Meyer (1998).

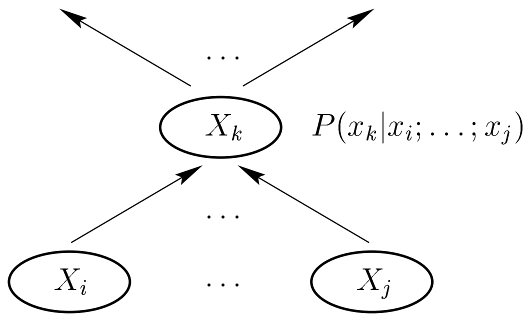

Abbildung 2: Einem Knoten des Bayes-Netzes beigeordnete bedingte Wahrscheinlichkeit.

ist eine bedingte Wahrscheinlichkeit notwendig. Alle direkten Vorgänger einer Knotenvariablen treten als Bedingung in der bedingten Wahrscheinlichkeit der Knotenvariablen auf, was in der Abbildung 2 veranschaulicht wird. Für die Wurzelknoten müssen mithin unbedingte marginale Wahrscheinlichkeiten gegeben sein. Je geringer die Anzahl der direkten Vorgänger ist, desto geringer ist die Anzahl der Bedingungen und somit auch die Dimensionalität.

Da die Zuordnung eines Bayes-Netzes zu einer Verteilung nur bezüglich einer gegebenen Knotenindizierung eindeutig ist, kommt dieser Indizierung eine besondere Bedeutung zu. Im Sinne einer einfachen Modellierung ist es das Ziel, ein Bayes-Netz mit möglichst wenig Pfeilen zu realisieren, was einer geringen Dimensionalität der bedingten Wahrscheinlichkeiten entspricht. Dieses Indizierungsproblem kann sich auf zweierlei Weisen stellen: zum einen kann eine gemeinsame Verteilung vorgegeben sein und man sucht eine möglichst optimale Indizierung zur Erzeugung eines einfachen Bayes-Netzes zwecks Wissensrepräsentation, zum anderen kann man ein Bayes-Netz zur Wissensaquisition verwenden, indem man zunächst ein möglichst einfaches Bayes-Netz aus der Erfahrung heraus aufstellt und hieraus die Indizierung und die benötigten Wahrscheinlichkeiten abliest.

Im Rahmen dieser Arbeit wird vor allem der synthetische Ansatz näher betrachtet. Die Aufgabe besteht dabei darin, das Wissen über ein Teilgebiet der realen Erfahrungswelt in einer Verbundwahrscheinlichkeit darzustellen. Hierzu sind zunächst alle relevanten Einflußgrößen mit sämtlichen Ausprägungen zu identifizieren. Diese bilden im probabilistischen Modell die Zufallsvariablen und die Menge ihrer Realisationen. Somit ist die Knotenmenge des Bayes-Netzes in ungeordneter Form gegeben. Auf dieser Grundlage sind die Abhängigkeiten zwischen den Einflußgrößen der Erfahrungswelt herauszuarbeiten und als Pfeile zwischen den Variablenknoten einzuzeichnen. Gemäß der Definition der Bayes-Netze ist hierbei auf Zyklenfreiheit zu achten. Um diese Bedingung zu erfüllen und den Graphen einfach zu halten, sollten vor allem wichtige Abhängigkeiten eingetragen werden, die sich auch gut quantifizieren lassen. Sinngemäß geht man hierbei von der Ursache in Richtung Wirkung vor, so daß die innerhalb des Modellrahmens unabhängigen Zufallsvariablen als Wurzelknoten des azyklischen gerichteten Graphen verbleiben. Hat man alle relevanten Abhängigkeiten erfaßt und azyklisch eingezeichnet, so kann man die Knotenmenge von den Wurzelknoten ausgehend durchnumerieren um eine Indizierung der Knotenmenge zu erhalten. Die Knotenmenge ist somit vollständig geordnet und der Graph bildet einen gerichteten azyklischen Graphen, der das Bayes-Netz der gesuchten Verteilung darstellt. Aus diesem Bayes-Netz kann man nun die zur Darstellung der Verbundwahrscheinlichkeit notwendigen bedingten Wahrscheinlichkeiten ablesen. Abschließend sind diese Wahrscheinlichkeiten aus der Erfahrung der Experten heraus zu bestimmen und zu tabellieren. Die ursprünglich gesuchte Verbundwahrscheinlichkeit zur Darstellung des erfaßten Wissens ergibt sich dann über die Kettenregel der Wahrscheinlichkeit (3) aus den bedingten Wahrscheinlichkeiten. Das zur Synthese der Verteilung konstruierte Bayes-Netz stellt gleichzeitig eine Repräsentation der Unabhängigkeiten in dieser Verteilung dar.

Anfragen an das Bayes-Netz als Wissensbasis werden als gesuchte Wahrscheinlichkeiten gestellt. Da bedingte Wahrscheinlichkeiten gemäß deren Definition (1) auch auf Randverteilungen zurückgeführt werden können, kommt es bei Anfragen in der Regel nur auf eine effiziente Berechnung von Randverteilungen an.

3.2 Anwendung

An einem Beispiel soll nun die Handhabung von Bayes-Netzen demonstriert werden. Zunächst wird die Aufstellung und anschließend die Nutzung aufgezeigt. Während die eigentliche Anwendung in der Praxis computergestützt mit Expertensystemen durchgeführt wird, ist die Aufstellung und Erfassung einer Wissensbasis eher manueller Natur. Dieses Vorgehen soll anhand des Lungenkrankheitsmodells von LAURITZEN/SPIEGELHALTER (1988) aufgezeigt werden.

3.2.1 Aufstellung des Bayes-Netzes

Die darzustellende Wissensbasis stellt sich wie folgt dar: Kurzatmigkeit (Dyspnoe) kann durch Tuberkulose, Lungenkrebs oder Bronchitis oder nichts von alledem oder etwas anderes verursacht sein. Ein Besuch in Asien erhöht das Risiko an Tuberkulose zu erkranken, während Rauchen ein Risikofaktor für Lungenkrebs und Bronchitis darstellt. Ein Röntgenbefund unterscheidet ebensowenig wie eine Dyspnoe zwischen Lungenkrebs oder Tuberkulose.

Die einfache Auflistung aller angesprochenen Objekte führt bereits auf eine Liste der in diesem Modell enthaltenen binären Zufallsvariablen:

$$\begin{aligned}
\text{(Asienbesucher) } \alpha &\in \{\text{wahr}, \text{falsch}\}, \\
\text{(Raucher) } \sigma &\in \{\text{wahr}, \text{falsch}\}, \\
\text{(Tuberkulose) } \tau &\in \{\text{wahr}, \text{falsch}\}, \\
\text{(Lungenkrebs) } \lambda &\in \{\text{wahr}, \text{falsch}\}, \\
\text{(Bronchitis) } \beta &\in \{\text{wahr}, \text{falsch}\}, \\
\text{(Röntgenbefund) } \xi &\in \{\text{wahr}, \text{falsch}\}, \\
\text{(Dyspnoe) } \delta &\in \{\text{wahr}, \text{falsch}\}.
\end{aligned}$$

Nun ist die Knotenmenge in Wirkungsrichtung zu sortieren und mit den Abhängigkeitspfeilen in azyklischer Weise zu versehen. Ursächlich für den Lungenzustand sind hier die externen Einflußgrößen Asienbesuch und Rauchen. Dementsprechend bilden α und σ die Wurzelknoten. Hiervon hängen im nächsten Schritt die Lungenbeschwerden τ, λ und β ab, die schließlich die Befunde ξ oder δ bewirken. Die azyklische Eintragung der anfangs beschriebenen Zusammenhänge führt zunächst auf das Bayes-Netz in der Abbildung 3, sofern alle vermuteten bedingten Abhängigkeiten bezüglich aller Vorgänger erfaßt worden sind. Aus diesem Bayes-Netz ist beispielsweise ersichtlich, daß die bedingten Wahrscheinlichkeiten $P(\xi|\tau;\lambda)$

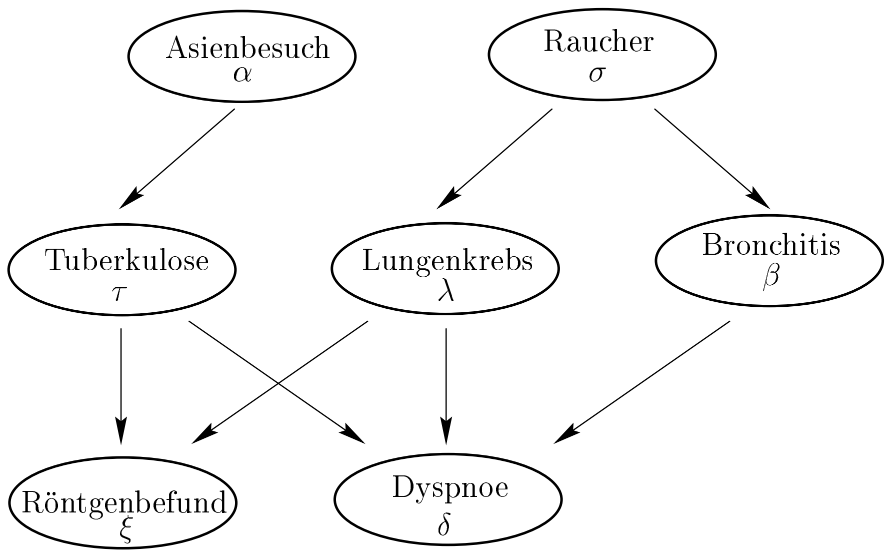

Abbildung 3: Erster Ansatz für ein Bayes-Netz der Lungendiagnose gemäß Lauritzen/Spiegelhalter (1988).

und $P(\delta|\tau;\lambda;\beta)$ benötigt werden. Hierbei ist es jedoch möglich, die Dimensionalität dieser durch die Einführung des Hilfsknotens

$$\text{Tuberkulose oder Lungenkrebs } \varepsilon \in \{\text{wahr}, \text{falsch}\}$$

weiter zu reduzieren. Ein solches Vorgehen kann in größeren Netzen deutliche Ersparnisse beim Speicher- und Rechenzeitbedarf erbringen, der wesentlich von der Dimensionalität der bedingten Wahrscheinlichkeiten abhängt. Das endgültige Bayes-Netz für die gesuchte probabilistische Wissensbasis nimmt somit die in der Abbildung 4 dargestellte Gestalt an.

Die vollständige Ordnung der Knotenmenge ergibt sich jetzt direkt aus einer Durchnumerierung in Pfeilrichtung. Wegen zweier Wurzelknoten ist diese jedoch nicht eindeutig. Die zur Darstellung der Gesamtverteilung mittels der Kettenregel der Wahrscheinlichkeit (3) benötigten bedingten Wahrscheinlichkeiten kann man nun aus dem Bayes-Netz ablesen: jedem Knoten wird eine bedingte Wahrscheinlichkeit für seine Zufallsvariable beigeordnet, wobei die einmündenden Pfeile von den Knoten der Zufallsvariablen der Bedingung ausgehen. Die Knotennumerierung sowie die beigeordneten Wahrscheinlichkeiten sind in der Abbildung 4 ebenfalls vermerkt.

Im letzten Aufstellungsschritt sind nun die Werte der einzelnen Wahrscheinlichkeiten, die durch das Bayes-Netz aufgezeigt werden,

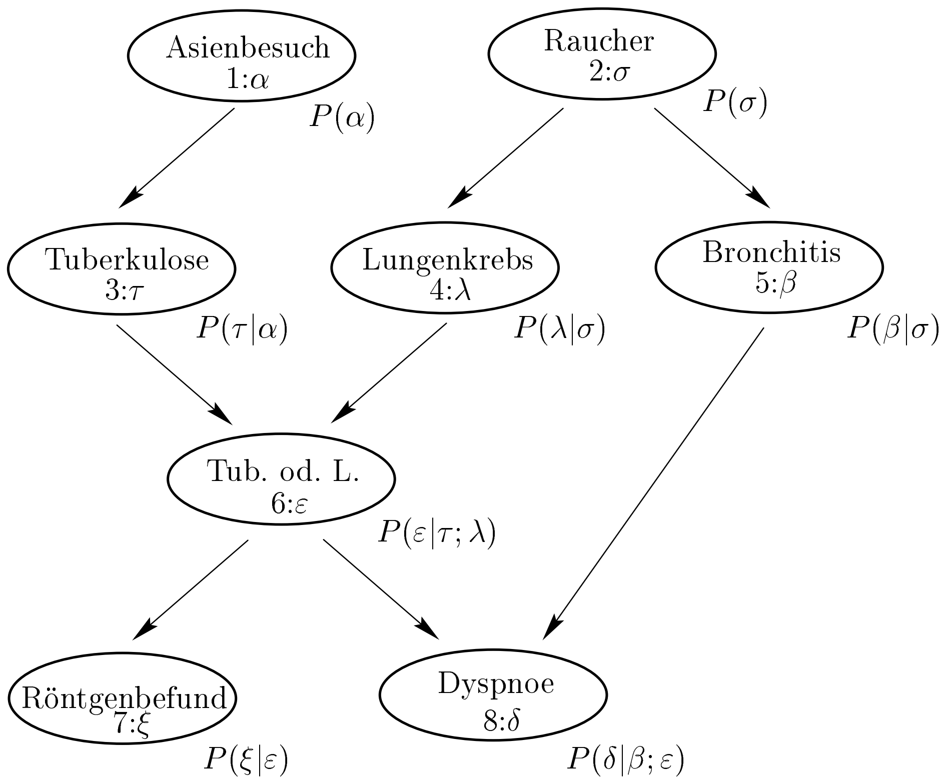

Abbildung 4: Bayes-Netz der Lungendiagnose gemäß LAURITZEN/SPIEGELHALTER (1988); die Knoten sind indiziert und die zu definierenden bedingten Wahrscheinlichkeiten sind angegeben.

Tabelle 1: Wahrscheinlichkeiten zum Bayes-Netz in Abbildung 4; α steht bspw. für $\alpha =$ wahr, die negierten Knotenwerte sind komplementär zu ergänzen.

Knoten	$P(\ldots)$	Wert
1:α	α	0,01
2:σ	σ	0,50
3:τ	$\tau\vert\alpha$	0,05
	$\tau\vert\neg\alpha$	0,01
4:λ	$\lambda\vert\sigma$	0,10
	$\lambda\vert\neg\sigma$	0,01
5:β	$\beta\vert\sigma$	0,60
	$\beta\vert\neg\sigma$	0,30
6:ε	$\varepsilon\vert\lambda;\tau$	1,00
	$\varepsilon\vert\lambda;\neg\tau$	1,00
	$\varepsilon\vert\neg\lambda;\tau$	1,00
	$\varepsilon\vert\neg\lambda;\neg\tau$	0,00
7:ξ	$\xi\vert\varepsilon$	0,98
	$\xi\vert\neg\varepsilon$	0,05
8:δ	$\delta\vert\beta;\varepsilon$	0,90
	$\delta\vert\beta;\neg\varepsilon$	0,80
	$\delta\vert\neg\beta;\varepsilon$	0,70
	$\delta\vert\neg\beta;\neg\varepsilon$	0,10

zu bestimmen. Das konsistente Vorgehen wäre hierbei der Rückgriff auf umfangreiches statistisches Zahlenmaterial zur Ermittlung dieser Größen. In der Praxis wird man jedoch häufig auf Schätzungen dieser Größen angewiesen sein. Die Werte für das hier dargestellte Beispiel sind ebenfalls Lauritzen/Spiegelhalter (1988) entnommen und in der Tabelle 1 aufgeführt worden.

Somit liegt nun eine probabilistische Wissensbasis vor, die durch das Bayes-Netz in der Abbildung 4 repräsentiert wird. Die Verbundwahrscheinlichkeit dieser Wissensbasis ergibt sich aus der Kettenregel der Wahrscheinlichkeit (3) zu

$$P(\alpha;\sigma;\tau;\lambda;\beta;\varepsilon;\xi;\delta) = P(\alpha)P(\sigma)P(\tau|\alpha)P(\lambda|\sigma)P(\beta|\sigma) \cdot P(\varepsilon|\tau;\lambda)P(\xi|\varepsilon)P(\delta|\beta;\varepsilon). \quad (8)$$

Es sind also lediglich die den Knoten zugeordneten bedingten Wahrscheinlichkeiten miteinander zu multiplizieren.

3.2.2 Nutzung des Bayes-Netzes

Nachdem man die Wissensbasis mit Hilfe eines Bayes-Netzes dargestellt hat, kann man zu deren Anwendung übergehen. Wie zuvor ausgeführt, laßt sich eine Anfrage an die Wissensbasis immer auf die Berechnung von Verbundwahrscheinlichkeiten zurückführen. Deshalb sei nur ein derartiges Beispiel betrachtet. Die exemplarisch zu klärende Frage lautet: Wie wahrscheinlich ist es, daß ein Patient zugleich an Tuberkulose und Bronchitis leidet? Mathematisch betrachtet wird die Verbundwahrscheinlichkeit $P(\tau;\beta)$ gesucht.

Diese berechnet man, indem man die Verteilung (8) über alle nicht instanziierten Zufallsvariablen summiert. Nach verschiedenen Vereinfachungen erhält man

$$P(\tau;\beta) = \left(\sum_{\alpha} P(\alpha)P(\tau|\alpha)\right)\left(\sum_{\sigma} P(\sigma)P(\beta|\sigma)\right), \tag{9}$$

woraus

$$P(\tau;\beta) = 0,00468$$

folgt. Somit beträgt die Wahrscheinlichkeit, daß ein Patient an Tuberkulose und Bronchitis leidet 0,00468.

Bei der Berechnung einer Randverteilung aus einem Bayes-Netz können zwei Vereinfachungen eingesetzt werden, die soeben demonstriert worden sind: Zum einen können alle Knoten, die keine instanziierte Zufallsvariable als Nachfolger besitzen, aus der weiteren Betrachtung ausgeschlossen werden, weil die Summationen über diese immer eins ergeben. Ergeben sich durch diese Auslassung getrennte Netze, so können die Wahrscheinlichkeiten der Instanziierungen in den Teilnetzen getrennt berechnet werden, da diese bedingt unabhängig sind. Die Verbundwahrscheinlichkeit der Randverteilung ergibt sich dann aus dem Produkt der Wahrscheinlichkeiten aus den Teilnetzen.

Diese Vereinfachungen sind umso eher möglich, je dünner das Bayes-Netz mit Pfeilen besetzt ist, und je näher die Instanziierungen an den Wurzelknoten liegen. Damit bei Bayes-Netzen höherdimensionaler Verteilungen kein erheblicher Rechenaufwand entsteht, ist also besondere Sorgfalt zum Entwurf eines Bayes-Netzes mit einer einfachen Struktur in diesem Sinne angebracht.

4 Schlußbemerkung

Im Rahmen dieser Arbeit sind Bayes-Netze als Werkzeug zur Handhabung probabilistischer Wissensbasen auf der Grundlage hochdimensionaler Verteilungen dargestellt worden. Dabei dienen die Bayes-Netze sowohl als Hilfsmittel zur Gewinnung der Verteilungsfunktion aus dem Erfahrungswissen heraus, als auch zur Darstellung der bedingten Unabhängigkeiten in der resultierenden Verteilung. Dieses Vorgehen kommt in ähnlicher Weise in einigen probabilistischen Expertensystemen wie bspw. HUGIN zum Einsatz.

Trotz der formal einfachen Handhabung dieses Ansatzes, weist dieser dennoch einige Schwächen auf. Bei einer bekannten Reihenfolge von Ursache und Wirkung ist die Aufstellung eines Bayes-Netzes relativ einfach. Ist die Abhängigkeitsstruktur jedoch nicht genau genug bekannt, so kann es bei ungünstiger Indizierung zu dicht besetzten Bayes-Netzen kommen. Dieses würde die Problematik der mitunter zahlreichen bedingten Wahrscheinlichkeiten, die zu ermitteln sind, noch weiter verschärfen. Die benötigten Wahrscheinlichkeiten müssen alle wenigstens näherungsweise bekannt sein. Bei der Anwendung des Entropieprinzips wie bspw. in SPIRIT kann hingegen unklares Wissen bei minimaler Fehlinformation konsistent aufgefüllt werden.

Aus der gerichteten Struktur des Bayes-Netzes folgt ferner, daß die bedingten Unabhängigkeiten jeweils nur bezüglich aller Vorgänger dargestellt werden können. Die azyklische Struktur verbietet darüber hinaus die Darstellung zyklischer Abhängigkeiten. Somit ist es im allgemeinen nicht möglich, alle vorhandenen Unabhängigkeiten der Ursprungsverteilung in Bayes-Netzen zu nutzen oder darzustellen.

Literatur

Borgelt, Christian; Timm, Heiko und Kruse, Rudolf (2000): *Unsicheres und vages Wissens*, http://fuzzy.cs.uni-magdeburg.de/studium/uvws/txt/uvws.pdf.

Bronstein, I. N. und Semendjajew, K. A. (1989): *Taschenbuch der Mathematik*, Verlag Harri Deutsch, Thun, 24. Aufl.

Fernuniversität Hagen (2001): *SPIRITshell 3.1*, Fernuniversität Hagen, Lehrstuhl für Betriebswirtschaftslehre, insb. Operations Research, 58084 Hagen.

Jensen, Finn V. (1996): *An introduction to Bayesian networks*, UCL Press Limited, University College London, Gower Street, London WC1E 6BT.

Koch, Karl-Rudolf (2000): *Einführung in die Bayes-Statistik*, Springer-Verlag Berlin Heidelberg.

Lauritzen, S. L. und Spiegelhalter, D. J. (1988): *Local Computations with Probabilities on Graphical Structures and their Application to Expert Systems*, J. R. Statist. Soc. B, 50(2):157–224.

Meyer, Carl-Heinz (1998): *Korrektes Schließen bei unvollständiger Information*, Doktorarbeit, Fernuniversität Hagen, Lehrstuhl für Betriebswirtschaftslehre, insb. Operations Research, 58084 Hagen, (http://www.fernuni-hagen.de/BWLOR/papers/1998mey.ps.gz).

Pearl, Judea (1991): *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan Kaufmann Publishers, Inc., revised second printing.

Rödder, Wilhelm; Kulmann, Friedhelm und Kopittke, Bruno Hartmut (1997): *Wissensbasierte Entscheidungsunterstützung mit SPIRIT*, Fernuniversität Hagen, Kurs 00840.

Whittaker, Joe (1990): *Graphical Models in Applied Multivariate Statistics*, Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons Ltd., Baffins Lane, Chichester, West Sussex PO19 1UD, England.