Bayessche Netze: Ein einführendes Beispiel

Bernhard Kipper, Thorsten Brants, Marcus Plach
Graduiertenkolleg Kognitionswissenschaft
Universität des Saarlandes
Postfach 15 11 50
66041 Saarbrücken
E-Mail: kipper@cs.uni-sb.de,
thorsten@coli.uni-sb.de,
marcus@cops.uni-sb.de

Ralph Schäfer
Fachbereich 14: Informatik
Universität des Saarlandes
Postfach 15 11 50
66041 Saarbrücken
E-Mail: ralph@cs.uni-sb.de

Zusammenfassung

Bayessche Netze stellen einen vielbeachteten Formalismus zur Repräsentation und Verarbeitung von unsicherem Wissen dar. Zum Formalismus der Bayesschen Netze existieren zwar einige einführende Arbeiten; was diesen Einführungen jedoch fehlt, ist eine Illustration der innerhalb von Bayesschen Netzen verwendeten Mechanismen an Hand konkreter (Zahlen-)Beispiele. Mit der vorliegenden Arbeit soll genau diese Lücke geschlossen werden: Die grundlegende Struktur Bayesscher Netze wird durch die Modellierung eines Beispielszenarios erläutert. In dem daraus resultierenden Beispielnetz werden ferner die probabilistischen Methoden, die bei Bayesschen Netzen Anwendung finden, mit konkreten Zahlenwerten durchgerechnet.

1 Einleitung

Bei **Bayesschen Netzen** (engl. *Bayesian networks*), die zuweilen auch als **Belief-Netze** (engl. *belief networks*) bezeichnet werden (z.B. in [Abramson 91] oder [Henrion 89]), handelt es sich um einen Formalismus zur Verarbeitung von unsicherem Wissen, der sowohl die Repräsentation unsicheren Wissens als auch die Durchführung von Inferenzen über dem repräsentierten unsicheren Wissen ermöglicht. In dieser Arbeit werden die grundlegenden Aspekte Bayesscher Netze erläutert und entsprechende Notationen eingeführt. Die Ausführungen werden dabei an einem durchgängigen Beispiel illustriert.

In Abschnitt 2 wird zunächst die grundlegende Struktur Bayesscher Netze erläutert. Dabei wird ein kleines Beispielnetz eingeführt. In Abschnitt 3 wird dann an Hand des gleichen Beispielnetzes die Initialisierung Bayesscher Netze erläutert. In Abschnitt 4 wird (ebenfalls an Hand des gleichen Beispielnetzes) illustriert, wie Wahrscheinlichkeiten zwischen den einzelnen Knoten innerhalb von Bayesschen Netzen propagiert werden. Abschnitt 5 schließlich enthält noch einige abschließende Bemerkungen zu Bayesschen Netzen.

Für eine Einführung in den Repräsentationsformalismus der Bayesschen Netze sei ferner auf [Charniak 91], [Kruse et al. 91] und [Pearl 86] verwiesen. Ausführliche Darstellungen Bayesscher Netze finden sich z.B. in [Neapolitan 90] und in [Pearl 91].

2 Die Struktur Bayesscher Netze

Bei Bayesschen Netzen handelt es sich um gerichtete, azyklische Graphen. Die Knoten des Graphen enthalten Informationen über die Wahrscheinlichkeit des Zutreffens bestimmter Ereignisse, die im folgenden als **Hypothesen** bezeichnet werden. Insofern repräsentieren sie das (unsichere) Wissen über das Zutreffen dieser Hypothesen.

Die Kanten des Graphen repräsentieren gewisse Abhängigkeiten zwischen den Hypothesen, die ihrerseits in verschiedenen Knoten repräsentiert sind. Genauer gesagt handelt es sich um Abhängigkeiten zwischen der Wahrscheinlichkeit des Zutreffens unterschiedlicher Hypothesen.

Diese grundlegende Struktur Bayesscher Netze soll im folgenden an einem kleinen Beispiel illustriert werden. Dieses Beispiel, das durch die Überlegungen in [Jameson 90] angeregt wurde, ist durch eine Prüfungssituation gegeben: Ein Prüfling hat eine Frage zu beantworten, und man ist an einer Einschätzung darüber interessiert, ob der Prüfling die Antwort auf die Frage weiß. Bei einer solchen Einschätzung sind zumindest die Faktoren *Wissensniveau des Prüflings* und *Schwierigkeitsgrad der Frage* zu berücksichtigen: Je höher das Wissensniveau des Prüflings ist bzw. je leichter die Frage ist, desto größer ist die Wahrscheinlichkeit, daß der Prüfling die Antwort auf die Frage kennt. Andererseits beeinflußt das Wissen bzw. das fehlende Wissen um die Antwort das Verhalten des Prüflings, genauer gesagt seine Gestik und seine Mimik: Kennt der Prüfling die Antwort, so wird seine Gestik viel eher ruhig und seine Mimik viel eher gelassen sein, als wenn er sie nicht kennt.

Die in diesem Beispielszenario auftretenden Zusammenhänge lassen sich auf die folgende Art und Weise mittels eines Bayesschen Netzes modellieren (vgl. Abbildung 1 auf der nachfolgenden Seite 3): Es werden fünf Knoten eingeführt, die die folgenden Hypothesen repräsentieren:

- Der Prüfling hat ein hohes bzw. ein niedriges **Wissensniveau** (Knoten N).
- Die Frage ist schwer, mittel bzw. leicht zu beantworten (Knoten **Schwierigkeit**, kurz S).
- Der Prüfling weiß die Antwort auf die Frage bzw. weiß sie nicht (Knoten **Wissen**, kurz W).
- Die **Gestik** des Prüflings ist hektisch bzw. ruhig (Knoten G).
- Die **Mimik** des Prüflings ist entsetzt bzw. gelassen (Knoten M).

Zwischen diesen fünf Knoten werden die in Abbildung 1 gezeigten vier Kanten eingeführt. Durch die Kanten wird repräsentiert, daß das Wissensniveau des Prüflings und der Schwierigkeitsgrad der Frage Auswirkungen auf die Einschätzung des Wissens um die Antwort auf die Frage hat und daß das (vorhandene oder nicht vorhandene) Wissen wiederum die Gestik und die Mimik des Prüflings beeinflußt.

Dieses kleine Beispiel macht bereits deutlich, auf welche Weise die Struktur eines Bayesschen Netzes die Zusammenhänge zwischen den zu repräsentierenden Sachverhalten wiederspiegelt. Im nachfolgenden Abschnitt 3 wird dann angegeben, wie diese zunächst noch „hohle“ Struktur eines Bayesschen Netzes mit konkreten Wahrscheinlichkeitswerten aufzufüllen ist.

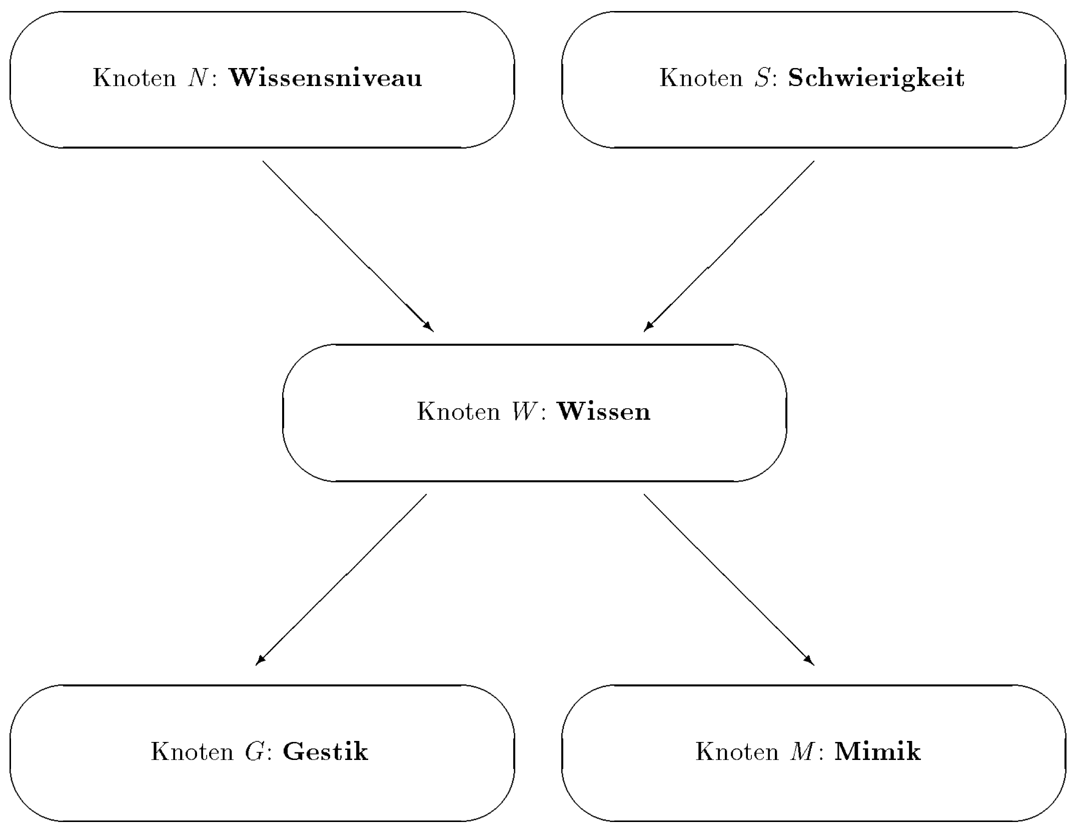


Abbildung 1: Die Struktur des Bayesschen Netzes zur Repräsentation der im Text beschriebenen Prüfungssituation.

3 Die Initialisierung Bayesscher Netze

Um ein funktionsfähiges Bayessches Netz zu erhalten, müssen zunächst A-priori-Wahrscheinlichkeiten für die durch die Knoten repräsentierten Hypothesen angegeben werden. Dabei wird ein Unterschied gemacht zwischen **Wurzelknoten** (d.h. Knoten, die in dem gerichteten Graphen keine Vorgängerknoten besitzen) und den restlichen Knoten (die im folgenden als **innere Knoten** bezeichnet werden):

- Bei Wurzelknoten werden direkt Wahrscheinlichkeitswerte für die jeweiligen Hypothesen angegeben.
- Bei inneren Knoten werden bedingte Wahrscheinlichkeiten spezifiziert, durch die festgelegt ist, wie wahrscheinlich das Eintreten einer Hypothese ist in Abhängigkeit von den Hypothesen, die durch den oder die Vorgängerknoten repräsentiert werden.

Nachdem die jeweiligen Wahrscheinlichkeiten für alle Knoten vorgegeben wurden, müssen die Knoten des Bayesschen Netzes noch mit den sich daraus ergebenden konkreten Wahrscheinlichkeitswerten initialisiert werden.

Dementsprechend werden für das in Abschnitt 2 eingeführte Beispielnetz zunächst die A-priori-Wahrscheinlichkeiten für die Wurzelknoten (Abschnitt 3.1) und die bedingten Wahrscheinlichkeiten für die inneren Knoten (Abschnitt 3.2) angegeben. In Abschnitt 3.3 wird dann die eigentliche Initialisierung des Netzes dadurch beschrieben, daß spezifiziert wird, welche Wahrscheinlichkeitswerte in den einzelnen Knoten gespeichert werden.

3.1 Die A-priori-Wahrscheinlichkeiten der Wurzelknoten

Das in Abschnitt 2 eingeführte Beispielnetz enthält zwei Wurzelknoten: Knoten N (Wissensniveau) und Knoten S (Schwierigkeit). In der Beispielsituation sei zunächst vorausgesetzt, daß das Wissensniveau des Prüflings hoch ist und daß die Frage eher schwer ist als leicht.

Welche Wahrscheinlichkeiten müssen nun angegeben werden, um diese Situation zu repräsentieren? Da beim Wissensniveau zwischen den beiden Alternativen *hoch* und *niedrig* unterschieden wird, sind im Knoten N zwei Hypothesen zu repräsentieren: die, daß das Wissensniveau des Prüflings hoch ist, und die, daß es niedrig ist.[1] Für diese beiden Hypothesen werden die folgenden Wahrscheinlichkeiten spezifiziert:

$$\begin{aligned} P(N = hoch) &= 0.8 \\ P(N = niedrig) &= 0.2 \end{aligned}$$

Diese beiden Wahrscheinlichkeiten stellen eine Wahrscheinlichkeitsverteilung über dem Wissensniveau des Prüflings dar, da in der Beispielsituation vorausgesetzt wird, daß sich die beiden Hypothesen *hohes Wissensniveau* und *niedriges Wissensniveau* gegenseitig ausschließen und daß sie gleichzeitig alle Möglichkeiten abdecken. In Vektorschreibweise erhält man somit:

[1] Die Hypothese, daß das Wissensniveau des Prüflings hoch ist, wird im folgenden zuweilen auch als Hypothese N_1 bezeichnet und die, daß sein Wissensniveau niedrig ist, als Hypothese N_2.

$$P(N) = \begin{pmatrix} 0.8 \\ 0.2 \end{pmatrix}$$

Analog dazu sind im Knoten S drei Hypothesen zu repräsentieren, da beim Schwierigkeitsgrad der Frage zwischen drei Alternativen unterschieden wird (*schwer*, *mittel* und *leicht*).[2] Für diese drei Hypothesen werden im Beispielnetz die folgenden Wahrscheinlichkeiten spezifiziert:

$$\begin{aligned} P(S = schwer) &= 0.6 \\ P(S = mittel) &= 0.3 \\ P(S = leicht) &= 0.1 \end{aligned}$$

Diese Wahrscheinlichkeitsverteilung läßt sich in Vektorschreibweise folgendermaßen darstellen:

$$P(S) = \begin{pmatrix} 0.6 \\ 0.3 \\ 0.1 \end{pmatrix}$$

Damit sind die A-priori-Wahrscheinlichkeiten für die Wurzelknoten des Beispielnetzes vollständig spezifiert.

3.2 Die bedingten Wahrscheinlichkeiten der inneren Knoten

In dem in Abschnitt 2 eingeführten Beispielnetz sind drei innere Knoten enthalten: Knoten W (Wissen), Knoten G (Gestik) und Knoten M (Mimik). Für jeden dieser drei Knoten müssen bedingte Wahrscheinlichkeiten spezifiziert werden, die angeben, wie wahrscheinlich die in dem jeweiligen Knoten repräsentierten Hypothesen sind unter der Bedingung, daß die in den Vorgängerknoten repräsentierten Hypothesen zutreffen bzw. nicht zutreffen. Im folgenden wird zunächst der Knoten W genauer betrachtet, dann G und zuletzt M.

In dem Beispielnetz für die Prüfungssituation soll der Knoten W die Einschätzung des Sachverhalts repräsentieren, ob der Prüfling die Antwort auf die ihm gestellte Frage weiß oder nicht. Diese Einschätzung wird im Beispiel davon abhängig gemacht, wie hoch das Wissensniveau des Prüflings ist und wie schwierig die gestellte Frage ist. Was diese Abhängigkeiten anbelangt, ist es plausibel, die folgenden Annahmen zu machen:

- Wenn der Prüfling ein hohes Wissensniveau hat und ihm eine leichte Frage gestellt wird, ist es wahrscheinlicher, daß er die Antwort weiß, als daß er sie nicht weiß.

- Umgekehrt ist es wahrscheinlicher, daß der Prüfling die Antwort nicht weiß, wenn er ein niedriges Wissensniveau hat und ihm eine schwere Frage gestellt wird.

[2] Die drei Hypothesen, daß die Frage schwer, mittel bzw. leicht ist, werden im folgenden zuweilen auch als S_1, S_2 bzw. S_3 bezeichnet.

- Wenn der Prüfling zwar ein hohes Wissensniveau hat, aber eine schwere Frage gestellt bekommt, dann ist es offen, ob er die Antwort weiß oder nicht (d.h. die Wahrscheinlichkeiten beider Hypothesen sind etwa gleich groß).

- Analog dazu ist es auch dann offen, ob der Prüfling die Antwort weiß, wenn er ein niedriges Wissensniveau hat und ihm eine leichte Frage gestellt wird.

Diese Annahmen dienten als Leitlinien bei der Festlegung der folgenden bedingten Wahrscheinlichkeiten, die angegeben, wie wahrscheinlich es ist, daß der Prüfling die Antwort auf die Frage weiß bzw. nicht weiß,[3] unter den Bedingungen, daß er ein hohes bzw. ein niedriges Wissensniveau hat und daß die Frage schwer, mittel bzw. leicht ist:

$$\begin{array}{llllll}
P\,(W = ja & | & N = hoch, & S = schwer) & = & 0.5 \\
P\,(W = nein & | & N = hoch, & S = schwer) & = & 0.5 \\
 & & & & & \\
P\,(W = ja & | & N = hoch, & S = mittel) & = & 0.6 \\
P\,(W = nein & | & N = hoch, & S = mittel) & = & 0.4 \\
 & & & & & \\
P\,(W = ja & | & N = hoch, & S = leicht) & = & 0.8 \\
P\,(W = nein & | & N = hoch, & S = leicht) & = & 0.2 \\
 & & & & & \\
P\,(W = ja & | & N = niedrig, & S = schwer) & = & 0.3 \\
P\,(W = nein & | & N = niedrig, & S = schwer) & = & 0.7 \\
 & & & & & \\
P\,(W = ja & | & N = niedrig, & S = mittel) & = & 0.4 \\
P\,(W = nein & | & N = niedrig, & S = mittel) & = & 0.6 \\
 & & & & & \\
P\,(W = ja & | & N = niedrig, & S = leicht) & = & 0.5 \\
P\,(W = nein & | & N = niedrig, & S = leicht) & = & 0.5
\end{array}$$

Diese bedingten Wahrscheinlichkeiten können auch in Form einer $2 \times 2 \times 3$-Matrix (also einer „dreidimensionalen“ Matrix) dargestellt werden, wobei die Hypothesen mittels der bereits eingeführten Kürzel W_i ($i \in \{1,2\}$), N_j ($j \in \{1,2\}$) und S_k ($k \in \{1,2,3\}$) bezeichnet werden:

$$P(W_1|N_j, S_k) = \begin{pmatrix} 0.5 & 0.6 & 0.8 \\ 0.3 & 0.4 & 0.5 \end{pmatrix} \qquad P(W_2|N_j, S_k) = \begin{pmatrix} 0.5 & 0.4 & 0.2 \\ 0.7 & 0.6 & 0.5 \end{pmatrix}$$

Da die in dieser Matrix enthaltenen bedingten Wahrscheinlichkeiten den Zusammenhang zwischen dem Knoten W und seinen Vorgängerknoten N und S spezifizieren, wird diese Matrix als **Verbindungsmatrix** (engl. *link matrix*) $M(W|N,S)$ bezeichnet, wobei die einzelnen Matrixeinträge festgelegt sind durch:

$$M(W|N,S)_{i,j,k} := P(W_i|N_j, S_k)$$

[3] Die beiden Hypothesen, daß der Prüfling die Antwort auf die Frage weiß bzw. nicht weiß, werden im folgenden auch als W_1 bzw. W_2 bezeichnet.

Eine solche Verbindungsmatrix muß für jeden inneren Knoten vorhanden sein. Hat ein innerer Knoten K die Vorgängerknoten $V_1, \ldots, V_n$, so ist die Anzahl der Dimensionen der Verbindungsmatrix $M(K|V_1, \ldots, V_n)$ von K gegeben durch die Anzahl n der Vorgängerknoten plus 1. Bei der Festlegung der in den Verbindungsmatrizen enthaltenen bedingten Wahrscheinlichkeiten ist zu beachten, daß diese so normiert sein müssen, daß für die in K repräsentierten Hypothesen K_i $(i = 1, \ldots, h)$ die folgende Bedingung erfüllt ist (dabei seien $V_{1,j_1}, \ldots, V_{n,j_n}$ beliebige (aber feste), in den jeweiligen Vorgängerknoten $V_1, \ldots, V_n$ repräsentierte Hypothesen):

$$\sum_{i=1}^{h} M(K|V_1, \ldots, V_n)_{i,j_1,\ldots,j_n} = 1$$

Im Falle des Knotens W aus dem Beispielnetz bedeutet dies folgendes: Da in W zwei Hypothesen repräsentiert sind und da W zwei Vorgängerknoten besitzt (nämlich die Knoten N und S), muß für jedes Paar N_{j_1}, S_{j_2} ($j_1 \in \{1,2\}$, $j_2 \in \{1,2,3\}$) von in N bzw. S repräsentierten Hypothesen gelten:

$$\begin{array}{ll} & \sum_{i=1}^{2} M(W|N,S)_{i,j_1,j_2} = 1 \\ \Leftrightarrow & \sum_{i=1}^{2} P(W_i|N_{j_1}, S_{j_2}) = 1 \\ \Leftrightarrow & P(W_1|N_{j_1}, S_{j_2}) + P(W_2|N_{j_1}, S_{j_2}) = 1 \\ \Leftrightarrow & P(W = ja|N_{j_1}, S_{j_2}) + P(W = nein|N_{j_1}, S_{j_2}) = 1 \end{array}$$

Diese Bedingung ist bei den oben spezifizierten bedingten Wahrscheinlichkeiten für den Knoten W erfüllt, da die Summe von jedem der 6 angegebenen Paare von bedingten Wahrscheinlichkeiten gleich 1 ist.

Nun zum Knoten G. Bei der Modellierung der Prüfungssituation wurde G vom Knoten W abhängig gemacht. Diese Abhängigkeit zwischen den beiden Knoten kann folgendermaßen umschrieben werden: Wenn der Prüfling die Antwort auf die ihm gestellte Frage weiß, dann ist es wahrscheinlicher, daß seine Gestik ruhig ist, als daß sie hektisch ist, und umgekehrt.[4] Daher wurden im Beispielnetz für den Knoten G die folgenden bedingten Wahrscheinlichkeiten festgelegt:

$$\begin{array}{llcl} P\,(G = hektisch & |\ W = ja) & = & 0.3 \\ P\,(G = ruhig & |\ W = ja) & = & 0.7 \\ & & & \\ P\,(G = hektisch & |\ W = nein) & = & 0.9 \\ P\,(G = ruhig & |\ W = nein) & = & 0.1 \end{array}$$

Auf Grund dieser bedingten Wahrscheinlichkeiten ergibt sich für den Knoten G die folgende Verbindungsmatrix:

$$M(G|W) = \begin{pmatrix} 0.3 & 0.9 \\ 0.7 & 0.1 \end{pmatrix}$$

[4] Die beiden Hypothesen, daß die Gestik des Prüflings hektisch ist bzw. daß sie ruhig ist, werden im folgenden auch als G_1 bzw. G_2 bezeichnet.

Analog zum Knoten G wurden auch die bedingten Wahrscheinlichkeiten für den Knoten M spezifiziert: M ist ebenfalls vom Knoten W abhängig, und man kann davon ausgehen, daß – falls der Prüfling die Antwort auf die Frage weiß – die Wahrscheinlichkeit, daß seine Mimik entsetzt ist, geringer ist als die, daß sie gelassen ist, und umgekehrt.[5] Diesen Umstand spiegeln die folgende Werte wieder:

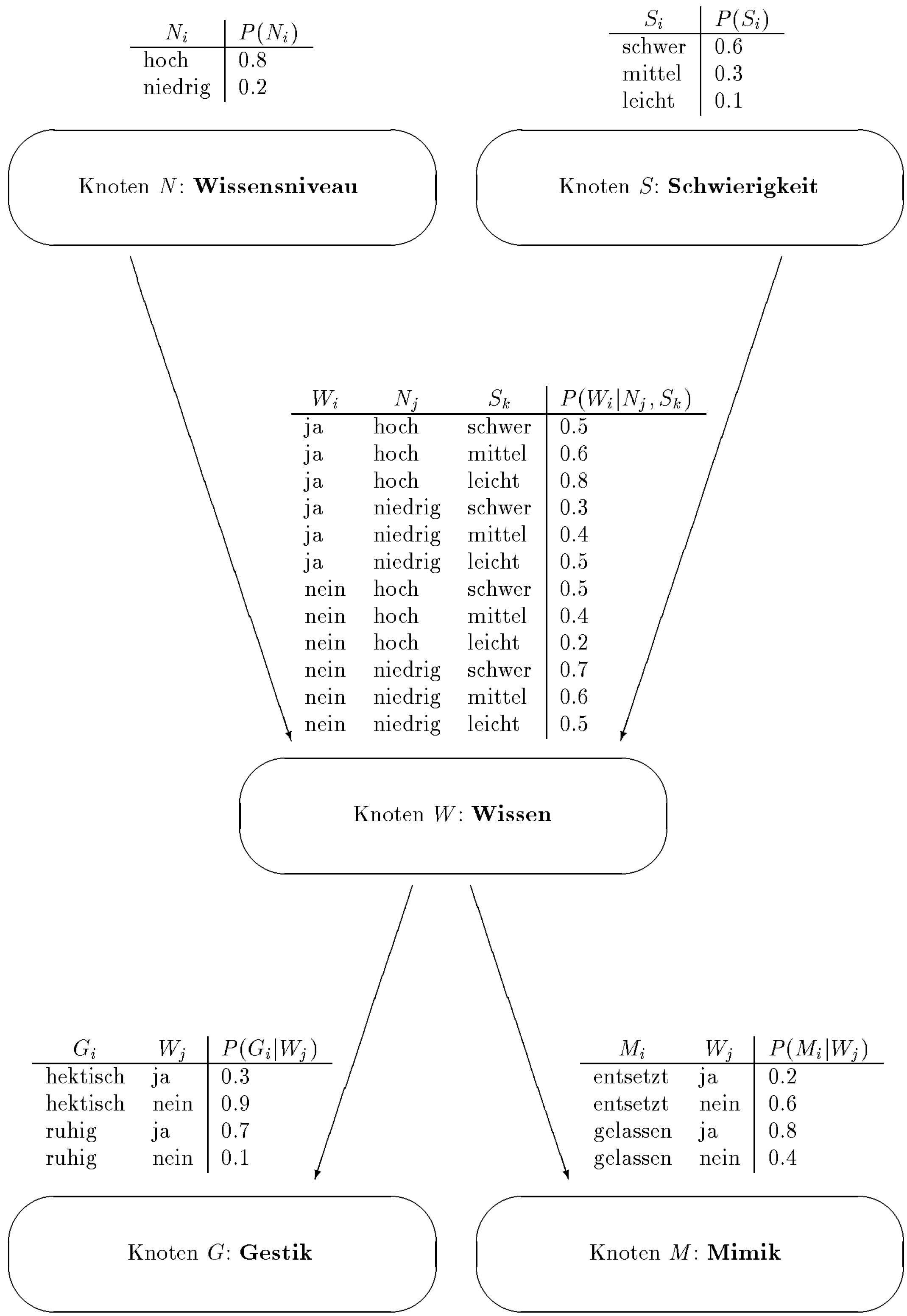
$$
\begin{array}{llll}
P\,(M = entsetzt & \mid \; W = ja) & = & 0.2 \\
P\,(M = gelassen & \mid \; W = ja) & = & 0.8 \\
\\
P\,(M = entsetzt & \mid \; W = nein) & = & 0.6 \\
P\,(M = gelassen & \mid \; W = nein) & = & 0.4
\end{array}
$$

Dementsprechend ergibt sich für den Knoten M die folgende Verbindungsmatrix:

$$
M(M|W) = \begin{pmatrix} 0.2 & 0.6 \\ 0.8 & 0.4 \end{pmatrix}
$$

Damit sind für alle inneren Knoten des Beispielnetzes die Verbindungsmatrizen spezifiziert. Zusammen mit den A-priori-Wahrscheinlichkeiten, die für die Wurzelknoten festgelegt wurden, werden sie in Abbildung 2 auf der nachfolgenden Seite 9 noch einmal im Überblick gezeigt.

[5]Die beiden Hypothesen, daß die Mimik des Prüflings entsetzt ist bzw. daß sie gelassen ist, werden im folgenden auch als M_1 bzw. M_2 bezeichnet.



N_i	$P(N_i)$
hoch	0.8
niedrig	0.2

S_i	$P(S_i)$
schwer	0.6
mittel	0.3
leicht	0.1

W_i	N_j	S_k	$P(W_i\|N_j, S_k)$
ja	hoch	schwer	0.5
ja	hoch	mittel	0.6
ja	hoch	leicht	0.8
ja	niedrig	schwer	0.3
ja	niedrig	mittel	0.4
ja	niedrig	leicht	0.5
nein	hoch	schwer	0.5
nein	hoch	mittel	0.4
nein	hoch	leicht	0.2
nein	niedrig	schwer	0.7
nein	niedrig	mittel	0.6
nein	niedrig	leicht	0.5

G_i	W_j	$P(G_i\|W_j)$
hektisch	ja	0.3
hektisch	nein	0.9
ruhig	ja	0.7
ruhig	nein	0.1

M_i	W_j	$P(M_i\|W_j)$
entsetzt	ja	0.2
entsetzt	nein	0.6
gelassen	ja	0.8
gelassen	nein	0.4

Abbildung 2: Die A-priori-Wahrscheinlichkeiten der Wurzelknoten und die Verbindungsmatrizen der inneren Knoten bei dem Beispielnetz für die Prüfungssituation.

3.3 Die Initialisierung der einzelnen Knoten

Es bleibt noch zu klären, in welcher Form die Knoten des Bayesschen Netzes die Wahrscheinlichkeitsverteilungen enthalten. Sei K ein beliebiger Knoten eines Bayesschen Netzes. In K seien die Hypothesen $K_1, \ldots, K_h$ repräsentiert. Zu jeder Hypothese K_j ($j \in \{1, \ldots, h\}$) im Knoten K werden die folgenden drei Wahrscheinlichkeitswerte spezifiziert:

- Der sogenannte **π-Wert**, der die gegenwärtige Stärke der kausalen Unterstützung einer Hypothese durch die Vorgängerknoten angibt.
- Der sogenannte **λ-Wert**, der die gegenwärtige Stärke der diagnostischen Unterstützung einer Hypothese durch die Nachfolgerknoten angibt.
- Der sogenannte ***BEL*-Wert**, der das Gesamtvertrauen in eine Hypothese angibt.

Der BEL-Wert einer Hypothese K_j des Knotens K berechnet sich dabei aus dem π-Wert und dem λ-Wert desselben Knotens gemäß der folgenden Formel:[6]

$$BEL(K_j) = \alpha \cdot \lambda(K_j) \cdot \pi(K_j) \tag{1}$$

Insofern stellen der π-Wert und der λ-Wert also Hilfsmittel zur Berechnung des Gesamtvertrauens BEL dar, die Aufschluß darüber geben, wie sich das Gesamtvertrauen in eine Hypothese zusammensetzt.

An dieser Stelle folgt ein kurzer Einschub, in dem die durch Formel (1) gegebene Berechnungsvorschrift für das Gesamtvertrauen in eine Hypothese begründet wird.

Zunächst ist festzuhalten, daß Formel (1) direkt auf der sogenannten **Bayesschen Formel** basiert.[7] Die Bayessche Formel, die sich ihrerseits aus der Definition von bedingten Wahrscheinlichkeiten ableitet,[8] lautet in ihrer allgemeinen Form (d.h. für beliebig viele, disjunkte Hypothesen $H_1, \ldots, H_n$) wie folgt:

$$P(H_j|E) = \frac{P(H_j) \cdot P(E|H_j)}{\sum\limits_{k=1}^{n} (P(H_k) \cdot P(E|H_k))} \qquad \text{mit} \quad \begin{array}{l} j \in \{1, \ldots, n\} \\ H_j \cap H_k = \emptyset \text{ für } j \neq k \text{ und} \\ P(H_j) > 0 \text{ für } j = 1, \ldots, n \end{array}$$

Im Prinzip ergibt sich Formel (1) aus der Bayesschen Formel, indem man folgende Einsetzungen vornimmt:

---

[6] Dabei ist α ein Normierungsfaktor, der so gewählt wird, daß die folgende Bedingung erfüllt ist:

$$\sum_{j=1}^{h} BEL(K_j) = 1$$

[7] Dieser Umstand liefert zusammen mit der Tatsache, daß die Berechnung der BEL-Werte mit Hilfe von Formel (1) von grundlegender Bedeutung innerhalb von Bayesschen Netzen ist, die Begründung dafür, daß der hier vorgestellte Formalismus zur Repräsentation und Verarbeitung von unsicherem Wissen den Namen *Bayessche Netze* trägt.

[8] Siehe beispielsweise [Bauer 91] oder [Hinderer 75] für eine explizite Herleitung der Bayesschen Formel.

- Als $P(H_j)$ wird der π-Wert einer Hypothese K_j, d.h. $\pi(K_j)$, genommen.
- Für die bedingten Wahrscheinlichkeiten $P(E|H_j)$ werden die λ-Werte einer Hypothese K_j, also $\lambda(K_j)$, eingesetzt.
- Der (konstanten) Summe im Nenner der Bayesschen Formel entspricht der Normierungsfaktor α in Formel (1).
- Der BEL-Wert einer Hypothese K_j stellt dann die bedingte Wahrscheinlichkeit $P(H_j|E)$ dar.

Allerdings unterscheidet sich Formel (1) von der Bayesschen Formel im Hinblick auf die beiden folgenden Punkte:

1. In der Bayesschen Formel wird vorausgesetzt, daß die Evidenz E tatsächlich eingetreten ist bzw. daß man weiß, daß E eingetreten ist. Bei Bayesschen Netzen muß man dies nicht voraussetzen. Bayessche Netze stellen somit eine Erweiterung des Bayesschen Ansatzes für den Fall unsicherer Evidenz dar: Zum einen ist es möglich, daß in den λ-Wert einer Hypothese die diagnostische Unterstützung mehrerer Evidenzen eingegangen ist, und zum anderen drücken die λ-Werte aus, daß die diagnostische Unterstützung eventuell auch durch nicht direkt beobachtbare Evidenzen (d.h. durch Nachfolgerknoten im Netz) vermittelt werden kann.

2. Der π-Wert einer Hypothese stellt – im Gegensatz zu den Wahrscheinlichkeitswerten $P(H_j)$ in der Bayesschen Formel – im allgemeinen keine A-priori-Wahrscheinlichkeit der Hypothese dar. Vielmehr fassen die π-Werte (in analoger Weise wie die λ-Werte) die kausale Unterstützung durch eventuell auch nicht direkt beobachtbare Ereignisse (d.h. durch Vorgängerknoten im Netz) zusammen. Dementsprechend gilt, daß die A-priori-Wahrscheinlichkeit einer Hypothese im allgemeinen nicht bekannt sein muß.

Infolgedessen bedeutet $\lambda(K_j)$ nur dann dasselbe wie die bedingte Wahrscheinlichkeit $P(E|H_j)$ in der Bayesschen Formel, wenn es sich bei K um einen Blattknoten handelt, und $\pi(K_j)$ bedeutet nur dann dasselbe wie die A-priori-Wahrscheinlichkeit $P(H)$ in der Bayesschen Formel, wenn K ein Wurzelknoten ist.

Nach diesem kurzen Einschub zurück zu den Wahrscheinlichkeitswerten, die in den Knoten eines Bayesschen Netzes gespeichert sind.

Da in einem Netzknoten mehrere Hypothesen repräsentiert werden (so viele, wie verschiedene Alternativen bei dem Sachverhalt unterschieden werden, der durch den Knoten repräsentiert werden soll) und da in den Knoten für jede Hypothese ein π-, ein λ- und ein BEL-Wert gespeichert wird, kann man insgesamt davon sprechen, daß jeder Knoten eines Bayesschen Netzes einen **π-Vektor**, einen **λ-Vektor** und einen **BEL-Vektor** enthält. Analog zu Gleichung (1) besteht auch zwischen den λ-, π- und BEL-Vektoren, in denen die einzelnen λ-, π- und BEL-Werte enthalten sind, der folgende Zusammenhang (dabei bezeichne das Symbol $\times$ die komponentenweise Multiplikation von Vektoren):

$$
\begin{aligned}
BEL(K) &= \alpha \cdot \quad \lambda(K) \quad \times \quad \pi(K) \\
&= \alpha \cdot \begin{pmatrix} \lambda(K_1) \\ \vdots \\ \lambda(K_h) \end{pmatrix} \times \begin{pmatrix} \pi(K_1) \\ \vdots \\ \pi(K_h) \end{pmatrix}
\end{aligned}
$$

In den nachfolgenden zwei Abschnitten 3.3.1 bzw. 3.3.2 wird angegeben, wie sich in dem Beispielnetz zu der Prüfungssituation die π-, λ- und BEL-Vektoren für die Wurzelknoten N und S bzw. für die inneren Knoten W, G und M berechnen.

Zu diesem Prozeß der Initialisierung der in den Knoten enthaltenen Wahrscheinlichkeitswerte sei an dieser Stelle noch die folgende Anmerkung gemacht: In der vorliegenden Beschreibung Bayesscher Netze wird unterschieden zwischen einer Initialisierungsphase, in der die π-, λ- und BEL-Vektoren zum ersten Mal mit Wahrscheinlichkeitswerten belegt werden, und der späteren Propagierung von Wahrscheinlichkeiten innerhalb des Bayesschen Netzes. Eine solche Unterscheidung ist bei Bayesschen Netzen jedoch nicht notwendig, da auch in der Initialisierungsphase die gleichen Mechanismen Anwendung finden können wie bei der Propagierung von Wahrscheinlichkeiten (indem in den in Abschnitt 4 vorgestellten Berechnungsvorschriften diejenigen Wahrscheinlichkeitswerte, die zwischen den Knoten ausgetauscht werden und die in der Initialisierungsphase noch nicht vorhanden sind, als 1 angenommen werden). Da es sich jedoch bei der Initialisierungsphase eines Bayesschen Netzes um einen Spezialfall handelt, der gewisse Vereinfachungen erlaubt, erscheint es didaktisch vorteilhafter, zunächst den einfacheren Spezialfall der Initialisierung zu beschreiben und dann erst den allgemeinen Fall der Propagierung von Wahrscheinlichkeiten.

3.3.1 Die Berechnung der π-, λ- und BEL-Vektoren der Wurzelknoten

Der π-Wert einer in einem Knoten repräsentierten Hypothese wurde soeben dahingehend charakterisiert, daß er die Stärke der Unterstützung einer Hypothese durch die Vorgängerknoten angibt. Da aber Wurzelknoten gemäß ihrer Definition keine Vorgängerknoten besitzen, hängt der π-Wert der in Wurzelknoten repräsentierten Hypothesen in keiner Weise von anderen Knoten ab, sondern nur von den vorgegebenen Wahrscheinlichkeiten. Das bedeutet, daß der π-Wert der in Wurzelknoten repräsentierten Hypothesen gleich den A-priori-Wahrscheinlichkeiten für diese Hypothesen ist.

In dem Beispielnetz für die Prüfungssituation sind also die π-Vektoren für die beiden Wurzelknoten N und S gleich den Wahrscheinlichkeitsverteilungen, die für die in diesen Knoten repräsentierten Hypothesen vorgegeben wurden:

$$\pi(N) = P(N) = \begin{pmatrix} 0.8 \\ 0.2 \end{pmatrix} \qquad \text{und} \qquad \pi(S) = P(S) = \begin{pmatrix} 0.6 \\ 0.3 \\ 0.1 \end{pmatrix}$$

Der λ-Wert einer in einem Knoten repräsentierten Hypothese wurde dagegen so charakterisiert, daß er die Stärke der Unterstützung einer Hypothese durch die Nachfolgerknoten angibt. Folglich errechnen sich die λ-Werte, die in Form einer **Likelihood** gegeben sind,[9] aus den Wahrscheinlichkeitswerten, die in den Nachfolgerknoten gespeichert sind. In der Initialisierungsphase tritt jedoch das Problem auf, daß die Wahrscheinlichkeitswerte der Nachfolgerknoten im allgemeinen noch nicht feststehen. Da somit noch keine spezifische Evidenz vorliegt und da die λ-Vektoren nicht normiert sein müssen (d.h. die Summe der im λ-Vektor enthaltenen λ-Werte braucht nicht 1 zu ergeben), werden in der Initialisierungsphase zunächst alle λ-Werte auf 1 gesetzt und erst später aktualisiert (siehe Abschnitt 4).

Gemäß Gleichung (1) berechnet sich der BEL-Vektor des Knotens N somit wie folgt, wobei die Vektoren komponentenweise miteinander multipliziert werden:

$$\begin{aligned} BEL(N) &= \alpha \cdot \lambda(N) \times \pi(N) \\ &= \alpha \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix} \times \begin{pmatrix} 0.8 \\ 0.2 \end{pmatrix} = \alpha \cdot \begin{pmatrix} 0.8 \\ 0.2 \end{pmatrix} \overset{\text{mit } \alpha=1}{=} \begin{pmatrix} 0.8 \\ 0.2 \end{pmatrix} \end{aligned}$$

In diesem Falle ist die Summe der Elemente des Vektors, der sich aus der Multiplikation des λ-Vektors mit dem π-Vektor ergibt, gleich 1. Daher braucht keine Normierung mehr vorgenommen zu werden, d.h. α wird auf 1 gesetzt.

Völlig analog zur Berechnung beim Knoten N läßt sich auch der BEL-Vektor beim Knoten S bestimmen:

$$\begin{aligned} BEL(S) &= \alpha \cdot \lambda(S) \times \pi(S) \\ &= \alpha \cdot \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \times \begin{pmatrix} 0.6 \\ 0.3 \\ 0.1 \end{pmatrix} = \alpha \cdot \begin{pmatrix} 0.6 \\ 0.3 \\ 0.1 \end{pmatrix} \overset{\text{mit } \alpha=1}{=} \begin{pmatrix} 0.6 \\ 0.3 \\ 0.1 \end{pmatrix} \end{aligned}$$

Damit sind alle Wahrscheinlichkeitswerte für die beiden Wurzelknoten des Beispielnetzes bestimmt. Diese Werte werden in Abbildung 3 auf Seite 18 noch einmal im Überblick gezeigt.

[9] Eine Likelihood ist eine bedingte Wahrscheinlichkeit $P(E|H)$, die die Wahrscheinlichkeit für das Eintreten einer Hypothese H unter der Voraussetzung einer beobachteten Evidenz E angibt.

3.3.2 Die Berechnung der π-, λ- und BEL-Vektoren der inneren Knoten

Genauso wie bei den Wurzelknoten tritt auch bei den inneren Knoten in der Initialisierungsphase das Problem auf, daß die Wahrscheinlichkeitswerte von Nachfolgerknoten im allgemeinen noch nicht feststehen. Aus den gleichen Gründen wie bei den Wurzelknoten werden somit auch bei den inneren Knoten alle λ-Werte zunächst mit 1 initialisiert und später aktualisiert (siehe Abschnitt 4).

Die Initialisierung der π-Vektoren der inneren Knoten unterscheidet sich jedoch von der der Wurzelknoten. Die π-Vektoren der inneren Knoten werden aus den π-Vektoren ihrer Vorgängerknoten berechnet. Diese Berechnung wird zum besseren Verständnis zunächst bei Knoten W beispielhaft durchgeführt, bevor die allgemeine Berechnungsmethode angegeben wird. Schließlich werden in diesem Abschnitt noch die π-Vektoren der Knoten G und M berechnet.

Grundlage für die Bestimmung der π-Werte innerer Knoten ist die Berechnung von Wahrscheinlichkeiten mittels der sogenannten **Formel von der totalen Wahrscheinlichkeit** (siehe z.B. [Dallmann & Elster 92], S. 455):

$$P(A) = \sum_{j=1}^{n} P(A|B_j) \cdot P(B_j) \qquad \text{mit } B_{j_1} \neq B_{j_2} \text{ für } j_1 \neq j_2$$

In dem Fall, daß das Ereignis A nicht nur von einer Reihe von Ereignissen $B_1, \ldots, B_n$ abhängig ist, sondern auch von einer weiteren Reihe von Ereignissen $C_1, \ldots, C_m$ (die unabhängig von $B_1, \ldots, B_n$ sein müssen), läßt sich die Formel von der totalen Wahrscheinlichkeit erweitern zu:

$$P(A) = \sum_{j=1}^{n} \sum_{k=1}^{m} P(A|B_j, C_k) \cdot P(B_j) \cdot P(C_k) \qquad \begin{array}{ll} \text{mit} & B_{j_1} \neq B_{j_2} \text{ für } j_1 \neq j_2 \\ \text{und} & C_{k_1} \neq C_{k_2} \text{ für } k_1 \neq k_2 \end{array}$$

Die π-Werte des Knotens W können mit Hilfe dieser Formel berechnet werden, indem folgende Werte in die Formel eingesetzt werden:

- Als Ereignisse $B_1, \ldots, B_n$ werden die Hypothesen genommen, die in einem der beiden Vorgängerknoten von W (nämlich N) repräsentiert sind: N_1 und N_2 (somit ist $n = 2$).
- Als Wahrscheinlichkeiten für das Eintreten dieser Ereignisse (d.h. als Werte für $P(B_1), \ldots, P(B_n)$) werden dann die π-Werte der Hypothesen N_1 und N_2 eingesetzt.
- Analog dazu werden als Ereignisse $C_1, \ldots, C_m$ die in dem anderen Vorgängerknoten von W (also in S) repräsentierten Hypothesen genommen: S_1, S_2 und S_3 (wodurch festgelegt ist, daß $m = 3$ ist).[10]

[10]Durch die Wahl von N_1 und N_2 für $B_1, \ldots, B_n$ bzw. von S_1, S_2 und S_3 für $C_1, \ldots, C_m$ ist gewährleistet, daß – wie oben gefordert – die $B_1, \ldots, B_n$ und die $C_1, \ldots, C_m$ unabhängig voneinander sind: Zwischen den Knoten N und S existiert im Beispielnetz kein Pfad, was in einem Bayesschen Netz gleichbedeutend damit ist, daß die in den jeweiligen Knoten repräsentierten Hypothesen als voneinander unabhängig angesehen werden.

- Dementsprechend werden als Werte $P(C_1), \ldots, P(C_m))$ die π-Werte der Hypothesen S_1, S_2 und S_3 eingesetzt.
- Als Ereignis A werden die Hypothesen W_1 bzw. W_2 genommen.
- Die Berechnung von $P(A)$ liefert dann die π-Werte von W_1 und von W_2.

Berücksichtigt man schließlich, daß die Matrixeinträge $M(W|N,S)_{i,j,k}$ definiert sind als $P(W_i|N_j,S_k)$, dann gestaltet sich die Berechnung der π-Werte des Knotens W wie folgt:

$$
\begin{aligned}
\pi(W_1) &= \sum_{j=1}^{2}\sum_{k=1}^{3} M(W|N,S)_{1,j,k}\cdot\pi(N_j)\cdot\pi(S_k)\\
&= \sum_{j=1}^{2}\sum_{k=1}^{3} P(W_1|N_j,S_k)\cdot\pi(N_j)\cdot\pi(S_k)\\
&= \quad P(W_1|N_1,S_1)\cdot\pi(N_1)\cdot\pi(S_1)\\
&\quad + P(W_1|N_1,S_2)\cdot\pi(N_1)\cdot\pi(S_2)\\
&\quad + P(W_1|N_1,S_3)\cdot\pi(N_1)\cdot\pi(S_3)\\
&\quad + P(W_1|N_2,S_1)\cdot\pi(N_2)\cdot\pi(S_1)\\
&\quad + P(W_1|N_2,S_2)\cdot\pi(N_2)\cdot\pi(S_2)\\
&\quad + P(W_1|N_2,S_3)\cdot\pi(N_2)\cdot\pi(S_3)\\
&= \quad P(W=ja|N=hoch,S=schwer)\\
&\qquad\qquad\qquad \cdot\pi(N=hoch)\cdot\pi(S=schwer)\\
&\quad + P(W=ja|N=hoch,S=mittel)\\
&\qquad\qquad\qquad \cdot\pi(N=hoch)\cdot\pi(S=mittel)\\
&\quad + P(W=ja|N=hoch,S=leicht)\\
&\qquad\qquad\qquad \cdot\pi(N=hoch)\cdot\pi(S=leicht)\\
&\quad + P(W=ja|N=niedrig,S=schwer)\\
&\qquad\qquad\qquad \cdot\pi(N=niedrig)\cdot\pi(S=schwer)\\
&\quad + P(W=ja|N=niedrig,S=mittel)\\
&\qquad\qquad\qquad \cdot\pi(N=niedrig)\cdot\pi(S=mittel)\\
&\quad + P(W=ja|N=niedrig,S=leicht)\\
&\qquad\qquad\qquad \cdot\pi(N=niedrig)\cdot\pi(S=leicht)\\
&= \quad 0.5\cdot0.8\cdot0.6+0.6\cdot0.8\cdot0.3+0.8\cdot0.8\cdot0.1\\
&\quad +0.3\cdot0.2\cdot0.6+0.4\cdot0.2\cdot0.3+0.5\cdot0.2\cdot0.1\\
&= 0.24+0.144+0.064+0.036+0.024+0.01\\
&= 0.518
\end{aligned}
$$

$$
\begin{aligned}
\pi(W_2) &= \sum_{j=1}^{2}\sum_{k=1}^{3} M(W|N,S)_{2,j,k}\cdot\pi(N_j)\cdot\pi(S_k)\\
&= 0.24+0.096+0.016+0.084+0.036+0.01\\
&= 0.482
\end{aligned}
$$

Der π-Vektor des Knoten W wird infolgedessen mit diesen Werten initialisiert:

$$\pi(W) = \begin{pmatrix} 0.518 \\ 0.482 \end{pmatrix}$$

Wie aber wird der π-Vektor eines inneren Knoten berechnet, der mehr als zwei Vorgängerknoten besitzt? Grundlage der Berechnung ist eine Verallgemeinerung der Formel von der totalen Wahrscheinlichkeit:

$$P(A) = \sum_{j_1=1}^{m_1} \cdots \sum_{j_n=1}^{m_n} P(A|B_{1,j_1},\ldots,B_{n,j_n}) \cdot P(B_{1,j_1}) \cdot \ldots \cdot P(B_{n,j_n})$$

$$\text{mit} \quad \begin{array}{c} B_{1,l} \neq B_{1,k} \text{ für } l \neq k \\ \vdots \\ B_{n,l} \neq B_{n,k} \text{ für } l \neq k \end{array}$$

Sei nun K ein innerer Knoten eines Bayesschen Netzes. K repräsentiere die Hypothesen $K_1,\ldots,K_h$ und besitze die Vorgängerknoten $V_1,\ldots,V_n$, die ihrerseits wiederum die Hypothesen $V_{1,1},\ldots,V_{1,m_1},V_{2,1},\ldots,V_{2,m_2},\ldots,V_{n,1},\ldots,V_{n,m_n}$ repräsentieren. Ferner sei dem Knoten K die Verbindungsmatrix $M(K|V_1,\ldots,V_n)$ zugeordnet. Durch Einsetzen in die verallgemeinerte Formel von der totalen Wahrscheinlichkeit ergibt sich dann die folgende Berechnungsvorschrift für die π-Werte $\pi(i)$ $(i=1,\ldots,h)$ von K:

$$\begin{aligned} \pi(K_i) &= \sum_{j_1=1}^{m_1} \cdots \sum_{j_n=1}^{m_n} M(K|V_1,\ldots,V_n)_{i,j_1,\ldots,j_n} \cdot \pi(V_{1,j_1}) \cdot \ldots \cdot \pi(V_{n,j_n}) \qquad (2) \\ &= \sum_{j_1=1}^{m_1} \cdots \sum_{j_n=1}^{m_n} P(K_i|V_{1,j_1},\ldots,V_{n,j_n}) \cdot \pi(V_{1,j_1}) \cdot \ldots \cdot \pi(V_{n,j_n}) \end{aligned}$$

Es bleibt noch, die π-Vektoren der Knoten G und M anzugeben. Da beide Knoten jeweils nur einen Vorgängerknoten besitzen, nämlich W, gestaltet sich bei G und M die Berechnung der π-Vektoren im Vergleich zum allgemeinen Fall relativ einfach:

$$\begin{aligned} \pi(G_1) &= \sum_{j_1=1}^{2} M(G|W)_{1,j_1} \cdot \pi(W_{j_1}) \\ &= \sum_{j_1=1}^{2} P(G_1|W_{j_1}) \cdot \pi(W_{j_1}) \\ &= \quad P(G_1|W_1) \cdot \pi(W_1) \\ &\quad + P(G_1|W_2) \cdot \pi(W_2) \\ &= \quad P(G = hektisch|W = ja,) \quad \cdot \pi(W = ja) \\ &\quad + P(G = hektisch|W = nein) \cdot \pi(W = nein) \\ &= 0.3 \cdot 0.518 + 0.9 \cdot 0.482 \\ &= 0.1554 + 0.4338 \\ &= 0.5892 \end{aligned}$$

$$\begin{aligned} \pi(G_2) &= \sum_{j_1=1}^{2} M(G|W)_{2,j_1} \cdot \pi(W_{j_1}) \\ &= 0.3626 + 0.0482 \\ &= 0.4108 \end{aligned}$$

$$
\begin{aligned}
\pi(M_1) &= \sum_{j_1=1}^{2} M(M|W)_{1,j_1} \cdot \pi(W_{j_1}) \\
&= \sum_{j_1=1}^{2} P(M_1|W_{j_1}) \cdot \pi(W_{j_1}) \\
&= \quad P(M_1|W_1) \cdot \pi(W_1) \\
&\quad + P(M_1|W_2) \cdot \pi(W_2) \\
&= \quad P(M = entsetzt|W = ja,) \quad \cdot \pi(W = ja) \\
&\quad + P(M = entsetzt|W = nein) \cdot \pi(W = nein) \\
&= 0.2 \cdot 0.518 + 0.6 \cdot 0.482 \\
&= 0.1036 + 0.2892 \\
&= 0.3928
\end{aligned}
$$

$$
\begin{aligned}
\pi(M_2) &= \sum_{j_1=1}^{2} M(M|W)_{2,j_1} \cdot \pi(W_{j_1}) \\
&= 0.4144 + 0.1928 \\
&= 0.6072
\end{aligned}
$$

Mit diesen π-Werten ergeben sich gemäß Formel (1) die folgenden BEL-Vektoren für die Knoten W, G und M:

$$
\begin{aligned}
BEL(W) &= \alpha \cdot \lambda(W) \times \pi(W) \\
&= \alpha \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix} \times \begin{pmatrix} 0.518 \\ 0.482 \end{pmatrix} = \alpha \cdot \begin{pmatrix} 0.518 \\ 0.482 \end{pmatrix} \text{mit} \underset{=}{\alpha=1} \begin{pmatrix} 0.518 \\ 0.482 \end{pmatrix}
\end{aligned}
$$

$$
\begin{aligned}
BEL(G) &= \alpha \cdot \lambda(G) \times \pi(G) \\
&= \alpha \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix} \times \begin{pmatrix} 0.5892 \\ 0.4108 \end{pmatrix} = \alpha \cdot \begin{pmatrix} 0.5892 \\ 0.4108 \end{pmatrix} \text{mit} \underset{=}{\alpha=1} \begin{pmatrix} 0.5892 \\ 0.4108 \end{pmatrix}
\end{aligned}
$$

$$
\begin{aligned}
BEL(M) &= \alpha \cdot \lambda(M) \times \pi(M) \\
&= \alpha \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix} \times \begin{pmatrix} 0.3928 \\ 0.6072 \end{pmatrix} = \alpha \cdot \begin{pmatrix} 0.3928 \\ 0.6072 \end{pmatrix} \text{mit} \underset{=}{\alpha=1} \begin{pmatrix} 0.3928 \\ 0.6072 \end{pmatrix}
\end{aligned}
$$

Die λ-, π- und BEL-Vektoren, die sich aus den für die Wurzelknoten vorgegebenen A-priori-Wahrscheinlichkeiten bzw. den für die inneren Knoten vorgegebenen bedingten Wahrscheinlichkeiten ergeben, sind in Abbildung 3 auf der nachfolgenden Seite 18 noch einmal im Überblick dargestellt. Damit ist die Initialisierungsphase abgeschlossen und ein stabiler Zustand innerhalb des Beispielnetzes hergestellt. Zu diesem Zeitpunkt kann das Beispielnetz bereits als Wissensbasis dienen, die Auskunft gibt über die Einschätzung der geschilderten Prüfungssituation. Welche Veränderungen sich innerhalb des Beispielnetzes ergeben, wenn zu den bisherigen Informationen neue hinzukommen, wird im nachfolgenden Abschnitt 4 erläutert.

Knoten N: **Wissensniveau**

N_i	$\lambda(N_i)$	$\pi(N_i)$	$BEL(N_i)$
hoch	1.0	0.8	0.8
niedrig	1.0	0.2	0.2

Knoten S: **Schwierigkeit**

S_i	$\lambda(S_i)$	$\pi(S_i)$	$BEL(S_i)$
schwer	1.0	0.6	0.6
mittel	1.0	0.3	0.3
leicht	1.0	0.1	0.1

Knoten W: **Wissen**

W_i	$\lambda(W_i)$	$\pi(W_i)$	$BEL(W_i)$
ja	1.0	0.518	0.518
nein	1.0	0.482	0.482

Knoten G: **Gestik**

G_i	$\lambda(G_i)$	$\pi(G_i)$	$BEL(G_i)$
hektisch	1.0	0.5892	0.5892
ruhig	1.0	0.4108	0.4108

Knoten M: **Mimik**

M_i	$\lambda(M_i)$	$\pi(M_i)$	$BEL(M_i)$
entsetzt	1.0	0.3928	0.3928
gelassen	1.0	0.6072	0.6072

Abbildung 3: Das Beispielnetz nach der Initialisierung der Wahrscheinlichkeitswerte aller Knoten.

4 Die Propagierung von Wahrscheinlichkeiten in Bayesschen Netzen

Im vorangegangenen Abschnitt 3 wurde die Initialisierung des Beispielnetzes zu der Prüfungssituation beschrieben. Nach der Initialisierungsphase befindet sich das Bayessche Netz in einem stabilen Zustand. Da es sich bei Bayesschen Netzen jedoch nicht um statische, sondern um dynamische Netze handelt, ist es möglich, dem Bayesschen Netz neue Informationen hinzuzufügen: Es lassen sich sowohl die in einem Knoten enthaltenen Wahrscheinlichkeitswerte verändern als auch die Struktur des Bayesschen Netzes selbst (d.h. es können dem Bayesschen Netz Knoten und Kanten hinzugefügt werden oder auch aus dem Bayesschen Netz herausgenommen werden).

In diesem Abschnitt wird an Hand des bereits eingeführten Beispielnetzes der Fall erörtert, daß die in den Knoten enthaltenen Wahrscheinlichkeitswerte auf Grund neuer Informationen modifiziert werden, wobei die Netzstruktur erhalten bleibt. Eine Beschreibung von dynamischen Strukturveränderungen an Bayesschen Netzen ist beispielsweise in [Goldman & Charniak 91] zu finden.

Die Prüfungssituation, die durch das Beispielnetz repräsentiert wird, wurde bisher aus der Perspektive beschrieben, wie sie sich vor dem Stellen der Frage darstellt: Es wurde eingeschätzt, wie hoch das Wissensniveau des Prüflings und der Schwierigkeitsgrad der Frage ist, und daraus wurden Annahmen darüber abgeleitet, wie wahrscheinlich es ist, daß der Prüfling die Antwort auf die Frage kennt, und wie die Gestik bzw. die Mimik des Prüflings sein wird. Im folgenden wird nun die gleiche Prüfungssituation zu einem späteren Zeitpunkt betrachtet: Es wird vorausgesetzt, daß die Frage bereits gestellt wurde, daß der Prüfling noch keine Antwort gegeben hat und daß beim Prüfling ein recht entsetzter Gesichtsausdruck zu beobachten ist.

Für jeden der fünf im Beispielnetz enthaltenen Knoten wird nachfolgend dargelegt, wie sich die in dem jeweiligen Knoten enthaltenen Wahrscheinlichkeitswerte durch das Hinzufügen dieser neuen Informationen verändern, d.h. wie die Werte auf Grund der Beobachtung aktualisiert werden.

4.1 Aktualisierung des Knotens M

Wie bereits in Abschnitt 3.3 erwähnt, gibt der λ-Wert einer Hypothese die gegenwärtige Stärke der diagnostischen Unterstützung dieser Hypothese an. Bei dem Umstand, daß der Prüfling einen recht entsetzten Gesichtsausdruck aufweist, handelt es sich um einen direkt beobachtbaren (man kann auch sagen: diagnostizierbaren) Sachverhalt. Infolgedessen führt die neue Information über die Mimik des Prüflings zu einer Modifikation der λ-Werte der Hypothesen M_1 und M_2.

Wie sieht die Modifikation am λ-Vektor des Knotens M konkret aus? In dem hier besprochenen Beispiel wird die Information *recht entsetzter Gesichtsausdruck* durch die folgende Wahrscheinlichkeitsverteilung über den Hypothesen M_1 und M_2 repräsentiert:

$$\begin{aligned} P(M_1) &= P(M = entsetzt) &= 0.9 \\ P(M_2) &= P(M = gelassen) &= 0.1 \end{aligned}$$

Diese beiden Wahrscheinlichkeitswerte werden als neue λ-Werte für die im Knoten M repräsentierten Hypothesen M_1 und M_2 genommen. Somit ergibt sich der folgende Vektor als neuer λ-Vektor des Knotens M:

$$\lambda(M) = \begin{pmatrix} 0.9 \\ 0.1 \end{pmatrix}$$

Auf Grund der Modifikation des λ-Vektors ergeben sich auch neue Werte für den BEL-Vektor des Knotens M:

$$\begin{aligned} BEL(M) &= \alpha \cdot \lambda(M) \times \pi(M) \\ &= \alpha \cdot \begin{pmatrix} 0.9 \\ 0.1 \end{pmatrix} \times \begin{pmatrix} 0.3928 \\ 0.6072 \end{pmatrix} \\ &= \alpha \cdot \begin{pmatrix} 0.35352 \\ 0.06072 \end{pmatrix} \overset{\text{mit } \alpha = \frac{1}{0.41424}}{=} \begin{pmatrix} 0.8534 \\ 0.1466 \end{pmatrix} \end{aligned}$$

Ein Vergleich des bisherigen BEL-Vektors von M mit dem modifizierten macht die starke Auswirkung der neu hinzugekommenen Information deutlich:

bisheriger Wert: $$BEL(M) = \begin{pmatrix} 0.3928 \\ 0.6072 \end{pmatrix}$$

neuer Wert: $$BEL(M) = \begin{pmatrix} 0.8534 \\ 0.1466 \end{pmatrix}$$

Die neue Information hat jedoch nicht nur Auswirkungen auf den Knoten M, sondern auch auf andere Knoten des Beispielnetzes: Wenn man beim Prüfling einen entsetzten Gesichtsausdruck festgestellt hat, ist es beispielsweise nicht mehr länger plausibel anzunehmen, daß die Wahrscheinlichkeiten für die beiden Hypothesen, daß der Prüfling die Antwort auf die Frage kennt bzw. nicht kennt, ungefähr gleich groß sind. Folglich sollte sich die neue Information auf den BEL-Vektor des Knotens W so auswirken, daß die Wahrscheinlichkeit für Hypothese W_1 (der Prüfling kennt die Antwort) sinkt, während sie für Hypothese W_2 (der Prüfling kennt die Antwort nicht) steigt. Diese Änderung am BEL-Vektor des Knotens W sollte ihrerseits wiederum Änderungen an den Nachbarknoten von W bewirken.

Da sich also die Auswirkungen, die sich auf Grund neu hinzugekommener Informationen ergeben, durch das Netz hindurch fortpflanzen, werden innerhalb Bayesscher Netze Informationen über geänderte Wahrscheinlichkeitswerte durch das Netz **propagiert**. Zu diesem Zweck wird ein Knoten **aktiviert**, wenn er neue Informationen erhält. Nachdem in einem aktivierten Knoten die entsprechenden Wahrscheinlichkeitswerte modifiziert wurden, schickt dieser Knoten Informationen über die vorgenommenen Änderungen an seine Nachbarknoten (sowohl Vorgängerknoten als auch Nachfolgerknoten), die auf Grund dessen aktiviert werden, und wird schließlich wieder deaktiviert.

In dem hier betrachteten Beispielszenario wird infolge der neuen Information über die Mimik des Prüflings zuerst der Knoten M aktiviert. Innerhalb von M werden dann – wie gerade beschrieben – die neuen Wahrscheinlichkeitswerte bestimmt. Da der Knoten M lediglich einen Nachbarknoten besitzt, nämlich seinen Vorgängerknoten W, wird M schließlich noch W über diese Änderungen informieren. Hierzu wird jedoch nicht der λ- oder der BEL-Vektor von M an W weitergeleitet, sondern es wird eine andere Information übermittelt, die im folgenden **λ-Nachricht** genannt wird. Nachfolgend wird erläutert, wie sich die λ-Nachrichten berechnen, die innerhalb von Bayesschen Netzen von inneren Knoten an die jeweiligen Vorgängerknoten geschickt werden.

Sei K wiederum ein innerer Knoten eines Bayesschen Netzes. K repräsentiere die Hypothesen $K_1, \ldots, K_h$ und besitze die Vorgängerknoten $V_1, \ldots, V_n$, die ihrerseits wiederum die Hypothesen $V_{1,1}, \ldots, V_{1,m_1}, V_{2,1}, \ldots, V_{2,m_2}, \ldots, V_{n,1}, \ldots, V_{n,m_n}$ repräsentieren. Ferner sei dem Knoten K die Verbindungsmatrix $M(K|V_1, \ldots, V_n)$ zugeordnet. Bei der λ-Nachricht, die der Knoten K an einen Vorgängerknoten V_i ($i \in \{1, \ldots, n\}$) schickt, handelt es sich um einen Vektor $\lambda_K(V_i)$ mit m_i Elementen:

$$\lambda_K(V_i) = \begin{pmatrix} \lambda_K(V_{i,1}) \\ \lambda_K(V_{i,2}) \\ \vdots \\ \lambda_K(V_{i,m_i}) \end{pmatrix}$$

Unter Verwendung der Verbindungsmatrix $M(K|V_1, \ldots, V_n)$ werden die einzelnen Vektorelemente $\lambda_K(V_{i,l})$ ($l \in \{1, \ldots, m_i\}$) dabei wie folgt berechnet[11] (dabei werden durch $\pi_K(V_{q,j_q})$ Elemente der Vektoren bezeichnet, die als π-Nachrichten verschickt werden und weiter unten noch genau erläutert werden):

$$\begin{aligned} \lambda_K(V_{i,l}) &= \\ &\sum_{g=1}^{h} \lambda(K_g) \cdot \Bigg(\sum_{j_1=1}^{m_1} \cdots \sum_{j_{i-1}=1}^{m_{i-1}} \sum_{j_{i+1}=1}^{m_{i+1}} \cdots \sum_{j_n=1}^{m_n} \quad M(K|V_1, \ldots, V_n)_{g,j_1,\ldots,j_{i-1},l,j_{i+1},\ldots,j_n} \\ &\qquad \cdot \prod_{q=1, q \neq i}^{n} \pi_K(V_{q,j_q}) \Bigg) \end{aligned} \tag{3}$$

Da der Knoten M des hier besprochenen Beispielnetzes nur einen einzigen Vorgängerknoten besitzt (d.h. $n = 1$ und somit auch $i = 1$) und sowohl im Knoten M als auch im Knoten W zwei Hypothesen repräsentiert sind (d.h. $h = 2$ bzw. $m_i = 2$), erhält man durch Einsetzen in obige Formel die folgende Berechnungsvorschrift für die beiden Vektorelemente $\lambda_M(W_l)$ ($l \in \{1, 2\}$), die als λ-Nachricht von M an W geschickt werden:

$$\lambda_M(W_l) = \sum_{g=1}^{2} \lambda(M_g) \cdot M(M|W)_{g,l}$$

Somit erhält man durch Einsetzen der konkreten Zahlenwerte:

$$\begin{aligned} \lambda_M(W) &= \begin{pmatrix} \lambda_M(W_1) \\ \lambda_M(W_2) \end{pmatrix} \\ &= \begin{pmatrix} \sum_{g=1}^{2} \lambda(M_g) \cdot M(M|W)_{g,1} \\ \sum_{g=1}^{2} \lambda(M_g) \cdot M(M|W)_{g,2} \end{pmatrix} \end{aligned}$$

[11]Zur Begründung dieser Berechnungsvorschrift siehe [Pearl 91].

$$
\begin{aligned}
&= \begin{pmatrix} \lambda(M_1) \cdot M(M|W)_{1,1} + \lambda(M_2) \cdot M(M|W)_{2,1} \\ \lambda(M_1) \cdot M(M|W)_{1,2} + \lambda(M_2) \cdot M(M|W)_{2,2} \end{pmatrix} \\
&= \begin{pmatrix} \lambda(M = entsetzt) \cdot P(M = entsetzt|W = ja) + \\ \lambda(M = gelassen) \cdot P(M = gelassen|W = ja) \\ \\ \lambda(M = entsetzt) \cdot P(M = entsetzt|W = nein) + \\ \lambda(M = gelassen) \cdot P(M = gelassen|W = nein) \end{pmatrix} \\
&= \begin{pmatrix} 0.9 \cdot 0.2 + 0.1 \cdot 0.8 \\ 0.9 \cdot 0.6 + 0.1 \cdot 0.4 \end{pmatrix} \\
&= \begin{pmatrix} 0.18 + 0.08 \\ 0.54 + 0.04 \end{pmatrix} = \begin{pmatrix} 0.26 \\ 0.58 \end{pmatrix}
\end{aligned}
$$

Nachdem diese λ-Nachricht vom Knoten M berechnet wurde, kann sie an den Knoten W geschickt werden und Knoten M wird wieder deaktiviert. Der Zustand des Beispielnetzes zu diesem Zeitpunkt ist in Abbildung 4 auf der nachfolgenden Seite 23 wiedergegeben.

Knoten N: **Wissensniveau**

N_i	$\lambda(N_i)$	$\pi(N_i)$	$BEL(N_i)$
hoch	1.0	0.8	0.8
niedrig	1.0	0.2	0.2

Knoten S: **Schwierigkeit**

S_i	$\lambda(S_i)$	$\pi(S_i)$	$BEL(S_i)$
schwer	1.0	0.6	0.6
mittel	1.0	0.3	0.3
leicht	1.0	0.1	0.1

Knoten W: **Wissen**

W_i	$\lambda(W_i)$	$\pi(W_i)$	$BEL(W_i)$
ja	1.0	0.518	0.518
nein	1.0	0.482	0.482

$$\uparrow \lambda_M(W) = \begin{pmatrix} 0.26 \\ 0.58 \end{pmatrix}$$

Knoten G: **Gestik**

G_i	$\lambda(G_i)$	$\pi(G_i)$	$BEL(G_i)$
hektisch	1.0	0.5892	0.5892
ruhig	1.0	0.4108	0.4108

Knoten M: **Mimik**

M_i	$\lambda(M_i)$	$\pi(M_i)$	$BEL(M_i)$
entsetzt	0.9	0.3928	0.8534
gelassen	0.1	0.6072	0.1466

Abbildung 4: Das Beispielnetz nach Aktualisierung des Knotens M: Neben der Neuberechnung des λ- und des BEL-Vektors, die innerhalb des Knotens M durchgeführt wurde, wurde vom Knoten M eine λ-Nachricht an den Knoten W geschickt.

4.2 Aktualisierung des Knotens W

Nach Erhalt der λ-Nachricht von Knoten M wird der Knoten W aktiviert. Innerhalb des Knotens W werden zuerst die Wahrscheinlichkeitswerte neu berechnet, die auf Grund der erhaltenen λ-Nachricht geändert werden müssen. Es handelt sich hierbei zunächst einmal um den λ-Vektor von W. Nach Abschluß der Initialisierung, bei der alle λ-Werte auf 1 gesetzt wurden, berechnet sich der λ-Vektor eines beliebigen Knotens K mit den Nachfolgerknoten $F_1, \ldots, F_n$ auf die folgende Weise aus den von den Nachfolgerknoten erhaltenen λ-Nachrichten:

$$\lambda(K) = \lambda_{F_1}(K) \times \cdots \times \lambda_{F_n}(K) \tag{4}$$

Durch Einsetzen in diese Formel ergibt sich als neuer λ-Vektor des Knotens W:[12]

$$\begin{aligned} \lambda(W) &= \lambda_G(W) \times \lambda_M(W) \\ &= \begin{pmatrix} 1 \\ 1 \end{pmatrix} \times \begin{pmatrix} 0.26 \\ 0.58 \end{pmatrix} = \begin{pmatrix} 0.26 \\ 0.58 \end{pmatrix} \end{aligned}$$

Auf Grund der geänderten λ-Werte von W muß auch der BEL-Vektor von W neu berechnet werden:

$$\begin{aligned} BEL(W) &= \alpha \cdot \lambda(W) \times \pi(W) \\ &= \alpha \cdot \begin{pmatrix} 0.26 \\ 0.58 \end{pmatrix} \times \begin{pmatrix} 0.518 \\ 0.482 \end{pmatrix} \\ &= \alpha \cdot \begin{pmatrix} 0.13468 \\ 0.27956 \end{pmatrix} \overset{\text{mit } \alpha=\frac{1}{0.41424}}{=} \begin{pmatrix} 0.3251 \\ 0.6749 \end{pmatrix} \end{aligned}$$

Vergleicht man den bisherigen BEL-Vektor des Knotens W mit dem aktualisierten,

$$\text{bisheriger Wert:} \qquad BEL(W) = \begin{pmatrix} 0.518 \\ 0.482 \end{pmatrix}$$

$$\text{neuer Wert:} \qquad BEL(W) = \begin{pmatrix} 0.3251 \\ 0.6749 \end{pmatrix}$$

so ist das plausible Ergebnis festzustellen, daß sich die Wahrscheinlichkeit für die Hypothese W_1 (der Prüfling kennt die Antwort) deutlich verringert hat, während sie für die Hypothese W_2 (der Prüfling kennt die Antwort nicht) deutlich gestiegen ist.

Nach der Aktualisierung der im Knoten enthaltenen Wahrscheinlichkeitswerte werden alle Nachbarknoten über die Änderungen informiert (ausgenommen hiervon ist lediglich der Knoten, durch den die Aktivierung bewirkt wurde). Im Falle des Knotens W werden also an alle Nachbarknoten außer M Nachrichten geschickt.

Zunächst zu den Vorgängerknoten von W. Da der Knoten W zwei Vorgängerknoten besitzt, nämlich N und S, wird an jeden dieser beiden Knoten eine λ-Nachricht

[12] Dabei läßt sich $\lambda_G(W)$, die λ-Nachricht von G an W, die bei dieser Berechnung benötigt wird, wiederum aus Formel (3) herleiten. Alle Werte dieser λ-Nachricht sind gleich 1, weil einerseits sämtliche Elemente des λ-Vektors von G gleich 1 sind und weil andererseits die Verbindungsmatrix $M(G|W)$ normiert ist.

geschickt. Die Vektoren, aus denen die λ-Nachrichten bestehen, berechnen sich wiederum nach Formel (3) (mit $K = W$, $h = 2$, $n = 2$, $V_1 = N$, $m_1 = 2$, $V_2 = S$, $m_2 = 3$):[13]

$$
\begin{aligned}
\lambda_W(N) &= \begin{pmatrix} \lambda_W(N_1) \\ \lambda_W(N_2) \end{pmatrix} \\
&= \begin{pmatrix} \sum\limits_{g=1}^{2} \lambda(W_g) \cdot \left(\sum\limits_{j_2=1}^{3} M(W|N,S)_{g,1,j_2} \cdot \prod\limits_{q=1,q\neq 1}^{2} \pi_W(V_{q,j_q}) \right) \\ \sum\limits_{g=1}^{2} \lambda(W_g) \cdot \left(\sum\limits_{j_2=1}^{3} M(W|N,S)_{g,2,j_2} \cdot \prod\limits_{q=1,q\neq 1}^{2} \pi_W(V_{q,j_q}) \right) \end{pmatrix} \\
&= \begin{pmatrix} \sum\limits_{g=1}^{2} \lambda(W_g) \cdot \left(\sum\limits_{j_2=1}^{3} M(W|N,S)_{g,1,j_2} \cdot \pi_W(S_{j_2}) \right) \\ \sum\limits_{g=1}^{2} \lambda(W_g) \cdot \left(\sum\limits_{j_2=1}^{3} M(W|N,S)_{g,2,j_2} \cdot \pi_W(S_{j_2}) \right) \end{pmatrix} \\
&= \begin{pmatrix} & \lambda(W_1) \cdot M(W|N,S)_{1,1,1} \cdot \pi_W(S_1) \\ + & \lambda(W_1) \cdot M(W|N,S)_{1,1,2} \cdot \pi_W(S_2) \\ + & \lambda(W_1) \cdot M(W|N,S)_{1,1,3} \cdot \pi_W(S_3) \\ + & \lambda(W_2) \cdot M(W|N,S)_{2,1,1} \cdot \pi_W(S_1) \\ + & \lambda(W_2) \cdot M(W|N,S)_{2,1,2} \cdot \pi_W(S_2) \\ + & \lambda(W_2) \cdot M(W|N,S)_{2,1,3} \cdot \pi_W(S_3) \\ & \\ & \lambda(W_1) \cdot M(W|N,S)_{1,2,1} \cdot \pi_W(S_1) \\ + & \lambda(W_1) \cdot M(W|N,S)_{1,2,2} \cdot \pi_W(S_2) \\ + & \lambda(W_1) \cdot M(W|N,S)_{1,2,3} \cdot \pi_W(S_3) \\ + & \lambda(W_2) \cdot M(W|N,S)_{2,2,1} \cdot \pi_W(S_1) \\ + & \lambda(W_2) \cdot M(W|N,S)_{2,2,2} \cdot \pi_W(S_2) \\ + & \lambda(W_2) \cdot M(W|N,S)_{2,2,3} \cdot \pi_W(S_3) \end{pmatrix} \\
&= \begin{pmatrix} & 0.26 \cdot 0.5 \cdot 0.6 + 0.26 \cdot 0.6 \cdot 0.3 + 0.26 \cdot 0.8 \cdot 0.1 \\ + & 0.58 \cdot 0.5 \cdot 0.6 + 0.58 \cdot 0.4 \cdot 0.3 + 0.58 \cdot 0.2 \cdot 0.1 \\ & \\ & 0.26 \cdot 0.3 \cdot 0.6 + 0.26 \cdot 0.4 \cdot 0.3 + 0.26 \cdot 0.5 \cdot 0.1 \\ + & 0.58 \cdot 0.7 \cdot 0.6 + 0.58 \cdot 0.6 \cdot 0.3 + 0.58 \cdot 0.5 \cdot 0.1 \end{pmatrix} \\
&= \begin{pmatrix} 0.078 + 0.0468 + 0.0208 + 0.174 + 0.0696 + 0.0116 \\ 0.0468 + 0.0312 + 0.013 + 0.2436 + 0.1044 + 0.029 \end{pmatrix} \\
&= \begin{pmatrix} 0.4008 \\ 0.468 \end{pmatrix}
\end{aligned}
$$

Analog dazu wird die λ-Nachricht berechnet, die der Knoten W an den Knoten S schickt:

$$
\lambda_W(S) = \begin{pmatrix} \lambda_W(S_1) \\ \lambda_W(S_2) \\ \lambda_W(S_3) \end{pmatrix}
$$

[13]Wie sich auf Grund der weiter unten noch vorgestellten Berechnungsvorschrift für π-Nachrichten ergibt, ist dabei infolge der Tatsache, daß N und S Wurzelknoten sind, $\pi_W(N) = \pi(N)$ sowie $\pi_W(S) = \pi(S)$.

$$
= \begin{pmatrix}
\sum_{g=1}^{2} \lambda(W_g) \cdot \left(\sum_{j_1=1}^{2} M(W|N,S)_{g,j_1,1} \cdot \prod_{q=1,q\neq 2}^{2} \pi_W(V_{q,j_q}) \right) \\
\sum_{g=1}^{2} \lambda(W_g) \cdot \left(\sum_{j_1=1}^{2} M(W|N,S)_{g,j_1,2} \cdot \prod_{q=1,q\neq 2}^{2} \pi_W(V_{q,j_q}) \right) \\
\sum_{g=1}^{2} \lambda(W_g) \cdot \left(\sum_{j_1=1}^{2} M(W|N,S)_{g,j_1,3} \cdot \prod_{q=1,q\neq 2}^{2} \pi_W(V_{q,j_q}) \right)
\end{pmatrix}
$$

$$
= \begin{pmatrix}
\sum_{g=1}^{2} \lambda(W_g) \cdot \left(\sum_{j_1=1}^{2} M(W|N,S)_{g,j_1,1} \cdot \pi_W(N_{j_1}) \right) \\
\sum_{g=1}^{2} \lambda(W_g) \cdot \left(\sum_{j_1=1}^{2} M(W|N,S)_{g,j_1,2} \cdot \pi_W(N_{j_1}) \right) \\
\sum_{g=1}^{2} \lambda(W_g) \cdot \left(\sum_{j_1=1}^{2} M(W|N,S)_{g,j_1,3} \cdot \pi_W(N_{j_1}) \right)
\end{pmatrix}
$$

$$
= \begin{pmatrix}
\begin{array}{cl}
 & \lambda(W_1) \cdot M(W|N,S)_{1,1,1} \cdot \pi_W(N_1) \\
+ & \lambda(W_1) \cdot M(W|N,S)_{1,2,1} \cdot \pi_W(N_2) \\
+ & \lambda(W_2) \cdot M(W|N,S)_{2,1,1} \cdot \pi_W(N_1) \\
+ & \lambda(W_2) \cdot M(W|N,S)_{2,2,1} \cdot \pi_W(N_2) \\
 & \\
 & \lambda(W_1) \cdot M(W|N,S)_{1,1,2} \cdot \pi_W(N_1) \\
+ & \lambda(W_1) \cdot M(W|N,S)_{1,2,2} \cdot \pi_W(N_2) \\
+ & \lambda(W_2) \cdot M(W|N,S)_{2,1,2} \cdot \pi_W(N_1) \\
+ & \lambda(W_2) \cdot M(W|N,S)_{2,2,2} \cdot \pi_W(N_2) \\
 & \\
 & \lambda(W_1) \cdot M(W|N,S)_{1,1,3} \cdot \pi_W(N_1) \\
+ & \lambda(W_1) \cdot M(W|N,S)_{1,2,3} \cdot \pi_W(N_2) \\
+ & \lambda(W_2) \cdot M(W|N,S)_{2,1,3} \cdot \pi_W(N_1) \\
+ & \lambda(W_2) \cdot M(W|N,S)_{2,2,3} \cdot \pi_W(N_2)
\end{array}
\end{pmatrix}
$$

$$
= \begin{pmatrix}
\begin{array}{cl}
 & 0.26 \cdot 0.5 \cdot 0.8 + 0.26 \cdot 0.3 \cdot 0.2 \\
+ & 0.58 \cdot 0.5 \cdot 0.8 + 0.58 \cdot 0.7 \cdot 0.2 \\
 & \\
 & 0.26 \cdot 0.6 \cdot 0.8 + 0.26 \cdot 0.4 \cdot 0.2 \\
+ & 0.58 \cdot 0.4 \cdot 0.8 + 0.58 \cdot 0.6 \cdot 0.2 \\
 & \\
 & 0.26 \cdot 0.8 \cdot 0.8 + 0.26 \cdot 0.5 \cdot 0.2 \\
+ & 0.58 \cdot 0.2 \cdot 0.8 + 0.58 \cdot 0.5 \cdot 0.2
\end{array}
\end{pmatrix}
$$

$$
= \begin{pmatrix}
0.104 + 0.0156 + 0.232 + 0.0812 \\
0.1248 + 0.0208 + 0.1856 + 0.0696 \\
0.1664 + 0.026 + 0.0928 + 0.058
\end{pmatrix}
$$

$$
= \begin{pmatrix}
0.4328 \\
0.4008 \\
0.3432
\end{pmatrix}
$$

Nun müssen auch noch die Nachfolgerknoten des aktivierten Knotens über die Modifikation der Wahrscheinlichkeitswerte innerhalb des Knotens informiert werden. Da der Knoten W zwei Nachfolgerknoten besitzt, einer davon aber die Aktivierung von W bewirkte (nämlich M), braucht nur eine Nachricht an den Knoten G geschickt zu werden.

In Anbetracht der Tatsache, daß Nachfolgerknoten nicht die gleichen Informationen übermittelt werden wie Vorgängerknoten, erfolgt an dieser Stelle zunächst eine allgemeine Charakterisierung der Nachrichten, die an Nachfolgerknoten geschickt werden und die im folgenden als **π-Nachrichten** bezeichnet werden.

Sei K ein aktivierter Knoten, der einen seiner Nachfolgerknoten $F_1, \ldots, F_n$ über Änderungen seiner Wahrscheinlichkeitswerte informieren will. Sei F_i ($i \in \{1, \ldots, n\}$) einer dieser Nachfolgerknoten. Der Inhalt der π-Nachricht $\pi_{F_i}(K)$, die K in einem solchen Fall an F_i schickt, ist festgelegt durch die folgende Berechnungsvorschrift (zu deren Erläuterung wiederum auf [Pearl 91] verwiesen sei):

$$\pi_{F_i}(K) = \pi(K) \times \left(\prod_{j=1, j \neq i}^{n} \lambda_{F_j}(K) \right)$$

Im Falle des Knotens W ist $K = W$, $n = 2$, $F_1 = G$, $F_2 = M$ und $i = 1$, und es ergibt sich durch Einsetzen:

$$\begin{aligned}
\pi_G(W) &= \pi(W) \times \left(\prod_{j=1, j \neq 1}^{2} \lambda_{F_j}(W) \right) \\
&= \pi(W) \times \lambda_M(W) \\
&= \begin{pmatrix} 0.518 \\ 0.482 \end{pmatrix} \times \begin{pmatrix} 0.26 \\ 0.58 \end{pmatrix} \\
&= \begin{pmatrix} 0.13468 \\ 0.27956 \end{pmatrix}
\end{aligned}$$

Damit sind alle Nachrichten bestimmt, die der Knoten W an seine Nachbarknoten schickt, und W wird deaktiviert. Der Zustand, in dem sich das Beispielnetz in dieser Situation befindet, ist in Abbildung 5 auf der nachfolgenden Seite 28 wiedergegeben.

Knoten N: **Wissensniveau**

N_i	$\lambda(N_i)$	$\pi(N_i)$	$BEL(N_i)$
hoch	1.0	0.8	0.8
niedrig	1.0	0.2	0.2

Knoten S: **Schwierigkeit**

S_i	$\lambda(S_i)$	$\pi(S_i)$	$BEL(S_i)$
schwer	1.0	0.6	0.6
mittel	1.0	0.3	0.3
leicht	1.0	0.1	0.1

$$\uparrow \lambda_W(N) = \begin{pmatrix} 0.4008 \\ 0.468 \end{pmatrix}$$

$$\uparrow \lambda_W(S) = \begin{pmatrix} 0.4328 \\ 0.4008 \\ 0.3432 \end{pmatrix}$$

Knoten W: **Wissen**

W_i	$\lambda(W_i)$	$\pi(W_i)$	$BEL(W_i)$
ja	0.26	0.518	0.3251
nein	0.58	0.482	0.6749

$$\downarrow \pi_G(W) = \begin{pmatrix} 0.13468 \\ 0.27956 \end{pmatrix}$$

Knoten G: **Gestik**

G_i	$\lambda(G_i)$	$\pi(G_i)$	$BEL(G_i)$
hektisch	1.0	0.5892	0.5892
ruhig	1.0	0.4108	0.4108

Knoten M: **Mimik**

M_i	$\lambda(M_i)$	$\pi(M_i)$	$BEL(M_i)$
entsetzt	0.9	0.3928	0.8534
gelassen	0.1	0.6072	0.1466

Abbildung 5: Das Beispielnetz nach Aktualisierung des Knotens W: Neben der Neuberechnung des λ- und des BEL-Vektors, die innerhalb des Knotens W durchgeführt wurde, wurden vom Knoten W λ-Nachrichten an die Knoten N und S sowie eine π-Nachricht an den Knoten G geschickt.

4.3 Aktualisierung des Knotens G

Da der Knoten G vom Knoten W eine π-Nachricht erhält, müssen auch die in G enthaltenen Wahrscheinlichkeitswerte aktualisiert werden. Zu diesem Zweck wird G aktiviert, und es wird zunächst der π-Vektor von G neu bestimmt.

Die Neuberechnung des π-Vektors eines inneren Knotens K mit Vorgängerknoten $V_1, \ldots, V_n$ erfolgt (genauso wie die ursprüngliche Berechnung der π-Vektoren innerer Knoten in der Initialisierungsphase) unter Verwendung einer Verallgemeinerung der Formel von der totalen Wahrscheinlichkeit. Bei der Neuberechnung werden jedoch an Stelle der π-Vektoren der Vorgängerknoten die von diesen erhaltenen π-Nachrichten genommen (vgl. Formel (2)):

$$\pi(K_i) = \sum_{j_1=1}^{m_1} \cdots \sum_{j_n=1}^{m_n} M(K|V_1, \ldots, V_n)_{i,j_1,\ldots,j_n} \cdot \pi_K(V_{1,j_1}) \cdot \ldots \cdot \pi_K(V_{n,j_n})$$

Da im Falle des Knotens G nur ein einziger Vorgängerknoten existiert (nämlich W), vereinfacht sich die Berechnungsvorschrift mit $K = G$, $n = 1$, $V_1 = W$ und $m_1 = 2$ zu:

$$\begin{aligned}
\pi(G_1) &= \sum_{j_1=1}^{2} M(G|W)_{1,j_1} \cdot \pi_G(W_{j_1}) \\
&= \sum_{j_1=1}^{2} P(G_1|W_{j_1}) \cdot \pi_G(W_{j_1}) \\
&= \quad P(G_1|W_1) \cdot \pi_G(W_1) \\
&\quad +P(G_1|W_2) \cdot \pi_G(W_2) \\
&= 0.3 \cdot 0.13468 + 0.9 \cdot 0.27956 \\
&= 0.040404 + 0.251604 \\
&= 0.292008
\end{aligned}$$

$$\begin{aligned}
\pi(G_2) &= \sum_{j_1=1}^{2} M(G|W)_{2,j_1} \cdot \pi_G(W_{j_1}) \\
&= 0.094276 + 0.027956 \\
&= 0.122232
\end{aligned}$$

Gemäß Formel (1) ergibt sich auf Grund dieser neuen π-Werte der folgende neue BEL-Vektor beim Knoten G:

$$\begin{aligned}
BEL(G) &= \alpha \cdot \lambda(G) \times \pi(G) \\
&= \alpha \cdot \begin{pmatrix} 1.0 \\ 1.0 \end{pmatrix} \times \begin{pmatrix} 0.292008 \\ 0.122232 \end{pmatrix} \\
&= \alpha \cdot \begin{pmatrix} 0.292008 \\ 0.122232 \end{pmatrix} \overset{\text{mit } \alpha=\frac{1}{0.41424}}{=} \begin{pmatrix} 0.7049 \\ 0.2951 \end{pmatrix}
\end{aligned}$$

Wie der Vergleich zwischen dem bisherigen und dem aktualisierten BEL-Vektor des Knotens G zeigt,

bisheriger Wert: $$BEL(G) = \begin{pmatrix} 0.5892 \\ 0.4108 \end{pmatrix}$$

neuer Wert: $$BEL(G) = \begin{pmatrix} 0.7049 \\ 0.2951 \end{pmatrix}$$

ist auf Grund der im Knoten W höher gewordenen Wahrscheinlichkeit für die Hypothese, daß der Prüfling die Antwort auf die ihm gestellte Frage nicht kennt, im Knoten G die Erwartung einer hektischen Gestik (Hypothese G_1) verstärkt worden.

Der Knoten G besitzt lediglich einen Nachbarknoten, nämlich seinen Vorgängerknoten W. Folglich wurde auch die Aktivierung von G durch W verursacht (durch Übermittlung der π-Nachricht). Da aber an den Knoten, der die Aktivierung ausgelöst hat, keine Nachricht zurückgeschickt wird, werden von G keinerlei Nachrichten an andere Knoten übermittelt. Aus diesem Grunde wird der Knoten G nach der (soeben erfolgten) Aktualisierung der in G enthaltenen Wahrscheinlichkeitswerte direkt deaktiviert. Abbildung 6 auf der nachfolgenden Seite 31 zeigt den Zustand des Beispielnetzes nach der Deaktivierung von G.

4.4 Aktualisierung des Knotens N

Infolge der λ-Nachricht, die vom Knoten W an den Knoten N geschickt wird, wird auch der Knoten N aktiviert. Die erhaltene λ-Nachricht führt innerhalb von N dazu, daß der λ-Vektor neu berechnet wird. Da W der einzige Nachfolgerknoten von N ist, ergibt sich nach Formel (4) als neuer λ-Vektor von N:

$$\lambda(N) = \lambda_W(N) = \begin{pmatrix} 0.4008 \\ 0.468 \end{pmatrix}$$

Auf Grund des aktualisierten λ-Vektors ergibt sich beim Knoten N auch ein neuer BEL-Vektor:

$$\begin{aligned} BEL(N) &= \alpha \cdot \lambda(N) \times \pi(N) \\ &= \alpha \cdot \begin{pmatrix} 0.4008 \\ 0.468 \end{pmatrix} \times \begin{pmatrix} 0.8 \\ 0.2 \end{pmatrix} \\ &= \alpha \cdot \begin{pmatrix} 0.32064 \\ 0.0936 \end{pmatrix} \overset{\text{mit } \alpha=\frac{1}{0.41424}}{=} \begin{pmatrix} 0.774 \\ 0.226 \end{pmatrix} \end{aligned}$$

Vergleicht man wiederum den bisherigen BEL-Vektor mit dem aktualisierten,

$$\text{bisheriger Wert:} \qquad BEL(N) = \begin{pmatrix} 0.8 \\ 0.2 \end{pmatrix}$$

$$\text{neuer Wert:} \qquad BEL(N) = \begin{pmatrix} 0.774 \\ 0.226 \end{pmatrix}$$

so ist in diesem Falle nur eine leichte Verschiebung der Wahrscheinlichkeitsverteilung über den Hypothesen zum Wissensniveau des Prüflings festzustellen. Dies ist insofern plausibel, als das wahrscheinliche Nicht-Wissen der Antwort auf eine einzige Frage nicht gleich bedeutet, daß das Wissensniveau des Prüflings wesentlich geringer ist als ursprünglich angenommen.

Da der Knoten W der einzige Nachbarknoten von N ist, W aber die Aktivierung von N durch Übermittlung der λ-Nachricht verursacht hat, braucht der Knoten N (genauso wie Knoten G) keinen anderen Knoten über die vorgenommenen Änderungen zu informieren. Daher wird der Knoten N nach der Neuberechnung der in N enthaltenen Wahrscheinlichkeitswerte direkt deaktiviert.

Knoten N: **Wissensniveau**

N_i	$\lambda(N_i)$	$\pi(N_i)$	$BEL(N_i)$
hoch	1.0	0.8	0.8
niedrig	1.0	0.2	0.2

Knoten S: **Schwierigkeit**

S_i	$\lambda(S_i)$	$\pi(S_i)$	$BEL(S_i)$
schwer	1.0	0.6	0.6
mittel	1.0	0.3	0.3
leicht	1.0	0.1	0.1

$$\uparrow \lambda_W(N) = \begin{pmatrix} 0.4008 \\ 0.468 \end{pmatrix}$$

$$\uparrow \lambda_W(S) = \begin{pmatrix} 0.4328 \\ 0.4008 \\ 0.3432 \end{pmatrix}$$

Knoten W: **Wissen**

W_i	$\lambda(W_i)$	$\pi(W_i)$	$BEL(W_i)$
ja	0.26	0.518	0.3251
nein	0.58	0.482	0.6749

Knoten G: **Gestik**

G_i	$\lambda(G_i)$	$\pi(G_i)$	$BEL(G_i)$
hektisch	1.0	0.292008	0.7049
ruhig	1.0	0.122232	0.2951

Knoten M: **Mimik**

M_i	$\lambda(M_i)$	$\pi(M_i)$	$BEL(M_i)$
entsetzt	0.9	0.3928	0.8534
gelassen	0.1	0.6072	0.1466

Abbildung 6: Das Beispielnetz nach Aktualisierung des Knotens G: Auf Grund der vom Knoten W erhaltenen π-Nachricht wurden innerhalb des Knotens G der π- und der BEL-Vektor modifiziert. Die λ-Nachrichten, die vom Knoten W an die Knoten N und S geschickt wurden, sind zu diesem Zeitpunkt noch nicht bearbeitet worden.

4.5 Aktualisierung des Knotens S

Die Aktualisierung des Knotens S verläuft völlig analog zu der des Knotens N: Auf Grund der λ-Nachricht des Knotens W wird der Knoten S aktiviert, und es erfolgt gemäß Formel (4) eine Neubestimmung des λ-Vektors von S:

$$\lambda(S) = \lambda_W(S) = \begin{pmatrix} 0.4328 \\ 0.4008 \\ 0.3432 \end{pmatrix}$$

Infolge der Änderung des λ-Vektors von S muß auch der BEL-Vektor von S neu berechnet werden:

$$\begin{aligned} BEL(S) &= \alpha \cdot \lambda(S) \times \pi(S) \\ &= \alpha \cdot \begin{pmatrix} 0.4328 \\ 0.4008 \\ 0.3432 \end{pmatrix} \times \begin{pmatrix} 0.6 \\ 0.3 \\ 0.1 \end{pmatrix} \\ &= \alpha \cdot \begin{pmatrix} 0.25968 \\ 0.12024 \\ 0.03432 \end{pmatrix} \text{mit } \overset{\alpha=\frac{1}{0.41424}}{=} \begin{pmatrix} 0.62688 \\ 0.29027 \\ 0.08285 \end{pmatrix} \end{aligned}$$

Genauso wie beim Knoten N ergibt auch beim Knoten S ein Vergleich des bisherigen BEL-Vektors mit dem modifizierten,

$$\text{bisheriger Wert:} \qquad BEL(S) = \begin{pmatrix} 0.6 \\ 0.3 \\ 0.1 \end{pmatrix}$$

$$\text{neuer Wert:} \qquad BEL(S) = \begin{pmatrix} 0.62688 \\ 0.29027 \\ 0.08285 \end{pmatrix}$$

daß der BEL-Vektor bei seiner Neuberechnung nur leicht abgeändert wurde. Analog zum Knoten N ist dieser Umstand dadurch plausibel, daß die im Knoten W vorgenommene Erhöhung der Wahrscheinlichkeit des Nicht-Wissens des Prüflings noch keine sehr große Evidenz dafür liefert, daß der Schwierigkeitsgrad der Frage zu niedrig eingeschätzt wurde.

Da auch der Knoten S keine Nachbarknoten über die vorgenommenen Änderungen informieren muß (aus demselben Grund wie beim Knoten N), wird S nach der Neuberechnung der Wahrscheinlichkeitswerte direkt deaktiviert.

Damit sind innerhalb des Beispielnetzes alle Änderungen durchgeführt, die sich auf Grund der Beobachtung des entsetzten Gesichtsausdruckes des Prüflings ergeben. Da zu diesem Zeitpunkt keine weiteren Nachrichten mehr bearbeitet werden müssen, ist das Beispielnetz damit bis zur nächsten Beobachtung in einem stabilen Zustand, der in Abbildung 7 auf der nachfolgenden Seite 33 dargestellt ist.

Knoten N: **Wissensniveau**

N_i	$\lambda(N_i)$	$\pi(N_i)$	$BEL(N_i)$
hoch	0.4008	0.8	0.774
niedrig	0.468	0.2	0.226

Knoten S: **Schwierigkeit**

S_i	$\lambda(S_i)$	$\pi(S_i)$	$BEL(S_i)$
schwer	0.4328	0.6	0.62688
mittel	0.4008	0.3	0.29027
leicht	0.3432	0.1	0.08285

Knoten W: **Wissen**

W_i	$\lambda(W_i)$	$\pi(W_i)$	$BEL(W_i)$
ja	0.26	0.518	0.3251
nein	0.58	0.482	0.6749

Knoten G: **Gestik**

G_i	$\lambda(G_i)$	$\pi(G_i)$	$BEL(G_i)$
hektisch	1.0	0.292008	0.7049
ruhig	1.0	0.122232	0.2951

Knoten M: **Mimik**

M_i	$\lambda(M_i)$	$\pi(M_i)$	$BEL(M_i)$
entsetzt	0.9	0.3928	0.8534
gelassen	0.1	0.6072	0.1466

Abbildung 7: Das Beispielnetz nach Aktualisierung der Knoten N und S: Auf Grund der vom Knoten W erhaltenen λ-Nachrichten wurden innerhalb der Knoten N und S jeweils der λ- und der BEL-Vektor modifiziert. Damit sind alle Nachrichten bearbeitet worden.

5 Schlußbemerkungen

In dieser Arbeit wurde an einem konkreten Beispiel die Struktur und die zugrundeliegenden Mechanismen von Bayesschen Netzen illustriert. Das in dem Beispielszenario verwendete Bayessche Netz besitzt dabei eine relativ einfache Struktur. Genauer gesagt handelt es sich bei dem Beispielnetz um ein sogenanntes **einfach verbundenes Bayessches Netz** (engl. *singly connected bayesian network*), d.h. um ein Bayessches Netz, bei dem zwischen zwei beliebigen Knoten höchstens ein Pfad entlang der gerichteten Kanten des Netzes existiert (im Gegensatz dazu stehen **mehrfach verbundene Bayessche Netze** (engl. *multiply connected bayesian networks*), bei denen zwischen zwei Knoten des Netzes beliebig viele Pfade existieren dürfen). Demzufolge handelt es sich bei den angegebenen Mechanismen zur Initialisierung und Modifizierung des Beispielnetzes um die Methoden für den einfacheren Fall der einfach verbundenen Bayesschen Netze. Ob und wie sich diese Methoden auf bzw. für den Fall von mehrfach verbundenen Bayesschen Netzen übertragen und erweitern lassen, wird z.B. in Abschnitt 4.4 *Coping with Loops* von [Pearl 91] und in [Lauritzen & Spiegelhalter 88] diskutiert.

Zur Implementierung von Bayesschen Netzen sei auf das System HUGIN[14] hingewiesen, bei dem es sich um eine professionelle Entwicklungsumgebung zur Erstellung von Bayesschen Netzen handelt (siehe [Andersen et al. 89]). HUGIN wurde aus dem System MUNIN[15] heraus entwickelt,[16] einer der allerersten Anwendungen Bayesscher Netze (siehe [Andreassen et al. 87]).

Ferner sei an dieser Stelle erwähnt, daß im Rahmen der vorliegenden Arbeit ein kleines Programmpaket zur Erstellung Bayesscher Netze mittels der Programmiersprache C implementiert wurde. Dieses Programmpaket wurde auf dem FTP-Server `ps-ftp.dfki.uni-sb.de` unter dem Verzeichnis `/pub/gk/Bayes` zur Verfügung gestellt und kann über anonymes FTP abgerufen werden.

[14] HUGIN ist eine Abkürzung für „Handling Uncertainty in General Inference Networks“.

[15] MUNIN ist ein Abkürzung für „MUscle and Nerve Inference Network“.

[16] Ein Zusammenhang wird zwischen diesen beiden Systemen auch durch ihre Namen hergestellt, da es sich bei den Bezeichnungen *Munin* und *Hugin* nicht nur um Abkürzungen für die angegebenen Wendungen handelt, sondern auch um die Namen der beiden Raben, die in der nordischen Göttermythologie dem Gott Odin als Ratgeber zur Seite stehen.

Literatur

[Abramson 91] B. **Abramson**. *On Knowledge Representation in Belief Networks.* In: B. Bouchon-Meunier, R. Yager und L. Zadeh (Hrsg.), Uncertainty in Knowledge Bases (Proceedings of the 3rd International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU 90), Paris, Frankreich, 2.-6. Juli 1990). Springer, Berlin, 1991. S. 86–96.

[Andersen et al. 89] S. **Andersen**, K. **Olesen**, Finn V. **Jensen** und Frank **Jensen**. *HUGIN – a Shell for Building Bayesian Belief Universes for Expert Systems.* In: N. Sridharan (Hrsg.), Proceedings of the 11th International Joint Conference on Artificial Intelligence (IJCAI 89), Detroit, USA, 20.-25. August 1989. Morgan Kaufmann, San Mateo (California), 1989. S. 1080–1085.

[Andreassen et al. 87] S. **Andreassen**, M. **Woldbye**, B. **Falck** und S. **Andersen**. *MUNIN – A Causal Probabilistic Network for Interpretation of Electromyographic Findings.* In: J. McDermott (Hrsg.), Proceedings of the 10th International Joint Conference on Artificial Intelligence (IJCAI 87), Mailand, Italien, 23.-28. August 1987. Morgan Kaufmann, Los Altos (California), 1987. S. 366–372.

[Bauer 91] H. **Bauer**. *Wahrscheinlichkeitstheorie.* 4., völlig überarbeitete und neugestaltete Auflage des Werkes: „Wahrscheinlichkeitstheorie und Grundzüge der Maßtheorie". de Gruyter, Berlin, 1991.

[Charniak 91] E. **Charniak**. *Bayesian Networks without Tears.* AI Magazine 12, Nr. 4, Winter 1991. S. 50–63.

[Dallmann & Elster 92] H. **Dallmann** und K.-H. **Elster**. *Einführung in die höhere Mathematik*, Band III. Gustav Fischer Verlag, Jena, 1992.

[Goldman & Charniak 91] R. **Goldman** und E. **Charniak**. *Dynamic Construction of Belief Networks.* In: P. Bonissone, M. Henrion, L. Kanal und J. Lemmer (Hrsg.), Uncertainty in Artificial Intelligence 6. North-Holland, Amsterdam, 1991. S. 171–184.

[Henrion 89] M. **Henrion**. *Some Practical Issues in Constructing Belief Networks.* In: L. Kanal, T. Levitt und J. Lemmer (Hrsg.), Uncertainty in Artificial Intelligence 3, North-Holland, Amsterdam, 1989. S. 161–173.

[Hinderer 75] K. **Hinderer**. *Grundbegriffe der Wahrscheinlichkeitstheorie.* Springer, Berlin, 1975.

[Jameson 90] A. **Jameson**. *Knowing What Others Know. Studies in Intuitive Psychometrics.* Doktorarbeit, Universität Amsterdam, Niederlande, 1990.

[Kruse et al. 91] R. **Kruse**, E. **Schwecke** und J. **Heinsohn**. *Uncertainty and Vagueness in Knowledge Based Systems.* Springer, Berlin, 1991.

[Lauritzen & Spiegelhalter 88] S. **Lauritzen** und D. **Spiegelhalter**. *Local Computations with Probabilities on Graphical Structures and their Application to Expert Systems.* Journal of the Royal Statistical Society (Series B) 50, Nr. 2, 1988. S. 157–224.

[Neapolitan 90] R. **Neapolitan**. *Probabilistic Reasoning in Expert Systems: Theory and Algorithms.* Wiley, New York, 1990.

[Pearl 86] J. **Pearl**. *Fusion, Propagation, and Structuring in Belief Networks.* Artificial Intelligence 29, 1986. S. 241–288.

[Pearl 91] J. **Pearl**. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* Revised second printing. Morgan Kaufmann, San Mateo (California), 1991.