Fortgeschrittene Themen der Wissensrepräsentation

Gabriele Kern-Isberner
LS 1 – Information Engineering

TU Dortmund
Sommersemester 2015

Kapitel 4

4. Revision in der Probabilistik

Kapitel 4

4. Revision in der Probabilistik

4.1 Probabilistische Revision – allgemein

Die älteste Revisionsmethode der Welt

(Epistemische) AGM-Revision behandelt das Problem, einen Wissenszustand Ψ durch neue, sicher geglaubte Information A zu verändern:

$$\Psi * A.$$

In einer probabilistischen Umgebung ist eine Wahrscheinlichkeitsverteilung P ein Wissenszustand:

$$P * A.$$

Eine Methode, die P dahingehend verändert, dass A nachher sicher geglaubt wird, d.h. mit Wahrscheinlichkeit 1, ist die Bayes'sche Konditionalisierung

$$P * A(\omega) = P(\omega|A) = \begin{cases} \frac{P(\omega)}{P(A)} & \text{falls } \omega \models A \\ 0 & \text{falls } \omega \models \overline{A} \end{cases}$$

(wenn $P(A) > 0$). (Success) ist gegeben durch

$$P * A(A) = 1.$$

Jeffrey's Regel 1/2

Konditionalisierung lässt sich verallgemeinern:

Seien $B_1, \ldots, B_n$ disjunkte Aussagen, über die zunächst Wahrscheinlichkeiten

$$P(B_1), \ldots, P(B_n)$$

bekannt sind; wir können annehmen, dass $B_1, \ldots, B_n$ auch erschöpfend sind, d.h. dass gilt

$$P(B_1) + \ldots + P(B_n) = 1.$$

Wie verändert sich die gesamte Verteilung P, wenn nun neue Informationen über die Wahrscheinlichkeiten der B_i bekannt werden?

D.h., P soll so zu P^* verändert werden, dass $P^*(B_i) = p_i$ vorgegebene Werte annimmt.

Jeffrey's Regel 2/2

Die Lösung zu diesem Problem liefert eine Annahme, die man als probability kinematics bezeichnet – nämlich, dass die neuen Wahrscheinlichkeiten der B_i keine der unter B_i bedingten Wahrscheinlichkeiten ändern sollte:

$$P^*(A|B_i) = P(A|B_i)$$

Daraus ergibt sich sofort mit dem Satz von der totalen Wahrscheinlichkeit die Regel von Jeffrey:

$$\begin{aligned} P^*(A) &= \sum_{i=1}^{n} P^*(A|B_i)P^*(B_i) \\ &= \sum_{i=1}^{n} P(A|B_i)P^*(B_i) \end{aligned}$$

Jeffrey's und Bayes Regel

Jeffrey's Regel verallgemeinert die Konditionalisierung nach Bayes:

Liegt nämlich nur ein Ereignis B mit Wahrscheinlichkeit $P^*(B) = 1$ vor, so ergibt Jeffrey's Regel:

$$P^*(A) = P(A|B)P^*(B) = P(A|B),$$

d.h. die posteriori Wahrscheinlichkeit ist nichts anderes als die nach B konditionalisierte priori Wahrscheinlichkeit; umgekehrt erhält man die bedingte Wahrscheinlichkeit als Spezialfall der Regel von Jeffrey, wenn die neue Information sicher ist, also Wahrscheinlichkeit 1 besitzt.

Multiple probabilistische Revision und mehr ...

Jeffrey's Regel kann als multiple probabilistische Revision (unter gewissen Bedingungen) betrachtet werden, die neue Information besteht hier aus einer Menge probabilistischer Fakten $\mathcal{S} = \{B_1[p_1], \ldots, B_n[p_n]\}$, und die Revisionsaufgabe besteht in der Bestimmung einer Verteilung

$$P * \mathcal{S} \quad \text{mit} \quad P * \mathcal{S} \models \mathcal{S} \ (\textit{Success}),$$

wobei

$$P \models B_i[p_i] \quad \text{gdw.} \quad P(B_i) = p_i.$$

Multiple probabilistische Revision und mehr . . . (Forts.)

In einem nächsten Schritt könnte die Revisionsaufgabe nun darin bestehen, P mit einer Menge probabilistischer Regeln

$$\mathcal{R} = \{(B_1|A_1)[x_1], \ldots, (B_n|A_n)[x_n]\}$$

so zu einer Verteilung $P * \mathcal{R}$ zu verändern, dass

$$P * \mathcal{R} \models (B_i|A_i)[x_i], 1 \leq i \leq n, \quad \textit{(Success)}$$

gilt, wobei

$$P \models (B|A)[x] \quad \text{gdw.} \quad P(A) > 0 \text{ und } P(B|A) = x.$$

Kapitel 4

4. Revision in der Probabilistik

4.2 Wissensrevision auf der Basis optimaler Entropie

Probabilistische Wissensrevision 1/3

Wenden wir uns dem Problem zu, Wissenszustände durch neue (konditionale) Informationen revidieren zu wollen.

Geeignete probabilistische Wissenszustände sind Wahrscheinlichkeitsverteilungen.

Geeignete neue Informationen sind Mengen probabilistischer Regeln.

Gesucht ist also ein Verfahren, um Wahrscheinlichkeitsverteilungen P durch Mengen probabilistischer Regeln $\mathcal{R}$ zu revidieren.

Probabilistische Wissensrevision 2/3

Mit welchen Problemen müssen wir rechnen?

- Auf der a priori-Seite (gegebenes Wissen) haben wir zuwenig Freiheitsgrade, da eine Wahrscheinlichkeitsverteilung vollständiges Wissen darstellt.
- Auf der a posteriori-Seite (revidierter Wissenszustand) haben wir zuviele Freiheitsgrade, da es – wie bei der *ME*-Inferenz – normalerweise unendlich viele passende Verteilungen gibt.

Probabilistische Wissensrevision 3/3

Was erwarten wir?

- Wir möchten genau eine revidierte Verteilung $P^* = P * \mathcal{R}$ berechnen.
- Diese Verteilung soll die neue Information darstellen: $P^* \models \mathcal{R}$ (Success).
- Unnötige Arbeit soll vermieden werden, d.h. wenn bereits $P \models \mathcal{R}$ gilt, so soll $P^* = P$ sein (Stabilität).
- Die Änderungen in P sollen minimal sein, d.h. die neue Verteilung P^* soll so nahe wie möglich an P liegen – wir benötigen also ein Abstandsmaß zwischen Verteilungen.

Relative Entropie 1/2

Problem: Die Verteilung P hat sich geändert zur Verteilung Q;
Gesucht: Maß für Informationsgewinn, dass man diese Änderung bemerkt hat

$$R(Q, P) = \sum_{\omega} Q(\omega) \log \frac{Q(\omega)}{P(\omega)}$$

Relative Entropie (cross entropy) von Q bezgl. P
Maß für den Informationsabstand von P zu Q

Relative Entropie 2/2

Eigenschaften der relativen Entropie:

- Die relative Entropie ist positiv definit, d.h. $R(Q, P) \geq 0$ und $R(Q, P) = 0$ genau dann, wenn $P = Q$;
- Die relative Entropie ist nicht symmetrisch: $R(Q, P) \neq R(P, Q)$ (i.Allg.);
- Sei $P = (p_1, \ldots, p_n)$ eine Wahrscheinlichkeitsverteilung, und sei $P_0 = (\frac{1}{n}, \ldots, \frac{1}{n})$ eine (passende) Gleichverteilung. Dann gilt

$$R(P, P_0) = \log n - H(P)$$

ME-Wissensrevision

- Gegeben: A priori-Verteilung P, neue Information $\mathcal{R}$;
- Gesucht: A posteriori-Verteilung P^* "nahe" zu P mit $P^* \models \mathcal{R}$;
- Lösung:

Prinzip der minimalen Relativentropie

Minimiere Informationsabstand zu P, gegeben
$\mathcal{R} = \{(B_1|A_1)[x_1], \ldots, (B_n|A_n)[x_n]\} \to$ Optimierungsproblem

$$P^* = \mathit{ME}(P, \mathcal{R}) = ((\arg) \min_{Q \models \mathcal{R}} R(Q, P) = \sum_{\omega} Q(\omega) \log_2 \frac{Q(\omega)}{P(\omega)}$$

eindeutig lösbar (für P-konsistentes[1] $\mathcal{R}$) mit Lösung $P^* = \mathit{ME}(P, \mathcal{R})$.

[1] $\mathcal{R}$ heißt P-konsistent, wenn es ein Modell Q von $\mathcal{R}$ gibt mit $P(\omega) = 0 \Rightarrow Q(\omega) = 0$; in diesem Fall sagt man auch, Q ist P-stetig.

Axiomatik des ME-Prinzips 1/2

Shore & Johnson, 1980:

Die relative Entropie ist die einzige Funktion, deren Optimierung das obige probabilistische Revisionsproblem in der Art löst, dass folgende Bedingungen erfüllt sind:

- Eindeutigkeit: Das Revisionsergebnis $P * \mathcal{R}$ ist eindeutig bestimmt.
- Invarianz: Die Lösung hängt nicht vom Koordinatensystem ab.
- System Independence: Sei $\mathbf{V} = \mathbf{V}_1 \cup \mathbf{V}_2$ eine Partitionierung der Variablenmenge, und seien jeweils $P_i, \mathcal{R}_i$ Verteilungen bzw. Regelmengen über $\mathbf{V}_i$. Dann gilt:

$$(P_1 P_2) * (\mathcal{R}_1 \cup \mathcal{R}_2) = (P_1 * \mathcal{R}_1)(P_2 * \mathcal{R}_2)$$

Axiomatik des ME-Prinzips 2/2

- Subset Independence: Seien $A_1, \ldots A_n$ exklusive und erschöpfende Formeln, d.h. $A_i A_j \equiv \perp$ für $i \neq j$ und $A_1 \vee \ldots \vee A_n \equiv \top$. Sei P eine Verteilung, und seien $\mathcal{R}_i$ Mengen probabilistischer Konditionale, deren Prämisse A_i impliziert (d.h. die $\mathcal{R}_i$ liefern jeweils Informationen über die bedingten Verteilungen $P(\cdot|A_i)$). Sei $\mathcal{R} = \mathcal{R}_1 \cup \ldots \cup \mathcal{R}_n$, und sei $\mathcal{S} = \{A_1[x_1], \ldots, A_n[x_n]\}$ mit $\sum_i x_i = 1$. Dann gilt

$$(P * (\mathcal{R} \cup \mathcal{S}))(\cdot|A_i) = P(\cdot|A_i) * \mathcal{R}_i$$

Insbesondere gilt dann

$$(P * \mathcal{R})(\cdot|A_i) = P(\cdot|A_i) * \mathcal{R}_i$$

ME-Wissensrevision und ME-Inferenz 1/2

Ein (zweistelliger) ME-Revisionsoperator $*_{ME}$ wird definiert durch

$$P *_{ME} \mathcal{R} := \mathit{ME}(P, \mathcal{R})$$

Nichtmonotone ME-Inferenzoperation (mit Hintergrundwissen P):

$$\begin{array}{rcl} C_P^{ME} & : & 2^{(\mathcal{L}|\mathcal{L})^{prob}} \to 2^{(\mathcal{L}|\mathcal{L})^{prob}} \\ C_P^{ME}(\mathcal{R}) & = & \begin{cases} \{\varphi \in (\mathcal{L} \mid \mathcal{L})^{prob} \mid P *_{ME} \mathcal{R} \models \varphi\} & \mathcal{R}\ P-\text{konsistent} \\ (\mathcal{L} \mid \mathcal{L})^{prob} & \text{sonst} \end{cases} \end{array}$$

ME-Wissensrevision und ME-Inferenz 2/2

Ist $P = P_0$ die Gleichverteilung, so ist $ME(P_0, \mathcal{R}) = ME(\mathcal{R})$, d.h. die zu $\mathcal{R}$ gehörige *ME*-Verteilung ist dasjenige $\mathcal{R}$-Modell, das minimalen Informationsabstand zur Gleichverteilung hat.

Für $P = P_0$ wird also

$$C^{ME}(\mathcal{R}) = C^{ME}_{P_0}(\mathcal{R})$$

durch die zugehörige *ME*-Verteilung $ME(\mathcal{R})$ bestimmt.

Eigenschaften der ME-Revision

Der *ME*-Revisionsoperator hat folgende Eigenschaften:

- $P *_{ME} \mathcal{R} = P$ genau dann, wenn $P \models \mathcal{R}$.
- Es ist i.Allg.

$$P *_{ME} (\mathcal{R} \cup \mathcal{S}) \neq (P *_{ME} \mathcal{R}) *_{ME} \mathcal{S},$$

aber es gilt:

$$(P *_{ME} \mathcal{R}) *_{ME} (\mathcal{R} \cup \mathcal{S}) = P *_{ME} (\mathcal{R} \cup \mathcal{S}) \quad \text{(Kohärenz)}$$

Eigenschaften der ME-Inferenz

Der *ME*-Inferenzoperator C_P^{ME} erfüllt die folgenden Eigenschaften (s. Commonsense Reasoning):

- Inklusion/Reflexivität: $\mathcal{R} \subseteq C_P^{ME}(\mathcal{R})$.
- Idempotenz: $C_P^{ME}(C_P^{ME}(\mathcal{R})) = C_P^{ME}(\mathcal{R})$.
- Kumulativität:

$$\mathcal{R} \subseteq \mathcal{S} \subseteq C_P^{ME}(\mathcal{R}) \quad \text{impliziert} \quad C_P^{ME}(\mathcal{R}) = C_P^{ME}(\mathcal{S})$$

- Supraklassizität, d.h. es gilt: $Cn_{prob}(\mathcal{R}) \subseteq C_P^{ME}(\mathcal{R})$.
- Loop: Sind $\mathcal{R}_1, \ldots, \mathcal{R}_m \subseteq (\mathcal{L} \mid \mathcal{L})^{prob}$ mit $\mathcal{R}_{i+1} \subseteq C_P^{ME}(\mathcal{R}_i)$, i modulo m, dann gilt

$$C_P^{ME}(\mathcal{R}_i) = C_P^{ME}(\mathcal{R}_j) \quad \text{für alle} \;\; i, j = 1, \ldots, m$$

Prinzip der konditionalen Erhaltung in der Probabilistik 1/3

Charakteristika von Konditionalen:

- Konditionale sind 3-wertig:

$$(B|A)(\omega) = \begin{cases} 1 & \text{wenn} \quad \omega \models AB \ (\textit{Verifikation}) \\ 0 & \text{wenn} \quad \omega \models A\overline{B} \ (\textit{Falsifikation}) \\ u & \text{wenn} \quad \omega \models \overline{A} \end{cases}$$

- Die Semantik von Konditionalen wird bestimmt durch Brüche:

$$P(B|A) = \frac{1}{1 + \frac{P(A\overline{B})}{P(AB)}}$$

Genauso wie für OCF benutzen wir beide Charakteristika zur Definition eines allgemeinen konditionalen Prinzips der Erhaltung in der Probabilistik.

Prinzip der konditionalen Erhaltung in der Probabilistik 2/3

ME Wissensänderung erfüllt folgendes

Probabilistisches Prinzip der konditionalen Erhaltung

Seien $\Omega = \{\omega_1, \ldots, \omega_m\}$ und $\Omega' = \{\omega'_1, \ldots, \omega'_m\}$ zwei Multimengen möglicher Welten (Welten also nicht notwendigerweise verschieden). Wenn für jedes probabilistische Konditional $(B_i|A_i)[x_i]$ in $\mathcal{R}$ sich Ω und Ω' gleich verhalten (s.u.), i.e., sie zeigen dieselbe Anzahl von Verifikationen bzw. Falsifikationen, dann gilt für die a priori P und die (*ME*-)posteriori P^* Folgendes:

$$\frac{P(\omega_1)\ldots P(\omega_m)}{P(\omega'_1)\ldots P(\omega'_m)} = \frac{P^*(\omega_1)\ldots P^*(\omega_m)}{P^*(\omega'_1)\ldots P^*(\omega'_m)}$$

Das Prinzip der konditionalen Erhaltung schafft also eine Balance zwischen der Erhaltung konditionaler Beziehungen in der priori Verteilung und der Herstellung konditionaler Beziehungen durch neue Informationen.

Prinzip der konditionalen Erhaltung in der Probabilistik 3/3

Ebenso wie für OCF kann man das Prinzip mit Hilfe konditionaler Strukturen noch kompakter darstellen:

Definieren wir dazu auch noch

$$P(\Omega) = \prod_{i=1}^{m} P(\omega_i)$$

für Multimengen $\Omega = \{\omega_1, \ldots, \omega_m\}$, so lässt sich das Prinzip der konditionalen Erhaltung kompakt so ausdrücken:

Prinzip der konditionalen Erhaltung

Gilt für zwei Multimengen $\sigma(\Omega) = \sigma(\Omega')$, so muss auch

$$\frac{P(\Omega)}{P(\Omega')} = \frac{P^*(\Omega)}{P^*(\Omega')}$$

mit $P^* = P * \mathcal{R}$ gelten, d.h. priori-Bruch = posteriori-Bruch.

ME-Wissensänderung unter der Lupe

Für $\mathcal{R} = \{(B_1|A_1)\,[x_1], \ldots, (B_n|A_n)\,[x_n]\}$ erhalten wir

$$ME(P,\mathcal{R})(\omega) = \alpha_0 P(\omega) \prod_{\substack{1 \leq i \leq n \\ \omega \models A_i B_i}} \alpha_i^{1-x_i} \prod_{\substack{1 \leq i \leq n \\ \omega \models A_i \overline{B_i}}} \alpha_i^{-x_i}$$

$$\text{mit} \quad \alpha_i = \frac{x_i}{1-x_i} \frac{\displaystyle\sum_{\omega \models A_i \overline{B_i}} P(\omega) \prod_{\substack{j \neq i \\ \omega \models A_j B_j}} \alpha_j^{1-x_j} \prod_{\substack{j \neq i \\ \omega \models A_j \overline{B_j}}} \alpha_j^{-x_j}}{\displaystyle\sum_{\omega \models A_i B_i} P(\omega) \prod_{\substack{j \neq i \\ \omega \models A_j B_j}} \alpha_j^{1-x_j} \prod_{\substack{j \neq i \\ \omega \models A_j \overline{B_j}}} \alpha_j^{-x_j}},$$

$$\text{und} \quad \alpha_i \begin{cases} > 0 & : \quad x_i \in (0,1) \\ = \infty & : \quad x_i = 1 \\ = 0 & : \quad x_i = 0 \end{cases} \quad , 1 \leq i \leq n.$$

ME-Prinzip, Bayes und Jeffrey

Mit den obigen Formeln rechnet man leicht nach:

- $\mathcal{R} = \{A[1]\}$:

$$P *_{ME} \{A[1]\} = P(\cdot|A)$$

  $\rightarrow$ Bayes-Konditionierung

- $\mathcal{R} = \{A[x]\}$:

$$P *_{ME} \{A[x]\})(\omega) = P(\omega|A)x + P(\omega|\overline{A})(1-x)$$

  $\rightarrow$ Jeffrey's Regel

Probabilistische Wissensrevision und AGM 1/2

Wissensmenge einer Wahrscheinlichkeitsverteilung P

$$K_P = \{A \in \mathcal{L} \mid P(A) = 1\}$$

$K_P \neq \emptyset$, $K_P \neq \mathcal{L}$; K_P ist deduktiv abgeschlossen.

Definiere Revision von K_P mit A durch

$$K_P * A = K_{P(\cdot|A)} = \{B \in \mathcal{L} \mid P(B|A) = 1\}$$

Probabilistische Wissensrevision und AGM 2/2

Diese Revision ist nur definiert, wenn $P(A) > 0$ ist, wenn also $\neg A \notin K_P$ ist, und ist damit eine Expansion. Sie ist allerdings eine AGM-Expansion, denn es ist

$$K_P * A = Cn(K_P \cup \{A\})$$

Beispiel Psychologe

Ein Psychologe fasst seine langjährigen Erfahrungen in der Drogenberatungsstelle in einer Verteilung P zusammen:

a : *addicted to alcohol*
d : *addicted to drugs*
y : *being young*

ω	$P(\omega)$	ω	$P(\omega)$	ω	$P(\omega)$	ω	$P(\omega)$
ady	0.050	$\overline{a}dy$	0.333	$ad\overline{y}$	0.053	$\overline{a}d\overline{y}$	0.053
$a\overline{d}y$	0.093	$\overline{a}\overline{d}y$	0.102	$a\overline{d}\overline{y}$	0.225	$\overline{a}\overline{d}\overline{y}$	0.091

Beispiel Psychologe (Forts.)

Wir beobachten in P die folgenden Wahrscheinlichkeiten:

$$
\begin{array}{ll}
P(d|a) = 0.242 & P(d|\overline{a}) = 0.666 \\
P(a|y) = 0.246 & P(d|y) = 0.662 \\
P(a|\overline{y}) = 0.660 & P(d|\overline{y}) = 0.251
\end{array}
$$

Beispiel Psychologe (Forts.)

Der Psychologe wechselt nun die Stelle – er wird in einer Klinik arbeiten, in der ausschließlich Alkohol- und Drogenabhängige behandelt werden. In seinem neuen Arbeitsbereich ist die Rate der alkoholabhängigen Personen, die gleichzeitig auch drogenabhängig sind, höher als üblich ist und liegt bei ca. 40 %;

der Psychologe wird also sein Wissen P mit der folgenden Information revidieren:

$$\mathcal{R} = \{a \vee d[1], (d|a)[0.4]\}$$

Beispiel Psychologe (Forts.)

Er erhält $P^* = P *_{ME} \mathcal{R}$:

ω	$P^*(\omega)$	ω	$P^*(\omega)$	ω	$P^*(\omega)$	ω	$P^*(\omega)$
ady	0.099	$\overline{a}dy$	0.425	$ad\overline{y}$	0.105	$\overline{a}d\overline{y}$	0.066
$a\overline{d}y$	0.089	$\overline{a}\overline{d}y$	0.0	$a\overline{d}\overline{y}$	0.216	$\overline{a}\overline{d}\overline{y}$	0.0

$$P^* \models \quad \begin{array}{ll} (d|a)[0.4] & (d|\overline{a})[1] \\ (a|y)[0.307] & (d|y)[0.855] \end{array}$$

Beispiel Psychologe (Forts.)

Nach ein paar Tagen bemerkt der Psychologe, dass es in der Klinik nur junge Leute gibt; er muss sein Wissen erneut revidieren und berechnet jetzt $P *_{ME} (\mathcal{R} \cup y[1]) =: P_1^*$:

ω	$P_1^*(\omega)$	ω	$P_1^*(\omega)$
ady	0.120	$\overline{a}dy$	0.700
$ad\overline{y}$	0.0	$\overline{a}d\overline{y}$	0.0
$a\overline{d}y$	0.180	$\overline{a}\overline{d}y$	0.0
$a\overline{d}\overline{y}$	0.0	$\overline{a}\overline{d}\overline{y}$	0.0

Beispiel Psychologe (Forts.)

Diese Wahrscheinlichkeit ist allerdings verschieden von der, die man durch Konditionieren nach y aus $P^* = P *_{ME} \mathcal{R}$ erhält:

ω	$P^*(\cdot\|y)(\omega)$	ω	$P^*(\cdot\|y)(\omega)$
ady	0.162	$\overline{a}dy$	0.693
$ad\overline{y}$	0.0	$\overline{a}d\overline{y}$	0.0
$a\overline{d}y$	0.145	$\overline{a}\overline{d}y$	0.0
$a\overline{d}\overline{y}$	0.0	$\overline{a}\overline{d}\overline{y}$	0.0

Diese (bedingten) Wahrscheinlichkeiten passen zur Anwendung des psychologischen Wissens auf den Fall einer jungen Person (Focusing).

Revision und Focusing 1/3

Unter Focusing versteht man das Anwenden generischen Wissens auf einen Fall. Wird der Fall durch die Proposition A beschrieben, so entspricht Focusing also einer Revisions- oder Update-Operation mit sicherem Wissen.

Im klassisch-logischen Rahmen der AGM-Revision bleibt einem folglich nichts anderes übrig, als mit A eine Revision oder ein Update durchzuführen:

$$K * A,$$

so dass Focusing von Revision bzw. Update nicht zu unterscheiden ist.

Revision und Focusing 2/3

Im *ME*-probabilistischen Rahmen hat man hingegen sehr viel mehr Modellierungsspielraum:

Im Psychologen-Beispiel gab es eine sehr deutliche Unterscheidung zwischen Revision und Focusing ($P^* =: P *_{ME} \mathcal{R}$)

Revision von P^* mit $y[1]$		Focusing von P^* auf $y[1]$
$P *_{ME} (\mathcal{R} \cup y[1])$	$\neq$	$(P *_{ME} \mathcal{R})(\cdot \vert y)$

Focusing entspricht also grundsätzlich dem üblichen Konditionieren.

Revision und Focusing 3/3

Man kann sich auch vorstellen, dass man nur über unsichere Information über einen Fall verfügt, z.B. A ist wahr mit Wahrscheinlichkeit 0.8. In diesem Fall muss man auf die unsichere Evidenz $A[0.8]$ fokussieren.

Wegen

$$P(\cdot|A) = P *_{ME} A[1]$$

ist hier eine passende Verallgemeinerung die Operation

$$P *_{ME} A[x],$$

in diesem Fall also für $x = 0.8$.

Beispiel Psychologe (Forts.)

In unserem Beispiel würde eine unsichere Fokussierung $P_2^* = P^* *_{ME} \{y[0.8]\}$ wie folgt aussehen:

ω	$P_2^*(\omega)$	ω	$P_2^*(\omega)$
ady	0.129	$\overline{a}dy$	0.555
$ad\overline{y}$	0.054	$\overline{a}d\overline{y}$	0.034
$a\overline{d}y$	0.116	$\overline{a}\overline{d}y$	0.0
$a\overline{d}\overline{y}$	0.112	$\overline{a}\overline{d}\overline{y}$	0.0

Fazit ME-Methodik

- Die *ME*-Methodik ermöglicht Wissensrevision bzw. Inferenz in ihrer allgemeinsten Form: Komplexe Wissenszustände (Wahrscheinlichkeitsverteilungen) können durch Anpassung an neue Information in komplexer Form (Mengen probabilistischer Konditionale) revidiert werden.
- Auch *ME*-Inferenz mit Hintergrundwissen erfüllt zahlreiche der Eigenschaften, die man an nichtmonotone Inferenzrelationen i.Allg. stellt, z.B. die Kumulativität.
- *ME*-Revision ist mit den Grundideen der AGM-Theorie (minimal change-Paradigma) verträglich; darüberhinaus gelten weitergehende Eigenschaften, die insbesondere für iterierte Revision relevant sind.

Ein Paradoxon ...

Das folgende Beispiel zeigt, wie leicht ein naiver Ansatz zur Modellierung von Nicht-Wissen mit den Gesetzen der Wahrscheinlichkeit in Konflikt kommen kann: Die Peter, Paul & Mary-Story

- Peter, Paul und Mary sind Profikiller, einer von ihnen hat Mr Jones ermordet;
- Inspektor Smith ist bisher noch vollkommen ratlos –
  - wer von den dreien war es? $P_1(Mary) = \frac{1}{3}$
  - war es ein Mann oder eine Frau? $P_2(Mary) = \frac{1}{2}$

Bei der Verwendung von Wahrscheinlichkeiten geht auch immer Wissen über die Grundgesamtheit betrachteter Merkmale mit ein!

. . . und ein Nicht-Paradoxon 1/4

Die Peter, Paul & Mary-Story (Forts.)

- Inspektor Smith weiß nun, dass Big Boss die Ermordung von Mr Jones durch Peter, Paul oder Mary in Auftrag gegeben hat;
- er weiß auch, dass Big Boss vorher eine (faire) Münze geworfen hat, um zu entscheiden, ob der Auftrag durch eine Frau oder einen Mann ausgeführt werden sollte;
- er weiß aber nichts über den Ausgang des Münzwurfs.

. . . und ein Nicht-Paradoxon 2/4

Damit lässt sich das Wissen von Inspektor Smith wie folgt formalisieren:

$$\mathcal{R}_1 = \{\mathit{Mary}\,[0.5], \mathit{Peter} \vee \mathit{Paul}\,[0.5]\}$$

Sei $P_1 = \mathit{ME}(\mathcal{R}_1)$ die zu $\mathcal{R}_1$ gehörige *ME*-Verteilung:

$$P_1(\mathit{Mary}) = 0.5,\ P_1(\mathit{Peter}) = P_1(\mathit{Paul}) = 0.25$$

Nun erfährt Inspektor Smith, dass Peter sich zur Tatzeit bei einer Polizeistation gemeldet hat, somit ein wasserdichtes Alibi hat und als Täter ausscheidet.

. . . und ein Nicht-Paradoxon 3/4

Üblicherweise wird dieses Wissen durch Konditionierung eingearbeitet, d.h. es wird als posteriori-Verteilung die Verteilung $P_2 = P_1(\cdot|\neg \mathit{Peter})$ berechnet, für die man die folgenden Wahrscheinlichkeiten berechnet:

$$\begin{aligned} P_2(\mathit{Mary}) &= 0.5/0.75 = 0.67 \\ P_2(\mathit{Paul}) &= 0.25/0.75 = 0.33 \end{aligned}$$

Dies wird i.Allg. als unintuitiv empfunden, viele wuerden als angemessene Wahrscheinlichkeiten unter der Voraussetzung, dass Peter nicht der Täter sein kann, erwarten

$$P_2{}'(\mathit{Mary}) = 0.5 = P_2{}'(\mathit{Paul})$$

. . . und ein Nicht-Paradoxon 4/4

Tatsächlich lässt sich dieses Ergebnis auch durch eine (richtige ME-) Revisionsoperation erreichen:

Das durch die priori-Verteilung P_1 ausgedrückte Wissen wird nicht auf Peter *angewendet* (genau das leistet i.Allg. die Konditionierung als Focusing-Operation), sondern das Wissen über Peter muss auf der gleichen Ebene wie $\mathcal{R}_1$ berücksichtigt werden, da es sich auf den Moment der Tat bezieht – d.h. P_1 wird zu

$$P_1{}' = \mathit{ME}(\mathcal{R}_1 \cup \{\neg \mathit{Peter}[1]\})$$

$$\text{mit} \quad P_1{}'(\mathit{Mary}) = 0.5 = P_1{}'(\mathit{Paul})$$

Das Paradoxon entsteht nicht ursächlich durch die Probabilistik, sondern durch eine unsachgemäße Verarbeitung von Information bzw. Modellierung von Wissen!

Zusammenfassung Kapitel 4

- Bayessche Netze modellieren bedingte Unabhängigkeiten, während die *ME*-Methodik sich auf die konsequente Ausnutzung bedingter Abhängigkeiten konzentriert.
- *ME*-Inferenz und *ME*-Revision sind mächtige Methoden für die probabilistische Wissensrepräsentation mit hervorragenden Eigenschaften, die eine angemessene, flexible Wissensmodellierung erlauben.
- Durch Abstraktion lassen sich aus Wahrscheinlichkeitsfunktionen qualitative Rangfunktionen gewinnen, auf die sich wichtige Aspekte der *ME*-Prinzipien übertragen lassen ($\rightarrow$ c-Repräsentation, c-Revision).