OR Spektrum (1997) 19: 41-46

Léa Sombé und entropie-optimale Informationsverarbeitung mit der Expertensystem-Shell SPIRIT

SPIRIT and Lea Sombe: A study in probabilistic reasoning

Wilhelm Rödder, Gabriele Kern-lsberner

FB Wiwi, LG für BWL, insb. OR, FernUniversität Hagen, Kleine Strasse 22, D-58084 Hagen (Fax-Nr. 023311987-335);

Eingegangen am 4. Mai 1995/ Angenommen am 29. August 1996

Zusammenfassung. Die Repräsentation und Verarbeitung unsicheren Wissens in einer mathematisch korrekten und berechenbaren Weise ist eines der zentralen Anliegen der Künstlichen Intelligenz. Am Beispiel der schon berühmt gewordenen Studentin Léa Sombé aus [27] zeigen wir in dieser Arbeit, wie subjektives, unvollständiges Wissen in einer probabilistischen Wissensbasis dargestellt und als Grundlage für Inferenzen dienen kann. Dies wird durch entropie-optimalen Abgleich von Verteilungen erreicht. Diese in der Expertensystem-Shell SPIRIT implementierte Methodik zeichnet sich nicht nur durch gute Berechenbarkeit aus, sondern gestattet auch dank der hier aufgezeigten Kompatibilität mit der Konditionallogik eine logisch profunde Informationsverarbeitung.

Abstract. In this paper, the famous Léa Sombé-example of [27] is re-examined by methods combining cross-entropy minimization with probabilistic conditional logic. Thus a knowledge base is built up which allows easy computations and inferences in a logically sound way. The concept is realized by the probabilistic expert system shell SPIRIT which is presented here, too. So the aim of this paper affects as much practical aspects as it concerns logical foundations of knowledge representation. As the Sombé-example illustrates, even incomplete knowledge based on subjective probabilities or statistical data may be represented and dealt with adequately.

Schlüsselwörter: Künstliche Intelligenz, unsicheres Schließen, probabilistische Logik, konditionale Logik, Entropie

Key words: Artificial Intelligence, uncertain reasoning, probabilistic logic, conditional logic, entropy

1. Einleitung und Übersicht

In [27] machte es sich eine Gruppe von Wissenschaftlern zur Aufgabe, anhand eines allgemeinverständlichen Beispiels die wichtigsten Vertreter der heute bekannten und verwendeten unsicheren Logiken auf ihre Eigenschaften, ihre Stärken und Schwächen hin zu untersuchen. Hauptperson dieses Beispiels ist die Kunstfigur „Léa Sombé“, nach der die Gruppe sich auch benannte. Léa ist ein Individuum einer Population, über deren Zusammensetzung einige Abhängigkeiten bekannt sind, und aus denen nun neue Erkenntnisse über Léa abgeleitet werden sollen. Dieses soziologische Beispiel wird in Abschn. 2 genauer beschrieben, um Ausgangslage und Fragestellung dieses Artikels zu konkretisieren. Abschnitt 3 skizziert die Rolle der Entropie für statistische Inferenzen und beschreibt die Methodik, auf der die probabilistische Expertensystem-Shell SPIRIT, mit Hilfe derer das Léa-Sombé-Beispiel gerechnet wurde, basiert (vgl. auch [21]). Die Abschn. 4 und 5 liefern hierfür das logisch-theoretische Konzept: Wir stellen die Grundzüge einer unsicheren Logik vor, die eine besonders adäquate und leistungsfähige Behandlung des Falles „Léa Sombé“ erlaubt, nämlich eine Kombination aus probabilistischer Konditionallogik und entropie-optimalem Vorgehen. Beide Methoden sind für sich genommen schon lange bekannt und werden in ihren Bereichen erfolgreich angewendet, doch gerade die Einbindung des entropie-optimalen Aspektes in logische Inferenzstrukturen erlaubt quasi „maßgeschneiderte“ Wissensrepräsentation und Inferenzen, die eingegebenes Wissen so gut wie möglich verarbeiten, ohne dem Modell eine starre Struktur aufzuzwingen oder vom Benutzer die Spezifikation fast unübersehbar vieler numerischer Werte zu verlangen. Abschnitt 6 zeigt formelmäßig, wie sich entropie-optimale Anpassung vollzieht. Die daraus erkennbare Übereinstimmung mit einem wesentlichen Charakteristikum konventioneller Konditionallogik verdeutlicht das qualitativ-logische Fundament dieser Vorgehensweise. Die Nützlichkeit der kompliziert anmutenden Formeln (3) und (4) wird in Abschn. 7 demonstriert, wo mit ihrer Hilfe in kleinen Beispielen die entropie-optimale Verteilung explizit berechnet werden kann. In Abschn. 8 kehren wir schließlich zum Ausgangspunkt unserer Arbeit zurück: Am Beispiel der Studentin Léa Sombé wird demonstriert, wie sich mit Hilfe von SPIRIT entropie-optimales Schlußfolgern aus unsicherem *bzw.* unvollständigem Wissen auf elegante und benutzerfreundliche Weise realisieren läßt. Ausblicke auf laufende Forschungsvorhaben beschließen in Abschn. 9 diese Arbeit.

2. Die Studentin Léa Sombé

In unserer fiktiven Gesellschaft seien folgende Eigenschaften relevant:
S: Student sein, *J*: jung sein, *L*: ledig sein, *E*: Eltern sein, *H*: verheiratet sein, *Z*: unverheiratet mit jemandem zusammenleben,

wobei folgende Interdependenzen bekannt seien:

(r0) „Ca. 90% der Studenten sind jung."

(r1) „Ca. 80% aller jungen Leute sind ledig."

(r2) „Umgekehrt sind ca. 70% der Singles jung."

(r3) „30% der jungen Leute studieren."

(r4) „Gesellschaftsmitglieder, die unverheiratet mit jemandem zusammen leben, sind zu 80% jung."

(r5) „Studenten mit Kindern leben zu 90% in einer Ehe oder eheähnlichen Gemeinschaft."

Für den Aufbau der Wissensbasis spielt es keine Rolle, ob diese prozentualen Anteile geschätzt oder einer Statistik entnommen wurden. Im letzteren Fall kann man natürlich die groben Angaben durch exakte Daten ersetzen, wobei allerdings geringe Zahlenabweichungen kaum zu Buche schlagen. So mögen hier diese gerundeten Zahlen genügen, auch um die Anwendbarkeit des Systems auf subjektiv gewonnene Erkenntnisse (z. B. Expertenmeinungen) zu demonstrieren. Diese Angaben beschreiben das Léa Sombé-Beispiel (vgl. [27, S. 11]). Die dort angegebenen „certainty factors" im Stile MYCINs (vgl. [26]) haben wir hier kurzerhand als prozentuale Anteile interpretiert, da sie realistisch genug erschienen. Um Mißverständnisse auszuschließen, weisen wir ausdrücklich darauf hin, daß es sich nicht um Zugehörigkeitsgrade im Sinne der Fuzzy-Logik handelt (vgl. hierzu z.B. [29]). Die Vagheit des Prädikates "jung" ist nicht Gegenstand dieser Untersuchung.
Mit Hilfe eines entropie-optimalen Abgleichs (s.u.) wird eine probabilistische Wissensbasis erstellt, die z. B, zur Beantwortung der Fragestellung

> „Wenn (das Gesellschaftsmitglied) Léa Student ist und ein Kind hat, ist sie dann jung?"

herangezogen werden kann. Als Antwort wird hier eine Wahrscheinlichkeit erwartet, die den Grad der Sicherheit spezifiziert, mit dem wir – unter Berücksichtigung des tatsächlich vorhandenen Wissens und der Evidenzen – die Frage bejahen können. Bevor wir nun anhand dieses Beispiels die Arbeitsweise des Systems SPIRIT erläutern, soll die Rolle der Entropie aus informationstheoretischer, probabilistischer und logischer Sicht beleuchtet werden.

3. Entropie-optimale Informationsverarbeitung – das System SPIRIT

Der aus der Thermodynamik stammende Begriff der Entropie wird in der Informationstheorie als ein Maß für den Informationsgehalt einer Wahrscheinlichkeitsverteilung bzw. als Maß der Unterschiedlichkeit zweier Verteilungen (relative Entropie) verwendet:

Seien *P*, *Q* Verteilungen über denselben Aussagevariablen. Der Ausdruck
$H(P) = -\sum_{\omega} P(\omega)\log P(\omega)$ wird als die Entropie von P bezeichnet, und die relative Entropie von Q bezüglich P ist definiert als
$R(Q,P) = \sum_{\omega} P(\omega)\log \frac{Q(\omega)}{P(\omega)}$, wobei die Summe über alle Elementarereignisse ω genommen wird (vgl. z.B. [3, 9, 23, 24]).

Schon mehrere Autoren, die sich mit probabilistischer Logik befassen (z.B. [17, 18]), wiesen auf die Möglichkeit hin, entropie-optimale Modelle zur Ableitung neuen Wissens aus bekannten Fakten und Regeln zu benutzen, ohne daß jedoch eine eingehende Untersuchung und Darstellung dieser entropie-optimalen Inferenzstrukturen erfolgt wäre. Malvestuto [15] zeigte, daß die entropie-optimale gemeinsame Verteilung mehrerer Randverteilungen mit gleicher zugrundeliegender Variablenmenge in gewissen Fällen eine Produktdarstellung besitzt und bedingte Unabhängigkeiten anlegt. Ein solches System von Randverteilungen, zu dem eine gemeinsame Verteilung de existiert, wird *LEG-Netz (local event group network)* genannt (vgl. [12, 21]). Shore und Johnson wiesen in [25] nach, daß die (relative) Entropie im wesentlichen das einzige Funktional ist, dessen Optimierung vier fundamentale analytisch-probabilistische Inferenzbedingungen respektiert. Grob gesagt, sichert die entropie-optimale Anpassung die Einbindung neuer Information, ohne unnötige Abhängigkeiten anzulegen.

Im Unterschied zu den Bayes-Netzen, die seit den Arbeiten von Lauritzen und Spiegelhalter (s. z.B. [11, 28]) als eine sehr effektive Art probabilistischer Wissensrepräsentation geschätzt werden, sind bedingte Unabhängigkeiten und gerichtete (kausale) Abhängigkeiten keine zwingenden Voraussetzungen für die Erstellung eines entropie-optimalen LEG-Netzes. Das Léa Sombé-Beispiel zeigt, daß die Behandlung wechselseitiger, zyklischer Abhängigkeiten nicht eindeutig kausalen Ursprungs in dieser Umgebung problemlos möglich ist. Graphentheoretisch handelt es sich bei einem LEG-Netz um einen ungerichteten Hypergraphen. Seine Ausdruckskraft ist geringer als die eines gerichteten Bayes-Netzes, dafür gestattet es die Darstellung sowohl komplexer als auch diffuser, nicht näher spezifizierter Abhängigkeiten.

Auf diesem intensionalen Repräsentations- und Inferenzkonzept basiert das probabilistische Expertensystem SPIRIT, das in Abschn. 8 genauer beschrieben wird. Vorhandenes, quantifizierters Wissen um Zusammenhänge wird entropie-optimal in ein LEG-Netz propagiert. Diese Vorgehensweise ist intensional in dem Sinne, daß die Netzstruktur aus den Regeln aufgebaut wird und sich flexibel an neue Erkenntnisse anpassen läßt; unbekannte Wahrscheinlichkeiten stellen sich dem Indifferenzprinzip folgend ein (vgl. [8]). Damit erspart man dem Benutzer die Vorgabe einer starren Netzstruktur ebenso wie die u.U. recht heikle Spezifikation zahlreicher (bedingter) Wahrscheinlichkeiten, wie es das Bayes-Netz erfordert. Tatsächlich läßt sich „Léa Sombé" nur mit den obigen Angaben, ohne weitere Informationen hinzuzufügen, nicht angemessen in einem Bayes-Netz repräsentieren (s.

hierzu wieder [27, S. 107 ff]).

Wir werden im folgenden auf die logischen Aspekte entropie-optimaler Inferenzstrukturen eingehen und Parallelen zur Konditionallogik (vgl. z.B. [1, 2, 5, 7]) aufzeigen. Zunächst sollen Syntax und Semantik dieser sog. *ME-Logik* (*ME* = maximal entropy oder minimal cross-entropy) im Rahmen einer probabilistischen Konditionallogik festgelegt werden.

4. Syntax und Semantik entropie-optimaler Logik (ME-Logik)

Wir unterscheiden syntaktisch zwischen ***Fakten*** und ***Regeln***: ***Fakten*** sind aussagenlogische Formeln, die als Konnektive lediglich $\wedge$, $\vee$ und $\neg$ enthalten (keine Implikation); (***unsichere***) ***Regeln*** sind Ausdrücke der Form $F_1 \xrightarrow{\approx} F_2$, wobei F_1, F_2 Fakten sind. ***Probabilistische Fakten*** bzw. ***probabilistische Regeln*** werden mit einer Wahrscheinlichkeit bzw. Schlußstärke $x \in [0,1]$ versehen.

Die semantischen Strukturen sind Populationen, auf denen gewisse Merkmale betrachtet werden bzw. Wahrscheinlichkeitsräume über entsprechenden propositionalen Variablen. Ist P eine Verteilung über eine Menge $V = \{A_1,..., A_n\}$ solcher Variabler, so bezeichnen wir mit p deren Wahrscheinlichkeitsfunktion $p(a_1,..., a_n)$ = $P(A_1 = a_1,..., A_n = a_n)$ wobei kleine Buchstaben die Ausprägungen der Variablen repräsentieren. Mod ($\triangle$) sei die Menge aller Wahrscheinlichkeitsverteilungen über n (binären) Aussagevariablen. Sei $P \in$Mod ($\triangle$), seien F, F_1, F_2 Fakten, die in einer geeigneten logischen Sprache formuliert sind.

Das (probabilistische) Faktum ***F(x) gilt in P***, in Zeichen: $P \models F(x)$, wenn $p(F) = x$ ist.

Die (probabilistische) Regel $F_1 \xrightarrow{\approx} F_2(x)$ ***gilt in P***, in Zeichen: $P \models F_1 \xrightarrow{\approx} F_2(x)$, wenn $p(F_2|F_1)$=x ist; das Implikationszeichen wird also als bedingte Wahrscheinlichkeit interpretiert. Insbesondere läßt sich jedes (probabilistische) Faktum als Regel mit tautologischer Prämisse auffassen.

Damit sind Syntax und Semantik einer probabilistischen Logik erklärt. Wenden wir uns nun dem Ableitungsbegriff zu: Welche Möglichkeiten einer „probabilistischen Folgerung" bieten sich an ?

5. Probabilistisches Folgern

Besteht die Wissensbasis aus einer (gegebenen) Verteilung $P \in$ Mod ($\triangle$), so ist klar, welche (probabilistischen) Aussagen wir aus P „folgern" können -nämlich genau diejenigen, die in P gültig sind. Der semantische Ableitungsoperator wird also wie üblich durch den Gültigkeitsbegriff initiiert. $Th(P) = \{rf(x) | P \models rf(x)\}$ sei die Menge aller in P gültigen bzw. aus P direkt ableitbaren probabilistischen Aussagen, wobei rf abkürzend für „Regel oder Faktum" stehe. Für das Folgende vgl. auch Makinsons Abhandlung über kumulative Inferenz [14].

Sei nun $\measuredangle$ eine Menge probabilistischer Fakten und Regeln.

Mit Mod($\measuredangle$) = $\{P \in \mathrm{Mod}(\triangle) \mid P \models \measuredangle\}$ bezeichnen wir die Menge aller probabilistischen Modelle von $\measuredangle$ d.h. aller Verteilungen (passenden Typs), in denen $\measuredangle$ gültig ist. Dem klassischen semantischen Ableitungsbegriff folgend können wir nun festsetzen:

$\measuredangle \models 0 \;\; rf(x)$ genau dann wenn $\mathrm{Mod}(\measuredangle) \subset \mathrm{Mod}(rf(x))$. Diese Art der Folgerung entspricht dem, was man unter dem eigentlichen probabilistischen Folgern versteht – Folgern nur unter Ausnutzung der Gesetze der Wahrscheinlichkeitstheorie. Die starke Bedingung $\mathrm{Mod}(\measuredangle) \subset \mathrm{Mod}(rf(x))$, die alle Modelle von $\measuredangle$ als Basis für eine Folgerung einbezieht, zeitigt allerdings einen recht schwachen Ableitungsbegriff. Der Gedanke liegt nahe, lediglich eine Teilmenge von Mod($\measuredangle$) oder sogar nur ein ausgewähltes Modell als Grundlage für Folgerungen zu nehmen. Ein solches Modell, das keine unnötigen Abhängigkeiten anlegt und in besonderer Weise die Bildung bedingter Wahrscheinlichkeiten respektiert, ist das entropie-maximale Modell ⁝*($\measuredangle$) (s. z. B. [25]), d. h. ⁝*($\measuredangle$) ist Lösung des folgenden Optimierungsproblems

$$\text{Max } H(Q) = -\sum_{\omega} Q(\omega) \log Q(\omega) \qquad (1)$$

unter der Nebenbedingung

Q ist Verteilung mit $Q \models \measuredangle$,

und wir schreiben $\measuredangle \models * \, rf(x)$ genau dann wenn ⁝*($\measuredangle$) $\models rf(x)$.

Wir betrachten im folgenden ⁝*($\measuredangle$) als Basis für probabilistische Inferenzen aus dem Regelsystem $\measuredangle$. Liegt außerdem noch apriori-Information in Form einer gegebenen Verteilung P vor, die an eine Menge $\measuredangle$ zusätzlich verfügbarer Regeln möglichst gut angepaßt werden soll, so haben wir das Optimierungsproblem

$$\text{Min } R(Q, P) = \sum_{\omega} Q(\omega) \log \frac{Q(\omega)}{P(\omega)} \qquad (2)$$

unter der gleichen Nebenbedingung wie oben zu lösen; dessen Lösung wollen wir mit ⁝*($\measuredangle$;P) bezeichnen. Wie man leicht nachrechnet, entspricht einer Maximierung der Entropie eine Minimierung der relativen Entropie zur Gleichverteilung, so daß das Minimierungsproblem (2) als die allgemeinere Aufgabenstellung angesehen werden kann.

Zur Berechnung von ⁝*($\measuredangle$) bzw. ⁝*($\measuredangle$;P) kann das aus der Statistik bekannte IPS- Verfahren (*i*terative *p*roportional *s*caling) herangezogen werden (vgl. z.B. [10]); damit ist die entsprechende Inferenzmaschine auf dem Rechner verfügbar. Wie aber läßt sich die entropieoptimale

Verteilung in Abhängigkeit von den eingegebenen Regeln und der evtl. vorliegenden apriori-Information darstellen?

6. Darstellung entropie-optimaler Wissensrepräsentation

Sei $P \in \mathrm{Mod}(\triangle)$, seien $r_i : F_{1i} \xrightarrow{\approx} F_{2i}(x_i), 1 \le i \le m$ probabilistische Regeln, an die die priori-Verteilung P entropie-optimal angepaßt werden soll. Aus den Kuhn-Tucker-Bedingungen dieses Optimierungsproblems sieht man, daß die entropie-optimale Lösung P* notwendig von der Form

$$p^*(\omega) = \alpha_0 P(\omega) \prod_{\substack{i: F_{1i}(\omega)=1,\\ F_{2i}(\omega)=1}} \alpha_i^{x_i - 1} \prod_{\substack{i: F_{1i}(\omega)=1,\\ F_{2i}(\omega)=0}} \alpha_i^{x_i} \tag{3}$$

ist, wobei die ω Elementarereignisse repräsentieren und die α_i nicht-negative reelle Zahlen sind (vgl. auch [22]). Wegen $p^*(F_{1i} | F_{2i}) = x_i$, $1 \le i \le m$ sind die Faktoren α_i für $i > 0$ Lösungen des (nichtlinearen) Gleichungssystems

$$\alpha_i = \frac{1 - x_i}{x_i} \tag{4}$$

$$\frac{\sum\limits_{\omega: F_{1i} F_{2i}(\omega)=1} p(\omega) \prod\limits_{(j \ne i, F_{1j} F_{2j}(\omega)=1)} \alpha_j^{x_j} \prod\limits_{(j \ne i, F_{1j} F_{2j}(\omega)=1)} \alpha_j^{x_j - 1}}{\sum\limits_{\omega: F_{1i} \overline{F}_{2i}(\omega)=1} p(\omega) \prod\limits_{(j \ne i, F_{1j} \overline{F}_{2j}(\omega)=1)} \alpha_j^{x_j} \prod\limits_{(j \ne i, F_{1j} F_{2j}(\omega)=1)} \alpha_j^{x_j - 1}}$$

und α_0 ergibt sich als Normierungsfaktor. Die α_i symbolisieren die durch die eingebrachten Regeln bewirkten Veränderungen in P. Ungeachtet des komplizierten Ausdrucks erkennt man aus (3) sehr deutlich, daß die Anpassung an eine Regel (abgesehen vom Normierungsfaktor) grundsätzlich nur die Wahrscheinlichkeitswerte derjenigen Elementarereignisse ändert, die die Prämisse erfüllen. Dabei gibt es zwei Änderungsvarianten, je nachdem, ob das Elementarereignis die Conclusio erfüllt oder nicht. Damit respektiert die entropie-optimale Propagation von Regeln die inhärente Dreiwertigkeit der Konditionallogik (vgl. z.B. [2, 4]), in der ein Konditional eine verallgemeinerte Indikatorfunktion (b | a) mit den Werten

$$(b|a)(\omega) = \begin{cases} 1, & \text{falls } \omega \in ab \\ 0, & \text{falls } \omega \in a\overline{b} \\ \text{undefiniert}, & \text{falls } \omega \notin a \end{cases}$$

ist. Dies ist ein weiterer wesentlicher Unterschied zur Arbeit mit Bayes-Netzen, die vom Benutzer bei Festlegung einer Abhängigkeit $p(b|a) = x$ auch die detaillierte Spezifikation aller Wahrscheinlichkeiten $p(b|\overline{a})$ verlangen, in denen eine angemessene Modellierung von Nichtwissen also nicht möglich ist.

Das Gleichungssystem (4) ist nur in Ausnahmefällen geschlossen lösbar, seine Komplexität verdeutlicht die Verflechtungen der Regelwirkungen.

Bevor wir den Fall „Léa Sombé“ in diesem probabilistisch-logischen Rahmen darstellen, soll die Wirkungsweise entropie-optimaler Anpassung anhand einfacher, aber typischer Beispiele verdeutlicht werden.

7. Beispiele zur Propagation von Fakten und Regeln

Sei $P = P(A, B)$ eine Verteilung über 2 binäre Variable mit den Werten $p_1 = p(a\,b)$, $p_2 = p(a\,\overline{b})$, $p_3 = p(\overline{a}\,b)$, $p_4 = p(\overline{a}\,\overline{b})$.

Beispiel 1. Propagation einer Regel. In diese Verteilung P soll nun die Regel $a \xrightarrow{\approx} b(x)$ propagiert werden. (3) liefert für $P^* = \vdots^*(a \xrightarrow{\approx} b(x); P)$

$$p^*(a\,b) = \alpha_0 \alpha^{x-1} p_1, \qquad p^*(\overline{a}\,b) = \alpha_0 p_3,$$
$$p^*(a\,\overline{b}) = \alpha_0 \alpha^{x} p_2, \qquad p^*(\overline{a}\overline{b}) = \alpha_0 p_4.$$

Dann ist

$$\frac{1-x}{x} = \frac{p^*(a\,\overline{b})}{p^*(a\,b)} = \frac{\alpha_0 \alpha^x p_2}{\alpha_0 \alpha^{x-1} p_1} = \alpha \frac{p^*(a\,\overline{b})}{p^*(a\,b)},$$

also $\alpha = \frac{1-x}{x} \div \frac{p^*(a\,\overline{b})}{p^*(a\,b)}$ und $\alpha_0 = \frac{p^*(\overline{a})}{p(\overline{a})}$,

d. h. α ist das Kreuzungsverhältnis von aposteriori–Verhältnis zu apriori-Verhältnis, und der Normierungsfaktor α_0 ist der Ausgleich der durch die Regel nicht betroffenen Menge $\overline{a}$.

Beispiel 2. Propagation eines Faktums. In die Verteilung P soll das Faktum $b(y)$ eingebracht werden. Da ein Faktum als Regel mit tautologischer Prämisse aufgefaßt wird, werden in diesem Fall die Wahrscheinlichkeiten aller Elementarereignisse von der Propagation berührt. Man erhält – wieder mit (3) – für die entropie-optimale Verteilung P*

$$p^*(a\,b) = \alpha_0 \alpha^{y-1} p_1 \qquad p^*(\overline{a}\,b) = \alpha_0 \alpha^{y-1} p_3$$
$$p^*(a\,\overline{b}) = \alpha_0 \alpha^{y} p_2 \qquad p^*(\overline{a}\,\overline{b}) = \alpha_0 \alpha^{y} p_4$$

wobei nun

$$\alpha = \frac{1-y}{y} \cdot \frac{p(b)}{p(\overline{b})} \quad \text{und} \quad \alpha_0 = \alpha^{1-y} \frac{y}{p(b)}$$

sind. Damit vereinfacht sich P* zu

$$p^*(a\,b) = p_1 \frac{y}{p(b)}, \qquad p^*(\overline{a}\,b) = p_3 \frac{y}{p(b)},$$
$$p^*(a\,\overline{b}) = p_2 \frac{1-y}{p(\overline{b})}, \qquad p^*(\overline{a}\,\overline{b}) = p_4 \frac{1-y}{p(\overline{b})},$$

d.h. die entropie-optimale Propagation eines Faktums liefert als Ergebnis dieselbe Verteilung wie die bedingungserhaltende Anpassung an eine geänderte Randverteilung (vgl. [19]). Dieselbe Formel stellen J.F. Lemmer und S.W. Barth in ihrer Arbeit über LEG-Netze als Verallgemeinerung der Bayesschen Inferenz vor (vgl. [13]).

8. SPIRIT und das Beispiel Léa Sombé

SPIRIT ist eine in C ++ programmierte Expertensystem-Shell, die die bisher theoretisch dargestellten Schritte

- Eingabe einer Menge ∡ von Fakten und Regeln mit geforderten Eintrittswahrscheinlichkeiten;
- Aufbau der Wahrscheinlichkeitsverteilung P* mittels Randverteilungen zu einem LEG-Netz;

- Propagationen von Evidenzen durch das LEG-Netz;
- Auswertung von als Anfrage formulierten Fakten und Regeln

in komfortabler Weise unter WINDOWS gestattet.

Der Anzahl zu erlernender Fakten und Regeln sind prinzipiell keine Grenzen gesetzt. Beschränkungen werden nur dann notwendig, wenn die Komplexität des Regelwerks das Anlegen einer Randverteilung erfordert, deren Mächtigkeit 2^{15} übersteigt.

Die Expertensystem-Shell ist menügesteuert, mehrsprachige Hilfe-Funktionen unterstützen den Benutzer. Sie wurde bereits in Wissensdomänen mit bis zu 30 involvierten Variablen eingesetzt (vgl. [16]). Wesentliche Grundzüge der Arbeitsweise des Systems lassen sich jedoch besser an überschaubaren Beispielen studieren. Wir kehren daher zu Léa Sombé und ihrem sozialen Umfeld zurück.

In der in Abschn. 4 beschriebenen Syntax haben die Aussagen (r0)-(r5) aus Abschn. 2 die folgende Form:

(r0) $s \xrightarrow{\approx} j$ (0.9)

(r1) $j \xrightarrow{\approx} l$ (0.8)

(r2) $l \xrightarrow{\approx} j$ (0.7)

(r3) $j \xrightarrow{\approx} s$ (0.3)

(r4) $z \xrightarrow{\approx} j$ (0.8)

(r5) $es \xrightarrow{\approx} \bar{l}$ (0.9)

Außerdem muß dem System noch mitgeteilt werden, daß die Variablen L, H und Z disjunkt und vollständig alle möglichen Ausprägungen des Attributes „Familienstand“ repräsentieren:

(f0) $l \dot{\vee} h \dot{\vee} z = 1$.

Dies geschieht in SPIRIT durch Anlegen einer dreiwertigen Variablen „Familienstand“ (F). Die anderen Variablen S, J und E sind binär. Damit ist $\measuredangle$ = {(r0), ..., (r5), (f0)} und $P^* = ⁝^*(\measuredangle)$ kann in der oben beschriebenen Weise berechnet werden. So erhält man eine Wissensbasis in Form eines LEG-Netzes, das durch die eingegebenen Regeln initiierte statistische Abhängigkeiten zwischen den beobachteten Merkmalen der Population widerspiegelt. Neue Erkenntnisse lassen sich nun leicht durch entsprechende Berechnungen ableiten. So liefert z.B. die bedingte Wahrscheinlichkeit $p^*(j \mid e\, s)$ die Antwort auf die eingangs gestellte Frage

> „Wenn Léa Student ist und ein Kind hat, ist sie dann jung?“

Dabei wird angenommen, daß Léa ein typisches Mitglied der Population ist, von dem lediglich bekannt ist, daß es studiert und ein Kind hat. Diese Art der Wissensdeduktion durch Abgleich von Evidenzen wird in [6] ausführlich als einsichtig und formal korrekt beschrieben.

Wir erhalten nun

$$P^* = M^*(R) \Rightarrow es \xrightarrow{\approx} j(0.8127),$$

d.h. Léa – als typisches Individuum mit den Eigenschaften „Eltern sein“ und „Student sein“ – ist mit einer Wahrscheinlichkeit von ca. 80% jung.

Abbildung 1 zeigt das mit SPIRIT berechnete und visualisierte Modell bei gleichzeitiger Evidenz (Instantiierung) von E und S. Das LEG-Netz ist in diesem Fall sehr klein, es besteht nur aus den beiden LEGs {E, S, F} und {S, F, J}. Nur aus Gründen der Übersichtlichkeit wurde eine Darstellung des Modells als Graph und nicht als Hypergraph gewählt.

Ebenso leicht lassen sich beliebige andere Fragestellungen, die als logische Fakten bzw. Regeln formalisiert werden können, beantworten. Wir erfahren z.B. außerdem, daß ca. 78% der Studenten kinderlos sind ($s \xrightarrow{\approx} \neg e(0.775)$) und daß ca. 84% von ihnen auch nicht verheiratet sind ($s \xrightarrow{\approx} \neg h(0.838)$). Umgekehrt studiert nur ca. 14% der verheirateten Bevölkerung ($h \xrightarrow{\approx} \neg s(0.860)$) Weiterhin bestätigt $⁝^*(\measuredangle)$ die Vermutung, daß junge Leute meistens nicht verheiratet sind ($j \xrightarrow{\approx} \neg h(0.928)$).

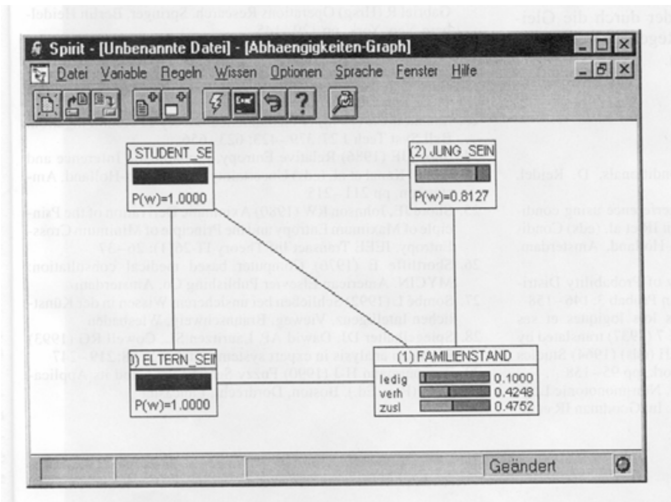

Abbildung 1. Das Léa Sombé-Beispiel in SPIRIT

Die beschriebene Vorgehensweise gestattet es, aus gegebenen Regeln eine gemeinsame Verteilung zu berechnen, die für weitere Anfragen über das zu behandelnde Universum als recht bequeme und gut handhabbare Wissensbasis zur Verfügung steht. In [20] wird demgegenüber ein Verfahren vorgestellt, das umgekehrt durch automatische Aggregation von Elementarkonjunktionen dem Benutzer eine Übersicht über wichtige, in einer Verteilung gültige Fakten und Regeln eines beliebigen Wahrscheinlichkeitsniveaus verschafft.

9. Einige abschließende Bemerkungen

Mit der Anwendung des in dieser Arbeit beschriebenen entropie-optimalen Konzeptes steht eine Art der Wissensrepräsentation und -verarbeitung zur Verfügung, die eine kompakte Handhabbarkeit auf dem Rechner mit einer soliden probabilistisch-logischen Grundlage verbindet und in dem vorgestellten System SPIRIT realisiert wurde. Die Modellierung des Problems erfolgt durch die Eingabe probabilistischer Regeln und Fakten, aus denen SPIRIT eigenständig ein LEG-Netz aufbaut. Dessen Darstellung durch einen Graphen visualisiert quantitative Abhängigkeitsbeziehungen. Die Bearbeitung von Anfragen seitens des Benutzers ist durch die Aktivierung der entsprechenden Knoten im Netz denkbar einfach (s. Abb. 1), daneben ist auch die probabilistische Auswertung komplexer logischer Ausdrücke im *Query editor* möglich.

An der optimalen und effizienten Anpassung der LEG-Netzwerkstrukturen an vorliegende Informationen wird weiterhin intensiv gearbeitet. In theoretischer Hinsicht sind die Vertiefung der oben angesprochenen Verbindung zwischen entropie-optimalem Vorgehen und konditionaler Logik sowie eine Beschreibung der durch die Gleichungen (3) und (4) ausgedrückten Regelverflechtungen Themen aktueller Forschungsarbeiten.

Literatur

1. Adams EW (1975) The Logic of Conditionals. D. Reidel, Dordrecht
2. Calabrese PG (1991) Deduction and interference using conditional logic and probability. In: Goodman IR et al. (eds) Conditional Logic in Expert Systems. North-Holland, Amsterdam, pp 71-100
3. Csiszar I (1975) I-divergence Geometry of Probability Distributions and Minimization Problems. Ann Probab 3: 146-158
4. De Finetti B (1937) La prevision, ses lois logiques et ses sources subjectives. Ann Inst H Poincare 7 (1937) translated by Kyburg H Jr. In: Kyburg H Jr., Smokler H (eds) (1964) Studies in Subjective Probability. Wiley, New York, pp 95-158
5. Dubois D, Prade H (1991) Conditioning, Nonmonotonic Logic and Nonstandard Uncertainty Models. In: Goodman IR et al. (eds), Conditional Logic in Expert Systems. North-Holland, Amsterdam, pp 115-158
6. Gerla G (1994) The Probability that Tweety is Able to Fly. Int J Intelligent Systems 9:403-409
7. Goodman IR, Nguyen HT (1988) Conditional objects and the modeling of uncertainties. In: Gupta MM, Yamakawa T (eds) Fuzzy Computing. North-Holland, Amsterdam, pp 119-138
8. Jaynes ET (1983) Papers on Probability, Statistics and Statistical Physics. Rosenkrantz RD (ed) D. Reidel, Dordrecht Holland
9. Kulback S (1968) Information Theory and Statistics. Dover, New York
10. Lauritzen SL (1982) Lectures on Contingency Tables. Aalborg University Press, Denmark
11. Lauritzen SL, Spiegelhalter DJ (1988) Local Computations with Probabilities in Graphical Structures and Their Applications to Expert Systems. J R Statist Soc B 50(2): 415-448
12. Lernmer J F (1983) Generalized Bayesian updating of incompletely specified distributions. Large Scale Syst 5: 51-68
13. Lernmer JF, Barth SW (1982) Efficient minimum information updating for Bayesian inferencing in expert systems. In: Proceedings of the National Conference on Artificial Intelligence. Morgan Kauffmann, Pittsburgh, PA, Los Altos, CA, pp 424-427
14. Makinson D (1989) General Theory of Cumulative Inference. In: Reinfrank M (ed) Nonmonotonic Reasoning. Springer, Ber- lin, pp 1-18
15. Malvestuto FM (1986) Decomposing complex contingency tables to reduce storage requirements. In: Proc. 3rd Intern. Workshop on Scientific and Statistical Database Management, pp 66- 71
16. Meyer CH, Kern-Isberner G, Rödder W (1996) Analyse medizinisch-soziologischer Daten mittels eines probabilistischen Expertensystems. In: Kleinschmid P et al. (Hrsg) Operations Re- search Proceedings 1995. Springer, Berlin et al., pp 347 -352
17. Nilsson NJ (1986) Probabilistic Logic. Artificial Intelligence 28(1): 71-87
18. Pearl J (1988) Probabilistic Reasoning in Intelligent Systems. Morgan Kauffmann, San Mateo, CA
19. Reidmacher HP (1994) Logisches Schließen bei Unsicherheit. Europäische Hochschulschriften, Reihe 5, Bd. 1519, Frankfurt am Main
20. Reidmacher HP, Kern-Isberner G (1996) Interpreting a Contingency Table by Rules. Int J Intelligent Syst 11: 327-346
21. Rödder W (1994) Symmetrical Probabilistic Intensional Reasoning in Inference Networks in Transition. In: Werners B, Gabriel R (Hrsg) Operations Research. Springer, Berlin Heidelberg New York, pp 129-145
22. Rödder W, Longgui X (1993) SPIRIT- Die Behandlung logischer Funktionen in einer probabilistischen Wissensbasis. In: Hansmann K-W et al. (Hrsg.) Operations Research Proceedings 1992. Springer, Berlin et al., pp 462-469
23. Shannon CE (1948) A mathematical theory of communication. Bell Syst Tech J 27: 379-423: 623-656.
24. Shore JE (1986) Relative Entropy, Probabilistic Inference and AI. In: Kanal et al. (eds) Uncertainty in AI. North-Holland, Amsterdam, pp 211-215
25. Shore JE, Johnson RW (1980) Axiomatic Derivation of the Principle of Maximum Entropy and the Principle of Minimum Cross- Entropy. IEEE Transact Inf Theory IT-26(1): 26-37
26. Shortliffe E (1976) Computer based medical consultation: MYCIN. American Elsevier Publishing Co, Amsterdarn
27. Sombe L (1992) Schließen bei unsicherem Wissen in der Künstlichen Intelligenz. Vieweg, Braunschweig, Wiesbaden
28. Spiegelhalter DJ, Dawid AP, Lauritzen SL, Cowell RG (1993) Bayesian analysis in expert systems. Statist Sci 8:219-247
29. Zimmermann H-J (1990) Fuzzy Set Theory -and its Applications (Rev. Ed.). Boston, Dordrecht, Lancaster