ZUR LOGIK VON BESTENLISTEN IN DER DIALOGMODELLIERUNG

Günther Wirsching[1], *Christian Kölbl*[2], *Markus Huber*[2]

[1]*Universität Eichstätt-Ingolstadt*, [2]*Universität Augsburg*
vorname.nachname@{ku-eichstaett|uni-augsburg}.de

Kurzfassung: Die Verwendung endlicher gewichteter Transduktoren als semantischen Träger setzt eine logisch-wahrscheinlichkeitstheoretische Analyse von Ergebnissen automatischer Spracherkennung voraus. In diesem Beitrag wird gezeigt, wie man aus einer Bestenliste, bei der zu jeder Erkennalternativeeine Semantik und ein Score mitgeliefert werden, Merkmalskonfidenzen und Wahrscheinlichkeitsfunktionen berechnen kann, die dann geeignete Gewichte für Transduktoren zur weiteren semantischen Analyse ergeben.

1 Einleitung

In der Ausagenlogik ordnet man jeder *Aussage* einen Wahrheitswert zu. Mögliche Wahrheitswerte sind *wahr* und *falsch*; in der Modallogik werden noch Werte wie *möglicherweise* oder *wahrscheinlich* oder ähnliche hinzugenommen. Wir bauen unsere Logik auf einer etwas anderen Situation, wie sie typisch ist für Ergebnisse automatischer Spracherkennung, auf: man erhält als Erkennergebnis eine *Bestenliste* bestehend aus einzelnen *Erkennalternativen*. wobei wir zusätzlich annehmen, dass zu jeder Erkennalternative eine *Semantik* und ein *Score* mitgeliefert werden. Dabei setzen wir voraus, dass die Semantik als Menge von *Merkmal-Wert-Paaren* [3] dargestellt ist, und dass sich die Scores zu Eins summieren.

In diesem Beitrag beschreiben wir eine logisch-wahrscheinlichkeitstheoretische Analyse von Bestenlisten, die eine Bestimmung von Merkmalskonfidenzen und Wahrscheinlichkeitsfunktionen aus den Merkmalstrukturen und den Scores erlaubt. Zusammen mit den in [4] beschriebenen Methoden ermöglicht das den Einsatz endlicher gewichteter Transduktoren als semantischen Träger.

2 Attributsabbildung und Wahrscheinlichkeitsvektoren

Nehmen wir an, das Ziel sei die eindeutige Identifikation einer Person, in einer Datenbank, wobei mehrere Personen den gleichen Nachnamen haben können. Nehmen wir weiter an, eine Bestenliste bestehe aus einer Anzahl möglicher Nachnamen mit Scores. Wie lässt sich nun aus einer (wahrscheinlichkeitstheoretischen) Information über einen Nachnamen Information zur Identifkation einer Person gewinnen? – Die hier angewandte Methode ist, eine Wahrscheinlichkeitsfunktion auf der Menge der Nachnamen entlang der Abbildung, die jeder Person ihren Nachnamen zuordnet, „zurückzuziehen“. Dieser Abschnitt ist einer mathematischen Beschreibung und Analyse dieses „Zurückziehens“ gewidmet.

2.1 Notationen

2.1.1 Attributsabbildung

Unser Ausgangspunkt sind die durch die Datenbank gegebenen Zuordnungen, z. B.:

$$\text{Datensatz} \quad \longmapsto \quad \text{Nachname.}$$

Etwas formaler und allgemeiner: Bezeichne D eine (endliche) Menge von *Datensätzen*, A ein *Attribut* und W_A die Menge möglicher Werte, die das Attribut A annehmen kann. Durch eine Datenbank ist dann eine *Attributsabbildung* zum Attribut A gegeben:

$$\phi_A : D \to W_A \tag{1}$$

ordnet jedem Datensatz $d \in D$ den Wert $\phi_A(d) \in W_A$ seines Attributs A zu. Handelt es sich z. B. um eine Personendatenbank, und bezeichnet A das Attribut *Nachname*, so ordnet ϕ_A jeder Person ihren Nachnamen zu.

2.1.2 Wahrscheinlichkeitsvektoren

Unter einer *Wahrscheinlichkeitsfunktion* auf einer endlichen Menge M verstehen wir eine Abbildung

$$p : M \to [0,1] \qquad \text{mit der Eigenschaft} \quad \sum_{x \in M} p(x) = 1. \tag{2}$$

Sind die Elemente von M nummeriert, $M = \{x_1, \ldots, x_n\}$ mit paarweise verschiedenen x_i, so kann man, indem man $p_i := p(x_i)$ für alle $i \in \{1, \ldots n\}$ setzt, die Wahrscheinlichkeitsfunktion auch als *Wahrscheinlichkeitsvektor* $p = (p_1, \ldots, p_n) \in \mathbb{R}^n$ auffassen. Die Menge der Wahrscheinlichkeitsvektoren ist der *Standard-Simplex* im Euklidischen Raum, den wir wie folgt bezeichnen:

$$\Delta^n := \left\{ (p_1, \ldots, p_n) \in \mathbb{R}^n \,\middle|\, \text{alle } p_i \geq 0, \sum_{i=1}^{n} p_i = 1 \right\}. \tag{3}$$

Außerdem bezeichnen wir den *offenen Simplex* als:

$$\Delta^n_+ := \left\{ (p_1, \ldots, p_n) \in \mathbb{R}^n \,\middle|\, \text{alle } p_i > 0, \sum_{i=1}^{n} p_i = 1 \right\}, \tag{4}$$

Des weiteren benutzen wir die *Normierungsabbildung*, die einem n-dimensionalen Vektor $v = (v_1, \ldots, v_n)$ mit nicht-negativen Koordinaten $v_i \in \mathbb{R}_+ = [0, \infty[$ mit Ausnahme des Nullvektors einen Wahrscheinlichkeitsvektor zuordnet:

$$\mathscr{N} : \mathbb{R}^n_+ \setminus \{0\} \to \Delta^n, \qquad \mathscr{N}(v_1, \ldots, v_n) = \left(\frac{v_1}{v_1 + \ldots + v_n}, \ldots, \frac{v_n}{v_1 + \ldots + v_n} \right). \tag{5}$$

Wir werden die Normierungsabbildung auch auf beliebige Funktionen $f : M \to [0, \infty[$, die auf einer endlichen Menge M definiert sind und an wenigstens einer Stelle $x \in M$ nicht verschwinden, anwenden:

$$\mathscr{N} f : M \to [0,1], \qquad \mathscr{N} f(x) := \frac{f(x)}{\sum_{\xi \in M} f(\xi)}. \tag{6}$$

2.2 Zurückziehen einer Wahrscheinlichkeitsfunktion

Ein naheliegender Ansatz zum Zurückziehen einer Wahrscheinlichkeitsfunktion $p : Y \to [0,1]$ entlang einer Abbildung $\phi : X \to Y$ ist die Hintereinanderschaltung

$$p \circ \phi : X \to [0,1], \quad x \mapsto p(\phi(x)). \tag{7}$$

Allerdings ist $p \circ \phi$ nicht in jedem Fall wieder eine Wahrscheinlichkeitsfunktion, da die Normierungsbedingung in (2) nicht unbedingt erfüllt ist; auch der Fall $p \circ \phi \equiv 0$ kann vorkommen, wenn nämlich alle $p(\phi(x)) = 0$ sind.

Jede der beiden folgenden Bedingungen ist hinreichend für die Normierbarkeit von $p \circ \phi$:

1. Die Wahrscheinlichkeitsfunktion $p : Y \to [0,1]$ besitzt keine Nullstelle.
2. ϕ ist surjektiv.

Ist eine dieser Bedingungen erfüllt, so ist die auf X *zurückgezogene* Wahrscheinlichkeitsfunktion definiert durch

$$\phi^* p := \mathscr{N}(p \circ \phi) : X \to [0,1]. \tag{8}$$

2.3 Zur Surjektivität von Attributsabbildungen

Im Fall automatischer Spracherkennung ist die Attributsabbildung $\phi_A : D \to W_A$ jedenfalls dann surjektiv, wenn dem Spracherkenner genau diejenigen Nachnamen als Muster vorliegen, die auch in der Datenbank vorkommen.

Es ist aber auch möglich, einen Spracherkenner so zu konfigurieren, dass Worte außerhalb einer definierten Liste, sogenannte *oov*-Worte (*out of vocabulary*), als solche erkannt werden können. Technisch hat man auch in diesem Fall eine surjektive Abbildung $\phi_A : \widehat{D} \to \widehat{W_A}$. Ihr Definitionsbereich $\widehat{D}$ enthält außer D noch eine geeignete Anzahl von „leeren" Datensätzen, und ihr Wertebereich besteht aus W_A aus der Datenbank und einigen zusätzlichen oov-Modellen, die als Bilder der „leeren" Datensätze betrachtet werden. Der Spracherkenner liefert im Erkennergebnis eine Wahrscheinlichkeitsverteilung auf dem erweiterten Wertebereich $\widehat{W_A}$, auf deren Grundlage das Dialogsystem dann entscheiden kann, ob der Benutzer einen neuen Datensatz anlegen möchte.

3 Das logische Oder

Unser nächstes Ziel ist, eine sinnvolle Definition für das *logische Oder* zweier Wahrscheinlichkeitsfunktionen mit demselben Definitionsbereich zu finden. Hierzu werfen wir zunächst einmal einen Blick auf das allgemeine Verhalten von Wahrscheinlichkeiten, wenn man die zugrunde liegenden Ereignisse mit *Oder* verknüpft.

3.1 Hintergrund und Formeln

3.1.1 Die wahrscheinlichkeitstheoretische Sicht

Seien D eine endliche Menge und $p : D \to [0,1]$ eine darauf definierte Wahrscheinlichkeitsfunktion, Nennt man eine Teilmenge $A \subseteq D$ ein *Ereignis*, so ordnet p jedem Ereignis seine *Wahrscheinlichkeit*

$$p(A) = \sum_{x \in A} p(x) \in [0,1] \tag{9}$$

zu. Für die Vereinigung zweier Ereignisse $A, B \subseteq D$, die dem *logischen Oder* der Element-Beziehung entspricht, gilt die Formel

$$p(A \cup B) = p(A) + p(B) - p(A \cap B). \tag{10}$$

Uns interessieren nun drei Spezialfälle:

1. Disjunkte Ereignisse. Hier ist $A \cap B = \varnothing$, also $p(A \cup B) = p(A) + p(B)$.
2. Stochastisch unabhängige Ereignisse. Dann ist $p(A \cup B) = p(A) + p(B) - p(A) \cdot p(B)$.
3. Das eine Ereignis ist Teil des anderen. Dann ist $p(A \cup B) = \max\{p(A), p(B)\}$.

Ist man in der Lage, in einem Problem eine in diesem Sinne wahrscheinlichkeitstheoretische Struktur zu erkennen, so geben diese Formeln einen Anhaltspunkt dafür, wie man die Wahrscheinlichkeit einer *Oder*-Verbindung berechnen kann.

3.1.2 Die entscheidungslogische Sicht

Falls ein Dialogsystem gezwungen ist, sich für eine Erkennalternative zu entscheiden, so ist es sinnvoll, einfach dasjenige mit maximalem Score zu nehmen.

3.1.3 Formel

Zur Oder-Verknüpfung zweier Wahrscheinlichkeitsfunktionen p, q ist nun die folgende Vorgehensweise naheliegend: Man setzt zunächst voraus, dass beide auf derselben Grundmenge gegeben sind, also $p, q : X \to [0,1]$. Dann definiert man koordinatenweise, also für jedes $x \in X$,

$$(p \vee q)(x) = p(x) \vee q(x) \in [0,1] \tag{11}$$

mittels einer der in Abschnitt 3.1.1 genannten Verknüpfungen. Um wieder eine Wahrscheinlichkeitsfunktion zu erhalten, muss man anschließend noch normieren:

$$p \oplus q := \mathcal{N}(p \vee q) : X \to [0,1]. \tag{12}$$

3.2 Übertragen einer Wahrscheinlichkeitsfunktion entlang einer beliebigen Relation

Wir gehen davon aus, dass die zugrundeliegende Datenbank - um in einem Dialogsystem Verwendung zu finden - einen geeigneten Ausschnitt der Welt modelliert. Dabei bezeichnet eine sinnvolle Zusammenfassung von Attributen einen Entitätstyp[1], z. B. Person. Eine konkrete Person wird als Entität bezeichnet. Weiter können Entitätstypen über Relationen miteinander verbunden, z.B. kann eine Person an einer Adresse wohnen. Modelliert man die im Laufe eines Dialogs aufgesammelte unsichere Information als Wahrscheinlichkeitsvektoren auf den Entitätslisten auf, so erhebt sich das Problem, eine gegebene Wahrscheinlichkeitsfunktion entlang einer Relation zu übertragen.

Gegeben seien also zwei Mengen E und D (z. B. zwei Entitätslisten), eine Relation $R \subset E \times D$, sowie eine Wahrscheinlichkeitsfunktion $p : D \to [0,1]$. Gesucht ist eine sinnvoll interpretierbare Formel zur Übertragung von p entlang der Relation R zu einer Wahrscheinlichkeitsfunktion $R^*p : E \to [0,1]$ (vgl. Formel (8)).

Analog zum Zurückziehen einer Wahrscheinlichkeitsfunktion entlang einer Abbildung setzen wir auch hier voraus, dass wenigstens eine der folgenden Bedingungen erfüllt ist:

[1] vgl. [2]

1. Die Wahrscheinlichkeitsfunktion p besitzt keine Nullstelle.
2. Die Relation R ist rechtstotal: zu jedem $y \in D$ gibt es wenigstens ein $x \in E$ mit $(x,y) \in R$.

Ist eine dieser Bedingungen erfüllt, so ist die Funktion

$$\bigvee(p \circ R) : E \to [0,1], \qquad x \mapsto \bigvee\{p(y) : (x,y) \in R\}$$

für wenigstens ein $x \in E$ von Null verschieden, und damit normierbar. Wir erhalten so die folgende Formel zum Übertragen eines Wahrscheinlichkeitsvektors p entlang einer Relation R:

$$R^* p := \mathscr{N}(\bigvee(p \circ R)). \tag{13}$$

4 Inkrementelles und gleichzeitiges Und

4.1 Inkrementelles Und: das Bayes'sche Vorgehen

Eine naheliegende Möglichkeit, zwei Wahrscheinlichkeitsvektoren $p, q \in \Delta^n$ zu verbinden, ist, den einen als *Bayes'schen Prior* und den anderen als *Likelihood-Vektor* zu betrachten, und dann den *Bayes'schen Posterior* als Ergebnis der Verknüpfung zu nehmen. Im Kontext von Dialogsystemen wurde diese Verknüpfung z. B. bei Schlangen et al. [5, 1] betrachtet. Dabei ging es darum, ein Objekt durch sukzessives Präzisieren von Merkmalen wie Farbe oder Form zu bestimmen, wobei die Information über Farbe oder Form als unsicher betrachtet wurde. Weil der Bayes'sche Update bei diesem Vorgehen immer wieder angewandt wird, bezeichnen wir ihn als *inkrementelles Und*. Weil zudem die Wahrscheinlichkeiten Unsicherheit modellieren, bezeichnen wir ihn auch als *weiches Und*.

Um den *Bayes'schen Posterior* zu erhalten muss die koordinatenweise Multiplikation mit der Likelihood normiert werden. Das Normieren ist allerdings nur definiert, wenn sich die beiden Wahrscheinlichkeitsvektoren beim Multiplizieren nicht gegenseitig auslöschen. Dies passiert genau dann, wenn in jeder Koordinate zumindestens einer der beiden Vektoren verschwindet.

In der automatischen Spracherkennung ist die Unsicherheit über das Erkennergebnis jedoch so groß, dass die Annahme, dass *keine* Koordinate des Wahrscheinlichkeitsvektors verschwindet, gerechtfertigt erscheint. Technisch kann man das durch „Glätten" eines Wahrscheinlichkeitsvektors $p = (p_1, \ldots, p_n)$ erreichen: man legt einen geeignetkleinen *Schwellenwert* $m > 0$ fest, und konstruiert aus p einen Wahrscheinlichkeitsvektor $\tilde{p} = (\tilde{p}_1, \ldots, \tilde{p}_n)$ mit der Maßgabe

$$\text{für alle } i \in \{1, \ldots, n\} : \qquad \tilde{p}_i \geq m. \tag{14}$$

Eine Diskussion von Glättungsmethoden sowie einen schnellen und flexiblen Glättungsalgorithmus findet man in [6]. Wir nehmen nun an, dass der Glättungsalgorithmus eine Normierung miteinschließt. Dann resultiert die Anwendung des Glättungsalgorithmus in einer *Glättungsfunktion*

$$\mathscr{G} : \mathbb{R}^n_+ \setminus \{0\} \to \Delta^n_+. \tag{15}$$

Für das Folgende gehen wir davon aus, dass die beteiligten Wahrscheinlichkeitsvektoren im offenen Inneren des Simplex, also in der Menge Δ^n_+ liegen.

Mit dieser Notation ist das inkrementelle oder weiche Und auf Wahrscheinlichkeitsvektoren die folgende Operation:

$$\circledast : \Delta^n_+ \times \Delta^n_+ \to \Delta^n_+, \qquad (p_1, \ldots, p_n) \circledast (q_1, \ldots, q_n) = \mathscr{N}(p_1 q_1, \ldots, p_n q_n). \tag{16}$$

4.2 Das gleichzeitige oder harte Und

Ein Problem des inkrementellen Und zeigt das folgende Beispiel: Angenommen, die in der Datenbank befindlichen Namen bilden die Menge

$$D := \{\text{Hans Maier, Hans Müller, Paul Maier, Paul Müller, Peter Maier, Peter Müller}\},$$

und der Spracherkenner hätte mit einem recht hohem Score den Namen „Hans Maier“ erkannt, während die fünf anderen möglichen Ergebnisse gar nicht in der Bestenliste auftreten.

Betrachtet man nun getrennt die beiden Merkmale „Vorname“ und „Nachname“ mit den Wertemengen

$$W_V := \{\text{Hans, Peter, Paul}\} \quad \text{und} \quad W_N := \{\text{Maier, Müller}\},$$

so erhält man nach Umrechnung die Wahrscheinlichkeitsfunktion auf den Vornamen

$$p_V : W_V \to [0,1], \qquad p_V(\text{Peter}) = 1, \quad p_V(\text{Hans}) = p_V(\text{Paul}) = 0,$$

sowie diejenige auf den Nachnamen

$$p_N : W_N \to [0,1], \qquad p_N(\text{Maier}) = 1, \quad p_N(\text{Müller}) = 0.$$

Wir vergleichen nun zwei Verfahren, aus p_V und p_N mit Hilfe der Attributsabbildungen

$$\phi_V : D \to W_V \quad \text{und} \quad \phi_N : D \to W_N,$$

der Normierung (6) und einer Glättung (15) eine Wahrscheinlichkeitsfunktion auf D zu berechnen:

1. *Weiches Und*: Beide Wahrscheinlichkeitsfunktionen auf D zurückziehen – beide glätten – mit dem Bayes'schen Verfahren verknüpfen:

$$p = \mathscr{N}(\mathscr{G}(p_V \circ \phi_V), \mathscr{G}(p_N \circ \phi_N)).$$

2. *Hartes Und*: Das Minimum der Kompositionen $p_V \circ \phi_V$ und $p_N \circ \phi_N$ normieren und glätten.

$$q = \mathscr{G}(\min\{p_V \circ \phi_V, p_N \circ \phi_N\}).$$

Merkmale		Wahrscheinlichkeitsfunktionen			
Vorname	Nachname	$p_V \circ \phi_V$	$p_N \circ \phi_N$	p	q
Hans	Maier	$\frac{1}{2} - 2m$	$\frac{1}{3} - m$	$1 - 7m + 12m^2$	$1 - 5m$
Hans	Müller	$\frac{1}{2} - 2m$	m	$3m - 12m^2$	m
Peter	Maier	m	$\frac{1}{3} - m$	$2m - 6m^2$	m
Peter	Müller	m	m	$6m^2$	m
Paul	Maier	m	$\frac{1}{3} - m$	$2m - 6m^2$	m
Paul	Müller	m	m	$6m^2$	m

Die Verschiedenheit der beiden Werteverläufe zeigt, dass das harte Und in dem Fall, dass in einer einzigen Äußerung sowohl Vor- als auch Nachname genannt wurden, eher angemessen ist als das weiche Und.

5 Die Verwendung von Merkmalskonfidenzen

Bisher haben wir eine Logik von Wahrscheinlichkeitsfunktionen bzw. -vektoren betrachtet. Aus Sicht der Spracherkennung mit Bestenlisten entspricht das dem Fall, in dem alle Erkennalternativen zu einem Merkmal gehören.

Nehmen wir nun an, der Spracherkenner liefere uns eine Bestenliste $L = \{E_1, \dots, E_n\}$, wobei jede Erkennalternative aus einer Menge S_i von Merkmal-Wert-Paaren und einem Score $p_i \in [0,1]$ besteht:

$$E_i = \big(S_i = \{(F_1, w_{i1}), \dots, (F_k, w_{ik})\}, p_i\big), \tag{17}$$

wobei $F_1, \dots, F_k$ die Merkmale bezeichnen, und die Werte aus den entsprechenden Wertemengen sind: $w_{ij} \in W_{F_j}$. Zur Illustration der Vorgehensweise betrachten wir das folgende Beispiel einer Bestenliste, bei dem der Einfachheit halber nur die beiden Merkmale *Vorname* und *Nachname* benutzt werden.

Erkennalternative	Wortfolge	Merkmal-Wert-Paare	Score
E_1	*Peter Maier*	Vorname:Peter und Nachname:Maier	$p_1 = 0.29$
E_2	*Peer Bayer*	Vorname:Peer und Nachname:Bayer	$p_2 = 0.27$
E_3	*Irlmayer*	Nachname:Irlmayer	$p_3 = 0.26$
E_4	*Peter*	Vorname:Peter	$p_4 = 0.08$
E_5	*Peter*	Nachname:Peter	$p_5 = 0.06$
E_6	...	*(kein Ergebnis)*	$p_6 = 0.04$

Unter der *Merkmalstruktur* einer Erkennalternative verstehen wir die Menge der Merkmale in der Semantik der betreffenden Erkennalternativen. Im Beispiel erhalten wir die folgenden Merkmalstrukturen, mit der Bezeichnung M_i für die Merkmalstruktur der i-ten Erkennalternative:

$$\begin{aligned} &M_1 = M_2 = \{\text{Vorname, Nachname}\}, &\quad& M_3 = M_5 = \{\text{Nachname}\},\\ &M_4 = \{\text{Vorname}\}, && M_6 = \varnothing. \end{aligned}$$

5.1 Parallelwelten

Ist eine Bestenliste von Erkennalternativen mit unterschiedlichen Merkmalstrukturen gegeben, so ist es nicht unbedingt sinnvoll, aus *allen* Erkennalternativen *eine* gemeinsame Wahrscheinlichkeitsverteilung auf einer Menge von Datensätzen zu ermitteln. Es ist aber ohne weiteres klar, dass es sinnvoll ist, Erkennalternativen mit gleicher Merkmalstruktur zusammenzufassen; im Beispiel sind das die Alternativen Nr. 1 und 2, sowie die Alternativen Nr. 3 und 5.

Technisch kann das so realisiert werden, dass man für jede der in der Bestenliste vorkommenden Merkmalstrukturen eine eigene *Parallelwelt* anlegt. Es ist dann Aufgabe des Dialogmanagers, zu entscheiden, welche Parallelwelten im weiteren Dialogverlauf weiter verfolgt werden – Parallelwelten mit zu geringer *Konfidenz* können abgeschnitten werden.

Im Folgenden wird nur der Fall behandelt, dass Erkennalternativen mit gleicher Merkmalstruktur zusammengefasst werden. Für eine Bestenliste L und eine Merkmalstruktur M bezeichne

$$P_L(M) := \{E_i = (S_i, p_i) \in L : \text{ die Semantik } S_i \text{ besitzt die Merkmalstruktur } M\ \} \tag{18}$$

die Menge der Erkennalternativen von L mit der Merkmalstruktur M.

5.2 Merkmalskonfidenzen und Wahrscheinlichkeitsverteilungen

Innerhalb einer Parallelwelt werden nur die Erkennalternativen einer Merkmalstruktur M berücksichtigt, also genau die E_i aus der Menge $P_L(M)$. Zu jedem Merkmal $F_j \in M$ definieren wir sein *Konfidenz* als

$$c(F_j) := \sum_{E_i \in P_L(M)} p_i. \tag{19}$$

Daraus ergibt sich für jedes $F_j \in M$ auch eine Wahrscheinlichkeitsfunktion

$$p_{F_j} : W_{F_j} \to [0,1], \qquad p_{F_j}(w) := \begin{cases} \frac{p_i}{c(F_j)} & \text{für } w = w_{ij}, \\ 0 & \text{sonst.} \end{cases} \tag{20}$$

Diese Wahrscheinlichkeitsfunktionen werden dann wie in Abschnitt 4.2 beschrieben auf die Menge D der Datensätze übertragen und mit dem *harten Und* verknüpft. Man erhält so zu jedem Datensatz ein *Konfidenz-Wahrscheinlichkeitspaar*. Weil die Menge der Konfidenz-Wahrscheinlichkeitspaare die Struktur eines Halbrings besitzt (siehe [7]), eignen sich diese als Gewichte zur weiteren Verarbeitung in einem endlichen Transduktor (siehe [4]).

Literatur

[1] BAUMANN, T., O. BUSS und D. SCHLANGEN: *InproTK in Action: Open-Source Software for Building German-Speaking Incremental Spoken Dialogue Systems*. In: *Beitrag zur ESSV 2010, Berlin, 8. bis 10. September 2010.*

[2] CHEN, P. P.-S.: *The Entity-Relationship Model—Toward a Unified View of Data*. 1(1):9–36, 3 1976.

[3] HUBER, M., C. KÖLBL, R. LORENZ, R. RÖMER und G. WIRSCHING: *Semantische Dialogmodellierung mit gewichteten Merkmal-Werte-Relationen*. In: *ESSV 2009, Tagungsband der 20. Konferenz, Dresden, 21. bis 24. September 2009*, S. 25–32.

[4] KÖLBL, C., M. HUBER und G. WIRSCHING: *Endliche gewichtete Transduktoren als semantischer Träger*. Beitrag zur ESSV 2011, Aachen, 28. bis 30. September 2011.

[5] SCHLANGEN, D., T. BAUMANN und M. ATTERER: *Incremental Reference Resolution: The Task, Metrics for Evaluation, and a Bayesian Filtering Model that is Sensitive to Disfluencies*. In: *Proceedings of SIGdial 2009, the 10th Annual SIGDIAL Meeting on Discourse and Dialogue*, London, UK.

[6] WIRSCHING, G. und H. FISCHER: *Probability Vector Estimation Under Constraints by Discounting*. Techn. Ber., Katholische Universität Eichstätt-Ingolstadt, 2011.

[7] WIRSCHING, G., C. KÖLBL und M. HUBER: *The confidence-probability semiring*. Techn. Ber., Angewandte Informatik, Universität Augsburg, 2010.