Sommersemester 2010: Soft Computing 2 — Fynn Feldpausch

Prof. Dr. Kerstin Schill — **20. Mai 2010**

Vertiefung in die Dempster-Shafer-Theorie

---

1 Grundlagen

Die in den 60er und 70er Jahren entwickelte *Dempster-Shafer-Theorie of Evidence* ist Teil der Wahrscheinlichkeitstheorie und stellt eine Alternative zur klassischen Bayes'schen Theorie dar. Im Kontext der Dempster-Shafer-Theorie gilt Θ als der Betrachtungsrahmen (*frame of discernment*) einer gegebenen Situation und stellt eine vollständige Menge sich gegenseitig ausschließender Elemente dar, welche als die Hypothesen der jeweiligen Situation interpretiert werden kann. Mit Hilfe einer Glaubensfunktion (*mass function*) kann den Hypothesen ein Glaubensmaß zugeordnet werden. Eine Funktion $m : 2^{\Theta} \to [0,1]$ heißt Glaubensfunktion (*mass function*), sofern sie folgenden Bedingungen genügt:

1. $m(\emptyset) = 0$
2. $\sum_{A \subseteq \Theta} m(A) = 1$

Werte unterschiedlicher Glaubensfunktionen lassen sich mittels Kombinationsregeln (▷ Abschnitt 2) verknüpfen. Die *gesamte* Unterstützung für eine Aussagenmenge lässt sich durch eine Vertrauensfunktion (*belief function*) $Bel : 2^{\Theta} \to [0,1]$ messen. Auch an sie sind verschiedene Bedingungen geknüpft:

1. $Bel(\emptyset) = 0$
2. $Bel(\Theta) = 1$
3. Für jede natürliche Zahl n und jede Sammlung $A_1, \ldots, A_n$ von Untermengen von Θ,

$$Bel(A_1 \cup \cdots \cup A_n) \geq \sum_{I \subset \{1,\ldots,n\}, I \neq \emptyset} (-1)^{|I|+1} Bel\left(\bigcap_{i \in I} A_i\right)$$

Üblicherweise ergibt sich die Vertrauensfunktion $Bel(A)$ als Summe der Glaubensfunktionen $m(B_i)$ aller ihrer Teilmengen ($B_i \subseteq A$). Die ausdruckslose Vertrauensfunktion (*vacuous belief function*) ordnet der Gesamtmenge Θ den Wert 1 und allen anderen Mengen den Wert 0 zu.

2 Kombinationsregeln

Werte unterschiedlicher Glaubensfunktionen m_1 und m_2 lassen sich durch verschiedene Kombinationsregeln miteinander verknüpfen. Die weit verbreitete Kombinationsregel von Dempster folgt dabei dem Prinzip, alle inkonsistenten Kombinationen zu verwerfen und den Rest zwischen allen konsistenten Kombinationen aufzuteilen. Diese Regel kann allerdings einer intuitiven Einschätzung widersprechen (▷ Abschnitt 3). Daher lohnt sich oft die Ergebnisbetrachtung weiterer Kombinationsregeln:

1. **Smets Kombinationsregel:** *»Kombiniere die Werte ohne spätere Normalisierung.«*

$$(m_1 \oplus m_2)(X) = \begin{cases} 1 - \sum_{A,B \subseteq \Theta, A \cap B = \emptyset} m_1(A) m_2(B), & \text{wenn } X = \emptyset \\ \sum_{A,B \subseteq \Theta, A \cap B = X} m_1(A) m_2(B), & \text{sonst} \end{cases}$$

2. **Yagers Kombinationsregel:** *»Weise den Rest nicht der leeren Menge sondern Θ zu.«*

$$(m_1 \oplus m_2)(X) = \begin{cases} 0, & \text{wenn } X = \emptyset \\ m_1(\Theta) m_2(\Theta) + \sum_{A,B \subseteq \Theta, A \cap B = \emptyset} m_1(A) m_2(B), & \text{wenn } X = \Theta \\ \sum_{A,B \subseteq \Theta, A \cap B = X} m_1(A) m_2(B), & \text{sonst} \end{cases}$$

Neben den vorgestellten Kombinationsregeln existieren weitere Möglichkeiten wie *Dubois–Prade's rule of combination*, *weighted operators*, *Transferable Belief Models* (*TBMs*), *disjunctive combinations*, ...

3 Umgang mit Widersprüchen

Die Verwendung verschiedener Kombinationsregeln kann insbesondere bei widersprüchlichen Szenarien zu sehr unterschiedlichen Ergebnissen führen. Der Eignungsgrad einer Kombinationsregel kann daher stark vom vorliegenden Szenario abhängen. Ein Beispiel zeigt die folgende Situation:

> Gegeben sei ein Mordfall mit drei Verdächtigen $\{A, B, C\}$ und zwei Zeugen $\{1, 2\}$. U_x bezeichne das Urteil mit x als Mörder. Zeuge 1 belastet den Verdächtigen A, während Zeuge 2 den Verdächtigen C für schuldig hält. Die zugehörigen Glaubensfunktionen lauten:
>
> » $\mathbf{m_1(U_A)} = 0,99$ » $\mathbf{m_1(U_B)} = 0,01$ » $\mathbf{m_1(U_C)} = 0$
>
> » $\mathbf{m_2(U_A)} = 0$ » $\mathbf{m_2(U_B)} = 0,01$ » $\mathbf{m_2(U_C)} = 0,99$

Eine Kombination beider Funktionen mittels der Regel von Dempster (**D**) führt erstaunlicherweise zu einem eindeutigen Schuldspruch des Verdächtigen B. Die Kombinationen nach Smet (**S**) und Yager (**G**) weisen durch ihre Ergebnisse hingegen auf mögliche Widersprüche im System hin:

	$(\mathbf{m_1} \oplus \mathbf{m_2})(\emptyset)$	$(\mathbf{m_1} \oplus \mathbf{m_2})(\mathbf{U_A})$	$(\mathbf{m_1} \oplus \mathbf{m_2})(\mathbf{U_B})$	$(\mathbf{m_1} \oplus \mathbf{m_2})(\mathbf{U_C})$	$(\mathbf{m_1} \oplus \mathbf{m_2})(\mathbf{\Theta})$
D	0	0	1	0	0
S	0,9999	0	0,0001	0	0
Y	0	0	0,0001	0	0,9999

4 Anwendungsbeispiel: Szenenanalyse

Ziel der Szenenanalyse ist es, mit einem Minimum an verfügbaren Ressourcen ein Maximum an Informationen über die eigene Umgebung zu sammeln. Da natürliche Systeme ein Optimum für Umgebungsexploration entwickelt haben, ist ein biologisch inspirierter Entwurf[1] naheliegend. In einem ersten Schritt werden dabei zunächst Merkmale einer Szene identifiziert und in eine **Feature-Karte** eingetragen. Zugehörige hochauflösende Informationen liegen als **Feature-Vektoren** $\vec{s} = [\vec{v_1}, \vec{e}, \vec{v_2}]$ ($\vec{v_i}$: Orientierungsdaten, $\vec{e}$: Bewegungsdaten) vor.

Abbildung 1: Feature-Karte

Die Menge dieser Feature-Vektoren bildet **sensomotorische Klassen** S_i, über die eine Szene im Wesentlichen repräsentiert werden kann. Jedes S_i unterstützt dabei eine Hypothese(nmenge) H_k eines Objektes der Szene mit dem Glaubensmaß $m_i(H_k)$. Die Kombination von Hypothesen kann mittels Dempster-Shafer durchgeführt werden. Aus Komplexitätsgründen wird dabei der Hypothesenraum 2^Θ auf eine hierarchische Struktur T beschränkt, welche Objekte und Objektklassen der Szene enthält.

Die *belief*-Werte dieser Objekte bzw. Objektklassen werden vom System erlernt, während Fixierungen vorgenommen werden. Die Wissenbasis wird so kontinuierlich erweitert und verbessert. Für die Datenauswahl wird **Inference by Information Gain (IBIG)** verwendet. Dabei wird stets das Datum ausgewählt, das bei der Aufnahme zum größten Informationsgewinn bezüglich der *belief*-Verteilung im Hypothesenraum führen würde:

1. kompatible sensomotorische Feature-Vektoren $\hat{S}_i = [\hat{\vec{v_1}}, \hat{\vec{e}}, \hat{\vec{v_2}}]$ identifizieren
   - $\hat{\vec{v_1}}$ stimmt mit $\vec{v_2}$ der letzten Bewegung überein
   - $\hat{\vec{e}}$ ist mit der Feature-Karte potentieller Ziele vereinbar
2. für jedes $\hat{S}_i$ alle Hypothesen H_k identifizieren, in denen $\hat{S}_i$ ein $\hat{m}_{T_i}(H_k) \neq 0$ herbeiführt
3. $\hat{m}_{T_i}(H_k)$ entsprechend der Kombinationsregel ändern
4. $\forall \hat{S}_i$: Informationsgewinn $I_i(H_k) = |m_T(H_k) - \hat{m}_{T_i}(H_k)|$ berechnen und Auswahl treffen

Auf diese Weise identifiziert das System »interessante« Regionen und kombiniert die Verarbeitung der Sensorinformationen (*bottom-up*) mit Informationen der Wissensbasis (*top-down*).

[1] Grundlage hier: Schill, K., Zetsche, C., Hois, J. (2008): *A belieb-based architecture for scene analysis*