

信念論理に基づく情報検索*

4E-10

村井哲也[†]
北海道教育大学函館校宮腰政明[‡]
北海道大学工学部新保 勝[§]
北海道大学工学部

1 まえがき

本論文では信念論理の Dempster-Shafer 理論 [6] (以下, D-S理論と略す) に基づくモデル [4, 5] を利用した情報検索のモデルを提案する. 本モデルは, 索引語の各文献に対する重みを D-S理論における証拠とみなし, 二値的な索引付けはその重みから定義される信念様相演算子に基づいて得ると考える. その結果, 検索過程を索引者の信念に基づくものと解釈することが可能となる. 更に, pertinence と relevance との区別 [1, 3] を扱うための非単調検索という考え方も提案する. 以下の議論の基礎となる, 様相論理に対するファジィ測度に基づく意味論に関しては, 文献 [4, 5] を参照されたい.

2 信念論理と情報検索

2.1 情報検索における重み

古典論理やファジィ論理に基づく情報検索において, 索引語の文献に対する重みは真理値とみなされる. しかし, これは情報検索ではあまり適当な考えとはいえない. なぜなら, 同義語・類義語の問題を考えれば分かるように, ある索引語が付されていない文献は必ずしもその索引語に不適合であるとはいえない場合がある.

同義語・類義語の問題は論理における否定演算子の定義に由来する: $V(\neg rel(d, t)) = 1 - V(rel(d, t))$. ここで, ' $rel(d, t)$ ' は「文献 d は索引語 t に適合 (relevant) する」を意味し, V は $[0, 1]$ または $\{0, 1\}$ の値を取る付値関数である. 確かに, 索引語 t の文献 d に対する適合性を表わす重み w は $rel(d, t)$ が真であることを支持する. しかし, $1-w$ は必ずしも, $rel(d, t)$ が偽であることを支持しない. それにもかかわらず, 否定演算子の定義は, 重み $1-w$ を $rel(d, t)$ が偽であること, すなわち $\neg rel(d, t)$ が真であることへの支持にすり替えてしまうのである.

重み $1-w$ が必ずしも偽に対する支持を意味しないという観点は, 実は, D-S理論の基本的考え方であり, 重み w は $rel(d, t)$ の真理値と考えるより, D-S理論における証拠とみなした方が適切である.

2.2 重み付き索引付け

$D = \{d_k\}$ および $T = \{t_k\}$ をそれぞれ, 文献および索引語の有限集合とし, $W_{d,t} = \{rel(d, t), \neg rel(d, t)\}$ とおく. このとき, 文献 d の索引語 t による重み付き索引付けを次の基本確率割当によって定義する:

$$m_{d,t}: \mathcal{P}(W_{d,t}) \rightarrow [0, 1]$$

$$\text{s.t. } \sum_{A \subseteq W_{d,t}} m_{d,t}(A) = 1, m_{d,t}(\phi) = 0.$$

文献 d と索引語 t に対して与えるべき重みは3つある:

$$m^+ = m(\{rel(d, t)\})$$

$$m^- = m(\{\neg rel(d, t)\})$$

$$m^0 = m(\{rel(d, t), \neg rel(d, t)\})$$

m^+ および m^- はそれぞれ $rel(d, t)$ が真および偽である程度を, m^0 はそのどちらともいえない無知の程度を表現する. 通常索引付けでは索引者は m^+ のみを評価すると考えられる. 重み $1 - m^+$ に対して $m^- = 1 - m^+$ と解釈すれば, ファジィ検索になる. 同義語・類義語の問題を考慮するときは, $1 - m^+$ は必ずしも否定的な支持ではないと解釈し $m^0 = 1 - m^+$ とする場合もある.

2.3 二値的索引付けのモデル

D-S理論に基づく信念論理 [4, 5] を利用すれば, 索引付けは索引者の信念に基づいて実行されるという観点を定式化できる. ' $ind(d, t)$ ' を ' t は d を索引付ける」を意味するものとするとき,

$$\models^M ind(d, t) \Leftrightarrow \models^M Brel(d, t)$$

$$\Leftrightarrow Pl(\|rel(d, t)\|) = 1$$

*Doxastic Document Retrieval

[†]Tetsuya MURAI, Hakodate Campus, Hokkaido University of Education[‡]Masaaki MIYAKOSHI, Faculty of Engineering, Hokkaido University[§]Masaru SHIMBO, Faculty of Engineering, Hokkaido University

によって二値的索引付けを定義できる。ここでは、重み付き索引付けを表わす基本確率割当から生成される plausibility 関数を利用して信念演算子を定義した。この B は様相論理体系 EMNP の演算子となり、一般に公理 $D_c: \neg Bp \rightarrow B\neg p$ を満たさない。したがって、

$$\neg ind(d, t) \rightarrow B\neg rel(d, t)$$

が一般に不成立であり、索引付けされていないことが不適合と信ずることにならない。

2.4 信念論理に基づく情報検索モデル

ブール検索式の集合 Q を再帰的に定義する：
 $q = rel(X, t) \Rightarrow q \in Q, q \in Q \Rightarrow \neg q \in Q,$
 $q_1, q_2 \in Q \Rightarrow q_1 \wedge q_2, q_1 \vee q_2 \in Q.$ ここで、 X は文献を値とする変数である。

ブール質問式 q に対して、 T_q を q に現われる索引語からなる集合とする。基本確率割当 $m_{d,t}$ を文献 d と $t \in T_q$ に対する重み付き索引付けとする。 $m_{d,t}$ は直積 [6] によって、 d の q に対する全体的索引付けに拡張できる：

$$m_{d,q} = \otimes_{t \in T_q} m_{d,t}$$

ブール質問式 q に対する信念論理による出力結果を $O(q)$ で表わす：

$$O(q) = \{d \mid \models^M Bq\}.$$

plausibility 関数に基づく様相演算子の性質から

$$O(q_1 \wedge q_2) \subseteq O(q_1) \cap O(q_2)$$

$$O(q_1 \vee q_2) \supseteq O(q_1) \cup O(q_2)$$

が成り立ち、また、ある q に対して

$$O(q) \cap O(\neg q) \neq \phi$$

となる。plausibility 関数が Dirac 測度になるとき、通常の論理検索と一致する。

3 非単調検索

文献と質問の間に成り立つ適合性を表わす概念として「relevance」の他に「pertinence」という概念があり、前者は文献と利用者の質問との間の客観的関係を表わし、後者は文献と利用者の必要性との間の主観的関係を表わすとされる [1, 3]。したがって、relevant でありながら、pertinent でない文献が存在し得る。pertinence に影響を与える理由として、Cooper[1] は次の諸点を指摘する：

- 利用者が文献の relevance を容易に理解できること

- 利用者にとって文献が重要であること
- 出版元、著者などに対する信頼性

などを挙げている。これらを考慮すれば検索はより効果的に実行できる。

検索過程において pertinence を考慮する一つの方法は非単調推論 [2] の利用であろう。すなわち、ある文献はそれが特に利用者にとって pertinent でないという証拠がない限り pertinent であることができる。これは典型的な非単調推論であり、次のように定式化できる：

$$Brel(d, t) \wedge \neg B\neg per(d, t) \rightarrow Bper(d, t),$$

ここで、'per(d,t)' は「d is pertinent to t」を意味する。これは

$$ind(d, t) \wedge \neg B\neg per(d, t) \rightarrow Bper(d, t)$$

と書くこともできる。本推論を検索過程に適用するためには、pertinence でないことを支持する証拠を集める必要がある。利用者の過去の検索結果およびそれに対する判断を記録したユーザ・モデルを構成すれば、より効果的な検索に役立つと思われる。

References

- [1] W.S.Cooper, A Definition of Relevance for Information Retrieval. *Information Storage and Retrieval*, 7(1971), 19-37.
- [2] M.L.Ginsberg, Readings in Nonmonotonic Reasoning. Morgan Kaufmann, 1987.
- [3] D.A.Kemp, Relevance, Pertinence, and Information System Development. *Information Storage and Retrieval*, 10(1974), 37-47.
- [4] T.Murai, M.Miyakoshi, and M.Shimbo, Measure-Based Semantics for Modal Logic. *R.Lowen and M.Roubens(eds.), Fuzzy Logic: State of the Art*, Kluwer Academic, 1993, 395-405.
- [5] T.Murai, M.Miyakoshi, and M.Shimbo, Soundness and Completeness Theorems between the Dempster-Shafer Theory and Logic of belief, submitted.
- [6] G.Shafer, Belief Function and Possibility Measures. *Analysis of Fuzzy Information*, Vol.1, 51-84, 1987.