

中心仮説選択法による対話型概念学習

5 L - 6

山口 尚吾 大川 剛直 馬場口 登 手塚 慶一

大阪大学 工学部

1. まえがき

対話型概念学習は、例題からの学習の一形態であり、学習者が自ら例題を選択し、その例題が教師の考える目標概念に属するかどうかを質問することにより学習を進めるものである。このとき、どのような例題から学習するかが、学習効率や学習結果の質を左右し、その選択手法が重要と言える。対話型学習において学習効率を決定付ける要因は、例題選択のための計算量と学習完了までの質問回数である。つまり、学習効率を向上させるためには、少ない質問回数で学習が可能となる良質の例題を、少ない計算量で選択することが必要となる。

これまでに提案されている代表的な例題選択手法として、二分法[1][2]と保存選択法[3][4]が知られている。まず、二分法では、最少の質問回数での学習が保証されるものの、その代償として例題選択のための計算量が大きくなっている。一方、保存選択法は、最初に与えられた例題から一般化により仮説を生成しながら、順次、例題を選択していく。生成される仮説が多数ある場合に対しても、一つの仮説のみに着目し、これが検証された際には、他の仮説は棄却されるため、質問回数を比較的抑えることができる。しかしながら、一般化のたびに仮説を生成するため、例えば1回の一般化で生成される仮説が少ないような状況においては、より少ない質問回数で学習できることが推察され、改善の余地が残されている。本稿では、以上のような観点から、より学習効率の向上を目指した中心仮説選択法(Middle Hypothesis Selection Method)を提案する。

2. 中心仮説選択法

2.1 諸定義

(1) 一般化木

《定義1》 (事例集合と記述集合)

ある概念に関して、その概念の学習において与えられ得る全ての事例を表現した記述の集合をその概念に関する事例集合という。また、目標概念の記述となり得る記述の集合を記述集合という。

《定義2》 (記述の一般-特殊関係)

記述A, Bについて、 $\{A\}$, $\{B\}$ をそれぞれA, Bに被覆される事例集合とする。このとき

$$\{A\} \supset \{B\}$$

が成り立つならば、記述Aは記述Bより一般的であるという。また、記述Bは記述Aより特殊であるという。

《定義3》 (一般化規則)

記述Aが記述Bより一般的であるとき、BからAを生成する規則を一般化規則という。

《定義4》 (一般化木)

記述Aに対し、一般化規則を適用して得られる記述、ならびに生成された記述に対し再帰的に一般化規則を適用して得られる全ての記述をノードとし、記述間の一般-特殊関係をアークとして表現される記述Aを根とする木を記述Aの一般化木という。

一般化木の生成に用いられる一般化規則には、例えば条件削除規則や連言選言化規則、階層上昇規則などがある。また、一般化木は記述集合内の全記述の一般-特殊関係が木構造をなすときに単純な直線構造となる。逆に、一つの記述に対して多くの一般化規則が適用できるとき一般化木の構造は複雑なものとなる。

(2) 中心仮説

《定義5》 (一般化段階数)

記述Aと、これより一般的な記述Bに対して、記述Aより一般的で記述Bより特殊な記述が、これらが属する記述集合内に存在しないとき、記述Bを記述Aの一段階一般化した記述であるという。さらに、記述Cが記述Aをm回一段階一般化して得られるとき、mを記述Aから記述Cまでの一般化段階数という。

《定義6》 (一般化木の深さ)

ある記述Aに対する一般化木に関して、各枝において記述Aから枝の末端に位置する記述までの一般化段階数を、その枝の深さという。また、一般化木の各枝の中で、最も深い枝の深さをその一般化木の深さという。

《定義7》 (中心仮説)

ある記述Aに対する一般化木において、その一般化木の深さがdのとき、記述Aを $(d/2)$ 段階一般化して得られる記述を、その一般化木の中心仮説という。

中心仮説は一般化木の深さが半分の所に位置する記述全てを指し、必ずしも唯一でない。

(3) 記述集合間の独立

《定義8》 (記述集合間の独立)

事例集合Iを持つある概念の記述集合の要素が、異なる二つの概念の記述集合A, Bの要素の連言によって表されているとする。記述集合Aの全ての記述aと記述集合Bの全ての記述bについて、aとbの連言記述 $a \wedge b$ に被覆される事例がI中に存在するならば、またそのときに限り、記述集合Aと記述集合Bは独立であるという。

2.2 学習手続き

Step I 教師により与えられた一つの正の例題を被覆する記述を生成

Step II I で求めた記述を独立な記述集合の要素に分割

Step III 各独立な記述集合に関して

- (1) 初期記述を根とする一般化木を構成
- (2) 一般化木から中心仮説を一つ選択
- (3) (2) で選択した仮説には被覆されるが、仮説を一段階特殊化した記述には被覆されない例題を事例集合から選択
- (4) (3) で選択した例題の正負を教師に質問
- (5) 例題の正負の如何により次の様に一般化木を再構成
例題が正の時：仮説を根として一般化木を構成
例題が負の時：別の仮説を選択し(3)へ。存在しない場合には根が同じで深さが半分の一般化木を構成
- (6) 一般化木中の記述が唯一に定まるまで(2)～(5)を反復
- (7) 唯一に定まった記述を記述集合中の概念記述に決定

Step IV 各独立な記述集合中の概念記述の連言を学習結果として出力

3. 学習実験及び検討

2.2の手続きを有機化合物の階層的分類を領域知識とする対象に適用し、学習実験した結果を図1に示す。ここで、目標概念 x は「他の物質と反応してエステルを作るもの」であり、入力として与えられる初期例題は「カテコール」、一般化規則は階層上昇規則、事例集合は{カテコール、フェノール、マレイン酸、ギ酸、...}である。また、システムに与えられる背景知識は図2に示す有機化合物の階層である。

システムの最後の出力記述(x (有機分類 水酸化化合物))が得られた学習結果であり、「他の物質と反応してエステルを作るもの」が水酸化化合物であることを学習している。

システム>Positive example of new concept?
教師 >(x (カテコール))
システム>(ベンズアルデヒド) is positive?
教師 >n
システム>(メチルアルコール) is positive?
教師 >y
システム>(n-ブタン) is positive?
教師 >n
システム>(x (有機分類 水酸化化合物))

図1 学習過程

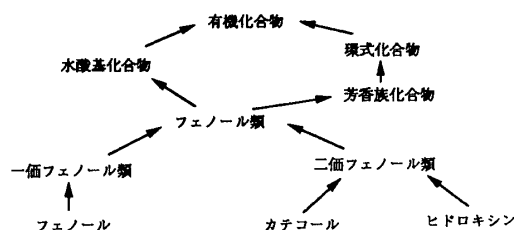


図2 有機化合物の階層関係の一部

表1 一般化木の深さと質問回数

一般化木の形	中心仮説選択法	保存選択法	二分法
深さ d の直線構造	$O(\log d)$	$O(d)$	$O(\log d)$
深さ d の二分木	$O(2^{d/2})$	$O(d)$	$O(d)$
有機分類に対する学習で生成された木	3. 4	3. 9	3. 1

表2 一般化木の深さと例題選択の全計算量

一般化木の形	中心仮説選択法	保存選択法	二分法
深さ d の直線構造	$O(n \log d)$	$O(nd)$	$O(nd)$
深さ d の二分木	$O(n 2^{d/2})$	$O(nd)$	$O(n 2^d)$

学習完了までの質問回数を従来手法と比較すると表1のようになる。ここで、学習実験に対する質問回数の値は、記述集合内の全ての記述を目標概念と考えたときの平均値である。

一般化木が直線構造の時は最少の保証がなされている二分法と同等の質問回数で学習が可能である。しかし、一般化木が二分木になると生成される仮説数が極めて多くなるため質問回数も増大する。なお、表1に示すように前述の有機化合物の階層的分類を領域知識とする対象についての学習においては平均して、保存選択法より少ない質問回数で学習可能であった。

次に、学習完了までの例題選択のための全計算量の従来手法との比較を表2に示す。但し、表中の n は事例集合の要素数である。

一般化木が直線構造の時は中心仮説選択法が最も効率良く学習可能である。しかし、一般化木が二分木になるときは保存選択法の方が効率が良い。

以上の結果から、中心仮説選択法は一般化木が枝別れの少ない、深さの深い領域、例えば木構造をなす階層的分類を領域知識とする対象に適用したときに有効であると考えられる。

4. まとめ

対話型の概念学習において、一般化木の深さを例題選択毎に半減させることにより効率化を図った中心仮説選択法を提案した。本手法の学習効率を、質問回数と例題選択のための計算量の面から、従来手法との比較により検討し、一般化木が枝別れの少ないとき、特に直線構造を持つときには非常に有効であることがわかった。

《参考文献》

- [1] Genesereth, M.R. and Nilsson, N.J.: Logical foundations of artificial intelligence, Morgan Kaufmann, Chapter 7 (1987).
- [2] Subramanian, D. and Feigenbaum, J.: Factorization in experiment generation, Proc. 5th AAAI, pp. 518-522 (1986).
- [3] Krawchuk, B. and Witten, I.: On asking the right questions, Proc. 5th MLC, pp. 15-22 (1988).
- [4] Raedt, L. and Bruynooghe, M.: Towards Friendly Concept-Learners, Proc. 11th IJCAI, pp. 894-853 (1989).