

背景知識を考慮した概念構造の対比操作

3AH-5

金子 雄一知 中島 誠 伊藤 哲郎
大分大学工学部 知能情報システム工学科

1 まえがき

種々の知識処理の分野においては、処理すべき対象同士を比較することを新しい問題の解を求めることや概念の生成のために役立てる。このとき、各処理単位を特徴づける背景知識としての属性や関係、その特徴づけに従う処理単位間の高次の関係を反映させることが重要であるが、これまでこの点に関しての認識が十分でなかった。

ここでは、背景知識を考慮しないとよい結果が得られない概念構造の対比（比較）問題について、概念学習でのMSC一般化記述を発展させて得たMSC被覆を用いて解くことを試みる。また、MSC被覆を求める効率的な方法についても考察する。

2 被覆とBK対応づけ操作

いま、2つの概念構造 S と T を取り上げ、概念に相当する各節点には共通の背景知識を通じた記述が与えられているとする。これら構造間の対比を、表面的ではなく背景知識を通じて見たときに関連深いとされる概念同士を、お互いの構造中での一般的/特定の関係を考慮しながら対応づけてゆくことと捉える。以下に被覆の考えに基づいた概念構造の対比操作（BK対比操作と呼ぶ）を示す。

- (bk1) S のすべての節点に高次の関係を保存するように T の節点を順に対応づけて、可能なすべての被覆を求める。
- (bk2) 得られた被覆の中から節点の対応づけに必要なMSC被覆を選ぶ。
- (bk3) MSC被覆を作った節点同士の対応づけを出力する。

ここで S の2つの節点間の一般的/特定の関係が、それぞれに対応づけられた T の2つの節点間にも見られるとき、高次の関係が保存されているという。一方の構造の節点数が少なかったり、幾つかの節点を除けば高次の関係を保存する対応づけが得られる場合は、仮の記述で規定された仮の節点に対応づけておく。

また、被覆とは S と T の個々に対応づけられた節点をひとまとめで言及する一般化記述を組にしたものである。いま、2つの被覆について、それぞれの被覆中の一般化記述を1対1に組み合わせてみたとき、一方の被覆中のすべての一般化記述が相手より一般的（特定の）になるなら、前者は後者より一般的（特定の）という。少なくとも他より一般的でない被覆がMSC被覆である。背景知識を介して見たとき、関連深い2つの節点はより特定の一般化記述を作り出す[1, 3]ことに注目すると、概念構造の対比には、種々の被覆の中からMSC被覆を選べばよいことになる。

3 効率化

BK対応づけ操作において可能なすべての被覆を求めることは、対象とする概念構造が大きくなると現実的でない。ここでは A^* アルゴリズムとビーム探索を用いた効率化を行なう。

まず、最適解探索のための A^* アルゴリズムを用いてレベル総数[3]の高い被覆を $N(\geq 1)$ 個だけ探すことを考える。レベル総数は、背景知識としての属性や関係について、それらの値を節点としたグラフ表現中で、下にある節点ほど高いレベルを設定し、被覆の各一般化記述中に現れる値のレベルを合計したものである。ある被覆より特定の（一般的）な被覆のレベル総数は前者のそれより大きい（小さい）。従って、MSC被覆を求めるには、レベル総数の大きい被覆を N 個だけ求め、各被覆についてそれより大きなレベル総数の被覆より一般的でないことを確かめればよい。但し、 N 個のすべてがMSC被覆にならないため、 N は必要な数より多めに設定しておく。

対象とする概念構造の節点数を n とすると、 A^* アルゴリズムによっても最悪の場合 $n!$ 個の被覆が生成されてしまう。それゆえ、ビーム探索による効率化を考える。最初に、 S と T の節点それぞれに対応づけたとき得られる一般化記述について、その中に現れる背景知識での値のレベルを求めて設定し、降順に並べたリストを用意しておく。このリストの上位にあるものから、高次の関係を保存するような対応づけによる一般化記述を、ビーム幅だけ取り出しながら被覆を求めていく。

ビーム探索では大きさが一定なら扱う概念構造に関わらず一定時間で探索が終了する。しかし、基本的なビーム探索そのままでは、節点の対応づけが、その結果得られる一般化記述に設定されたレベルが大きいものから試みられることから、局所解に陥りやすい。ここでは、遺伝的アルゴリズムでの局所解からの脱出法 [5] を参考に、得られた複数の被覆からそれらに共通する対応づけをスキーマとして抜き出し、これを種々変化させたものを元に新たな被覆を作成することを繰り返すことで対処する。

4 実験

まず、A* アルゴリズムによって求めるレベル総数の大きな被覆がどれくらいの割合で MSC 被覆になるかを、 N を 40 とし、5 人の被験者から抽出した個々の持つ 5 つの色概念構造 (I ~ V と書く) [2] を種々組み合わせ調べて。色合い、明るさ、鮮やかさの 3 つの属性が背景知識となっている。結果を図 1 に示す。レベル総数の降順に見た順位が大きくなるに従って、MSC 被覆と判断される割合が急に低くなっている。これより、 N を比較的小さな値に設定しても十分な数の MSC 被覆が取り出せると言える。

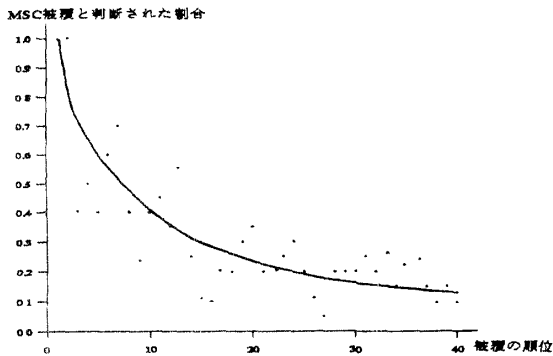


図1 被覆の順位とMSC被覆の関係

次に BK 対比操作の効率化の有効性を構造 I ~ V から種々の大きさの部分構造 (節点数で 10, 14 あるいは 20) を抽出して調べた。コストは、高次の関係を保存するかどうかの検査を試みた回数で測った。実験より、可能な被覆すべてを求める場合には節点の対応のさせ方の種類の数が 2 倍になるとコストは 2^5 倍になるが、A* アルゴリズムを用いると $2^3 \sim 2^4$ 倍で済んだ。ビーム幅を 20 としたビーム探索では、スキーマを利用して 5 回繰り返して被覆を求めた場合でも 2^2 倍で済んだ。

さらに、ビーム探索による被覆の探索の特性を調べた。決定木学習システム ID3 に正の整数を値とする幅、高さの 2 つの属性によって表された様々な長方

形からなる人工的な事例集合を与えて得た分類木を概念構造と捉えて用いた [4]。

20 通りの分類木の組み合わせで、ビーム探索 (ビーム幅=20) で得られる被覆の中でレベル総数で見て上位 3 位までの被覆が、A* アルゴリズム ($N=20$) で求められた被覆の中で何位になるか調べた。結果を図 2 に示す。横軸は、ビーム探索での探索の過程で得

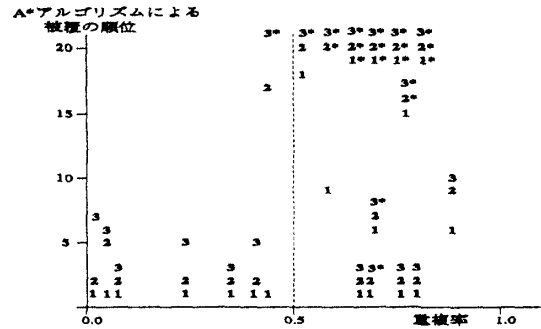


図2 ビーム探索による被覆
1,2,3 はビーム探索での被覆の順位。
* は対応する被覆が 20 位以内にならなかったことを表す。

られた被覆のうち、同じレベル総数となったものの割合 (重複率) を示している。図より、重複率が小さく (0.5 以下) なるような構造の対比では、ビーム探索でも A* アルゴリズムと同様の被覆が求められると考えられる。このことから、節点の対応づけの違いがレベル総数の差異として現れるような構造間の対比では、ビーム探索の利用が有効であると言える。

5 むすび

背景知識を考慮に入れながら概念構造全体の対比を行う BK 対応づけ操作を、MSC 被覆の考えを導入して定式化した。また A* アルゴリズムとビーム探索を利用した BK 対応づけ操作の効率化も行い、その有効性も示した。一般に MSC 被覆は複数求められる。その中からどの MSC 被覆を選べばよいかを決める課題が残されている。

参考文献

- [1] Dietterich, T. G., and Michalski, R. S.: A comparative review of selected methods for learning from examples, R.S. Michalski, J.G. Carbonell and Y.M. Mitchell (eds.), Machine Learning, Morgan Kaufmann, Los Altos, CA (1983).
- [2] 伊藤哲郎: 分類による概念形成, 人工知能学会誌, Vol.6, No.6, pp.881-890 (1991).
- [3] Kolodner, J. L.: Case-Based Reasoning, Morgan Kaufmann, San Mateo, CA (1993).
- [4] Quinlan, J. R.: C4.5: Programs for Machine Learning, Morgan Kaufmann, San Mateo, CA (1993).
- [5] 筒井 茂義, 藤本 好司: 個体群探索分岐型遺伝的アルゴリズム fGA (Forking GA) の提案, 人工知能学会誌, Vol.9, No.5, pp.741-747 (1994).