

知識ベース獲得のための帰納学習法の拡張

4 J-2

森田 快太郎 淵 一博

東京大学工学部*

1 はじめに

現在ではエキスパートシステムは広く認知され、多くの分野でその有用性を発揮している。しかし、すべてにおいて問題点がないわけではなく、知識ベースの構築などいくつかの問題点が指摘されるとともに、それらを解決することを目指した新しい研究も行われている。本発表では帰納論理プログラミングにおける背景知識の使用という点に着目し、帰納論理プログラミングを学習によるエキスパートシステムの知識ベースのための知識獲得の手段としてとらえたときに生じる問題点について検討を行う。

2 学習による知識ベースの構築

2.1 学習による知識獲得

知識獲得はエキスパートシステム構築の最重要課題の一つである。エキスパートからの知識（ルール）抽出は困難で、知識獲得のボトルネックと呼ばれており、可能な限りの自動化が望まれている。人工知能の分野では古くから学習に関する研究が行われてきたが、決定木の学習を除いて知識ベースの構築に役立つようなものはなかった。ところが、最近注目されている帰納論理プログラミングは従来の学習とは一線を画している。背景知識を B 、正負の事例を E^+, E^- とすると、

$$B \wedge H \models E^+ \quad (1)$$

$$B \wedge H \not\models E^- \quad (2)$$

を満たす仮説 H を求める、という枠組を使用している。すなわち、帰納論理プログラミングは対象とする領域に関する事前知識を持ち、それに基づいて幾つかの例題から学習を行うもので、決定木の学習とともに知識ベースの構築に応用できる大きな可能性を持っている。

* "Acquisition of knowledge base for expert systems using Inductive Learning"

Kaitaro MORITA, Kazuhiro FUCHI,
University of Tokyo, Department of Engineering,
7-3-1 Hongou, Bunkyo-ku, Tokyo 113, Japan

2.2 帰納学習を用いた知識ベースの構築

帰納論理プログラミングを用いた実世界の問題に対する応用例として、有限要素法 [2], 故障診断 [2] などがあるが、いずれの場合も、

- 専門家の協力をそれほど必要としない一般的な知識

を背景知識として用いているところに特徴がある。すなわち、専門家からの知識の抽出に大きく依存することなしに知識ベースが得られる可能性がある。背景知識に一般的な知識を用いて学習を行う点から考えると、従来の例題のみからの学習方法と比較しても帰納論理プログラミングがエキスパートシステムの知識ベースの獲得に対して有効な学習手法である可能性があると考えられることができる。

3 帰納学習法の拡張

3.1 本研究の着目点

C4.5 に代表される決定木アルゴリズムは数値の学習を可能にしている。実際に、エキスパートシステムの知識ベースを構築するためには、医学的な測定結果など数値を扱う対象も多く、記号と数値を同等に扱える機能が要求される。そこで、帰納論理プログラミングの工学的な応用のために、帰納論理プログラミングを用いて数値を含む問題領域の学習を行う手法の検討が要求される。また、実世界の問題においては例外に関する表現が必要な場合があると考えられる。このため例外を表現する方法も実際の応用には検討されなければならない。

3.2 提案する手法

本節では、帰納論理プログラミングシステムの一つである GOLEM を用いて数値データを含む実世界の例題を学習させるための手法および、例外の表現を用いた学習手法との組合せた手法について提案する。以下に学習アルゴリズムを述べる。

3.3 クラスタリングを用いた離散化

例題の目的は、ある例が与えられたときに、それを分類するようなルールを獲得することである。そこで、train 用の例題の属性において、それぞれのクラスについて次元クラスタリングを行い、すべてのクラスの閾値をその属性の閾値とし、離散化を行ってから学習させる。クラスタ数はクラスタ評価基準 J を

$$\begin{cases} J = \sum_{i=1}^{\alpha} J_i \\ J_i = \sum_{x \in X_i} \|x - m_i\|^2 \\ m_i = \frac{1}{N_i} \sum_{x \in X_i} x_i \end{cases} \quad (3)$$

とすると、

$$J_s = \{J(1), J(2), \dots, J(\alpha), J(\alpha+1)\} \quad (4)$$

が最小になるような $J(\alpha)$ のときの α となる。

3.4 ルールの一般化

クラスタリングを行うと、図 1 のように構成要素数の少ないクラスタが負の例として存在するために、ルールが特殊化されてしまうことがある。構成要素数の少ないクラスタは外れ値として考えることができるので、負の例における構成要素数の少ないクラスタは、例外として負の例から削除する。負の例におけるクラスタの集合を N 、構成要素数の少ないクラスタの集合を N^- とすると、

$$N' = N - N^- \quad (5)$$

が実際に用いられる負の例の集合となる。以上の様子を図であらわすと、図 1 のようになる。

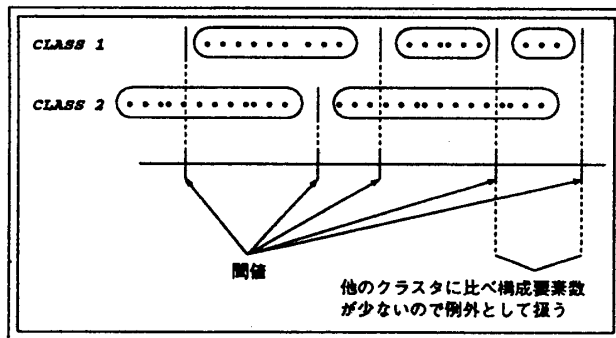


図 1: 閾値の設定と一般化

3.5 例外の取り扱い

Bain[2] によって示されている非単調学習の手法が適用できる。まず、例題をランダムに 2 つに分類し、1 つ

の例題の集合から得られたルールの集合を T_i とする。特殊化すべきルールを $C_i \in T_i$ とし、 C_i によってカバーされる正・負の例の集合をそれぞれ P, N 、次に C_i の変数の集合を V とし、 $r, s \in P, t, u \in N$ とする。ここで、 V を変数として持つ新しい述語 L に対して、代入 $L\theta_r, \dots, L\theta_u$ を行い、 $L\theta_t, L\theta_u$ を正の例、 $L\theta_r, L\theta_s$ を負の例とした一般化を行う。この一般化によって得られた仮説 X に対して X のヘッド部 h を not 述語を用いて C_i に付加する。これを C_{i+1} とすると、

$$C_{i+1} = C_i \vee h \quad (6)$$

となり、新しいルールの集合は

$$T_{i+1} = (T_i - C_i) \cup C_{i+1} \cup X \quad (7)$$

となる。

以上の操作を行うことによって例外の取り扱いが可能となる。

ここで、提案した手法と従来の機械学習の手法との違いをまとめると次のようになる。

- 背景知識を用いているため、問題に関する知識の使用が可能。
- 数値を含む例題からの学習が可能。
- 例外の取り扱いが可能。

という特徴がある。

4 まとめ

本発表では帰納論理プログラミング、及び、機械学習を用いたエキスパートシステムの知識ベースの獲得について説明を行い、次に、機械学習を用いたエキスパートシステムのための知識ベースの獲得の一手法として、帰納論理プログラミングを用いた知識ベースの獲得方法について考察・検討を行った。

今後の課題としては、例外として取り扱うクラスタの構成要素数の検討、および、これらの手法を用いた場合における正答率の比較・検討などが考えられる。

参考文献

[1] M. Anderberg. *Cluster Analysis for Applications*. Academic Press, New York, 1973.
 [2] S. Muttleton ed. *Inductive Logic Programming*. Academic Press, 1992.