

6P-6

# 統計的情報の可視化機構をもつ 帰納論理プログラミングシステムの構築

向川 慎治      大原 剛三      馬場口 登      北橋 忠宏

大阪大学 産業科学研究所

## 1 はじめに

帰納論理プログラミング (Inductive Logic Programming: **ILP**) [1, 3] による知識獲得は、目標概念の正例、負例、及び背景知識が与えられた時に、1つの負例も導くことなく、全ての正例を導くような目標概念を説明するルール (仮説) を獲得するものであり、その獲得方法は仮説候補からなる探索空間の探索問題と捉えることができる [3]。しかし、実世界のデータベースへの適用を考えた場合、仮説候補数が爆発し、探索に膨大な時間がかかる傾向がある。従って、探索空間を縮小する技術が必要となるが、これを論理的、統計的な条件に基づいて、計算機内で自動的に実行するには限界がある。

そこで本研究では、ユーザ自身が個人のもつ経験的知識などに基づいて、仮説生成に必要な属性をある程度取捨選択できるように、学習過程におけるデータの統計的情報を可視化するシステムを提案する。

## 2 可視化システムの概要

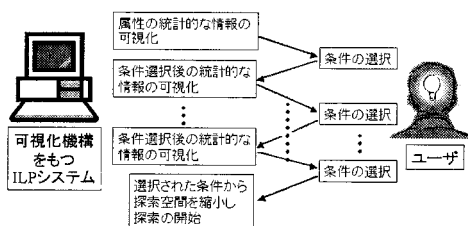


図 1: システムの概要

可視化システムの概要を図 1 に示す。提案システムでは、ILP システムが探索する探索空間のボトムとなる仮説候補に現れる属性に関して、統計的な情報を可視化することで、ユーザがシステムと対話的に概念記述に必要な、不必要な属性を取捨選択することを支援する。提案システムが対象とする ILP システムとして、本研究では筆者らが開発した G-REX [4] を用いている。G-REX は探索空間のボトムとなる最も特殊な仮説 (Most Specific Clause: **MSC**) を Inverse Entailment (**IE**) [2] を用いて作成し、最も評価値の高い仮説 (最適仮説) を見つける

Visualizing Statistical Information in Inductive Logic Programming  
Shinji MUKAIGAWA, Kouzou OHARA,  
Noboru BABAGUCHI, and Tadahiro KITAHASHI  
I. S. I. R., Osaka University

ために、探索空間を最も一般的な仮説候補から特殊な仮説候補へとトップダウンに探索する。

尚、以下では  $p$  を目標概念を表すリテラルとしたとき、 $MSC, p \leftarrow p_1, \dots, p_N$  に関して、その条件部を構成するリテラル  $p_i$  は事例  $X$  が属性  $q_i$  について属性値  $v_i$  をとることを意味する  $q_i(X, v_i)$  という形式であると仮定する。ここで、記号「 $\leftarrow$ 」は含意記号を意味し、以後、 $p_i$  を単に属性と参照する。

## 3 属性の相関度と可視化

### 3.1 属性の相関度

一般に ILP における仮説候補の評価値は、導く正例数が多い仮説候補ほど高くなる。そのため、MSC 中のある属性  $p_k$  が仮説候補の条件部に加えられたときに、導く正例数がどのように変化するかが重要となる。

そこで、2つの属性  $p_i, p_j$  が共通に導く正例の割合に注目し、その割合を  $p_i, p_j$  の相関度  $h$  として、以下のよう

$$h(p_i, p_j) = \frac{p \leftarrow p_i, p_j \text{ の導く正例数}}{p \leftarrow p_i \text{ の導く正例数}} \quad (1)$$

さらに、相関度を用いて MSC の条件部中の各  $p_k$  を特徴づけるために、次のような相関度ベクトルを定義する。

$$p_k(h(p_k, p_1), h(p_k, p_2), \dots, h(p_k, p_N)). \quad (2)$$

この相関度ベクトルは、 $p_k$  が他の属性と組み合わせられたとき、どれだけ導く正例数を変化させるかという特徴を表している。

### 3.2 属性間関係の可視化

前節で定義した相関度ベクトルに基づいて、属性間関係をユーザに直感的に示すため、主成分分析を用いて 2次元平面上に属性を可視化する。主成分分析では、相関度ベクトルで表現された各  $p_k$  を個体、相関度ベクトルの各成分を変量として扱い、その第一主成分を  $x$  座標、第二主成分を  $y$  座標とする。その結果、得られた 2次元平面上では、導く正例数の変化が類似している属性同士が近くに現れることになる。

ユーザは、このように可視化された属性間関係およびデータベースに対する個人のもつ経験的知識に基づいて、仮説に使用する、もしくは使用しないであろう

表 1: 属性  $p_i$  選択時の図の変化

$p_i$ と他の属性の関係 図の変化	相関度が 全体的に低い 直線的 になる	相関度が 全体的に高い 直線的にならず に変化する	相関度の 偏りがない 小さい (変らない)
------------------------	------------------------------	------------------------------------	--------------------------------

属性  $p_i$  を選択する。このユーザの選択に対してシステムは、 $p_i$  が他の属性にどのような影響を与えているかを示すために、以下のように相関度ベクトルを再計算し、可視化した図を更新する。

$$p_k(h(p_k, p_1 \wedge p_i), h(p_k, p_2 \wedge p_i), \dots, h(p_k, p_N \wedge p_i)). \quad (3)$$

ここで、 $h(p_k, p_1 \wedge p_i)$  は式 (1) の右辺の分子に  $p \leftarrow p_k, p_1, p_i$  の導く正例数を用いることを意味する。

上記のように相関度ベクトルを再計算した場合、属性  $p_i$  と他の各属性との相関度に基づき、可視化した 2 次元平面上の属性には、表 1 に示すような変化が生じる。従って、ユーザは選択後の図の変化に応じて、次のように当該属性の使用、削除を判断できる。

属性が直線的になる：  $p_i$  は他の属性と全体的に相関度が低いので、MSC から削除可能であると判断。

直線的にならずに変化：  $p_i$  と他の属性との相関度には明確な特徴が見られないため、削除、使用のいずれともしない。

変化が小さい：  $p_i$  は他の属性と全体的に相関度が高いので、仮説に使用することが妥当であると判断。

## 4 実行例

UCI により公開されているマッシュルームデータベースに対して、提案手法を適用した結果の一部を図 2 に示す。なお、マッシュルームデータベースの詳細は表 2 に示す。

図 2(a) は、学習過程で最初に得られた MSC に関して、その条件部中の 21 個の属性を可視化した結果である。図 2(b) は属性の選択により図が直線的になった例であり、選択した属性は図 2(a) における属性 A である。実際には、属性 A は元の MSC を用いて生成される最適仮説には含まれない属性であり、A の周辺に位置する属性に関しても同様の変化が見られた。次に、図 2(c) は属性の選択により図の変化が小さかった例であり、選択した属性は図 2(a) における属性 B である。B は元の MSC を用いて生成される最適仮説に現れる 4 つの属性の内の 1 つであり、他の 3 つの属性も B の周辺に現れ、それらを含め B の周辺に現れた属性は同様の変化を示した。最後に、図 2(d) は属性の選択により図が直線的にはならないが変化が起こった例であり、選択した属性は図 2(a) における属性 C である。C と同様に属性 A、B いずれの周辺にも位置しない属性に関しても同様の変化が見られた。

以上のように、本実行例は、提案システムにより可視化された図の変化が、最適仮説における属性の使用状況を反映していることを示している。

表 2: マッシュルームデータベース

目標概念	正例数	負例数	属性数	属性値
毒がある	3,916	4,208	22	2~12

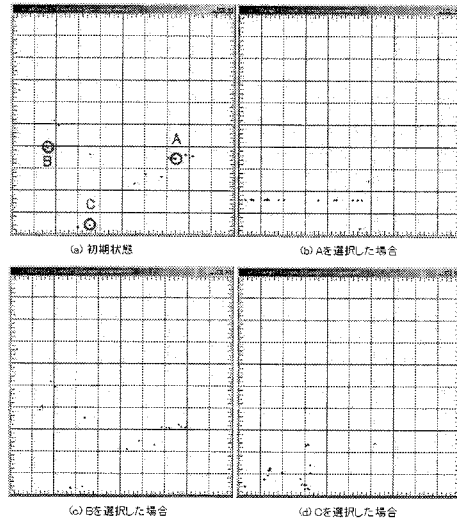


図 2: 各属性を選択した後に可視化した図

## 5 まとめ

本稿では、属性の統計的情報を用いて属性間の関係を可視化することで、ILP における探索空間の縮小を行う手法を提案し、その実行例を示した。提案手法は、大規模なデータベースからの仮説獲得時など、探索数が膨大になる場合において、探索時間を削減する為に有効であると考えられる。

今後の課題としては、ユーザに対して属性の削除、使用の判断に有益なさらなる情報の提示、及び対象とする背景知識の形式の多様化、システムの評価実験などが挙げられる。

## 参考文献

- [1] L.de Raedt : Advances in Inductive Logic Programming, IOS Press(1996).
- [2] S.Muggleton : Inverse Entailment and Prolog, New Generation Computing Journal Vol.13, pp.245-286(1995).
- [3] 古川 : 帰納論理プログラミング-チュートリアル-, 人工知能学会誌 Vol.12, pp.655-664(1997).
- [4] 高, 大原, 馬場口, 北橋 : 例外関係に着目した不完全知識の獲得システムの実装と実験的評価, 人工知能学会研究報告 (SIG-FAI-9803-3), pp.11-17(1998).