

帰納学習ツール I S L によるケース分析 (I)

1 D - 8

— 設計 —

杉本 勉* 大熊 喜之** 大和田 勇人** 溝口 文雄**

*NTTデータ通信(株) **東京理科大学理工学部

1. はじめに

現在、企業において戦略情報システム (S I S) やエキスパートシステムの構築が盛んに行なわれている。戦略情報システムを構築する際には対象を詳細に分析するケース分析によって問題点を抽出し適切な戦略を立案し、システム化することが要求される。この分析結果によってシステムの方向性が左右されるためその責任は大きい。

本研究は帰納学習をこのようなケース分析に適用することによって帰納学習の新たな応用分野を得ることを目的としている。そこで本稿では、その基本設計として適用可能性を明らかにし、帰納学習ツール I S L (Inductive System for Learning) を用いた場合のシステム構成および他手法との比較を行う。

なお、実現方法は[1]を参照のこと。

2. ケース分析と帰納学習

ケース分析とは集められた事例データがどのような傾向をもっているのかを知り、対象の全体像や特徴を把

握することが中心である。ケース分析にはこれまで、定量データには多変量解析や主成分分析などの統計アプローチが、非定量データには I S M (Interpretive Structure Modeling, 階層構造化モデル) や A H P (Analytic Hierarchy Process, 階層分析法)、数量化理論などが用いられてきた。

帰納学習は与えられた事例データから何らかの規則を導き出すことであり、それを新たな知識の学習としている。帰納学習はエキスパートシステム構築における知識獲得ボトルネックが指摘され始めた頃から活発に研究されている。帰納論理プログラミング (Inductive Logic Programming, 以下 I L P) は帰納学習と論理プログラミングを融合した枠組であり、QuinlanのFOILやMuggletonのGOLEMなどがこれを実現するシステムとして知られ、知識ベースの自動構築や自動プログラミングに適用されている[3]。I L Pと先に述べたいくつかの方法との比較を表1に示す。

本研究は、帰納学習の新たな適用分野としてケース分析を提案するものである。その主な理由は、事例データから一般的な知識を得る帰納学習は、事例データの特徴や規則性を把握するケース分析と見なすことができるからである。

表1 他の分析手法との比較

| | 帰納論理プログラミング | 多変量解析 | I S M | A H P |
|--------|------------------|--------------------------|---------------------|-------------------------------|
| 特徴 | 事例データから関係やルールを導出 | 統計数学によって複雑な現象を解析する手法の集まり | 要素間の階層構造をアルゴリズム的に導出 | 解候補の選好順序を主観的判断とシフトアプローチによって導出 |
| 入力 | 正/負事例データ 背景知識 | 観測データ | 有向グラフ隣接行列 | 問題決定要素 解候補 |
| 出力 | 関係/ルール | 相関係数 回帰直線など | 多階層構造 | 選好順序 |
| データタイプ | 定量/非定量 | 定量 | 非定量 | 非定量 |
| 基本手法 | 一般化 逆導出 | 統計数学 | 一対比較 行列計算 | 階層化 ペア比較 固有値 |
| 長所 | 論理的に健全 | 最も使われている | 汎用性が高い | データがないときに有効 |
| 短所 | ノイズを含むデータの扱いが困難 | 手法が難解 | 関係の強さが評価に使われない | 階層構造の決定が主観的 |

3. I S L を使ったシステム構成

I L P で用いられているアルゴリズムは論理プログラミングにおける融合原理の逆操作を定式化したものである。

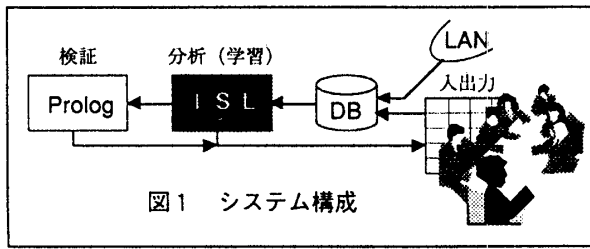
帰納学習ツール I S L はGOLEMのアプローチを多ソート帰納学習の観点で拡張したものであり、以下の機能を持っている[2]。

- (1) 数値データの帰納学習
- (2) 記号データの帰納学習
- (3) 大規模数値データの帰納学習

観測言語は事例の集合からなり、事例は属性-値のベ

An application of inductive learning tool, ISL, to knowledge analysis and synthesis

Tsutomu SUGIMOTO, NTT DATA Comm. Sys. Corp.
Yoshiyuki OOKUMA, Hayato OHWADA, Fumio MIZOGUCHI, Science University of Tokyo



アである。それぞれの事例にはクラスが与えられている。

仮説言語はプロダクションルールの集合であり、条件部は属性が満たすべき条件の連言によって表され、結論部はクラスである。

図1はシステム構成を示している。ISLは分析結果をPrologプログラムとして出力するので結果の検証系としてPrologとリンクしている。データの inputs は応用例のようにオンラインで得る場合とユーザが入力する場合がある。

4. 他手法との比較

4.1 多変量解析とILP

多変量解析はもっとも確立している分析手法の一つである。これは相関係数や回帰分析などによって集められた数値データに隠された特徴を浮かび上がらせる方法である。例えば、多変量解析で表2のようなデータを解析する場合、各クラスごとに回帰直線を求める(図2)。ILPは例えば「 $(X < 3 \text{ かつ } Y > 1)$ または $(Y > 2)$ ならばクラスa。それ以外はクラスb」というルールを導出する。このルールは各クラスが属する空間(凸包)を表す(図3)。これは新たに入力されたデータがどのクラスに属するかを求める際に有効である。すなわち、多変量解析はデータをもっとうまく表現する方程式を求めるのに対し、ILPは分類規則を求める。

4.2 ISM, AHPとILP

ISMは数学モデルを用いてより客観的な方法で階層構造を決定する。具体的には要素間ごとに影響があるかないかを関係行列として表しこれから可到達行列を求め、さらに構造化行列を求めて階層構造を決定する。AHPは問題を最終目標、評価基準、代替案の三階層にして代替案の重要度を計算しそれらの順位を導くものである。これら2つとILPに共通していることはいずれも非定量データを扱うことで

表2 データ

| No | 属性 | | クラス |
|----|----|---|-----|
| | X | Y | |
| 1. | 1 | 2 | a |
| 2. | 2 | 1 | b |
| 3. | 2 | 3 | a |
| 4. | 3 | 2 | b |
| 5. | 3 | 3 | a |
| 6. | 4 | 2 | b |

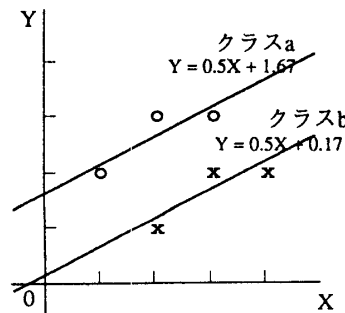


図2 多変量解析の場合

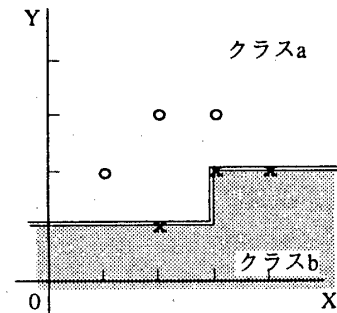


図3 ILPの場合

ある。もっとも大きな違いは前者は事例データがないまたはほとんどない場合に使われるということである。後者は事例データがある場合に使われる。これは、AHPでは階層構造と一対比較などで代替案の重要度を計算するのに対して、ILPは事例データから各代替案が選ばれるルールを抽出することに注眼をおくためである。また、ISMは階層構造を求めるのに対してILPでは各要素間の関係を求めることが異なる。階層構造も要素間の上下関係に対応付けできるのでISMが求める階層構造も事例データがあればILPで求められる。

4.3 比較のまとめ

目的によって使う手法は当然異なるが、事例データが十分ある場合はILPや多変量解析を、ないまたは少ない場合はISMやAHPと分類できる。

5. おわりに

帰納学習をケース分析に適用することを提案した。ケース分析のなかでも、事例データの分類規則を求める場合にILPは有効である。

これまでにSIS向けのケース分析と研究室LAN管理支援に本アプローチを適用した[1]。その結果、事例データからの規則性の抽出に対する有効性は確認できた。しかし、以下の問題点が明らかになり今後検討する予定である。

- (1) どの知識に焦点を当てるかわからない。すなわち、知識間の重み付けやプライオリティがない。
- (2) ISMやAHPに比べ結果の解釈が難しい。
- (3) ユーザにとって必要でない知識も導出される。

参考文献

[1] 大熊他: 帰納学習ツールISLによるケース分析 (II) - 実現方法 -, 情報処理学会第46回全国大会予稿集, 1993/3.
 [2] 大和田他: 多ソート帰納学習システムの設計 (I) - 一般化アルゴリズム -, 情報処理学会第46回全国大会予稿集, 1993/3.
 [3] S.Muggleton, Inductive Logic Programming, in S.Muggleton (ed.) Inductive Logic Programming, Academic Press, 1992/6.
 [4] 木下, わかりやすい意思決定論入門, 啓学出版, 1992.