

帰納論理プログラミングと関係データベースの結合*

1 J-4

溝口文雄† 大和田勇人† 石井雅子†

東京理科大学 理工学部‡

1 はじめに

近年、データベースが大規模化している。データベースの効率的利用にとって、内在する特性をルール化したり概念クラスに分類することは重要であるとともに、表面に現れない隠れた意味的構造を発見することも価値ある知識を獲得するという点で重要である。また最近、帰納論理プログラミング(Inductive Logic Programming)が注目されている。帰納論理プログラミングは機械学習と論理プログラミングを結ぶものであり、論理プログラムの形式で記述された例と背景知識から、第一階確定節を構成するものである[3]。論理プログラミングの特徴は、関係を記述する点である。論理プログラムで表現された述語形式のデータベースに着目した場合、関係データベースとしてとらえることが可能である。本稿では帰納論理システムにおける相対最小汎化と関係データベースにおける結合演算との間の類似性とともに、関数従属性[1]との関連を示し、関係データベースの枠組での帰納的一般化法を提案する[2]。本方法により、従来のドラッグデザインやレイアウト問題などにおける大規模データベースが関係データベースの枠組で扱うことが可能となり、そのため効率的に一般化規則が導出できる。

2 関数従属性による閉包

関数従属性(Functional Dependency)とは、ある関係データベースにおいて属性集合 $\{A_1, A_2, \dots, A_l, B_1, B_2, \dots, B_m\}$ が与えられている場合に

$$A_1, A_2, \dots, A_l \rightarrow B_1, B_2, \dots, B_m$$

が一意に決定されることをいう。またこの場合 A_1, A_2, \dots, A_l を決定子、 B_1, B_2, \dots, B_m を被決定子と呼ぶ。また、ある関係データベースに既知の関数従属性の集合 F があるとする。 X, Y を属性集合として、例えばその F に $X \rightarrow Y$ なる関数従属性が成立するかどうかは X を決定子として F により被決定子を全て求め(これを F のもとで

の X の閉包といい X^+ で表す)、 X^+ に Y が含まれているかどうかを調べればよい。

3 相対最小汎化と結合演算

帰納論理プログラミングにおいて、相対最小汎化(Relative Least General Generalization)について述べる。相対とは与えられた背景知識の元でという意味である。二つの概念 C_1 と C_2 の最小汎化は、その両概念を包摂する概念の中で最小なものと定義されるが、背景知識 K の元での C_1 と C_2 の相対最小汎化は、 K に付け加えるとそこから演繹的に C_1 と C_2 がともに導かれるような θ -包摂中の概念のなかで最小なものと定義される。

関係データベースの枠組で結合演算(join operation)を用いることによる相対最小汎化について、関数従属性との関連から述べる。

事例集合を

$$E = \{\wedge_j A_j, B\}$$

とする。ただし、ここで $\wedge_j A_j$ を入力属性、 B を出力属性とする。また背景知識を次のような関数従属性をもつものとする。

$$C_1 \leftarrow \wedge_{m_1} D_{1m_1} \quad C_2 \leftarrow \wedge_{m_2} D_{2m_2} \quad \dots$$

ここで、 $\wedge_{m_1} D_{1m_1}, \wedge_{m_2} D_{2m_2}$ を入力属性、 C_1, C_2 を出力属性とする。

$$\wedge_{m_1} D_{1m_1} \subseteq \wedge_j A_j$$

であるとして、関数従属性 $C_1 \leftarrow \wedge_{m_1} D_{1m_1}$ を用いて事例と背景知識を結合すると、

$$C_1 \leftarrow (\wedge_{m_1} D_{1m_1} \wedge \wedge_j A_j)$$

となり、ここで $C_1 \wedge B = \emptyset$ ならば、 C_1 も入力属性として新たに加えて

$$(C_1 \vee \wedge_j A_j) = \wedge_{j_1} A_{1j_1}$$

とおく。そしてさらに

$$\wedge_{m_2} D_{2m_2} \subseteq \wedge_{j_1} A_{1j_1}$$

であるとして、関数従属性 $C_2 \leftarrow \wedge_{m_2} D_{2m_2}$ により事例と背景知識を結合すると、

$$C_2 \leftarrow (\wedge_{m_2} D_{2m_2} \wedge \wedge_{j_1} A_{1j_1})$$

となる。そして、 $C_2 \wedge B \neq \emptyset$ ならば、

$$(B \leftarrow (\wedge_{m_2} D_{2m_2} \wedge \wedge_{j_1} A_{1j_1}))$$

$$\wedge (C_1 \leftarrow (\wedge_{m_1} D_{1m_1} \wedge \wedge_j A_j))$$

*Unifying Inductive Logic Programming and Relational Database

†Fumio MIZOGUCHI, Hayato OHWADA, Masako ISHII

‡Faculty of Sci. and Tech., Science University of Tokyo

となり、GOLEM[3]における決定性という制限を用いることにより、意味のないリテラルを削除し相対最小汎化が導出できる。

この過程においては、まず背景知識のもとで $\wedge_j A_j$ から結合演算により閉包を求めている。そして、この閉包のうち出力属性と一致するものがあるとき、事例における入出力属性間の関数従属性が求められている。そのことにより相対最小汎化が導出可能である。よって、関係データベースにおける結合演算は、帰納論理プログラミングにおける相対最小汎化と一致している。つまり、関係データベースからの帰納的一般化において、ある関係と他の関係の結合により、相対最小汎化が実行可能である。

例として第一引数と第二引数の積算結果が第三引数となる $mult/3$ をあげる。事例関係データベースとして図 1 における属性名 A, B, C の $mult/3$ を、そして背景知識関係データベースとして属性名 DA, DB の $dec/2$, MA, MB, MC の $mult/3$, PA, PB, PC の $plus/3$ を与える。ここで、 $dec/2$ は第一引数から 1 引いた減算結果を第二引数とするものであり、 $plus/3$ は第一引数と第二引数の加算結果を第三引数とするものである。

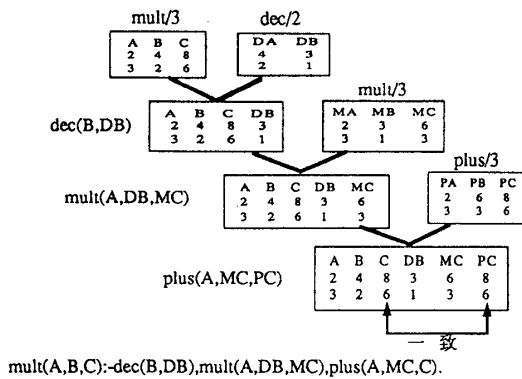


図 1: $mult/3$ の相対最小汎化の実行過程

以下のような結合で、相対最小汎化が導出される。

- B と DA について $\rightarrow dec(B, DB)$
- A と MA 及び DB と MB について $\rightarrow mult(A, DB, MC)$
- A と PA 及び MC と PB について $\rightarrow plus(A, MC, PC)$

よって、この場合 $mult/3$ の相対最小汎化は次のようになる。

$$mult(A, B, C) :- \begin{matrix} dec(B, DB), \\ mult(A, DB, MC), \\ plus(A, MC, C). \end{matrix}$$

つまり、 $B \rightarrow DB$ と $A, DB \rightarrow MC$, 及び $A, MC \rightarrow C$ という関数従属性が得られている。

4 アルゴリズム

関係データベースの枠組での相対最小汎化の導出についてのアルゴリズムの概略を述べる。まず事例関係デー

タベースと背景知識関係データベースの結合を行なう。各導出表の事例数、出力変数の決定数、及び“評価値 = 結合による事例数 - ステップ数”をチェックする。そして、出力変数の決定数の最大数が結合前の事例数と一致していれば、第 1 一般化規則として導出し、第 1 一般化規則によって一般化不可能である事例関係データベースを基底事例として残す。そうでなければ、評価値が最大の事例関係データベースを選択して、あらかじめ設けた制限回数を越えていなければ、再び同様な処理を繰り返す。制限回数を越えていれば、第 1 一般化規則によって一般化不可能である事例関係データベースを基底事例として残す。

5 アルゴリズム評価

本アルゴリズムにより GOLEM が一般化可能であるような基本的論理プログラム及び、補助的な論理プログラムが必要である $append/3$ なども制約関数なる組み込み関数を利用することにより導出可能である。また大規模データベースを必要とする、ドラッグデザイン、レイアウト問題などにおいて、関係データベースの枠組による高速化技術を利用することにより効率的に一般化規則が導出できる。

6 おわりに

本稿では、帰納論理プログラミングにおける相対最小汎化を用いた関係データベースからの帰納的一般化法を提案した。本方法では、論理プログラムを関係データベースとしてとらえることにより、帰納論理プログラミングにおける相対最小汎化を、関係データベース間の結合として実行する。それは関数従属性と深い関わりがある。関係データベースの枠組を利用することにより、大規模データベースからの帰納的一般化が効率的に実行できる。また、背景知識として単なる事例集合のみではなくてユーザによる制約関数を用いて入力として定義された変数値から新たな属性を導出することも可能である。

参考文献

- [1] E.F.Codd: Further Normalization of the Data Base Relational Model, Data Base Systems, Courant Computer Science Symposia 6, R.Rustin 編, Prentice-Hall, Englewood Cliffs, N.J., pp.33-64, 1972.
- [2] 大和田 勇人, 石井 雅子, 溝口 文雄: 帰納論理システムによる発見的学習～関係データベースからの帰納的一般化～, 人工知能学会全国大会 (第 8 回) 論文集, pp133-136, 1994.
- [3] S.Muggleton: Inductive Logic Programming: Theory and Methods, 1993.(to appear).