5R－4

極小多重汎化と評価関数による論理プログラム頭部の推測

森田將敬　山本章博
北海道大学工学部

1. はじめに

論理プログラムを用いた帰納推論では，最小モデルの部分集合を与えてそれを説明するプログラムを構成する．その場合，プログラムの頭部を推測することは重要な操作の一つである．頭部を推測する方法として，本稿では極小多重汎化アルゴリズム[1]と評価関数を組合わせた方式を考案し，その方式に基づいて実際に計算機上で行った推測実験とその結果について報告する．

2. 極小多重汎化

準備として木パターンに関するいくつかの用語を[1]に基づいて定義しておく．Σを関数記号の有限集合，Xを変数の集合とし，T_ΣをΣを用いた一階論理の項全体の集合，TP_Σを$\Sigma \cup X$を用いた一階論理の項全体の集合とする．T_Σの要素を定木，TP_Σの要素をΣに関する木パターンという．TP_Σ上の二値関係$\leqq'$を$p=\theta(q)$なる代入θが存在するとき$p \leqq' q$と定義する．木パターンpによって定義される言語とは集合$L(p)=\{t \in T_\Sigma | t \leqq' p\}$である．$T_\Sigma$の部分集合$L$が$\Sigma$に関する木パターン言語であるとは，$p \in TP_\Sigma$が存在して$L=L(p)$のときをいう．木パターン言語の和とはある木パターンの組$\{p_1,\ldots,p_n\}$が存在して$L=L(p_1)\cup\cdots\cup L(p_n)$となることをいい，言語$L$を$L(\{p_1,\ldots,p_n\})$と書く．また$\Sigma$上の$k$個の木パターン言語の和からなる族を$TPL^k$で表す．

Plotkinの最小汎化では与えられた定木の集合を一つの木パターンで一般化するのに対して，k-極小多重汎化はk個の木パターンで一般化する．

定義　Sを定木の集合とする．木パターンの組$\{p_1,\ldots,p_k\}$は，その定義する言語$L(\{p_1,\ldots,p_k\})$がSを含むTPL^k中の極小言語であるとき，Sのk-極小多重汎化(k-*mmg*)という．

一般に定木の有限集合Sの極小多重汎化は一意には定まらない．k-*mmg*アルゴリズム[1]はSのk-極小多重汎化が一つ以上存在する場合にそのうちの一つを計算する．Sの全てのk-極小多重汎化を求めるには，異なるk-極小多重汎化が求まらなくなるまでk-*mmg*アルゴリズムを繰り返すことで実現できる．

3. 評価関数

$P=\{C_1,\ldots,C_n\}$を論理プログラム，SをPの最小モデルの部分集合とする．ただし$C_i(1\leqq i\leqq n)$は節を表す．$head(C)$を節Cの頭部とするとき，

$$S\subseteq L(\{head(C_1),\ldots,head(C_n)\})$$

なる$\{head(C_1),\ldots,head(C_n)\}$に等しい$k$-極小多重汎化$TP^k=\{p_1,\ldots,p_k\}$を推測するため，$TP^k$に関する評価関数を三つ定義する．以下では$\sigma^2(x_1,\ldots,x_n)$を$x_1,\ldots,x_n$の分散値とする．

(1)　$SIZE(TP^k)=\sigma^2(size(p_1),\ldots,size(p_k))$

$$size(p)=os(p)-ov(p)$$

$os(p)$: pに出現する記号の数
$ov(p)$: pに出現する変数の種類数

ただし$size$はReynolds[2]による定義である．

(2)　$OVERLAP(TP^k)$
$=\sigma^2(overlap(p_1),\ldots,overlap(p_k))$

$$overlap(p) = \sum_{i=1}^{n} \frac{1}{nv_i(p)}$$

$nv_i(p)$: pに含まれる変数v_iの個数
v_i : pに含まれるn個の変数$(1\leqq i\leqq n)$

(3)　$COMB(TP^k)$
$=\sigma^2(SIZE(TP^k), OVERLAP(TP^k))$

An Inference Method for Heads of Logic Programs with the Minimal Multiple Generalization and Evaluations
Masahiro MORITA, Akihiro YAMAMOTO
Hokkaido University

4. 実験

著者らの推論方式は論理プログラムの最小モデルの部分集合Sを例として***k*-mmg**アルゴリズムに与えて，求められたSのk-極小多重汎化に評価関数を適用することで論理プログラムの頭部に相当する極小多重汎化を推測するものである．

4-1. 実験の対象

実験は論理プログラムとしてPrologに関する様々な文献に掲載されているプログラムのうち二つの節からなり，

$$\begin{cases} append([\,],X,X). \\ append([A|X],Y,[A|Z]) \leftarrow \\ \qquad\qquad append([A|X],Y,[A|Z]). \end{cases}$$

$$\begin{cases} even(0). \\ even(s(s(X))) \leftarrow even(X). \end{cases}$$

のようなリストや後者関数に関する純Prologプログラム37個について頭部の推測実験を行った．

4-2. 2-mmgに関する実験

はじめに例Sとしてプログラムから最小モデルの部分集合であるアトムの集合を生成した．これは適当な2-極小多重汎化(以下2-*mmg*とする)が得られるようにプログラムごとにアトムを100個程度用意した．次に2-*mmg*アルゴリズムにSを入力として与えてSの2-*mmg*を全て求め，その中にプログラムの頭部に相当する2-*mmg*が含まれているかを調べた．その結果，33個のプログラムについて頭部に相当する2-*mmg*が含まれていた．

4-3. 評価関数に関する実験

頭部に相当する2-*mmg*が含まれているプログラム33個のうち複数の2-*mmg*が求められた24個について，プログラムごとに各2-*mmg*に評価関数を適用してその値を比較した．その結果，評価値最小の2-*mmg*が頭部に相当している場合が*SIZE*で12個，*OVERLAP*で18個，*COMB*が18個と高い割合でみられた．そこでさらに三つの評価関数のうちの二つ以上が最小値である2-*mmg*と頭部との関係を調べた(表1)．ここで比較不能であるとは評価値が一致して二つ以上の最小値をもつ2-*mmg*がない場合である．

実験対象のプログラム数	24個
2-*mmg*が頭部に相当する数	19個
2-*mmg*が頭部に相当しない数	2個
比較不能である	3個

表1　評価関数の最小値による評価

5. まとめ

実験の結果をもとに，次のような論理プログラムの頭部の推測方式を提案する．

入力は最小モデルの部分集合S，出力はSを説明する論理プログラム$P=\{C_1,C_2\}$(ただしC_1,C_2は節を表す)の頭部の対$\{head(C_1), head(C_2)\}$である．

1. Sを入力として***k*-mmg**アルゴリズムにより全ての2-*mmg*を求める．
2. 全ての2-*mmg*について評価関数*SIZE*, *OVERLAP*, *COMB*を適用する．
3. 三つの評価関数のうち二つ以上が最小値を示す2-*mmg*を出力する．

この方式は，4節で述べた結果をまとめると表2となり，七割以上の確率で正しい論理プログラムの頭部を推測できる．

実験対象のプログラム数	37個
頭部を正しく推測できた数	28個
正しく推測できた割合	75.7%

表2　本方式による推測

参考文献

[1]Arimura, H., Shinohara, T. and Otsuki, S. : A Polynomial Time Algorithm for Finding Finite Unions of Tree Pattern Languages, LNAI 659, 118-131. Springer-Verlag, 1991.

[2]Reynolds, J.C.: Transformational Systems and the Algebraic Structure of Atomic Formulas, Machine Intelligence 5, 135-151. Edinburgh University Press, 1970.