2J−6

# 帰納論理プログラムにおける背景知識を用いた多項式時間一般化アルゴリズム

有村博紀, 篠原武

九州工業大学情報工学部

## 1 はじめに

機械学習や帰納推論, 仮説推論等における重要な推論手法の一つが, 概念の汎化である. 概念の汎化とは, いくつかの具体例が与えられたときに, それらを説明するより一般的な概念を見つけることであり, 構成的学習において有効な手法である. とくに帰納論理プログラミング等の論理プログラムの学習の分野では, 汎化にもとづく学習システムについて活発な研究がおこなわれている.

本研究では, 論理プログラムを背景知識の存在のもとで汎化する問題について議論する. とくに, 高々 $k$ 個の制約確定節からなる集合のうちで, 例として与えられたすべての確定節を導き, 背景知識 $B$ のもとで含意に関して極小なものを見つける問題を考察し, この問題を多項式時間で解くアルゴリズムを与える. これは [2] の結果を複数の節からなる論理プログラムに拡張するものである. さらに, 論理プログラムの学習への応用について述べる.

## 2 複数の節による汎化

ページ＆フリッシュ [2] は, 論理式の集合 $B$ を背景知識として仮定したときに, 例として与えられたすべての確定節を含意し, 同時に背景知識のもとで論理的に最小な確定節を多項式時間で計算するアルゴリズムを与えた.

ここで「象はフレッドを押しつぶす」のような知識を `crush(X,fred):- ELEPHANT(X)` のといった確定節で表現するものとし, 背景知識 $B$

```
ELEPHANT(jumbo).  HUMAN(fred).
ELEPHANT(clyde).  HUMAN(joe).
VEHICLE(car1).  HOUSE(fred's).
VEHICLE(car2).  HOUSE(joe's).
VEHICLE(jumbo).
ELEPHANT(son(X)):- ELEPHANT(X).
HUMAN(son(X)):- HUMAN(X).
BIGGER(X,Y):- ELEPHANT(X),HUMAN(Y).
BIGGER(X,Y):- VEHICLE(X),HUMAN(Y).
BIGGER(X,Y):- HOUSE(X),HUMAN(Y).
```

と述語 crush に関する例の集合 E

```
crush(X,fred):- ELEPHANT(X).
crush(clude,Y):- HUMAN(Y).
crush(jumbo,Y):- HUMAN(Y).
crush(car1,Y):- HUMAN(Y).
crush(car2,Y):- HUMAN(Y).
```

が与えられたと仮定する. このとき, ページ＆フリッシュのアルゴリズムによって例の集合 $E$ を単一の確定節で汎化するとつぎの確定節を得る.

`crush(X,Y):- BIGGER(X,Y).` $(c)$

「大きなものはより小さいものを押しつぶす」

しかし, 未知の概念が複数の確定節からなるときには, この確定節が未知の概念を過剰に一般化している可能性がある. そこで, 例の集合 $E$ を高々2つの確定節で汎化し, 背景知識のもとでできるだけ具体的な汎化を求めてやると, つぎの確定節 $c1, c2$ を得る.

`crush(X,Y):- ELEPHANT(X),HUMAN(Y).` $(c_1)$

`crush(X,Y):- VEHICLE(X),HUMAN(Y).` $(c_2)$

「象は人を押しつぶし, 乗り物も人を押しつぶす」

背景知識によって象は人よりも大きく, また乗り物も人よりも大きいことからわかるように, 節 $c1, c2$ は節 $c$ よりも含意に関して真に具体的であり, より詳細な概念を表現している. このような高々 $k$ 個の節からなる集合で, すべての例を導き, 背景知識のもとで含意について可能な限り具体的なものを, ここでは $k$ 相対極小汎化と呼ぶ.

## 3 確定節上の汎化関係

確定節 $c = \alpha \leftarrow \beta_1, \ldots, \beta_n (n \geq 0)$ は, 節本体に出現する任意の項が, 部分項として節頭部に出現するならば, 遺伝的であるという. ここに, $\alpha, \beta_1, \ldots, \beta_n$ は原子式であり, $head(c)$ で節の頭部 $\alpha$ を, $body(c)$ で節の本体の集合 $\{\beta_1, \ldots, \beta_n\}$ をそれぞれ表わす. 本稿では, 遺伝的な節のみを考える.

**背景知識**とは任意の遺伝的確定節の集合 $B$ である. 遺伝的節 $\alpha \leftarrow \beta_1, \ldots, \beta_n (n \geq 0)$ は, 本体のすべての述語が $B$ に出現しており, 頭部の述語が $B$ に出現しないならば, $B$ における**制約確定節**と呼ばれる. 定義から, 明らかに制約確定節は非再帰的な節である.

直感的には, 背景知識 $B$ のもとで節 $c$ が節 $d$ よりも一般的であるのは, $B$ のもとで $c$ が $d$ より多くの論理式を含意するとき, すなわち, $B \cup \{c\} \models d$ が成立するときと考えられる. しかし, 意味的な汎化を求めることは, 構文的な汎化に比べて難しいことが多い.

そこで, バンティン [1] は, 構文的な汎化関係である $\theta$

---

A polynomial-time relative generalization algorithm for learning logic programs with background knowledge
Hiroki Arimura and Takeshi Shinohara
Department of Artificial Intelligence, Kyushu Institute of Technology, Kawazu 680-4, Iizuka 820, Japan

包摂を

$$c \succeq_B d \iff (\exists\theta : 代入)\ head(c)\theta = head(d),\ かつ\ B \cup body(d)\phi \models body(c)\theta\phi$$

によって背景知識のもとでの汎化関係 $\succeq_B$ に拡張した．ここに，$\phi$ は $d$ 中の異なる変数をすべて異なる新しい定数で置き換えるような代入である．

ページ＆フリッシュ [2] は，制約確定節上に制限したとき，この関係 $\succeq_B$ が背景知識のもとでの論理的含意に一致することを示した．さらにこのとき，例の集合 $S$ が与えられた場合に，$\succeq_B$ に関する相対最小汎化 $lgg_B(S)$ が常に存在し，$S$ のサイズの多項式時間で計算可能であることを示した．この関係 $\succeq_B$ は，

$$C \sqsupseteq_B D \iff (\forall c \in C)(\exists d \in D)\ d \succeq_B c$$

によって節集合の間の汎化関係 $\sqsupseteq_B$ に自然に拡張できる．$C \sqsupseteq_B D$ のとき，背景知識 $B$ のもとで $C$ は $D$ より一般的であるという．定義より $C \sqsupseteq_B D$ ならば $B \cup C \models D$ が成立する．制約確定節に対してはこの逆も成り立つ．

## 4 汎化手続きと正当性，時間計算量

以下に説明するアルゴリズム **RMMG** (relative minimal multiple generalization) は，例の集合 $S$ の相対最小汎化 $lgg_B(S)$ から始めて，各確定節を具体化したり，あるいは分割したりしながら，すべての例を含意する極小な相対汎化を求める．ここで用いられる確定節の基本操作 $\rho$ は精密化演算子とよばれ，関係 $\sqsupseteq$ を一般なものから具体的なものへと段階的に計算するものである．確定節 $c$ に対して，以下のいづれかの操作を適用して得られる確定節 $d$ 全体を $\rho(c)$ とする．

- $d$ は $c$ 中のある変数 $x$ を $f(y_1, \ldots, y_n)$ の形の項 ($y_1, \ldots, y_n$ は新しい変数) で置き換えるか，$c$ 中の異なる変数 $x$ と $y$ を単一化して得られる確定節．
- $d$ は $c$ の本体に，$B \cup body(c)$ から含意されないような原子式 $\beta = p(t_1, \ldots, t_n)$ を付け加えて得られる確定節．ただし，$p$ は $B$ で使われている述語記号であり，$t_1, \ldots, t_n$ は頭部 $head(c)$ の部分項である．

このとき，$c$ より真に具体的な任意の節は，$c$ に $\rho(c)$ を有限回適用することで得られる．さらに任意の $e \in \rho(c)$ に対し，$e \succeq d \succeq c$ ならば $e = d$ または $d = c$ であり，$e$ と $c$ の中間の節は存在しない．$\rho(c)$ は $c$ のサイズの多項式時間で計算できる．

節集合 $P$ が $S$ に関して既約であるとは，$B \cup P \models S$ かつ任意の $P' \subset P$ に対して $B \cup P \not\models S$ であることをいう．$B \cup \{c\} \models S$ であるような節 $c$ が $S$ に関して $k$ 分割可能であるとは，高々 $k$ 個の $p$ より具体的な節の集合 $P$ で，$B \cup P \models S$ をみたし，既約なものが存在することをいう．そうした $P$ を $p$ の $S$ に関する $k$ 分割という．

**RMMG**$(k, B, S)$;
  $P := \{lgg_B(S)\}$; $\Delta k := k$;
  **while** $\left(\begin{array}{l}\Delta k \geq 2,\ かつ\ S - L(P \setminus p)\ に関\\ して\ \Delta k\ 分割可能な\ p \in P\ が存在\end{array}\right)$ **do**
    $\Delta k$ 分割可能な $p \in P$ を選ぶ;
    $\Delta S := S - L(P - \{p\})$;
    $\Delta P :=$ **Divide**$(p, \Delta k, \Delta S)$;
    $\Delta P :=$ **Tighten**$(\Delta P, \Delta S)$;
    $\Delta P := (P - \{p\}) \cup \Delta P$;
    $\Delta k := k - \sharp P + 1$;
  **endwhile**
  return $P$;

**Divide**$(p, k, B, S)$;
  $\rho_B(p)$ の部分集合で，$\sharp P \leq k$ かつ $P$ が $S$ に関して既約なものを選ぶ;
  return P;

**Tighten**$(P, B, S)$;
  **while** 精密化可能な $p \in P$ が存在 **do**
    $\Delta S := S - L(P - \{p\})$ ;
    $q := lgg_B(\Delta S)$;
    $P := (P - \{p\}) \cup \{q\}$;
  **endwhile**;

図 1: 多項式時間相対汎化アルゴリズム

図 1 に $k$ 相対極小汎化を求めるアルゴリズムの概略を示す．**RMMG**$(k, B, S)$ は，節集合 $S$ の $k$ 相対極小多重汎化を求める．**Divide**$(p, k, B, S)$ は，$p$ の $S$ に関する $k$ 分割を求める．**Tighten**$(P, B, S)$ は，$P$ 中の節を $B \cup P \models S$ である限りできるだけ具体的なものに置き換えた集合を求める．

**定理 1** $B$ を背景知識とする．確定節の有限集合 $S$ が例として与えられたとき，図 *1*のアルゴリズム **RMMG** は，高々 $k$ 個の制約確定節からなる集合のうちで，$S$ のすべての例を導き，$B$ のもとで含意に関して極小なものの一つを，$S$ の例のサイズの総和の多項式時間で計算する．

含意にもとづく学習 (learning from entailment) [2] では，例として任意の確定節 $c$ を用いる．例 $c$ は，$B \cup P \models c$ が成立するとき $P$ の含意による正例という．

**定理 2** $B$ を背景知識とし，自然数を $k$ とする．このとき，高々 $k$ 個の制約確定節からなる任意の集合 $P$ を，無矛盾で保守的な多項式時間仮説更新によって含意による正例から極限同定する学習アルゴリズムが存在する．

## 参考文献


[1] W. Buntine. Generalized subsumption and its applications to induction and redundancy. *Artificial Intelligence*, **36**, 149–176, 1988.

[2] C. D. Page Jr. and A. M. Frisch. Generalization and Learnability: A Study of Constrained Atoms. *Inductive Logic Programming*, 29–61, Academic Press,1992.