

## 2C-5 概念の重みづけを用いた帰納推論アルゴリズム — 評価

野呂 寛洋 大野 泉 正木 彰一 熱田 清明 峯崎 俊哉 近藤 正三  
東海大学

### 1. はじめに

著者らは概念の重みづけを用いた帰納推論アルゴリズムを提案している [1]。それは Michalski の S T A R 方法論に基づいている [2]。提案したアルゴリズムでは共有度、典型度という2つの尺度を導入し、概念の質的評価を与えている。

本件では、提案したアルゴリズムと S T A R 方法論に関して、解として得られた概念記述の質、安定性、および概念記述が得られるまでの速度について比較評価を行い、その結果をまとめる。

### 2. 概念の重みづけ

#### 2. 1 帰納推論アルゴリズム

帰納推論とはいくつかの事例から一般法則、すなわち概念記述を導き出す過程を意味する。

著者らが提案した帰納推論アルゴリズムは Michalski の S T A R 方法論 (Star Methodology) に基づいている [2]。帰納推論過程では P O S (正の事例) と N E G (負の事例) の記述の集合が用いられる。事例を記述する言語は A P C (Annotated Predicate Calculus) とよばれる。A P C は帰納推論に適するように、通常の述語論理に新しい形式を付加して表現力を増加させたものである。事例は A P C を用いてセレクトの連言で記述される。セレクトとは特徴記述の形式である。S T A R 方法論ではまず P O S からランダムに事例を選択し、その事例を形作っているセレクトに分解する。分解されたセレクトに一般化規則等を適用し、解となる概念記述を構成して行く。これらの概念記述をある評価基準に基づいて並べ替え、一番目の概念記述を取る。この推論の最初における事例の選択は過程の速度、解となる概念記述の質に大きな影響を与える。

提案したアルゴリズムでは、セレクトに関する尺度である共有度と事例に関する尺度である典型度という二つの尺度を用いている。S T A R 方法論がランダムに事例を選んでくるのに対し、提案したアルゴリズムでは典型度を用いて事例を選択する。この典型度を用いた選択によって、解となる概念記述の安定性が得られ、質を制御することができる。また、概念記述を得るまでの速度を一定にすることができる。

#### 2. 2 概念の共有度・典型度

共有度はセレクトに対する尺度である。それはセレクトを持っている事例の数として定義される。共有度が大きいということは、そのセレクトが多くの事例に含まれていることを示している。

典型度は事例に対する尺度である。それは事例を構成するセレクトの共有度の合計である。事例を記述するセレクトが大きな共有度を持つほど、またそれらのセレクトの数が多いほど典型度は大きくなる。これは、典型的な概念記述となるものを多く含んでいる。しかし、無駄な物を多く含んでいる (冗長度が高い) ともいえる。逆に典型度が小さいということは、共有度の大きなセレクトを多く持たない、セレクトの数が少ないなどの要因が考えられる。これは、その事象が例外的であるということの意味している場合もあり、また、無駄な物をあまり含んでいないともいえる。

### 3. 図形識別による評価実験

#### 3. 1 評価実験

P O S と N E G の事象を構成するセレクトはランダムに作成した。P O S の事象は 12 個、N E G の事象は 5、12、20 個の 3 種類を作成した。実験は、同じ P O S に対してこれら 3 種類の N E G を用いて行った。この実験における事例の選択は概念記述が得られるまでの速度および概念記述の質に影響を与えている。

Inductive Inference Algorithm using Measures for Concepts - Evaluation

Hiromi NORO, Izumi OHNO, Shoichi MASAKI, Kiyooki ATSUTA, Toshiya MINEZAKI, Shozo KONDO  
Tokai University

Michalski の S T A R 方法論では、この選択はランダムである。著者らの方法では典型度の大きい物から選択する場合（これを M A X とよぶ）と典型度の小さな物から選択する場合（これを M I N とよぶ）の 2 通りを行う。

なお、本実験は SUN3/260C 上の SUN Common Lisp を使用した。

### 3. 2 速度とその安定性

実際の実験結果を図 1～3 に示す。これらから Michalski の S T A R 方法論は速度に安定性がないことがわかる。それに対して著者らの方法では速度が安定していることが分かる。また M I N は Michalski よりも速度が早い場合がある。M A X の速度が M I N に比べて遅いのは選び出す事例のセレクト数が多いためそれらを分解・並べ替えをする時間がよりかかってしまうためである。

### 3. 3 解概念の安定性と質

S T A R 方法論では事例をランダムに選んでくるため、そのつど獲得される概念記述が変わってしまい同じ答えを得ることができない。それに対して著者らの方法では、常に選んでくる事例が同じであるため、答えが変わらずに安定した結果が得られた。また、獲得される概念記述は選言で結ばれた記述であるが、M A X ではその記述が少ないのに対して M I N はその記述が多い場合や冗長な記述を含むことが多くなる。

S T A R 方法論では、獲得する概念記述がそのつど変わるため解の質を制御することが非常に困難である。しかし、著者らの方法では解が安定しているので、M A X と M I N を使い分けることにより獲得する概念記述の質を制御することができる。

### 4. 結論

本件は典型度・共有度という 2 つの尺度を用いたアルゴリズムを利用した図形識別への応用の結果について述べた。今回の実験では、解の安定性や速度の安定性の両面からみて優れた結果が得られたと考えられる。しかし、単純に典型度の大きい事例または小さい事例から選択するということは不十分である。そのため、概念に対するこれらの重みづけの改善があげられる。また、獲得された解概念の質に対する評価基準が曖昧である。よって、これらのことは今後に残された課題である。

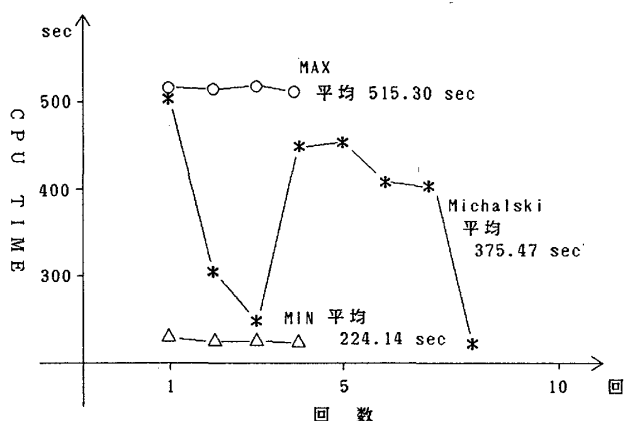


図 1 NEGの事象数 5

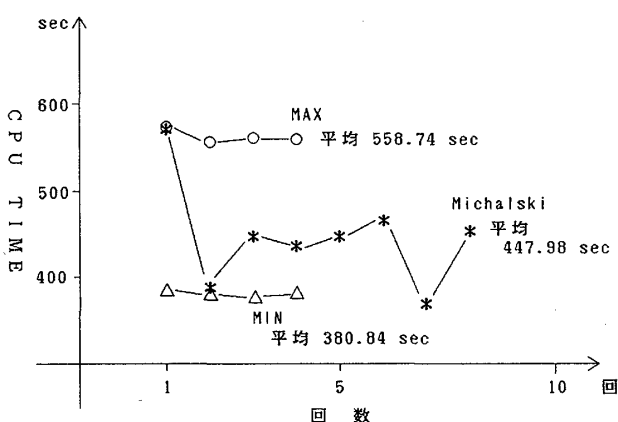


図 2 NEGの事象数 12

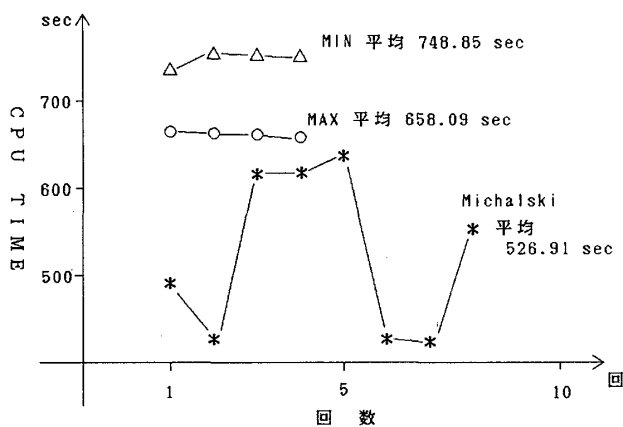


図 3 NEGの事象数 20

### 参考文献

- [1] 峯崎, 熱田, 近藤, 概念の新しい重みづけを用いた帰納推論アルゴリズム, 情報処理学会第 36 回全国大会 2R-8, 1988.
- [2] Michalski, R. S., "A Theory and Methodology of Inductive Learning", Michalski, R. S., Carbonell, J. G., Mitchell, T. N., eds. "Machine Learning", Springer-Verlag, 1984.