

# 遺伝的アルゴリズムを用いた概念学習に関する考察<sup>†</sup>

2R-1

遠藤 聡志 野沢 慎吾 大内 東

北海道大学工学部

## 1 はじめに

機械学習の一分野である帰納学習において、事例からの概念学習に関する様々な手法が提案されている。代表的なアルゴリズムに Mitchell.T.M の Version Space がある [1]。このアルゴリズムは、学習される概念の完全性・極大性を完備した厳密なアルゴリズムである反面、網羅的探索が必要である、ノイズに対応できない、概念同定までに必要な事例数の予測が難しい等いくつかの問題点を含む。本稿では、様々な最適化問題に対する有効な解法として注目されている遺伝的アルゴリズムを取り上げ、このアルゴリズムの概念学習に対する定式化及び従来法の概念学習法が持つが持つ問題点に対する有効性の検討を行う。

## 2 概念形成問題

帰納学習の分野において単一概念形成を扱った代表的なアルゴリズムに、Mitchell の Version Space (VS) がある。このアルゴリズムでは、概念形成問題は以下のように定式化される。

概念:

その概念にあてはまる事例に適用した場合に真、あてはまらない事例に適用した場合に偽となるような、あらかじめ規定された記述言語で記述される述語。

概念 [属性1, 属性2, ..., 属性n]

記述言語:

概念を表現する言語。概念は、複数の属性によって規定される。VSにおいて各属性は、その取り得る値、すなわち属性の集合要素に対して、その一般性の度合いから半順序を与えた木構造として表現される。

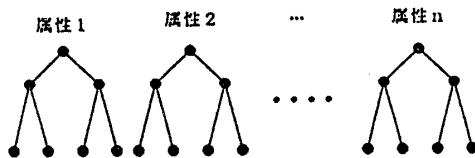


図1: 概念記述言語

事例集合:

目的とする概念と同種の属性によって構成される概念例の集合。目的概念にあてはまる正例、及びあてはまらない負例からなる。

概念形成:

概念記述言語によって規定される問題空間から、正例を満足し負例を排除するような概念を探索すること。VSにおいては、目的概念候補が存在する仮設空間を最も一般的なG、および最も特殊なSの二つの境界集合で規定し、正例による一般化および負例による特殊化をG=Sとなるまで繰り返す。このとき得られる境界概念が目的概念である。

## 3 遺伝的アルゴリズム

J.Holland によって提案された遺伝的アルゴリズム (GA) は、自然界における遺伝、淘汰のメカニズムを模倣した多点探索アルゴリズムである [2]。以下に概念形成のためのGAの定式化およびその有効性を検討する。

### 3.1 概念形成に対するGA

単一概念形成問題に対するGAは以下の各Stepで構成される。その詳細を以下に示す。

Step0: 概念記述言語および事例の設定

対象とする概念を構成する各属性集合を決定し、各属性集合の要素に一般性に関する半順序を与える。また、対象概念の正例及び負例を用意する。

Step1: 遺伝子型の設定

遺伝子の型はその遺伝子を構成する要素によって決まる。概念形成問題では、学習すべき概念が遺伝子に対応する。従って、遺伝子の構成要素は、概念を構成する属性となる。

Step2: 集団の設定

Step1で決定した遺伝子型を持つ遺伝子の集団を設定する。各遺伝子は、Step0で設定した概念記述言語によって規定される各属性の組み合わせをランダムに決定し、複数個の同型遺伝子を設定する。

loop

Step3: 適応度評価

母集団中の各遺伝子に対して、以下のルールによって適応度を評価する。

- ・ 遺伝子が正例を満たす。 → +1 point
- ・ 遺伝子が負例を満たす。 → -α point

あらかじめ用意された全ての事例に対してポイントを加算する事によって適応度が計算される。すなわち、

Fitness = (満たす正例数) - (α · 満たす負例数) となる。

<sup>†</sup>Genetic Algorithm for Concept Learning  
Satoshi ENDOH, Shingo NOZAWA and Azuma OHUCHI  
Faculty of Engineering, Hokkaido University

但し、遺伝子が例を満たすとは、遺伝子を構成する全ての属性値が Step0 で設定した半順序に対して上位に存在するかまたは等しいことを言う。

ここで、 $\alpha$  の値が正例を満たす場合に与えられるポイント 1 に対して十分に大きい場合に GA は、VS の様な候補消去アルゴリズムとしての意味合いが強くなる。一方、 $\alpha$  を 1 に近い値に設定すると、負の事例のノイズに対しても、ある程度対応することができる。

#### Step4:淘汰

Step3 によって計算された各遺伝子の適応度をもとに、低得点の遺伝子から順に一定割合で母集団から削除する。

#### Step5:増殖

Step4 によって減じられた分の遺伝子を現在の遺伝子集合から一定比率で抽出し、増殖させる。増殖させる遺伝子の選択は、集団の適応度を上げるために、遺伝子個々の適応度の高いものを優先的に選ぶ。

#### Step6:交差

母集団から任意の遺伝子対を選択しその特定部位の入れ替えを行う。交差法に関しては、様々な方法が提案されているが [3]、一般的な交差法としては One Point Crossover が挙げられる。

また、効果的な解の探索のために、概念形成問題の特徴を生かした次のような交差法も考えられる。

#### 一般化交差

Step3 の適応度評価において、 $\alpha$  の値は負例を受け入れた場合のペナルティである。 $\alpha$  の値を大きく設定すると、ひとつでも負例を受け入れる遺伝子は、Step4 の淘汰において取り除かれる。従って、次世代に残る遺伝子は、負例を受け入れることはないが、受け入れることができない正例を持つような遺伝子が多くなる。そこで、現段階で存在する遺伝子よりも少し一般性の高い遺伝子を生成することを考える。交差を行う遺伝子対の各属性に対して、概念記述言語上で双方の属性を包含する要素のうち最も特殊な要素を属性値に持つような遺伝子を生成する。

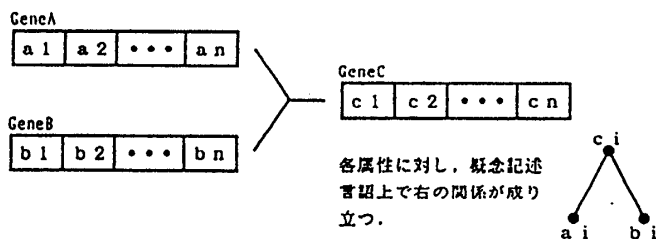


図1：一般化交差

#### Step7:突然変異

ある遺伝子の特定部位の属性値をランダムに変更する。この操作により突発的に評価値の高い遺伝子が生まれる可能性がある。

この操作においても、Step6 の一般化交差での議論がそのまま適用できる。すなわち、変化させる属性を現在のものよりも一般性の高いものに置き換えるような操作が考えられる。

end

GA では一般に、Step3 ~ Step7 までを一定回数繰り返し、最も評価値の高いものを解とする。しかし、本稿で提案する概念形成に対する GA では、遺伝子が取り得る評価値の最大値は設定した正例数であるから、このような値をとる遺伝子を解の候補と考えればよい。解の候補が複数得られた場合は、設定した事例が十分ではないことを意味するので、その時点で新しい事例を追加し、再びループさせる。

## 3.2 有効性

種々の学習機構によって得られた単一概念は、分類、予測、データ圧縮等のタスクを持つアプリケーションの中で用いられる。VS では、アルゴリズムの途中においては解の存在する範囲を与えることになるが、概念形成 GA ではアルゴリズム実行中の各時点においてもっともらしい解の候補を与える。この点で、対話的にエキスパートをサポートするような柔軟なアプリケーション向けの学習機構と言うことができる。

また VS では、

- ・概念形成のために問題空間の網羅的探索が必要である。従って、概念を規定する属性数の増加が探索計算に組み合わせ的に影響する。
- ・ノイズに対応できない。
- ・探索終了までの事例数の予測がつかない。

等の問題点が指摘されている。これに対して、概念形成 GA では、

- ・アルゴリズムが、比較的計算の軽いステップで構成されている。
- ・事例数が十分に多ければ、ノイズの影響は小さなものとして扱うことができる。
- ・あらかじめ与えられた事例から探索される候補概念を具体例の形で示すことができる。さらに解の候補が示されることによって、不足している事例を捕うことが容易になる。

等の優位点を挙げる事ができる。

## 4 おわりに

概念形成問題に対する遺伝的アルゴリズムの枠組みを提案した。また、従来法との比較から有効性を述べた。今後の課題としては、概念形成問題固有の特徴から提案したアルゴリズムの効率改善等がある。

## 参考文献

- [1] Mitchell, T.M.  
"Version Spaces: A Candidate Elimination Approach to Rule Learning"  
Proc. of the Fifth IJCAI, 1977
- [2] Goldberg, D.E.  
"Genetic Algorithms in Search, Optimization and Machine Learning"  
Addison-Wesley, 1989
- [3] 須貝康雄・平田廣則  
"組合せ最適化アルゴリズムとその応用"  
計測と制御 Vol.29 No.12, Dec.1990