



Title	遺伝的アルゴリズムによる概念学習の枠組みについて
Author(s)	遠藤, 聡志; Endoh, Satoshi; 野沢, 慎吾 他
Citation	北海道大學工学部研究報告, 164, 41-48
Issue Date	1993-05-28
Doc URL	https://hdl.handle.net/2115/42370
Type	departmental bulletin paper
File Information	164_41-48.pdf



遺伝的アルゴリズムによる概念学習の枠組みについて

遠 藤 聡 志 野 沢 慎 吾 大 内 東

(平成 4 年 12 月 25 日受理)

Genetic Algorithm for Concepts Learning.

Satoshi ENDOH, Shingo NOZAWA and Azuma OHUCHI
(Received December 25, 1992)

Abstract

Numerous research work concerning the induction are performed in the field of the artificial intelligence.

The Version space algorithm developed by Mitchell is known well as a typical method for learning single concepts.

The single concept learning problem is an essentially combinatorial optimization problem of the attribute value.

Genetic algorithms are paid close attention as an effective method of the combinatorial optimization problem in recent years.

In this paper, we proposed a genetic algorithm as a learning mechanism of the single concept learning.

Further, we described the expansion of this algorithm and discussed the effectiveness.

1. ま え が き

人工知能の中心的な研究課題である機械学習の分野において、例題に基づく帰納学習に関する様々な研究事例が発表されている。中でも、単一概念の学習(Learning single concepts)に関する Mitchell, T. M の Version Space(1)は、この分野の典型的な問題設定およびその問題に対する解決法を与える枠組みとしてよく知られている。Version Space 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という特長を持つ一方で、探索途中での妥当な解を与えられない、網羅的探索が必要である、ノイズに対応できない、等いくつかの問題点が指摘されている。

単一概念学習問題は、本質的に概念を構成する属性の取り得る値を、組み合わせ的に決定する問題と捉えることができる。近年、このような組み合わせ最適化の問題に対する有効なアルゴリズムとして遺伝的アルゴリズム(Genetic Algorithms)(3)(4)(5)が注目されてきている。遺伝的アルゴリズムは、その探索機構の簡潔さが特長となっている。このため環境の変化(問題領域の

変化)に柔軟に対応できる、機能拡張を容易に行える等、複雑な問題を扱うための学習機構としての優位性が期待される。このことから、本稿では単一概念学習問題に対する学習機構として、遺伝的アルゴリズムを導入することを試みる。はじめに、取り扱う単一概念学習問題について解説する。次に、遺伝的アルゴリズムによる問題の定式化を行う(6)(7)。さらに、アルゴリズムの機能追加に関して、新しい遺伝子オペレータの提案および複数概念学習問題(Learning multiple concepts)への拡張について言及する。

2. 単一概念学習問題(Learning single concepts)

帰納学習の分野において単一概念学習を扱った代表的なアルゴリズムとして、Mitchell. T. M の Version Space がある。このアルゴリズムにおいて、単一概念学習問題は以下のように定式化されている。

概念：その概念にあてはまる実例に適用した場合に真、あてはまらない実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として表現される。表現形式として以下のような概念記述ベクトルが用いられる。

概念 [属性 1, 属性 2, ..., 属性 n]

図1：概念記述ベクトル

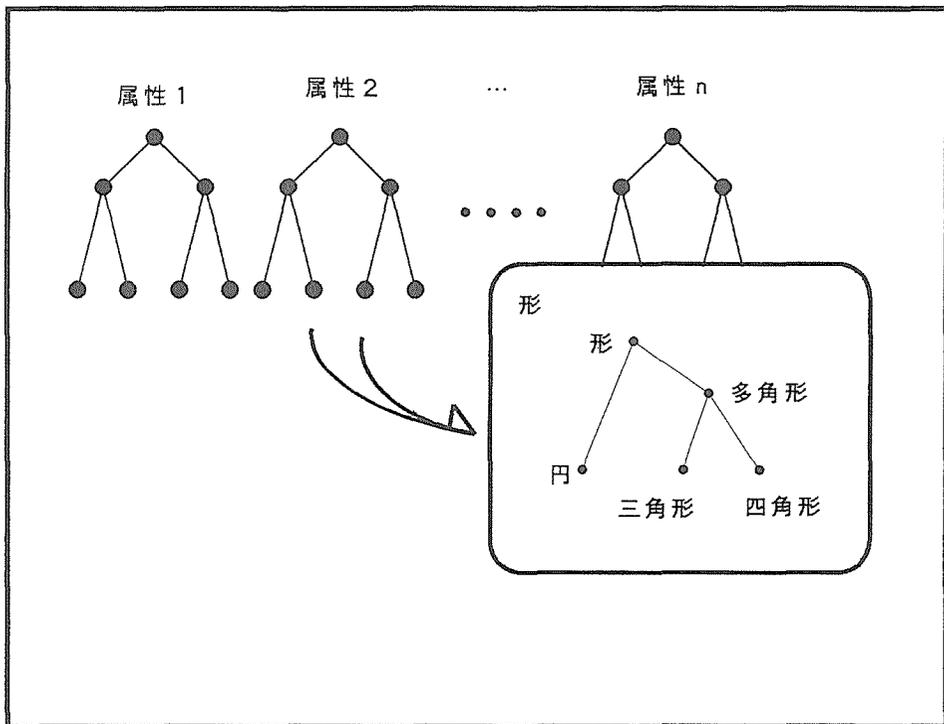


図2：概念記述言語

概念記述言語：概念を表現する言語。定義より、概念は複数の属性によって規定される。Version Space の枠組みにおいて各属性の取り得る属性値の上には、その一般性の度合いから半順序が与えられ、これによって各属性は木構造として表現される。例えば、形という属性の記述言語は図 2 のように与えられる。

事例集合：目的とする概念と同じ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。

概念学習：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。Version Space 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による単一概念学習

J. Holland によって提案された遺伝的アルゴリズムは、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。その概要は、

1) コーディング

扱う問題のパラメータ集合を文字配列の形式にコード化する

2) 遺伝的オペレータによる探索

コード化された個体の集合に対して、淘汰・増殖(Selection・Reproduction)、交叉(Crossover)、突然変異(Mutation)の各遺伝子操作を繰り返し行うこと集団の環境適応度を上昇させることで解を探索する

によって構成される。

以下で単一概念学習のための遺伝的アルゴリズムについて述べる。

3. 1 単一概念学習に対する遺伝的アルゴリズム

単一概念学習問題に対する遺伝的アルゴリズムは以下の各 Step で構成される。その詳細を以下に示す。ここで、Step 1-* はコード化フェイズ、Step 2-* は探索フェイズを示す。

Step 1-1：染色体型の設定

遺伝的アルゴリズムにおいて染色体の型は、その染色体を構成する遺伝子によって決まる。概念

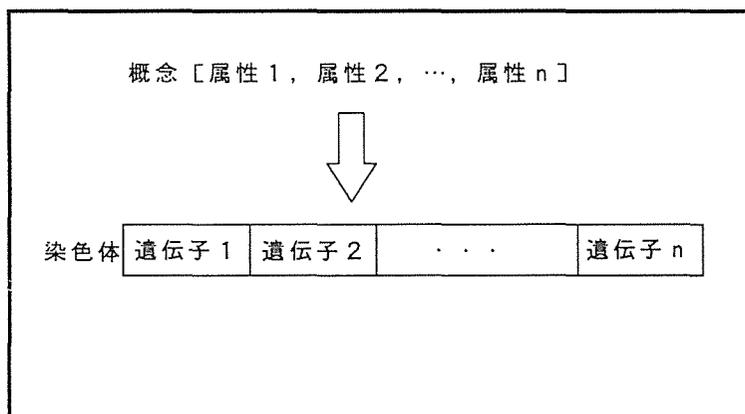


図 3：ダイレクトコーディング

学習問題では、学習すべき目的概念が染色体に対応する。従って、染色体の各遺伝子は、概念を構成する属性となる。このステップでは、目的概念を構成する属性およびその配列順序を決定する。

Step 1-2：概念記述言語および事例の設定

Step 1-1 によって定められた目的概念を構成する各属性のうち、属性木として未定義のものがあれば、それらを作成する。また、目的概念の具体例と思われる正例及びそうでない負例を用意する。

Step 1-3：集団の設定

Step 1-1 で決定した染色体型を持つ染色体の集団を設定する。各染色体は、Step 1-2 で設定した概念記述言語によって規定される各属性の属性値の組み合わせをランダムに決定し、複数個の同型染色体を設定する。また、集団設定に対してあらかじめ与えられた正例を初期の集団構成要素として取り入れることも適応度の高い初期集団を生成するために効果的である。

Step 2-1：適応度評価

母集団中の各染色体に対して、以下のルールによって適応度を評価する。

- ・染色体が正例を満たす。→ +1 point
- ・染色体が負例を満たす。→ $-\alpha$ 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、

$$\text{Fitness} = (\text{満たす正例数}) - (\alpha \cdot \text{満たす負例数})$$

となる。

ここで、染色体が事例を満たすとは、染色体を構成する全ての遺伝子の取る値が、事例の遺伝子の取る値に対して Step 1-2 で設定された概念記述言語の半順序の上で、可到達かまたは等しいことを言う。また、 α の値の設定に関して、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値を設定した場合には、負例を満たす染色体は致死として扱われる。この場合に遺伝的アルゴリズムは、Version Space と同様に候補消去アルゴリズムとして働く。一方、 α を 1 に十分に近い値に設定すると、負事例のノイズが多く含まれる場合に対して、ある程度対応することができると考えられる。これらのことから、 α に関しては、通常は $\alpha \gg 1$ 設定し、事例に曖昧なものが多い場合に限って $\alpha \simeq 1$ とすることとする。

以下のステップは遺伝的オペレータに関するものである。各遺伝的オペレータに関しては、扱う問題に応じて様々な方法が提案されているがここでは最も一般的な方法によってアルゴリズムを構成することにする。

Step 2-2：淘汰

Step 2-1 によって各染色体はそれぞれの適応度が計算される。この適応度に基づいて次世代に残る染色体が決定される。すなわち、各染色体の適応度比が次世代にその染色体が残る確率となる。

Step 2-3：交叉

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。交叉法として、OnePoint_Crossover を用いることにする。

Step 2-4：突然変異

ある染色体の特定部位の遺伝子の値を、値が取り得る範囲内で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。

遺伝的アルゴリズムでは一般に、Step 2-1～Step 2-4 までを一定回数繰り返し、最も評価値の高いものを近似解として採用する。しかし、本稿で提案する単一概念学習に対する遺伝的アルゴリズムでは、染色体が取り得る適応度の最大値は設定した正例数であるから、このような値を適応度とする染色体が発見された時点で探索を終了する事ができる。この解が利用者にとって満足のいかないものであれば、さらに事例を加えてアルゴリズムを実行することになる。

以上が単一概念学習に対する遺伝的アルゴリズムの詳細である。概要を図4にまとめる。

3.2 遺伝的アルゴリズムによる単一概念学習の優位性

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。Version Space 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、遺伝的アルゴリズムではアルゴリズム実行中の各時点において、現在最も妥当な解を得ることができる。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、Version Space では一切対応できなかった事例に対するノイズに関しても正当な事例が十分に多ければ吸収することが可能であると考えられる。さらに、事例の追加／削除や概念記述言語の変更といった、いわゆる環境の変化に対しても、変化以前の集団を初期集団として用いることで探索の効率化が期待される。従来法に対する優位点をまとめると、

- ・妥当な解の逐次表示
- ・ノイズへの対応

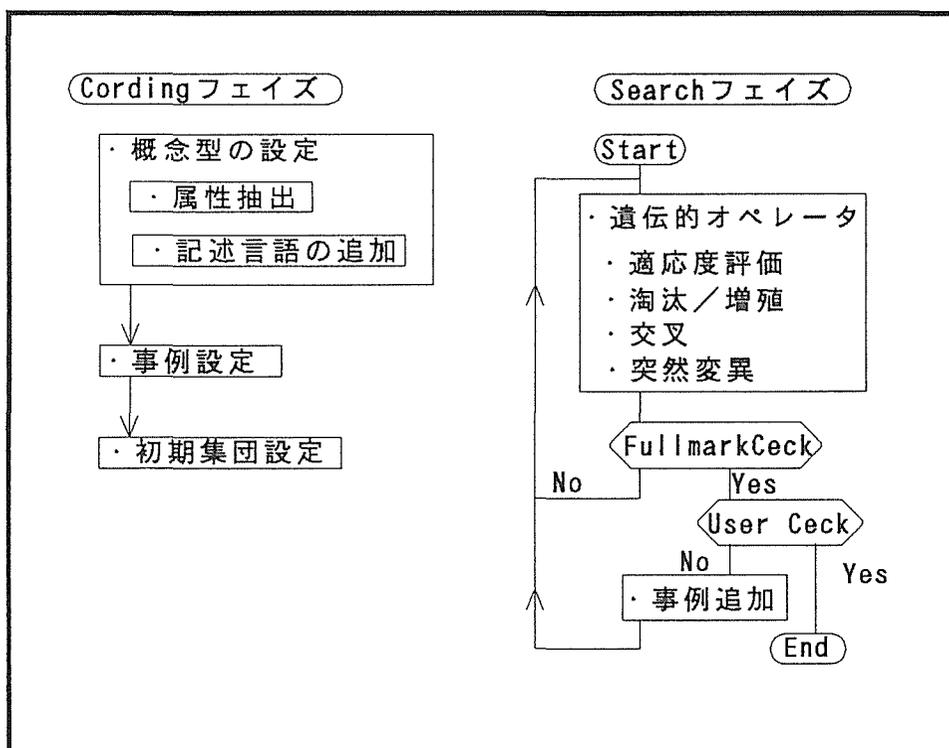


図4：概念学習問題に対する遺伝的アルゴリズムの構成

・環境変化への対応
となる。

さらに、遺伝的アルゴリズムの持つ特長として機能拡張の容易性があげられる。次章では、機能拡張の具体例として、新しい遺伝的オペレータの導入および複数概念学習問題へ問題を拡張する場合について説明する。

4. アルゴリズムの機能拡張

4.1 一般化交叉

この節では新しい遺伝的オペレータの導入について考察する。新しいオペレータは交叉に関するものである。

定義：以下の方法による遺伝的オペレータを一般化交叉と呼ぶ。

まず、任意の染色体対をランダムに選択する。選択された染色体の各遺伝子に対して、概念記述言語上で双方の値を含意する要素のうち最も特殊な要素を遺伝子の値とするような染色体を生成する(図5)。

Step 2-1の適応度評価において、 α の値は負例を受け入れた場合のペナルティーと考えられる。 α の値は通常 $\alpha \gg 1$ と設定されている。よって、負例を受け入れる染色体はStep 2-2の淘汰において致死として集団から取り除かれる。従って、アルゴリズムがある程度進行した後の世代では、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いくつかの正例を持つような染色体が集団を支配すると考えられる。このような集団から、より適応度の高い染色体を効率よく発見するためには、過度にならない程度に現在の染色体を一般化するとよい。一般化交叉はこのような考え方に基づいた交叉法である。

この交叉を、集団の適応度を尺度として、探索過程に組み込むことで、解探索に対する効率化が期待できる(8)(9)。

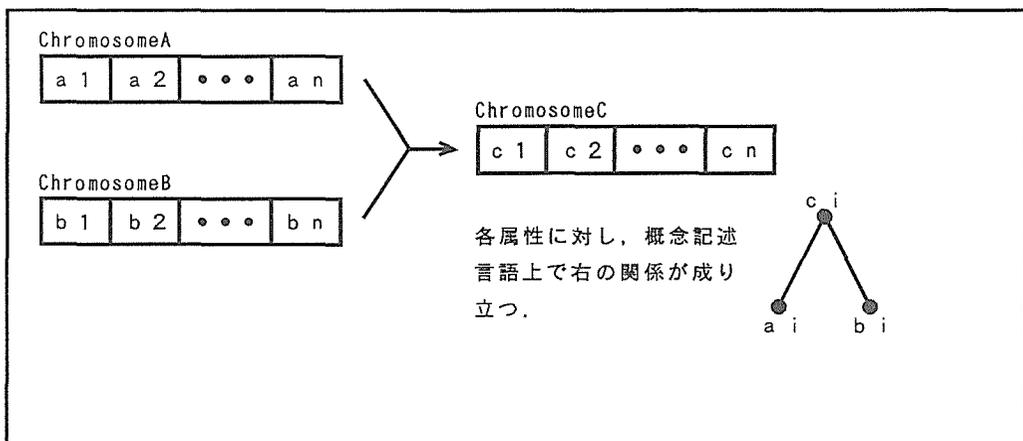


図5：一般化交叉

4.2 選言概念の獲得

獲得すべき目的概念が一つの連言概念記述で表現しきれない場合、これまでの議論の枠組みでは問題を扱うことができない。このような複数概念獲得の問題に対して、Version Space に対

する選言概念獲得のための拡張に関する報告(2)が成されている。このような問題の拡張に関して、遺伝的アルゴリズムを用いた概念学習の枠組みでは比較的簡単に対処することができる(10)。

この問題で獲得すべき目的概念は、複数の連言概念を論理和で結んだ形式を取る。染色体の型は連言形概念記述とする。染色体は各遺伝的オペレータによって、問題空間を並列的に探索する。単一概念学習における Step 2-1 の淘汰によって生存した複数の優良な遺伝子を論理和で結び、選言的な概念記述を形成する。ただし、獲得すべき目的概念を選言形式で得るために環境適応度の式に変更を加える。

環境適応度とは、その時点において満たす事例集合に基づいた、仮説(解)の“もっともらしさ”と考えることができる。

- (1) 包含する正事例の個数が多いこと
- (2) 包含する正事例の密度が高いこと

の2つを“もっともらしさ”の条件とし、環境適応度の関数には以下のような式が考えられる。

$$\text{Fitness}(m, n) = n/m \cdot (1 + km) = n/m + kn$$

この時、 n は染色体(仮説)が包含する正事例の個数、 m は染色体が表す空間の広さを示す。また、 k は上記の条件1と2の優先度を与える定数で、大きいほど条件1が重視される。この適応度関数は、図6のような問題空間において、破線で示した概念記述に優先して、二つの実線で示した概念記述を獲得することになる。

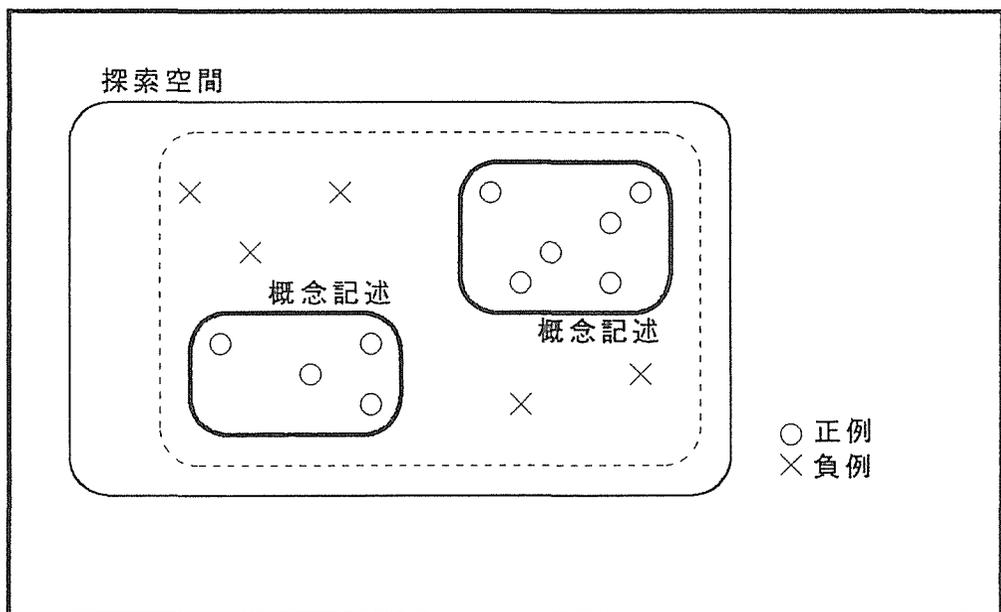


図6：探索空間上の選言概念

5. おわりに

本報告では、事例からの学習に関する典型的な問題の一つである単一概念学習問題に対して、遺伝的アルゴリズムによる解法を示した。さらに、遺伝的アルゴリズムによる解法が持つ優位性

について指摘し、特に機能拡張性について、新しいオペレータの導入、複数概念学習問題への適応の二つの拡張例を示した。このアルゴリズムに基づき現在、ワークステーション上での支援ツールを開発中である。

参 考 文 献

- 1) Mitchell, T. M. : "Version Spaces: A Candidate Elimination Approach to Rule Learning", Proc. of the Fifth IJCAI, 1977.
- 2) Murray, K. S. : "Multiple convergence: An Approach to Disjunctive Concept Acquisition", Proc. of the Tenth IJCAI, 1987.
- 3) Holland, J. H. : "Adaptation in Natural and Artificial Systems", Ann Arbor, MI: University of Michigan Press, 1975.
- 4) Goldberg, D. E. : "Genetic Algorithms in Search, Optimization and Machine Learning", Addison-Wesley, 1989.
- 5) 北野宏明 : "遺伝的アルゴリズム", 人工知能学会誌, Vol. 7, No.1, 1992, (pp. 26-37)
- 6) 遠藤, 野沢, 大内 : "遺伝的アルゴリズムを用いた概念学習に関する考察", 情報処理学会第44回全国大会講演論文集, 2R-1, 1992.
- 7) 野沢, 遠藤, 大内 : "帰納と対話の併用による概念知識ベースの構築", 情報処理学会第44回全国大会講演論文集, 4Q-5, 1992.
- 8) 遠藤, 野沢, 大内 : "概念学習に対する遺伝的アルゴリズムの局所探索に関する考察", 情報処理学会第45回全国大会講演論文集, 5E-1, 1992.
- 9) 遠藤, 野沢, 大内 : "遺伝的概念学習の局所探索法に関する検討", 電気関係学会北海道支部連合大会, 224, 1992.
- 10) 野沢, 遠藤, 大内 : "遺伝的アルゴリズムによる選言概念の獲得", 電気関係学会北海道支部連合大会, 223, 1992.