



Title	遺伝的アルゴリズムによる選言概念の獲得
Author(s)	野沢, 慎吾; Nozawa, Shingo; 遠藤, 聡志 他
Citation	北海道大學工學部研究報告, 165, 83-92
Issue Date	1993-07-30
Doc URL	<a href="https://hdl.handle.net/2115/42374">https://hdl.handle.net/2115/42374</a>
Type	departmental bulletin paper
File Information	165_83-92.pdf



## 遺伝的アルゴリズムによる選言概念の獲得

野沢 慎吾 遠藤 聡志 大内 東

(平成5年3月30日受理)

### Disjunctive Concept Aquisition using Genetic Algorithm

Shingo NOZAWA Satoshi ENDOH and Azuma OHUCHI

(Received March 30, 1993)

#### Abstract

As shown in Mitchell's paper, concept aquisition may be viewed as a search problem in a hypothesis space defined by the representation language. Its goal is to find the concept description that satisfies the completeness and consistency. Most of concept aquisition methods assume that concepts are independently separable and be must aquired as a single conjunctive description. However, real life concepts do not always hold this assumption and often need disjunctive concept description. This paper presents a new method for aquiring disjunctive concept description. We use genetic algorithm to search in a hypothesis space. This method has been implemented in the concept aquisition system COAST/GA, and a detailed example of its execution is presented.

#### 1. はじめに

概念獲得とは、与えられた事例集合から、それらに内在する規則を取り出す過程である。その枠組みは、Mitchellらによって定められ<sup>(1)(2)</sup>、知識工学や人工知能における重要なテーマとして注目されてきた。しかし、概念獲得を現実的なシステムへ組み込むためには、解決しなければならない様々な問題が残っており、実用化への見通しが十分に得られていない。

概念獲得の実用的な利用分野は、分類(classification)であると考え事ができる。つまり、獲得された概念記述を、以降に直面する事例がその概念に当てはまる事例なのか否かを判断するために用いる。しかし、この獲得された概念記述の位置付けの違いによって、利用分野は更に2つに分けられ、それに伴い概念獲得の方法も2種類に分化できると思われる。

まず、1つは、事例集合の圧縮(data compression)として概念記述を利用する場合である。エキスパートシステムの知識ベース等に用いるには、獲得される概念の必然性が要求されることが多い。つまり、事実として正確さが保証されなくてはならない。

もう1つの利用分野は、概念記述を未知の事例の予測(prediction)として用いる場合である。獲

得された概念の正確さの検証は後の処理にゆだね、仮説的な知識として推論に利用する。このような時、必然性よりも蓋然性が重要となる。

概念獲得法として最も代表的な、Mitchellの候補排除アルゴリズムは、極大性と無矛盾性を順守した最小干渉の帰納推論技法である。獲得される概念の正確さが保証され、必然性が要求される場合に有効である。しかしまた、そのために概念記述の低次性やノイズへの弱さなど様々な問題が生じている。

本研究は、蓋然的仮説生成を目的とし、遺伝的アルゴリズムを用いた選言概念獲得法を提案した。また、提案した方法に基づいてシステムを構築し、その有効性を検証した。

## 2. 概念獲得

概念獲得における概念という言葉には、外延的意味と内包的意味を考えることができる。

### 【定義1】概念

いろいろな事例がはじめにあって、それらの集合で定義された概念の意味を、概念の外延的意味という。また、いろいろな属性がはじめにあって、それらの集合で定義された概念の意味を、概念の内包的意味という。

### 【定義2】概念記述

概念の内包的意味の記述を概念記述と言う。

概念記述はあらかじめ定められた表現言語に基づいて表される。

### 【定義3】表現言語

概念や事例を記述するための言語を表現言語と言う。表現言語は概念を特徴的にとらえるための属性の集合から成り、属性の取る値を属性値と呼ぶ。

例えば、属性として“形”を考えたとする。“形”という属性の属性値として、円や三角がある。形式的には属性は属性値の集合として表現できる。すなわち、属性“形” = {任意形, 多角形, 三角, 四角, 円} 等である。

表現言語が与えられたとき、属性値を論理積で結んだ形式で概念を記述できる。

### 【定義4】連言概念

$v_1, v_2, \dots, v_m$ を属性値とする。このとき、 $[v_1 \text{ and } v_2 \text{ and } \dots \text{ and } v_m]$ , すなわち属性値の論理積で定義される概念を、連言概念 (conjunctive concept) と言う。

実世界に存在する概念により広く対応するためには、論理和を概念記述に導入する必要がある。

### 【定義5】選言概念

$c_1, c_2, \dots, c_m$ を連言概念とする。このとき、 $c_1 \text{ or } c_2 \text{ or } \dots \text{ or } c_m$ , すなわち連言概念の論理和

で定義される概念を、選言概念 (disjunctive concept) と言う。

【定義 6】事例

事例とは最も特殊な概念、すなわち、属性木において葉ノードとなる属性値のみを用いた連言概念である。

【定義 7】一般性

任意の属性の属性値集合には一般性という二項関係 $\geq$ が成り立つ。一般性 $\geq$ は木構造表現される特殊な半順序の性質を持つ。この属性の木構造を属性木と呼ぶ。

また、一般性 $\geq$ は、概念記述にも定義される。連言概念  $c_1=[v_{11} \text{ and } \dots \text{ and } v_{1m}]$ ,  $c_2=[v_{21} \text{ and } \dots \text{ and } v_{2m}]$  に対し、すべての  $i(1, \dots, m)$  において  $v_{1i} \geq v_{2i}$  が成り立つとき、またその時に限り  $c_1 \geq c_2$  である。同様に、選言概念を  $dc=c_1 \text{ or } c_2 \text{ or } \dots \text{ or } c_n$ , 連言概念を  $cc$  に対し、ある  $i(1, \dots, n)$  において、 $c_1 \geq cc$  が成り立つとき、またその時に限り  $dc \geq cc$  である。

図 1 は属性 “形” の属性木である。  $x \leftarrow y$  は  $x \geq y$  を示している。

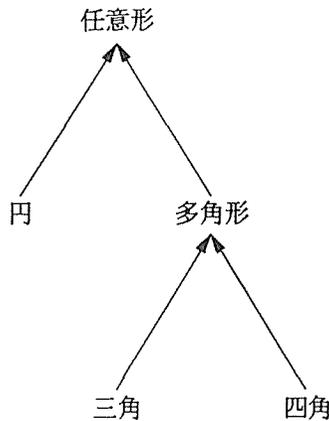


図 1 属性木

事例の集合は、目的の概念に当てはまる事例と、当てはまらない事例に分けて扱われる。

【定義 8】正事例・負事例

事例  $p$  が、概念  $c$  に対して  $c \geq p$  が成り立つならば、 $p$  は  $c$  の正事例であると言い、 $c \geq p$  が成り立たないならば、 $p$  は  $c$  の負事例であると言う。

以上の準備のもと、概念獲得は次のように定義できる。

【定義 9】概念獲得

有限個の正事例  $p_1, p_2, \dots, p_m$ , および負事例  $q_1, q_2, \dots, q_n$  ( $m \geq 0, n \geq 0$ ) が与えられているとする。このとき、

極大性 : すべての正事例  $p_i$  に対し  $c \geq p_i$  が成り立つ。

無矛盾性：すべての負事例  $q_i$  に対し  $c \geq q_i$  が成り立たない。  
 の2つの性質が満たされる論理式  $c$  を求めることを、事例  $p_1, p_2, \dots, p_m, q_1, q_2, \dots, q_n$  に対する概念獲得と言う。すなわち概念獲得とは、外延の意味で与えられた概念の内包的意味を獲得する過程である。

### 3. 従来法の問題点

Mitchellが提案する候補排除アルゴリズムは (candidate-elimination algorithm) は、概念記述が一般性による半順序関係になっていることを利用して、極大性と無矛盾性を満たす概念記述集合(バージョン空間)を2つの境界で表し、それらの最小限の更新によって目的の概念記述を絞り込む手法である。このアルゴリズムは、漸進的な探索法であり、獲得された概念の正しさを保証するが、次のような問題点が指摘されている<sup>(3)(4)</sup>。

1. 選言概念に対応できない。
2. 表現言語の変更に対処できない。
3. ノイズを含んだ事例集合に弱い。
4. 正事例のみからの獲得ができない。

### 4. 遺伝的アルゴリズムによる概念獲得

近年、最適化手法の一つとして注目を浴びている遺伝的アルゴリズム (Genetic Algorithm) は、生物が持つ適応性に学ぼうという立場から提案された手法である<sup>(5)</sup>。本研究では、人類の仮説洗練過程と生物の進化の過程の類似性に着目し、蓋然性を重視した選言概念獲得技法に遺伝的アルゴリズムを適用する<sup>(6)</sup>。

#### 4. 1 遺伝子型

連言概念を遺伝的アルゴリズムにおける個体として設定する。遺伝子型は、連言概念の各属性の値を遺伝子とする構成となる。

$$[a_1 \text{ and } b_5 \text{ and } c_3 \text{ and } d_6] \implies \begin{array}{|c|c|c|c|} \hline a_1 & b_5 & c_3 & d_6 \\ \hline \end{array}$$

図2 遺伝子型の設定

#### 4. 2 適応度

個体が表す概念が、獲得の対象である实在概念にどの程度近いかが、すなわち蓋然性の度合を、個体の環境適応度として用いる。当然ながらシステムにとって实在概念は未知であり、直接それを適応度の計算に用いる事は不可能である。そこで、实在概念を間接的に反映している情報として、システムが受け取った事例集合を用い、個体の適応度を計算する事にする。

適応度を設定するにあたり、連言概念の大きさというものを導入する。

【定義10】 連言概念の大きさ

任意の属性値  $a$  に対し、属性木の上で  $a \geq b$  となる葉ノード (子を持たないノード)  $b$  の個数を  $l(a)$  とする。この時、連言概念  $c = [c_1 \text{ and } \dots \text{ and } c_m]$  の大きさは、 $l(c_1) \times \dots \times l(c_m)$  である。

個体の適応度は、個体が表す概念の大きさ  $spc$  と、個体が受け入れる正事例の個数  $pos$  (重複は数えない) をもとに、次のような指針で設定する。

1. 受け入れる正事例の個数 ( $pos$ ) が多いほど適応度が高い。
2. 受け入れる正事例の密度 ( $pos/spc$ ) が大きいほど適応度が高い。

その時点までにシステムに与えられた正事例の個数を  $N$  とすると、最も適応度が高くなるべき個体は  $spc=N, pos=N$  である。すなわち、すべての正事例を最も高い密度 ( $pos/spc$ ) で受け入れる連言概念の適応度  $fitness(spc, pos) = 1$  である。これを最高値とし、密度一定 ( $pos/spc=1$ ) で正事例の個数減少の方向と正事例の個数一定 ( $pos=N$ ) で密度減少の方向とで同じ傾斜で減少するような適応度を考える。

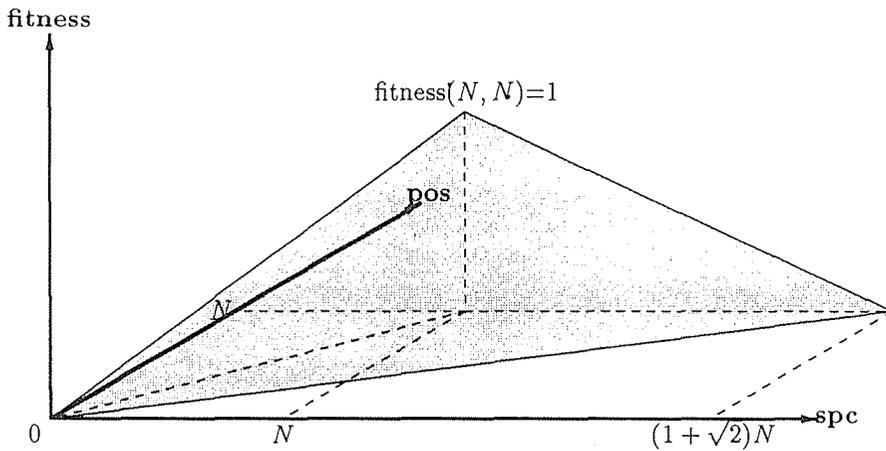


図3 適応度

この時、適応度は以下のように定められる。

$$fitness(spc, pos) = \max\left(0, \frac{1}{N}\left(pos - \frac{spc - pos}{\sqrt{2}}\right)\right)$$

また、負事例は致死条件として環境を設定し、1つでも負事例を受け入れてしまう個体は死滅することにする。

4. 3 交差

概念獲得における交差とは、仮説と仮説の合成として位置付ける事ができる。そこで、本研究では仮説空間の持つ一般性の半順序構造を交差に利用し、より効果的な操作を行うような2種類の交差を設定する。ある2個体を包含するように一般化した個体を生成する事を一般化交差と言い、2個体の共通部分をもとに特殊化した個体を生成する事を特殊化交差と言う。図4に示す例では、 $[a_1 \text{ and}$

$b_1$  and  $c_3$  and  $d_4]$  と  $[a_3$  and  $b_2$  and  $c_5$  and  $d_3]$  から、一般化交差によって  $[a_1$  and  $b_0$  and  $c_1$  and  $d_0]$  が生成される。また、図5の例では、 $[a_0$  and  $b_1$  and  $c_3$  and  $d_0]$  と  $[a_2$  and  $b_2$  and  $c_1$  and  $d_5]$  から、特殊化交差によって  $[a_2$  and  $b_1$  and  $c_3$  and  $d_5]$  が生成される。

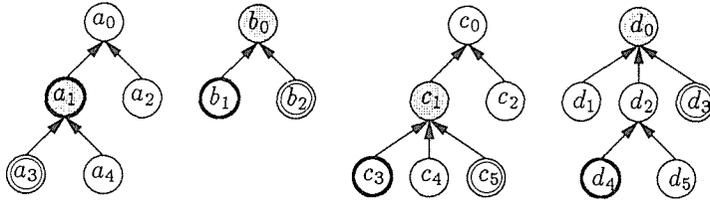


図4 一般化交差

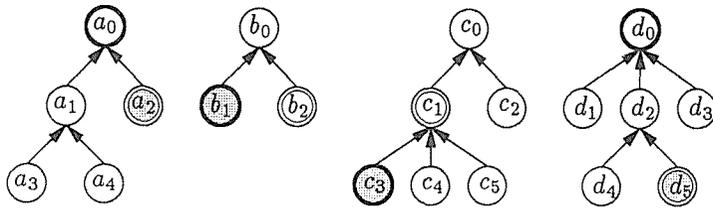


図5 特殊化交差

#### 4. 4 突然変異

概念獲得における突然変異とは、突発的な仮説の飛躍と考える事ができる。交差と同様に、仮説空間の持つ一般性を突然変異に適用し、遺伝的アルゴリズムの操作が、より概念獲得に対して効果的に作用するよう設定する。ある個体が表す概念から、乱数によってある属性値を選択し、属性木に従ってその上位要素または下位要素に変更した個体を生成する。

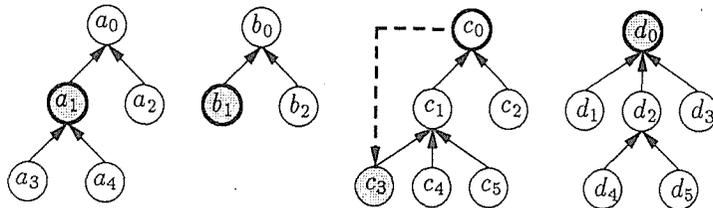


図6 突然変異

### 5. 実験

以上の設定に基づいて、選言概念獲得システムCOAST/GAを作成した。アルゴリズムは次の通りである。

Step 1 : 環境Env =  $\langle P, Q \rangle$  を設定する。ここで、Pは正事例の集合、Qは負事例の集合である。

獲得概念Cを空に初期化する。

Step 2: 世代数 $t=0$ とし, 初期世代の個体の集団  $I = \{i_1, i_2, \dots, i_{size}\}$  を生成する。初期の個体は, 正事例集合の中からランダムに選ぶ。

Step 3: 各個体に対し, 適応度 fitness を計算する。

Step 4: 適応度 fitness に比例した確率で, 2つの個体を選択する。

Step 5: 選択された2つの個体に対し, 一般化交差, 特殊化交差を適用し, 新たな2つの個体を生成する。さらに, 確率 $r$ で突然変異を適用する。

Step 6: Step 5 で生成された2つの個体が致死条件(負事例を受け入れる)を満たさなければ, 集団中の適応度の小さい2つの個体と置き換える。

Step 7: 世代数  $t=t+1$ 。  $t$ が設定値 $T$ より小さければ, Step 3へ。

Step 8: 最も適応度の高い個体 $i_x$ を抽出し, 獲得概念に論理和で加える ( $C=C \text{ or } i_x$ )。

$i_x$  が受け入れている正事例を正事例集合 $P$ から取り除く。 $P$ が空でなければ, Step 2へ。

Step 9: 獲得概念Cを出力する。

実験結果の例を表1に示す。扱った問題は, 6つの属性がそれぞれ7, 8, 8, 8, 9, 10個の属性値を持つような表現言語を用いた。すなわち,  $7 \times 8 \times 8 \times 8 \times 9 \times 10 = 322560$ 個の仮説からなる空間の探索である。また, 遺伝的アルゴリズムのパラメータは, 集団の大きさ $SIZE=20$ , 世代の上限を $T=50$ , 突然変異の起きる確率 $r=10\%$ とした。獲得概念の欄は, システムが獲得した概念記述を構成する連言概念を大きさとし, それを受け入れる正事例数の分数形式で, 列挙している。

表 1

正事例	負事例	獲得概念 ( <i>pos/spc</i> )
8	5	2/3, 2/2, 1/1, 1/1, 1/1, 1/1
14	5	4/6, 4/4, 3/3, 2/2, 1/1
20	10	9/12, 6/8, 5/6
26	10	12/12, 8/8, 6/6

表から分かるように, 事例の増加に伴って一般性の高い概念記述が獲得されている。仮説洗練の過程をグラフに示す (図 7)。最初の連言概念を獲得する間に最優良個体の適応度と集団の適応度の平均がどのように変化するかを表したものである。いずれも乱数の初期値を変更した30回の実験の平均である。優良な仮説が次々と生成され, それに伴って集団全体が進化し, 優良な集団へと収束している。

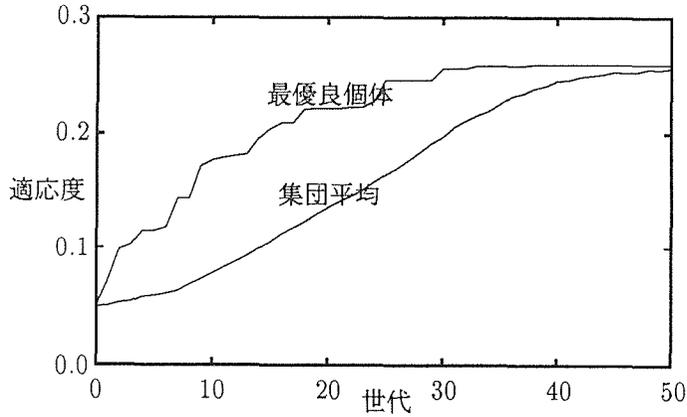


図7

次に、同じ問題に対し、集団の大きさを20から10に減らした場合とで、最優良個体の適応度を比較する(図8)。初期集団を正事例から生成するため、集団の大きさが正事例数に比べて小さい時、初期集団の遺伝子情報が不足し、十分な一般化が行われていないと思われる。

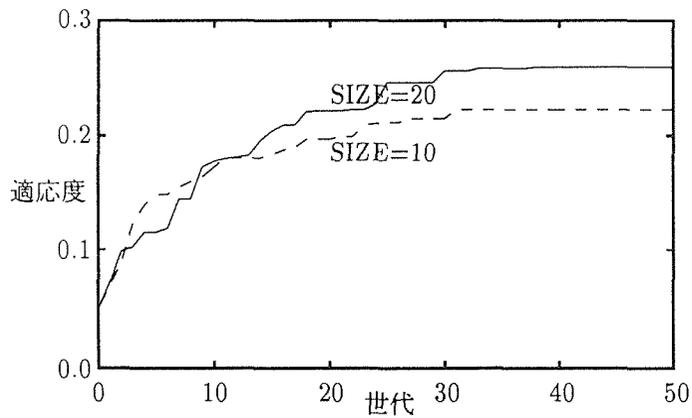


図8

集団の大きさを10にしたまま、突然変異の発生する確率を変えて実験を行ってみた(図9)。この結果は、突然変異によって集団の大きさの不足を補っていることを示している。しかし、集団の大きさが十分にある場合、突然変異はむしろ適切な交配を破壊してしまい、好ましい結果が得られなかった。

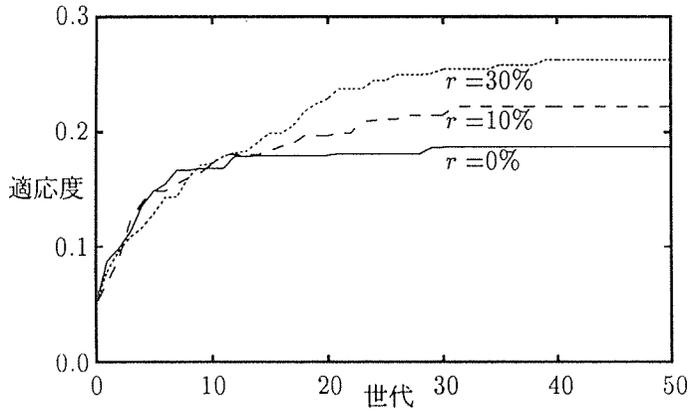


図9

表2に示す例では、負事例を与えず正事例のみからの概念獲得を試みた。本手法では、密度を取り入れた適応度が一般化し過ぎを抑制し、正事例のみでも十分な概念獲得が行われている。

表2

正事例	負事例	獲得概念 ( <i>pos/spc</i> )
8	0	3/4, 2/3, 1/1, 1/1, 1/1
14	0	5/8, 4/6, 2/3, 2/2
20	0	9/12, 6/8, 5/6
26	0	12/12, 8/8, 6/6

## 6. おわりに

本研究において、筆者は以下の事を行った。

1. 遺伝的アルゴリズムによる選言概念獲得法の提案。
2. アルゴリズムの実験的評価。
  - ・ 蓋然性の高い選言概念の獲得に、ほぼ期待通りの効果が得られた。
  - ・ 正事例のみの事例提示からも一般的過ぎない概念が得られた。

## 参考文献

- (1) Mitchell, T. M. : Version Spaces : A Candidate Elimination Approach to Rule Learning, Proceedings of the Fifth IJCAI, 305-310, (1977)
- (2) Mitchell, T. M. : Generalization as Search, Artificial Intelligence, Vol.18, 203-226, (1982)
- (3) Ditterich, T.G. : Learning and Inductive Inference, The Handbook of Artificial Intelligence, Vol.3, (1982)

- (4) Murray, K. S. : Multiple Convergence : An Approach to Disjunctive Concept Acquisition, Proceedings of the Tenth IJCAI, 297-300, (1987)
- (5) Goldberg, D. E. : Genetic Algorithms in Search, Optimization and Machine Learning, Addison-Wesley, (1989)
- (6) 野沢, 遠藤, 大内 : 遺伝的アルゴリズムを用いた選言概念の学習, 第25回計測自動制御学会北海道支部学術公演会論文集, 153-154, (1993)