

8D-8

遺伝的概念学習のための支援ツールに関する検討[†]

遠藤 聡志 野沢 慎吾 大内 東

北海道大学工学部

1 はじめに

組み合わせ最適化問題の解法として、J.H.Holland の遺伝的アルゴリズム (GA) が注目されている [1][2]。我々は、GA の応用分野の一つとして、帰納学習の枠組みである事例からの概念学習問題に着目し、単一連言概念学習問題に対する遺伝的アルゴリズムの適用についての検討を行ってきた [3][4]。また、探索効率化のために効果的に局所探索を行う新しい交叉について検討を行っている [5]。本稿では、これらの内容についての実験的な考察および汎用的な学習機構の構築の要求に対応するためのツールの構築について検討する。また、このツールでは宣言概念学習の問題への拡張を行っており、この問題についても合わせて検討する。

2 遺伝的概念学習

2.1 単一概念形成問題

単一概念形成問題とは、概念を属性の連言で規定し、目標概念を適切に表現する属性値の組み合わせを決定する問題である。各属性の取り得る値は、一般性に基づいた木構造 (表現言語) として与えられる。よって、問題空間は属性値集合の包含関係に基づく束をなす。この問題空間上で、目標概念の正例及び負例を与えながら、正例を満足し負例を排除するような解の探索を行う。この問題は、Mitchell の Version Space[7] によって定式化された問題である。単一概念形成は基本的に各属性の属性値の組み合わせ問題であり、組み合わせ問題の確率的探索解法である GA を容易に適用することが出来る。

2.2 概念形成 GA の手順

・コーディング

学習すべき目的概念は、それを構成する各属性の連言で表現される。この表記は属性ベクトルと呼ばれる。この属性ベクトルの各属性値を GA おける遺伝子に対応させることによって単一概念形成問題はダイレクトコーディングが可能となる。

概念 [属性 1 , 属性 2 , ..., 属性 n]



染色体 [遺伝子 1 , 遺伝子 2 , ..., 遺伝子 n]

また、初期集団の染色体は正例集合で構成することによって、より高い適応度から探索を始めることが出来る。

・適応度評価

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。→ -α point

あらかじめ設定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。

$$Fitness(n, \bar{n}) = n - \alpha \cdot \bar{n}$$

ここで、 n は満たす正事例数、 \bar{n} は満たす負事例数とする。但し、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値が設定した表現言語上で可到達であるかまたは等しいことを言う。

また、 α の値に関して $\alpha \gg 1$ と設定すればアルゴリズムは候補消去型、 $\alpha \approx 1$ とすれば負例ノイズ吸収型の性格を持つことになる。

・遺伝子操作

・選択 (Selection) ・交差 (CrossOver) ・突然変異 (Mutation)

この問題では、ダイレクトコーディングがなされているため、OnePointCrossover 等の、GA が与える一般的な方法による遺伝子操作の構成が可能である。

・アルゴリズムの終了条件

概念形成に対する GA では、適応度評価の定義から、各染色体が取り得る評価値の最大値は設定した正例数に等しい。従って、正例数に等しい評価値をとる染色体が、あらかじめ設定された事例集合から決定しうる最適解となる。ここで、最大値を取る解が複数個得られた場合は、設定した事例が解を同定するのに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、再びループさせることになる。

3 遺伝的概念学習ツール

[目的]

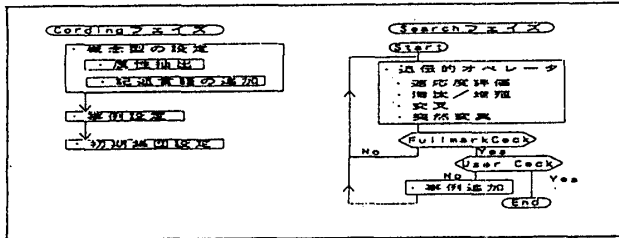
このツールは、遺伝的アルゴリズムによる概念学習モデルの実現を目的としており、汎用的な知識ベースシステムや知識獲得シス

[†]Support system for GAs of Concept Learning
Satoshi ENDOH, Shingo NOZAWA, and Azuma OHUCHI
Faculty of Engineering, Hokkaido University

テムの学習機構としての利用を考えている。

[構成]

このツールは, attribute クラス, example クラス, chromosome クラス, population クラス等の遺伝的概念学習を構成する基本的なクラス型によって構成しており, C++によって実現されている。処理の流れを図に示す。



[拡張]

このツールでは, 2節で示した基本的なアルゴリズムの拡張として, 2つの機能を追加している。

拡張機能1: 一般化交叉

適応度評価において, α の値は染色体が負例を受け入れた場合のペナルティーである。 α の値を十分に大きく設定してアルゴリズムを実行した場合, 負例を受け入れた染色体は積極的に淘汰される。従って, 局所探索が効果的であるような後世代の集団に残ることが出来る染色体は, それが最適解でなければ以下の性質を持つこととなる。

- (1) ただ一つの負例も受け入れない。
- (2) 設定された大部分の正例を受け入れるが, 一部の受け入れることができない正例がある。

このような性質を持つ集団から最適解を得るには, 集団中の染色体をオーバージェネラライズにならないように一般化する事が効果的である。このことから, 一般化交叉では, 任意に選択された2つの染色体の各遺伝子に対して, 表現言語上で双方の属性を含意する要素のうち最も特殊な要素を遺伝子に持つような染色体を生成する。

ツールにおいて, この機能は population クラスのメンバ関数として用意される。

拡張機能2: 選言概念獲得

獲得すべき目的概念が一つの選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 [6]。この問題で獲得すべき目的概念は, 複数の選言概念を論理和で結んだ形式を取る。染色体の型は選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適応度の式に変更を加える。

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, 解の“もっともらしさ”と考えることができる。

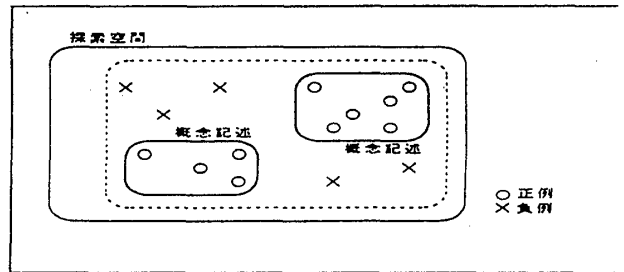
- (1) 包含する正事例の個数が多いこと
- (2) 包含する正事例の密度が高いこと

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。

$$Fitness(m, n) = \frac{n}{m}(1 + km) = \frac{n}{m} + kn$$

この時, n は染色体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。

ツールにおいてこの機能は, chromosome クラスのメンバ関数として用意される。



4 おわりに

本稿では, 遺伝的アルゴリズムによる単一選言概念学習のためのツールについて報告した。このプログラムでは, 遺伝的概念学習の基本的な部分に加えて, 一般化交叉の利用や, 選言概念学習のための機能拡張を行っており, この機能についてもあわせて報告した。

参考文献

- [1] Goldberg, D.E. "Genetic Algorithms in Search, Optimization and Machine Learning" Addison-Wesley, 1989
- [2] 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)
- [3] 遠藤聡志 野沢慎吾 大内東 "遺伝的アルゴリズムを用いた概念形成に関する検討" SICE北海道支部学術講演会講演論文集, 1992(pp.65-66)
- [4] 遠藤聡志 野沢慎吾 大内東 "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
- [5] 遠藤聡志 野沢慎吾 大内東 "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
- [6] 野沢慎吾 遠藤聡志 大内東 "遺伝的アルゴリズムによる選言概念の獲得" 平成4年電気関係学会北海道支部連合大会講演集 No.223, 1992
- [7] Mitchell, T.M. "Version Spaces: A Candidate Elimination Approach to Rule Learning" Proc. of the Fifth IJCAI, 1977