1311 遺伝子の特徴抽出による新しい GA 法の開発研究

The study of New GA technique by Special Feature Extracting of Gene

正 尾田 十八 (金沢大) ○学 金井 亮 (金沢大院) 河登 洋介 (金沢大)

Juhachi ODA, Kanazawa University, 2-40-20, Kodatsuno, Kanazawa, Ishikawa

Ryo KANAI, Dept. Kanazawa University of Graduate School

Yosuke KAWATO, Human and Mechanical Engineering, Kanazawa University.

The Genetic Algorithm (GA), which is optimum method based schema theorem, is applicable on a wide scale problem. So, various optimum methods derived from the GA are studied.

In this study, one of them, that is the Special Feature Extracting GA (SFE-GA), is proposed. The SFE-GA is based on the idea that genes possessing suitable environment extend to all over the mass, and is GA with an operation of stepping up the function.

This method is applied to a problem that searches of the maximum value in a simple function, and the validity is checked. The behavior in search is clarified and it is confirmed to extract the effective schema sequentially.

Key Word: Computational Mechanics, Genetic Algorithm, Special Feature Extracting, Finite Element Method

1. 緒言

メタヒューリスティクスとは物理現象や生物の進化過程、生物の集団・社会行動などから知識を得て提案された最適化手法である。その特徴として、有効性が経験的に明らかになっている知識(ヒューリスティクス)を取り込んでいることや、履歴を利用して新たな解を探索し、解を評価した後に探索にフィードバックすることが挙げられる。これはスキーマ説に基いている。

遺伝的アルゴリズム(GA)もまたメタヒューリスティクスの一つで、適応範囲が非常に広い。そのため GA から派生した非常にバラエティに富んだ最適化手法が現在も研究されている。

本研究はその一つである特徴抽出型 GA(Special Feature Extracting: SFE-GA)を提案し、以下で詳述する。

2. 特徴抽出型 GA

SFE-GA は環境に適した形質を発現する遺伝子配列が徐々に集団全体へと浸透する考えを基にし、この作用を促進させる操作を遺伝オペレータとして従来知られている GA に新たに加えた手法⁽¹⁾である。またその操作にはゲノム解析の際、塩基配列の類似性からその遺伝子座の持つ意味を推定するという手法⁽²⁾を参考にしている。

Fig.1: Special features extracting of genes

2.1 特徴抽出 それぞれの遺伝子配列には必ず意味があり、適合度の高い染色体において配列の傾向を調べ、その傾向を他の遺伝子にも適用することで集団の適合度を上げていく。このことを、図1を用いて具体的に説明する。

染色体を目的関数の降順に並べ、上位 $n\%$ を各 bit 毎にその類似性を調べる。つまり各 bit における遺伝子座[0,1]の傾向から特徴を得る。その後得られた特徴を目的関数下位 $m\%$ の染色体に抽入する。類似性を調べる際には、 t 検定を用いることとした。この操作を遺伝オペレータのルーレット選択の代わりに用いる。

2.2 t 検定 特徴抽出に用いるのは、統計学で2つの群に差があるか否かを判断するのに用いられる t 検定を利用した。有意確率 α で求める際、帰無仮説 H_0 : 「2 群の母平均値に差はない」に対し、対立仮説 H_1 : 「2 群の母平均値に差がある」とする。それぞれの群に対して平均値 x_m, y_n , 不偏分散 s_x, s_y を算出し、検定統計量 t_0 , 自由度 ν を求める。自由度と有意確立に準ずる値と検定統計量を比較してどちらの仮説を採択するかを決める。この検定は、2 つの群の等分散性が明確なときに用いる。しかし本手法で用いる場合、実際には規模、分散性共に、どのような群になるかは想定不可能なため、Welch の t 検定を用いる。

2.3 Welch の t 検定 この Welch の t 検定では等分散性を仮定せずに行う。そのため検定統計量と自由度は通常と異なる式(1)によって求められる。その点以外は同様である。

$$t_0 = \frac{|x_m - y_n|}{\sqrt{\frac{s_x^2}{m} + \frac{s_y^2}{n}}}, \quad \nu = \frac{\left(\frac{s_x^2}{m} + \frac{s_y^2}{n}\right)^2}{\frac{s_x^4}{m^2(m-1)} + \frac{s_y^4}{n^2(n-1)}} \quad (1)$$

本手法では対象の遺伝子座に対し、『0』、『1』の群に分けて検定を行う。2群が等しいと判断された場合には『0』、『1』の変化による差異がないとする。逆に異なる場合には目的関数値が優秀な遺伝子群の値をスキーマとし、抽出する。

3. 探索におけるスキーマの発見

まず、検定に用いるパラメータの選択可能域を決定するため、これらの値を変化させ、その傾向を確認する。また、本手法ではスキーマを発見し、それを集団に伸展させる操作を加えている。そこで、その操作が十分に機能しているか否かを例題を通して確認する。

3.1 パラメータによる差異 SFE-GA 特有のパラメータによる差異を明らかにするため、式(2)における最大値探索を行った。

$$f(x) = x^3 \quad \{0 \leq x < 1\} \quad (2)$$

x を、上位ビットほど f の値に強い影響を与える遺伝子長 18bit の染色体へとコーディングして最大値を求める。GA のパラメータは人口 300、交叉率 40%、突然変異率 0.5% とする。この条件のもと、同一の初期集団に対して特徴抽出のパラメータを変更し、解に変化が現れるか、最大値を得た世代の早さなどを考察する。変化させる検定パラメータの値域は以下のようにする。

- ・検定を行う母集団 $n\%$ (10~100)
- ・特徴抽出する下位集団 $m\%$ (5~25)
- ・有意差 $\alpha\%$ (1,2,5,5)

まず基準として、 $n=40\%$ の上位集団において有意差 $\alpha=5\%$ で検定を行い、下位 $m=10\%$ に特徴を抽出するものとする。この結果に対し、それぞれの値を変更して比較を行う。

3.2 結果考察 最大値を得た世代と、母集団 n の関係を図 2 に示す。母集団 n を 10% 刻みで変化させた場合、最大値が得られたもの(図中■)と得られないもの(図中◆)に差が現れ、出現した世代数にも傾向が見られた。

図からもわかるように、 $n\%$ の値を大きくするにつれ、探索に要する時間が長くなっていることが確認できる。また、70%を超えると一気に時間が短くなっているが、これは最大値を見つけることができずに探索を終了しているためである。80%以降も同様に、 n の増加に伴い、探索時間が長くなっている。この原因としてはやはり、母集団を大きくした結果、目的関数値の低いものが入ってしまい、検定がしにくくなる(2群に差がなくなる)ためと考えられる。

Fig.2:Defference from domains to refer

次に、下位集団 m の規模を 5% 刻みで変化させた場合、解の探索時間および最大値に変化は見られなかった。また、有意差 α を変化させた場合においても、値を大きくすると、探索時間が長くなる傾向にあると考えられるが、その変化は母集団の変化に比べて顕著ではなかった。この結果、検定では元になるデータが探索結果に大きく影響するため、その影響力の方が(下位集団の変化の影響よりも)強かったと考えられる。

3.3 スキーマの発見状況 次に、最大値が得られるまでの遺伝子解析を考察する。3.1 の結果を用いて、最大値を得るまでの遺伝子解析結果を図 3 に示す。図中黒い部分が遺伝子座を確定した場所で、白い部分は未確定部である。

問題設定が簡単であるためか、初期の段階で複数箇所の遺伝子の確定が行われた。それは特に重要な、上位ビットに多く見られた。しかし、どの試行においても、優位遺伝子座の確定と共に、劣性遺伝子座の確定も早期に生じている。そして、最大値を得る世代前後で、全ての遺伝子座において『1』を選択していた。

この理由として、下位遺伝子座の影響は非常に微々たるもので、式(2)に示すように少ない差異を 3 乗するため、目的関数値に反映され難い。そのため、初期世代における下位遺伝子座の『1』の確定は、優位遺伝子座の影響によって生じた、言わば『釣られた』結果と考えることができる。

以上より、ほぼ影響を与えない下位遺伝子座が混ざるものの、重要なスキーマを抽出していることが確認された。

4. 結言

遺伝子の特徴抽出を用いた新しい GA 法を提案し、以下を明らかにした。

- (1) 特徴抽出という新しい遺伝オペレータを用いて最大値探索問題を解くことができた。
- (2) 探索過程において、有効なスキーマから抽出しているのを確認した。

参考文献

- (1) 尾田,金井,河登:日本機械学会北陸信越支部第 40 期総会・講演会講演論文集, No.037-1,(2003-3),133-134.
- (2) 榊 佳之:ヒトゲノム,岩波書店,(2001),83-94

Fig.3:Process of SFE