

遺伝的アルゴリズムにおける突然変異率の 遺伝子座ごとの設定に関する調査

青山大樹* 栗原正仁 小山聡 佐藤晴彦
(北大情報科学)[†]

1 はじめに

遺伝的アルゴリズムは生物の遺伝のメカニズムを参考に、交叉と突然変異を繰り返すことで最適解を探索する手法である。遺伝的アルゴリズムはその性質上、局所解に陥りやすく、そこから抜け出すために、突然変異というプロセスを導入している。突然変異は確率的に発生する。突然変異率は解の収束までの時間や解の精度に影響を与える重要な値のため、決定するための手法が多く提案されている[1]。

実際の生物の突然変異率に関する研究では、2012 年に Luscombe らによって、大腸菌における突然変異の発生はランダムではなく遺伝子座（遺伝子の位置）ごとに突然変異率が異なり、生存するために必要な遺伝子ほど突然変異確率は低く、そうでない遺伝子の突然変異率は高いままであることが確認されている[2]。このような突然変異率の違いによって、生命活動を維持できないような子孫を作らないようにリスクマネジメントを行っていると考えられる。

遺伝的アルゴリズムにおける突然変異率においても、重要な遺伝子の突然変異率を下げることで探索範囲を狭め収束までの時間が短くなると考えられる。

また、重要な遺伝子とそうでない遺伝子とを分けて考えることは、個体の持つ遺伝子を部分ごとに評価することになる。このように、個体の持つ遺伝子を全体ではなく部分ごとに注目する手法は近年多く報告されている[3]。

本研究では重要な遺伝子とそうでない遺伝子を持つ問題について考え、突然変異率をすべての遺伝子座で一定の場合と、遺伝子座ごとに異なるように設定した場合について、実験を行った結果の比較と考察について報告する。

2 遺伝的アルゴリズム

自然界における生物の適応では、ある世代を形成する個体群の中で、環境への適応度が高い個体が生き残り、生き残った個体の交叉や突然変異によって次の世代が

形成される。このメカニズムをモデルとして最適解を探索するアルゴリズムが遺伝的アルゴリズムである[4]。

遺伝的アルゴリズムでは、個体の特徴はそれぞれが持つ染色体によって決まる。染色体は遺伝子（通常では 0 または 1）の集合によって構成される。また、各遺伝子の置かれている位置を遺伝子座と呼ぶ。

次に遺伝的アルゴリズムの手順について説明する。

1. **初期世代の生成:** ランダムな遺伝子を持つ初期集団を生成する。
2. **評価:** 各個体の遺伝子から適応度を計算する。
3. **選択:** 適応度の高い個体を複数体選ぶ。
4. **交叉:** 選択した個体を元に次の世代を生成する。
5. **突然変異:** 突然変異確率にしたがって遺伝子座から一つ、あるいは複数選びその遺伝子を反転させる。
6. **終了判定:** 再度評価を行い、終了条件を満たしていれば終了する。満たしていない場合は手順 3 選択に戻る。

手順 3 の選択では一般的にエリート保存戦略が用いられる。エリート保存戦略では評価値の高い個体を複数個選び、選んだ個体を次の世代に残す選択方法である。

手順 4 の交叉は一点交叉、多点交叉、一様交叉などさまざまな交叉方法がある[5]。

手順 5 の突然変異での突然変異率は通常、すべての遺伝子座で一定の突然変異率である。また、染色体の長さを L としたときの突然変異率は経験上 $1/L$ の場合が最も良いとされている[5]。

3 提案手法の原理

提案手法である遺伝子座ごとに突然変異率を設定する遺伝的アルゴリズムについて説明する。

本手法では重要な遺伝子とそうでない遺伝子に分ける必要がある。遺伝的アルゴリズムにおいて重要な遺伝子とは、他の遺伝子よりも評価に大きく影響を与える遺伝子である。そのため評価に大きく影響を与える遺伝子座における突然変異率を下げ、その他の遺伝子座は従来

*aoyama_daiki@complex.ist.hokudai.ac.jp

[†]札幌市北区北 14 条西 9 丁目北海道大学大学院情報科学研究

の遺伝的アルゴリズムと同様に $1/L$ とする遺伝的アルゴリズムが提案手法である。

4 実験

ここでは実験の概要について説明する。

実験では評価に大きく影響を与える遺伝子とそうでない遺伝子を持つ問題として次のような問題について考える。染色体を4等分し、それぞれについて評価を行い、その合計によって評価が決まる。また、4等分したうちのいくつかは評価関数に大きく影響を与えている。

$$f(x) = \omega_1 f_1(x) + \omega_2 f_2(x) + \omega_3 f_3(x) + \omega_4 f_4(x) \quad (1)$$

$f(x)$ は個体の評価関数、 $f_k(x)$ ($k=1, 2, 3, 4$)は4等分したそれぞれの染色体の評価値を表している。 ω_k は評価の重みであり、この値が高くなるとその部分の遺伝子は個体の評価値により多く影響を与えているといえる。それぞれの染色体は x_{ki} ($i=1, 2, \dots, N$)の遺伝子(1 or 0)を持っている。

$f_k(x)$ は次の式で定義される。

$$f_k(x) = \begin{cases} N - \sum_{i=1}^N x_{ki} - 1 & \text{if } \sum_{i=1}^N x_{ki} < \frac{N}{2} \\ \sum_{i=1}^N x_{ki} & \text{if } \sum_{i=1}^N x_{ki} \geq \frac{N}{2} \end{cases} \quad (2)$$

式(2)は、すべての遺伝子が0の場合に局所解となり、すべての遺伝子が1の場合に最適解となる。また、0と1の遺伝子が混在する場合には低い値となる。このよう関数は、だまし問題としてよく知られている[6]。

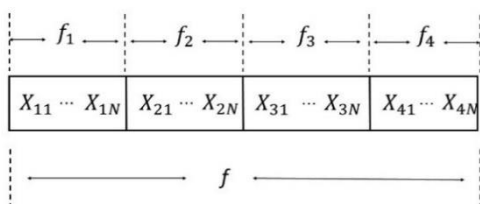


Fig.1 問題の概念図

Fig.1 は4等分した染色体と、それぞれの持つ遺伝子とその評価値の関係を表している。

このような問題に対して従来の遺伝的アルゴリズムと提案手法によって探索を行う。

交叉方法は一点交叉で、交差点は染色体の中心に設定する。突然変異率は、従来の遺伝的アルゴリズムではす

べての遺伝子座で $1/L$ とし、提案手法では f_2, f_3 部分の遺伝子座で $1/4L$ と従来法と比べ低い値に設定し、その他の遺伝子座では従来法と同じ $1/L$ と設定する。選択はエリート保存戦略を用いる。

このような設定で実験を行い、それぞれについて解の収束までにかかる時間や局所解に陥りやすいかどうかについて調べ、それぞれのアルゴリズムの特徴について検討する。

5 おわりに

本研究では重要な遺伝子とそうでない遺伝子を持つ問題について考え、突然変異率をすべての遺伝子座で一定の場合と、遺伝子座ごとに異なるように設定した場合について、実験を行った結果の比較と考察について報告した。

遺伝的アルゴリズムが提案された1970年代に比べて、近年では遺伝学が発展し、自然界の遺伝の仕組みについてより多くのことがわかっている。また、計算機の性能も格段に向上し、複雑な処理も比較的短時間で実行できるようになった。

遺伝的アルゴリズムはもともと生物の遺伝の仕組みをモデルとして考案されたアルゴリズムではあるが、ここで再び自然界の遺伝を参考にすることで、遺伝的アルゴリズムを改良できるだろうと考えている。

今後の課題としては、重要な遺伝子座を動的に特定し、突然変異率を動的に設定することや、並列分散遺伝的アルゴリズムと併用することで解の収束にどのような影響を与えるのかについて検討することなどがあげられる。

参考文献

- [1] 北野宏明 (1997), 遺伝的アルゴリズム③, 産業図書
- [2] Iñigo Martincorena, Aswin S. N. Seshasayee & Nicholas M. Luscombe, Nature **485**, 95-98 (03 May 2012) doi:10.1038/nature10995
- [3] 棟朝雅晴, エピスタシス尺度に基づくリンケージ同定手法の提案, 情報処理学会論文誌「数理モデル化と応用」, Vol. 43, No. SIG10(TOM7), pp.6-13(2002)
- [4] 北野宏明 (1993), 遺伝的アルゴリズム, 産業図書
- [5] 伊庭斉志 (2002), 遺伝的アルゴリズム, 医学出版
- [6] 松井, 小杉, 報酬付きマルコフ過程による遺伝的アルゴリズムの基礎的な解析, 電子情報通信学会論文誌(D-II), vol. J80-D-II, No. 2, pp.607-617(1997)