

EMACI: 遺伝的アルゴリズムの探索解の時系列可視化手法の提案

大山 喜冴 (指導教員: 伊藤貴之)

1. はじめに

遺伝的アルゴリズム(GA: Genetic Algorithm)は、解空間構造が不明で、全探索が不可能なほど広大な解空間を持つ問題に有効な最適化手法である。GA はダーウィンの進化論をモチーフに、環境に適合する度合い(適合度)を数値で表し[1][2], その適合度を最適化問題の目的関数と考える。適合度の値が進化と共に大きくなるのが、最適化問題における最適解に近づくことに相当する。

本論文では GA の反復処理の過程に関する可視化手法を提案する。本手法では GA の解探索過程を可視化し、ユーザが可視化結果を観察しながら、意図的に突然変異確率・交叉確率を調節する機能を備える。この機能により、ユーザはより短時間で準最適解を発見できると考えている。

2. 関連研究

遺伝的アルゴリズム

GA は解の候補を遺伝子で表現した「個体」を複数用意し、世代が進むに連れ適合度の高い個体を優先的に選択して、交叉・突然変異などの操作を繰り返しながら解を探索するアルゴリズムである[1]。GA の解探索性能は遺伝的演算手法だけでなく、そのパラメータの個体数・突然変異確率・交叉確率などにも大きく依存する。そのため、遺伝的演算手法やパラメータが解探索性能に及ぼす効果を知るために、適合値の改善効果を指標とするのが一般的である。

Two-Tone Pseudo Coloring

2つの色を帯領域に割り当て、その各々の色の占める幅を制御することで、任意のスカラ値を表現する可視化手法が提案されている[3]。特に時系列データにおいて、数値変化を正確に、また微小な変化が読み取りやすいように可視化できるのが特徴である。この手法による表示例を図1に示す。

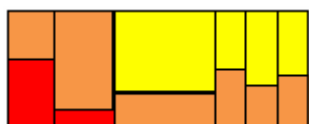


図1: Two-Tone Pseudo Coloring による表示例

平行座標プロット

平行座標プロット(PCP: parallel coordinate plot)は

多変数データを視覚化するのに有用な統計グラフの一つである。PCP では変数の座標軸を平行に配置し、全ての変数で最小値が下端に、最大値が上端になるように変数値をプロットする。そして隣接する座標軸上の変数値をそれぞれ線分で結ぶ。平行座標軸の間隔を調節することにより、すべての変数の座標軸を一面に表示できることから、同時に全ての辺量を表示する事が可能である[4][5]。

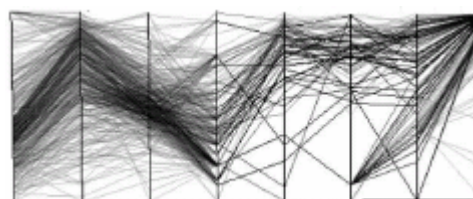


図2: 平行座標プロットによる可視化結果

3. 提案手法

本手法は以下の3つの機能を搭載するものである。

時系列表示部: 世代ごとのエリート解のスコア値とパラメータ値を時系列で表示。

世代全表示部: 各世代の全ての遺伝子で表現した個体のパラメータ値を表示。

GUI: ユーザが可視化結果を観察しながら、意図的に突然変異確率・交叉確率を調節するためのGUI。

本手法は、時系列表示部と世代全表示部を同じウィンドウに搭載し、反復処理過程に応じて GUI が起動するようになっている。図3と図4に実行結果を示す。図3の実行結果の上方が時系列表示部、下方が世代全表示部であり、図4が GUI である。

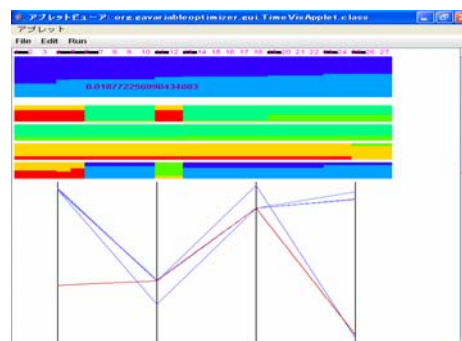


図3: 提案手法による可視化例

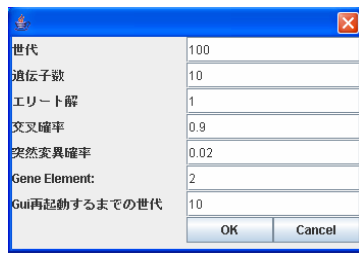


図4：GUI

時系列表示部

本手法では、解の候補を遺伝子で表現した「個体」の中で、最も適合度の高い個体をエリート解とみなす。また本手法では、遺伝子の適合度をスコア値とし、遺伝子を構成する値をパラメータ値とする。時系列表示部では、世代が進むごとに更新されるエリート解のスコア値とパラメータ値の変化を、Two-Tone Pseudo Coloring を用いて表示する。

世代全表示部

ある一世代について、非エリート解を含めた全遺伝子の全パラメータを、PCP を用いて表示する。ユーザが選んだ世代の解候補の遺伝子を観測できるので、個体同士の相対的位置関係も理解する事ができ、解の収束や集散の様子などを把握できるようになる。

本手法では座標軸の数は、パラメータの次元数と同じであり、座標軸上の変量値は各パラメータの変数としている。

GUI

ユーザが GA の反復過程を観察しながら、状況に応じて意図的に、突然変異確率・交叉確率・遺伝子数を対話的に再設定できる。ユーザが指定した一定の世代数にわたって、エリート解のパラメータ値・スコア値に変化が見られなかった場合に、この GUI を起動する。

4. 実行結果

本論文では以下の問題について、本可視化手法を用いて準最適解を求めた。

パラメータ数	10
パラメータ範囲	0.0~1.0
スコアの極大値	6

上記の問題を 25 回ずつ、一回につき 150 世代まで繰り返し実験した結果の平均値を、表 1 に示す。以下では、実行結果とその考察を述べる。

結果として、GA の反復処理過程中に状況に応じて突然変異確率・交叉確率を意図的に調節する方が良い結果が得られた。理由としては、再設定をしない場合、GA の問題点でもある初期収束に陥りやすいか

らだと考えられる。ここで、初期収束とは初期世代で偶然、他の個体より適合度が圧倒的に高い個体が生まれたとき、その個体の子孫ばかりが増えて初期段階で局所解に収束してしまう現象である。

また、初期収束が起こってしまった場合は最初から計算をやり直す方が良い結果が生まれる。そのため、各世代の解候補を表示することは初期収束の早期発見にも繋がるので、より所要時間も短縮される。

表1：実行結果

	局所解回避確率	スコア平均値
再設定無し	0.48	0.9781
交叉確率を再設定	0.52	0.9911
突然変異確率を再設定	0.88	0.9729

4. まとめ

本論文では GA の反復処理の過程で得られる解探索過程を可視化させ、可視化結果を観察しながら意図的に突然変異確率・交叉確率を調節することができる可視化手法を提案した。

本手法により、解候補である個体同士の相対的位置関係も理解する事ができ、解の収束や集散の様子などを把握できるようになった。そのため、解の収束や集散に応じて突然変異確率や交叉確率を操作できたため、通常の GA の反復処理より早い世代で準最適解を発見する事が出来た。

謝辞

指導教員の伊藤准教授。副査の河村教授。京都大学の小山田教授と山下准教授に感謝の意を示します。

参考文献

- [1] D. E. Goldberg, Genetic Algorithms in Search, Addison-Wasley Publishing Company, 1989.
- [2] D. E. Goldberg, Optimization and Machine Learning, Addison-Wasley Publishing Company, 1989.
- [3] T. Saito, H. N. Miyamura, M. Yamamoto, H. Saito, Y. Hoshiya, T. Kaseda, Two-Tone Pseudo Coloring: Compact Visualization for One-Dimensional Data, IEEE Information Visualization '05, pp. 173-180, 2005.
- [4] H. Y. Lee, H. L. Ong, Visualization Support for Data Mining, IEEE EXPERT, 1996.
- [5] A. Inselberg, B. Dimsdale, Parallel Coordinate: A Tool for Visualizing Multi-Dimensional Geometry, IEEE Visualization, 1990.