

進化計算シンポジウム 2019 に参加して

川 畑 忠 宏

Tadahiro KAWABATA

電子情報学専攻修士課程 2019 年度修了

1. はじめに

私達のチームは2018年12月14日から15日にかけて、兵庫県の淡路島南あわじ市 Hotel & Resorts MINAMIWAJII で開催された進化計算シンポジウム 2019 に参加し、「優良部分ルールを保存する文法進化」という題目で発表した。



図 1 ポスター発表の様子

2. 優良部分ルールを保存する文法進化

2.1 研究背景

ウェブ上に蓄積された大容量データから有用なルールやパターンを見つけ出す技術に注目が集まっている。その解法の一つである進化計算手法の Grammatical Evolution (GE) は、進化により性能の高いルールを生成する手法であり、あらかじめ記述したプロダクションルールに従いルールを生成する。

その高度化として、マッピング法や初期集団の生成方法の改良が考えられている。GE では、マッピング法を用いて遺伝子から表現型のルールを生成するが、このマッピングでは 1 対 1 の射影になってい

る。改良法として π GE があり、これは遺伝子にインデックス情報を付与することで、効率的なマッピングを実現しているが、個体の構造情報は考慮していない。また、初期集団の生成を C 4.5 によって生成する先行研究など、初期化に関する研究報告はある。しかし、探索の効率化に関する報告は少ない。

本研究では遺伝子型と表現型とのマッピング手法に着目し、優良解に頻出したモチーフが継承されるマッピング手法を提案する。また、GE の解探索は多様性がなくなる恐れがあり、汎化性を高める Lexicase selection を導入し、その性能を分析する。

2.2 Grammatical Evolution

GP では木構造によって個体が表現されているが、遺伝的操作によって文法的に正しくない個体が生み出される恐れがある。その為 GP では、そのような無効個体を生成しないように遺伝的操作に制約をかけるか、もしくは生成した無効個体を修正するように設定しなければならない。そこで、そのような制約を解消するために提案されたのが Grammatical Evolution (GE) である。

GE とは、C. Ryan らによって提案された機械学習の一種である。主に GE は決定木のような文法ルールを記述するために用いられ、例えばデータから有益な情報を抽出するために用いられる。

GE の特徴として個体は、遺伝子型 (genotype) と表現型 (phenotype) の二つの構造情報を所有していることが挙げられる。生物における染色体を模している遺伝子型は、可変長の二進数配列で表現されているため、Genetic Algorithm (GA) と同様の柔軟性のある遺伝的操作が可能である。

2.3 提案法

GE には有益な特徴を失うことなく次世代へと継続させるアルゴリズムが必要となる。そこでユーザーが選好した個体に頻出しているパターンを学習し、マッピングの際に出現した構造を活用する手法を提案する。前世代で求められた個体の有益な特徴

を壊すことなく継承することで、次世代の解探索に活かすことができると考えられる。

提案法では、一定の世代ごとに GE の表現型から motif と呼ばれる共通するパターンを抽出する。本研究では、motif には重要な情報が保存されていると考え、抽出した motif を以降のマッピングルールの候補へと加えることで、重要な構造の保存を行う。motif の抽出には、配列データベースから共通するパターンを抽出する modified prefixspan 法を用いた。

3. lexicase selection

lexicase selection とは、Lee Spector らによって提案された個体選択の一種である。個体選択の多くは全体を通して評価値が高い個体を選択するが、lexicase selection はランダムなデータの評価値を考慮することで選択を行う。lexicase selection のアルゴリズムを図 2 で示す。



図 2 lexicase selection

4. 評価実験

提案したマッピング法の有効性を確認するために、UCI machine learning Repository から 10 種類の分類問題を用いて性能評価を行った。

4.1 実験パラメータ

実験におけるパラメータは、試行数が 20、個体数 100、世代数 50 である。また、5 世代ごとに母集団から評価値が高い順に 50 個体から motif 抽出を行い、50 個体中 5 個体以上で出現している構造を motif として数える。

4.2 比較手法

提案法と lexicase selection を評価するための実験として、tournament selection を行う GE (normal)、マッピングの際に motif 抽出を行う GE (motif)、lexicase selection を用いる GE (lexicase)、マッピングの際に motif 抽出を行い lexicase selection を行う GE (lexicase selection + motif) の 4 種類の手法を用いて、それぞれのデータセットに対して解探索する比較実験を行う。

4.3 結果

10 種類の分類問題の結果を、表 1 に示す。

表 1 提案手法の評価結果

data set	normal	motif	lexicase	lexicase+motif
Iris	0.970792	0.971782	0.979208	0.978713
Digits	0.766760	0.766899	0.767875	0.767387
Wine	0.954348	0.956522	0.965942	0.971377
Diabetes	0.816406	0.816211	0.812402	0.811719
ILP	0.748936	0.748453	0.746905	0.746905
Glass	0.734731	0.732036	0.742216	0.740719
Bupa	0.822500	0.821500	0.811000	0.814000
WQR	0.637922	0.640426	0.627785	0.628098
WQW	0.521555	0.521505	0.517042	0.516792
BCW	0.958297	0.960991	0.964009	0.965517

4 種類の手法を比較したところ、あまり性能の差は出なかった。しかし、motif 抽出を行う手法の方が探索速度が早く、探索途中では良い評価値を得ることができている。motif 抽出により重要な個体の構造を保存することにより、探索の効果が motif 抽出によって向上していることが分かる。

5. おわりに

進化計算シンポジウムに参加して、私の研究内容について発表した。参加者の方々の意見を聞くことで、非常に有意義な時間を過ごすことができた。

最後に、今回の発表を行うにあたって、多大なご指導を頂いた小野景子講師、研究室の皆様深く感謝いたします。