

2 進木の進化と統計力学

日立、東大先端研^A

甲元 洋、岡部 洋一^A

1. はじめに

遺伝的プログラミング (以下 GP と略す。) は、進化論的な木構造データの探索手法である。特に、木構造のデータ自体を遺伝子と捉え、世代交代として突然変異あるいは交叉を施すことに特徴がある。問題として取り上げるのは、次の2点である。

(1) GP は、ランダムサーチを超える能力を有するか？

(2) GP において、交叉演算は有効であるか？

本稿では、簡略化した2進木モデルにおいて、確率差分方程式を立てて探索成功確率を計算した結果について報告する。計算結果は、定数オーダーの有効性を示す。

2. モデル化と確率差分方程式

進化の対象を2進木として、非終端記号を1種類 (*)、終端記号を1種類 (1) とする簡略モデルを考える。本モデルでは、木の評価はそのトポロジーのみで決まる。問題が自明にならないよう、演算 * は非可換とした。また、2進木の木の深さは、制限をかけて3とする。そうすると、考えるべき木の種類は、終端記号単独を含めて26種となる。2進木 i の評価には、次のような関数 E を用いる。

$$E_{*,1,seed}[i]; \quad \{i\} \rightarrow \{n \mid 0 \leq n \leq 1000, n \in \mathbb{N}\} \quad i \rightarrow n \quad \dots (1)$$

この関数は、乱数で与えられ、seed は乱数シードを表す。

次に、2進木の集合が十分大きく、次世代の2進木の存在確率が現世代の存在確率のみに依存して決定されるとして、 t 世代後の2進木 i の存在確率 $p_i(t)$ と、各世代に1個体ずつ2進木を取り出した場合に2進木 i を得る確率を探索成功確率 $q_i(t)$ と定義する。すると、 $p_i(t)$ と $q_i(t)$ は、以下の確率差分方程式の解となる。

$$p_i(t+1) = (1 - \alpha - \beta)p_i(t) + \alpha \left[\sum_{j \neq i} M[j][i]p_j(t) + (1 - \sum_{j \neq i} M[i][j])p_i(t) \right] + \beta \left[\sum_{j \neq i} C[j][i]p_j(t) + (1 - \sum_{j \neq i} C[i][j])p_i(t) \right] \quad \dots (2)$$

$$q_i(t+1) = (1 - q_i(t))p_i(t+1) + q_i(t) \quad \dots (3)$$

ここで、各記号は次のように定義する。

$p_i(t)$: t 世代目の2進木 i の存在確率 $q_i(t)$: t 世代目迄に2進木 i を得る確率

α : 突然変異比率 β : 交叉比率

$M[i][j]$: 突然変異による2進木 $i \rightarrow j$ への遷移確率行列

$$M[i][j] = \sum_{i \rightarrow j \text{ の可能な遷移}} 1/N[i] * \text{部品生成確率} * \text{遷移確率} \quad \dots (4)$$

$C[i][j]$: 交叉による2進木 $i \rightarrow j$ への遷移確率行列

$$C[i][j] = \sum_{i \rightarrow j \text{ の可能な遷移}} 1/N[i] * \left(\sum_k B[k][i] * p_k(t) / N[k] \right) * \text{遷移確率} \quad \dots (5)$$

ここで、(4),(5)式の遷移確率は、2進木の存在確率が Boltzmann-Gibbs 分布で決まると考えた際の条件付き確率に等しいとしている。(Tは系の温度パラメータである。)

$$\text{遷移確率} = \exp(-E[j]/T) / \{ \exp(-E[i]/T) + \exp(-E[j]/T) \} \quad \dots(6)$$

また、(4),(5)式の $N[i]$ は 2 進木 i のノード数、(4)式の部品生成確率は終端記号と非終端記号を確率 1/2 で割り当てた際 2 進木 j のために必要な部品としての 2 進木ができる確率を、(5)式の $B[k][i]$ は 2 進木 j のために必要な 2 進木 k の内部に存在する部品としての 2 進木 i の員数を表す(図 1 参照)。

3. 計算結果

図 1 に $B[k][i]$ を示す。図 2 に初期条件として各 2 進木に 1/26 の確率を与えた場合の 100 世代後までの計算結果を示す。この計算は $\alpha=0.1$, $\beta=0.8$ であり、 $\alpha=0$, $\beta=0$ の場合がランダムサーチに対応する。また、 $\alpha=0.9$, $\beta=0$ の場合には、図 2 の各曲線はランダムサーチの曲線に近づく傾向にある。

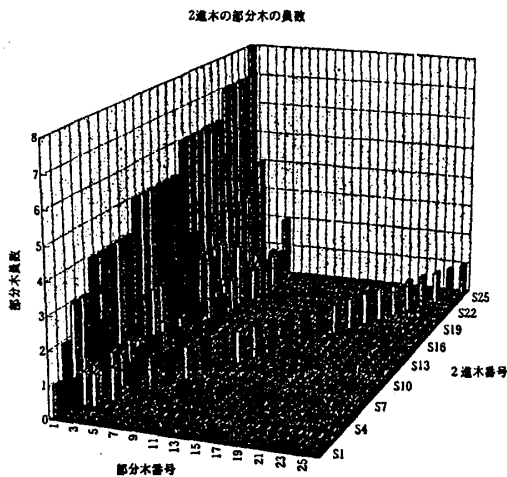


図 1 2 進木中の部分 2 進木員数
・対角成分は 2 進木本体
・部分 2 進木は 3 層以下

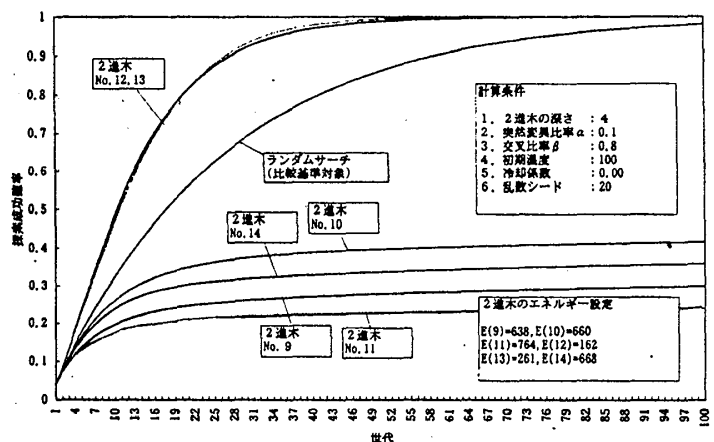


図 2 探索成功確率の時間発展
・突然変異 10%、交叉 80%
・初期存在確率は 1/26

4. 結論

簡略化した 2 進木モデルによって、2 進木の集合が大きな理想的な系という仮定のもと、探索成功確率の時系列計算を行った。その結果、次の結論を得た。

(1) エネルギー E の値が小さければ、探索成功確率はランダムサーチに比較して、定数オーダー程度の有効性が認められる。

(2) 交叉は、突然変異のみの場合に比較してわずかながら有効性を示す。

但し、両者とも、エネルギーと探索成功確率の間には、順序関係がそのまま成り立つわけではなく、問題依存性(乱数によるエネルギー設定)を示す。

文 献

- 1) M. Michell: An Introduction to Genetic Algorithm, MIT(1996)175-183
- 2) 伊庭斉志: 遺伝的プログラミング, 東京電機大学出版(1996)43-72.