

Extremal Optimization を用いた調停グラフの交差数減少

田村 慶一[†] 森 康真[†] 北上 始[†]

系統樹を比較する手法の1つとして調停が提案されている。調停では、系統樹や分類樹を順序木とみなし、2つの順序木を向き合わせた調停グラフを作成する。調停を潤滑に進めるためには、順序木の葉節点間に交差がないか、もしくはできるだけ交差が少ない調停グラフを作成する必要がある。本論文では、Extremal Optimization (EO) による調停グラフの交差減少を提案する。EO を適用するにあたり、EO に対して、(1) ルーレット選択により状態遷移の候補となる構成要素を選択する、(2) 複数の近傍解中の最良解を次世代の解とする、という2つの改良を加えた、評価実験の結果、提案手法による調停グラフの交差数減少の方が、EO、SA やタブーサーチと比較して優れていることを確認した。

Reducing Crossovers in Reconciliation Graphs with Extremal Optimization

KEIICHI TAMURA,[†] YASUMA MORI[†] and HAJIME KITAKAMI[†]

The reconciliation is a one of techniques for comparing phylogenetic trees. In the reconciliation, a phylogenetic tree is considered to be an ordered tree, order trees are made to be opposite, and the reconciliation graph is made. There is no intersection between leaf nodes to advance reconciliation to lubrication or it is necessary to make reconciliation graph where intersection is a little. This paper proposes the reducing crossovers in reconciliation graphs with Extremal Optimization (EO). There are two characteristics in the proposed method. The construct which serves as a candidate of a change state by roulette selecting is chosen. The best solution in two or more neighborhoods is chosen as the next solution. The experimental results show that the proposed method is excellent as compared with EO, tabu search, SA and GP.

1. はじめに

生物の多様性を解明するために、系統樹を比較する研究が行われている¹⁾。系統樹を比較する手法の1つに調停²⁾がある。調停では、系統樹をそれぞれ順序木とみなし、比較する2つの順序木を向き合わせた調停グラフを作成する。そして、作成した調停グラフのノードを一对一に対応付け、比較する2つの順序木の違いを明らかにしていく。

調停を潤滑に進めるためには、順序木の葉ノード間に交差がない状態か、または、交差が少ない状態にする必要がある。調停グラフの交差数減少とは、交差数がより少ない調停グラフを求める問題である。この問題は、組合せ最適化問題であり、厳密解を求めようとすると非常に大きな計算量を必要とする。

本論文では、Extremal Optimization (EO)³⁾による調停グラフの交差数減少を提案する。EO を調停グラフの交差数減少に適用するにあたり、EO に対して、(1) ルーレット選択により状態遷移の候補となる構成要素を選択する、(2) 複数の近傍解中の最良解を次世

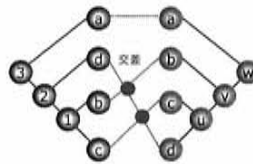


図1 調停グラフの例

代の解とする、という2つの改良を加えた。

評価実験の結果、EO と比較して、提案手法の方が、より早い段階で最良な解を求めることができることを確認した。また、提案手法と、タブーサーチと SA とを比較した結果、提案手法の方が優れていることを確認できた。

本論文の構成は以下の通りである。第2章では、調停グラフの交差数減少について説明する。第3章では、EO の概要を示す。第4章では、提案手法を説明し、第5章では、関連研究について述べる。第6章で評価実験の実験結果を示し、最後に、第7章において、本論文をまとめる。

2. 調停グラフの交差数減少

調停グラフは、2つの順序木、順序木 $T1 = (V1, E1)$,

[†] 広島市立大学大学院情報科学研究科
Graduate School of Information Sciences, Hiroshima
City University

順序木 $T2 = (V2, E2)$ から構成される。 $V1$ と $V2$ とはノード集合であり、 $E1$ と $E2$ とは辺集合である。 ルートノード以外のノードの入次数は 1 であり、出次数が 0 のノードを葉ノードと呼ぶ。

図 1 に調停グラフの簡単な例を示す。 図 1 では、生物種 'a', 'b', 'c', 'd' を別々の観点で分類した 2 つの順序木を図示している。 図のように 2 つの順序木を向き合わせたグラフが調停グラフとなる。 同じ葉ノード間を線で結ぶと、葉ノード間に 2 つの交差が生じる。 この葉ノード間の交差を削減する問題が調停グラフの交差数減少である。

葉ノードの順序が次の配列として与えられたとする。

$$OL1 = [ol1_1, ol1_2, \dots, ol1_n]$$

$$(ol1_i \in L1, \mathcal{L}(ol1_i) \in L_{leaf})$$

$$OL2 = [ol2_1, ol2_2, \dots, ol2_n]$$

$$(ol2_i \in L2, \mathcal{L}(ol2_i) \in L_{leaf})$$

\mathcal{L} はノードのラベル名を返す関数である。 このとき、結合行列 M の要素 $m_{i,j}$ を以下のように、

$$m_{i,j} = \begin{cases} 1 & \text{if } \mathcal{L}(ol1_i) = \mathcal{L}(ol2_j), \\ 0 & \text{otherwise.} \end{cases}$$

と表すと、葉ノード間の交差数は、以下の式で求めることができる。

$$\sum m_{j,\beta} m_{k,\alpha} [1 \leq j < k \leq n, 1 \leq \alpha < \beta \leq n]$$

葉ノードの順序を変化させると結合行列が変化し、交差数が変化する。 交差数を削減するにあたり、順序木のノード間で交差を発生させてはならない、順序木が示す分子進化系統樹の系統や生物分類樹の分類を壊してはならない、という制約を守る必要がある。

3. Extremal Optimization

Extremal Optimization (EO) は、生物進化の Bak-Sneppen モデル⁴⁾ の考えに基づき提案されている発見的探索手法である。 EO では、適応度 λ_i が最も小さい構成要素 O_i か、もしくは、適応度のランキングのべき分布に従って (べき分布に従う EO を τ EO と呼ぶ)、構成要素を状態遷移の候補として選択する。 そして、当該構成要素を状態遷移をさせた近傍解をランダムに生成し、次世代の解とする。 この選択と状態遷移を繰り返しながら、解の探索を進めていく。

4. 提案手法

4.1 構成要素

構成要素は、以下のように、同じラベル名を持つ葉ノードのペアと定義する。

$$O_i = \{ol1_i, ol2_{\delta(i)}\} \quad (\mathcal{L}(ol1_i) = \mathcal{L}(ol2_{\delta(i)}))$$

関数 $\delta(i)$ は、 $ol1_i$ のラベル名と同じラベル名を持つ $OL2$ の添字番号を返す関数とする。

4.2 適応度

葉ノード $ol1_i$ と葉ノード $ol2_{\delta(i)}$ の間に存在する交差数 $C(M, i)$ は、以下の計算式で求めることができる。

$$C(M, i) = \sum_{l=i+1}^n \sum_{m=1}^{\delta(i)-1} \frac{m_{l,m}}{2} + \sum_{l=1}^{i-1} \sum_{m=\delta(i)+1}^n \frac{m_{l,m}}{2}$$

$C(M, i)$ を使用して、構成要素 O_i の適応度 λ_i を以下のように、

$$\lambda_i = \frac{C(M) - C(M, i)}{C(M)},$$

と定義する (ただし、 $C(M) = 0$ の場合を除く)。 $C(M)$ は調停グラフで生じている交差数である。

4.3 状態遷移

構成要素 O_i の状態遷移の手順を以下に示す。

- (1) 順序木 $T1$, 順序木 $T2$ をランダムに 1 つ選択する。
- (2) 順序木 $T1$ が選択されたならば、 $AS(T1, ol1_i)$ を候補集合 CS とする。 順序木 $T2$ が選択されたならば、 $AS(T2, ol2_j)$ を候補集合 CS とする。 ここで、 $AS(T1, ol1_i)$ は、順序木 $T1$ について、葉ノード $ol1_i$ からルートノードへの道に存在する中間ノードの集合を示す。
- (3) 候補集合 CS の中から子ノードの数が 2 つ以上の中間ノードをランダムに 1 つ選択し、選択したノードの子ノードの順序関係を入れ替える。

4.4 アルゴリズム

Algorithm1 に、提案手法における交差数減少の流れを示す。 まず、ルーレット選択により、構成要素を 1 つ選択する。 次に、前節で示した状態遷移の方法で、選択した構成要素を状態遷移することで近傍解を 1 つ生成する。 そして、指定した近傍解の生成数に達するまで、ルーレット選択と近傍解の生成を繰り返す。 指定した近傍解の生成数に達すると、生成した近傍解の中から最良解を選択し、次世代の解とする。

調停グラフの交差数減少では、構成要素の適応度が一定値以上大きくならない場合が考えられる。 そこで、ルーレット選択により状態遷移の候補となる構成要素を選択し、適応度が小さい構成要素も、ある程度、状態遷移の候補として選ばれやすくしている。

また、状態遷移により、局所的な変化が発生するこ

とや大域的な変化が発生することがある。そこで、解が急激に改悪の方向に進まないようにするために、近傍解を複数生成して、複数の近傍解の中から最良の解を次世代の解としている。

Algorithm 1 EO_CROSS

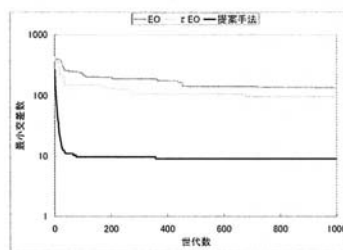
- 1: 初期解 S_{init} として、ランダムに調停グラフを生成。
 - 2: $S_{best} := S_{init}; S_{current} := S_{init};$
 - 3: **while** 指定した世代数以下ならば **do**
 - 4: $S_{current}$ のすべての構成要素 O_i に関して、適応度 λ_i を算出。 $N := \phi;$
 - 5: **while** 指定した近傍解生成数以下ならば **do**
 - 6: ルーレット選択により、構成要素 O_i を1つ選択。
 - 7: 構成要素 O_i を対象として、 $S_{current}$ に対して状態遷移を行った近傍解 $N(O_i)$ を生成。
 - 8: $N := N \cup N(O_i);$
 - 9: **end while**
 - 10: N の中から最良解 N_{best} を選択し、次世代の解 S_{next} とする。
 - 11: $S_{current} := S_{next};$
 - 12: もし、 $\mathcal{F}(S_{current}) < \mathcal{F}(S_{best})$ ならば、
 $S_{best} := S_{current};$
 - 13: **end while**
-

5. 関連研究

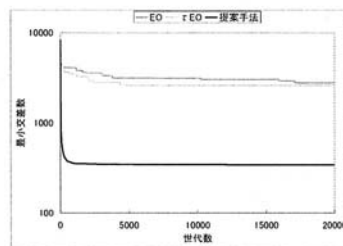
調停グラフ交差数減少に対して、結合行列の行・列を探索的に入替える手法¹⁾と、遺伝的アルゴリズムを用いた手法^{5),6)}が提案されている。遺伝的アルゴリズム(GA)と局所探索とを組み合わせた手法は、Simple GAを使用して求めた最良解に対して、さらに、局所探索を行い、交差数を削減する手法である。しかしながら、この手法では、解が初期集団に依存し、局所最適解におちいりやすいという課題が存在する。一方、世代交代モデルとしてMGGを適用した手法では、課題となっていた多様性の維持と、局所最適解に陥る問題を解決することができた。MGGは、ある処理時間をかけても多様性を維持することを目的とする手法ではあるが、調停グラフの規模が大きくなるほど、実用的な時間で問題を解くことができない。

6. 評価実験

提案手法の有効性を示すために、評価実験では、次の2つの実験を行う。実験1では、提案手法と、EO



(a) hashimoto97



(b) コケ

図2 世代数と最小交差数

と τ EOの比較を行う。実験2では、提案手法、タブーサーチ、もしくは、SAを使用して調停グラフの交差減少を行ったときをそれぞれ比較する。

図2(a)と図2(b)に、データセット「hashimoto97」、データセット「コケ」を使用したときの、世代数と最小交差数(最良解の交差数)を示す。提案手法における近傍解の生成数は50である。 τ EOにおける τ は1.5とする。測定は、5回行い、最良の3つの結果の平均を測定値として示す。また、最大世代数はそれぞれ1000、20000とする。

データセット「hashimoto97」では、EOと τ EOとを使用した場合、最小交差数は100より下がることはないが、提案手法では、最適解である交差数9を求めることができている。これは、交差数が少なくなるほど改悪になる可能性が高くなるため、近傍解を1つしか生成しないEOと τ EOとでは、100前後で、解が改悪になったり改善したりと、堂々巡りをしていると考えられる。同様に、データセット「コケ」においても、提案手法の方が良い結果を示している。

実験2では、測定を200回行い、各測定の最後に得られた最小交差数の度数分布を求めた。最大世代数は5000回で測定を行った。提案手法とタブーサーチでは、近傍解を50個生成する。また、SAにおけるパラメータである初期温度は1000、温度係数は0.9とした。

図3に、データセット「hashimoto97」を使用したときの測定結果を示す。グラフは、棒グラフが頻度を示し、線が累積分布を示している。提案手法、タブーサーチとSAともに最適解を求めることができている

表 1 データセット

データ名	生物分類樹		分子進化系統樹		最小交差数
	非葉ノード数	葉ノード数	非葉ノード数	葉ノード数	
hashimoto97	201	40	39	40	9
コケ	83	207	187	207	-

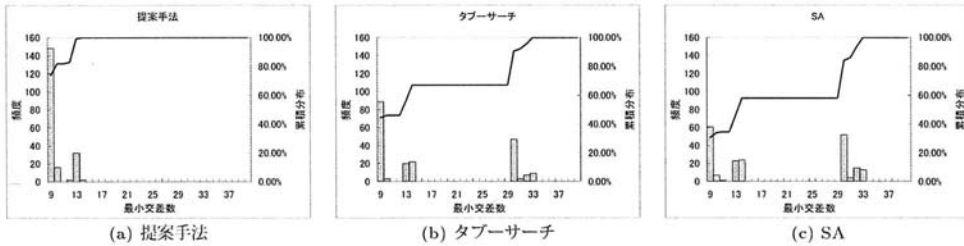


図 3 最小交差数の度数分布 (hashimoto97)

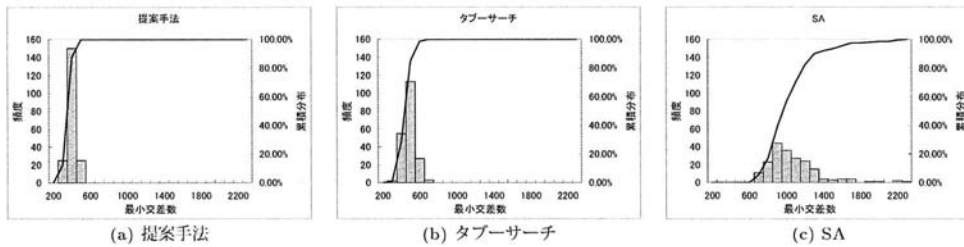


図 4 最小交差数の度数分布 (コケ)

が、最も頻度が高いのが提案手法である。また、タブーサーチと SA とには、最小交差数 30 周辺にいくつかの山が存在する。これは、最小交差数が 30 の周辺に局所最適解が存在するためである。この 30 前後の山は、提案手法では発生していない。同様に、図 4 に示すように、データセット「コケ」においても提案手法の方が良い結果を示している。

7. まとめ

本研究では、EO による調停グラフの交差数減少を提案した。提案手法では、EO に対して、ルーレット選択と複数近傍解生成という 2 つの改良を加えた。評価実験の結果、提案手法は、EO、 τ EO、タブーサーチや SA を使用したときよりも、評価の高い結果を示した。これからの課題として、最適な近傍解生成数の自動設定方法の検討を行う予定である。

謝辞 本研究の一部は、日本学術振興会・科学研究費補助金（基盤研究 (C)、(一般)、課題番号：17500097）、文部科学省・科学研究費補助金（課題番号：18700094）の支援によりおこなわれた。

参考文献

1) 北上始, 森康真, 太田聡史, 斎藤成也: 異種系統樹間の調停のためのゼロ交差制約の充足, 情報処理

学会論文誌データベース, Vol. 40, No. 5, pp. 1-14 (1999).

2) Stege, U.: Gene Trees and Species Trees: The Gene-Duplication Problem in Fixed-Parameter Tractable, *Workshop on Algorithms and Data Structures*, pp. 288-293 (1999).
 3) Boettcher, S. and Percus, A.: Nature's way of optimizing, *Artificial Intelligence*, Vol. 119, No. 1-2, pp. 275-286 (2000).
 4) Bak, P. and Sneppen, K.: Punctuated Equilibrium and Criticality in a Simple Model of Evolution, *Physical Review Letters*, Vol. 71, pp. 4083-4086 (1993).
 5) Kitakami, H. and Mori, Y.: Reducing Crossovers in Reconciliation Graphs using the Coupling Cluster Exchange Method with a Genetic Algorithm, *Active Mining, IOS press*, Vol. 79, pp. 163-174 (2002).
 6) 札埜有香, 田村慶一, 森康真, 北上始: 遺伝的プログラミングによる調停グラフの交差数減少とその並列分散処理, DEWS2005, 電子情報通信学会データ工学研究専門委員会, Online Proceedings (2005).