

1 はじめに

生物系統学や文化系統学の学問において様々な生物、文化、製品などの系統を明らかにする研究が行われてきた。今日では対象の形質データなどからその対象群の姉妹関係を枝分かれした図（分岐図）として表現する分岐図推定法が主として使用されてきた。そこで太田らは対象の生年情報に着目して対象群の祖先-子孫関係を表現する系統樹を推定する手法を提案した [1]。しかし実際の対象の系統にはそれぞれ異なる特性があり、太田らの系統樹推定法が全ての系統に適用できるか明らかではない。そこで本研究では様々な対象集合に対して系統樹推定法を適用し評価をするために、異なる特性の系統関係をもつ対象集合を生成する系統発生モデルを構築する。

2 系統発生モデル

系統発生モデルはある対象が進化して新たな対象が生成される、もしくはある対象に影響を受けて新たな対象が生成されることを仮定して対象集合を生成するモデルである。本モデルは遺伝的アルゴリズムをベースにして、図 1 に示す流れで対象集合を生成しており、以下のことが考慮される。

初期集団の生成： N 個の個体を生成する。これらの個体は適応度属性とダミー属性をもつ。ダミー属性は、生存に影響を与えない形質として定められている。個体のすべての形質が生存に影響を与えるわけではないという考えに基づく。

適応度の評価～複製選択：適応度属性から適応度を計算する。良い対象ほど生存する可能性を高くし子孫を残しやすくする。

子個体生成：複製選択された対象は突然変異により属性を変化させ次世代に残る。

エリート選択：良い対象ほど記録に残りやすい可能性があることを表現している。

3 実験

系統発生モデルのパラメータを変更していくつかの対象集合を生成して、生成した対象集合に対して太田らの系統樹推定法を適用する。そして推定した系統樹と真の系統樹との一致度を計算する。図 2 は表 1 のパラメータ設定の系統発生モデルで作成した対象集合の系統樹の一例である。

3.1 太田らの系統樹推定法

太田らの系統樹推定法では対象集合の分岐図と、その対象集合の生年情報を基に、対象群のクラスタ関係と生年順について無矛盾な系統樹を推定する。

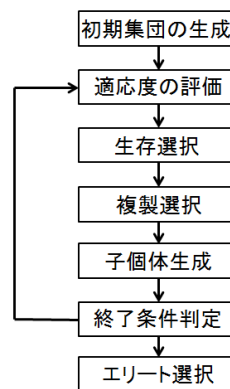


表 1: パラメータ初期値

パラメータ	値
適応度属性数 l	500
ダミー属性数 m	500
属性変化率 r_i	
初期個体数 N	100
適応度属性の分割数 f	10
生存選択率	0.1
複製選択率	0.5
終了世代数 g	30
エリート選択率	0

図 1: 系統発生モデル

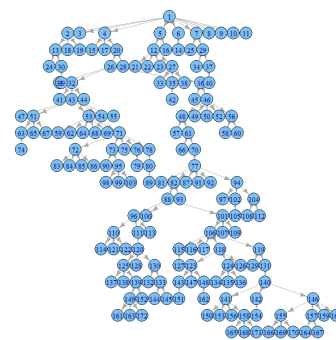


図 2: モデルにより生成した系統樹の一例

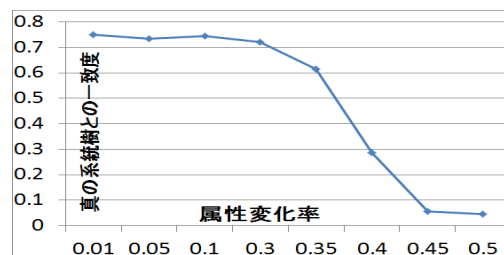


図 3: 属性変化率による系統樹の一致度

3.2 実験結果

図 3 に属性変化率による系統樹の一致度を示す。属性変化率が一定値を超えると系統樹の一致度が大幅に減少することから、親と子の属性の差が大きいと推定が難しいことがわかる。

4 まとめ

本研究では、さまざまな系統関係を作り出す系統発生モデルを構築した。いくつかの系統関係を作成し、それらに太田らの系統樹推定法を適用したときの客観的事実をある程度明らかにし考察した。今後は系統発生モデルの改善と、各種分岐図推定法での追加実験が挙げられる。

参考文献

- [1] 太田 章悟, 武藤 敦子, 犬塚 信博. "生年情報とクラスタ構造からの系統樹推定", 第 11 回情報学ワークショップ (WiNF2013) 論文集, 2013.