

# 遺伝的プログラミングによる特徴的な TTSP グラフパターン獲得手法の改良

## An Improvement of the Method for Acquiring Characteristic TTSP Graph Patterns by Genetic Programming

大塚喜明 鈴木祐介 宮原哲浩 内田智之  
Yoshiaki Otsuka Yusuke Suzuki Tetsuhiro Miyahara Tomoyuki Uchida

広島市立大学情報科学研究科  
Graduate School of Information Sciences, Hiroshima City University

Machine learning and data mining from graph structured data are studied intensively. TTSP (Two-Terminal Series Parallel) graphs are used as data models for electric networks and scheduling. We propose an improvement of the method for acquiring characteristic TTSP graph patterns from positive and negative data by using Genetic Programming.

### 1. はじめに

グラフ構造データからの機械学習やデータマイニングの研究が注目されている。TTSP グラフ (Two-Terminal Series Parallel graph) は、電気ネットワークやスケジューリングをコンピュータで扱う際にデータモデルとして用いられることが多い。TTSP グラフに構造的な変数の概念を導入したグラフパターンを **TTSP グラフパターン** または **TTSP 項グラフ** という。高味ら [Takami 09] は正事例から極小一般化 TTSP 項グラフを効率的に学習する手法を提案している。河野ら [Kono 11] は正事例から極大頻出 TTSP 項グラフを枚挙する手法を提案している。遺伝的プログラミング (Genetic Programming, GP) とは木構造やグラフ構造などの構造的表現を扱えるようにした進化的計算手法である [Poli 08]。長嶺ら [Nagamine 07] は、正事例と負事例の木構造データから、特徴的な木構造パターンを遺伝的プログラミングにより獲得する手法を提案している。有向グラフ構造を扱う進化的計算手法に遺伝的ネットワークプログラミング (Genetic Network Programming, GNP) がある [Katagiri 00]。先行研究である永井ら [Nagai 12] は TTSP グラフの正事例と負事例から特徴的な TTSP 項グラフを獲得する進化的手法を提案している。本研究は、永井らの提案した進化的手法の改良を行った。

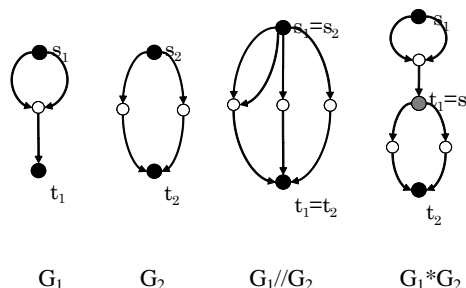


図 1 TTSP グラフ  $G_1, G_2, G_1//G_2, G_1*G_2$

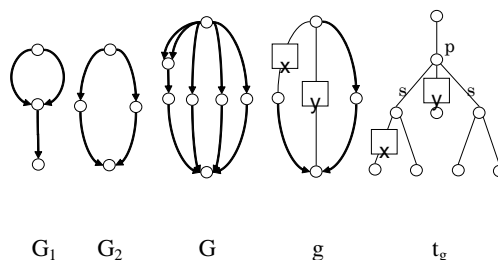


図 2 TTSP グラフ  $G_1, G_2, G, g$ , TTSP 項グラフ  $g$ , 構文木  $t_g$

で、 $g$  と  $G$  はマッチする。 $g$  の構文木は  $t_g$  である。構文木  $t_g$  の  $x$  と  $y$  は変数を表し、 $p$  は並列操作、 $s$  は直列操作を表す。

### 2. TTSP 項グラフ

次の(1),(2)で帰納的に定義される多重辺を許す有向グラフを **TTSP グラフ** という。(1)2つの頂点  $u, v$  と  $u$  から  $v$  への1つの辺から成る有向グラフは、ソースとして  $u$  を持ちシンクとして  $v$  を持つ TTSP グラフである。(2)  $G_1, G_2$  をソース  $s_1, s_2$  とシンク  $t_1, t_2$  をそれぞれ持つ TTSP グラフとする。このとき、 $s_1$  と  $s_2, t_1$  と  $t_2$  を同一視する操作を並列操作 (得られるグラフは  $G_1//G_2$  と表現する)、 $s_2$  と  $t_1$  を同一視する操作を直列操作 (得られるグラフは  $G_1*G_2$  と表現する) という。並列操作、直列操作のどちらかによって得られるグラフは TTSP グラフである。この操作を図 1 に示す。

TTSP グラフのいくつかの辺を変数にしたものを **TTSP 項グラフ** という。TTSP 項グラフ  $g$  と TTSP グラフ  $G$  に対して、 $g$  の変数を適切な TTSP グラフで置き換えることで  $G$  が得られるならば、 $g$  と  $G$  は **マッチする** という。TTSP グラフや TTSP 項グラフを定義する操作を木で表現したものを構文木という。例えば、図 2 において、TTSP 項グラフ  $g$  の変数  $x$  を TTSP グラフ  $G_1$  で、変数  $y$  を TTSP グラフ  $G_2$  で置き換えると TTSP グラフ  $G$  が得られるの

### 3. 進化的手法による特徴的な TTSP 項グラフの獲得

GP による TTSP 項グラフ発見問題を次のように定義する。

入力: 正事例と負事例からなる TTSP グラフの有限集合  $D$

問題: GP により、適合度の高い TTSP 項グラフ  $g$  を発見する

入力の TTSP グラフは構文木で与えられるものとし、GP の過程では TTSP 項グラフの構文木を個体として遺伝操作を行う。TTSP 項グラフ  $g$  の適合度  $fitness_D(g)$  を ( $g$  が  $D$  の正事例にマッチする割合 +  $g$  が  $D$  の負事例にマッチしない割合) / 2 と定義する。

TTSP 項グラフ発見問題に対する GP の手順を次のようにする [Nagai 12]。(1)  $D$  の正事例から、TTSP 項グラフで使用するキーワードの有限集合  $KW$  を求める。(2)  $KW$  に含まれるキーワードを辺ラベルとしてランダムに初期 TTSP 項グラフを生成する。(3) TTSP 項グラフの適合度  $fitness_D(g)$  を求める。(4) 適合度の大きさに比例した確率によって TTSP 項グラフの選択を行う。(5) 遺伝操作 (複製, 逆位, 交叉, 突然変異) により、次世代の集団を生成する。(6) 終了条件が満たされているときは終了する。そうで

連絡先: 宮原哲浩, 広島市立大学情報科学研究科, 〒731-3194 広島市安佐南区大塚東 3-4-1, miyares12@info.hiroshima-cu.ac.jp

なければ(3)へ戻る。遺伝操作の逆位(TTSP 項グラフの対象とする部分の順序を入れ替える操作), 交叉(2つの TTSP 項グラフの対象とする部分を入れ替える操作), 突然変異の add-TTSP(TTSP 項グラフにランダムに生成した TTSP 項グラフを加える操作)の適用例を図 3, 4, 5 に示す。

本研究では, 遺伝操作と TTSP 項グラフの構文木の標準形を求める計算を 2 段階に分けることで, 遺伝操作の複雑さを無くし, 遺伝操作の正当性の確認を行いやすくした。また, 遺伝操作における交叉オペレータを適用する際の部分 TTSP 項グラフの選択手法の拡張を行った。

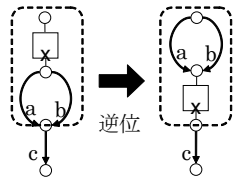


図 3 逆位の適用例

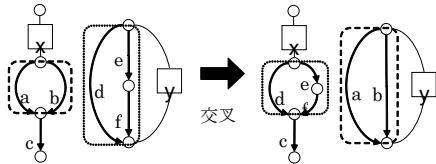


図 4 交叉の適用例

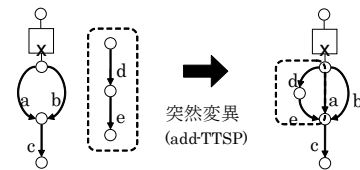


図 5 突然変異(add-TTSP)の適用例

#### 4. 実験結果

本研究の提案手法を実装し, 先行研究の手法と比較する評価実験を行った。実験に用いた GP のパラメータは以下の通りである。個体数:50, 複製確率:0.05, 逆位確率:0.05, 交叉確率:0.45, 突然変異確率:0.45, 最大世代数:200

ある TTSP 項グラフにマッチするもの 100 個を正事例, マッチしないもの 100 個を負事例とする。正事例と負事例にはそれぞれ 5% のノイズを入れた。提案手法と先行研究の手法で行ったそれぞれの 10 試行の平均値を表 1 と図 6 に示す。表 1 は実行時間と最終世代の適合度の最も高い個体(最良個体)のサイズと辺の数および割合の比較である。ここで例または個体のサイズとは構文木における葉の数である。つまり, TTSP グラフの場合は辺の総数であり, TTSP 項グラフの場合は辺(変数を除く)の総数+変数の総数である。図 6 は各世代の最良個体の適合度の平均値(1stfit), 各世代の集団の全ての個体の適合度の平均値(avefit)の比較である。

本研究の提案手法は先行研究と比較して, 表 1 より具体的な TTSP 項グラフを獲得することができたと考えられる。図 6 より, 先行研究よりも適合度の高い TTSP 項グラフを獲得することができたと見える。1stfit と avefit についての推移を比較すると, 1stfit と avefit 共に先行研究よりも高い値で推移していることが

わかる。また 1stfit については先行研究より早く収束していることがわかる。先行研究では初期個体のサイズの上限を小さくしていたが, 提案手法では初期個体のサイズをランダムな大ききで生成するようにした。これにより早い段階で適合度の高い個体が生成されたことが要因として考えられる。

表 1 実行時間と最終世代の最良個体のサイズ

	本研究	先行研究
実行時間(s)	612	699
個体のサイズ	8.6	7.3
個体の辺の数	4.1	2.4
個体の辺の割合(%)	47	32

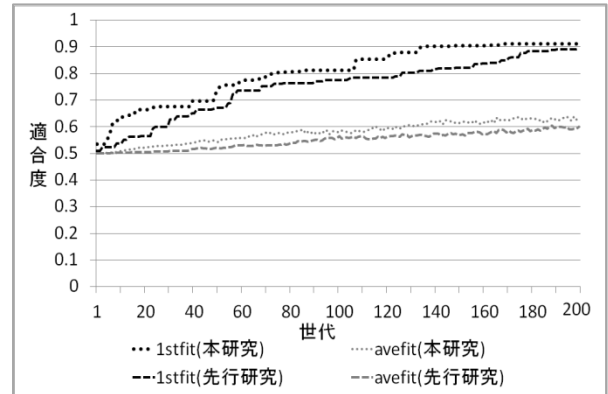


図 6 各世代の適合度の平均値の推移

さらにシステムを高速化するために, TTSP 項グラフのマッチングアルゴリズムの改良を行った。高味ら[Takami 09]は TTSP 項グラフのマッチング判定問題に対して, TTSP 項グラフを構文木に変換し, 構文木のマッチング判定問題を解くマッチングアルゴリズムを提案している。TTSP 項グラフの構文木では葉にあたる部分にのみ変数が存在する。そこで本研究では構文木の性質を用い, 根からの深さを考慮した頂点間の対応付けを調べることで, 構文木のマッチングをより高速に行えるよう改良を行った。特徴的 TTSP 項グラフ獲得システムの TTSP 項グラフマッチング判定プログラムを改良版に置き換えて実験をしたところ, 10 試行の実行時間の平均時間は置き換える前のシステムの約 10% となった。

#### 5. おわりに

本研究では, 先行研究で提案された TTSP グラフの正事例と負事例から特徴的な TTSP 項グラフを獲得する進化的手法を改良した。今後の課題として, 実データを入力することが考えられる。

#### 参考文献

- [Katagiri 00] H.Katagiri et al., Genetic Network Programming - Application to Intelligent Agents, Proc. IEEE Int. Conf. Systems, Man, and Cybernetics, pp.3829-3834, 2000
- [Kono 11] T.Kono et al., Enumerating Maximally Frequent TTSP Graph Patterns, Proc. 7th LLLL Workshop, pp.43-50, 2011
- [Nagai 12] S.Nagai et al., Acquisition of Characteristic TTSP Graph Patterns by Genetic Programming, Proc. 2012 IIAI International Conference on Advanced Applied Informatics, pp.340-344, 2012
- [Nagamine 07] M.Nagamine et al., A Genetic Programming Approach to Extraction of Glycan Motifs Using Tree Structured Patterns, Springer-Verlag LNAI 4830, pp.150-159, 2007
- [Poli 08] R.Poli et al., A Field Guide to Genetic Programming, Lulu Press, 2008
- [Takami 09] R.Takami et al., Polynomial Time Inductive Inference of TTSP Graph Languages from Positive Data, IEICE Transactions, E92-D(2), pp.181-190, 2009