

グラフパターンの木表現を用いた遺伝的プログラミングによる 特徴的なブロック保存型外平面的グラフパターンの獲得 Acquisition of Characteristic Block Preserving Outerplanar Graph Patterns using Genetic Programming and Tree Representation of Graph Patterns

大内山 祐斗^{*1} 宮原 哲浩^{*1} 鈴木 祐介^{*1} 内田 智之^{*1} 久保山 哲二^{*2} 徳原 史也^{*1}
Yuto Ouchiya ^{*} Tetsuhiro Miyahara Yusuke Suzuki Tomoyuki Uchida Tetsuji Kuboyama Fumiya Tokuhara

^{*1} 広島市立大学情報科学部

Faculty of Information Sciences, Hiroshima City University

^{*2} 学習院大学計算機センター

Computer Centre, Gakushuin University

Machine learning and data mining from graph structured data are studied intensively. We use block preserving outerplanar graph patterns having structured variables for expressing structural features of outerplanar graphs. We propose a method for acquiring characteristic block preserving outerplanar graph patterns from positive and negative outerplanar graph data, by using Genetic Programming and tree representation of graph patterns. We report experimental results on applying our method to synthetic outerplanar graph data.

1. はじめに

グラフ構造を持つデータに関する機械学習やデータマイニングが注目されている。例えば、多くの化合物は外平面的グラフの構造を持っており、化合物データからグラフ構造的特徴を獲得する研究が知られている[Horvath 10]。

遺伝的プログラミング (Genetic Programming, GP)[Koza 92, Banzhaf 98]とは、遺伝的アルゴリズムの遺伝子型を木構造のような構造的表現を扱えるように拡張した進化的手法である。外平面的グラフのグラフ構造的特徴を表現するブロック保存型外平面的グラフパターン (Block Preserving Outerplanar graph pattern, BPO グラフパターン) が提案されている[Sasaki 07, Yamasaki 09]。正事例と負事例の外平面的グラフの集合から特徴的な BPO グラフパターンを獲得する遺伝的プログラミングによる手法[大内山 13]を実装し、人工データを用いた評価実験を行ってその手法の有効性を確認したので[Ouchiya ^{*} 15]、本稿で報告する。

先行研究[Sasaki 07, Yamasaki 09]では、BPO グラフパターンにより定義されるグラフ言語のクラスが正事例から多項式時間帰納推論可能であることが示され、正事例からの頻出 BPO グラフパターン列挙アルゴリズムが提案されている。関連研究として、正事例と負事例からなる木構造データから特徴的な木パターンを獲得する遺伝的プログラミングによる手法[Nagamine 07, Nakai 13, Miyahara 14]、正事例と負事例からなるグラフ構造データから特徴的なグラフパターンを獲得する遺伝的プログラミングによる手法[Nagai 12]、及びグラフ構造データに対する進化的手法[Katagiri 00, Shirakawa 07]がある。

2. 準備

ラベル付きグラフとは、頂点と辺がラベル付けされたグラフをいう。すべての頂点が外平面に接するように平面埋め込み可能であるグラフを外平面的グラフとよぶ。本稿では、連結なラベル付き外平面的グラフのみを扱うため、以後連結なラベル付き外平面的グラフを単に外平面的グラフということにする。連結グラフ G において、削除することで非連結にできる頂点と辺をそれぞれ G のカット点とブリッジという。本稿では、頂点数 3 以上のカッ

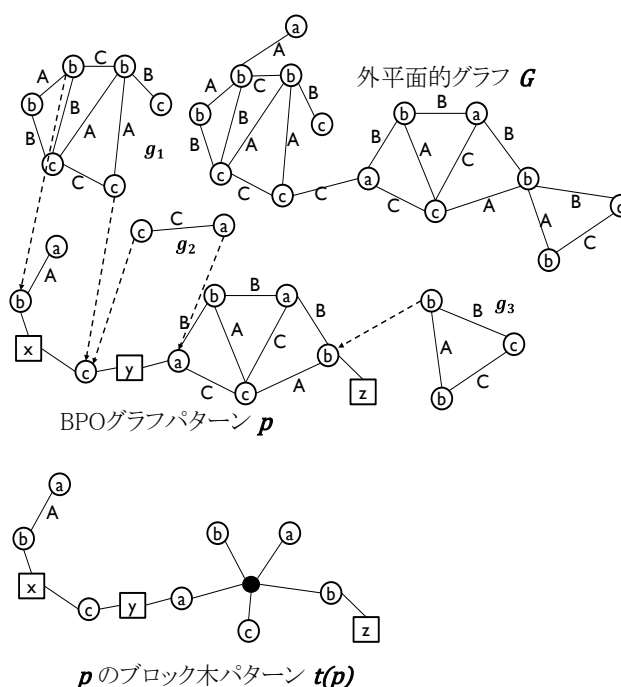


図 1: BPO グラフパターン p とマッチする外平面的グラフ G .
 p のブロック木パターン $t(p)$.

ト点をもたない極大な部分グラフ (2 重連結成分) を G のブロックという。外平面的グラフのいくつかのブリッジを 2 頂点からなる変数 (ブリッジ変数) で置き換え、1 頂点からなる変数 (末端変数) を頂点に付加して得られるグラフパターンを **BPO グラフパターン** とよぶ。BPO グラフパターンの例を図 1 に示す。外平面的グラフ G と BPO グラフパターン p に対し、 p のすべての変数を適当な外平面的グラフで置き換えることによって G と同型な外平面グラフが得られるとき、 p と G は **マッチする** という。図 1 において、BPO グラフパターン p と外平面的グラフ G はマッチする。なぜならば、 p の変数 x, y, z を、それぞれ外平面的グラフ g_1, g_2, g_3 と点線で対応付けを示している頂点どうしを同一視して置き換えることで G と同型な外平面グラフを得ることができるからである。

連絡先: 宮原哲浩, 広島市立大学情報科学部, 〒731-3194

広島市安佐南区大塚東 3-4-1, miyares16@info.hiroshima-cu.ac.jp

BPO グラフパターン p の各ブロック B を次の操作(1)-(3)を適用して置き換えることにより得られるグラフパターンを p の**ブロック木パターン**[Sasaki 07, Yamasaki 09]とよび、 $t(p)$ と表す。特に、変数を持たないブロック木パターンを単にブロック木とよぶ。(1)ブロック B 内の辺の接続情報をラベル(ブロックラベル)としてもつ新しい頂点(ブロック頂点) v_B を追加する。(2) B 中のすべての辺を削除する。(3) B 中の各頂点と頂点 v_B とを辺で結ぶ。図1にブロック木パターンの例を示す。

3. 遺伝的プログラミングによる特徴的な BPO グラフパターンの獲得

3.1 特徴的な BPO グラフパターン獲得問題と手法

本節で、本研究で扱う特徴的な BPO グラフパターン獲得問題を定式化し、さらにその問題を解く遺伝的プログラミングによる手法を与える。

特徴的な BPO グラフパターン獲得問題

入力: 正事例, 及び負事例からなる外平面的グラフの有限集合 D .

問題: D に関する適合度の高い BPO グラフパターンを獲得せよ.

BPO グラフパターン p の D に関する適合度 $fitness_D(p)$ を、

$$fitness_D(p) = (D \text{の正事例に} p \text{がマッチする割合} + D \text{の負事例に} p \text{がマッチしない割合})/2$$
と定義する。特徴的な BPO グラフパターン獲得問題を解く遺伝的プログラミングによる手法を以下に与える。

遺伝的プログラミング(GP)による BPO グラフパターン獲得手法

1. s をユーザが与える集団サイズとする。 D 中の正事例を用いて、ランダムに生成された s 個の BPO グラフパターン p_1, \dots, p_s の集団 $P = \{p_1, \dots, p_s\}$ を生成する。
2. P 中の各 BPO グラフパターン p_i ($i = 1, \dots, s$)に対して、その適合度 $fitness_D(p_i)$ を求める。
3. 次世代の BPO グラフパターンを s 個生成するまで、以下の手順を繰り返して、次世代の BPO グラフパターンの集団 P' を作成する。
 適合度の大きさによって集団 P から選ばれた BPO グラフパターンに遺伝操作(交叉・突然変異・複製)を適用し新しい BPO グラフパターンを生成する。
4. P' 中の各 BPO グラフパターン p'_i ($i = 1, \dots, s$)に対して、その適合度 $fitness_D(p'_i)$ を求める。
5. 終了条件である最大世代数まで到達したら終了する。そうでなければ $P = P'$ とし 3 に戻る。

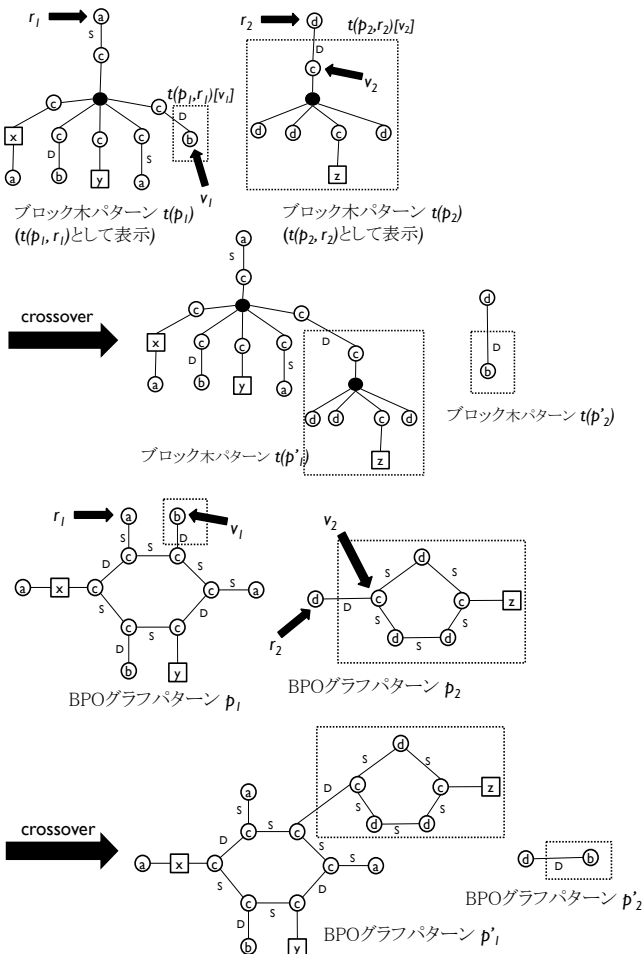


図 2: ブロック木パターン $t(p_1), t(p_2)$ と対応する BPO グラフパターン p_1, p_2 に対する交叉オペレータの適用。

3.2 ブロック木パターンに対する遺伝操作を行う GP オペレータ

BPO グラフパターン p とそのブロック木パターン $t(p)$ は 1 対 1 に対応するため、GP による BPO グラフパターン獲得手法における、BPO グラフパターン p に対する GP オペレータの代わりに、 p に対応するブロック木パターン $t(p)$ に対する GP オペレータを与える。ブロック木パターン $t(p)$ は根なし無順序木パターンであるので、GP オペレータでは最初に $t(p)$ で仮の根 r をランダムに選択し、根が r である無順序木パターンである $t(p, r)$ に変換し遺伝操作を行う。GP オペレータは、ブロック木パターンのブロックラベルとそのブロック頂点が表すブロック情報に変化を加えないように設計した。なお、 $t(p, r)[v]$ は、 $t(p, r)$ における頂点 v とそのすべての子孫と v から根へ向かう辺からなる部分木パターンを表す。

(1)交叉(crossover)オペレータ

ブロック木パターン $t(p_1), t(p_2)$ に対して、次の操作を行う。

1. $t(p_1), t(p_2)$ に対して、ランダムにそれぞれ仮の根 r_1, r_2 を選ぶ。
2. $t(p_1, r_1), t(p_2, r_2)$ に対して、ランダムにそれぞれ頂点 v_1, v_2 を選ぶ。
3. v_1, v_2 が根、ブロック頂点、もしくはブロック頂点の子であれば選びなおす。
4. $t(p_1, r_1)$ の $t(p_1, r_1)[v_1]$ と $t(p_2, r_2)$ の $t(p_2, r_2)[v_2]$ を交換して、根ありブロック木パターン $t(p'_1, r_1)$ と $t(p'_2, r_2)$ を生成する。
5. ブロック木パターン $t(p'_1)$ と $t(p'_2)$ を返す。

図 2 に、ブロック木パターン $t(p_1), t(p_2)$ に対する交叉オペレータの適用例を与える。なお、ブロック木パターン $t(p_1), t(p_2)$ に対応する BPO グラフパターン p_1, p_2 に対する交叉オペレータの適用例も与える。

(2)突然変異オペレータ

以下の 5 つの GP オペレータ(部分パターン交換, 部分パターン追加, 部分パターン削除, 頂点ラベル交換, 辺ラベル交換)を確率的に適用する。

(a)部分パターン交換(change-subpattern) オペレータ

ブロック木パターン $t(p)$ に対して, 次の操作を行う。

1. $t(p)$ に対して, ランダムに仮の根 r を選ぶ。
2. $t(p, r)$ に対して, ランダムに頂点 v を選ぶ。
3. v が根 r , ブロック頂点, もしくはブロック頂点の子であれば選びなおす。
4. 新しい根ありブロック木パターン t_0 を生成する。
5. $t(p, r)$ の $t(p, r)[v]$ と t_0 を交換して, 根ありブロック木パターン $t(p', r)$ を生成する。
6. ブロック木パターン $t(p')$ を返す。

(b)部分パターン追加(add-subpattern)オペレータ

ブロック木パターン $t(p)$ に対して, 次の操作を行う。

1. $t(p)$ に対して, ランダムに仮の根 r を選ぶ。
2. $t(p, r)$ に対して, ランダムに頂点 v を選ぶ。
3. v がブロック頂点であれば選びなおす。
4. 新しい根ありブロック木パターン t_0 を生成する。
5. t_0 の根が v の子になるように t_0 を $t(p, r)$ に付加して, 根ありブロック木パターン $t(p', r)$ を生成する。
6. ブロック木パターン $t(p')$ を返す。

(c)部分パターン削除(del-subpattern)オペレータ

ブロック木パターン $t(p)$ に対して, 次の操作を行う。

1. $t(p)$ に対して, ランダムに仮の根 r を選ぶ。
2. $t(p, r)$ に対して, ランダムに頂点 v を選ぶ。
3. v が根 r , ブロック頂点, もしくはブロック頂点の子であれば選びなおす。
4. $t(p, r)$ から $t(p, r)[v]$ を削除して, 根ありブロック木パターン $t(p', r)$ を生成する。
5. ブロック木パターン $t(p')$ を返す。

(d)頂点ラベル交換(change-vertex-label) オペレータ

ブロック木パターン $t(p)$ のランダムに選ばれた頂点 v の頂点ラベルをランダムに選ばれた他の頂点ラベルに変更したブロック木パターンを返す。なお, v がブロック頂点, もしくはブロック頂点に隣接する頂点であれば選びなおす。

(e)辺ラベル交換(change-edge-label) オペレータ

ブロック木パターン $t(p)$ に対して, ランダムに仮の根 r を選ぶ。 $t(p, r)$ の頂点 v をランダムに選ぶ。頂点 v が根 r , ブロック頂点, もしくはブロック頂点の子であれば選びなおす。 $t(p, r)$ において v とその親の間に辺があれば, その辺を他の辺ラベルをもつ辺またはブリッジ変数に置き換える。 v とその親の間にブリッジ変数があれば, そのブリッジ変数を辺に置き換える。このようにして, 根ありブロック木パターン $t(p', r)$ を生成する。ブロック木パターン $t(p')$ を返す。

4. 実験

BPO グラフパターンの適合度の計算に, BPO グラフパターンマッチングアルゴリズム[Sasaki 07, Yamasaki 09]を用いて提案手法を実装した。設定した BPO グラフパターン π について, 人工的に生成した外平面グラフのうち π にマッチするものを正事例,

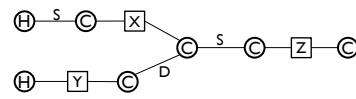
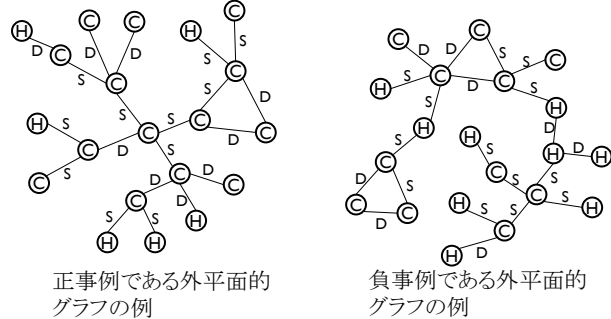
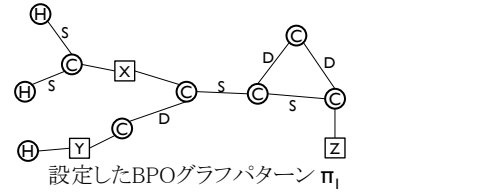


図 3: 実験 1 で設定した BPO グラフパターン π_1 , 正事例と負事例の例, 獲得した BPO グラフパターン。

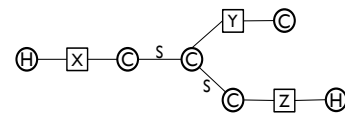
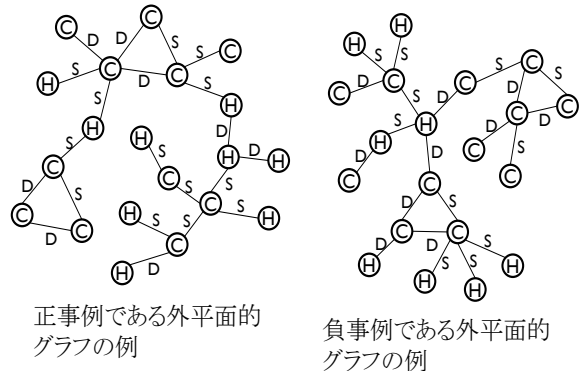
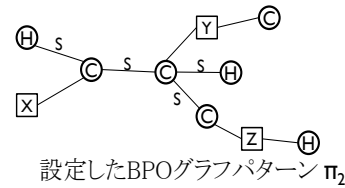


図 4: 実験 2 で設定した BPO グラフパターン π_2 , 正事例と負事例の例, 獲得した BPO グラフパターン。

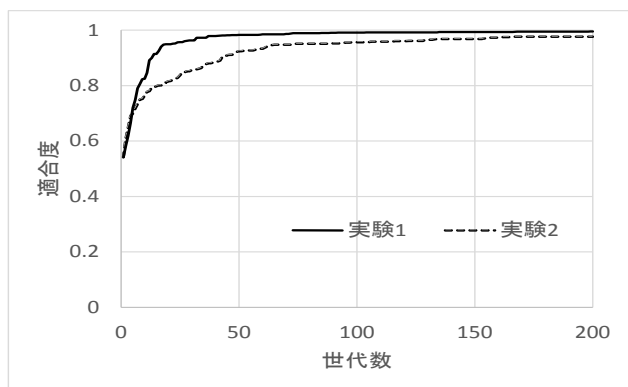


図 5: 実験 1, 2 における最良個体の適合度の推移。

マッチしないものを負事例として、100 個の正事例と 100 個の負事例からなる集合 $D(\pi)$ を作成した。この人工データ $D(\pi)$ を用いて提案手法の評価実験を行った。実験 1 と実験 2 では、それぞれ図 3 と図 4 に示す BPO グラフパターン π_1 と π_2 に関する人工データ $D(\pi_1)$ と $D(\pi_2)$ を用いた。各実験で、それぞれ 20 試行の特徴的 BPO グラフパターン獲得を行った。

本実験における GP のパラメータは以下のとおりである。集団サイズ:50, 複製確率:0.05, 交叉確率:0.50, 突然変異確率:0.45 (部分パターン交換:0.40, 部分パターン追加:0.10, 部分パターン削除:0.40, 頂点ラベル交換:0.05, 辺ラベル交換:0.05), 選択方法:ルーレット選択, トーナメント選択(サイズ 2), エリート保存(サイズ 3), 最大世代数:200。

実験 1 における 20 試行の最終世代の最良個体の適合度の平均は 0.996 であり、実行時間の平均は 1787 秒である。獲得した最終世代の最良個体である BPO グラフパターンを図 3 に示す。実験 2 における 20 試行の最終世代の最良個体の適合度の平均は 0.977 であり、実行時間の平均は 1596 秒である。獲得した最終世代の最良個体である BPO グラフパターンを図 4 に示す。実験 1, 2 における各世代の最良個体の適合度の 20 試行の平均値の推移を図 5 に示す。どちらの実験においても入力データを生成するために設定した BPO グラフパターンに近い、特徴的な BPO グラフパターンを獲得することができた。

5. おわりに

本研究では、正事例と負事例の外平面的グラフから遺伝的プログラミングによる特徴的な BPO グラフパターンを獲得する手法を実装し、人工データにより提案手法の有効性を確認した。本研究の発展として、正事例と負事例に化合物の実データを用いて、正事例のラベルの情報を利用した特徴的な BPO グラフパターンを獲得する進化的手法を実現している[徳原 16]。提案手法を大規模なグラフ構造データに適用して有効性を確認することなどが今後の課題である。

参考文献

[Banzhaf 98] W.Banzhaf et al., Genetic Programming: An Introduction: On the Automatic Evolution of Computer Programs and Its Applications, Morgan Kaufmann,1998.
 [Horvath 10] T.Horvath et al., Frequent subgraph mining in outerplanar graphs, Data Mining and Knowledge Discovery, vol.21,pp.472-508,2010.

[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.
 [Koza 92] J.R. Koza, Genetic Programming: On the Programming of Computers by Means of Natural Selection, MIT Press, 1992.
 [Miyahara 14] T.Miyahara and T.Kuboyama, Learning of glycan motifs using genetic programming and various fitness functions, Journal of Advanced Computational Intelligence and Intelligent Informatics, vol.18(3),pp.401-408,2014.
 [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, Proc. AI 2007, Springer-Verlag LNAI 4830, pp.150-159, 2007.
 [Nakai 13] S.Nakai et al., Acquisition of characteristic tree patterns with VLDC's by genetic programming and edit distance, Proc. 2013 IIAI International Conference on Advanced Applied Informatics, pp. 113 - 118, 2013.
 [大内山 13] 大内山祐斗ほか, 遺伝的プログラミングによる特徴的なブロック保存型外平面的グラフパターンの獲得, 火の国情報シンポジウム 2013 論文集, A-4-3,2013.
 [Ouchiyama 15] Y.Ouchiyama et al., Acquisition of characteristic block preserving outerplanar graph patterns from positive and negative data using genetic programming and tree representation of graph patterns, Proc. 2015 IEEE 8th International Workshop on Computational Intelligence and Applications, pp. 95-101, 2015.
 [Sasaki 07] Y.Sasaki et al., Mining of frequent block preserving outerplanar graph structured patterns, Proc. ILP-2007, Springer-Verlag LNAI 4894,pp.239-253,2007.
 [Shirakawa 07] S.Shirakawa et al., Graph structured program evolution, Proc. Genetic and Evolutionary Computation Conference 2007,pp. 1686-1693,2007.
 [徳原 16] 徳原 史也ほか, ラベルの情報を利用した遺伝的プログラミングによる特徴的なブロック保存型外平面的グラフパターンの獲得, 火の国情報シンポジウム 2016 論文集, 6C-3,2016.
 [Yamasaki 09] H.Yamasaki et al., Learning block-preserving graph patterns and its application to data mining, Machine Learning, vol.76,pp.137-173,2009.