The 33rd Annual Conference of the Japanese Society for Artificial Intelligence, 2019

# 遺伝的プログラミングによるワイルドカード付きブロック保存型外平面的グラフパターンの獲得

Acquisition of Block Preserving Outerplanar Graph Patterns with Wildcards by Genetic Programming

沖永 志帆 *1 Shiho Okinaga　徳原 史也 *1 Fumiya Tokuhara　宮原 哲浩 *1 Tetsuhiro Miyahara　鈴木 祐介 *1 Yusuke Suzuki　久保山 哲二 *2 Tetsuji Kuboyama　内田 智之 *1 Tomoyuki Uchida

*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. Many chemical compounds can be expressed by outerplanar graphs. The purpose of this paper is to propose a learning method for obtaining characteristic graph patterns from positive and negative outerplanar graph data. We propose an evolutionary learning method for acquiring characteristic block preserving outerplanar graph patterns with wildcards from positive and negative outerplanar graph data. Our evolutionary learning method is based on Genetic Programming applied to tree representations of block preserving outerplanar graph patterns with wildcards. We report preliminary experimental results on our evolutionary learning method.

## 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おり，正負の分類ラベルを持つグラフデータを分類するグラフ分類問題 [Rehman 12] が知られている．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].

我々は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）[Sasaki 08, Yamasaki 09] を構造的特徴表現として，外平面的グラフに対するグラフ分類問題に取り組み，正事例と負事例の外平面的グラフから，ブロック木パターンに対する遺伝的プログラミングを用いて特徴的な 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 [Ouchiyama 15, Tokuhara 16a, Tokuhara 16b]．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．ブロック木パターン [Sasaki 08, Yamasaki 09] とは BPO グラフパターンの木表現である．

本稿では，分類精度向上のため，ワイルドカード付き BPO グラフパターンを考案し，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を用いて特徴的なワイルドカード付き BPO グラフパターンを獲得する進化的手法を提案する．ワイルドカード付き BPO グラフパターンとは，頂点ラベルのワイルドカード（任意の頂点ラベルとマッチする）と辺ラベルのワイルドカード（任意の辺ラベルとマッチする）を持つように表現を拡張した BPO グラフパターンである．

関連研究として，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 [Sasaki 08, Yamasaki 09]．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．

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，miyares19@hiroshima-cu.ac.jp

## 2. ワイルドカード付き BPO グラフパターン

BPO グラフパターンとブロック木パターン [Sasaki 08, Yamasaki 09] の説明をする [大内山 16, Tokuhara 16a, Tokuhara 16b, 徳原 17, Tokuhara 17, 徳原 18].

$G$ をグラフ，$\Lambda$ と $\Delta$ をアルファベットとする．ラベル付きグラフとは，頂点集合 $V(G)$ と辺集合 $E(G)$ の各要素が，それぞれ $\Lambda$ と $\Delta$ の要素によってラベル付けされたグラフをいう．ラベル付きグラフのすべての頂点が外平面に接するように平面埋め込みが可能であるとき，そのグラフを外平面的グラフとよぶ．連結グラフにおいて，削除することでグラフを非連結にすることができる頂点をカット点という．外平面的グラフのブロックとは，頂点数 3 以上のカット点をもたない極大な二重連結成分をいう．ブロックに属さない辺をブリッジとよぶ．

**BPO グラフパターン**とは，ブリッジ変数および末端変数とよばれる 2 種類の構造的変数を持つ連結な外平面的グラフの構造をしたグラフパターンである．本稿では，連結な外平面的グラフのみを扱う．以後，連結な外平面的グラフを単に外平面的グラフとよぶ．“?”をワイルドカードを表す記号とする．$\Lambda$ と $\Delta$ が “?”を含むときの BPO グラフパターンを**ワイルドカード付き BPO グラフパターン**とよぶ．$\Lambda$ と $\Delta$ が “?”を含まないときの BPO グラフパターンを**ワイルドカードを持たない BPO グラフパターン**とよぶ．BPO グラフパターンのブリッジ変数は頂点数 2 以上の外平面的グラフと置き換えることができ，末端変数は頂点数 1 以上の外平面的グラフと置き換えることができる．BPO グラフパターン $p$ と外平面的グラフ $G$ に対し，$p$ の変数を適当な外平面的グラフで置き換えて，$p$ のワイルドカードを適当な辺ラベルまたは頂点ラベルで置き換えることによって $G$ が得られるとき，$p$ は $G$ と**マッチする**という．図 1 に BPO グラフパターンと外平面的グラフのマッチ関係の例を示す．ワイルドカード付き BPO グラフパターン $p_1$ は外平面的グラフ $G_1$ と外平面的グラフ $G_2$ とマッチするが，ワイルドカードを持たない BPO グラフパターン $p_2$ は外平面的グラフ $G_2$ としかマッチしない．

BPO グラフパターン $p$ のブロック部分を，ブロックの辺ラ

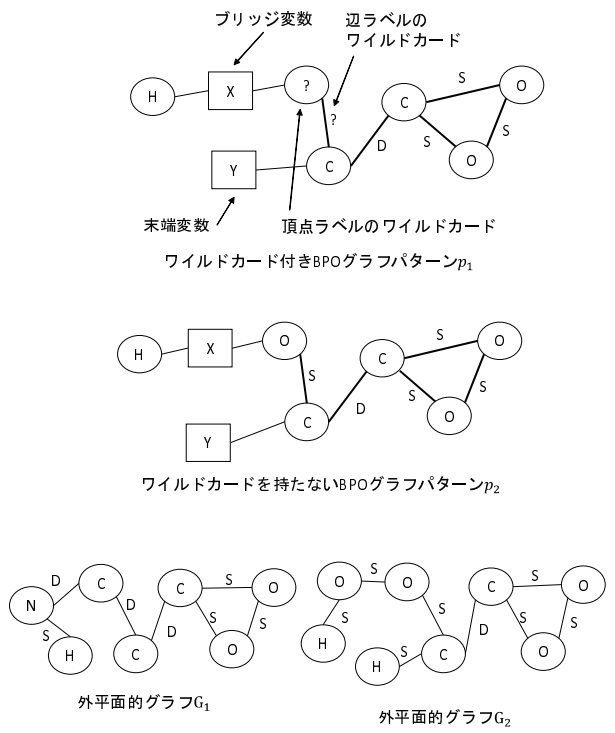


図 1: BPO グラフパターンと外平面的グラフのマッチ関係

ベルの情報を保持したブロック頂点へと置き換えることで得られる，根なし無順序木の構造を持つグラフパターンを $p$ のブロック木パターンとよび，$t(p)$ で表す．ブロック木パターン $t(p)$ は，BPO グラフパターン $p$ の無順序木表現である．

## 3. 遺伝的プログラミングによるワイルドカード付き BPO グラフパターンの獲得

本研究では，以下に述べる特徴的な BPO グラフパターン獲得問題に取り組む．

**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する．

ここで，BPO グラフパターン $p$ の $D$ に関する適合度 $fitness_D(p)$ は，$fitness_D(p) = (p$ が $D$ の正事例にマッチする割合 $+\ p$ が $D$ の負事例にマッチしない割合 $)/2$ で定義する．我々は特徴的な BPO グラフパターン獲得問題を解く，ワイルドカードを持たない BPO グラフパターンを個体とする遺伝的プログラミングによる手法 [Ouchiyama 15, Tokuhara 16a, Tokuhara 16b] を与えている．ワイルドカードを持たない BPO グラフパターンに対する従来手法を拡張して，特徴的な BPO グラフパターン獲得問題を解く，ワイルドカード付き BPO グラフパターンを個体とする遺伝的プログラミングによる手法を実現したので本稿で報告する．遺伝的プログラミングの遺伝操作は BPO グラフパターン $p$ の木表現であるブロック木パターン $t(p)$ に適用する．

ワイルドカードを持たない BPO グラフパターンに対する従来手法における遺伝操作には，交叉・突然変異・複製の 3 種類がある．突然変異には部分木変換 (change-subpattern)・部分木追加 (add-subpattern)・部分木削除 (del-subpattern)・頂点ラベル変更 (change-vertex-label)・辺ラベル変更 (change-edge-label) の 5 つの操作がある．提案手法では，突然変異の遺伝操作のうちで，部分木追加・頂点ラベル変更・辺ラベル変更の遺伝操作を適用する際に，確率的に頂点ラベルまたは辺ラベルをワイルドカードに置き換えて，ワイルドカード付き BPO グラフパターンを個体として生成する．このようにして，特徴的な BPO グラフパターン獲得問題を解く，ワイルドカード付き BPO グラフパターンを個体とする遺伝的プログラミングによる手法を実現した．頂点ラベルまたは辺ラベルをワイルドカードに置き換える確率を，突然変異ワイルドカード設定率とよぶ．提案手法における，ワイルドカード付き BPO グラフパターンとそれに対応するブロック木パターンに対する交叉の適用例を図 2 に示す．図中のブロック木パターンの黒丸はブロック頂点を表す．

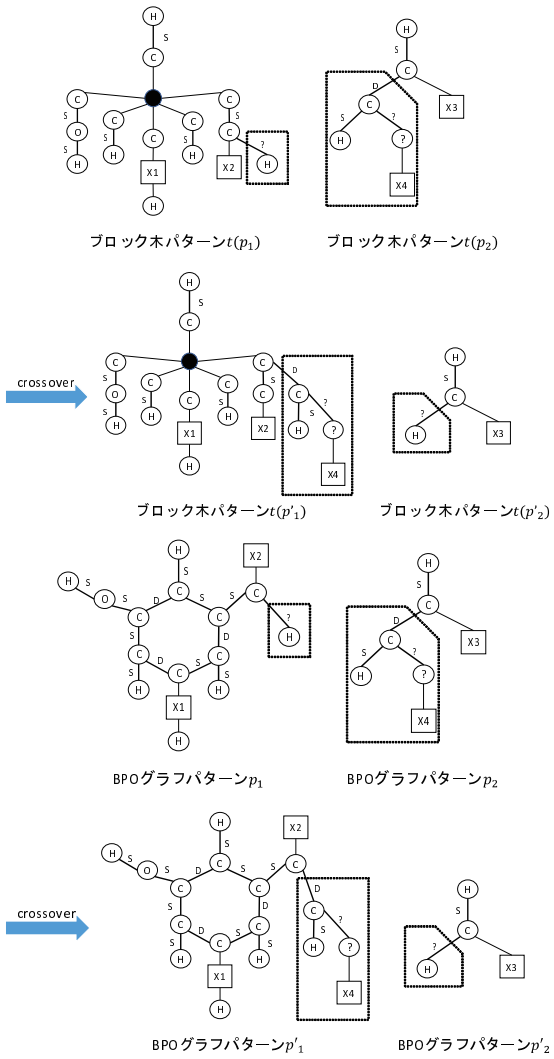


図 2: ワイルドカード付き BPO グラフパターンとそれに対応するブロック木パターンに対する交叉の適用

## 4. 実験

遺伝的プログラミングを用いて特徴的なワイルドカード付き BPO グラフパターンを獲得する提案手法を，Intel Xeon CPU W-2133 3.60GHz プロセッサ，実装メモリ 32.0GB の Windows10 Pro 64bit OS 上に Java 言語で実装した．比較のため，遺伝的プログラミングを用いて特徴的なワイルドカードを持たない BPO グラフパターンを獲得する従来手法も同じ環境で実装した．

提案手法と従来手法に共通するパラメータ設定は以下のとおりである．遺伝的プログラミングのパラメータを次のように設定した．集団サイズ:50，エリートサイズ:3，最大世代数:200，トーナメントサイズ:2，複製確率:0.05，交叉確率:0.50，突然変異確率:0.45 (change-subpattern:0.30, add-subpattern:0.30, del-subpattern:0.30, change-vertex-label:0.05, change-edge-label:0.05)．正事例のラベル情報は利用しない．提案手法，従来手法を用いて，実データまたは人工データから，特徴的な BPO グラフパターンを獲得する 10 試行の実験を行った．

実験 1 では次の実データを用いた．NCI データベース [NCI 99] のファイル “CAD2DA99.sdz”から CA(Confirmed Active) または CM(Confirmed Moderate) とラベル付けされた外平面的グラフとみなせるデータ 434 個を正事例とし，CI(Confirmed Inactive) とラベル付けされた外平面的グラフとみなせるデータから 434 個のデータを選んで負事例とした．実験 2 では次の人工データを用いた．ワイルドカードを持たない BPO グラフパターンを 3 つ設定し，人工的に生成した外平面的グラフのうちで，設定した BPO グラフパターンの少なくとも 1 つとマッチするものを正事例とし，そうではないものを負事例として，500 個の正事例と 500 個の負事例からなる外平面的グラフの集合を作成した．

表 1 に，実験 1(実データから特徴的 BPO グラフパターンを獲得)，実験 2(人工データから特徴的 BPO グラフパターンを獲得) の結果を示す．提案手法で獲得するのは特徴的なワイルドカード付き BPO グラフパターンであり，従来手法で獲得するのはワイルドカードを持たない BPO グラフパターンである．提案手法のワイルドカード設定率とは，突然変異ワイルドカード設定率のことである．最終世代の最良個体の 10 試行平均適合度，最終世代の最良個体の 10 試行平均頂点数，最終世代の全個体の平均頂点数の 10 試行平均値，10 試行平均実行時間を示している．ここで，個体（BPO グラフパターン）の頂点数は，対応するブロック木パターンの頂点数とする．頂点数の大きい BPO グラフパターンは，具体的な BPO グラフパターンであると考えられる．表 1 の実験結果より，提案手法は従来手法よりも，適合度が大きくてより具体的な特徴的 BPO グラフパターンを獲得したといえる．

## 5. おわりに

本稿では，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を用いて特徴的なワイルドカード付き BPO グラフパターンを獲得する進化的手法を提案した．実データ，人工データに適用した結果，提案手法は遺伝的プログラミングを用いて特徴的なワイルドカードを持たない BPO グラフパターンを獲得する従来手法よりも，適合度が大きくてより具体的な特徴的 BPO グラフパターンを獲得することができた．今後の課題として，獲得した BPO グラフパターンの適合度の向上などが挙げられる．

表 1: 最良個体の平均適合度と平均頂点数，最終世代の個体の平均頂点数，平均実行時間

| 実験 1(実データから特徴的 BPO グラフパターンを獲得) | 提案手法 (ワイルドカード設定率:0.1) | 従来手法 |
|---|---|---|
| 最良個体の平均適合度 | 0.598 | 0.593 |
| 最良個体の平均頂点数 | 2.1 | 1.5 |
| 最終世代の個体の平均頂点数 | 3.0 | 2.8 |
| 平均実行時間 (秒) | 188.3 | 179.5 |
| **実験 2(人工データから特徴的 BPO グラフパターンを獲得)** | **提案手法 (ワイルドカード設定率: 0.2)** | **従来手法** |
| 最良個体の平均適合度 | 0.712 | 0.687 |
| 最良個体の平均頂点数 | 4.8 | 4.5 |
| 最終世代の個体の平均頂点数 | 5.3 | 5.0 |
| 平均実行時間 (秒) | 4169.0 | 4262.7 |

### 謝辞

本研究は JSPS 科研費 JP15K00312, JP15K00313, JP17K00321 の助成を受けたものです．

## 参考文献

[Banzhaf 98] Banzhaf, W., Nordin, P., Keller, R., and Francone, F.: *Genetic Programming: An Introduction : On the Automatic Evolution of Computer Programs and Its Applications*, Morgan Kaufmann (1998)

[Horvath 10] Horvath, T., Ramon, J., and Wrobel, S.: Frequent Subgraph Mining in Outerplanar Graphs, *Data Mining and Knowledge Discovery*, Vol. 21, pp. 472–508 (2010)

[Katagiri 00] Katagiri, H., Hirasawa, K., and Hu, J.: Genetic Network Programming - Application to Intelligent Agents, *Proc. IEEE Int. Conf. Systems, Man, and Cybernetics*, pp. 3829–3834 (2000)

[Koza 92] Koza, J. R.: *Genetic Programming: On the Programming of Computers by Means of Natural Selection*, MIT Press (1992)

[Miyahara 14] Miyahara, T. and Kuboyama, T.: Learning of Glycan Motifs Using Genetic Programming and Various Fitness Functions, *Journal of Advanced Computational Intelligence and Intelligent Informatics (JACIII)*, Vol. vol.18(3), pp. 401–408 (2014)

[Nagai 12] Nagai, S., Miyahara, T., Suzuki, Y., and Uchida, T.: Acquisition of Characteristic TTSP Graph Patterns by Genetic Programming, *Proc. IIAI AAI 2012*, pp. 340–344 (2012)

[Nagamine 07] Nagamine, M., Miyahara, T., Kuboyama, T., Ueda, H., and Takahashi, K.: A Genetic Programming Approach to Extraction of Glycan Motifs using Tree Structured Patterns, *Proc. AI-2007, Springer-Verlag, Vol. 4830*, pp. 150–159 (2007)

[Nakai 13] Nakai, S., Miyahara, T., Kuboyama, T., Uchida, T., and Suzuki, Y.: Acquisition of Characteristic Tree Patterns with VLDC's by Genetic Programming and Edit Distance, *Proc. IIAI AAI 2013*, pp. 147–151 (2013)

[NCI 99] NCI: National Cancer Institute, The NCI Open Database, *Release 1 Files* (1999)

[Ouchiyama 15] Ouchiyama, Y., Miyahara, T., Suzuki, Y., Uchida, T., Kuboyama, T., and Tokuhara, F.: Acquisition of Characteristic Block Preserving Outerplanar Graph Patterns from Positive and Negative Data using Genetic Programming and Tree Representation of Graph Patterns, *Proc. IWCIA 2015*, pp. 95–101 (2015)

[Rehman 12] Rehman, S., Khan, A., and Fong, S.: Graph mining: A survey of graph mining techniques, *Proc. ICDIM 2012*, pp. 88–92 (2012)

[Sasaki 08] Sasaki, Y., Yamasaki, H., Shoudai, T., and Uchida, T.: Mining of Frequent Block Preserving Outerplanar Graph Structured Patterns, *Proc. ILP-2007, Springer-Verlag LNAI 4894*, pp. 239–253 (2008)

[Shirakawa 07] Shirakawa, S., Ogino, S., and Nagao, T.: Graph Structured Program Evolution, *Proc. Genetic and Evolutionary Computation Conference (GECCO) 2007*, pp. 1686–1693 (2007)

[Tokuhara 16a] Tokuhara, F., Miyahara, T., Suzuki, Y., Uchida, T., and Kuboyama, T.: Acquisition of Characteristic Block Preserving Outerplanar Graph Patterns by Genetic Programming using Label Information, *Proc. IIAI AAI 2016*, pp. 203–210 (2016)

[Tokuhara 16b] Tokuhara, F., Miyahara, T., Suzuki, Y., Uchida, T., and Kuboyama, T.: Using Canonical Representations of Block Tree Patterns in Acquisition of Characteristic Block Preserving Outerplanar Graph Patterns, *Proc. IWCIA 2016*, pp. 93–99 (2016)

[Tokuhara 17] Tokuhara, F., Miyahara, T., Kuboyama, T., Suzuki, Y., and Uchida, T.: Acquisition of Multiple Block Preserving Outerplanar Graph Patterns by an Evolutionary Method for Graph Pattern Sets, *Proc. IWCIA 2017*, pp. 191–197 (2017)

[Yamasaki 09] Yamasaki, H., Sasaki, Y., Shoudai, T., Uchida, T., and Suzuki, Y.: Learning Block-Preserving Graph Patterns and Its Application to Data Mining, *Machine Learning*, Vol. 76, pp. 137–173 (2009)

[大内山 16] 大内山 祐斗, 宮原 哲浩, 鈴木 祐介, 内田 智之, 久保山 哲二, 徳原 史也：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, 2016 年度人工知能学会全国大会論文集,3I4-4 (2016)

[徳原 17] 徳原史也，宮原哲浩，久保山哲二，鈴木祐介，内田智之： 遺伝的プログラミングによる複合的なブロック保存型外平面的グラフパターンの獲得, 2017 年度人工知能学会全国大会論文集,4H2-2 (2017)

[徳原 18] 徳原史也，宮原哲浩，久保山哲二，鈴木祐介，内田智之：外平面的グラフのクラスタリングによる複合的なブロック保存型外平面的グラフパターンの進化的獲得, 2018 年度人工知能学会全国大会論文集,3E2-02 (2018)