

令和元年6月10日現在

機関番号：25403

研究種目：基盤研究(C)（一般）

研究期間：2015～2018

課題番号：15K00312

研究課題名（和文）強力な表現能力を持つグラフ構造パターンによるグラフ構造データからの深層知識の発見

研究課題名（英文）Discovery of Deep Knowledge from Graph-Structured Data using Expressive Graph-Structured Patterns

研究代表者

宮原 哲浩（MIYAHARA, Tetsuhiro）

広島市立大学・情報科学研究科・准教授

研究者番号：90209932

交付決定額（研究期間全体）：（直接経費） 3,500,000円

研究成果の概要（和文）：強力な表現能力を持つグラフ構造パターンによるグラフ構造データからの深層知識の発見について研究した。正事例と負事例の外平面的グラフから特徴的なブロック保存型外平面的グラフパターンを獲得する機械学習手法を提案した。正事例と負事例のグラフデータから特徴的な複合的ブロック保存型外平面的グラフパターンおよび、特徴的な複合的TTSPグラフパターンを獲得する手法を提案した。これらの学習手法は、構造的表現を扱うことのできる進化的学習手法である遺伝的プログラミングに基く。グラフ構造データから知識を発見するための他の学習手法も提案した。

研究成果の学術的意義や社会的意義

グラフ構造データから知識を発見して活用するためのデータマイニングと機械学習における基盤技術を確立するために、従来手法では表現できなかったパターン型深層知識を発見する手法を開発した。

研究成果の概要（英文）：We have studied discovery of deep knowledge from graph-structured data using expressive graph-structured patterns. We have proposed a machine learning method for acquiring characteristic block preserving outerplanar graph patterns from positive and negative outerplanar graph data. We have proposed machine learning methods for acquiring characteristic multiple blockblock preserving outerplanar graph patterns and characteristic multiple TTSP graph patterns from positive and negative graph data. These methods are based on Genetic Programming, which is an evolutionary learning method dealing with tree structures. We have proposed other learning methods for discovering knowledge from graph-structured data.

研究分野：総合領域

キーワード：データマイニング 機械学習 グラフ構造データ 木構造データ 遺伝的プログラミング

### 1. 研究開始当初の背景

トランザクションデータ, 数値データやテキストデータだけではなく, 化合物データ, 分子生物学データ, ウェブデータ, ソーシャルネットワークデータなどのグラフ構造を持つデータが大規模に収集されている. このような大規模なグラフ構造データから知識を発見し, 発見した知識を統合して利用するための手法やシステムの研究開発が求められている.

グラフ構造データや木構造データのような構造的データからのデータマイニングが大きな研究課題となっている. さらに, このような複雑なデータに隠されている深層知識(ディープナレッジ)の発見がデータマイニングにおける重要な研究課題として認識されている. 本研究では, この2つのデータマイニング課題に取り組み, グラフ構造データからのパターン型深層知識発見システムを開発する.

### 2. 研究の目的

本研究では, 大規模なグラフ構造データに隠されている深層知識(ディープナレッジ)を, 強力な表現能力を持つグラフ構造パターンを用いて表現し, パターン型の深層知識として発見するシステムを開発することを目的とする. 化合物データ, 分子生物学データ, ウェブデータ, ソーシャルネットワークデータなどのグラフ構造(ネットワーク構造)を持つデータが爆発的に増大している. このような大規模なグラフ構造データから知識を発見して活用するためのデータマイニングと機械学習における基盤技術を確立するために, 従来手法では表現できなかったパターン型深層知識を発見することを目指す.

### 3. 研究の方法

以下では, 正事例と負事例の外平面的グラフの集合から特徴的なブロック保存型外平面的グラフパターンを獲得する機械学習手法について述べる.

外平面的グラフの構造的特徴を表現するグラフ構造パターンであるブロック保存型外平面的グラフパターンおよび, その木パターン表現であるブロック木パターンが提案されている [Sasaki et al. 2008, Yamasaki et al. 2009]. 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ブロック保存型外平面的グラフパターン(BPO グラフパターン)という. BPO グラフパターン  $p$  と外平面的グラフ  $G$  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  $G$  と同型な外平面的グラフが得られるとき,  $p$  と  $G$  はマッチするという. 遺伝的プログラミング(Genetic Programming, GP)とは, 木構造のような構造的表現を扱う進化的学習手法である. 以下で定義する問題を対象とする.

#### 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する.

ここで, BPO グラフパターン  $p$  の  $D$  に関する適合度を  $(p$  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 と定義する. 適合度の高い BPO グラフパターンは,  $D$  の多くの正事例にマッチし,  $D$  の負事例にあまりマッチしないような, 特徴的な BPO グラフパターンであると考えられる. 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 による特徴的な 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. 図1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示す. 図2に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を示す.

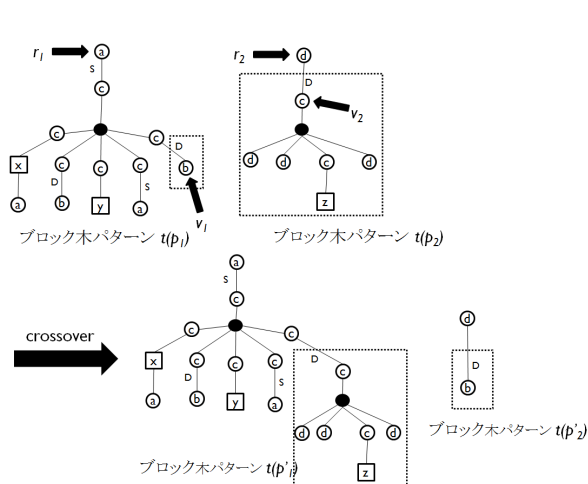


図1 ブロック木パターンに対する交叉

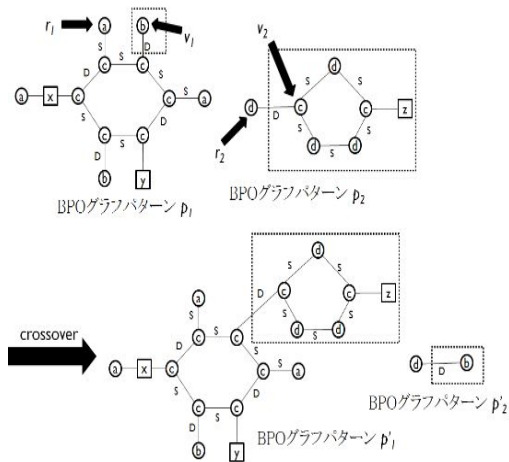


図2 BPO グラフパターンに対する交叉

#### 4. 研究成果

大規模なグラフ構造データに隠されている深層知識を強力な表現能力を持つグラフ構造パターンを用いて表現し、パターン型の深層知識として発見するシステムについて研究を行った。グラフ構造パターン型知識発見、木構造パターン型知識発見の2つのサブテーマで研究を展開し、以下の成果を得た。

遺伝的プログラミングとブロック保存型外平面的グラフパターンの木構造表現を用いて、正事例と負事例の外平面的グラフから特徴的なブロック保存型外平面的グラフパターンを獲得する進化的学習システムを実現した。正事例から抽出したラベルの情報を利用した遺伝的操作を用いる特徴的なブロック保存型外平面的グラフパターンを獲得する進化的学習手法も考案した。ブロック木パターンの正規形表現から定義される深さラベル列を利用して実行時間を削減する方法を提案した。正事例と負事例のグラフデータから特徴的なグラフパターンを獲得する進化的学習システムにおいて、特徴選択に基づいて相互に関連する個体の適合度を評価する手法を提案した。正事例と負事例の外平面的グラフから特徴的なブロック保存型外平面的グラフパターンを獲得する実験、および正事例と負事例の TTSP グラフから特徴的な TTSP グラフパターンを獲得する実験により、提案手法の有効性を確認した。ブロック保存型外平面的グラフパターン集合を個体とする二段階構造の進化的計算を用いて、正事例と負事例の外面的グラフから特徴的な複合的ブロック保存型外平面的グラフパターンを獲得する手法を提案し、その有効性を確認した。TTSP グラフパターン集合を個体とする進化的計算を用いて、正事例と負事例の TTSP グラフから、特徴的な複合的 TTSP グラフ構造パターンを獲得する手法を提案し、グラフパターン表示機能を実現した。

タグ木パターンは、任意の木構造データを代入できる構造的変数を持つ木構造パターンである。タグ木パターン集合に作用する遺伝的プログラミングによる、正事例と負事例の木構造データから特徴的な複合的木構造パターンを獲得する進化的学習手法を考案した。順序ワイルドカード木パターンは、任意の順序木を代入できる構造的変数と辺ラベルのワイルドカードを持つ木構造パターンである。順序タグ木パターンは、さらにタグとキーワードを持つ木構造パターンである。順序木構造データの構造的特徴を抽出するために、極大頻出順序ワイルドカード木パターンと極大頻出順序タグ木パターンを枚挙する手法を提案した。構造的変数を持つ木構造パターン、グラフ構造パターンの帰納推論手法、マッチング手法について研究した。

#### 5. 主な発表論文等

[雑誌論文](計 10 件)

Yuuki Yamagata, Fumiya Tokuhara, Yuito Inaba, Kouki Muratani, [Tetsuhiro Miyahara](#), Yusuke Suzuki, [Tomoyuki Uchida](#), [Tetsuji Kuboyama](#), Evolutionary Learning of Multiple TTSP Graph Structured Patterns from Positive and Negative TTSP Graph Data: Its Graph Pattern Display System and Performance Evaluation, Proc. 22nd Asia Pacific Symposium on Intelligent and Evolutionary Systems (IES 2018), pp.120-127, 2018. (査読有)

Fumiya Tokuhara, [Tetsuhiro Miyahara](#), [Tetsuji Kuboyama](#), Yusuke Suzuki, [Tomoyuki Uchida](#), Aggregative Context-Aware Fitness Functions based on Feature Selection for Evolutionary Learning of Characteristic Graph Patterns, Vietnam Journal of Computer Science, Vol.5, pp.229-239, 2018. (査読有) DOI:10.1007/s40595-018-0118-8

Fumiya Tokuhara, [Tetsuhiro Miyahara](#), [Tetsuji Kuboyama](#), Yusuke Suzuki, [Tomoyuki Uchida](#), Acquisition of Characteristic Sets of Block Preserving Outerplanar Graph Patterns by a Two-Stage Evolutionary Learning Method for Graph Pattern Sets, International Journal of Computational Intelligence Studies, Vol.7, pp.270-288, 2018. (査読有) DOI: 10.1504/IJCISTUDIES.2018.096191

Fumiya Tokuhara, [Tetsuhiro Miyahara](#), [Tetsuji Kuboyama](#), Yusuke Suzuki, [Tomoyuki Uchida](#), Acquisition of Multiple Block Preserving Outerplanar Graph Patterns by an Evolutionary Method for Graph Pattern Sets, Proc. IEEE 10th International Workshop on Computational Intelligence and Applications (IWCI A 2017), pp. 191-197, 2017. (査読有) DOI: 10.1109/IWCI A.2017.8203583

[Tetsuhiro Miyahara](#), Yusuke Suzuki, Takayoshi Shoudai, [Tomoyuki Uchida](#), [Tetsuji Kuboyama](#), Enumeration of Maximally Frequent Ordered Tree Patterns with Wildcards for Edge Labels, IPSJ Transactions on Mathematical Modeling and Its Applications, Vol.10, pp.59-69, 2017. (査読有)

Yuuki Yamagata, [Tetsuhiro Miyahara](#), Yusuke Suzuki, [Tomoyuki Uchida](#), Fumiya Tokuhara, [Tetsuji Kuboyama](#), Acquisition of Multiple Graph Structured Patterns by an Evolutionary Method using Sets of TTSP Graph Patterns as Individuals, Proc. 6th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI 2017), pp.459-464, 2017. (査読有) DOI: 10.1109/IIAI-AAI.2017.198

Fumiya Tokuhara, [Tetsuhiro Miyahara](#), [Tetsuji Kuboyama](#), Yusuke Suzuki, [Tomoyuki Uchida](#), A Context-Aware Fitness Function Based on Feature Selection for Evolutionary Learning of Characteristic Graph Patterns, Lecture Notes in Computer

Science Vol.10191, pp.748-757, 2017. (査読有)

Fumiya Tokuhara, Tetsuhiro Miyahara, Yusuke Suzuki, Tomoyuki Uchida, Tetsuji Kuboyama, Using Canonical Representations of Block Tree Patterns in Acquisition of Characteristic Block Preserving Outerplanar Graph Patterns, Proc. IEEE 9th International Workshop on Computational Intelligence and Applications (IWCIA 2016), pp. 93-99, 2016. (査読有) DOI: 10.1109/IWCIA.2016.7805755

Fumiya Tokuhara, Tetsuhiro Miyahara, Yusuke Suzuki, Tomoyuki Uchida, Tetsuji Kuboyama, Acquisition of Characteristic Block Preserving Outerplanar Graph Patterns by Genetic Programming using Label Information, Proc. 5th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI 2016), pp.203-210, 2016. (査読有) DOI:10.1109/IIAI-AAI.2016.212

Yuto Ouchiya, Tetsuhiro Miyahara, Yusuke Suzuki, Tomoyuki Uchida, Tetsuji Kuboyama, Fumiya Tokuhara, Acquisition of Characteristic Block Preserving Outerplanar Graph Patterns from Positive and Negative Data using Genetic Programming and Tree Representation of Graph Patterns, Proc. IEEE 8th International Workshop on Computational Intelligence and Applications (IWCIA 2015), pp. 95-101, 2015. 査読有) DOI: 10.1109/IWCIA.2015.7449469

[学会発表](計 11 件)

山縣佑貴, 徳原史也, 稲葉唯人, 村谷光輝, 宮原哲浩, 鈴木祐介, 内田智之, 久保山哲二, 進化的手法による複合的 TTSP グラフパターン獲得システムのパラメータ設定と性能評価, 2018 IEEE SMC Hiroshima Chapter 若手研究会, 2018.

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, 2018.

山縣佑貴, 宮原哲浩, 鈴木祐介, 徳原史也, 内田智之, 久保山哲二, 特徴的な複合的 TTSP グラフパターンの獲得における深さラベル列の利用, 2017 IEEE SMC Hiroshima Chapter 若手研究会, 2017.

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 2017.

徳原史也, 宮原哲浩, 鈴木祐介, 内田智之, 久保山哲二, 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におけるブロック木パターンの深さラベル列の利用, 2017 年度 人工知能学会全国大会, 2017.

山縣佑貴, 宮原哲浩, 鈴木祐介, 内田智之, 久保山哲二, TTSP グラフパターン集合を個体とする進化的手法による複合的グラフ構造パターンの獲得 火の国情報シンポジウム 2017, 2017. Tetsuhiro Miyahara, Yusuke Suzuki, Takayoshi Shoudai, Tomoyuki Uchida, Tetsuji Kuboyama, Enumeration of Maximally Frequent Ordered Tree Patterns with Wildcards for Edge Labels, 情報処理学会第 112 回数数理モデル化と問題解決研究発表会, 2017

徳原史也, 宮原哲浩, 鈴木祐介, 内田智之, 久保山哲二, 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におけるブロック木パターンの正規形表現の利用, 2016 IEEE SMC Hiroshima Chapter 若手研究会, 2016.

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, 2016.

重松卓也, 宮原哲浩, 鈴木祐介, 内田智之, タグ木パターン集合に作用する遺伝的プログラミングによる特徴的な複合的木構造パターンの獲得, 火の国情報シンポジウム 2016, 2016.

徳原史也, 宮原哲浩, 鈴木祐介, 内田智之, 久保山哲二, 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, 2016.

## 6. 研究組織

### (1) 研究分担者

内田 智之 (UCHIDA, Tomoyuki)

広島市立大学・情報科学研究科・准教授

研究者番号: 7 0 2 6 4 9 3 4

久保山 哲二 (KUBOYAMA, Tetsuji)

学習院大学・計算機センター・教授

研究者番号: 8 0 3 0 2 6 6 0

廣渡 栄寿 (HIROWATARI, Eiju)

北九州市立大学・基盤教育センター・教授

研究者番号: 6 0 2 7 4 4 2 9