

Title	Study on Mining General Structured Patterns
Author(s)	猪口, 明博
Citation	大阪大学, 2004, 博士論文
Version Type	
URL	<a href="https://hdl.handle.net/11094/44905">https://hdl.handle.net/11094/44905</a>
rights	
Note	著者からインターネット公開の許諾が得られていないため、論文の要旨のみを公開しています。全文のご利用をご希望の場合は、 <a href="https://www.library.osaka-u.ac.jp/thesis/#closed">〈a href="https://www.library.osaka-u.ac.jp/thesis/#closed"〉</a> 大阪大学の博士論文について <a href="https://www.library.osaka-u.ac.jp/thesis/#closed">〈/a〉</a> をご参照ください。

***Osaka University Knowledge Archive : OUKA***

<https://ir.library.osaka-u.ac.jp/>

Osaka University

氏名	猪口 明博
博士の専攻分野の名称	博士(工学)
学位記番号	第 18767 号
学位授与年月日	平成 16 年 3 月 25 日
学位授与の要件	学位規則第 4 条第 1 項該当 工学研究科通信工学専攻
学位論文名	Study on Mining General Structured Patterns (一般的構造を有するパターンのマイニングに関する研究)
論文審査委員	(主査) 教授 元田 浩 (副査) 教授 馬場口 登 教授 北山 研一 教授 小牧 省三 教授 塩澤 俊之 教授 河崎善一郎

#### 論文内容の要旨

本論文は、大量のラベル付きグラフデータから特徴的な部分構造グラフをパターン、ルール、あるいは知識として効率良く抽出する手法に関する研究の成果をまとめたもので、7章により構成されている。

第1章は序論であり、本研究の背景となるグラフデータに対するデータマイニング手法について概説し、本研究の意義と位置付けを明確にしている。

第2章では、本論文が取り扱うグラフ構造データからの多頻度グラフパターンマイニング問題について定式化している。多頻度グラフパターンマイニングとは、複数のラベル付きのグラフデータが与えられたときに、それらに頻出する全ての部分グラフをパターンとして出力する問題である。さらに第2章では、問題の解法として **Apriori-based Graph Mining** (以下、AGM) アルゴリズムを提案し、そのデータ構造とアルゴリズムについて詳細に説明している。

第3章では、第2章で提案した AGM アルゴリズムの改良について述べている。AGM アルゴリズムは入力として与えられた多数のグラフデータから頻出する部分構造をパターンとして出力するが、出力されるパターンには連結グラフのパターンとともに非連結グラフパターンも含まれ、ユーザが意図するパターンのみを取り出すためには後処理を必要とする。第3章では、バイアスと呼ばれる探索空間の制御方法を AGM アルゴリズムに導入した **Biased Apriori-based Graph Mining** (以下、B-AGM) アルゴリズムを提案し、グラフ構造以外の順序木データや一般の木構造データなど様々なタイプの入力データから特定の出力パターンのみを効率良く、高速に探索する手法について説明している。

第4章では、入力となるグラフがクラスラベルを持つ入力データを対象としている。第2、3章で扱った手法が出力する頻出パターンが入力データを分類するのに必ずしもより良い知識表現ではない例を挙げ、入力データを分類するために改良する手法を提案している。提案手法では、パターンの組み合わせに着目し、取り出されるルールの評価関数を最大化するルールを効率良く取り出す手法について述べている。

第5章では、提案手法の実験的評価について論じている。はじめに人工的に作成されたデータによる定量的なアルゴリズムの評価をしており、提案手法が十分に高速に動作し、有効であることを示している。次に、実データを用いて、提案手法と既存手法について比較評価しており、提案手法が既存手法と同等以上に効率良く動作することを述べ

ている。さらに、第4章で提案された手法が大規模グラフ構造データから有用な知識を効率良く発見できていることを分類問題による知識発見という観点から示すことでアルゴリズムの有用性を評価している。

第6章は、手法の提案と比較実験の結果を受けて、関連研究と相違点や提案手法の特徴について議論している。

第7章は結論であり、本研究で得られた結果の総括をしている。

## 論文審査の結果の要旨

本論文は、現在該当分野の主に行われている研究がグラフのような構造を持つデータから効率良く知識を取り出すことができないという観点から、構造を有する大規模データからより良い知識を効率良く発見する汎用的な手法を提案している。その主な成果は以下のように要約される。

1. グラフ構造から有用な部分構造パターンを抽出する問題は、頂点、辺、およびそれらのラベルの組み合わせ問題であり、非常に計算コストを要する問題であるが、パターンの表現方法、パターンの生成手法、パターン表現形式上の変換手法、および実装方法などを提案することで、非常に高速に頻出する全ての部分構造パターンを列挙するアルゴリズムを提案している。
2. AGM アルゴリズムの汎用手法である B-AGM アルゴリズムは、様々なデータ構造に適用でき、さらにユーザが意図する特定のパターンのみを抽出できるため、多くの問題事例に対し適用できる可能性を有している。さらに B-AGM アルゴリズムが、特定のデータ構造にのみ適用できるように提案された他の手法と比較され、提案された汎用手法が既存手法と同等以上に効率良く動作することを述べている。
3. パターンの組み合わせを求める問題では、パターンを評価する評価関数の特性を利用し、評価値の上限を見積もり、不要なパターンを効率良く枝狩りする手法を提案し、既存手法に後処理を追加した方法よりも高速に知識を抽出できることを、実データを用いて示している。ルールの有用性をカイ2乗検定により評価し、抽出されたルールが既存手法で得られたルールより良いルールを取り出していることを示している。
4. 上記で述べた手法を詳細に解説し、まとめている。
5. B-AGM アルゴリズムについて、人工データを用いた計算時間等の定量的な評価実験により、提案手法が高速であり、大規模なグラフ構造データにも計算時間の観点からも十分適用できる手法であることを示している。
6. 提案手法を実データに適用することで、現実世界の大規模なグラフ構造データからも高速に知識を抽出できることを示している。さらに、抽出された知識の有用性を薬理活性の問題により評価し、抽出された知識の信頼性という点からも提案手法が有用であることを確認している。

以上のように、本論文はグラフ構造に関するデータマイニング手法に関する多くの知見を得ており、大規模構造データ解析の分野の発展に寄与するところが大きい。よって本論文は博士論文として価値あるものと認める。