

Title	大規模グラフ構造データからの高速マイニングに関する研究
Author(s)	松田, 喬
Citation	大阪大学, 2003, 博士論文
Version Type	
URL	https://hdl.handle.net/11094/44287
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"〉 大阪大学の博士論文について 〈/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	まつ だ たかし 松 田 喬
博士の専攻分野の名称	博 士 (工 学)
学位記番号	第 17851 号
学位授与年月日	平成 15 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	大規模グラフ構造データからの高速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 河崎善一郎 教授 森永 規彦 教授 小牧 省三 教授 塩澤 俊之 教授 北山 研一

論 文 内 容 の 要 旨

本論文は、以下の 5 章により構成される。

第 1 章では、本研究の背景について説明し、データマイニングの現状と課題について述べた。

第 2 章では、まず最初の節で、グラフ構造データ中に現れる特徴的な構造を高速に抽出することができる GBI 法を提案した。GBI 法はグラフを“二つの頂点とその間をつなぐ辺”を最小単位とすることで、入力グラフを分解し、逐次的にそれらをひとつの頂点にまとめることでグラフから特徴的なパターンを抽出する。さらに、まとめられた頂点の内部と外部の接続情報を失わないためにグラフ表現を拡張した新しいデータ構造を提案した。また、他の関連研究と比較して、本研究の位置付けをさらに明確にした。

第 3 章では、GBI 法の改良について述べた。探索空間を指数的にではなく線形に増加させるために Beam 探索の導入が有効であることを説明した。別の改良として、GBI 法でより良いパターンを抽出するためには、二つの評価基準でパターンを評価すればよいことを提案した。また、ペアを正確に数えるために Canonical Label を用いてペアが表現するグラフ構造をチェックする仕組みを取り入れ、その効率的な計算方法を提案した。さらに、GBI 法により部分グラフではなく誘導部分グラフで評価する手法を提案した。最後に、以上の改良を実装したアルゴリズムを Beam-wise Graph Induction と新たに名付け、そのアルゴリズムをまとめた。

第 4 章では開発手法の実験的評価について論じた。まずはじめに計算時間を実験的に評価した。World Wide Web のアクセス記録を対象問題として、アクセス記録から多くのユーザが頻繁にたどる興味深い URL のパターンを抽出した。DNA データからはこれらのデータを分類できるルールを抽出した。このデータセットからは特徴的なパターンが抽出でき、さらにそれを用いた予測精度は他の主要な分類規則学習法と同程度であることが確認できた。さらに、化学物質の毒性をその化学構造から予測するための特徴的なパターンの抽出を試みた。そして、選び出されたパターンを用いて分類規則学習を行わせ、その予測精度を評価した。その結果、このデータセットが持つおおまかな特徴を捉えることができていることがわかった。

第 5 章は結言であり、本研究で得られた結論の総括をした。本研究により膨大なグラフ構造データから高速にかつ効率よく有用な知識発見をすることができる。

論文審査の結果の要旨

本論文では、現在該当分野で主に行われている研究では大規模な構造データを取り扱うことができないという観点から、大規模グラフ構造データから高速に有用な知識を発見するためのアルゴリズムを提案しており、さらにそのアルゴリズムにより抽出される知識の質を向上させるためのいくつかの提案をしている。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を抽出することは非常に計算コストがかかるため、グラフを二つ頂点とその間の辺からなるペアという単位で分解し解析することで、近似的にはあるが、非常に高速にグラフ構造から有用な部分構造を抽出するアルゴリズム GBI 法を提案している。
- (2) GBI 法が Greedy 探索を採用していることにより探索空間が限定されるという問題点に対し、Beam 探索を採用することで、探索空間の爆発的な増加を回避しながら線形的に探索空間を広げることが可能となることについて述べている。
- (3) GBI 法により抽出される知識をより対象ドメインに特化した質の高いものにするための問題点について述べ、評価指標を分離することでその問題点を改善できることを提案している。
- (4) GBI 法において、同じグラフ構造を異なる構造として取り扱うことがあることを述べ、この問題を解消するためにグラフ構造の同一性を判定するための手法である Canonical Label を提案手法に組み込み、またそれを高速に計算する方法について提案している。
- (5) 対象ドメインやグラフ構造によっては冗長な部分グラフが抽出される場合があり、それを解消するためには誘導部分グラフの抽出が有効であることを説明し、基本アイデアを拡張することにより GBI 法で誘導部分グラフのみを抽出することが可能となることを述べている。
- (6) 上で述べた改良を施したアルゴリズムを B-GBI 法と新たに名付け、そのアルゴリズムを詳細に解説し、まとめている。
- (7) GBI 法および B-GBI 法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十分耐えうる手法であることを示している。
- (8) GBI 法および B-GBI 法を実データに適用することで、現実世界の大量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。また、抽出された知識の有用性を分類規則学習の予測精度により評価し、抽出された知識の質という点からも提案手法が有用なものであることを確認している。

以上のように、本論文は大規模なグラフ構造データから高速に有用な知識を発見する手法についての多くの知見を含んでおり、大規模構造データ解析の分野の発展に大きく寄与するものである。よって本論文は博士論文として価値あるものと認める。