

科学研究費助成事業（科学研究費補助金）研究成果報告書

平成25年6月7日現在

機関番号：25403

研究種目：若手研究（B）

研究期間：2010～2012

課題番号：22700153

研究課題名（和文）

地図・罫線文書等を含むウェブデータを対象としたマイニング手法の開発

研究課題名（英文）

Development of data mining methods for web data containing map and ruled paper

研究代表者

鈴木 祐介（SUZUKI YUSUKE）

広島市立大学・情報科学研究科・助教

研究者番号：10398464

研究成果の概要（和文）：

本研究の目的は、地図や罫線文書などの平面上に描画可能な構造を持つ平面データを対象としたマイニング手法の開発、及びそのための理論的基盤についての確立である。TTSP グラフ、外平面的グラフ、順序グラフでモデル化される平面データを対象に、それらの特徴を表現するグラフパターンの提案を行った。さらに、それらのグラフパターンに対する多項式時間機械学習アルゴリズムを提案した。またその結果を応用したグラフマイニング手法について考察した。

研究成果の概要（英文）：

In this research, we aim to develop data mining system for graph structured data drawing in the plane such as Map and ruled paper. We introduced TTSP graphs patterns, BPO-graphs patterns and ordered graphs patterns for representing common structural features of such graph structured data. We considered polynomial time learning algorithms for finding a characteristic graph patterns. As applications of this results, some graph mining methods have proposed for such planar graphs.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2010年度	1,000,000	300,000	1,300,000
2011年度	900,000	270,000	1,170,000
2012年度	800,000	240,000	1,040,000
年度			
年度			
総計	2,700,000	810,000	3,510,000

研究分野：統合領域

科研費の分科・細目：情報学 知能情報学

キーワード：知識発見とデータマイニング

1. 研究開始当初の背景

ウェブデータからのデータマイニングに関しては国内外で盛んに研究がなされている。データが明示的で平坦な構造を持つ関係データベースやテキストデータなどからの情報抽出については活発に研究がなされており、その一部についてはすでにウェブ上の文書検索エンジンなど商業的に実用化され

ているものもある。

近年のコンピュータやネットワーク技術の発達によって、ウェブ上には多くの罫線文書や地図データ、CAD、LSIフロアプラン等の平面上に描画可能な、構造を持つデータが増加している。本研究では上記のような紙やディスプレイなどの平面上に描画可能

な構造を持つデータを平面データと呼ぶこととする。これら平面データに対して画像工学技術を応用した画像の類似性判定や、数学的な図形の相似性などを利用した特徴抽出手法を用いた研究がなされている。

2. 研究の目的

本研究では地図や罫線文書などの平面上に描画可能な、構造を持つデータ（平面データ）を対象としたマイニング手法の開発、及びそのための理論的基盤についての研究を目的とする。

平面グラフとは平面上の頂点と頂点同士を交差することなく結ぶ辺からなるグラフである。本研究では平面グラフを用いて、紙やディスプレイ上に描画可能な構造を持つ平面データを表現し、平面データに対する機械学習手法を用いた共通パターンの発見およびその学習アルゴリズムを用いたデータマイニング手法の開発を目的とする。

3. 研究の方法

平面データを表現できる平面グラフパターンの提案、及び効率のよいパターン照合アルゴリズムとパターン発見アルゴリズムの提案を行う。その結果を元に平面グラフパターンの機械学習可能性の考察を行う。また提案したアルゴリズムを用いて平面データからの特徴抽出プログラムの開発、及びデータマイニングシステムへの応用の研究を行う。

平面グラフには TTSP グラフや外平面的グラフなど様々な種類があるが、単一の平面グラフで全ての平面データを表現するのではなく、地図データ、罫線文書など平面データの集合ごとに適した平面グラフを提案する。提案した平面グラフに対して構造的変数を持たせた平面グラフパターンを提案し、その機械学習可能性を考察する。

4. 研究成果

(1) ウェブ上に存在する平面データの表現方法および、共通するパターンの発見に関する研究として、TTSP グラフに変数の概念を導入した TTSP グラフパターンのマイニング手法について考察を行った。TTSP グラフは電気ネットワークやスケジューリングなどの平面データを表現可能なグラフ構造である。本研究では、TTSP グラフ構造を有するデータに頻出する TTSP グラフパターンを枚挙するマイニングアルゴリズムを提案した。また提案したアルゴリズムをコンピュータ上に実装し、その評価実験を行った。さらに TTSP グラフパターンのマイニング手法に関して、

TTSP グラフと TTSP グラフパターンの照合アルゴリズムの改良と実装、および評価実験を行った。またその成果を用いた遺伝的プログラミングによる TTSP グラフパターンからの特徴抽出について発表を行った。

(2) 過去の研究で、分子化合物等に共通するパターンを表現可能な BPO グラフパターン（ブロック保存型外平面的グラフパターン）というグラフパターンと、それに対する照合アルゴリズムを提案している。分子化合物の多くは外平面的グラフとして表現可能であり、BPO グラフパターンを用いることで、外平面的グラフに共通するパターンが発見可能である。本研究では、BPO グラフパターンに対する照合アルゴリズムの実装と評価実験を行った。また BPO グラフパターンの応用について考察を行った。

(3) グラフパターンの研究に関して、補可約グラフパターンという新しいグラフパターンの提案を行った。補可約グラフは1頂点のグラフから和および補をとるという2つの演算を再帰的に繰り返すことのできるグラフである。補可約グラフはそのままでは平面データの表現に向かないが、補可約グラフの性質を利用することで平面データを表現するグラフパターンへの応用が可能であると考えられる。本研究では補可約グラフパターンに対して、照合アルゴリズムと共通パターン発見アルゴリズムの提案を行い、機械学習可能性についての考察を行った。本研究の成果について ILP2011 において発表を行った。

(4) ウェブ上に存在する地図や罫線文書等の平面データの表現方法および、共通するパターンの発見に関する研究として平面図パターンのマイニング手法について考察を行った。平面図パターンは地図などの平面データを表現するのに適したデータ構造である。本研究では、平面図パターンと平面データの照合問題を解く照合アルゴリズムの改良とコンピュータ上での実装を行い、その評価実験を行った。

さらに本研究の拡張として、順序グラフパターンという新しいグラフパターンの提案を行った。順序グラフはデータの位置情報や接続情報を表現可能なデータ構造であり、各頂点間の接続関係に順序を与えたものである。順序グラフを用いることで、地図や罫線文書の表現がより容易になることが考えられる。順序グラフパターンは順序グラフに共通する構造を表現可能なグラフパターンであり、平面図パターンの拡張にあたる。順序グラフパターンを用いることで、平面図パターンでは表現できなかった構造が表現可能になり、新たな知識発見に応用可能であると

考えられる。本研究では、順序グラフパターンの機械学習可能性と、それを用いたデータマイニング手法について考察を行った。機械学習可能性の考察、およびデータマイニングへの応用に関して順序グラフパターンと順序グラフの照合問題を解く照合アルゴリズムの提案を行った。照合アルゴリズムはグラフパターンがグラフの特徴を表現しているかどうかを判定する非常に重要なアルゴリズムである。本研究の成果について国際会議 ILP2012 において発表を行った。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計 4 件)

河野達也, 鈴木祐介, 内田智之, 宮原哲浩, Enumerating Maximally Frequent TTSP graph patterns, Proceedings of the Seventh Workshop on Learning with Logics and Logics for Learning (LLLL 2011), 査読有, pp.43-50, 2011

吉村友太, 正代隆義, 鈴木祐介, 内田智之, 宮原哲浩, Polynomial Time Inductive Inference of Cograph Pattern Languages from Positive Data, Proceedings of the 21st International Conference on Inductive Logic Programming (ILP2011), 査読有, pp389-404, 2012

日野隆博, 鈴木祐介, 内田智之, 糸川裕子, Polynomial Time Pattern Matching Algorithm for Ordered Graph Patterns, Proceedings of the 22nd International Conference on Inductive Logic Programming (ILP2012), 査読有, pp.86-101, 2013

永井聡, 宮原哲浩, 鈴木祐介, 内田智之, Acquisition of Characteristic TTSP Graph Patterns by Genetic Programming, Proceedings of 2012 IIAI International Conference on Advanced Applied Informatics (IIAI AAI 2012), 査読有, pp.340-344, 2012

[学会発表] (計 5 件)

河野達也, 鈴木祐介, 内田智之, 宮原哲浩, Enumerating Maximally Frequent TTSP graph patterns, Seventh Workshop on Learning with Logics and Logics for Learning (LLLL2011), 2011/3/30, 日本・大阪府

吉村友太, 正代隆義, 鈴木祐介, 内田智之, 宮原

哲浩, Polynomial Time Inductive Inference of Cograph Pattern Languages from Positive Data, The 21st International Conference on Inductive Logic Programming (ILP 2011), 2011/8/3, イギリス・ウインザー

日野隆博, 鈴木祐介, 内田智之, 糸川裕子, Polynomial Time Pattern Matching Algorithm for Ordered Graph Patterns, 22nd International Conference on Inductive Logic Programming (ILP2012), 2012/9/17, クロアチア・ドブロブニク

永井聡, 宮原哲浩, 鈴木祐介, 内田智之, Acquisition of Characteristic TTSP Graph Patterns by Genetic Programming, IIAI International Conference on Advanced Applied Informatics (IIAI AAI 2012), 2012/9/20, 日本・福岡県

大塚喜明, 鈴木祐介, 宮原哲浩, 内田智之, 遺伝的プログラミングによる特徴的な TTSP グラフパターン獲得手法の改良, 人工知能学会全国大会 2013, 2013/6/5, 日本・富山

[図書] (計 件)

[産業財産権]

○出願状況 (計 件)

名称 :
発明者 :
権利者 :
種類 :
番号 :
出願年月日 :
国内外の別 :

○取得状況 (計 件)

名称 :
発明者 :
権利者 :
種類 :
番号 :
取得年月日 :
国内外の別 :

[その他]

ホームページ等

6. 研究組織

(1) 研究代表者

鈴木祐介 (SUZUKI YUSUKE)

広島市立大学・情報科学研究科・助教

研究者番号：10398464

(2) 研究分担者

()

研究者番号：

(3) 連携研究者

()

研究者番号：