

平成 21 年 5 月 22 日現在

研究種目：若手研究 B

研究期間：2006～2008

課題番号：18700014

研究課題名（和文） データ解析における階層的知識構造の抽出に関する研究

研究課題名（英文） Extracting Hierarchical Knowledge Structures in Logical Analysis of Data

研究代表者

小野廣隆 (Hiroataka ONO)

九州大学・大学院システム情報科学研究所・助教

研究者番号：00346826

研究成果の概要：

本研究課題のテーマはデータからの知識獲得について、特に知識の持つ階層構造の抽出の立場から高性能アルゴリズムの開発、もしくは計算複雑度の解明を明らかにすることにある。人間が理解しやすい構造（たとえば階層性）を持った知識はある特別なクラスに属する論理関数の抽出問題として捉えることができる反面、それらを的確かつ厳密に捉えるアルゴリズムの構築は理論上・実際上難しいものであった（計算困難）。本研究ではこれらの計算困難な知識抽出にまつわる問題群に対し、現実的な制約の下でなら、もしくは厳密ではなく近似的にならば、これらの問題が実際上・また理論上十分高速に解決可能であることを示した。

交付額

（金額単位：円）

	直接経費	間接経費	合計
2006 年度	900,000	0	900,000
2007 年度	900,000	0	900,000
2008 年度	800,000	240,000	1,040,000
総計	2,600,000	240,000	2,840,000

研究分野：理論計算機科学

科研費の分科・細目：情報学基礎

キーワード：データ解析, 論理関数, 知識階層, 最適化問題, グラフモデル, 近似精度保証

## 1. 研究開始当初の背景

本研究課題のテーマはデータからの知識獲得について、特に知識の持つ階層構造の抽出の立場から高性能アルゴリズムの開発、もしくは計算複雑度の解明を明らかにすることにあつた。近年の情報・通信技術の発展、ゲノム解析など生物情報分野の発展から、大規模データから意味ある情報（ここでは「知識」と呼ぶ）を効率的に抽出する技術は、ますます重要性を増している。本研究では、申請者がこれ以前に取り組んできた論理関数を用いたデータ解析について研究の立場から見

ると、人間が理解しやすい構造（たとえば階層性）を持った知識はある特別なクラスに属する論理関数の抽出問題として捉えることができる。概念のもつ階層性は、知識構造における本質的な性質の一つであり、既存の手法によって得られた知識整理などの面においても重要な役割を果たすと考えられている。一方で、アルゴリズム論的な考察は十分なされておらず、緩やかなモデル化とヒューリスティクス解法による解決に基づく研究が大半であつた。

## 2. 研究の目的

本研究では、データ解析における知識階層構造の抽出を、(1)汎用性を持つ形でのモデル化、ならびに最適化問題としての定式化、(2)その問題の計算複雑度の解析、(3)実用的なアルゴリズムの設計と実装の観点から取り組むものであった。(1)の部分では、文字列データからの特徴抽出など、バイオインフォマティクス等への適用も可能な、汎用的なモデル化を目指すことにより、より有意義な結果が得られると考えている。また(2)において、問題を厳密に解くことが困難である(例えばNP困難)ことが明らかになった場合、(3)ではメタ解法などに基づく実用的な意味での高性能なアルゴリズムの開発を目指す。

## 3. 研究の方法

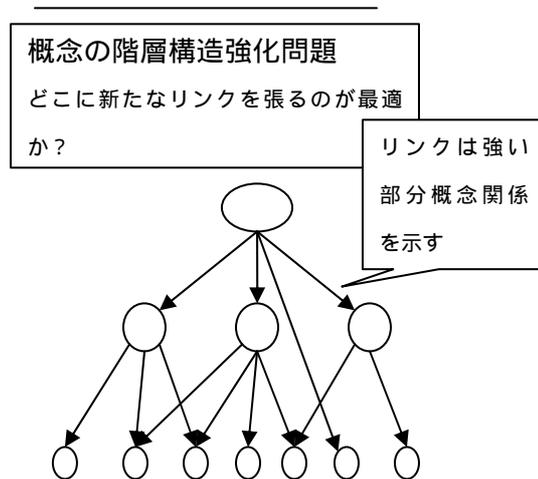
本研究では、データ解析における知識階層構造の抽出問題を組合せ最適化の立場から取り組むことにより新しい知見の獲得を目指す。具体的には、研究目的の項で述べた通り、(1)汎用性を持つ形でのモデル化、ならびに最適化問題としての定式化、(2)その問題の計算複雑度の解析、(3)実用的なアルゴリズムの設計と実装、の三視点から取り組む。基本的に、(1) (2) (3)の順に研究を進めていくことを考えているが、(2)と(3)の計算複雑度の解析とアルゴリズム設計は裏表の関係にあるため、必ずしもこの順に従うわけではない。また、得られる結果のモデルへのフィードバックは随時行うことを考えている。

## 4. 研究成果

研究期間を通して得られた結果を要約すると以下ようになる：知識モデルをグラフモデルへと展開した際に現れる種々の計算困難な問題に対して、知識が非巡回階層構造を持つ場合、最適な強化を行うことがNP困難(かつ近似困難)であることを示した。一方で、 $P \neq NP$ の仮定の下で、最適な近似を行う多項式時間アルゴリズムを提案した。これにより、本問題に対しては理論的には最善のアルゴリズムが得られたこととなる。最適構造抽出に関する問題について、関連知識をクラスタ化する問題において、そのクラスタサイズをバランスさせる問題をバランス化尺度の最適化問題として定義することにより、高速な(現実的な多項式時間の)アルゴリズムを設計することに成功した。得られた知識構造を蓄えた知識ベースから、いかに効率的かつロバストに演繹を行うかについて、一般には(残念ながら)計算困難であることが判明したが、実際上の利用でよく現れると考えられる「短い表現」の知識活用に限定すると、十分高速な演繹が可能であることが判明した。

(1) 平成 18 年度： 知識構造をグラフとしてモデル化した際に現れる最適構造抽出に関する問題について取り組んだ。まず問題の計算複雑度を明らかにし、それに対する精度保証付き近似アルゴリズムを開発すると共に、現実的な求解の観点から局所探索に基づく高精度アルゴリズムの開発を行った。

(2) 平成 19 年度： 知識構造をグラフとしてモデル化した際に現れる最適構造抽出に関する問題について取り組んだ。具体的には、知識が非巡回階層構造を持つ場合(たとえばWebなどが持つ重複を持つ階層構造)について、その構造強化問題を定式化し、最適な強化を行うことがNP困難(かつ近似困難)であることを示した。一方で、 $P \neq NP$ の仮定の下で、最適な近似を行う多項式時間アルゴリズムを提案した。これにより、本問題に対しては理論的には最善のアルゴリズムが得られたこととなる。



(3) 平成 20 年度： 知識構造をグラフとしてモデル化した際に現れる最適構造抽出に関する問題について、また得られた知識構造を蓄えた知識ベースから、いかに効率的かつロバストに演繹を行うかについて、の2点に焦点を絞り取り組んだ。前者に関しては知識構造をグラフとして表現した際に、関連知識をクラスタ化する問題において、そのクラスタサイズをバランスさせる問題をバランス化尺度の最適化問題として定義することにより、高速なアルゴリズムを設計することに成功した。後者に関しては、ロバストな演繹は一般には(残念ながら)計算困難であることが判明したが、実際上の利用でよく現れると考えられる「短い表現」の知識活用に限定すると、十分高速な演繹が可能であることが判明した。この結果は具体的には以下のような表としてまとめられる。

知識ベースからのロバスト演繹の  
計算複雑度

	内包	外包	外包 近似
式表現	P	coNP-C(1)	coNP-C
モデル表現	NP-C(2)	coNP-C(3)	P

(1) ただし, ロバストパラメータが定数の場合, またはクエリの負リテラル数が知識ベースの  $\log$  オーダーならば多項式時間.

(2) ただし, ロバストパラメータが定数の場合多項式時間.

(3) ロバストパラメータが定数の場合, またはクエリの正リテラル数が定数, またはクエリの負リテラル数が知識ベースの  $\log$  オーダーならば多項式時間.

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 8 件)

K. Makino, H. Ono, Deductive Inference for the Interiors and Exteriors of Horn Theories, the Proceedings of Algorithms and Computation, 19th International Symposium, ISAAC2008, Lecture Notes in Computer Science, 5369, 390-401, 2008.

Y. Harada, H. Ono, K. Sadakane, M. Yamashita, The Balanced Edge Cover Problem, the Proceedings of Algorithms and Computation, 19th International Symposium, ISAAC2008, Lecture Notes in Computer Science, 5369, 246-257, 2008.

T. Hasunuma, T. Ishii, H. Ono, Y. Uno, An  $O(n^{1.75})$  Algorithm for  $L(2, 1)$ -Labeling of Trees, Proceedings of 11th Scandinavian Workshop on Algorithm Theory, SWAT2008, Lecture Notes in Computer Science, 512, 185-197, 2008.

Y. Asahiro, E. Miyano and H. Ono, Graph Classes and the Complexity of the Graph Orientation Minimizing the Maximum Weighted Outdegree, Proceedings of Computing: the Twelfth Australasian Theory Symposium (CATS 2008), Wollongong, NSW, Australia. CRPIT, 77. Harland, J. and Manyem, P., Eds., ACS. 97-106, January 2008.

Y. Asahiro, E. Miyano, T. Murata and H. Ono, On Approximation of Bookmark Assignments, Proceedings

of 32nd International Symposium on Mathematical Foundations of Computer Science, MFCS 2007, Lecture Notes in Computer Science 4708, 115-124, 2007.

Y. Asahiro, J. Jansson, E. Miyano, H. Ono and K. Zenmyo, Approximation Algorithms for the Graph Orientation Minimizing the Maximum Weighted Outdegree, Proceedings of The Third International Conference on Algorithmic Aspects in Information and Management, AAIM 2007, Lecture Notes in Computer Science 4508, 167-177, 2007.

Y. Asahiro, E. Miyano, H. Ono, K. Zenmyo, Graph Orientation Algorithms to minimize the Maximum Outdegree. International Journal Foundation Computer Science 18(2), 197-215, 2007.

Y. Asahiro, T. Horiyama, K. Makino, H. Ono, T. Sakuma, and M. Yamashita, How to Collect Balls Moving in the Euclidean Plane, Discrete Applied Mathematics Vol 154, no. 16, 2247-2262, 2006.

[学会発表](計 5 件)

小野廣隆, 蓮沼徹, 石井利昌, 宇野裕之  
A Linear Algorithm for  $L(2, 1)$ -labeling of Trees, 日本応用数理学会 2008 年度年会, 2008 年 9 月 17 日 東京大学柏キャンパス.

小野廣隆, 蓮沼徹, 石井利昌, 宇野裕之,  
An  $O(n \log^2 n)$  Algorithm for  $L(2, 1)$ -labeling of Trees, 日本オペレーションズ・リサーチ学会平成 20 年度秋季研究発表会, 2008 年 9 月 11 日, 北海道.

橋本英樹, 小野 廣隆, 宇野毅明, 漆原秀子, 柳浦睦憲, 近傍ハッシュ法によるエラー許容頻出パターン列挙, 情報処理学会研究報告 第 13 回バイオ情報学研究会, 2008 年 6 月 27 日琉球大学.

H. Ono (T. Hasunuma, T. Ishii, Y. Uno), An Efficient Algorithm for  $L(2, 1)$ -Labeling of Trees, The First AAAC Annual Meeting AAAC08, 2008 年 4 月 26 日, 香港.

K. Maemura, J. Jansson, H. Ono, K. Sadakane and M. Yamashita. Approximation Algorithms for Constructing Evolutionary Trees from Rooted Triplet, 2007 Korea-Japan Joint Workshop on Algorithms and Computation. 56-63.

2007 (11<sup>th</sup> August 2008)

6 . 研究組織

(1)研究代表者

小野廣隆 (ONO Hirotaka)

九州大学大学院システム情報科学研究

院・助教

研究者番号 : 00346826