

機関番号：12612

研究種目：若手研究（B）

研究期間：2009年度～2010年度

課題番号：21700010

研究課題名（和文） 高機能な問合せをサポートする実践的グラフ圧縮アルゴリズムの開発

研究課題名（英文） Practical encodings of graphs with highly functional query supports

研究代表者

山中 克久 (YAMANAKA KATSUHISA)

電気通信大学・大学院情報システム学研究科・助教

研究者番号：60508836

研究成果の概要（和文）：グラフ圧縮アルゴリズムの開発を行った。とくに、極大平面グラフや格子方形描画に着目して研究を進めた。極大平面グラフに関しては、圧縮アルゴリズムを設計し、高速に問合せをサポートする方法を設計した。問合せとしては(1) 隣接判定、(2) 次数判定、(3) 時計回りに次に現れる頂点の計算、の3種類をサポートすることができる。とくに(3)の問合せは、平面グラフに関するアルゴリズムで頻繁に使われる命令であるため実用性が非常に高い。また、格子方形描画に対して圧縮アルゴリズムを設計し、格子方形描画に対するコンパクトな表現を与えた。

研究成果の概要（英文）：We designed encoding algorithms of graphs. We mainly focuses on plane triangulations and grid rectangular drawing, and developed encoding algorithms of them. The algorithm can encode a plane triangulation with compact space and supports the following queries: (1) adjacency, (2) degree, and (3) clockwise neighbour query. The clockwise neighbor query is frequently used in algorithms for plane graphs. Hence it is an extremely practical query. We also developed an algorithm which encodes a grid rectangular drawing with compact space.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2009年度	600,000	180,000	780,000
2010年度	600,000	180,000	780,000
年度			
年度			
年度			
総計			

研究分野：総合領域

科研費の分科・細目：情報学・情報学基礎

キーワード：アルゴリズム理論，グラフ理論，圧縮アルゴリズム，クエリのサポート

## 1. 研究開始当初の背景

データ圧縮とは、データの内容を全く、または、ほとんど損なうことなくデータの量を小さくすることである。近年、様々なデータ圧縮技術が広く利用されるようになった。データ圧縮には、大別して可逆圧縮と非可逆圧縮の2種類がある。圧縮したデータを完全に

元のデータに復元できる圧縮手法は可逆圧縮と呼ばれ、データの圧縮・復元時に情報の損失が発生するものは非可逆圧縮と呼ばれる。

本研究は、超大規模なグラフ構造の可逆圧縮に関するものを扱う。

近年、カーナビの地図データや、Webのリンク構造、遺伝子データ、XMLデータ、3

次元物体の精密なメッシュデータ等、様々な分野で超大規模なグラフ構造を高速に扱う技術が求められている。これら大規模グラフを自然なデータ構造で格納すると、膨大なメモリが必要となり、データの保存や送受信に膨大なコストが必要になってしまう。一方、これらのデータを圧縮することができるならば、保存や送受信のコストを節約できる。

しかし、圧縮したデータを使用する際には、自然なデータ構造に戻してから通常利用する。この際、グラフのデータは高速なメインメモリに収まらず、ハードディスク等の低速な2次記憶装置を使って処理することになるかもしれない。ハードディスク上のデータへのアクセスは、メインメモリへのアクセスより数十倍以上遅いので、大規模グラフデータを扱うシステムの速度は極めて遅いものになってしまう。

もし、大規模なグラフのデータを圧縮したまま利用することができるならば、ハードディスクを使用することなくメインメモリのみを使用して扱えるグラフの規模が大きくなる。その結果、計算速度の大幅な改善が期待できる。本研究では、そのような圧縮したまま利用できる、新しいグラフのデータ構造を設計・開発する。

現実世界では、大規模なグラフデータが頻繁に出現するようになったため、上記のような問題が応用分野で数多く発生している。そのため、高機能なグラフの圧縮アルゴリズムは、大規模グラフデータに対する基盤技術として早急に研究されるべき課題である。

一方、近年の情報爆発により、グラフに限らず、超大規模なデータを取り扱うことが現実問題として発生しているため、巨大データを扱うための様々な研究が行われている。その代表的なもの1つとして、DNA等に対する“巨大な文字列データ”を高速に処理するための技術が盛んに研究されている(suffix tree, suffix array など)。本研究はこれの“グラフ版”と言える。

グラフの圧縮に関する研究は、既にいくつか行われている。それらは“理論的な興味”という観点から研究されており、実用性が極端に低い段階にある。本研究では実用的な性能を考慮したアルゴリズムを開発する。また、既存研究で考慮されているクエリの種類は、グラフの頂点の隣接判定・次数判定の2つが主であった。本研究では、応用分野を考慮し、より高機能なクエリの実現を目指す。

## 2. 研究の目的

本研究では、大規模データを効率良く扱うことを目的として、高機能なグラフ圧縮アル

ゴリズムを開発する。

既存研究が、理論的な性能を重視していることに對し、本研究では、実際の性能を考慮した圧縮アルゴリズムを設計する。

まず、実用上、重要なグラフクラスに着目して圧縮アルゴリズムを設計する。また、実用上重要な情報を圧縮データから直接に取り出す方法を開発する。

例えば、3次元物体を表現する三角メッシュをモデル化したグラフとして、極大平面グラフというものが知られている。このグラフをコンパクトに圧縮して表現する方法と、圧縮した表現から高速に元のグラフの情報を取り出すための方法を開発する。このようにすることで、圧縮したままでも、元の極大平面グラフの情報を取り出すことができる。応用的には、元の三角メッシュの情報を得ることができる。三角メッシュを扱う分野ではどのようなクエリが求められているかを調査し、応用分野にあったクエリの開発を目指す。

その他にも、様々なグラフのクラス、または離散構造に対して、圧縮アルゴリズムと圧縮後のデータをそのまま利用する技術を開発する。

## 3. 研究の方法

現在主流になっているグラフ圧縮アルゴリズムのアイデアは次の通りである。

はじめに、グラフを小さなパーツに分割し、パーツ同士の隣接関係を覚える。同様に、各パーツをさらに小さなパーツに分割し、パーツ同士の隣接関係を覚える。最小のパーツに対してどうするかと言うと、それ自身を覚えておくことはしない。代わりに、パーツとして取り得る全パターンをテーブルに記憶する。各パーツについて、対応するテーブル番号さえ覚えておけば、テーブルを参照してパーツの詳細を得ることができる。ここで、テーブル番号を記憶するために必要な記憶容量は、パーツを覚えるための記憶領域としては情報理論的に最適になっていることに注意されたい。さらに、各パターンに対して、クエリの解を計算するために必要な情報を前もって計算してテーブルに保存しておく。この情報をいくつか組み合わせることによりクエリの解を計算する。このような“2段階の分割”と“解情報のテーブル”のおかげで、コンパクトにグラフを圧縮したままでも高速な問合せをサポートしている。

この手法では、パーツ1つ当たりのサイズが巧妙に設定されているため、理論的には効率の良い結果になっている。しかし、実際に扱うようなサイズのグラフに対しては非効率である、という短所がある。

上記のものとは別のアイデアに従って、既

に、方形描画というグラフに関して高機能な圧縮アルゴリズムを提案している。ここで、方形描画とは、全ての面が四角形であるような平面グラフのことである。詳細は省略するが、このアルゴリズムのアイデアは次の通りである。まず、方形描画の面を頂点に、面同士の隣接関係を辺に置き換えることによって、双対グラフを作成する。得られた双対グラフに対して全域木を“上手”に定義し、全域木と全域木以外の部分に分けて符号化を行う。全域木に関しては、全域木の親子関係を、括弧の入れ子構造に対応付けることにより、括弧からなる2進符号へ変換する。全域木以外の部分については、詳細は省略するが、全域木以外の部分にも入れ子構造を定義することができ、これを同様に括弧の入れ子構造で表現する、というものである。このアルゴリズムにより方形描画を  $5n$  bit で表現でき、さらに、面同士の隣接判定と面に関する次数判定を、圧縮表現からそれぞれ  $O(1)$  時間で計算できる。ここで、 $n$  は方形描画の頂点数である。

圧縮アルゴリズムのアイデアを2つ説明したが、両者の手法を他のグラフクラスへ適用できないか深く考察し、新しい圧縮アルゴリズムの開発を試みる。実際の研究成果としては、後者のアイデアに基づいたものが得られた。これについては、4. 節で詳しく紹介する。また、方形描画の圧縮アルゴリズムを一般化し、より応用に近い形で圧縮アルゴリズムを開発した。これについても詳細は4. 節を参照されたい。

次に研究体制について述べる。

基本的に単独で行う研究課題であるため、研究遂行が滞ったときに、時間が浪費されることが懸念される。そこで、できる限り他研究者と交流するように研究活動を行う。

得られた研究成果は積極的に国内外の学会等で発表し、他研究者からの意見を募り、さらなるアルゴリズムの改良に役立てる。また、積極的に他研究者とディスカッションすることで、単独では思いつかないようなアイデアを発見できるように工夫する。

研究会等の学会に積極的に参加し、「応用分野ではどのようなクエリが求められているのか？」を調査し、クエリの種類として実用性の高いものを設計できるように工夫する。

#### 4. 研究成果

主に、極大平面グラフや格子方形描画に着目して研究を進めた。

極大平面グラフは、3次元物体を表す三角メッシュをモデル化したグラフであり、応用的にも重要なグラフのクラスである。本研究

では、極大平面グラフを圧縮するアルゴリズムを設計した。具体的には、与えられた極大平面グラフを  $6n$  bit の2進文字列へ符号化・復号化する方法を与えた。ここで  $n$  は極大平面グラフの頂点の個数である。開発したアルゴリズムはシンプルなアイデアに基づいており、実装も容易である。既存のアルゴリズムは、非常に複雑で、実装が困難であったが、本研究では、実装が容易で、かつ、コンパクトな表現を与えることができた。この結果は、当該テーマにおいて大きなインパクトをもつ。コンパクトな表現を与えただけでなく、様々なクエリをサポートする方法も設計した。サポートしたクエリは、(1)隣接判定、(2)次数判定、(3)時計回りに次に現れる頂点の計算、の3種類である。とくに(3)の間合せは、平面グラフに関するアルゴリズムで頻繁に使われる命令であるため実用性が非常に高い。それぞれ、 $O(1)$  時間で計算できることが理論的に保証されており、非常に高速にクエリの解を返すことができる。

また、格子方形描画を圧縮してコンパクトに表現する方法を設計した。方形描画というグラフは、VLSI 等に応用をもつ重要なグラフクラスの1つとして知られているが、方形描画の各点を格子上に配置し、各辺に長さの情報をもたせたものが格子方形描画である。方形描画を圧縮する技術はすでに知られているが、長さの情報をもった格子方形描画を圧縮する方法は今までに知られていなかった。格子方形描画は、各頂点が格子上に配置された方形描画である。各辺が長さの情報をもっているため、方形描画よりも応用に近いクラスである。よって、既存の方形描画に対する圧縮アルゴリズムよりも、より実用性が高い圧縮アルゴリズムを開発できたとと言える。設計した圧縮アルゴリズムにより、 $m$  本の辺をもち、長さの総和が  $L$  であり、 $m_H$  本の極大水平線分をもつような格子方形描画を高々  $m+L-(m_H-1)$  bit で表現できる。ただし、クエリに答える方法は未開発であり、機能性が低い圧縮アルゴリズムになっている。隣接判定・次数判定を高速に圧縮データから計算できるようにすることが今後の課題として挙げられる。

その他にも、指定した個数の葉をもつ根つき順序なし木とあみだくじに対して列挙アルゴリズムを、2部置換グラフ・真区間グラフに対して数え上げ・ランダム生成・列挙を行うアルゴリズムを設計した。各離散構造について、列挙や数え上げによって個数を計算すれば、符号の上限を知ることができるので、これらの成果は、今後の圧縮アルゴリズムの開発のために非常に重要であると言える。この中でもとくに、あみだくじに関しては良い成果が得られた。あみだくじは、日本人にとって馴染み深いだけでなく、実は、代数学や

幾何学の分野に登場する重要なオブジェクトである。あみだくじを数え上げることにより、幾何学の分野に登場する擬直線アレンジメントの個数を数え上げることに成功した。この成果はオンライン整数列大辞典 (<http://oeis.org/A006245>) に掲載されている。

あみだくじをはじめとする、これらの離散構造に対してコンパクトな符号を設計することが今後の課題として挙げられる。

## 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 3 件)

① 山中克久, 中野眞一, A Compact Encoding of Plane Triangulations with Efficient Query Supports, Information Processing Letters, 18-19 巻, 803 頁~809 頁, 2010 年, 査読有

① 斎藤寿樹, 山中克久, 清見礼, 上原隆平, Random Generation and Enumeration of Proper Interval Graphs, IEICE Transactions on Information and Systems, E93-D 巻, 1816 頁~1823 頁, 2010 年, 査読有

① 山中克久, 中野眞一, 松井泰子, 上原隆平, 仲田研登, Efficient Enumeration of All Ladder Lotteries and Its Application, Theoretical Computer Science, 411 巻, 1714 頁~1722 頁, 2010 年, 査読有

[学会発表] (計 6 件)

① 須田亮平, 格子方形描画のコンパクトな符号, 第 73 回 情報処理学会 全国大会, 東京工業大学, 2011 年 3 月 2 日

② 山中克久, Enumerating All Rooted Trees Including k Leaves, 情報処理学会 第 131 回 アルゴリズム研究会, 函館工業専門学校, 2010 年 9 月 22 日

③ 斎藤寿樹, Random Generation and Enumeration of Bipartite Permutation Graphs, International Symposium on Algorithms and Computation (ISAAC 2009), ホノルル, アメリカ合衆国, 2009 年 12 月 18 日

④ 斎藤寿樹, Random Generation and Enumeration of Bipartite Permutation

Graphs, 電子情報通信学会 コンピューテーション研究会, 鳥取環境大学, 2009 年 9 月 14 日

⑤ 斎藤寿樹, Random Generation and Enumeration of Bipartite Permutation Graphs, 夏の LA シンポジウム, 松島(宮城県), 2009 年 7 月 24 日

⑥ 山中克久, Efficient Enumeration of All Pseudoline Arrangements, 情報処理学会 第 124 回 アルゴリズム研究会, 東京大学 医科学研究所, 2009 年 5 月 5 日

## 6. 研究組織

### (1) 研究代表者

山中 克久 (YAMANAKA KATSUHISA)  
電気通信大学・大学院情報システム学研究科・助教  
研究者番号: 60508836