

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24年 5月29日現在

機関番号：14301

研究種目：基盤研究(C)

研究期間：2009～2011

課題番号：21500014

研究課題名（和文） 巨大情報からの超高速情報抽出アルゴリズムの研究

研究課題名（英文） Hypervelocity information extraction from huge informations

研究代表者

伊藤 大雄 (ITO HIRO)

京都大学・大学院情報学研究科・准教授

研究者番号：50283487

研究成果の概要（和文）：データが膨大である対象をあつかうための、高速アルゴリズムとその計算量の下界値に対して研究を行った。特に、対象となるデータのうちのごく一部、定数個しか見ないで問題を解くという定数時間アルゴリズムを中心に研究し、グラフの最大独立集合や最大マッチングのサイズの近似、外平面性の検査、疎性マトロイドのランクの近似、ナップザック問題の解の近似に対して、高速な定数時間アルゴリズムを与えた。また、線形下界値を示す技法をいくつか与えた。他に、ユニットディスクグラフを描画領域でパラメータ化し、いくつかの問題に対し FPT アルゴリズムと下界値を与えた。

研究成果の概要（英文）：We investigated fast algorithms for treating big data, and lower bounds about them. Especially, algorithms that see only constant number of data of the given instance are mainly investigated. We obtained fast constant-time algorithms for approximating the size of the maximum independent set and the maximum matching of graphs, a testing outerplanarity of graphs, approximating the rank of sparsity matroids, and approximating the solution of the knapsack problem. And we presented new tools for showing linear lower bounds for testing algorithms. Moreover, we considered to parametrize unit disk graphs by the area, and gave some FPT algorithms and lower bounds on some problems on them.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2009年度	1,100,000	330,000	1,430,000
2010年度	900,000	270,000	1,170,000
2011年度	1,300,000	390,000	1,690,000
年度			
年度			
総計	3,300,000	990,000	4,290,000

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、定数時間アルゴリズム、グラフアルゴリズム

1. 研究開始当初の背景

計算機能力の向上に伴い、我々の回りには前世紀のそれとは比較にならない程巨大な情報がひしめいている。例えば、インターネ

ットとそれに関連するデータベース、ゲノム情報、POS システム等の管理システムの情報等、様々である。しかし、いくら情報量が膨大でもそれを適切に扱う方法を知らなけれ

ば使い物にならない。例えば、インターネットも、その初期においては我々一部の専門家が電子メールやファイル転送に利用していたぐらいであったが、Googleなどの検索エンジンの出現により格段に便利さが増し、一気に一般家庭に浸透し、今や、研究活動はもとより、日常生活に不可欠な物とまでなっている。

巨大情報を扱う上で重要なのは、その情報からその本質をいかに素早く見抜き、いかにコンパクトに表現するか、という技術である。例えばインターネット検索は、ウェブページの重要度を、簡便なアルゴリズムから得られるページランクという数値で表現したことの勝利である。

この考えに基づき、申請者らは科学研究費基盤研究(C)「巨大情報のアルゴリズム的超圧縮技法の研究」(平成18年度～20年度)において、クリークの線形時間列挙、家系図グラフの対数的圧縮表現、一般化ゴスパー曲線の効率の列挙等で重要な成果を出した。しかしそれと同時に新たな問題点も見えてきた。

例えば、クリークの列挙は、応用として、インターネットをモデル化したウェブグラフがある。我々の成果は理論上最高速の線形時間を達成しており、このことは大きな成果であるが、一方、ウェブグラフそのものが途轍もなく巨大(数え方にも依存するが、2008年夏の時点でノード数が数百億～一兆)なので、それを全て見るだけで時間がかかってしまうという問題点がある。これまでの常識では、対象物を全て見なければ話にならないので、線形時間は理論上の最高速と考えられていた。しかし、近年その常識が崩れつつある。すなわち、線形より少ない計算時間(劣線形時間)で計算をするアプローチの出現である。

その代表的なものが性質検査(property testing)という枠組みである。「対象物(例えばグラフG)が性質P(例えば平面性)を持つか?」という問題に対し、性質検査アルゴリズムは、対象物が性質Pを持てば高確率で受理し、性質Pから遠ければ(定義は略するが、パラメータ $0 < \epsilon \leq 1$ を用いて「 ϵ -遠隔」と表現される)拒否する。

性質検査の歴史は遡っても10年あまりであるが、最近急激に注目を浴び、ここ2～3年では国際会議の最重要トピックの一つとなっている。性質検査の計算時間は、対象物をどこまで見るか(これを「質問(クエリ)計算量」という)で評価できるが、多くの場合、劣線形は言うに及ばず、定数時間アルゴリズムが主流である。対象物を定数個しか見ない、ということは、ウェブグラフやゲノム情報等の超巨大情報を扱うのいうってつけのアプローチである。この技術を採用することで、巨大情報からの情報抽出技術が飛躍的に進歩すると考えられる。

2. 研究の目的

主にグラフに対する定数時間性質検査アルゴリズムの研究を行う。アルゴリズムは定められたクエリによってグラフの形状を把握するが、そのクエリの違いによって大きく分けて二通りのモデルが存在する。一つは「隣接行列モデル」で、もう一つは「次数制限モデル」である。前者は密な(枝数が多い)グラフに使われ、後者は疎なグラフに用いられる。本研究では、主に次数制限グラフに主眼を置く、その理由は、ウェブグラフは一般に疎になるからである。そして次数制限モデルで定数時間検査可能なものと、そうでないものを明らかにしていく。まずは個別の問題を解いていくことになるが、最大マッチング問題、最大独立集合問題、ハミルトン閉路問題等が最初の考慮の対象となる。これらの結果を元に、定数時間で検査できるものと、出来ないもののそれぞれの特徴づけを行っていき、期間内に、ある程度の一般的な結果を出すことを目標とする。

隣接行列モデルにおいては、Regularity Lemmaの活用によって、多くの性質が定数時間で検査できることが既に分かっている。しかし、次数制限モデルにおいては、このような万能な方法は知られておらず、断片的な成果しかない。また、定数時間アルゴリズムが知られているものについても、現時点では数学的な興味に基づく研究がほとんどであるため、非常に大きな定数のものが多く、実用性が低い。これらに対し実用的な検査アルゴリズムが開発されれば、工学的な意味でも大きな進展に繋がると考えられる。

3. 研究の方法

次数制限モデルを主な対象とし、定数時間検査可能な問題と線形時間下界値を持つ問題との峻別を行うことを目標とする。(但しその中間のものも存在する。)さらに、定数時間検査可能なものについては、その検査アルゴリズムの高速化(定数を小さくする)を行う。実際、多くの定数時間検査アルゴリズムの定数は、パラメータの指数になっているものが多く(中には2段階以上の指数のものもある)、実用性は疑問である。この原因は、これまでは数学的な研究対象であったからであり、本研究では応用(実用性)を意識して、定数であっても、パラメータの多項式で表される定数を目指す。(これを我々は多項式的定数と呼ぶことにする。)多項式的定数時間アルゴリズムの開発には、いわゆる「力づく」の総当たりは、例えば定数個の要素に対してでも許されない。例えば極大マッチングを求める定数時間検査アルゴリズム[Nguyen & Onak, FOCS08]においては、無作為に選ん

だ枝から乱歩によって周辺を検索する技法を用いているが、この方法では指数定数になるのは免れない。我々は、これに対し、枝の検索順序を工夫することで、多項式的定数時間が達成できるのでは無いかと考えている。

(実際計算機によるシミュレーションでは、このことは確認できている。) 定数時間アルゴリズムは必然的に乱数を使ったものになり、評価も確率的なものになるため、この様に計算機シミュレーションによって結果を予測することが時に重要である。

また、線形下界値については、古典的計算量理論で用いられてきた多項式時間帰着に代表される「帰着」の概念を用いることが有力である。すなわち、なんらかの帰着手法を開発し、それに基づいて、一つのクラスを定義するというアプローチである。

4. 研究成果

主な成果は以下の通りである。

(I) 定数時間アルゴリズム

主にグラフの次数制限モデル上の定数時間アルゴリズムについて研究をし、以下の成果を得た。

(1) 最大独立集合の定数時間近似

グラフの節点を一様ランダム抽出し、その周りを探索することによって独立集合を求めるアルゴリズムについて、探索方法をランダム探索では無く、最良値探索を導入することで、従来のもより高速なアルゴリズムを得た。具体的には、節点数 n 、次数上限 d のグラフに関して、極大マッチングの $(1, \epsilon n)$ 近似を $O(d^4/\epsilon^2)$ 時間で得ることができる。これは 2008 年の FOCS で Nguyen と Onak の与えた算法の計算時間が $O(2^d/\epsilon^2)$ 時間と、 d に関して指数だったのを多項式にした点で画期的である。このアルゴリズムを用いることで、最小節点被覆の $(2, \epsilon n)$ 近似を同じ計算時間で得ることができるほか、さらに本アルゴリズムをサブルーチンとして用いることで、最大マッチングの $(1, \epsilon n)$ 近似を $d^O(1/\epsilon^2) (1/\epsilon)^{O(1/\epsilon)}$ 時間で求めることができる。これは上記の従来算法では $2^d O(1/\epsilon)$ 時間であったものをやはり指数的に改善している。なお、このアルゴリズムの改良自体は単純であるが、その計算時間を正しく評価するのは難解であった。この結果は STOC2009[14] に採録され、高く評価されている。

(2) グラフの外平面性の検査

グラフが外平面グラフであるか否かを検査する $O(\epsilon^{-13} + d\epsilon^{-2})$ 時間のアルゴリズムを提案した。このアルゴリズムは与えられたグラフが外平面性を持つなら

ば確率 $2/3$ 以上で受理し、外平面性より遠ければ (ϵ 遠隔ならば) 確率 $2/3$ で拒否するという確率的アルゴリズムである (従って、その遠さを示すパラメータ $0 < \epsilon \leq 1$ が計算時間に入る)。重要な点は、この計算時間はグラフのサイズとは無関係だという点である。すなわち、いかに巨大なグラフが与えられても常に同じ計算時間で結果を出すという定数時間アルゴリズムであり、ウェブグラフなど巨大なグラフに対して有効である。さらにこのアルゴリズムを修正することで、グラフが (外平面グラフの部分クラスである) カクタスであるか否かの検査も同じ計算時間で可能である [12]。

(3) 性質検査の下界値

この分野では有用な帰着法の開発が研究分野の進捗に対する重要な鍵となるが、我々はクエリ数を線形の範囲で保存する 2 種類の新たな帰着法である「強間隔保存局所帰着 (strong gap-preserving local reduction)」と「強 L 帰着 (strong L-reduction)」を提案し、それを用いて、「3 辺彩色」「有向 (無向) ハミルトン閉路」「3 次元マッチング」「Schaefer の 3SAT の諸問題」の各問の次数制限版での検査がすべて線形数のクエリを必要とすることを証明した [10]。

(4) 疎性マトロイドのランクの定数時間近似

グラフ $G=(V, E)$ は空でない任意の辺部分集合 $F \subseteq E$ が $|F| \leq k|V(F)| - \ell$ を満たすとき (k, ℓ) -sparse と言われる (ただし $V(F)$ は F に属する辺に接続する頂点の集合)。 (k, ℓ) -sparse な辺集合を独立集合と看做することでマトロイドが定義できる。これを G 上の疎性マトロイド (sparsity matroid) と呼ぶ。疎性マトロイドは、連結性、剛性、木のパッキングなど様々な重要な問題の共通の一般化であり、極めて重要な概念である。本研究において、疎性マトロイドのランクの定数時間近似アルゴリズムを与えた。このような一般性の高いマトロイドに対する定数時間近似アルゴリズムはこれまでは無く、定数時間アルゴリズムに対する極めて重要な一歩である。本結果は理論計算機科学の欧州最高の会議である ICALP2012 (7 月, Coventry, UK) に採録が決定している [1]。

(5) ナップザック問題の定数時間近似

ナップザック問題に対する ϵ 近似アルゴリズムは過去にも見つかっているが、線形より少ない時間でのアルゴリズムはこれまでは無かった。しかも過去に得られている NP 完全問題に対する定数時間アルゴリズムはすべてグラフや論理式の次数を制限するなど、制限付きのものであったが、本アルゴリ

ズムは一般のナップザック問題に対して ϵ 近似を定数時間で与えている [2]。

(II) ユニットディスクグラフ上の FPT アルゴリズム

ユニットディスクグラフの描画範囲を面積 α の正方領域に収まっている場合、 α をパラメータとしたパラメータ化計算量についても研究し、ハミルトン閉路問題と k 彩色問題が FPT (つまりパラメータが小さいときには簡単に解ける問題) であり、最大独立集合問題と最小支配集合問題とが $W[1]$ 困難であることを示した。なお、これらの問題はパラメータ α を与えなければ、ユニットディスクに限定しても NP 困難であることが分かっている。この結果は、アドホックネットワークなどの比較的狭い領域に限定した無線ネットワーク上では、ハミルトン閉路問題や k 彩色問題などは解きやすいということを意味している [13]。

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 15 件)

- (1) ITO Hiro TANIGAWA Shin-ichi, YOSHIDA Yuichi, Constant-time algorithms for sparsity matroids, The 39th International Colloquium on Automata, Language and Programming (ICALP 2012), The University of Warwick, Coventry, UK, July 9--13, 2012. (to appear)
- (2) ITO Hiro, KIYOSHIMA Susumu, YOSHIDA Yuichi, Constant-time approximation algorithms for the knapsack problem, The 9th Annual Conference on Theory and Applications of Models of Computation (TAMC 2012), LNCS, #7287, 2012, pp. 131-142.
- (3) YOSHIDA Yuichi and ITO Hiro, Property testing on k -vertex connectivity of graphs, Algorithmica, Vol. 62, No. 3--4, 2012, pp. 701--712.
- (4) ITO Hiro, TERUYAMA Junichi, and YOSHIDA Yuichi, An Almost Optimal Algorithm for Winkler's Sorting Pairs in Bins, Progress in Informatics, No. 9, 2012, pp. 3--7.
- (5) Wolfgang BEIN, HATTA Naoki, Nelson HERNANDEZ-CONS, ITO Hiro, KASAHARA Shoji, and KAWAHARA Jun, An online algorithm optimally self-tuning to congestion for power management problems, Proceedings of the 9th Workshop on Approximation and Online Algorithms (WAOA 2011), LNCS, Springer, #7164, pp. 35--48, 2012.
- (6) AKIYAMA Jin, ITO Hiro, KOBAYASHI Midori and NAKAMURA Gisaku, Arrangements of n points whose incident-line-numbers are at most $n/2$, Special Issue of JCCGG2009, Graphs and Combinatorics, Springer, Vol. 27, No. 3, 2011, pp. 321--326.
- (7) ITO Hiro and TAKATA Satoshi, PSPACE-completeness of the weighted poset game, Proceedings of the 10th International Symposium on Operations Research and Its Applications (ISORA 2011), Lecture Notes in Operations Research, #14, pp. 89--93, WORLD PUBLISHING CORPORATION, 2011.
- (8) Erik D. DEMAINE, Martin L. DEMAINE, Andrea HAWKSLEY, ITO Hiro, Po-Ru LOH, Shelly MANBER and Omari STEPHENS, Making polygons by simple folds and one straight cut, Proceedings of the China-Japan Joint Conference on Computational Geometry, Graphs and Applications (CGGA 2010), LNCS, Springer, # 7033, pp. 27--43, 2011.
- (9) UMESATO Takuya, SAITOH Toshiki, UEHARA Ryuhei, and ITO Hiro, Complexity of the stamp folding problem, Proceedings of The 5th Annual International Conference on Combinatorial Optimization and Applications (COCOA 2011), LNCS, #6831, Springer, 2011, pp. 311--321.
- (10) YOSHIDA Yuichi and ITO Hiro, Query-number preserving reductions and linear lower bounds for testing, IEICE Transactions, Vol. E93-D, No. 2, 2010, pp. 233--240.
- (11) YOSHIDA Yuichi and ITO Hiro, Testing k -edge-connectivity of digraphs, Journal of System Science and Complexity, Vol. 23, 2010, pp. 91--101.
- (12) YOSHIDA Yuichi and ITO Hiro, Testing outerplanarity of bounded degree graphs, Proceedings of the 14th Intl. Workshop on Randomization and Computation (RANDOM 2010), LNCS, #6302, 2010, pp. 642--655.
- (13) ITO Hiro and KADOSHITA Masakazu, Tractability and intractability of problems on unit disk graphs parameterized by domain area, Proceedings of the 9th International Symposium on Operations Research and Its Applications (ISORA 2010), Lecture Notes in Operations Research,

- #12, pp. 120--127, WORLD PUBLISHING CORPORATION, 2010.
- (14) YOSHIDA Yuichi, YAMAMOTO Masaki, and ITO Hiro, An improved constant-time approximation algorithm for maximum matchings, Proceedings of the 41st ACM Symposium on Theory of Computing (STOC 2009), pp. 225--234.
- (15) ITO Hiro and IWAMA Kazuo, Enumeration of isolated cliques and pseudo-cliques, ACM Transactions on Algorithms, Vol. 5, Issue 4, Oct. 2009, Article 40.

[学会発表] (計 12 件)

- (1) ITO Hiro, Constant-time algorithms and fundamental theorems, Oct. 21, 2011, Colloquium, Dept. Comput. Science, Univ. Nevada, Las Vegas.
- (2) 伊藤大雄, 定数時間アルゴリズムとその基本定理 (前編) --- 密グラフモデルと正則性補題, ERATO セミナー, 2011/08/23, 15:30--17:30, 北大.
- (3) 伊藤大雄, 楠本充, 定数時間アルゴリズムと BDD, ERATO 湊離散構造処理系プロジェクト 2011 年度 初夏のワークショップ, 北大, 2011/06/10--11.
- (4) ITO Hiro, TANIGAWA Shin-ichi, YOSHIDA Yuichi, Testing algorithms for (k, l) -sparsity and (k, l) -edge-connected-orientability, The 7th Hungarian-Japanese Symposium on Discrete Mathematics and Its Applications, 京都大学 数理解析研究所, 2011/05/31--06/03.
- (5) W. BEIN, HATTA Naoki, ITO Hiro, KAWAHARA Jun, An Online Algorithm Optimally Self-tuning to Congestion for Power Management Problems, 信学技報, Vol. 111, No. 20, COM2011-11 (2011-4), pp. 29--36 (電子情報通信学会 コンピューテーション研究会 (COMP), 京都大学, 2011/04/22.)
- (6) 伊藤大雄, 清島 奨, 吉田悠一, ナップサック問題に対する定数時間近似アルゴリズム, 信学技報, Vol. 110, No. 464, COM2010-51 (2011-3), pp. 29--36 (電子情報通信学会 コンピューテーション研究会 (COMP), 中央大学 後楽園キャンパス, 2011/03/09.)
- (7) 伊藤大雄, 大規模ネットワーク問題のための定数時間アルゴリズム, 甲南大学 知的情報通信研究所 シンポジウム「ネットワークと情報処理」, 甲南大学, 2009/12/26.
- (8) 伊藤大雄, 無秩序の代償 (price of

anarchy) の理論への招待, FIT 2009, イベント企画「アルゴリズム・サイエンスにおけるゲーム理論の新展開」, 東北工業大学 八木山キャンパス, 2009/09/02--04.

- (9) ITO Hiro, TERUYAMA Junichi, and YOSHIDA Yuichi: An almost optimal algorithm for Winkler's sorting pairs in bins, AAAC2010, 浦項 (pohang), 韓国, 2010/04/17--19.
- (10) ITO Hiro, TERUYAMA Junichi, and YOSHIDA Yuichi: An almost optimal algorithm for Winkler's sorting pairs in bins, 信学技報, COM2009-45 (2010-01), pp. 45--49 (電子情報通信学会 コンピューテーション研究会 (COMP), 九州大学西新プラザ, 2010/01/25) .
- (11) YOSHIDA Yuichi, YAMAMOTO Masaki, and ITO Hiro, Constant-time approximations using minimum value search for independent sets and matchings, The 2nd Annual Meeting of Asian Association for Algorithms and Computation (AAAC 2009), Hangzhou (杭州), 中国, April 11-12, 2009.
- (12) 吉田悠一, 山本真基, 伊藤大雄, 最大独立集合と最大マッチングに対する定数時間近似アルゴリズムの改善, 電子情報通信学会総合大会, 愛媛大 (松山市), 2009/03/17--20.

[図書] (計 1 件)

- (1) 伊藤大雄・宇野裕之 編著, 「離散数学のすすめ」, 現代数学社, 2010 年 5 月 15 日 刊行.

6. 研究組織

(1) 研究代表者

伊藤 大雄 (ITO HIRO)

京都大学・大学院情報学研究科・准教授
研究者番号：50283487

(2) 研究分担者

岩間 一雄 (IWAMA KAZUO)

京都大学・大学院情報学研究科・教授
研究者番号：50131272

(3) 連携研究者

なし