

平成 22 年 5 月 30 日現在

研究種目：基盤研究 (C)
 研究期間：2007～2009
 課題番号：19500010
 研究課題名 (和文) 最大クリーク抽出アルゴリズムの効率化・拡張と計算量評価および応用
 研究課題名 (英文) Improvement and extension of maximum-clique-finding algorithms with complexity analysis and their applications
 研究代表者
 富田 悦次 (TOMITA ETSUJI)
 電気通信大学・名誉教授
 研究者番号：40016598

研究成果の概要 (和文)：最大クリークを抽出する新しいアルゴリズム MCS を開発し、格段に高速であることを明らかにした。これにより、従来では 100 日以上かかっても解けなかった幾つかの問題を 100 秒以内で解くことに成功した。最大クリーク問題が多項式時間的に可解となる基本的結果も確立した。また、最大クリーク抽出アルゴリズムがハイパーグラフにおいても効率的に稼働する様に拡張した。更に、これらのアルゴリズムをデータマイニングなどの実問題に応用して有効な結果を得た。

研究成果の概要 (英文)： We have developed a remarkably fast algorithm MCS for finding a maximum clique. It is confirmed that MCS can solve some problems in less than 100 seconds which other algorithms require more than 100 days to solve. We have established a basic condition under which the maximum clique problem can be solved in polynomial time. An efficient algorithm is devised for finding a maximum clique in a hyper-graph. Successful applications are also obtained for data mining and others.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2007 年度	1,100,000	330,000	1,430,000
2008 年度	1,000,000	300,000	1,300,000
2009 年度	1,300,000	390,000	1,690,000
年度			
年度			
総計	3,400,000	1,020,000	4,420,000

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：最大クリーク，極大クリーク，分枝限定法，並列処理，NP 完全問題，時間計算量，多項式時間的解可能性，データマイニング

1. 研究開始当初の背景

最大クリーク抽出問題は、NP 困難な難しい問題であるが、そのアルゴリズムの効率化は着実に進められてきた。特に近年、バイオインフォマティクス、画像処理、クラスタリング、データマイニング、薬剤設計 など多

くの実問題への応用が可能であることが明らかにされ、その結果最大クリーク関係のアルゴリズムの更なる効率化が一層強く要請されている。

これらのアルゴリズムは、Rutgers 大学、Princeton 大学等の運営による DIMACS

(Discrete Mathematics and Theoretical Computer Science)プロジェクトによるベンチマーク問題と基準プログラム dfmax を通して、国際的推進・評価がなされている。

ここにおいて、我々は新しい最大クリーク抽出アルゴリズム **MCR** を開発し、発表している。本アルゴリズム **MCR** は従来の他アルゴリズムに対してほぼ全面的に**格段の優位性**が確認されており、当論文に対しては国内外から幾つもの問い合わせを受けており、またその新しい応用結果も発表されている。

最大クリーク抽出を拡張した問題として、極大クリークを全列挙する問題があるが、これもデータマイニングなどの実際の問題を解くための基本となるため、大きい注目を浴びている。このためのアルゴリズムとして、我々は新しいアプローチによるアルゴリズム **CLIQUES** を下記論文に発表し、先ず、それが **理論的に最適** な時間計算量を達成していることを証明した。

[CLIQUES] E. Tomita, A. Tanaka, H. Takahashi : The worst-case time complexity for generating all maximal cliques and computational experiments, (Invited paper for the special issue on COCOON 2004), Theoretical Computer Science, 363, pp.28-42 (2006). (当論文は、TOP25 Hottest Articles within the journal: Theoretical Computer Science の 2010年January-March期の16位にランクされた。)

本アルゴリズムは理論的に最適であるだけでなく、実際にも広範囲のグラフに対して従来のアルゴリズムよりも非常に高速であることを確認しており、本アルゴリズムも広く注目を集めている。

以上の様に、これまでに理論的にも実験的にも着実な成果を上げてきた。

2. 研究の目的

前記の様にこれまでに蓄積されてきた実績を最大限活用して、最大クリーク抽出アルゴリズムを一層効率化すると共に、それに対する実験的および理論的に明確な優位性評価を与える。更にそれを拡張し、より広い応用への有効性を明確にし、本研究に先立つ科学研究費基盤研究 (B)「効率的な組合せ最適化アルゴリズムの開発と応用」の成果を大きく発展させることを本研究の目的とする。

3. 研究の方法

先ず、先行研究成果である最大クリーク抽出アルゴリズム **MCR** において最も重要な要

素である分枝限定効果を徹底的に検討・評価し、それを強化することにより、一層効率的な最大クリーク抽出アルゴリズムを開発する。それを基にして、拡張したグラフにおける効率的な最大クリーク抽出アルゴリズムも開発する。

これらのアルゴリズムは、大量のデータに対する計算機実験により、その優位性を明確に示す。また、理論的な評価も出来る限り明らかにする。

これらのアルゴリズムを用いて、データマイニング等の具体的な応用にも適用し、その有効性を明らかにする。

4. 研究成果

[1] 最大クリーク抽出アルゴリズムの効率化

(1-1) 分枝限定アルゴリズムの効率化

最大クリーク抽出のための分枝限定アルゴリズムにおいて最も重要である分枝限定のための近似彩色法を改良し、更に節点の整理を適切化することにより、従来よりも大幅に効率的な最大クリーク抽出アルゴリズム **MCS** を開発し、計算機実験によりその高速性を確認した。この結果、従来アルゴリズムに対して、対象グラフによっては10倍、100倍以上に高速であり、特に枝が超高密度であるランダムグラフにおいては10万倍以上の高速化を達成していることを確認した。また、枝が疎である大規模グラフに対しても、基本アルゴリズム dfmax に対して初めて2倍以上の高速化達成に成功した。

アルゴリズム **MCS** の実行例を表1に示す。

表1 実行時間比較 [sec]

Graph	dfmax	New	MCR	MCS
r200.8	193	147	12.3	4.5
r200.98	>10 ⁵		30.9	0.2
r300.98			284,534	2,623
r500.994	>1.5·10 ⁷		>10 ⁷	39
brock400_4	10,633		639	248
MANN_a27	>10 ⁵	>2,232	2.5	0.8
p_hat500-2	133	96	3.1	0.7
p_hat500-3	>10 ⁵		1,788	150
san200_0.9_3	42,648		0.16	0.06
san400_0.7_2	>10 ⁵	113	0.3	0.1
san400_0.9_1	>10 ⁵		3.4	0.1
gen400_p0.9_55			5,846,951	58,431

[dfmax] D.S. Johnson, M.A. Trick (eds.): Cliques, Coloring, and Satisfiability, DIMACS Series in Discrete Mathematics

and Theoretical Computer Science, 26, American Mathematical Society (1996)

[New] P. R. J. Östergård: A fast algorithm for the maximum clique problem, Discrete Applied Mathematics, 120, pp.197-207 (2002).

[MCR] E. Tomita, T. Kameda: An efficient branch-and-bound algorithm for finding a maximum clique with computational experiments, Journal of Global Optimization, 37, pp.95-111 (2007).

[MCS] E. Tomita, Y. Sutani, T. Higashi, S. Takahashi, M. Wakatsuki: A simple and faster branch-and-bound algorithm for finding a maximum clique, Lecture Notes in Computer Science, 5942, pp.191-203 (2010)

(1-2) 並列処理

実働的に最大クリーク抽出をより高速化する一つの手法として、共有メモリ型並列計算機上において並列処理を実行する新しい手法を開発した。これにより、8 CPU 計算機による実験結果として、多くの問題に対して1CPU 計算機におけるよりもほぼ8倍に近い高速化を達成できることを明らかにした。更に、32CPU 共有メモリ計算機においても、ほぼ CPU 数比例に近い高速化が達成出来ることの基礎的な実験的確認結果も得た

[2] 最大クリーク抽出アルゴリズムの理論的時間計算量評価

(2-1) 一般グラフに対する最大クリーク抽出問題の時間計算量評価

最大クリーク抽出問題の理論的時間計算量評価の改良を行い、従来の他の研究結果と比べて大幅に改善した評価を保証するアルゴリズムと評価結果を得た。また、その時間計算量の評価結果は理論的に良いだけでなく、実働上においても優れていることを確認し、理論的結果と実験的結果の良好性の融合を達成した。

ただし、時間計算量解析には膨大な場合分けと非常に複雑な解析を必要としており、その単純化は今後の課題として残されているが、その基礎は確立することが出来た。

(2-2)最大クリーク問題の多項式時間的
可解性

与えられたグラフの最大次数が節点数の対数オーダーである場合には、最大クリーク問題を多項式時間で解くことができる非常に単純な分枝限定アルゴリズムと、その計算量を証明する単純な解析法を与えた。これは、典型的な NP 困難問題である最大クリーク問題を多項式時間的に可解とする極めて自然な結果であり、他の NP 困難問題の多項式時間的
可解条件の基本とすることができる。

[3] 最大クリーク抽出アルゴリズムの拡張

前記[1]と同様の手法を、より高度な構造を有するハイパーグラフに対しても活用出来る様に、ハイパーグラフ独自の特徴に基づいた考察を行うことにより拡張し、ハイパーグラフ中の最大クリークを従来よりも大幅に効率良く抽出するアルゴリズムを開発した。これにより、DNA コンピューティングのための配列設計の基礎をより強力とした。

[4] 極大クリーク抽出アルゴリズムの応用

(4-1)企業間関係の構造分析

極大クリーク全列挙アルゴリズムをデータマイニングに活用する一つの結果として、企業間の関係の構造解析を行い、企業間関係の構造をより全体的な特徴から分析する新たな手法を提唱し、その有効性を確認した。

(4-2) バイオ情報のデータマイニング

これまでの研究成果として得ていた極大クリーク全列挙法を基礎として、バイオ医療データベースにおけるデータマイニング法を開発し、メタボリック症候群の肥満、高血圧、糖尿病、高脂血症に
関係する遺伝子間の関係を自動抽出し、有意義な関係を明らかにすることに成功した。

[5] その他

上記に関連して、下記 5. 主な発表論文等に示した様に、画像処理、量子計算、学習アルゴリズム、計算量解析についての有効な研究成果を得た。

5. 主な発表論文等

[雑誌論文] (計 26 件)

1. 中西裕陽, 富田悦次: 最大クリーク問題の多項式時間的
可解性の一結果, 電子情報通信学会論文誌 D, J93-D, 査読有 (採録決定).
2. J. Tarui: Smallest formulas for the parity of 2^k variables are essentially unique, Theoretical Computer Science, 査読有 (採録決定).
3. K. Amano, J. Tarui: A well-mixed function with circuit complexity $5n$: Tightness of the Lachish-Raz-type bounds, Theoretical Computer Science, 査読有 (採録決定).
4. E. Tomita, Y. Sutani, T. Higashi, S. Takahashi, M. Wakatsuki: A simple and faster branch-and-bound algorithm for finding a maximum clique, Lecture Notes in Computer Science, 5942, pp.191-203, 査読有 (2010)
5. T. Matsunaga, C. Yonemori, E. Tomita, M. Muramatsu: Clique-based data mining for related genes in a

- biomedical database, BMC Bioinformatics, 査読有, Online Journal (2009)
6. 米森力, 松永務, 関根純, 富田悦次: クリークを用いた企業間関係の構造分析, 日本データベース学会論文誌, 7, pp.55-60, 査読有 (2009)
 7. 富田悦次: 先進的アルゴリズムに向けて (招待論文), 電気通信大学紀要, 21, 査読無 (2009)
 8. R. Debnath, M. Muramatsu, H. Takahashi: Implementation issues of second-order cone programming approaches for support vector machine learning problems, IEICE Trans. Fundamentals, E92-A, pp. 1209-1222, 査読有 (2009)
 9. K. Kawabata, S. Morishita, H. Takemura, K. Hotta, T. Mishima, H. Asama, H. Mizoguchi, H. Takahashi: Development of an automated microscope for supporting qualitative asbestos analysis by dispersion staining, Journal of Robotics and Mechatronics, 21, 査読有 (2009)
 10. 鈴山有紀, 堀田一弘, 高橋治久: 文脈情報に基づく対象が存在する事前確率の推定, 電気学会論文誌, 129-C, pp.832-837. 査読有 (2009)
 11. J.O. Mercado, K.Hotta, H. Takahashi, M.N. Miyatake, K.T. Medina, H.P. Meana: Improveing the eigenphase method for face recognition, IEICE Electronics Express, 6, pp.1112-1117, 査読有 (2009)
 12. Y. Kakishita, K. Sasahara, T. Nishino, M. Takahasi, K. Okanoya: Ethological data mining: An automata-based approach to extract behavioral units and rules, Data Mining and Knowledge Discovery, 18, pp. 446-471, 査読有 (2009)
 13. H. Morizumi, and J. Tarui: Linear-size log-depth negation-limited inverter for k -tonic binary sequences, Theoretical Computer Science, 410, pp. 1054-1060, 査読有 (2009)
 14. K. Iwama, H. Morizumi, J. Tarui: Negation-limited complexity of parity and inverters, Algorithmica, 54, pp. 256-267, 査読有 (2009)
 15. M. Wakatsuki, E. Tomita: Polynomial time identification of strict deterministic restricted one-counter automata in some class from positive data, IEICE Trans. on Inform. Syst., E91-D, pp.1704-1718, 査読有 (2008)
 16. 中西裕陽, 富田悦次: 最大クリーク問題の多項式時間的可解性 —基本的結果—, FIT2008, pp.1-4, 査読有 (2008)
 17. 清野和司, 富田悦次, 若月光夫: 実時間空スタック受理式決定性限定ワンカウンタ変換器対の多項式時間等価性判定, 電子情報通信学会論文誌 D, J91-D, pp.1188-1201, 査読有 (2008)
 18. 関口涼平, 高橋治久, 堀田一弘: カーネル主成分分析を用いた学習機械のパラメータ自動決定法, 情報処理学会論文誌: 数理モデル化と応用, 49, pp79-87, 査読有 (2008).
 19. 本多武尊, 山崎匡, 田中繁, 西野哲朗: 小脳顆粒層をモデル化したスパイクングネットワークの研究: NMDA受容体を介した同期発火状態と時間表現状態の遷移, 電子情報通信学会論文誌 D, J91-D, pp. 1-10, 査読有(2008).
 20. J. Tarui: On the minimum number of completely 3-scrambling permutations, Discrete Mathematics, 308, pp.1350-1354, 査読有 (2008)
 21. K. Iwama, H. Morizumi, J. Tarui: Reductions for monotone Boolean circuits, Theoretical Computer Science, 408, pp. 208-211, 査読有 (2008)
 22. E. Tomita, T. Kameda: An efficient branch-and-bound algorithm for finding a maximum clique with computational experiments, Journal of Global Optimization, 37, pp.95-111, 査読有 (2007)
 23. H. U. Simon, E. Tomita(Eds.): Guest editors' foreword, Theoretical Computer Science, 387, pp.1-3, 査読無 (2007)
 24. 清野和司, 富田悦次, 若月光夫: ε -推移を許したある決定性プッシュダウン変換器対の等価性判定, 電子情報通信学会論文誌 D, J90-D, pp.2675-2690, 査読有 (2007)
 25. T. Nishino: How to derive a quantum complexity lower bound, Electronics and Communications in Japan, Part 3, 90, pp.9-17, 査読有 (2007)
 26. 眞鍋秀聡, 西野哲朗, 山崎匡, 田中繁: インターナルクロックモデルに基づくロボット制御法の実現, 情報処理学会論文誌:数理モデル化と応用, 48, pp. 139-154, 査読有 (2007)
- [学会発表] (計 82 件)
1. 中西裕陽, 富田悦次, 若月光夫: 最大クリーク抽出の単純な最大時間計算量と多項式時間的可解性, 冬の LA シンポジウム, 2010年2月3日, 京都大学.
 2. 若月光夫, 清野和司, 富田悦次, 西野哲朗: 空スタック受理式決定性限定ワンカウンタ変換器の多項式時間等価性判定アルゴリズム, 冬の LA シンポジウム, 2010年2月2日, 京都大学.
 3. E. Tomita, Y. Sutani, T. Higashi, S. Takahashi, M. Wakatsuki: A simple

- and faster algorithm for finding a maximum clique, 情報処理学会アルゴリズム研究会, 2010年1月26日, 九州大学.
4. 中西裕陽, 富田悦次, 若月光夫: 最大クリーク抽出の単純な最大時間計算量評価, 夏のLAシンポジウム, 2009年7月22日, かんぼの宿 松島.
 5. E. Tomita, H. Nakanishi: Polynomial-time solvability of the maximum clique problem, Proc. European Computing Conference 2009, pp.203-208, 査読有, 2009年6月26日, Tbilisi, Georgia.
 6. 富田悦次: 組合せ最適化問題の新展開 - 最大クリーク抽出問題の高速化とその応用 -, 緊急地震速報システムの利活用研究会, 2009年3月27日, 晴海トリトンスクウェア.
 7. 若月光夫, 富田悦次, : 準同型写像によって拡張されたある言語クラスに対する正例からの極限同定, 情報処理学会数理モデル化と問題解決研究会, 2009年3月5日, 沖縄科学技術研究基盤機構.
 8. 若月光夫, 高橋真也, 富田悦次: 最大クリーク抽出アルゴリズムの共有メモリ型並列計算機上での並列化, 情報処理学会数理モデル化と問題解決研究会, 2008年9月18日, 電気通信大学.
 9. 中西裕陽, 富田悦次: 最大クリーク抽出問題の理論計算量評価について-グラフの次数を限定した場合, 情報処理学会アルゴリズム研究会, 2008年9月12日, 名古屋工業大学.
 10. 富田悦次: 最大クリーク抽出問題とその応用, 北海道大学大学院情報科学研究科コンピュータサイエンス専攻 離散構造とアルゴリズム研究会, 2008年8月28日, 千歳市支笏湖丸駒温泉.
 11. 中西裕陽, 富田悦次: 最大クリーク抽出問題の理論計算量評価について-多項式時間計算機量の場合, 夏のLAシンポジウム, 2008年7月23日, 南紀勝浦国民休暇村.
 12. 中西裕陽, 富田悦次: 最大クリークを抽出する時間計算量 $O(2^{0.1912^n})$ の多項式領域アルゴリズム, 情報処理学会アルゴリズム研究会, 2008年1月23日, 法政大学.
 13. E. Tomita, : The maximum clique problem and its applications (Invited Lecture), 情報処理学会数理モデル化と問題解決研究会/バイオ情報学研究会, 2007年12月20日, 産総研生命情報工学研究センター.
 14. 玉田和洋, 富田悦次, 中西裕陽: 理論評価付き最大クリーク抽出アルゴリズムの実験的評価, 情報処理学会数理モデル化と問題解決研究会/バイオ情報学研究会, 2007年12月20日, 産総研生命情報工学研究センター.
 15. 米森力, 松永務, 富田悦次: 極大クリーク全列挙アルゴリズムを用いた企業コミュニティ理解, 情報処理学会数理モデル化と問題解決研究会/バイオ情報学研究会, 2007年12月20日, 産総研生命情報工学研究センター.
 16. S. Urabe, E. Tomita: NetMCQ: A distributed exact maximum clique solver, 情報処理学会数理モデル化と問題解決研究会/バイオ情報学研究会, 2007年12月20日, 産総研生命情報工学研究センター.
 17. 中西裕陽, 富田悦次: 最大クリークを抽出する時間計算量 $O(2^{0.19669^n})$ の多項式領域アルゴリズム, 2007年11月30日, 新潟大学.
 18. 中西裕陽, 富田悦次: 最大クリークを抽出する時間計算量 $O(2^{0.2495^n})$ の多項式領域アルゴリズム, 2007年10月29日, 東北大学.
 19. E. Tomita: Advanced algorithms and their applications, Advanced Information and Communication Technology, 2007年9月20日, Beijing, China.
 20. M. Wakatsuki, E. Tomita: Polynomial time identification of finite state transducers in some class, Advanced Information and Communication Technology, 2007年9月20日, Beijing, China.
 21. 須谷洋一*, 富田悦次, 小林聡: ハイパーグラフを対象とした最大クリーク抽出の分枝限定アルゴリズム, 情報処理学会数理モデル化と問題解決研究会, 2007年9月4日, 名古屋大学 (*情報処理学会山下記念研究賞受賞).
 22. 中西裕陽, 富田悦次: 最大クリークを抽出する単純なアルゴリズムの最大次数4のグラフにおける計算量, 電子情報通信学会コンピュータシミュレーション研究会, 2007年6月29日, 北海道大学.
 23. E. Tomita, Y. Sutani, T. Higashi: A more efficient algorithm for finding a maximum clique with an improved approximate coloring, Proc. PDPTA 2007, pp.719-725, 査読有, 2007年6月25日, Las Vegas, USA.
- [図書] (計7件)
1. E. Tomita, T. Akutsu, T. Matsunaga, "Efficient algorithms for finding maximum and maximal cliques - Effective tools for bioinformatics - " in "Biomedical Engineering, Trends, Researches and Technologies,"

- INTECH, Vienna, Austria (to appear).
2. 中森眞理雄, 富田悦次, 数理モデルの応用技術 (情報処理学会 50 周年史, 第 3 編, 第 17 章), 情報処理学会 (出版予定).
 3. 富田悦次, 他, 丸善, 第 2 版 現代数理科学辞典, 2009, 1450 ページ.
 4. 富田悦次, 横森貴, 森北出版, オートマトン・言語理論 (第 21 刷・改訂増刷), 2009, 205 ページ.
 5. 高橋治久, 堀田一弘, コロナ社, 学習理論, 2009, 215 ページ.
 6. 西野哲朗, 井元信之(監訳), 共立出版, 量子情報の物理-量子暗号, 量子テレポーテーション, 量子計算-, 2007, 400 ページ.
 7. 西野哲朗, ナツメ社, 図解雑学・量子コンピュータ, 2007, 216 ページ.

[その他]

ホームページ等

http://www.etlab.ice.uec.ac.jp/~tomita/index_e.html

6. 研究組織

(1) 研究代表者

富田 悦次 (TOMITA ETSUJI)
電気通信大学・名誉教授
研究者番号 : 40016598

(2) 研究分担者

高橋 治久 (TAKAHASHI HARUHISA)
電気通信大学・電気通信学部・教授
研究者番号 : 90135418

西野 哲朗 (NISHINO TETSURO)
電気通信大学・電気通信学部・教授
研究者番号 : 10198484

若月 光夫 (WAKATSUKI MITSUO)
電気通信大学・電気通信学部・助教
研究者番号 : 30251705

垂井 淳 (TARUI JUN) (2008-2009 年度)
電気通信大学・電気通信学部・准教授
研究者番号 : 00260539