

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6月10日現在

機関番号：32657

研究種目：基盤研究（C）

研究期間：2008～2011

課題番号：20500021

研究課題名（和文） 計算困難な問題への混成アプローチ：近似、乱択、並列化

研究課題名（英文） Hybrid Approaches to Computationally Hard Problems: Approximation, Randomization, and Parallelization

研究代表者

陳 致中（CHEN ZHI-ZHONG）

東京電機大学・理工学部・教授

研究者番号：00242933

研究成果の概要（和文）：何らかの利益を最大化する問題や何らかのコストを最小化する問題は実世界に遍在している。このような問題は最適化問題と呼ばれ、そのほとんどが計算困難である。計算困難な最適化問題へのアプローチとして、近似（最適な解に近い解を効率よく求めること）という手法がよく使われている。本研究でこのアプローチに乱択・並列化という2つの手法を融合して、いくつかの計算困難な最適化問題に対して既知のアルゴリズムより性能の良いアルゴリズムを設計した。

研究成果の概要（英文）：Problems that aim to maximize certain kinds of profit or minimize certain kinds of cost are ubiquitous in the real world. Such problems are called optimization problems and most of them are computationally hard. A well-used approach to hard optimization problems is approximation (which aims to find near-optimal solutions efficiently). In this project, we combine this approach with two other powerful approaches (namely, randomization and parallelization) and apply the resulting hybrid approach to several hard optimization problems. As the results, we have obtained more efficient algorithms for the hard problems than the previous bests.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
2008年度	500,000	150,000	650,000
2009年度	500,000	150,000	650,000
2010年度	500,000	150,000	650,000
2011年度	500,000	150,000	650,000
年度			
総計	2,000,000	600,000	2,600,000

研究分野：総合領域

科研費の分科・細目：情報学、情報学・情報学基礎

キーワード：近似アルゴリズム、確率アルゴリズム、並列アルゴリズム、NP 困難性

1. 研究開始当初の背景

計算困難な問題へのアプローチとして、近似（もっともほしい解に極めて近い解を求める方法）と乱択（乱数発生機能を持った計算機でほしい解を高い確率で求める方法）と並列化（多数のプロセッサを持った計算機でほしい解を極めて速く求める方法）が知られて

いたが、今までの研究の中にこの3つのアプローチを融合して使うものは多くはなかった。研究代表者は近年の研究で、この3つのアプローチを融合した混成アプローチが計算困難な問題の解決に極めて有効であることを明らかにした。この混成アプローチをさらに発展させて、より多くの計算困難な問題

の解決に応用できるかは興味深い課題である。

2. 研究の目的

混成アプローチをさらに発展させ、より多くの重要で計算困難な問題の解決を目指すとともに、計算困難な問題を解くアルゴリズムの新しい設計技法の開発を目指す。具体的には下記の計算困難な最適化問題を混成アプローチで解こうとする。

- (1) 最小補パス集合問題：与えられた無向グラフ G から最小本数の辺を削除して G が互いに交わらないパスの集まりにしたい。この問題は計算困難であるが、遺伝子の染色体における位置を特定するのに使われる放射線ハイブリッド法と関係している。そのため、アルゴリズムがすでに提案されている。本研究でよりよい性能を持つアルゴリズムを設計する。
- (2) 最大 2-辺彩色問題：与えられた無向グラフ G において 2 つの交わらないマッチングを求めてその合計サイズを最大化したい。この問題は、有名な最大マッチング問題と違って計算困難であるが、大規模な衛星に基づく通信ネットワークの構築などに応用がある。そのため、いくつものアルゴリズムが提案されている。本研究でよりよい性能を持つアルゴリズムを設計する。
- (3) 容量付きマルチキャスト・樹形ルーティング問題：与えられたネットワークにおいて、ソースノードからすべての目的ノードにメッセージを送りたい。ただし、毎回のキャストに、予め決められた個数（たとえば、 k 個）の目的ノードにしかメッセージを送信できない。また、毎回のキャストに樹形部分ネットワークを使用しなければならない。これらの制限の下ですべての目的ノードにメッセージを送信するのに複数の樹形部分ネットワークが必要となる。目的は送信用に使われる樹形部分ネットワークの合計コストを最小化することである。この問題は計算困難であるが、WDM 方式光ネットワークにおけるキャストに関係している。以前の研究で、この問題に対する近似アルゴリズムで近似率がそれぞれ $(2 + \rho)$ と $(1.5 + \rho/4)$ のものが設計された。ここで、 ρ は最小スタイナー木問題を解く既知の近似アルゴリズムの最も良い近似率である。本研究でよりよい性能を持つアルゴリズムを設計する。
- (4) 最寄文字列問題： n 個の同じ長さ L の DNA 列または蛋白質列 s_1, s_2, \dots, s_n および自然数 d が与えられたとき、 $s_1,$

s_2, \dots, s_n のどれとのハミング距離も d 以下の文字列を 1 つ求めたい。この問題は計算生物学とバイオインフォマティクスにおいて極めて基本的な問題であり、PCR プライマー設計や遺伝的プローブ設計やアンチセンスドラッグ設計やモチーフ検索などに応用があるが、計算困難である。この問題の重要性から様々なアルゴリズムが設計されてきた。その中に近似・乱択・並列化の 3 つの手法を使ったものも複数含まれている。しかし、実際の応用で近似解よりも厳密解が好ましいことがある。厳密解を求める既知のアルゴリズムはいくつかあるが、そのどれの時間量も d に重く依存しているので実用的ではない。本研究で、厳密解を求めて d への依存度が小さいアルゴリズムを設計する。

- (5) 最寄部分列問題： n 個の長さ L 以上の DNA 列または蛋白質列 s_1, s_2, \dots, s_n および自然数 d と L が与えられたとき、「 s_1, s_2, \dots, s_n のどれも t とのハミング距離が d 以下であるような長さ L の部分列を持つ」ことを満たす長さ L の文字列 t を 1 つ求めたい。この問題は最寄文字列問題の拡張であるので計算困難である。また、最寄文字列問題と並んで計算生物学とバイオインフォマティクスにおいて極めて基本的な問題である。この問題に対して厳密解を求める既知のアルゴリズムはいくつかあるが、そのどれの時間量も d に重く依存しているので実用的ではない。本研究で、厳密解を求めて d への依存度が小さいアルゴリズムを設計する。

3. 研究の方法

基本的に 2 で挙げた計算困難な問題を混成アプローチで解く。

- (1) 最小補パス問題：この問題に対して確率近似アルゴリズムを設計してから脱ランダム化する。また、アルゴリズム設計の際、局所探索法と動的計画法を活用する。その結果、2 つの解（補パス集合）が求められ、サイズのより小さい方を出力するようにする。
- (2) 最大 2-辺彩色問題：この問題に対して既知のアルゴリズムがいくつかあった。その中の 1 つは研究代表者によるもので、近似・乱択・並列化に基づいていた。本研究では、そのアルゴリズムを改善するために、局所探索法と動的計画法と再帰的呼び出しを取り入れる。また、新しいアルゴリズムを解析する際、近似解による誤差を課金方式 (charging scheme) で分散させる手法を用いる。

- (3) 容量付きマルチキャスト・樹形ルーティング問題: この問題に対する既知の近似アルゴリズムの設計に「コスト平均化技法」と「最小スタイナー木のルーティング木への分割技法」が利用された。本研究でもこの2つの技法を使うが、そのまま使うのではなく、うまく拡張したバージョンを使用する。
- (4) 最寄り文字列問題: 既知の厳密アルゴリズムと同じように、どれかの入力文字列(たとえば、 s_1)を選んでその高々 d 文字を変更することによって解を得ることができるかを調べる。現在の s_1 がまだ解になっていなければ、 s_1 とのハミング距離が d より大きい s_i があるはずである。 s_1 と s_i の異なる桁が $2d$ より大きければ解がないことがわかる。さもなければ、 s_1 と s_i の異なる桁が $2d$ 以下であり、その異なる桁から高々 d 桁を選んで変更する方法をすべて試して解を得ることができるかを調べる。以前のアルゴリズムは、選んだ d 以下の桁を変更するとき s_1 におけるそれらの桁の文字を参考にしなかった。本研究では、 s_i におけるそれらの桁の文字を参考にすることによって、より効率的なアルゴリズムを設計する。また、設計したアルゴリズムを OpenMP で並列化する。
- (5) 最寄部分列問題: 定義から分かるように、この問題は最寄文字列問題の拡張である。そのため、最寄部分列問題のアルゴリズムを得るのに、最寄文字列問題のアルゴリズムを拡張すればよい。

4. 研究成果

2 で挙げた各問題に対して下記の成果を上げることができた。

- (1) 最小補パス集合問題: この問題に対して近似率 $10/7$ と計算時間 $O(n^{1.5})$ の近似アルゴリズムを設計した。ここで、 n は入力グラフの頂点数である。以前の近似アルゴリズムは近似率 2 と計算時間 $O(n^2)$ であった。したがって、本研究で提案したアルゴリズムは計算時間と近似率の両面で以前のアルゴリズムに勝っている。この結果より、乱択と近似を融合した混成アプローチはこの問題に対して非常に有効であることがわかった。
- (2) 最大 2-辺彩色問題: この問題に対して近似率 0.842 と計算時間 $O(mn^3)$ の近似アルゴリズムを設計した。ここで、 n と m はそれぞれ、入力グラフの頂点数と辺の本数である。以前の最も良い近似アルゴリズムは近似率 $5/6$ と計算時間 $O(mn^3)$ であった。したがって、本研究で提案したアルゴリズムは以前の最も良いアルゴリズムの計算時間を維持した

ままその近似率を改善したことになる。また、新しいアルゴリズムを解析するのにトレーニンググラフと星状グラフのある種の構造を明らかにした。その構造が他の類似した計算困難な問題の解決にも応用できると思われる。

- (3) 容量付きマルチキャスト・樹形ルーティング問題: この問題に対して、近似率 $(1.6 + \rho/5)/4$ を達成する近似アルゴリズムを設計した。現在知られている ρ の最小値が 1.55 であるので、この新しい近似アルゴリズムは既知の近似アルゴリズムよりいつもよい近似率を達成する。
- (4) 最寄文字列問題: この問題に対して2つの厳密アルゴリズムを設計した。その1つは DNA 列に対して時間量 $O(nL+nd \cdot 17.97^d)$ を達成して、蛋白質列に対して時間量 $O(nL+nd \cdot 61.86^d)$ を達成する。このアルゴリズムは最寄部分列問題に拡張できる。もう1つは DNA 列に対して時間量 $O(nL+nd \cdot 13.92^d)$ を達成して、蛋白質列に対して時間量 $O(nL+nd \cdot 47.21^d)$ を達成する。このアルゴリズムは最寄部分列問題に拡張できない。この新しい2つのアルゴリズムに対し、以前の最も効率的なアルゴリズムは DNA 列に対して時間量 $O(nL+nd \cdot 28.54^d)$ を達成して、蛋白質列に対して時間量 $O(nL+nd \cdot 180.75^d)$ を達成する。この時間量の改善によって、以前よりも広範囲の d について実用的な時間内に厳密解を求めることができるようになった。また、得られたアルゴリズムを OpenMP で並列化して、複数のコアを持ったパソコン上で実行した結果、並列化による実行時間の短縮を確認できた。
- (5) 最寄部分列問題: 最寄文字列問題に対して設計した1つ目のアルゴリズムを拡張することによって、最寄部分列問題を厳密に解くアルゴリズムを得た。このアルゴリズムは DNA 列に対して時間量 $O(nm^2(L+d \cdot 17.97^d \cdot m^{\log d}))$ を達成して、蛋白質列に対して時間量 $O(nm^2(L+d \cdot 61.86^d \cdot m^{\log d}))$ を達成する。ここで、 m は最も長い入力文字列の長さである。この新しいアルゴリズムは厳密解を1つ見つけるだけでなく、すべての厳密解を見つめることもできる。以前、すべての厳密解を見つめる最も効率的なアルゴリズムが Davila らによって設計されたが、そのアルゴリズムの見つけた解と本研究で得たアルゴリズムの見つけた解を比べることによって彼らのアルゴリズムにバグがあることが判明した。

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 6 件)

- ① Zhi-Zhong Chen and L. Wang. Fast Exact Algorithms for the Closest String and Substring Problems with Application to the Planted (L, d)-Motif Model. IEEE/ACM Transactions on Computational Biology and Bioinformatics, Vol. 8, pp. 1400-1410, 2011. DOI 10.1109/TCBB.2011.21
- ② Zhi-Zhong Chen, G. Lin, and L. Wang. An Approximation Algorithm for the Minimum Co-Path Set Problem. Algorithmica, Vol. 60, pp. 969-986, 2011. DOI 10.1007/s00453-010-9389-x
- ③ Zhi-Zhong Chen and L. Wang. HybridNET: a Tool for Constructing Hybridization Networks. Bioinformatics, Vol. 26, No. 22, pp. 2912-2913, 2010.
- ④ Zhi-Zhong Chen, S. Konno, and Y. Matsushita. Approximating Maximum Edge 2-Coloring in Simple Graphs. Discrete Applied Mathematics, Vol. 158, pp. 1894-1901, 2010. DOI 10.1016/j.dam.2010.08.010
- ⑤ Z. Cai, Zhi-Zhong Chen, and G. Lin. A 3.4713-Approximation Algorithm for the Capacitated Multicast Tree Routing Problem. Theoretical Computer Science, Vol. 410, No. 52, pp. 5415-5424, 2009. DOI 10.1016/j.tcs.2009.05.013
- ⑥ Zhi-Zhong Chen and L. Wang. Improved Approximation Algorithms for Reconstructing the History of Tandem Repeats. IEEE/ACM Transactions on Computational Biology and Bioinformatics, Vol. 6, pp. 438-453, 2009. DOI 10.1109/TCBB.2008.122

[学会発表] (計 6 件)

- ① Zhi-Zhong Chen, B. Ma, and L. Wang. A Three-String Approach to the Closest String Problem. Proceedings of 16th Annual International Computing and Combinatorics Conference (COCOON' 2010), Lecture Notes in Computer Science, Vol. 6196, pp. 449-458, Nha Trang, Vietnam, July 19-21, 2010.
- ② Zhi-Zhong Chen, S. Konno, and Y. Matsushita. Approximating Maximum Edge 2-Coloring in Simple Graphs. Proceedings of 6th International Conference on Algorithmic Aspects in Information and Management (AAIM' 2010), Lecture Notes in Computer

Science, Vol. 6124, pp. 78--89, Weihai, China, July 19-21, 2010.

- ③ Zhi-Zhong Chen, M. R. Fellows, B. Fu, H. Jiang, Y. Liu, L. Wang, and B. Zhu. A Linear Kernel for Co-Path/Cycle Packing. Proceedings of 6th International Conference on Algorithmic Aspects in Information and Management (AAIM' 2010), Lecture Notes in Computer Science, Vol. 6124, pp. 90--102, Weihai, China, July 19-21, 2010.
- ④ Z. Cai, Zhi-Zhong Chen, G. Lin, and L. Wang. An Improved Approximation Algorithm for the Capacitated Multicast Tree Routing Problem. Proceedings of 2nd International Conference on Combinatorial Optimization and Applications (COCOA' 08), Lecture Notes in Computer Science, Vol. 5165, pp. 286-295, Newfoundland, Canada, August 21-24, 2008.
- ⑤ Zhi-Zhong Chen and R. Tanahashi. Approximating Maximum Edge 2-Coloring in Simple Graphs via Local Improvement. Proceedings of 4th International Conference on Algorithmic Aspects in Information and Management (AAIM' 08), Lecture Notes in Computer Science, Vol. 5034, pp. 84-96, Shanghai, China, June 23-25, 2008.
- ⑥ Zhi-Zhong Chen, R. Tanahashi, and L. Wang. An Improved Randomized Approximation Algorithm for Maximum Triangle Packing. Proceedings of 4th International Conference on Algorithmic Aspects in Information and Management (AAIM' 08), Lecture Notes in Computer Science, Vol. 5034, pp. 97-108, Shanghai, China, June 23-25, 2008.

[その他]

ホームページ

<http://rnc.r.dendai.ac.jp/~chen/downloadable.html>

6. 研究組織

(1) 研究代表者

陳 致中 (CHEN ZHI-ZHONG)
東京電機大学・理工学部・教授
研究者番号：00242933

(2) 研究分担者 なし

(3) 連携研究者 なし