

令和 5 年 6 月 6 日現在

機関番号：14301

研究種目：基盤研究(C) (一般)

研究期間：2020～2022

課題番号：20K04978

研究課題名(和文) 離散問題のモデリングとアルゴリズム ～部分問題からのアプローチ～

研究課題名(英文) Modeling and Algorithms for Discrete Problems

研究代表者

原口 和也 (Haraguchi, Kazuya)

京都大学・情報学研究科・准教授

研究者番号：80453356

交付決定額(研究期間全体)：(直接経費) 3,300,000円

研究成果の概要(和文)：特に離散列挙問題に関して顕著な成果を挙げた。極大解生成オラクルによって暗に与えられるアイテム付き集合システム(ハイパーグラフ)に対し、共通アイテム集合に関して極大な解をすべて列挙する問題を考えた。集合システムに一切の仮定がない場合、合流性を仮定した場合のそれぞれについて、多項式遅延アルゴリズムを開発した。これらの成果は一般性が高いものであり、アイテム付きグラフにおける共通アイテム集合に関して極大な連結誘導部分グラフの列挙アルゴリズムの実装、様々な連結条件を満たす部分グラフの列挙問題(たとえば二辺ノ点連結誘導部分グラフ、強連結誘導部分グラフ)など、新しい研究への展開につながることができた。

研究成果の学術的意義や社会的意義

共通アイテム集合に関して極大解を列挙する問題を、従来まったく取り扱われてこなかった集合システム上に拡張し、多項式遅延列挙が可能であることを示した。この成果は、様々な連結性の条件について、共通アイテム集合に関して極大かつ当該連結条件を満たす誘導部分グラフを列挙することが、多項式遅延で可能なことを意味している。この問題は遺伝生物学において有意なタンパク質構造の発見や、DWAS分析などに応用を持ち、これら個別の問題を解決するための効率の良いアルゴリズムの存在が示唆されている。

研究成果の概要(英文)：It is the enumeration problem that we have made remarkable achievements. Suppose that a set system (i.e., a hypergraph) with an item set is implicitly given by means of an oracle that generates maximal solutions that are subsets of an input subset. We consider a problem that asks to enumerate all maximal solutions with respect to the common item set. We have developed polynomial-delay algorithms for a case when there is no assumption on the set system and a case when the set system is "confluent". These results are so general that enabled us to find a way to new research problems that include implementation of algorithms that enumerate all maximal connected induced subgraphs with respect to the common itemset and enumeration of subgraphs that satisfy various connectivity conditions (e.g., 2-edge/vertex-connected induced subgraphs in undirected graphs and strongly-connected induced subgraphs in digraphs).

研究分野：数理工学

キーワード：列挙 アルゴリズム理論 離散最適化

## 様式 C - 19、F - 19 - 1、Z - 19 (共通)

### 1. 研究開始当初の背景

(1) オペレーションズ・リサーチ (OR) や社会工学において離散問題 (最適化問題・決定問題・列挙問題など) は幅広い応用を持つが、残念ながらその多くは計算困難であることが知られている。離散最適化に関して言えば、整数最適化 (IP)・充足可能性問題 (SAT)・制約充足問題 (CSP) といった汎用モデルを解くソルバの性能は飛躍的に高まっているものの、サイズの大規模化や定式化の難しさ (問題構造の対称性など) から、実際のインスタンスを実用的な時間内で厳密に解くことは難しい。

(2) 厳密解であれ近似解であれ、実用的な時間内で所望の解を得るには問題のモデリングが重要である。モデリングとは解きたい問題 P の部分問題への分解であり、そして問題 P を解くアルゴリズム (全体アルゴリズム) は、各部分問題を「部分アルゴリズム」によって系統立てて解き、その結果を統合し、P の解として出力するものである。

(3) 多くの離散問題では、モデルとアルゴリズムの開発・改善の余地がまだまだ残されている。また世の進展に伴い、新しい問題が次々と提起され、実際に解くべきインスタンスの巨大化も進んでいる。そのような中で、できるだけ広汎な問題群に適用できるようなモデリングおよび効率的な部分問題の解決手法とは一体どのようなものだろうか。部分アルゴリズムはどう全体アルゴリズムに組み込めばよいのだろうか。またアルゴリズムは実装しなければ絵に描いた餅同然である。実装をどのように実現すればよいだろうか。

### 2. 研究の目的

上記 1 (3) で述べたリサーチクエスチョンに対する答えを見出すべく、以下の (1) から (3) に関する研究を行う。

(1) 典型的な「計算困難問題」を精選した上で、様々なモデリング手法を検討し、その部分問題を効率良く解くためのアルゴリズムについて研究する。

(2) 得られた部分アルゴリズムを基本部品とし、問題そのものを解くための全体アルゴリズムを開発する。また関連問題への拡張可能性を検討する。

(3) 全体アルゴリズムを実装し、計算機実験に基づいた適切な評価を行う。

### 3. 研究の方法

これまでの研究の経緯にしたがい、上記 2 (1) の「計算困難問題」として頂点彩色問題 (NP 困難問題) およびアイテム付きグラフに関する部分グラフ列挙問題の一般化を取り扱う。

頂点彩色問題は、与えられた無向グラフ上のすべての頂点に色を割り当てることを問う問題である。ただし任意の割当てが許容されるわけではなく、隣接する頂点同士には同じ色を割り当てることはできないという条件の下で、使用する色の数を最小化したい。オイラーの時代から知られるグラフ理論の古典的な問題だが、様々なスケジューリング問題を記述できることが知られている。

またアイテム付きグラフとは頂点にアイテム (の集合) が付されたグラフだが、たとえば生物学では、タンパク質を頂点、相互作用を枝、病気をアイテムとみなしたアイテム付きグラフから、共通アイテム集合の意味で極大な連結誘導部分グラフを列挙することの応用が知られている。

上記 2 (2) の拡張可能性について、研究を通じて開発してきたモデリングやアルゴリズムはできるだけ一般的な形で記述することを目指す。

(1) 頂点彩色問題に対して用いられる典型的な戦略の一つとして、 $k$ -固定型が知られる。ある整数  $k$  に対して  $k$ -彩色可能性問題 ( $k$  色で彩色可能かどうかを問う決定問題) を解き、YES の証拠が得られれば  $k$  を 1 減じて、再び同問題を解くというものである。すなわち頂点彩色問題の部分問題として  $k$ -彩色可能性問題が含まれている。同問題はそれ自体が NP 完全問題であるため、ヒューリスティクスを用いて解くことが一般的である。探索を開始する解 (初期解) はヒューリスティクスの性能を大きく左右するため、どのような初期解生成法が良いのかについて、調査を行う。ここではリサイクル法という手法に着目し、従来手法 (単純な欲張り法や DSATUR 法など) との実験的比較を行う。

(2) アイテム付きグラフにおける列挙問題を一般化し、研究代表者が過去に開発してきた手法 [Haraguchi and Nagamochi, Proceedings of ISAAC 2019] を拡張できないかを検討する。

#### 4. 研究成果

##### (1) 頂点彩色問題

k-彩色可能性問題それ自体は決定問題だが、任意の k-彩色を実行可能解とし、その違反枝(両端点に同じ色が割り当てられた枝)の本数を最小化する問題とみなすことができる。このとき、リサイクル法によって得られる初期解の違反枝の本数が、頂点数、最大次数および k によって表される式を上界に持つことを示した。

またヒューリスティクス of 初期解生成法としてのリサイクル法の有用性を、TabuCol など広く用いられるヒューリスティクス of 初期解生成法として実際に用いることで、実験的に示した。下の表は典型的な結果で、[Uchida, Yajima and Haraguchi, JORSJ, 2021] から抜粋したものである。

この表はベンチマーク問題集として知られる DIMACS サイトの中で、特に難しいとされる 20 個のグラフに対して TabuCol を適用した結果である。最左列「Instance」はグラフの名前を示し、「First」は DSATUR 法という頂点彩色問題に対して広く知られる構築型解法によって決定された最初の k の値を示し、以下「Rec」から「Rlf」は、リサイクル法、単純な欲張り法、ランダム割当、DSATUR 法を k-彩色可能性問題に適応させたアルゴリズム、RLF 法をそれぞれ k-彩色可能性問題の初期解生成法として用いたときの、TabuCol の出力値を示す(カッコ外の値が達成した k の最小値で、カッコ内の値は全 50 回の試行のうち、最小値を達成した試行回数)。また「Best」「 $\chi$ 」はそれぞれ best-known 値と最適値(もしくはその下界)である。各行で太字になっているエントリは、その行(すなわちグラフ)における最良の結果を表す。また最後の行には五つの初期解生成法の中での平均順位を示している。

この表から、初期解生成法としてのリサイクル法の有用性が読み取れる。また同様の結果が、他のベンチマーク問題やヒューリスティクスについても観察されたことを追記しておく。

k-固定型やリサイクル法の考え方は、たとえば最大独立集合問題など他の離散最適化問題に対して応用することも可能である。このことは、離散最適化問題に対するヒューリスティクスを設計する上でリサイクル法を選択肢の一つに入れることのメリットを示唆している。

表: [Uchida, Yajima and Haraguchi, JORSJ, 2021]から抜粋

Instance	First	REC	GRD	RND	DSA	RLF	Best	$\chi$
							[10, 30, 32]	
DSJC1000.1	25 (1)	20 (6)	20 (4)	20 (6)	<b>20 (11)</b>	20 (5)	20	$\geq 10$
DSJC1000.5	113 (1)	88 (3)	88 (2)	88 (3)	89 (37)	88 (1)	82	$\geq 73$
DSJC1000.9	297 (1)	<b>225 (6)</b>	226 (32)	226 (31)	226 (19)	225 (1)	222	$\geq 216$
DSJC500.1	15 (11)	12 (50)	12 (50)	12 (50)	12 (50)	12 (50)	12	$\geq 5$
DSJC500.5	63 (3)	<b>49 (13)</b>	49 (8)	49 (3)	49 (7)	49 (9)	47	$\geq 43$
DSJC500.9	161 (3)	126 (2)	126 (1)	<b>126 (3)</b>	127 (50)	127 (50)	126	$\geq 123$
DSJR500.1c	87 (3)	86 (11)	87 (3)	87 (3)	86 (7)	<b>85 (1)</b>	85	84
DSJR500.5	129 (9)	<b>124 (4)</b>	127 (1)	128 (3)	125 (7)	126 (2)	122	122
R1000.1c	103 (3)	<b>98 (33)</b>	98 (11)	98 (12)	98 (23)	98 (24)	98	98
R1000.5	250 (50)	<b>240 (4)</b>	247 (1)	249 (9)	240 (2)	243 (2)	234	234
R250.1c	65 (34)	64 (26)	65 (34)	65 (34)	<b>64 (28)</b>	64 (12)	64	64
R250.5	66 (8)	66 (43)	66 (11)	66 (8)	<b>66 (46)</b>	66 (8)	65	65
flat1000.50_0	111 (1)	<b>50 (50)</b>	56 (2)	56 (1)	58 (1)	58 (1)	50	50
flat1000.60_0	112 (4)	<b>60 (50)</b>	73 (2)	74 (5)	75 (3)	75 (1)	60	60
flat1000.76_0	112 (1)	87 (1)	<b>87 (2)</b>	87 (1)	87 (1)	88 (47)	81	76
flat300.28_0	40 (1)	<b>28 (4)</b>	30 (1)	30 (1)	31 (49)	31 (50)	28	28
le450.15c	23 (12)	15 (2)	<b>15 (3)</b>	15 (1)	15 (2)	15 (1)	15	15
le450.15d	23 (2)	16 (50)	16 (49)	16 (50)	15 (1)	16 (50)	15	15
le450.25c	28 (4)	26 (50)	26 (50)	26 (50)	26 (49)	26 (50)	25	25
le450.25d	28 (19)	26 (50)	26 (50)	26 (50)	26 (50)	26 (50)	25	25
Averaged rank		1.40	2.85	2.90	2.80	2.95	-	-

##### (2) アイテム付き集合システムにおける列挙問題

集合システムとは、要素集合  $V$  およびその部分集合族  $\mathcal{C}$  の対である。ハイパーグラフとも呼ばれ、グラフの一般化とみなすことができる ( $\mathcal{C}$  に属するすべての  $C$  についてそのサイズが  $2$ )。アイテム付き集合システム  $(V, \mathcal{C}, I, \pi)$  とは、集合システム  $(V, \mathcal{C})$ 、アイテム集合  $I$ 、アイテムの割当  $\pi: V \rightarrow 2^I$  から成る。アイテム付き集合システムは、何らかのオラクルを用いて暗に与えられるものとする。本研究では以下の条件を満たす部分集合  $C \in \mathcal{C}$  (解) すべてを列挙することを問う問題を考える。その条件とは、 $C, C' \in \mathcal{C}, C \supset C'$  ならば  $\pi(C) \subset \pi(C')$ 。ただし  $\pi(C)$  は  $C$  に属する要素  $v$  のアイテム集合  $\pi(v)$  すべての積集合として定義される。

合流システム (Confluent system) が極大オラクルによって与えられるとき、すべての解を入力サイズおよびオラクル計算時間に関して列挙する多項式遅延アルゴリズムを開発した。なお極大オラクルとは、任意の部分集合  $S \subseteq V$  を入力とし、 $S$  に含まれる極大な成分  $C \in \mathcal{C}$  すべてを出力するものである。この研究成果の意義は以下の二点である。一点目は、従来全く研究されてこなかった合流システムに対して多項式遅延列挙の可能性を示したことである。これまでの研究は独立性システム (Independent system) およびその拡張である Strongly-accessible system、Accessible-system に関するものに集中していたが、合流システムを取り扱ったものはほとんど見られなかった。なお合流システムはグラフにおける様々な連結性を一般化した概念であり、その多項式遅延列挙可能性を示したことは、部分グラフ列挙の新しい可能性を切り拓いたと言える。二点目はオラクルの可能性である。従来研究はメンバシップオラクル (任意の部分集合  $S \subseteq V$  を入力とし、それが  $c$  に属するか否かを返す) を仮定するものがほとんどであった。極大オラクルは、その設計が列挙問題を解くことに相当し、一般には容易ではないが、合流システムにおいてはそれが多項式時間で可能であることも少なくなく、オラクルの柔軟な定義により、「解ける」クラスの新しい可能性を見出せることを示したのは大きな意義がある。

一連の研究成果をとりまとめた論文は、Algorithmica 誌および Theoretical Computer Science 誌に掲載された。集合システムの関係を表したベン図を下記に示す。

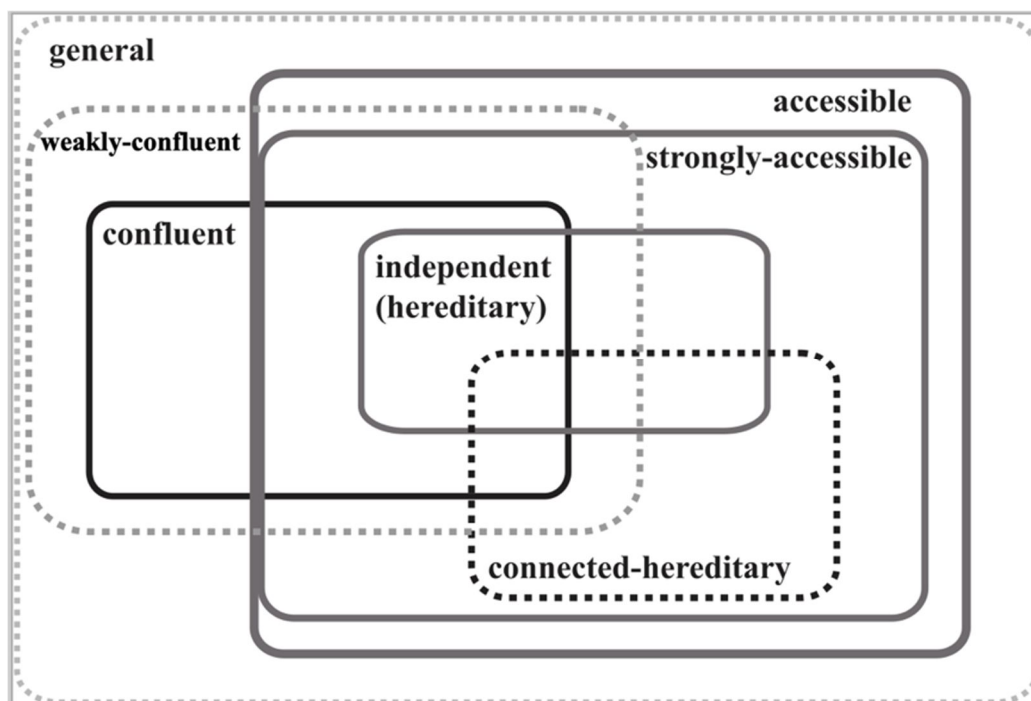


図: [Haraguchi and Nagamochi, Algorithmica, 2022] から抜粋

## 5. 主な発表論文等

〔雑誌論文〕 計6件（うち査読付論文 6件/うち国際共著 0件/うちオープンアクセス 0件）

1. 著者名 Tada Takumi, Haraguchi Kazuya	4. 巻 13889
2. 論文標題 A Linear Delay Algorithm for Enumeration of 2-Edge/Vertex-Connected Induced Subgraphs	5. 発行年 2023年
3. 雑誌名 LNCS	6. 最初と最後の頁 368 ~ 379
掲載論文のDOI (デジタルオブジェクト識別子) 10.1007/978-3-031-34347-6_31	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -
1. 著者名 Haraguchi Kazuya, Nagamochi Hiroshi	4. 巻 963
2. 論文標題 Polynomial-delay enumeration algorithms in set systems	5. 発行年 2023年
3. 雑誌名 Theoretical Computer Science	6. 最初と最後の頁 113963 ~ 113963
掲載論文のDOI (デジタルオブジェクト識別子) 10.1016/j.tcs.2023.113963	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -
1. 著者名 Haraguchi Kazuya, Nagamochi Hiroshi	4. 巻 -
2. 論文標題 Enumeration of Support-Closed Subsets in Confluent Systems	5. 発行年 2022年
3. 雑誌名 Algorithmica	6. 最初と最後の頁 -
掲載論文のDOI (デジタルオブジェクト識別子) 10.1007/s00453-022-00927-x	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -
1. 著者名 Yasutaka Uchida, Kaito Yajima, Kazuya Haraguchi	4. 巻 64 (3)
2. 論文標題 Recycling Solutions for Vertex Coloring Heuristics	5. 発行年 2021年
3. 雑誌名 Journal of the Operations Research Society of Japan	6. 最初と最後の頁 -
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Shi Yu, Zhu Jianshen, Azam Naveed Ahmed, Haraguchi Kazuya, Zhao Liang, Nagamochi Hiroshi, Akutsu Tatsuya	4. 巻 22
2. 論文標題 An Inverse QSAR Method Based on a Two-Layered Model and Integer Programming	5. 発行年 2021年
3. 雑誌名 International Journal of Molecular Sciences	6. 最初と最後の頁 2847 ~ 2847
掲載論文のDOI (デジタルオブジェクト識別子) 10.3390/ijms22062847	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Haraguchi Kazuya, Torii Kotaro, Endo Motomu	4. 巻 287
2. 論文標題 Maximum weighted matching with few edge crossings for 2-layered bipartite graph	5. 発行年 2020年
3. 雑誌名 Discrete Applied Mathematics	6. 最初と最後の頁 40 ~ 52
掲載論文のDOI (デジタルオブジェクト識別子) 10.1016/j.dam.2020.07.017	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

[学会発表] 計5件 (うち招待講演 0件 / うち国際学会 2件)

1. 発表者名 Jianshen Zhu, Kazuya Haraguchi, Hiroshi Nagamochi, Tatsuya Akutsu
2. 発表標題 Adjustive Linear Regression and Its Application to the Inverse QSAR
3. 学会等名 BIOINFORMATICS 2022 (13th International Conference on Bioinformatics, Models, Methods and Algorithms) (国際学会)
4. 発表年 2022年

1. 発表者名 Jianshen Zhu, Naveed Ahmed Azam, Kazuya Haraguchi, Liang Zhao, Hiroshi Nagamochi, Tatsuya Akutsu
2. 発表標題 A Method for Molecular Design Based on Linear Regression and Integer Programming
3. 学会等名 12th International Conference on Bioscience, Biochemistry and Bioinformatics (ICBBB 2022)
4. 発表年 2022年

1. 発表者名 Naveed Ahmed Azam, Jianshen Zhu, Kazuya Haraguchi, Liang Zhao, Hiroshi Nagamochi, Tatsuya Akutsu
2. 発表標題 Molecular Design Based on Artificial Neural Networks, Integer Programming and Grid Neighbor Search
3. 学会等名 2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)
4. 発表年 2021年

1. 発表者名 Kouki Tanaka, Jianshen Zhu, Naveed Ahmed Azam, Kazuya Haraguchi, Liang Zhao, Hiroshi Nagamochi, Tatsuya Akutsu
2. 発表標題 An Inverse QSAR Method Based on Decision Tree and Integer Programming
3. 学会等名 Intelligent Computing Theories and Application (ICIC 2021)
4. 発表年 2021年

1. 発表者名 Jianshen Zhu, Naveed Ahmed Azam, Kazuya Haraguchi, Liang Zhao, Hiroshi Nagamochi, Tatsuya Akutsu
2. 発表標題 An Improved Integer Programming Formulation for Inferring Chemical Compounds with Prescribed Topological Structures
3. 学会等名 IEA/AIE 2021 (国際学会)
4. 発表年 2021年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究協力者	永持 仁  (Nagamochi Hiroshi)		

6. 研究組織（つづき）

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究協力者	橋本 英樹  (Hashimoto Hideki)		

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関