

令和 4 年 6 月 3 日現在

機関番号：12601

研究種目：基盤研究(B)（一般）

研究期間：2016～2020

課題番号：16H02781

研究課題名（和文）簡潔データ構造の理論と応用

研究課題名（英文）Theory and Practice of Succinct Data Structures

研究代表者

定兼 邦彦（SADAKANE, KUNIHICO）

東京大学・大学院情報理工学系研究科・教授

研究者番号：20323090

交付決定額（研究期間全体）：（直接経費） 5,600,000円

研究成果の概要（和文）：理論的にも実用的にも優れた簡潔データ構造を複数開発した。例えば、集合族を表現するZDDの使用メモリを削減し、高速に演算を行える手法を開発した。実験により、頻出アイテム集合を求める問題において、通常のZDDと比較して使用メモリが1/3になり、実行時間の増加は40%のみであった。また、高次元直交領域探索のための省メモリのデータ構造を開発した。

研究成果の学術的意義や社会的意義

ビッグデータ処理で問題になるのはデータが大量にあるため計算機のメモリ上での処理が行えないということである。古典的な手法ではデータをハードディスク上に格納して処理を行うI/Oアルゴリズムを用いていた。しかしハードディスクのランダムアクセスはメモリへのアクセスとは比べ物にならないほど低速であり、ビッグデータ処理には時間がかかっていた。簡潔データ構造では、データ構造のサイズは $o(n)$ 、つまりデータ量の劣線形となっており、データを計算機のメモリ上に格納することができ、スケーラブルなアルゴリズムが開発できる。

研究成果の概要（英文）：We developed several succinct data structures that have good performances in both theory and practice. For example, we reduced the working space for ZDD that represents set families. We also developed space-efficient data structures for multi-dimensional orthogonal range searches.

研究分野：アルゴリズムとデータ構造

キーワード：簡潔データ構造

1. 研究開始当初の背景

ビッグデータを高速に処理することの必要性については説明するまでもないが、その際に問題になるのはデータが大量にあるため計算機のメモリ上での処理が行えないということである。古典的な手法ではデータはハードディスク上に格納し、限られたメモリを使ってディスクを読み書きしつつ処理を行う I/O アルゴリズムを用いていた。しかしハードディスクのランダムアクセスはメモリへのアクセスとは比べ物にならないほど低速であるため、ランダムアクセスの回数が少なくなるようにアルゴリズムを設計する必要があった。その後、メモリ価格の低下から計算機に大量のメモリを搭載できるようになり、多くのデータをメモリ内に格納できるようになったため、通常アルゴリズムを用いることが可能になったが、今度はこれまで無視されていた別の問題が生じるようになった。それは、古典的なデータ構造のサイズはデータ量に対して超線形であるということである。

簡潔データ構造では、データ構造のサイズは $o(n)$ 、つまりデータ量の劣線形となっている。よって、データ量が増えてもデータ構造が必要とするメモリはあまり増えない。なおかつ従来の超線形データ構造と同じ処理を同じ時間計算量で行えるため、スケーラブルなアルゴリズムが開発できる。

2. 研究の目的

研究代表者はこれまでビットベクトル、木構造、文字列検索のデータ構造などの多くの基本的な簡潔データ構造を開発し、それらの理論的解析と実験的評価を行ってきた。これらのデータ構造を基盤として、様々な問題を解くためのデータ構造の開発と評価を行う。具体的には、主に次の問題について取り組む。

(a) 動的な簡潔 BDD/ZDD とその応用

BDD (binary decision diagram) と ZDD (zero-suppressed BDD) は論理関数や集合族をコンパクトに表現する手法であり、それらを用いて組み合わせ最適化問題を解く手法が提案されている。本研究では、静的な簡潔 BDD と動的な通常の BDD を併用する手法を用い、実用的な動的簡潔 BDD/ZDD を開発する。これにより、組み合わせ最適化問題を少ない使用メモリで高速に解くことが可能になる。

(b) 幾何データの簡潔データ構造

2次元、3次元の点座標を表現する通常のデータ構造は R 木、4分木、kd 木など様々なものが存在するが、それらのサイズは超線形のためスケーラブルでは無い。例えば日本全体の道路地図(道路の座標データ等)を表現するデータ構造はサイズが非常に大きくなってしまふ。幾何データの表現で問題になるのは、点座標が実数値であるためそれ自体の格納領域も大きくなることである。座標値の圧縮を行う際に用いられる手法の 1 つに、値を昇順にソートして各値は隣の値との差分で表現する手法がある。これにより値を圧縮することができるが、別の問題が生じる。それは座標値の検索が行いにくくなるということである。データの検索を効率よく行うためには、データに索引を付ける必要がある。検索はその索引を用いて行うが、データが差分で表現されていると索引を付けることが難しくなるという問題がある。また、点集合を表現する場合には順序は関係ないため並び替えを行えるが、線分のデータでは並び替えを行えないため圧縮がしにくくなる。これらの問題を考慮した簡潔データ構造を開発する。

3. 研究の方法

簡潔データ構造の応用について、主にハイブリッド簡潔 ZDD の効率的アルゴリズム、幾何データの簡潔データ構造と最短路探索の簡潔索引について研究を行う。計算機実験により、提案したデータ構造の効率を評価する。従来手法よりも使用メモリを大幅に削減しつつ、計算時間があまり増えないようなアルゴリズムを開発する。

4. 研究成果

主な研究成果について挙げる。

・ハイブリッド DenseZDD の開発

ZDD は集合族をコンパクトに表現するデータ構造である。静的な DenseZDD と通常の動的な ZDD を組み合わせ、少ない使用メモリで高速に演算を行える手法を開発した。実験により、頻出

アイテム集合を求める問題において、通常の ZDD と比較して使用メモリが $1/3$ になり、実行時間の増加は 40% のみであった。

・高次元直交領域探索問題に対する省スペースデータ構造

3次元以上の空間内の n 点の集合に対する直交領域探索に対して 2 つのデータ構造を提案した。1 つ目は、 $dn \log n + o(n \log n)$ ビットの領域を使用し、問い合わせを $O((n^{(d-2)/d} + occ) \log n / \log \log n)$ 時間で行える。2 つ目は、 $dn \log U - n \log n + o(n \log n)$ ビットの領域を使用し、実データに対して問い合わせを高速に実行できる。このデータ構造の使用する領域は最適である。

5. 主な発表論文等

〔雑誌論文〕 計18件（うち査読付論文 17件 / うち国際共著 6件 / うちオープンアクセス 6件）

1. 著者名 Sankardeep Chakraborty, Kunihiko Sadakane	4. 巻 LIPIcs 138
2. 論文標題 Indexing Graph Search Trees and Applications	5. 発行年 2019年
3. 雑誌名 Proceedings of 44th International Symposium on Mathematical Foundations of Computer Science (MFCS)	6. 最初と最後の頁 67:1--67:14
掲載論文のDOI (デジタルオブジェクト識別子) 10.4230/LIPIcs.MFCS.2019.67	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -
1. 著者名 Kazuki Ishiyama, Kunihiko Kunihiko	4. 巻 Online First
2. 論文標題 Compact and succinct data structures for multidimensional orthogonal range searching	5. 発行年 2020年
3. 雑誌名 Information and Computation	6. 最初と最後の頁 Online First
掲載論文のDOI (デジタルオブジェクト識別子) 10.1016/j.ic.2020.104519	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 Manas Jyoti Kashyop, Tsunehiko Nagayama, Kunihiko Sadakane	4. 巻 23
2. 論文標題 Faster algorithms for shortest path and network flow based on graph decomposition	5. 発行年 2019年
3. 雑誌名 Journal of Graph Algorithms and Applications	6. 最初と最後の頁 781 ~ 813
掲載論文のDOI (デジタルオブジェクト識別子) 10.7155/jgaa.00512	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -
1. 著者名 Kentaro Sumigawa, Kunihiko Sadakane	4. 巻 13
2. 論文標題 Storing Partitions of Integers in Sublinear Space	5. 発行年 2019年
3. 雑誌名 The Review of Socionetwork Strategies	6. 最初と最後の頁 237 ~ 252
掲載論文のDOI (デジタルオブジェクト識別子) 10.1007/s12626-019-00044-2	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 Kentaro Sumigawa and Kunihiko Sadakane	4. 巻 LNCS 10979
2. 論文標題 An Efficient Representation of Partitions of Integers	5. 発行年 2018年
3. 雑誌名 Proceedings of IWOCA	6. 最初と最後の頁 361 ~ 373
掲載論文のDOI (デジタルオブジェクト識別子) 10.1007/978-3-319-94667-2_30	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 Paniz Abedin, Arnab Ganguly, Wing-Kai Hon, Yakov Nekrich, Rahul Shah, Sharma V. Thankachan, Kunihiko Sadakane	4. 巻 LNCS 10976
2. 論文標題 A Linear-Space Data Structure for Range-LCP Queries in Poly-Logarithmic Time	5. 発行年 2018年
3. 雑誌名 Proceedings of COCOON	6. 最初と最後の頁 615 ~ 625
掲載論文のDOI (デジタルオブジェクト識別子) 10.1007/978-3-319-94776-1_51	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 該当する

1. 著者名 Alessio Conte, Gaspare Ferraro, Roberto Grossi, Andrea Marino, Kunihiko Sadakane, Takeaki Uno	4. 巻 KDD '18
2. 論文標題 Node Similarity with q -Grams for Real-World Labeled Networks	5. 発行年 2018年
3. 雑誌名 Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining	6. 最初と最後の頁 1282 ~ 1291
掲載論文のDOI (デジタルオブジェクト識別子) 10.1145/3219819.3220085	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 該当する

1. 著者名 Shuhei Denzumi, Jun Kawahara, Koji Tsuda, Hiroki Arimura, Shin-ichi Minato, and Kunihiko Sadakane	4. 巻 11(8)
2. 論文標題 DenseZDD: A Compact and Fast Index for Families of Sets	5. 発行年 2018年
3. 雑誌名 Algorithms	6. 最初と最後の頁 128 ~ 128
掲載論文のDOI (デジタルオブジェクト識別子) 10.3390/a11080128	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 Kengo Nakamura, Kunihiko Sadakane	4. 巻 12(3)
2. 論文標題 Space-Efficient Fully Dynamic DFS in Undirected Graphs	5. 発行年 2019年
3. 雑誌名 Algorithms	6. 最初と最後の頁 52 ~ 52
掲載論文のDOI (デジタルオブジェクト識別子) 10.3390/a12030052	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 Ishiyama Kazuki, Sadakane Kunihiko	4. 巻 IEEE
2. 論文標題 A Succinct Data Structure for Multidimensional Orthogonal Range Searching	5. 発行年 2017年
3. 雑誌名 Proceedings of IEEE Data Compression Conference	6. 最初と最後の頁 270-279
掲載論文のDOI (デジタルオブジェクト識別子) 10.1109/DCC.2017.47	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 Patrick Dinklage, Johannes Fischer, Dominik Koppl, Marvin Lobel, Kunihiko Sadakane	4. 巻 75
2. 論文標題 Compression with the tudocomp Framework	5. 発行年 2017年
3. 雑誌名 Proceedings of International Symposium on Experimental Algorithms (SEA)	6. 最初と最後の頁 13:1-13:22
掲載論文のDOI (デジタルオブジェクト識別子) 10.4230/LIPICS.SEA.2017.13	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 該当する

1. 著者名 Ishiyama Kazuki, Sadakane Kunihiko	4. 巻 LNCS 10609
2. 論文標題 Practical Space-Efficient Data Structures for High-Dimensional Orthogonal Range Searching	5. 発行年 2017年
3. 雑誌名 Proceedings of SISAP	6. 最初と最後の頁 234-246
掲載論文のDOI (デジタルオブジェクト識別子) 10.1007/978-3-319-68474-1_16	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 Ishiyama Kazuki, Kobayashi Koji, Sadakane Kunihiko	4. 巻 LNCS 10609
2. 論文標題 Succinct Quadrees for Road Data	5. 発行年 2017年
3. 雑誌名 Proceedings of SISAP	6. 最初と最後の頁 262-272
掲載論文のDOI (デジタルオブジェクト識別子) 10.1007/978-3-319-68474-1_18	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 Taito Lee, Shuhei Denzumi, Kunihiko Sadakane	4. 巻 LNCS 9685
2. 論文標題 Engineering Hybrid DenseZDDs	5. 発行年 2016年
3. 雑誌名 Proceedings of SEA	6. 最初と最後の頁 201-216
掲載論文のDOI (デジタルオブジェクト識別子) 10.1007/978-3-319-38851-9_14	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 Takuya Takagi, Shunsuke Inenaga, Kunihiko Sadakane, Hiroki Arimura	4. 巻 LNCS 9843
2. 論文標題 Packed Compact Tries: A Fast and Efficient Data Structure for Online String Processing	5. 発行年 2016年
3. 雑誌名 Proceedings of IWOCA	6. 最初と最後の頁 213-225
掲載論文のDOI (デジタルオブジェクト識別子) 10.1007/978-3-319-44543-4_17	査読の有無 無
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 Arnab Ganguly, Wing-Kai Hon, Kunihiko Sadakane, Rahul Shah, Sharma V. Thankachan, Yilin Yang	4. 巻 54(2)
2. 論文標題 Space-Efficient Dictionaries for Parameterized and Order-Preserving Pattern Matching	5. 発行年 2016年
3. 雑誌名 Proceedings of CPM	6. 最初と最後の頁 1-12
掲載論文のDOI (デジタルオブジェクト識別子) 10.4230/LIPIcs.CPM.2016.2	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 該当する

1. 著者名 Jose Salavert Torres, Andres Tomas, Ignacio Medina, Kunihiko Sadakane, Ignacio Blanquer	4. 巻 11(4)
2. 論文標題 Pair-End Inexact Mapping on Hybrid GPU Environments and Out-Of-Core Indexes	5. 発行年 2016年
3. 雑誌名 Current Bioinformatics	6. 最初と最後の頁 459-469
掲載論文のDOI (デジタルオブジェクト識別子) 10.2174/1574893611666160212235359	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 該当する

1. 著者名 Dinghua Li, Ruibang Luo, Chi-Man Liu, Chi-Ming Leung, Hing-Fung Ting, Kunihiko Sadakane, Hiroshi Yamashita, Tak-Wah Lam	4. 巻 102
2. 論文標題 MEGAHIT v1.0: A Fast and Scalable Metagenome Assembler driven by Advanced Methodologies and Community Practices	5. 発行年 2016年
3. 雑誌名 Methods	6. 最初と最後の頁 3-11
掲載論文のDOI (デジタルオブジェクト識別子) 10.1016/j.ymeth.2016.02.020	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 該当する

〔学会発表〕 計0件

〔図書〕 計1件

1. 著者名 定兼 邦彦	4. 発行年 2018年
2. 出版社 共立出版	5. 総ページ数 232
3. 書名 簡潔データ構造	

〔産業財産権〕

〔その他〕

https://researchmap.jp/sada/

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
連携研究者	伝住 周平 (DENZUMI SHUHEI) (90755729)	東京大学・大学院情報理工学系研究科・助教 (12601)	
連携研究者	小池 敦 (KOIKE ATSUSHI) (20639166)	一関工業高等専門学校・未来創造工学科・准教授 (51201)	

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