

**科学研究費助成事業 研究成果報告書**

平成 29 年 6 月 13 日現在

機関番号：17104

研究種目：挑戦的萌芽研究

研究期間：2014～2016

課題番号：26540119

研究課題名(和文)データ圧縮が発見する結び目理論の新定理

研究課題名(英文)Finding new mathematical knot theory using data compression

研究代表者

坂本 比呂志 (Sakamoto, Hiroshi)

九州工業大学・大学院情報工学研究院・教授

研究者番号：50315123

交付決定額(研究期間全体)：(直接経費) 2,800,000円

研究成果の概要(和文)：ミルナー不変量は、絡み目の分類を与える重要な概念であるが、計算が困難であるため他の不変量解析と比較して研究が進んでいない。本研究は、アルゴリズムによって絡み目の理論やその周辺領域を発展させる解析ツールを提供し、学術上の新発見を支援する。本研究では、データを少しずつ読み込みながらデータを圧縮することで省スペースで圧縮が可能なオンライン文法圧縮を応用して、これまでは計算できなかった深さの展開式まで実行可能なミルナー不変量解析ツールを開発し、専門家のコミュニティに提供した。

研究成果の概要(英文)：Milnor invariant is one of most important notions for classifying link groups in the low-dimensional topology. The invariant is, however, not easy to compute for large links. Thus, in this research, we propose a useful tool for this analysis using the grammar compression that makes us possible to compute Milnor invariant for a large depth not computable so far. We developed such a tool and supplied it to the community of researchers for discovering new knowledge of the link groups.

研究分野：データ圧縮

キーワード：文法圧縮 結び目理論 低次元トポロジー ミルナー不変量

### 1. 研究開始当初の背景

学術上の新発見のいくつかは、アルゴリズムを活用することで成し遂げられたものである。アップルとハーケン(とコンピュータ)による4色定理の証明はその代表例である。この問題は、地図を平面グラフによって抽象化することでグラフ理論の問題として定式化された。位相幾何学における結び目グラフのように抽象化することではじめて解析が可能となる研究分野である。具体的な結び目の描き方は無数に存在し、例えば、図1の複雑な結び目は、それを囲む楕円と同値である。

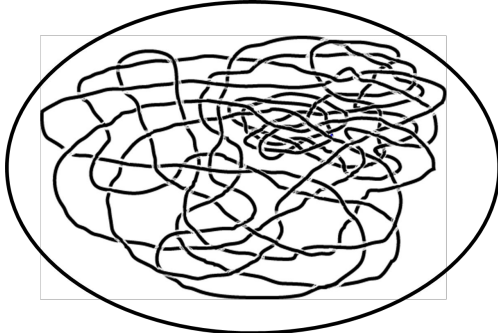


図 1

絡み目は両端が閉じていない複数の紐からなる結び目をより一般化させたものであるが、その効率的な等価性判定法は知られていないため、これまでに絡み目の構造を調べるための様々な不変量が提案されている。ミルナー不変量は、絡み目の重要な分類を与える概念であるが、計算が困難であるため他の不変量解析と比較して研究が進んでいない。

### 2. 研究の目的

本研究は、アルゴリズムによって絡み目の理論やその周辺領域を発展させる解析ツールを提供し、学術上の新発見を目指す。実際にミルナー不変量を求めるとき、単項式と呼ばれる中間生成物を經由して最終的な多項式不変量を得る。ここで、単項式や多項式を構成する変数は非可換(例えば  $xy$  と  $yx$  を区別する)であるため、比較的簡単な場合でも数GB超のデータサイズに膨れ上がってしまい、これまでの素朴なアルゴリズムでは事実上計算は不可能である。本研究では、データ圧縮の理論を駆使することで、計算のボトルネックとなる単項式を陽に生成することなく直接不変量を求めるアルゴリズムを実現し、その成果を研究者に提供する。そして、この分野の新しい問題の定式化や未解決問題の予想を支援する。

### 3. 研究の方法

本研究は、ミルナー不変量の大規模計算の実現のための3つの基本問題から構成される。すなわち、アルゴリズムの設計、データ構造

の開発、そして専用ツールの開発である。そしてそれぞれは、計算機科学の技法を用いてそれぞれ次のような問題に帰着される。

#### 【文法圧縮による単項式の圧縮】

研究代表者らは、これまでに文字列の文法圧縮について精力的に研究を行ってきた。その結果、出力サイズとはほぼ同じ領域で最適な圧縮を可能にするアルゴリズムを開発した。文法圧縮とは、文脈自由文法(CFG)の生成規則で文字列を圧縮する方法である(図2)。

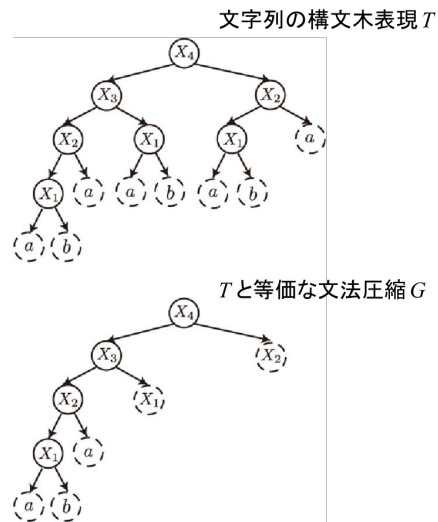


図 2

このアルゴリズムは、文字列を逐次的に処理しながら出力するオンラインアルゴリズムであるため、メモリ使用量が小さいことが特徴である。一方、非可換単項式  $F$  は、変数が逆元を持つという点以外は文字列と見なせる。そこで、文法圧縮を  $F$  の圧縮表現として採用し、同一パターンに対するマグナス展開の重複を可能な限り防ぐことで高速化する。また、既存の文法圧縮はCFGの構文木が2分木に制限されているが、圧縮率のさらなる向上のため、多分木表現の文法圧縮を検討する。

#### 【高速多項式展開】

非可換単項式  $F$  の変数に多項式を代入して展開することでミルナー不変量が得られる。 $n$  項からなる多項式同士の積は、FFTを用いて  $O(n \log n)$  時間で計算可能である。しかし、以下に説明するように、単項式  $F$  の展開にはFFTをそのまま用いることはできない。 $F$  が与えられたとき、ある紐に対応する変数  $x$  には逆元  $x^{-1}$  が定義される。このとき、 $F$  のマグナス展開とは、以下の代入を行うことである。

$$x = 1 + X; x^{-1} = 1 - X + X^2 - X^3 + \dots$$

ただし、この場合も変数は非可換であり、

$x^{n-1}$  のマクローリン展開は適当な次数で打ち切ってよい。マグナス展開によって生成される例えば  $XXYY$ ,  $YXYX$ ,  $YYXX$ ,  $YXXY$  などの項はすべて区別される(可換の場合は  $X^2Y^2$  と等価)。したがって、FFT が威力を発揮する畳み込み計算の効果はあまり期待できず、予備実験でもその傾向が見られた。したがって、前述の文法圧縮による集約された  $F$  に対して計算を実行するアルゴリズムを設計し、非可換多項式の積の計算を高速化する。

#### 【文字列集合の格納】

ミルナー不変量の実体は、単項式  $F$  のマグナス展開によって得られる各項の係数である。したがって、すべての項は区別して格納し、それぞれの項ごとに係数を集計する必要がある。これは変換した文字列中に出現するパターンを集計することであるが、これは文法圧縮で表現された構文木の頂点のラベルを数え上げることで近似できる。この技法を用いて、マグナス展開を省スペースで効率的に計算する手法を開発する。

#### 4. 研究成果

本研究では、データを少しずつ読み込みながらデータを圧縮することで省スペースで圧縮が可能なオンライン文法圧縮を応用して、これまでは計算できなかった深さの展開式まで実行可能なミルナー不変量解析ツールを開発し、専門家のコミュニティに提供した。展開の深さは全体の計算量に最も影響を与えるパラメータであるが、本研究で提案した手法は、素朴な手法では不可能であった深さが 15 を超える値まで実用的な時間で計算できることが示されており、今後はこれをより大規模化・多機能化することで、これまでは知られていなかった不変量に対する新しい知見が得られるものと期待できる。さらに、本研究では、完成までは到達できなかったものの、研究者がフリーハンドで描いた絡み目を画像として取り込んで、それを不変量に変換する機能を実現した(図3)。これによって研究者の思いついたアイデアを直ちに計算してみることが可能となった。このツールも将来的に公開する予定である。

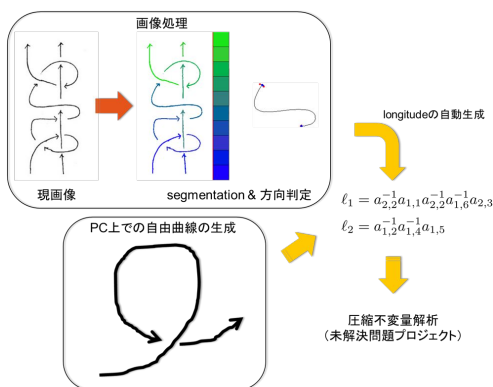


図 3

#### 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 2 件)

Y. Takabatake, K. Nakashima, T. Kuboyama, Y. Tabei, H.Sakamoto. siEDM: an efficient string index and search algorithm for edit distance with moves. *Algorithms*, 9(2) article 26 (18 pages), 2016 (査読あり)

Shouhei Fukunaga, Yoshimasa Takabatake, Tomohiro I, Hiroshi Sakamoto. Online Grammar Compression for Frequent Pattern Discovery. *Proceedings of Machine Learning Research*, 57:93-104, 2016 (査読あり)

〔学会発表〕(計 21 件)

大西 孝典, 坂本 比呂志. 文法圧縮を用いた類似度計算の大規模データへの適用. 第 100 回 SIG-FPAI 研究会, 2016 年 03 月 27 日~2016 年 03 月 28 日, 熊本市市民会館(熊本県熊本市)

徳永 啓太郎, 坂本 比呂志. 定数領域の頻度計算を用いたオンライン文法圧縮アルゴリズム. 第 100 回 SIG-FPAI 研究会, 2016 年 03 月 27 日~2016 年 03 月 28 日, 熊本市市民会館(熊本県熊本市)

福永 祥平, 坂本 比呂志. 文法圧縮における逆引き辞書の省スペース化. 第 100 回 SIG-FPAI 研究会, 2016 年 03 月 27 日~2016 年 03 月 28 日, 熊本市市民会館(熊本県熊本市)

青山 友紀, 高島 嘉将, 坂本 比呂志. ストリームデータからの頻出パターンの近似発見. 第 99 回 SIG-FPAI 研究会, 2016 年 01 月 21 日~2016 年 01 月 22 日, 湯の原ホテル(宮城県仙台市)

水野 仁人, 高島 嘉将, 坂本 比呂志. 文法圧縮のハッシュ領域の削減. 第 99 回 SIG-FPAI 研究会, 2016 年 01 月 21 日~2016 年 01 月 22 日, 湯の原ホテル(宮城県仙台市)

Takako Hashimoto, Tetsuji Kuboyama, Basabi Chakraborty. Topic extraction from millions of tweets using singular value decomposition and feature selection. *APSIPA 2015*, 2015 年 12 月 16 日~2015 年 12 月 19 日, Hong Kong (香港(中国))

Takako Hashimoto, Dave Shepard, Tetsuji Kuboyama, Kilho Shin. Event Detection from Millions of Tweets Related to the Great East Japan Earthquake Using Feature Selection Technique. *ICDM Workshops 2015*, 2015 年 12 月 14 日~2015 年 12 月 17 日, Atlantic City, NJ, USA (アトランティックシティ(米国))

Kilho Shin, Tetsuji Kuboyama, Takako Hashimoto, Dave Shepard . Super-CWC and super-LCC: Super fast feature selection algorithms . Big Data 2015 , 2015 年 10 月 29 日 ~ 2015 年 11 月 01 日 , Santa Clara, CA, USA ( サンタクララ (米国))

Tomoya Yamazaki, Akihiro Yamamoto, Tetsuji Kuboyama . Tree PCA for Extracting Dominant Substructures from Labeled Rooted Trees . DS 2015 , 2015 年 10 月 04 日 ~ 2015 年 10 月 06 日 , Banff, AB, Canada ( バンフ (カナダ))

Yoshimasa Takabatake, Yasuo Tabei, Hiroshi Sakamoto . SPIRE 2015 , 2015 年 09 月 01 日 ~ 2015 年 09 月 04 日 , London, UK ( ロンドン (英国))

Shinichi Yamagiwa, Koichi Marumo, Hiroshi Sakamoto . Stream-Based Lossless Data Compression Hardware Using Adaptive Frequency Table Management . BPOE 2015 , 2015 年 08 月 31 日 ~ 2015 年 09 月 04 日 , Kohala, HI, USA ( ハワイ (米国))

高島 嘉将, 坂本 比呂志 . 文法圧縮のための逆引き辞書の省スペース化. 第 98 回 SIG-FPAI 研究会 , 2015 年 08 月 07 日 ~ 2015 年 08 月 08 日 , 和歌山大学 (和歌山県和歌山市)

岩崎 暁, 坂本 比呂志 . ストリーム中の頻出アイテムの発見のための Simple Algorithm の高速化 . 第 98 回 SIG-FPAI 研究会 , 2015 年 08 月 07 日 ~ 2015 年 08 月 08 日 , 和歌山大学 (和歌山県和歌山市)

高島 嘉将, 中島 健太, 田部井 靖生, 坂本 比呂志 . siEDM: 移動付き編集距離の効率的な索引 . 第 98 回 SIG-FPAI 研究会 , 2015 年 08 月 07 日 ~ 2015 年 08 月 08 日 , 和歌山大学 (和歌山県和歌山市)

高島 嘉将, 田部井 靖生, 坂本 比呂志 . 移動付き編集距離のオンラインパターンマッチング . 第 97 回 SIG-FPAI 研究会 , 2015 年 03 月 22 日 ~ 2015 年 03 月 23 日 , 別府国際コンベンションセンター (大分県別府市)

中島 健太, 前田 幸司, 高島 嘉将, 坂本 比呂志 . 移動付き編集距離に基づく曖昧検索が可能な圧縮索引 , 第 97 回 SIG-FPAI 研究会 , 2015 年 03 月 22 日 ~ 2015 年 03 月 23 日 , 別府国際コンベンションセンター (大分県別府市)

高島 嘉将, 田部井 靖生, 坂本 比呂志 . 文法圧縮に基づく自己索引のオンライン構築について . 第 96 回 SIG-FPAI 研究会 , 2015 年 01 月 13 日 ~ 2015 年 01 月 14 日 , 名古屋工業大学 (愛知県名古屋市)

前田 幸司, 高島 嘉将, 田部井 靖生, 坂本 比呂志 . 文法圧縮を応用したハミング距離計算の高速化 . 第 95 回 SIG-FPAI 研究会 , 2014 年 10 月 10 日 ~ 2014 年 10

月 10 日大阪大学産業科学研究所(大阪府豊中市)

Hiroshi Sakamoto . Grammar Compression: Grammatical Inference by Compression and Its Application to Real Data . 12th International Conference on Grammatical Inference (招待講演), 2014 年 09 月 17 日 ~ 2014 年 09 月 19 日 , Kyoto University (京都府京都市)

前田 幸司, 高島 嘉将, 田部井 靖生, 坂本 比呂志 . 文法圧縮を応用したハミング距離の短い文字列列挙アルゴリズム . 第 94 回 SIG-FPAI 研究会 , 2014 年 07 月 24 日 ~ 2014 年 07 月 24 日 , 根室市総合文化会館 (北海道根室市)

- ② 宮木 亮洋, 坂本 比呂志 . 乱択アルゴリズムを用いた頻出文字列の近似数え上げに基づく省スペース文法圧縮 . 第 94 回 SIG-FPAI 研究会 , 2014 年 07 月 24 日 ~ 2014 年 07 月 24 日 , 根室市総合文化会館 (北海道根室市)

[その他]

ホームページ等

<http://www.donald.ai.kyutech.ac.jp/~hiroshi>

## 6. 研究組織

### (1) 研究代表者

坂本 比呂志 (SAKAMOTO, Hiroshi)

九州工業大学・大学院情報工学研究院・教授

研究者番号 : 50315123

### (2) 研究分担者

安原 晃 (YASUHARA, Akira)

津田塾大学・学芸学部・教授

研究者番号 : 60256625

### (3) 研究分担者

久保山 哲二 (KUBOYAMA, Tetsuji)

学習院大学・計算機センター・教授

研究者番号 : 80302660