

令和 6 年 5 月 7 日現在

機関番号：12612

研究種目：研究活動スタート支援

研究期間：2022～2023

課題番号：22K21273

研究課題名（和文）動的文字列処理に対するアルゴリズム技法の開発と計算限界の解明

研究課題名（英文）Development of algorithms for dynamic string processing and elucidation of computational lower bound.

研究代表者

三重野 琢也（Mieno, Takuya）

電気通信大学・大学院情報理工学研究科・助教

研究者番号：60953388

交付決定額（研究期間全体）：（直接経費） 2,200,000円

研究成果の概要（和文）：本研究では、時間とともに変化する文字列データに対するアルゴリズム技法の開発に取り組んだ。そして、準動的な文字列上での計算問題や、効率的な動的文字列処理に資する計算問題に対して、省領域かつ高速なデータ構造とアルゴリズムを提案した。特に、入力文字列中の出現回数が少ない回文や、長い回文部分文字列を計算する問題に主に取り組み、成果を得た。

研究成果の学術的意義や社会的意義

世の中に溢れるあらゆる電子的なデータは、計算機上では記号の羅列、すなわち文字列として表現される。よって、文字列データ処理の研究は計算機科学において最も基本的かつ重要な役割を担っている。近年の科学技術の進歩により、大量かつ多様な電子データ（文字列データ）が日々生み出されているため、そのような大量かつ多様な文字列データを効率的に処理する技法の開発は重要である。電子データの中には時間とともに変化する動的データと、それ以外の静的データが存在する。本研究では、前者の動的データを効率的に扱うためのいくつかのアルゴリズム技法を開発した。

研究成果の概要（英文）：We worked on developing algorithmic techniques for dynamic string data. We proposed space-efficient and fast data structures and algorithms for problems on semi-dynamic strings and problems that contribute to efficient dynamic string processing. In particular, we worked on the problem of computing palindromes that occur less frequently in the input strings and long palindromic substrings, and obtained some research results.

研究分野：離散アルゴリズム

キーワード：アルゴリズム データ構造 動的文字列処理 回文

1. 研究開始当初の背景

近年のネットワーク技術の発展や多様な情報システムの普及を背景に、大規模なデータが日々生成され続けている。そして、大規模データを解析することでそこに内在する価値ある情報を発見する重要性が、多くの分野で指摘されている。しかし、そのような大規模データは一般に静的なものではなく、時間とともにデータが増減する、あるいはデータが書き換えられることがある。例えば、センサ機器から得られる観測データは時間経過に従い増加を続ける。さらには、メモリ資源の制約のために、過去の観測データが破棄される場合がある。また、プログラムのソースコードや SNS の投稿などのデータに対しては、人為的な編集操作が行われることがある。そのような動的に変化するデータから価値ある情報を適切に発見するためには、データの変化に対して頑健で効率的なデータ処理手法が求められる。上記の例を含むあらゆる動的データは、計算機の上では記号や文字の列、すなわち文字列とみなすことができる。従って、動的文字列データ処理に対するアルゴリズム技法の開発は、あらゆる動的データ処理技術の中で最も基本的かつ重要な要素のひとつである。

2022 年に理論計算機科学のトップ国際会議 STOC にて、動的文字列処理の分野における飛躍的進歩があった (引用文献 [1])。その論文では、接尾辞配列と呼ばれる文字列索引構造を動的文字列上で扱う問題について、データ構造の更新時間と配列要素へのアクセス時間がともに $O(\text{polylog}(n))$ である手法が提案された。ここで n は文字列の長さであり、 $\text{polylog}(n)$ は n の対数についての多項式を表す。接尾辞配列は多くの文字列アルゴリズム中で利用される強力なデータ構造のひとつである。ゆえに、動的文字列に対する接尾辞配列 (動的接尾辞配列) の設計は、動的文字列上の多くの問題の解法に多大な影響を与える。これまで、動的接尾辞配列の更新時間及びアクセス時間には $O(\text{polylog}(n))$ の壁があったが、文献 [1] によりその壁が初めて破られた。この結果の出現により、動的文字列アルゴリズムの研究は今後さらに加速・発展することが予想される。

2. 研究の目的

本研究の目的は、動的文字列データ処理のための効率的なアルゴリズムを開発すること、そしてそれに関連する計算問題の理論的境界、すなわち時間・領域計算量の下界を示すことである。動的文字列に対して、入力文字列長 n に依存した計算時間を要する静的文字列アルゴリズムをそのまま利用すると、文字列が変化するとともに n に依存した時間がかかる。そのため、文字列長が非常に長く編集操作が複数回適用される場合に非効率である。そこで、本研究における動的文字列アルゴリズムの設計では、入力文字列長には陽には依存せず、編集によって生じる計算対象である離散構造の差分のみに依存した計算時間の達成を目指す。さらに、動的文字列に対する各種問題に対して、その解法の理論的境界、すなわち計算量の下界を示すことも目指す。

3. 研究の方法

本研究では、動的文字列上のユニーク部分文字列計算問題、動的文字列上の極大一致ユニーク部分文字列計算問題、動的文字列上の不在文字列計算問題、そしてその他の文字列構造の計算問題に取り組む。

各計算問題に対しては、以下のように研究に取り組む：動的文字列処理では、入力文字列に対して基本編集操作(文字の挿入・削除・置換)が複数回適用される状況が一般的である。本研究ではまず、文字列の任意の1文字を置換したときの文字列構造の変化を解析し、その変化を高速に検出する手法の提案を目指す。それができた後に、2回以上の編集操作を許す問題設定について研究し、最終的には任意の回数の編集操作を許す設定にまで一般化をする。それと並行して、当該問題の難しさの本質を文字列の組合せ的性質を駆使して解明し、計算量の上下界を漸近的に一致させることを目指す。

4 . 研究成果

本研究で得られた主な成果は以下のとおりである。

1) 準動的文字列上でのユニーク回文部分文字列計算データ構造の提案

文字列中にちょうど1度だけ出現する回文を、ユニーク回文部分文字列と呼ぶ。生物学分野において、DNA配列中のあるユニークな回文構造の存在がオリゴヌクレオチドの免疫活性化活動に影響がするという報告があり、それがユニーク回文部分文字列計算の動機である。本研究ではユニーク回文部分文字列の新たな組合せ的性質を解明した上で、静的文字列上で動作する省領域アルゴリズムを提案した。それに加え、準動的文字列に対する初めてのアルゴリズムを提案した。いずれの手法も、理論的に高速であるといえる。

2) 部分文字列中のトップ k 最長回文問い合わせに対するデータ構造の提案

Amirら(引用文献 [2])によって提案された動的文字列処理のための汎用的な枠組みの構成要素のひとつとして、部分文字列問い合わせがある。部分文字列問い合わせでは、入力文字列の任意の部分文字列に対して所望の値や構造を高速に計算することが要求される。本研究では、トップ k 最長回文計算問題の部分文字列問い合わせ版を扱った。そして、任意の部分文字列問い合わせに最適時間で応答する線形領域のデータ構造を提案した。そのデータ構造の構築時間は文字列長 n に対して $O(n \log n)$ 時間であり、さらに $k=1$ の場合には構築時間を $O(n)$ に高速化することに成功した。

<引用文献>

- [1] Dominik Kempa and Tomasz Kociumaka. 2022. Dynamic suffix array with polylogarithmic queries and updates. In Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing (STOC 2022). Association for Computing Machinery, New York, NY, USA, 1657–1670.
- [2] Amihood Amir, Panagiotis Charalampopoulos, Solon P. Pissis and Jakub Radoszewski. Dynamic and Internal Longest Common Substring. *Algorithmica* **82**, 3707–3743 (2020).

5. 主な発表論文等

〔雑誌論文〕 計2件（うち査読付論文 2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 Mieno Takuya, Funakoshi Mitsuru	4. 巻 86
2. 論文標題 Data Structures for Computing Unique Palindromes in Static and Non-Static Strings	5. 発行年 2023年
3. 雑誌名 Algorithmica	6. 最初と最後の頁 852 ~ 873
掲載論文のDOI（デジタルオブジェクト識別子） 10.1007/s00453-023-01170-8	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 Mitani Kazuki, Mieno Takuya, Seto Kazuhisa, Horiyama Takashi	4. 巻 979
2. 論文標題 Finding top-k longest palindromes in substrings	5. 発行年 2023年
3. 雑誌名 Theoretical Computer Science	6. 最初と最後の頁 114183 ~ 114183
掲載論文のDOI（デジタルオブジェクト識別子） 10.1016/j.tcs.2023.114183	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1件（うち招待講演 0件 / うち国際学会 1件）

1. 発表者名 Kazuki Mitani, Takuya Mieno, Kazuhisa Seto, Takashi Horiyama
2. 発表標題 Shortest cover after edit
3. 学会等名 35th Annual Symposium on Combinatorial Pattern Matching (CPM 2024). (国際学会)
4. 発表年 2024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
---------------------------	-----------------------	----

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
---------	---------