

自己参照ありLZ分解とLyndon分解の項数の関係について

On the Size of Overlapping Lempel-Ziv and Lyndon Factorizations

H31海自9

派遣先 30th Annual Symposium on Combinatorial Pattern Matching
(イタリア・ピサ)

期 間 2019年6月16日～2019年6月23日(8日間)

申請者 九州大学 大学院システム情報科学府

博士前期課2年 浦部 裕 貴

海外における研究活動状況

研究目的

近年、センサ技術の発達やモバイル端末の普及により膨大なデータが生み出されている。このようなデータには社会や経済などの問題解決に寄与し得る情報が潜在していることから、その利活用が求められている。大規模なデータに対する処理を高速に行うために新しいデータ処理技術の確立が求められている。このようなデータの多くは記号、単語、数値の列とみなせ、文字列として捉えることができる。文字列を対象とするアルゴリズムは文字列が潜在的に持つ代数的・組み合わせ的構造を利用することでさらなる高速化を図ることができる。今回、文字列の分解表現として知られる自己参照ありLZ分解の項数およびLyndon分解の項数が満たす関係を新たに解き明かした。

海外における研究活動報告

1. 会議概要

Symposium on Combinatorial Pattern Matching (略称CPM) は文字列情報処理に関する国際学会である。文字列の圧縮表現、およびパターン照合問題に関する発表が行われ

る。本学会には約50人が参加し、2回の招待講演および31回の論文発表が行われた。今年で30回目を迎え、ピサ大学で開催された。

2. 発表概要

申請者は、CPM2019にて「On the Size of Overlapping Lempel-Ziv and Lyndon Factorizations (自己参照ありLZ分解とLyndon分解の項数の関係について)」という題目で発表を行った。

本発表では文字列の分解表現として知られる自己参照ありLZ分解の項数とLyndon分解の項数が満たす関係を述べた。LZ分解とは、文字列の圧縮表現に利用され、自己参照なしLZ分解や自己参照ありLZ分解など様々な種類が存在する。自己参照なしLZ分解とは各項が、以前に出現する最長の文字列、もしくは出現しない文字のどちらかを満たす文字列の列である。さらに自己参照ありLZ分解とは、各項が以前に開始位置を持つ最長の文字列もしくは出現しない文字のどちらかを満たす文字列の列である。一方で、Lyndon分解とは、辞書式順序が降順であるLyndon文字列の列に分解した列のことである。ここで、Lyndon文字列とは接尾辞の中で辞書式順序最小の文字列が自身で

ある文字列である。LZ分解およびLyndon分解は共に文字列情報学において様々な問題を解く上で活用され、どちらも任意の文字列で一意に定まることが知られている。これらの分解は定義が大きく異なるが、文字列の圧縮表現として知られるStraight Line Programのサイズの下界であったり、文字列の繰り返し構造の一般的な定義として知られる“連”を計算するアルゴリズムに関係があったりと共通点が多く存在した。しかし、直接の関係は知られていなかった。

LZ分解とLyndon分解の関係として、同一の文字列に対して自己参照なしLZ分解の項数を z_{no} 、Lyndon分解の項数を m としたとき、それらの項数が $2z_{no} > m$ を満たすことが2017年にkärkkäinenらによって示された。この結果には自己参照なしLZ分解を持つ性質を利用しており、さらには $z < z_{no}$ が任意の文字列で成り立つため、自己参照ありLZ分解の項数とLyndon分解の項数のタイトな関係は自明には得ることができない。本発表では、任意の文字列において自己参照ありLZ分解の項数 z とLyndon分解の項数 m が $4z > m$ を満たすことを示した。

3. 会議全体について

本発表のための入念な準備や、多くの研究者との交流、さらには現地での英語を用いたコミュニケーションは大きな刺激であり、自身の成長の糧となった。本会議では、申請者が研究対象としてきたLyndon文字列に関連する発表が招待講演として「Some Variations on Lyndon Words」という題目で行われ、Lyndon文字列についてさらに理解を深めることができた。また、同じ文字列情報学であっても、今まで研究対象としてこなかった様々な研究テーマについて、当分野の著名な方々から発表を聞くことができ、昨今の文字列処理分野の動向について多大な知見を得ることができた。

謝辞

文字列情報学の主要国際会議であるCPMにて発表を行うことで当分野の発展に寄与することができ、国際会議の参加の意義を肌を感じる事ができた。これらは一重に貴財団のおかげであり、心より深く感謝申し上げます。

この派遣の研究成果等を発表した 著書、論文、報告書の書名・講演題目

Yuki Urabe, Yuto Nakashima, Shunsuke Inenaga, Hideo Bannai, and Masayuki Takeda. On the Size of Overlapping Lempel-Ziv and Lyndon Factorizations. In CPM 2019.