数学と情報科学で解き明かす多様な対象の数理構造と活用
2019 年度採択研究者

2020 年度
年次報告書

稲永　俊介

九州大学大学院システム情報科学研究院
准教授

文字列学的手法によるシーケンシャルデータ解析

§1. 研究成果の概要

本研究では，文字列，生物学的配列，数列，時系列などの多様な系列データに加え，2次元列やラベル付き木・グラフをも包含する「広義文字列」を対象とする．また，広義文字列の数理構造の解明を行い，大規模な広義文字列データを高速かつ省領域に処理する基盤アルゴリズムを開発した．n を入力長，σ をアルファベットサイズとする．

(a) 動的文字列データ処理：複数ストリームデータ処理の基盤たる，複数テキスト索引構造の逐次構築法を確立した．本手法は，O(nlog σ)最適時間で動作する．また，連長 BW 圧縮の最悪時感度の下界がΩ(log n)であることを証明した．

(b) 広義文字列の数理とアルゴリズム：ラベル付き木に対する索引構造のサイズのタイトな上界・下界を与え，さらにラベル付き木に対する初めての線形領域 DAWG を開発した．また，DAWG を，文字列構造を反映したパラメタ化照合に拡張することに成功した．加えて，時系列処理のための，高速 DTW 計算法と等間隔パターン照合法を開発した．

(c) 文字列反復性指標と圧縮：代表的な反復性指標である文字列アトラクタと，実用上最も広く用いられる LZ77 圧縮のサイズが，大きく異なる入力文字列を示した．続いて，文法圧縮に対して，以下の成果を得た．まず，LZ78, Bisection, RePair などの文法圧縮の出力サイズに関するタイトな下界を与えた．さらに，RePair と比較し実験的により小さい文法を出力する MR 圧縮法と，Lyndon 文字列に基づく高速圧縮自己索引データ構造を開発した．次に，ミスマッチを許容する擬似平方分割問題を効率的に解く手法を与えた．また，最短ユニーク部分文字列に対する 2.6n+o(n)ビット領域の簡潔データ構造を開発した．

受賞：情報処理学会60周年記念論文，SPIRE 2020 Best Paper，SOFSEM 2021 Best Paper

【代表的な原著論文情報】

1) Shunsuke Inenaga, Towards a complete perspective on labeled tree indexing: new size bounds, efficient constructions, and beyond, Journal of Information Processing, 29:1-13, IPSJ 60th Anniversary Best Paper
2) Hideo Bannai, Momoko Hirayama, Danny Hucke, Shunsuke Inenaga, Artur Jeż, and Markus Lohrey, The Smallest Grammar Problem Revisited, IEEE Transactions on Information Theory, 67(1):317-328, January 2021.
3) Sara Giuliani, Shunsuke Inenaga, Zsuzsanna Lipták, Nicola Prezza, Marinella Sciortino, and Anna Toffanello, Novel Results on the Number of Runs of the Burrows-Wheeler-Transform, Proc. 47th International Conference on Current Trends in Theory and Practice of Computer Science (SOFSEM 2021), LNCS 12607, pp. 249-262, January 2021. Best Paper Award
4) Yoshifumi Sakai and Shunsuke Inenaga, A reduction of the dynamic time warping distance to the longest increasing subsequence length, Proc. 31st International Symposium on Algorithms and Computation (ISAAC 2020), 6:1-6:16, December 2020.
5) Kanaru Kutsukake, Takuya Matsumoto, Yuto Nakashima, Shunsuke Inenaga, Hideo Bannai,

and Masayuki Takeda, On repetitiveness measures of Thue–Morse words, Proc. 27th International Symposium on String Processing and Information Retrieval (SPIRE 2020), pp. 213–220, October 2020. Best Paper Award