

準動的でコンパクトな索引構造および 簡潔なvan Emde Boas木について

Semi-Dynamic Compact Index for Short Patterns and Succinct van Emde Boas Tree

H27海自59

派遣先 第26回 組合せパターン照合に関する国際会議
(イタリア・イスキア)

期 間 平成27年6月27日～平成27年7月3日(7日間)

申請者 九州大学 大学院システム情報科学府 情報学専攻

修士2年 松 岡 禎 明

海外における研究活動状況

研究目的

本研究で扱う問題は、文字列情報処理における基本的な問題の1つであるパターン検索問題である。検索可能なパターン長に上限があるという制約のもと、高速かつ省領域な索引構造を提案することを目的とする。

海外における研究活動報告

1. 会議の動向

2015年6月29日から7月1日にかけて開催された、26th Annual Symposium on Combinatorial Pattern Matching (CPM 2015)に参加した。この会議は文字列情報処理分野の国際会議であり、文字列検索などの文字列処理アルゴリズムや文字列の組合せ論的性質、バイオインフォマティクスなどを対象とするものである。

本会議には計83件の論文の投稿があり、例年と同程度の34件が採択され、発表が行われた。そのほか、3件の招待講演が行われた。また、会議には17ヶ国から67人が参加し、活発な議論が行われた。

他の研究者の発表の中には、私と同じく索引

構造に関するものも多数含まれていた。そのほか、文字列の組合せ論に関する発表もいくつかあり、応用分野は不明であるものの非常に興味深い内容であった。今後は、それらの研究内容を精読し、本研究を組み合わせることにより良いアルゴリズムを構築できないか、検討していきたいと考える。

2. 発表概要

文字列検索問題は、あるテキストとパターンが与えられたときに、テキスト中におけるパターンの出現を求める問題である。文字列検索を高速に行うためのデータ構造である索引構造は、実用上極めて重要なデータ構造であるため、多くの研究者が盛んに研究を行っており、様々なアプローチによる索引構造が多数提案されている。

索引構造の評価指標としては、索引構造構築やパターン検索に要する時間のほか、索引構造の表現領域などが挙げられる。一般にこれらの指標はトレードオフの関係にあることが多く、時間と領域ともに良い性能となる索引構造の実現は難しい。そこで、本研究では、検索可能なパターン長に上限を設けることで、

これら2つの指標の両方において良い性能となる索引構造の実現を目指した。

本研究では、検索できるパターン長が短いという条件のもと、高速かつ省領域な索引構造を提案した。また、既存研究として存在するいくつかの省領域な索引構造と実験的に比較することで、構築時間については7倍以上、パターン検索時間については3倍以上高速であることを確認し、本索引構造の優位性を示した。テキストが極めて長い場合において、本索引構造は有用であることが期待される。

さらに、本研究の副産物として、整数集合を管理するためのデータ構造である van Emde Boas 木について、その簡潔表現を提案した。

以上の内容を、会議最終日である7月1日に発表した。本発表内容に関し、2名の研究者から質問およびコメントを頂いた。うち1名のコメントは、本アルゴリズムの手法の別の問題に適用できないかという内容の意見であり、参考になるものであった。このアドバイスを基に、本研究内容の更なる発展を考えていくつもりである。

3. 終わりに

今回の会議参加により、自らの研究内容を

対外的に発表する機会を得ることができ、それにより研究内容を進展させることができた。また、他の研究者の発表を聞いたり、あるいは議論したりすることで、最新の研究を知ることができ、さらに発表技術などを参考にすることができた。一方で、英語でのコミュニケーションにはまだまだ課題を感じる部分も多く、さらなる努力の必要性を認識することとなった。

この会議は、私にとって非常に貴重かつ有意義なものであったと考える。最後に、本会議参加のための渡航費を助成してくださった村田学術振興財団に対し、厚くお礼申し上げます。

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

[著者]

Yoshiaki Matsuoka, Tomohiro I, Shunsuke Inenaga, Hideo Bannai, Masayuki Takeda

[題目]

Semi-dynamic compact index for short patterns and succinct van Emde Boas tree

[会議名および講演概要集]

Combinatorial Pattern Matching (CPM) 2015, pp.355-366