

バイツールを並列処理するための導入提案

株式会社レベルファイブ

福家 優

fukuie@level-five.jp

開発における問題点

バイオインフォマティクス分野における背景

- ・処理を独立して実行できる
- ・マルチコア非対応のツールが多く普及
- ・対象データが大量であるため、マルチコア非対応のツールを使用した場合に膨大な時間を要する

手法・ツールの適用による解決

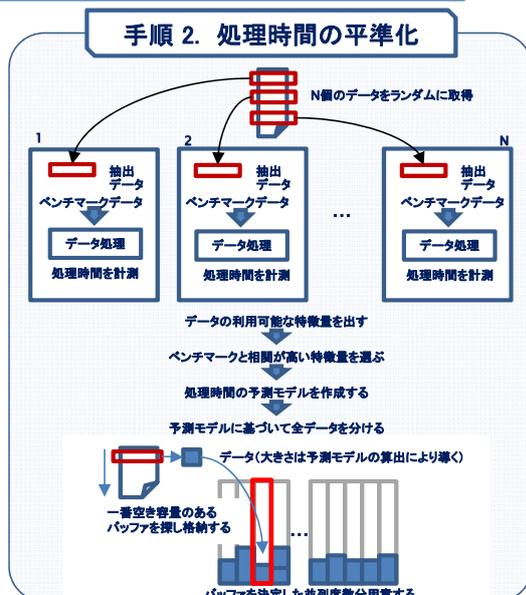
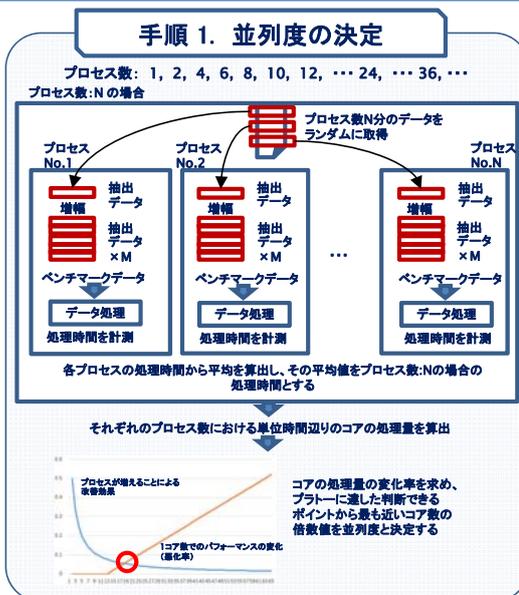
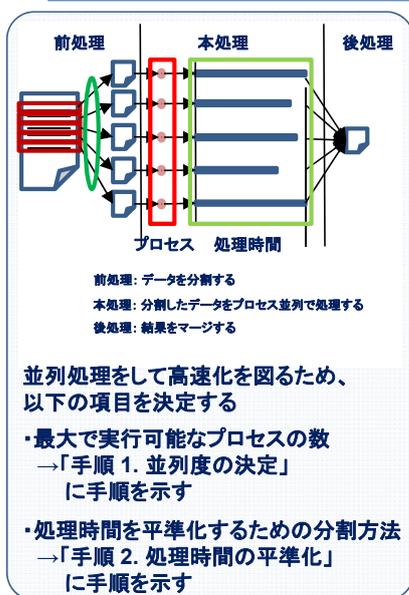
本研究の目的

「マルチコア非対応のツールをプロセス並列処理し解析の高速化を図る手法を提案する」

解決した課題

- ・最大で実行可能なプロセスの数の決定方法
- ・処理時間を平準化するための分割方法

提案手法概要



評価と考察

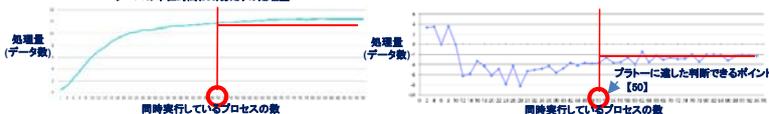
マルチコア非対応であるFrog2(立体配座作成ツール)に適用し評価を実施。

◆ 実行環境

CPU: Intel(R) Xeon(R) CPU X5650 @ 2.67GHz(12コア、24スレッド)、Memory: 96GB

◆ 並列度の決定(手順1)

サーバの単位時間あたりの処理量

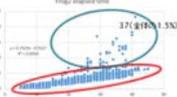


◆ 処理時間の平準化(手順2)

原子数と処理時間に相関あり(重相関0.797)

以下の回帰直線を予測モデルとして使用

$$y = 0.1529x - 0.5527$$



◆ 考察

対象データ: Enamine1309_hts_collection_7.sdf(含有化合物数: 62360)

平準化効果の確認実験:

上記対象データにおいて、平準化を経て実行したデータ群(平準化群)と一定数でデータを分割したデータ群(非平準化群)の各プロセスの処理時間(分)の比較

本手法を適用し並列処理を導入することで、並列処理を適用しない場合と比較して、95%速度向上する。また、平準化効果の確認実験から、平準化を行うことで、適用しない場合と比較して、7%速度向上する。よって、本手法は有効である。

課題

◆ Future Work

今後、以下の項目をFuture Workとして実施する必要がある

- ・並列度の決定の際、飽和した後、スループットが減少しなかった理由の追究
- ・ベンチマークの取り方
- ・Frog2以外のツールでの検証
- ・別環境での検証