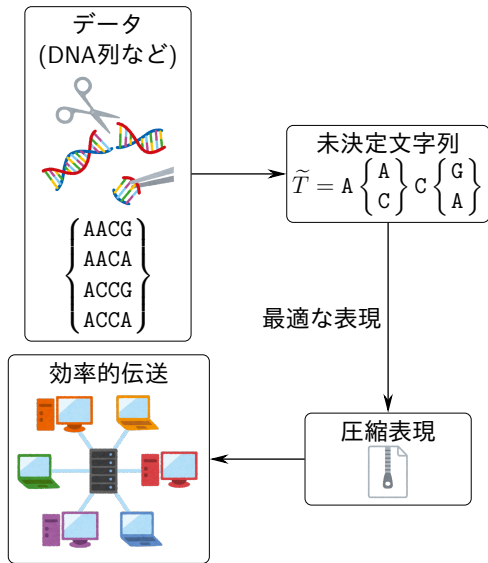


一次元データの効率的な 処理手法の開発

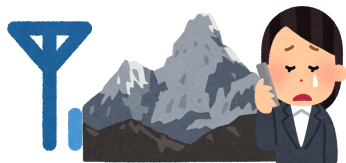
クップル ドミニク

山梨大学



背景・目的

- ▶ 遺伝子データやテキストなど一次元データの処理量が急増
- ▶ 特に山梨県のような山間地域では通信環境が制限され、**データをいかに小さく・効率よく扱うか**が重要



課題

- ▶ 効率的なデータ圧縮と処理が求められている
 - ▶ 文字数に依存しない新たなデータ処理手法の開発
 - ▶ 大規模データの高速かつ圧縮的な解析
- ⇒ モバイル・医療データ伝送にも応用可能

研究の方法・内容

本研究では、遺伝子配列を表す
未決定文字列を対象とした。

未決定文字列

- ▶ 各位置に複数の文字を許容する文字列.
- ▶ 例: $A \begin{Bmatrix} C \\ G \end{Bmatrix} T \begin{Bmatrix} A \\ T \end{Bmatrix}$ は ACTA, ACTT, AGTA, AGTT を表す.
- ▶ 遺伝的変異や測定誤差を柔軟に表現可能.

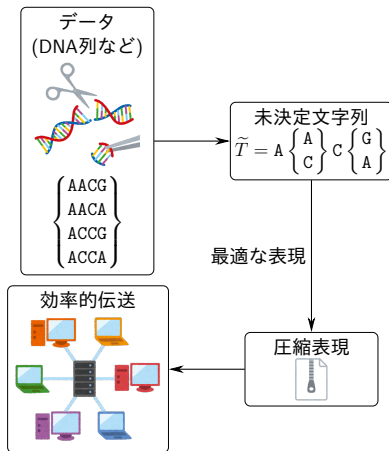
この上で、欠如単語の探索問題を定式化し、その計算量を解析した。

欠如単語

- ▶ 未決定文字列に現れない単語.
- ▶ 例: 上記未決定文字列に対し、CG は欠如単語.

研究の手法

- ▶ 計算量を特定
- ▶ 実用化を目指して解集合プログラミングを用いた探索アルゴリズムを実装



従来の手法では文字列長に依存する計算制限が存在したが，本研究ではその制約を取り除く**新しい理論的枠組み**を提案した．

- ▶ 未決定文字列上での欠如単語探索問題が NP 困難であることを証明
- ▶ ASP を用いた実装により，実際的な規模での探索が可能であることを確認
- ▶ 未決定文字列の一般化である弾性退化文字列への応用可能性を示唆

今後の展望

- ▶ ASP 符号化の最適化と弾性退化文字列への拡張
- ▶ また、山梨県における医療データ通信やゲノム医療の分野への応用
- ▶ 限られた通信環境下でも効率的なデータ処理を実現

論文投稿

- ▶ Dominik Köppl, Jannik Olbrich: 未決定文字列における一意単語の検索の困難さ. 情報処理学会, 研究報告アルゴリズム巻 2024-AL-201, 号 4, ページ 1-5. (2025)
- ▶ Dominik Köppl, Jannik Olbrich: 未決定文字列における欠如単語の検索の困難さ. 情報処理学会, 研究報告アルゴリズム巻 2024-AL-200, 号 15, ページ 1-5. (2024)