

どんな研究

データ分析において、データから重要な特徴のグループを見つけることは基本的なタスクの一つです。しかし、データの規模が大きいとその計算に時間がかかってしまいます。本研究では、**特徴グループの選択において精度を劣化させずに高速化を実現しました。**

どこが凄い

本技術は不要な計算を安全にスキップし、重要そうなグループを重点的に最適化することで、標準的な手法と比較して**精度を維持したまま最大35倍の高速化**に成功しました。また、本技術は追加のハイパーパラメータがないため、**追加のチューニングコストも不要**です。

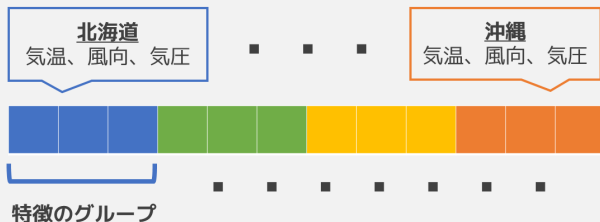
めざす未来

グループをはじめとした複雑な構造を持つデータの分析を高速化することで、**複雑かつ大規模なデータから価値を創出**できるようになります。本研究を発展させることにより、**多種多様なデータから価値を創出し社会へ還元できる未来の実現**をめざします。

背景：データ分析と特徴グループの選択

□ データから重要な特徴グループを選択することは、データ分析では基本的なタスクの一つです。

例：全国の都道府県の天候データから、東京の天気予報にとって重要な都道府県を見つけたい。



どの特徴グループが重要か？

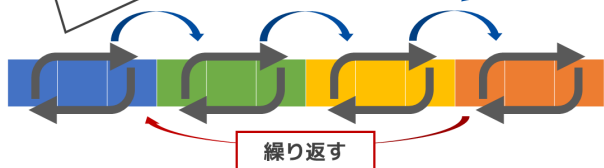
課題：選択の計算に時間がかかります

□ ビッグデータではグループの選択に時間がかかります。

グループが重要か否かを
スコアを計算してチェックし最適化

スコア ≤ 閾値 → 重要でないグループ
スコア > 閾値 → 重要なグループ

順番に各グループの
スコアを計算



「スコア計算の回数」と「全体の繰り返しの回数」
がボトルネックとなり時間がかかります。

提案手法：精度を劣化させずに高速化します

□ 2つのアイデアを利用し高速化します。

1. 不要なスコア計算を安全にスキップします。

重要でないグループのスコア計算を安全に
スキップすることでスコア計算の回数を減らします。



スコアをより高速に計算できるスコアの上界で近似
することで重要でないグループを高速に特定します。

スコアの上界 ≤ 閾値 → 重要でないグループ

2. 重要なグループを重点的に最適化します。



目的関数を最小化するように重要なグループを重点的に
最適化することで全体の繰り返し回数の削減を
狙います。

スコアをより高速に計算できるスコアの下界で近似
することで重要なグループを高速に特定します。

スコアの下界 > 閾値 → 重要なグループ

実験では標準的な手法と比べて
精度を維持したまま最大で35倍の高速化
に成功しました。

関連文献

- [1] Y. Ida, Y. Fujiwara, H. Kashima, "Fast Sparse Group Lasso," in *Proc. Neural Information Processing Systems (NeurIPS)*, 1700-1708, 2019.
[2] 井田安俊, 藤原靖宏, 鹿島久嗣, "Sparse Group Lassoのための高速なBlock Coordinate Descent," *人工知能学会論文誌*, 36(1), 2021.

連絡先

井田 安俊 (Yasutoshi Ida) ソフトウェアイノベーションセンタ
Email: cs-openhouse-ml@hco.ntt.co.jp