

02

データに隠れた関係性を賢く抜き出します

～無限バイクラスタリングによる特徴的部分行列の抽出～

どんな研究

行列の形式で与えられた**関係データ**の中から、**周りとは大きく違う値をもつ部分（バイクラスタ）**だけを**自動的に**見つける研究です。そのような部分は「もっとも売れる顧客層と商品群」や「特定の治験で効果のあった遺伝子群」などのように、**有益な発見**となる可能性が高い部分です。

どこが凄い

これまでの技術では、バイクラスタの数を事前に指定する必要があります。しかし、**事前に正確な数を知ることが困難**です。そこで、任意数のバイクラスタを表現可能な機械学習モデルを用いて、**各データに対して適切なバイクラスタ数を自動的に推定して抽出**できます。

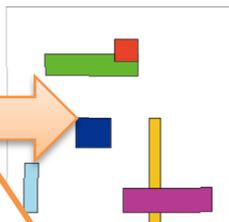
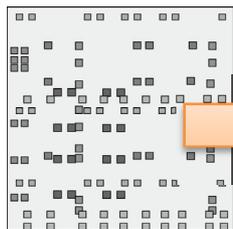
目指す未来

これまでの関係データ解析は、データを表現する隠れたパターンを発見するだけで、「**どれが興味深いパターンか**」は**人が判断する必要**がありました。今後はこの研究のように、「**興味深いと思われるパターン**」だけを発見する、より使いやすい関係データ解析技術を目指します。

バイクラスタリング: 与えられた関係データから 周囲と異なる特徴的な値をもつバイクラスタ(部分行列)を発見する

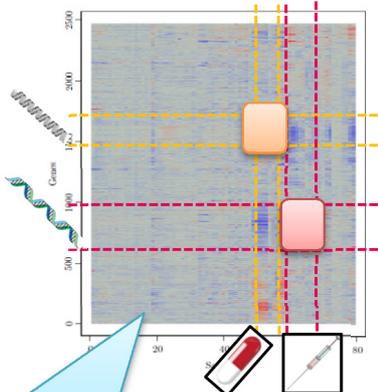
行のクラスタ x 列のクラスタ
= バイクラスタ(部分行列)

関係データ(行列形式)



並び替え
+クラスタリング

遺伝子発現データ



[治療・投薬パターン] x
[特異的に反応する遺伝子群]

購買データ



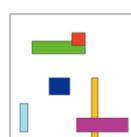
[優良顧客パターン] x
[各パターンに売れる商品群]

従来法: 隠れたバイクラスタの数を
指定しないとイケない☹

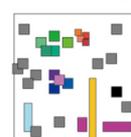
K = 3?



K = 6?



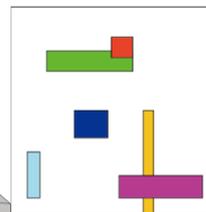
K = many?



提案法: 任意のデータからバイクラスタを個数を含めて推定し、発見する機械学習モデル

☺ 最適バイクラスタ数をデータから自動推定
→ 簡単にバイクラスタリング!

K = 6!!



【関連文献】

- [1] K. Ishiguro, I. Sato, M. Nakano, A. Kimura, N. Ueda, "Infinite plaid models for infinite bi-clustering," in *Proc. 30th AAAI Conference on Artificial Intelligence (AAAI)*, 2016.
- [2] 石黒勝彦, 佐藤一誠, 中野允裕, 木村昭悟, 上田修功, "Infinite bi-clusteringのための無限Plaid Model," 第18回情報論的学習理論ワークショップ (IBIS2015), 2015.

【連絡先】

石黒 勝彦 (Katsuhiko Ishiguro) 協創情報研究部 知能創発環境研究グループ
E-mail : ishiguro.katsuhiko(at)lab.ntt.co.jp