

質量分析データの解析による微生物同定に関する研究

情報科学研究科 情報科学専攻
データサイエンス・人工知能領域 博士前期課程
2024年3月修了

浅野公平

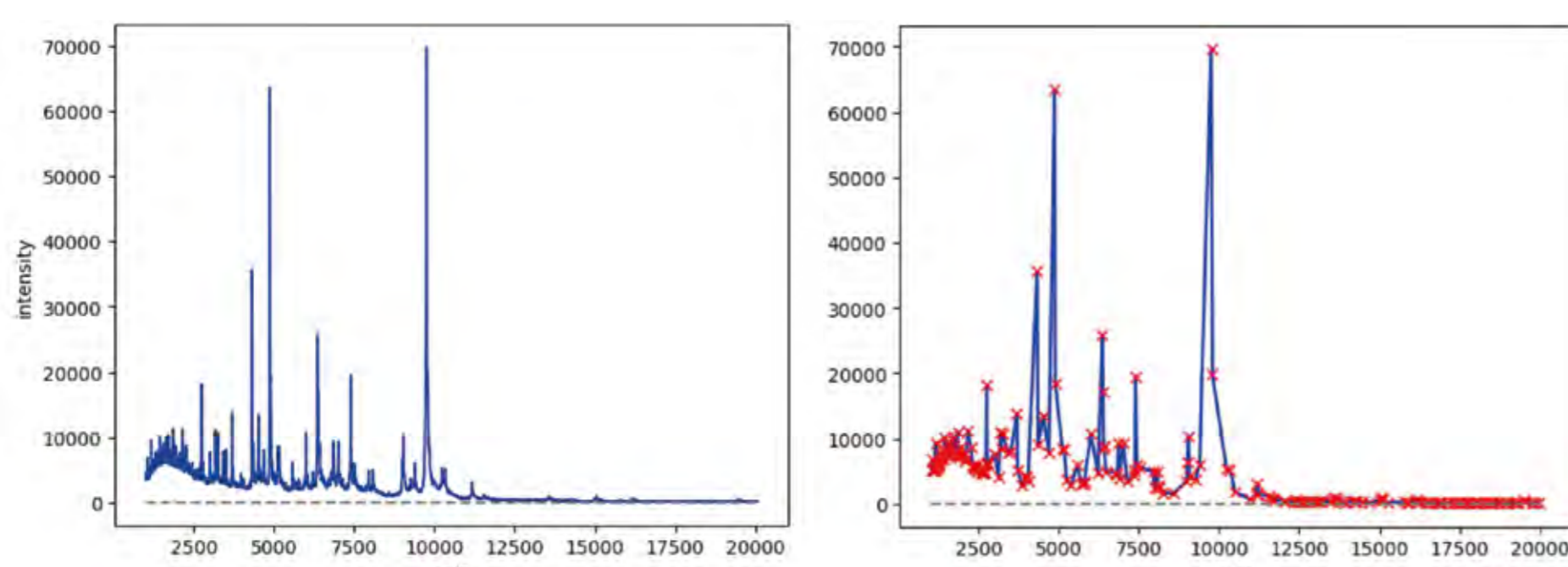
主査 成凱 副査 仲隆 前田誠

研究背景

食品の腐敗や食中毒の原因となる微生物を特定することが必要不可欠。微生物同定のために様々な技術が開発されたが時間やコストを要することや、菌種によっては同定できないか同定精度が低いという問題が存在する。近年、質量分析法で得られるマススペクトルと呼ばれるデータを用いた微生物同定が普及してきたが、試料のイオン化に伴うノイズの存在等により、高精度の微生物同定にはまだ課題が多く残っている。

研究概要

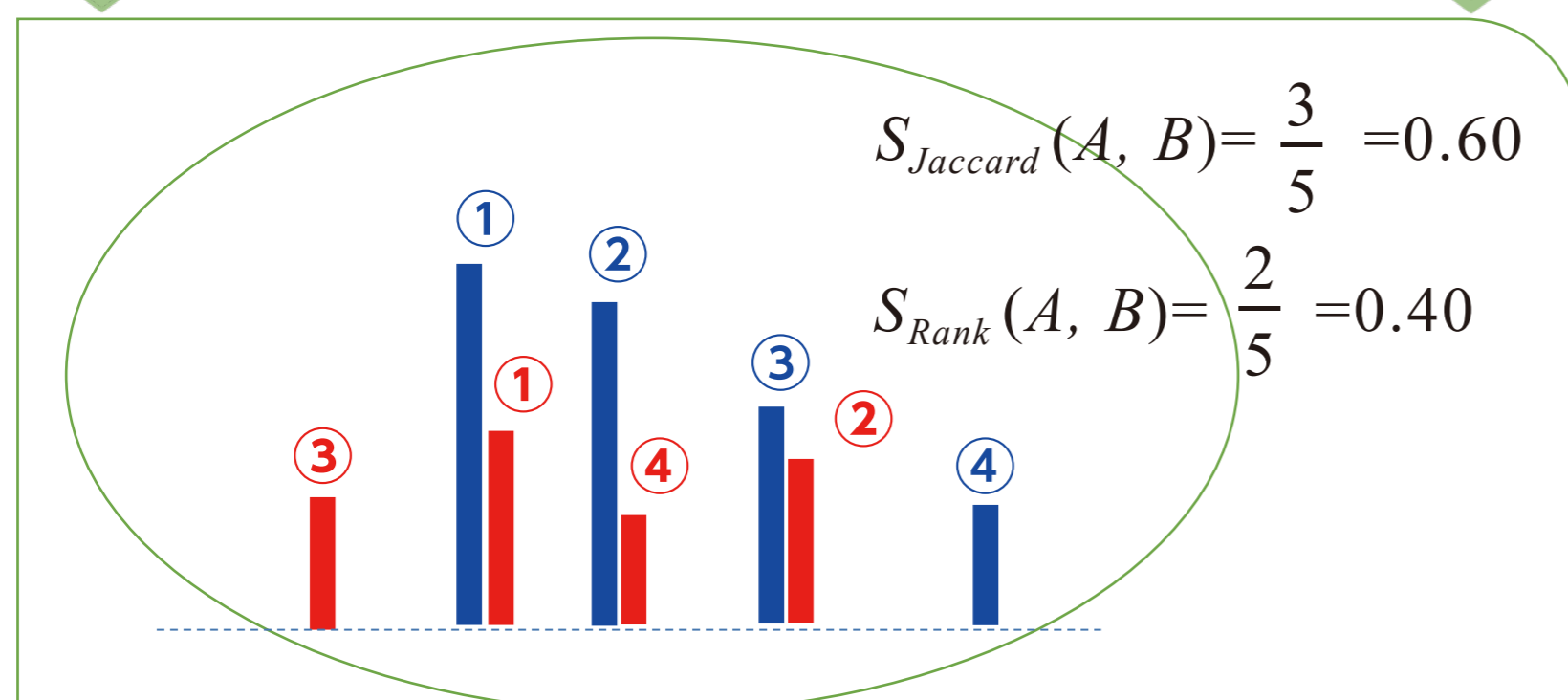
STEP 1. ピーク抽出



STEP 2. ピークアライメント

| ピークID | m/z | ピークID | m/z |
|-------|----------------|-------|---------------|
| 1: | 7600 1000.871 | 3: | 4691 1016.902 |
| 2: | 8866 1008.478 | 4: | 4768 1024.57 |
| 4: | 8345 1024.834 | 6: | 5576 1054.559 |
| 9: | 8516 1045.199 | 7: | 4928 1085.871 |
| 6: | 10569 1054.823 | 9: | 8731 1166.515 |
| 8: | 9882 1147.183 | 10: | 5186 1173.905 |
| 9: | 15204 1166.777 | 11: | 5806 1206.618 |
| 12: | 11945 1217.43 | 12: | 5380 1216.75 |
| 14: | 10820 1247.237 | 13: | 5509 1225.243 |
| 15: | 11625 1262.862 | 14: | 5917 1245.811 |
| 16: | 13042 1319.352 | 15: | 6115 1262.388 |
| 17: | 16232 1331.915 | 16: | 7154 1318.984 |
| 19: | 13732 1379.865 | 17: | 6916 1322.805 |

STEP 3. ピークパターンの類似度計算



研究目的

本研究では、微生物同定の精度向上に向けてマススペクトルの適切な前処理手法、ピークパターンの類似度計算及びそれに基づく微生物同定の手法を提案し、実データを用いた実験により提案手法の評価を行う。これにより、高精度の微生物同定の課題を特定することや精度向上のためのヒントを得ることを目的とする。

評価実験

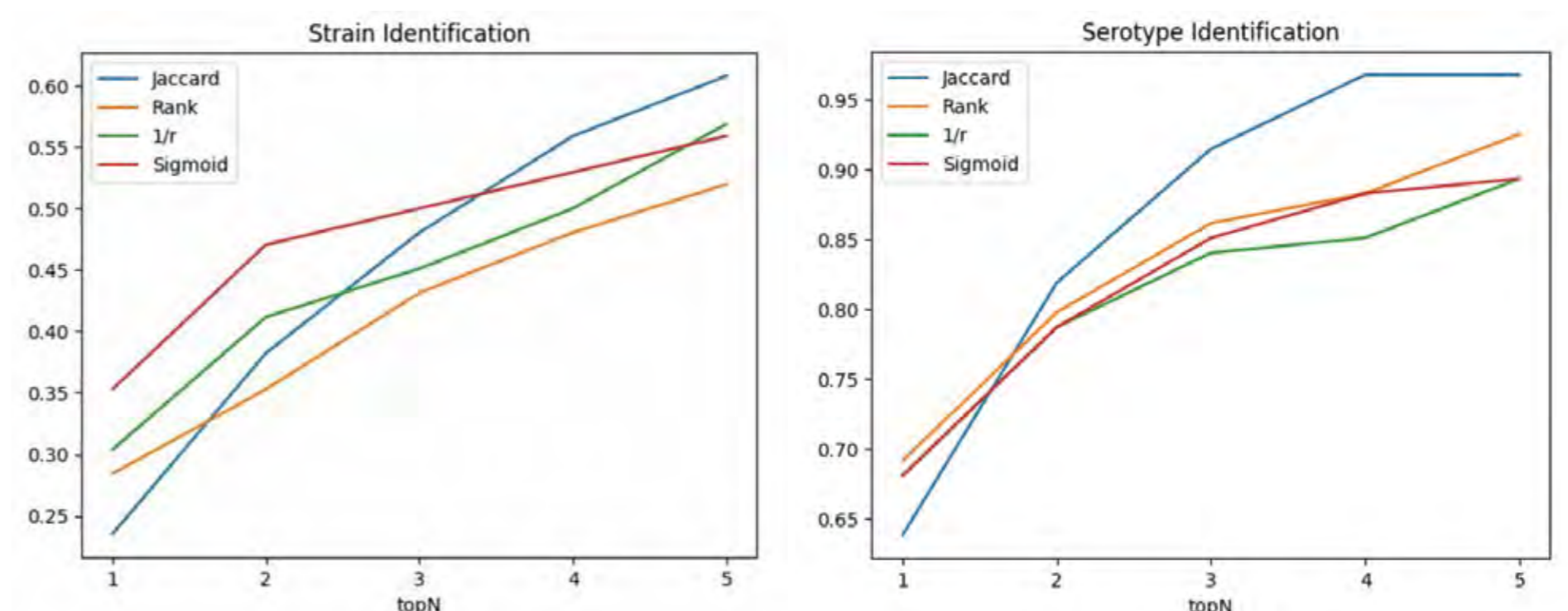
実験データ

- マススペクトル実データ102個
- 菌株数51 (うち47個血清型ラベル付き)

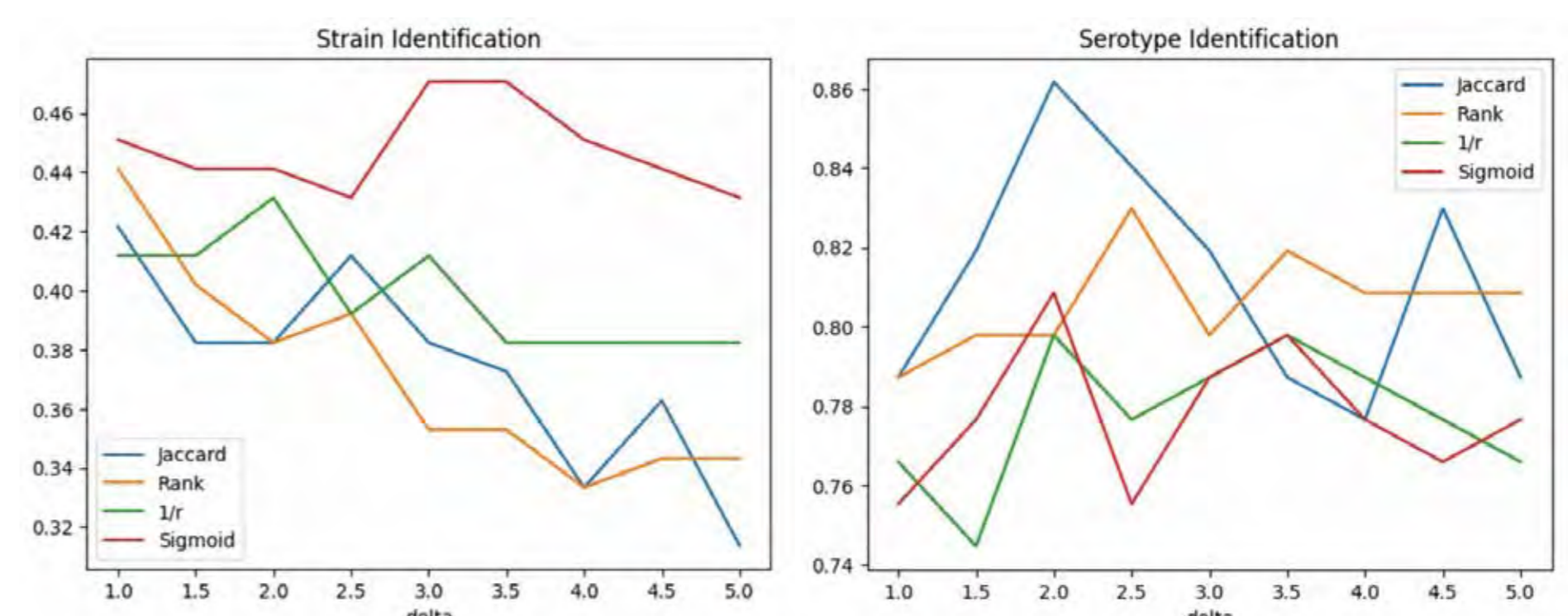
評価対象:

- 類似度計算方式 (Jaccard, Rank, Weighted)
- 重み付け関数 (逆関数, Sigmoid関数)
- ピークアライメント (Pairwise, Global)
- 正解率 = $\frac{\text{未知株と正解株の類似度が TOPN以内の数}}{\text{菌株総数}}$

実験1. 正解と認める類似度の許容値 N と同定精度



実験2. ピークアライメント閾値 δ と同定精度



成果・まとめ

本研究では、微生物同定の精度向上に向けて、マススペクトルの適切な前処理手法、ピークパターン類似度の計算方式及びそれに基づく微生物同定の方法を提案し、実データによる評価実験を通して提案手法の評価を行った。結果として、 $\delta=3, N=1$ のとき Rank 類似度を用いた菌種同定の正解率が最も高く、 $\delta=2, N=2$ のときには Jaccard 類似度を用いた菌種同定が最良であることがわかった。



指導教員コメント

食の安全、食品ロス問題に対処するために、原因となる微生物を特定することが重要である。本研究は、情報科学やデータサイエンスのアプローチで微生物同定の精度を向上させる方式を開発したとともに、今後更なる精度向上のヒントを得たと思う。