質量分析データの類似度評価による菌株識別

浅野　公平 * 豊坂　祐樹 ** 成　凱 *
(九州産業大学 * 情報科学研究科 ** 学術研究推進機構)

1　はじめに

食の安全を守るため，食中毒の原因になる微生物の同定を迅速にかつ正確に行う必要がある．近年，質量分析で得られるマススペクトル (Mass Spectrum: MS) と呼ばれるデータを用いて微生物を効率よく同定する手法が普及してきている [1]．しかし，試料のイオン化に伴うノイズが存在し，高精度の微生物同定にはまだ課題が多く残っている．本研究ではマススペクトルのピークパターンの類似度計算による菌株識別方式を提案し，評価実験を行った．

2　質量分析とマススペクトル

質量分析 (Mass spectrometry) とは，気体状にイオン化された粒子を真空中で運動させ電磁気力を用いて，あるいは飛行時間差によりそれらイオンを質量電荷比 (m/z) に応じて分離・検出することである．質量電荷比に応じて分離・検出されたイオンをもとに，横軸に m/z，縦軸にイオンの信号強度をとったグラフをマススペクトル (Mass spectrum) と呼ぶ．

マススペクトルは，どの分子がどのくらいの信号強度でどのタイミングで現れているのかを示されており，菌種または菌株によってパターンが異なり，微生物同定に利用できる．一方，試料分子をイオン化する過程においては後続反応によって試料が分解するフラグメンテーションといった現象等の発生により未知試料のマススペクトルを読む際の障害となる場合が多いと知られており，高精度の微生物同定が困難な要因である．

3　ピークパターンによる菌株識別

微生物同定の主なタスクとして，菌種同定と菌株識別があると知れれている．菌種同定とは，未知の分類菌株がすでに記載されたどの菌種に最も近いかを決定する作業である．一方，菌株識別とは，登録された複数の菌株に最も近い菌株を識別する作業である．菌株識別で得られた知見は，菌種同定にも応用できる．

3.1　ピーク・ピーク検出

データセット内の局所的最大値のことをピーク (Peak) といい $< m/z, intensity >$ で表す．m/z は，質量電荷比 (m/z 値) であり，intensity は信号強度である．データセットから信号強度の局所的最大値とその位置を検出することは，ピーク検出 (Peak Detection) といい，スムージングとベースライン補正を行なった後に行う．ピークと強度のパターンは菌種によって異なるため，ピークパターンが菌種同定のポイントとなっている．

3.2　ピークアライメント

質量分析で得られたマススペクトルにも，横軸の m/z 値に僅かな誤差が生じることがあるため，同一ピークかどうかを判断する際に，このような誤差を補正するピークアライメントを行う必要がある．本研究では，ピーク検出で得られたピークに対し，m/z 値の差が一定範囲内であれば，同一ピーク ID を付与することで，$< peak\_id, intensity >$ のピーク列を利用する．

3.3　ピーク類似度

本研究では，ピークパターン類似度による菌株識別を提案する．二つのピーク列のピーク ID の集合をそれぞれ A, B するとき，ピーク列の類似度は次のように定義する．

$$s(A,B) = \frac{\sum_{x \in A \cap B} \sigma(x)}{|A \cup B|}$$

評価関数 σ の定義によって異なる類似度評価方式が得られる．(1)*Jaccard* **類似度**，共通ピークを区別せず一律 1，つまり $\sigma(x) = 1$ とカウントする．(2)*Rank* **類似度**, 共通ピークの信号強度順（「ランク」）が近いときのみ 1 とカウントする．(3) *Weighted Rank* **類似度**, 共通ピークのランクが近いとき，ランクによって違う重み w を加算する．例えば，$\sigma(x) = 1/r_A(x) + 1/r_B(x)$．ここで $r_A(x)$，$r_B(x)$ はそれぞれ x の A，B における信号強度ランクである．

4　実験結果

提案手法を検証するために評価実験を行った．実験では 51 菌株の計 102 のマススペクトルデータを使用した．ピーク検出に連続ウェブレット変換 (CWT) を使用し，各マススペクトルデータからおよそ 200 のピークを抽出した．ピークアラインメントの閾値を 3 とした．各菌株に対して，もっとも類似度の高い TopN 個の菌株を抽出し，そのうち，同一株のものが含まれた場合は正解と判定した．実験結果は図 1 に示している．全体的に Jaccard と Weighted がより良い結果が得られた．

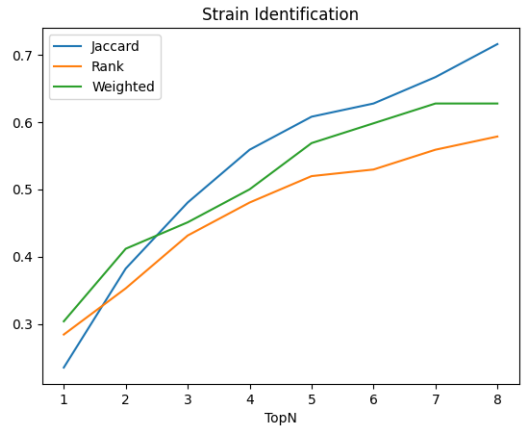

図 1: 各種類似度に基づく菌株識別の実験結果

5　まとめ

本研究では，質量分析データの類似度評価による微生物同定手法についてまとめた．[2] で我々が提案した内容に、ピークアライメントを追加し、さらに追加の評価実験を実施した．

謝辞

本研究の一部は，KSU 基盤研究費 K060300 助成を受けたものである．実験に九州産業大学総合機器センターのご協力を得ている．

参考文献

[1] 大楠清文，質量分析技術を利用した細菌の新しい同定法，モダンメディア 58 巻 4 号 2012, pp.113-122
[2] 浅野　公平，豊坂　祐樹，成　凱，質量分析データの機械学習による微生物同定 情報処理学会第 85 回全国大会，2023 年 3 月．