

ピーク形状に着目した機械学習による高分解能MALDI-TOFMSのマススペクトルからのピーク抽出方法開発と合成高分子分析への応用

関連製品: 質量分析計(MS)

はじめに

マトリックス支援レーザー脱離イオン化飛行時間質量分析計 (MALDI-TOFMS) は、ポリマーの分析において強力なツールである。高分解能MALDI-TOFMSを使えば、繰り返し単位や末端基の組成によるポリマーシリーズの識別が容易となり、イオン強度の分布からポリマーの分子量分布を算出できる。実際の工業材料分析においては、分子量分布や末端基の異なるポリマーの混合物が分析対象となり、複雑なマススペクトルを俯瞰できるケンドリックマスディフェクト (KMD) 解析の活用が進んでいる。KMD解析はKMDプロットと呼ばれる図上にポリマーシリーズが直線上に表現されることから、複雑なマススペクトル中に含まれるポリマーシリーズの数や相対量を可視化できる。また微量成分の発見も容易となるのも特徴である。KMDプロットの作成ではマススペクトルからピークを抽出するため、解析対象ピークとノイズピークを適切に識別することが重要となる。MALDI-TOFMSのマススペクトルでは、 m/z の増加にしたがって指数関数的にイオン強度が減少するノイズピーク群がしばしば観測される。これらはピーク幅が広く、形状が歪で再現性に乏しい。高分解能MALDI-TOFMS JMS-S3000 “SpiralTOF™”シリーズを用いた測定では、解析対象ピークは高分解能でピーク幅が狭いため、目視ではノイズピークとの識別が可能であるが、微量成分も含みマススペクトル全域にわたり識別を実施することは非効率である。また一般的な自動ピーク判定では、ピーク面積値をイオン強度とする。そのため幅が広いノイズピークは、解析対象ピークと高さが同等の場合、イオン強度がより大きくなるため閾値で一律に選別することが難しい場合もある。Figure 1は、プロファイルマススペクトルと一般的なピーク判定後の解析対象ピークとノイズピークについて図示したものである。ピークリスト中には便宜的に解析対象ピークを赤、ノイズピークを緑で色分けしている。プロファイルマススペクトルでは微量成分として1 u毎にノイズピークが観測された。プロファイルマススペクトルでは解析対象ピークは分解能から識別できるものの、ピーク判定後はノイズピークのイオン強度 (ピーク面積) が相対的に大きくなるため解析対象ピークの識別が難しくなっている。本報告ではこの課題を解決するため、ピーク形状に注目した教師データあり機械学習を用いて、マススペクトル中のピークが解析対象ピークかノイズピークか識別する手法を開発したので報告する。

実験

機械学習用のデータには、平均分子量400, 600, 1000, 2000のポリエチレングリコール (PEG) を10 mg/mLで調製し、1:1:2:4 (v/v/v/v) で混合した (PEG混合物)。また、低濃度PEG混合物として、PEG混合物を100倍希釈したものを準備した。マトリックスにはDCTB (10 mg/mL)、カチオン化剤にはトリフルオロ酢酸ナトリウム (1 mg/mL) を用いた。マススペクトルの取得には、JMS-S3000 “SpiralTOF™-plus”を用いてSpiralTOF正イオンモードを用いた。機械学習によるノイズ除去機能は、msPeakFinderに実装されている。またKMD解析はmsRepeatFinderで行った。

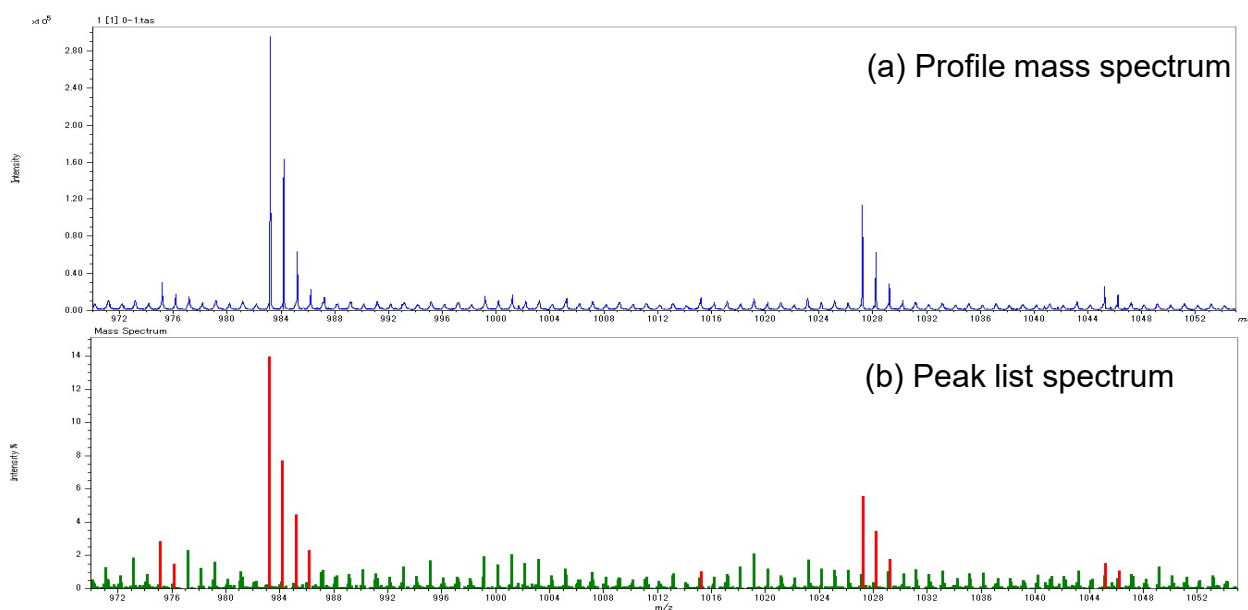


Figure 1 Profile mass spectrum of high-resolution MALDI-TOFMS (a) and peak list spectrum using conventional peak detection method.

機械学習の方法

機械学習には、条件付き敵対的生成ネットワーク (cGAN: Conditional Generative Adversarial Network) を採用した。cGANは、入力した条件データに応じた生成データを出力するため、条件データから生成データへの変換と考えることができる。本手法では、観測したマススペクトルを入力してノイズピークを除去した疑似マススペクトル (pseudo-mass spectrum) を出力することをコンセプトとして、ノイズピーク除去に応用した。Figure 2に本手法の機械学習モデル作成手順をフローチャートで示す。なおフローチャートで背景黄色は人の作業が介在し、緑は自動で行われる。まず、教師データ用にPEG混合物のマススペクトルを1つ取得した (Figure 3a)。得られたマススペクトルに対して従来法でピーク判定を行いピークリストを作成した後、熟練者がピーク形状から解析対象ピークを判断し抽出した (Figure 3b赤矢印)。抽出した解析対象ピークに対し、イオン強度にかかわらず一定の高さとなるようにガウス分布でピーク形状を作成し、疑似マススペクトルとした (Figure 3c)。本手法では、測定マススペクトルと疑似マススペクトルを対にして教師データの元データとした。さて、教師データの数を増やすためにマススペクトルを多数取得するには時間と労力を必要とする。そこで元データを1,024点ごとに分割、またその分割する始点を5回変更することで、1つの元データから合計1,600対の教師データを作成した。このようにして作成した教師データを用いて機械学習モデルを生成した。その概念図をFigure. 4に示す。測定マススペクトルはGeneratorにより疑似マススペクトルに変換される。この測定マススペクトルとGeneratorを通じて変換された疑似マススペクトルと、測定マススペクトルと教師データの疑似マススペクトルの組み合わせをDiscriminatorで真偽判定を行いながら、Generatorの質を高めていった。

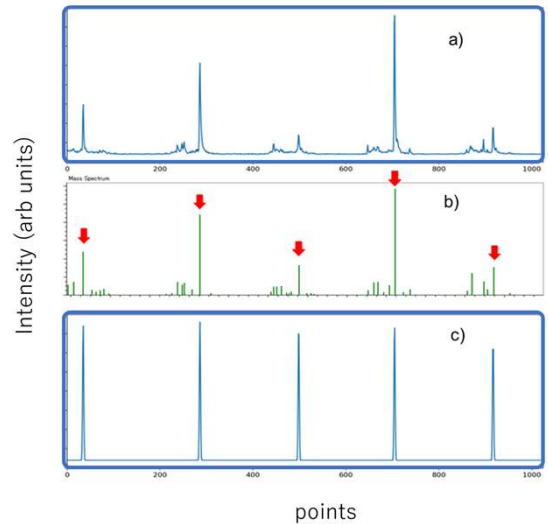
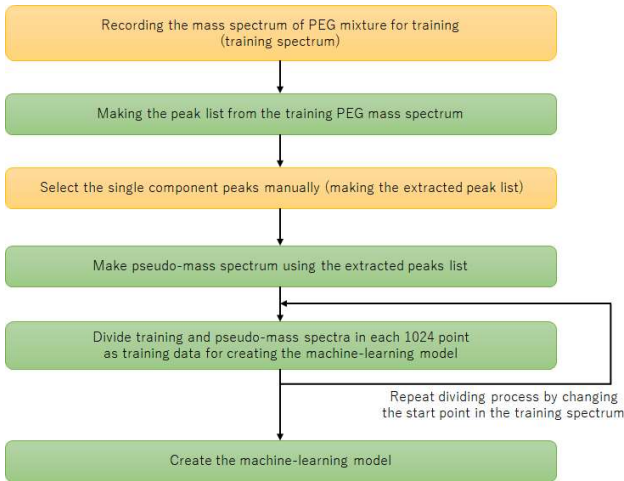


Figure 3 The relationship between profile mass spectrum(a), peak list(b) and pseudo-mass spectrum(c).

Figure 2 Flowchart of making the machine learning model.

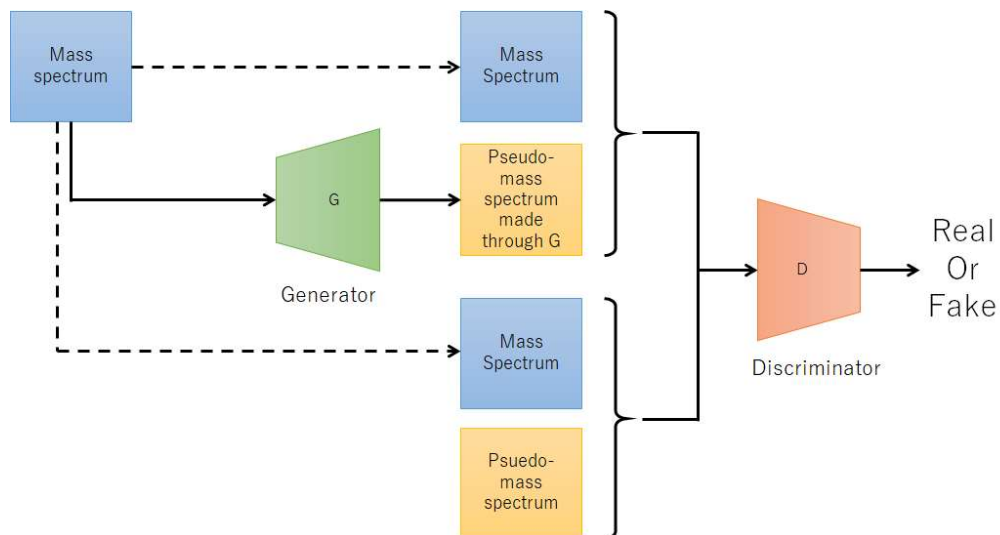


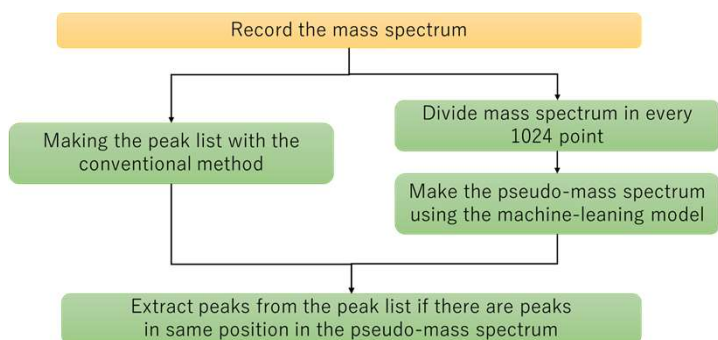
Figure 4 The scheme of making the machine learning model using cGAN.

機械学習モデルの妥当性確認と応用

次に生成した機械学習モデルを用いて実際のノイズ除去をする手順を示す (Figure 5)。なおフローチャートで背景黄色は人の作業が介在し、緑は自動で行われる。取得したマススペクトルに対し、従来法によるピーク判定を実施すると並行して1,024点ごとに分割し機械学習モデルを用いて疑似マススペクトルに変換する。従来法によりピーク判定されたピークのうち、疑似マススペクトルのピークの位置と一致するもののみを残しノイズ除去したピークリストを生成する。つまり、本手法により抽出されたピークリストの m/z やイオン強度は従来法のものを採用している。ここで、教師データ作成に用いたPEG混合物のマススペクトルを用いてノイズ除去を実施した。その結果をTable 1にまとめる。PEG混合物のマススペクトルからは従来法により合計4,390本のピークが検出された。そのうち左上の1,265本と3,105本(合計で全体の99.5%)は、教師データ作成時に行った判定と機械学習モデルでの判定結果が同等のものである。右上の14本は機械学習モデル作成時に解析対象ピークとして判断したが、機械学習モデルによりノイズピークとして判定されたものである。これらは確認の結果ピーク形状が少し歪で熟練者でも判断が難しいものであった。左下の6本は教師データ作成時にノイズピークとして判断したが、機械学習モデルにより解析対象ピークとして判定されたものである。確認の結果これらは教師データ作成時の人為的なミスによるものであることが分かった。その後、このミスを修正した教師データで再度機械学習を実施した。このように機械学習モデルの作成に用いたマススペクトルでモデルの検証をすることは有効であると考えられる。最後に、低濃度PEGのマススペクトルを使用してピーク抽出を行い、KMDプロットに展開した結果をFigure 6に示す。Figure 6aは測定したマススペクトルであり、Figure 6bはKMDプロットである。KMDプロットの赤い点は機械学習により解析対象ピークとして、緑の点はノイズピークとして判定されたものである。この結果から、特に $m/z < 1,500$ においてノイズが除去されることにより、PEGのシリーズが可視化できているのが分かる。

まとめ

以上のように、高分解MALDI-TOFMSのデータから低分子領域にしばしば観測されるノイズピークを機械学習モデルにより除去することで、KMD解析がより効率的に実施できることが示すことができた。



Result		Single component peaks	Noise peaks
		Training data	
Single component peaks		1265	14
Noise peaks		6	3105

Figure 5 Flowchart of making the extract peak list by the machine learning model.

Table 1 Comparison between the peak lists of PEG mixture used as the training data and the one extracted by the machine learning model.

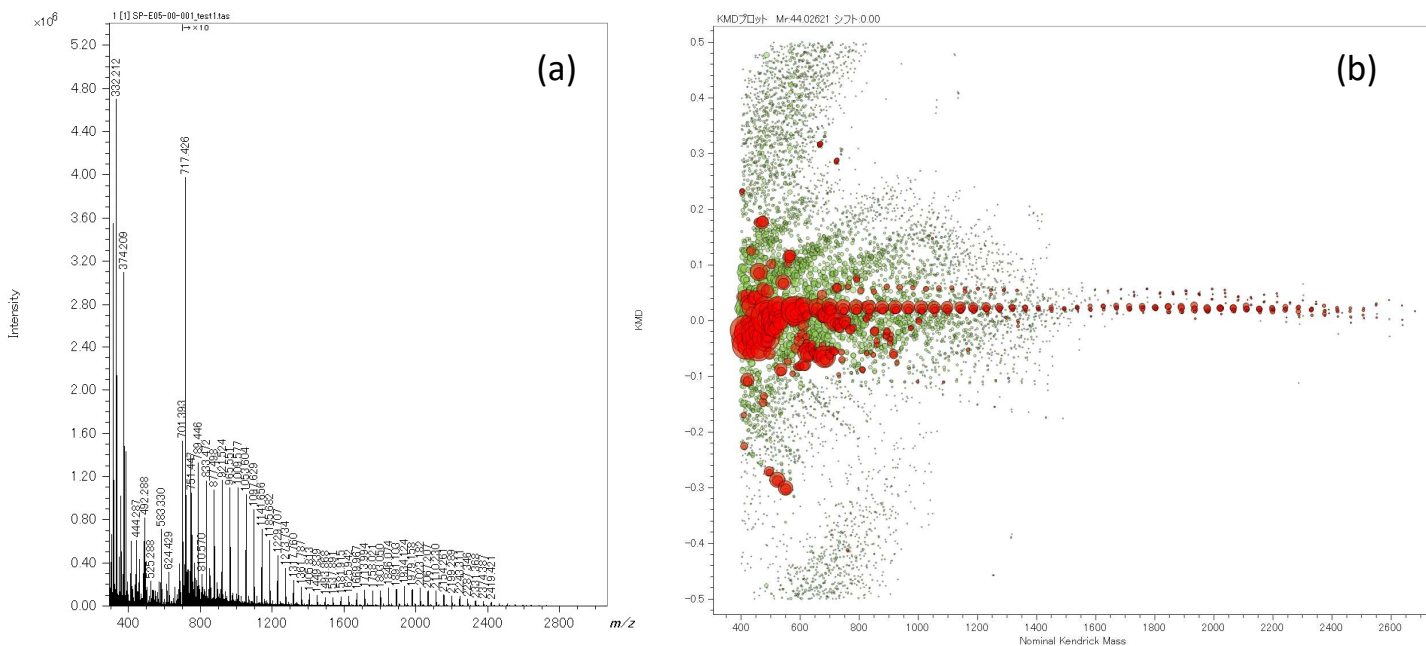


Figure 6 Mass spectrum of low concentration PEG mixture(a) and the KMD plot of the extracted peak list (red) and the noise peak list (green) separated by the machine learning model.

このカタログに掲載した商品は、外国為替及び外国貿易法の安全輸出管理の規制品に該当する場合がありますので、輸出するとき、または日本国外に持ち出すときは当社までお問い合わせください。 Copyright © 2021 JEOL Ltd.

