

博士學位論文審査要旨

学位申請者氏名	松田 和大
論文題目	生体試料の TOF-SIMS 分析における教師なし機械学習の応用
審査委員 (職名・氏名・印)	
主査	教授 青柳 里果
審査委員	教授 富谷 光良
	教授 中野 武雄
	教授 武藤 俊介
論文審査結果 (合 否)	合 格
論文審査の要旨	<p>本論文では、飛行時間型二次イオン質量分析法 (time-of-flight secondary ion mass spectrometry : TOF-SIMS) によって得られた生体試料などの複雑な計測データの自己符号化器 (オートエンコーダー) を用いた解析に成功し、次のような知見を得ている。</p> <p>1. 毛髪試料の TOF-SIMS による深さ方向分析データの自己符号化器による解析</p> <p>TOF-SIMS は、固体試料の表面の化学構造情報や3次元分子イメージングが得られる手法であり、特に最近では生体組織の組成分布を可視化するための質量イメージング手法としての活用が進んでいる。TOF-SIMS からはスペクトルおよびイメージデータとして試料に関する膨大な情報が得られるため、TOF-SIMS データの解釈が難しい場合が多く、これまでは主成分分析などの多変量解析手法が TOF-SIMS データの解析に応用されてきた。しかし、装置開発の発展により TOF-SIMS データはさらに複雑かつ大規模となり、より柔軟で応用性の高い解析手法の適用が望まれている。ただし、TOF-SIMS データの解析では、一般に、注目する成分の2次元もしくは3次元分布イメージと注目する成分に関連する物質を示す質量ピーク情報 (スペクトル) を得ることが重要である。そのためには TOF-SIMS データを分類するだけでなく、解析過程が明らかとなる解析手法が要求される。例えば、スペクトルとイメージを併せ持つ TOF-SIMS データの解析に主成分分析を応用した場合は、一般に、主成分得点として分布イメージが得られ、負荷量として質量ピーク情報が得られる。このような結果が得られる手法として、本論文では人工ニューラルネットワークを用いた教師なし手法の一つである自己符号化器に注目した。</p> <p>自己符号化器は、数値化した測定データを入力層とし、入力層をエンコードした中間層 (隠れ層) で特徴抽出し、出力層が入力層を再現できるように学習を進める。中間層の各ノード (ニューロン) が特徴成分の分布イメージを示し、入力層から中間層にエンコードする際の重みもしくは中間層から出力層にデコードする際の重みが、各ノードが示す特徴成分に重要な質量ピークを示す。そこで、毛髪試料を深さ方向分析した1次元の TOF-SIMS データを自己符号化器で解析した結果、毛髪のキューティクルの層状構造に対応した特徴が抽出できることが示された。</p>

論文審査の要旨（続）

2. 皮膚試料の TOF-SIMS データの自己符号化器による解析

次に、2次元のイメージを持つ皮膚試料の TOF-SIMS データを解析対象として、L1 正則化もしくは KL ダイバージェンスによる正則化を導入したスパース自己符号化器の応用を検討した。対象試料は生体内には存在しない成分（ジクロフェナク）を塗布した皮膚試料であり、ジクロフェナクに由来する成分や代表的な生体分子に関連する成分の分類が適切に行われるかどうか評価した。

本試料の TOF-SIMS の質量スペクトル上には約 500 個の質量ピークが存在するが、各純成分に関連するのはそれらの質量ピークのうちの一部分である。しかし、自己符号化器でエンコードする際には、全ての質量ピークに重みが割り当てられて、特徴抽出される。L1 正則化はおもに重みに対する制約となり、各特徴に寄与できる質量ピークが制限される。また、本試料の場合は、成分のほとんどは試料内に均一に存在するのではなく、一部にのみ分布するため、分布イメージはスパース（疎）となる。KL ダイバージェンスによる正則化は中間層をスパースにする効果が期待できる。

皮膚試料の TOF-SIMS データを自己符号化器で解析する際に、正則化の種類や正則化の程度を変化させたところ、正則化なしの場合と比べて、L1 正則化を加えることによって、ジクロフェナクに関する成分が他の成分と明確に分離され、信号ノイズ比 (S/N) もより高い質量ピーク情報が得られた。KL ダイバージェンスによる正則化を加えた場合も同様にジクロフェナクが明確に分離され、さらに脂肪酸の炭素鎖長の違いも分類可能な結果が示された。このように、本論文では、適切な条件の下では、スパース自己符号化器は、単純な自己符号化器に比べて特徴抽出性能が顕著に向上することを示した。

3. 有機物積層試料の TOF-SIMS データの自己符号化器による解析

TOF-SIMS による試料測定では、共存物質や試料状態によって質量ピーク強度が変化するマトリックス効果が課題の一つとして挙げられる。一般的な試料の TOF-SIMS データにおいて、マトリックス効果によって、多くの質量ピーク強度が濃度に対して非線形な応答を示すため、定量分析が難しい場合が多い。そこで、二種類の有機物を異なる濃度比で混合した層を積層したモデル試料を TOF-SIMS で深さ方向分析して得られた 3次元データを試料断面についての 2次元データとしたデータセットを対象にし、スパース自己符号化器による解析で濃度応答性が改善できるかどうか検討した。結果として、スパース自己符号化器によって、マトリックス効果の影響を軽減し、濃度依存性の高い強度を示す特徴成分が抽出できることを示した。これは非線形近似が可能な人工ニューラルネットワークに基づく解析手法による TOF-SIMS データ解析の有用性を示した結果と考えられる。

このように、本論文では、人工ニューラルネットワークに基づいた教師なし手法である自己符号化器を生体試料の TOF-SIMS データに対して適用し、詳細に検討することにより、正則化を加えることによって、複雑な試料データから適切かつ効果的に特徴抽出ができることを示した。さらに、非線形な濃度応答性を持つデータの濃度応答性を改良できる可能性も示唆した。本論文で得られた知見は、今後の大規模分析データへの機械学習の適用に大きく役立つと考える。

(以上)