2021 年度

博士論文

生体試料の TOF-SIMS 分析における教師なし機械学習の応用

成蹊大学大学院理工学研究科

理工学専攻

松田和大

目次 2
第一章 序論・研究背景 5

第二章 既往研究 10
2.1 飛行時間型二次イオン質量分析法 11
2.1.1 二次イオン質量分析法の概要 11
2.1.2 飛行時間型二次イオン質量分析法の原理と特徴 12
2.1.3 TOF-SIMS に使用されるイオンビームの種類と最近の装置開発動向 16
2.1.4 TOF-SIMS のデータ構造 18
2.2 機械学習（Machine learning） 19
2.2.1 概要 19
2.2.2 ニューラルネットワークの基礎 20
2.2.3 活性化関数（Activation function） 22
2.2.4 損失関数（Loss function） 23
2.2.5 最適化関数（Optimizer） 24
2.2.5.1 勾配降下法（Gradient Descent） 24
2.2.5.2 Deep neural network 用に開発された高速な最適化関数 26
2.2.6 自己符号化器（Autoencoder） 27
2.3 多変量解析（Multivariate analysis: MVA） 28
2.3.1 概要 28
2.3.2 主成分分析（Principal component analysis: PCA） 28
2.3.3 多変量スペクトル分解（Multivariate curve resolution: MCR） 29
2.4 データ前処理 30
2.5.生体試料（Biological samples） 33
2.5.1 生体を構成する主要成分 33
2.5.2 生体試料の組成イメージング手法 33

第三章 ヒト毛髪のデプスプロファイルデータに対するオートエンコーダーの適用検討 35
3.1 はじめに 36
3.2 実験方法 39
3.2.1 分析試料調整 39
3.2.2 TOF-SIMS 測定条件 39
3.3 データ解析 39
3.3.1 データ前処理 39
3.3.2 データ解析条件 40

3.3.3 計算実行条件 41
3.4 結果と考察 41
3.4.1 TOF-SIMS 測定結果 41
3.4.2 オートエンコーダーによる学習 44
3.4.3 抽出された特徴の可視化 46
3.4.4 Encoder Weighs と Decoder Weights の比較 48
3.4.5 抽出された特徴の妥当性の検証 54
3.4.6 PCA との結果の比較 57
3.4.7 抽出された特徴に対する中間層サイズの影響 59
3.5 結論 61

第四章 TOF-SIMS 二次元イメージデータに対するスパースオートエンコーダーの適用検討 62

4.1 はじめに 63
4.1.1 スパースオートエンコーダー 63
4.1.2 ヒト皮膚の構造と機能 64
4.2 実験方法 65
4.2.1 分析試料調整 65
4.2.2 TOF-SIMS 測定条件 65
4.3 データ解析 66
4.3.1 データ前処理 66
4.3.2 データ解析条件 66
4.4 結果と考察 67
4.4.1 TOF-SIMS 測定結果 67
4.4.2 オートエンコーダー（正則化なし）による解析 69
4.4.3 スパースオートエンコーダー（L1 正則化）による解析 74
4.4.4 スパースオートエンコーダー（KL-divergence 正則化）による解析 80
4.4.5 正則化項の違いが特徴抽出性能に与える影響 83
4.4.6 学習時のバッチサイズの影響評価 84
4.5 結論 85

第五章 スパースオートエンコーダーと他の特徴抽出法の比較 88
5.1 はじめに 89
5.2 実験方法 89
5.3 データ解析 89

5.3.1 データ前処理 89
5.3.2 データ解析条件 89
5.4 結果と考察 90
5.4.1 三種類の特徴抽出法により抽出された特徴 90
5.4.2 スパースオートエンコーダーと他の特徴抽出法の結果の比較 97
5.5 結論 104

第六章 スパースオートエンコーダーによるマトリックス効果補正の検証 105
6.1 はじめに 106
6.2 解析データ 106
6.3 データ解析 107
6.3.1 データ前処理 107
6.3.2 データ解析条件 107
6.4 結果と考察 108
6.5 結論 118

第七章 結論 119

参考文献 122
研究業績一覧 129
謝辞 等 130

第一章
序論・研究背景

細胞・組織レベルで生体分子や金属元素の分布を捉え可視化するバイオイメージング技術は、生体の機能の理解するうえで欠かすことのできない技術である。一方で、医学・薬学分野においては、近年、抗体医薬品や核酸医薬品といった従来の低分子医薬品に比べて特異性の高い（ターゲットとなる病原あるいは生体分子を正確に認識する）新しいタイプのモダリティの開発や、薬物を必要とする部位に必要な量だけ送達するための薬物送達システム（Drug Delivery System: DDS）の開発・実用化も進んでいる[1]。これら新モダリティや DDS 技術の開発を行うにあたっては、生体組織中、場合によっては単一細胞レベルで薬物や生体分子の分布を調べる必要がある。そのような要求を叶えるために様々な新規のバイオイメージング技術の開発が行われている。

バイオイメージングの技術として最も一般的に用いられているのは、光学顕微鏡をベースとした技術である。その中でも蛍光顕微鏡法は、光励起によって蛍光を発する分子（蛍光プローブ）でターゲットとなる薬物分子を標識することで、感度良く薬物分子の局在を評価することが可能である。また、生体組織中に内在する生体分子についても、蛍光プローブで標識した抗体を用い、抗原抗体反応を利用して選択的に蛍光標識することで可視化することが可能である（所謂、免疫染色法）。しかしながら蛍光を利用したイメージング手法では、蛍光標識により薬物動態に影響が出る可能性や、光毒性により周囲の細胞にダメージを与える懸念が指摘されている[2]。また、多くの成分の分布状態を網羅的に評価することはその原理から困難である。さらに、生体分子そのものが持つ自家蛍光の影響や蛍光の消光といった原因により、目的とする分子のイメージングが十分でない場合もある[3]。蛍光プローブを用いずに他成分を網羅的に評価する手法として、近年、質量イメージング技術（Mass Spectrometry Imaging: MSI）が注目を集めている[4]。MSI はレーザーや帯電液滴、イオンビームを用いて試料を直接イオン化させ、質量分析計で定性する手法である。MSI は分子そのものを検出するため高い選択性を持ち、例えば薬剤の未変化体と代謝物を区別してイメージを取得することも可能である。この点は、蛍光プローブを用いた蛍光顕微鏡法や、放射性同位体標識を行うオートラジオグラフィーにない特徴である。MSI の種類としては、試料表面に照射するプローブの種類によって、マトリックス支援レーザー脱離イオン化を用いたもの（MALDI-MS）や脱離エレクトロスプレーイオン化を用いたもの（DESI-MS）、イオンビームを用いたもの（SIMS）などがあり、それぞれの特徴を活かした評価に用いられている[5]。

イオンビームを用いた MSI の一手法である飛行時間型二次イオン質量分析法（TOF-SIMS）は、固体表面の約 1〜3 nm に存在する微量成分（元素・分子）を、高い空間分解能（〜300 nm）でイメージングできることから、半導体[6, 7]や高分子材料[8, 9]、生体関連材料[10-12]、生体試料[13-18]といった、様々な試料の表面分析手法として広く使用されている。特にその高い空間分解能は、上述の単一細胞レベルでの薬物分布の解析にとって非常に魅力的な能力として認知されている。しかし、TOF-SIMS のデータは主に次の 3 点の理由から複雑になりがちである。

① イオンビーム照射によるイオン化は、分子構造を保持した状態でイオン化した分子イオンのほかに、イオン化の過程で断片化されたフラグメントイオンなどが多く生成されること。
② 二次イオンピークの重ね合わせが生じやすく、一つのピークが複数の物質に由来すること。

③ 固体の最表面は一般に複数の成分が存在する混合系となっていること。
④ 混合状態によっては、共存物質の影響によって、二次イオン強度が増大されたり抑制されたりするマトリックス効果が生じ、二次イオン強度と物質の濃度の関係が非線形になること。
⑤ 装置構成上の問題や試料表面から発生したイオンの総量の少なさから、質量分析の前段に成分分離用の機構を備えることが困難であること。

これらの問題は、特に生体試料の分析において顕著である。すなわち、生体試料は金属や無機塩、低分子量の脂質、高分子量のタンパク質や核酸など、様々な生体分子が混在しているため、TOF-SIMS によって得られるデータは一般的な工業材料のそれと比べ一層複雑なものとなりやすい。近年、クラスターイオン（C_{60} [19-21]、Ar[22-24]、H_2O[25,26]、CO_2[25]など）を一次イオンとして用いることでフラグメントイオンの生成を抑制できることがわかり、これらのイオンビームの使用はデータの複雑さの低減に貢献している。しかしながら一方でクラスターイオンによるイオン化では、従来使用されていたイオンビームではイオン化が不可能であった高分子量（*m/z* 2000 以上）の分子の感度の増加という効果もあることから、データの量としては却って増加する傾向がある。それに加え、質量イメージングとイオンエッチングを組み合わせた三次元イメージング技術の開発や、画像解像度や質量分解能の向上といった装置の仕様の向上が組み合わさり、TOF-SIMS のデータ量は飛躍的に増加する傾向にある。このような背景から、TOF-SIMS データを効率的に解釈可能なデータ解析方法が求められている。

TOF-SIMS のデータ解析法としては 2000 年頃より、主成分分析（Principal Component Analysis: PCA）[27-29]や、多変量スペクトル分解（Multivariate Curve Resolution: MCR）[27, 29-31]を代表とした非負行列因子分解（Non-negative Matrix Factorization: NMF）といった、多変量解析手法の適用が検討されてきた。線形データを対象とした線形モデルであるこれらの解析手法は、マトリックス効果などによって非線形の応答を示す TOF-SIMS データの解析には最適ではないものの、一定の成果をあげてきた。一方で、動物の脳内のシグナル伝達を模した人工ニューラルネットワークは、各ニューロンを繋ぐ伝達関数（活性関数）に非線形関数を用いることで、非線形近似が可能であり、マトリックス効果の影響を含む TOF-SIMS データをより適切に扱える可能性がある。

Hinton らは 2006 年に人工ニューラルネットワークを用いた次元圧縮法としてオートエンコーダー（自己符号化器）を提案した[32]。このオートエンコーダーは複雑なデータからの特徴抽出法としての用途や、異常検知などに応用されている。更に 2012 年に行われた ImageNet Large Scale Visual Recognition Challenge 2012（ILSVRC2012）において、Hinton らは人工ニューラルネットワークを積層したディープラーニングによって画像認識を高精度に行えることを示し、画像認識分野におけるディープラーニングの有用性を示した[33]。これらの成果から、医療分野やインフラ分野をはじめとした多くの分野において、大規模データの分類や回帰などを高精度かつ高速に行えるシステムとして、人工ニューラルネットワークの応用研究が多数行われている[34-35]。この流れは分析データの解析においても同様であり、分光スペクトルデータの解析[36]や質量スペクトル・質量イメージングの解析[37-41]などへの適用検討が行われている。

質量イメージングに関しては、生物試料の質量イメージングデータに対して、人工ニューラルネットワークに基づく教師なしの解析手法であるオートエンコーダーを用いることで、PCA や NMF に比べて、生物学的に意味のある特徴抽出がで実現できる可能性が 2016 年に Thomas らによって示されている[38]。しかしながら、その詳細については検討されていない。TOF-SIMS データに対するオートエンコーダーの適用については、2020 年の Kawashima らの論文[37] に示されているが、この研究でも、オートエンコーダーの結果は MCR との比較に用いられているだけで、オートエンコーダーがどのように TOF-SIMS データの解析に有効なのか明らかにはされていない。そのほか、ニューラルネットワークに基づいた TOF-SIMS データの解析としては、Pigram らによる自己組織化マップを用いた検討[40, 41] がある。この検討において自己組織化マップの有用性は示されているが、解析対象は生体試料に比べるとかなり単純化された高分子や有機物のモデル試料の解析にとどまっている。

以上のことから、TOF-SIMS データの解析への人工ニューラルネットワークの適用については、意義があると考えられる。しかし上述した通り、体試料を対象とした TOF-SIMS 画像データに対する人工ニューラルネットワークの適用に際して、パラメーターの条件検討をはじめとした、詳細な検討はほとんどなされていない。そこで本研究では、人工ニューラルネットワークを利用したオートエンコーダーを生体試料の TOF-SIMS 画像データからの特徴抽出に応用し、生物学的な知見や従来から使用されてきた多変量解析手法による特徴抽出結果との比較を通じて、オートエンコーダーを実施するうえでのパラメーター設定などの指針を得ることを目的とした。

はじめに第三章では、二次元の画像データの解析を行う前段として、ヒト毛髪の一次元デプスプロファイルデータに対し、入力層、中間層、出力層の各一層ずつによって構成される最も単純な構造のオートエンコーダーを適用し、学習の結果として得られる各種パラメーターの中から特徴の解釈に有用なパラメーターを明らかにするとともに、中間層のサイズが特徴抽出結果に与える影響について検証した。また、特徴抽出結果を生物学的に既知の知見と照らし合わせて、有用な結果が得られているかどうかについても検証を行った。

第三章の結果を基に第四章において、ヒト皮膚組織の二次元画像データに対してオートエンコーダーによる特徴抽出を行った。さらに、特徴抽出性能を向上させるための方法としてスパースオートエンコーダーに着目した。オートエンコーダーによって中間層に抽出される特徴（すなわち本研究の場合は TOF-SIMS によって得られる試料中の物質を分類した成分イメージにあたる）が特定の分布を持つスパース性の高いイメージであることから、中間層にスパース性を課すオートエンコーダーが有効であると考えられる。人工ニューラルネットワークを用いた解析では、PCA に比べて検討すべきパラメータが多い上に、スパース性も課すことによってさらにパラメータが増える。実際にオートエンコーダーを TOF-SIMS データをはじめとする質量イメージングやその他の化学イメージングが可能な手法に応用する際に、こうした複雑なパラメーター設定はこの手法の応用の妨げとなる可能性もある。そこで、本研究では、パラメーター設定が解析結果に及ぼす基本的な影響を、単純なオートエンコーダーとの特徴抽出性能の違いなども踏まえて検討し、解析条件設定の指針を導き出すことを目標とした。

第五章では、第四章で最終的に得られたスパースオートエンコーダーによる特徴抽出結果と、従来から使用されている多変量解析手法（PCA、MCR）によって得られた特徴抽出結果を比較し、スパースオートエンコーダーの特徴と応用の方向性について議論した。

第六章では、非線形近似が可能な人工ニューラルネットワークを利用するオートエンコーダーが、マトリックス効果によって濃度への応答が非線形となるような試料の TOF-SIMS データについて、適切な濃度応答性を示すことができるか検討した。

第七章では、第三章から第六章の検討結果を踏まえて本研究で得られた知見をまとめると共に、人工ニューラルネットワークを活用した TOF-SIMS データの解析の将来性について述べ、本研究における結論とした。

＜備考＞

本研究を遂行するにあたり、実施したヒト由来組織を用いた実験（測定）は全て、株式会社東レリサーチセンターの倫理委員会の承認の下に行った。また、毛髪・皮膚組織の提供者からは書面によるインフォームド・コンセントを得ている。

第二章
既往研究

2.1 飛行時間型二次イオン質量分析法

2.1.1 二次イオン質量分析法の概要

真空中に置かれた固体試料表面に対し、数～数十 keV に加速させたイオンビーム（一次イオン）を照射すると、スパッタリングによって試料表面から様々な粒子が放出される。特に放出されたイオン（二次イオン）を質量分析計により検出することで、試料の組成分析を行う分析手法を二次イオン質量分析法（Secondary ion mass spectrometry: SIMS）という[42]。SIMS は一次イオンの電流密度により、Dynamic-SIMS（D-SIMS）と Static-SIMS（S-SIMS）に大別される。D-SIMS は高い電流密度により試料表面がエッチングされるため、表面数 nm～数十 μm の領域の元素組成情報を深さ方向に調べること（デプスプロファイル測定）が可能であり、半導体素子や材料における不純物分析法の中でも必須の手法として、広く工業的に利用されている。D-SIMS は更に二重収束型と四重極型の質量分析計を用いた 2 つのタイプに細分され、それぞれ感度や深さ分解能、二次イオン検出の特異性（質量分解能）などが異なることから、それぞれの持つ特徴を活かした評価に用いられている。一方で後者の S-SIMS は、一次イオンをパルス化し低電流密度で照射することで、試料最表面（約 1～3 nm）に存在する原子および分子を、選択的にイオン化させることが可能である。S-SIMS は質量分析計に飛行時間型（Time-of-flight: TOF 型）を採用したものが現在、主流である。これは、D-SIMS に比べて試料表面から発生する二次イオン総量が桁違いに少ないことから、一定質量範囲内のイオンを平行（同時）検出することが可能な飛行時間型との相性が良いことによる。この、低い電流密度の一次イオンと飛行時間型質量分析計の組み合わせによって、試料最表面の無機物・有機物の組成情報を得ることに盛んに利用されている。

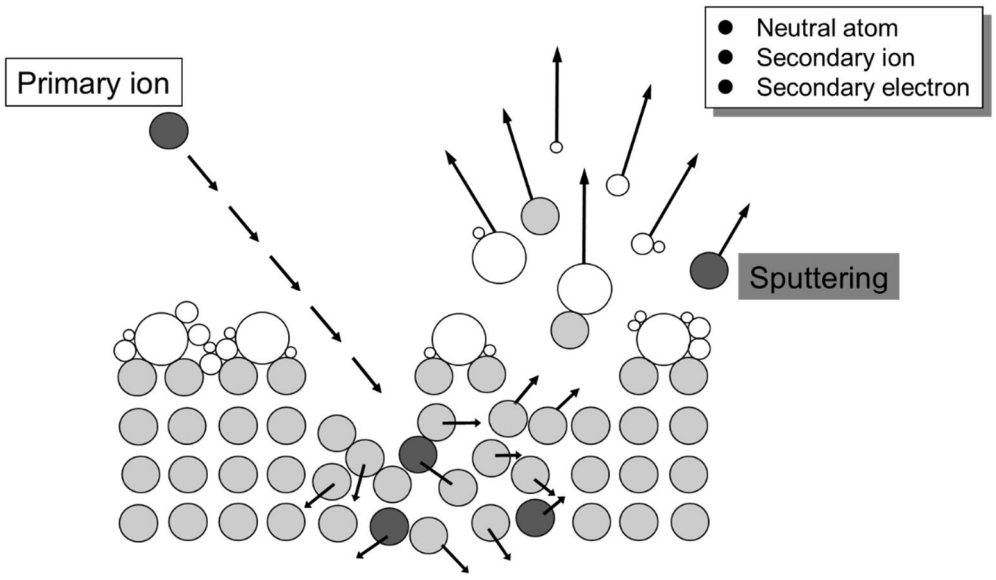


Figure 2.1 SIMS の原理図

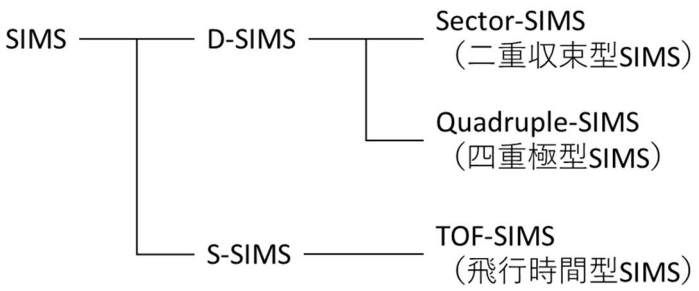


Figure 2.2 代表的な SIMS の分類

Table 2.1 SIMS の質量分析計による性能の違い [43]

Type	Mass resolution ($m/\Delta m$)	Mass range (m/z)	Transmission (%)	Mass detection	Relative sensitivity
Quadrupole	10^2 - 10^3	$<10^3$	1 – 10	Sequential	1
Sector	10^4	$>10^4$	10 – 50	Sequential	10
Time-of-flight	$>10^3$	$<10^3$ - 10^4	50 - 100	Parallel	10^4

2.1.2 飛行時間型二次イオン質量分析法の原理と特徴

1969 年に A. Benninghoven らは、イオン照射量を極めて低く抑えることによって、固体試料の最表面（1～3 nm）に存在する有機物の構造情報を取得できること（S-SIMS）が見いだした[44]。イオン照射により表面の有機物の分子構造は破壊されるが、一部の分子は破壊を免れ分子イオンとなるったり、ある程度の構造を保ったフラグメントイオンとなるため、それらを高質量分解能の質量分析計で検出することで、有機物の構造解析を行うことができる。表面のすべての有機分子を破壊するまでの一次イオン照射量は Static-Limit（～1×10^{13} ions/cm^2）と呼ばれ、最表面の分析のためには照射量をそれ以下に抑える必要がある。現在、市販されている多くの S-SIMS では、Static-Limit の範囲内で発生した二次イオンを効率よく検出するために、透過率が高く（50～100 %）、広い質量範囲の二次イオンを一斉（平行）検出可能な TOF 型質量分析計との組み合わせが一般的である。そのため「S-SIMS ≒ TOF-SIMS」として広く受け入れられている。

TOF-SIMS では、試料表面から放出された二次イオンは一定の電圧(V_{accl})で加速され、質量分析計内を質量(m)に応じた速度(v)で飛行し、検出器まで導かれる。このとき、二次イオンの持つエネルギー(E)は各パラメーターを用いて次式で表される。ここで、z はイオンの電荷数、e は電気素量(1.602×10^{-19})である。

$$E = z \cdot e \cdot V_{accl} = \frac{1}{2}mv^2$$

検出器までの距離(l)とV_{accl}は一定であることから、検出器に到達するまでの飛行時間(t)は質量(m)と電荷数(z)の関数となり次式で表されるが、TOF-SIMS で通常観測される 2 次イオンは一価のイオンであることから、$z = 1$である。すなわち、飛行時間(t)の分布を精密に計測することによって二次イオンの質量分布、すなわち質量スペクトルが得られることがわかる。

$$t = \frac{l}{v} = l \cdot \left(\frac{m}{2zeV_{accl}}\right)^{\frac{1}{2}}$$

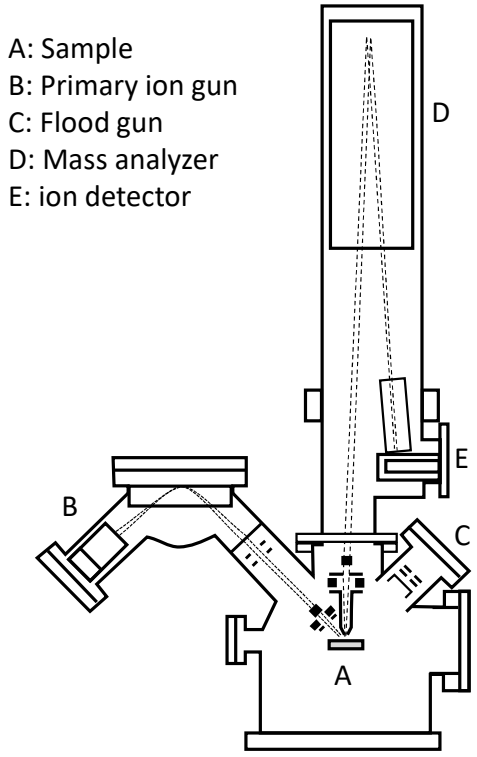


Figure 2.3 Reflectron 型の TOF-SIMS の一般的な構成

（表面分析：SIMS －二次イオン質量分析法の基礎と応用－[45]、p.91 の図を基に作成）

前述の通り TOF-SIMS は他の SIMS に比べて、①最表面（1～3 nm）分析が可能、②有機物の構造情報を取得可能という特徴を持つ。さらに TOF-SIMS では一次イオンにビーム径を細く収束させることが可能な液体金属イオン源（Liquid metal ion gun: LMIG）を使用することで、水平方向の空間分解能が高く、他の SIMS に比べてより微小領域の測定や、マッピング分析が可能という特徴も持つ。一例として Bi 一次イオンビームを用いて、$Al_{0.7}Ga_{0.3}As$ / GaAs のライン＆スペース試料（BAM L200）をイメージングしたデータを Figure 2.4 に示した。幅幅 20 nm 程度まで良好な解像結果が得られている様子がよくわかるデータである。

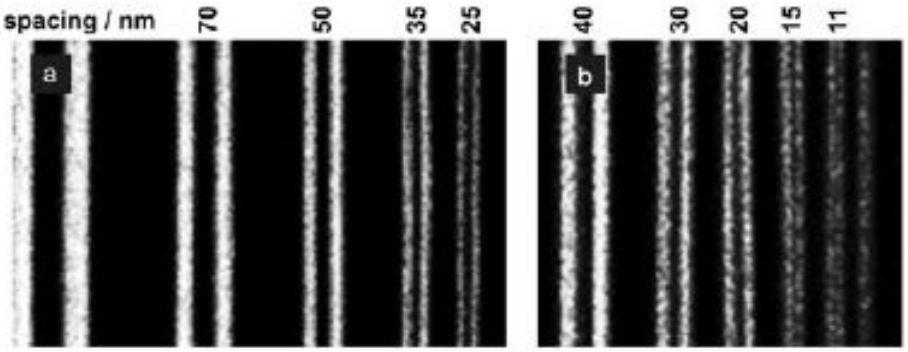


Figure 2.4　標準試料（BAM L200）を用いた Al パターンのマッピング
測定領域　(a):1.8 µm×1.0 µm、(b):0.9 µm×0.7 µm [46]

TOF-SIMS は元々、最表面分析に利用されていたが、近年、深さ方向分析への活用事例も多く報告されている[47-49]。この場合、一次イオンの電流密度が小さく、それ自体で試料をエッチングすることが困難であることから、試料のエッチング専用のイオンビームを別途装備する必要がある。エッチング用のイオンビームとしては、D-SIMS に使用される Cs^+や O_2^+のほか、後述するクラスターイオン（C_{60}^+, Ar_{1500}^+）が用いられている。この 2 本のイオンビームを用いて深さ方向分析を行うことを特にデュアルビームデプスプロファイリングという。TOF-SIMS ではエッチング用イオンビームを目的に応じて使い分けることが可能であり、元素情報を高感度に取得したい場合は Cs^+または O_2^+を使用し、有機物の構造情報を取得したい場合は後述するクラスターイオン（Ar_{1500}^+など）を使うなど、柔軟なセッティングが可能である。さらに一次イオンの高い空間分解能を活かした 3 次元マッピングも可能であり、多数報告されている[50-51]。深さ方向分析の例として Figure 2.5 に有機物の積層膜の分析事例を、Figure 2.6 にポリマー材料中の顔料分布を 3 次元マッピングで可視化した事例を示した。

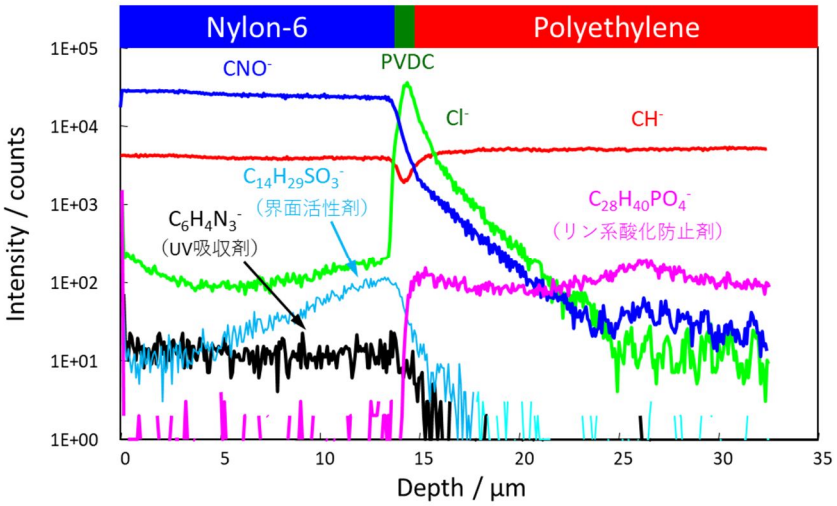


Figure 2.5 包装用フィルム（Nylon-6、ポリ塩化ビニリデン：PVDC、ポリエチレンの積層）の TOF-SIMS デプスプロファイル

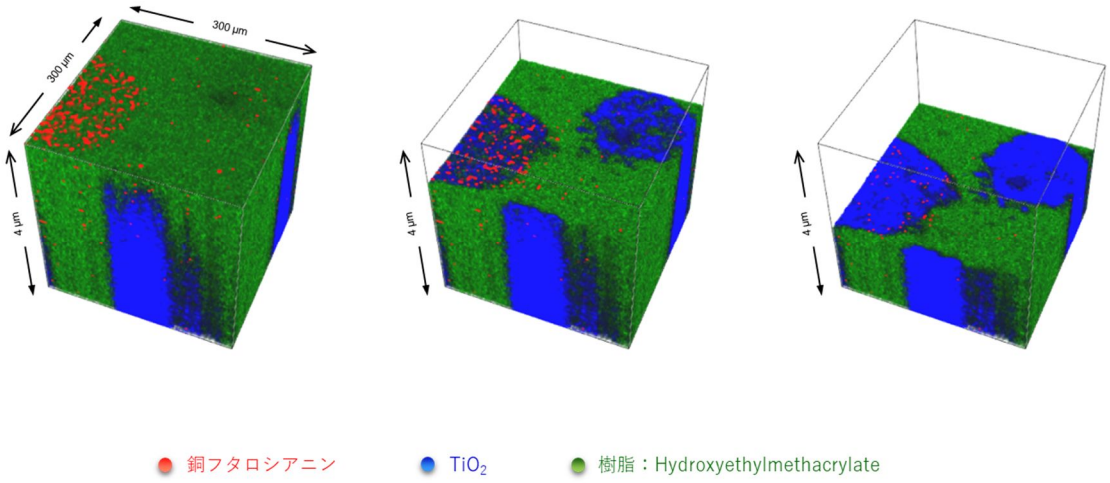


Figure 2.6 装飾用カラーコンタクトレンズの TOF-SIMS 3D イメージ

2.1.3 TOF-SIMS に使用されるイオンビームの種類と最近の装置開発動向

TOF-SIMS は表面分析から深さ方向分析、無機物（元素分析）から有機物までと、測定方法や測定対象が広いことから、原型の発明からこれまで継続的な要素技術開発がなされてきた。その中で、特に多くの開発が活発な分野がイオンビームの種類である。

TOF-SIMS の黎明期には、Cs^+や Ar^+といった単原子イオンが主に使用されていたが、微小部分析のために Ga-FIB を用いる装置が一般化した[52]。しかしながら、Ga イオン源では質量（*m/z*）が 200 程度までの元素や分子の検出については一定の結果が得られたものの、それ以上の質量を持った分子についてはイオン化率が非常に低く、感度が著しく低下してしまうという問題があった。このような有機分子の分析感度を高めるためのアプローチとして、多原子一次イオン（クラスターイオン）の開発がある。1987 年、Appelhans らは、非導電性ポリマー表面への帯電効果を避けるため、中性の SF_6 ビームを適用したが、この際、二次イオンの収率が同等エネルギーの原子イオンと比較して数桁向上したという副次的かつ予想外の結果が得られた[53]。この報告から、高質量側のイオン化率を向上するために一次イオンにクラスターイオンを用いるというアイデアが生まれ、以降、様々なクラスターイオンの開発が行われてきている。クラスターイオンにより有機物の二次イオンの高質量側で収率が向上した原因については、クラスターイオンの衝突時にエネルギーが横方向に広がり、その結果として Figure 2.7 に示すフラグメントイオンや分子イオンの発生するエネルギーが与えられた領域が広くなったためと定性的に理解されている。

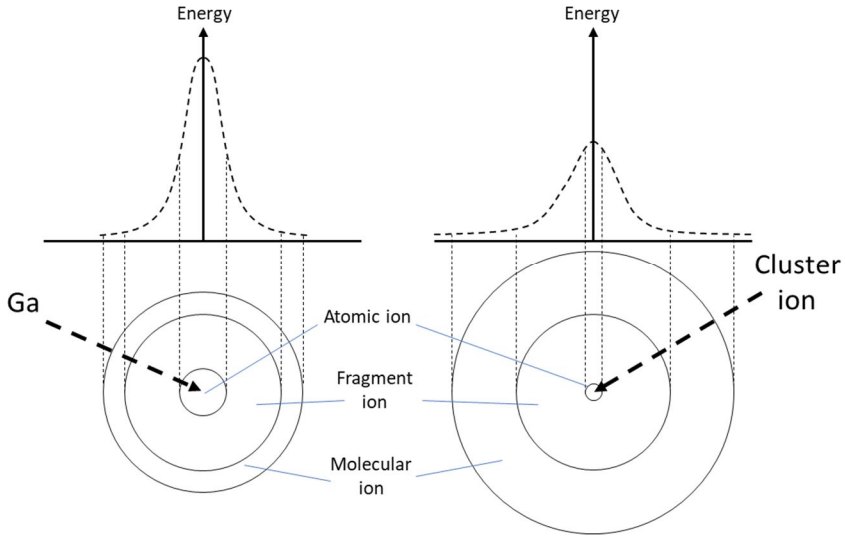


Figure 2.7　Ga^+イオンとクラスターイオンそれぞれの衝突によって固体試料表面に与えられるエネルギー分布の概念図（同心円の各領域はそれぞれ原子イオン、フラグメントイオン、分子イオンが生成される領域を表す）

クラスターイオンには一次イオンとしての用途のほかに、大電流密度によって試料をエッチングする用途がある。その際、試料に与えるダメージの影響が小さいという利点がある。この原因は、C_{60}^{+}と Ga^{+}の Ag(111)面への衝突をシミュレーションした結果（Figure 2.8）[54]からわかるように、Ga^{+}は試料深くに入り込み試料内部の原子を攪拌を多く起こすのに対して C_{60}^{+}ではイオンの入り込みが浅く試料表面にエネルギーが多く分散して試料表面の原子を効果的にスパッタするため、ダメージ（原子の攪拌など）の影響がない面が残ることによると考えられている。

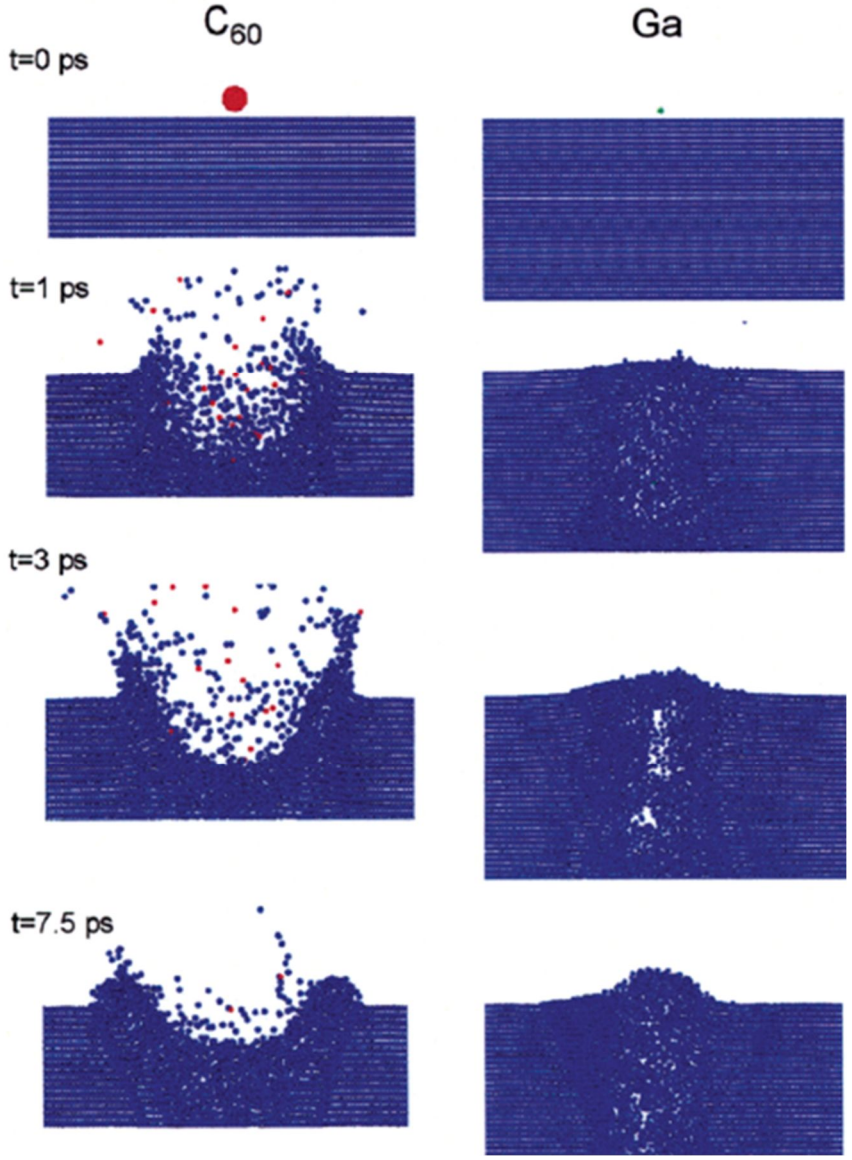


Figure 2.8　クラスターイオン衝突時のシミュレーション画像 [54]

クラスターイオンは金属イオン（Au_3^+, Bi_3^+, Bi_5^+など）を用いたものと、それ以外に大別される。金属イオンは液体金属イオン銃によりビーム径を細く収束することができるため、Ga イオンと同様に、一次イオンとして用いることで微小部分析が可能である。特にBiと特定の元素からなる共晶合金を使用したものは、現在の TOF-SIMS 装置のデファクトスタンダードとして使用されている[55]。一方で、C_{60}^+や Ar_n^+（Ar_{1500}^+など）はビームの収束は金属イオンに劣るが、よりクラスターサイズを大きくすることが可能であることから、より高分子量のイオン化率の向上やダメージ蓄積の低減、高スパッタ率という特性があり、スパッターイオンとしての利用が多く行われている[56-57]。

最近では生体分子などのより高質量側の二次イオン収率の向上のために、クラスターイオンを一次イオンとして使用する研究が盛んに行われている。生体試料の場合は元々の TOF-SIMS の特徴でもあった最表面分析が求められることは少なく、むしろ一定の体積中に含まれる分子をより高感度に分析・イメージングしたいという要求が大きい。そのような要求に対応する装置として、クラスターイオン（Ar_n^+）による連続イオンビームを試料に照射し、連続的に試料表面から発生する二次イオンを質量分析計側でパルス化して TOF 型質量分析計へと導入する Orthogonal-SIMS（直交加速型 SIMS）が、京都大学の松尾らにより開発されている[58, 59]。同様に Ar_n^+の連続イオンビーと正電場を用いた Orbitrap 型質量分析計を組み合わせた装置がイギリス国立物理学研究所の Gilmore らにより開発されている[60]。

2.1.4 TOF-SIMS のデータ構造 [61]

TOF-SIMS では、*m/z* = 1～約 2,000 の質量範囲に 2,000～10,000 のピークが観測される質量スペクトルが、一般に 256×256 個のピクセルの一つ一つに格納されたデータとなる。そのため、TOF-SIMS のデータは 1.3×10^8 以上のデータサイズを持つことになる。さらに前項で紹介した、「クラスターイオンの連続ビーム＋高分解能質量分析計」の組み合わせでは、256×256×500 以上のボクセルに、20,000～2,000,000 チャンネルの質量スペクトルが含まれるデータ（6.5×10^{13} 以上のデータサイズ）となるなど、一層のデータサイズの増大が予想される。

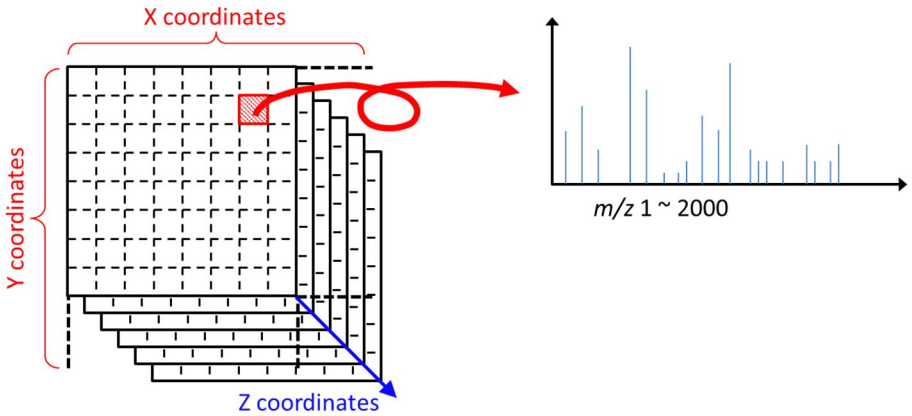


Figure 2.9　TOF-SIMS のデータ構造の模式図

2.2 機械学習（Machine learning）

2.2.1 概要

機械学習は人工知能（Artificial intelligence: AI）の一分野として発展した技術であり、明示的な指示（プログラム）をすることなく、機械がデータから自動的に学習を行うモデルを構築する統計的手法の総称である。機械学習の一分野で、生物の脳内における情報伝達を模したニューラルネットワークを重層的に使用し、データから直接的に表現やタスクを学習する手法は、特に深層学習（ディープラーニング）と呼ばれる[62]。

・AI：　　コンピューター自らが状況認識、決断、行動の一連の動作を行うシステム。

・機械学習：　　データから自動的に学習するモデルを構築する技術。ディープラーニングも包含した技術であるが、特にディープラーニングと区別して呼称される場合は、「特徴を手動で選択して学習させる」ものを機械学習と呼ぶことが多い。代表的な手法として、決定木、ランダムフォレスト、サポートベクターマシン、アンサンブル法などがある。

・深層学習：　　機械学習の一分野であり、脳の神経回路（ニューラルネットワーク）をモデル化したものである。大きな特徴として、アルゴリズムが自動的にどのパラメーターが特徴として有用かを判断して学習を行う点が挙げられる。画像解析分野で一般的に使用される畳み込みニューラルネットワーク（Convolutional neural network: CNN）や、音声などの時系列的データを扱うことが可能な再帰型ニューラルネットワーク（Recurrent neural network: RNN）などがある。

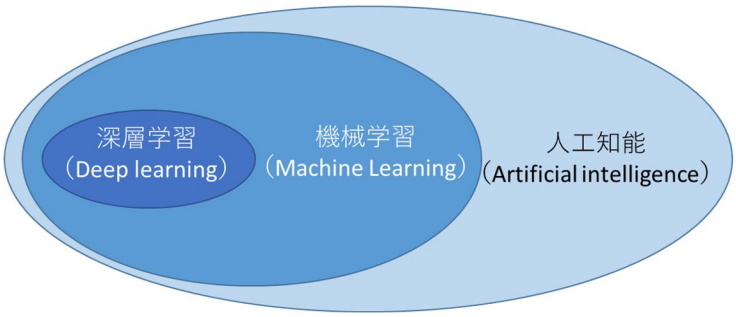


Figure 2.10　機械学習と深層学習の概略　[62]

一方で、機械学習（＋深層学習）の別の区分として、入力データと出力データ（解）が揃っており、入力データから出力データを予測するモデルを作成する「教師あり学習（Supervised learning）」と、入力データのみで学習を行い、データの背景にある隠れたパターン（特徴）を抽出する「教師なし学習（Unsupervised learning）」がある。

2.2.2 ニューラルネットワークの基礎 [63, 64]

生物の大脳皮質に存在するニューロンは、核をはじめとした細胞小器官を内包した細胞体の周囲から樹状突起や軸索と呼ばれる長い突起が生えた構造をしている。軸索の末端にはシナプスと呼ばれる構造があり、これが他のニューロンの樹状突起に接続している。ニューロンは他のニューロンからシナプスを介して信号（電気刺激）を受け、その量が十分な量になると自身からも信号を発し、他のニューロンへと伝えていく。このような刺激の伝達を数十億個のニューロンからなる巨大なネットワークで行うことで、脳内活動が行われている。この生物学的ニューロンをモデルとして複雑な論理演算を行う仕組みが人工ニューラルネットワーク（Artificial neural network: ANN）である。

Figure 2.11 の図はパーセプトロンと言われ、ANN の構造で最も単純なものである。入力に対して重み付けされた値の総和（$z = w_1x_1 + w_2x_2 + \cdots + w_nx_n = W^T \cdot X$）にステップ関数を適用した結果$h(X) = f_{step}(W^T \cdot X)$を出力する（活性化する）。$h(X)$がステップ関数の閾値を超えている場合は 1 が出力され、閾値より小さい場合は 0 を出力する。

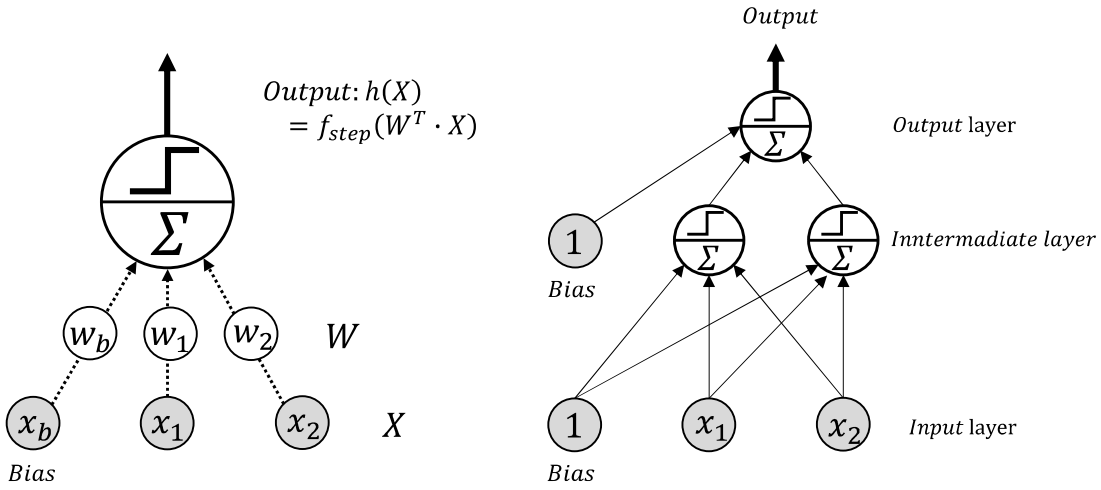


Figure 2.11 （左）パーセプトロンの構成単位である LTU、（右）マルチレイヤーパーセプトロン[64]

※右の図では重み付け（W）の表示を省略している。

この一つのパーセプトロンより成る単純な構成でも、論理演算のうち OR 型、AND 型、NAND 型については対応できるが、XOR 型（排他的 OR 型）について対応できない点が重大な欠点として挙げられていた（Table 2.2、Figure 2.12）。しかしこの問題については、パーセプトロンを二層積み上げる、つまり OR 型と NAND 型のパーセプトロンを AND 型のパーセプトロンで結合することで解決された（Figure 2.12）。このような、一つの入力層（Input）と一層以上の中間層（Intermediate）、一層の出力層から成るネットワークを Multi-layer perceptron: MLP と呼ぶ。

パーセプトロンの学習は、出力値とターゲットとなる値（正解）との誤差を最小化するように重み（W）を更新することで行われる。重み更新を効率的に行うためには後述する勾配降下法が使用される。勾配降下法は微分を用いた最適化アルゴリズムであるため、微分不可なステップ関数は使

用ができない。そのため、シグモイド関数（ロジスティック関数）などの様々な活性化関数の適用が検討されてきている。

Table 2.2 代表的な論理演算の種類

論理演算	論理式	概要
論理和（OR）	$A + B$	入力値のいずれかが 1 のとき 1 を出力する。 それ以外のときは 0 を出力する。
論理積（AND）	$A \cdot B$	入力値がすべて 1 のとき 1 を出力する。 それ以外のときは 0 を出力する。
否定論理積（NAND）	$\overline{A \cdot B}$	入力値がすべて 1 のとき 0 を出力する。 それ以外は 1 を出力する。
排他的論理和（XOR）	$\bar{A} \cdot B + A \cdot \bar{B}$	二つの入力値が異なるときに 1 を出力する。 それ以外は 0 を出力する。

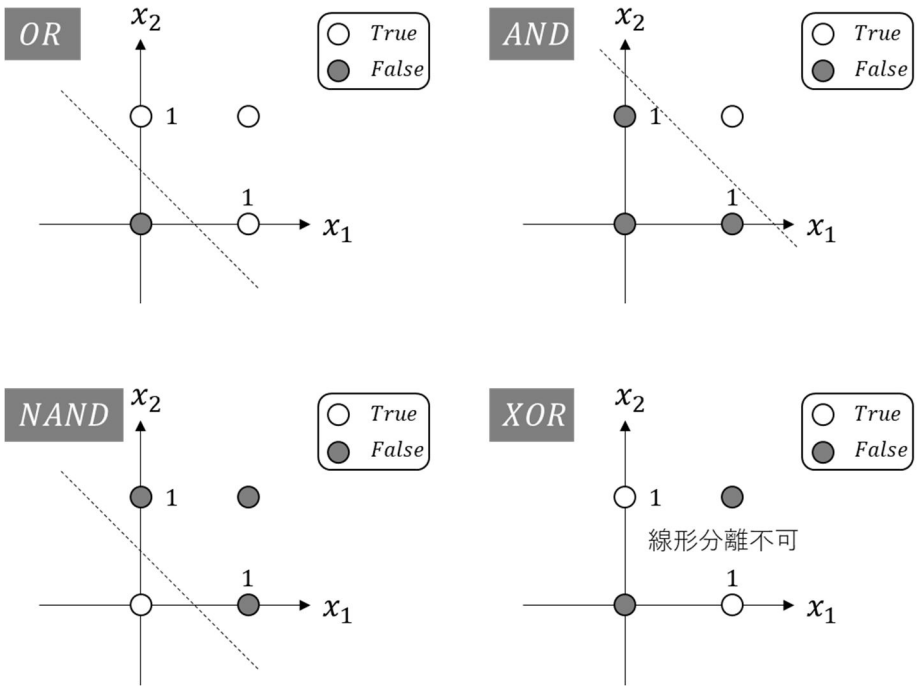


Figure 2.12 代表的な論理演算の模式図 [64]

2.2.3 活性化関数（Activation function）[64, 65]

単純なパーセプトロンや MLP ではステップ関数が用いられたが、実際の ANN では目的に応じて様々な活性化関数が用いられている。ここでは、基本的なものを示した。Figure 2.13 に各活性化関数を図示した。

（1）恒等写像（Identify）

線形関数（Linear function）とも呼ばれ、入力値と同じ値を出力する関数である。線形回帰に使用される。

$$\phi(z) = z$$

（2）ステップ関数

入力データの和（z）が指定された閾値（θ）よりも大きい場合は 1 を出力し、それ以外の場合は 0 を出力する。（パーセプトロンに使用）

$$\phi(z) = \begin{cases} 0, & z < 0 \\ 0.5, & z = 0 \\ 1, & z > 0 \end{cases}$$

（3）シグモイド

恒等写像やステップ関数と異なり、S 字型の非線形関数である。ロジスティック回帰や多層ニューラルネットワークに使用される。

$$\phi(z) = \frac{1}{1 + e^{-z}}$$

（4）双曲線正接（Hyperbolic tangent: tanh）

シグモイド関数と類似のS字型かつ連続で微分可能な関数であるが、出力範囲が-1 から+1 であり、原点を通過する。シグモイドに比べて微分係数の値（導関数の出力値）の最大値が大きくなるため、学習時間が低減される効果がある。シグモイドと同様に多層ニューラルネットワークに使用される。

$$\phi(z) = \frac{e^{z} - e^{-z}}{e^{z} + e^{-z}}$$

（5）ReLU（Rectified liner unit）

連続だが z＝0 で可微分ではない。出力に負値を含まない点と、最大値がないことが特徴である。特に後者の特徴は後述する勾配法アルゴリズムでしばしば問題となる「勾配消失」の問題を緩和できることから、近年、深層学習の活性化関数として ReLU（またはその亜種：Leaky-ReLU, ELU, SELU など）が多く用いられている。

$$\phi(z) = \begin{cases} 0, & z \leq 0 \\ z, & z > 0 \end{cases}$$

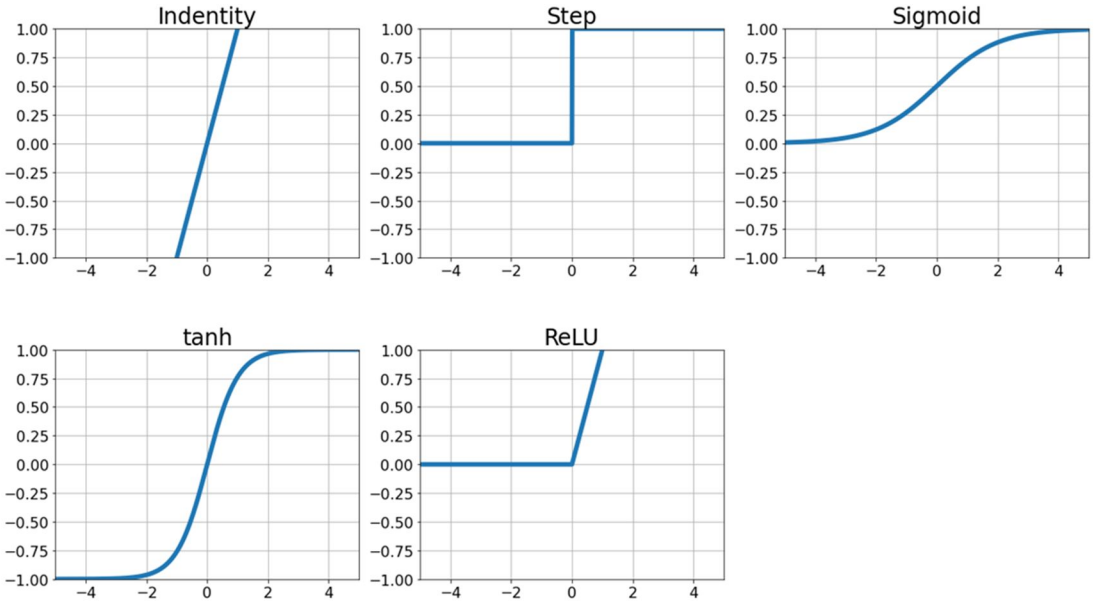


Figure 2.13　機械学習に使用される代表的な活性化関数（伝達関数）

2.2.4 損失関数（Loss function）[64, 65]

ニューラルネットワークの学習の基本は、モデルの予測（出力）値と正解の誤差を最小化するように各ニューロン間の重みを更新することである。この際、誤差を定義する関数を損失関数と呼び様々な関数が用いられている。ここではニューラルネットワークで一般的な損失関数を挙げた。

（1）平均二乗誤差（Mean squared error: MSE）

モデルの予測値（$\hat{y_i}$）と正解（y_i）の差分の二乗を平均したもの。最も一般的に損失関数として用いられる。

$$MSE = \frac{1}{n}\sum_{i=0}^{n}(\hat{y_i} - y_i)^2$$

（2）平均絶対誤差（Mean absolute error: MAE）

モデルの予測値（$\hat{y_i}$）と正解（y_i）の差分の絶対値を平均したもの。主に回帰問題における出力層の損失関数として用いられる。

$$MAE = \frac{1}{n}\sum_{i=0}^{n}|\hat{y_i} - y_i|$$

（3）クロスエントロピー（Cross entropy）

モデルの予測値の確率分布を$q(x)$、正解の確率分布を$p(x)$としたとき、クロスエントロピーは次式で与えられる。主に分類問題における損失関数として使用される。

$$H(p, q) = -p(x)\log q(x)$$

(4) カルバックーライブラーダイバージェンス(Kullback-Leibler divergence)

モデルの予測値の確率分布を$Q(x)$、正解の確率分布を$P(x)$としたとき、KL-ダイバージェンスは次式で与えられる。

$$D_{KL}(P \parallel Q) = \sum_{x} P(x)\log P(x) - P(x)\log Q(i)$$

2.2.5 最適化関数(Optimizer)

損失関数を最小化するアルゴリズムとして、勾配降下法およびその亜種が一般的に用いられている。ここでは代表的な最適化関数について概要を記した。数学的な詳細については原著論文および専門書を参照されたい。

2.2.5.1 勾配降下法(Gradient Descent) [63, 64]

勾配降下法は、さまざまな問題の最適解を求めることができる汎用的な最適化アルゴリズムであり、これまで様々な種類のものが考案されてきた。勾配降下法の一般的なアイデアは、損失関数を最小化するためにパラメーター(重み)を反復的に更新することである。つまり、パラメータベクトルθに関する損失関数の局所的な勾配を測定し、勾配が降下するようにθを更新していき、勾配がゼロになる(損失関数が最小値になる)位置を探索する。この際に重要なパラメーターは、損失関数の曲線上を移動するステップの幅である学習率(η)である。ηが小さいと損失関数が収束(最小値)となるまでの時間が増加する。一方でηが大きいと、最小値を飛び越えてしまい、容易に収束しない事態に陥る。Figure 2.14 (A), (B)には単純な一次元のパラメーター(θ)に対して、ηを変えた際の損失関数の変化の様子を示した。実際には損失関数の形状は単純な凹型ではなく、複雑なうねりや局所的に深い穴やプラトーな領域を多数持つ。そのため、ηを小さくしすぎると局所的な最小値(Local Minimum)につかまってしまうため、全体の最小値(Global minimum)に到達できない(Figure 2.14(C))。そのため、適切な学習率の設定が重要となる。

勾配降下法の一般的な手順としては以下の①～④がある。

① ニューラルネットワークに入力データをすべて入れて予測値を出力する、
② 正解と予測値の誤差を定義された損失関数($J(\theta)$)を用いて算出する。
③ 損失関数を各パラメーター(θ)で微分(偏微分)
④ 微分して出た値(∇_θ)と学習率(η)を用いてパラメーターを更新する
⑤ 必要な分(決められた分)だけ①～④を繰り返す

この一連の手順は次式を繰り返し行うことに相当する。

$$\theta \leftarrow \theta - \eta \nabla_\theta J(\theta)$$

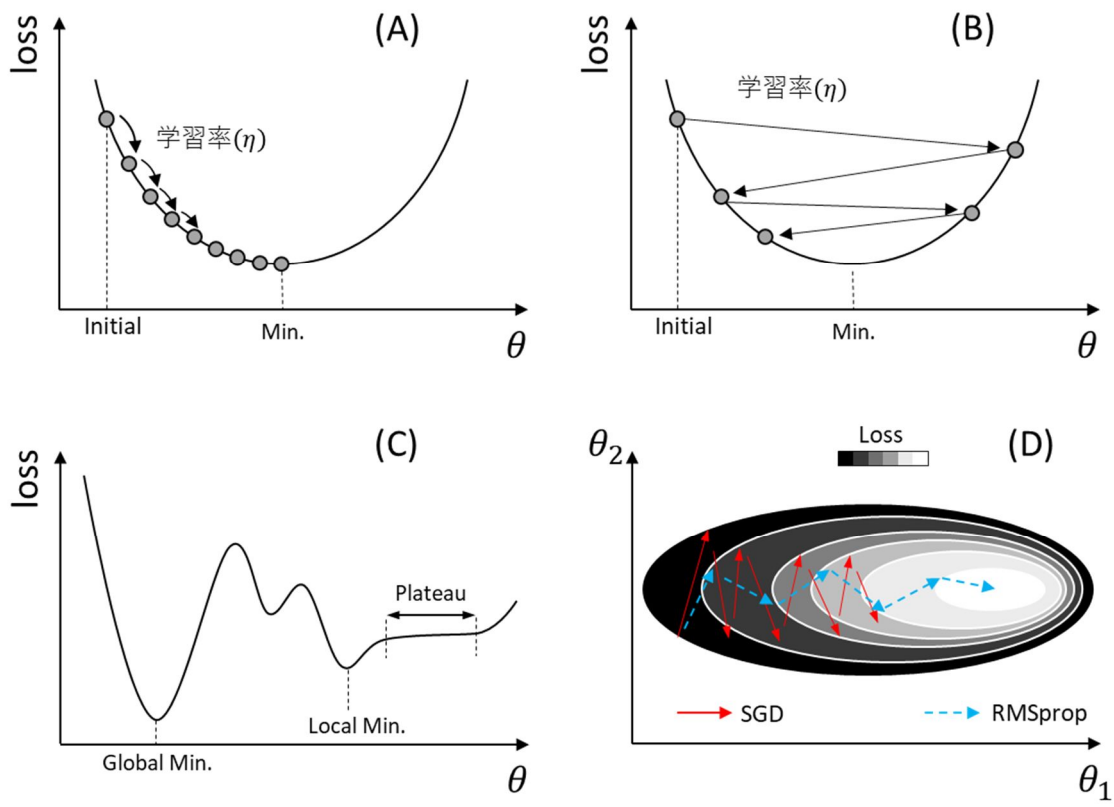


Figure 2.14　勾配降下法の原理図　(A)単純な凹型の損失関数においてηを小さくした場合、(B)ηを大きくした場合、(C)複雑な損失関数における局所解（Local Min.）と全体の最適解（Global Min.）、(D)改善された最適化関数における振る舞いの違い

勾配降下法は一回のパラメーター更新に用いるデータ量によって、次の三つに分けられる。

（1）バッチ勾配降下法

一回のパラメーター更新を行う際に、データ全てを用いて勾配計算を行う。すべてのパラメーター更新を一度で行うことが可能だが計算量が多くなる。しかし、この問題は並列計算を行うことで計算時間を短縮することが可能である。一方で、局所解に捕まった際にパラメーター更新時の変化量が小さく、局所解から抜け出すことが困難になるという欠点がある。

（2） 確率的勾配降下法（Stochastic gradient descent: SGD）

一回のパラメーター更新の際に、ランダムに選ばれた一つのデータのみを使用する。このランダム性によって、直前のパラメーター更新で局所解に捕まったとしても、次にランダムに選ばれたデータでは損失が非常に大きなものとなり、パラメーターが大きく更新される可能性があり、局所解から抜け出すことができる。欠点として、一つのデータでパラメーター更新が終わらないと次のデータに移れないため並列計算ができず、計算処理時間が長くなる。

（3） ミニバッチ勾配降下法（Mini-batch gradient descent）

バッチ勾配降下法と SGD の中間的な手法である。まず、データ全体を一定のデータ数を持つデータ群に分ける。その中からランダムに一つを選び、各データを用いた勾配計算を並列で行う。それにより、ランダム抽出から得られる頑健性と計算速度の両立を図った方法である。本研究ではこのミニバッチ勾配降下法、具体的にはミニバッチ勾配降下法をベースとした後述する「Adam」を採用した。

2.2.5.2 Deep neural network 用に開発された高速な最適化関数

大規模データの訓練や、深層ニューラルネットワークの訓練は、非常に多くの計算リソースを必要とする。更に、通常の勾配降下法（SGD など）では Figure 2.14 (D)に示したように、損失関数の形状によっては、勾配が大きな方向に振動してしまい、最小値に到達するのに多くの時間を要する場合がある。そこで、より高速に最小値に収束するように勾配降下法に対して様々な工夫を加えた最適化関数が開発されている。

（1）Momentum

単純な勾配降下法では設定した学習率（η）にしたがって、一定のステップで関数の勾配を移動していく。そのため、局所的に勾配が緩やかな領域では収束速度が低下する。Momentum では以前の勾配の履歴を慣性ベクトル(m)として利用して、それを現在の勾配の向きに足し合わせることで振動を抑制し高速な収束を可能とする。これは勾配の移動平均を用いてパラメーター（重み）を更新していることに相当する。

（2）RMSProp

RMSProp は Momentum と同様に勾配降下法の振動の抑制を目的としているが、こちらは学習率（η）を調整する。つまり勾配の急激な位置においてηを下げることによって振動を抑制する。

（3）Adam

上記の Momentum と RMSProp の組み合わせであり、ディープラーニングにおいて現在、最も使用されている。本研究でも、Adam を最適化関数に採用した。

2.2.6 自己符号化器(Autoencoder)

オートエンコーダーは ANN を用いた特徴抽出(次元削減)アルゴリズムの一つであり、エンコーダー(Encoder)とデコーダー(Decoder)という 2 つの連結されたネットワークより構成される(Figure 2.15)。2006 年に Hinton らによって開発された[32]。エンコーダーはd次元の入力データ(X)をp次元のデータ($\hat{X}$)に変換する。つまりエンコーダーは$\hat{X} = f(X)$をモデル化する方法を学習する。一方でデコーダーはp次元データを基のd次元データ($\bar{X}$)に復元する方法を学習する。この際、$d > p$として中間層を"ボトルネック"とすることで、$\hat{X}$はXの次元圧縮された特徴であるとみなせる[65]。

実際のオートエンコーダーの学習では、入力データ(X)と出力(再構成)データ($\bar{X}$)の差分を損失関数とし、勾配降下法アルゴリズムを用いて重み(W)の更新を行う。この際、エンコードされた特徴($\hat{X}$)とデコードされた出力($\bar{X}$)は、入力データ($\bar{X}$)と活性化関数f, g、重み(W)、バイアス(B)を用いてそれぞれ次式で表される[66]。

$$\hat{X} = f(W_{enc}X + B_{enc})$$

$$\bar{X} = g(W_{dec}\hat{X} + B_{dec}) = g(W_{dec}(f(W_{enc}X + B_{enc})) + B_{dec})$$

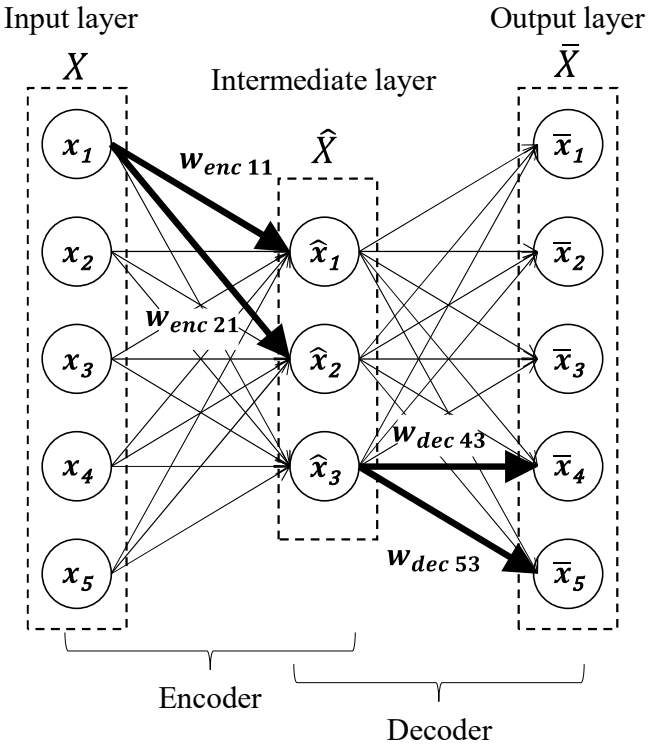


Figure 2.15 オートエンコーダーの概略図 [66]

オートエンコーダーは後述する主成分分析（PCA）や多変量スペクトル分解（MCR）と同様の特徴抽出器としての利用法のほか、ノイズ除去（Noise reduction）や異常検知（Anomaly detection）に用いられている。更に近年注目を集めている学習データから新たなデータを生成する「生成モデル」である敵対的生成ネットワーク（Generative adversarial network: GAN）[67]にも応用されている。

2.3 多変量解析（Multivariate analysis: MVA）

2.3.1 概要

高分子材料や生体高分子について TOF-SIMS 分析を行う場合、分子イオンが直接観察されることは分子量の大きさからほとんどなく、通常は低質量側の複数のフラグメントイオンを用いて物質の同定を行う。また、TOF-SIMS は最表面分析であり、意図しないコンタミネーションの影響がデータに含まれることが当然であることから、一般にそのデータは複雑な混合物データとなる。そのため、データの解釈を容易にするために、多変量解析手法の適用検討が多くなされてきた[28-31]。

適用される解析手法としては、主に多変数データを人が理解しやすい少ない変数（低次元）のデータに変換する次元削減（特徴抽出）法が用いられている。その代表的な手法として主成分分析（PCA）と多変量スペクトル分解（MCR）の概要を以下に記す。

2.3.2 主成分分析（Principal component analysis: PCA） [68]

PCA は 1901 年に Karl Pearson によって開発された手法であり、多変数（$x_1, x_2, x_3, x_4 \cdots\cdots x_p$）のデータを元データより少ない変数（$PC_1, PC_2, PC_3 \cdots\cdot,\ PC_q$）に、情報量を最大限に維持しながら縮約する手法である。理解しやすいように二次元の場合の概念図を Figure 2.16 に示す。はじめにデータの分散が最大になる方向に第一主成分軸をとり、次に第一主成分軸と直交する（無相関）方向に対して、分散を最大化するように第二主成分軸をとる。ここで Score はデータの各主成分軸上での得点（主成分軸への写像）であり、Loading は各主成分軸の向きを表す基底ベクトルである。Loading（e1, e2）について e1>e2 の場合、その主成分軸は$x1$ の影響が強いことを意味する。

主成分分析の式は次式で示され、元データの行列（X）は主成分得点（U）と負荷量（重み）（V^T）という 2 つの行列に分解される。添え字の T は行列の転置を意味し、Rは残差行列である。

$$X = UV^T + R$$

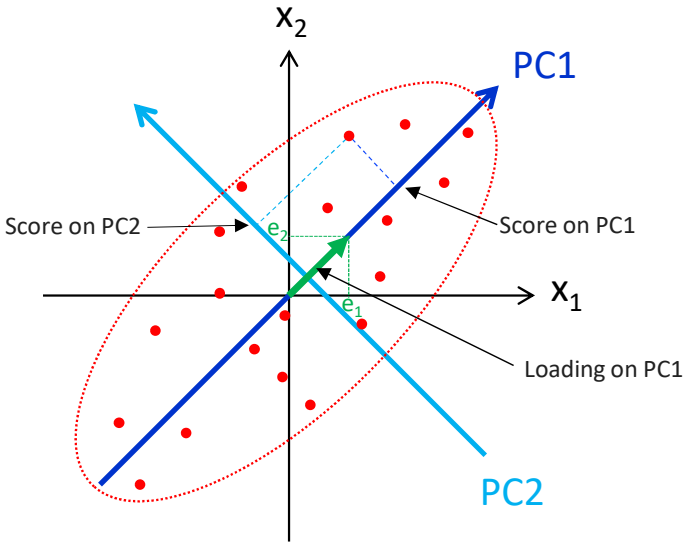


Figure 2.16　主成分分析（PCA）の原理図

主成分分析の計算は、分散共分散行列の固有値問題として解くことが可能であり、固有値が Score、固有ベクトルが Loading である。固有値の大きな順に第一、第二、・・・第 N 主成分となるが、どこまでを有意なものとして取り上げるかは、一概には判断できない。目安としては、寄与率（固有値を全主成分の固有値の総和で割った値）の累積値（累積寄与率）が 80 %以上とする場合や、寄与率が急激に変化する主成分までとする場合など、が提案されている。

2.3.3　多変量スペクトル分解（Multivariate curve resolution: MCR）[69]

MCR はデータ行列 D から純成分のスペクトル（S）とその相対濃度（分布）（C）を分離する手法であり、次式で表される。E はモデルで説明しきれない残差行列である。

$$D = CS^T + E$$

MCR-ALS（Alternating least square）では、交互最小二乗法により残差（E）が決められた値を下回るまで上式を繰り返し解き、それにより実験データ行列Dに最適な濃度行列Cと純粋なスペクトルであるS^Tを算出する。この際、「負の値をとらない（非負性）」や「成分数の指定」といった制約（拘束条件）を設定する必要がある。前述の PCA と比べると、MCR で得られる結果は非負の制約により実際の測定データ（分光スペクトルや質量スペクトル）の形に近く、理解が容易である。そのため、分析化学の観点から意味のある結果を得ることができる[30, 31]。しかしながら成分数の設定が適切でないと、純成分のスペクトルが抽出されない場合や、データを過剰に分離してしまうなどの結果へと繋がるため、注意が必要である。

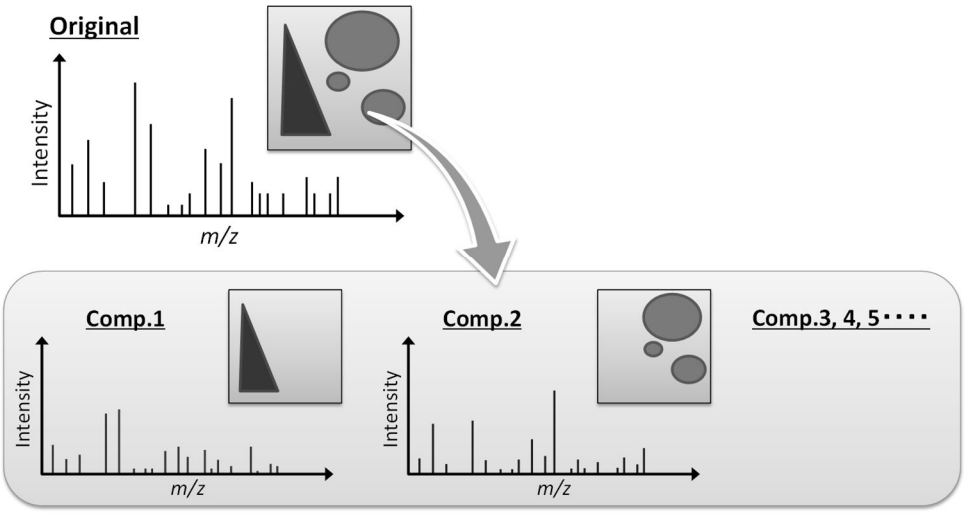


Figure 2.17　多変量スペクトル分解（MCR）の原理図

2.4 データ前処理

多変量解析を行う際に適切なデータ前処理方法を行うことは、複雑なデータからより良いデータを抽出するうえで重要である。ここでは TOF-SIMS データの解析や多変量解析に用いられる代表的なデータ前処理方法について記述する。

（1） Mean-centering [70]

すべての変数の値から各変数の平均値を差し引く処理である。Mean-centering を実行すると、サンプル（データ）間の相対的な位置関係を変えずに、データセットの中心（重心）に原点を移動させることができる。原点移動することによって次元が一つ下がる効果があることから、PCA の前処理として一般的によく使用される前処理方法である。

（2） Auto-scaling [70]

解析データによっては、変量（変数）によって単位やノイズレベルが異なる場合がある。このような場合、測定値の大きさは情報の重要性と必ずしも一致しない。そのため、データ行列の各変数の情報量が互いに同等となるようにするために、変数のスケーリングを行う場合がある。

$$X_{scale} = X \cdot S$$

ここで、Xは入力データ、Sはスケーリング係数の対角線行列、X_{scale}はスケーリング後のデータの行列である。Auto-scaling は Mean-centering を実施した後に、各変数をその列の標準偏差で割る方法であり、Sの対角線行列は各変数の標準偏差の逆数となる。したがって、$X \cdot S$の各列の平均値は 0 で、標準偏差は 1 となる。Auto-scaling は各変数の分散の主な原因がノイズではなくシグナルである場合、変数の尺度や単位の違いを補正するための有効な方法だが、ある変数がノイズの影

響を大きく受けていたり（すなわち、*S/N* 比が低い）、標準偏差がゼロに近い場合、Auto-scaling はノイズの影響を強調することになる。このような状況では、このような変数を除外するなどの対応が必要となる。

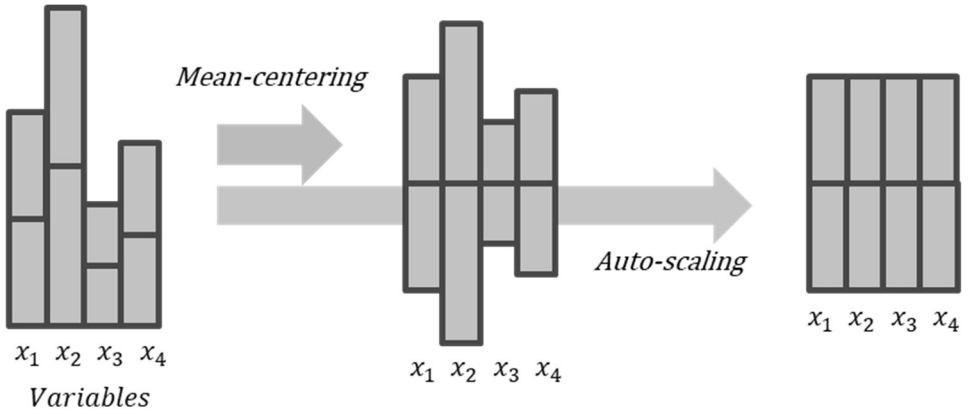


Figure 2.18　Mean-centering と Auto-scaling の概念図

（3） Poisson scaling

光子・電子・イオン計数器については、不感時間による数え落としを考慮しなくてよい場合、計数値の分布は Poisson 分布に従うとされている。電気的ノイズなどによる統計的な不確実さの程度は一様ではなく、計数値の強度が大きいほど相対的に増加する。変量（質量ピーク）と試料（画像データではピクセル）から成る質量イメージングデータにおいて、試料由来の不確実性を無視し、変量由来の不確実さのみを考慮すると、Poisson 分布ではピーク強度の誤差は分散の平方根に等しいとみなされるため、Poisson scaling で処理されたデータは次式で表される[27]。

$$\bar{X} = \frac{X}{\sqrt{\bar{V}}}$$

ここで Poisson scaling では次式の近似がされる。

$$X = V$$

ここで X は前処理後のデータ、X は前処理前のデータ、V はデータの分散である。Poisson scaling で処理されたデータは、電気的ノイズの影響が軽減され本来のスペクトルに近いデータが得られる効果が期待されることから、質量イメージングデータへ適用することで良好な結果が得られることが実証されている[27]。

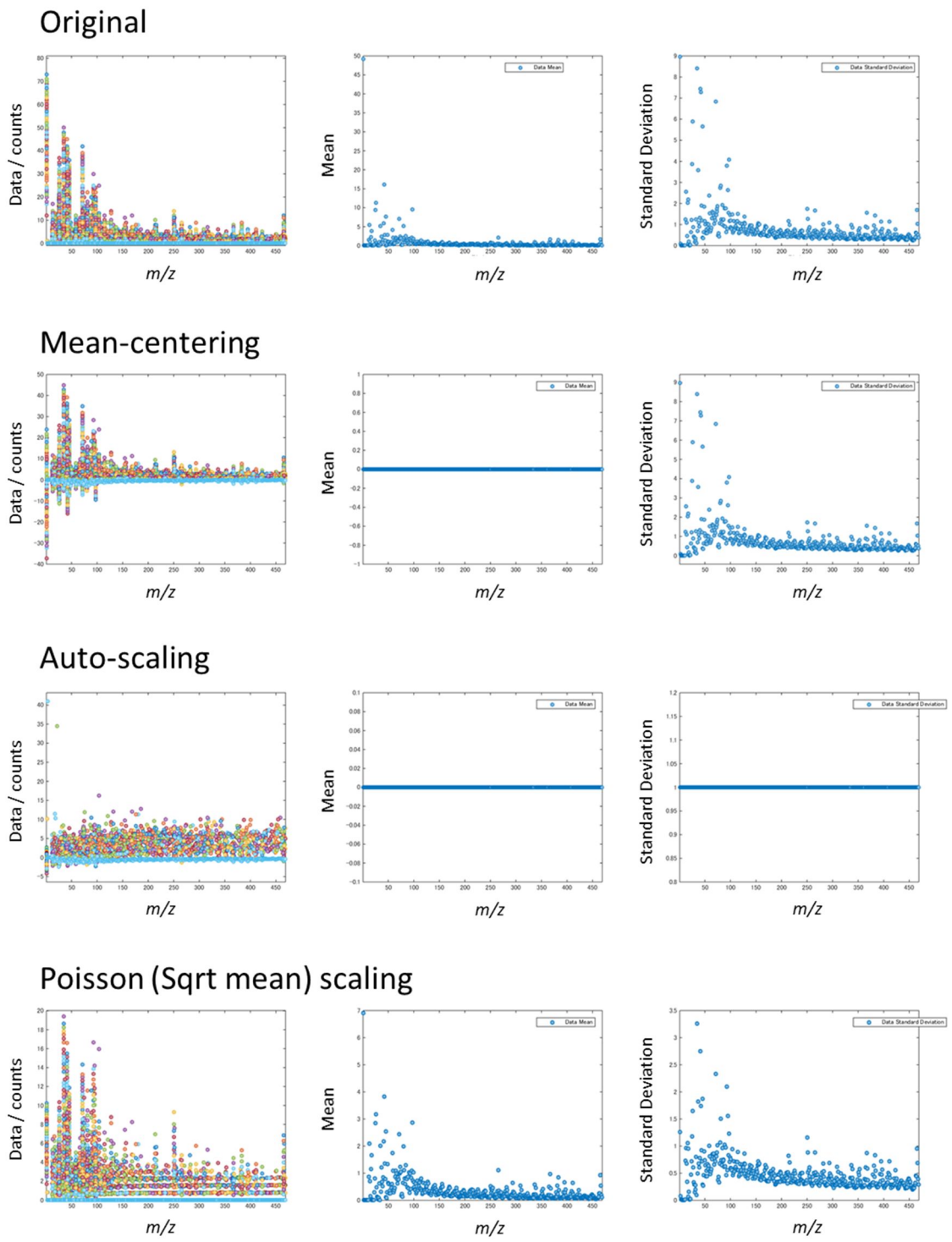


Figure 2.19　Mean-centering と Auto-scaling、Poisson scaling の効果

2.5 生体試料（Biological samples）

2.5.1 生体を構成する主要成分 [71]

生体を構成する物質は大きくは、炭水化物（糖質）、タンパク質、核酸および脂質の 4 つに大別される。炭水化物はエネルギー源や構造材料としての役割のほか、タンパク質や脂質に結合して極性を調整する役割や、他の分子に対する認識部位となる役割を併せ持つ。

タンパク質はアミノ酸を構成単位としたポリマー（生体高分子）であり、生体の構造材料としての役割のほか、生体内で起こる様々な反応における触媒としての役割（酵素）、生理活性に関わる受容体としての役割などを担っている。

核酸は、塩基と糖、リン酸で構成されるヌクレオチドが連なった生体高分子であり、増殖や遺伝に関与する。糖がリボースである RNA とデオキシリボースである DNA の 2 種類が存在する。DNA の塩基配列情報は主な遺伝情報を含んでおり、細胞核内にて情報の保存・蓄積を担っている。一方で RNA は DNA の塩基配列を基に転写により合成され、細胞中でのタンパク質合成に必要な情報の伝達を担っている。

脂質はエネルギーの貯蔵物質としての役割と、細胞と細胞を隔てる細胞膜の構成単位という役割を担っている。

生体を構成する上記の 4 物質は、主に 6 種の元素（水素、炭素、酸素、窒素、リン、硫黄）の組み合わせにより構成されている。そのため、各物質を分子レベルで分離・識別してその分布を評価するためには、各分子に特徴的な構造（官能基）や大きさ（分子量）を測定できる手法が必要となる。

Table 2.3　主な生体分子の機能と構成元素

	主要成分	主な機能	主な構成元素
①	炭水化物（糖）	エネルギー源（アデノシン三リン酸：ATP の合成原料）	C, H, O
②	タンパク質	生体構造の形成、生体反応の触媒、輸送や貯蔵、免疫など	C, H, O, N, S
③	核酸	情報の保存、蓄積、伝達	C, H, O, P
④	脂質	エネルギーの貯蔵、細胞膜の構成成分	C, H, O, P

2.5.2 生体試料の組成イメージング手法

生体試料のイメージング手法として最も代表的な手法は、光学的な顕微鏡観察手法である。古典的な明視野、暗視野観察から位相像、微分干渉像による観察、蛍光プローブを用いた蛍光顕微鏡法、レーザー光源を用いた共焦点顕微鏡法など、様々な観察機器が考案されてきた。近年では光の回折限界を超えた超解像顕微鏡として、誘導放出抑制顕微鏡（Stimulated Emission Depletion: STED）や、光活性化局在性顕微鏡（Photoactivated localization microscopy: PALM）といった手法も開発されてきている[72]。そのほか、蛍光プローブを導入した抗体と試料表面に存在する特定タンパク質（抗原）との特異的結合を利用した免疫染色法などの観察技術も一般によく用

いられている。これらの光学顕微鏡観察手法は、生体の様々な成分の分布や機能を可視化することができることから、生物学の発展に多大な貢献をしてきた。

一方で近年、TOF-SIMS をはじめとした質量イメージングが注目を集めている。質量イメージング法の光学顕微鏡法に対する利点としては、測定対象成分（原子・分子）を直接（蛍光プローブなどの標識なしで）観察できること、分子量情報を検出することから薬剤の原型と代謝物の識別が可能など、成分の選択性に優れる点が挙げられる。イオン化法としては、SIMS のほかにマトリクス支援レーザー脱離イオン化法（Matrix-assisted laser desorption ionization: MALDI）や脱離エレクトロスプレーイオン化法（Desorption electrospray ionization: DESI）が代表的である。MALDI はレーザーを吸収して熱エネルギーとして放出するマトリックスを試料に塗布することで、測定対象をソフトイオン化することができる。ペプチドやタンパク質など、SIMS ではイオン化が困難な高分子量成分のイオン化が可能である[73]。一方、DESI はキャピラリーから発生させた一次帯電液滴を試料表面に照射することで、試料に含まれる成分を溶媒抽出するとともにプロトン移動によって二次帯電液滴としてイオン化させる。大気圧下でマトリクス塗布などの特別な処理をすることなくイオン化できることや、フラグメンテーションをほとんど生じないことから、遊離のアミノ酸や脂肪酸の評価が可能という特徴がある。[74]

MALDI や DESI は表面分析ではないため、生物組織切片のイメージング手法という観点から見た場合、一度の測定でイオン化する試料の絶対量が SIMS に比べて多いことがメリットである。組み合わせる質量分析計の自由度も高く、TOF 型以外にもフーリエ変換型（FT-ICR 型）や静電場を用いた Orbitrap 型、四重極型と TOF 型の組み合わせ（Q-TOF 型）などが使用されている。さらにそれらの優れた質量分析計の前段として衝突誘起乖離（CID）セルなどを備えることで MS/MS 分析を行えるという優れた特性を有している。一方で SIMS では、MALDI と DESI に対して勝っている空間分解能にさらに特化した装置（NanoSIMS）も開発され、単一細胞レベルのイメージングに活用されている[75, 76]。また、SIMS においても高分子量の分子をイオン化可能な新規イオンビームの開発や、表面分析に拘らない装置の開発が継続的に行われている。そのため、MALDI や DESI、SIMS の使い分けが一層曖昧なものになってくるかもしれない。

第三章
ヒト毛髪のデプスプロファイルデータに対するオートエンコーダーの適用検討

3.1 はじめに

本章では、ニューラルネットワークを利用した特徴抽出法である自己符号化器（オートエンコーダー）の、TOF-SIMS データ解析における有用性を検討することを目的とした。序章にて述べた通り、分析データからオートエンコーダーを用いて特徴抽出を行う試みは、FT-IR や MALDI-TOF-MS イメージングにて報告されている[38]。しかしながら、それらは抽出された特徴がどのような分布をしているのかを評価することを主眼としており、抽出された個別の特徴についてどの様な情報が含まれているのかといった、特徴の解釈については十分に行われていない。そこで本章では、オートエンコーダーによって TOF-SIMS データから特徴抽出（組成分布の抽出）を行い、各特徴に含まれる組成情報をネットワークの重みパラメーターを詳細に解析することで明らかにすることを目的とした。さらに、従来から TOF-SIMS のデータ解析に使用されてきた多変量解析（PCA）との比較に着目し、オートエンコーダーの優位性の有無についても評価を行った。

特徴抽出の性能を評価するうえで、解析するデータはある程度組成が複雑であることが好ましく、その点で様々な生体分子より構成される生体組織は適当と考えられる。これまで TOF-SIMS を用いて、その組成分布が調べられた生体組織としてはマウス脳[55, 61]・ヒト皮膚[77-81]・ヒト毛髪[66, 82]が代表的である。その理由としては、それぞれ次に示す理由が考えられる。

① 脳組織については TOF-SIMS で分析対象となる比較的低分子量の脂質成分に富み、また、生体組織イメージング手法として先行する MALDI-TOF-MS との比較に適するため。
② ヒト皮膚やヒト毛髪については比較的入手がし易いヒト由来組織であると共に、薬剤や有効成分を直接浸透させる組織であることから、分析対象としての需要が高いため。

本検討では②の理由から、ヒト毛髪内部に浸透したヘアケア剤成分の分布をオートエンコーダーで特徴として抽出できるかに主眼を置いて検討を実施した。そのためのモデルデータとして、Ar クラスターイオンビーム（Ar-GCIB）によるイオンエッチングを併用した TOF-SIMS 測定による毛髪表面のデプスプロファイルデータを解析データとして用いた。

ここで、毛髪試料の構成について概要を記す。哺乳類の毛髪は、扁平な細胞が重なり合った外側の層（キューティクル）が、細胞が集まった内側のコア（コルテックス）を取り囲んでいる（Figure 3.1）。キューティクルは、毛髪内部への有害物質等の侵入を防ぐバリアとしての役割のほか、毛髪繊維の様々な物理特性（強度・しなやかさなど）を維持するのに役立つ水分の出入りの調節の機能も担っている。このキューティクルを構成する細胞の層数や厚さは種によって大きく異なり、羊毛の場合は通常一層だが人毛の場合は 6～8 層程度となる。キューティクルの細胞は更に複数の部位に分かれていることが、透過型電子顕微鏡（TEM）による観察（Figure 3.2）や元素分析（TEM-EDX, EELS）の結果（Figure 3.3）から明らかになっている。これらの層は、表面側より a 層（厚み：100 nm 以下）、エクソキューティクル：Exocuticle（厚み：約 100～300 nm）、エンドキューティクル：Endocuticle（厚み：約 50～300 nm）と呼ばれる。エンドキューティクルは深部側で細胞膜複合体

（CMC）と呼ばれる層と接している。キューティクルを含む毛髪組織は主にケラチンと呼ばれるタンパク質より構成されているが、a 層、エクソキューティクル、エンドキューティクルの 3 層ではケラチンを構成するアミノ酸組成が異なり、特に硫黄含有アミノ酸であるシステイン（またはシステイン 2 分子がジスルフィド結合で繋がったシスチン）の比率に差があることがわかっている[83]。

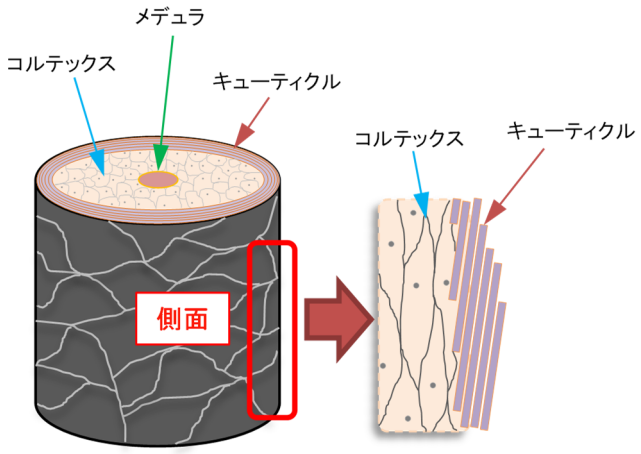


Figure 3.1　毛髪の模式図

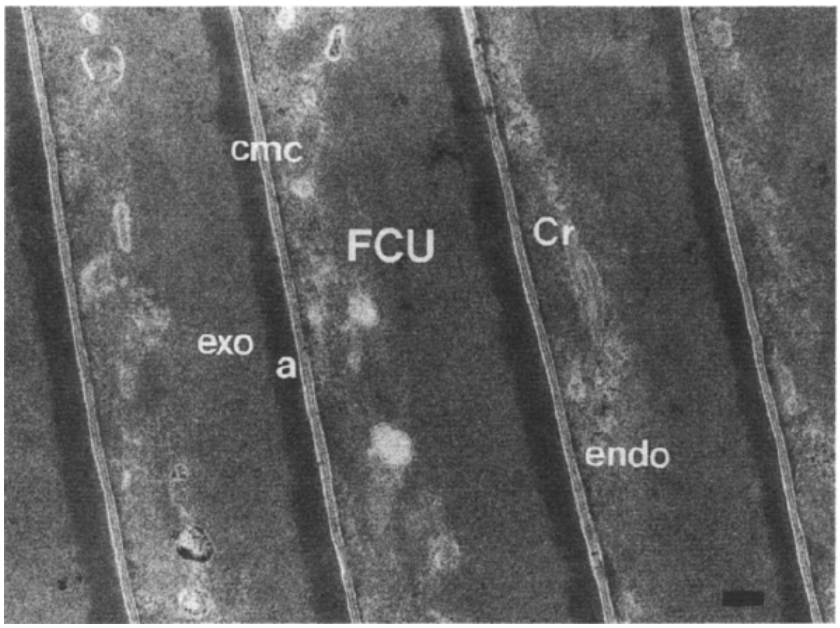


Figure 3.2　ヒト毛髪のキューティクル断面の透過型電子顕微鏡像（四酸化オスミウム固定、酢酸ウラニルおよびクエン酸鉛による後染色後）[83]。a は a-layer、exo は Exocuticle、endo は Endocuticle、cmc は細胞膜複合体を表す。スケールバーは 100 nm。

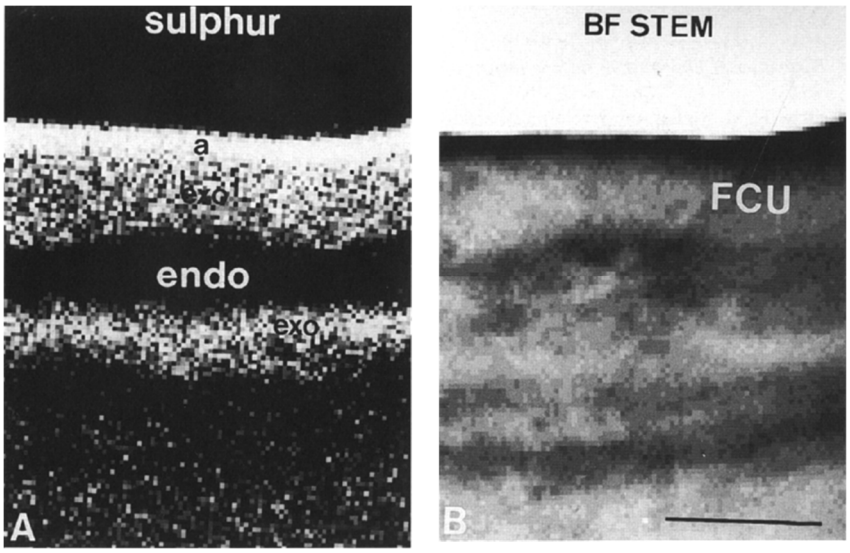


Figure 3.3　キューティクル断面の硫黄のマッピング（**TEM-EDX**）[83]。**a-layer** にて硫黄濃度が最も高く、**Endocuticle** にて硫黄濃度が非常に低い様子を示している。スケールバーは **100 nm**

3.2 実験方法

3.2.1 分析試料調整

成人男性から提供された毛髪をアセトン（> 99.5 %, Sasaki Chemical, Kyoto, Japan）、エタノール（> 99.5 %, Fujifilm Wako, Osaka, Japan）、蒸留水（LC/MS Grade, Fujifilm Wako, Osaka, Japan）の順に、それぞれ 5 分間浸漬し、毛髪表面に付着した油分や汚染物質を除去した。その後、市販のヘアケア剤に 10 分間浸漬し、含有成分を毛髪内部に浸透させた。さらに蒸留水中で緩やかに撹拌を行い、毛髪表面に過剰に付着したヘアケア剤を除去し、分析試料とした。なお、参照試料としてヘアケア剤への浸漬を行わない試料も用意した。

3.2.2 TOF-SIMS 測定条件

キューティクル層のデプスプロファイルデータの測定には、TOF.SIMS 5（IONTOF GmbH, Münster, Germany）を用いた。一次イオンとしてパルス化された Bi_3^{2+}ビームを用い、試料表面のエッチング用として Ar クラスターイオン（Ar_{1500}^{+}）を用いた。二次イオン極性は、タンパク質の構成アミノ酸由来の二次イオンが感度良く多数検出される正 2 次イオン（Positive）を選択した。その他の測定条件については Table 3.1 に示した。

Table. 3.1　TOF-SIMS の主な測定パラメーター

	Primary ion	Sputtering ion
Ion specie	Bi_3^{2+}	Ar_{1500}^{+}
Energy	60 keV	2.5 keV
Current	Approx. 0.1 pA	0.1 nA
Field of view	50 µm × 50 µm	500 µm × 500 µm
Pixel number	64 × 64	64 × 64
Dose density	approx. 1×10^{11} ions/cm^2 /cycle	approx. 4×10^{13} ions/cm^2 /cycle
Charge compensation	ON（Low-energy electron flooding）	
Scan number (data points)	10,000 scans	

3.3 データ解析

3.3.1 データ前処理

TOF-SIMS 装置付属の解析ソフトウェアである、SurfaceLab 6.5（IONTOF GmbH, Münster, Germany）を用いて、質量のキャリブレーションを実施した（$^{15}CH_3^+$, $^{27}C_2H_3^+$, $^{39}C_3H_3^+$, $^{53}C_4H_5^+$の 4 つの既知のピークを用いた）。また同ソフトウェアの Auto peak search 機能を用いて、*m/z* 12 ~ 500 の質量範囲から 626 のピークを自動選択し、デプスプロファイルデータを作成した。なお、10,000 scan

のデータをすべて積分した総二次イオンスペクトル（Total secondary ion spectrum）における最低ピーク強度が 100 counts、S/N が 1.0 以上のピークを対象として Auto peak search を実施した。更に 10 cycle の移動平均を用いてスムージングを行った後、最終的に得られた 626 peaks×9,999 cycles のデータをオートエンコーダーと PCA の解析に使用した。

データの前処理として、PCA では通常、Mean-centering や Poisson scaling を用いる場合が多く、本検討でも Mean-centering を実施した。一方で、TOF-SIMS データのオートエンコーダーによる解析について、最適な前処理方法は明らかでない。そのため、本検討では、オートエンコーダーの解析に使用するデータに対し、前処理は実施せず、強度データをそのまま入力データとした。

3.3.2 データ解析条件

高速な行列計算ライブラリや機械学習ライブラリが存在し、現在、データサイエンスに盛んに利用されているプログラミング言語である Python を、オートエンコーダーの実行環境として用いた。オートエンコーダーの実行には、深層学習ライブラリである KERAS[84]を用い、KERAS のバックエンドとして Tensorflow を用いた。使用した主な Python ライブラリの詳細を Table 3.2 に示した。

Table 3.2　オートエンコーダーの解析に使用した Python ライブラリ

Library name	Version	Description
Python	3.7.6	General purpose programming language
Numpy	1.17.3	Array processing for numbers strings records, and objects
KERAS-gpu	2.3.1	Deep learning library for Theano and Tensorflow
KERAS-applications	1.0.8	Applications module of the keras deep learning library
KERAS-preprocessing	1.1.0	Data preprocessing and data augmentation module of the keras deep learning library
Tensorflow-gpu	2.0.0	Machine learning library

オートエンコーダーの具体的な構造として、Figure 2.15 に示したエンコーダーとデコーダーの 2 つの部分から成るシンプルなネットワーク構造を採用した。中間層のサイズ（ニューロンの数）は、既知の情報である、キューティクルを構成する層の数や浸透させたヘアケア剤組成を考慮して、10 に設定した。したがって、626 次元の入力データを中間層にて 10 次元に圧縮し、出力層にて再度 626 次元に再構成した。エンコーダーとデコーダーそれぞれの活性化関数には ReLU を採用した。「2.2.3　活性化関数」の項で示した通りに、ReLU は負値を出力せず、出力上限値を持たない。この特徴は負値を持たない TOF-SIMS データについて Scaling を行わずに解析できる点で適当と考えられる。損失関数には平均二乗誤差（MSE）、最適化関数には Adam[85]を採用し、バッチサイズは 128 とした。

PCA の実行は MATLAB R2015b（Mathworks, Inc., USA）上で動作する多変量解析ソフトウェアである PLS-toolbox 8.0.2（Eigenvector Research, Inc., USA）を用いて行った。

3.3.3 計算実行条件

オートエンコーダーおよび PCA の計算環境としては、市販のノートパソコンを用いた。計算に関わる主な使用について以下に示した。

Table. 3.3　解析に使用したコンピューターの主な使用

Component type	Specifications
CPU	Intel® Core™ i7-8850H 2.6GHz
GPU	NVIDA® Quadro® P2000 with Max-Q Design
RAM (Graphic mamory)	6 GB
RAM (Main memory)	PC4-21333 DDR4 SDRAM SODIMM 32 GB
Strage	512GB solid state drive (M.2 PCIe NVMe)

3.4 結果と考察

3.4.1 TOF-SIMS 測定結果

TOF-SIMS 測定によって得られた、毛髪試料の総二次イオン質量スペクトルを Figure 3.4 に、毛髪内部に浸透させたヘアケア剤標品の二次イオン質量スペクトルを Figure 3.5 に示した。Figure 3.4 からは、毛髪を構成するタンパク質由来のアミノ酸フラグメントが特徴的に検出された。$C_4H_8N^+$はプロリンやアルギニン、ロイシン、$C_5H_{12}N^+$はロイシンとイソロイシン、$^{100}C_4H_{10}N_3^+$, $^{111}C_4H_{11}N_3^+$はアルギニン、$^{110}C_5H_8N_3^+$はヒスチジン、$^{120}C_8H_{10}N^+$はフェニルアラニン、$^{107}C_7H_7O^+$, $^{136}C_8H_{10}NO^+$はチロシンに特徴的な二次イオンピークである。特に $^{76}C_2H_6SN^+$は毛髪中に多く含まれるシステイン（またはシスチン）に対応するピークである[28, 86-87]。そのほか、*m/z* 300 以上の質量域において、$^{312}C_{21}H_{46}N^+$, $^{368}C_{25}H_{54}N^+$がわずかに検出された。Figure 3.5 より、これらのピークはヘアケア剤中に含まれる長鎖アルキル四級アンモニウムカチオン（トリメチルステアリルアンモニウム、トリメチルベヘニルアンモニウム）と考えられる。したがって、毛髪内部にヘアケア剤成分が浸透していることが確認された。なお、ヘアケア剤標品からはシリコーンオイル成分であるポリジメチルシロキサン由来の $^{73}Si(CH_3)_3^+$や $^{147}Si_2O(CH_3)_5^+$も観測された。

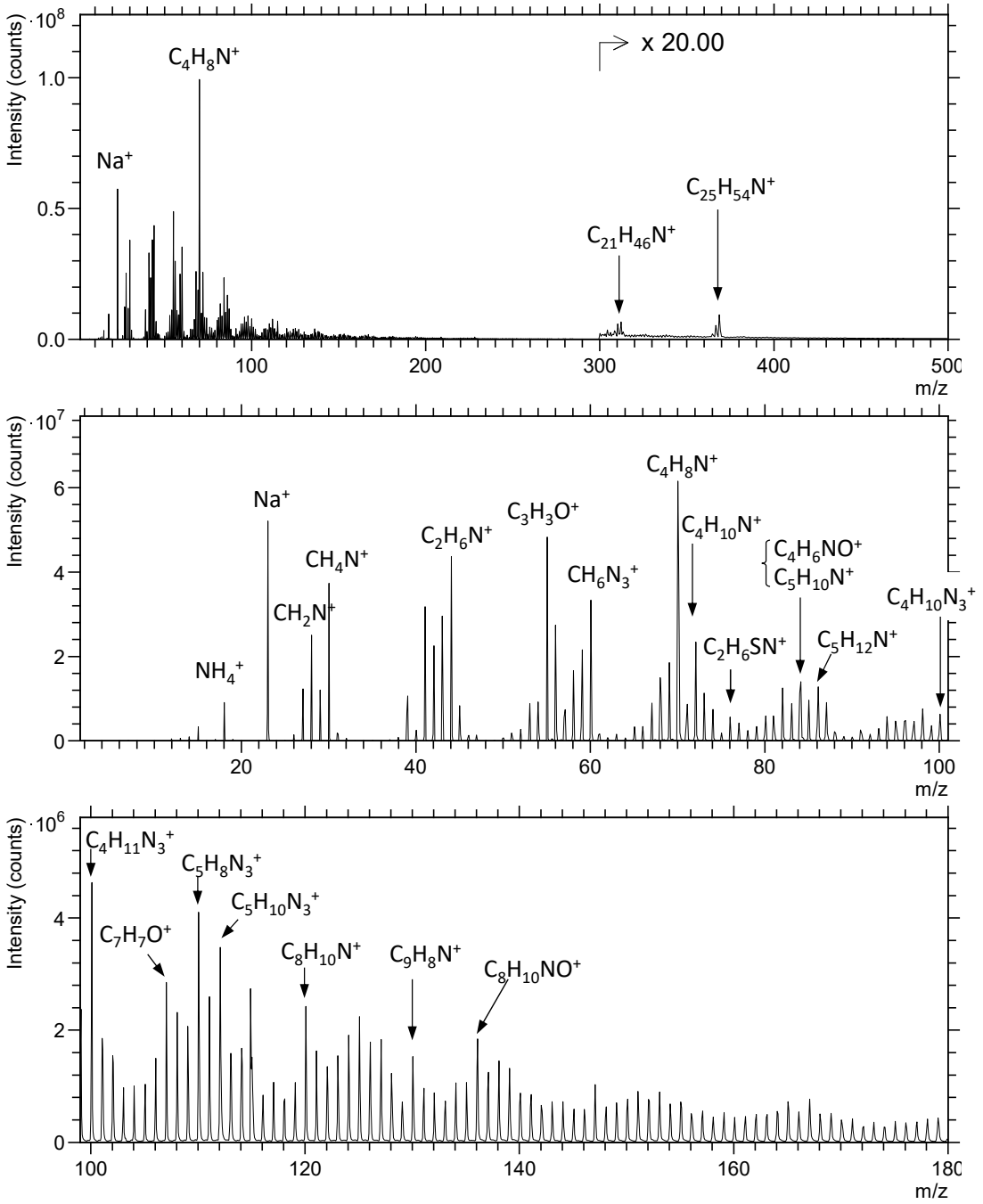


Figure 3.4　TOF-SIMS デプスプロファイル測定を通じて得られた質量スペクトル（正 2 次イオン）

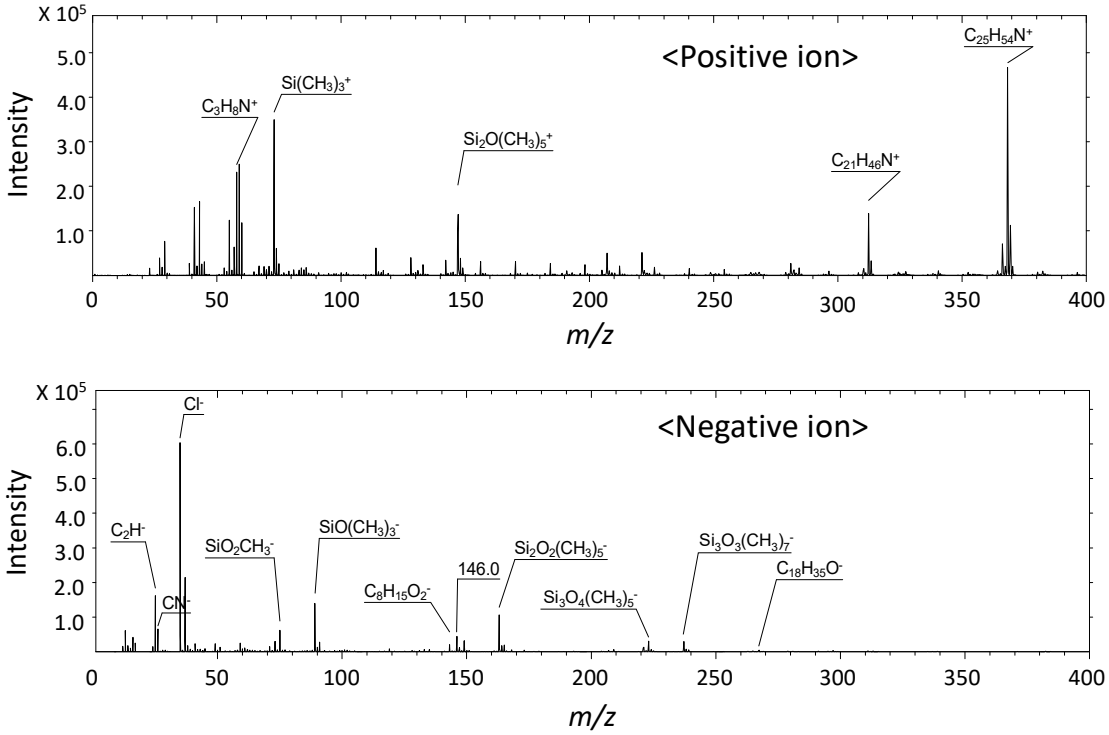


Figure 3.5　毛髪に浸透させた市販ヘアケア剤の **TOF-SIMS** 質量スペクトル
（上段：正二次イオン、下段：負二次イオン）

3.4.2 オートエンコーダーによる学習

TOF-SIMS 測定によって得られたデプスプロファイルデータ（626 peaks×9,999 cycles）に対して、オートエンコーダーの学習を行った際の、学習回数（Epoch）に対する損失関数（MSE）の変動を Figure 3.6 に示した。学習回数の増加と共に、入力データと出力データの誤差である損失関数の値が減少してく様子が認められた。190 epoch と 7200 epoch 付近にて急激に損失関数の値が減少するが、それ以降は変化が小さく、4×10^3 で一定値となることから、10000 epoch にて学習は十分に進行しているものと判断した。そこで 10000 epoch の学習を実行した後のデコーダーの出力データ（再構成データ）より、100 ピークのデプスプロファイルをランダムに選択・抽出し、入力データ（オリジナルの TOF-SIMS データ）と比較した（Figure 3.7）。その結果、横軸 0～9999 cycle の全範囲で強度が 30 counts 以上の二次イオンピークについては、出力データにて概ね再現されていることが確認されたが、それ以下の強度の二次イオンピークについては横軸の全範囲で 0 を出力した。この結果より、本検討に用いた単純なオートエンコーダーでは強度の弱いデータについては再現が十分ではなく、それが入力と出力の誤差である損失関数が 4×10^3 以下に低下しない原因と考えられる。なお、ヘアケア剤から特徴的に検出された長鎖アルキル四級アンモニウムカチオン（$^{312}C_{21}H_{46}N^+$, $^{368}C_{25}H_{54}N^+$）についても確認を行ったところ、表面側（1～約 100 cycle）で強度が強い様子は、デコーダーの出力結果においても再現されていることが確認された。しかし、横軸 400 cycle 以降の強度が低い領域では、10 count 程度の強度を境界として再現が十分ではない。

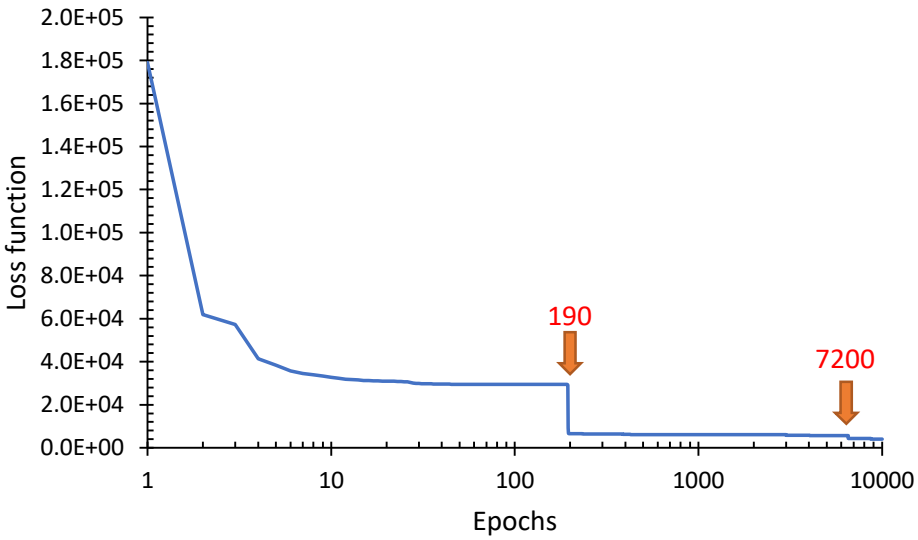


Figure 3.6　訓練回数（Epoch）に対する損失関数の値の推移

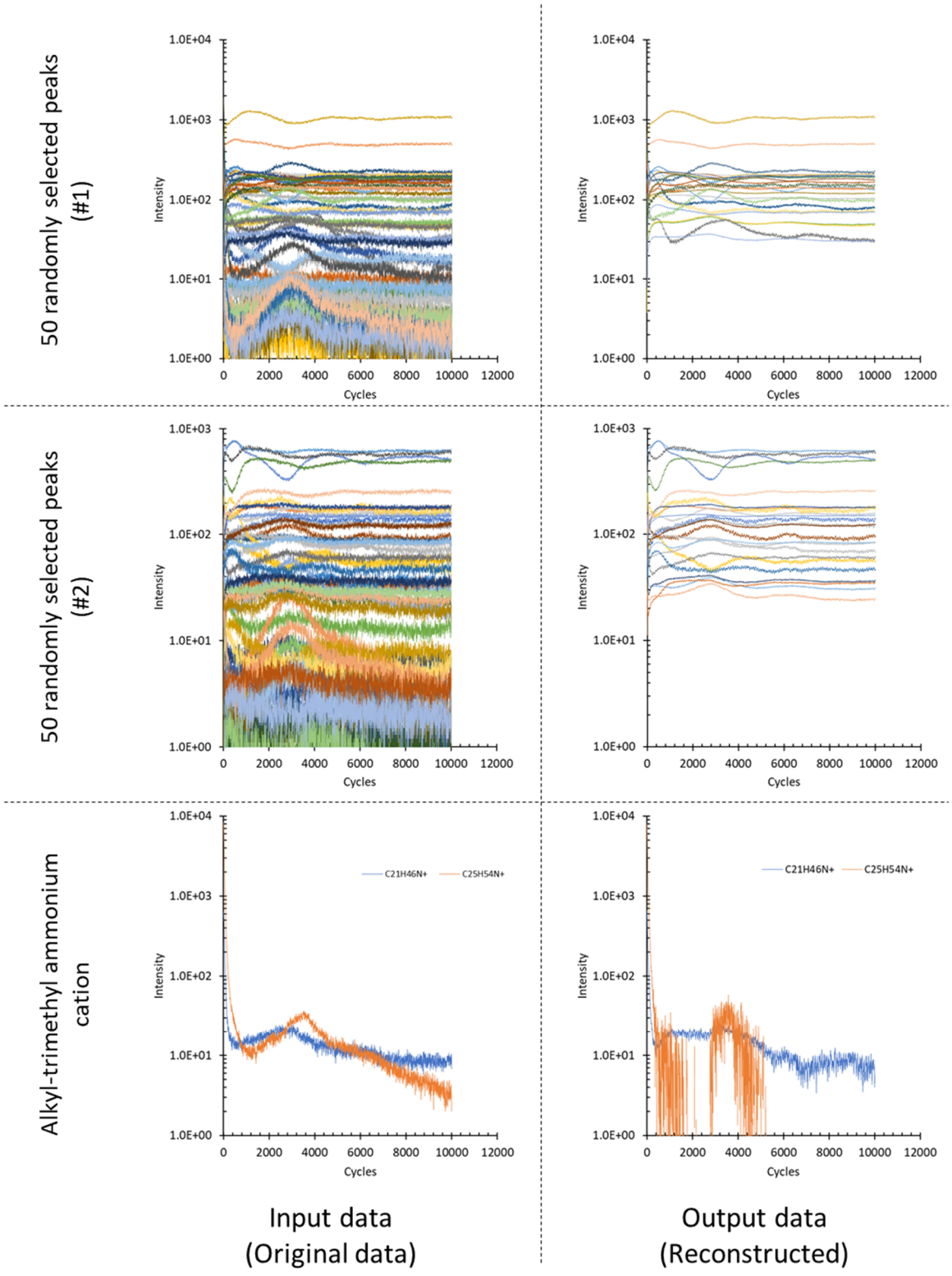


Figure 3.7　代表的な二次イオンのデプスプロファイルの入力データと出力データの比較.

3.4.3 抽出された特徴の可視化

オートエンコーダーによって中間層に抽出される特徴のデータサイズは（10 features × 9999 cycles）であり、20 個のデプスプロファイル形式で表すことができる。中間層の 20 個のニューロンそれぞれに対応する特徴を$\hat{X}1, \hat{X}2, \hat{X}3 \cdots\cdots \hat{X}10$ として Figure 3.8 に示した。$\hat{X}2$と$\hat{X}3$を除く 8 個のニューロンに抽出された特徴は深さ方向にそれぞれ強度変動が認められ、組成分布を反映した特徴が抽出されていると推察された。一方で$\hat{X}2$と$\hat{X}3$では完全に 0 が出力されており、入力情報を何も反映していない。$\hat{X}2$と$\hat{X}3$が 0 を出力した理由としては、学習過程で重み（W）が負側に大きく更新された結果、これらのニューロンに入るデータ（$W_{enc} \cdot X$）が負値となり、活性化関数：ReLU で出力された値が 0 になったためと考えられる（ReLU は負値を 0 として出力する。（第 2 章「2.2.3 活性化関数」参照）。これは「Dying ReLU」として知られており[63, 64]、活性化関数に ReLU を用いた場合にしばしば生じる問題である。

Dying ReLU が起こるニューロンの個数が適当な場合、中間層での次元削減がより進行するため、特徴抽出にプラスに働く可能性もある。しかしながら、中間層のニューロン数に対して Dying ReLU の状態になるニューロンが多すぎると、勾配が 0 となるニューロンが多くなる（ReLU は 0 での勾配も 0 である）ことから、勾配降下法による学習がうまく進行せず良好な特徴が得られなくなる可能性がある。今回の結果は中間層のニューロンの個数の 80 %には特徴が抽出されたことから、許容範囲内と判断した。なお、Dyingn ReLU を回避する方法としては次式および Figure 3.9 に示した、活性化関数 ReLu の負の入力範囲にわずかに勾配を持つ Leaky ReLU の使用が挙げられる（α は一般に 0.01 が使用される）。Leaky ReLU では中間層のニューロンへの入力値が負値であったとしても、勾配が 0 にならないため、学習が停滞することなく進むことが期待される。ただし、中間層の出力データ（$\hat{X} = f(W_{enc}X + B_{enc})$に負値を持つ可能性があるため、特徴の解釈が難しくなるというデメリットもある。そのため、本検討では Leaky ReLU の採用は行わなかった。

$$\phi(z) = \begin{cases} \alpha \cdot z, & z \leq 0 \\ z, & z > 0 \end{cases}$$

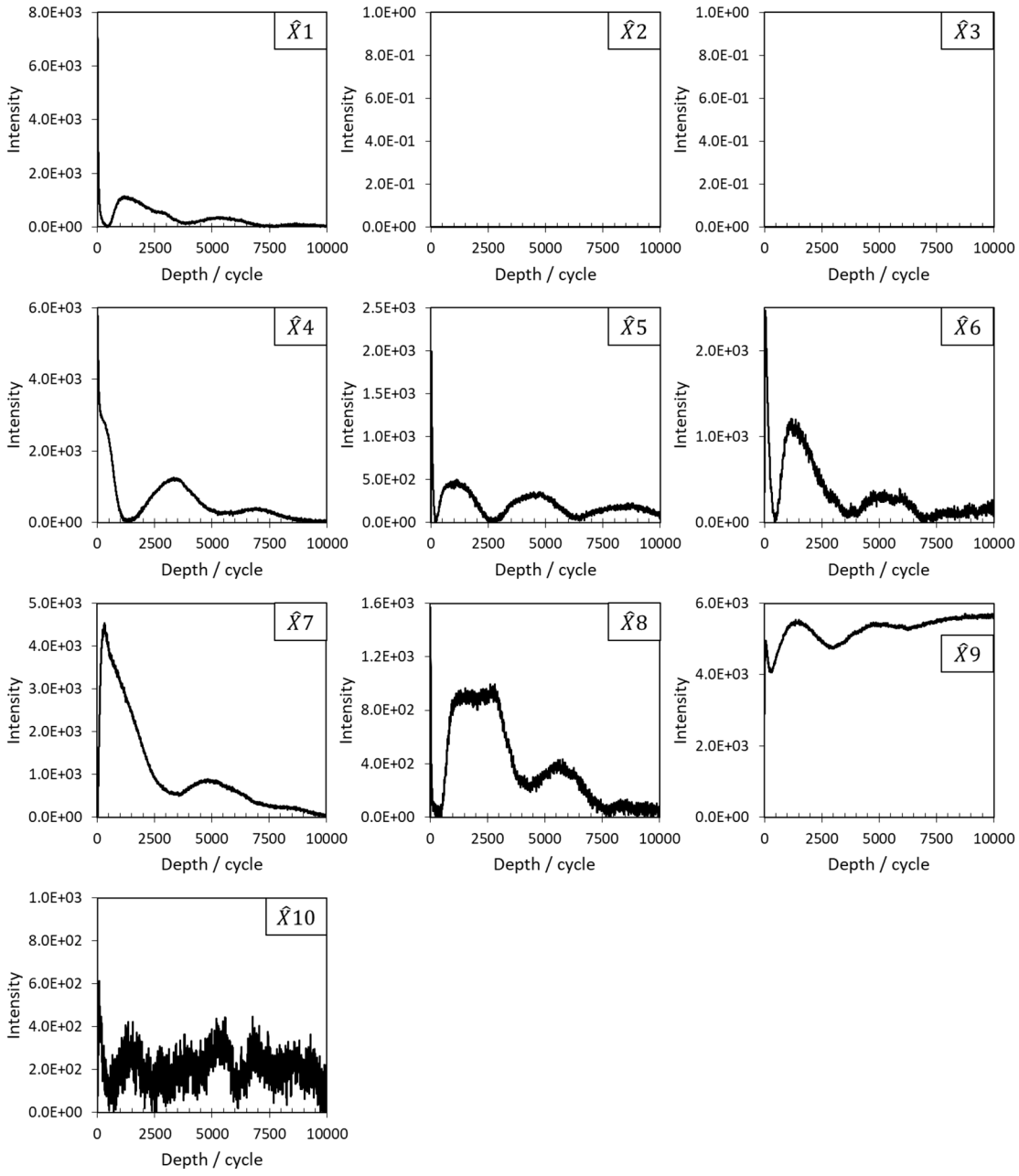


Figure 3.8 中間層（10 ニューロン）に抽出された特徴（デプスプロファイル）

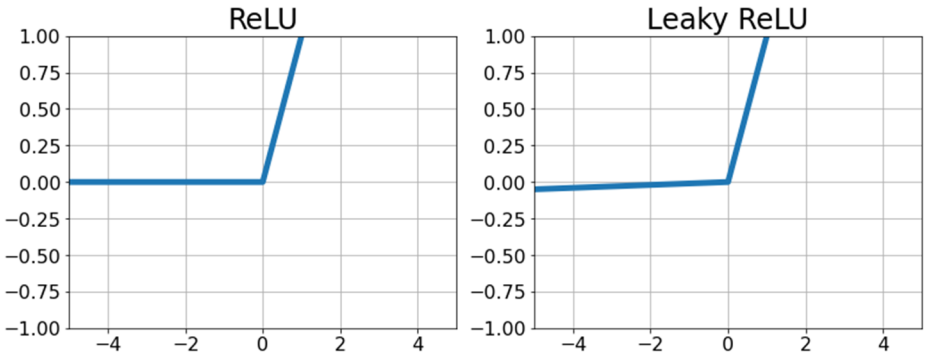


Figure 3.9　活性化関数 ReLU と Leaky ReLU の比較

3.4.4　Encoder Weighs と Decoder Weights の比較

中間層ニューロンに特徴として抽出されたデータの中身については、そのニューロンに繋がる重み（ベクトル）を調べることで、どのような変数（二次イオンピーク）の寄与が高いかを調べることで可能である。重みには入力データを特徴に変換するエンコーダーの重み（W_{enc}）と、変換された特徴を入力データに再構成するデコーダーの重み（W_{dec}）の 2 つが存在する（Figure 3.10）。W_{enc}は（20×468）の行列、W_{dec}は（468×20）の行列形式で得られ、W_{enc}の列ベクトルとW_{dec}の行ベクトルのそれぞれは、468 本のピークから成る質量スペクトル形式で表すことができる。Figure 3.11 および Figure 3.12 にエンコーダーの重み（W_{enc}）とデコーダーの重み（W_{dec}）を各 10 個の質量スペクトル形式で示した。ここで、$W_{enc}1$と$W_{dec}1$は同じ中間層ニューロンへに関連する重みである。

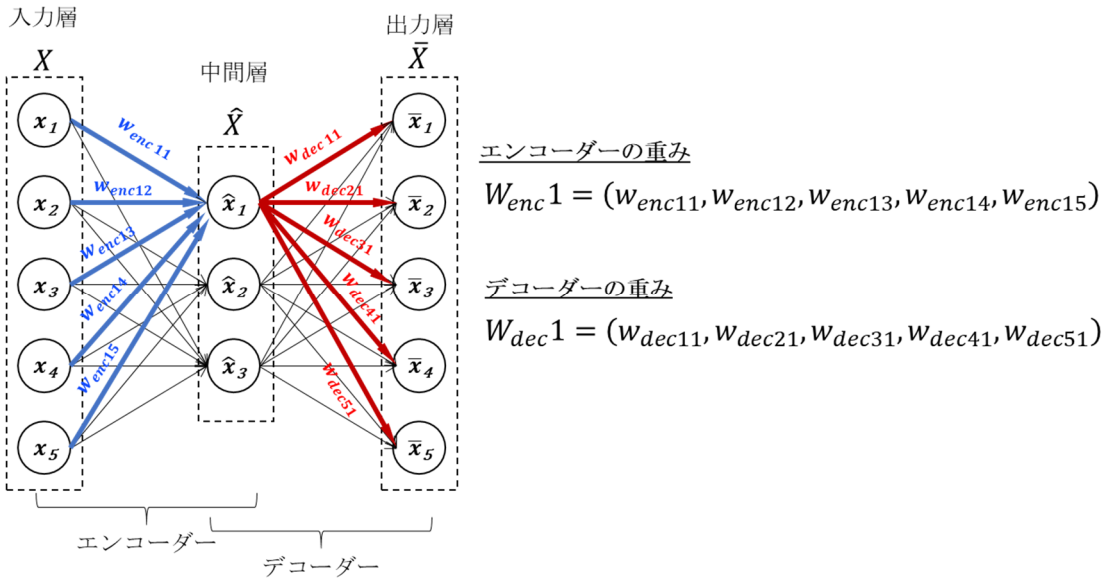


Figure 3.10　入力層と中間層、出力層から成るオートエンコーダーの略図

Figure 3.11, 3.12 からわかる通り、重みデータは正負両方の値を持つ。しかしながら、活性化関数に負値を出力しない ReLU を使用しているため、中間層に抽出された特徴に寄与するのは正の重みのみとなる。そのため、特徴の解釈には重みの正値のみを考慮すれば良い。

中間層の同一ニューロンに関連するW_{enc}とW_{dec}を比較すると、W_{dec}の方で元データにて観測された主要なピークが多く表れる傾向が認められた。例として Figure 3.12-2 の$W_{dec}8$では、$^{44}C_2H_6N^+$, $^{70}C_4H_8N^+$, $^{86}C_5H_{12}N^+$, $^{110}C_5H_8N_3^+$, $^{120}C_8H_{10}N^+$などのタンパク質から観測される代表的なフラグメントイオン（タンパク質の構成アミノ酸に対応）が高強度で表された。毛髪は主にケラチンタンパク質より構成されるため、この結果は妥当と考えられる。一方で、Figure 3.11-2 の$W_{enc}8$では、$^{1}H^{34}S^+$, $^{44}Ca^+$, $^{65}Cu^+$などの比率の低い同位体を含むピークがより比率の高い同位体を含むピークより強く観測され、*m/z* 398.42 などの帰属不明なピークも高強度で観測された。これらの$W_{enc}8$で特徴的に観測されたピークは、元の TOF-SIMS データ（Figure 3.4）では強度が非常に弱く、情報の有用性が相対的に低い。また、Figure 3.12-1 の$W_{dec}1$ではヘアケア剤中に含まれる長鎖アルキル四級アンモニウムカチオン由来の $^{58}C_3H_8N^+$や $^{112}C_7H_{16}N^+$、$^{368}C_{25}H_{54}N^+$、シリコーンオイル由来の $^{73}SiC_3H_9^+$, $^{147}Si_2OC_5H_{15}^+$などが強く観測されているのに対し、Figure 3.11-1 の$W_{enc}1$では $^{58}C_3H_8N^+$や $^{73}SiC_3H_9^+$, $^{147}Si_2OC_5H_{15}^+$なども主要なピークとして認められるものの、その強度は小さく、代わりに *m/z* 34.03, 376.39 などが特徴的に観測された。

このように、エンコーダー側の重み（W_{enc}）でノイズのような小さいピークが特徴的に表れる原因は次のように考えられる。つまり、エンコーダーに入力される元データには元々ノイズ様のピークが多く含まれているため、必然的に低次元表現へと落とし込む際にそれらのピークにも重み付けがなされる。一方でデコーダー側の重み（W_{dec}）では、一度、低次元表現に落とし込まれたことによってノイズの影響が軽減したものから元データを復元しようとするため、元データにて強度が比較的強いピークの重みが相対的に大きくなるように学習される（入力を再現するのに必要なデータを優先的に重み付が多くなされる）。 そこで、本研究では主にW_{dec}を用いて抽出された特徴（Figure 3.8）の解釈を行った。

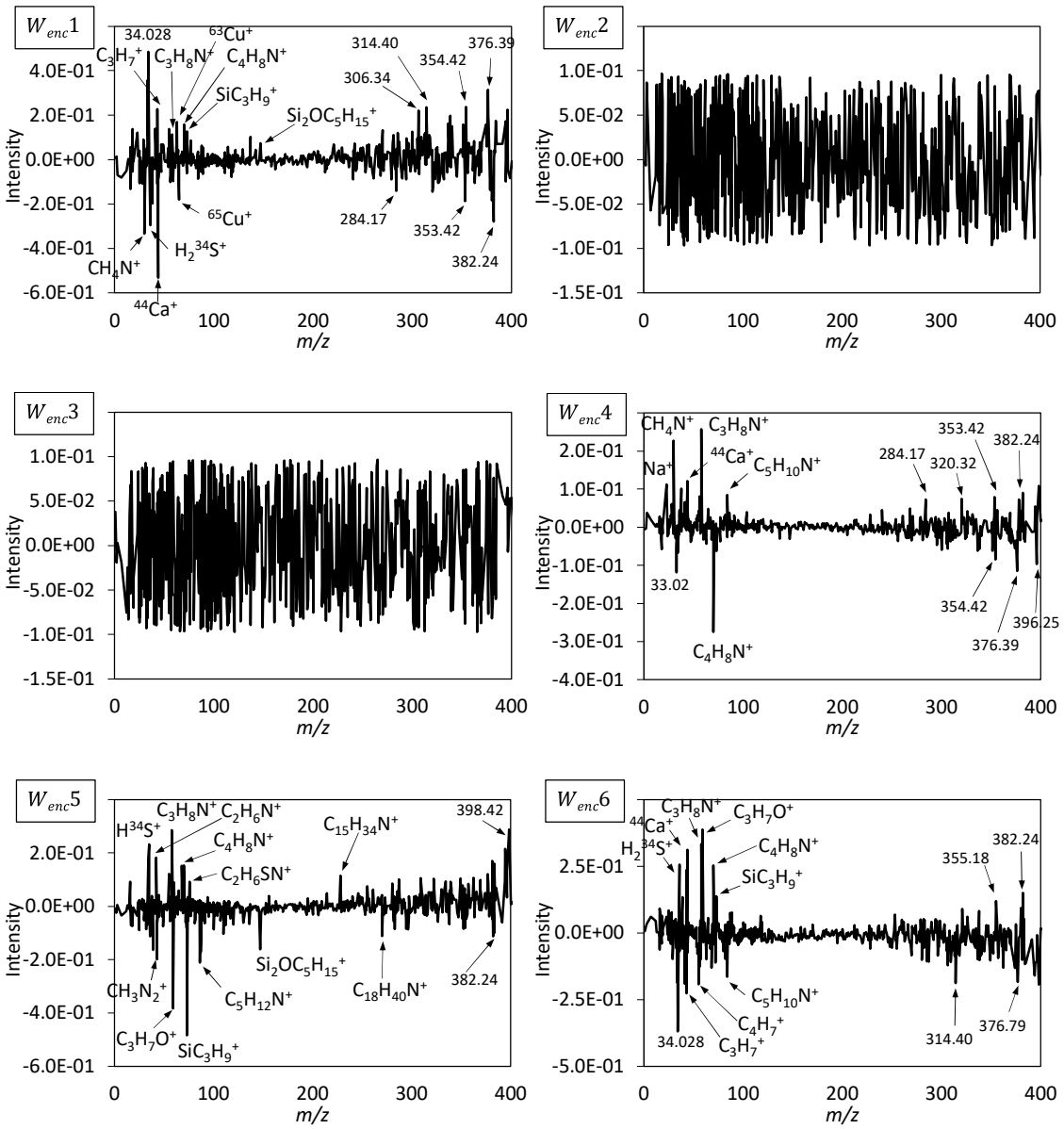


Figure 3.11-1　エンコーダーの重み（$W_{enc}1$ から $W_{enc}6$）

＜活性化関数:ReLU によって中間層には正値のみが出力されるため、負の重みについては無視できる＞

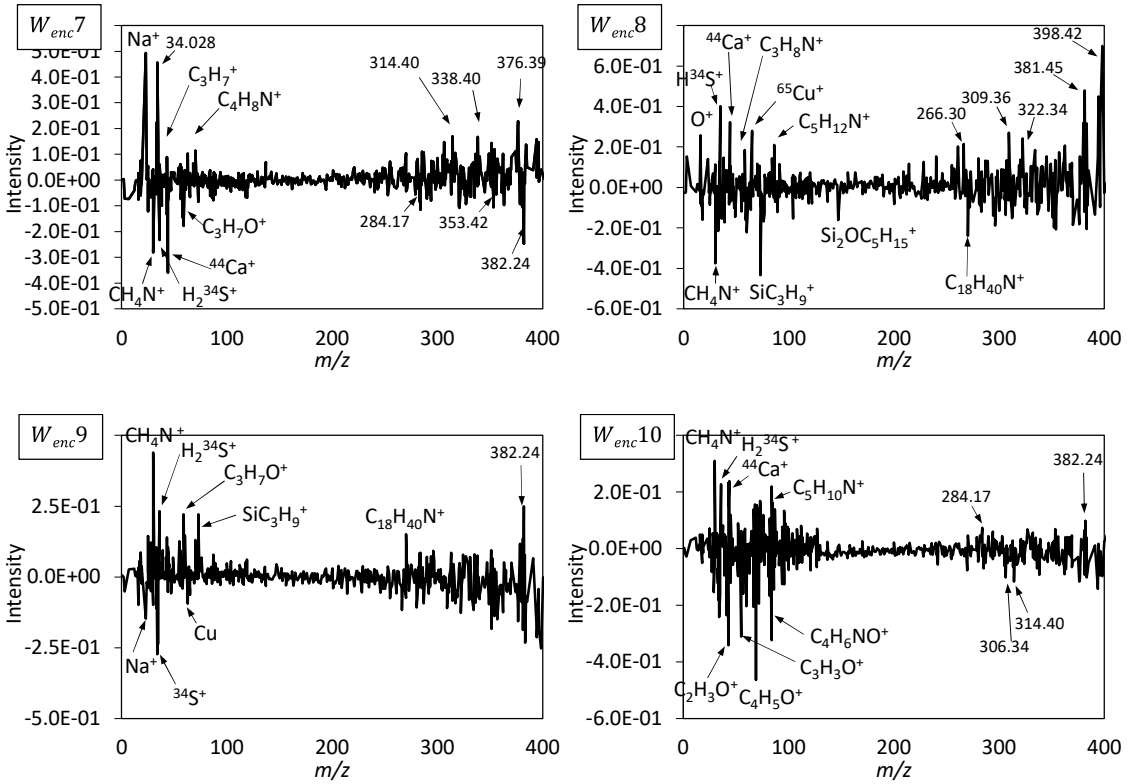


Figure 3.11-2　エンコーダーの重み（W_{enc}7 から W_{enc}10）

＜活性化関数:ReLU によって中間層には正値のみが出力されるため、負の重みについては無視できる＞

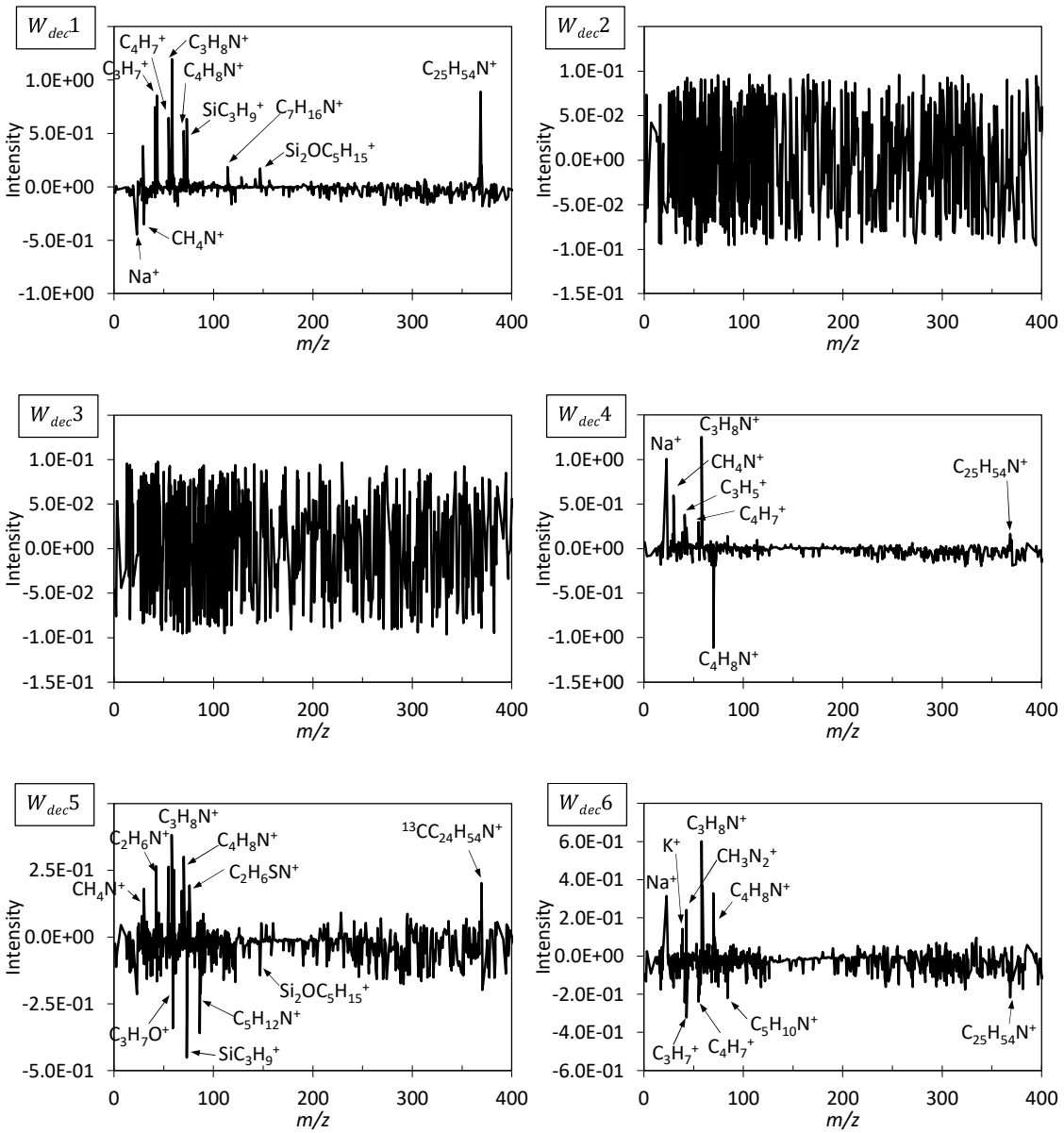


Figure 3.12-1　デコーダーの重み（$W_{dec}1$ から $W_{dec}6$）

＜活性化関数:ReLU によって中間層には正値のみが出力されるため、負の重みについては無視できる＞

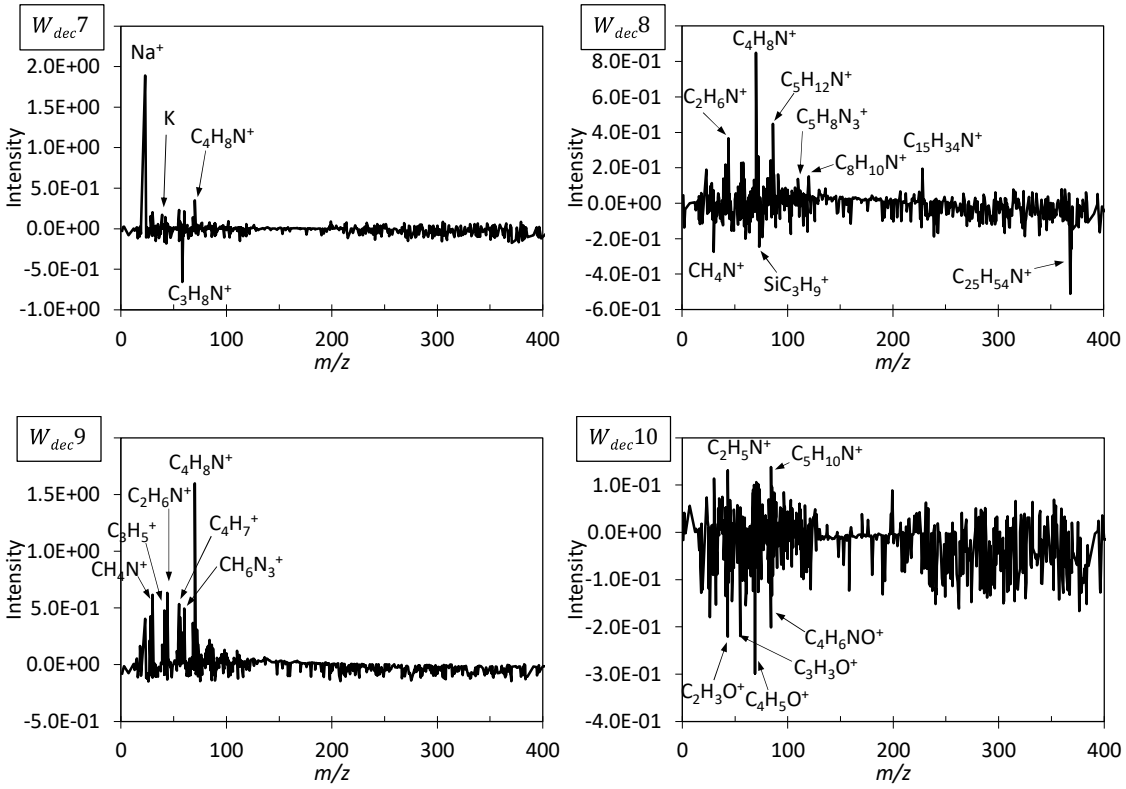


Figure 3.12-2　デコーダーの重み（$W_{dec}7$ から $W_{dec}10$）

＜活性化関数:ReLU によって中間層には正値のみが出力されるため、負の重みについては無視できる＞

Figure 3.12-1 の$W_{dec}1$には、前述の通り長鎖アルキル四級アンモニウムカチオン（$^{368}C_{25}H_{54}N^+$）やシリコーンオイル（$^{73}SiC_3H_9^+$, $^{147}Si_2OC_5H_{15}^+$）が高強度で表れることから、これらの成分が特徴$\hat{X}1$に大きく寄与していることがわかる。同様に、$W_{dec}4$と$W_{dec}7$を見ると、特徴$\hat{X}4$ではナトリウム（$^{23}Na^+$）と長鎖アルキル四級アンモニウムカチオンのピークが、特徴$\hat{X}7$では Na が大きく寄与していることがわかる。 また，$W_{dec}5, 6, 8, 9$には，タンパク質の構成アミノ酸由来のピーク（$^{30}CH_4N^+$, $^{44}C_2H_6N^+$, $^{70}C_4H_8N^+$, $^{76}C_2H_6SN^+$, $^{86}C_5H_{12}N^+$, $^{110}C_5H_8N_3^+$, $^{120}C_8H_{10}N^+$）が顕著に認められた。特に $^{76}C_2H_6SN^+$はシステインやシスチンに特徴的であり、$^{110}C_5H_8N_3^+$はヒスチジン、$^{120}C_8H_{10}N^+$はフェニルアラニンに特徴的なイオン種である。つまりW_{dec}の解析より、各特徴はそれぞれ下表に示す成分の分布を示していると考えられる。

Table 3.4 オートエンコーダーにより抽出された特徴の帰属

特徴	主に寄与する成分
$\hat{X}1$	長鎖アルキル四級アンモニウムカチオン、シリコーンオイル、ナトリウム
$\hat{X}4$	長鎖アルキル四級アンモニウムカチオン、ナトリウム
$\hat{X}5$	タンパク質（システイン／シスチンの比率が大）
$\hat{X}6$	タンパク質、ナトリウム
$\hat{X}7$	ナトリウム
$\hat{X}8$	タンパク質（ヒスチジン、フェニルアラニンの比率が大）
$\hat{X}9$	タンパク質

3.4.5 抽出された特徴の妥当性の検証

各特徴の解釈の結果をもとに、長鎖アルキル四級アンモニウムカチオンの分布について詳細な分析を行った（Figure 3.13（A））。$\hat{X}4$（長鎖アルキル四級アンモニウムカチオン、ナトリウム）と$\hat{X}7$（ナトリウム）の比較から、長鎖アルキル四級アンモニウムカチオンは最表面（1～20 サイクル）と約 3300 サイクルの深さに分布していると考えられる。$\hat{X}5$の強度が小さくなる約 2700 サイクルが、システイン／シスチンの濃度が低い領域と考えられることから、長鎖アルキル四級アンモニウムカチオンはその領域よりもやや深部側に局在していると考えられる。ここで Figure 3.2, 3.3 に示した TEM 観察および TEM-EDX による硫黄（S）のマッピング分析の報告を参照すると、システイン／シスチンの濃度が低い領域（2700 サイクル）がエンドキューティクルであり、システイン／シスチンの濃度が高い領域（4000～5000 サイクル）が a 層～エクソキューティクルと考えられる。したがって、長鎖アルキル四級アンモニウムカチオンの局在領域は、エンドキューティクルと a 層の間に位置する細胞膜複合体（CMC）に対応すると考えられる。

CMC は β 層／δ 層／β 層の三層より構成されており、β 層が脂質を多く含み疎水性、δ 層が親水性であることが知られている[83, 88]。さらに村越によるイオン性蛍光色素を用いた観察結果より、キューティクル表面は疎水性であるが、キューティクルのエッジ部分では親水性の δ 層が表面に露出していると考えられる[88]（Figure 3.14 参照）。そのため、キューティクルエッジから δ 層を通じて長

鎖アルキル四級アンモニウムカチオンが毛髪内部に浸透することは妥当と考えられる。つまり、オートエンコーダーによって、生物学的知見に照らし合わせて妥当な結果が抽出されたと言える。

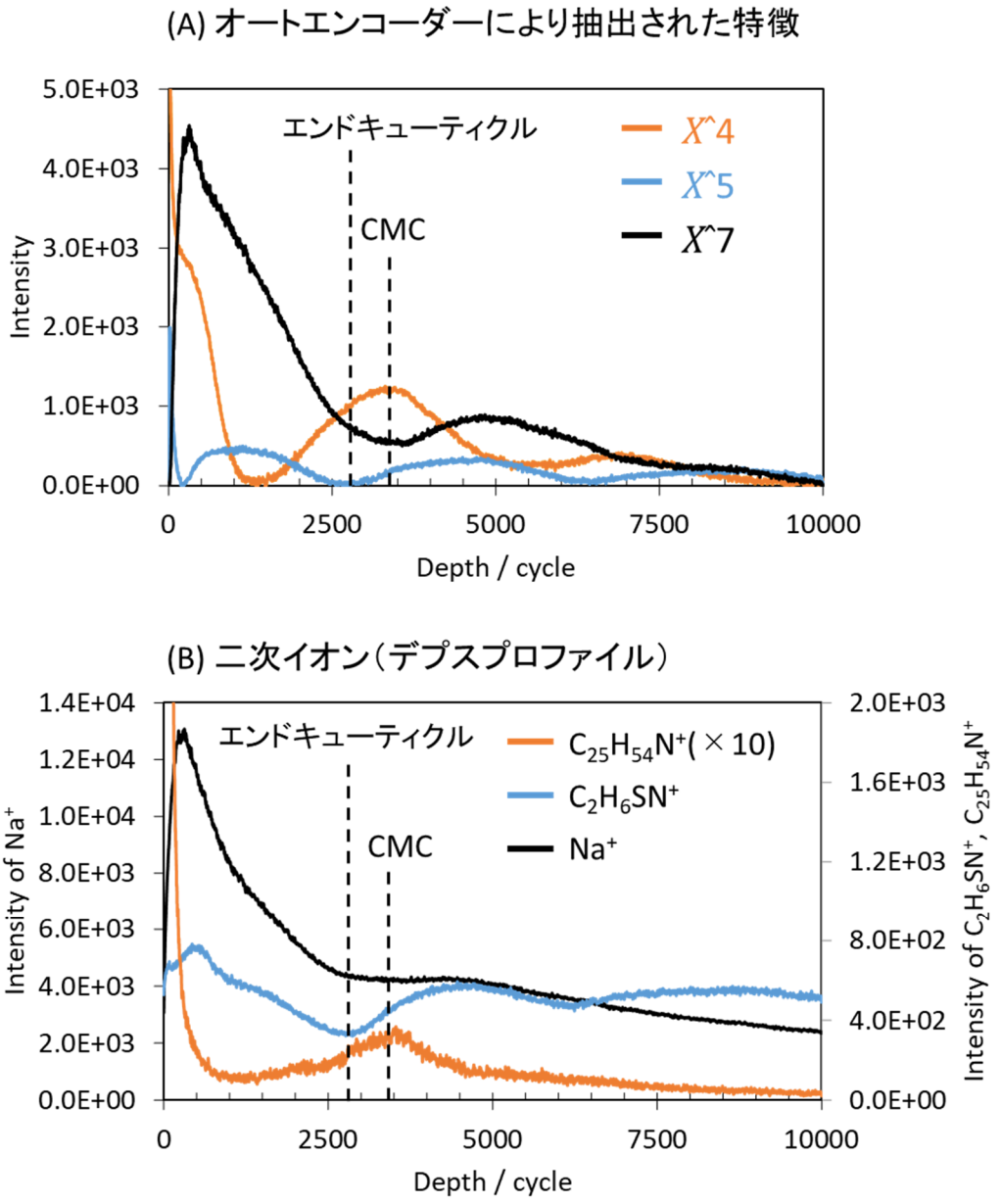


Figure 3.13　オートエンコーダーにより抽出された特徴と二次イオン強度の比較　(A) 長鎖アルキル四級アンモニウムカチオン、ナトリウム、システイン／シスチン含有タンパク質の寄与が大きな特徴　(B) 二次イオンデプスプロファイル：Na^+, $C_2H_6SN^+$ (Cystine/Cysteine), and $C_{25}H_{54}N^+$ (cationic surfactant).

更にオートエンコーダーで抽出された特徴の妥当性を評価するために、TOF-SIMS データから直接、各成分に特徴的な二次イオンの分布（デプスプロファイル）を作成し、オートエンコーダーで抽出された特徴との比較を行った。

Figure 3.13（B）に長鎖アルキル四級アンモニウムカチオン（$^{368}C_{25}H_{54}N^+$）とシステイン／シスチン（$^{76}C_2H_6SN^+$）、ナトリウム（$^{23}Na^+$）の二次イオンのデプスプロファイルを示した。ナトリウムが表面側から深部側にかけて減少していく様子や、システイン／シスチンの増減の周期構造、長鎖アルキル四級アンモニウムカチオンが CMC と推測される領域で多いことについては、オートエンコーダーによって抽出された特徴とよく一致した。しかしながら、システイン／シスチン（$^{76}C_2H_6SN^+$）が横軸約 500 cycle においてピーク状に強度が増す点など、オートエンコーダーの特徴に反映されていない情報も認められた。つまり、オートエンコーダーによって、各成分の分布が完全に抽出できるわけではない（これは他の特徴抽出法でも同じである）。そのため、抽出された特徴から概要把握や注目すべき成分の見当をつけた後、必要に応じて元の TOF-SIMS データに戻って解析を行うことが望ましいと考えられる。

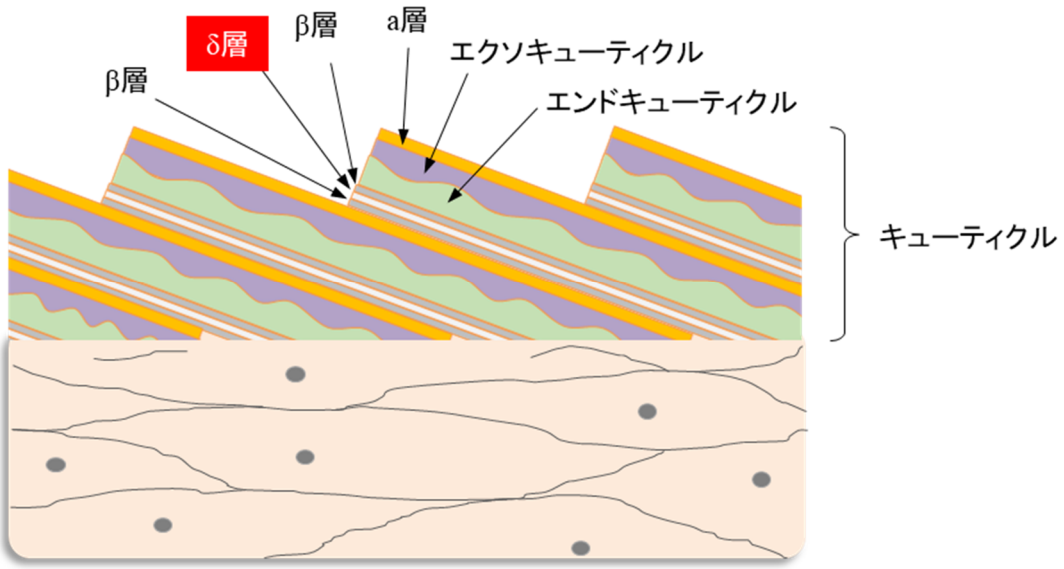


Figure 3.14　毛髪表面（キューティクル部位）の模式図（[88]の文献を基に作成）
＜キューティクルエッジでは親水性の δ 層が表層に露出している＞

3.4.6 PCA との結果の比較

オートエンコーダーと同じデータについて、主成分分析を行った。各主成分の寄与率の推移（各主成分の分散とデータ全体の分散の比）を示したスクリープトットを Figure 3.15 に示した。第一主成分（PC1）から第四主成分（PC4）までで、元データの 99 %以上の情報を含むことから、本検討では PC4 までの結果を前述の Autoencoder の特徴抽出結果と比較した。なお、PCA により得られる主成分得点（スコア）は圧縮されたデータ、つまり「特徴」を表す。また、ローディングは初期変数（すなわち TOF-SIMS ピーク）と新しく生成された変数との関係を表し、これを調べることでスコアの内容を解釈することができる。ただし、PCA ではスコアおよびローディングに正値と負値を持つ点に注意する必要がある。

PCA によって得られたスコアプロットを Figure 3.16 に、ローディングプロットを Figure 3.17 に示した。PC1 の正値にはアルカリ金属（$^{23}Na^+$, $^{39}K^+$）の寄与が大きく、ヘアケア剤成分の長鎖アルキル四級アンモニウムの寄与もわずかに含まれる。負値には多くのアミノ酸に共通する $^{70}C_4H_8N^+$が表れ、タンパク質の寄与が大きいと考えられる。同様に PC2 では正値に長鎖アルキル四級アンモニウム（$^{368}C_{25}H_{54}N^+$）の寄与が大きく、負値にタンパク質（$^{70}C_4H_8N^+$）やアルカリ金属（$^{23}Na^+$）の寄与が大きい。PC4 の正値にはタンパク質由来のピーク（$^{30}CH_4N^+$, $^{42}C_2H_4N^+$, $^{60}CH_6N_3^+$, $^{76}C_2H_6SN^+$）が表れたが、特にシステイン／シスチンに特徴的な $^{76}C_2H_6SN^+$の寄与が大きいことから、シスチン／システイン比率の高いタンパク質の寄与が大きいと考えられる。また、長鎖アルキル四級アンモニウム（$^{368}C_{25}H_{54}N^+$）も、PC4 の正値に大きく寄与している。以上のローディングプロットの解釈結果を Table 3.5 にまとめ、Figure 3.16 のスコアプロット中にも記載した。

Table 3.5 の結果からスコアプロット（Figure 3.16）を解釈すると、PC2 は、長鎖アルキル四級アンモニウムが主に最表面（1～20 サイクル）と約 3300 サイクルの深さに分布していることを示している。また、PC2 と PC4 のスコアプロットの比較から、長鎖アルキル四級アンモニウムがシステイン／シスチン比率の低いタンパク質層（約 2700 サイクル）よりもやや深い領域に分布していると考えられる。この結果は Figure 3.13(A)に示したオートエンコーダーの解析結果と一致している。

Table 3.5　PCA ローディングプロットの解析より得られた各主成分に特徴的な成分

主成分	正値	負値
第一主成分（PC1）	アルカリ金属（Na、K） 長鎖アルキル四級アンモニウム	タンパク質
第二主成分（PC2）	長鎖アルキル四級アンモニウム	アルカリ金属、タンパク質
第三主成分（PC3）	長鎖アルキル四級アンモニウム シリコーンオイル	
第四主成分（PC4）	タンパク質（システイン／シスチンの比率が大）	タンパク質（ロイシン／イソロイシンの比率が大）

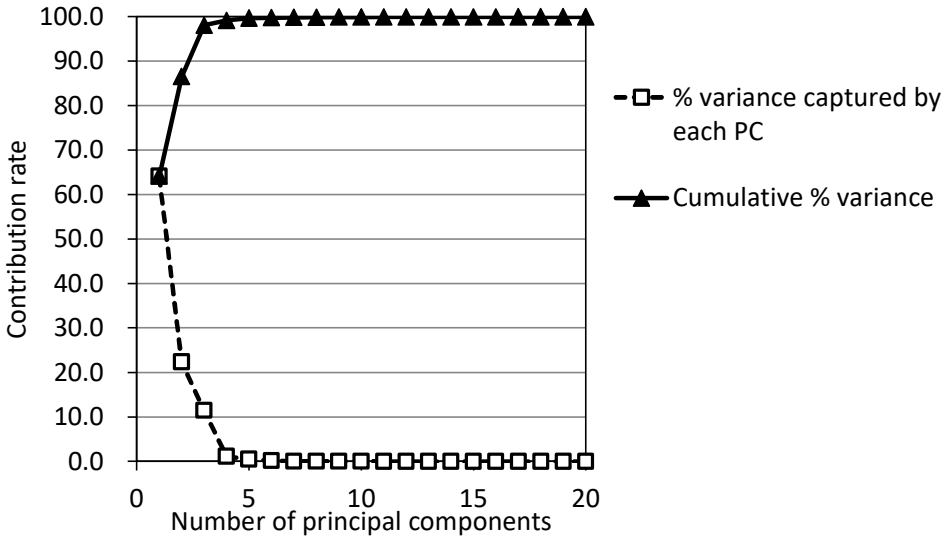


Figure 3.15　スクリープロット

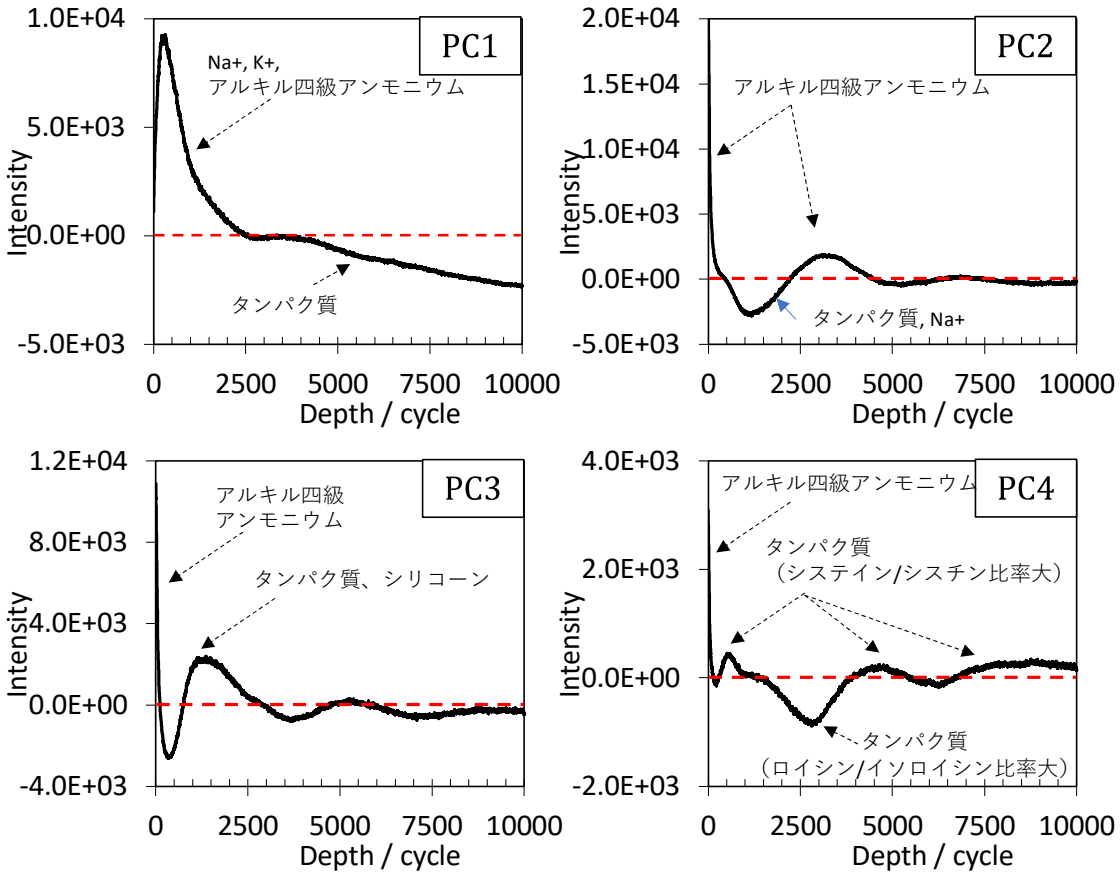


Figure 3.16　PCA により得られた主成分得点（スコア）のプロット（各グラフ中の成分の記載は Figure 3.15 のローディングプロットの結果を踏まえて記載した）

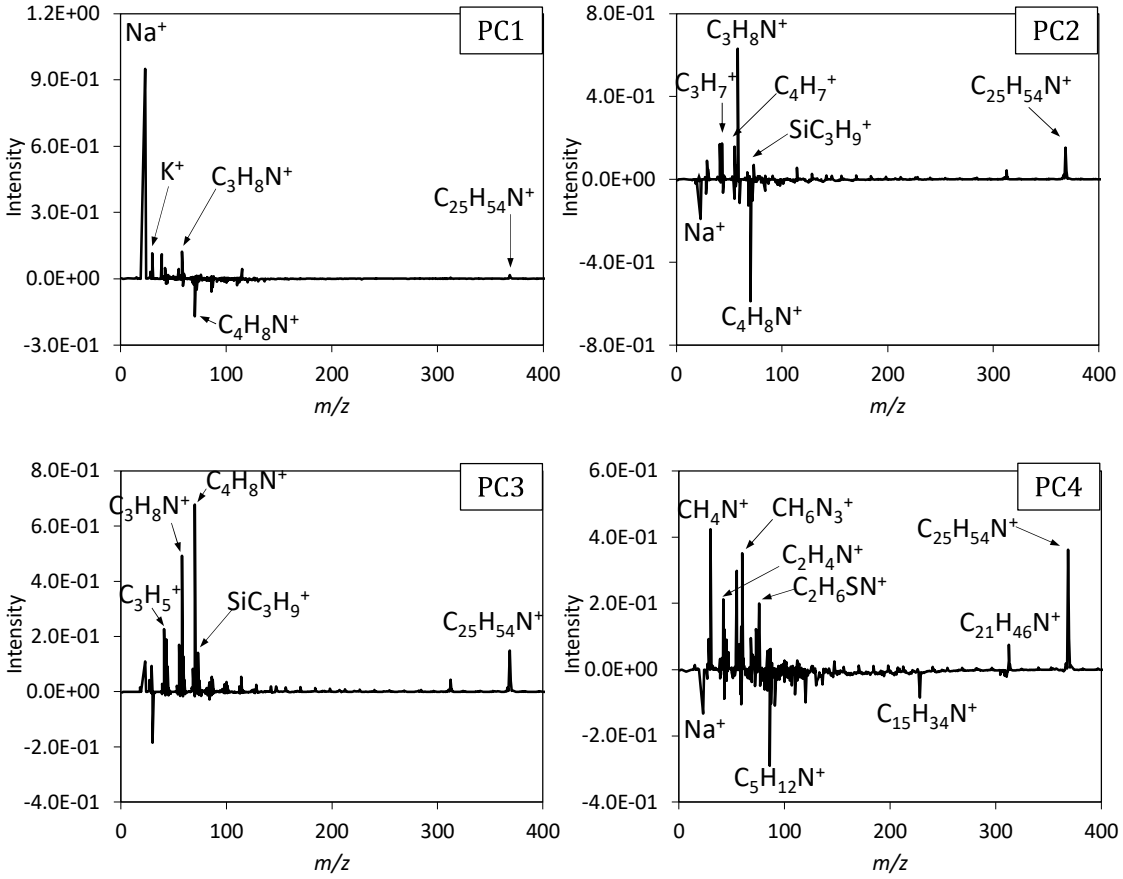


Figure 3.17　PCA により得られた各スコアに関連するローディングのプロット

3.4.7　抽出された特徴に対する中間層サイズの影響

3.4.2～3.4.5.の検討においては、オートエンコーダーの中間層のサイズ（ニューロンの数）は、既知の情報（キューティクルを構成する層の数や浸透させたヘアケア剤組成）を考慮して、10 に設定した。しかし、常に中間層のサイズ決定の参考となる情報が入手できているとは限らない。そこで中間層サイズを、想定される特徴（本研究の場合は、試料中で異なる空間分布を持った成分（群））の数より過剰に大きく設定した場合に、抽出される特徴にどのような影響が現れるか検証することを目的として、中間層サイズを 20 に設定してオートエンコーダーによる解析を行った（Figure 3.18）。なお、中間層サイズ以外の条件については、3.4.2～3.4.5.の検討と同一とした。

Figure 3.18 の結果を中間層サイズが 10 の結果（Figure 3.8）と比較したところ、Figure 3.8 で抽出された特徴に類似した特徴は、Figure 3.18 においても抽出されていることが確認された。したがってオートエンコーダーでは、抽出したい特徴の数が不明の場合であっても、中間層サイズを大きく設定しておくことで、漏れなく特徴を抽出できる可能性がある。

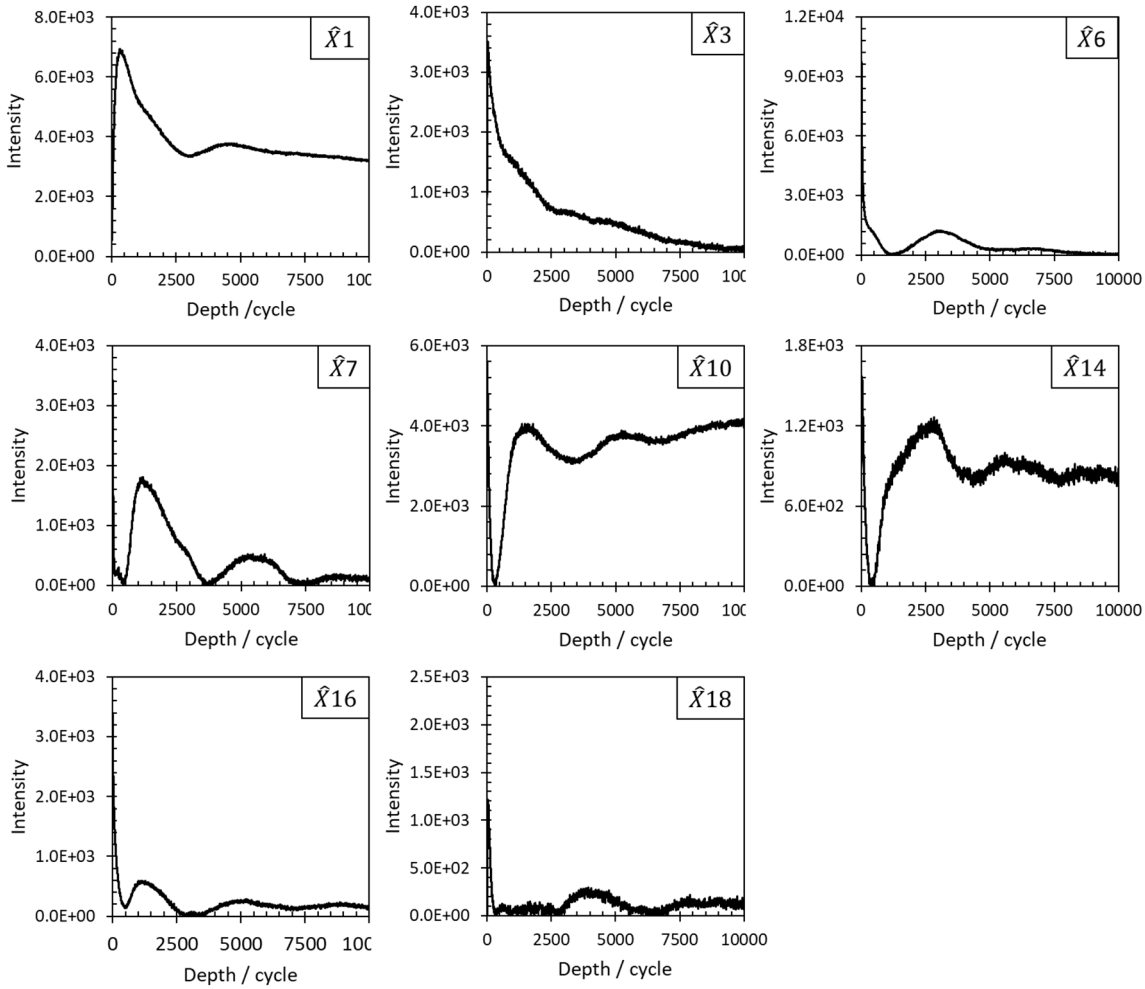


Figure 3.18　中間層（20 ニューロン）に抽出された特徴（デプスプロファイル）

＜$\hat{X}2, 4, 5, 8, 9, 11, 12, 13, 15, 17, 19, 20$については 0 が出力された。＞

Table 3.6　中間層サイズの

中間層サイズ：10（Figure 3.8）	中間層サイズ：20（Figure 3.17）
$\hat{X}1$	$\hat{X}16$
$\hat{X}4$	$\hat{X}6$
$\hat{X}5$	$\hat{X}16$
$\hat{X}6$	$\hat{X}7$
$\hat{X}7$	$\hat{X}1$
$\hat{X}8$	$\hat{X}14$
$\hat{X}9$	$\hat{X}10$
$(\hat{X}10)$※	$(\hat{X}18)$※

※S/N が悪く、明瞭な分布を反映していない特徴である。

3.5 結論

本章ではヒト毛髪のデプスプロファイルデータをモデルデータとして、オートエンコーダーで組成が複雑な生体試料の TOF-SIMS データから、試料中の空間（深さ・面内）分布が異なる成分（群）の分布状態を、特徴として抽出できるかどうかについて検証を行った。その結果、比較的単純なネットワーク構造のオートエンコーダーであっても、毛髪内部に浸透した界面活性剤（長鎖アルキル四級アンモニウム）や、特徴的なアミノ酸組成の分布を抽出することができることが確認された。オートエンコーダーを用いて得られた結果を、特徴抽出法として実績のある PCA を用いて得られた結果と比較したところ、PCA により得られる重要な特徴について、オートエンコーダーはすべて網羅できていることが確認された。特に、本研究ではオートエンコーダーの活性化関数に、正値のみを出力する関数（ReLU）を採用したことで、PCA よりも結果の解釈が容易な結果を得ることができた。したがって、生体試料の TOF-SIMS データ解析に対して、オートエンコーダーによる特徴抽出は有用であると考えられる。また、オートエンコーダーに特徴的なハイパーパラメーターである中間層のサイズについては、未知試料などで分析試料の組成情報が十分ではない場合については、中間層サイズを大きく設定することが、特徴を漏れなく抽出することに有効という、解析の指針を得ることができた。

第四章
TOF-SIMS 二次元イメージデータに対するスパースオートエンコーダーの適用検討

4.1 はじめに

4.1.1 スパースオートエンコーダー[63]

前章の検討において、オートエンコーダーによる特徴抽出は、生体組織（毛髪）の構造と照らし合わせて妥当な結果を抽出できる潜在的な能力を有していることが示された。また、PCA などの既存手法と比べ、解釈が容易な結果を出力できるという知見も得ることができた。しかしながら、TOF-SIMS データの解析では、よりデータ数の大きい二次元画像データの解析が一般に多く行われている。そこで、本章では二次元画像データに対して同様にオートエンコーダーによる特徴抽出を行い、その有用性の検証を行うと共に、特徴抽出性能を向上するための改良として、スパースオートエンコーダーの適用の検討を行った。また、ミニバッチ勾配降下法（p.26、第二章、「2.2.5 最適化関数」を参照）によるバッチサイズの違いが抽出された特徴に与える影響についても検証した。

スパースオートエンコーダーではより良い特徴を抽出するために、ネットワークにある種の制約を課すことによって特徴抽出層（中間層）をスパース（疎）にする。つまり、例えば中間層のニューロンのうち 50 %しか活性化しない場合、ネットワーク全体としてはこの 50 %のニューロンの組み合わせで入力を再現しようと学習し、その結果として有用な特徴が抽出されてくることが期待できる、という原理である。制約としては損失関数に正則化項と呼ばれるペナルティ項を加える方法が知られている。正則化項は次式で示す L1 ノルム（重みの絶対値の和）、L2 ノルム（重みの二乗和）が一般的に用いられており、特に L1 ノルムを使用した場合、「損失関数＋正則化項」の解は、損失関数の解に比べて原点に近づき、その結果、いくつかの重みの値が軸上（$w_i = 0$）に存在しやすくなる[64]。特定の中間層ニューロンに接続するすべての重みが 0 になると、そのニューロンは 0 を出力することから、中間層を疎にする効果が表れる。例としてパラメーターが二次元の場合を Figure 4.1 に示した。

$$L1\ ノルム \cdots \lambda \sum_{i=1} |w_i|$$

$$L2\ ノルム \cdots \lambda \sum_{i=1} |w_i|^2$$

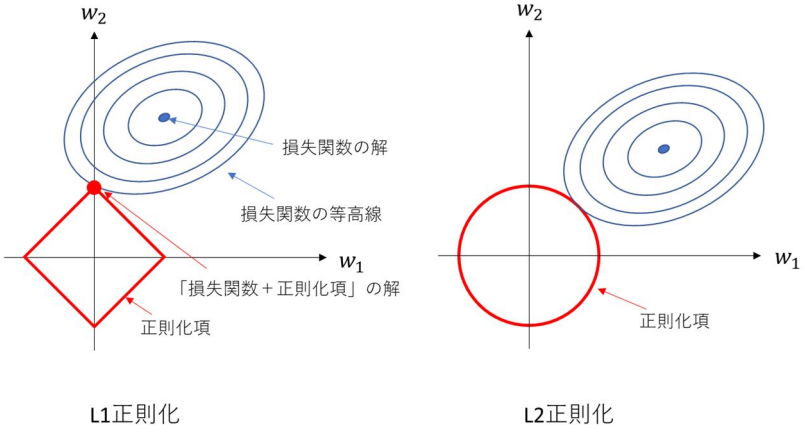


Figure 4.1　正則化の原理図

また、別の方法として中間層のスパース性の目標値と実際の差を訓練毎に求め、その差を損失関数に加える方法がある。差としては平均二乗誤差を用いるほかに、より強い勾配を持つ次式で表されるカルバック―リーブラーダイバージェンス（Kullback-Leibler divergence）がよく用いられている。

$$D_{KL}(p \parallel q) = p \log \frac{p}{q} + (1-p) \log \frac{1-p}{1-q}$$

ここで、pは中間層のニューロンの目標（平均）活性化率、qは実際の（平均）活性化率である。本研究のイメージデータでは各ニューロンにイメージのピクセル数個のデータがあるため、中間層の各ニューロンについて、そのニューロンの各ピクセルの出力値の平均値が平均活性率に当たる（次式）。目標活性化率 pの値を低い値に設定することで中間層で活性化するニューロンの数を減少するため、限られたニューロンに有用な特徴が抽出されやすくなる。

$$q = \frac{1}{n} \sum_{j=1}^{n} f\left(w_i x_j + b_i\right)$$

本章の検討では中間層を疎にするために、L1 ノルムおよび KL-divergence を正則化項（ペナルティ項:$\Omega_{sparsity}$）として損失関数に導入した。次式の第一項が、出力層と入力層の出力の平均二乗誤差であり、第二項が正則化項である。ペナルティの程度はハイパーパラメーター（λ）により調整した。

$$Loss = \frac{1}{N} \sum_{n=1}^{N} \sum_{k=1}^{K} \|\bar{x}_{kn} - x_{kn}\|^2 + \lambda \cdot \Omega_{sparsity}$$

4.1.2 ヒト皮膚の構造と機能

解析用の画像データとしてはヒト皮膚表面に存在する角質層について測定したものを使用した。ヒトの皮膚は表皮、真皮、皮下組織の三つの部位から成り、さらに表皮は基底層、有棘層、顆粒層、角質層に分かれる（Figure 4.2）。表皮は絶えず紫外線や外力にさらされることから、頻繁に再生を繰り返すことにより、その機能や構造を維持している。具体的には、基底層に存在する幹細胞から発生した角化細胞（ケラチノサイト）は、皮膚の表面側に移動した後に細胞核を失い角質層を形成する[89]。角質層は核を消失した死細胞が 10～20 層程度積み重なり、その間を細胞間脂質が埋めた構造をしている。この角質層は外部から有害物質が侵入するのを防ぐバリア機能と、水分の過剰な蒸散を阻止する保湿機能を司っている。

皮膚組織を透過して薬物を投与する方法として経皮吸収製剤が注目されている。経費吸収製剤は、肝臓での代謝を受けずに薬物を血中に簡便に投与できる利点がある。角質層のバリア機能を如何にして突破するかということは、製剤開発を行う上で重要であり、そのための検証として、TOF-SIMS を用いた皮膚組織

Figure 4.2　皮膚の模式図

内の薬物分布の可視化事例が多く報告されている[77-81]。薬物を浸透させた皮膚組織（角質層）の TOF-SIMS イメージデータは、薬物や細胞間脂質に注目することで、特徴抽出性能を評価するのに有用な試料であると考えられる。そのため、スパースオートエンコーダーの解析用モデルデータとして採用した。

4.2 実験方法

4.2.1 分析試料調整

年齢 30 代の成人男性の前腕部内側を市販のハンドソープを用いて洗浄後、水道水でハンドソープ成分をよく洗い流した。自然乾燥を行った後、ジクロフェナクナトリウム（$C_{14}H_{10}Cl_2NO_2Na$）の 10 wt%イソプロパノール溶液（Voltaren AC lotion, GlaxoSmithKline plc, UK）を約 2 cm ×2 cm の範囲に約 10 µL 塗布した。30 分間後にポリエステル基材の粘着テープ（No.31 series, Nitto Denko Corporation, Osaka, Japan）の粘着面を上記溶液塗布部に密着させ、直後に剥がすことで皮膚表面の角質層を採取した（テープストリップ法：Figure 4.3 参照）。同じ部位に対して同様の作業を繰り返すことで段階的に角質層を採取することが可能であり、本検討では 3 回目に採取した角質層を TOF-SIMS 測定に用いた。

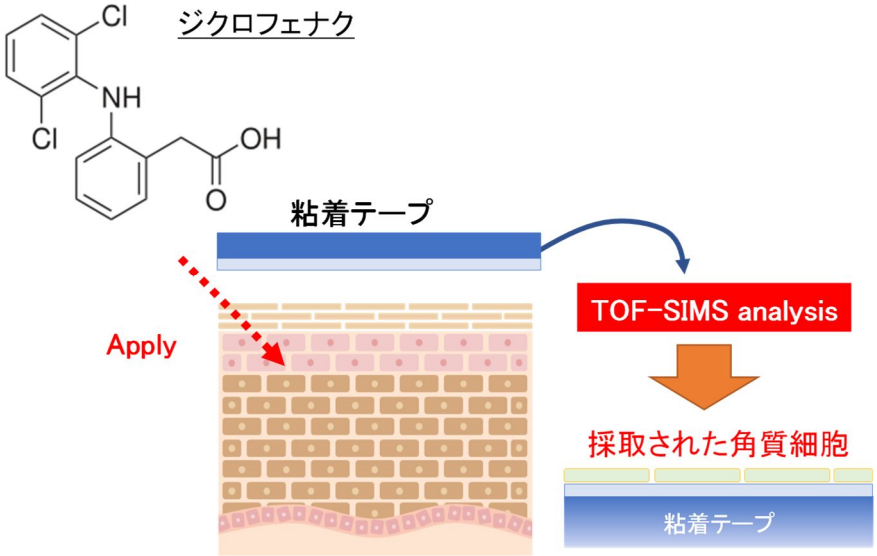


Figure 4.3 テープストリップ法による角質細胞の採取

4.2.2 TOF-SIMS 測定条件

粘着テープ上に採取したヒト皮膚角質の測定には、TOF.SIMS 5（IONTOF GmbH, Münster, Germany）を用いた。一次イオンとしてパルス化された Bi_3^{2+}ビームを用い、二次イオン極性は角質層に多く含まれる脂質を感度良く観測可能な負 2 次イオン（Negative ion）とした。測定モードには空間分解能を重視した Burst alignment mode を用いた（※質量分解能は低下する）。その他の、測定条件については Table 4.1 に示した。なお、質量スペクトルのピークの帰属の参考とするため、高質量分解能（※空間分解能は低下する）条件による 500 µm×500 µm の測定を別途行った。

Table 4.1　TOF-SIMS の主な測定パラメーター

	Primary ion
Ion specie	Bi_3^{2+}
Energy	50 keV
Current	Approx. 0.1 pA
Field of view	100 μm × 100 μm
Pixel number	256 × 256 （= 65536）
Dose density	approx. 1.3×10^{12} ions/cm^2
Charge compensation	ON（Low-energy electron flooding）
Scan number	64 scans

4.3 データ解析

4.3.1 データ前処理

TOF-SIMS 装置に付属する制御・解析ソフトウェアである、SurfaceLab 6.5（IONTOF GmbH, Münster, Germany）を用いて、質量のキャリブレーションを実施した（$^{13}CH^-$, $^{25}C_2H^-$, $^{49}C_4H^-$の 3 つの既知のピークを用いた）。Burst alignment mode の使用による低い質量分解能のデータであることを考慮して、質量軸を *m/z* 1 単位にビニングした。結果として得られた 468 peaks×65536 pixels のデータをスパースオートエンコーダーの解析に使用した。なおデータ前処理は実施せず、強度データをそのまま入力データとした。

4.3.2 データ解析条件

ライブラリおよびハードウェアについては第三章の検討と同一である（Table 3.2 参照）。

オートエンコーダーの具体的な構造としては第二章と同様に、Figure 3.10 に示したエンコーダーとデコーダーの 2 つの部分から成るシンプルなネットワーク構造を採用した。中間層のサイズは、前章の検討結果を参考として、20 に設定した。468 次元の入力データを 20 次元に圧縮し、出力層にて再度 468 次元に再構成した。エンコーダーとデコーダーの活性化関数と最適化関数には、第三章と同様に、それぞれ ReLU と Adam をを採用した。損失関数には、「4.1.1 スパースオートエンコーダー」で述べたように、平均二乗誤差（MSE）に L1 ノルムおよび KL-divergence を正則化項（ペナルティ項）として導入したものを使用した。バッチサイズは 64 を基準として 16～65536 の間で値を振り、解析結果に与える影響について検証した。なお KERAS ではニューラルネットワークの重みとバイアスの初期値を乱数で与えることができるが、本検討ではハイパーパラメーターの特徴抽出性能への影響を正確に評価するために、初期値として常に同じ値（固定値）を用いた。

4.4 実験・解析結果と考察

4.4.1 TOF-SIMS 測定結果

TOF-SIMS 測定によって得られた、ヒト角質層の総二次イオン質量スペクトルを Figure 4.4 に示した。質量スペクトル上にて観測された主なピークに対して帰属した結果を Table 4.2 に示した。なお、ジクロフェナクナトリウムと粘着テープの粘着剤については、別途行ったジクロフェナクナトリウム溶液の蒸発乾固物、粘着剤表面の測定結果からそれぞれに特徴的なピークを把握した。

Table 4.2　角質層の TOF-SIMS 質量スペクトルのピーク帰属

試料	特徴的な負 2 次イオンピーク
ジクロフェナクナトリウム （浸透薬剤）	$^{35}Cl^-$, $^{58}C_2H_2O_2^-$, $^{93}NaCl_2^-$, $^{214}C_{13}H_9ClN^-$, $^{250}C_{13}H_{10}Cl_2N^-$, $^{281}NaC_{14}H_9ClNO_2^-$, $^{316}NaC_{14}H_9Cl_2NO_2^-$, $^{352}NaC_{14}H_{10}Cl_3NO_2^-$
ポリブチルアクリレート（粘着剤）	$^{41}C_2HO^-$, $^{55}C_3H_3O^-$, $^{71}C_3H_3O_2^-$, $^{81}C_5H_5O^-$, $^{115}C_6H_{11}O_2^-$
硫酸コレステロール	$^{80}SO_3^-$, $^{97}SO_4H^-$, $^{465}C_{27}H_{45}SO_4^-$
脂肪酸（またはエステル）	$^{255}C_{16}H_{31}O_2^-$, $^{281}C_{18}H_{33}O_2^-$, $^{283}C_{18}H_{35}O_2^-$, $^{339}C_{22}H_{43}O_2^-$, $^{367}C_{24}H_{47}O_2^-$, $^{381}C_{25}H_{49}O_2^-$, $^{395}C_{26}H_{51}O_2^-$, $^{423}C_{28}H_{55}O_2^-$
アミド結合を含む成分 （主にタンパク質）	$^{42}CNO^-$
ポリオキシエチレンアルキル硫酸 エステル（界面活性剤）	$^{265}C_{12}H_{25}SO_4^-$, $^{293}C_{14}H_{29}SO_4^-$, $^{309}C_{14}H_{29}SO_5^-$, $^{337}C_{16}H_{33}SO_5^-$

注 1）ジクロフェナクナトリウムと粘着剤は標準試料の測定を基にピークを帰属した。

注 2）高質量分解能測定によるデータを参考に二次イオンピークの帰属を行った。

質量スペクトルのピーク帰属結果（Table 4.2）より、

①塗布したジクロフェナクナトリウムが角質層内に浸透していること
②測定面内の一部にテープストリップに使用した粘着テープの粘着剤が露出していること
③細胞間脂質（硫酸コレステロールや脂肪酸）が検出されていること

が確認された。Figure 4.5 にジクロフェナクナトリウムと粘着剤、アミド（タンパク質）に特徴的な二次イオンの面内分布（二次イオン像）を示した。アミドの二次イオン像で観測される多角形の形状が角質細胞の形状に対応しており、ジクロフェナクナトリウムは特定の細胞の位置で高強度に観測されている（局在している）ことが分かった。そこで、まずはジクロフェナクナトリウムやアミド、粘着剤の分布が、どのように特徴として抽出できるのかに着目して、オートエンコーダーおよびスパースオートエンコーダーの性能を検証することとした。

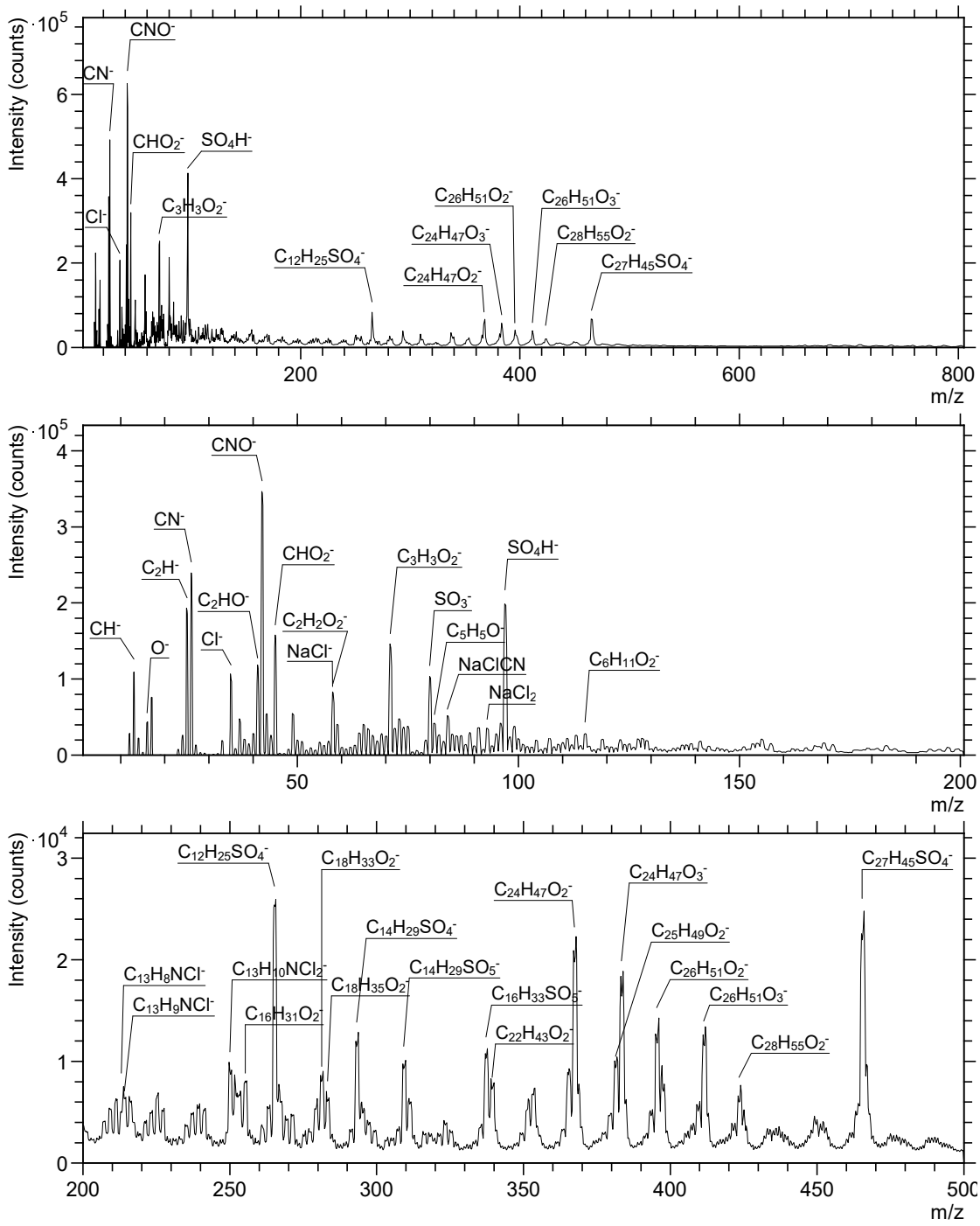


Figure 4.4 角質層表面の **TOF-SIMS** 質量スペクトル（負二次イオン）

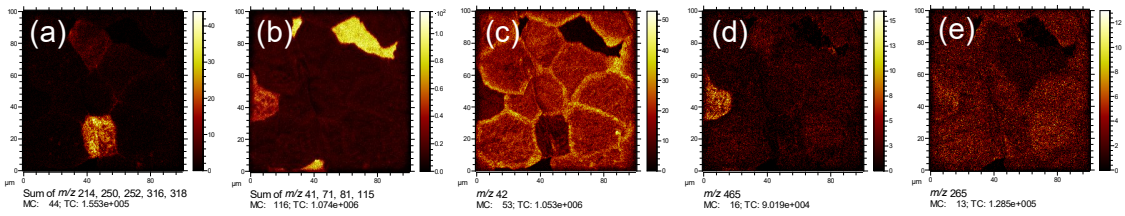


Figure 4.5 主な負二次イオン像((a):ジクロフェナクナトリウム、(b):粘着剤、(c)アミド(タンパク質)、(d)硫酸コレステロール、(e)ポリオキシエチレンアルキル硫酸エステル

4.4.2 オートエンコーダー(正則化なし)による解析

スパースオートエンコーダーの性能を評価するうえで、比較対象となる正則化なしのオートエンコーダーにより、二次元画像データ(468 peaks×65536 pixels)からどのような特徴が抽出されるかについて確認を行った。Figure 4.6 に示した損失関数の推移より、20 epoch で損失関数の値は一定値となりそれ以上の低下は認められなかったことから、学習は十分に進行したものと判断した。

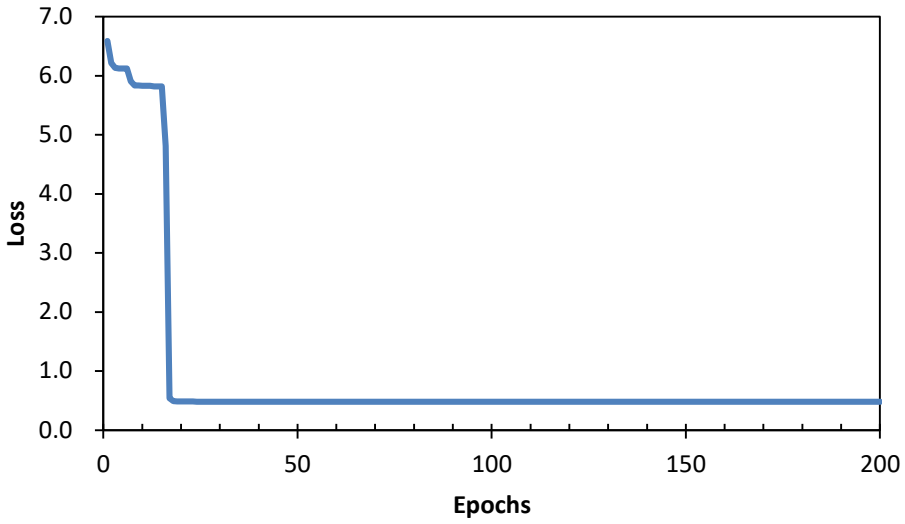


Figure 4.6 訓練回数(Epoch)に対する損失関数の値の推移

実際に学習結果を確認するために、*m/z* 12～111 までの 100 個の二次元画像データについて、入力データとデコーダーの出力(再構成)データをそれぞれ Figure 4.7、4.8 に示した。両者の比較から、入力データで強度が弱く分布が不明瞭なデータ(青枠で示した画像)については、デコーダーでの再現ができていない。また、入力データにて明瞭な分布が観測されたデータ(緑枠で示した画像)についても、一部で再現できていないものの存在が認められた。

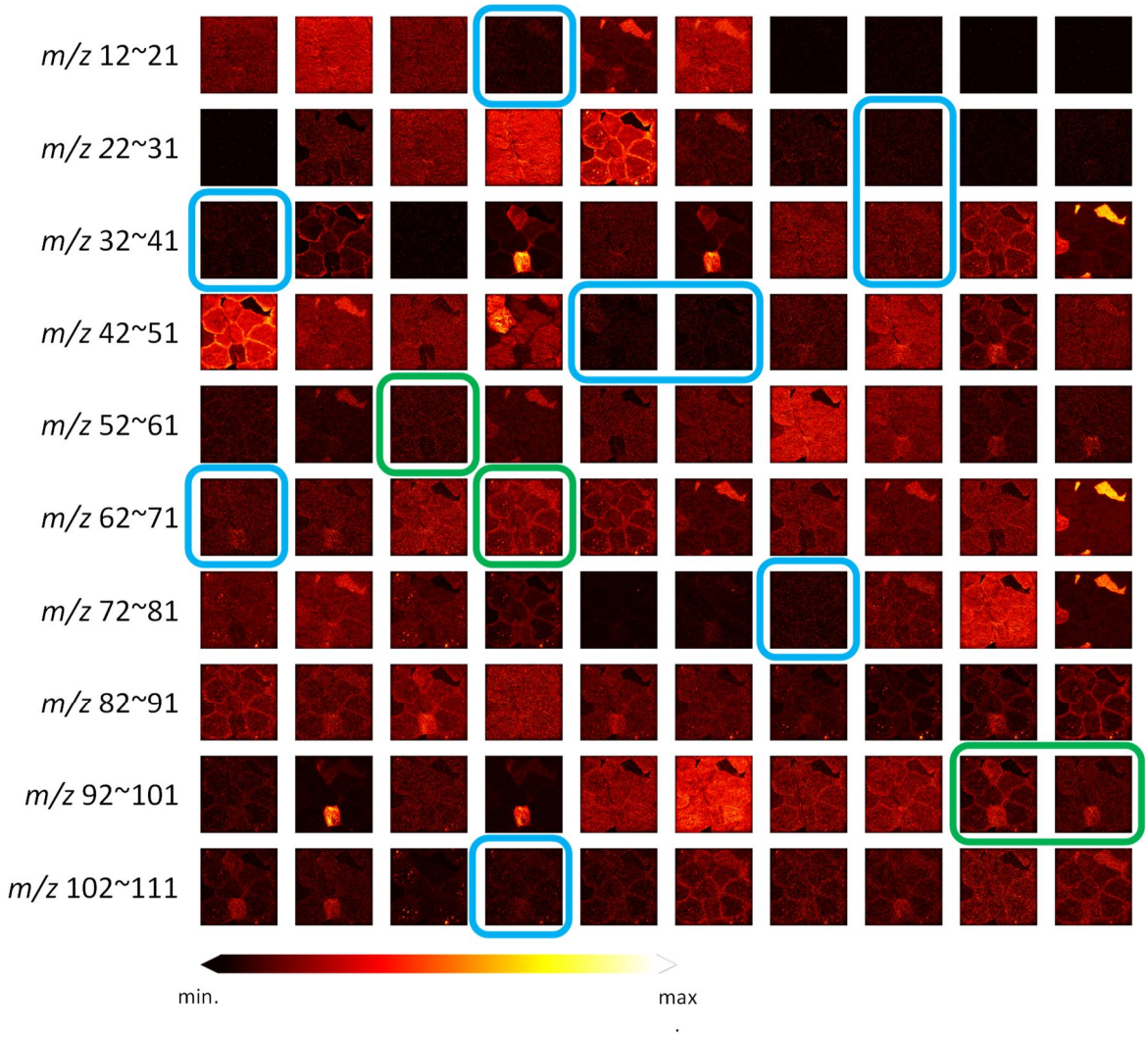


Figure 4.7　100 個の入力データ（オリジナルの正二次イオン像、*m/z* 12 ~ 111）

青枠：強度が弱く出力データにて再現できなかった二次イオン

緑枠：強度が比較的強いが再現できなかった二次イオン

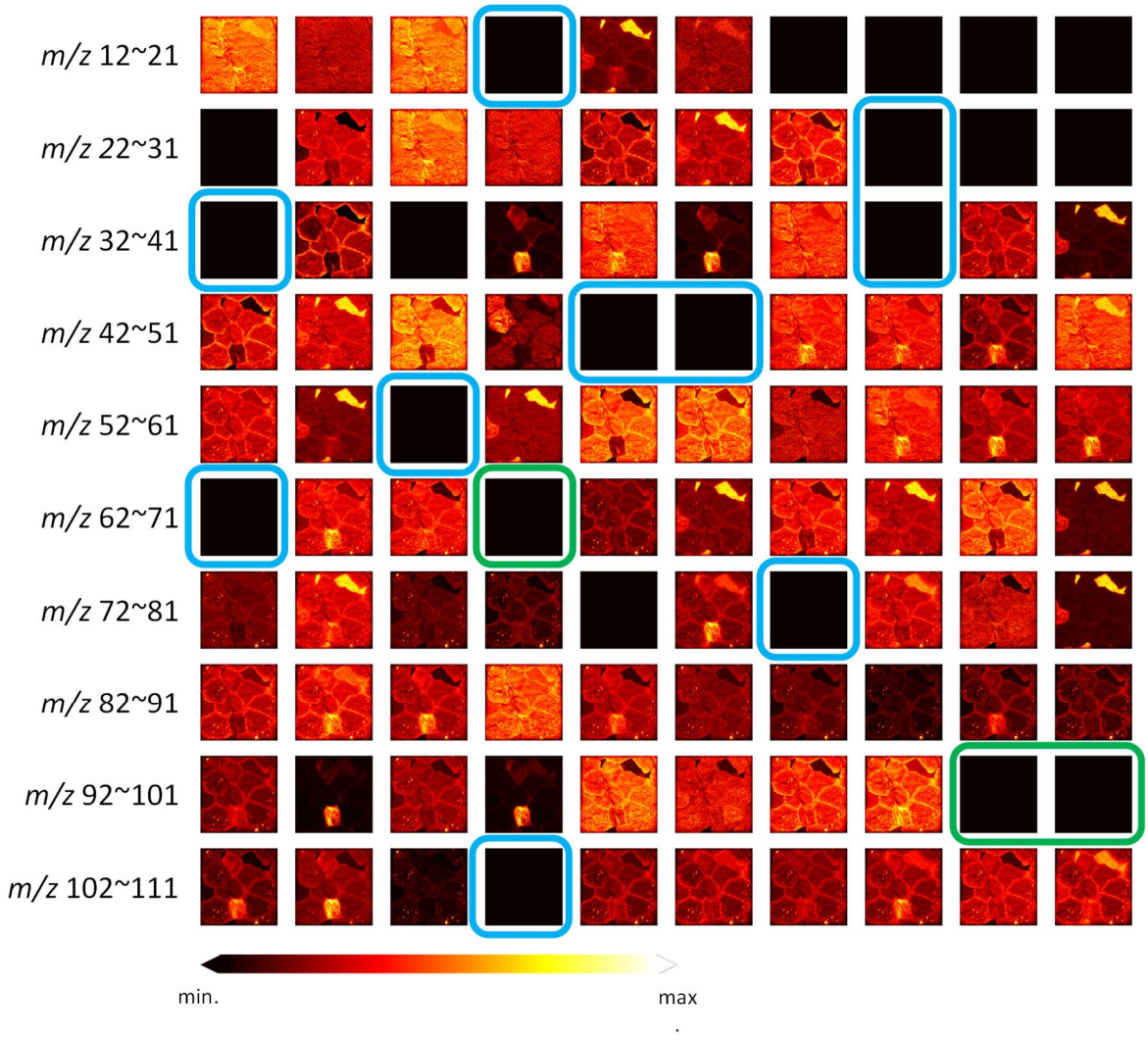


Figure 4.8　100 個の出力データ（デコーダーによる再構成画像、*m/z* 12 ~ 111）

青枠：入力データで強度が弱く出力データにて再現できなかった二次イオン

緑枠：入力データで強度が比較的強いが再現できなかった二次イオン

オートエンコーダー（正則化なし）で抽出された特徴を Figure 4.9 に示した。20 個の中間層ニューロンのうち、5 個では 0 が出力された。これは「3.4.3 抽出された特徴の解釈」にて述べた Dying ReLU の影響であると考えられる。Figure 4.5 に示した負 2 次イオン像との比較より、$\hat{X}2$, $\hat{X}7$, $\hat{X}8$がそれぞれアミド（タンパク質）、ジクロフェナクナトリウム、粘着剤の分布を抽出した特徴であると判断した。実際に$\hat{X}7$と $\hat{X}8$に対応するデコーダーの重み（$W_{dec}7, W_{dec}8$）を確認すると、Table 4.1 に示したジクロフェナクナトリウムと粘着剤（ポリブチルアクリレート）に特徴的なピークが確認された（Figure 4.10）。$\hat{X}8$には粘着剤のほかに硫酸コレステロール（$^{465}C_{27}H_{45}SO_4^-$）に特徴的なピークも観測されたが、これは粘着剤の分布と硫酸コレステロールの分布が一部で一致しているためと考えられる（Figure 4.5 (b), (d)参照）。

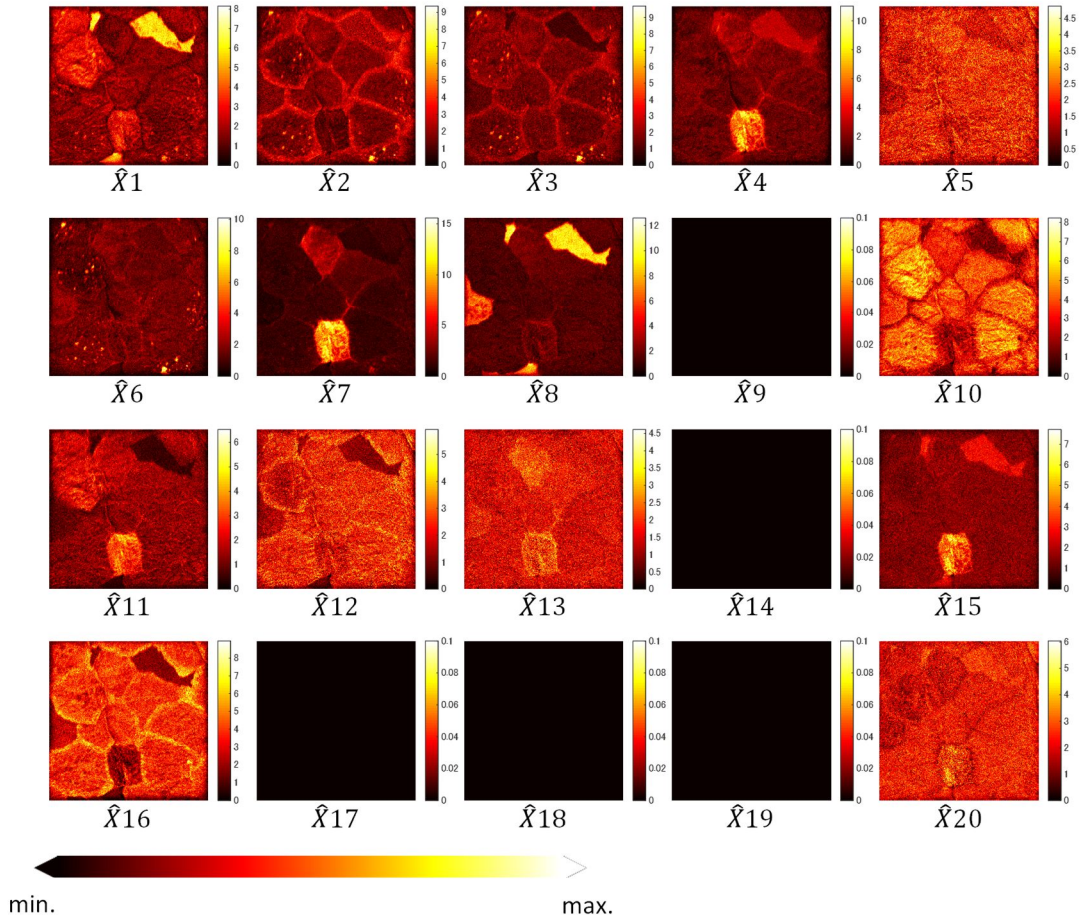


Figure 4.9　中間層（20 ニューロン）からの出力画像（エンコーダーにより圧縮された特徴画像）

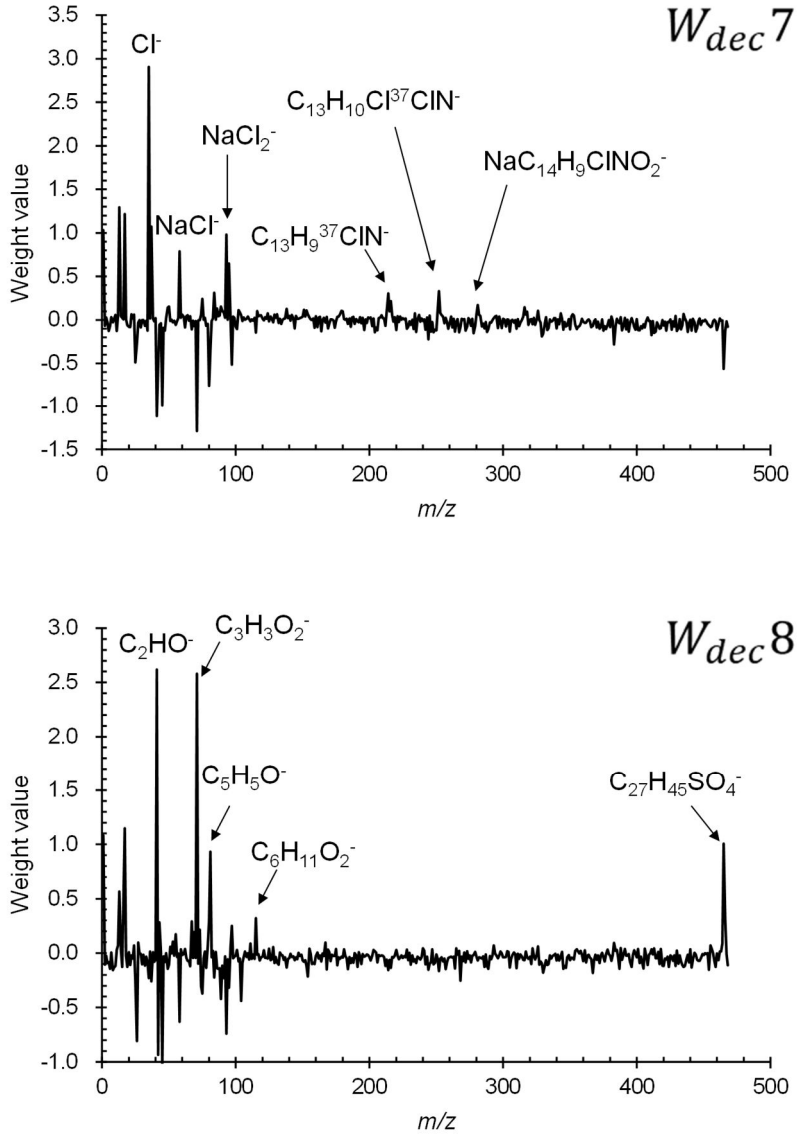


Figure 4.10　代表的な特徴画像に関連するデコーダーの重み（$W_{dec}7$: 主にジクロフェナクの分布、$W_{dec}8$: 主に粘着剤） ＜活性化関数:ReLU によって中間層には正値のみが出力されるため、負の重みについては無視できる＞

4.4.3 スパースオートエンコーダー（L1 正則化）による解析

4.4.2 の結果から、二次元画像データからもオートエンコーダーである程度の特徴が抽出できていることが確認された。本項では更に L1 正則化項を損失関数に加えたスパースオートエンコーダーにより、特徴抽出性能がどのように変わるかを調べた。

正則化の程度を調整するハイパーパラメーター（λ, p.64 参照）を 0, 1×10^{-5}, 1×10^{-4}, 1×10^{-3} と変化させた際の、損失関数の推移を Figure 4.11 に示した。どの結果も 200 epoch までで損失関数の値は一定値となったが、最終的に到達した値はλが大きい場合ほど大きくなった。これは損失関数に正則化項の値分が上乗せされたためと理解できる。

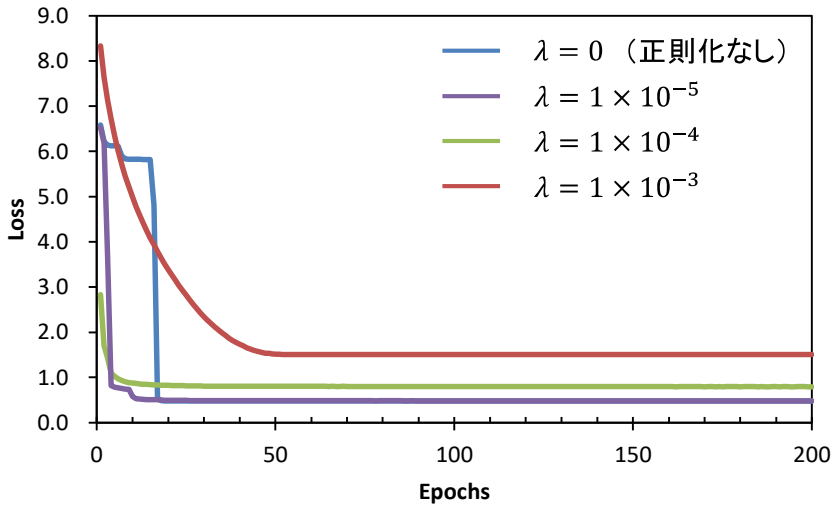


Figure 4.11　スパースオートエンコーダー（L1 正則化）の損失関数の推移

次に、抽出された特徴を示す中間層ニューロン 20 個の出力結果を Figure 4.12 に示した。正則化なし（λ =0）と比べると、λの値が増加するにつれて 0 を出力するニューロンの数が増加（5→6→14→20）しており、スパースオートエンコーダーによって中間層がスパースになる効果が確認された。ニューロン全てが 0 を出力したλ =1×10^{-3}を除いた 3 条件について、詳細な特徴の比較を行った。これらの条件で抽出された特徴からは、いずれからもジクロフェナクナトリウムの分布を示すニューロンと、粘着剤の分布を示すニューロン、顆粒状の分布を示すニューロンが認められた（Figure 4.12 中にそれぞれ赤枠、青枠、緑枠で示した）。ジクロフェナクナトリウムの分布を示す特徴を負二次イオン像（Figure 4.5 (a)）と比較すると、正則化なし（λ =0）では負二次イオン像で強度が弱い領域からも強度が観測されているが、λ =1×10^{-5}, λ =1×10^{-4} では負二次イオン像と比較的近い分布である。さらに、デコーダー重み（Figure 4.13）の比較からは、正則化なし（λ =0）に比べでジクロフェナクナトリウムに特徴的なピークの S/N が良く、また高質量側のピーク（$^{352}NaC_{14}H_{10}Cl_3NO_2^-$）も現れた。これらの結果から、$\lambda$ =1×10^{-5}, λ =1×10^{-4} の結果ではジクロフェナクナトリウムの分布をより単成分の分布として抽出できているものと考えられる。

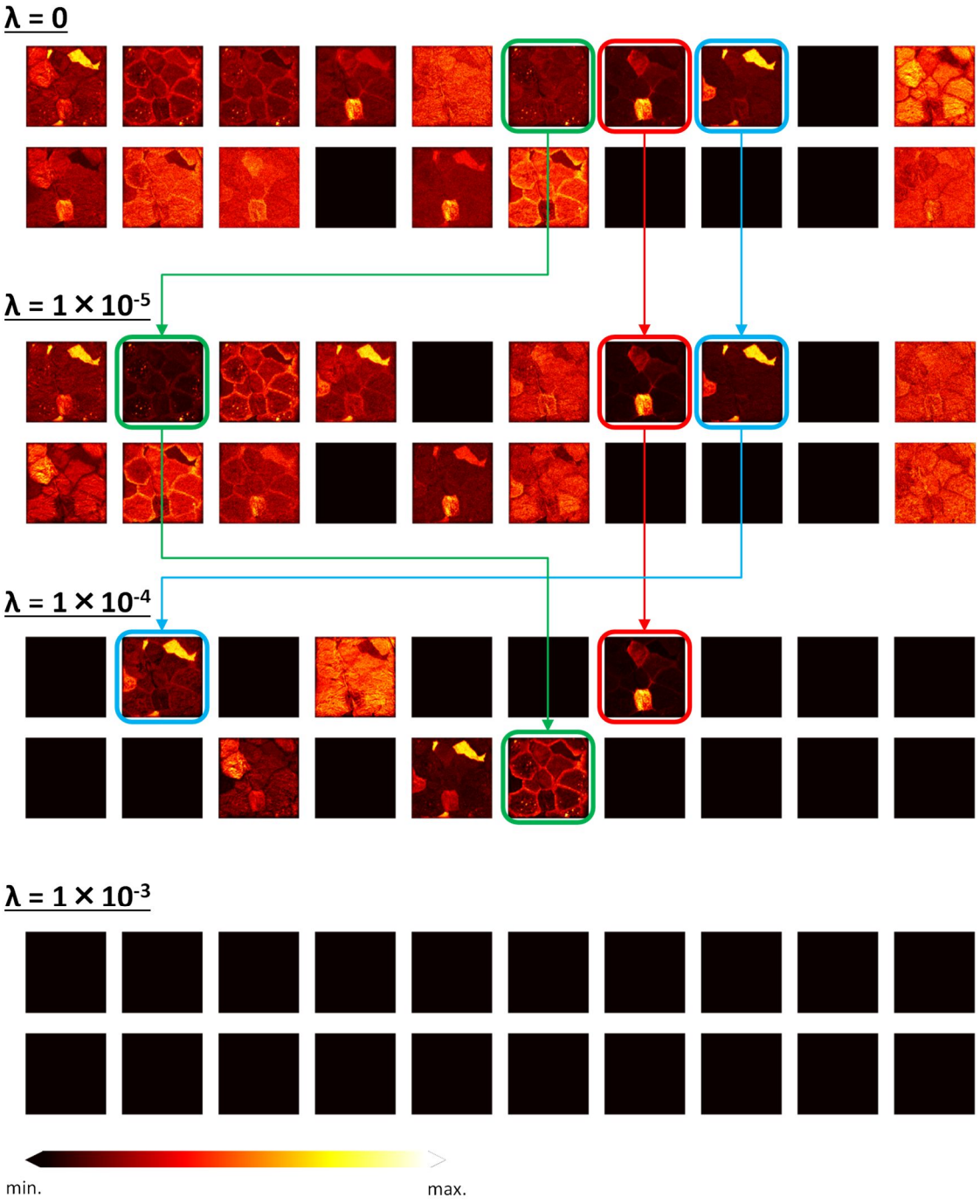


Figure 4.12　抽出された特徴に与える L1 正則化の程度の影響（赤枠：主にジクロフェナクナトリウムの分布を反映した特徴、青枠：主に粘着剤の分布を反映した特徴、緑枠：顆粒上の分布が確認される特徴）

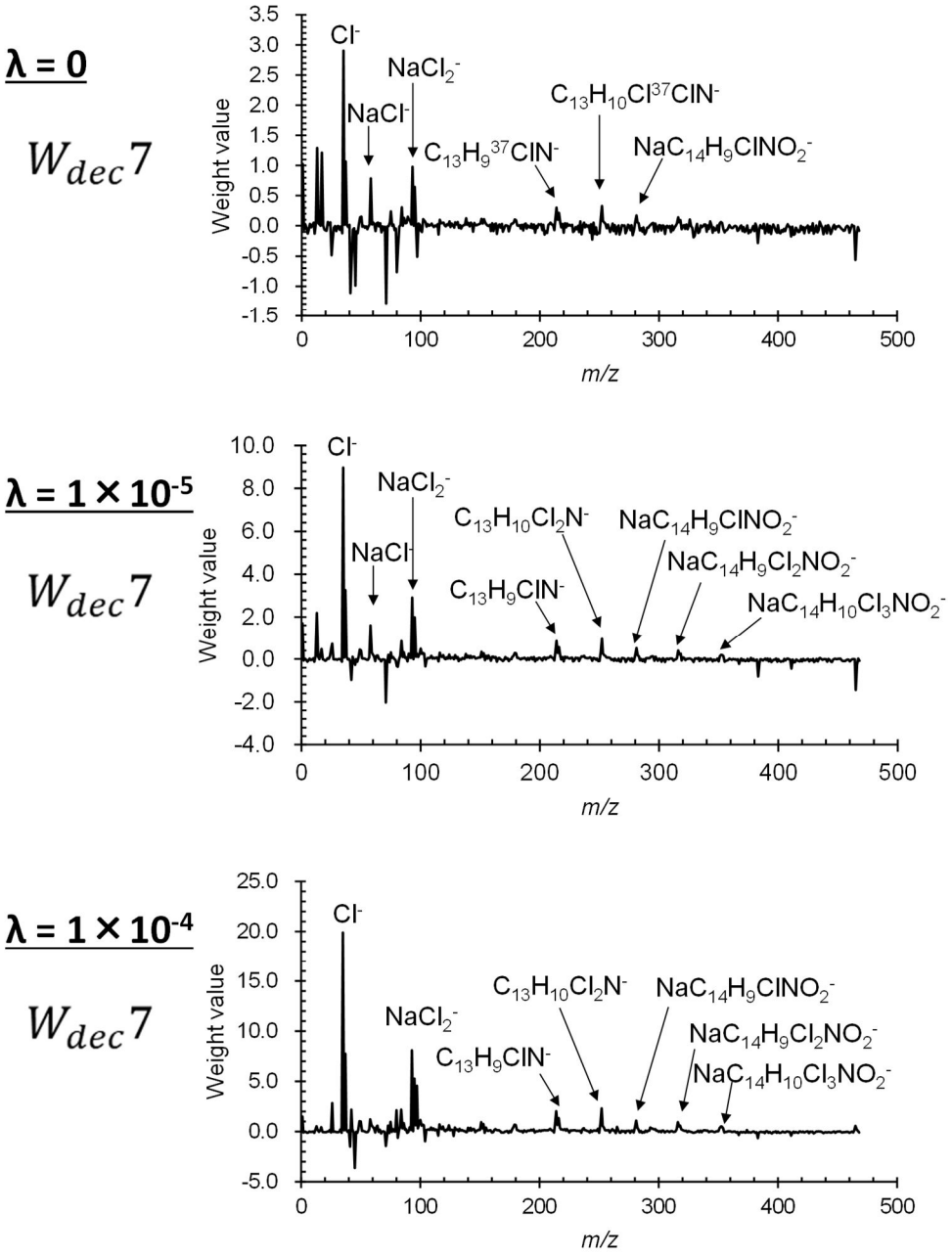


Figure 4.13 **主にジクロフェナクナトリウムの分布を表す特徴に対応するデコーダー重み（L1 正則化の程度による変化）** ＜活性化関数:ReLU によって中間層には正値のみが出力されるため、負の重みについては無視できる＞

一方で、顆粒状の分布（Figure 4.12；緑枠）は$\lambda = 0$ に比べて$\lambda = 1 \times 10^{-5}$ では、より強調された分布で抽出されたが、$\lambda = 1 \times 10^{-4}$ ではアミド（主にタンパク質）に特徴的な多角形の形状を示す分布が支配的な特徴と一緒に抽出された。同様に粘着剤の分布（Figure 4.12；青枠）についても、$\lambda = 1 \times 10^{-5}$ においては負二次イオン像（Figure 4.5 (b)）と一致する分布であるが、$\lambda = 1 \times 10^{-4}$ ではアミド（主にタンパク質）に特徴的な多角形の形状が混在した分布である。このように抽出された特徴が変化した理由としては、$\lambda = 1 \times 10^{-4}$ では中間層がスパースになりすぎたため、各成分を単一の特徴として抽出するには中間層のニューロンの数が不足し、本来、分布が異なる複数の成分を一つのニューロンに抽出したためと推察される。したがって、特徴抽出性能の向上（特徴的な分布を持つ成分を、単独の特徴として抽出する）には正則化の程度（λの値の大きさ）をどの程度に設定するかが重要であると言えるが、最適値は元データや中間層のサイズなどに依るところが大きい。そのため、一律に最適値を設定することは困難と考えられ、必要に応じて複数のλ値で解析を行い、結果を比較することが望ましいと考えられる。

ところで、中間層をスパースにするために正則化項を損失関数に付与することは、正則化項を加えずに単純に中間層のサイズを小さくすることととどのように違うのだろうか。この点について調べるために、正則化なし（$\lambda = 0$）の条件で中間層のニューロン数を 20、15、10 とし、抽出された特徴を比較した。Figure 4.14 に示した中間層ニューロンの出力では、どの条件からもジクロフェナクナトリウムの分布に対応した特徴（赤枠）が抽出された。これらの特徴は、中間層ニューロンの数が減少しても大きな分布の変化は認められず、スパースオートエンコーダー（$\lambda = 1 \times 10^{-5}$）のように、負二次イオン像に類似してくる、すなわちよりジクロフェナクナトリウム単成分の分布を反映した特徴として抽出される傾向は認められなかった。デコーダー重み（Figure 4.15）の比較からも、各条件で顕著な変化は認められず、スパースオートエンコーダーのようにジクロフェナクナトリウムに特徴的なピークがより明瞭に（S/N が良く）観測されることはなかった。学習終了時の損失関数の値（Table 4.3）を確認すると、中間層のサイズが 10 の場合は損失関数が学習によって減少せず、200 epoch の学習後でも他の条件に比べて顕著に大きな値を示すことがわかった。これは中間層サイズが小さくなることによって、中間層での表現力が低下したため、入力層の再現に必要な情報を出力層に伝達できなくなったためと推察される。

Table 4.3　中間層サイズと損失関数の対応

中間層のサイズ（ニューロンの数）	1 epoch 目の損失関数の値	学習後の損失関数の値
20	6.56	0.67
15	2.33	0.55
10	7.13	6.15

以上の結果から、スパースオートエンコーダーでは、中間層のサイズを十分に大きく保ち、ニューラルネットワークの表現力を確保しながら、学習の過程で中間層をスパースにすることが、単純なオートエンコーダーに比べ、特徴抽出性能が向上する要因であると推察される。

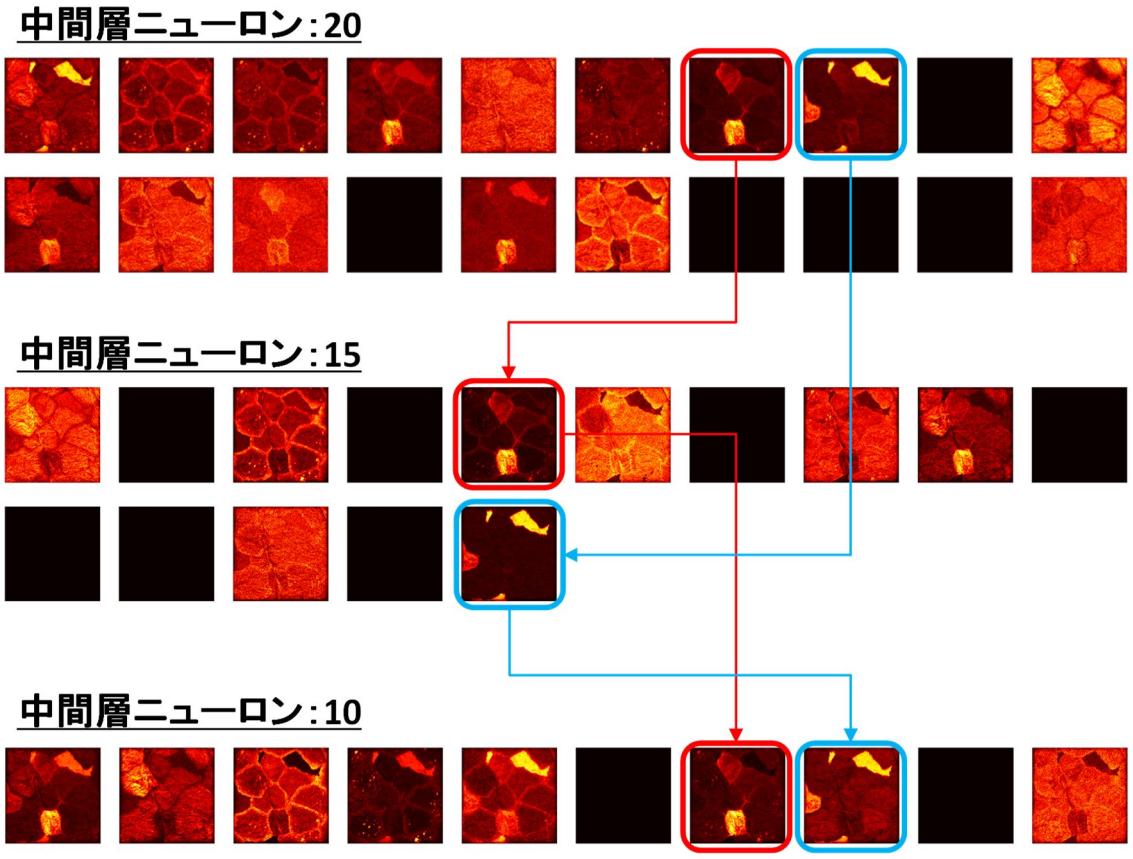
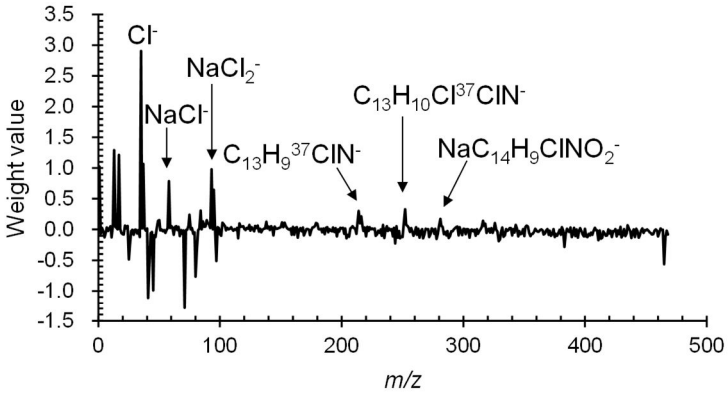


Figure 4.14　オートエンコーダー（正則化なし）の中間層サイズを変更した際の抽出された特徴の変化

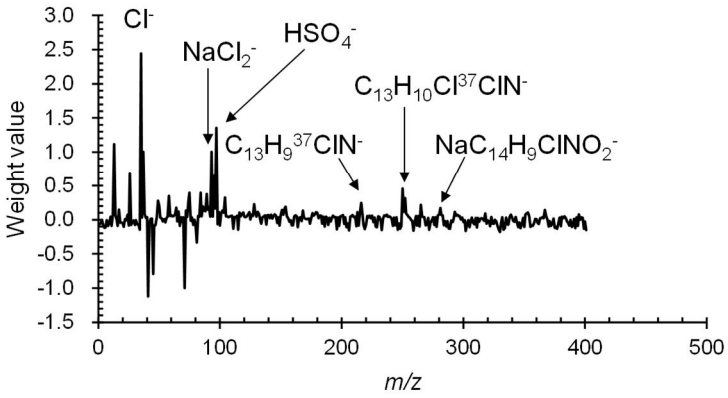
中間層ニューロン：20

$W_{dec}7$



中間層ニューロン：15

$W_{dec}5$



中間層ニューロン：10

$W_{dec}7$

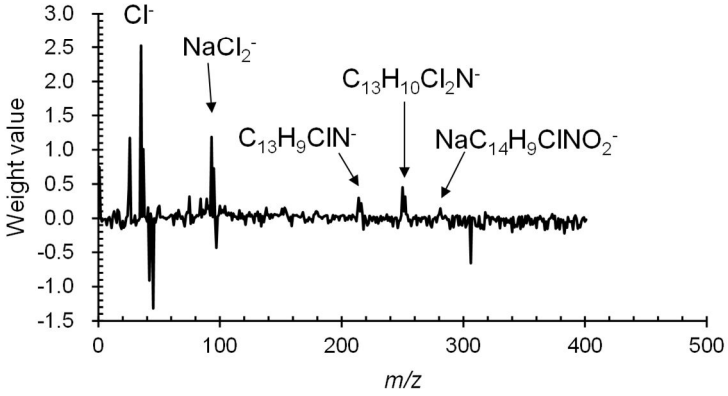


Figure 4.15　ジクロフェナクナトリウムの分布に対応した特徴のデコーダー重み（中間層サイズによる変化）＜活性化関数：ReLU によって中間層には正値のみが出力されるため、負の重みについては無視できる＞

4.4.4 スパースオートエンコーダー（KL-divergence 正則化）による解析

中間層のスパース度合いを評価する指標としての KL-Divergence を求めるには、スパース度合いの目標値（Target sparsity）を設定する必要があるが、本検討では文献[63]の設定値を参考として 0.1 に設定した。そのうえで、正則化の程度を調整するハイパーパラメーター（λ）を 0.05, 0.1, 0.5 と変化させ、特徴抽出性能がどのように変わるかを調べた。

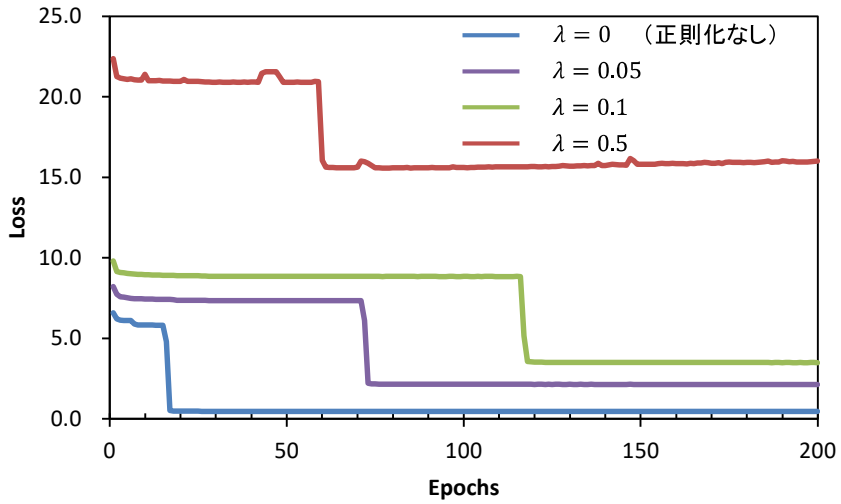


Figure 4.16　スパースオートエンコーダー（KL-Divergence 正則化）の損失関数の推移

損失関数の推移を Figure 4.16 に示した。L1 正則化の場合と同様に、どの結果も 200 epoch までで損失関数の値は概ね一定値となったが、最終的に到達した値はλが大きい場合ほど大きな値となった。次に、抽出された特徴を示す中間層ニューロン 20 個の出力結果を Figure 4.17 に示した。前項の L1 正則化の場合、λの値が増加するにつれて 0 を出力するニューロンの数の増加が認められたが、KL-Divergence 正則化ではλ =0.05, λ =0.1 では逆に 0 を出力するニューロンの減少が認められた。この理由としては、正則化項の値が損失関数に上乗せされたため、Dying ReLU の原因となる「重みが負値側に大きく更新される」ことが抑制されたことによる可能性が考えられるが、詳細については十分に判断できない。さらに正則化の程度を上げλ =0.5 となると、0 を出力するニューロンが増加した。これらの条件で抽出された特徴からは、L1 正則化の場合と同様に、いずれからもジクロフェナクナトリウムの分布を示すニューロンと、粘着剤の分布を示すニューロン、顆粒状の分布を示すニューロンが認められた（Figure 4.17 中にそれぞれ赤枠、青枠、緑枠で示した）。ジクロフェナクナトリウムの分布を示す特徴を負二次イオン像（Figure 4.5 (a)）と比較すると、λ =0.1 の結果で最も負二次イオン像と近い分布が抽出された。さらに、デコーダー重み（Figure 4.18）の比較からも、λ =0.1 においてジクロフェナクナトリウムに特徴的なピークが最も S/N が良く現れた。さらに顆粒状の分布を示す特徴についても、λ =0.1 において最も明瞭なコントラストで抽出された（粘着剤はλ =0.05 とλ =0.1 で同程度である）。これらの結果から、λ =0.1 において最もジクロフェナクナトリウムの分布を単成分の分布として抽出できていると判断した。

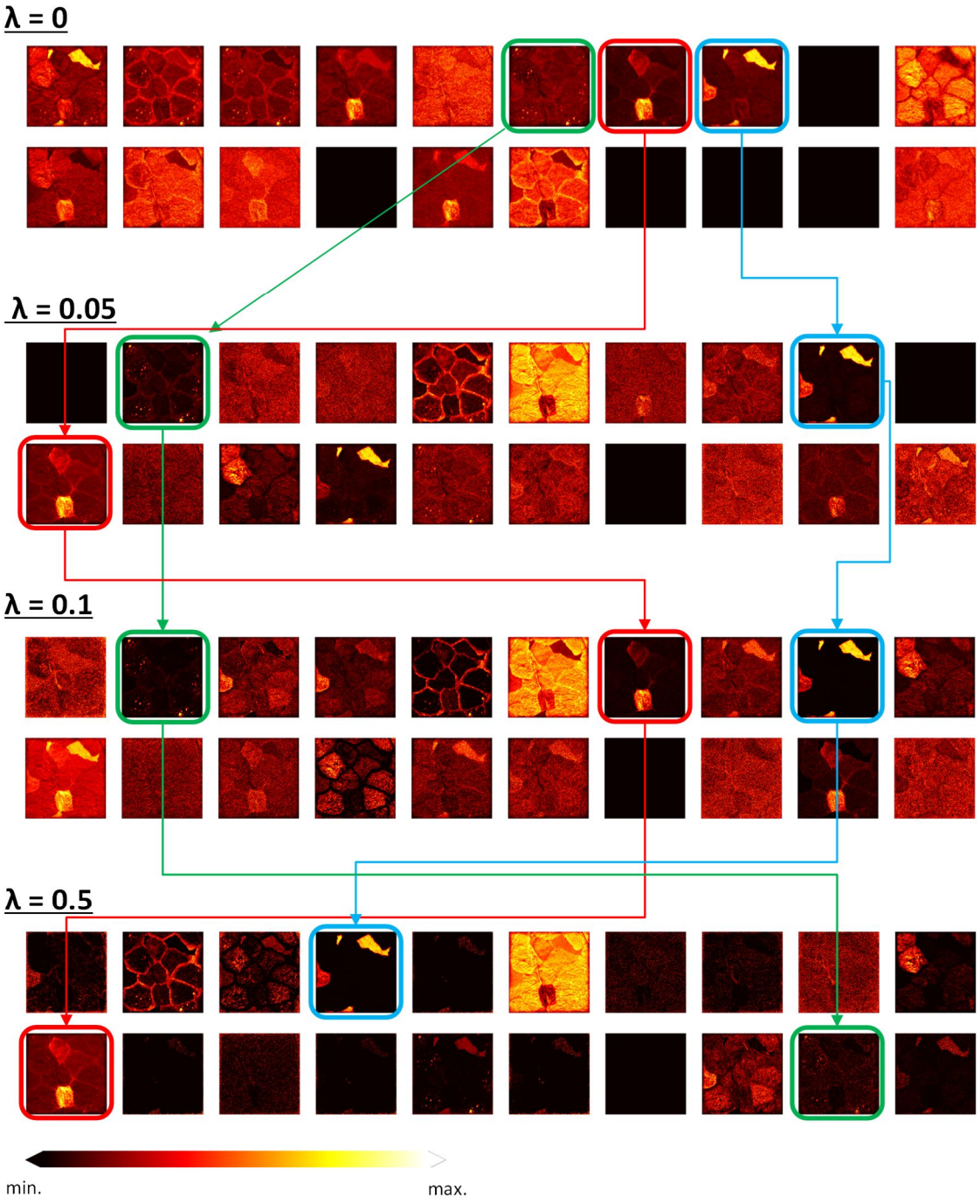
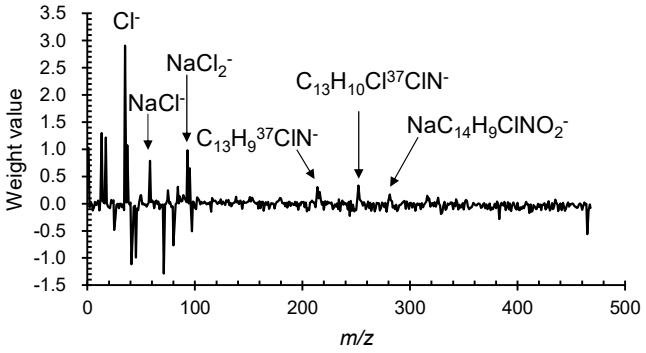


Figure 4.17　抽出された特徴に与える KL-Divergence 正則化の程度の影響（赤枠：主にジクロフェナクナトリウムの分布を反映した特徴、水色枠：主に粘着剤の分布を反映した特徴、緑枠：顆粒上の分布が確認される特徴）

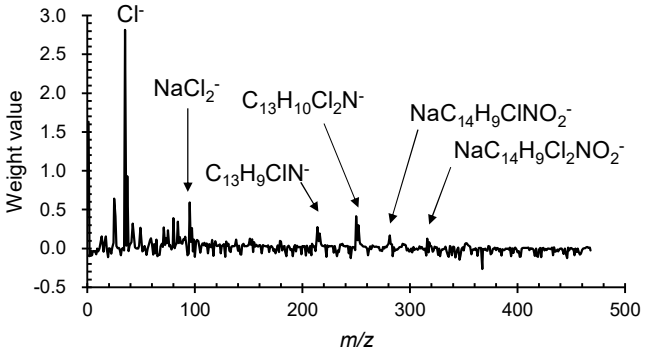
λ = 0

$W_{dec}7$



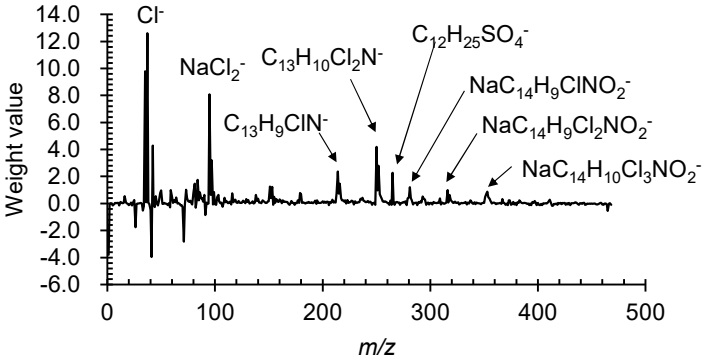
λ = 0.05

$W_{dec}11$



λ = 0.1

$W_{dec}7$



λ = 0.5

$W_{dec}11$

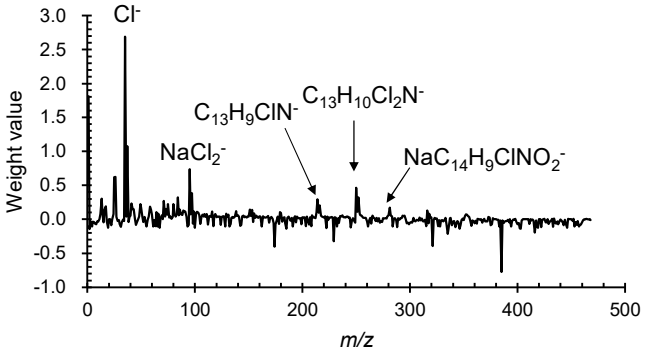


Figure 4.18　主にジクロフェナクナトリウムの分布を表す特徴に対応するデコーダー重み（KL-Divergence 正則化の程度による変化）＜活性化関数：ReLU によって中間層には正値のみが出力されるため、負の重みについては無視できる＞

4.4.5 正則化項の違いが特徴抽出性能に与える影響

L1 正則化および KL-Divergence 正則化のどちらを使用した場合でも、正則化の程度を適当な値にした場合に、正則化なしに比べて特徴抽出性能が向上することが確認された。これまで特徴抽出性能の評価のために着目してきたジクロフェナクナトリウム、粘着剤、顆粒状の分布については、L1 正則化と KL-Divergence 正則化で明確な差は認められなかった（Figure 4.12 および Figure 4.17、Fugure 4.13 および Figure 4.18）。そこで、他の比較可能な特徴として、（C9：0）、（C10：0）、（C16：0）、（C18：0）脂肪酸（エステル）の分布を反映した特徴に着目した。これらの脂肪酸の分布を反映した特徴は、L1 正則化（Figure 4.19 (C)）と KL-Divergence 正則化（Figure 4.19 (D)）のどちらにおいても、二次イオン像（Figure 4.19 (E)）と類似の分布として抽出された。しかし、デコーダー重み（Figure 4.19 (A), (B)）を確認すると、L1 正則化では（C24：0）、（C26：0）脂肪酸（エステル）に対応する $^{367}C_{24}H_{47}O_2^-$, $^{395}C_{26}H_{51}O_2^-$のピークが混在していることが確認された。（C24：0）、（C26：0）脂肪酸（エステル）の負二次イオン像（Figure 4.19 (F)）を踏まえると、L1 正則化では（C9：0）、（C10：0）、（C16：0）、（C18：0）脂肪酸（エステル）と、（C24：0）、（C26：0）脂肪酸（エステル）の分布が混ざって抽出されていると考えられる。

上記の結果より、一概にどちらの正則化項がより有用であるかは断定はできないものの、本検討に使用したデータ、条件に限っては、KL-Divergence による正則化を用いた方がより適当な結果が得られたと判断した。

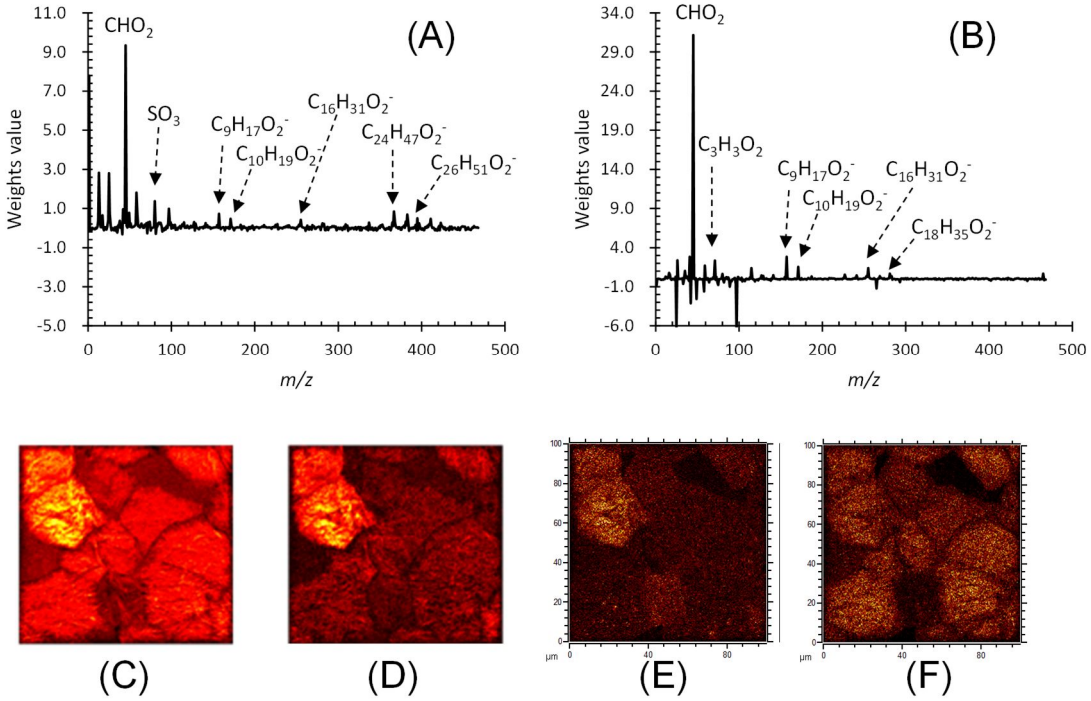


Figure 4.19　L1 正則化（λ=1×10^5）と KL-Divergence 正則化（λ=0.1）の比較例

(A), (C)：L1 正則化を行った場合のデコーダー重みと中間層ニューロンの出力結果

(B), (D)：KL-Divergence 正則化を行った場合のデコーダー重みと中間層ニューロンの出力結果

(E)：（C9：0）、（C10：0）、（C16：0）、（C18：0）脂肪酸の負 2 次イオン像の足し合わせ像

（$^{157}C_9H_{17}O_2^-$, $^{171}C_{10}H_{19}O_2^-$, $^{255}C_{16}H_{31}O_2^-$, $^{283}C_{18}H_{35}O_2^-$）

(E)：（C24：0）、（C26：0）脂肪酸の負 2 次イオン像の足し合わせ像（$^{367}C_{24}H_{47}O_2^-$, $^{395}C_{26}H_{51}O_2^-$）

4.4.6　学習時のバッチサイズの影響評価

第二章（2.2.5.1　勾配降下法）で述べた通り、ニューラルネットワークの学習の際のバッチサイズは、局所解の捕まりにくさに影響し、最終的な学習データ（特徴抽出性能）に影響を与える。また、同時に計算コストにも影響する。そこで、実際にバッチサイズを変更したことによって特徴抽出性能にどのような影響が現れるかを、Epoch 数を 200 に固定して検証した。Figure 4.20, Figure 4.21 に KL-Divergence 正則化を適用し、バッチサイズを 16 から 65536 まで変更した際の中間層ニューロンの出力結果を示した。

バッチサイズが 16 の場合、0 を出力する（特徴が抽出されない）ニューロンが多く観測された。これは、一回の重みの更新の際に用いる情報量が少なすぎて、有意な特徴を抽出できなかったためと推測される。一方で、バッチサイズを大きく 4096、16384、65536 とした場合、ジクロフェナクナトリウムと粘着剤の分布が中間層ニューロンの出力の多くを占める結果となった（Figure 4.21 の赤枠、

水色枠)。ここで、ジクロフェナクナトリウムと粘着剤が分布する領域は、負 2 次イオン像(Figure 4.5、Figure 4.19(E, F))にて細胞や細胞間脂質(脂肪酸など)の検出量が少なく、組成が周囲と比べて大きく異なる。さらに、バッチサイズを大きく設定することは、重みの更新時に使用する情報の平均化の程度を強めることを意味する。そのため、情報が平均化された際にジクロフェナクナトリウムと粘着剤の局在という明瞭な分布に、他の脂質などの微妙な分布の違いが埋もれてしまい、結果としてジクロフェナクナトリウムと粘着剤以外の成分の分布が特徴としてうまく抽出されなかったものと推察される。なお、バッチサイズが上記の中間である 64～1024 では、ジクロフェナクナトリウムや粘着剤、脂肪酸、顆粒状に分布する成分が、各条件で類似した分布として抽出されていることから、特徴抽出性能に顕著な差はないと考えられる。したがって本結果より、256 pixels × 256 pixels (= 65536 pixels)の TOF-SIMS イメージデータについては、バッチサイズをまずは 64～1024 の範囲で設定して学習を行うことを提案する。

4.5 結論

本章では、薬剤を浸透させた皮膚組織の TOF-SIMS の二次元画像データに対し、オートエンコーダーを適用することにより、薬剤分布や細胞間脂質の特徴的な分布を抽出できることを示した。さらに、スパースオートエンコーダーは単純なオートエンコーダーに比べて、特徴的な分布を持った成分(群)を、単独の特徴として抽出する(特徴抽出性能が向上する)傾向が認められた。これは、スパースオートエンコーダーでは学習過程で中間層をスパースにすることによって、中間層の表現力を極度に低下させずに、抽出される特徴の数を減らすことが可能なためと推察される。また、正則化の程度を左右するハイパーパラメーター(λ)の値や学習時のバッチサイズが、特徴抽出結果に与える影響についても調べ、正則化の程度については実際に出力された中間層のスパース性を考慮して複数の値で比較することが、最適な結果を得るためには重要であること、バッチサイズについては 64～1024 の範囲で設定することが望ましいといった、パラメーター設定の指針を得ることができた。

なお、本章の検討では各種のハイパーパラメーター(正則化パラメーター: λ、KL-divergence 正則化における目標活性律:q、バッチサイズなど)の値の検証をマニュアルで行ったが、オープンソース機械学習ライブラリである“scikit-learn”などには、最適化ハイパーパラメーター値の組み合わせを交差検証で評価する機能(Grid search、Random search など)が備わっている。多変量解析手法に比べて機械学習の難しい点としてハイパーパラメーターの数の多さが挙げられることから、必要に応じて上記の機能を使用することは、有用であると考えられる。

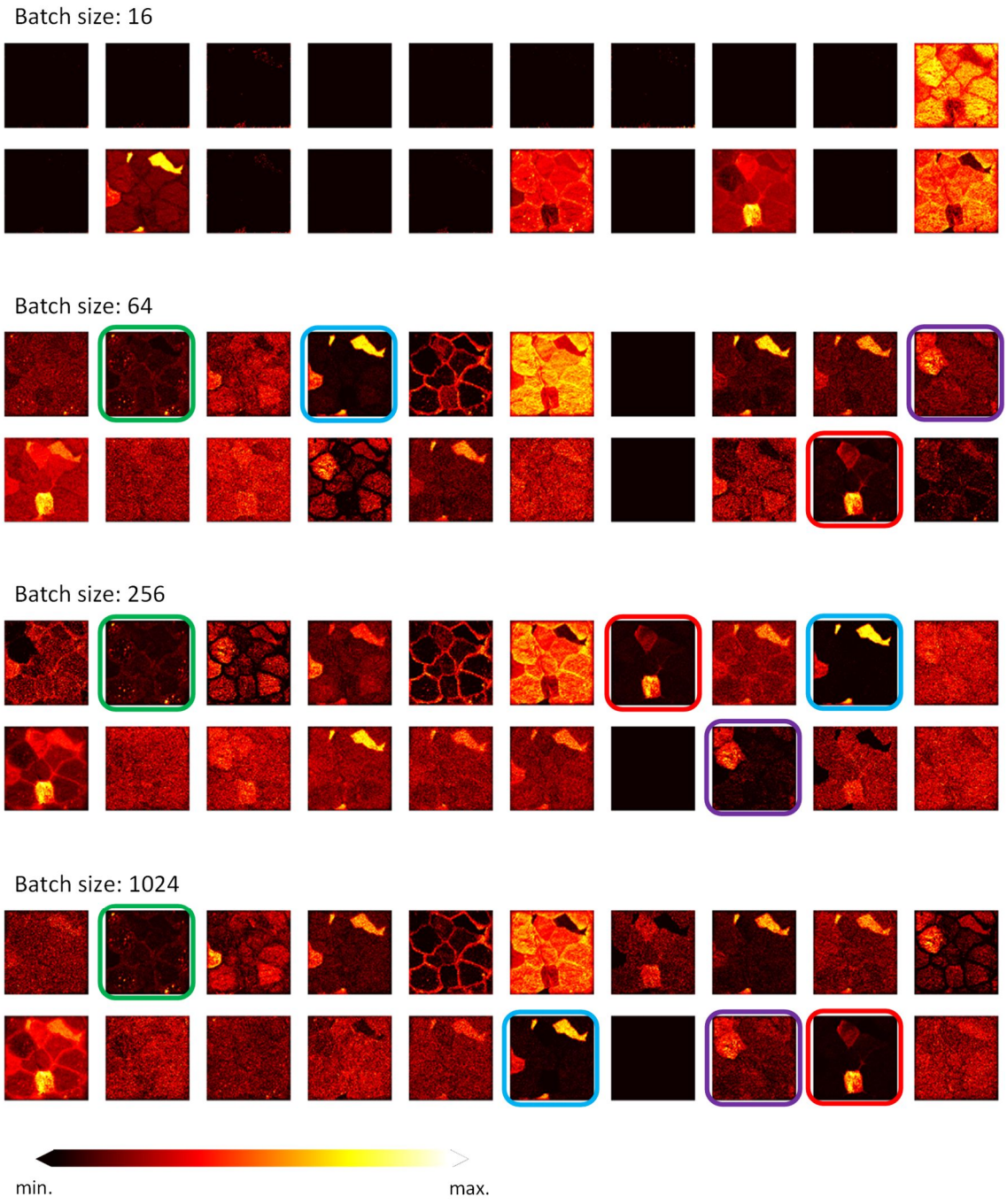


Figure 4.20　スパースオートエンコーダー（KL-Divergence 正則化）による中間層ニューロンの出力結果（バッチサイズ変更の影響）。＜赤枠：主にジクロフェナクナトリウム、水色枠：主に粘着剤、緑枠：顆粒上の分布、紫枠：（C9：0）、（C10：0）、（C16：0）、（C18：0）脂肪酸＞

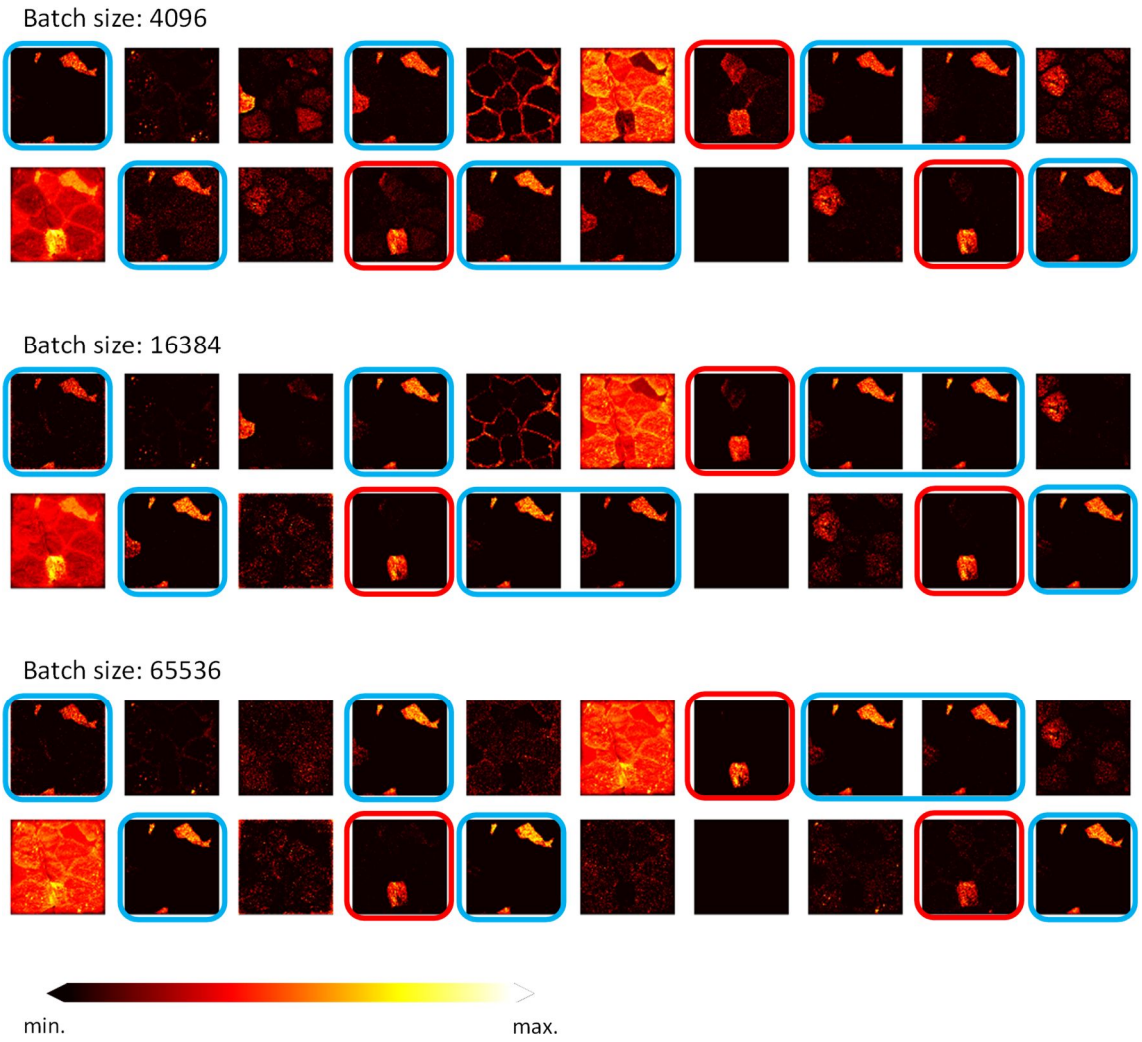


Figure 4.21　スパースオートエンコーダー（KL-Divergence 正則化）による中間層ニューロンの出力結果（バッチサイズ変更の影響）。

第五章
スパースオートエンコーダーと他の特徴抽出法の比較

5.1 はじめに

前章の検討によって、オートエンコーダーに正則化の概念を取り入れたスパースオートエンコーダーは、正則化の程度を調整することで、単純なオートエンコーダーに比べて浸透薬剤や細胞間脂質の分布を、単一成分の分布に近い状態で抽出できる性能を有することが示された。一方で、TOF-SIMS イメージデータの解析においては従来より主成分分析（PCA）や多変量スペクトル分解（MCR）といった多変量解析手法が使用されており、一定の有用な成果をもたらしてきた。しかしながら、それらの解析手法は解析データが純成分のデータの線形結合で構成されているとの仮定に基づいているという本質的な問題がある（TOF-SIMSデータは一般にマトリクス効果による非線形性を有するデータである）。人工ニューラルネットワークを用いたスパースオートエンコーダーは、非線形データへも対応が可能であることから、TOF-SIMS データを PCA、MCR に比べて処理できる潜在的能力がある。そこで本章では、同一 TOF-SIMS データに対してスパースオートエンコーダーと、PCA、MCR を適用することで、抽出されてきた特徴に差が現れるかどうか検証した。データとしては、前章で用いたヒト角質層の TOF-SIMS イメージデータを用いた。

5.2 実験方法

解析用データは第四章にて用いたヒト皮膚角質層の TOF-SIMS イメージデータを使用した（データの取得条件などについては、4.2.1, 4.2.2 参照）。

5.3 データ解析

5.3.1 データ前処理

TOF-SIMS データの質量軸を *m/z* 1 単位にビニングすることによって得られた 468 peaks×65536 pixels のデータを、スパースオートエンコーダーおよび多変量解析（PCA, MCR）の解析データとして使用した。スパースオートエンコーダーについては、前章と同様に前処理は実施せず、強度データをそのまま入力データとした。PCA の解析には Mean-centering を、MCR の解析には Poisson scaling を前処理として適用した。これらの前処理方法は、それぞれの解析手法を行う上で有用であることが知られている。

5.3.2 データ解析条件

ライブラリおよびハードウェアについては第三章の検討と同一である（Table 3.2 参照）。

スパースオートエンコーダーの具体的な構造としては第二章と同様に、Figure 3.10 に示したエンコーダーとデコーダーの 2 つの部分から成るシンプルなネットワーク構造を採用した。スパースオートエンコーダーの正則化項については、第四章の検討結果（4.4.5 正則化項の違いが特徴抽出性能に与える影響）より KL-Divergence 正則化を採用した。また。バッチサイズについても前章の検討結果（4.4.6 学習時のバッチサイズの影響評価）より 64 に設定した。解析条件を Table 5.1 に示した。

Table 5.1　スパースオートエンコーダーの解析条件

中間層サイズ		20
活性化関数（エンコーダーとデコーダーで共通）		ReLU
損失関数		MSE with KL-Divergence
正則化パラメーター	Target sparsity	0.1
（KL-Divergence）	Weight（λ）	0.1
最適化関数		Adam
バッチサイズ		64

PCA および MCR の実行は、MATLAB R2015b（Mathworks, Inc., USA）上で動作する多変量解析ソフトウェアである PLS-toolbox 8.0.2 および MIA-toolbox 2.9.2（Eigenvector Research, Inc., USA）を用いて行った。MCR はスパースオートエンコーダーの結果を参考にして、成分数を 10 および 20 とした 2 つのパターンについて解析を実施した。

5.4 結果と考察

5.4.1 三種類の特徴抽出法により抽出された特徴

第四章の検討により得られた、スパースオートエンコーダー（KL-Divergence 正則化）による特徴抽出結果（中間層の出力）を Figure 5.1 に示した。

MCR の成分数を 10 として得られたスコアプロット（二次元プロット）を Figure 5.2 に、成分数を 20 として得られたスコアプロットとローディングプロットをそれぞれ Figure 5.3、Figure 5.4 に示した。

PCA については、スクリープロット（Figure 5.5）より第一主成分から第十主成分までで元データの分散の 68 %を占め、第九主成分以降は寄与率が 1 %以下であることから、第十主成分までで元データの主要な特徴がすべて抽出できていると判断し、第一から第十主成分までのスコアプロットを Figure 5.6 に示した。実際に第十一主成分以降についても確認したが、第十主成分までのスコアプロットと明確に異なる分布を示す主成分は認められなかった。なお、PCA のスコアプロットは、スパースオートエンコーダーの中間層からの出力結果や、MCR のスコアプロットと異なり負値を持つ。そのため、正値を暖色のカラースケールで、負値を寒色のカラースケールで表示した（0 が黒色）。

各特徴抽出法によって抽出された特徴の分布と、前章の Figure 4.5, Figure 4.19 に示した負二次イオン像との比較から、抽出された特徴がそれぞれどの化合物分布に対応しているのかを調べ、それらを Table 5.2－5.5 に示した。

MCR の成分数を 10 とした場合と 20 とした場合を比較すると（Figure 5.2、Figure 5.3）、互いに類似の分布が得られているが、MCR の成分数を 20 とした場合では類似の分布を示す特徴が複数存在した。例として Component 2, 3 は共に粘着剤（ポリブチルアクリレート）の分布に対応する特徴と見做せる。しかし、それらのローディングプロット（Figure 5.4）を確認すると、Component 2 にてポリブチルアクリレートに特徴的なピークのうち主要な $^{41}C_2HO^-$, $^{71}C_3H_3O_2^-$, $^{81}C_5H_5O^-$, $^{115}C_6H_{11}O_2^-$が現れ

るが、酸素を含む化合物に共通の $^{16}O^-$は観測されない。一方で Component 3 にて $^{16}O^-$が特徴的に観測され、他のポリブチルアクリレートに特徴的なピークは非常に弱い。同様に、ジクロフェナクナトリウムの分布に対応する Component 15, 19 においても、ジクロフェナクナトリウムに特徴的なピークのうち、$^{93}NaCl_2^-$, $^{214}C_{13}H_9ClN^-$, $^{250}C_{13}H_{10}Cl_2N^-$などは Component 15 で強く、$^{35}Cl^-$は Component 19 で強い。その他、炭化水素に共通の $^{13}CH^-$, $^{25}C_2H^-$, $^{49}C_4H^-$がそれぞれ単独で強く観測される特徴が現れた。これらは MCR の成分数の設定が 20 では多すぎたため、同じ成分に由来し同じ分布を持つピーク（変数）を別の成分として過剰に分割して抽出してしまっているものと考えられる。このように、MCR では成分数というハイパーパラメーターの設定が非常に重要となる。そのため、実際に異なる分布として存在する化合物の数が不明の状態では最適な結果を得るのは困難であり、事前に PCA を行うなどして適切な成分数の予想をする必要がある。今回は成分数を 20 とした場合で成分数を 10 とした場合に比べて新たに異なる分布を持つ特徴が抽出されてこなかったことから、成分数は 10 で適当であると判断した。

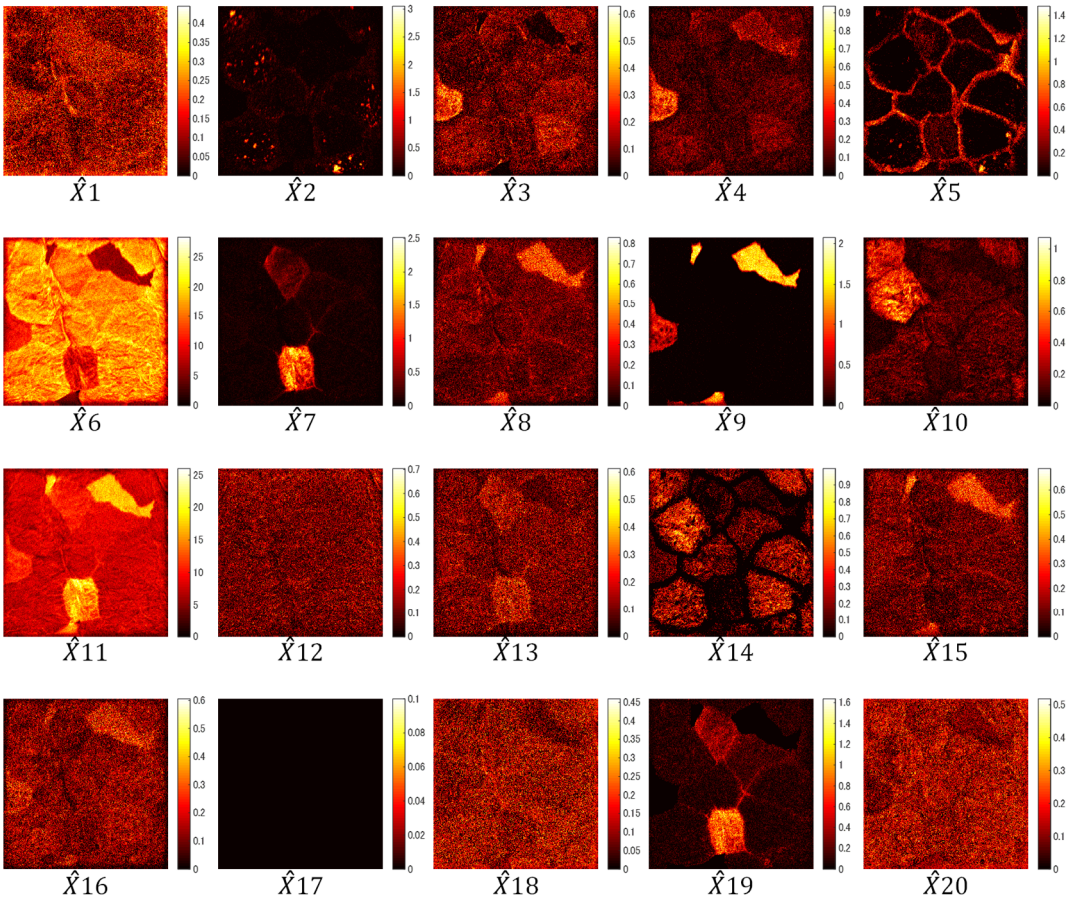


Figure 5.1　スパースオートエンコーダー（KL-Divergence 正則化）による学習後の中間層からの出力画像（エンコーダーにより圧縮された特徴画像、「Figure 4.16 の λ=0.1」と同一データ）。

Table 5.2　スパースオートエンコーダーにより抽出された特徴

ニューロンの番号	主な化合物の分布
$\hat{X}2$	顆粒状に分布する成分
$\hat{X}3, \hat{X}4$	硫酸コレステロール
$\hat{X}5$	アミド（角質細胞の境界）
$\hat{X}7, \hat{X}19$	ジクロフェナクナトリウム
$\hat{X}9$	粘着剤（ポリブチルアクリレート）
$\hat{X}10$	(C9:0), (C10:0), (C16:0), (C18:0)脂肪酸またはエステル
$\hat{X}14$	(C24:0), (C26:0)脂肪酸またはエステル

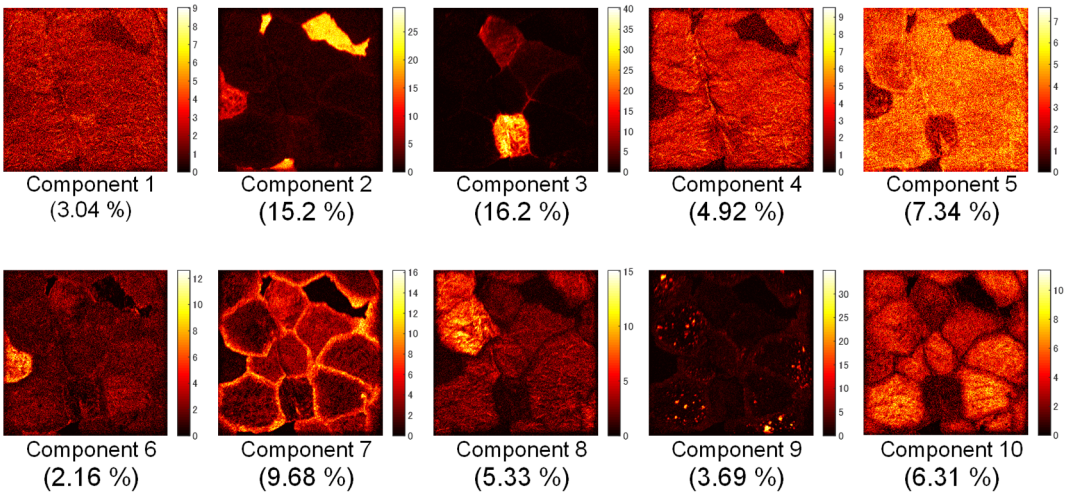


Figure 5.2　MCR（Number of component =10）によって得られたスコアの二次元プロット（括弧内の数字は各成分の寄与率）

Table 5.3　MCR（Number of component: 0）により抽出された特徴

主成分	主な化合物の分布
Component 2	粘着剤（ポリブチルアクリレート）
Component 3	ジクロフェナクナトリウム
Component 6	硫酸コレステロール
Component 7	アミド（角質細胞の境界）
Component 8	(C9:0), (C10:0), (C16:0), (C18:0)脂肪酸またはエステル
Component 9	顆粒状に分布する成分
Component 10	(C24:0), (C26:0)脂肪酸またはエステル

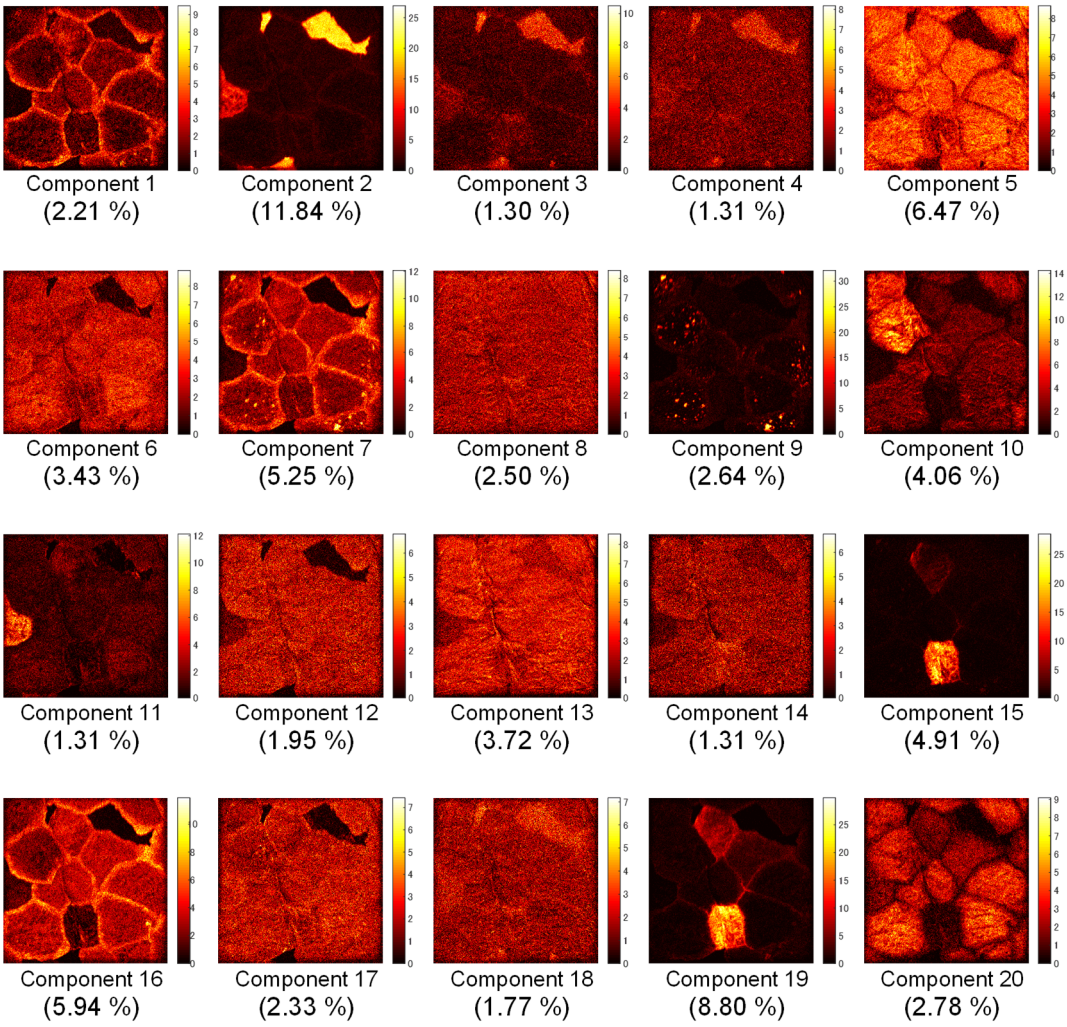


Figure 5.3　MCR（Number of component =20）によって得られたスコアの二次元プロット（括弧内の数字は各成分の寄与率）

Table 5.4　MCR（Number of component: 0）により抽出された特徴

主成分	主な化合物の分布
Component 1, 7, 16	アミド（角質細胞の境界）
Component 2, 3	粘着剤（ポリブチルアクリレート）
Component 5, 20	(C24:0), (C26:0)脂肪酸またはエステル
Component 9	顆粒状に分布する成分
Component 10	(C9:0), (C10:0), (C16:0), (C18:0)脂肪酸またはエステル
Component 11	硫酸コレステロール
Component 15, 19	ジクロフェナクナトリウム

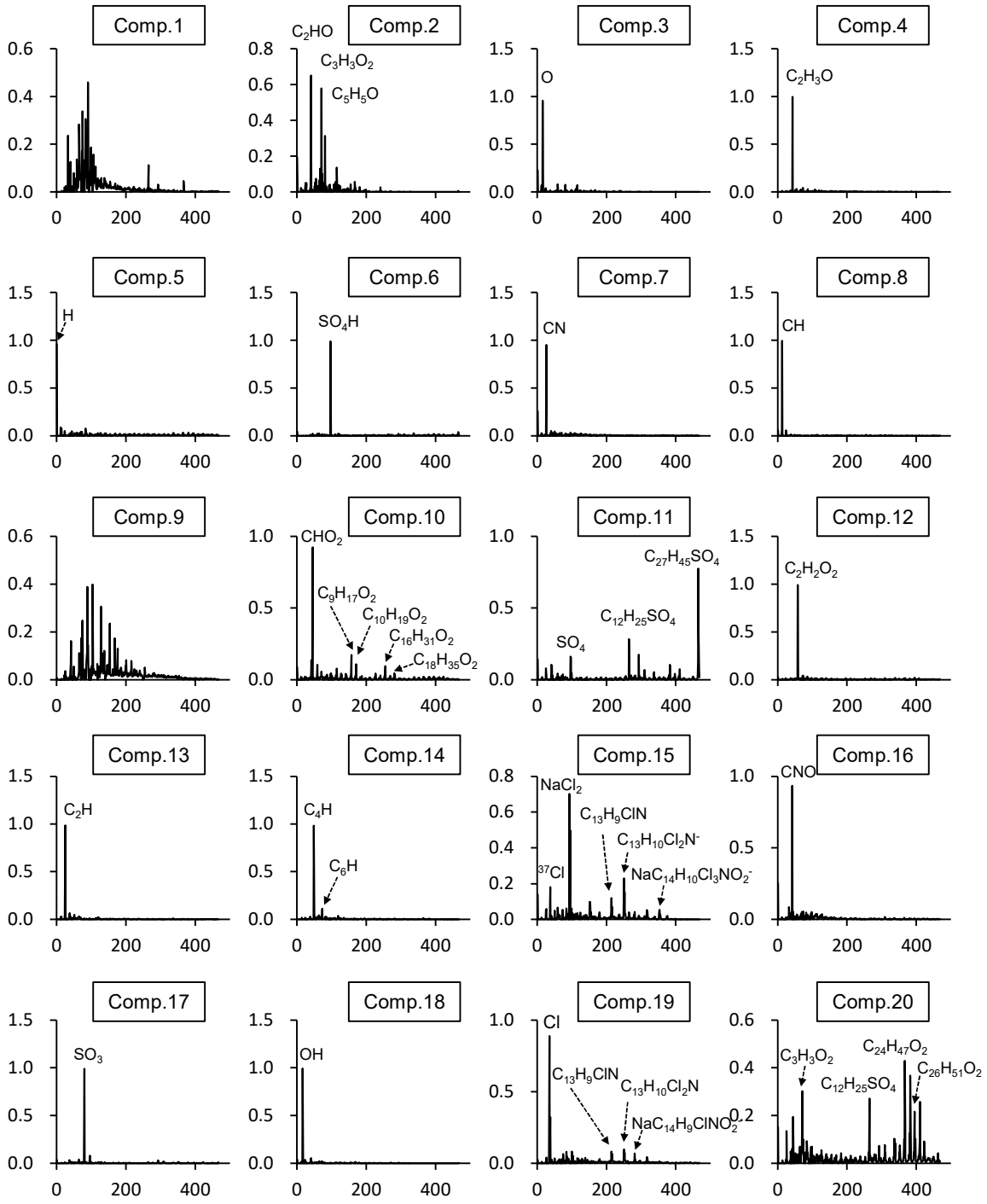


Figure 5.4　MCR（Number of component =20）によって得られたローディングプロット

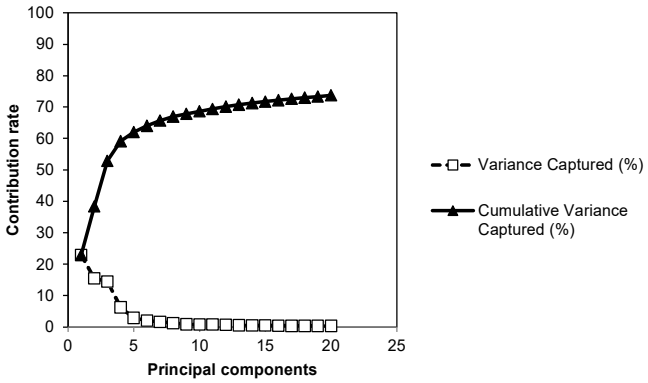


Figure 5.5　PCA 結果（スクリープロット）

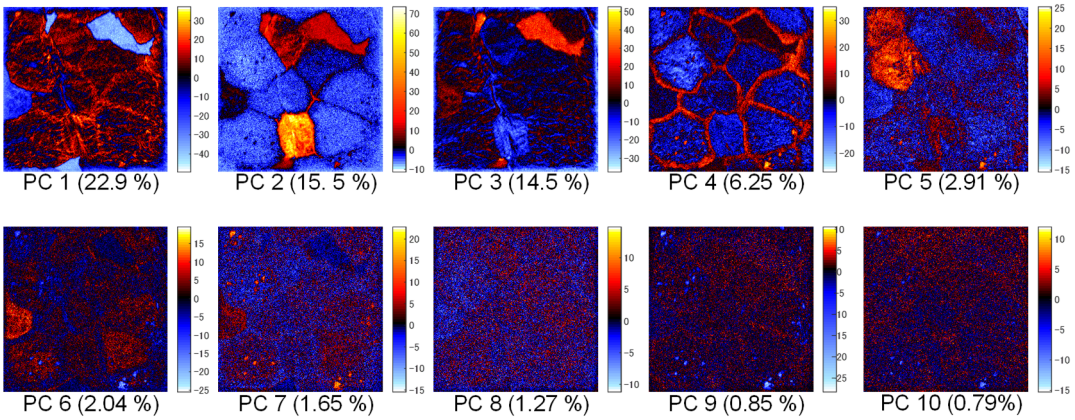


Figure 5.6　PCA により得られた主成分得点（スコア）の二次元プロット（括弧内の数値は各主成分の寄与率を表す）。

Table 5.5　PCA により抽出された特徴（括弧内の＋とーは正値および負値を表す）

主成分		主な化合物の分布
PC3 (+)	PC1 (−)	粘着剤（ポリブチルアクリレート）
PC2 (+)	PC3 (−)	ジクロフェナクナトリウム
	PC2 (−)	(C24:0), (C26:0)脂肪酸またはエステル
PC4 (+)	PC6 (−)	アミド（角質細胞の境界）
PC5 (+)	PC4 (−)	(C9:0), (C10:0), (C16:0), (C18:0)脂肪酸またはエステル
PC6 (+)		硫酸コレステロール
PC7 (+)	PC9 (−) PC10 (−)	顆粒状に分布する成分

5.4.2 スパースオートエンコーダーと他の特徴抽出法の結果の比較

三種の特徴抽出法によって抽出された各特長の中身を詳細に比較するために、代表的な特長について質量スペクトルに対応するデータ(スパースオートエンコーダーは重み、MCR と PCA はローディングプロット)において、どのような質量ピークが観測されているのかを調べた。なお、スパースオートエンコーダーでは入力層から中間層への重みであるエンコーダー側の重み(W_{enc})と、中間層から出力層への重みであるデコーダー側の重み(W_{dec})の二つが存在する。第三章にて示した通り、W_{dec}には入力データに含まれる微小なピークの影響を含むのに対し、W_{dec}では一度、特徴として抽出されたものから入力層を再現するためのパラメーターでありノイズの影響が少ない。そのため、W_{enc}に比べ解釈が容易なデータとなることから、本検討ではW_{dec}を用いて特徴の解釈を行った。代表的な特徴としては、皮膚内部に浸透させた薬剤であるジクロフェナクナトリウムの局在の分布と、細胞間脂質である脂肪酸の分布に着目した。

はじめに、Figure 5.7 のジクロフェナクナトリウムの分布に対応した特徴について見ると、どの解析手法の結果からもジクロフェナクナトリウムの分子構造に特徴的な $^{214}C_{13}H_9NCl^-$, $^{250}C_{13}H_{10}NCl_2^-$, $^{316}C_{14}H_9NO_2NaCl_2^-$, $^{352}C_{14}H_{10}NO_2Cl_3Na^-$が現れた(赤▼で表示)。特にスパースオートエンコーダー(a)と MCR(b)の結果では、低質量側の $^{35}Cl^-$や $^{93}NaCl_2^-$のピークに比べて $^{214}C_{13}H_9NCl^-$, $^{250}C_{13}H_{10}NCl_2^-$などの強度比が大きく、分布(d, e)も負二次イオン像 (g)に近いことから、よりジクロフェナクナトリウムの分布を正確に抽出できているものと考えられる。一方で PCA の解析結果(c)においてジクロフェナクナトリウムの分布を反映していると考えられる負値には、ジクロフェナクナトリウム由来のピークのほかに粘着剤(ポリブチルアクリレート)由来の $^{41}C_2HO^-$, $^{71}C_3H_3O_2^-$, $^{81}C_5H_5O^-$が現れた(青■で表示)。これらのピークは、特徴分布(f)に白矢印で示した領域に起因すると考えられる。この領域は負 2 次イオン像(Figure 5.7(g))との比較より、粘着剤露出部に相当する。つまり、スパースオートエンコーダーと MCR では、ジクロフェナクナトリウムが局在している領域をうまく特徴として抽出できたが、PCA では不十分な結果であったと言える。なお、MCR の結果(Figure 5.7(b))ではジクロフェナクナトリウム由来のピークのみが主要なピークとして観測されたが、スパースオートエンコーダーの結果(Figure 5.7(a))ではジクロフェナクナトリウムに加えてアミド由来の $^{42}CNO^-$とポリオキシエチレンアルキル硫酸エステル由来の $^{265}C_{12}H_{25}SO_4^-$が観測された。負二次イオン像(Figure 4.5 (c), (e))より、これらのピークは測定面内の大部分の領域で観測されており、ジクロフェナクナトリウムが多い領域においても観測されている。そのため、主にジクロフェナクナトリウムと混ざって特徴として抽出されたと考えられる。なお、これらのピークは MCR においては後述の脂肪酸の分布と共に抽出されている(Figure 5.8)。

続いて、Figure 5.8 に示した(C24:0), (C26:0)脂肪酸の分布に対応した特徴に着目すると、スパースオートエンコーダーと MCR の質量スペクトルデータ(Figure 5.8 a, b)では、 (C24:0), (C26:0)脂肪酸に特徴的な $^{367}C_{24}H_{47}O_2^-$, $^{395}C_{26}H_{51}O_2^-$(灰色●で表示)が特徴的に観測された。また、スパースオートエンコーダーと MCR の特徴分布(Figure 5.8 d, e)では、負二次イオン像(Figure 5.8 e)と類似した結果が得られた。したがって、これら二手法ではこれらの脂肪酸の分布をよく抽出できていると言える。一方で、PCA の結果では、質量スペクトルデータ(Figure 5.8 c)では負値に $^{367}C_{24}H_{47}O_2^-$,

$^{395}C_{26}H_{51}O_2^-$が特徴的に観測されたものの、特徴分布（Figure 5.8 f）ではジクロフェナクナトリウムが局在する領域において、脂肪酸の寄与が存在しないことになり、負二次イオン像と一致しない結果である。これは PCA は正値と負値にそれぞれ意味を持つように特徴抽出がなされるため、単成分の分布の抽出には原理上、適していないためと考えられる。なお、スパースオートエンコーダーについても、活性化関数に正値と負値の両方を出力する関数を選択した場合には、PCA と同様に単成分の分布の抽出に不適な手法となる。そのため、活性関数に非負性を持つ関数を選択することが、TOF-SIMS のデータ解析については重要と言える。そのほか、Figure 5.9 の質量スペクトルデータにおいては、スパースオートエンコーダーでは(C9:0), (C10:0), (C16:0), (C18:0)脂肪酸に対応するピーク（$^{157}C_9H_{17}O_2^-$, $^{171}C_{10}H_{19}O_2^-$, $^{255}C_{16}H_{31}O_2^-$, $^{283}C_{18}H_{35}O_2^-$）のみが特徴的に観測されたのに対し、MCR ではそれらに加えて(C24:0), (C26:0)脂肪酸に特徴的な $^{367}C_{24}H_{47}O_2^-$, $^{395}C_{26}H_{51}O_2^-$も観測され、分布の異なる二種の脂肪酸類が混在して抽出された。

以上の結果から、スパースオートエンコーダーと MCR では細かな違いがあるものの、透過した薬物や細胞間脂質の分布を反映した特徴については、二次イオン像と非常に類似した分布を示し、特徴的な面内分布を持つ成分（群）を単独の特徴として抽出することができることが確認された。なお、MCR の成分数はあらかじめ設定しなければならないハイパーパラメーターであり、成分数を過剰に設定してしまうと、単一成分の分布を分割して別の特徴として抽出してしまい、解釈が困難な結果となってしまう。スパースオートエンコーダーでは中間層サイズを余裕のあるサイズに設定することで、MCR で同様の懸念は回避できるが、第四章の検討より、正則化の程度によって抽出される特徴に変化があることがわかっているため、必要に応じて正則化の程度は調整する必要がある。

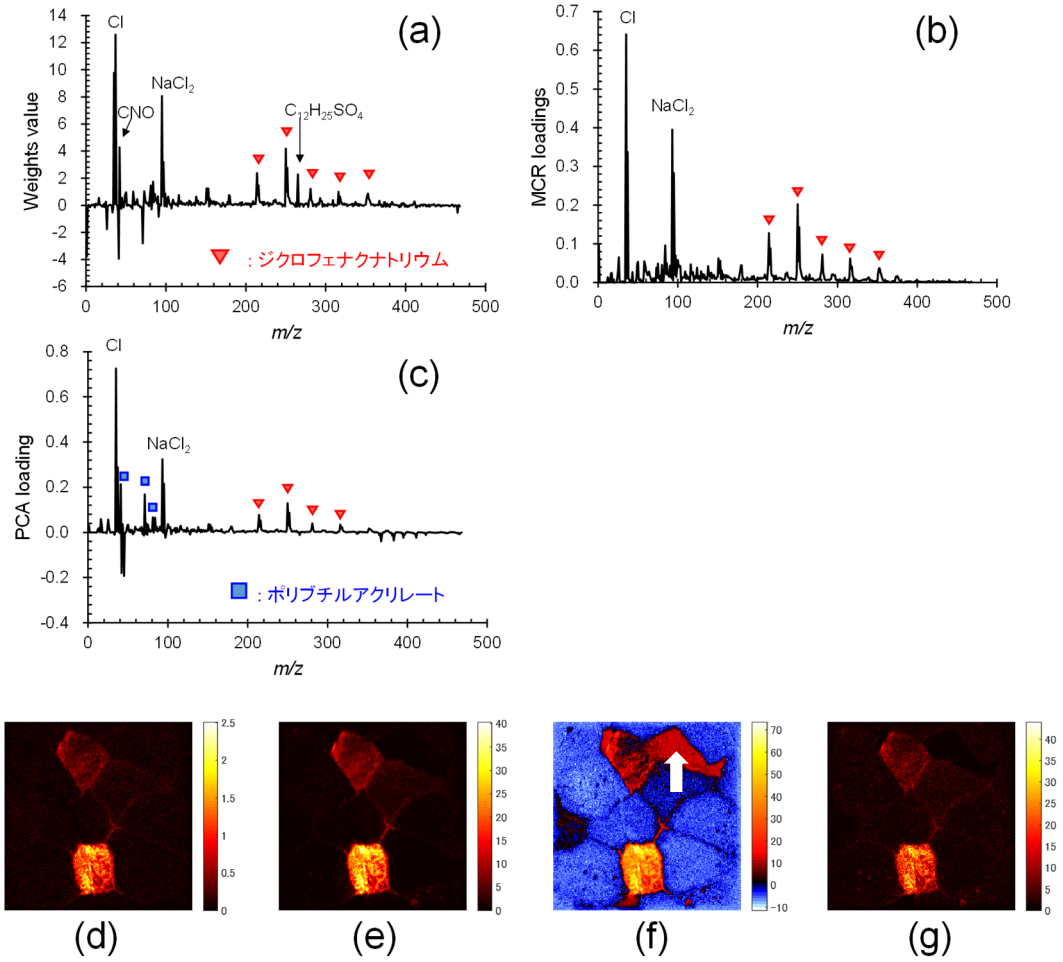


Figure 5.7　異なる特徴抽出法によって得られたジクロフェナクナトリウムの分布に対応した特徴（(a-c)は質量スペクトル、(d-f)は二次元プロット）。

(a), (d): スパースオートエンコーダー（KL-Divergence 正則化）

(b), (e): MCR（Number of components = 10）

(c), (f): PCA

(g): ジクロフェナクナトリウム由来の二次イオン像

(sum of $^{214}C_{13}H_9ClN^-$, $^{250}C_{13}H_{10}Cl_2N^-$, $^{316}NaC_{14}H_9Cl_2NO_2^-$)

＜活性化関数：ReLU によって中間層には正値のみが出力されるため、オートエンコーダーの負の重みについては無視できる＞

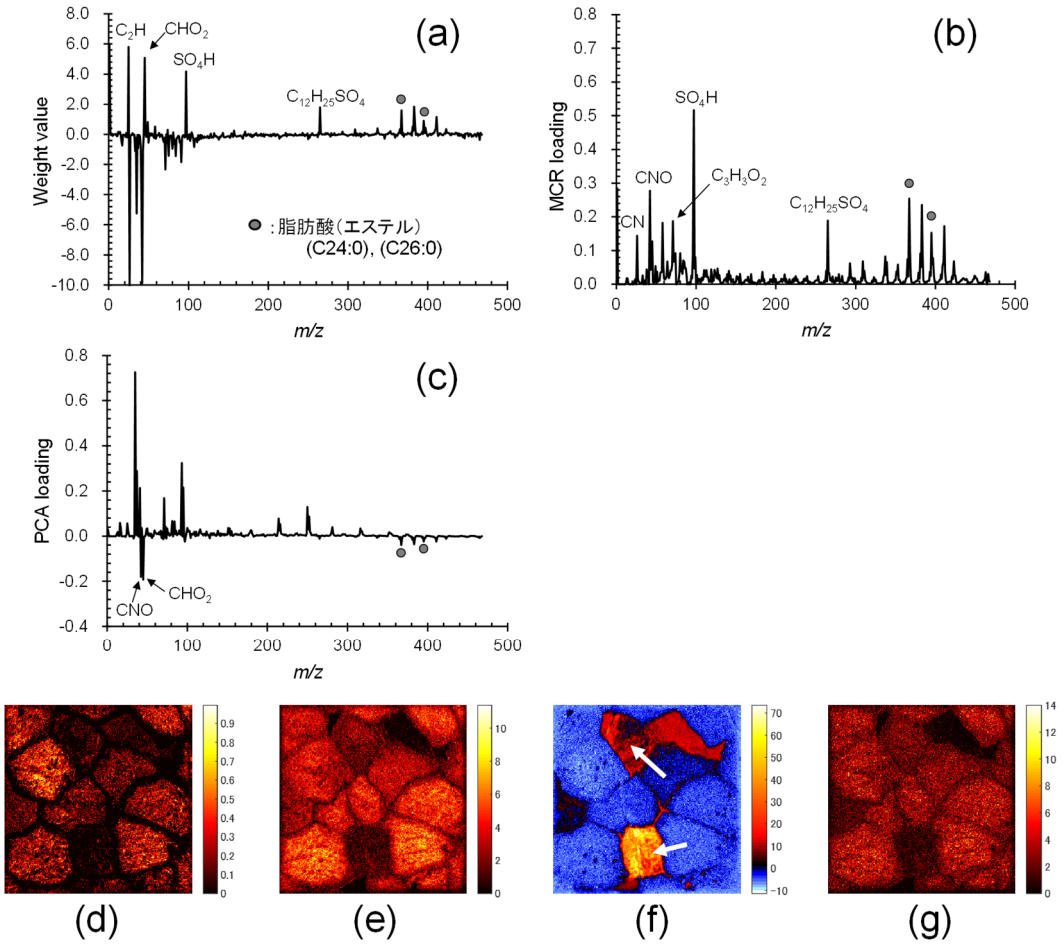


Figure 5.8　異なる特徴抽出法によって得られた（C24：0），（C26:0）脂肪酸の分布に対応した特徴（(a-c)は質量スペクトル、(d-f)は二次元プロット）。

(a), (d): スパースオートエンコーダー（KL-Divergence 正則化）

(b), (e): MCR（Number of components = 10）

(c), (f): PCA（白矢印部分はジクロフェナクナトリウムの局在する領域に対応）

(g): （C24：0），（C26:0）脂肪酸由来の二次イオン像

(sum of $^{367}C_{24}H_{47}O_2^-$, $^{395}C_{26}H_{51}O_2^-$)

＜活性化関数：ReLU によって中間層には正値のみが出力されるため、オートエンコーダーの負の重みについては無視できる＞

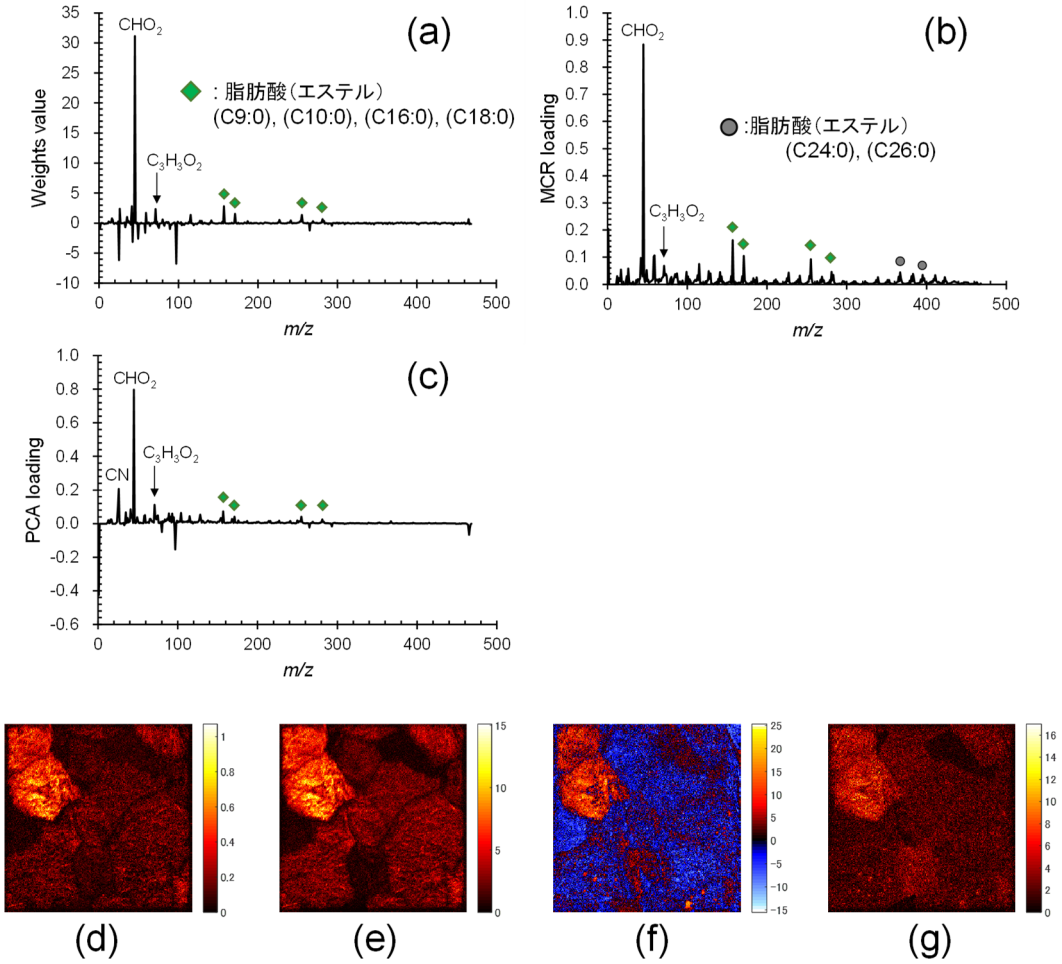


Figure 5.9　異なる特徴抽出法によって得られた（C9:0），（C10:0），（C16:0），（C18:0）脂肪酸の分布に対応した特徴（(a-c)は質量スペクトル、(d-f)は二次元プロット）。

(a), (d): スパースオートエンコーダー（KL-Divergence 正則化）

(b), (e): MCR（Number of components = 10）

(c), (f): PCA

(g): （C9:0），（C10:0），（C16:0），（C18:0）脂肪酸由来の二次イオン像

(sum of $^{157}C_9H_{17}O_2^-$, $^{171}C_{10}H_{19}O_2^-$, $^{255}C_{16}H_{31}O_2^-$, $^{283}C_{18}H_{35}O_2^-$)

＜活性化関数：ReLU によって中間層には正値のみが出力されるため、オートエンコーダーの負の重みについては無視できる＞

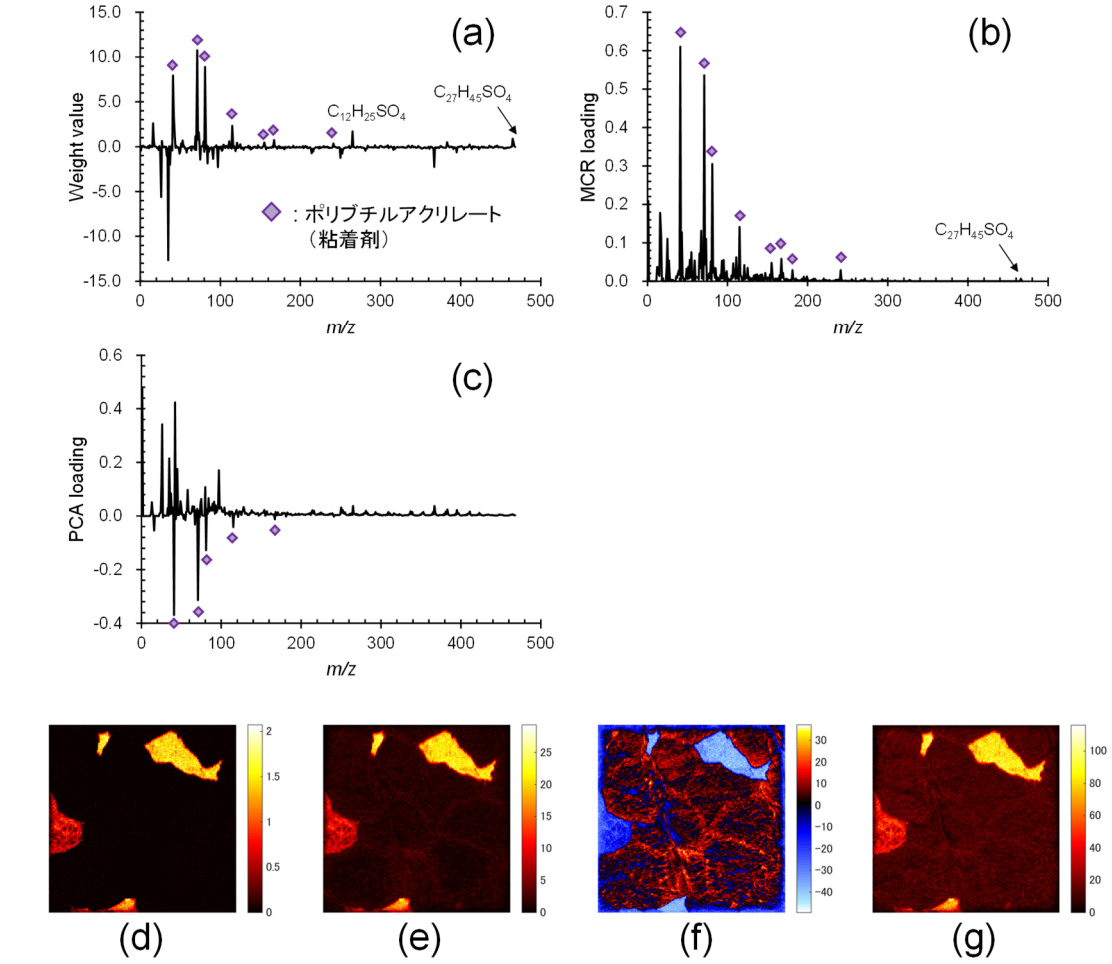


Figure 5.10　異なる特徴抽出法によって得られた粘着剤（ポリブチルアクリレート）の分布に対応した特徴（(a-c)は質量スペクトル、(d-f)は二次元プロット）。

(a), (d): スパースオートエンコーダー（**KL-Divergence** 正則化）

(b), (e): MCR（**Number of components = 10**）

(c), (f): PCA

(g): 粘着剤（ポリブチルアクリレート）由来の二次イオン像

(sum of $^{41}C_2HO^-$, $^{71}C_3H_3O_2^-$, $^{81}C_5H_5O^-$, $^{115}C_6H_{11}O_2^-$)

＜活性化関数：ReLU によって中間層には正値のみが出力されるため、オートエンコーダーの負の重みについては無視できる＞

特徴的な分布を持った成分の分布を、個別の特徴として抽出できるかどうかという点は、特徴抽出法としての重要な性能である。それと同時に、特徴抽出に必要な計算コストも、実際の分析の現場においては重要な性能の一つであると言える。そこで、スパースオートエンコーダーと MCR、PCA の三種の特徴抽出法の性能について、計算コストの面から比較を行った。同じハードウェア環境において解析に要した時間を測定した結果を Table 5.6 に示した。

Table 5.6 各手法における結果を得るまでにかかる所要時間の比較

	スパースオートエンコーダー（KL-Divergence 正則化）				PCA	MCR	
	Batch size					成分数	
	16	64	256	1024	—	10	20
Time (sec)	2260	600	200	140	5 <	210	1400

上表に示した所要時間の値より、スパースオートエンコーダーは、バッチサイズの大きさによって計算時間が変わるが、第四章の結果よりバッチサイズが 64 から 1024 においては抽出された特徴に顕著な差は認められないことから、MCR よりも短時間で結果を得ることも可能である。MCR は成分数の仮定によって計算時間が大きく変わり、この点からも MCR を実施するうえで事前の成分数の見積もりが重要であることがわかる。一方で、PCA は計算時間が非常に短いため、データの概略を迅速に把握するという用途では有用な解析手法であると言える。

5.5 結論

本章では、第三章、第四章にて特徴抽出法としての有用性を確認したスパースオートエンコーダーについて、従来よりTOF-SIMSデータの解析に用いられてきた特徴抽出法であるPCA、MCRと同一データを用いて直接的な特徴抽出性能の比較を行った。具体的には皮膚角質層内に浸透させた薬剤（ジクロフェナクナトリウム）と細胞間脂質の脂肪酸（異なる分布を持つ鎖長違いの二つのグループ）、下地の粘着剤の分布に着目し、それらが各特徴抽出法にてどのように抽出されてくるのかを比較した。

スパースオートエンコーダーとMCRは特徴が正値のみを出力することから、類似した面内分布を持つ成分の情報が、各特徴に抽出された。一方、PCA では各特徴（主成分）に正値と負値の両方が含まれることから、スパースオートエンコーダーと MCR に比べて解釈が困難である。そのため、特徴的な面内分布を持つ成分（群）を、それぞれ個別の特徴として抽出できるかどうかという観点で判断すると、スパースオートエンコーダーと MCR の方が PCA に比べて有用である。スパースオートエンコーダーと MCR の比較では、本検討で実施した条件では、類似した特徴が抽出され、明らかな優劣については判断できない。しかし、スパースオートエンコーダーでは鎖長違いの脂肪酸のグループについても、別の特徴として明確に分離することができており、MCR と同等かそれ以上に性能を有する可能性を示した。

特徴抽出に必要な計算コストの観点から優位性を論じると、PCA は他の二手法に比べて明らかに高速なアルゴリズムであることが確認された。そのため、データの概略を迅速に把握するという用途においてはPCAを用い、試料中に含まれる成分がどのような面内分布をしているかといった、具体的な情報を取得したい場合は、スパースオートエンコーダーまたは MCR を用いるといった使い分けが提案される。

以上のことから、スパースオートエンコーダーは現状でも、これまでTOF-SIMSデータの解析に用いられてきた PCA、MCR という二手法と比較して同等以上の性能を持つと言える。更に、多様な TOF-SIMS データに対してスパースオートエンコーダーの適用検討を行い、ネットワーク構造の改良や、汎用的な正則化パラメーターなどを導出することによって、TOF-SIMS データの主要な解析手法となることが期待される。

第六章
スパースオートエンコーダーによるマトリックス効果補正の検証

6.1 はじめに

第三、四、五章では、生体試料の TOF-SIMS データからのオートエンコーダーによる特徴抽出を検討した。抽出された特徴が生物学的知見や元データ（二次イオン像に依る分布）と照らし合わせて妥当と見做せる特徴が抽出できるかどうかに着目して、その有用性を評価した。本章では、定量的な視点からオートエンコーダーの有用性を検証するために、マトリックス効果によって濃度への応答が非線形となるような試料の TOF-SIMS データについて、オートエンコーダーが適切な濃度応答性を示すことができるか検討する。モデル試料として組成比が明らかな二成分混合系の有機薄膜試料を用い、その TOF-SIMS 測定データをスパースオートエンコーダーを用いて解析し、得られた特徴がマトリックス効果の影響をどのように反映するか検証する。

6.2 解析データ

解析用データとして、2014 年に行われた Versailles Project on Advanced Materials and Standards(VAMAS) Technical Working Areas (TWA) 2 におけるラウンドロビンテストで取得された、低分子有機物の積層蒸着膜のデプスプロファイルデータを用いた[90]。

積層蒸着膜は、Figure 6.1 に示した 2 種の低分子化合物（Irganox 1010 と Fmoc-PFLPA）の体積比（v/v）を変えた層を複数含む。各層における Irganox1010 の体積分率（$\varphi_{Irganox}$）は Figure 6.2 に示した。なお組成比の正確さについては、Shard らの報告（同一試料に対して XPS 分析を行い窒素およびフッ素の原子数比を算出）より、設計値の～%以内に収まっていることが確認されている。上記構造の積層蒸着膜に対して、TOF-SIMS 測定（128 pixels×128 pixels）と Ar-GCIB（ガスクラスターイオンビーム）による表面エッチングを交互に 160 cycle 行うことで、Figure 6.2 の層 1 から層 8 までの深さ方向分析データが得られた。更に面データ（128 pixels×128 pixels）を積分して 128 pixels×（1 pixel）×160 cycles の形に変形された。このような形のデータを質量スペクトルにて観測された 1134 個の二次イオンピークについて作成することで、最終的に「1134 ions ×128 pixels×160 cycles」のデータを得た（Figure 6.3）。

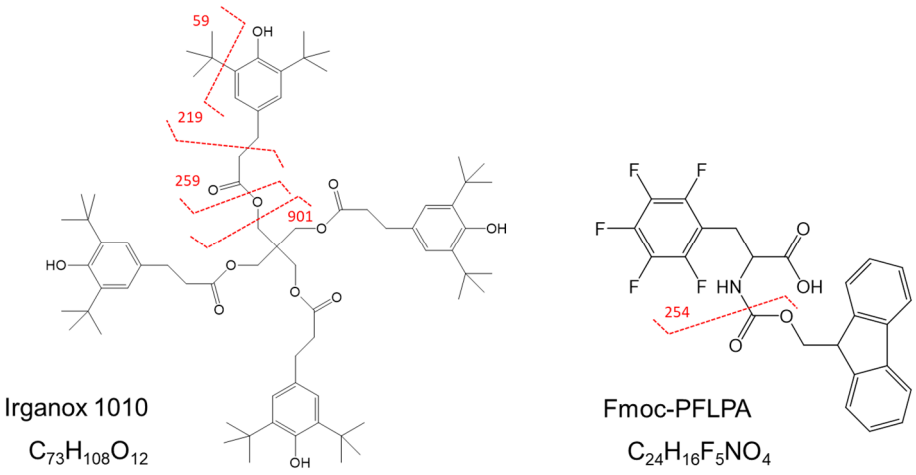


Figure 6.1 Irganox 1010 と Fmoc-PFLPA の分子構造＜赤破線はフラグメントイオン（Table 6-1 参照）の開裂部位＞

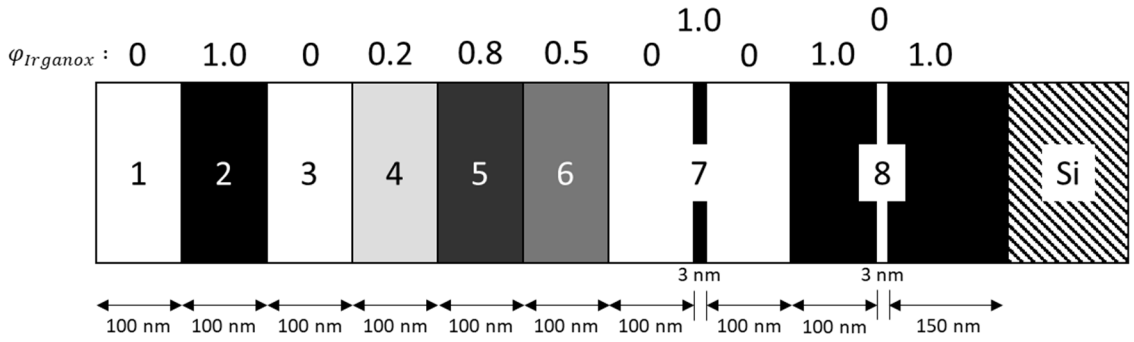


Figure 6.2　積層蒸着膜の構造（$\varphi_{Irganox}$: Irganox 1010 の体積分率）

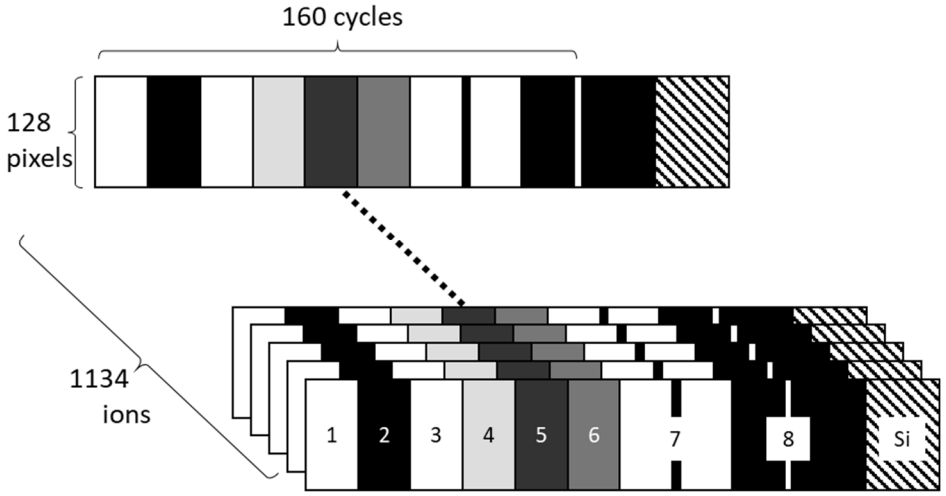


Figure 6.3　オートエンコーダー解析データの構造（1134 ions×128 pixels×160 cycles）

6.3 データ解析

6.3.1 データ前処理

Figure 6.3 に示した 1134 ions×128 pixels×160 cycles のデータを、1134 ions×20480 datapoints に変形したものを、スパースオートエンコーダーおよび多変量解析（PCA, MCR）の解析データとして使用した。

6.3.2 データ解析条件

ライブラリおよびハードウェアについては第三章の検討と同一である（Table 3.2 参照）。更に、スパースオートエンコーダーの構造としては、第二章と同様に、Figure 3.10 に示したエンコーダーとデコーダーの 2 つの部分から成るシンプルなネットワーク構造を採用した。スパースオートエンコーダーの正則化項については、第四章の検討結果（4.4.5 正則化項の違いが特徴抽出性能に与える影響）より KL-Divergence 正則化を採用した。解析条件を Table 6.1 に示した。

Table 6.1　スパースオートエンコーダーの解析条件

中間層サイズ		10
活性化関数（エンコーダーとデコーダーで共通）		ReLU
損失関数		MSE with KL-Divergence
正則化パラメーター（KL-Divergence）	Target sparsity (p)	0.1
	Weight（λ）	0.1
最適化関数		Adam
バッチサイズ		128
学習回数		1000 epochs
データ前処理		なし

MCR の実行は、MATLAB R2015b（Mathworks, Inc., USA）上で動作する多変量解析ソフトウェアである PLS-toolbox 8.0.2 および MIA-toolbox 2.9.2（Eigenvector Research, Inc., USA）を用いて行った。データ前処理として MCR の解析には Poisson scaling を採用した。また、成分数を 2 および 5 として解析を実施した。

6.4 結果と考察

積層蒸着膜の構成材料である Irganox 1010 と Fmoc-PFLPA に由来する正二次イオンとしては、標準試料の測定データと分子構造より、下表に示すものとが特徴的に観測されることがわかっている。これらの正二次イオン種について強度分布を可視化したものを Figure 6.4 に示した。Irganox 1010 の体積分率（$\varphi_{\mathrm{Irganox}}$）が異なる層 1 から層 8 において、$\varphi_{\mathrm{Irganox}}$ の値に応じて強度が変化している様子が認められた。

Table 6.2　Irganox 1010 と Fmoc-PFLPA に特徴的な正二次イオン種

化合物	質量（m/z）	正二次イオン
Irganox 1010	59	$C_4H_9^+$
	219	$C_{15}H_{23}O^+$
	259	$C_{17}H_{23}O_2^+$
	899	$C_{56}H_{83}O_9^+$
Fmoc-PFLPA	256	$C_9F_5H_7NO_2^+$
	476	$C_{24}F_5H_{15}NO_4^+$ ($[M\text{-}H]^+$)

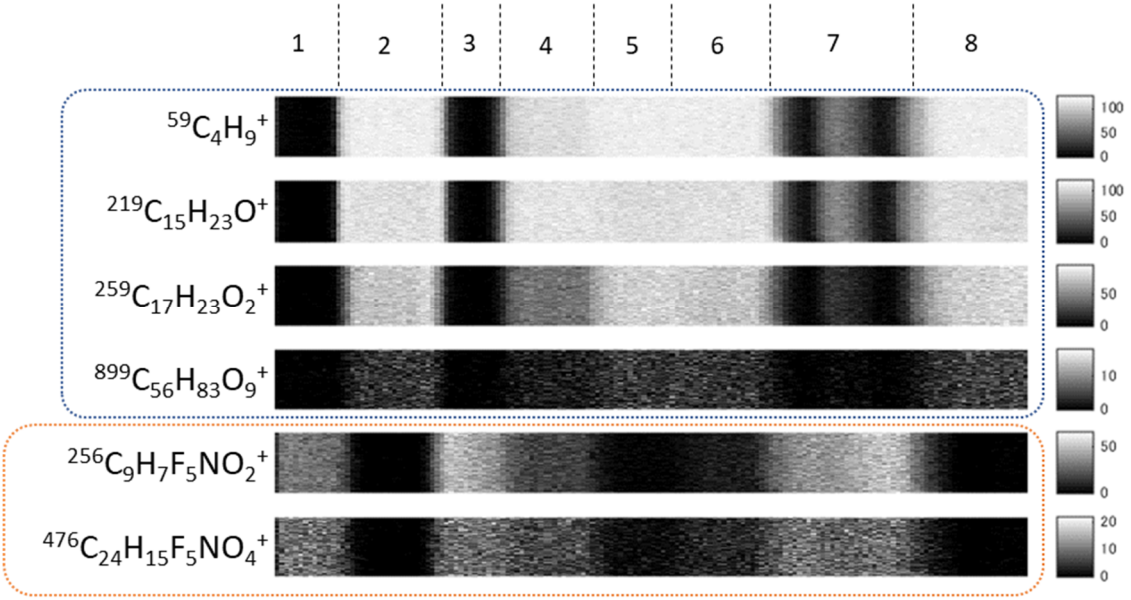


Figure 6.4　Irganox 1010 および Fmoc-PFLPA に特徴的な正二次イオンの深さ方向の強度分布

$C_4H_9^+$, $C_{15}H_{23}O^+$, $C_{17}H_{23}O_2^+$, $C_{56}H_{83}O_9^+$: Irganox 1010

$C_9H_7F_5NO_2^+$, $C_{24}H_{15}F_5NO_4^+$: Fmoc-PFLPA

Irganox 1010 の体積分率（$\varphi_{Irganox}$）　層 1, 3：0、層 2：1.0、層 4：0.2、層 5：0.8、層 6：0.5

体積分率（$\varphi_{Irganox}$）の変化に応じて、各成分に特徴的な正二次イオンの強度がどのように変化しているのかを明確にするために、Table 6.2 に示したイオン種について層 2～層 6（$\varphi_{Irganox}$ = 0, 0.2, 0.5, 0.8, 1.0）の各層について平均強度を求めた。なお、イオン種間での比較を容易にするために、$\varphi_{Irganox}$ が 0 および 1 における平均強度をそれぞれ I_0, I_1 とし、以下の式によりN_φ（規格化された相対強度）を算出し、$\varphi_{Irganox}$ に対してプロットした（Figure 6.5）。

$$N_\varphi = \frac{I_\varphi - I_0}{I_1 - I_0}$$

N_φの 0 から 1 を繋いだ破線は、マトリックス効果が存在しない場合の各イオンの強度の推移を示している。この「理想状態」に対して実際のデータ（プロット）を見ると、Fmoc-PFLPA 由来の二次イオン（$C_9F_5H_7NO_2^+$, $C_{24}F_5H_{15}NO_4^+$）については$\varphi_{Irganox}$ が 0.2～0.8 の範囲においてN_φが理想値より大きいことから、マトリックス効果により強度が「増強」されていることを示している。逆に Irganox 1010 由来のイオン（$C_4H_9^+$, $C_{15}H_{23}O^+$, $C_{17}H_{23}O_2^+$, $C_{56}H_{83}O_9^+$）については、$\varphi_{Irganox}$ が 0.2～0.8 の範囲において強度が破線を下回っていることから、マトリックス効果により強度が「抑制」されていることを示している。

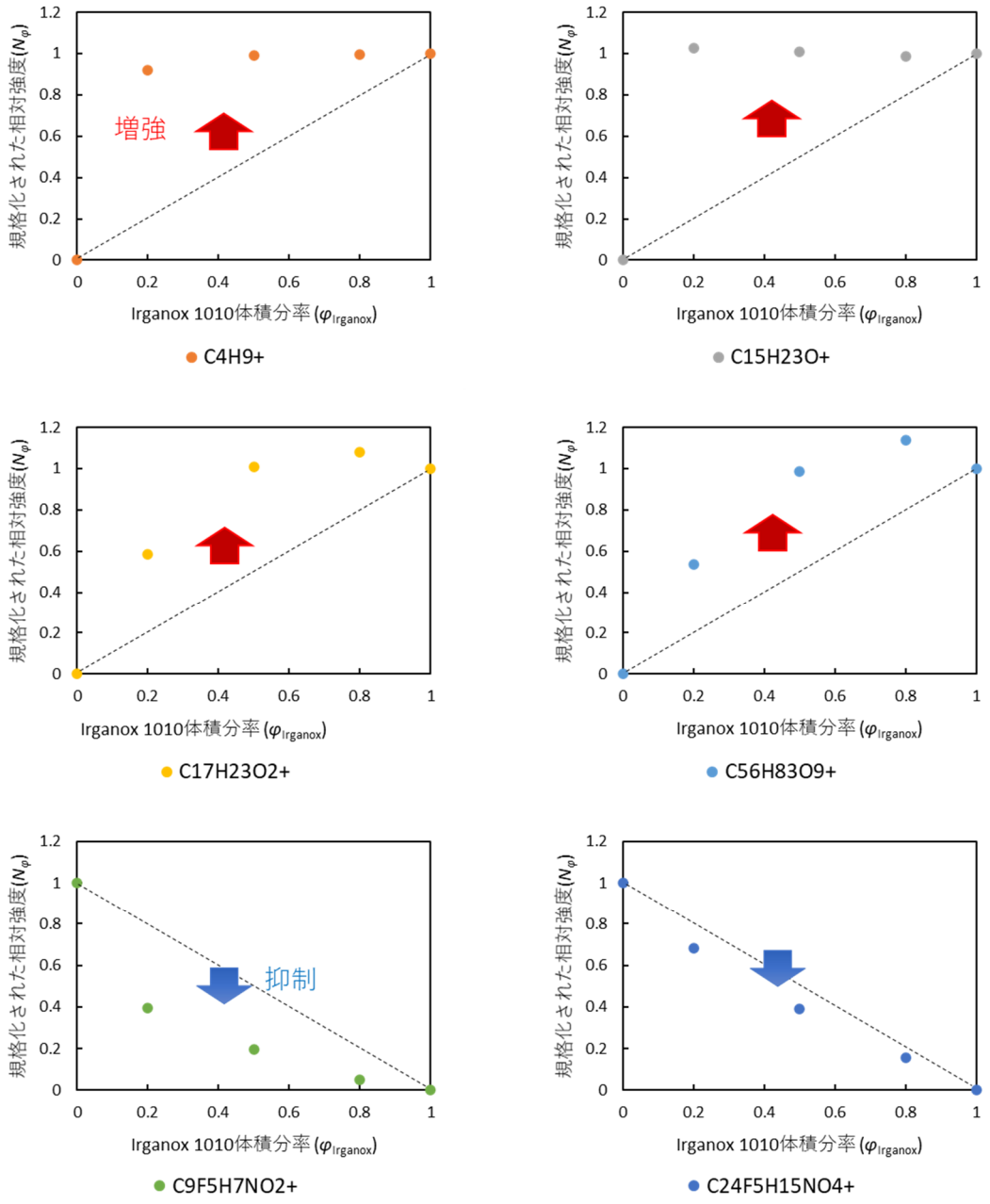


Figure 6.5　各正二次イオンの各層ににおける平均強度＜$\varphi_{Irganox}$=0（層 3）における強度を 0、$\varphi_{Irganox}$=1.0（層 2）における強度を 1 として規格化した強度。破線はマトリックス効果が存在しない理想状態における$\varphi_{Irganox}$と規格化強度との関係を示す。＞

このようなマトリックス効果の影響が大きいデータに対して、スパースオートエンコーダーによる特徴抽出を行った結果を Figure 6.6 に示した。中間層のサイズ 10 に対して 6 個のニューロンに特徴が抽出された（$\hat{X}1, \hat{X}3,\ \hat{X}8, \hat{X}9$については 0 が出力された）。このうち、$\hat{X}4$と$\hat{X}7$に$\varphi_{\mathrm{Irganox}}$ の変化に対応したコントラストが顕著に認められた。そこで、$\hat{X}4$と$\hat{X}7$について正二次イオン種（Figure 6.5）の場合と同様に層 2～6 の各層について規格化された相対強度（N_{φ}）を算出し、$\varphi_{\mathrm{Irganox}}$ に対してプロットした（Figure 6.7）。$\hat{X}4$の結果を、マトリックス効果による増強が認められた Irganox 1010 由来のイオン（$C_4H_9^+$, $C_{15}H_{23}O^+$, $C_{17}H_{23}O_2^+$, $C_{56}H_{83}O_9^+$）の結果（Figure 6.5）と比較すると、$\varphi_{\mathrm{Irganox}}$＝0.2, 0.5 において理想値（破線）からの乖離が顕著に減少した。また$\hat{X}7$の結果をマトリックス効果による抑制が認められた Fmoc-PFLPA 由来の二次イオン（$C_9F_5H_7NO_2^+$, $C_{24}F_5H_{15}NO_4^+$）の結果と比較すると、理想値との乖離の減少が認められた。この結果より、スパースオートエンコーダーがマトリックス効果による二次イオン強度の増感および抑制を補正できる性能を有していることが示唆された。

同様の検討を MCR についても実施した。成分数を 2 および 5 として MCR により抽出された特徴を可視化した図を Figure 6.8、6.9 に示した。また層 2～6 の各層の平均強度を用いて、N_{φ}を$\varphi_{\mathrm{Irganox}}$ に対してプロットした図を Figure 6.10, 6.11 に示した。成分数 2 の結果（Figure 6.10）では、マトリックス効果による増強が認められた Irganox 1010 については、Comp.1 の結果において全組成範囲にて理想値（破線）からの乖離が顕著に減少した。しかしながらマトリックス効果による抑制が認められた Fmoc-PFLPA については、Comp.2 の結果において増強される方向に理想値からの乖離が増加した。成分数を 5 として解析した場合においても、Fmoc-PFLPA の分布を反映した特徴（Comp. 4）は理想値からの乖離が大きい。

TOF-SIMS データから二次イオン強度をそのまま用いた場合と、スパースオートエンコーダーおよび MCR により抽出された特徴を用いた場合のそれぞれで、マトリックス効果の影響を詳細に比較するために、次式で表される残差平方和（RSS）をそれぞれ算出し、理想値からの乖離の程度を比較した（Figure 6.12, 6.13）。

$$RSS = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

n: データ点数

y_i: 実際の値、　$\hat{y}_i$: 理想値（マトリックス効果による増強・抑制なし）

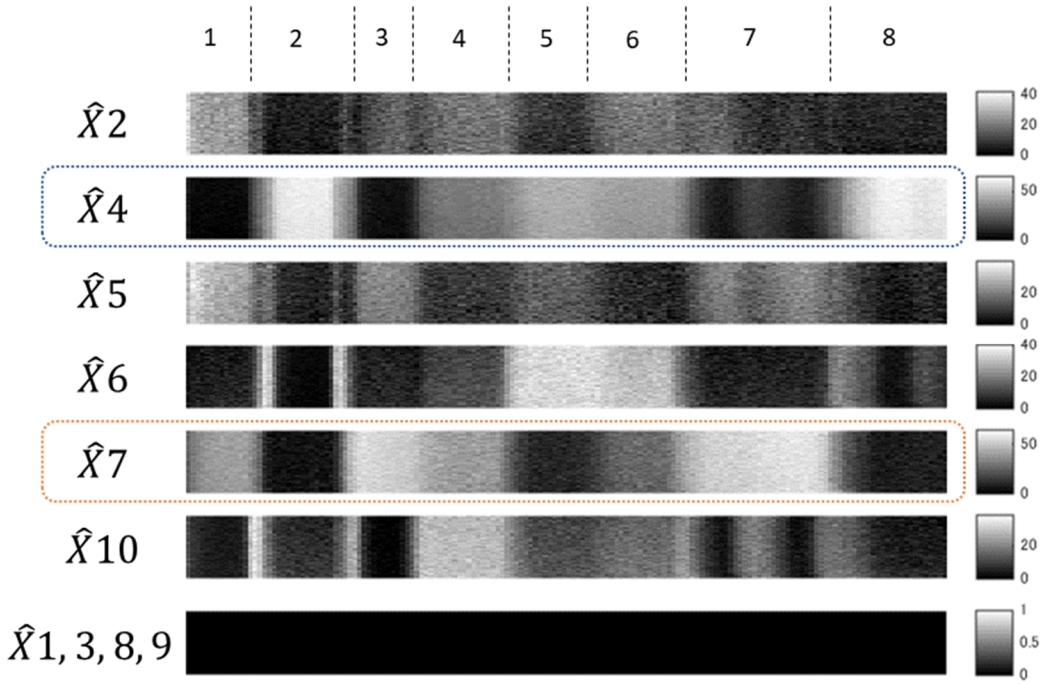


Figure 6.6　スパースオートエンコーダー（KL-Divergence 正則化）によって抽出された特徴

$\hat{X}_4$: Irganox 1010 の分布を反映した特徴、$\hat{X}_7$: Fmoc-PFLPA の分布を反映した特徴
Irganox 1010 の体積分率（$\varphi_{\mathrm{Irganox}}$） 層 1, 3：0、層 2：1.0、層 4：0.2、層 5：0.8、層 6：0.5

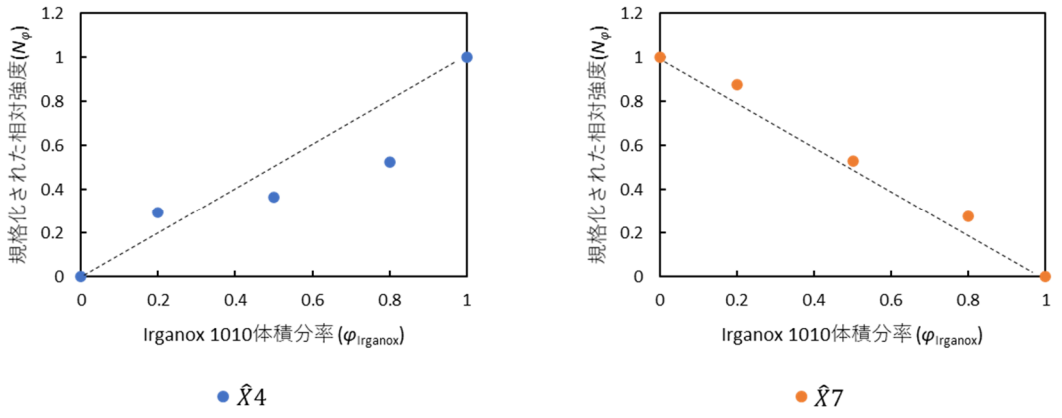


Figure 6.7　スパースオートエンコーダー（KL-Divergence 正則化）によって抽出された特徴における各層の平均強度＜ $\varphi_{\mathrm{Irganox}}$=0（層 3）における強度を 0、$\varphi_{\mathrm{Irganox}}$=1.0（層 2）における強度を 1 として規格化した強度。破線はマトリックス効果が存在しない理想状態における$\varphi_{\mathrm{Irganox}}$と規格化強度との関係を示す。＞

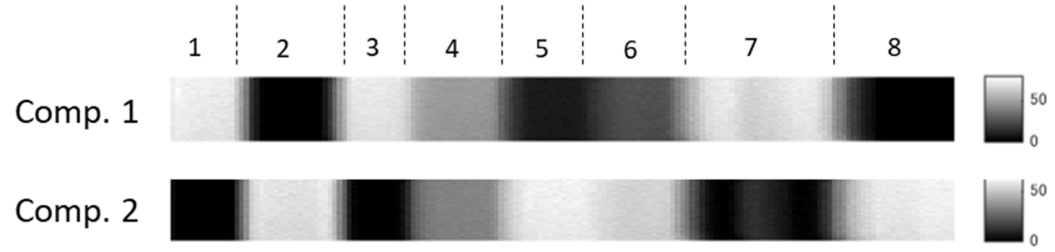


Figure 6.8　MCR（成分数：2）によって抽出された特徴

Comp. 1: Fmoc-PFLPA の分布を反映した特徴

Comp. 2: Irganox 1010 の分布を反映した特徴

Irganox 1010 の体積分率（$\varphi_{\mathrm{Irganox}}$） 層 1, 3：0、層 2：1.0、層 4：0.2、層 5：0.8、層 6：0.5

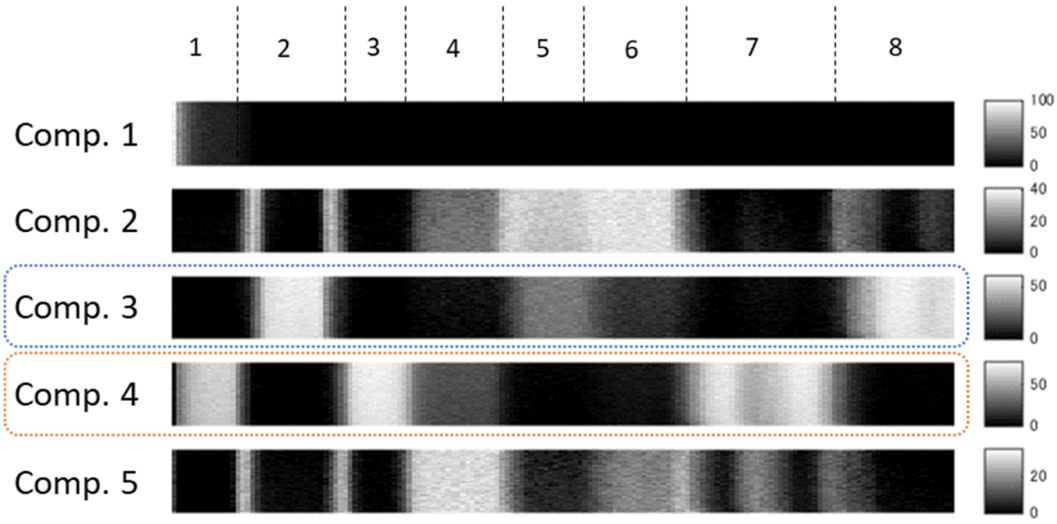


Figure 6.9　MCR（成分数：5）によって抽出された特徴

Comp. 3: Irganox 1010 の分布を反映した特徴

Comp. 4: Fmoc-PFLPA の分布を反映した特徴

Irganox 1010 の体積分率（$\varphi_{\mathrm{Irganox}}$） 層 1, 3：0、層 2：1.0、層 4：0.2、層 5：0.8、層 6：0.5

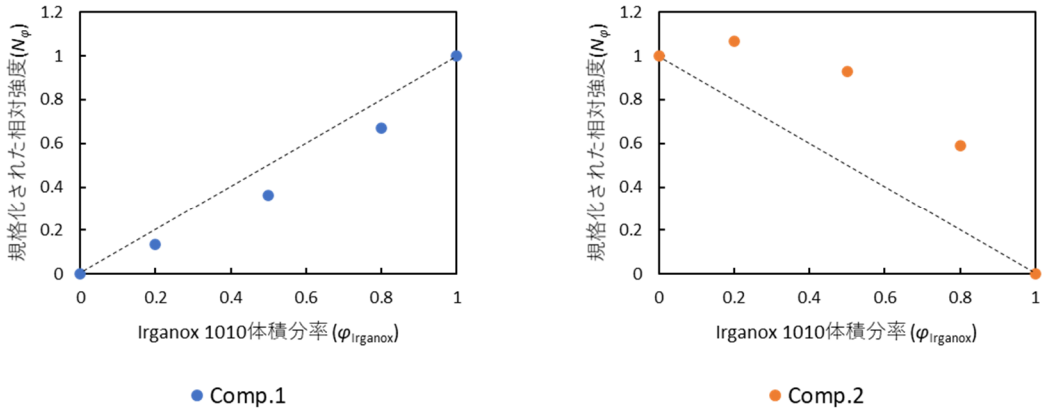


Figure 6.10　MCR（成分数：2）によって抽出された特徴における各層の平均強度＜ $\varphi_{Irganox}$=0（層 3）における強度を 0、$\varphi_{Irganox}$=1.0（層 2）における強度を 1 として規格化した強度。破線はマトリックス効果が存在しない理想状態における$\varphi_{Irganox}$と規格化強度との関係を示す。＞

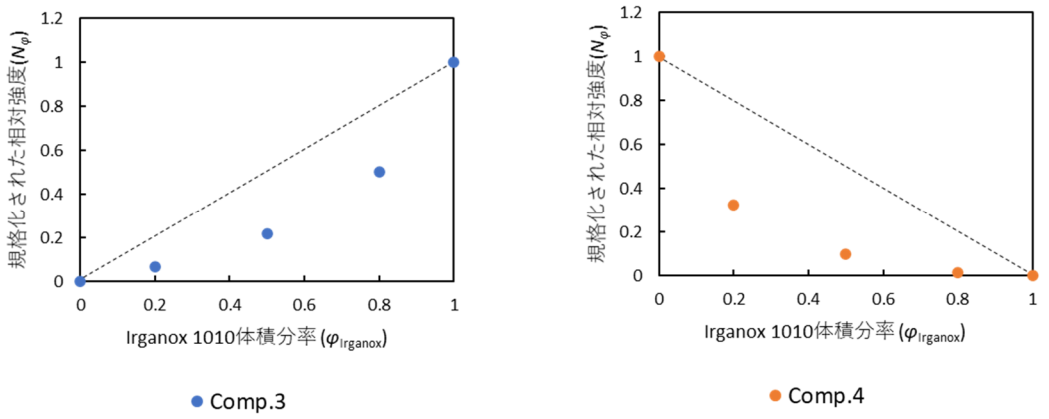


Figure 6.11　MCR（成分数：5）によって抽出された特徴における各層の平均強度＜ $\varphi_{Irganox}$=0（層 3）における強度を 0、$\varphi_{Irganox}$=1.0（層 2）における強度を 1 として規格化した強度。破線はマトリックス効果が存在しない理想状態における$\varphi_{Irganox}$と規格化強度との関係を示す。＞

Table 6.3　Irganox 1010 関連の二次イオンおよび特徴の理想値からの乖離の程度

$\varphi_{Irganox}$		20	50	80	残差平方和(RSS)
理想値		0.2	0.5	0.8	0
二次イオン強度	$C_4H_9^+$	0.919	0.990	0.995	0.795
	$C_{15}H_{23}O^+$	1.02	1.01	0.985	0.972
	$C_{17}H_{23}O_2^+$	0.586	1.01	1.08	0.485
	$C_{56}H_{83}O_9^+$	0.536	0.988	1.14	0.465
ｽﾊﾟｰｽｵｰﾄｴﾝｺｰﾀﾞｰ		0.291	0.365	0.523	0.103
MCR(成分数:2)		0.133	0.365	0.670	0.040
MCR(成分数:5)		0.069	0.224	0.515	0.175

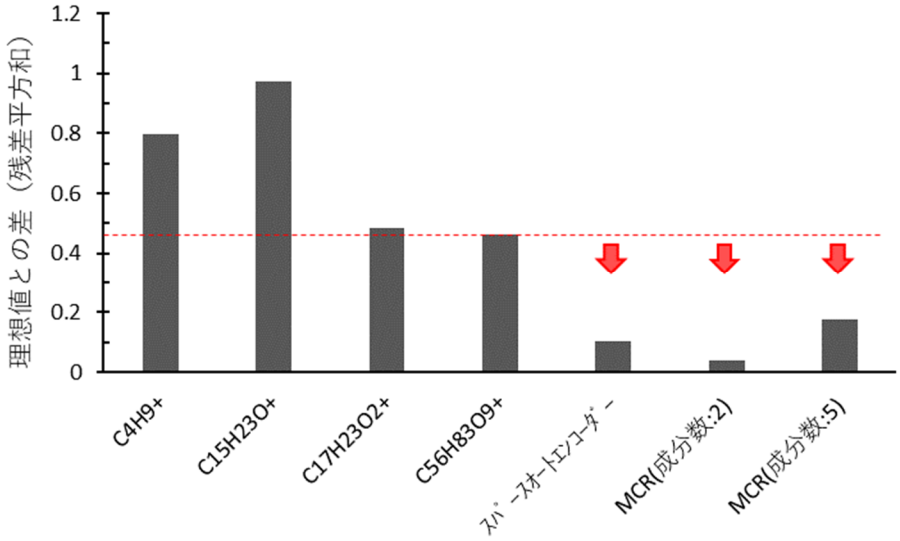


Figure 6.12　　Table 6-3 の残差平方和（RSS）のグラフ化

＜赤線は、最も理想値との差が小さい二次イオンである $C_{56}H_{83}O_9^+$ の値を表す＞

Table 6.4　Fmoc-PFLPA 関連の二次イオンおよび特徴の理想値からの乖離の程度

$\varphi_{Irganox}$		20	50	80	残差平方和(RSS)
理想値		0.8	0.5	0.2	0
二次イオン強度	$C_9F_5H_7NO_2^+$	0.395	0.195	0.0485	0.280
	$C_{24}F_5H_{15}NO_4^+$	0.683	0.393	0.158	0.0270
ｽﾊﾟｰｽｵｰﾄｴﾝｺｰﾀﾞｰ		0.874	0.527	0.274	0.0118
MCR(成分数:2)		1.07	0.929	0.590	0.407
MCR(成分数:5)		0.324	0.0864	0.00920	0.434

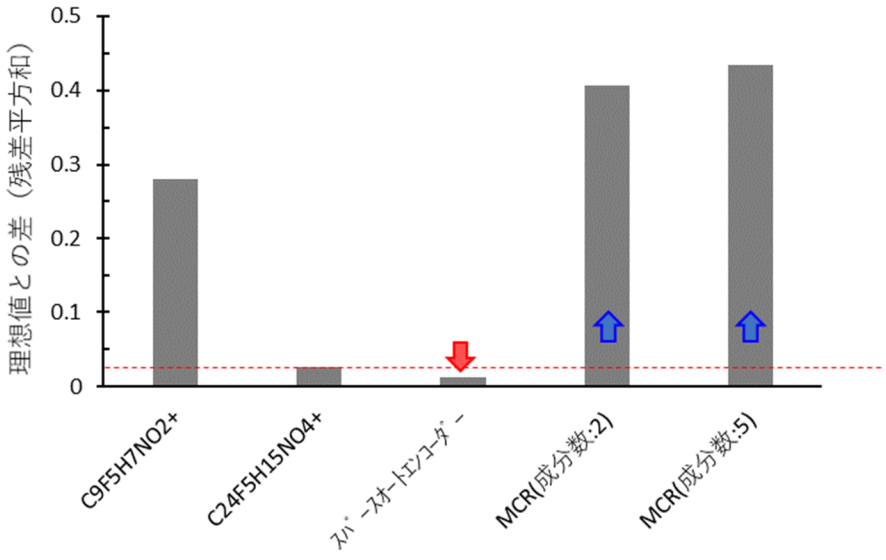


Figure 6.13　Table 6-4 の残差平方和（RSS）のグラフ化

＜赤線は、最も理想値との差が小さい二次イオンである $C_{24}F_5H_{15}NO_4^+$ の値を表す＞

残差平方和（RSS）を比較すると、スパースオートエンコーダーによって抽出された特徴の RSS は、最もマトリックス効果の影響が小さな二次イオン（Irganox 1010 : $C_{56}H_{83}O_9^+$、Fmoc-PFLPA : $C_{24}F_5H_{15}NO_4^+$）と比較して、Irganox 1010 で 20 %、Fmoc-PFLPA で 43 %であり、どちらの成分においてもマトリックス効果の影響の軽減効果が認められた。MCR（成分数：2）は Irganox 1010 で 8.6%、 Fmoc-PFLPA で 1500 %であり、Irganox 1010 についてはスパースオートエンコーダーより RSS が小さいが、Fmoc-PFLPA について RSS が非常に大きな値を取り、濃度依存的な評価は困難である。MCR（成分数：5）の場合でにおいても、RSS は Irganox 1010 で 37 %、Fmoc-PFLPA で 1600 %であり、成分数を大きく設定しても、RSS の値の改善は認められなかった。

ここで、Shard らは、増強の場合は下記の（1）式を、抑制の場合は（2）式を適用することで、良好な補正が可能となることを報告している[91]。これらの式は、標準試料が用意できる試料の場合はマトリックス効果補正は有効である。

$$I_A(\varphi_a) = \varphi_a I_A(1) + \varphi_b[I_A(0) + I_A(1)\alpha\{1 - exp(-\beta\varphi_a)\}] \qquad (1)$$

$$I_A(\varphi_a) = \varphi_b I_A(0) + \varphi_a I_A(1)[1 - \alpha\{1 - exp(-\beta\varphi_b)\}] \qquad (2)$$

A, B: 二種の有機成分
成分 A、成分 B の体積分率: φ_a, φ_b $(\varphi_a + \varphi_b = 1)$
成分 A、成分 B の由来の二次イオンの強度: I_A, I_B
α, β: 組成比既知の標準試料から求めるフィッティングパラメーター

MCR では原理的に元データは抽出された特徴の線形結合で表されることから、元データにマトリックス効果の影響が含まれている以上、マトリックス効果を受けた純成分の情報が抽出されると考えられる。今回の結果において、各試料の代表的な二次イオン強度から示された通り、マトリックス効果が小さい Irganox 1010 については RSS が減少した一方で、大きなマトリックス効果を受ける Fmoc-PFLPA については RSS が増加したことは、この考察を支持するものと考えられる。一方で、スパースオートエンコーダーによってマトリックス効果の影響が軽減された理由としては、次のように考えられる。すなわち、TOF-SIMS データで主要な二次イオンピーク（強度が比較的強いピーク）では非線形性が認められるものの、微小な強度のピークも含めると、組成に応じて線形の応答を示しているピークも多く存在すると予想される。オートエンコーダーでは非線形近似が可能であることから、マトリックス効果の影響を受けたデータの中から Irganox 1010 と Fmoc-PFLPA の両成分について、組成に線形応答する特徴を抽出できた可能性がある。ただし、スパースオートエンコーダーのマトリックス効果補正効果について議論するには、更なる検証が必要と思われる。

6.5 結論

本章では、二種の低分子量化合物から成る有機積層蒸着膜の分析データを用いることで、スパースオートエンコーダーによる特徴抽出結果にマトリックス効果の影響がどのように表れるか検証した。

組成比（体積分率）の異なる各層における、各構成成分（Irganox 1010 と Fmoc-PFLPA）に帰属される特徴の強度データを、マトリックス効果のない理想状態と比較した。その結果、スパースオートエンコーダーを用いた場合、二成分共にマトリックス効果の影響を低減し、理想状態との乖離が小さい特徴抽出結果が得られることが確認された。この結果より、前章までに実施した生体試料の TOF-SIMS データからの特徴抽出においても、マトリックス効果の影響を小さく抑え、構成成分の深さ方向・面内方向の分布状態を、より実際の分布に近く抽出できているもの考えられる。

一方で、MCR による特徴抽出では、二種の構成成分のうちマトリックス効果の影響が小さい成分（Irganox 1010）については、元の TOF-SIMS データを反映して理想値からの乖離が少ない特徴を抽出したが、マトリックス効果の影響が大きい成分（Fmoc-PFLPA）に関しては、元の TOF-SIMS データと同様にマトリックス効果の影響がより顕著に表れる特徴が抽出された。そのため、定量的な評価を行う場合は、MCR の結果に対して、Shard らの提案している補正式を参考にマトリックス効果補正を行うことが望ましいと考えられる。

第七章

結論

本論文では、近年、分析需要が増加している生体試料の組成イメージングデータ、中でも TOF-SIMS による質量イメージングデータから、効率よく有益な情報を引き出すための手段として、人工ニューラルネットワークをベースにした特徴抽出手法である、自己符号化器（オートエンコーダー）の有用性の実証、および適切なネットワーク構造（パラメーター設定）の指針について述べた。

第一章では、生体試料の組成イメージングにおける TOF-SIMS の有用性と共に、データ解析法の重要性について述べた。特に人工ニューラルネットワークを用いた解析法としてオートエンコーダーは、非線形データに対応可能である点が、マトリックス効果に起因する非線形性を含む TOF-SIMS データの解析に有用であることを述べた。

第二章では、本研究を行う上での前提知識となる TOF-SIMS の原理やその現状と、オートエンコーダーをはじめとした人工ニューラルネットワークのアルゴリズムとその実装のために必要な要素（関数）について記述した。

第三章では、実際の生体試料へのオートエンコーダーの適用検討として、Ar クラスターイオンによるイオンエッチングを併用して取得したヒト毛髪の TOF-SIMS デプスプロファイルについて、エンコーダーとデコーダーの二つの部分より構成された単純なオートエンコーダーを適用した。オートエンコーダーにより抽出された特徴は、毛髪のキューティクルを構成するアミノ酸比率（シスチン／システイン含有率）の異なるタンパク質のレイヤー構造の存在や、ヘアケア剤中に含まれる界面活性剤が親水性部位を有する CMC（細胞膜複合体）に局在することを示した。これらの結果は、オートエンコーダーによって抽出された特徴が、既知の毛髪の構造情報に照らし合わせて妥当な結果であり、オートエンコーダーの有用性を示した結果と言える。

第四章では、ヒト皮膚表層の角質層の TOF-SIMS イメージデータをモデルデータとし、第三章で使用した単純なオートエンコーダーと、正則化の概念を取り入れたスパースオートエンコーダーによる特徴抽出結果を比較した。それにより、スパースオートエンコーダーは単純なオートエンコーダーに比べて特徴抽出性能が向上し、データ内に含まれる特徴的な空間（面内・深さ）分布を持つ成分（群）を、他の分布を持つ成分と分離して個別の特徴として抽出する傾向があることが確認された。その要因としては、中間層サイズを大きくして設定したうえで、正則化によって学習過程で中間層をスパース（疎）にする方が、中間層サイズを小さくした場合や正則化を行わない場合に比べて、ネットワークの表現力を維持しつつ特徴を少数の中間層ニューロンに抽出できるためと考えた。さらに正則化項の種類（L1 正則化、KL-Divergence 正則化）やその程度、学習時のバッチサイズが特徴抽出結果に与える影響についても調べ、パラメーター設定の指針を得た。具体的には、正則化の程度については実際に出力された中間層のスパース性を考慮して複数の値で比較することが、最適な結果を得るためには重要であり、バッチサイズについては 64～1024 の範囲で設定することが望ましいと考えらえた。

第五章ではスパースオートエンコーダーと、従来法である PCA、MCR の比較を行った。スパースオートエンコーダーと MCR は類似した特徴抽出性能を示し、分析試料中に含まれる成分の特徴的な分布を情報として個別に抽出したい場合においては、PCA よりも有用であることを確認した。一方で、PCA はスパースオートエンコーダーと MCR に比べ計算コストが低いことから、データの概

要を迅速に把握するうえでは有用であることを確認した。

第六章では、マトリックス効果の影響が強く現れた、二種の低分子量化合物から成る有機積層蒸着膜の分析データについて、スパースオートエンコーダーによる特徴抽出を行い、抽出された特徴においてマトリックス効果の影響がどのように表れるか検証を行った。マトリックス効果のない理想的な状態との比較から、スパースオートエンコーダーによって抽出された特徴は、マトリックス効果の影響を顕著に軽減し、定量的な評価に使用できる可能性があることを示した。これはスパースオートエンコーダーが人工ニューラルネットワークをベースとしており、非線形近似に対応していることが要因と考えられた。

以上の検討結果から、スパースオートエンコーダーは TOF-SIMS データの解析に有用な手法として、主要な位置を占める潜在的な能力を持つことが示された。

人工ニューラルネットワークに基づくオートエンコーダーは、従来より TOF-SIMS データからの特徴抽出法として使用されてきた多変量解析手法（PCA や MCR）に比べて、非線形近似が可能であるという原理的に優れた点がある。しかしながら、オートエンコーダーによる TOF-SIMS データからの特徴抽出に関しては、本研究以前にはほとんど検討されてきていない。類似の検討として、脳組織の MALDI-TOF-MS イメージデータについて、オートエンコーダーにより特徴抽出を検討した事例が報告されてはいたものの、重み（W）パラメーターの解析による特徴量の解釈や、ネットワークパラメーター（中間層サイズ、正則化、バッチサイズなど）が特徴抽出の性能に与える影響の評価については本研究にて初めてなされたものである。今後、益々の発展がなされていくであろう人工ニューラルネットワークによるアルゴリズムを、TOF-SIMS をはじめとした質量イメージングデータの解析に応用していくうえで、本研究によって得られた知見は基礎的な知見として意義があるものと考えられる。今後、更に多くの種類のデータに対して検討を行い、正則化の程度について汎用的な値（推奨値）を求めたり、ネットワーク構造のチューニングを行うことを予定している。それにより、誰もが簡易に使用可能な汎用的な解析手法とすることが、分析化学として重要であると考えている。

また、人工ニューラルネットワークについては日進月歩で進化を続けており、アルゴリズムのほかネットワークを構成する各関数のような要素技術も新しいものが登場してきている。それらを用いて、更なるネットワークの改良・最適化を実施していくことで、オートエンコーダーの特徴抽出性能の継続的な向上が期待される。また、TOF-SIMS データからの特徴抽出の次ステップとして、抽出された各特徴のスペクトルパターンを事前に収集された標準試料データのスペクトルパターンとマッチングさせることについても、人工ニューラルネットワークをはじめとした機械学習アルゴリズムを適用していくことが重要と考えられる。実際に、非常に多くのサンプルについて迅速な定性分析が求められる環境モニタリングにおいて、クロマトグラフィーで単離した成分の質量スペクトルデータを、NIST の標準データベースとマッチングさせることに、機械学習アルゴリズムを適用する検討が行われている [92]。それらの研究内容も踏まえると、最終的には TOF-SIMS データの解析の自動化を視野に入れた研究もおこなわれていくことが期待される。また、TOF-SIMS に限らず機器分析全般において取得される情報量の増加は今後も継続し、大量の情報の活用に AI 技術の活用がより進展していくと予想される。本研究はそれらの研究・技術開発の一助となるものと考えられる。

参考文献

1. 伊藤裕子（2019）.「新しい創薬モダリティとしての核酸医薬の動向」,『STI Horizon』, 5(4) pp.21-25.
2. 松田知己、「続・生物工学基礎講座　バイオよもやま話　蛍光タンパク質―知っておきたい性質―」,『生物工学会誌』, 94(9), pp.555-8.
3. 水島昇、鈴木邦律（2009）.「蛍光顕微鏡データの誤った解釈」,『蛋白質 核酸 酵素』, 54(2)
4. 小川潔、竹内貞夫、出水秀明、原田高宏、吉田佳一、瀬藤光利（2006）.「顕微質量分析装置の開発」,『島津評論』、5(3・4), pp.125-135
5. A. R. Buchberger, K. DeLaney, J. Johnson, L. Li, Mass Spectrometry Imaging: A Review of Emerging Advancements and Future Insights, Anal. Chem. 2018, 90, 1, 240–265
6. Douglas MA, Chen PJ. Quantitative Trace Metal Analysis of Silicon Surface by ToF-SIMS. Surf Interface Anal. 1998;26(13):984-94.
7. Zanderigo F, Ferrari S, Queirolo G, Pello C, Borgini M. Quantitative TOF-SIMS analysis of metal contamination on silicon wafers. Mat Sci Eng B-Solid. 2000;73(1):173-7.
8. Auditore A, Samperi F, Puglisi C, Licciardello A. ToF-SIMS investigation of the thermally induced processes at the surface of polyester based polymer blends. Compos Sci Technol. 2003;63(8):1213-19.
9. Oran U, Ünveren E, Wirth T, Unger WES. Poly-dimethyl-siloxane (PDMS) contamination of polystyrene (PS) oligomers samples: a comparison of time-of-flight static secondary ion mass spectrometry (TOF-SSIMS) and X-ray photoelectron spectroscopy (XPS) results. Applied Surface Science. 2004;227(1-4):318-24.
10. Belu AM, Graham DJ, Castner DG. Time-of-flight secondary ion mass spectrometry: techniques and applications for the characterization of biomaterial surfaces. Biomaterials. 2003;24(21):3635-53.
11. Kessler L, Legeay G, Coudreuse A, Bertrand P, Poleunus C, Vanden Eynde X, Mandes K, Marchetti P, Pinget M, Belcourt A. Surface treatment of polycarbonate films aimed at biomedical application. J Biomater Sci, Polym Ed. 2003;14(10):1135-53.
12. Kingshott P, Andersson G, McArthur SL, Griesser HJ. Surface modification and chemical surface analysis of biomaterials. Curr Opin Chem Biol. 2011;15(5):667-76.
13. Benabdellah F, Seyer A, Quinton L, Touboul D, Brunelle A, Laprevote O. Mass spectrometry imaging of rat brain sections: nanomolar sensitivity with MALDI versus nanometer resolution by TOF-SIMS. Anal Bioanal Chem. 2010;396(1):151-62.
14. Brunelle A, Touboul D, Laprevote O. Biological tissue imaging with time-of-flight secondary ion mass spectrometry and cluster ion sources. J Mass Spectrom. 2005;40(8):985-99.
15. Kim JH, Kim JH, Ahn BJ, Park JH, Shon HK, Yu YS, Moon DW, Lee TG, Kim KW. Label-free calcium imaging in ischemic retinal tissue by TOF-SIMS. Biophys J. 2008;94(10):4095-102.

16. Kojima T, Yamada H, Yamamoto T, Matsushita Y, Fukushima K. Dyeing regions of oxidative hair dyes in human hair investigated by nanoscale secondary ion mass spectrometry. Colloids Surf B Biointerfaces. 2013;106:140-4.
17. Mains J, Wilson CG, Urquhart A. ToF-SIMS analysis of dexamethasone distribution in the isolated perfused eye. Invest Ophthalmol Vis Sci. 2011;52(11):8413-9.
18. Sjövall P, Rossmeisl M, Hanrieder J, Kuda O, Kopecky J, M. B. Dietary uptake of omega-3 fatty acids in mouse tissue studied by time-of-flight secondary ion mass spectrometry (TOF-SIMS). Anal Bioanal Chem. 2015;407(17):5101-11.
19. Fletcher JS. Latest applications of 3D ToF-SIMS bio-imaging. Biointerphases. 2015;10(1):018902.
20. Passarelli MK, Winograd N. Characterizing in situ Glycerophospholipids with SIMS and MALDI Methodologies. Surf Interface Anal. 2011;43(1-2):269-71.
21. Weibel DE, Lockyer N, Vickerman JC. C_{60} cluster ion bombardment of organic surfaces. Appl Surf Sci. 2004;231-232:146-52.
22. Angerer TB, Velickovic D, Nicora CD, Kyle JE, Graham DJ, Anderton C, Gamble LJ. Exploiting the Semidestructive Nature of Gas Cluster Ion Beam Time-of-Flight Secondary Ion Mass Spectrometry Imaging for Simultaneous Localization and Confident Lipid Annotations. Anal Chem. 2019;91(23):15073-80.
23. Fujii M, Nakagawa S, Matsuda K, Man N, Seki T, Aoki T, Matsuo J. Study on the detection limits of a new argon gas cluster ion beam secondary ion mass spectrometry apparatus using lipid compound samples. Rapid Commun Mass Spectrom. 2014;28(8):917-20.
24. Mohammadi AS, Phan NT, Fletcher JS, Ewing AG. Intact lipid imaging of mouse brain samples: MALDI, nanoparticle-laser desorption ionization, and 40 keV argon cluster secondary ion mass spectrometry. Anal Bioanal Chem. 2016;408(24):6857-68.
25. Dimovska Nilsson K, Karagianni A, Kaya I, Henricsson M, Fletcher JS. $(CO_2)_n^+$, $(H_2O)_n^+$, and $(H_2O)_n^+$ (CO_2) gas cluster ion beam secondary ion mass spectrometry: analysis of lipid extracts, cells, and Alzheimer's model mouse brain tissue. Anal Bioanal Chem. 2021;413(16):4181-94.
26. Razo IB, Sheraz S, Henderson A, Lockyer NP, Vickerman JC. Mass spectrometric imaging of brain tissue by time-of-flight secondary ion mass spectrometry - How do polyatomic primary beams C_{60}^+, Ar_{2000}^+, water-doped Ar_{2000}^+ and $(H_2O)_{6000}^+$ compare? Rapid Commun Mass Spectrom. 2015;29(20):1851-62.
27. Lee JLS, Gilmore IS, Fletcher IW, Seah MP. Multivariate image analysis strategies for ToF-SIMS images with topography. Surf Interface Anal. 2009;41(8):653-65.
28. Wagner MS, Castner DG. Characterization of Adsorbed Protein Films by Time-of-Flight Secondary Ion Mass Spectrometry with Principal Component Analysis. Langmuir. 2001;17(15):4649-60.

29. Wagner MS, Graham DJ, Castner DG. Simplifying the interpretation of ToF-SIMS spectra and images using careful application of multivariate analysis. Appl Surf Sci. 2006;252(19):6575-81.
30. Aoyagi S, Matsuzaki T, Takahashi M, Sakurai Y, Kudo M. Evaluation of reagent effect on skin using time-of-flight secondary ion mass spectrometry and multivariate curve resolution. Surf Interface Anal. 2012;44(6):772-5.
31. Lai H, Liu Q, Deng J, Wen S, Liu Z. Surface chemistry study of Cu-Pb sulfide ore using ToF-SIMS and multivariate analysis. Appl Surf Sci. 2020;518.
32. Hinton GE, Salakhutdinov RR. Reducing the Dimensionality of Data with Neural Networks. 2006;313(5786):504-507.
33. ImageNet Large Scale Visual Recognition Competition 2012（ILSVRC2012）. http://image-net.org/challenges/LSVRC/2012/results.html
34. Almuqhim F, Saeed F. ASD-SAENet: A Sparse Autoencoder, and Deep-Neural Network Model for Detecting Autism Spectrum Disorder (ASD) Using fMRI Data. Front Comput Neurosci. 2021;15:654315.
35. Aslam MA, Xue C, Chen Y, Zhang A, Liu M, Wang K, Cui D. Breath analysis based early gastric cancer classification from deep stacked sparse autoencoder neural network. Sci Rep. 2021;11(1):4014.
36. Huang G, Yuan L-M, Shi W, Chen X, Using one-class autoencoder for adulteration detection of milk powder by infrared spectrum, Food Chem., 2022;372(15):131219.
37. Kawashima T, Aoki T, Taniike Y, Aoyagi S. Examination of beauty ingredient distribution in the human skin by time-of-flight secondary ion mass spectrometry. Biointerphases. 2020;15(3):031013.
38. Thomas SA, Race AM, Steven RT, Gilmore IS, Bunch J. Dimensionality reduction of mass spectrometry imaging data using autoencoders. Proceedings of the 2016 IEEE Symposium Series on Computational Intelligence (SSCI). 2016.
39. Matsuda K, Aoyagi S. Time-of-flight secondary ion mass spectrometry analysis of hair samples using unsupervised artificial neural network. Biointerphases. 2020;15(2):021013.
40. Madiona RMT, Bamford SE, Winkler DA, Muir BW, Pigram PJ, Distinguishing chemically similar polyamide materials with TOF-SIMS using self-organizing maps and universal data matrix. Anal. Chem. 2018;90:12475-84.
41. Gardner W, Cutts SM, Muir BW, Jones RT, Pigram PJ, Visualizing TOF-SIMS hyperspectral imaging data using color-tagged toroidal self-organizing maps. Anal. Chem. 2019;91:13855-65.
42. 日本表面科学会　編、(1999).「表面分析技術選書　二次イオン質量分析法（第二版）」、丸善
43. Vickerman JC, Briggs D, TOF-SIMS: Surface Analysis by Mass Spectrometry, IM Publication and SurfaceSpectra Limited, Chichester and Manchester. 2001.

44. Benninghoven A., Surface investigation of solids by the statical method of secondary ion mass spectroscopy (SIMS), Surf. Sci. 1973;35:427-57.
45. D. Briggs, M. P. Seah 編、清水隆一、二瓶好正 監訳、(2003).「表面分析:SIMS―二次イオン質量分析法の基礎と応用―(第一版)」、アグネ承風社
46. Kollmer F, Paul W, Krehl M, Niehuis E, Ultra high spatial resolution SIMS with cluster ions - approaching the physical limits, Surf. Interface Anal. 2013;45:312-14
47. Marseilhan D, Barnes JP, Fillot F, Hartmann JM, Holliger P, Quantification of SiGe layer composition using MCs^+ and MCs_2^+ secondary ions in ToF-SIMS and magnetic SIMS, Appl. Surf. Sci., 2008;255(4):1412-14.
48. Mine N, Douhard B, Brison J, Houssiau L, Molecular depth-profiling of polycarbonate with low-energy Cs^+ ions, Rapid commun. mass sp.,2007;21(16):2680-4.
49. Wehbe N, Tabarrant T, Brison J, Mouhib T, Delcorte A, Bertrand P, Moelllers R, Niehuis E, Houssiau L, TOF-SIMS depth profiling of multilayeramino-acid films using large Argon cluster Ar_n^+,C_{60}^+ and Cs^+ sputtering ions: A comparative study. Surf. Interface Anal. 2013;45:178-80.
50. Henss A, Otto SK, Schaepe K, Pauksch L, Lips KS, Rohnke M, High resolution imaging and 3D analysis of Ag nanoparticles in cells with ToF-SIMS and delayed extraction, Biointerphases. 2018;13(31):03B410
51. Kim SH, Lee J, Jang YJ, Lee KB, Lee Y, TOF-SIMS and AFM Characterization of Brown Cosmetic Contact Lenses: From Structural Analysis to the Identification of Pigments, J. Anal. Methods Chem., 2020:6134627
52. Schwieters J, Cramer HG, Heller T, Jöürgens U, Niehuis E, Zehnpfenning J, Benninghoven A, High mass resolution surface imaging with a time-of-flight secondary ion mass pectroscopy scanning microprobe, J. Vac. Sci. Technol., A, 1991;A9(6):2864.
53. Appelhans AD, Delmore JE, Dahl DA, Focused, rasterable, high-energy neutral molecular beam probe for secondary ion mass spectrometry, Anal. Chem., 1987;59(13):1685-91.
54. Postawa Z, Czerwinski B, Szewczyk M, Smiley EJ, Winograd N, Garrison BJ, Enhancement of Sputtering Yields Due to C60 versus Ga Bombardment of Ag{111} As Explored by Molecular Dynamics Simulations, Anal. Chem., 2003;75(17):4402-7.
55. Touboul D, Kollmer F, Niehuis E, Brunelle A, Laprévpte O, Improvement of biological time-of-flight secondary ion mass spectrometry imaging with a bismuth cluster ion source, J. Am. Soc. Mass Spectrom., 2005;16:1608-18.
56. Weibel D, Wong S, Lokyer N, Blenkinsopp P, Hill R, Vickerman JC, A C_{60} primary ion beam system for time of flight secondary ion mass spectrometry: Its development and secondary ion yield characteristics, Anal. Chem., 2003;75(7):1754-64.
57. Ninomiya S, Nakata Y, Ichiki K, Seki T, Aoki T, Matsuo J, Measurements of secondary ions emitted from organic compounds bombarded with large gas cluster ions, Nucl. Instrum. Methods

Phys. Res., 2007;256(1):493-6.
58. Ichiki K, Tamura J, Seki T, Aoki T, Matsuo J, Development of gas cluster ion beam irradiation system with an orthogonal acceleration TOF instrument, Surf. Interface Anal., 2013;45(1):522-4.
59. 松尾二郎(2021).「飛躍的な発展を遂げる二次イオン質量分析(SIMS)法の展望 有機・生体高分子分野への新展開」,『表面と真空』, 64(10), pp. 458-65.
60. Passarelli MK, Pirkl A, Moellers R, Grinfeld D, Kollmer F, Havelund R, Newman CF, Marshall PS, Arlinghaus H, Alexander MR, West A, Horning S, Niehuis E, Makarov A, Dollery CT, Gilmore IS, The 3D OrbiSIMS - Label-free metabolic imaging with subcellular lateral resolution and high mass-resolving power, Nat. Methods, 2017;14(12):1175-83.
61. Winograd N, The magic of cluster SIMS, Anal. Chem, 2005;77(7):142-49.
62. The MathWorks, Inc.,『ディープラーニングと従来の機械学:適切なアプローチの選択 』https://jp.mathworks.com/content/dam/mathworks/ebook/gated/jp-deep-learning-vs-machine-learning-ebook.pdf
63. Géron A, Hands-on Machine Learning with Scikit-Learn, Keras & TensorFlow, Second edition, O'Reilly Media Inc., 2019.
64. Aurelian Geron 著、下田倫大 監訳、長尾高弘 訳、(2018).『scikit-learn と TensorFlow による実践機械学習(初版)』、オライリージャパン
65. Raschka S, Mirjalili V 著、株式会社クイープ 訳、福島真太朗 監訳、(2020).『達人データサイエンティストによる理論と実践 Python 機械学習プログラミング(第3版)』、インプレス
66. Matsuda K, Aoyagi S, Time-of-flight secondary ion mass spectrometry analysis of hair samples using unsupervised artificial neural network, Biointerphases, 2020;15:021013.
67. Goodfellow IJ, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y, Generative Adversarial Nets, NeurIPS Proceedings, https://proceedings.neurips.cc/paper/2014/file/5ca3e9b122f61f8f06494c97b1afccf3-Paper.pdf
68. 宇田明史、寺田勝英、(2006).「品質管理のための主成分分析 Principal Component Analysis for Quality Control」、『PDA Journal of GMP and Validation in Japan』, 8(2), pp.94-105.
69. Jaumot J, Gargallo R, Juan AD, Tauler R, A graphical user-friendly interface for MCR-ALS: a new tool for multivariate curve resolution in MATLAB, Chemometrics and Intelligent Laboratory Systems, 2005;76:101-10.
70. Advanced Preprocessing: Variable Centering, Eigenvector Research Documentation Wiki; [updated 2021 Octber 24 at 16:23; cited 2021 November 17]. Available from https://wiki.eigenvector.com/index.php?title=Advanced_Preprocessing:_Variable_Centering
71. Conn EE, Stumpf PK, Bruening G, Doi RH 著、田宮信雄、八木達彦 訳、(1991).『生化学(第5版)』、東京化学同人
72. 藤田克昌(2010).「超解像顕微鏡の進展」,『生物物理』, 50(4), pp.174-179.

73. Caprioli RM, Farmer TB, Gile J, Molecular Imaging of Biological Samples: Localization of Peptides and Proteins Using MALDI-TOF MS, Anal. Chem. 1997;69:4751-60.
74. Takáts Z, Wiseman JM, Cooks RG, Ambient mass spectrometry using desorption electrospray ionization (DESI): Instrumentation, mechanisms and applications in forensics chemistry, and biology, J. mass spectrum., 2005;40(10):1261-75.
75. Lechene C, Hillion F, McMahon G, Benson D, Kleinfeld AM, Kampf JP, Distel D, Luyten Y, Bonventre J, Hentschel D, Park KM, Ito S, Schwartz M, Benichou G, Slodzian G, High-resolution quantitative imaging of mammalian and bacterial cells using stable isotope mass spectrometry, J. Biol. 2006;5(6):20.
76. Rovira-Clavé X, Jiang S, Bai Y, Zhu B, Barlow G, Bhate S, Coskun AF, Han G, Ho CMK, Hitzman C, Chen S, Bava FA, Nolan GP, Subcellular localization of biomolecules and drug distribution by high-definition ion beam imaging, Nat. comun. 2021;12(1):4628.
77. Starr NJ, Abdul Hamid K, Wibawa J, Marlow I, Bell M, Perez-Garcia L, Barrett DA, Scurr DJ. Enhanced vitamin C skin permeation from supramolecular hydrogels, illustrated using in situ ToF-SIMS 3D chemical profiling. Int J Pharm. 2019;563:21-9.
78. Cizinauskas V, Elie N, Brunelle A, Briedis V. Fatty acids penetration into human skin ex vivo: A TOF-SIMS analysis approach. Biointerphases. 2017;12(1):011003.
79. Holmes AM, Scurr DJ, Heylings JR, Wan KW, Moss GP. Dendrimer pre-treatment enhances the skin permeation of chlorhexidine digluconate: Characterisation by in vitro percutaneous absorption studies and Time-of-Flight Secondary Ion Mass Spectrometry. Eur J Pharm Sci. 2017;104:90-101.
80. Sjövall P, Skedung L, Gregoire S, Biganska O, Clément F, Luengo GS. Imaging the distribution of skin lipids and topically applied compounds in human skin using mass spectrometry. Sci Rep. 2018;8(1):16683.
81. Starr NJ, Johnson DJ, Wibawa J, Marlow I, Bell M, Barrett DA, Scurr DJ. Age-Related Changes to Human Stratum Corneum Lipids Detected Using Time-of-Flight Secondary Ion Mass Spectrometry Following in Vivo Sampling. Anal Chem. 2016;88(8):4400-8.
82. Ishikawa K, Okamoto M, Aoyagi S, Structural analysis of the outermost hair surface using TOF-SIMS with gas cluster ion beam sputtering, Biointerphases, 2016;11(2):02A315.
83. Jones LN, Rivett DE, The role of 18-methileicosanoic acid in the structure and formation of mammalian hair fibres, Micron, 1997;28(6):469-85.
84. Chollet F, et al., Keras. see https://keras.io (2015).
85. Kingma DP, Ba JL, Adam: A method for Stochastic Optimization, preprint arXiv:1412.6980, 2014.
86. Tidwell CD, Castner DG, Golledge SL, Ratner BD, Meyer K, Hagenhoff B, Benninghoven A, Static time‐of‐flight secondary ion mass spectrometry and x‐ray photoelectron spectroscopy

characterization of adsorbed albumin and fibronectin films, Surf. Interface Anal. 2001;31:724-33.

87. Canavan HE, Graham DJ, Cheng X, Ratner BD, Castner DG, Comparison of Native Extracellular Matrix with Adsorbed Protein Films Using Secondary Ion Mass Spectrometry, Langmuir 2007; 23:50-6.
88. 村越紀之、「蛍光でみる毛髪の構造とダメージ」、J.Soc.Cosmet.Chem.jpn., 2015;49(2):87-94.
89. 駒崎伸二、(2020).『ヴァーチャルスライド　組織学』、羊土社
90. Shard AG, Havelund R, Spencer SJ, Gilmore IS, Alexander MR, Angerer TB, Aoyagi S, Barnes J-P, Benayad A, Bernasik A, Ceccone G, Counsell J.D.P, Deeks C, Fletcher JS, Graham DJ, Heuser　C, Geol Lee T, Marie C, Marzec MM, Mishra G, Rading D, Renault O, Scurr DJ, Kyong Shon H, Spampinato V, Tian H, Wang F, Winograd N, Wu K, Wucher A, Zhou Y, Zhu Z, Measuring Compositions in Organic Depth Profiling: Results from a VAMAS Interlaboratory Study, J. Phys. Chem. B 2015;119(33):10784–97.
91. Shard AG, Spencer SJ, Smith SA, Havelund R, Gilmore IS, The matrix effect in organic secondary ion mass spectrometry, Int. J. mass spectrom., 2015;377:599-609.
92. Bendik J, Kalia R, Sukumaran J, Richardot W.H, Hoh E, Kelley S.T, Automated high confidence compound identification of electron ionizarion mass spectra for nontargeted analysis, J. Chromatogr. A, 2021;1660:462656.

研究業績一覧

【査読付き論文】

I. 筆頭著者として(2 報)

1. **Matsuda K.**, Aoyagi S., Time-of-flight secondary ion mass spectrometry analysis of hair samples using unsuperevised artificial neural network. Biointerphases. 2020;15(2):021013. https://doi.org/10.1116/6.0000044
2. **Matsuda, K.**, Aoyagi, S., Sparse autoencoder-based feature extraction from TOF-SIMS image data of human skin structures. Anal. Bioanal. Chem. 2021;414(2):1177-86. https://doi.org/10.1007/s00216-021-03744-3

II. 共著者として(2 報)

3. Aoyagi S., Fujiwara Y., Takano A., Vorng J.-L., Gilmore I.S., Wang Y.-C., Tallarek E., Hagenhoff B., Iida S.-I., Luch A., Jungnickel H., Lang Y., Shon H.K., Lee T.G., Li Z., **Matsuda K.**, l,Mihara I., Miisho A., Murayama Y., Nagatomi T., Ikeda R., Okamoto M., Saiga K., Tsuchiya T., Uemura S., Evaluation of Time-of-Flight Secondary Ion Mass Spectrometry Spectra of Peptides by Random Forest with Amino Acid Labels: Results from a Versailles Project on Advanced Materials and Standards Interlaboratory Study, Anal. Chem. 2021;93(9):4191-7., https://doi.org/10.1021/acs.analchem.0c04577
4. 伊藤　克、松田　和大、青柳　里果、自己符号化器(autoencoder)を用いた高分子試料の TOF-SIMS データ解析、J. surf. anal. 2021, **(Accepted)**

【学会発表】

I. 国際学会

1. **Matsuda, K.**, Sameshima, J., Aoyagi, S., Characterization of the outermost hair surface using TOF-SIMS depth profiling with multivariate analysis and machine learning. The 22nd International Conference on Secondary Ion Mass Spectrometry (poster), October, 2020, Kyoto, Japan
2. **Matsuda, K.**, Aoyagi, S., Investigation of the feature extraction on TOF-SIMS data of human corneocytes by sparse autoencoder (poster). The 13th International Symposium on Atomic Level Characterization for New Materials and Devices '21, October, 2021, Online

II. 国内学会・研究会

1. 松田　和大、青柳　里果、機械学習を用いた TOF-SIMS データからの特徴量抽出の基礎検討(ポスター)、 学振マイクロビームアナリシス第 141 委員会　第 177 回研究会 2019 年 8 月、兵庫県尼崎市
2. 松田　和大、青柳　里果、ヒト皮膚組織の質量イメージングデータからの機械学習を活用した特徴抽出の検討、2020 年日本表面真空学会　学術講演会(口頭)、2020 年 11 月、オンライン開催

謝辞

まず、社会人博士として受け入れてくださった、成蹊大学理工学部、青柳里果教授に厚く御礼申し上げます。

本研究は株式会社東レリサーチセンターにおいて、青柳里果教授の御指導を受けながら 2019 年より研究を行った内容を中心にまとめたものです。社会人として業務と並行して研究を進める必要があること、また在学時の大半がコロナ禍ということもあり、研究の進捗が十分でない時期もありましたが、そのような状況下においても青柳先生には適宜、WEB 会議などを通じてフォロー頂きました。三年間で無事に研究成果を挙げ、本学位論文をまとめることができたことも、青柳先生の懇切なご指導、ご鞭撻、そして細心のフォローによるものです。誠に感謝しております。

本学位論文の審査において副査を御担当いただきました、成蹊大学理工学部の富谷光良教授、中野武雄教授、名古屋大学未来材料・システム研究所の武藤俊介教授に、厚く御礼申し上げます。富谷教授より賜りました物理シミュレーションに関する御講義、中野教授より賜りました各種の成膜プロセスに関する御講義、どちらの知識についても日々の業務で求められる機会が多くあるものの、これまで十分に学ぶことができていなかった分野であり、非常に為になりました。

株式会社東レリサーチセンター専務取締役・研究部門長、吉川正信氏と、表面科学研究部部長、鮫島純一郎氏には、博士号取得を強く勧めていただいたと共に、研究活動を行う時間や費用面でのサポートを頂きました。厚く御礼申し上げます。本学位論文を執筆するために学んだ知識を、今後は業務においてもフィードバックしていきたいと考えております。

最後に、本研究の遂行のために休日不在とすることが多かったにも関わらず、不満も言わずに協力してくれた家族に心から感謝いたします。

2022年1月　松田　和大

最後に、本学位論文の第三章は、Biointerphases 誌（AIP Publishing LLC 社）に掲載された以下の 1. に示した学術論文に掲載された内容を基に作成された。また、第四章、五章については、Analytical and Bioanalytical Chemistry 誌（Springer Nature 社）に掲載された以下の 2. に示した学術論文に掲載された内容を基に作成された。

1. **Matsuda K**., Aoyagi S., Time-of-flight secondary ion mass spectrometry analysis of hair samples using unsuperevised artificial neural network. Biointerphases. 2020;15(2):021013. DOI: https://doi.org/10.1116/6.0000044
2. **Matsuda, K**., Aoyagi, S., Sparse autoencoder-based feature extraction from TOF-SIMS image data of human skin structures. Anal. Bioanal. Chem. 2022;414(2):1177-86. DOI: https://doi.org/10.1007/s00216-021-03744-3