

## 深層学習を用いた波長優先度推定による蛍光指紋識別 Excitation-Emission Matrix Classification by Wavelength Priority Estimation using Deep Learning

林田 純弥<sup>†</sup> 柿下 容弓<sup>†</sup> 服部 英春<sup>†,‡</sup>  
Junya Hayashida Yasuki Kakishita Hideharu Hattori

### 1. はじめに

蛍光指紋(Excitation-Emission Matrix, EEM)とは、分光蛍光光度計を用いて、波長を変えながら試料に励起光を照射し、試料から放出される蛍光スペクトルを記録した画像形式のデータである。蛍光指紋は特定の励起波長帯に対する蛍光スペクトルと比較して情報量が多く、化学・医薬品、食品、工業製品等、様々な分野で試料の解析に用いられている。

蛍光指紋を用いた試料の解析において、ユーザー利便性向上のために蛍光指紋測定時間を短縮し、解析のスループットを向上することが求められている。蛍光指紋は励起波長あるいは蛍光波長単位で測定を行うため、解析に用いる励起あるいは蛍光波長帯数を削減することが、測定時間短縮につながる。このような背景から、現行の解析手法には、Lasso 回帰[1]と呼ばれる線形回帰手法が主に用いられる。Lasso 回帰は、蛍光指紋を構成する波長帯の内、少数の波長帯のみを解析に使用するように正則化をかけつつ、解析モデルを構築する手法である。以降、解析に使用される波長帯を有効波長帯と呼ぶ。モデル構築後は、有効波長帯の蛍光強度のみを測定することで解析するため、運用時の蛍光指紋測定時間短縮が可能となる。

蛍光指紋を構成する各波長帯の蛍光スペクトルには、試料が放出する蛍光以外にも、ラマン散乱等の散乱光が含まれており、識別におけるノイズになる場合がある。Lasso 回帰を用いた識別モデルでは、蛍光指紋を構成する全ての波長帯の蛍光強度を説明変数としているため、上記ノイズの影響を受けやすく、選択された波長帯を用いて識別した場合、識別精度が不十分な場合がある。そこで、ノイズの

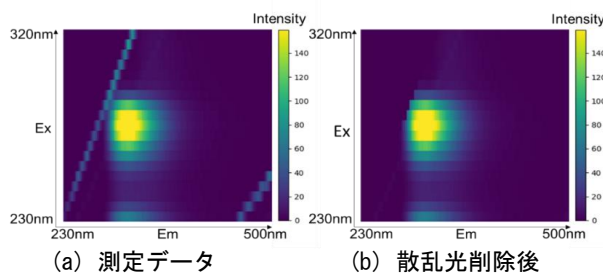


図 1 蛍光指紋の可視化例

影響を軽減するために、識別に有効な波長帯を手で選定する場合がある。しかし、有効波長帯の選定は専門家が自身の経験に基づき、試行錯誤しながら実施するため、属人性が高く、作業時間増大の要因となる。

本稿では、蛍光指紋を構成する各波長帯の、識別に対する重要度(波長優先度)を推定し、波長優先度に基づいて有効波長帯を選択することで、ノイズの影響を抑えつつ、運用時の蛍光指紋測定時間を短縮した、解析モデルを作成する手法を提案する。図 1 に提案手法の概要図を記載する。波長優先度推定器は、識別タスクの情報として、蛍光指紋と識別ラベルの集合を受け取り、クラス間の関係性を抽出することで、ノイズ成分のような識別に不要な波長帯に対しては低い波長優先度を推定し、識別に有効な波長帯に対しては高い波長優先度を推定する。波長優先度推定器は、様々な識別タスクを用いて、識別に有効な波長帯を学習(経験)することで、未知の識別タスク(蛍光指紋とラベルの集合)においても、波長優先度を推定する。波長優先度推定器により、ユーザーによる有効波長帯選定時の属人性を軽減し、迅速化することが可能となる。

試料に含まれる化合物を識別する実験において、一部の化合物情報のみで学習した波長優先度推定器を用いて選定した有効波長帯を用いて識別モデルを構築した結果、従来手法(Lasso 回帰)と比較して、提案手法は高精度かつ少ない有効励起波長帯数での識別を達成した。

### 2. 蛍光指紋について

蛍光指紋とは、分光蛍光光度計を用いて測定される、励起波長(Excitation)、蛍光波長(Emission)、蛍光強度(Intensity)の 3 軸で構成される画像形式のデータである。本研究で用いる蛍光指紋の可視化例を図 2 に示す。図 2(a)は測定後の蛍光指紋の可視化例、図 2(b)は図 2(a)に対して前処理を適用した結果であり、横軸を蛍光波長、縦軸を励起波長とし、蛍光強度を色で表現している。蛍光指紋の形状やピークを解析することで、試料の性質の推定が行われる。

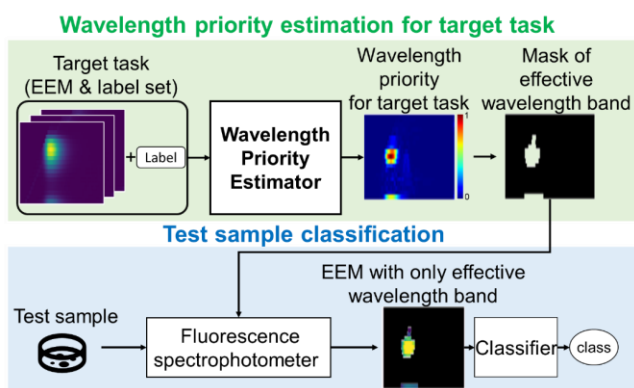


図 2 提案手法の概要

<sup>†</sup> 日立製作所 研究開発グループ Hitachi, Ltd. Research & Development Group

<sup>‡</sup> 日立ハイテク アナリティカルソリューション事業統括本部 Analytical & Medical Solutions Business Group, Hitachi High-Tech Corporation

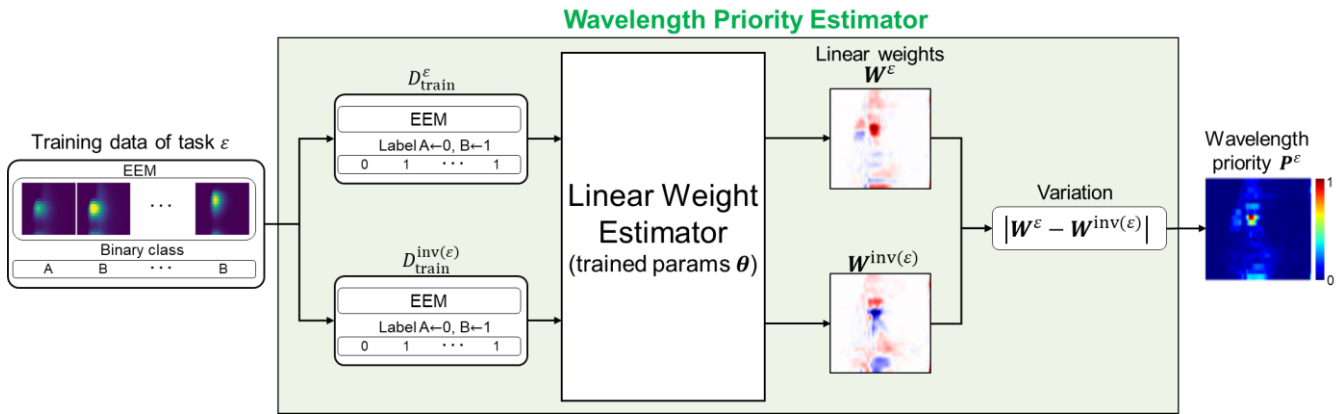


図 3 波長優先度の推定

1 つの励起波長帯における、蛍光波長帯毎の蛍光強度のスペクトルを蛍光スペクトル、1 つの蛍光波長帯における励起波長帯毎の蛍光強度のスペクトルを励起スペクトルと呼ぶ。試料に対し、照射する励起波長帯を変化させながら蛍光スペクトルを測定する、あるいは受光する蛍光波長帯を変化させながら励起スペクトル単位での測定を繰り返すことで、蛍光指紋が得られる。そのため、測定する励起波長帯あるいは蛍光波長帯を限定することが、蛍光指紋の測定時間短縮に繋がる。

図 2(a)において、蛍光指紋は、励起波長 $\geq$ 蛍光波長となる非蛍光領域、励起光が試料表面で反射しそのまま検出された、励起波長=蛍光波長 $\pm \beta$ の散乱光、および回折格子によって生じる散乱光の 2、3、 $\dots$ 、 $n$ 次光(励起波長 $=n \times$ 蛍光波長 $\pm \beta$ および  $1/n \times$ 蛍光波長 $\pm \beta$ )を含む。ただし、 $\beta$ は散乱光の帯域幅を表すハイパーパラメータである。非蛍光領域および散乱光は、試料が放出する蛍光とは異なり、試料特性に関係しないノイズを含むため、図 2(b)のように、解析前に削除することが一般的である。今回の実験における前処理については 5.1 節にて説明する。

### 3. 関連手法

#### 3.1 Lasso 回帰

Lasso 回帰[1]とは、線形回帰手法の一つであり、目的関数に正則化項を追加することで、幾つかの係数が 0 に設定される、スパース性を持つことを特徴とする。0 に設定された係数に対応する説明変数は、推定時には測定不要となる。つまり、Lasso 回帰は係数推定と、説明変数(蛍光指紋における波長帯)の選択を同時に行う回帰手法である。

上記スパース性から、蛍光指紋を用いた識別には Lasso 回帰が広く用いられている。訓練用の蛍光指紋に対し、各蛍光強度を説明変数、試料が属するクラスを目的変数として、変数(波長帯)選択を行う。運用時には、評価対象の試料に対し、選択された波長帯の蛍光強度のみを測定し、定量や識別を行うことで、測定時間の短縮が可能となる。

#### 3.2 集合を入力とする深層学習モデル

本提案手法は、波長優先度を入力されたタスク(蛍光指紋とラベルの集合)に対して一意に推定する。入力集合のため、波長優先度推定器の入力は、要素数が不定で、か

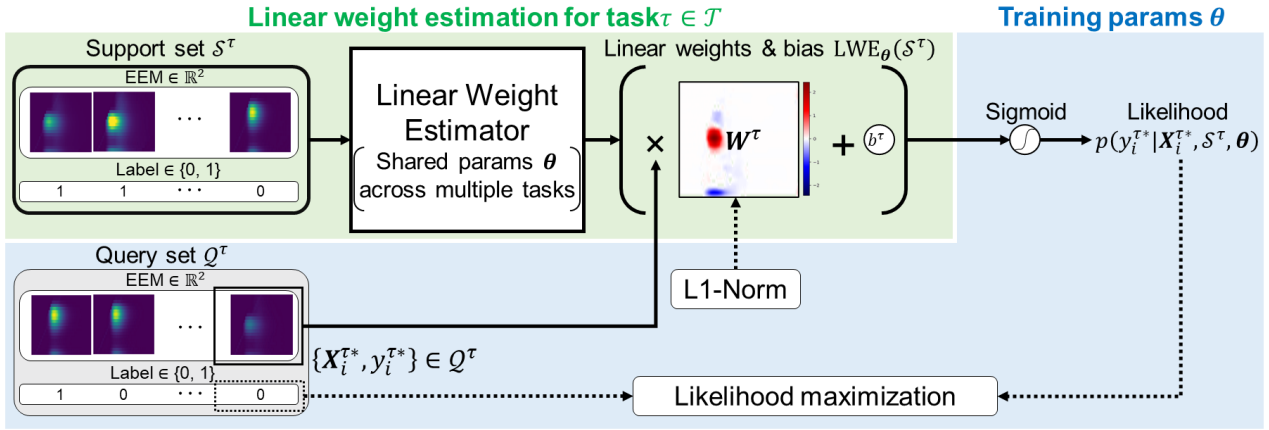
つ各要素は順番を持たない。Zaheer らはこのような、集合を入力とするモデルとして permutation invariant model[2]を提案している。permutation invariant model とは、入力集合の各要素に対し、適当な関数 $\phi$ と $\rho$ を用いて $\rho(\sum_{x \in X} \phi(x))$ と表せるモデルのことであり、入力集合の順番に依存しない結果を出力することが出来る。上記、permutation invariant model に該当する深層学習モデルとして Neural Process と呼ばれるメタ学習モデルが提案されている[3][4]。Neural Process は、パラメータを $\theta$ とする深層学習モデルによって尤度関数 $p_\theta(y|x; C)$ を推定するモデルである。 $p_\theta(y|x; C)$ は、入力 $x$ に対する出力 $y$ の確率分布が、コンテキスト集合 $C$ によって決定されることを示しており、 $C$ は $x, y$ と同じタスクのデータの集合 $\{(x_n, y_n)\}_{n=1}^N$ である。提案手法である波長優先度推定器においても、モデルの一部が Neural Process の構造を持っており、Neural Process と同様に、尤度関数を最大化することで学習を行う。

## 4. 提案手法

### 4.1 概要

本提案手法は、入力されたタスク(蛍光指紋と 2 クラス識別ラベルの組)に対し、深層学習を用いて線形回帰の重みを推定し、その重み情報から波長優先度を算出する。図 3 に波長優先度推定器(Wavelength Priority Estimator)を用いた波長優先度の推定過程を示し、図 4 に線形回帰の重みを推定する線形重み推定器(Linear Weight Estimator)の学習過程を示す。

図 3 にあるように、波長優先度推定器は、入力として蛍光指紋と 2 クラス識別ラベルの集合を受け付け、最終的な出力は波長優先度となる。波長優先度推定器の内部では、線形重み推定器を用いた線形重み推定処理を行う。線形重み推定器は、入力されるタスクに対して、それらを識別するための線形重みを推定する。線形重みは、前記入力データの一方のクラスを 0、もう片方のクラスを 1 としたデータセットと、クラスを逆転させたデータセットの 2 通りに対して推定する。入力する 2 クラスラベルを逆転させた場合、識別に重要な箇所の線形重みは正負が逆転することが期待される。しかし、ノイズの影響が大きい波長帯ではクラスラベルを逆転させた前後で重みが逆転しない傾向がある。そこでクラスラベル逆転前後の線形重みの変化量を算出することで、より識別に有効な波長帯を波長優先度とし

図 4 タスク  $\tau \in \mathcal{T}$  に対する線形重み推定器の学習過程

て抽出する方式を提案する。図 1 にあるように、推定した波長優先度を用いて識別モデルに入力する波長帯を選択することで識別精度を維持しつつ、有効波長帯数を削減する。

図 4 にあるように、線形重み推定器は、線形重みを推定するためのサポート集合(Support Set)と、推定した線形重みの精度を測るためのクエリ集合(Query Set)を用いて学習を行う。様々なタスクに対して、クエリ集合を識別する線形重みをサポート集合から推定するようにモデルを学習させることで、未知のタスクに対しても、識別のための線形重みを推定することをめざす。

#### 4.2 波長優先度の推定

後述する線形重み推定器を用いて、蛍光指紋を構成する各波長帯の識別タスクに対する重要度を表す、波長優先度を推定する。図 3 はある識別タスク  $\varepsilon$  に対する、波長優先度の推定を示した図である。識別タスク  $\varepsilon$  の訓練用データセット  $D_{\text{train}}^{\varepsilon} = \{(X_n^{\varepsilon}, y_n^{\varepsilon})\}_{n=1}^{N_{\varepsilon}}$  を学習済みの線形重み推定器に入力し、得られた線形重みを用いて波長優先度を求める。ただし、 $X_n^{\varepsilon}$  および  $y_n^{\varepsilon}$  は、蛍光指紋  $X \in \mathbb{R}^2$  および 2 クラスラベル  $y \in \{0,1\}$  を示す。

識別タスク  $\varepsilon$  の訓練用データセット  $D_{\text{train}}^{\varepsilon} = \{(X_n^{\varepsilon}, y_n^{\varepsilon})\}_{n=1}^{N_{\varepsilon}}$  を線形重み推定器に入力する際、2 クラスラベルを (0, 1) とする場合と、(1, 0) とする(ラベルを逆転させる)場合の 2 通り行い、それぞれの線形重み  $W^{\varepsilon}$ 、 $W^{\text{inv}(\varepsilon)}$  の変化量を正規化した以下の値を、タスク  $\varepsilon$  の波長優先度  $P^{\varepsilon}$  とする。

$$P^{\varepsilon} = \frac{|W^{\varepsilon} - W^{\text{inv}(\varepsilon)}|}{\max(|W^{\varepsilon} - W^{\text{inv}(\varepsilon)}|)} \quad (1)$$

ただし、 $\text{inv}(\varepsilon)$  は、タスク  $\varepsilon$  のラベルを逆転させたタスクを表す。線形重みは、入力されたタスクを識別するために必要な波長帯程、高値になる傾向がある。入力されるラベルを逆転した場合、重みの正負も逆転すると期待されるが、ノイズを多く含む波長帯の場合、重みがラベル逆転前後で変化しにくい傾向があることを発見した。そこで、入力されるラベルを逆転させた際の線形重みの変化量を波長優先度として算出する。また、3 クラス以上のマルチクラス識別の場合は、クラス毎に、自身のクラスと、それ以外のクラスの 2 クラス識別タスクとして波長優先度を推定し、各クラスで得られた波長優先度を最大値で統合することで、マルチクラス識別用の波長優先度を算出する。

#### 4.3 線形重み推定器

線形重み推定器について説明する。図 4 は、ある 2 クラス識別タスク  $\tau$  の蛍光指紋集合を入力した際の、線形重み推定器の学習過程を示した図である。図 4 に示すように、線形重み推定器は、2 クラス識別タスク(蛍光指紋と 2 クラスラベル)を入力とし、識別するための線形重みとバイアスを推定する深層学習モデルである。学習では、重みの推定に用いた集合(サポート集合)とは別に用意した集合(クエリ集合)に対する線形重みの識別精度を基準に、線形重み推定器のパラメータを最適化する。上記を、様々な 2 クラス識別タスクの蛍光指紋集合に対して行うことで、線形重み推定器を学習する。ただし、3 クラス以上の識別データセットであっても、クラスを 2 グループに分けることで、2 クラス識別データセットとして扱うことができる。

以降、4.3.1 節にて学習に用いる蛍光指紋集合について、4.3.2 節にて線形重み推定器を用いた線形重みの推定過程について、4.3.3 節にて線形重み推定器のパラメータの最適化方法について説明する。

##### 4.3.1 蛍光指紋集合の準備

線形重み推定器の学習には複数の 2 クラス識別データセット(蛍光指紋とラベルの組)を用いる。ここで、学習に用いる 2 クラス識別タスクの集合を  $\mathcal{T}$ 、各タスクのデータセットの集合を  $\mathcal{D} = \{D^{\tau \in \mathcal{T}}\}$  とする。また、 $\mathcal{D}$  を構成する各データセット  $D^{\tau \in \mathcal{T}}$  は、線形重み推定に用いるサポート集合  $S^{\tau} = \{(X_n^{\tau}, y_n^{\tau})\}_{n=1}^{N_{\tau}}$  と、線形重み推定器のパラメータ  $\theta$  を最適化するために用いるクエリ集合  $Q^{\tau} = \{(X_m^{\tau}, y_m^{\tau})\}_{m=1}^{M_{\tau}}$  を含んでいる。ただし、 $X_n^{\tau}$ 、 $X_m^{\tau}$  および  $y_n^{\tau}$ 、 $y_m^{\tau}$  は、蛍光指紋  $X \in \mathbb{R}^2$  および 2 クラスラベル  $y \in \{0,1\}$  を示す。学習において、2 クラスに割り当てるラベル  $\{0,1\}$  は、エポック毎にランダムに決定する。

##### 4.3.2 線形重みの推定

線形重みの推定過程を、サポート集合を入力とした場合として説明する。学習後の線形重み推定器は、4.2 節で説明したように、対象の識別タスクの訓練データセットを入力として線形重みを推定する。

ある学習タスク  $\tau \in \mathcal{T}$  において、サポート集合  $S^{\tau} = \{(X_n^{\tau}, y_n^{\tau})\}_{n=1}^{N_{\tau}}$  を線形重み推定器に入力した場合、線形重み推定器は線形重み  $W^{\tau}$  およびバイアス  $b^{\tau}$  を推定する。 $W^{\tau}$ 、 $b^{\tau}$  は各蛍光指紋  $X_n^{\tau}$  の各要素を説明変数とした際の線形回帰の

重みとバイアスである。学習可能なパラメータを $\theta$ とした線形重み推定器を関数 $LWE_{\theta}(\cdot)$ とした場合、サポート集合 $S^{\tau}$ と線形重み $W^{\tau}$ 、バイアス $b^{\tau}$ の関係は以下の通りである。

$$(W^{\tau}, b^{\tau}) = LWE_{\theta}(S^{\tau}) \quad (2)$$

$LWE_{\theta}$ は、集合を入力とするため、permutation invariant model[2]の構造を有している。サポート集合 $S^{\tau}$ の内、クラス 0 に属する部分集合 $S_0^{\tau}$ を、クラス 1 に属する部分集合を $S_1^{\tau}$ とした場合、各蛍光指紋に対して特徴量抽出を行う関数 $\phi$ と、 $S_0^{\tau}, S_1^{\tau}$ 毎に平均した特徴量に対して特徴量抽出を行う関数 $\rho$ を用いて、 $LWE_{\theta}$ は以下の式で表すことができる。

$$LWE_{\theta}(S^{\tau}) = \rho \left( \frac{1}{|S_0^{\tau}|} \sum_{X \in S_0^{\tau}} \phi(X), \frac{1}{|S_1^{\tau}|} \sum_{X \in S_1^{\tau}} \phi(X) \right) \quad (3)$$

実験では、 $\phi$ は蛍光指紋の空間的な特徴を抽出するために、CNN(Convolutional Neural Network)[5]、 $\rho$ はクラス毎の平均特徴量間の関係性を抽出するために、Transformer[6]におけるEncoder Layerを用いた。Transformerは、主に自然言語処理の分野で用いられるモデルであり、Encoder Layerは、入力される各特徴量の関係性を、非線形変換と内積を組み合わせた演算(Attention)を用いて抽出する。本提案手法においては、クラス 0 の平均特徴量と、クラス 1 の平均特徴量の関係性を抽出するためにEncoder Layerを用いた。このように、サポート集合から、クラス間の関係性を抽出することで、タスクに応じた重みとバイアスを推定するモデルをめざす。

#### 4.3.3 パラメータの最適化

線形重み推定器 $LWE_{\theta}$ のパラメータ $\theta$ は、各学習タスクのサポート集合から得られた重みとバイアス、及び各クエリ集合を用いて最適化される。目的関数は以下のような尤度関数と正則化関数の式で表すことができる。

$$\mathbb{E}_{\tau \in \mathcal{T}} \left[ \left[ \prod_{\{(X_m^{\tau*}, y_m^{\tau*}) \in \mathcal{Q}^{\tau}\}} p(y_m^{\tau*} | X_m^{\tau*}, S^{\tau}, \theta) \right] - \|W^{\tau}\|_1 \right] \quad (4)$$

ここで $p(y_m^{\tau*} | X_m^{\tau*}, S^{\tau}, \theta)$ は、 $X_m^{\tau*}$ のラベルが $y_m^{\tau*}$ である確率であり、Bernoulli 分布と sigmoid 関数を用いて表すと、Bernoulli( $\text{sigmoid}(W^{\tau} \cdot X_m^{\tau*} + b^{\tau})$ )となる。不要な波長帯の重みを 0 にするために、正則化項として L1 ノルム(各線形重みの絶対値の和)を設けている。この目的関数を最大化するように、線形重み推定器のパラメータ $\theta$ を最適化する。

#### 4.4 有効波長帯選定と識別モデル構築

波長優先度を 2 値化することで、識別に使用する波長帯数を削減しながら識別モデルを構築する。ここでは波長優先度を 2 値化する際の閾値の決定方法および識別モデル構築方法について説明する。波長優先度推定及び識別モデル構築に用いる訓練用データセットを $D_{\text{train}}^{\varepsilon}$ 、検証用データセットを $D_{\text{valid}}^{\varepsilon}$ 、推定した波長優先度 $P^{\varepsilon}$ とした場合、アルゴリズム 1 に示す手順で、有効波長帯数削減と識別モデル構築を行う。

アルゴリズム 1 において、Binalize は波長優先度を閾値 threshold で 2 値化し、有効波長帯を決定する処理、Mask は蛍光指紋を有効波長帯に限定する処理、Build は任意の識別モデルを構築する処理、Validate は検証データに対する識別性能を測る処理を意味する。アルゴリズム 1 は、閾値を徐々に上げながら、検証データに対する識別性能を基準に閾値と識別モデルを選択する方式である。今回の実験に

おいては、検証データに対する識別精度を基準として選択した。

#### Algorithm 1. 識別モデル構築

```

1: score ← 0
2: for threshold ∈ {0.00, 0.05, 0.10, ..., 1.00} do
3:   wave ← Binalize( $P^{\varepsilon}$ , threshold)
4:   classifier ← Build(Mask( $D_{\text{train}}^{\varepsilon}$ , wave))
5:   if score ≤ Validate(classifier,  $D_{\text{valid}}^{\varepsilon}$ ) then
6:     score ← Validate(classifier,  $D_{\text{valid}}^{\varepsilon}$ )
7:     result ← classifier
8:   else then
9:     break
10: return result

```

#### 4.5 構築した識別器による識別

構築した識別タスク $\varepsilon$ の識別器を用いて、実際に試料から識別を行う方法について説明する。図 1 に示すように、分光蛍光光度計(Fluorescence spectrophotometer)を用いて、試料(Test sample)から有効励起波長帯の蛍光強度のみを測定し、評価用蛍光指紋を得る。得られた評価用蛍光指紋を構築した識別器に入力することで識別を行う。このように、試料から、有効励起波長帯のみを測定することで識別が可能なため、提案手法は蛍光指紋の測定時間短縮の効果があ

### 5. 実験

#### 5.1 データセット

実験に用いるオープンデータセットとして、化合物が含まれた溶液を試料とする蛍光指紋(化合物蛍光指紋) [7]を用いた。化合物蛍光指紋は 6 種の成分の濃度(含まない場合は濃度=0)がラベルされており、このデータセットに対し、提案手法を用いて、成分の有無(濃度>0 か濃度=0 か)を識別する実験を行った。各蛍光指紋は、励起波長 230nm ~ 320nm を 5nm 間隔(計 19 波長)、蛍光波長 230nm ~ 500nm を 2nm 間隔(計 136 波長)で測定されており、また、1 つの試料に対して同条件で 5 回測定されている。

6 種の成分はそれぞれ、Catechol、Hydroquinone、Indole、Resorcinol、Tryptophane、Tyrosine である。しかし、Resorcinol に関しては、Resorcinol を含んでいる(濃度>0 である)サンプル数が、全体の約 9%と少なく、正例/負例の割合が偏っているため、データセットから除外した。

化合物蛍光指紋は、識別対象とした成分の内、一つの成分のみを含む蛍光指紋(単成分)および、複数の成分を含む蛍光指紋(複成分)に分けられる。単一成分および複成分の蛍光指紋の例を図 5 に記載する。前処理として、非蛍光領域および散乱光(1 次光, 2 次光)に相当する領域の蛍光強度を 0 にする前処理を施した。

表 1 に単一成分のデータセット構成を示す。各数値は、対象成分毎の訓練、検証、評価用データセットに含まれる試料の数を示す。各成分の試料において、訓練、検証、評価用データセットに含まれる試料の割合が 3:1:1 になるようにデータセットの分割を行った。ただし、同一の試料に対し蛍光指紋は 5 回計測されているため、表 1 の値の 5 倍の蛍光指紋サンプルが用意されている。訓練および検証用データセットは波長優先度を用いた識別器の構築時に用いる。後述の線形重み推定器の学習においては、表 1 の緑でハイライトしたサンプルのみを用いた。

表 2 に複合成分のデータセット構成を示す。表 1 と異なり、各数値は、対象成分を含む試料数/含まない試料数を示す。成分毎に、その成分を含む試料数と含まない試料数それぞれにおいて、訓練、検証、評価用データセットに含まれる試料の割合が 3:1:1 になるようにデータセットの分割を行った。単一成分と同様に、試料毎に 5 件の蛍光指紋が用意されている。

表 1 単一成分のデータセット構成  
(対象成分の試料数)

	Train	Valid	Eval
Catechol	7	2	2
Hydroquinone	9	3	3
Indole	9	3	3
Tryptophane	4	2	2
Tyrosine	11	3	3
Total	40	13	13

表 2 複合成分のデータセット構成  
(対象成分を含む試料数/含まない試料数)

	Train	Valid	Eval
Catechol	79/35	26/12	26/12
Hydroquinone	73/41	24/14	25/13
Indole	77/37	25/13	26/12
Tryptophane	55/58	18/20	18/20
Tyrosine	53/61	17/21	18/20

## 5.2 線形重み推定器の学習

5 種の成分の内、3 種の成分(Catechol、Hydroquinone、Indole)の単一成分の訓練用および検証用試料(表 1 の緑でハイライトしたサンプル)から測定した蛍光指紋(試料 25 件×5 回測定分の計 125 件)のみを用いて、線形重み推定器の学習を行った。

1 回のパラメータ更新(イテレーション)毎に、2 クラスレベルのパターンをバッチサイズの数だけ定義する。パターンは、「Catechol、Hydroquinone、Indole の内、ランダムに 1 つの成分を選択し、その成分が含まれているか否か」とし、訓練用、検証用データセットから、サポート集合およびクエリ集合をサンプリングする。上記において、バッチサイズは 8、サポート集合のサイズは 16、クエリ集合のサイズは 24、イテレーション回数は 14000 回とした。

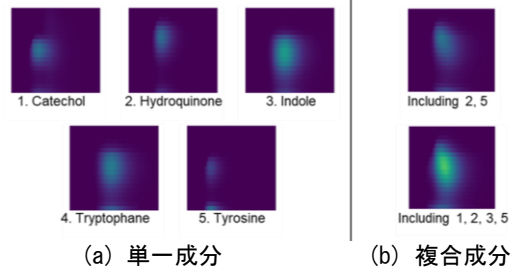


図 5 化合物蛍光指紋の可視化例

## 5.3 識別モデルの構築

5.2 節で得られた線形重み推定器を用いて、単一および複合成分の訓練用蛍光指紋から、波長優先度を推定し、識別性能を検証データの識別精度としたアルゴリズム 1 の方式で識別モデルを構築した。識別モデルには、蛍光強度を説明変数とした Lasso 回帰と、筆者らが蛍光指紋向けに提案済みの深層学習による特徴量抽出器を用いた識別手法[8]の 2 パターンを検証した。[8]は共通特徴量抽出器(Common Feature Extractor)と呼ばれる CNN ベースの特徴量抽出器を用いて、励起波長帯毎(蛍光スペクトル毎)に特徴量抽出を行う。

## 5.4 識別モデルの評価

識別モデルに対する評価用データセット(表 1、2 の Eval)の評価結果を表 3、表 4 に記載する。表 3 において、「Lasso」は Lasso 回帰による識別モデルを表し、「Priority+Lasso」は Lasso 回帰に対してアルゴリズム 1 を適用した提案手法を表す。表 4 において、「CFE」は特徴量抽出識別手法[8]を表し、「Priority+CFE」は CFE に対してアルゴリズム 1 を適用した提案手法を表す。「Acc.」は識別精度(Accuracy)[%]、「Eff.」は有効励起波長帯数(Number of effective excitation wavelength bands)を表す。ただし、全励起波長帯数は 19 である。

表 3 Lasso 回帰における提案手法の適用結果

	Lasso		Priority + Lasso	
	Acc.	Eff.	Acc.	Eff.
Catechol	83.0	15	<b>98.1</b>	<b>3</b>
Hydroquinone	92.5	13	<b>98.1</b>	<b>2</b>
Indole	75.9	14	<b>96.3</b>	<b>3</b>
Tryptophane	92.3	15	<b>92.3</b>	<b>4</b>
Tyrosine	96.1	17	<b>98.0</b>	<b>2</b>
Average	88.0	14.8	<b>96.6</b>	<b>2.8</b>

表 4 特徴量抽出識別手法[8]における提案手法の適用結果

	CFE[8]		Priority + CFE	
	Acc.	Eff.	Acc.	Eff.
Catechol	<b>96.2</b>	8	<b>96.2</b>	<b>3</b>
Hydroquinone	<b>100.0</b>	5	<b>100.0</b>	<b>2</b>
Indole	<b>96.3</b>	10	94.3	<b>7</b>
Tryptophane	94.2	10	<b>96.2</b>	<b>4</b>
Tyrosine	<b>98.0</b>	7	<b>98.0</b>	<b>6</b>
Average	<b>97.1</b>	8.0	<b>97.1</b>	<b>4.4</b>

表 3 から分かる通り、提案手法は従来手法と比較して平

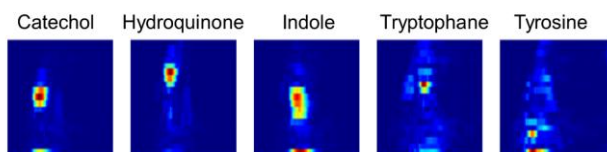


図 6 波長優先度の可視化結果

均識別精度が 8.8% 向上し、平均有効励起波長帯数を 81% 削減した。また、表 4 から分かる通り、提案手法は CFE を識別モデルとした場合においても、平均識別精度を維持しつつ、平均有効励起波長帯数を 45% 削減した。

### 5.5 考察

波長優先度の可視化結果を図 6 に記載する。図 6 より、学習に用いていない成分(Tryptophane、Tyrosine)も含めて、波長優先度がそれぞれ異なる波長帯に反応を示していることが分かる。学習に用いていない成分(Tryptophane、Tyrosine)においても有効波長帯数を削減しつつ、Lasso 回帰あるいは特徴量抽出識別手法[8]以上の識別精度を得たことから、波長優先度が、識別に有効な波長帯を選択することが出来たと考える。また、運用時は有効波長帯のみを測定することで識別が可能のため、提案手法を用いることで、蛍光指紋の測定時間の短縮が可能となる。

本提案方式では、クラスラベルを逆転した線形回帰の重みの差分を用いて、波長優先度を算出する方法を採用した。検証のために、推定した重みを用いて直接識別したところ、平均識別精度は 71.4% だった。これは推定した線形回帰の重みがそのままではノイズの影響を含んでいるためと考える。一方、提案手法ではクラスラベルを逆転した場合の重みの差分に着目することでノイズの影響を軽減し、識別に有効な波長帯をより選択的に推定している。

表 4 にある通り、提案手法を用いることで、特徴量抽出識別手法[8]の識別精度を維持しつつ、有効波長帯数を削減することが出来た。特徴量抽出識別手法を組み合わせた際の更なる精度向上・有効波長帯数削減のためには、励起波長帯毎に得られた特徴量に対する波長優先度が必要であると考え。提案手法における線形重みは、蛍光指紋の各蛍光強度を説明変数とした重みであり、特徴量を説明変数とした重みでは無い。そこで、特徴量を説明変数と想定した波長優先度推定器を学習することで、精度向上と有効励起波長帯数削減を両立可能と考えている。

### 6. おわりに

本稿では、分光蛍光光度計を用いて試料から得られる蛍光指紋を用いた識別タスクにおいて、識別に有効な波長帯を波長優先度として推定する方式を提案した。波長優先度を用いることで、識別に有効な波長帯のみを用いた識別モデルの構築を可能とし、高識別精度かつ短測定時間な識別を実現する。試料に含まれる化合物を識別する実験において、一部の化合物情報のみを用いて波長優先度推定器を学習し、従来手法と組み合わせた識別モデルを構築した結果、提案手法は高精度かつ少ない有効波長帯数での識別を達成した。

### 参考文献

- [1] R. Tibshirani, "Regression Shrinkage and Selection via the Lasso", *Journal of the Royal Statistical Society. Series B (Methodological)*, Vol.58, No.1 (1996).
- [2] Z. Manzil, et al., "Deep Sets", *NeurIPS*, (2017).
- [3] G. Marta, et al., "Conditional Neural Process", *ICML*, (2018).
- [4] R. James, et al., "Fast and Flexible Multi-Task Classification Using Conditional Neural Adaptive Processes", *NeurIPS*, (2019).
- [5] K. Alex, et al., "ImageNet Classification with Deep Convolutional Neural Networks", *NeurIPS*, (2012).
- [6] V. Ashish, et al., "Attention Is All You Need", *NeurIPS*, (2017).
- [7] B. Rasmus, et al., "Standard error of prediction for multilinear PLS 2. Practical implementation in fluorescence spectroscopy", *Chemometrics and Intelligent Laboratory Systems*, Vol.75, No.1 (2005).
- [8] J. Hayashida, et al., "Excitation-Emission Matrix Classification by Common Feature Extraction using Deep Learning", 第 21 回情報科学技術フォーラム, (2022).