# Zero-Shot Learning 講義ノート

劉慶豊 Hosei University 2025年8月7日

## 1 はじめに:医用画像解析の例

Zero-Shot Learning (ZSL) は、訓練時に観測されなかったクラス(未学習クラス)に対しても、正確な分類を行うことを目指す機械学習の枠組みである。従来の教師あり学習では、訓練時とテスト時のクラスが一致していることを前提とするが、ZSLではこの前提を取り払う。

このような汎化を実現するために、ZSL では各クラスに対応する**意味特徴ベクトル**(semantic attribute vector)を導入し、視覚空間と意味空間の橋渡しを行う必要がある。

### 1.1 医用画像解析の設定

本講義では特に、医用画像解析における ZSL の応用を基礎例として設定する。

- 訓練時に観測された疾患(既知クラス):肺がん、肺炎、結核、気胸、サルコイドーシスなど 20 種類
- 未知クラス(診断すべき疾患):ランゲルハンス細胞組織球症、リンパ脈管筋腫症、肺胞蛋白症など
- 視覚特徴(入力空間):胸部 CT 画像から抽出された放射線特徴ベクトル  $x \in \mathbb{R}^{d_x}$
- **意味特徴(意味空間)**:疾患の臨床記述(例:「上肺野の多発結節」「喫煙歴との関連」「10 歳以下での発症」など)を BioBERT 等でベクトル化した  $a_y \in \mathbb{R}^{d_z}$
- 目的:新しい症例画像が与えられたときに、訓練に使われていない疾患名に分類・診断すること

本例に基づき、以下の章では ZSL の理論と構造について詳述する。

## 2 問題設定と記号

入力空間を  $\mathcal{X} \subset \mathbb{R}^{d_x}$ 、出力空間(クラス集合)を  $\mathcal{Y}$  とする。 $\mathcal{Y}$  は既知クラス(学習済みクラス)と未知クラス(未学習クラス)に分割される:

$$\mathcal{Y} = \mathcal{Y}^{\text{seen}} \cup \mathcal{Y}^{\text{unseen}}, \quad \mathcal{Y}^{\text{seen}} \cap \mathcal{Y}^{\text{unseen}} = \emptyset$$

訓練データ:

$$D_{\text{train}} = \{(x_i, y_i)\}_{i=1}^n, \quad x_i \in \mathcal{X}, \quad y_i \in \mathcal{Y}^{\text{seen}}$$

各クラス  $y\in\mathcal{Y}$  に対して、意味特徴ベクトル  $a_y\in\mathbb{R}^{d_z}$ (例えば属性ベクトルや word2vec など)が与えられていると仮定する。

## 3 手法の基本構造

ZSL では、視覚特徴ベクトル  $x \in \mathbb{R}^{d_x}$  を意味空間  $\mathbb{R}^{d_z}$  に写像する線形変換  $W \in \mathbb{R}^{d_z \times d_x}$  を学習する。

### 3.1 マッピングの定義

 $\hat{z} = Wx$ 

ここで、 $\hat{z}$  は視覚特徴ベクトル x を意味空間へ射影したベクトルである。

## 4 What is Zero-Shot Learning

Zero-Shot Learning(ZSL)は、訓練時に観測されなかったクラス(未学習クラス)に対しても、正確な分類を行うことを目指す機械学習の枠組みである。従来の教師あり学習では、訓練時とテスト時のクラスが一致していることを前提とするが、ZSLではこの前提を取り払う。

このような汎化を実現するために、ZSLでは各クラスに対応する**意味特徴ベクトル**(semantic attribute vector)を導入し、視覚空間と意味空間の橋渡しを行う必要がある。

## 5 問題設定と記号

入力空間を  $\mathcal{X} \subset \mathbb{R}^{d_x}$ 、出力空間(クラス集合)を  $\mathcal{Y}$  とする。 $\mathcal{Y}$  は既知クラス(学習済みクラス)と未知クラス(未学習クラス)に分割される:

$$\mathcal{Y} = \mathcal{Y}^{\text{seen}} \cup \mathcal{Y}^{\text{unseen}}, \quad \mathcal{Y}^{\text{seen}} \cap \mathcal{Y}^{\text{unseen}} = \emptyset$$

訓練データ:

$$D_{\text{train}} = \{(x_i, y_i)\}_{i=1}^n, \quad x_i \in \mathcal{X}, \quad y_i \in \mathcal{Y}^{\text{seen}}$$

各クラス  $y \in \mathcal{Y}$  に対して、意味特徴ベクトル  $a_y \in \mathbb{R}^{d_z}$ (例えば属性ベクトルや word2vec など)が与えられていると仮定する。

## 6 手法の基本構造

ZSL では、視覚特徴ベクトル  $x \in \mathbb{R}^{d_x}$  を意味空間  $\mathbb{R}^{d_z}$  に写像する線形変換  $W \in \mathbb{R}^{d_z \times d_x}$  を学習する。

## 6.1 マッピングの定義

$$\hat{z} = Wx$$

ここで、 $\hat{z}$  は視覚特徴ベクトル x を意味空間へ射影したベクトルである。

## 7 学習:射影行列 W の最適化

訓練データに対し、以下の目的関数を最小化することでWを学習する:

$$\min_{W} \sum_{i=1}^{n} \|Wx_i - a_{y_i}\|^2 + \lambda \|W\|_F^2$$

- x<sub>i</sub>:画像から抽出された視覚特徴ベクトル
- $a_{y_i}$ : クラス  $y_i$  に対応する意味ベクトル

λ:正則化項の係数

•  $\|\cdot\|_F$ : Frobenius ノルム

## 8 推論:未知クラスの予測

テスト時には、未知クラス  $y \in \mathcal{Y}^{\text{unseen}}$  に対して以下を最大化することで予測を行う:

$$\hat{y} = \arg\max_{y \in \mathcal{Y}^{\text{unseen}}} F(x, y) = \arg\max_{y} \langle Wx, a_y \rangle$$

ここで内積  $\langle Wx, a_y \rangle$  は、x の意味空間への射影 Wx とクラス y の意味ベクトル  $a_y$  との類似度を表す。

## 9 スコア最大化と損失最小化の等価性

損失最小化とスコア最大化は、以下の変形により等価であることが示されている:

$$||Wx_i - a_{y_i}||^2 = (Wx_i)^\top (Wx_i) - 2(Wx_i)^\top a_{y_i} + ||a_{y_i}||^2$$
(1)

$$= 条件によってほぼ一定 - 2\langle Wx_i, a_{y_i} \rangle + 定数$$
 (2)

### 9.1 補足:スコア最大化と損失最小化の関係とその解釈

ZSL の損失関数とスコア最大化の関係は、以下のように整理できる。

#### 9.1.1 二乗損失の展開

損失最小化問題:

$$\min_{W} \sum_{i=1}^{n} \|Wx_i - a_{y_i}\|^2 + \lambda \|W\|_F^2$$

損失項の展開:

$$||Wx_i - a_{y_i}||^2 = ||Wx_i||^2 - 2(Wx_i)^{\top} a_{y_i} + ||a_{y_i}||^2$$
(3)

$$= x_i^{\top} W^{\top} W x_i - 2x_i^{\top} W^{\top} a_{y_i} + ||a_{y_i}||^2$$
 (4)

ここで:

- 第1項  $x_i^\top W^\top W x_i = \|W x_i\|^2$  は W に依存する
- 第2項  $-2x_i^\top W^\top a_{y_i}$  はスコア項に対応
- 第3項  $\|a_{y_i}\|^2$  は W に依存しないため、定数として無視できる

### 9.1.2 「 $W^{\top}W$ が一定」という表現の解釈

文献によっては「 $W^\top W$  が一定」または「第1項が定数」と記述されるが、これは厳密には正しくない。以下のような条件下において近似的に成り立つと考えられる:

- (1) W に対して強い正則化  $\lambda \|W\|_F^2$  を課しており、W の変動が小さい場合
- (2) 入力ベクトル  $x_i$  がすべて  $||x_i||=1$  に正規化されており、 $x_i^\top W^\top W x_i$  のばらつきが小さい場合
- (3)  $W^{\mathsf{T}}W = I$  (直交性制約) などが課されている場合

したがって、「 $W^\top W$  が一定」とは、実際には W の構造が安定しており、第 1 項  $\|Wx_i\|^2$  の寄与が目的関数全体に対してあまり変化を与えない状況を指している。

### 9.2 なぜ直交性制約 $W^{\mathsf{T}}W = I$ を通常は課さないのか?

Zero-Shot Learning (ZSL) において、写像行列 W に直交性制約  $W^{\top}W=I$  を課すことは一般的ではない。主な理由は以下のとおりである:

#### 1. 直交性制約は最適化が難しくなる

直交性制約  $W^\top W = I$  は非線形かつ非凸な制約条件であり、最適化が複雑になる。特に深層学習や大規模データセットを扱う場合、このような制約を厳密に満たすように W を更新することは計算コストが高く、効率的でない。

#### 2. 汎用性の高いモデル設計に向かない

ZSL では、視覚特徴を意味空間へ柔軟にマッピングすることが求められる。W に直交性のような構造的制約を課すと、写像の自由度が制限され、必要な表現力が損なわれる可能性がある。

#### 3. 通常は正則化項で十分

実際の応用では、Frobenius ノルム  $\|W\|_F^2$  を目的関数に加えることで、W のサイズや振る舞いを効果的に制御できる。このため、明示的に直交性を強制する必要性は低く、より簡便で汎用的な正則化で十分な性能が得られる。

#### 9.3 より正確な記述例

「 $\|Wx_i\|^2$  は厳密には W に依存するが、W のノルムを強く制約したり、 $W^\top W$  に構造的な制約を課すことで、その影響は抑えられる。このような仮定のもとでは、スコア項の最大化と損失関数の最小化はほぼ等価な目的となる。」

このため、最大スコアに基づく推論と二乗誤差による訓練が本質的に整合している。

## 10 閉形式解 (Closed-form Solution)

最小二乗法によるWの解析解は以下の通りである:

$$W^* = AX^{\top}(XX^{\top} + \lambda I)^{-1}$$

- $X = [x_1, \dots, x_n] \in \mathbb{R}^{d_x \times n}$
- $A = [a_{y_1}, \dots, a_{y_n}] \in \mathbb{R}^{d_z \times n}$

## 11 図解:意味空間への写像

### 12 応用

ZSL は、学習時に観測されなかったクラスに対しても、意味空間を介して予測・分類を行う技術である。本節では、代表的な応用分野を3つ取り上げ、それぞれの構造と目的を具体的に解説する。

### 12.1 画像認識 (未知の物体カテゴリの識別)

● **目的:** テスト時に初めて出現するカテゴリ(例:「アフリカスイギュウ」「ホバークラフト」など)を、 ラベルなし画像から自動で識別する。

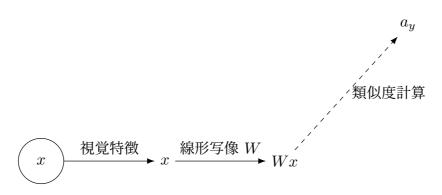


図1 意味空間への写像プロセス

• 入力: 画像 x (JPEG 等) を CNN で処理し、視覚特徴ベクトル  $Wx \in \mathbb{R}^{d_z}$  に変換。

• **意味空間:** クラス名(日本語または英語)を word2vec や CLIP Text Encoder でベクトル化した  $a_y$  を用意。

• **推論:** 画像特徴と意味ベクトルとの内積  $\langle Wx, a_y \rangle$  を計算し、最も近いクラスを出力:

$$\hat{y} = \arg\max_{y \in \mathcal{Y}^{\text{unseen}}} \langle Wx, a_y \rangle$$

• 評価: Top-1 / Top-5 精度により、未知クラスへの分類精度を評価する。

## 12.2 文書分類 (新しいトピックの自動分類)

● 目的: AI 倫理、宇宙資源、量子社会など、学習データに含まれていないトピックを自動分類する。

• **入力**: 文書 x (ニュース記事や論文) を BERT などでエンコードし、文埋め込みベクトルに変換。

• **意味空間:** クラス名(トピック名)および代表キーワード(例:「倫理」「リスク」「透明性」)からベクトル  $a_u$  を構成。

ullet 推論: 文書ベクトル Wx と意味ベクトル  $a_y$  の類似度を比較し、最も近いトピックを予測:

$$\hat{y} = \arg\max_{y \in \mathcal{Y}^{\text{unseen}}} \langle Wx, a_y \rangle$$

• 評価: 新トピック文書に対して分類精度 (accuracy) や F1 スコアを用いて評価。

#### 12.3 医用画像解析(稀少疾患の識別)

• 目的: 学習時に観測されていない稀な疾患に対しても、意味的知識を介して正確に診断する。

#### 構造とプロセス:

● 訓練時に観測された疾患:肺がん、肺炎、結核、気胸、サルコイドーシスなど一般的な疾患(計 20 種類)

- **未知クラス(診断すべき疾患)**: ランゲルハンス細胞組織球症、リンパ脈管筋腫症、肺胞蛋白症など稀 少疾患
- 入力(視覚特徴):CT 画像から放射線特徴を抽出(3D CNN や radiomic feature)し、ベクトル $x \in \mathbb{R}^{d_x}$  を得る
- 意味空間(疾患記述): 各疾患の臨床的記述を日本語または英語でまとめたテキスト(例: 「10 歳以下で発症」「上肺野に多発性結節」「間質性陰影」「喫煙歴との関連」)を医療用言語モデル(BioBERT、PubMedBERT など)でベクトル化して  $a_u \in \mathbb{R}^{d_z}$  を構成

• 推論:入力画像 x を意味空間へ射影 Wx したうえで、未知疾患候補  $\mathcal{Y}^{\mathrm{unseen}}$  の意味ベクトルと比較:

$$\hat{y} = \arg\max_{y \in \mathcal{Y}^{\text{unseen}}} \langle Wx, a_y \rangle$$

● **評価**:テストセットには診断済みの症例(ただし訓練時に未使用)が含まれており、Top-1 精度や診断 一致率で評価される。

### ZSL が有効となる医学的背景と意義:

- 稀少疾患の症例は非常に少なく、従来の教師あり学習では対応困難
- 新型感染症や変異株など、「前例のない病態」への即時対応が求められる
- 意味空間を用いた類似性ベース分類は、既知疾患の知識を活かして未知に対処できる
- 医師の補助判断ツールとして、診断支援の信頼性向上に貢献

### ZSL の究極的な目標(医療応用):

「どのような病気か全くわからない医用画像を与えられたときに、訓練データに一度も登場しなかった 疾患名に対しても、意味情報をもとに正確な分類・診断を行う」

これは、医療 DX の一環として AI 診断補助システムにとって極めて重要な要素となる。