

汎化性能を考慮した食品パッケージの画像分類

20X4110 鴫田 優太

法政大学 理工学部
経営システム工学科

2024年2月1日

背景と問題

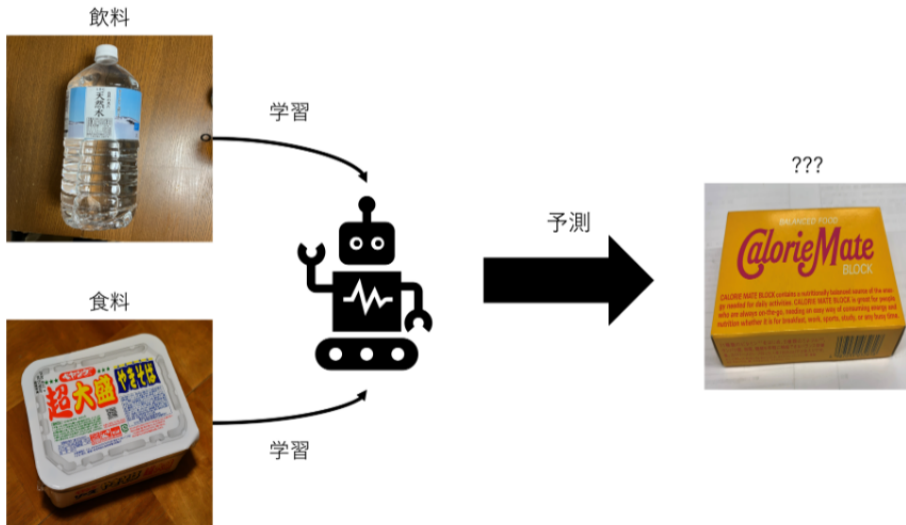
画像分類技術を応用することで、人間が視覚を通して物を判別するような作業を自動的に判別することが可能になる。しかし、簡単な分類モデルでは、人間の視覚を通じた判別に劣ってしまう問題が発生する。

このような背景から、食料品の工場向けのデータ分析に注目し、汎化性能に重点を置いた、商品のパッケージが映った画像を食料か飲料かに自動分類するアルゴリズムを開発することが望まれている。

汎化性能 未知のデータに対して正しく予測できる性能。

背景と問題

<https://signate.jp/competitions/1106> から引用



先行研究

- ▶ 吉次研二・中本幸一 (2021)[1] は、画像分類において、CNN より Vision Transformer モデルの方が有効な場合があることを転移学習を用いて示している。
- ▶ Liu et al(2022)[2] は、ConvNet を改善した Convnext を提案し、87.8 パーセントの ImageNet トップ 1 精度を達成している。
- ▶ 石田直也 (2021)[3] は、不均衡データの取り扱いに着目した深層学習による運転者の眠気推定における提案モデルにおいて、StratifiedKFold($K=5$) を分割に用いている。

CNN 畳み込み層やプーリング層で構成される深いモデル。

Vision Transformer 自然言語処理で使われてきた Transformer を使うモデル。

転移学習 あるタスクで学習したモデルである学習済みモデルを異なるタスクに転用する手法。

StratifiedKFold($K=5$) クラスの割合を保ちながらデータを分割する方法。K は分割数で、各分割セットをフォールド (Fold) という。

目的

- ▶ 目的
汎化性能を考慮したモデルを提案したい。
- ▶ アプローチ
StratifiedKFold($K=5$) を用いて、データを5つのフォールドに分割し、各フォールド毎に転移学習を通して、学習・検証を行い、最適な重みに更新していく。こうして得られた5つのモデルの予測を TTAch を通して平均することで、1つの予測値を求める方法を行う。

データ

本研究では、「SIGNATE」にて、「テクノプロ・デザイン社 食品パッケージ画像解析チャレンジ」から提供されたデータを扱う。2023/8/11 から 2023/9/29 までの期間で開催で参加人数は 640 人であった。訓練データは 2176 枚の画像データと対応するラベルデータ、テストデータは 2180 枚の画像データのみである。

▶ 訓練データ

訓練データは答えが与えられてるデータで、ラベル 1(食料) が 1182 枚、ラベル 0(飲料) が 994 枚と不均衡なクラス比率である。これを学習用データと検証用データに分割する。

▶ テストデータ

テストデータは、予測するためのデータのことで、Public データと Private データに分けられる。Public データは、コンペティション期間中に精度を確認できるデータをいい、Private データは、期間後に確認できるデータである。また、評価指標には二値分類モデルの性能を評価できる AUC(Area Under the Curve) を使用する。

データ

1785.png



1797.png



0648.png



0166.png



Figure: ラベル 0 の画像の一部.

モデル

提案する深層学習によるモデル構成は以下の表の通りである。

Table: 提案モデル

| 段階 | 項目 | 手法 |
|-----|---------|--------------------------|
| 学習前 | データ前処理 | 6つのデータ拡張/Skf(K=5) |
| 学習前 | ミニバッチ処理 | ImbalancedDatasetSampler |
| 学習 | 訓練方法 | AMP |
| 学習 | 転移学習 | EVA-02 |
| 学習 | 重みの更新 | Ranger21 |
| 学習 | 損失関数 | CCE |

ミニバッチ処理 データをミニバッチというセットに分けて学習する処理。

転移学習 あるタスクで学習したモデルである学習済みモデルを異なるタスクに転用する手法。

損失関数 最適なモデルの重みを見つけるために最小化される関数。

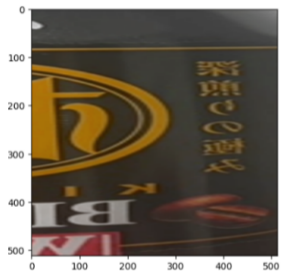
データ前処理 (データ拡張)

データ拡張 画像に変形を施す処理のこと.

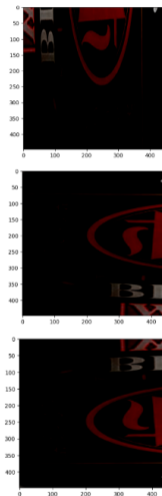
→汎化性能の向上を期待.

- ▶ HorizontalFlip: 水平方向に回転.
- ▶ VerticalFlip: 垂直方向に回転.
- ▶ RandomRotate90: 90 度回転.
- ▶ Normalize: 標準化のため.
- ▶ Resize: モデル構造が求める画像サイズに変換するため.
- ▶ ToTensorV2: モデル構造を Pytorch ライブラリで扱うため.

データ拡張の例



元の画像

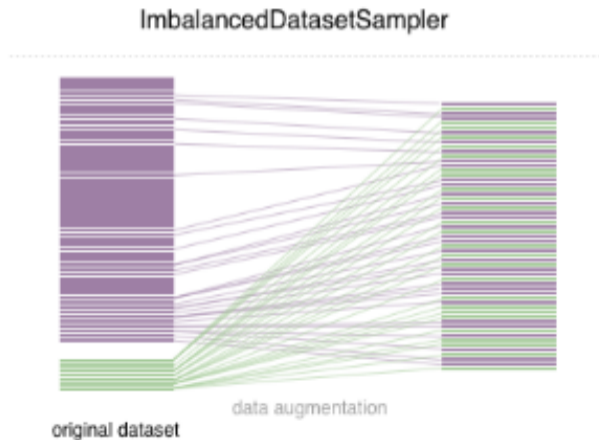


データ拡張後

Figure: データ拡張の例.

ミニバッチ処理 (ImbalancedDatasetSampler)

ミニバッチ処理 データをミニバッチというセットに分けて学習する処理のこと。
ImbalancedDatasetSampler を用いることで、ラベルの比率を考慮してミニバッチのサンプリングを行い汎化性能の向上を期待する。



学習項目

- ▶ 訓練方法 (AMP(Automatic Mixed Precision))
半精度の浮動小数点数を使用して訓練する手法 [4].
 - ・モデルの精度を落とすことなく、メモリ要件を半減できる。
- ▶ 転移学習 (EVA-02)
Vision Transformer 系のモデルで、学習済みモデルの中で一番の精度を達成している [5].
- ▶ 重みの更新 (Ranger21)
既存の更新法を改善し、検証精度と学習速度を大幅に向上している [6].
 - ・学習率を徐々に小さくすることで、最適なパラメータに安定して到達しやすくする。
- ▶ 損失関数 (CCE(Complement Cross Entropy))
既存の損失関数を改善し、不均衡な画像分類の予測精度の向上を期待できる。 [7].

予測 (TTAch)

予測のアプローチでは，StratifiedKFold を用いて学習した 5 つのモデルそれぞれで予測させた後，5 つの結果を平均化することで最終的な予測結果を取るアプローチを行う．ここで，5 つのモデルそれぞれで予測する際に，TTAch を用いる．

なお，実際に扱ったデータ拡張は，HorizontalFlip, VerticalFlip, そして Rotate90 である．

<https://github.com/qubvel/ttach> から引用



Figure: TTAch の構造.

学習結果と予測 (ハイパーパラメータ)

なお、学習率は、初期段階では 6×10^{-5} 、学習 72 パーセント地点で 3×10^{-5} に下げている。

ハイパーパラメータ モデルの挙動を制御する値。

Table: 提案するハイパーパラメータの値

| | |
|------------|---|
| エポック数： | 2 |
| バッチサイズの高さ： | 17 |
| 学習率： | $6 \times 10^{-5} \rightarrow 3 \times 10^{-5}$ |

学習結果と予測 (検証結果)

Table: エポックによる AUC の変化

| Fold | Train Auc | | Valid Auc | |
|------|-----------|--------|-----------|--------|
| | エポック 1 | エポック 2 | エポック 1 | エポック 2 |
| 0 | 0.7800 | 0.9380 | 0.8990 | 0.9620 |
| 1 | 0.8200 | 0.9620 | 0.9142 | 0.9667 |
| 2 | 0.8175 | 0.9510 | 0.8946 | 0.9716 |
| 3 | 0.8155 | 0.9642 | 0.8978 | 0.9702 |
| 4 | 0.8127 | 0.9587 | 0.8837 | 0.9379 |

Public データにおける AUC スコアは 0.888, Private データにおける AUC スコアは 0.9052 で参加者 640 人の中で 2 位の精度を達成している。

Private データにおける AUC スコアが Public データに劣らず伸びていることから、高い汎化性能を持つモデルを構築できていることがわかる。

混合行列

モデルがラベル 0 を正しく 0 と予測する真陽性は 900, 反対にできなかった場合の偽陽性は 94 つである. モデルがラベル 1 を正しく 1 と予測する真陰性は 1047, 反対にできなかった場合である偽陰性は 135 つである. どちらかのラベルに偏った予測をしていないことがわかる.

<sklearn.metrics._plot.confusion_matrix.ConfusionMatrixDisplay at 0x78bb08757550>

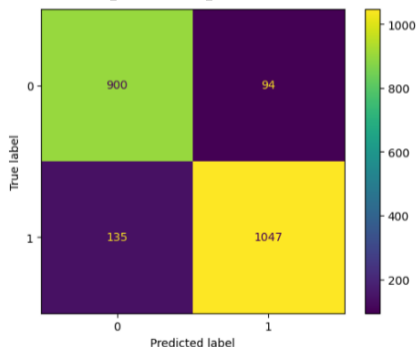


Figure: 訓練データにおける混合行列.

ImbalancedDatasetSampler 無しとの比較

Table: ImbalancedDatasetSampler あり

| | |
|--------------------|--------|
| 検証スコア (平均) | 0.9596 |
| Public データにおけるスコア | 0.8880 |
| Private データにおけるスコア | 0.905 |

Table: ImbalancedDatasetSampler なし

| | |
|--------------------|--------|
| 検証スコア (平均) | 0.9558 |
| Public データにおけるスコア | 0.8916 |
| Private データにおけるスコア | 0.8999 |

ミニバッチ処理が特に効き、最終 10 位の精度から最終 2 位の精度まで更新できた。

CNN モデルとの比較

CNN モデルの転移学習には，Convnext を用いた．

Table: Convnext で用いたハイパーパラメータの値

| | |
|-------------|---|
| エポック数： | 3 |
| バッチサイズの大きさ： | 8 |
| 学習率： | 7×10^{-5} から 3×10^{-5} に Warmdown |

Table: Convnext におけるスコア

| | |
|--------------------|--------|
| 検証スコア (平均) | 0.9583 |
| Public データにおけるスコア | 0.8762 |
| Private データにおけるスコア | 0.8861 |

本研究でも，先行研究同様に CNN 系のモデルより Vision Transformer 系のモデルの転移学習の方が有効であった．

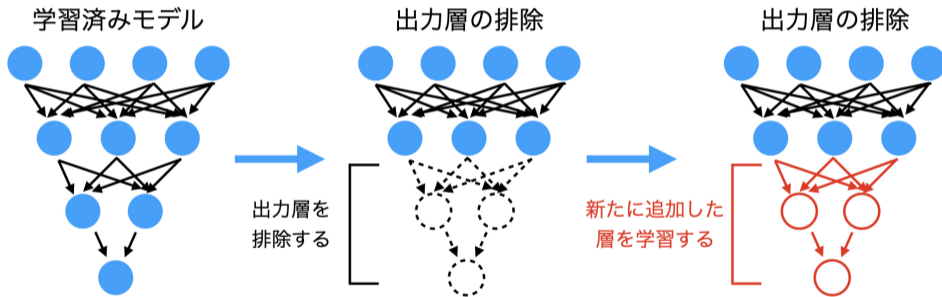
おわりに

本研究で提案しているアプローチを取ることで、AUC スコアが 0.9 という高い精度を保ちつつ、未知な Private データに対しても十分な汎化性能を持つモデルを構築することに成功した。

先行研究

-  吉次研二・中本幸一，画像分類における CNN と Vision Transformer の精度比較，2021 年情報処理学会関西支部 支部大会，2021.
-  Z. Liu, H. Mao, C. Wu, C. Feichtenhofer, T. Darrell, S. Xie, A ConvNet for the 2020s, arXiv, 2022.
-  石田直也，不均衡データの取り扱いに着目した深層学習による運転者の眠気推定，大学院研究年報理工学研究科編，2021.
-  P. Micikevicius, S. Narang, J. Alben, G. Diamos, E. Elsen, D. Garcia, B. Ginsburg, M. Houston, O. Kuchaiev, G. Venkatesh, H. Wu, Mixed Precision Training, arXiv, 2018.
-  ⁴ Y. Fang, Q. Sun, X. Wang, T. Huang, X. Wang, Y. Cao, EVA-02: A Visual Representation for Neon Genesis, arXiv, 2023.
-  ⁴ L. Wright, N. Demeure, Ranger21: a synergistic deep learning optimizer, arXiv, 2021.
-  ⁴ Y. Kima, Y. Leea, M. Jeona, Imbalanced Image Classification with Complement Cross Entropy, arXiv, 2021.

転移学習のイメージ



学習済みモデルの入力付近の層は特徴を捉えることができるので
流用することが出来る！

付録

▶ 損失関数

<https://atmarkit.itmedia.co.jp/ait/articles/2104/15/news030.html> から引用

