

汎化性能を考慮した食品パッケージの画像分類

鍋田 優太 (20X4110) 指導教員 作村 建紀

1. はじめに

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機械で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。なお、汎化性能とは、未知のデータに対して正しく予測できる性能のことを言う。

先にキーワードを簡単に定義する。

- 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。

吉次研二・中本幸一は、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 [1]。

本研究では、汎化性能を考慮したモデルを提案することを目的とする。その目的を達成するアプローチとして、StratifiedKFold($K=5$) を用いて、データを5つのフォールドに分割し、各フォールド毎に学習済みモデルの転移学習を通して、学習・検証を行い、最適なハイパーパラメータ、パラメータを調整、また保存する。こうして保存された5つのモデルを TTA 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。

2. データについて

データには、訓練データ、テストデータに分けられる。訓練データは答えが与えられてる

2186 枚の画像データと対応するラベル情報を持つデータで、ラベル 1 が 1182 枚、ラベル 0 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して、モデルのパラメータを調整していく。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、期間後に精度を確認することが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。

3. モデル

提案するモデルは以下の通りである。

表 1: 提案モデル

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

データ拡張とは、画像に変形を施す処理のことで、それぞれ水平方向、垂直方向、90度回転させる HorizontalFlip, VerticalFlip, そして RandomRotate90 を用いる。他に、標準化のために Normalize, モデル構造が求める画像サイズに変換する Resize, モデル構造を Pytorch ライブラリで扱うために tensor 型に変換する ToTensorV2 を用いる。

ミニバッチ処理とは、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を言う。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う。

訓練方法では、AMP を用いる。AMP を用いることで、モデルの精度を落とすことなく、メモリを半減しつつ訓練ができる [2]。

転移学習には、学習済みモデルに EVA-02 のラージモデル [3] を利用することで、精度の高

表 2: 提案するハイパーパラメータの値

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

いモデルの構築に繋がることを期待する。

重みの更新には Ranger21[4] を用いることで、精度向上を期待する。

損失関数とは、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である。本研究では、CCE[5] を用いることでモデルの汎化性能の向上に期待する。

また、予測部分では、TTAch を用いる。この手法では、学習したモデルにデータ拡張した予測画像を数回表示する。そして、対応する各画像の予測値を平均化することで、それを最終的な予測結果とする手法である。実際に扱ったデータ拡張は、HorizontalFlip, VerticalFlip, そして Rotate90 で、学習時に施した回転系データ拡張と同じである。これらを用いることで、画像の汎化性とモデル精度の向上を期待する。

4. 学習結果と予測

4.1 ハイパーパラメータ

ハイパーパラメータはモデルの挙動を制御する。提案する最適なハイパーパラメータを表 2 にまとめる。

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。

4.2 検証結果

検証データの結果を以下の表 3 に記載する。どの Fold においても、AUC スコアを伸ばせている。また、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 位の精度を達成することに成功している。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。

表 3: 検証結果

Fold0 : 0.8990 \rightarrow 0.9620
Fold1 : 0.9142 \rightarrow 0.9667
Fold2 : 0.8946 \rightarrow 0.9716
Fold3 : 0.8978 \rightarrow 0.9702
Fold4 : 0.8837 \rightarrow 0.9379

表 4: 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

表 5: 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

4.3 ImbalancedDatasetSampler 無しとの比較

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。

5. 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。

参考文献

- [1] 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.
- [2] 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.
- [3] 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.
- [5] Y. Kima, Y. Leea, M. Jeona, Imbalanced Image Classification with Complement Cross Entropy, arXiv, 2021.