

EfficientNet と LSTM の統合モデルによる異なるデータへの汎化能力の検証

Evaluation of the Generalization Capability of an Integrated EfficientNet and LSTM Model on Different Datasets

友光 祐輔¹⁾望月 久稔¹⁾

Yusuke TOMOMITSU Hisatoshi MOCHIZUKI

1 はじめに

機械学習は多様なデータ形式に対して適用されており、画像や音声、表形式データなどのそれぞれの構造に応じたモデル設計が進められている [1]。なかでも、センサーログや経済指標などの時系列の表形式データは広く利用されており、予測や異常検知、分類などにおいて重要な役割を果たしている。このようなデータには、ある時点における複数の特徴量間の関係性（空間的情報）と、時間に沿った連続的な変化（時系列情報）の双方が存在することが多い。よって、これら両方の情報を統合的に捉えるモデルが求められている。先の研究 [2] では画像特徴抽出に優れた EfficientNetB0 [3] と時系列情報の学習に適した LSTM [4] を統合したモデルを提案した。しかし、この統合モデルがデータ構造や分類、回帰の違いによってどの程度適応できるかは十分に検証されていない。そこで、EfficientNetB0 と LSTM を組み合わせた統合モデルの異なるデータにおける汎化性能を検証する。

2 長方形カーネルを用いた EfficientNetB0 と LSTM の統合モデル

本研究の対象となるモデルとして、先の研究で提案した EfficientNetB0 と LSTM の統合モデルの構成と動作原理の概要と、特徴の一つである長方形カーネルの導入について説明する。

2.1 統合モデルの構成要素

本モデルは、単一の画像や表形式データが持つ静的な特徴と、それらの連続から生じる時間的な依存関係を同時に学習することを目的として設計している。モデルの全体構成として、まず、入力データから EfficientNetB0 により空間的な特徴を抽出する。その後、得られた特徴ベクトルを、LSTM が時間的または順序的な依存関係を学習できるように、適切な形状に変換して入力する。

また、データセットの規模や複雑さに応じて最適なモデルの表現能力は異なるので、過学習の抑制と予測精度の向上を両立させるため、本研究では EfficientNetB0 の深さを調整したモデル群を用いる。基本となる

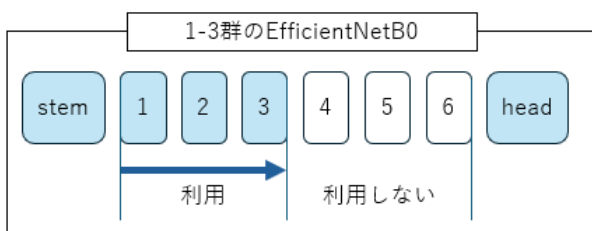


図1 1-3群 EfficientNetB0 の構成

EfficientNetB0 における6段階のブロックに基づき、使用するブロック数を変えた6種類(1-1群~1-6群)の深さのモデルを定義し、検証に用いる。1-3群の構成例を図1に示す。図1より1-3群は、EfficientNetB0 の構造をもとに、6つの段階的なブロックのうち後半3ブロックを省略し、入力層である stem、前半3ブロック、および出力に対応する head で構成される。

2.2 長方形カーネルへの拡張

空間的な特徴抽出には EfficientNetB0 を用いるが、多様なデータへの適応を目的としてカーネルの形状を変更する。一般的な CNN は画像認識に用いられるが、行を各種特徴量、列を時間とするような表形式のデータを、画像のような2次元の行列として扱うことで、CNN アーキテクチャへの適用が可能となる。しかし、このようなデータにおいて、行方向は特徴量間の相関を、列方向は時間的な相関を意味するため、両者の性質は本質的に異なる。そこで、特定の方向性を持つ特徴をより効率的に抽出するために、EfficientNetB0 内の一般的な正方形カーネル ($k \times k$) を、非対称な長方形カーネル ($m \times n$) に拡張する。

3 異種データに対する汎化性能の実験的評価

EfficientNetB0 と LSTM の統合モデルに対し、異なる種類のデータや、分類と回帰の違いによる汎化性能を実験的に評価する。分類には正解率を、回帰には決定係数を評価指標とし、後述する5種類のデータセットを用いる。

分類には画像データである CIFAR-10 [5] と MNIST [6] に加え、表形式データである Adult データセット [7] を利用する。CIFAR-10 は自然画像を入力とし、飛行機や犬など、写っている物体の種類を10クラスに分類する。MNIST は手書き数字の画像を入力とし、0から9までの10クラスに分類する。Adult データセットは、個人の職業や年齢、学歴などの属性からその人物の年収が5万ドル以上か否かを2値分類する。

一方、回帰にはいずれも表形式かつ時系列情報を持つ為替予測データ [8] と成績予測に用いるデータを使用する。前者は、過去の為替レートや経済指標などから将来の為替価格を予測し、後者は2022年度のICTの授業における学習過程に取った小テストの結果から最終的な成績を予測する。

3.1 分類予測

まず、CIFAR-10 と MNIST、Adult データセットを用いて、分類予測の性能を検証する。

各群における最も精度の高かったものを表1に示す。表1より、MNIST においては1-4群が最も高い正解率である0.60を示し、モデルの深さがある程度までは性能向上に寄与することが示唆される。一方で、1-5群以降では精度が低下しており、過度な深さにより学習が困難

1) 大阪教育大学 Osaka Kyoiku University

表 1 各群の分類予測における正解率

群	MNIST	CIFAR-10	Adult
1-1	0.30	0.17	0.80
1-2	0.36	0.17	0.80
1-3	0.44	0.18	0.80
1-4	0.60	0.18	0.80
1-5	0.55	0.12	0.79
1-6	0.30	0.12	0.76

表 2 各群の回帰予測における決定係数

群	為替予測	成績予測
1-1	0.74	-6.60
1-2	0.98	-6.32
1-3	0.96	-6.51
1-4	0.87	-6.45
1-5	-1.66	-6.34
1-6	-0.49	-5.46

になり過学習を引き起こした可能性がある。

また、CIFAR-10 では全体的に正解率が低く、最高でも 1-3 群、1-4 群で 0.18 にとどまった。これは、CIFAR-10 のような多様かつ高次元な画像データに対して十分な表現力を発揮できないことが示唆される。さらに、MNIST と同様に、深さの増加に伴って精度が低下する傾向が確認された。

MNIST は画像から文字を認識して予測し、CIFAR-10 は画像から物体を認識して予測する。MNIST より CIFAR-10 の分類精度が低い理由として、文字は垂直や水平方向に重要な情報が多く含まれるからであると考ええる。

以上のように、画像分類においては性能が限定的であることが確認された。これは、本モデルが時系列構造を前提とした設計であることから、静的な画像データに対しては構造的に不適合である可能性がある。

続いて、Adult データセットでは、1-4 群以下の群で最も高い正解率である 0.80 を示し、構造化された表形式データに対しては本モデルが一定の分類性能を発揮することが示唆された。Adult データは各特徴量の意味が明確で単純な表形式データであり、特徴量間の局所的な関係性を EfficientNetB0 の畳み込み層が適切に捉えたことで、高い分類性能に寄与した可能性がある。また、データの構造が明瞭でノイズが少なかったため、深いモデルにおいても過学習の影響が小さく、全体として安定した精度が得られたと考える。この結果は、静的な画像データに対して分類性能が限定的であったのに対し、構造を持つ表形式データにおいては本モデルの構成が有効に機能したことが示唆される。

3.2 回帰予測

次に、為替予測と成績予測の 2 種類のデータセットを用いて、回帰予測における性能を評価する。

各群における最も精度の高かったものを表 2 に示す。表 2 より、為替予測においては 1-2 群で最も予測精度が高く、0.98 であった。これは本モデルが時系列かつ表形式のデータ構造に適合していることが示唆される。具体的には、EfficientNetB0 により各時点の特徴量間の関係性を効果的に抽出し、LSTM が時間方向の依存性を捉えることで、特徴間・時系列間の双方の情報を統合的に学習できたと考える。為替データのように明確な時間順の構造と、複数の関連指標が並行して変化する表形式データに対しては、このようなモデル構成が特に有効であり、モデルの構成要素がデータの特性と高い整合性を持っていたことが、高精度の要因であると考えられる。また、1-2 群のような中程度の深さを持つ構造は、モデルの表現力と汎化能力のバランスが取れており、データの複雑さに対して適切な構造であった可能性がある。

次に成績予測では、表 2 より、全ての決定係数が負の値となり、為替予測と比較して明らかに精度が低く、十分な予測性能は得られなかった。主な要因として、学習に用いたデータ数の少なさが挙げられる。本実験では、学習データが 60 件、検証データが 29 件と、深層学習モデルの学習に必要なサンプル数としては極めて少ない規模であり、モデルの表現力を活かすには不十分であったと考える。また、成績予測は他の予測と異なり、個人の学習履歴や環境要因など、外的・内的に複雑な要因が影響する可能性がある。このような非線形性やノイズ性の高いデータに対して、サンプル数が限られていたことが重なり、モデルの学習が困難になったと考える。

4 おわりに

本研究では、EfficientNetB0 と LSTM を統合したモデルに対して、分類および回帰の異なるタスクを含む 5 種類のデータセットを用い、汎化性能を実験的に評価した。

分類予測では、MNIST において中程度の深さのモデルが最も高い精度を示した一方、CIFAR-10 では全体的に性能が低く、本モデルが複雑な画像データには適さない可能性が示された。一方、Adult データに対しては比較的高い精度を記録し、表形式データと相性が良いことが示唆された。回帰予測では、為替予測において高い決定係数が得られ、時系列表形式データに対して本モデルが有効であることが示唆された。成績予測では、データ量の少なさにより、十分な性能は得られなかった。

以上の結果から、本モデルは時系列構造を持つ表形式データに適しており、画像などの静的なデータやサンプル数の少ないデータには不向きである可能性がある。

参考文献

- [1] Y. LeCun, Y. Bengio, and G. Hinton : Deep learning, Nature, vol. 521, no. 7553, pp. 436–444, (2015).
- [2] 友光祐輔, 望月久稔 : データサイズに応じた CNN の畳み込み層の深さ分析とカーネルの形状変化による比較, 情報処理学会 第 86 回全国大会, (2024).
- [3] Tan, M., Le, Q. V. : EfficientNet : Rethinking Model Scaling for Convolutional Neural Networks, (online), DOI : 10.48550/ARXIV.1905.11946, (2019) .
- [4] Hochreiter, S. and Schmidhuber, J. : Long short-term memory, Neural computation, Vol.9, No.8, pp.1735–1780, (1997) .
- [5] A. Krizhevsky : Learning multiple layers of features from tiny images, Technical Report, University of Toronto, (2009).
- [6] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner : Gradient-based learning applied to document recognition, Proceedings of the IEEE, vol. 86, no. 11, pp. 2278–2324, (1998).
- [7] Adult - UCI Machine Learning Repository, 入手先 <https://archive.ics.uci.edu/dataset/2/adult> , 参照 (2025-06-11).
- [8] HistData.com, 入手先 <http://www.histdata.com/> , 参照 (2025-06-11).