

深層学習におけるネットワーク構造比較による重みパラメータ特性の考察

Consideration of weight parameter characteristics by comparing network structures in deep learning

○井上直哉¹, 塚本新²*Naoya Inoue¹, Arata Tsukamoto²

Abstract : For estimating the character of complex network structures, we study the weighting parameters dependency on the network configured by CNN and LSTM. In this report, we compared each weights on a simplified 2 models on progression. We suggested that the image recognition model embedded in the network may have feature extraction that depends on its structure.

1 はじめに

深層学習では、入力から出力までの導出が不透明であるため、機械による判別や分類の理由が説明不可能となるブラックボックス問題がある^[1]。近年では、この問題に対して、Grad-CAM^[2]のような学習した重みに着目し、学習モデルを解析する研究が行われている。しかし、ネットワーク構造が複雑な場合、重みパラメータの特性や傾向の読み取りが困難となる。

本研究では、ネットワークの構造と、反応する特徴量の関係について検討することを目的として、複数のネットワークを組み合わせた複雑なモデルと、そのネットワークを分割したモデルの学習後の重みを比較した。

2 実験方法

2.1 対象タスク

本研究では、2つの入力画像の前後関係から等差数列となるような、3つ目の数値を予測するモデルタスクを扱う。対象タスクの概要を Figure 1 に示す。

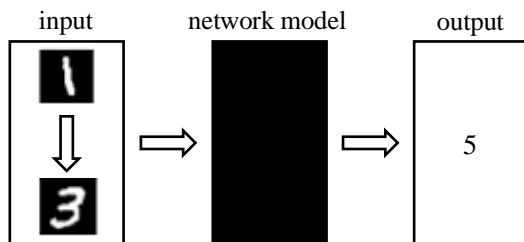


Figure 1. Overview of the target task

2.2 使用ネットワークモデル

ネットワークモデルには、画像特徴量抽出や画像認識・分類に特化した畳み込みニューラルネットワーク (Convolutional Neural Network : CNN) と、順序や時間に依存する系列データを得意とする長・短期記憶 (Long Short Term Memory : LSTM) を使用する。CNN のネットワーク構成を Figure 2 に示す。畳み込み層のフィルタを 16 個とし、過学習を抑制するため、Pooling 層の後、Dropout 率 0.25 の Dropout 層を追加する。

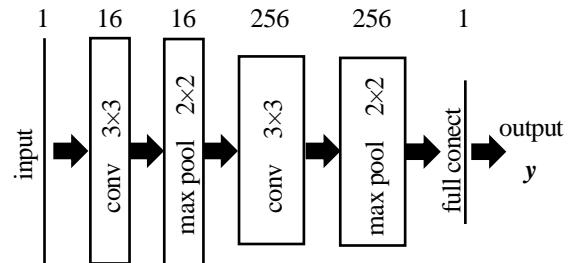
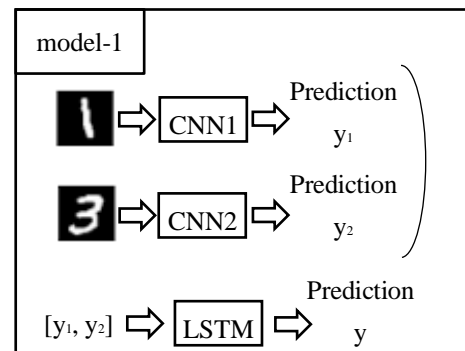


Figure 2. CNN configuration

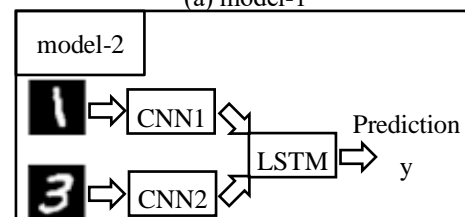
2.3 ネットワーク構造

Figure 3 に今回使用するネットワーク構造を示す。2つの CNN を学習した後、学習データ画像に対する予測ラベルを LSTM の学習データとし、数値予測を行う model-1, 2つの CNN 出力を直接 LSTM に入力し、全体を単一のネットワークとして学習する複合ネットワークを model-2 とする。

Table 1 に各 CNN, LSTM, model-2 の複合ネットワークの学習時におけるハイパーパラメータを示す。



(a) model-1



(b) model-2

Figure 3. Network structure

Table 1. Hyperparameters

batch	epoch	loss function	optimizer
128	100	cross entropy	sgd (lr = 0.01)

2.3 使用データ

データセットは、MNIST 手書き数字画像^[2]70000 枚を使用する。これを学習用 42000 枚、検証用 18000 枚、テスト用 10000 枚に分配する。その後、分配した3つの順序を反転したものを元の画像配列と結合し、画像2枚の組み合わせを入力データとし、2枚に対して等差数列となる3つ目の数値を正解ラベルとする。

3 結果

3.1 モデルによる予測精度の違い

model-1, model-2 の精度比較のため、各モデル各ネットワークにおけるテストデータの予測精度を Table 2 に示す。Table 2 より、対象タスクに対し、model-2 の方が高い予測精度を有することを確認した。

Table 2. Predictive accuracy for each network

	Network	Accuracy
model-1	CNN1	(0.982)
	CNN2	(0.983)
	LSTM	<u>0.494</u>
model-2		<u>0.867</u>

3.2 重みパラメータ比較

学習後の各 CNN のフィルタ比較を行う。本稿では、簡単のため各 CNN の第一畳み込み層のフィルタ 16 個を扱う。また、比較方法として、基準となる CNN の i 番目のフィルタ F_i に対し、比較対象の CNN の j 番目のフィルタとのユークリッド距離を E_{ij} とする。この時、基準となる F_i と比較対象となる CNN の 16 フィルタ全体に対する非類似度として、 S_i を以下の式で定義する。

$$S_i = \min(E_{ij}) \quad (1 \leq j \leq 16) \quad (1)$$

同モデル内の CNN1, CNN2 のフィルタ比較による非類似度を、ヒートマップを用いて表したものを Figure 4 に示す。Figure 4 では、model-1 において、CNN1 基準、比較対象 CNN2 の場合を (a)、CNN2 基準、比較対象 CNN1 の場合を (b) とする。model-2 も同様に、CNN1 基準、比較対象 CNN2 の場合を (c)、CNN2 基準、比較対象 CNN1 の場合を (d) とする。

次に、モデル間における CNN1 同士、または CNN2 同士のフィルタ比較による非類似度を、ヒートマップを用いて表したものを Figure 5 に示す。Figure 5 では、互いの CNN1 において、model-1 基準、比較対象 model-2 の場合を (a)、model-2 基準、比較対象 model-1 の場合を (b) とする。CNN2 も同様に、model-1 基準、比較対象 model-2 の場合を (c)、model-2 基準、比較対象 model-1 の場合を (d) とする。

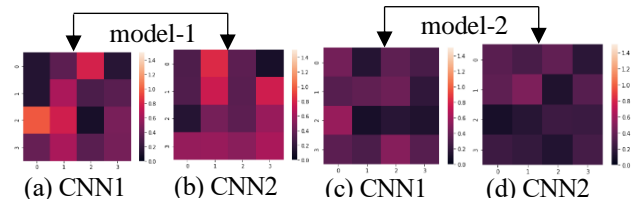


Figure 4. Filter comparison for the same model

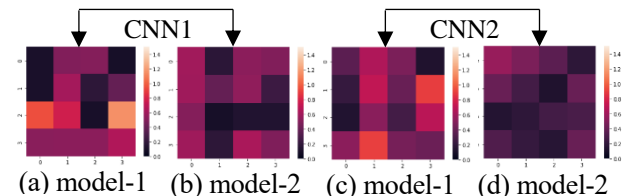


Figure 5. Filter comparison between models

4 考察

3.1 の model-1 における CNN1, CNN2 の予測精度の差は 0.1% であり、Figure 4 (a), (b) より、非類似度の高いフィルタが model-1 の CNN1, CNN2 それぞれに存在することから、同等の予測精度を出す画像認識モデルにおいても、フィルタの学習結果は一意に定まらなないと考えた。また、Figure 5 の (b), (d) より、model-2 の CNN における全てのフィルタに対し、類似度の高いフィルタが model-1 の CNN に存在することや、Figure 4 の (c), (d) より、model-2 内でお互いに類似度の高いフィルタが存在することが確認できた。これらから、model-2 の CNN のフィルタによる数値予測としての画像認識モデルの分類力は、model-1 よりも低く、画像を認識・分類するのではなく、LSTM の学習を正しく行う入力を得るための特徴量抽出を行っていることが示唆できる。

5. まとめ

本研究では、入力画像の前後関係から等差数列となる数値予測タスクに対し、ネットワーク構造の違いによる予測精度確認、及び CNN フィルタ比較を行った。

以上より、同等の認識精度であっても、重みによって抽出する特徴量が一意に定まらないこと、ネットワーク構造によって、CNN 単体で抽出する特徴量とは異なる特徴抽出を行っている可能性を示唆した。

6. 参考文献

- [1] 福島俊一, 他: 「ビッグデータ×機械学習の展望: 最先端の技術的チャレンジと広がる応用」, 情報管理, Vol.60, No.8, pp.543-554, 2017
- [2] R. Selvaraju: “Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization”, arXiv: 1610.02391, 2016
- [3] Yann LeCun: “The MNIST database of handwritten digits”, <http://yann.lecun.com/exdb/mnist/>, 19 SEP 2021