

モデル再適用手法の背景と CNN の特徴抽出の関係に関する検討 Examination of the Relationship between Background of Model Reapplication Techniques and CNN Feature Extraction

外川 宙¹⁾ 神野 健哉¹⁾
Sora Togawa Kenya Jin'no

概要

近年、注目を集めている深層学習モデルの学習方法に model merge と呼ばれる手法がある。model merge は複数の学習済モデルの重みを組み合わせることでタスクの処理能力の向上を実現させる手法である。本研究では、事前学習済みモデルの再適用手法の成功の背景を探究すると同時に、畳み込みニューラルネットワーク (CNN) の動作をより深く理解することを目的とする。異なるドメインで学習した CNN モデルの比較を行い、畳み込み層には共通のカーネルが生成されることが明らかにする。この結果は、ImageNet で事前学習されたモデルが広く応用可能な理由が、特徴抽出層が適切に構築されているためであることを示唆している。

1 まえがき

画像処理分野において CNN は優れた性能を発揮する。画像分類タスクを解くためのモデルとして CNN が提案されるようになった当初は、畳み込み層の数やカーネルサイズ、残差接続の有無など、どのように構造を設計し、ハイパーパラメータを組み合わせるかを探求することで、高性能なモデルを追求してきた。また、それに伴いモデルの複雑さや解釈の不透明さも増しており、説明可能性の観点から問題視されている。一方、近年では fine tuning や transfer learning のようなモデル再適手法を利用して、より高性能なモデルを生成することがある。特に、ImageNet[1] で学習されたモデルを、対象のドメインで再学習させると良い性能が得られることが多い。これらの手法とは異なり、近年注目を集めているモデル再適用手法に、model merge と呼ばれる手法がある。自然言語処理分野では、model merge によって優れた性能を発揮する Large Language Models が生成される [2]。

本研究では、model merge を CNN に適用し、model merge の成功要因と CNN の特徴抽出処理の関係性を調査する。model merge に用いる 2 つのモデルが異なるデータセットで学習されても、互いのカーネルの形状に類似性が現れる可能性について考える。また、それらのモデルを用いて model merge で生成されたモデルの性能が、前述した 2 つのモデルの性質と関連しているかを考

察する。これらの考察をもとに、CNN の特徴抽出の不透明性を解消していくことを目的とする。

2 model merge

CNN における model merge にはいくつかの方法がある。例えば、2 つのモデルを持つ各層を組み替えて、1 つの新たな大きいモデルを生成する手法や、同じ構造を持つ 2 つのモデルの重みを合成する手法 (図 1) がある。本研究では後者の方法を採用する。

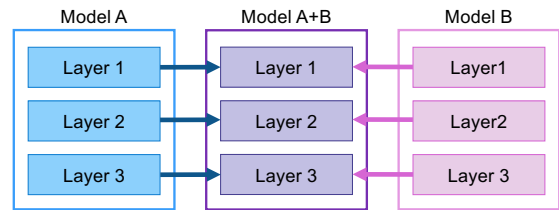


図 1 model merge の方法

画像分類タスクにおいて、CNN は分類に必要な何らかの特徴を、対象のデータセットから抽出し、その情報を変換、伝達している。しかし、データセットに含まれる画像のどのような特徴が抽出されているのかを解釈するのは困難であり、モデルの説明可能性を引き下げている。一方で、ImageNet で学習したモデルを用いた fine tuning や transfer learning は、モデルのより良い性能を引き出すことがあるが、その成功要因はわかっていない。そこで我々は、分類タスクに必要な情報はドメインによらず共通しているという仮定について考える。本研究では、model merge に 2 つの CNN を用いる。これらのモデルの構造は同じであるが、学習データが異なっている。仮定をもとに考えると、model merge に用いる 2 つのモデルは、浅い層において、特徴抽出の役割を担うカーネルの形状に共通性が存在すると考えられる。したがって、我々は model merge に用いる 2 つのモデルのカーネルの類似度を測定し、類似度行列を観察する。また、model merge で生成したモデルの性能が、仮定と関係しているかを考察する。

2.1 各モデルについて

model merge に用いる 2 つの CNN は、それぞれ CIFAR-10[3], CIFAR-100[3] で学習する。ここで、CIFAR-100 は

1) 東京都市大学大学院 総合理工学研究科 情報専攻
Informatics, Integrative Science and Engineering, Graduate
School of Tokyo City University

本来 100 クラス分類のデータセットだが、今回の実験では、実験の簡略化のために 20 クラスのスーパーラベルで分類を行った。また、model merge で生成したモデルは、CIFAR-100 を用いて fine tuning した。3 つのモデルの性能を表 1 に示す。

表 1 各モデルの性能

model	data	train accuracy	test accuracy
(a)	CIFAR-10	0.95	0.85
(b)	CIFAR-100	0.71	0.61
merge((a)+(b))	CIFAR-100	0.83	0.68

3 実験結果

model merge の元となる 2 つのモデル (a), (b) と、これらを用いた model merge によって生成されたモデルについて、それぞれの組み合わせで重みの類似度を測定した。図 2 には、各モデルの 1 層目の畳み込み層同士のコサイン類似度を測定した結果の類似度行列を表している。図中では、コサイン類似度を測定したのち、0.5 に設定した閾値を超えていたら 1、そうでなければ 0 として 2 値化している。縦軸、横軸はそれぞれ畳み込み層のフィルタの数に対応している。フィルタとは、カーネルをチャンネル単位で分解したものであり、1 層目は入力チャンネルが 3、出力チャンネルが 128 であるため、両軸の範囲は 0 から 383 である。

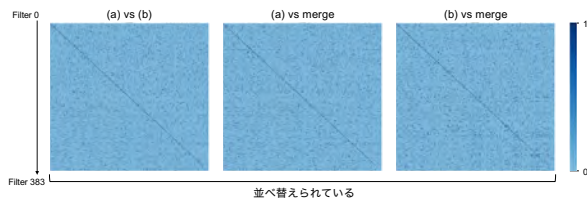


図 2 各モデルの重み同士の類似度行列

類似度行列の対角成分で類似度が高くなっていることが 2 から確認できる。この結果は、どの組み合わせでも似た形状のフィルタが存在していることを示している。ただし、この類似度行列は、片方のモデルで並べ替えが行われている。例えば (a) と (b) の行列については、モデル (a) の 0 番目のフィルタから順番に、最も類似度の高いモデル (b) のフィルタを探し出し、対角成分にその値が対応するように列を並べ替えた。この並べ替えを行わない場合、このような対角線は現れないことを我々は確認している。また、seed 値が同じ場合は、並べ替えを行わなくても対角線が現れることを確認している。したがって、1 層目の畳み込み層では、データセットによら

ず似た形状のカーネルが生成され、特徴を抽出していると示唆される。また、seed 値はフィルタの順番を決定していると考えられる。

表 1 に示す model merge によって生成されたモデルの性能に着目すると、CIFAR-100 で学習したモデル (b) よりも分類精度が向上していることがわかる。しかしながら、重みの類似度行列の結果からは、モデル (b) と model merge によって生成されたモデルの組み合わせにおいて特殊な結果は得られなかった。したがって、今回の実験では、model merge による性能向上と、分類タスクに必要な情報はドメインによらず共通しているという仮定には関係性が見られなかったが、少なくともなんらかの利点が存在していると言える。現状の結果から考えられる成功の要因は、CIFAR-100 からでは学習することができなかったカーネルを、CIFAR-10 のカーネルと混ぜることで獲得できたという現象であるが、より詳しい実験が必要である。

4 まとめ

本研究では、モデル再利用手法である model merge を利用し、CNN が画像分類を行うために必要な特徴について調査を行った。データセットによらず、画像分類に必要な情報が同じであるという仮定に関しては、今回の実験ではそのような傾向が示唆されたが、より詳細な調査が必要である。また、model merge によって分類精度が向上した要因と、model merge の元となるモデル同士の重みが類似している傾向の間には、特徴的な関係を見つけることができなかった。今後も引き続き、CNN の特徴抽出の傾向と、model merge の成功要因について、さらなる実験と解析を行い、解明を目指す。

謝辞

本研究の一部は JSPS 科研費 23K11266, 23H03387, 24K15115, 東北大学電気通信研究所共同プロジェクト研究, 東京都市大学重点推進研究未来知能ユニットの助成によるものです。

参考文献

- [1] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, L. Fei-Fei, "ImageNet Large Scale Visual Recognition Challenge," International Journal of Computer Vision, vol. 115, pp. 211–252, 2015.
- [2] T. Akiba, M. Shing, Y. Tang, Q. Sun, D. Ha, "Evolutionary Optimization of Model Merging Recipes," arXiv:2403.13187, 2024.
- [3] <https://www.cs.toronto.edu/~kriz/cifar.html>