

進化的手法に基づく深層学習の構築に関する研究

1 はじめに

近年、機械学習の発展は目覚ましく、広範な分野に適用され多くの研究がなされている。例えば深層ニューラルネットワーク (deep neural network; DNN) を用いた深層学習では、音声・画像・自然言語を対象とした問題で既存手法を大きく上回る高い性能を示し、強い関心を集めた。

一方で DNN は言葉通り層を深く重ねるため、層の種類、パラメータ、数、層同士の接続など多数のハイパーパラメータを持つ。このため問題に合わせて適切なパラメータを選択しなければ良好な精度が出ないが、そのネットワークやアーキテクチャの設計に明確な指針はなく試行錯誤する必要があった。

この問題を解決するために自動化された機械学習 (Automated Machine Learning; AutoML) と呼ばれる分野が注目されている。AutoML ではアーキテクチャ全体を対象とした最適化を目指す。本研究では AutoML を用いる前段階として、論文を紹介する。まず AutoML の分類を見た後、その例である Neural Architecture Search (NAS) と Auto Augment について示す。また予備的に実施した数値実験も示す。

2 要素技術

2.1 AutoML

自動化された機械学習 (Automated Machine Learning; AutoML)[1] は特定の技術ではなく、機械学習モデルの設計を自動化する全般的な手法、または概念を指す。

問題ごとにパラメータを適切に設定する必要があるというパラメータ設定問題の解決を目的とする。確率的なブラックボックス最適化であるこの問題の解決によって、

- 時間コストの削減・性能の向上
- アルゴリズムの評価、比較のためパラメータの最適性の影響を緩和
- パラメータが与える影響の知識の必要性を排除

の 3 つの利益が得られる。

AutoML はそのアルゴリズムによって、以下の 3 つの主要カテゴリと 5 つのサブカテゴリに分類される [1]。

主要カテゴリ 構造や構成による分類

1. 単純な生成・評価法
生成段階で候補となる設定を生成、評価段階で評価して最適な設定を見つける手法。
2. 反復的生成・評価法
少数の設定を生成して、最も優れたものを発見し、反復的に生成する新しい設定の指針とする手法。
3. 高レベルの生成・評価方法
高レベル生成機構として既存の自動パラメータ調整手法や探索手法を持ち、少数精鋭の設定を生成し評価段階では慎重に評価する手法。

サブカテゴリ 評価方法による分類

1. 繰り返し評価法
複数回の評価を、平均するなど統計的に評価する手法。
2. F-Racing
統計的に劣る評価の設定を段階的に排除し、有望な候補に計算を集中する手法。繰り返し評価より効率的になる。
3. インテンシフィケーション (強化)
問題のリストで候補設定と暫定設定の評価を次々比較し、劣る場合は途中で排除し、そうでなければ候補が暫定設定となる手法。
4. シャープニング
少ないテスト数で評価を始め、将来性のある設定のテスト数を 2 倍にすることで素早く探索できる手法。
5. アダプティブキャッピング
有望でない設定の実行を中断して計算量を削減できる手法。

2.2 Neural Architecture Search

従来の機械学習では手作業によって設計されたモデルをデータセットで学習して重みの最適化を行うが、アーキテクチャの設計には、高度な専門知識と手作業による構築が必要である。Neural Architecture Search(NAS)[2] は、ニューラルネットワークのアーキテクチャ自体を最適化する。

アーキテクチャの探索には3つの段階がある。まず最初にコントローラと呼ばれる再帰型ニューラルネットワーク (Recurrent Neural Network; RNN) で、アーキテクチャのハイパーパラメータを生成する。例えば畳み込み層を利用するネットワークでは、レイヤーごとにフィルタの高さ・幅、ストライドの高さ・幅、フィルタ数が必要となる。次にハイパーパラメータから子ネットワークを構築し、通常のように重みを訓練して検証データセットの精度を得る。最後に得られた精度で報酬を計算し、方策勾配法 (Policy gradient method) によってコントローラのネットワークを更新する。これらの手順を繰り返すことで、子ネットワークのアーキテクチャが最適化される。

NASでは子ネットワークとして、畳み込みネットワークと RNN を探索をし従来のネットワークより高い精度と少ないパラメータ数を達成したが、数百台のGPUと1ヶ月の時間がかかり計算量に問題があった。以降の研究では計算量の削減を目的とした手法が提案された。

佐藤ら [3] の研究では、NASを改良した先行研究と比較して、より高速なアーキテクチャ探索アルゴリズムを開発した。

特徴としてコントローラで生成していたカテゴリカルなアーキテクチャパラメータから、アーキテクチャを確率分布で連続的に表現し、分布推定を行う手法に変えることで、微分可能となった。これによって貢献度と呼ぶアーキテクチャの構成要素ごとの評価を導入して、一様な更新ではなく構成要素ごとの勾配で更新をすることができ、高速化ができた。

またNASではアーキテクチャ候補ごとに重みを学習し直していたが、ネットワークの重みを再利用することで大幅な学習時間の削減を達成した。

重みの再利用 深さや構造が異なるネットワークでの重みの再利用するとき、どの重みを使うかは自明でない。そこで非巡回完全有向グラフのノードを特徴、エッジを混合演算子としたネットワークをセルとして、アーキテクチャ構造を決めておく。この演算子には畳み込み層やプーリング層、活性化関数や恒等写像などが含

まれる。エッジの演算子候補に零写像を加えることで、ノード間の切断ができ様々なネットワークを表現できる。ネットワークのとり得る構造を網羅しているため、各エッジの各演算子ごとに学習した重みを使用することが可能となる。

2.3 AutoAugment

NAS に対してデータセットを対象とした AutoML も提案されている。画像識別問題で利用されるデータセットでは、画像データの左右反転や回転、色相や輝度の操作することでデータ数を増やすデータ拡張 (Data Augmentation; DA) が精度の向上に効果的であることが分かっている。しかし操作の種類と強度、操作の順序の組合せは膨大であり、現在データ拡張の実装は手動で設計されているため試行錯誤に時間がかかる。AutoML の一つである Auto Augment[4] はデータ拡張の空間を探索し、最高の検証精度となる最適な方策を見つける。

3 実験

今回は計算時間の関係から 2.3 で述べた Auto Augment に関わる実験をした。Auto Augment はデータセットを対象とした操作を最適化するが、データセットの持つ特徴を解析する実験として、次の予備実験をした。

以下では、画像データの分類問題を考える。異なるクラスラベルを持つサブデータセット集合と、それらを1つに結合したデータセットを作成する。集合内の各サブデータセットで学習した各分類器を用いて、結合したデータセット全体のクラス分類をする。

関連手法として、複数のネットワークの多数決をとり1つの識別器を構成するアンサンブル学習がある

3.1 データセット

データセットは、10クラスの画像データが含まれるCifar-10[5]を利用した。サブデータセットは、Cifar-10から10クラスの内のいくつかを抜き出し作成した。

3.2 モデルの構築

分類器には、16層の畳み込みニューラルネットワーク (Convolutional Neural Network; CNN) を用いた。実験1では作成したサブデータセットで、5クラス分類を

表 1: 実験の設定

Optimizer	SDG
momentum	0.9
lr	0.001
Loss	Cross Entropy Loss
batch size	64
epoch	100

学習した後、同様に学習した他のネットワークと組合せて 10 クラス分類をする。図 1 にこの関係を示した。

3.3 実験 1

実験 1 ではデータセットを 2 つの異なる 5 クラスサブデータセットに分割した。このデータセットはそれぞれ 16,000 枚の画像データが含まれる。クラスを表す番号の前半 (airplane, mobile, bird, cat, deer) で学習したモデルを A, 後半 (dog, frog, horse, ship, truck) で学習したモデルを B とする。

結合したモデル A + B は、10 クラス分類問題を解く。結合の際は 10 クラスのデータセットを A, B に入力し、得られた出力の次元を 10 クラスに拡張して和を求めた。

3.3.1 結果

図 2 にモデル A, B のテスト精度の結果を示した。モデル A が 86.5%, B が 91.2% となった。この精度はやや低いため、データ拡張による改善が望まれる。

結合モデル A + B の 10 クラス分類の結果は、69.2% となった。片方のモデルのみで分類した場合、未学習のクラスは識別できないため 46.5% となる。モデルの結合によってデータセットのクラス拡張に対応していることが分かる。

3.4 実験 2

あるクラスの分類を複数のモデルで実行することで、実験 1 より高い精度の獲得を目指した。7 クラスを持つサブデータセットを 3 つ作成し、学習するモデルをそれぞれ A, B, C とした。各データセットのクラスは、2 つ以上の分類器でクラスを推定できるように振り分けた。表 2 には、クラスの番号と共に選択したクラスを o で示した。

表 2: モデルごとの 7 クラスの振り分け

model	0	1	2	3	4	5	6	7	8	9
A	o	o	o	o	o	o	o			
B	o	o	o	o				o	o	o
C	o				o	o	o	o	o	o

3.4.1 結果

図 3 に 7 クラス分類の精度を示した。

7 クラス分類では、正答率が平均 80% 程度で、結合した結果 10 クラス分類では 86.7% となった。1 つのモデルで分類した場合は、58.6% となった。クラスが増え問題が困難になっているが、精度が高まる結果となった。

4 考察

モデルの結合によって、一部モデルが学習していないクラスでも正答できることが示された。特に実験 2 では、7 クラス分類の精度を 10 クラス分類の精度が上回り、アンサンブル学習的な結合の有効性が読み取れた。

図 2, 3 とともにインデックスが前半のクラスを持つモデルでは、正答率が低い傾向が見られた。再度学習を行っても変わらなかったため誤差の影響ではなく、困難なクラスの分類によって精度が下がっていると考えられる。

データセットにはクラスの潜在的な識別難度や、類似クラス間の識別における相互関係があり、複雑な作用が働いている。今回は均等になるようモデルにクラスを割り振ったが、困難なクラスは多くのモデルで推定し、簡単なクラスは 1 つのモデルだけにすることで、より結合の効果を高めることが考えられる。表 2 の組合せをバイナリで表現し、適切な組合せを進化型計算で求めるなどの AutoML による応用が考えられる。

5 まとめと今後の課題

ネットワークによって自動的にネットワークをメタ最適化する AutoML について紹介した。従来では人が行っていた工程を機械学習によって自動化しようという試みがある。一方でネットワークを複数回学習するため計算時間が問題となったが、探索時間を削減する様々な工夫が提案された。

本研究では時間の都合上予備実験をした。実験 1 ではクラスを拡張したデータセットでの、モデルの結合の

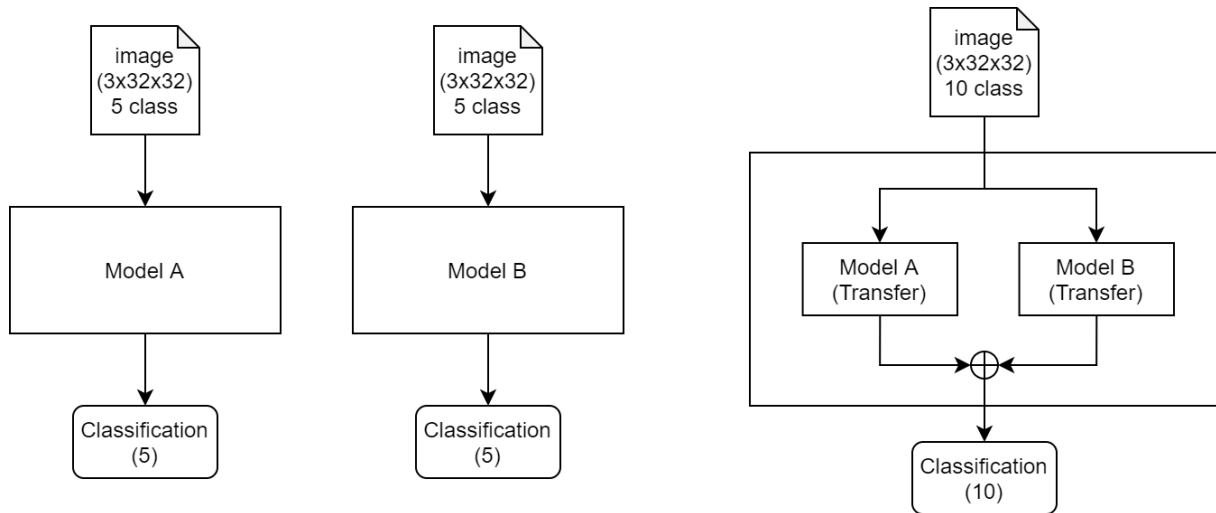


図 1: モデルの簡略図 () 内はデータの次元数

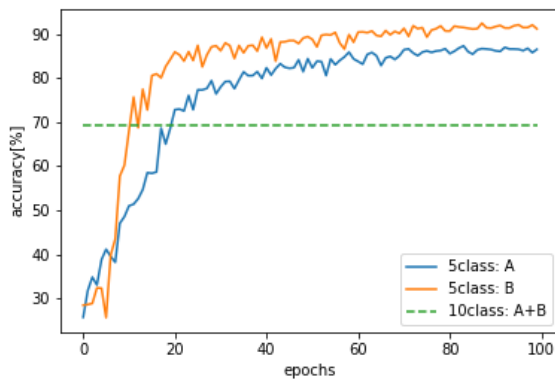


図 2: 実験 1 の 5 クラス分類器の正答率

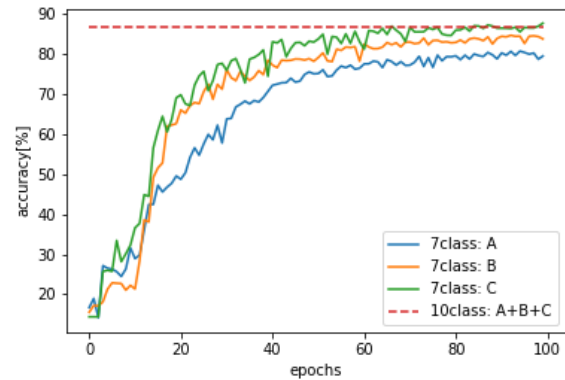


図 3: 実験 2 の 7 クラス分類器の正答率

有効性を見た。実験 2 では分類の結果を重ね合わせることで、テスト精度が非常に高まることが分かった。

今後の課題として、紹介したネットワークを探索する NAS ファミリのシステムを実装する。NAS にはそれを改良する様々な手法が提案されているが、佐藤ら [3] の確率的なアーキテクチャ表現をしたシステムを構築することを目標とする。

参考文献

[1] Changwu Huang, Yuanxiang Li, and Xin Yao. A survey of automatic parameter tuning methods for metaheuristics. *IEEE Transactions on Evolutionary Computation*, 24:201–216, 06 2019.

[2] Barret Zoph and Quoc V. Le. Neural architecture search with reinforcement learning. *CoRR*, abs/1611.01578, 2016.

[3] 佐藤 怜, 秋本 洋平, and 佐久間 淳. 貢献度分配を導入した方策勾配による neural architecture search の高速化. *人工知能学会全国大会論文集*, JSAI2019:2P3J202–2P3J202, 2019.

[4] Ekin Dogus Cubuk, Barret Zoph, Dandelion Mané, Vijay Vasudevan, and Quoc V. Le. Autoaugment: Learning augmentation policies from data. *CoRR*, abs/1805.09501, 2018.

[5] Alex Krizhevsky, Vinod Nair, and Geoffrey Hinton. Cifar-10 (canadian institute for advanced research).