

入力の変化を捉える深層学習モデルの定量化指標とそ
の構造の提案

2023年 3月

太田 望

入力の変化を捉える深層学習モデルの定量化指標とそ
の構造の提案

太田 望

理工情報生命学術院
システム情報工学研究群
筑波大学

2023年 3月

博士論文概要

学籍番号 202030165

氏名 太田望

第1章：序論

深層学習は2012年の第3次AIブームに伴い、研究開発が活発に進んでいる。国内のニューラルネットワーク関連の特許申請件数は2020年には1191件となっている。深層学習の中でも特に利用されているのが畳み込みニューラルネットワーク(CNN)である。CNNは畳み込み層とプーリング層によって繰り返し画像を処理することによって、画像中の線や点、物体の形状などの特徴を抽出することができる。中間層の出力を特徴マップと呼び、より出力側に近い特徴マップでは複雑で広域的な特徴を抽出しているとされる。抽出された特徴をもとに画像分類や特徴マッチングなどの画像処理タスクを高い精度で実行することができる。

CNNがコンピュータビジョンにおいて大きな成功を収め、実用化が進んだ原因として、CNNがシフト不変性を持つことが挙げられる。シフト不変性とは、入力画像を平行移動しても特徴マップが変化しない性質であり、これによって画像を少ないパラメータで処理することができ、過学習を回避し、少ないデータで高い精度を達成することができる。

CNNの問題点としては、画像の平行移動以外の変化、例えば回転や色の変化にはそのような不変性がないことである。そこで、本論文のリサーチクエスチョンを「入力画像の変化を捉える深層学習の構造はどのようなものか?」とする。この問題を解決するにあたって2つのアプローチが研究されている。

第2章：関連研究

1つはCNNに代わる新しい画像認識アーキテクチャを提案する試みである。カプセルネットワークは画像中の物体の姿勢の変化に対する頑健性を持ったネットワークとしてGoogle Brainによって提案された。カプセルネットワークは画像の部分的なプロパティを保存するカプセルを構成要素として持ち、複数の下位カプセルをより大局的な情報を持つ上位カプセルにつなげることで、画像全体のクラスとプロパティを予測するモデルである。本研究ではカプセルネットワークの問題点である、パラメータ数が多いために学習に計算コストがかかることと、上位カプセルが狭い範囲に分布するという問題点を距離学習手法の1つであるArcFaceの損失関数をカプセルネットワークの導入により軽減することである。

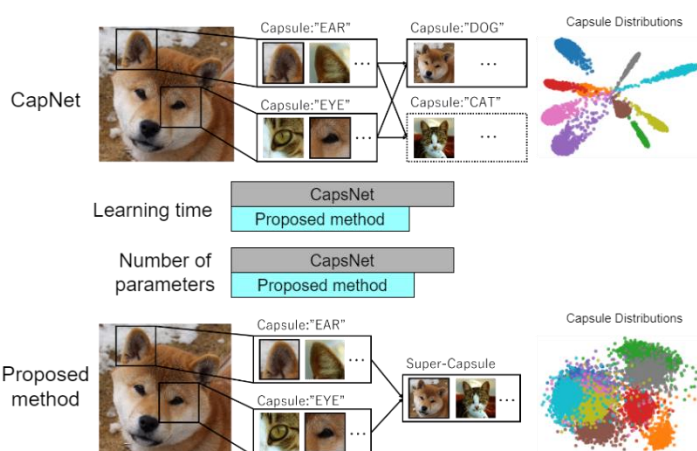
もう1つのアプローチはCNNに間接的に入力画像が変化した場合に特徴マップが変化するように学習させる手法である。この手法は自己教師あり学習の一種である。自己教師あり学習とは、教師データのない画像について自動的に生成できる疑似ラベルを用意し、疑似タスクをCNNに解かせることで特徴抽出器を学習させることである。現実的な問題の場合、学習に使える教師データ付きの画像は少ない場合が多いため、自己教師あり学習

によって CNN の大半のパラメータを予め学習させておくことにより、学習データが少なくとも過学習を防ぐことができる。このような自己教師あり学習の一種として、回転させた画像をモデルに入力し、画像が何度回転したものを疑似ラベルとする回転予測タスクや、RGB 画像を白黒画像に変換した後でモデルに入力し、各ピクセルの RGB 値を当てさせる色予測タスクなどが挙げられる。しかし、これらのタスクによって学習された特徴抽出器がどのような性質を持っているのかは自明でないため、本来のタスクに有用な特徴を抽出するためには試行錯誤が必要である。本論文では回転予測タスクにおいて、特徴マップがどのような性質を持つかを同変性という概念を用いて分析する。

第3章：距離学習に基づいたカプセルネットワークの拡張

【研究概要】

カプセルネットワーク (CapsNet) は、Google Brain によって提案された、画像中の物体の姿勢の変化に対する頑健性を備えた画像分類のための深層学習モデルである。カプセルとは、その方向が物体の存在、位置、大きさ、姿勢など、物体の特徴を表すベクトルである。



しかし、CapsNet では、カプセルの分布がクラスごとに集中し、クラスの数に応じてカプセルの数と計算量が増加する。そこで、1つのカプセルで複数の物体クラスを表現できるようにすることで、カプセルの方向の多様性を高め、CapsNet の学習の計算量を削減する手法を提案する。クラス間の距離を学習するために ArcFace と呼ばれる損失関数を用いる。

【背景】

深層学習による画像処理の重要な分野として、特徴抽出器による画像の低次元圧縮がある。高次元な画像データの情報を低次元の特徴ベクトルに圧縮することができれば、クラス分類や特徴マッチングなど、あらゆる画像処理タスクに応用が可能であるためである。しかし、現在の主流の深層学習モデルである畳み込みニューラルネットワークはプーリング層によって情報の一部を捨てている。そこで Google Brain によってプーリング層を使わない CapsNet が提案されたが、後述するように CapsNet には未解決の問題があるため広く使われていない。

【問題点・提案手法】

CapsNet の問題点は2つある。カプセルの特徴空間上の分布がクラスごとに直線上に集中すること、クラスの数に応じてカプセルの数と計算量が増加することである。そこで、1つのカプセルで複数の物体クラスを表現できるようにすることで、カプセルの方向の多様性を高め、CapsNet の学習の計算量を削減する手法を提案する。具体的には、クラスごとに代表ベクトルを用意し、代表ベクトルとの位置関係に分類と画像の特徴の表現を両立

するために ArcFace と呼ばれる損失関数を用いる。

【実験・結果】

提案手法の検証のため、MNIST, fashion-MNIST, EMNIST, SVHN, CIFAR-10 データセットと、それに対応するアフィン変換されたデータセットを用いて、提案手法とオリジナルの CapsNet の精度と学習時間を測定した。提案手法の精度は CIFAR10 データセットで 8.91% 向上し、学習時間は各データセットで 19% 以上短縮されることが確認された。また、主成分分析を用いて従来手法と提案手法のカプセルの分布を可視化し、主観評価を行った。

第 4 章：正準相関分析に基づいた同変性指標

【研究概要】

自己教師あり学習は、人工的に生成したラベルデータを用いて、汎用的な特徴抽出器を学習するプロセスである。画像の特徴表現には、用途に応じて不変性と同変性が望まれる。不変性とは、入力画像に画像の拡大縮小や回転などの変換を加えても、その前後で特徴表現が変わらない性質のことで、これに対して同変性とは、入力画像に画像の拡大縮小や回転などの変換に応じて、特徴マップが変化する性質のことである。自己教師あり学習では、ある変換に対して不変であるか同変であるかは、後続のターゲットタスクに大きく影響する。しかし、同変性を測定する方法はまだ提案されていない。本研究では、特徴量の表現観察と正準相関分析に着想を得て、新たな同変性指標を提案した。MNIST, fashion-MNIST, CIFAR-10 データセットを用いて、自己教師あり学習の一種である回転予測タスクで学習したエンコーダに対する、提案する同変性指標の挙動を検証した。

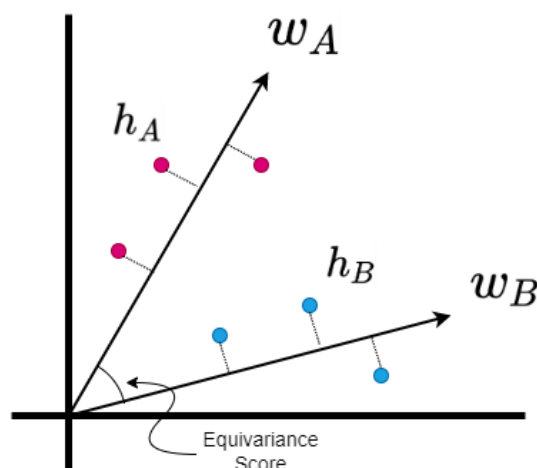
【背景】

自己教師あり学習では画像の特徴抽出器を学習させる。これらの特徴抽出器は人手で作った教師データを用いることなく、様々なタスクに転用することができる。特徴抽出器のパラメータは自動生成可能なラベルによって事前に学習され（疑似タスク）、目的のタスク（後続タスク）に利用される。疑似タスクでは元のラベルデータを利用できないため、自己教師あり学習は少数ショット学習やドメイン適応など、データが限られた状況での学習にも応用される。自己教師あり学習は入力データの拡張に対して不変性を持つものと同変性を持つものに分類される。同変な自己教師あり学習では、データの変換前と変換後のデータの違いを学習させる。同変な特徴マップはいくつかのコンピュータビジョンタスクで有用であることが示されている。

【問題点・提案手法】

しかし、ある変換に対して不変あるいは同変な自己教師あり学習は、後続のタスクにとって有害な場合がある。例えば、色変換に不変なように学習した特徴抽出器は、物体を色で分類するような後続タスクでは性能が悪くなる可能性がある。したがって、自己教師あり学習の設計者は、後続のタスクがどのような変換に対して不変または同変であるべきかを理解する必要がある。不変性と同変性を組み合わせた自己教師あり学習に関する研究では、後続タスクがどのような変換に対して不変であるか、どの程度同変であるかによって性能が大きく変化することが明らかにされている。これら複数の変換を組み合わせた自己教師あ

り学習の問題点は、特徴空間における同変性の測定方法が確立されていないため、データ拡張のためのハイパーパラメータの探索が試行錯誤になることである。しかし、変換に必要な同変性は未知のままである。そこで本研究では、正準相関分析に基づき、同変性の強さを表す尺度を定義した。次に、疑似タスクに回転予測を用い、後続タスクに分類を用いた場合の自己教師あり学習した特徴量の分布と性質について考察した。



【実験・結果】

提案する同値性スコアを検証するため、MNIST、fashion-MNIST、CIFAR-10 という 3 つのデータセットにおいて、回転予測タスクを疑似タスクとして用い、特徴抽出器の学習を行った。後続タスクの分類精度と特徴抽出器から得られた同変性指標と比較した。結果として、同変性指標は中間の大きさのものが最も後続タスクの精度が高かった。同変性が大きいものは疑似タスクが簡単で後続タスクに有用な特徴を抽出できず、同変性が小さい場合は疑似タスクがデータセットによって難しく疑似タスクの学習に失敗していると考えられる。また、主観評価としては、特徴マップを t-SNE によって低次元可視化し、提案指標との整合性を確認した。

第 5 章：結論

第 3 章では入力画像の変化を特徴マップに反映しやすいモデルの構築という観点からカプセルネットワークの改良を行った。特徴マップの分布を制御するために ArcFace と呼ばれる距離学習手法を損失関数に導入し、入力画像の微小変形に対する頑健性を向上させることができた。また、学習時間やパラメータの削減にもつながった。

第 4 章では同変性誤差と同変性感度の 2 つの指標の提案を行い、その妥当性を検証した。また、同変性感度の大きさが後続タスクの精度に関係するという知見が得られた。

本論文のリーディング・レスキューは「入力画像の変化を捉える深層学習の構造はどのようなものか？」であった。第 4 章の回転予測タスクのように、入力画像の変化を教師データとして利用する場合は CNN でも同変性が確認できた。それに対し、第 3 章の提案手法では教師データを利用せずに回転に対する同変性が得られたため、入力画像の変化を捉える構造と言える。また、同変性感度が大きすぎる、または小さすぎる場合に分類精度の悪化につながる、同変性感度を調整するためにスケール係数が有効であるなどの知見が得られた。

本論文の残された課題としては、回転以外の変換に対する同変性について実験を行っていないこと、データセットを変えた場合や分類以外のタスクでの最適な同変性感度については未解決であることが挙げられる。今後の展開としては、上記の課題の解決や、同変性指標をもとに学習法の開発、またはモデルを設計することを目指す。