

複数階層のカテゴリー分類を行うニューラルネットワーク

○渡邊響輝（日大工・情報・院生） 田中宏卓（日大工・情報）

和泉勇治（日大工・情報） 加瀬澤正（日大工・情報）

1. まえがき

近年、機械学習の技術向上により、画像中の物体認識技術は目覚ましく向上している。

しかしながら、そのほとんどは既に学習された物体を分類するための技術・手法であり、我々が日常的に行っている、未知の物体についての分類問題は未解決のままである。

そのため、このような問題の解決を目的として、最近の研究では、画像を用いて、複数の階層において同時に認識するための手法に関する検討が進められている。この手法が確立できれば、未学習の物体であっても、その物体が上位階層のどのカテゴリーに属するかを推論したり、少ない画像を用いた学習のみにより、その物体を認識対象として新たに追加したりすることができる。

そこで、本研究では、複数階層のカテゴリー分類を行うニューラルネットワークの構築を目的とし、いくつかのモデルを用いて、物体の上位階層の表現学習について検討を行う。

2. 複数階層のカテゴリー分類を行うニューラルネットワーク

物体は階層的なカテゴリー構造を持っている。猫を例に取って考えるとき、そのカテゴリーの構造は、たとえば図1のように表すことができる。

以下では、図中の最下層、すなわち「猫」を含む階層を小分類と呼ぶことにする。また、「哺乳類」を含む階層を中分類、「動物」を悩む階層を大分類と呼ぶことにする。

本研究の目的は、このような階層的なカテゴリー構造の大分類・中分類、小分類の3つの階層において同時にカテゴリー推論を行うニューラルネットワークモデルを構成することである。

このような目的のもとに筆者が提案したネットワークモデル¹⁾を図2に示す。このモデルはCNNをベースとしており、始めに小分類についてのカ

テゴリー推論結果を得る。その後、プーリングを介して得られるサイズの小さな抽象度の高い特徴マップに基づき、中分類、大分類の順にカテゴリー推定結果を得る。

3. 実験1

3. 1 概要

実験1では、既存のニューラルネットワークモデルが、潜在空間において、上位階層の共通概念の表象である大分類・中分類をどの程度表現できているのかについて調べる。

今回想定する階層構造を表1に示す。実験では、表1の大分類3クラス、中分類14クラスのいずれであるかをそれぞれ出力するように、前述した筆者が提案したネットワークモデルと、出力層のみをファインチューニングした Inception v3^[1]、MobileNet v2^[2]、NasNetMobile^[3]、ResNet50^[4]、の4種類のモデルを使用する。

今回の実験では、学習には使用していない未知物体を集めたテスト用のデータセットを入力として用いることにする。小分類クラスひとつにつき、テスト用画像は50枚を使用する。サイズは112×112とし、カラー画像とする。また、潜在空間のベクトルとして、各モデルの出力層の直前の層のベクトルを使用し、大分類に注目する。

3. 2 実験結果

ファインチューニング後の各モデルの、大分類の各クラスに属する物体の特徴ベクトルの平均ベクトルを用いて算出した大分類におけるクラス間距離を表3に、標準偏差ベクトルのノルムを表4に示す。ここで、標準偏差ベクトルのノルムとは、当該クラスに属する物体の特徴ベクトルを用いて算出した多次元分散ベクトルのノルムである。

また、クラス間距離を標準偏差ベクトルのノルムで除算した値を表5に示す。この値を分離度と呼ぶことにする。

4. 3 考察

今回実験に使用したモデルの中で、特に分離度が大きくなったのは提案ネットワークモデルと MobileNet v2 で、特に小さくなったのは Inception と NasNetMoibile であった。出力ベクトルを用いて計算したクラス間の距離や分散について、クラス間距離が大きいほど各クラスが潜在空間上で分かれており、分散が小さいほど分類結果のばらつきが小さくなると考えられるので、分離度は大きくなるほどクラス間の分類が明確にできていることになる。

今回の実験結果によれば、提案モデルや MobileNet v2 のように、通常の畳み込み層や空間方向とチャンネル方向で分けて行う畳み込み処理を利用した構造をもつモデルが複数階層のカテゴリ分類には適しているものと考えられる。

一方で、同様の構造を ResNet も持つのに分離度が小さくなったのは、畳み込み処理の出力値に入力値を足し合わせる、残差ブロックと呼ばれる構造を接ぎ合わせた構造をしていることが影響していると考えられる。

しかし、このことを確かめるためには、類似した構造をしている DenseNet や、残差ブロックの概念を Inception に取り入れたモデルである Inception-ResNet について同様の実験を行うことが必要だと考えた。

また、中分類についてはどのような結果になるのかについて今後検証したい。

5. 実験 2

5. 1 実験概要

実験 2 では、実験 1 で用いたネットワークモデルのカテゴリ分類の精度を確かめる。実験 1 で用いたテスト用のデータセットを当該モデルに入力し、大分類、中分類階層における認識性能を検証する。

5. 2 実験結果

未知物体の画像を入力したときの、大、中分類における正答率をそれぞれ表 6、7 に示す。

縦軸がモデル名、横軸がカテゴリを表す。

5. 3 考察

大分類において、最も平均正答率が高くなったのは提案モデルで、次点が ResNet であった。

しかし、これらのモデル、特に ResNet はクラスごとの正答率に大きな差が生じており、すべてのクラスの特徴を捉えているとは言い切れない。

一方で、MobileNet の平均正答率は低いものの、カテゴリ間の正答率の差は他のモデルよりも比較的小さい。次点でカテゴリ間の正答率の差の小さなモデルは提案モデルであることから、実験 1 で算出した分離度が大きいモデルはカテゴリ間の正答率の差が小さくなる傾向があると考えられる。

これらの仮説を検証すべく、他の既存モデルや新しい構造を持つ自作モデルに対して同様の実験を行い、この評価方法の妥当性を調べる必要があると考えた。

前述した傾向が正しいとわかれば、モデルやその層の出力ベクトルの分離度に注目することで、複数階層のカテゴリ分類に適したモデル構造が発見できるはずである。

また、同様に複数階層のカテゴリ構成そのものにも最適なものが存在するはずであるので、Calinski-Harabasz といった最適なクラスター数を求めることに利用できる指標を用いて、カテゴリ構成も最適なものを求める必要があると考えた。

6. 結言

今回の実験では、提案モデルと既存のネットワークモデルが、上位階層である大分類・中分類についてどの程度表現できているのかということと、それぞれのモデルのカテゴリ正答率ほどの程度であるか確かめた。

今後は今回比較に使用したモデルと似た構造のものを含む複数のモデルで、出力ベクトルを得る層の位置を変えるなどしながら同様の実験を行う。

それらを通して、実験で用いた分離度の妥当性や、分離度を用いた新しい構造のモデルの作成を行っていきたい。

また、並行して最適なクラスター数などについて

でも検討し、最適な階層的カテゴリー構造についても模索していきたい。

参考文献

- 1) Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jonathon Shlens, Rethinking the Inception Architecture for Computer Vision, CVPR 2016
- 2) Andrew G. Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, Hartwig Adam, MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications (<https://arxiv.org/abs/1704.04861>)
- 3) Barret Zoph, Vijay Vasudevan, Jonathon Shlens, Quoc V. Le, Learning Transferable Architectures for Scalable Image Recognition, CVPR 2018
- 4) Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian SunDeep, Residual Learning for Image Recognition, CVPR 2016

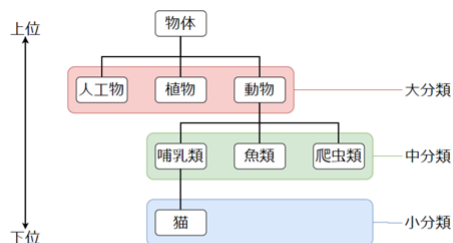


図1 階層的なカテゴリーのイメージ

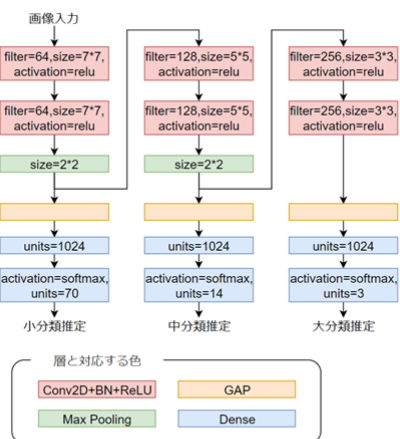


図2 提案ネットワークモデルの構成

表1 実験で想定する階層構造

大	中	小分類		
①	①	ボーダーテリア, ペルシアン猫, バグ, 豚, シマウマ		
	②	コモドドラゴン, ナイルワニ, ニシキヘビ, アカウミガメ, エリマキトカゲ		
	③	金魚, シビレエイ, オニカマス, ホホジロザメ, クマノミ		
	④	クジャク, ワシ, ニワトリ, ダチョウ, コンドル		
	⑤	コガネグモ, テントウムシ, ハエ, バッタ, トンボ		
②	⑥	パイナップル, レモン, ザクロ, イチジク, バナナ		
	⑦	キャベツ, キュウリ, イチゴ, トウモロコシ, ビーマン		
	⑧	キノコ, マイタケ, ハラタケ, シロソウメンタケ, スッポンタケ		
	⑨	ヒナギク, たんぽぽ, バラ, ひまわり, チューリップ		
	⑩	刑務所, 高架橋, レストラン, 灯台, 噴水		
③	⑪	新幹線, マウンテンバイク, 旅客機, ミニバン, タクシー		
	⑫	本棚, 四柱式ベッド, 机, ロッキングチェア, 扇風機		
	⑬	シャベル, ハンマー, 手斧, 定規, ボールペン		
	⑭	Tシャツ, ジーンズ, ブラジャー, セーター, ジャケット・コート		
中分類		大分類		
①	哺乳類	⑩ 建築物	①	動物
②	爬虫類	⑪ 乗り物	②	植物
③	魚類	⑫ 家具	③	人工物
④	鳥類	⑬ 道具		
⑤	昆虫	⑭ 衣類		

表2 テスト用画像詳細

大	中	小分類
①	①	キットギツネ, レッサーパンダ, ゴリラ
	②	イモリ, キスイガメ, ウィップテール・リザード
	③	ウナギ, ダツ, アカエイ
	④	フクロウ, メキシコカワガラス, 鴨
	⑤	ハンミョウ, オサムシ, 蜂
②	⑥	オレンジ, グラニースミス, パラミツ
	⑦	ブロッコリー, ズッキーニ, バターナッツかぼちゃ
	⑧	シャグマアミガサタケ, ボルチーニ, テングタケ
	⑨	アジサイ, ハイビスカス, ユリ
③	⑩	納屋, ダム, 城
	⑪	飛行船, モーターボート, リムジン
	⑫	食器棚, 折り畳みいす, 洗濯機
	⑬	傘, モップ, カップ
	⑭	毛皮のコート, トレーナー, ボンチョ

クラス間距離	動物～植物	動物～人工物	植物～人工物
自作	0.38027121	0.06407065	0.380271212
Inception v3	30.5076876	5.458157031	36.38737193
MobileNet v2	0.06257862	0.062675881	0.279726569
NasNetMobile	3.93936268	3.447949952	2.41758824
ResNet 50	0.00166849	0.001911893	0.040907822

表3 大分類におけるクラス間距離

標準偏差ベクトルのノルム	動物	植物	人工物
自作	0.483363	0.62871	0.472321
Inception v3	957.3242	1122.771	933.8516
MobileNet v2	0.429106	0.369319	0.49825
NasNetMobile	170.3208	179.9185	183.181
ResNet 50	0.302118	0.297112	0.313789

表 4 標準偏差ベクトルのノルム

分離度	動物～植物	動物～人工物	植物～人工物
自作	0.34194809	0.067041655	0.345377477
Inception v3	0.017493126	0.002886118	0.017692778
MobileNet v2	0.350347866	0.067585549	0.322425738
NasNetMobile	0.006902674	0.009753697	0.006658197
ResNet 50	0.068267329	0.003104188	0.066963095

表 5 大分類の分離度

正答率	動物	植物	人工物	平均
自作	0.529333333	0.821666667	0.713333333	0.688111111
Inception v3	0	0	1	0.333333333
MobileNet v2	0.442666667	0.448333333	0.282666667	0.391222222
NasNetMobile	0.108	0.853333333	0.136	0.365777778
ResNet 50	0.456	0.025	0.828	0.436333333

表 6 大分類における各モデルの正答率

正答率	哺乳類	爬虫類	魚類	鳥類	昆虫
自作	0.166667	0.106667	0.42	0.173333	0.313333
Inception v3	0	0	0	0	0
MobileNet v2	0.106667	0.113333	0.033333	0.066667	0.013333
NasNetMobile	0	0	0	1	0
ResNet 50	0.046667	0.02	0.033333	0.38	0.093333

	果物	野菜	キノコ	花	建築物
自作	0.38	0.38	0.826667	0.646667	0.666667
Inception v3	0	0	0	1	0
MobileNet v2	0.08	0.526667	0	0.04	0.093333
NasNetMobile	0	0	0	0	0
ResNet 50	0	0.033333	0	0.013333	0.113333

	乗り物	家具	道具	衣類	平均
自作	0.533333	0.4	0.34	0.18	0.395238
Inception v3	0	0	0	0	0.071429
MobileNet v2	0.093333	0.006667	0.08	0.033333	0.091905
NasNetMobile	0	0	0	0	0.071429
ResNet 50	0.033333	0.26	0.313333	0.466667	0.129048

表 7 中分類における各モデルの正答率