6Q-08

画像の劣化・ノイズによるDNNモデルの較正と分布外汎化への影響

Measuring the Effect of Image Corruption and Perturbation
through the Lens of Calibration and Out-of-Distribution Generalization

多田 圭吾 *1　長沼 大樹 *2*3
Tada Keigo　Hiroki Naganuma

*1立命館大学 Ritsumeikan University
*2モントリオール大学 Université de Montréal
*3Mila Mila - Quebec AI Institute

カメラ映像による人流解析や自動運転などのシナリオでは、用いるカメラレンズの経年劣化や、気象条件の変化などの環境変化によって、 深層ニューラルネットワークの性能を劣化させることが課題である。我々は、これらの環境変化をデータの分布シフトとしてキャストし、分布外汎化問題に取り組む。特に分布シフト下での影響を 19 種類の画像の劣化パターンとして分類し、性能劣化と不確実性の推定への影響への調査を行った。

1. はじめに

深層ニューラルネットワーク（DNN）の学習において、ERM (Empirical Risk Minimization) などの既存の学習アルゴリズム [Vapnik 91] は、推論時のデータが学習データに含まれている場合においては高い性能を発揮することが知られている。しかしながら、学習データとは異なる環境からサンプルされたデータ（分布外データ）に対してはその性能が劣化することが知られている [Arjovsky 21]。

また、DNN の別の側面として、高い予測のランキング性能を持つモデルであっても、古典的な統計的モデルに比べ、その不確実性が信頼できるものでないことが問題視されている [Guo 17]。特に、自動運転などのシナリオでは、意思決定において不確実性の信頼性が不可欠である。

本研究では、自動運転などのシナリオにおける画像の劣化・ノイズの影響を、分布シフトの問題にキャストし、各シフトがモデル性能 (分布外汎化性能・不確実性) に及ぼす影響について調査する。また、現在広く利用されている不確実性の較正手法を用いて、各シフトに対してどの程度較正が有効であるのか検証を行う。

2. 背景

2.1 分布外汎化

深層ニューラルネットワークにおける教師あり学習は、一般的に、訓練データとテストデータが同じ分布から抽出されるという条件を仮定する。ERM では、この独立同一分布 (IID) の仮定に基づき、正則化付き経験損失の最小化を行うことで、テストデータの損失を最小限にすることを期待する [Vapnik 91]。しかしながら、実社会のアプリケーションにおいては、テストデータの分布が訓練データの分布と異なることが一般的であり、その分布の差異、すなわち学習・推論時の環境変化は、従来の教師あり学習の前提に反する。例えば、自動運転技術への応用において、晴天時・降雪時の道路状況は異なるため、訓練データに晴天時の道路画像しか含まれていない場合、降雪時の運用において画像内に雪や雨風が映り込むと認識精度が著しく劣化する。IID の仮定に従わない条件でのテスト環境での汎化性能を分布外汎化と呼び、この性能を向上させることが実応用における喫緊の課題である。

2.2 キャリブレーション (Calibration Metric)

統計的モデルの評価指標として、精度として知られるランキング性能だけでなく、自動運転などのシナリオにおいては不確実性の正しさが重要となる [Abdar 21]。例えば，同じ認識精度のモデルであっても，走行時に人の出現を推定する Segmentation タスクなどのシナリオでは、確信度が 99%と 50%では意思決定における意味合いが異なる。

ここで、不確実性の正しさ（すなわち確信度 90% のデータが 100 件あった場合、90 件は正解である）が意思決定には欠かせないが、昨今の深層ニューラルネットワークモデルは、確信度と精度にギャップがあることが実応用への障壁となっている [Guo 17]。この差異を定量化する指標として、キャリブレーション (ECE: Expected Calibration Error) が広く用いられており [Roelofs 21]、我々もこの指標を実験に用いている。

2.3 較正手法 (Calibration Method)

不確実性の正しさの問題に対して、モデルが出力する確信度と精度の間に生じるギャップを較正することを考える。較正には、モデルの学習時に補正を行う In-training 手法と、モデルの出力を事後解析し、不確実性の較正を行う Post-hoc 手法が存在する。本研究では、Post-hoc 手法の代表例として広く用いられている Temperature Scaling(TS) を用いる。[Guo 17]

$$\hat{\mathrm{q}}_{\mathrm{i}} = \max_{k} \sigma_{\mathrm{SM}} \left(\boldsymbol{p}_i / T \right)^{(k)} \tag{1}$$

ここで、T はモデル出力を較正するパラメータであり、$\boldsymbol{p}_i$ はモデル出力を示す。

3. 分布シフト

3.1 画像の劣化・ノイズによる分布シフト

画像認識モデルを実応用する際に想定される分布シフトを対象として、モデルの分布外汎化性能と不確実性の評価を行う。シフトを大きく以下の 3 つに分けて実験を行う。(i)Corruption Shift: ブレや輝度変化がノイズとなり生じるシフト (ii)Perturbation Shift: 回転やスケールなどの摂動が加わることで生じるシフト (iii)Sampling Bias: 分類対象は同じであるが、偏った標本抽出によって生じるシフト。我々は (i) ～ (iii) を想定して作成されたデータセットを使用し、モデルの頑健性を評価する。

3.2 データセット

CIFAR10, ImageNet を用いて学習を行なったモデルの分布外汎化性能と不確実性を評価する。それぞれの分布外データセットは次のものを使用する。CIFAR10 をベースとした分布外データセット：CIFAR10.1 [Recht 18], CIFAR-10-C, CIFAR-10-P [Hendrycks 19], ImageNet をベースとした分布外データ

連絡先: is0463hx@ed.ritsumei.ac.jp

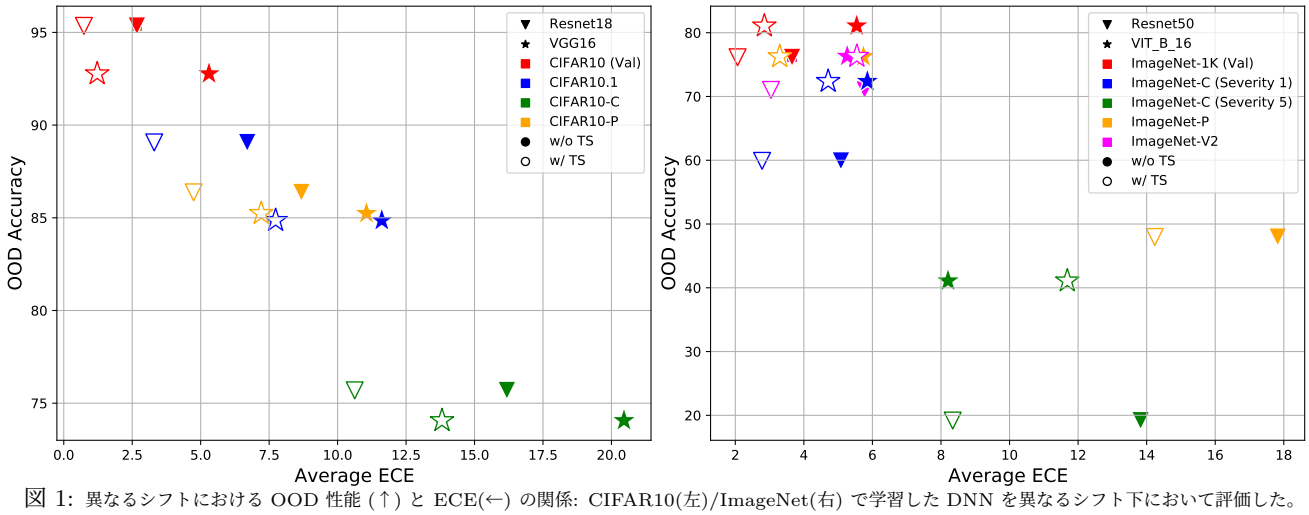

図 1: 異なるシフトにおける OOD 性能 (↑) と ECE(←) の関係: CIFAR10(左)/ImageNet(右) で学習した DNN を異なるシフト下において評価した。

セット：ImageNet-V2 [Recht 19], ImageNet-C, ImageNet-P [Hendrycks 19]。

4. 実験

異なる画像の劣化・ノイズがモデルのランキング性能・キャリブレーションに及ぼす影響を調査するため、元分布として CIFAR10, ImageNet データセットで ERM により学習を行った、DNN モデル (各 2 種類) に対し、分布シフト先での汎化性能 (ランキング性能) と、キャリブレーションの評価を行う。また、2.3 章で紹介した TS を用いて ECE の較正を行い、各シフトに対する TS の有効性を検証する。

5. 結果・考察

図 1 に示す通り、CIFAR10 の実験においては、分布外環境におけるランキング精度 (OOD Accuracy) と ECE には相関があることが明らかになった。つまり、OOD Accuracy と ECE どちらかを計測できれば、もう一方を推定することが可能であることを示唆している。さらに、実験を行った全ての分布シフトにおいて TS 後の ECE が減少しており、TS 後の ECE についても OOD Accuracy と順位相関を保っている。

対して、ImageNet の実験においては、必ずしも OOD Accuracy と ECE の順位相関が保たれないが傾向として比例関係を持つ結果となった。さらに、Perturbation Shift 下では、全ての分布外データで TS による較正が有効であったものの、Corruption Shift 下では、TS による較正が必ずしも有効でないことが示唆さる結果となった。これは、TS の原理的に temperature を求めるために用いる学習ドメインの Val ECE は下がることが保証されるものの、OOD 環境である Test ECE が下がるとは限らないことに起因する。

6. おわりに

本研究では、深層ニューラルネットワークが実社会のアプリケーションにおいて避けられない分布シフトを対象に、異なるシフトを含むデータセットを用いて分布外汎化性能を評価した。また、各分布シフト下における ECE に較正を施し、その改善度合いについて検証を行なった。実験の結果、CIFAR10 を用いた実験では、分布外データにおいて OOD Accuracy と ECE の間に高い相関関係があることが示され、TS による較正後もその関係性が保たれていることが判明した。一方、ImageNet を用いた実験においては、OOD Accuracy と ECE の間にある程度の相関は見られたものの、CIFAR10 の実験と比較して分散が大きくなる結果となった。特に Corruption Shift 下において TS 後の ECE が悪化するケースが複数見られたことから、これらのシフトを対象にした較正手法が深層ニューラルネットワークの応用には必須であることが示唆された。本研究では TS を用いてモデル出力の較正を行なったが、モデルの信頼度を学習プロセスにおいて補正する In-training 手法を用いて学習されたモデルが分布シフト下においてどのような振る舞いをするのか検証し、各手法が有効に働く環境を調査する必要がある。

参考文献

[Abdar 21] Abdar, M., Pourpanah, F., Hussain, S., Rezazadegan, D., Liu, L., Ghavamzadeh, M., Fieguth, P., Cao, X., Khosravi, A., Acharya, U. R., et al.: A review of uncertainty quantification in deep learning: Techniques, applications and challenges, *Information Fusion*, Vol. 76, pp. 243–297 (2021)

[Arjovsky 21] Arjovsky, M.: Out of Distribution Generalization in Machine Learning (2021)

[Guo 17] Guo, C., Pleiss, G., Sun, Y., and Weinberger, K. Q.: On calibration of modern neural networks, in *International Conference on Machine Learning*, pp. 1321–1330PMLR (2017)

[Hendrycks 19] Hendrycks, D. and Dietterich, T.: Benchmarking Neural Network Robustness to Common Corruptions and Perturbations, *Proceedings of the International Conference on Learning Representations* (2019)

[Recht 18] Recht, B., Roelofs, R., Schmidt, L., and Shankar, V.: Do CIFAR-10 Classifiers Generalize to CIFAR-10? (2018), https://arxiv.org/abs/1806.00451

[Recht 19] Recht, B., Roelofs, R., Schmidt, L., and Shankar, V.: Do imagenet classifiers generalize to imagenet?, in *International Conference on Machine Learning*, pp. 5389–5400PMLR (2019)

[Roelofs 21] Roelofs, R., Cain, N., Shlens, J., and Mozer, M. C.: Mitigating bias in calibration error estimation (2021)

[Vapnik 91] Vapnik, V.: Principles of risk minimization for learning theory, *Advances in neural information processing systems*, Vol. 4, (1991)