ISSN 2186-5647

OS-2

Wasserstein 距離による特徴マップの異常性スコアを基準とした CNN 特徴量の低次元化アルゴリズム

日大生産工　○大前 佑斗　　柿本 陽平　　豊谷 純

1. CNNのホワイトボックス化技術

畳み込みニューラルネットワーク（Convolutional Neural Networks; CNN）は入力画像から特徴抽出を行うことで多様な表現能力を獲得することができる。従来の CNN は推定根拠がブラックボックスであるという欠点があったものの、近年はホワイトボックス化技術（例えば、Class/Regression Activation Map; CAM [1] /RAM [2]など）が発展している。これは、Global Average Pooling レイヤ [3] 直前の特徴マップと、それにより獲得された CNN 特徴量ベクトルに対し、回帰ないしはクラス分類問題を解くように学習されたウェイトパラメータを利用することで、推定根拠の可視化を実現する。従来の CAM/RAM にはモデルの形状が限定されてしまうという欠点があったが、現在ではその問題が解消された Grad-CAM/RAM [4, 5] も登場している。

CNNのホワイトボックス化技術は重要であり、特に推定根拠の信頼性が要求される領域では、標準的に利用されている（例えば、医療支援に活用する CNN [6] など）。

2. 汎化誤差の評価に関する問題

従来の CNN の信頼性評価のプロセスでは、学習に使用しないテストデータを別途用意しておき、それにより汎化誤差を推定するというアプローチが取られる。これにより、テストデータの誤差の平均値が低ければ良好なモデルと解釈する。しかし、良好なモデルと判断された CNN についても、個々の推定時に得られた推定根拠を確認すると、不適切なものが含有されることがある。例えば、Saito et al. [7] は、クロスバリデーションエラー最小化基準により、患者の胸部 X 線画像から心臓の状態を推定する CNN を構築した。CNN に付与したタスクの特性上、心臓に推定根拠が集中していることが望ましい。多くの入力画像についてはこれを満たしたが、一部の入力画像については、不適切な可視化根拠が現れた（例えば、心臓の状態推定であるにも関わらず、肩や食道を根拠にしてしまっている、など）。Omae et al. [8] は不適切な根拠に基づく推定は、適切な箇所を根拠とした推定よりも誤差が大きいという仮説を立て、これを統計的に検証した。その結果、上述の仮説を肯定する結果が得られた。

この結果は、平均誤差が低いモデルであろうとも、新規に入力した画像の根拠が不適切であれば、誤差が高いという可能性を示唆するものである。そのため、実応用場面で CNN を活用するならば、推定根拠の適切さを確認することが必要である。

3. 不適切な特徴マップの除去

推定根拠が不適切であれば誤差が大きいという統計的な検証結果は、推定根拠を改善させるような仕組みを学習済みのCNNに導入すれば、回帰推定値ないしはクラス分類精度が向上する可能性を示唆するものである。CAMやRAMはウェイトパラメータにより重みづけられた特徴マップの重ね合わせにより得られるため、不適切な特徴マップの存在が、不適切な推定根拠を生み出す原因となる。そのため、このような特徴マップを発見・除去することができれば、

・ホワイトボックス型CNNの推定根拠の改善
・汎化誤差の低減

という2つの効果が期待できる。そのため本研究では、最適輸送問題により得られる Wasserstein 距離 [9] を活用し、特徴マップの異常性スコアを算出し、不適切な特徴マップを自動的に検出・除去するアルゴリズムを提案する。この詳細については、原著論文において公表する。

Dimension Reduction Algorithm for CNN Feature Vectors by Feature Map Anomaly Scores based on Wasserstein Distance

Yuto OMAE, Yohei KAKIMOTO, and Jun TOYOTANI

4. CNN 特徴量空間の次元削減との関係

特に Global Average Pooling レイヤ [3] により特徴抽出を行うCNNにおいては、特徴マップの枚数が、抽出される特徴量ベクトルの次元数と等価になるという性質がある。すなわちこれにより構成される特徴量空間の次元も同様である。そのため、不適切な特徴マップの削減は、特徴量ベクトル・特徴量空間の次元削減と等価となる。この性質を考慮すれば、本研究で提案する手法は、CNN特徴量ベクトルの低次元化アルゴリズムと見做すことも可能である。なお、ここでいう低次元化とは、圧縮的な手法（主成分分析、特異値分解、オートエンコーダ など）ではなく、複数の特徴量の中から適切なものを選び取る、選択的手法である。

従来の特徴量選択問題を解くアルゴリズムに関しては、多様な方法が提案されており [10]、CNN特徴量に対して適用される事例も多い [11]。そのため本手法は、特徴量選択アルゴリズムを新規に提案する側面としての価値もある。

5. 提案手法の有効性

提案手法を学習済みのCNNに対し適用した結果、推定根拠の信頼性と汎化性能が大きく向上した。また、ベーシックな次元削減手法であるL1正則化と対立させた結果、特に汎化性能の観点から同等の特徴量が採択されることを確認している。この詳細については、原著論文にて公表する。

参考文献

1. B. Zhou et al., Learning deep features for discriminative localization, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp.2921-2929, 2016.
2. Z. Wang et al., Diabetic retinopathy detection via deep convolutional networks for discriminative localization and visual explanation, Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, pp.514-521, 2018.
3. A. Al-Sabaawi et al. Amended convolutional neural network with global average pooling for image classification, International Conference on Intelligent Systems Design and Applications, 171-180, 2020.
4. R. R. Selvaraju et al., Grad-cam: Visual explanations from deep networks via gradient-based localization, Proceedings of the IEEE international conference on computer vision, pp.618-626, 2017.
5. G. Qu et al., Ensemble manifold regularized multi-modal graph convolutional network for cognitive ability prediction, IEEE Transactions on Biomedical Engineering, vol.68, no.12, pp.3564-3573, 2021.
6. M. Kara et al., Covid-19 diagnosis from chest CT scans: a weakly supervised CNN-LSTM approach, AI, vol.2, pp. 330-341, 2021.
7. Y. Saito et al., Quantitative estimation of pulmonary artery wedge pressure from chest radiographs by a regression convolutional neural network, Heart and Vessels, vol.37, pp.1387-1394, 2022.
8. Y. Omae et al., Reliability metrics of explainable CNN based on Wasserstein distance for cardiac evaluation, Research Square, 2022. doi: 10.21203/rs.3.rs-1965782/v1
9. S. Kolouri et al., Optimal Mass Transport: Signal processing and machine-learning applications, IEEE Signal Processing Magazine, vol.34, no.4, pp.43-59, 2017.
10. R. Yao et al., Feature selection based on random forest for partial discharges characteristic set, IEEE Access, vol. 8, pp.159151–159161, 2020.
11. T. Saba et al., Categorizing the students' activities for automated exam proctoring using proposed deep L2-graftnet CNN network and aso based feature selection approach, IEEE Access, vol.9, pp.47639–47656, 2021.