

不均衡回帰問題に対する重み付きバッチ正規化法の検証

Yeo Xian En[†]山形大学大学院理工学研究科[†]安田宗樹[‡]山形大学大学院理工学研究科[‡]

1 はじめに

深層学習の発展により、パターン認識など様々な人工知能的課題の性能が比較的向上した [1, 2]. 良質な学習のためには大量の良質な訓練データを必要とするのだが、不均衡データの存在が実用的な場面でしばしば問題となる。不均衡データを学習すると、データ数が相対的に多い入出力関係が支配的となってしまう、データ数の相対的に少ない入出力関係をうまく再現することができなくなる。不均衡データの問題を解決する方法の一つに cost sensitive (CS) 学習がある。CS 学習は、各訓練データに適切な重みを追加することで疑似的にデータを複製して不均衡を解消する方法である。

近年、ディープニューラルネットワーク (deep neural network (DNN)) に対する CS 学習にバッチ正規化 [4] を単純に組み合わせると学習がうまく進まないことがあることが指摘され、その問題を解消する方法として、重み付きバッチ正規化 (weighted batch normalization (WBN)) が提案された [5]. しかしながら、WBN の有効性は不均衡クラス分類問題に対する CS 学習でのみ検証されており、不均衡回帰問題に対する適切な使用法及び有効性に関しては調べられていない。本研究では、不均衡回帰問題に対する WBN の使用方法を提案すると共に、その有効性について検証する。

2 CS 学習

入力データ \mathbf{x} と対応する教師データ \mathbf{t} の組を一つのデータ点 (\mathbf{x}, \mathbf{t}) として、 N 個のデータ点で構成される訓練データ集合 $D := \{(\mathbf{x}_\mu, \mathbf{t}_\mu) \mid \mu = 1, 2, \dots, N\}$ が与えられたとする。この訓練データ集合に対して、DNN 学習は損失関数

$$L(\theta) := \frac{1}{N} \sum_{\mu=1}^N \ell(\mathbf{x}_\mu, \mathbf{t}_\mu, \theta) \quad (1)$$

の θ に関する最小化の問題として定式化される。ここで、 θ は学習パラメータであり、 $\ell(\mathbf{x}_\mu, \mathbf{t}_\mu, \theta)$ は μ 番目の訓練データに対する損失を表わしている。

CS 学習は次のような重み付き損失関数

$$L_w(\theta) := \frac{1}{Z} \sum_{\mu=1}^N w_\mu \ell(\mathbf{x}_\mu, \mathbf{t}_\mu, \theta) \quad (2)$$

の最小化問題として定式化される。 w_μ は μ 番目の訓練データに対する正の重み係数であり、 $Z := \sum_{\mu=1}^N w_\mu$ は規格化定数である。重み係数 w_μ は対応する損失 $\ell(\mathbf{x}_\mu, \mathbf{t}_\mu, \theta)$ の相対的な大きさを調整している。つまり、対応するデータの実質的なデータ数を調整していると思わせる (例えば、 $w_\mu = 10$ の場合、 μ 番目の訓練データが 10 個に複製されていると考えることができる)。重み係数の値を適切に設定すれば、データ不均衡の影響を緩和した学習が可能となる。重み係数の決定法は様々であるが、クラス分類問題に対しては inverse class frequency (ICF) がしばしば用いられる [3]. ICF は、重み係数 w_μ を該当するクラスの存在確率の逆数:

$$w_\mu = \left(\frac{n(\mu)}{N}\right)^{-1} \quad (3)$$

と設定する。ここで、 $n(\mu)$ は μ 番目の訓練データが属するクラスに属するデータ数である。ICF はクラス毎のデータ数のバランスを整え、疑似的に均衡データの学習を実現することができる。

3 重み付きバッチ正規化

バッチ正規化は、DNN の学習において過学習の問題や共変量シフトの問題などの問題を回避することが可能とする [4]. バッチ正規化は、中間層の出力信号をミニバッチ毎に標準化し、平均・分散を一定に保つようする。これにより、中間層の出力信号の分布が中間層を経るごとに大きく変化してしまうこと (共変量シフト) を防ぎ、結果、学習の質を向上させることができる。

しかしながら、バッチ正規化における標準化は CS 学習と相性が悪い。標準化の計算はバッチサイズに依存するのだが、CS 学習の重み係数 w_μ によってデータの実質的なサイズは変化してしまっている。通常のバッチ正規化は重み係数 w_μ によるデータサイズの変化を考慮しないので、結果としてデータサイズの解釈に不一致が生じてしまうのである。そこで、その不一致の問題を解消する方法として WBN が提案された [5]. WBN を使用することによって、CS 学習の重み係数 w_μ を考慮した標準化計算が可能となり、データサイズの解釈を CS 学習とバッチ正規化の両方において一致させることができる。

クラス分類問題に対して WBN の有効性は調べられており、実際、不均衡クラス分類問題において CS 学習の性能の向上に成功している [5].

4 不均衡回帰問題への応用

CS 学習と WBN の組み合わせ自体はどのような種類のデータにも適用することができるが、クラス分類問題以外のクラス概念がないような問題では重み係数 w_μ の設定法が問題と

Evaluation of weighted batch normalization for data-imbalanced regression problem

[†] Yeo Xian En; Graduate School of Science and Engineering, Yamagata University

[‡] Muneki Yasuda; Graduate School of Science and Engineering, Yamagata University

なる。クラス分類問題の場合は式 (3) で示した ICF が一つの強力な常套手段となるが、回帰問題においては ICF に対応するような常套手段が自明ではない。回帰問題においては、教師データは一般に実数ベクトルであり、クラス概念はない。

本節では、まず、ICF の考え方を拡張して、回帰問題に対する重み係数 w_μ の適切な設定法をカーネル密度推定 (kernel density estimation (KDE)) [6] を用いて提案する。そして、不均衡回帰問題に対する WBN を用いた CS 学習に提案法により算出された重み係数を適用する。

4.1 カーネル密度推定による重み係数決定

教師データが m 次元の実数ベクトルであるとする。 N 個の教師データの集合 $D_{\text{target}} := \{\mathbf{t}_\mu \in \mathbb{R}^m \mid \mu = 1, 2, \dots, N\}$ を用いて教師データ $\mathbf{t} \in \mathbb{R}^m$ に対する確率密度分布 $P(\mathbf{t})$ を推定する。KDE を用いると、目的の分布は

$$P(\mathbf{t}) = \frac{1}{Nh^m} \sum_{\mu=1}^N K\left(\frac{\mathbf{t} - \mathbf{t}_\mu}{h}\right) \quad (4)$$

となる。ここで、 $K(\mathbf{t})$ はカーネル関数であり、 h はバンド幅 (分布の平滑性) を表すハイパパラメータである。本研究では、カーネル関数としてガウス分布

$$K(\mathbf{t}) = \frac{1}{(2\pi)^{m/2}} \exp\left(-\frac{\mathbf{t}^t \mathbf{t}}{2}\right) \quad (5)$$

を用い、 h の値を尤度を基準とした最尤交差検証で決定する。

ICF の考え方を応用して、式 (4) により得られた分布から重み係数 w_μ を決定する。 μ 番目の訓練データに対する重みを、次のように分布 $P(\mathbf{t})$ の逆数 (inverse KDE (IKDE)) で設定する。

$$w_\mu = P(\mathbf{t}_\mu)^{-1} \quad (6)$$

$P(\mathbf{t}_\mu)$ は $\mathbf{t} = \mathbf{t}_\mu$ となる確率密度であり、全体に占める \mathbf{t}_μ の存在割合に類似した意味をもつ。したがって、IKDE は ICF と類似の概念によるものと解釈できる。実際、 \mathbf{t}_μ がクラス分類用の 1-of- K ベクトルの場合、 h が十分小さいとき、IKDE は ICF とほとんど一致する。このことから、IKDE が ICF の自然な拡張の一つであることが伺える。

4.2 数値実験

UCI で公開されているワイン品質に関するデータ集合 (wine quality dataset) [7] を用いて提案法を検証する。複数種の赤ワイン (1599 種) と白ワイン (3299 種) それぞれに対する品質を 0 ~ 10 で評価したものである。入力には酒石酸濃度、酢酸濃度、クエン酸濃度など、計 12 次元であり、出力は 1 次元である。このデータ集合を回帰問題として扱う。評価の分布を見ると、極端に評価が低いものや高いものは数が圧倒的に少ない。つまり、不均衡回帰問題の 1 例となっている。

このデータ集合に対して、5 層 DNN を用いて CS 学習をする。層のサイズは入力層から出力層にかけてそれぞれ 12-11-8-4-2-1 としており、中間層の活性化関数は ReLU、出力層の活性化関数は恒等関数を使用している。損失関数は平均 2 乗誤差とし、最適化法はミニバッチサイズ 128 の adam を利用した。白ワインのデータを訓練データとし、赤ワインのデータをテストデータとした。

表 1 90 エポック学習後の DNN でテストデータを評価したときの損失値。10 回の実験の平均値である。

| 手法 | 損失値 |
|--------------------|------|
| BN | 0.63 |
| CS(IKDE)+BN | 1.17 |
| CS(IKDE)+WBN (提案法) | 0.55 |

表 1 にテストデータ (赤ワイン) に対する損失値を示す。表は、それぞれ 3 つの学習法により得られた DNN でテストデータを評価したときの損失値を表わしている。表中の“BN”は CS 学習を使わないで通常のパッチ正規化のみを使用する方法であり、“CS(IKDE)+BN”は IKDE を用いた CS 学習と通常のパッチ正規化を組み合わせた方法、そして、“CS(IKDE)+WBN”が IKDE を用いた CS 学習と WBN を組み合わせた方法 (提案法) である。不均衡クラス分類問題のときと同様に、不均衡回帰問題においても CS 学習と WBN の組み合わせの有効性を示唆する結果となっている。これより、提案の IKDE は不均衡回帰データのバランス補正に成功し、クラス分類問題のときと同様に、WBN の効果を引き出せるものになっていると期待される。

5 まとめ

本研究では、KDE を用いて、不均衡回帰問題に対する CS 学習における重み係数の決定法 (IKDE 法) を提案した。数値実験の結果より、提案法は不均衡回帰データのバランス補正に役立ち、結果、WBN が不均衡回帰問題にも有効であることが確認できた。しかしながら、現状ではまだ限られたデータに対する結果しか検証していない。別の複数のデータに対して同様の検証を行い、提案法の信頼性を高めることが今後の課題である。

謝辞

本研究は科研費 (18K11459, 18H03303)、JST-CREST (JP-MJCE1312) 及び JST COI プログラム (JPMJCE1312) の助成を受けたものである。

文献

- [1] 麻生ら: 深層学習, 近代科学社, 2015.
- [2] I. Goodfellow, Y. Bengio and A. Courville: Deep Learning, MIT Press, 2016
- [3] C. Huang, Y. Li, C. C. Loy, and X. Tang: Learning Deep Representation for Imbalanced Classification, *In Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.5375-5384, 2016
- [4] S. Ioffe and C. Szegedy: Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift, *In Proc. of the 32nd International Conference on International Conference on Machine Learning (ICML)*, vol.37, pp.448-456, 2015
- [5] M. Yasuda, Y. X. En, and S. Ueno: Consistent Batch Normalization for Weighted Loss in Imbalanced-Data Environment, *Nonlinear Theory and its Applications (NOLTA), IEICE*, Vol.11, Issue 4, pp.454-465, 2020.
- [6] C. M. Bishop: Pattern Recognition and Machine Learning, Springer, 2006
- [7] D. Dua and C. Graff: UCI Machine Learning Repository, 2019