

制約付き最適化を用いた非負 PNU 半教師学習による過学習の抑制

○石川 周*
Shu Ishikawa

全 眞嬉*
Jinhee Chun

徳山 豪*
Takeshi Tokuyama

1 はじめに

近年の通信技術や情報サービスの発展と多様化により、膨大な量と多種にわたるデータが利用可能となった。それらを用いることで、様々な分野や目的に対して機械学習によるパターン認識の活用が期待されている。本研究では、機械学習での分類問題における、多層パーセプトロンや畳み込みニューラルネットワークなどをはじめとした複雑なモデルに対して過学習を抑えられ、学習データの生成にかかるコスト削減の実現を目指す半教師あり学習の方法について取り組む。

一般に半教師あり学習はラベルあり正例データ、ラベルあり負例データ、ラベル無しデータの3種の学習データを用いて経験損失と呼ばれる目的関数を最小化するようなモデルパラメータを求める。経験損失は学習データに対してどのくらい正しく分類できるかの度合いを表す。本研究及び先行研究はこの最適化式の設計についてである。

2 先行研究

2.1 PNU 学習

PNU 学習 [1] は半教師あり学習の手法の一つで、ラベルあり正例データとラベルあり負例データによる PN 学習、ラベルあり正例データとラベル無しデータによる PU 学習、ラベルあり負例データとラベル無しデータによる NU 学習を組み合わせることで構成される。PNU 学習は既存する他の半教師あり学習の方策と比べて、学習に大きく影響を与えるモデル過程を必要としないため、より安定的な学習が期待できる手法である。ここで、PU 学習と NU 学習についてこれらが複雑なモデルの学習で過学習をする傾向があると指摘された [2]。

2.2 非負 PU 学習

非負 PU 学習 [2] は、PU 学習について過学習を抑制するために変更が加えられた方法である。これは PU 学習と同様ラベルあり正例データとラベル無しデータによる学習である。PU 学習が過学習をする原因として、ラベルあり正例データについて詳しくなりすぎるよう

な最適化をしてしまうことであると指摘された。そこで非負 PU 学習はこの最適化の対象となる経験損失に max 関数を用いて、正例データについての最適化を緩くすることで達成される。

3 提案手法

本研究では PNU 学習へ PU 学習や NU 学習の代わりに非負 PU 学習の方策を組み込むことで、複雑なモデルの学習において過学習を抑制できる半教師あり学習の方法を提案する。提案手法を記す手順として、はじめに非負 PU 学習を提案した先行研究の論文 [2] の補正の提案をし、次にそれをもとに提案手法の最適化式の構築をする。

3.1 先行研究の補正の提案

先行研究の論文 [2] 中では非負 PU 学習についてこれの経験損失の設計と、これを実装して実験した際のアルゴリズムが記されている。ここで設計された経験損失は前述の通り max 関数を用いた式である。しかし実験に用いられたアルゴリズムは max 関数ではなく、正例データに関する項が小さくなりすぎないように制約条件が付加されたものであると考えられる。つまりこれら2種類の最適化は、① max 関数を用いた最適化、②制約条件を用いた最適化となるように異なるものである。

ここで2種類の最適化について、それぞれがどのような学習をするのか計算機を用いて比較実験を行った。この結果から、②の制約付き最適化の方が①の max 関数を用いた最適化よりも、評価データにおいて安定的に小さな損失値と大きな正答率をもたらすことが確認できた。したがって、以降では制約条件を付加する方針を用いて提案手法を構築する。

3.2 提案手法の最適化式

提案手法は先行研究の PNU 学習の方策をもとに、ラベルあり正例データとラベルあり負例データによる PN 学習、ラベルあり正例データとラベル無しデータによる学習に制約条件を付加した非負 PU 学習、ラベルあり負例データとラベル無しデータによる学習に制約条

*東北大学 大学院情報科学研究科

件を付加した非負 NU 学習の 3 種の学習を組み合わせることにより構築した。ここで、非負 NU 学習は前述の非負 PU 学習について正例と負例を反転させることで改めて定められる。

具体的な構造として、この手法は最適化式の構築にあり最適化の対象となる経験損失は PN 学習と PU 学習、NU 学習の経験損失がハイパーパラメータ η によって重み付きで足し合わされ、そこへ制約条件が加えられている。要素となっているそれぞれの経験損失の比重は η によって調節可能である。以後、この提案手法を非負 PNU 学習と呼ぶ。非負 PNU 学習の最適化式は

$$\begin{cases} \text{minimise } (1 - \eta) \hat{R}_{PN}(g) + \eta \hat{R}_{PU}(g) \\ \text{subject to } \hat{R}_{U^-}(g) - \pi_P \hat{R}_P^-(g) \geq -\pi_P \beta & \text{if } \eta \geq 0 \\ \text{minimise } (1 + \eta) \hat{R}_{PN}(g) - \eta \hat{R}_{NU}(g) \\ \text{subject to } \hat{R}_{U^+}(g) - \pi_N \hat{R}_N^+(g) \geq -\pi_N \beta & \text{otherwise} \end{cases} \quad (1)$$

である。ただし、これに記される各記号及び関数は [2] に記されるものと対応する。

4 実験

提案した非負 PNU 学習と先行研究の PNU 学習 [1]、非負 PU 学習 [2] がそれぞれどのような学習をするのか、計算機による比較実験を行って確かめた。記述できるスペースの都合上、実験に用いた複数種類あるデータセットのうち MNIST に対する実験について記す。

学習するモデルは多層パーセプトロンを用いた。このモデルは入力層 1 層、中間層 5 層、出力層 1 層で設計し、このモデルパラメータの総数は 505,500 である。モデルパラメータ数が非常に多いこのモデルは複雑であり、過学習しやすいと言える。また、本実験はミニバッチ勾配降下法をもとにした更新アルゴリズムを用いた。評価方法はホールドアウト検証を採用した。MNIST はもとよりデータ全てにラベルの付いたものであるが、半教師あり学習のために各学習ごとでラベルありデータから無作為にラベルを除くことでラベル無しデータを生成した。学習は 30 回繰り返し、各エポックごとに損失値の平均値と標準誤差を算出した。

訓練データにおける損失値を Fig. 1、評価データにおける損失値を Fig. 2 に示す。各線は値の平均値、薄い領域は標準誤差を示す。それぞれの図について、これらの値は学習の経過における分類器がどれだけ訓練データ、または評価データを正しく分類できているかを意味し、値が小さいほど正しく分類できていることを示す。PNU 学習の結果において、訓練データの損失値は他の方策と比べて大きく減少しているが評価デー

タでは増加している。これは PNU 学習が過学習したことで生じたと考えられる。これと比べて非負 PNU 学習は評価データの損失値は小さいまま留まり、過学習を抑制することができたと見られる。また非負 PU 学習と比較したときこれより小さい損失値を達成していることが確認できた。他のデータセットについても同様の結果が得られた。

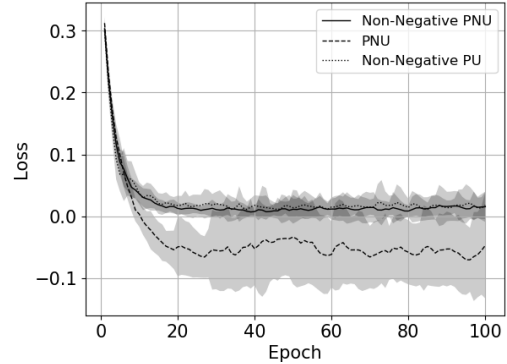


Fig. 1: MNIST 訓練データ損失値

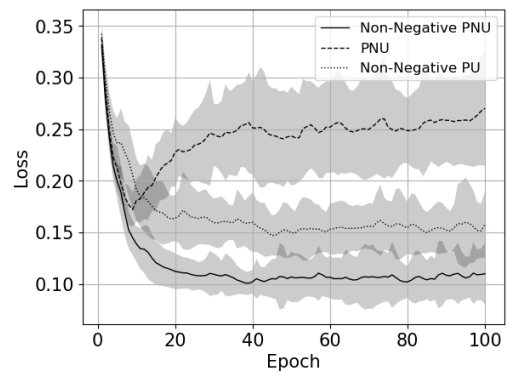


Fig. 2: MNIST 評価データ損失値

5 まとめ

PNU 学習へ非負 PU 学習の手法を組み込むことで、複雑なモデルに対して過学習を抑制できる新たな半教師あり学習の方策、非負 PNU 学習を提案した。また、この方策が上記 2 つの方策よりも良い分類器が生成できることが計算機実験より確認できた。

謝辞

本研究は、革新的研究開発推進プログラム (ImPACT) の一環として実施されました。

参考文献

- [1] Tomoya Sakai et al. Semi-supervised classification based on classification from positive and unlabeled data. *ICML 2017*, pages 2998–3006.
- [2] Kiryo Ryuichi et al. Positive-unlabeled learning with non-negative risk estimator. In *Advances in NIPS*, pages 1675–1685, 2017.