

多層ニューラルネットワークを弱学習器として用いた Boosting手法の検討

アドバイスや意見
などは非お願い
いたします！

○齊藤 優也, 松尾 信之介, 内田 誠一 (九州大学), 末廣 大貴 (九州大学/理研AIP)

概要

- ✓ Boostingの弱学習器として多層ニューラルネットワークを使用
- ✓ Boostingの理論に基づく弱学習器の条件を, 最低限度満たす多層ニューラルネットワークを使用
- ✓ 様々なパラメータやデータを用いた実験を行い, 理論・実用上の課題について調査

背景

- 画像分類問題での最悪クラス誤差抑制手法 [Saito+,2023] でDNNを弱学習器として用いていたが, その効果については未検証であった
- 実験結果では弱学習器でDNNを用いるBoosting [Hazan,2022]が通常の分類学習器 (CE) を超える性能を示す

Datasets	CIFAR-10	CIFAR-100	Tiny ImageNet	LT-CIFAR-10	EMNIST	TissueMNIST
Metrics	Avg. Error	Avg. Error	Avg. Error	Avg. Error	Avg. Error	Avg. Error
CE (ERM)	14.9	46.7	42.8	21.0	12.4	33.9
OCO Boosting	13.1	35.4	37.2	19.4	12.9	34.3

本研究の狙い

- 着眼点1 | DNNは単体でも十分に訓練データを学習できる分類器だが, Boostingとの相性を確認
- 着眼点2 | 弱学習器として用いたDNNの学習挙動を確認し, 汎化性能に影響を及ぼす要因を検証

弱学習器としてのDNNの設計

$$\sum_i w_{i,l}(x_i, y_i, f_e) \leq \frac{1}{2} - \gamma$$

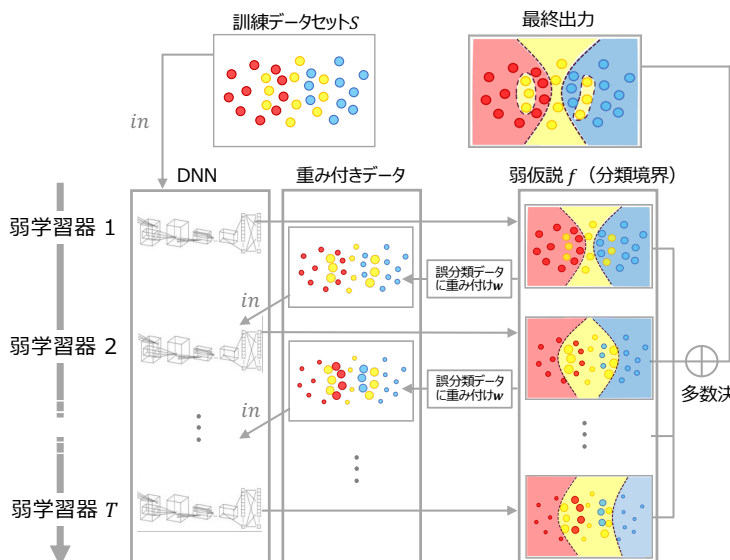
を満たす f_e のうち e が最小のもの

x : 訓練データ
 y : ラベル
 f_e : エポック数 e の学習器
 S : データ集合
 w : データ重み

- 理論背景 | 弱学習器の仮説集合の複雑性を抑えと, ラマツハ複雑度による一様バウンドにより汎化性能が向上 [Mohri+,2018]
- アイデア | 弱学習器の条件を “最低限度満たす” DNNの性能を弱めて, 仮説集合の複雑性を抑えられないか
 - 万能近似定理により, 必ず弱学習可能性を満たすDNNが存在

概略 | DNNを用いたBoostingアルゴリズム

- 弱仮説 f_1, \dots, f_T のアンサンブルによって強力な学習器を得る手法

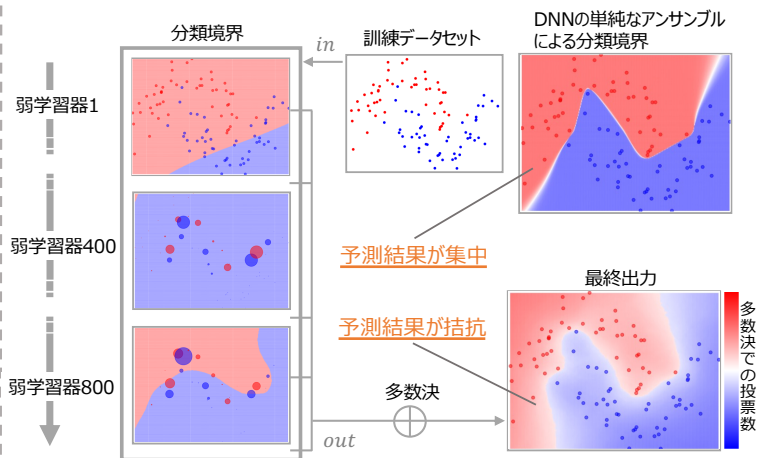


● Boostingによる理論保証

重み付け方法	$w_{t+1,i} = \frac{w_{t,i} e^{-\eta r_{t,i}}}{\sum_{j=1}^n w_{t,j} e^{-\eta r_{t,j}}}$ $r_{t,i}(f_t) = \begin{cases} 1, f_t(x_i) = y_i \\ 0, \text{otherwise} \end{cases}$ η : 定数
弱学習器の条件 (弱学習可能性)	どのようなデータ重み w に対しても以下を満たす $\sum_{i=1}^n w_{t,i} I(f_t(x_i) \neq y_i) \leq \frac{1}{2} - \gamma$
定理	$\text{error}_S(f) = 0 \quad f(x) = \text{sign}(\sum_{t=1}^T f_t(x))$ 訓練データを全て分類可能

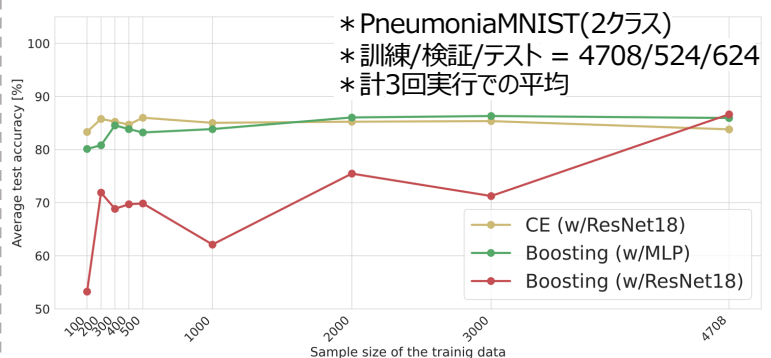
実験1 | 人工データでの分類境界の可視化

- 弱仮説 f_1, \dots, f_{922} と多数決後の分類境界を可視化 ($\gamma = 0.1$)



- 弱学習器の条件を満たす程度に性能を弱めたDNNを実現している
 - 各弱学習器は一部のデータに特化した学習を行っている
- Boostingでは境界線付近の多数決票が割れており汎化性能に期待

実験2 | MedMNISTによる汎化性能の検証



- 単体のDNN(ResNet18)に比べて僅かな性能向上を確認
 - データ数が十分あれば, Boosting手法が有効となることを示唆

γ	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45
Avg. Acc	0.821	0.801	0.834	0.857	0.859	0.851	0.852	0.834

- 弱学習器の “弱さ ($= \gamma$)” による汎化性能への影響を検証
 - ただ性能の良い弱学習器を用いるだけでは不十分であることを確認

議論

- データ全体を網羅する強力な表現学習よりも, 一部のデータに特化した表現学習をアンサンブルする方が良いのか
- ラウンド数が増えると計算量が大きくなるが, 解決策はあるのか