

動的サンプリングを使用した勾配ブースティング決定木の連合追加学習 Federated Continuous Learning of Gradient Boosting Decision Trees Using Dynamic Sampling

三浦 啓吾 * 王 立華 † 小澤 誠一 ‡
Miura Keigo Lihua Wang Seiichi Ozawa

キーワード 連合学習, データストリーム, 勾配ブースティング決定木, 追加学習, サンプリング

あらまし

勾配ブースティング決定木^[1] (GBDT) は, 様々な分野において高い性能を発揮しており近年注目されている機械学習モデルのひとつである. 既存の学習器をインクリメンタルに更新する追加学習は, 逐次的に新たなデータが与えられる環境 (データストリーム) においてのモデルの学習に効果的であるが, GBDT モデルの追加学習では, 学習が進むにつれて弱学習器の数が増えてしまいメモリを圧迫してしまったり, 古い弱学習器を棄却することで過去の入出力関係を失う^[2]といったリスクを伴う. また, データストリーム環境においてはデータのわずかな変化にモデルを対応させるという観点から, 単一組織だけでなく複数組織間のデータを利活用することも重要であるといえる. そういった手法では特に, 複数組織間でデータを分散させたまま利用し, 各組織の持つ情報のプライバシーを保護しつつ横断的なモデルの学習を行う連合学習が昨今注目されているが, 前述した GBDT モデルの追加学習における課題は, 単一組織での学習だけではなく連合学習においても同様である.

そこで本研究では, モデルの学習時に訓練データをサンプリングし, また訓練データに対するモデルの精度をもとに, イテレーション毎にサンプリング確率を更新することで, 効率的に新たなデータに適応した弱学習器を

生成するモデルを提案する. サンプリング手法に関しては, AdaBoost におけるデータの重みに基づいてサンプリングを行う手法と, 勾配情報をベースにしたサンプリングを行う 2 通りの手法を実装し, 単一組織での学習と連合学習に提案手法を導入して実験を行った.

実験結果として, 単一組織, 複数組織共に, 提案手法は従来の GBDT の追加学習と比較して少ない弱学習器で高い精度を出すことができ, 従来の手法では精度が落ちた場面でも安定した性能をキープしていた. また, 新たなデータのクラスタを追加した場合においても, 過去の入出力関係を維持しつつテストデータに対する精度を向上させることができた.

参考文献

- [1] T. Chen and C. Guestrin: “Xgboost: A scalable tree boosting system,” in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ser. KDD ’16. New York, NY, USA: ACM, 2016, pp. 785-794.
- [2] G. A. Carpenter and S. Grossberg: “The ART of adaptive pattern recognition by a self-organizing neural network”, IEEE Computer, 21(3), pp. 77-88 (1988)

* 神戸大学大学院工学研究科 〒 657-8501
神戸市灘区六甲台町 1-1. Graduate School of Engineering,
Kobe University. 1-1 Rokkodai-cho, Nada-ku, Kobe, 657-8501,
213t251t@stu.kobe-u.ac.jp

† 国立研究開発法人 情報通信研究機構 〒 184-8795
東京都小金井市 貫井北町 4-2-1. National Institute of
Information and Communications Technology.
4-2-1 Nukui-Kitamachi, Tokyo, 184-8795,
lh-wang@nict.go.jp

‡ 神戸大学数理・データサイエンスセンター 〒 657-8501
神戸市灘 区六甲台町 1-1. Center for Mathematical and
Data Sciences, Kobe University. 1-1 Rokkodai-cho,
Nada-ku, Kobe, 657-8501,
ozawasei@kobe-u.ac.jp