

# 野生動物の生態観察に向けた非同期分散型勾配ブースティング

山下 優衣<sup>†</sup> 田谷 昭仁<sup>‡</sup> 下徳 大祐<sup>♭</sup> 小林 博樹<sup>♭</sup> 戸辺 義人<sup>‡</sup>

青山学院大学大学院理工学研究科理工学専攻<sup>†</sup>

青山学院大学工学部情報テクノロジー学科<sup>‡</sup>

東京大学情報基盤センター<sup>♭</sup>

## 1. はじめに

野生動物を追跡し行動や活動範囲を把握することは、生態系の理解に繋がるため重要である。生態観察には、センサ端末から取得したデータを機械学習で訓練し、構築された予測モデルを使用することができるが、野生動物を監視するような環境は、特殊な制約が課されることがある。例えば、インターネットへの接続ができず、LPWA (Low Power Wide Area) のようなマルチホップネットワークが必要になること、デバイスへの給電機会が限られ、消費電力を抑えなければならないことなどである。

本稿では、このような環境下においても生態観察を可能にするために、NN (Neural Network) に比べて計算量が少ない機械学習手法の一つである GBDT (Gradient Boosting Decision Tree) [1] を用いた分散アルゴリズムを提案する。提案アルゴリズムは、サーバを用いないマルチホップネットワーク下で、データではなく GBDT の学習モデルを共有することで一般化されたモデルの訓練を行う。また、各端末間の学習速度差を考慮しモデルの送受信を非同期処理で実現する。

## 2. 関連研究

ローカルデータではなく機械学習モデルを共有する分散学習アルゴリズムに FL (Federated Learning) [2] がある。FL ではクライアントごとに NN モデルを学習し、サーバを利用してモデルの集約とグローバルモデルの構築をしている。集約と更新を繰り返すことで、機械学習の精度を向上することができる。

本稿では、FL アルゴリズムを元に各端末での機械学習手法に GBDT を適用する。GBDT は、決定木を弱学習器とし、複数のモデルを組み合わせ

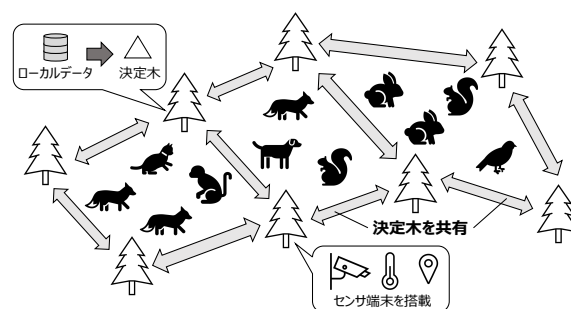


図1. システムモデル

て学習をするアンサンブル手法である。FL 条件下において決定木を用いた分散学習を行なっている Pivot[3]では、企業ごとに所持している異なるユーザの特徴量からモデルを構築し、これらのモデルを共有することで分散学習がなされているが、サーバの役割を担う正解ラベルを持つ端末が必要となる。本研究では、端末間で特徴量を分割するのではなく、各端末で同じ特徴量から異なる学習モデルを構築し、サーバレス状況下でモデルを共有することを考える。

## 3. システムモデル

提案アルゴリズムは、インターネット接続が難しいような環境下においても使用可能とするため、LPWA を用いて端末同士が直接通信を行うマルチホップネットワーク接続を想定する。野生動物が生息する周囲の環境にセンサ端末を設置し、カメラなどから得られたデータを用いて野生動物の行動モデルを機械学習により学習することができる。また、構築した学習モデルを、取得したデータの代わりに端末間で共有することで、送受信するデータ量を削減する。さらに、センサ端末が設置される場所ごとに自然環境が異なり、観察できる動物や動物の行動にも差がある non-IID (Independent and Identically Distributed) データであると考えられる。このシステムモデルを図1に示す。

Asynchronous Distributed Gradient Boosting Decision Tree for Biologging Wild Animals

<sup>†‡</sup> Yui Yamashita, Akihito Taya, Yoshito Tobe / Aoyama Gakuin University

<sup>♭</sup> Daisuke Shimotoku, Hiroki Kobayashi / Tokyo University

#### 4. GBDTによるサーバレス FL

学習速度は端末ごとに異なり、また分散型ネットワークでは同期をとることが困難であるため、本稿では非同期分散型の分散学習手法を提案する。

はじめに各端末で取得したデータから弱学習器である決定木を構築し、構築した木を隣接端末へ送信する。端末*i*が所持するローカルデータセットを $\mathcal{D}_i$ 、端末*i*がブースティング回数*j*回目に学習した決定木を $h_{i,j}$ と定義すると、1回目にローカルデータから構築した木は、

$$h_{i,1} \leftarrow \text{train}(\mathcal{D}_i), \quad (1)$$

と表せる。訓練された木は分散学習を行うために隣接端末へ送信される。この時、端末*i*は自身が学習した木とともに、まだ送信していない他端末から受信した木も送信する。

端末*i*がブースティング回数*j*回目に更新した、複数の決定木から成る GBDT モデルを $f_{i,j}$ とする。 $h_{i,1}$ の学習後、端末*i*はデータセット $\mathcal{D}_i$ を訓練して得られた決定木 $h_{i,1}$ と隣接端末が学習した決定木を合算し GBDT モデル $f_{i,1}$ を更新する。 $f_{i,1}$ は以下の通りとなる。

$$f_{i,1} \leftarrow h_{i,1} + \mathcal{T}_{i,1}. \quad (2)$$

ここで、 $\mathcal{T}_{i,j}$ は端末*i*が $f_{i,j}$ の更新時に既に他端末から受信している決定木の集合である。また、2回目以降は残差を用いて次の決定木を学習するため学習率を $\eta$ とすると、 $j \geq 2$ 回目における決定木 $h_{i,j}$ と GBDT モデル $f_{i,j}$ はそれぞれ以下のようになる。

$$h_{i,j} \leftarrow \text{train}(\eta(\mathcal{D}_i - f_{i,j-1})) \quad (j \geq 2), \quad (3)$$

$$f_{i,j} \leftarrow \sum_{k=1}^j h_{i,k} + \sum_{k=1}^j \mathcal{T}_{i,k}. \quad (4)$$

以上のように決定木の訓練と GBDT のモデルの更新を繰り返すことを繰り返す。 $M$ 回後を最終モデルとし $\mathcal{F}_i$ と表すと、GBDT は全ての決定木の合計が最終的な予測となるため、集約式は次式となる。

$$\mathcal{F}_i \leftarrow \{f_{i,1}, f_{i,2}, \dots, f_{i,M}\}. \quad (5)$$

#### 5. 評価

提案アルゴリズムを、画像のデータセット MNIST を用いて評価した。図 2 に評価に用いたトポロジーと各端末の所持ラベルを示す。図 2 に示した端末間の繋がりや Raspberry Pi を用いて形成した。各端末の所持するラベルは、近くにある端末同士が似たデータを持ち、それらに偏りがあることを想定し、グループごとに同じ 2

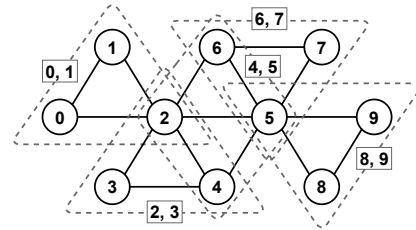


図 2. 評価トポロジーと所持ラベル

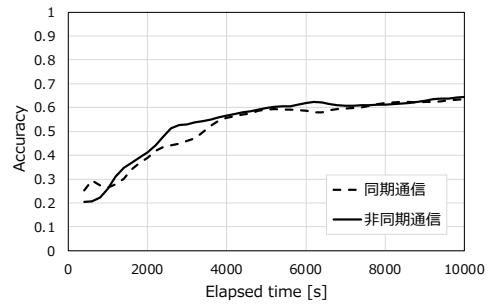


図 3. 評価結果

種類のラベルデータを持つように設定した。複数のグループに属する端末はグループ数に応じて 4 または 6 種類のデータを所持するものとした。また、端末間のリソースや通信タイミングのばらつきを考慮するため、端末間で計算速度が異なるよう設定し、同期処理と非同期処理の学習速度を比較した。

正解率の学習経過を図 3 に示す。グラフより 40%, 50%, 60%の精度に達する時間を比較したとき、同期処理による通信に比べて非同期処理の方が先に到達していることがわかる。この結果より、非同期処理により計算速度の速い端末が他端末の計算を待たずに学習することで、より短時間で高い精度のモデルを構築することができると思われる。

#### 謝辞

本研究は JSPS 科研費 JP18H04089 の助成を受けたものである。

#### 参考文献

- [1] Friedman H. J.: Greedy Function Approximation: A Gradient Boosting Machine, The Annals of Statistics, Vol.29, No.5, pp. 1189--1232 (2001).
- [2] Konečný, J., et al.: Federated Learning: Strategies for Improving Communication Efficiency, NIPS Workshop on Private Multi-Party Machine Learning (2016).
- [3] Wu, Y., et al.: Privacy preserving vertical federated learning for tree-based models, Proc. of the VLDB Endowment, Vol.13, No. 12 (2020).