

無線センサネットワークにおける深層学習の分散実装方式

Distributed Implementation Method of Deep Learning on Wireless Sensor Networks

梅田 果凜[†]
Karin UMEDA三好 匠[†]
Takumi MIYOSHI山崎 託[†]
Taku YAMAZAKI堀江 輝樹^{††}
Teruki HORIE[†] 芝浦工業大学システム理工学部電子情報システム学科
College of Systems Engineering and Science, Shibaura Institute of Technology^{††} 早稲田大学理工学術院総合研究所
Waseda Research Institute for Science and Engineering, Waseda University

1. まえがき

近年、無線センサネットワークにおいて、エッジコンピューティングや端末間協調を用いた深層学習の分散処理技術が注目されている。しかし、既存方式ではデータの処理速度は向上するが、通信量や消費電力が増大するという課題がある。本稿では、通信量の削減とクラウドの負荷軽減を目的とし、各センサに深層学習の中間層を分割する分散実装方式について検討する。

2. 従来手法

文献 [1] では、深層学習の計算をエッジサーバで最初に行った後、モバイル端末で残りの計算を行う手法を提案している。処理順序のほか、通信量やサーバ負荷の点で課題がある。文献 [2] は、畳み込み層の計算処理をモバイル端末の GPU にオフロードすることで、ビデオフレームの連続処理を行う手法である。モバイル端末のメモリ、CPU と GPU リソースだけを使用しており、連続するビデオフレームの中で前のフレームから変化したところを畳み込み層の計算を行う。認識率、処理遅延、消費電力を評価している。文献 [3] では、複数のモバイル端末でデータを並列処理する分散処理手法が提案されている。本手法では、モバイル端末の計算能力に応じて入力データを分割し、各端末で畳み込み層の計算を並列処理している。その計算結果を集約し、集約した結果をモバイル端末に再度送信し計算を繰り返すことで、畳み込み層の実行効率を向上させている。並列処理によってデータの処理速度は向上するが、各畳み込み層の計算のたびにデータを更新する必要があるため、モバイル端末間の通信量が大きくなる。

3. 提案手法

本稿では、無線センサネットワークにおける深層学習の分散実装方式を提案する。本手法では深層学習の中間層を複数のブロックに分割し、これらを複数のセンサに実装してデータ演算を分散処理することで、通信量の削減とクラウドの負荷軽減を実現する。

図 1 に提案手法の動作手順を示す。まず、データを取得するノード n_1 には、入力層、及び中間層の最初の複数層が割り当てられる。中継ノード n_2 、 n_3 には、残りの中間層が分割して割り当てられ、最終ノード n_4 には、全結合層と出力層が割り当てられる。各ノードは、演算結果を次ノードに無線通信を用いて送信し、 n_4 は出力結果をサーバに送信する。マルチホップでデータを伝送している間に中間層の計算が終了した場合、それ以降のセンサでは演算処理を行わず、シンクノードにマルチホップで転送する。一方、ホップ数が不足しシンクノード到達時にデータ処理が終了していない場合には、残りの中間層の演算処理をサー

バで実行する。

4. 実装実験と結果

本稿では、実装例として TensorFlow 上で物体を検出するプログラム VGG16 と Inception-v3 を使用する。センサノードとして、Raspberry Pi を 4 台を使用する。提案手法の性能評価のため、総送信データ量および計算処理時間に着目し、 n_1 で深層学習の全ての演算処理を行う場合と提案手法を比較する。

総送信データ量に着目すると、 n_1 で全ての計算処理を行う場合、出力結果のみを送信するためデータ量は小さくなる。提案手法は、演算処理の途中の結果を伝送するためデータ量は大きくなる。次に、計算処理時間に着目する。 n_1 で全ての演算処理を行う場合、計算処理能力を考慮すると遅くなる。提案手法では、分割処理によりそれぞれのノードの処理する演算処理が小さくなるため、全体の計算処理時間は小さくすると予測できる。消費電力の観点からは、1 台のセンサで全ての演算処理を行うのに比べて、提案手法の方が分散化を実現しているため、電力消費の集中を抑えられると考えられる。

5. むすび

本稿では、無線センサネットワークにおける深層学習の分散処理手法として中間層の分散処理を提案し、実装した。今後は、実装実験による評価や、センサを増やした場合の通信環境やルーティングを考慮した手法について検討する予定である。

文献

- [1] E. Li, Z. Zhou, and X. Chen, "Edge intelligence: On-demand deep learning model co-inference with device-edge synergy," MECCOM'18, pp. 31–36, Aug. 2018.
- [2] L.N. Huynh, Y. Lee, and R.K. Balan, "DeepMon: Mobile GPU-based deep learning framework for continuous vision applications," ACM MobiSys 2017, pp. 82–95, June 2017.
- [3] J. Mao, Z. Yang, W. Wen, et al, "MeDNN: A distributed mobile system with enhanced partition and deployment for large-scale DNNs," IEEE/ACM IC-CAD 2017, pp. 751–756, Nov. 2017.

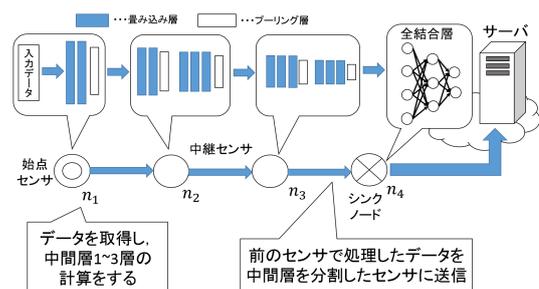


図 1 提案手法の概要