

## TSUBAME 共同利用 平成30年度 学術利用 成果報告書

## 利用課題名 近似ベイズ推定を用いた分散並列深層学習

英文: Parallel Distributed Deep Learning Using Approximate Bayesian Inference

## 利用課題責任者

Mohammad Emtiyaz Khan

## 所属

理化学研究所 革新知能統合研究センター

## URL

[http://www.riken.jp/en/research/labs/aip/generic\\_tech/approx\\_bayes\\_infer/](http://www.riken.jp/en/research/labs/aip/generic_tech/approx_bayes_infer/)

## 邦文抄録(300 字程度)

深層学習では、ノイズが多く不完全で欠損のあるデータを用いて汎化性能の高いモデルを学習することが最大の課題である。近似ベイズ推定は従来の点推定による深層学習とは異なり、不確かさを含めて学習することでデータが少なかったり、ノイズが多かったり、欠損があったりする場合にも高い汎化性能を保ちつつ正しく学習することができる。ただし、近似ベイズ推定による深層学習は計算コストが高く、膨大な計算資源を要する。そこで、本研究では TSUBAME3.0 を用いることで、従来では不可能であった実用規模での近似ベイズ推定による学習を行うことを目的とする。

## 英文抄録(100 words 程度)

One of the main challenges in deep learning is to improve the generalization capability of models when trained on noisy and incomplete training data. Bayesian inference can learn the uncertainty in the data and has good generalizability even for noisy and incomplete data. However, Bayesian inference is computationally expensive and a fast approximation method is needed to train in a realistic timeframe. Using TSUBAME3.0 we aim to train on real-world datasets using approximate Bayesian inference.

*Keywords:* Deep learning, Bayesian inference, Data-parallel, Sample-parallel

## 背景と目的

近年の深層学習手法と計算機技術の発展により画像認識、音声認識、自然言語処理の分野で人間の能力を超える性能を発揮できるようになってきている。画像認識技術は監視カメラ、顔認証、自動運転の分野で応用され、音声認識は音声入力、自然言語処理は自動翻訳や自動文章生成などで応用されており、いずれも、今後の AI 社会の基盤となる技術である。

深層学習の問題点は膨大なデータを必要とすることである。単にデータを集めただけでは学習に使用せず、それに正解のラベルを人間が付与する必要があるため、学習用のデータを大量に揃えることは容易ではない。そのため、深層学習分野ではラベルのないデータから特徴量を抽出する教師なし学習や、少ないラベル付きのデータと大量のラベルなしデータから学習する半教師あり学習などが盛んに研究されている。また、これとは異なるアプローチでベイズの定理にもとづく不確かさを含めた学習を行う事でノイズの多い不完全なデータからも学

習できるようにする手法がある。従来の学習ではニューラルネットの重みを点推定する手法が用いられてきたが、ベイズ推定では重みの期待値と分散の両方を学習することで不確かさも含めて学習することができる。これにより、データが少ないときは間違った答えを出すのではなく、予想に自信がないということを明示的に示すことができるようになる。ただし、ベイズ推定は点推定に比べて膨大な計算時間を要するため、非常に小さなデータとモデルに対してしか用いられてこなかった。

本プロジェクトでは、ベイズ推定の新たな近似解法を提案し、その分散メモリ並列化を行うことで TSUBAME3.0 の計算性能を活かし、既存研究では不可能であった、実用レベルのデータセットとモデルでベイズ推定を行うことができた。

## 概要

本プロジェクトでは、ベイズ推定で必要になるモンテカルロサンプルとミニバッチ勾配降下法で必要になるデー

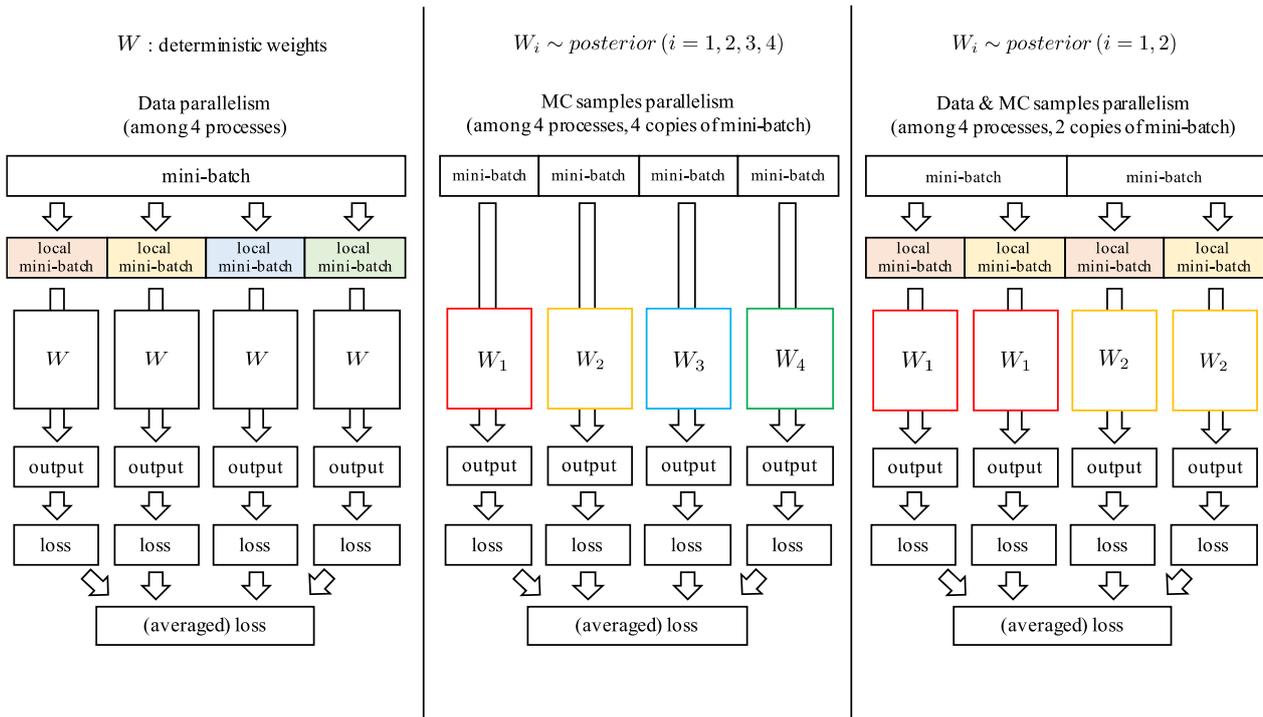


図1 データ並列とサンプル並列の組み合わせによるベイズ推定の分散並列化の様子

タサンプルの両方を同時に分散並列化の枠組みに組み込む新たな手法を提案した。図1の左側にデータ並列のみの場合、中央に MC サンプル並列のみの場合、右側にデータ・MC サンプル同時並列の場合のデータフローの様子を示す。左側のデータ並列のみの場合は重み  $W$  は全プロセスで共通しており、各プロセスに分散されたローカルミニバッチのデータが異なる。一方、中央の MC サンプル並列のみの場合は、ミニバッチのデータは全プロセスで共通しており、MC サンプルした重み  $W$  が少しずつ異なる。右側の図ではこの2つの並列化を組み合わせた状態になっている。

### 結果および考察

図2に本手法(VOGN)と従来法(Adam)と普通のモンテカルロ法(MC)の陽性と偽陽性の指数を示す。この曲線が左上の隅に寄っているほど精度が高い。従来法は普通のモンテカルロ法より精度が高く、提案手法はそれらよりはるかに高い精度が得られていることが分かる。



図2 本手法(VOGN)と従来法(Adam)と普通のモンテカルロ法(MC)の陽性と偽陽性の指数

### まとめ、今後の課題

図2の実験で用いたデータセット CIFAR10 とモデル LeNet5 はまだまだ小さいため、TSUBAME3.0 の計算能力が十分に発揮できている結果とは言えない。ただし、手法の有効性は確かめられたため今後はより大規模なデータセットとモデルを用いて学習を行っていく予定である。