

令和元年6月26日現在

機関番号：17104

研究種目：挑戦的研究（萌芽）

研究期間：2017～2018

課題番号：17K20010

研究課題名（和文）回路化の際に生じる制約を積極活用するニューラルネットワークの高性能化

研究課題名（英文）Performance improvement on neural networks that actively employ the constraints of hardware circuits

研究代表者

田向 権（Tamukoh, Hakaru）

九州工業大学・大学院生命体工学研究科・准教授

研究者番号：90432955

交付決定額（研究期間全体）：（直接経費） 5,000,000円

研究成果の概要（和文）：回路化の際に生じる演算誤差を乱数として活用することでのRBMの学習や、乱数の代わりとなる単純なルールでのDropout実装により、MLP、CNN、LSTMといった様々なニューラルネットワークの学習が可能であることを見いだした。これにより、乱数生成器のために必要であった回路規模の削減などが可能となり、ニューラルネットワークをハードウェア実装した際の高性能化が期待できることを示した。

研究成果の学術的意義や社会的意義

深層学習が大きな注目を集めるにつれて、多数の回路研究者が深層学習のアクセラレータ開発へと集結している。しかし、ニューラルネットワークの理論やアルゴリズムにまで踏み込んだ回路実装に関する研究領域は未開拓で、特に消費電力や排熱が重要となる組込み化はこれからの領域である。本研究成果により、乱数生成に係わる一部分ではあるが、ニューラルネットワークを理論面から軽量化することに成功し、回路化への道筋を付けることができた。本成果により、我が国が得意とする組込みシステムや自動車・ロボット分野への深層学習応用について大きな貢献が期待できる。

研究成果の概要（英文）：We employ rounding errors that occur in hardware circuits as random numbers for training restricted-Boltzmann Machine (RBM) neural networks. We also propose a modified dropout algorithm that employs a simple rule for training multi-layer perceptron (MLP), convolutional neural networks (CNN) and long-short-term memory (LSTM) neural networks. By using these proposed methods, we can reduce hardware resources for random number generators and improve the performance of neural networks implemented by hardware circuits.

研究分野：脳型計算機システム

キーワード：ニューラルネットワーク 演算誤差 デジタルハードウェア FPGA

1. 研究開始当初の背景

生物の脳の演算方式を模擬したニューラルネットワークを多層化した深層学習 (Deep Learning) が大きな着目を浴びている。現状、クラウド側での大型計算機による研究開発や応用が先行しているが、ロボットや自動車といったエッジ側での組込み分野への応用も大きく期待されている。組込み化には、ロボットなどの自らの身体が持つセンサから得られるデータを超低消費電力かつ実時間処理できる能力が求められるが、クラウド側で深層学習の高速化に用いられる Graphic Processor Unit (GPU) は消費電力と排熱の面で不向きなので、ニューラルネットワークの回路化が必須となる。

研究代表者は博士課程時よりニューラルネットワークの回路化に従事してきた。この中で、図 1 に示す、粗い比較で勝者ニューロンを決定する“粗い勝者決定”という競合学習則を提案、比較演算のビット精度を落としつつ演算速度とベクトル量子化性能の両方を同時に向上できることを示した (H. Tamukoh, IEICE Trans, 2004)。この成果から“ソフトコンピューティング”では粗い演算が許容され、場合によっては性能向上につながることを既に見出していた。

また研究代表者は、動画圧縮伸長コーデックや TCP/IP プロトコルの論理回路設計に携わってきた (田向, 電子情報通信学会論文誌, 2010)。これらは 1 ビットのエラーですら許容しない“ハードコンピューティング”であり、誤り訂正や同期方式を用いてノイズや通信遅延に対応することが必須である。このような一般的なアルゴリズムの回路化では、図 2 に示すように、理論研究者と回路研究者はアルゴリズムを境に分断されており、互いに影響を与えることは無く、良くも悪くも分業が出来ている現状である。

一方で、生物の脳は、電氣的・化学的に生じるノイズを積極活用した情報処理を行う。脳という構造があるがゆえの情報処理様式であり、これに倣うと、ニューラルネットワークの回路化では理論とアルゴリズム、回路を同時に考える必要性が示唆される。しかし、分野間の大きな壁から、理論から回路までを含む脳型計算機システムの研究は空白地帯になっている。そこで、ソフトコンピューティング分野の回路研究者という、応募者の特異性とこれまでの研究成果を合わせることで、理論研究者やハードコンピューティングの回路研究者のみでは思い至らない脳型計算機システムの実現が可能になると考え本提案に至った。

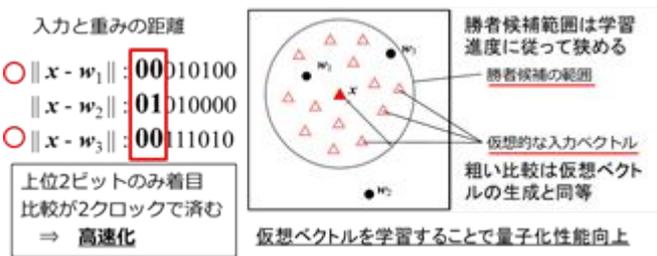


図 1：粗い勝者決定による高速化と量子化性能向上。

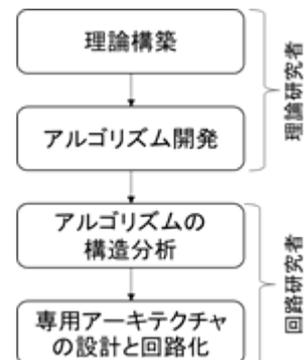


図 2：回路化の流れ。

2. 研究の目的

本研究では、アルゴリズムを回路化する際に生じる演算誤差や通信遅延など、一般的なアルゴリズムでは性能に悪影響を与える制約を積極的に活用し、ニューラルネットワークの高性能化や小面積化を目指す。これにより、現状困難な深層学習の組込み機器への実装に道筋を付けると共に、これまで長年にわたって体系化されてきたハードコンピューティングの回路設計手法について、ソフトコンピューティングの領域ではその方法論を大きく転換することを狙う。

3. 研究の方法

回路化の際に生じる誤差を積極的に活用する例として、図 3 に Restricted Boltzmann Machines (RBM) における隠れ層ニューロンの発火確率を求める計算過程を示す。結合荷重と発火確率を、それぞれ固定小数点 16 ビット (整数部 8 ビット、小数部 8 ビット) で表現した場合、まず乗算でビット範囲が倍になり、更に加算による桁上げで整数部のビット範囲が増大する (210 ニューロンの加算の場合 10 ビット増える)。この値を乱数生成器で生成した乱数と比較し、大きければ発火、小さければ非発火とする。この際、次の計算ステップで使うために元の 16 ビットへ戻さねばならないが、ここで切り捨てられるビットが誤差となる。

アルゴリズムを回路化すると、回路内の至る所にこのような誤差が現れる。そこで研究代表者は、回路化の際に現れる誤差を乱数の代わりに使えないかと考えた。もし、このような計算途中に現れる誤差を積極活用出来れば、例えば乱数生成器の削減や、それにより余った面積でコア数を増やし並列度を上げることで性能向上につながる。

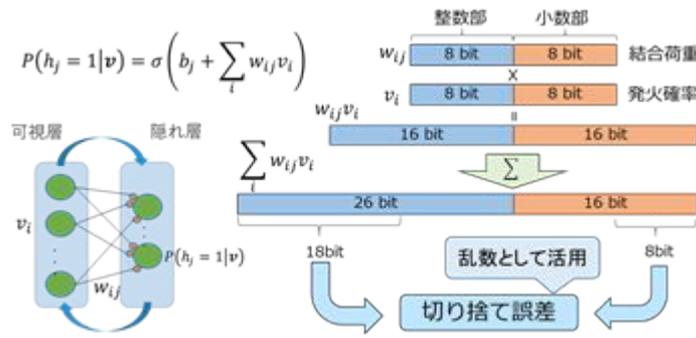


図 3 : RBM の計算過程 . 誤差を乱数として活用可能 .

4 . 研究成果

演算誤差を積極活用する例として , RBM の計算過程に現れる小数部の切り捨て誤差を乱数として活用し , MNIST データセットによる大規模実験により実際に RBM に適用したところ , 正常に学習が進むことを確認できた (図 4 , 図 5) . これにより , 発火・非発火を判定する乱数を生成する乱数生成器をニューロン回路から削除できることを示した . (主な発表論文 [3])

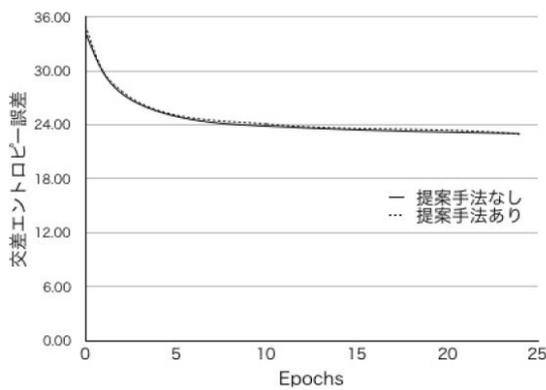


図 4 : RBM の学習結果 . 両手法とも同等の学習結果 . 図 5 : 提案手法により得られた出力画像 .

乱数を用いる学習手法として代表的な Dropout について RBM と同様のアナロジーでの実現が可能かどうか検討を行った . その結果 , 事前定義した Dropout Mask をシフトするのみの , より簡便な仕組みでの Dropout 実現が可能であることを示し , Multi-Layer Perceptron (MLP) や Convolutional Neural Networks , Long-Short-Term Memory (LSTM) での性能を確認 , 乱数生成器を用いる方法に比べて大幅に回路規模を削減できることを示した (図 6 , 7) . (主な発表論文 [5][7])

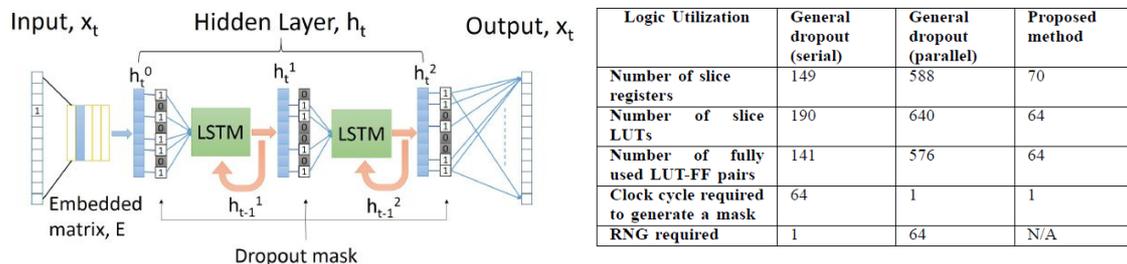


図 6 : LSTM への提案 Dropout の実装例 . 図 7 : 64 ビット Dropout Mask 実装時の回路規模比較 .

5 . 主な発表論文等

〔雑誌論文〕(計2件)

- [1] Akihiro Suzuki, Takashi Morie and Hakaru Tamukoh, “A shared synapse architecture for efficient FPGA implementation of autoencoders,” PLoS ONE. 2018; 13(3):e0194049. <https://doi.org/10.1371/journal.pone.0194049>
- [2] Dinda Pramanta, Takashi Morie, Hakaru Tamukoh, “Synchronization of Pulse-Coupled Phase Oscillators over Multi-FPGA Communication Links,” Journal of Robotics, Networking and Artificial Life, Vol. 4, No. 1, pp. 91–96, June 2017.

〔学会発表〕(計5件)

- [3] 堀 三晟, 田向 権, “ニューラルネットワークのハードウェア実装に向けた乱数生成手法の提案と検証,” 電子情報通信学会スマートインフォメディアシステム研究会(SIS), SIS2019-1, 2019年6月13-14日(13), 長崎, 福江文化会館.
- [4] Ninnart Fuengfusin, Hakaru Tamukoh, “Mixed Precision Weight Networks: Training Neural Networks with Varied Precision Weights,” 25th International Conference on Neural Information Processing (ICONIP2018), Siem Reap, Cambodia, December 13-16(15), 2018.
- [5] Yoeng Jye Yeoh, Hakaru Tamukoh, “Alternative Dropout for Hardware Implementation in Recurrent Neural Networks,” 2018 International Workshop on Smart Info-Media Systems in Asia (SISA2018), RS-13, Kanagawa, Japan, December 13-14(14), 2018. SISA Best Student Paper Award
- [6] Akihiro Suzuki, Hakaru Tamukoh, “Reverse Reconstruction of Anomaly Input Using Autoencoders,” 2018 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS2018), FM1B-3, Okinawa, Japan, November 27-30(30), 2018.
- [7] Yoeng Jye Yeoh, Takashi Morie and Hakaru Tamukoh, “A Hardware Oriented Dropout Algorithm for Efficient FPGA Implementation,” Proceedings of 24th International Conference on Neural Information Processing (ICONIP2017), Lecture Notes in Computer Science, Vol. 10638, pp.821-829, Guangzhou, Nov. 14-18(17), 2017.

〔図書〕(計0件)

〔産業財産権〕

出願状況(計0件)

〔その他〕

ホームページ等

九州工業大学 田向研究室 : <https://www.brain.kyutech.ac.jp/~tamukoh/>

6 . 研究組織

(1)研究分担者

無し

(2)研究協力者

研究協力者氏名 : 鈴木 章央

ローマ字氏名 : SUZUKI, Akihiro

研究協力者氏名 : 堀 三晟

ローマ字氏名 : HORI, Sansei

研究協力者氏名：プラマンタ ディンダ
ローマ字氏名：PRAMANTA, Dinda

研究協力者氏名：ヨンジェ ヨー
ローマ字氏名：YOENG JYE, Yeoh

研究協力者氏名：フエンフシン ニンナート
ローマ字氏名：FUENGFUSIN, Ninnart