



HOKKAIDO UNIVERSITY

Title	エッジ環境における知的情報処理に向けた低消費電力プロセッサアーキテクチャに関する研究 [論文内容及び審査の要旨]
Author(s)	肥田, 格
Degree Grantor	北海道大学
Degree Name	博士(情報科学)
Dissertation Number	甲第13516号
Issue Date	2019-03-25
Doc URL	https://hdl.handle.net/2115/74196
Rights(URL)	https://creativecommons.org/licenses/by-nc-sa/4.0/
Type	doctoral thesis
File Information	Itaru_Hida_abstract.pdf, 論文内容の要旨



学 位 論 文 内 容 の 要 旨

博士の専攻分野の名称 博士（情報科学） 氏名 肥田 格

学 位 論 文 題 名

エッジ環境における知的情報処理に向けた低消費電力プロセッサアーキテクチャに関する研究
(Low-power processor architectures for intelligent information processing in edge environment)

本研究は、低電力で人工知能演算を実行できる汎用プロセッサの実現に向けた、CPU アーキテクチャおよびニューラルネットワーク・アクセラレータに関するものである。

今日の人工知能を活用した情報処理機構は、データセンタの高性能なコンピュータが、エッジで収集されたデータを集中的に処理する仕組みを取っている。しかし、集積されるデータ量の爆発的な増加に伴い、データセンタの大電力化や通信レイテンシ問題が顕在化し始めており、エッジ側にも人工知能演算の機能を持たせて負荷を分散させる必要性が唱えられている。エッジ端末の多くは電池容量および計算性能が限られており、それらに人工知能を搭載するためには、電力的に効率よく推論・学習処理を行うハードウェアが欠かせない。そこで本研究では、エッジに組み込まれるプロセッサへの応用を見込み、(1)CPU 自体の電力効率を向上させる汎用アクセラレータおよび学習型分岐予測器の開発、(2) 不揮発性メモリをニューラルネットワーク・アクセラレータとして利用するための学習則の開発に取り組んだ。

CPU はあらゆる演算を実行できるが、その消費電力のうち演算由来の電力が占める割合は 10% ほどで、ほとんどの電力はメモリアクセスや CPU 自体の制御で生じており、演算性能に対する電力効率が低いことが知られている。動的再構成アクセラレータは、大量の演算器を用いてプログラムの一部を CPU の代わりに並列処理することで、これらの非演算依存電力を縮小し、さらに実行速度も加速させる。また、プログラム実行中に演算器間の配線情報が動的に切り替えられるため多様なプログラムに対応できるが、この再構成時に生じる電力が却って全体の消費電力を引き上げてしまうことがあった。本研究では、プログラムのコントロール・データフローに着目して、演算器間配線のうちデータの依存性が変化し得るデータパスのみを動的に再構成し、柔軟性と電力効率を両立するアクセラレータを開発した。

演算以外の電力を削減する手法として、分岐予測器の高精度化も有効である。今日の RISC CPU は、およそ 10 段以上の深い命令パイプラインを有している。分岐予測の失敗は、パイプラインを一時停止して処理途中の命令をクリアし、さらにメモリから適切な命令やデータを読み出し直す必要が生じるため、パイプラインが深いプロセッサほど電力効率の低下を招く。本研究では、ベイズの定理に基づく統計的機械学習の手法を取り入れた分岐予測器を実際のソフトコア CPU ヘレジスタ転送レベルで組み込み、予測精度と消費電力をシミュレーションした。その結果、非学習型の分岐予測に対する高精度化に成功し、それに伴い消費電力が大きく削減され得ることを示した。

上述したような CPU の高電力効率化をもってしても、低電力な人工知能プロセッサの実現には不十分である。これは、ニューラルネットワークの演算のほとんどがニューロンの出力信号とシナプス荷重の積和演算の繰り返しであり、ノイマン型アーキテクチャとの相性が良くないことに起因する。そこで次に、不揮発性抵抗変化型メモリをニューラルネットワーク専用アクセラレータとして利用するための、学習手法の開拓に取り組んだ。

抵抗変化型メモリの各抵抗素子をシナプス荷重と見なすと、メモリセル内の電圧と抵抗素子の間に成り立つオームの法則、およびメモリセル間の電流に成り立つキルヒホッフの法則により、一度のメモリ読み出し動作で推論演算を実行することができ、低電力なアクセラレータとしての応用が期待されている。本研究では、三次元積層されたメモリの空間的特徴を深層学習に利用するための、畳み込み深層ネットワークの学習法を開発した。本手法ではネットワークに対して layer-wise な教師なし事前学習を施し特徴抽出器としての機能を持たせ、転移学習への応用の可能性も示された。

また、より汎用的な全結合ネットワークのために、シナプス荷重の正負符号を初期値で固定して学習する手法を開発し、符号非制約下の場合と同等の汎化能力を学習し得ることを示した。本来のニューラルネットワークの結合重みは学習次第で正の値も負の値も取りうるのに対し、実世界に負の抵抗値は存在しない。そこで、これまでの先行研究では、1 個の重みを 2 個の抵抗素子で表現し、同一の入力電圧に対する出力電流の差分をとることで、重みが正の場合と負の場合の両方に対応する必要があった。本学習法を用いると 1 個の重みを 1 個の抵抗素子で表現可能となり、面積効率の大きな改善が期待される。

以上、本論文に記した研究は、回路のアーキテクチャを切り口にプロセッサの低消費電力化を目指したものである。今日の人工知能技術の飛躍的進歩に代表されるように、情報技術の進化は常に集積回路技術の進化に支えられてきたが、半導体プロセスの点では、今世紀中に微細化の限界に達することが明白とされている。アーキテクチャの見直しにより大幅に電力性能を向上させられることを示したこれらの研究は、微細化限界後における集積回路研究の方向性を示唆する点で、重要であると考えられる。