(402) アーキテクチャ

動的再構成を用いたニューラルネットワークプロセッサ（DRNNP）の開発

Dynamically Reconfigurable Neural Network Processor ：DRNNP

山口拓哉 † 森下賢幸 † 小椋清孝 † 伊藤信之 †
† 岡山県立大学大学院情報工学研究科
Takuya Yamaguchi † Takayuki Morishita † Kiyotaka Komoku † Nobuyuki Itoh †
† Okayama Prefectural University Graduate School

1 はじめに

近年、ニューラルネットワーク（NN）の研究が盛んに行われている。実用規模の NN は、規模が大きくなり、認識や学習に多大の時間を要する。NN の処理を専用ハードウェア化することで演算速度を加速し、この問題を解決することを目指した研究が進められている。当研究室でも、CNN（畳み込みニューラルネットワーク）専用プロセッサの開発を進めているが、計算アルゴリズムを固定化しているため、多様なモデルには対応できず、モデルが変更されるたびに再設計が必要となる。

一方で、当研究室では多数の演算セルを有するセルアレイ構造を持つ DRCAP2(Development of 2nd generation Dynamically Reconfigurable Cell Array Processor)[1]の開発も進めている。このプロセッサは、動的再構成技術による命令レベル並列処理と自由なパイプライン処理を組み合わせて高速処理を実現する。DRCAP2 に用いられている動的再構成技術で作成されたセル部を基にして、CNN プロセッサのシナプス部やニューロン部を組み込むことで、階層型 NN(ニューラルネットワーク)の多様なモデルに対応でき、しかも NN 専用の処理部を高いパフォーマンスで処理できる NN プロセッサが開発できるのではないかと期待される。

本研究では、上記のような NN プロセッサのアーキテクチャを提案し、階層型 NN の基本的なモデルに対する動作クロック数や回路規模を評価する。

2 提案アーキテクチャ

最初に、階層型 NN のプロセッサについて、アーキテクチャを提案する。ただし 3 層の階層型 NN[2]を対象として、各変数の値や数値を供給するメモリ部の構成は無視して、各変数の値や数値が必要に応じて適宜与えられるものとする。提案するアーキテクチャを図 1 に示す。

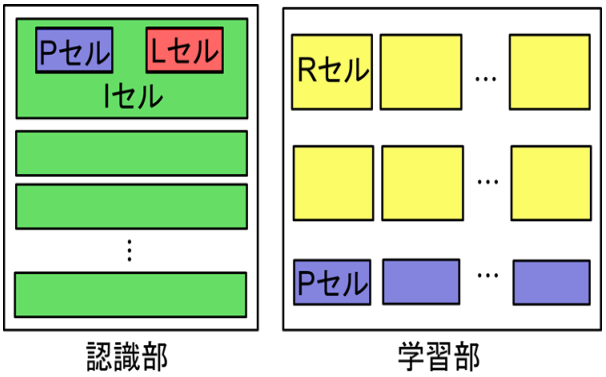

図 1 提案アーキテクチャの構造

I セルは、P セル 1 個と L セル 1 個で構成され、階層型 NN のフォワード処理を行い、認識部を構成する。R セルは、動的再構成を用いた並列処理演算器で構成され、プログラムに従って、構造を変更することができる。パイプライン処理と並列処理により、NN のバックワード処理を高速に行うことができる。I セルでは積和演算と活性化関数の計算を行っている。積和演算には P セルを用い、活性化関数の計算には L セルを用いる。P セルの構造を図 2 に、L セルの構造を図 3 に示す。

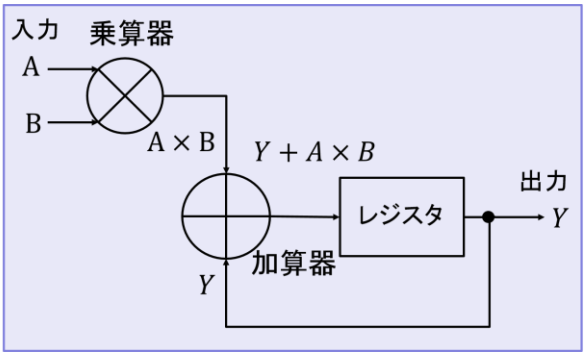

図 2 P セル(積和演算器)の構造

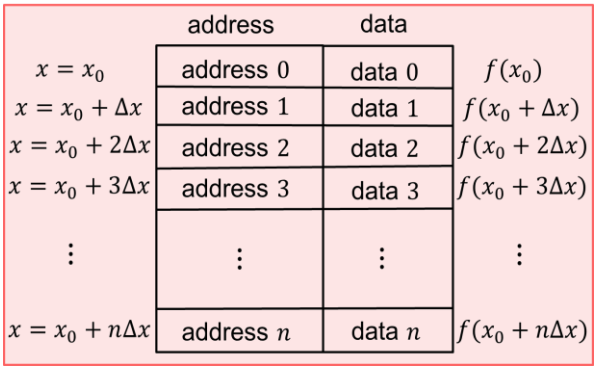

図 3　L セル(ルックアップテーブル)の構造

次に、R セルの構造を図 4 に示す。

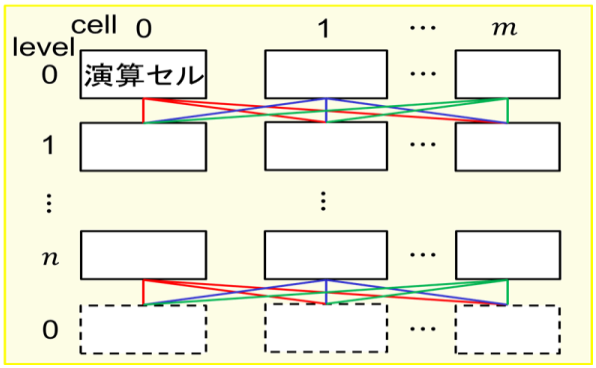

図 4　R セル(動的再構成セル)の構造

3　提案アーキテクチャの評価

提案アーキテクチャの動作クロック数を計算し、逐次処理を行ったときと比較し、逐次処理に対する加速率で評価する。加速率は逐次処理のクロック数を並列処理のクロック数で割ったものである。ただしメモリは無限にあり、データ受け渡しが瞬時に行われるとする。また L セル、P セル、R セルの演算セルでの計算にかかる時間は、全て 1 クロックとする。各部の処理毎にかかるクロック数を表 1 にまとめる。表 1 の各部の計算の並列度と加速率の関係を図 5 に示す。a は入力層のニューロン数、b は中間層のニューロン数、c は出力層のニューロン数を表し、l1 から l5 は各部の並列度を表す。

表 1　各部の並列度と加速率のまとめ

	中間層の出力	出力層の出力	出力層の重み	中間層の微係数	中間層の重み
逐次処理	$ab + b$	$bc + c$	$7(b+1)c$	$9bc$	$(a+1)bc$
パイプライン処理			$3 + (b+1)c$	$3 + bc$	
並列処理	$\frac{ab+b}{l_1}$	$\frac{bc+c}{l_2}$	$3 + \frac{(b+1)c}{l_3}$	$3 + \frac{bc}{l_4}$	$\frac{(a+1)bc}{l_5}$
加速率	l_1	l_2	$\frac{7(b+1)c}{3+\frac{(b+1)c}{l_3}}$	$\frac{9bc}{3+\frac{bc}{l_4}}$	l_5

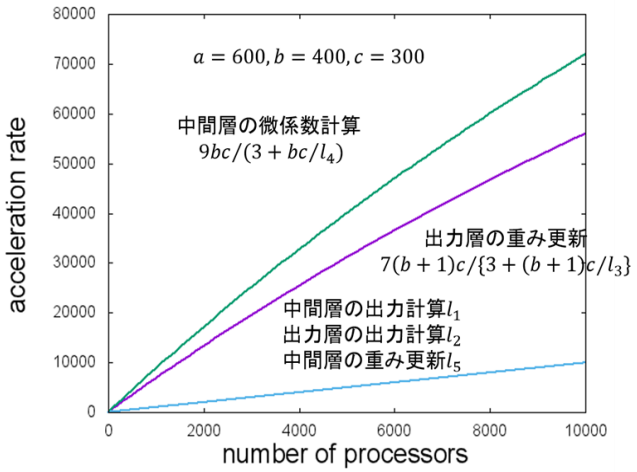

図 5　各部の並列度と加速率の関係

4　まとめ

DRNNP のアーキテクチャを提案し、階層型 NN に対する動作クロック数を計算して評価した。並列度と加速率はほぼ比例関係にあることがわかった。今後の課題としては、メモリ部や制御部の構成を考えることと、DNN（Deep Neural Network）への対応を行うことである。

参考文献

[1]森下 賢幸, 古賀 健一, 小椋 清孝, 伊藤 信之, “動的再構成可能なセルアレイプロセッサ DRCAP2 の開発”、第 25 回回路とシステムワークショップ論文集、pp.408-413、2012.
[2]小高 知宏、機械学習と深層学習、オーム社出版、2016