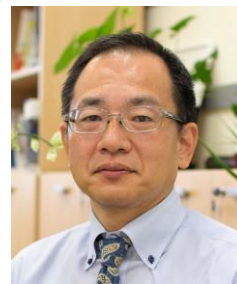


AI コンピューティング アーキテクチャの研究

教授 本村 真人

研究分野：ディープラーニング/アニーリング等のプロセッサ

ホームページ：<http://www.artic.iir.titech.ac.jp>



研究目的・内容

深層ニューラルネット(DNN)技術の勃興とともに、人工知能(AI)コンピューティングの分野が大きく進展しています。従来型のコンピューティングが「手続き型」であるのに対し、AI コンピューティングの分野は「構造型」であることを大きな特徴としています。その違いをアーキテクチャ(=処理方式)の革新に活かすことで、これまでよりも大幅にエネルギー効率や処理速度が高いコンピューティングシステムの実現が可能となります(図 1)。このような観点から世界中でアーキテクチャ変革の大規模競争が始まっており、我々はその中で独創的なアイデアで革新的なアーキテクチャを産み出す研究を続けています。

この領域の研究にはソフトウェアとハードウェアの間の接点を最適化する発想が重要です。この観点から、我々は「システムウェアなハードウェア」というキーワードを標榜し、AI コンピューティング分野を広くカバーする共通アーキテクチャプラットフォームの確立を目指して研究しています(図 2)。

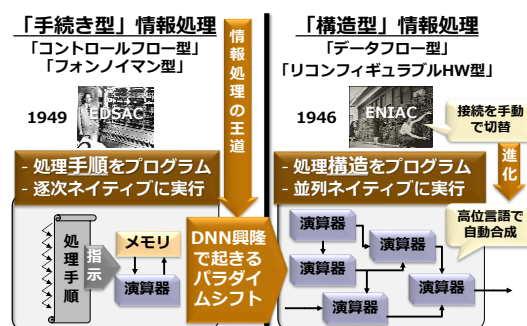


図 1. AI コンピューティング時代の
情報処理パラダイムの変革

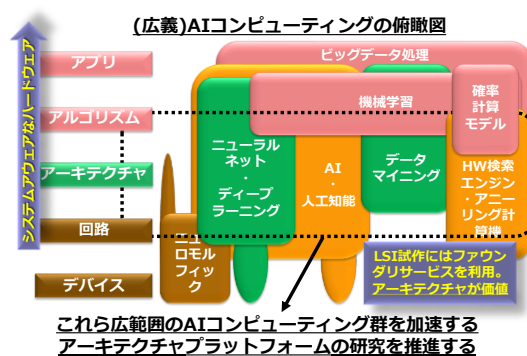


図 2. 研究コンセプト

研究テーマ

1. DNN アクセラレータの研究

DNN を支える深層学習(ディープラーニング)の世界では、量子化、枝刈、強化学習、転移学習、人間の脳に学んだ新しい学習手法など、日進月歩でアルゴリズムの進化が続いています。その応用範囲も、データセンター内のビッグデータ処理だけには留まらず、車やスマートフォンなどの身の回りのリアルな機器のスマート化へと急速に広がりを見せており、DNN 処理の低電力化・高速化するアクセラレータの研究が求められています。

我々は、バイナリ DNN アルゴリズム(DNN の重み係数や活性化値を 2 値で表現)に一早く注目して LSI アーキテクチャの研究を進め、2017 年に世界初のバイナリ DNN 推論エンジン LSI を発表し、大きな注目を集めました(図 3)。またバイナリから対数量子化(2 のべき乗表現の重み係数/ニューロン値の指数を量子化)までカバーする 3 次元積層リコンフィギュラブル型 DNN 処理エンジンを 2018 年に世

界で初めて発表し、いくつもの賞を受賞しました(図 4)。現在は、これらの成果をベースにして更に学習アルゴリズムとアーキテクチャの協創に踏み込んだ DNN アクセラレータの研究を続けています。

2. アニーリングマシンの研究

スマート化が進む社会においては、様々な制約が存在する中で、溢れ出るデータを活かして常に最適な判断や制御を続けていくことが求められます(交通、社会インフラ、農業、等)。このような問題は、数学的な枠組みとしては「組合せ最適化問題」として知られており、物理的な「エネルギー最小化原理」を活かしたアニーリング処理の考え方で解くことができることが知られています。いわゆる量子アニーリングマシンもその一つのアプローチですが、我々は DNN の学習プロセスとアニーリングによる求解プロセスの類似性に着目し、両者に共通に適用できる構造型のアーキテクチャプラットフォームとその LSI 実証の研究を進めています。

3. 動的再構成ハードウェア技術の研究

構造型の情報処理を支える基本技術が、「柔軟かいハードウェア」を実現するための動的再構成技術です。この技術のパイオニアとして AI コンピューティング分野への応用展開と技術革新を進めています。

● 教員からのメッセージ

2018 年度まで北海道大学 大学院情報科学研究科でこの分野の活動を続けてきました。2019 年度からは科学技術創成研究院(すずかけ台)に「AI コンピューティング」研究ユニット(ArtIC)を新設し、同研究ユニットを共同運営する劉載勲准教授とともに、日本の中核研究拠点となることを目指して活動を広げています。新しい「構造」を考えることが好きな人、「設計」が好きな人、ソフトとハードの両方をわかりたい人、情報処理の変革に自ら携わりたい人など、何かを創り出すことへの情熱を持った学生の皆さんを歓迎します。国家プロジェクトへの参画、企業共同研究、海外研究機関との協力等も積極的に行っており、一線級の国際会議で研究成果を発表する機会を提供できます。

● 関連する業績、プロジェクトなど

1. 研究活動全般の紹介(ArtIC ホームページ/研究活動)
<http://www.artic.iir.titech.ac.jp/wp/research/>
2. 東工大ニュース: 組合せ最適化問題を高速に解く新しいアニーリングマシンを開発 (2020/02/18)
<https://www.titech.ac.jp/news/2020/046309.html>
3. NEDO セミナー講演: 「AI コンピューティングがアーキテクチャにもたらすもの」 (2018/10/31)
<https://www.nedo.go.jp/content/100885737.pdf>

受賞 92 年 IEEE JSSC Best Paper Award、99 年 IPSJ 年間最優秀論文、11 年 IEICE 業績賞、18 年 ISSCC Silkroad Award(指導学生)、19 年 JSPS 育志賞(指導学生)、等

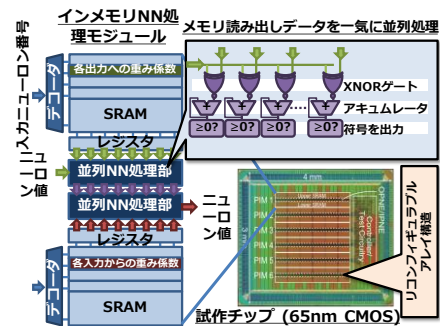


図 3. バイナリ DNN 推論チップ (2017 年)

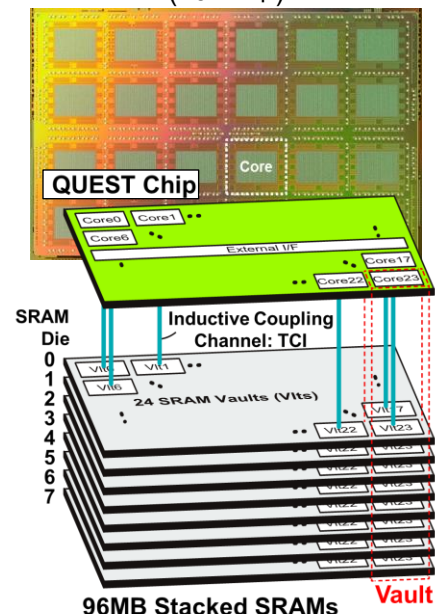


図 4. 3 次元積層・対数量子化 DNN チップ(2018 年)