

HDCアクセラレータとRISC-Vを組み合わせたエッジサーバの開発

— 高速で低消費電力なAIを身近な存在に —

1. 背景

複雑な認知タスクを処理する従来の機械学習技術は進化が目覚ましく、様々なタスクを実行できる高精度なアルゴリズムが多く考案されている。しかし、それらの技術は強力なマシンパワーを必要とし大量のGPUやCPUを動かすため、莫大なエネルギー消費量と計算量を要求する。これはハードウェアリソースが限られる組み込みデバイスやIoTデバイスでの利用に不向きであるだけでなく、環境課題の観点から見ても持続可能性が低いことが指摘できる。

この課題に対する解決策として、Hyperdimensional Computing (HDC) と呼ばれる手法を従来の機械学習手法の代替として注目する動きがある。HDCは、数千から数万次元のベクトルのHyper Vectors (以下、超次元ベクトル) を用いた計算モデルであり、様々な認知タスクに応用可能だ。超次元ベクトルに対しての独自のベクトル演算が定義されており、バイナリの超次元ベクトルとして応用することでコンピュータからも扱いやすくなる。そして、超次元の特性により入力されたデータに対してロバストであることや、回路を最適化することでHDCアルゴリズムの高速化がしやすいことや、その小型化による省電力化が行えることなど、様々な恩恵を期待できる。また、すでに一部の認識タスクでは深層学習と比較しても低消費電力で高精度な結果を出すケースが報告されており、実用性も期待できる。

一方で、HDCには独自のベクトル演算が定義されているため超次元ベクトルを十分に高速に実行するには様々な工夫が必要になる。この点がボトルネックとなり、現在も他の機械学習に比べると実際の開発現場で使われることがほとんどない。そのため、「ロバスト」「高速」「低消費電力」な特性を持つHDCを私たちの生活に普及させるためには、HDCに最適化された簡単に使える実用的なアクセラレータの開発が鍵となる。

2. 目的

本プロジェクトは、簡単にHDCを使ったアプリケーションの開発が行える環境を整備し、HDCの優れた点を活用したアプリケーションの普及を促進することが究極的な目的である。そのために、既存のものよりも使い勝手の良い、CPUと協調動作可能なHDCアクセラレータ (Hyperdimensional Computing Processor Unit, HPU) とそのアセンブラやライブラリを開発し、そしてそれらを使った実用例を示す。

提案したアーキテクチャやライブラリを使って、様々なアプリケーションを実装することは本プロジェクトの重要な目的に繋がる。すなわち、それは本プロジェクトの成果をアピールするだけでなく、HPUの使い方や実用例を示すことで実際にHDCをアプリケーションへ応用することを開発者に強く勧める効果があると期待している。

3. 開発の内容

本プロジェクトは当初、次の4つを目標とした。1) HDCアクセラレータを開発する。2) HDCアクセラレータを活用したエッジサーバを開発する。3) 開発者向けライブラリを開発する。4) 応用的なアプリケーションを開発し実用可能性を検証する。

一つ目のHDCアクセラレータの開発に関しては、図1で示すような、HDC独自の演算である3種の演算（Bind, Bound, Permutation）全てに対して高速に処理できる演算器を実装したアーキテクチャの開発に成功した。

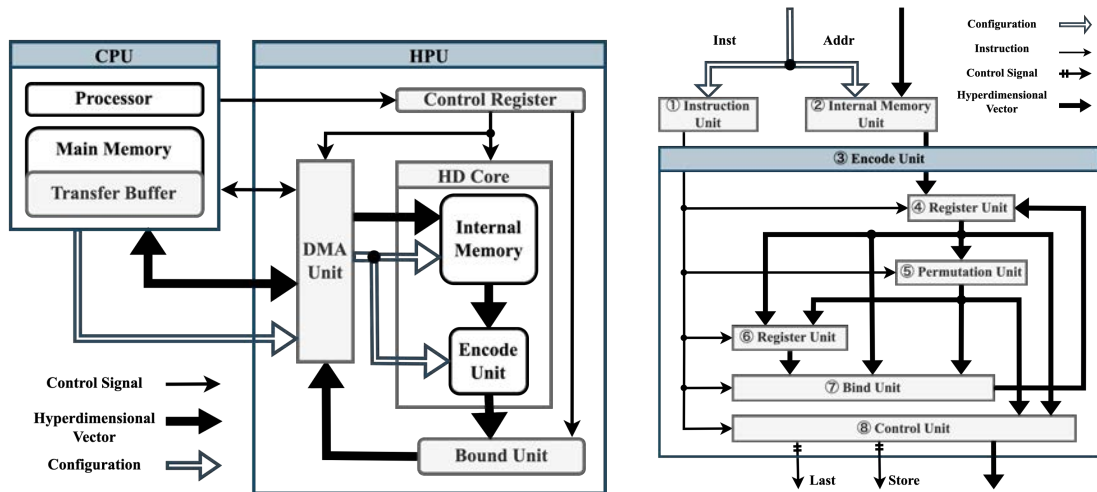


図1 提案するHPUの概要図

エッジサーバは図1の左側で示すように、CPUと協調動作するアーキテクチャとして開発した。このアーキテクチャによりHDCアクセラレータの応用先が広がる。本HPUはCPUと協調動作可能であるため、エッジサーバとして応用可能なだけでなく、IoTデバイスや一般消費者向けのCPUと協調動作できることを目指した設計となった。

そして、このアーキテクチャをC言語を使って簡単にCPUから実行できるようなライブラリの開発も成功した。これにより、ユーザはライブラリを使用すれば関数形式でCPUからHPUに指示を出すことも可能になっている。例えば、N-gramの符号化方式は図2で示すコードのようにC言語で書くことになる。

最後の実用可能性の検証には、既存の論文で紹介されている手法を使った画像認識・音声認識・言語認識のアプリケーションを開発し、本アクセラレータを搭載した低消費電力向けCPUで実行速度を計測した。いずれの結果でもM1 Max (3.00GHz)、Intel Core i7 (1.80GHz)、ARM-v7 Cortex-A9 (0.666GHz)と比較した場合、大幅な速度改善が確認され、最大で169倍の高速化に成功している。また、電力効率を示す指標の一つであるエネルギー遅延積（ED積）は最大13,469倍も改善した。それぞれの認識タスクにおける各CPUとの速度比較のグラフは図3の通りである。

```

① Load_1(0); // アドレス0からレジスタ1にロード
② Load_2(1); // アドレス1からレジスタ2にロード
③ Permute_2(1); // レジスタ2の値を1回Permutation
④ Bind_p11(); // レジスタ1とPermutationの結果をBindしレジスタ1に格納
⑤ Load_2(2); // アドレス2からレジスタ2にロード
⑥ Permute_2(2); // レジスタ2の値を2回Permutation
⑦ Bind_p12(); // レジスタ1とPermutationの結果をBindしレジスタ2に格納
⑧ Bound_2() // レジスタ2の値をBound

```

図2 N-gramの符号化をHPUに指示するコード例

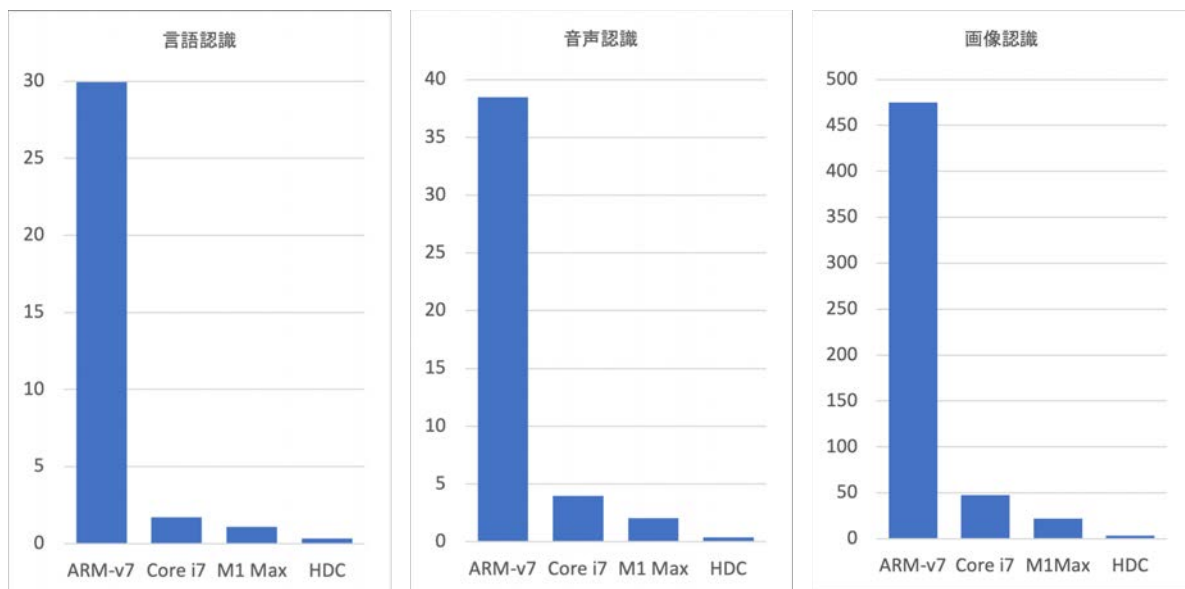


図3 アプリケーションテストにおける速度比較

以上により、周波数で圧倒的に優っている汎用的なCPUと比較しても、HDCプロセッサが最も高速にHDCタスクを実行できることを示しており、HPUの実用性をアピールできた。

4. 従来の技術（または機能）との相違

従来の技術としては、CPUを使ったものや、IoT向けに開発されたHDCアクセラレータ、そしてプログラマブルなHDC計算基盤などがある。しかし、これらは表1でまとめるように、「計算速度」「電力効率」「プログラマビリティ」「CPUと協調動作可能か」という観点から評価するとどこかに課題が存在している。本プロジェクトが開発するHPUはこれらの課題を包括的に解決しており、さらにその速度や電力効率を定量的に評価することで実用性へのアピールも行なっている。

表1 既存手法と提案手法の比較

	速度	電力効率	プログラマビリティ	CPUとの協働性
CPU	×	×	◎	—
HDCアクセラレータ	◎	◎	×	×
プログラマブルな HDC計算基盤	—	○	○	×
HPU	○	○	○	◎

5. 期待される効果

本HPUの開発をさらに進めることで、低消費電力かつ高速なHDCアクセラレータを低コストで提供できる可能性を示すことができた。そのため、さらに多くのエッジコンピューティングがHDCアプリケーションを採用することを期待している。これにより従来の機械学習を採用した場合よりも高速、低消費電力、かつ高精度な分類器が普及する。そして、そのようなエッジサーバやIoTデバイスはあらゆる分野に応用でき、様々な効果をもたらすことができるだろう。自律型のロボットが街中でも普及したり、交通整備や自動運転にも使われる。あるいは、ARやVRの分野でも活躍が期待できる。

6. 普及（または活用）の見通し

IoTのようなエッジデバイスの開発者とHDCに関するアルゴリズムの研究者に活用してもらうことを目指す。今回の成果を広めることができれば、これまでよりも手軽で高速にHDCを実行できるようになるため、HDCアプリケーションの研究者や開発者は実験や開発に必要な試行錯誤のサイクルを大幅に改善できるようになる。また、安価でCPUと協調動作することもできるため、既存のIoTアプリケーションの拡張としても利用できる。

7. クリエータ名（所属）

井阪友哉（奈良先端科学技術大学院大学）