

科学研究費助成事業 研究成果報告書

平成 29 年 6 月 9 日現在

機関番号：14401

研究種目：基盤研究(C) (一般)

研究期間：2014～2016

課題番号：26330065

研究課題名(和文) 柔軟な構成変更が可能な機械学習専用VLSIの実装

研究課題名(英文) VLSI Implementation of Generic Hardware for Machine Learning

研究代表者

尾上 孝雄 (Onoye, Takao)

大阪大学・情報科学研究科・教授

研究者番号：60252590

交付決定額(研究期間全体)：(直接経費) 3,700,000円

研究成果の概要(和文)：本研究では、応用システム要求に応じて柔軟に構造を変更しつつ、高速な識別処理を実現する機械学習専用のハードウェアエンジンを実装した。サポートベクタマシンの識別能力を維持しつつ大幅な高速化が可能であるソフトカスケード処理を適用し、データの次元数、演算精度を変更することができる再構成可能アーキテクチャとして実装した。FPGAならびに45nmテクノロジーで実装した結果、要求性能に応じた構成が採れることを確認するとともに、FPGA版175.7MHz動作にて、毎秒79VGAフレーム7HDフレームの処理能力を持ち、45nm版では、最大毎秒361VGAフレーム35HDフレームの処理が可能であることを確認した。

研究成果の概要(英文)：In this research, design of generic hardware for high speed classification in machine learning is proposed with providing flexible adaptability of its hardware organization. Specifically, a Support Vector Machine, which has inherent versatility for applications is accelerated by soft-cascade processing while maintaining its classification capability. Dimension of feature vectors and bit precision in calculation can be controlled without re-designing hardware. VLSI implementation is coordinated with the use of FPGA and 45nm technology to confirm the ability of adaptive reconfiguration. The designed FPGA and 45nm circuit attain 79VGA frames/7HD frames processing and 361VGA frames/35HD frames processing per second, respectively.

研究分野：応用集積システム

キーワード：機械学習 VLSI アーキテクチャ

1. 研究開始当初の背景

我々の周りにさまざまな情報が氾濫する、いわゆるビッグデータの時代に突入し、2020年には約40ゼタバイトと10年間で約50倍に増大するといわれている。我が国でも、平成25年6月に相次いで閣議決定された、「科学技術イノベーション総合戦略」「日本再興戦略」「世界最先端IT国家創造宣言」の中で、新たなビッグデータ利活用技術に関する研究開発を重点的に進めるとしている。しかしながら、Google、Microsoft、Appleなど米国の巨大データセンターを有する企業における研究開発は群を抜いており、国内での研究開発は特定の応用システムを対象としたものや一部の処理技法に着目したものに限定されているのが現状である。これらの技術の実用化は実際にデータを保持する機関との協調が必須であるため、不況からの業績回復を目指す国内の電機メーカ、情報通信関連業界の参入には障壁が高い。

2. 研究の目的

ポストユビキタス環境とされているアンビエント情報環境は、このようなビッグデータ利活用の好例であるといえる。空間内で多数の人と多数の環境デバイスがインタラクションを起こすことが想定されており、そのバックエンドには取得した大量のデータを処理する計算サーバが存在する。環境状況を認識するための主要な処理の一つとして機械学習による識別処理が挙げられる。例えば画像センサから入力される環境映像中の物体の認識や、端末操作による行動認識がこれにあたる。従来、このような処理では、端末は入出力のみを担当し、識別処理はサーバ側で行われていた。これは、識別処理に用いるデータが膨大であり機械学習の処理負荷も大きいためである。そこで本研究では、

- (1) サポートベクタマシンの小型低電力高速化アーキテクチャ
- (2) 再構成可能アーキテクチャの実装
- (3) ソフトウェア・ハードウェア協調による識別処理

について検討する。

3. 研究の方法

下記の方法で研究を実施した。

- ・サポートベクタマシンの小型低電力高速化アーキテクチャ

特定の応用システムに特化せず高い汎化性能を持ち、1995年に提案されてから20年近く経った現在でも広く利用されているサポートベクタマシンについて、一般的な線形サポートベクタマシンを高速化するソフトカスケードを適用する機構を、小型低電力で実装するVLSIアーキテクチャを構築する。特に、特徴ベクトルの次元数や演算ビット数をプログラマブルとしつつ、モジュール構成として複数活用による演算速度要求要件を満たすことが可能なようなアーキテクチャ

構成とする。

- ・再構成可能アーキテクチャの実装
- 前述のアーキテクチャを、実際にハードウェアとして実装する。特に、特徴ベクトルの次元数や演算ビット精度は応用システムや特徴抽出手法により異なってくるため、ハードウェア実装の場合でも、これらをパラメータ化して実装することが可能なよう再構成可能アーキテクチャとして粗粒度(多ビット)要素を中心とした実装とする。
- ・ソフトウェア・ハードウェア協調による識別処理

一般に、機械学習を利用する場合には、対象とする応用システムに適切な入出力管理・機能動作管理などのアルゴリズム処理を執り行う必要がある。ここでは、上記の再構成アーキテクチャを発展させたハードウェア実装に対して、パラメータセットや機能切替えを含むソフトウェア処理との協調動作を行い、その機能を確認する。

4. 研究成果

まず、図1に示すソフトカスケード構造を持つサポートベクタマシンのアーキテクチャを構築した。

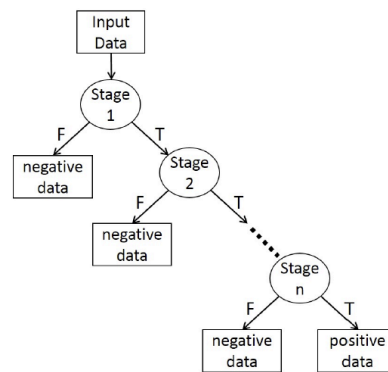
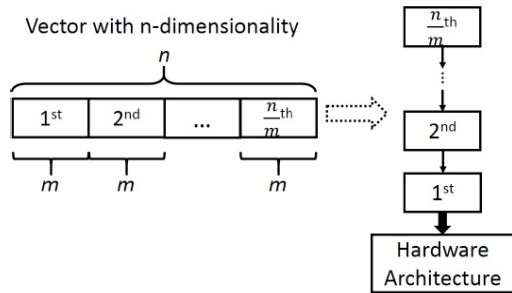


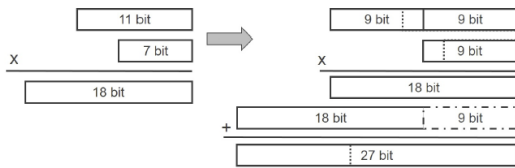
図1: ソフトカスケード構造

アーキテクチャ構築に当たっては、特徴ベクトルの次元数と演算ビット精度に対する自由度を担保する必要がある。特徴ベクトルの次元数については、ソフトカスケード処理との整合性を担保する必要上、図2(a)に示すように、全体の特徴ベクトル(n次元)をm次元ごとに均等に分割する構成を採った。これにより、分割された格段の処理は独自に行うことができ、スケーラビリティと高速化の両立が可能となる。一方で、ビット精度については、補数表現の場合の拡張符号ビットの関係上、均等に分割することは困難である。このため、図2(b)に示すように、LSB (Least Significant Bit)優先で、ワード分割する構成を採った。上記の方針のもと、実装の拡張性と数十万次元という非常に大きい次元数を持つ特徴ベクトルの場合でも、問題なく処理が可能なように、モジュール構造を採用し、複数のモジュールに同時に、あるいは、時分割で繰り返し特徴ベクトルを入力できるアーキテクチャを構築した。図3に全体構成を示すように、乗算器モジュール部、累算器部、

比較器部、制御部を有するハードウェアを実装した。



(a) 特徴ベクトルの分割



(b) 演算ビット数の分割

図 2: 柔軟な汎用性を担保する分割構造

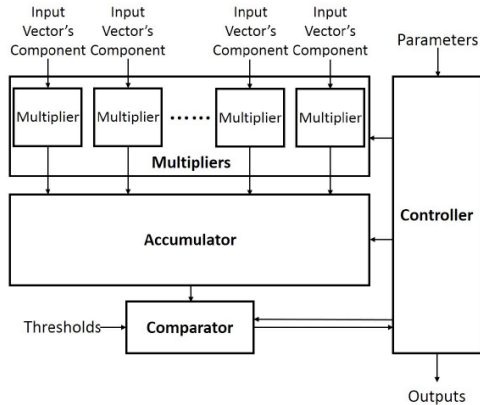


図 3: SVM ハードウェアエンジンのアーキテクチャ

上記のアーキテクチャを Xilinx Kintex-7 XC7K325T FPGA ならびに 45nm Nangate ライブラリに VLSI 実装した。その結果、図 4 に示すように、乗算モジュール数と演算ビット精度を変化させることでハードウェアリソースの使用数を制御できることを確認した。

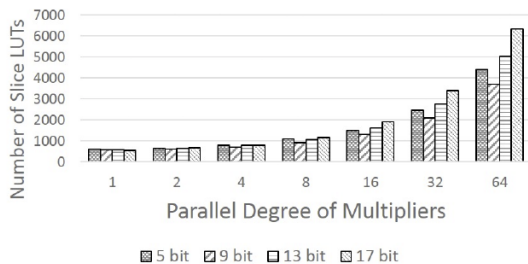


図 4: FPGA スライス LUT の使用状況

実行性能については、FPGA 版 175.7MHz 動作にて、毎秒 79VGA フレーム 7HD フレームの処理能力を持ち、45nm 版では、最大毎秒 361VGA フレーム 35HD フレームの処理が可能であることを確認した。

さらに、上記の概念を発展させ、ACF (Aggregated Channel Features)+ BDT (Binary Decision Trees) 構造を持つ、新しい処理エンジンを設計し、図 5 に示すような、ソフトウェアとハードウェアを協調動作させるシステムとして、Kintex7 FPGA と ARM Cortex-A9 プロセッサを集積する Zynq7000 (XC7Z045) 上に統合化フレームワークとして実装した。本フレームワークでは、運転者補助をアプリケーションとして、自動車(前方車・対向車)、歩行者、信号(青・黄・赤、矢印)を同時に認識しつつ、HD フレームをリアルタイムに処理可能である。

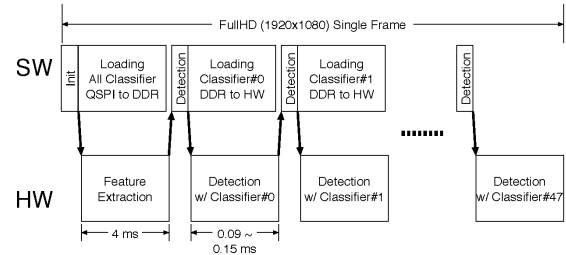


図 5: ハードウェア・ソフトウェア協調動作

5. 主な発表論文等

[雑誌論文] (計 4 件)

- ① H. Konoura, D. Alnajjar, Y. Mitsuyama, H. Shimada, K. Kobayashi, H. Kanbara, H. Ochi, T. Imagawa, K. Wakabayashi, M. Hashimoto, T. Onoye, H. Onodera: "Reliability-Configurable Mixed-Grained Reconfigurable Array Supporting C-based Design and Its Irradiation Testing," IEICE Trans. Fundamentals, E-97-A, pp. 2518-2519, 2014.
- ② K. Tomita, M. Hatanaka, T. Onoye: "Implementation of Viterbi Decoder toward GPU-based SDR Receiver," IEICE Trans. Fundamentals, E98-A, pp. 2246-2253, 2015.

[学会発表] (計 7 件)

- ① R. Miyamoto, J. Yu, T. Onoye: "Normalized Channel Features for Accurate Pedestrian Detection," in Proc. 2014 Int'l Symp. Commun., Ctrl, and Signal Processing, 2014.
- ② E. Aliwarga, J. Yu, M. Hatanaka, T. Onoye: "Design of Generic Hardware for Soft Cascade-based linear SVM Classification," in Proc. Int'l Symp. Intelligent Signal Processing and Communication Systems, 2015.
- ③ T. Onoye: "Implementation Technology of Media Processing Engines - Traditions and Trends," Int'l Conf. Digital Arts, Media and Technology, 2016.
- ④ M. Koga, T. Onoye, J. Yu, T. Azuma, E.

Aliwarga: “Vision-based
Comprehensive Framework for Senior
Driver Assistance,” in Proc. 11th ITS
European Congress, 2016.

6. 研究組織

(1) 研究代表者

尾上 孝雄 (ONOYE, Takao)

大阪大学・大学院情報科学研究科・教授

研究者番号：60252590