

研究終了報告書

「高性能ストリームデータ圧縮技術の開発」

研究期間:2020 年 11 月～2024 年 3 月

研究者: 山際 伸一

1. 研究のねらい

インターネットのコモディティ化に伴い、すべての情報機器がそれにつながり、新たな社会インフラを構築する Society 5.0 の時代が到来しようとしている。数百万といった数の IoT 端末からデータが爆発的に生成される状況はこれまでの情報処理技術に新たな挑戦を投げかけている。その莫大なデータ量を扱うことが必要であるが、従来からのプロセッサの性能を向上するだけでは対応できる規模ではない。そこで、その莫大なデータをロスレス圧縮することで削減し、システムの負荷を下げることで、IoT のシステムでの処理の許容可能性を高める方法を提案した。このような IoT のシステムが扱うデータは常に流れ続けるデータストリームであることが特徴であり、リアルタイムに圧縮できる方法を適用すれば良い。しかし、従来からの方法では、データをバッファリングすることで圧縮処理を施す。データ圧縮の元となる情報理論であるシャノンの平均情報量を使った方法では、データ全体のビットパターンの出現頻度を把握しないと符号化できない。ストリームデータは無限遠であり、このような統計手法を扱えず、さらに、ハードウェアによる高速化を狙うとしても、完全なハードウェア化が不可能という問題がある。そこで、ストリームデータを扱える第3世代のロスレス圧縮技術を開発した。2つの単位データ(シンボルと呼ぶ)を1つのシンボルに変換することで、処理を定量化し、コンパクトにハードウェア化が可能な LCA-DLT を開発した。LCA-DLT はそのモジュールを数珠つなぎにして、1つのモジュールが圧縮可能なシンボルの個数より長いパターンを圧縮できるユニバーサル性をもっていたが、出力される圧縮データはシンボルよりも少ないビット列に削減する可変長圧縮ができず、圧縮率が良好ではないことが弱点であった。それを克服する可変長圧縮が可能な ASE Coding を開発し、単位データを最少で1ビットにまで圧縮できた。しかし、これらのストリームデータ圧縮技術は可変長圧縮とユニバーサル圧縮のそれぞれを別なアルゴリズムで実現するまでには到達したが、双方の利点である、圧縮シンボルのビット長が可変長で、さらに、ユニバーサル性を併せ持つストリームデータ圧縮技術を新たに開発することが本研究の狙いである。

2. 研究成果

(1) 概要

本さがけ研究活動では IoT のデータ流通やリソース、および処理時間の短縮といった圧縮に関する研究に注力し、ストリームデータ圧縮技術のエントロピー符号化とユニバーサル化の融合を目指した。その中で、大きく 4 つのテーマに関して成果を得ている。本研究課題はハードウェア化可能なストリームデータ圧縮技術の開発を狙いとしており、その基盤技術である IoT 向けのプロセッサ開発に取り組んだ。研究開始当初から RISC-V が新たな潮流として、その命令セットのロイヤリティフリーの性質やソフトウェア開発環境に注目していた。その仕様に沿ったプロセッサを開発した。RISC-V プロセッサを中心としたストリームデータ圧縮技術を備えた SoC を

開発し、キャッシュ部分について、ストリームデータ圧縮技術の特性を活かした I/O 技術の特許出願した。また、本さがけ研究の開始時点では ASE Coding の特性に関する知見が十分ではなかった。特にソフトウェアにおける実装面での高速化に未知なる部分が多く、ハードウェア実装との組合せにおける処理速度のミスマッチが問題であったが、解凍側でのバッファリング技術による処理の多重化に加え、圧縮処理の並列実行によって高速化する手法を発表した (Kato, Yamagiwa, ISCC2023, BigData2023)。一方で、IoT アプリケーションに関する研究を行った。従来、音声に適用されていた ADPCM を動的にビット数変更が可能な Variable Bitrate ADPCM を開発し、映像伝送による有効性を示した (Yamagiwa, Ichinomiya, Sensors, 2023)。ロスレスデータ圧縮では圧縮後のデータサイズが予測不可能である問題に対し、深層学習で最良のアルゴリズムを選択する手法を提案した (Yamagiwa, Yang, Electronics, 2023)。また、身体科学への IoT の応用となる基盤技術も研究し、運動の直前体制からその因果と成功/不成功を予測する機械学習手法を開発した (Kato, Yamagiwa, Sensors, 2023, Computation 2022)。また、深層学習におけるネットワークモデル圧縮技法をヒューリスティックに組み合わせることでモデルを最小化する手法を開発した (Danhe, Yamagiwa, Wada, Sensors, 2022, IEEE Access, 2024)。さらに、ストリームデータ圧縮技術における圧縮データストリームに例外処理を組み込む手法も開発した (Yamagiwa, Marumo, Kuwabara, Electronics, 2021)。そして、本研究の真の狙いである可変長圧縮が可能で、ユニバーサル化されたストリームデータ圧縮技術を開発した。そのアルゴリズムや性能に関する考察を加え、原著論文として発表した (Yamagiwa, Kato, IEEE Access, 2024)。上述の RISC-V を使った SoC として組み込み、システムとして実証する FPGA による実験システムを開発し、性能および省電力化についての実証環境を整えた。以上の成果をもって、本さがけ研究の目的を達成し、研究の狙いを達成できた。

(2) 詳細

上述の概要での成果についてそれぞれを説明する。

(1) ストリームデータ圧縮技術を搭載した SoC の開発

ストリームデータ圧縮技術と協働する基盤技術である IoT 向けのプロセッサ開発として図 1 に示す SoC を開発した。RISC-V プロセッサをコントローラとして、周辺機器としてバスに接続された ASE Coding のモジュールを介し、メモリの内容や通信データを圧縮する。このメモリ制御部分に関して特許出願した。この構成をハードウェア記述言語によって RTL での実装を行い、FPGA での動作検証を行った。

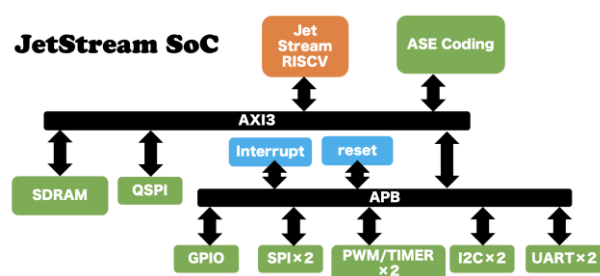


図 1: SoC の構成

(2) ASE Coding のソフトウェア実装における高速化技法の開発

ロスレス圧縮技術では、解凍後のデータサイズが予見できず、解凍用のバッファメモリを解凍

後のサイズで確保できないため、小さなバッファを繰り返し用いて解凍する作業が必要であり、その粒度がストレージへの書き出しのサイズとミスマッチを起こす問題があり、それに対する解決策として、SOB-EOB と呼ばれるバッファサイズを圧縮データチャンクに埋め込む方法を開発した。この方法をストリームデータ圧縮に適用し、ASE Coding の解凍側のスループットが向上した。さらに、ストリームデータ圧縮は、圧縮処理のアルゴリズムに並列化の余地がないため、図 2 に示すように、圧縮器をスレッドで並列化し、ストリームをチャンクに分割して並列実行することで、高速化する手法を開発した。並列化の際に、その並列度によって 3 つの手法を提案した。この方法ではチャンクへの分割がデータエントロピーの連続性を失うことによる圧縮率への影響が懸念されるが、

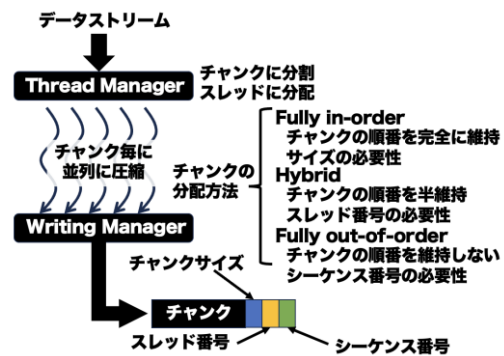


図 2: ストリームデータ圧縮の並列化

その影響は少ないことを実験的に示した。さらに、並列度の高い手法ではチャンクに情報を付加するため圧縮率に影響が出るが、全体に対して微少である事を示した。

(3) IoT アプリケーションに関する研究

映像をデータストリームとして高速伝送する方法として、量子化技術がある。映像の場合、各ピクセルのビット数を削減し、その前後のピクセルの関係からの元の情報量に近い色情報を復元するが、その際にデータが劣化してしまう。その劣化具合をデータ伝送の動的な性能変化に合わせ、伝送スピードが速いタイミングでは高精細に伝送する新しい手法を提案した。従来から電話などの音声に適用されていた ADPCM を動的にビット数変更が可能な Variable Bitrate ADPCM を開発し、映像伝送による有効性を示した。

また、多くのロスレスデータ圧縮のアルゴリズムが提案されているが、用いるアルゴリズムとその対象となるデータのエントロピーの組合せで、圧縮後のデータサイズが決まる。すなわち、アルゴリズムにより同じデータであっても同一の圧縮率を得られない。最も良好な圧縮率を得るには全てのアルゴリズムで検査するしかないが、それは NP 困難問題であり、大きなデータに対しては現実的な時間で完了できない。このように圧縮後のデータサイズが予測不可能である問題に対し、同一のデータパターンに対して複数のアルゴリズムの圧縮率に関するランキングを正解とする深層学習で、最も効率よく圧縮出来るアルゴリズムを選択する手法を提案した。画像に対して実験を行い、そのデータブロック毎に推論することで、データ全体に対して単一のアルゴリズムを選択する場合に比べ良好な圧縮率を維持できることを示した。

さらに、IoT センシングアプリケーションへの基盤技術として、身体科学への IoT の応用となる基盤技術も研究した。運動の直前体制に関して、映像からその特徴点を抽出し、その因果と成功/不成功を予測する機械学習手法を開発した。特に因果に関する予測においては、複数の運動の結果が予測されるため、柔道の投げ技を実験データとして用いる事で、組手の段階でどの技に移行するかを予測できる事を見出した。この発見によって、投げ技のトレーニング

段階での運動の要点をデータのみから発見することが可能なことを提示した。

データ圧縮と同様に、深層学習ではネットワークモデルを縮小することが、ハードウェア実装においてはリソース量の削減につながり、ソフトウェア実装においては計算量の削減につながる重要な技術である。畳み込み深層学習におけるネットワークモデル圧縮技法をヒューリスティックに組合せることでネットワークモデルを最小化する手法を開発した。Structured/unstructured pruning、そして量子化技法の3つの異なる特性を持つモデル圧縮技術を使って、それぞれの適用率を制御することによって、現実的な処理時間で最小化できる手法を開発した。

最後に、ストリームデータ圧縮技術における圧縮データストリームに例外処理を組み込む手法も開発した。ロスレスデータ圧縮では、圧縮シンボルに割り当てられるコードは原データと一対一対応していなければならない、そのコードに意味を持たせることができない。そこで、圧縮器で故意にエラーを発生させる圧縮コードを出力することによって、解凍側に例外の発生を検知させる方法を開発し、ASE Coding に組み込んだ。この方法により、圧縮データストリームの末尾を知らせる、解凍側に圧縮側のパラメタの変更を知らせる、といった特殊な操作をデータストリームに埋め込むことが可能になった。

(4) エントロピー符号化とユニバーサル化を併せ持つ圧縮技術の開発

さきがけ研究の狙いであるストリームデータ圧縮技術のエントロピー符号化とユニバーサル化を融合した新圧縮技術を開発した。本圧縮技術はデータストリームを滞りなく、たった一度のスキャンをすることで複数シンボルを1ビットにまで圧縮することができる。この圧縮方式はLCA-DLTで2シンボルから1シンボルに圧縮する方式でのユニバーサル性とASE Codingにおけるエントロピーをリアルタイム解析し、シンボルを1ビットにまで圧縮できる利点とを融合し、複数シンボルからなるデータパターンに対し拡張した本研究の狙いに合致する研究成果である。無限遠のデータストリームを扱うには一度のデータスキャンで圧縮操作を実施しなければならない。しかし、従来からZIPなどで利用されているDeflateのように複数の圧縮技法を組み合わせ、さらにメモリを用いて何度もデータスキャンが可能な方式と比べても、その性能に匹敵する圧縮効率を示すことを確認している。すなわち、十分にユニバーサル性の利点を活かした圧縮性能を実現できていることを確認し、この圧縮技法に関する論文を発表した。また、ハードウェアでの実装に親和性があるようにアルゴリズムを最適化した。その性能の実証、および、データ圧縮によってデータ伝送路の電力を下げられることを実証するため、評価用ハードウェア基板を作成し、実証実験を行っている。特に8Gbyte/secの超高速伝送を実現する圧縮、解凍が可能なPCI Expressボードでの実装はデータセンターなどのクラウド側での利用に十分に耐えうるスピードで、データ圧縮の遅延を隠蔽して通信量の削減とストレージの小型化が可能になる。IoTエッジデバイスを模したシステムも開発したため、エッジデバイス側での処理に関して、省電力化の可能性を示す実証実験を行っている。

3. 今後の展開

本さきがけでの成果として開発したエントロピー符号化とユニバーサル性を同時にもつ圧縮技術を半導体、および、データセンターを保持する企業への利用を推進し、その効果について実感し

でもらうと共に、ビジネス利用へ向け、アプリケーション開発を行う産業界との共同研究をすすめ、社会実装を行っていく。しかしながら、本研究課題で開発した圧縮技術は、現在はまだ、その情報理論的側面に関する特性が十分に解明できていない部分がある。圧縮率に影響を与えるパラメタ設定や制御に関わるタイミング制御に関してのノウハウを蓄積する必要があると共に、データエントロピーの高効率な把握ができる制御方法を解明していくことが必要である。今後は最大性能が発揮できるハードウェア実装の最良のハードウェア/ソフトウェアデザインを見出す作業を行っていく事を考えている。

4. 自己評価

本研究課題の採択直後にコロナ禍となり、研究者間の交流が途絶え、領域会議も全てオンラインとなったため、他の研究テーマの研究者との間での交流が困難となり、さきがけ研究らしい活動ができなかった。さらに、当初、LSI 化を計画していたが、全世界的な半導体不足、特にデジタル回路の半導体の不足に巻き込まれ、ハードウェアの研究の遂行がほぼ不可能となった。しかし、最終年度の後半で市場状況が改善されたため、ハードウェアによる実験基板の開発を急遽行い、残りの研究期間が少ない中で成果の巻き返しができたことは自ら評価できる。上記のような研究遂行上の避けがたい環境の変化があったものの、研究の狙いにあるストリームデータ圧縮技術のエントロピー符号化とユニバーサル性の融合した新たなデータ圧縮技法を実現できたことで本研究の狙いを完遂できたことは今回のさきがけ研究は成功したと考えている。さらに、コロナ禍後に欧州ショートビジットを通じてさきがけ研究者間での交流ができ、また、さきがけ研究の成果展開の可能性を試すため、ニュージーランド、ポルトガル、スウェーデンの研究者へのショートビジットを実施した。それぞれの研究者との交流は短期間ではあるが、コロナ禍前の互いの親交を思いだし、今後も続く長い交流を約束できた。このような人的交流を通じ、今後の研究活動の幅を国外にも広げ、世界トップクラスの成果を得るための活動基盤を再起動できた。このような努力を通して自らの研究力をレベルアップできたことは大きな成果であると自己評価する。

5. 主な研究成果リスト

(1) 代表的な論文(原著論文)発表

研究期間累積件数: 7 件

1. Shinichi Yamagiwa, Yuma Ichinomiya. Stream-Based Visually Lossless Data Compression Applying Variable Bit-Length ADPCM Encoding, Sensors, 21(13), 4602, (2021-07-05), DOI:10.3390/s21134602
ADPCMを動的にビット数変更が可能な Variable Bitrate ADPCMを開発し、映像伝送による有効性を示した。
2. Danhe Tian, Shinichi Yamagiwa, Koichi Wada. Heuristic Compression Method for CNN Model Applying Quantization to a Combination of Structured and Unstructured Pruning Techniques, IEEE Access (volume 12), pp. 66680 - 66689, (2024-05-09), DOI:10.1109/ACCESS.2024.3399541
畳み込み深層学習におけるネットワークモデル圧縮技法をヒューリスティックに組合せることで

ネットワークモデルを最小化する手法を開発した。Structured/unstructured pruning、そして量子化技法の3つの異なる特性を持つモデル圧縮技術を使って、それぞれの適用率を制御することによって、現実的な処理時間で最小化できる手法を開発した。

3. Shinichi Yamagiwa, Taiki Kato, Universal Adaptive Stream-Based Entropy Coding, IEEE Access (volume 12), pp. 98768 - 98786, (2024-07-16), DOI:10.1109/ACCESS.2024.3429389

エントロピー符号化とユニバーサル性を融合することにより、複数のデータシーケンスを一度のスキャンによって最少で1ビットにまで圧縮可能なストリームデータ圧縮方式を開発した。

(2) 特許出願

研究期間全出願件数：2件(特許公開前のものは件数にのみ含む)

(3) その他の成果(主要な学会発表、受賞、著作物、プレスリリース等)

国際会議論文・学会発表

1. Taiki Kato, Shinichi Yamagiwa, Koichi Marumo. Performance Enhancement of Stream-Based Decompression Process by Notifying Compression Buffer Size, 2023 IEEE Symposium on Computers and Communications (ISCC), 491-494, (2023-08-28), DOI:10.1109/iscc58397.2023.10218261

2. Taiki Kato, Shinichi Yamagiwa, Koichi Marumo. Performance Enhancement of Stream-Based Decompression Process by Notifying Compression Buffer Size, 2023 IEEE Symposium on Computers and Communications (ISCC), 491-494, (2023-08-28), DOI:10.1109/iscc58397.2023.10218261

著書

Sublinear Computation Paradigm: Algorithmic Revolution in the Big Data Era (Chapter: Stream-Based Lossless Data Compression), Springer Singapore, October 2021. DOI: 10.1007/978-981-16-4095-7

プレスリリース

1. 科学技術振興機構・共同発表 流れるデータをリアルタイムに圧縮・解凍する新技術を開発・2024年7月19日

<https://www.jst.go.jp/pr/announce/20240719/index.html>

2. 筑波大学・流れるデータをリアルタイムに圧縮・解凍する新技術を開発・2024年7月19日

<https://www.tsukuba.ac.jp/journal/technology-materials/20240719141500.html>

3. 筑波大学・画像を認識する人工知能をコンパクトに実装できる新手法を開発・2024年5月22日

<https://www.tsukuba.ac.jp/journal/technology-materials/20240522140000.html>

4. 筑波大学・柔道の投技の成功要因を直前の体勢から発見する手法を開発 ～スポーツ技能上達のコツをビッグデータから探る～・2021年9月1日

<https://www.tsukuba.ac.jp/journal/technology-materials/20210901160000.html>

報道

組込みシステム技術協会機関誌 **Bulletin JASA**「[大学訪問] 筑波大学 情報システム系 “エッジ”と言われる前からテーマとして追究 課題を先取り最前線で研究開発に取り組む」 2024 年 1 月号 (Vol.088,2024)

学会講演

第 31 回日本運動生理学会大会 (筑波大学 つくばキャンパス) キーノートレクチャー「スポーツビッグデータからの知識発見」・2023 年 8 月 24 日