

# Computation Reuse と Approximate Computing

COMPUTATION REUSE AND APPROXIMATE COMPUTING

津邑 公暁<sup>1</sup>

Tomoaki TSUMURA

名古屋工業大学<sup>1</sup>

Nagoya Institute of Technology

## 1 はじめに

情報システムで扱うべきデジタルデータは、画像や音声など旧来のメディアデータはもちろんのこと、RFIDやセンサにより収集されるデータ、ソーシャルメディアを介して発信されるデータなどが加わり、爆発的な増大を見せている。今後も、いわゆる「超可観測」が進行することで、扱うべきデータ量はさらに超大規模化していくと考えられる。これに伴い、計算性能の向上はもちろんのこと、その計算形態自体の再考が迫られている。

特にこのような大規模データを対象とする代表的アプリケーションである、信号処理やマイニング、機械学習などにおいては、個々の計算結果の高信頼性と高精度を必ずしも強く要求しない場合が多く、必要最低限の精度で、いかにより早く結果を出力するか、あるいはより多くの演算を行うかが重要となる。

このような背景から、出力の精度に重大な影響を及ぼさない範囲で誤差を許容しつつ、適切に近似計算を行い処理を高速化・省電力化する、Approximate Computing というパラダイムが期待されている [1]。本講演では、主に Computation Reuse (計算再利用) が持つ Approximate Computing との親和性、およびこれを活用した既存研究について紹介した上で、著者らが進める Approximate Computing 基盤の構想について紹介する。

## 2 Computation Reuse の活用

著者はこれまで、計算再利用の概念に基づくプロセッサ高速化手法を研究してきた。計算再利用とは、関数やループなどの処理ブロックに対し、計算に用いた入力と出力をペアで記憶しておき、再実行時、過去と同じ入力であった場合は、記憶された過去の出力 (計算結果) を再利用することで、当該処理ブロックの計算自体を省略する考え方である。

この計算再利用において、記憶された過去の入力と現在の入力との一致比較に曖昧性を持たせる、すなわち、厳密に一致しなくとも近い値であれば一致したとみなすことによって、近い入力に対する出力を再利用し、処理量を削減できることに、著者らは非常に早くから着目していた [2]。近年、この計算再利用と Approximate Computing の親和性に着目した研究が数多くなされており、著者らと同様に画像処理アプリケーションに対して応用した DSA の研究 [3] や、CNN 計算のエネルギー効率改善に応用した研究 [4] など、さまざまな試みがなされている。

## 3 Approximate Computing 基盤の実現に向けて

Approximate Computing は省電力と高性能を両立するパラダイムとして期待を集めているものの、これをアプリケーションに適用する統一的な計算基盤は今のところ存在していない。そもそも、各計算の近似化が最終的な出力に及ぼす誤差は強くアプリケーションに依存し、また近似化の方法自体もさまざまである。

これに対し著者らは、Computation Reuse の際の入力一致における曖昧性の調節により、出力誤差も調節可能であるという発想に基づき、プログラマが出力誤差の許容度を指定するだけで、動的に計算近似度を調整し、指定された誤差の範囲内で最大のパフォーマンスを引き出す、可用性の高い Approximate Computing 基盤の実現を目指した研究を行っている。

言語仕様とプロセッサ拡張とから成るこの計算基盤について、既にプロトタイプを設計し、シミュレーションによる評価を通じて、いくつかのドメインのアプリケーションにおいて、このアイデアが効果的に適用可能であることを確認しており [5, 6]、今後さらに機能拡張および詳細な評価を行っていく予定である。

謝辞 本成果の一部は、JSPS 科研費 JP17K19971 の助成による。

### 参考文献

- [1] R. Nair, “Technical perspective: Big data needs approximate computing,” *Communication of the ACM*, vol.58, no.1, p.104, 2015.
- [2] 津邑公暁, 中島康彦他, “ステレオ画像処理を用いた曖昧再利用の評価,” *情報処理学会論文誌: コンピューティングシステム*, vol.44, no.SIG 11(ACS 3), pp.246–356, Aug. 2003.
- [3] H. Osawa and Y. Hara-Azumi, “Approximate Data Reuse-based Processor: A Case Study on Image Compression,” *Proc. 15th IEEE/ACM Symp. on Embedded Systems for Real-Time Multimedia (ESTIMedia’17)*, pp.32–40, Oct. 2017.
- [4] X. Jiao, et al., “Energy-efficient neural networks using approximate computation reuse,” *Proc. Design, Automation & Test in Europe Conference & Exhibition (DATE’18)*, pp.1223–1228, March 2018.
- [5] Y. Sato, T. Tsumura, and Y. Nakashima, “An Approximate Computing Stack based on Computation Reuse,” *Proc. 3rd Int’l Workshop on Computer Systems and Architectures (CSA’15)*, pp.378–384, Dec. 2015.
- [6] 井内悠太, 津邑公暁, “カーネルテンプレート化と計算再利用による CNN の計算量削減に関する検討,” *情処研報 (SWoPP2018)*, pp.1–7, July 2018.