

省電力なAIモデルの実現に向けて

<p>どんな研究</p>	<p>AIの高性能に伴う計算量や消費電力の増大に対し、疎なネットワーク構造の活用や量子化などの効率化手法が提案されてきました。本研究では学習方法を工夫することで、省電力アナログハードウェアでも柔軟に活用できるニューラルネットワーク設計の新しい考え方を紹介します。</p>
<p>どこが凄い</p>	<p>従来の効率化手法は疎な構造と量子化の併用を前提とする場合が多く、疎な計算を効率的に扱えるデジタルハードウェアなどに適用が限られることがありました。本研究では疎性を必須とせず、量子化と両立可能な学習枠組みを示し、多様な実装技術への展開可能性を広げました。</p>
<p>めざす未来</p>	<p>学習方法の工夫によって、特定の計算方式や実装技術に依存しないAI設計を可能にします。これにより、将来は省電力で再構成可能なAIが、デジタルからアナログまで幅広い計算基盤やデバイス上で活用される社会をめざします。</p>

ニューラルネット (NN) の効率化

- NNは大量の計算とメモリを必要とし、消費電力が大きい
→ **枝刈り (疎性)** と **量子化** は効率化に不可欠

高精度・密なNN

全ての結合と細かい重み
→ 計算・メモリコストが膨大
例: 32ビットx9重み

枝刈り

不要な結合を削除し疎性を導入
→ メモリ・演算コストを削減
例: 32ビットx5重み

量子化

重みの値を制約しビット数を削減
→ メモリ・演算コストを削減
例: 3ビットx9重み

枝刈り + 量子化

構造と数値を同時に簡素化
→ 効率を大幅向上
例: 3ビットx5重み

本研究：再利用可能なアナログAI

- デジタルとアナログでは、**学習すべきものが違う**：
デジタルは「結合」、アナログは「符号と大きさ」を学習

デジタルデバイス

重みの値：任意 (固定)
結合構造：学習
AND
既定 (0ビット) → 例: 1ビットx9重み

高効率な疎なNN
正確なゼロで表現される疎性

アナログデバイス

重みの値：学習
結合構造：任意 (固定)
AND
既定 (0ビット)

高効率な疎なNN
物理的な接続で実装される疎性

学習・固定対象の反転

提案手法：固定された結合構造の上で、重みの「符号」と「大きさ」を学習

- ★ 構造制約下で量子化込みの学習 (QAT)
- ✓ 既存手法の利点をすべて保持
- ✓ NNを更新しても構造は不変 → 実用的
- ✓ 結合構造を学習しないので、トポロジー最適化が可能となる → 効率向上

アナログデバイスにおける疎性の壁

- 枝刈りはデジタルデバイスの正確なゼロを前提とするが、アナログデバイス (例: 光) ではノイズにより成立しない

デジタルデバイス

不要な結合が正確なゼロで表現
→ 正しく計算

アナログデバイス

不要な結合がノイズにより非ゼロ化
→ 計算が崩れる

- 既存手法：製造時に枝刈りを物理的に反映する
- ✓ トポロジーから接続を削除して、ゼロを完全に実現
- ✓ ゼロの保存が不要 → 圧縮率向上
- ✓ ハードウェア削減 = 面積、ノイズ、消費電力削減
- ✗ 代償としてデバイスの再利用性を失う → 実用性が低い

学習 → NNを更新すると、結合構造が変わる → 正しいゼロ → 再学習や調整など → 再製造が必要

物理的枝刈り → 結合構造が変わると、デバイスと不一致 → 再利用不可

- デバイスの結合構造を物理的に枝刈りした後、符号や大きさだけを学習することで、**再製造なしに再利用が可能**

学習 → NN結合に反映 → 物理的枝刈り → 再利用不可

学習 → 重みの値のみ学習 → NN結合に反映 → 書込み (プログラム) → 再利用可能

アナログ実装 値が更新されても、結合構造はデバイスと一致

関連文献

[1] Á. López García-Arias, Y. Okoshi, H. Otsuka, D. Chijiwa, Y. Fujiwara, S. Takeuchi, M. Motomura, "The Trichromatic Strong Lottery Ticket Hypothesis: Neural Compression With Three Primary Supermasks," *Workshop on Machine Learning and Compression, Conference on Neural Information Processing Systems (NeurIPS)*, 2024.

[2] Á. López García-Arias, Y. Okoshi, H. Otsuka, D. Chijiwa, Y. Fujiwara, S. Takeuchi, M. Motomura, "The Trichromatic Strong Lottery Ticket Hypothesis: A Unifying View of Supermask-Based Learning," *IEEE International Joint Conference on Neural Networks (IJCNN)*, 2026.

連絡先

ロペス アンヘル (Angel Lopez) メディア情報研究部 メディア認識研究グループ

共同研究先・外部資金

本展示の成果は東京科学大学との共同研究によるものです。