

論文 / 著書情報

Article / Book Information

題目(和文)	高効率・高精度なエッジAIアンサンブル手法の研究
Title(English)	A Study on Efficient and Accurate Ensemble Methods for Edge AI
著者(和文)	熊澤峻悟
Author(English)	Shungo Kumazawa
出典(和文)	学位:博士(学術), 学位授与機関:東京科学大学, 報告番号:甲第382号, 授与年月日:2025年3月26日, 学位の種別:課程博士, 審査員:本村 真人,高橋 篤司,CHU VAN THIEM,ISLAM A K M MAHFUZUL,佐々木 広
Citation(English)	Degree:Doctor (Academic), Conferring organization: Institute of Science Tokyo, Report number:甲第382号, Conferred date:2025/3/26, Degree Type:Course doctor, Examiner:,,,,
学位種別(和文)	博士論文
Category(English)	Doctoral Thesis
種別(和文)	審査の要旨
Type(English)	Exam Summary

論文審査の要旨及び審査員

報告番号	甲第		号	学位申請者氏名		熊澤 峻悟	
論文審査 審査員		氏 名		職 名		氏 名	職 名
	主査	本村 真人		教授	審査員	佐々木 広	准教授
	審査員	高橋 篤司		教授			
		Chu Van Thiem		准教授			
		ISLAM Mahfuzul		准教授			

論文審査の要旨（2000 字程度）

本論文は、A Study on Efficient and Accurate Ensemble Methods for Edge AI（高効率・高精度なエッジ AI アンサンブル手法の研究）と題し、英文 5 章から構成されている。 第 1 章 Introduction（序論）では、本研究で取り組む問題の紹介が行われている。近年、IoT デバイスの急速な普及により、応答性やエネルギー効率の向上をもたらすエッジ AI が注目されている。その中でも、増加するデバイス群を効率的に活用することで、単一の端末の性能向上の限界に対処可能な協調推論手法の重要性が高まっている。その有望な解決策であるエッジアンサンブルは、周囲の使用可能なデバイスと動的に協調グループを形成し、複数のエッジデバイスの予測結果を統合することで、高精度な推論を実現するスケーラブルな協調推論手法である。しかし、連携するデバイスの増加に伴う通信負荷やレイテンシ、使用可能デバイスの変動といった問題が存在し、計算資源や電力に制約があり、即応性が要求されるエッジ環境における推論システムへの適用には課題が残る。この論文では、限られたデバイス数において高精度な推論を可能にするためのアンサンブル手法基盤の高精度化・高効率化手法を提案する。 第 2 章 Background（研究背景）では、この論文の基盤となる基本的な研究について紹介されている。具体的には、協調的なエッジ AI の手法が論じられ、特にエッジアンサンブルの利点とその重要性が他の手法と比較して説明されている。また、アンサンブル手法の理論や既存研究を解説し、エッジアンサンブルが直面する効率性に関する課題として軽量の学習・推論手法と効率的なモデル統合手法の必要性が挙げられている。 第 3 章 Extra Ferns: An Edge-oriented Fern Ensemble Learning Method with Minimal Memory Access（Extra Ferns: 低メモリアクセス数でエッジ向けな Fern アンサンブル学習手法）では、軽量のエッジ環境向けアンサンブル学習手法である「Extra Ferns」が提案されている。Extra Ferns は、決定木において同じ深さを共有した Fern を基盤モデルとしたアンサンブル手法で、限られた計算資源しか持たないエッジデバイス向けに設計されたもので、エッジアンサンブルを支える学習技術の 1 つとして提案されている。この手法は既存の決定木アンサンブルの強みを活かしつつ、その課題を克服する独自の方法を採用している。具体的には、Extra Trees と同様のランダムなノード生成法と、Random Ferns と同様の Fern 構造を用いることで、ランダムな少数ノードによるメモリアクセス数や計算量を削減する。さらに、推論精度改善のためノードの閾値の最適化といった特徴や、従来の Fern アンサンブルで問題となった葉ノードのメモリ使用量の削減といった特徴をもつ。実験結果によると、Extra Ferns は従来の Fern アンサンブルと比較して最大で 24.9%と、大幅に推論精度を向上させた。また従来の決定木アンサンブルと比較して、モデル学習時のメモリアクセス数を最大で 96.5%削減しつつ、推論精度低下を 5%以内に抑えている。また、画像データセットにおいてランダム射影を前処理として適用することで、さらに最大 2.0%の精度向上を報告している。 第 4 章 Improving Collaborative Effectiveness in Ensemble-based Inference at the Edge（エッジにおけるアンサンブル推論の協調効果の向上）では、深層学習を基盤としたエッジアンサンブルにおけるモデル統合の効率性と精度を向上させる手法について取り組んでいる。本章では、エッジアンサンブルの協調効率を向上させるため、従来のモデル統合手法であるカスケード、重み付き平均、および Test Time Augmentation (TTA) の有効性を評価し、それらをエッジアンサンブル向けに最適化した新たな手法を提案している。カスケード手法は推論に必要なモデル数を削減できる一方で、逐次的な処理が原因でレイテンシが増大する問題があった。推論処理のレイテンシの増大は即応性を損ない、即時の意思決定が求められるアプリケーションへの適用を困難にする。これに対し、本章では同時に処理するモデル数を調整することで、レイテンシ問題を解決している。また、TTA と重み付き平均については、学習データの正解ラベルの代わりにアンサンブル予測ラベルを利
--

用してこれらのパラメータを学習する新しい手法を提案し、エッジ側で学習データを使用せずにアンサンブルの協調効果の向上を実現している。

最後に第 5 章 Conclusion (結論)では、この論文がエッジ環境における協調推論システムの進展に向けた新しい手法を提案し、それを検証した成果を総括する。今後、IoT デバイスの増加がさらに加速する中で、協調推論手法の重要性はますます高まり、その実現が強く期待される。本研究はその基盤となるアンサンブル手法を用いた協調推論において、第 1 章で述べた限られたデバイス数において高精度な推論を可能にするための効率性の課題を解決するために 2 つの新たな提案を行った。これらの効率性向上に関する提案と分析は、単にエッジアンサンブルのパフォーマンスを向上させるだけでなく、モデルの学習手法や協調形態に関する議論を含んでおり、実際に協調推論システムを設計する際の指針の一つとなる。

以上を要するに、本論文はエッジ環境における協調推論システムを対象に、軽量のアンサンブル手法である Extra Ferns や、協調効果を向上させるアンサンブル技術を提案・検証し、リソースが限られた環境において高精度・高効率な推論処理を可能にするための新たな知見を提示しており学術的貢献は大きい。よって、審査員は本論文が博士（学術）の学位論文として十分に価値があるものと認める。

注意：「論文審査の要旨及び審査員」は、東京科学大学リサーチリポジトリ(T2R2)にてインターネット公表されますので、公表可能な範囲の内容で作成してください。