

学習モデルの段階化による 分散機械学習の効率的実現手法

芝浦工業大学

システム理工学部 電子情報システム学科

マルチメディア情報通信研究室

教授 三好 匠

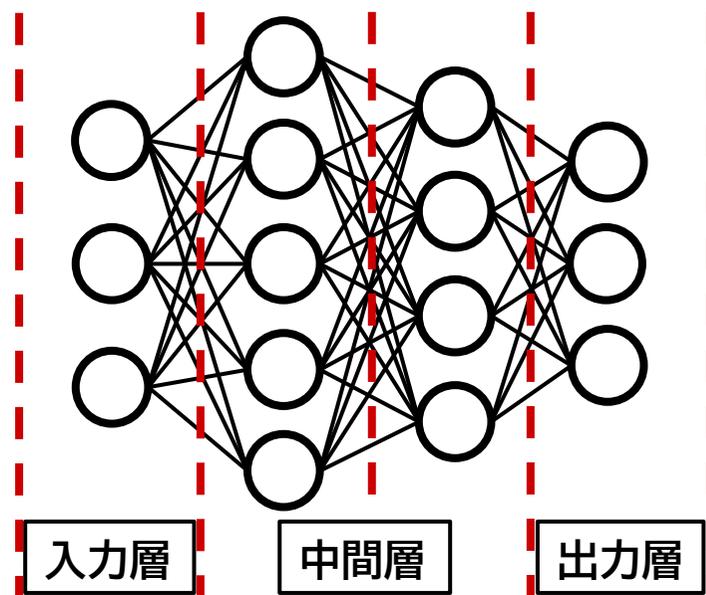
2024年9月19日

研究の背景

- 近年、スマートシティ実現への期待が高まる
 - Society 5.0、サイバーフィジカルシステム、デジタルツインなど
 - 政府・自治体を挙げた研究開発が進展
- 多種多様なセンサを用いて、まちから収集したビッグデータを分析し、機械学習により未来予測を行うシステムの開発に注目
- まちサービスにおける機械学習の具体例
 - 画像認識、画像識別
 - 自動運転、渋滞予測、人流予測、天候予測など

機械学習(深層学習)とは

- 入力データと出力データのパターンを発見・分析し、それに基づいて予測を行う仕組み
- ニューラルネットワークや深層ニューラルネットワークと呼ばれる数理モデルにより実現
 - ビッグデータ(過去データ)の入力と出力が整合するように「学習」を行い、パラメータを調整(学習フェーズ)
 - 学修済みモデルを用いて、入力データに対して出力を算出(運用フェーズ)
 - 学習フェーズに膨大な計算時間が必要

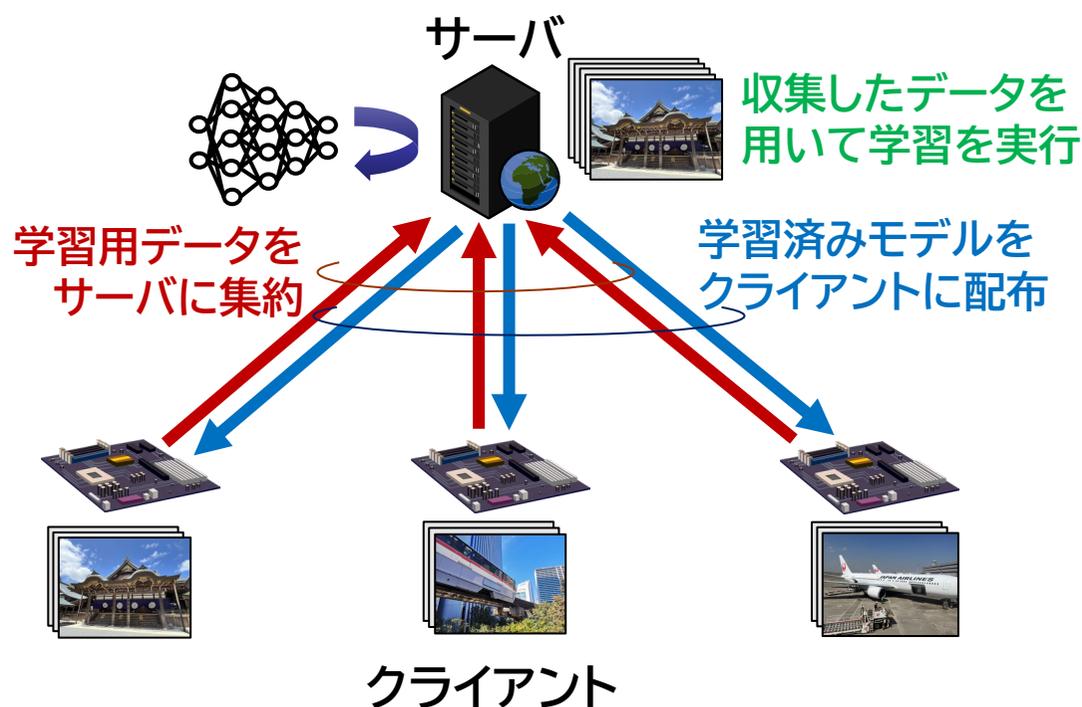


学習モデル(以下、モデル)

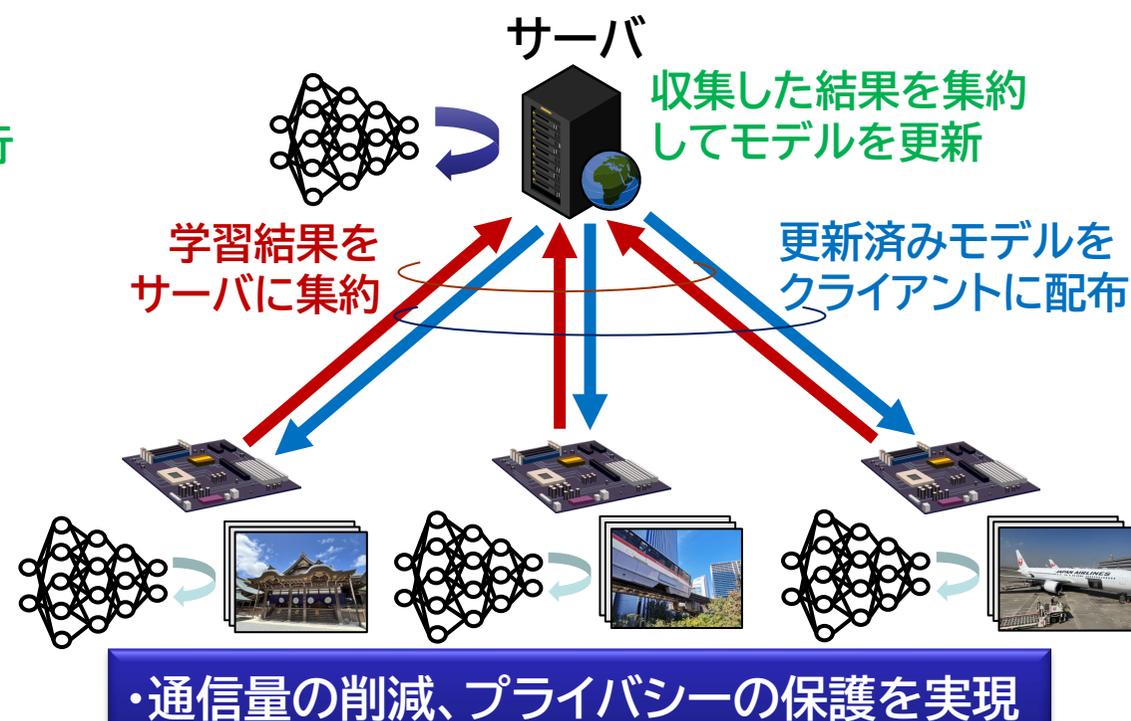
分散型機械学習(連合学習)

- 機械学習の学習フェーズに必要な計算を複数のコンピュータで分散協調的に行う手法

一般的な機械学習



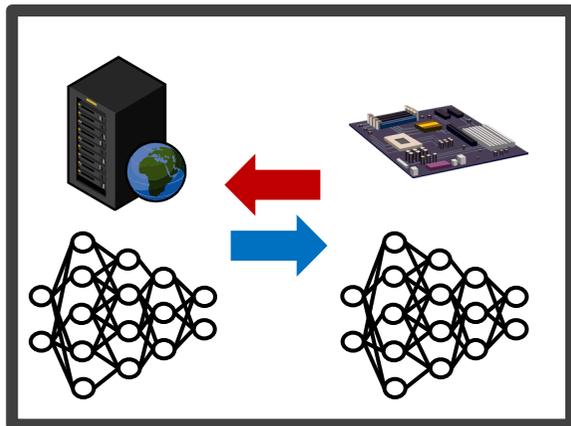
連合学習



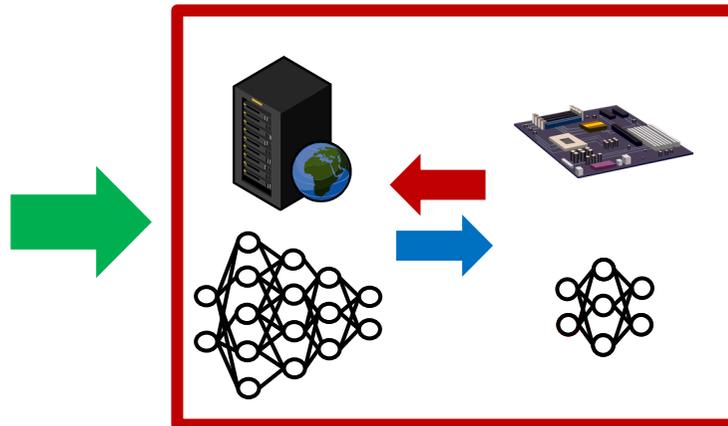
従来技術：小型モデルを使用した連合学習

- 連合学習において大規模なモデルを使用する場合には、クライアントの計算負荷が増大
- 「蒸留」と呼ばれる技術を用いて **モデルを小型化** することで、**クライアントの計算負荷、通信量を削減**
 - サーバで再学習が必要となるが、それを差し引いても高速化が可能

一般的な連合学習



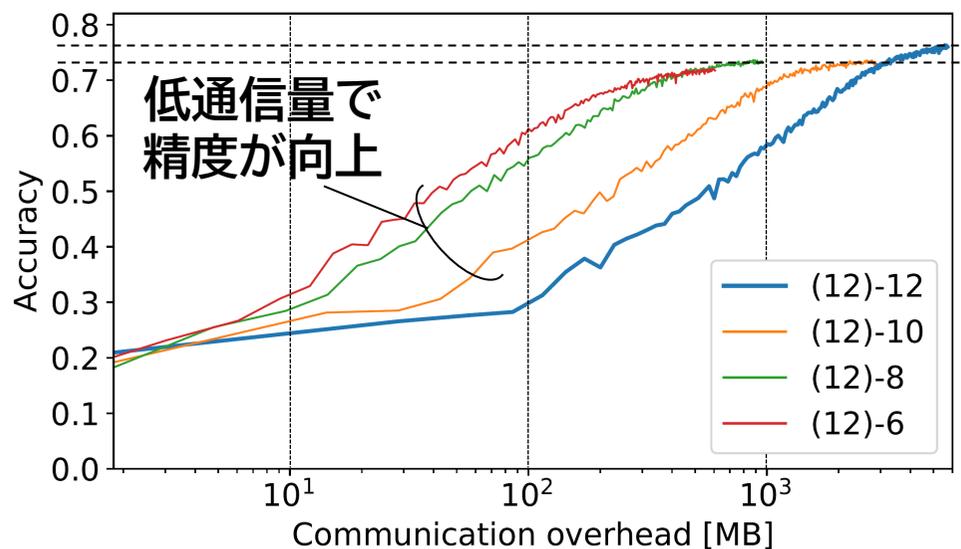
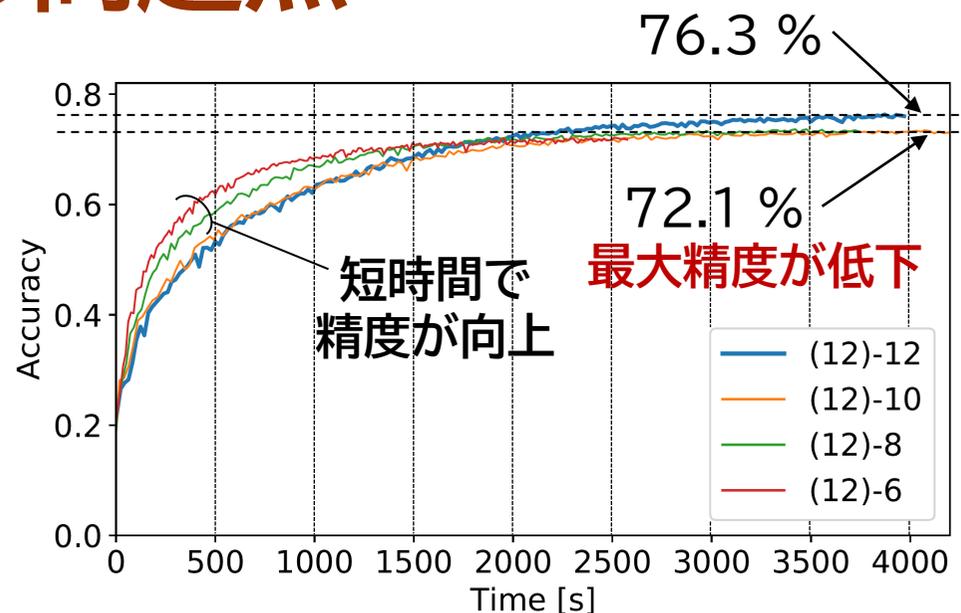
小型モデルを用いた連合学習



	学習モデル	
サイズ	大 > 小	
学習精度	高	低
計算負荷	高	低
通信量	高	低

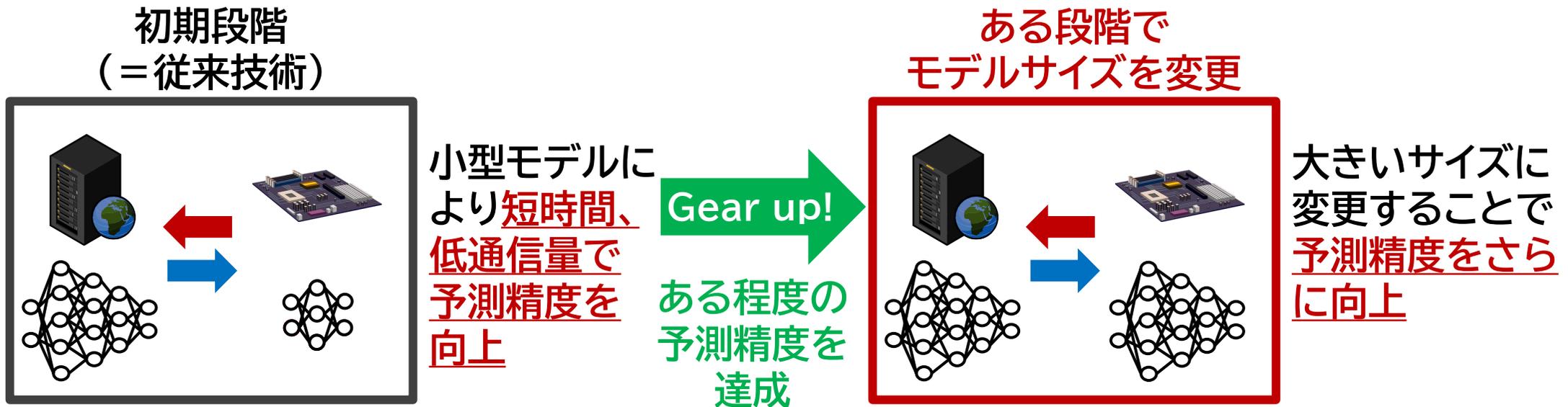
従来技術の問題点

- 小型モデルを用いることで計算負荷が低下し、**短い時間、少ない通信量でモデルの予測精度が向上する**
- 小型モデルの基本的特性として、**最終的に到達できる予測精度に限界**がある



新技術：学習モデルを段階化した連合学習

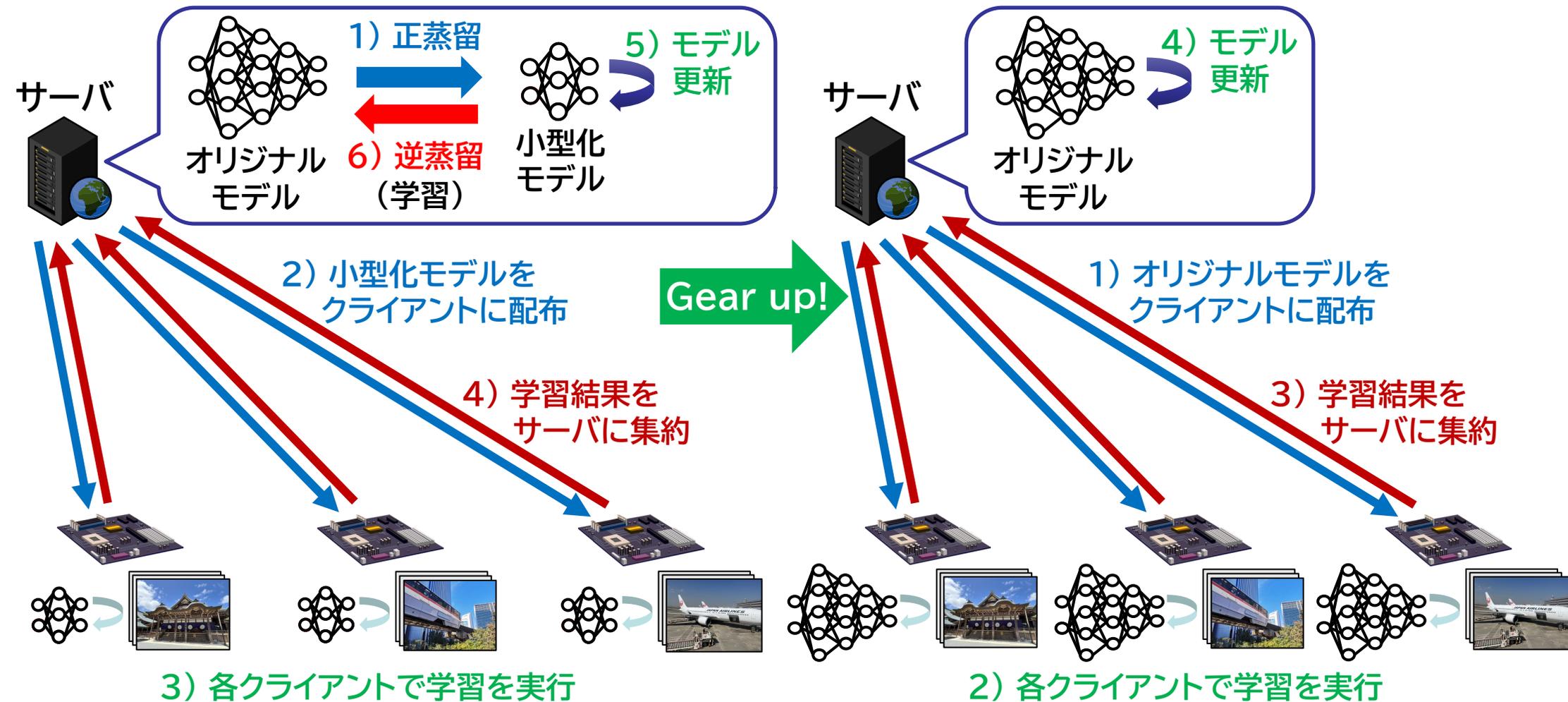
- モデルサイズを段階的に変更する技術を開発
 - 最初は従来技術と同じ小型モデルで学習を開始
 - ある段階でモデルサイズを大きいものに変更
- 最終的に到達できる予測精度の向上を実現**



新技術の動作手順

初期段階(従来技術)

サイズ変更後(通常の連合学習)

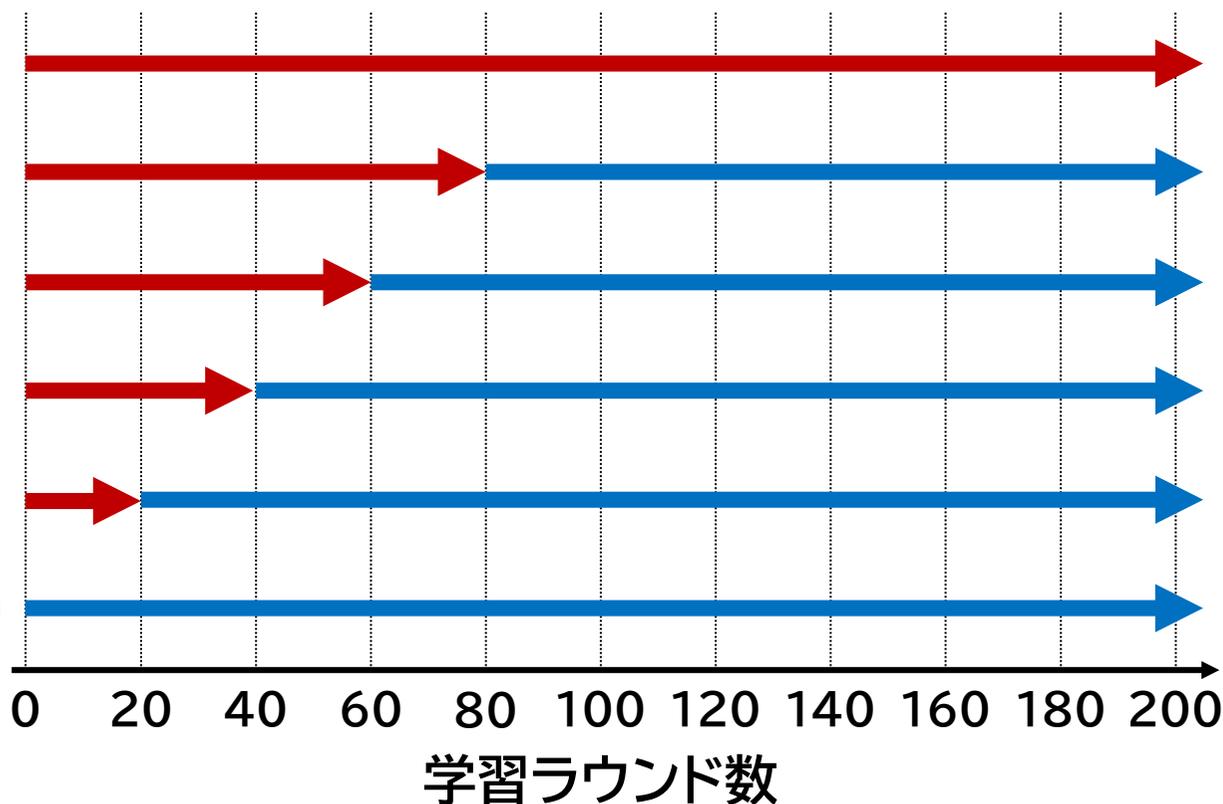


サイズ段階化の例

- サイズ段階化のパターン例と比較対象

- 12-6 (従来技術)
- Gear 80
- Gear 60
- Gear 40
- Gear 20
- 12-12 (通常連合学習)

新技術の
領域



→ オリジナルモデル: 12層
小型化モデル: 6層

→ オリジナルモデル: 12層

新技術の評価環境

- サーバ1台、クライアント3台による連合学習
- 機械学習でよく用いられる画像データセットCIFAR-10
 - **学習用データ: 50,000枚**
 - サーバ: 20,000枚(蒸留処理にのみ使用)
 - クライアント: 10,000枚 x 3
 - **評価用データ: 10,000枚**
 - サーバ、クライアントに同一の評価用データを使用
 - 学習用データ50,000枚を等分してクライアントに持たせ、通常の連合学習を行う**C50000をリファレンスモデルとして使用**
- 画像認識タスクを200ラウンド実行

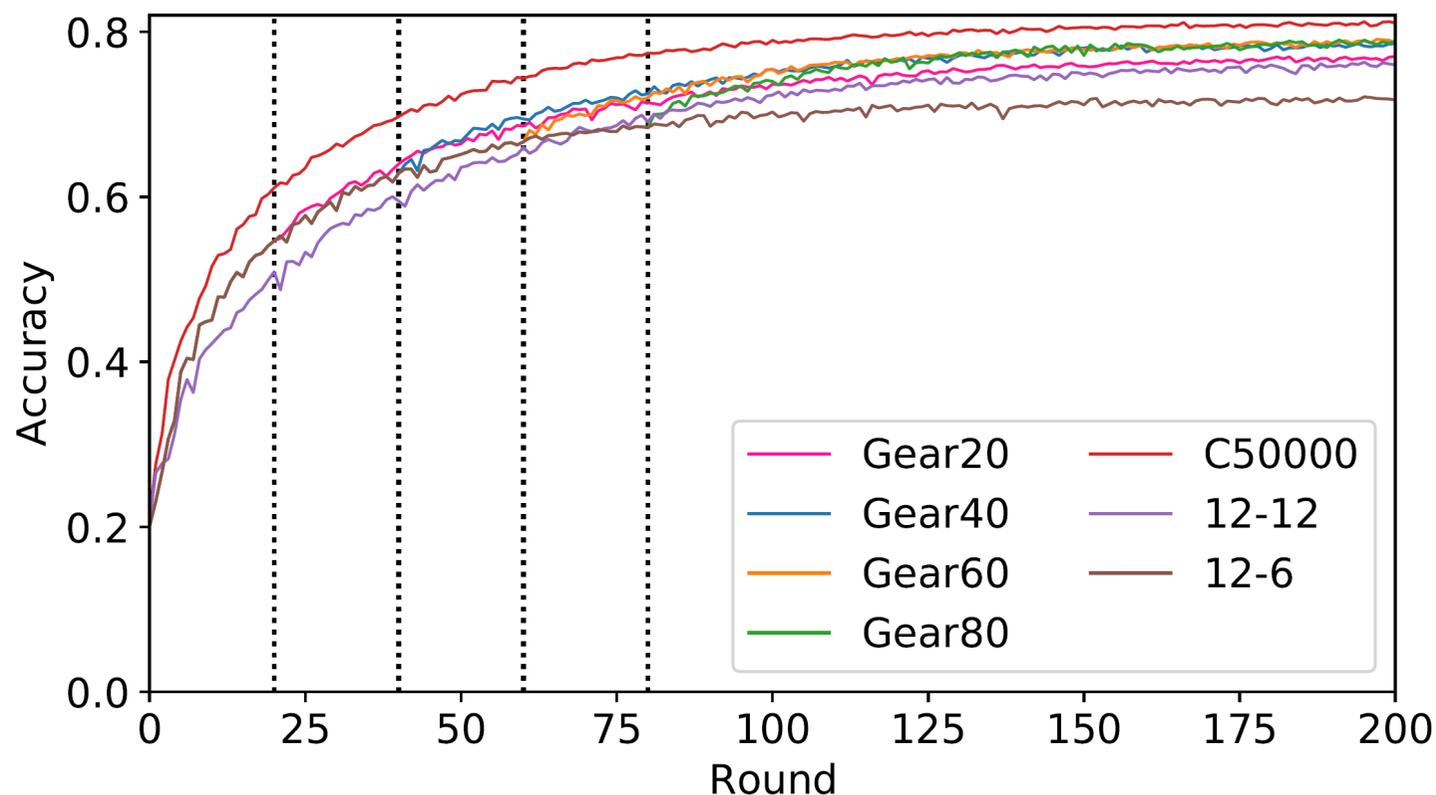
性能評価(まとめ)

	通常の連合学習		従来技術	新技術			
	12-12	12-12 C50000		12-6	Gear 20	Gear 40	Gear 60
学習モデルサイズ	12-12	12-12 C50000	12-6	Gear 20	Gear 40	Gear 60	Gear 80
1ラウンドの通信量 [MB]	28.63	28.63	3.03	3.03 ⇒ 28.63			
1ラウンドの時間[秒]	19.8	31.6	12.9	—	—	—	—
最大精度[%]	76.3	81.2	72.1	77.0	78.6	79.1	79.0
60%到達時間[秒]*	360	288	166	166	166	166	166
70%到達時間[秒]*	642	514	334	340	334	334	334
80%到達時間[秒]*	1190	895	644	846	601	644	644
90%到達時間[秒]*	2,238	1,759	—	2,174	1,322	1,242	1,315

* C50000の最大精度81.2%のx%に到達するまでの時間を計測。

性能評価(ラウンド数と最大精度)

- 新技術は、途中からオリジナルモデルに移行することにより最終的に到達する予測精度が向上することを確認



最終的に到達する予測精度

通常の連合学習

12-12 : 76.3%

C50000 : 81.2%

従来技術

12-6 : 72.1%

新技術

Gear 20 : 77.0%

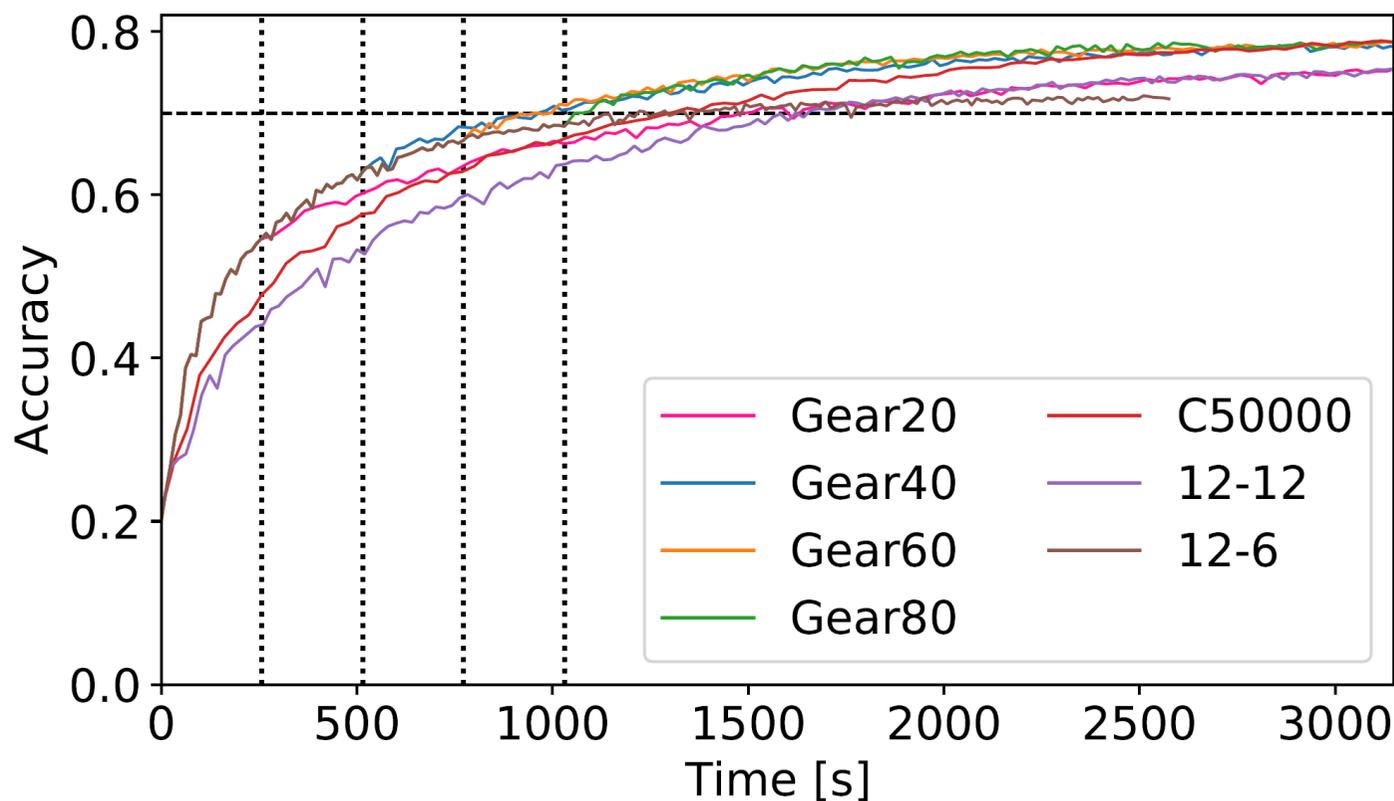
Gear 40 : 78.6%

Gear 60 : 79.1%

Gear 80 : 79.0%

性能評価(計算時間)

- 新技術は、従来技術や通常の連合学習よりも短時間で予測精度が向上することを確認



精度が70%に到達した時間[秒]

通常の連合学習

12-12 : 1,662

C50000 : 1,313

従来技術

12-6 : 1,222

新技術

Gear 20 : 1,510

Gear 40 : 965

Gear 60 : 957

Gear 80 : 1,057

性能評価(通信量)

- 新技術は、従来技術や通常の連合学習よりも少ない通信量で予測精度が収束することを確認

500MB通信時の予測精度

通常の連合学習

12-12 : 48.1%

C50000 : 57.8%

従来技術

12-6 : 71.3%

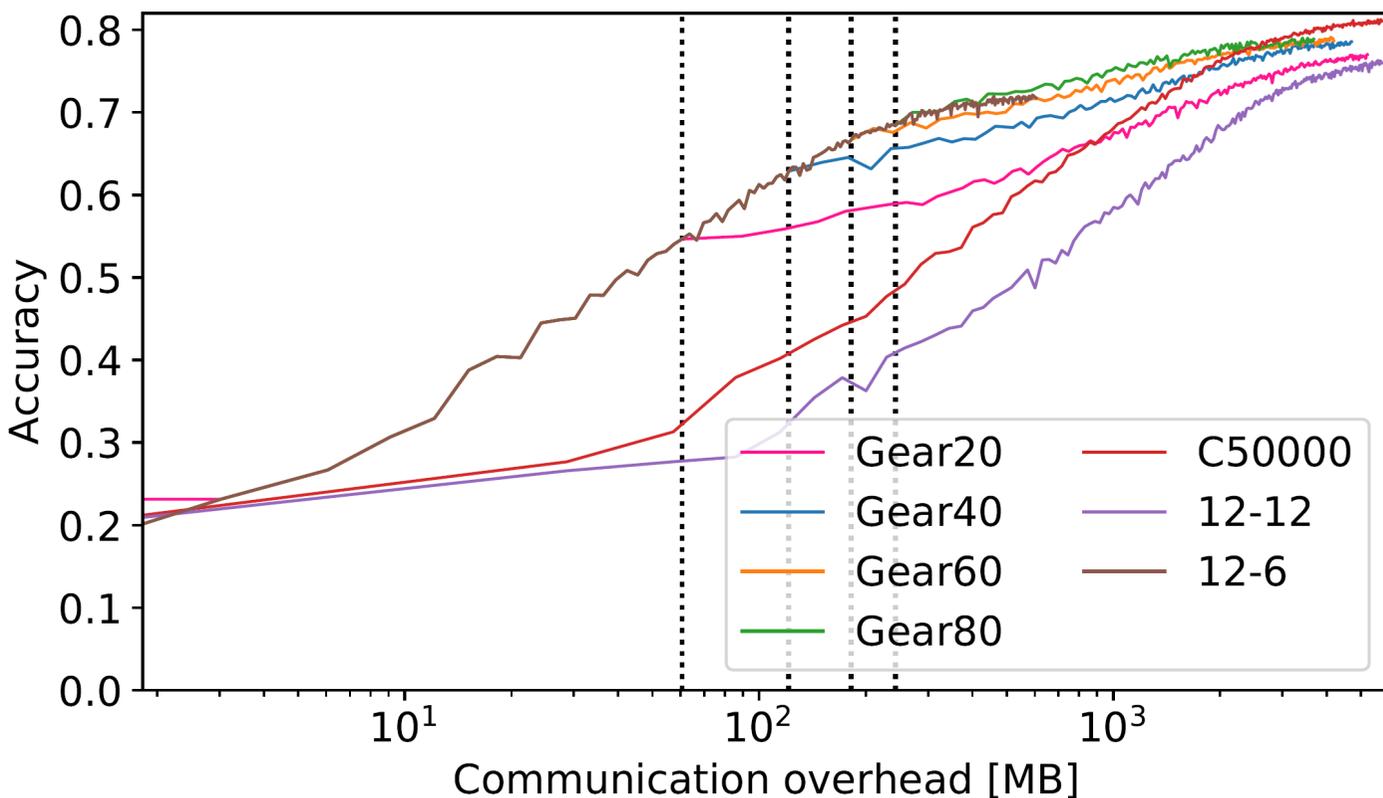
新技術

Gear 20 : 52.1%

Gear 40 : 68.2%

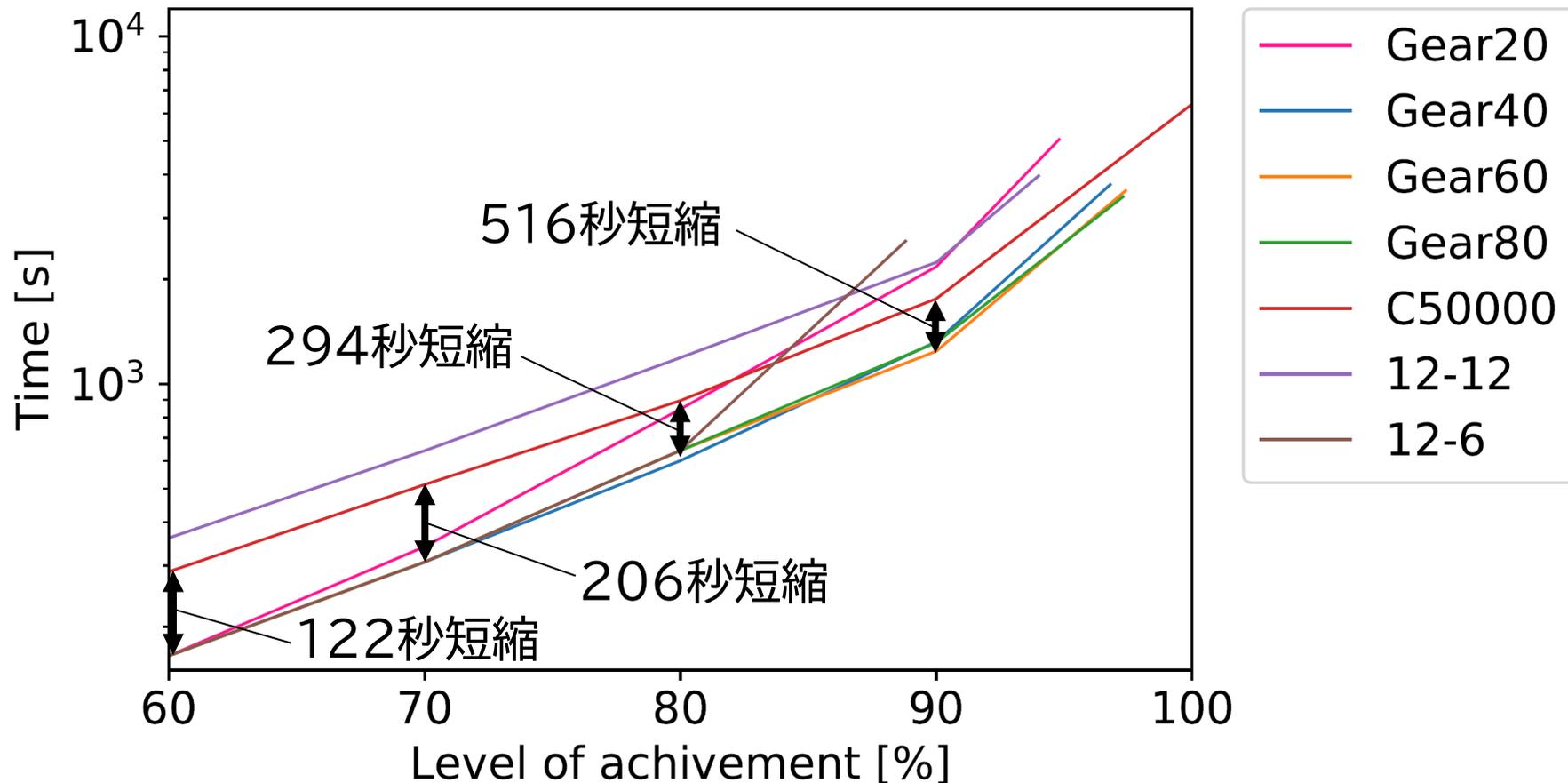
Gear 60 : 71.6%

Gear 80 : 72.2%



性能評価(計算時間と精度の関係)

- 新技術はリファレンスモデルと比較して、同じ予測精度に到達するまでの計算時間が短縮されることを確認



新技術の特徴と従来技術との比較

- 従来技術では、小型モデルを用いることで、短い時間、少ない通信量で連合学習モデルの予測精度を向上できるが、最終的に到達できる予想精度に限界があった
- 新技術では、学習フェーズの途中でモデルのサイズを変更することにより、計算時間と通信量を抑えながら最終的に到達できる予測精度を向上させることに成功した
- 複雑な機械学習を行うような場合には、計算時間と通信量の削減効果はより顕著に現れると考えられ、新技術の適用により計算資源や通信資源、消費電力の削減に対して大きな効果が得られるものと推測される

想定される用途

- 機械学習の分野においては、その効率を高めるために蒸留を用いてモデル圧縮を行う手法が知られているが、学習フェーズにおいて小型化したモデルを段階的に利用する手法は見当たらず、**分野を問わず利用可能であり、極めて汎用性の高い技術**であると考えられる
- 特に、**継続的に学習を行わなければならない環境**、例えばインターネットにおける**セキュリティ対策**や**侵入検知システム**、まちでの**人流動態予測**、**交通事故防止システム**などでの利用が想定される

実用化に向けた課題

- 学習モデルの段階化手法の一例として画像認識システムの実装を行い評価を行ったが、用途に応じて使用される学習モデルが異なるため、用途に合わせた学習モデルの検討が必要である
- 今回は、モデルサイズの変更を行うタイミングを数種類実施したが、サイズ変更の自動化が課題である

企業様への貢献、及び共同研究のお誘い

- 本技術は、連合学習を用いるさまざまな分野において利用可能であり、極めて汎用性の高い技術である
- スマートシティの実現に向けた新たなまちサービスの開発に貢献できるものと考えます
- 機械学習、特に継続的に学習を行う必要のある環境での機械学習導入企業様との共同研究を希望
- さまざまな利用用途での実装を、本学とともに検討いただける企業様のご支援を期待しております

本技術に関する知的財産権

発明の名称: 連合学習に関する情報処理方法、
サーバ、情報処理プログラム、及び、
学習済みモデル

出願番号: 特願2024-010135

出願人: 芝浦工業大学

発明者: 三好 匠、矢島 大路、山崎 託

産学連携の経歴

- 2006～2007 インターネット関連企業と共同研究を実施(代表者)
「安全なP2Pルータシステムの研究開発」
- 2008～2010 情報・通信関連企業と共同研究を実施(代表者)
「自律分散型P2Pトラフィック制御システム」
- 2020～2022 シンクタンク関連企業と共同研究を実施(代表者)
「COVID-19感染拡大・抑制シミュレーションにおける
検査・医療リソースシミュレーション」
- 2021～2022 情報・通信関連企業と共同研究を実施(分担者)
「IoTサービス品質に基づくセンサデータ通信制御技術
の研究」
- 2022～2025 NICT Beyond5G研究開発促進事業に採択(分担者)
「City as a Serviceを支えるデジタルツインを持続
可能な状態で自己成長させるエコシステム」

お問い合わせ先

芝浦工業大学

研究推進部 研究企画課

TEL: (03)5859-7180

Email: sangaku@ow.shibaura-it.ac.jp