

ディープラーニングにおける 適応的学習係数調整手法

神戸大学 大学院 科学技術イノベーション研究科
科学技術イノベーション専攻
教授 川口 博

2021年10月14日

ディープラーニング学習の問題点

- 一部の専門家のみ利用可能
 - 演算時間(資源)がかかる
 - 学習係数(LR)の設計が大変
 - 高度な専門知識が必要

従来手法

・学習係数(LR)とは...

SGD(確率的勾配降下法)

$$w^{(t+1)} = w^{(t)} - \eta \nabla E_n$$

w : 重み, バイアス

η : 学習係数

∇E_n : 勾配

・学習係数選択手法

e.g. AdaGrad, AdaDelta, Adam, etc...

➤パラメータ数が増加

➤タスクやネットワーク毎に選択が必要

➡ 試行を繰り返して選択, 設計する必要がある

LRの設計を低コストなアルゴリズムによる探索する必要

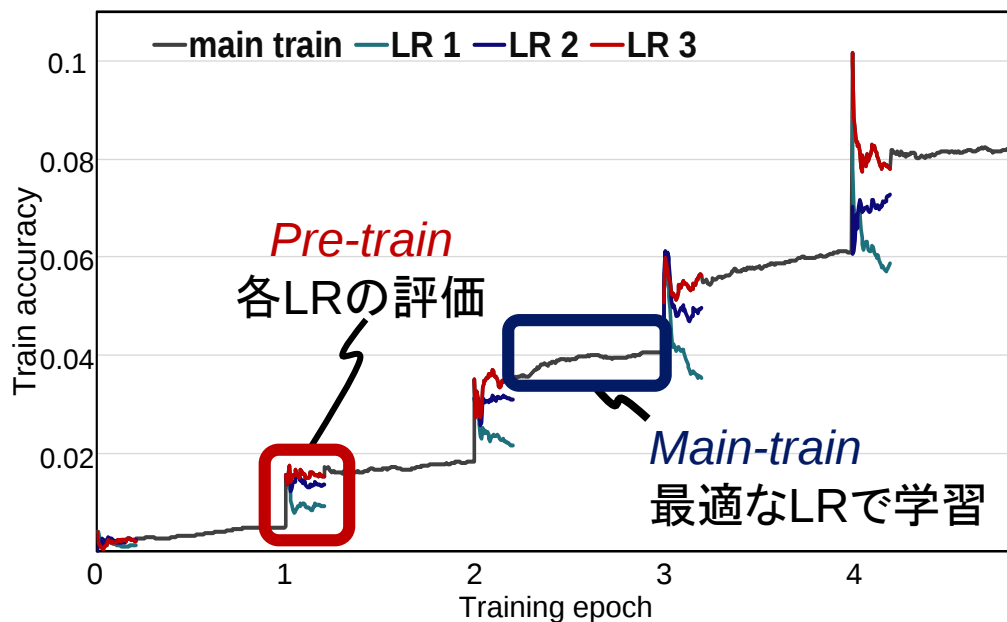
提案手法

①学習係数の評価を短時間で行う(STPT: Short-Term Pre-Training)

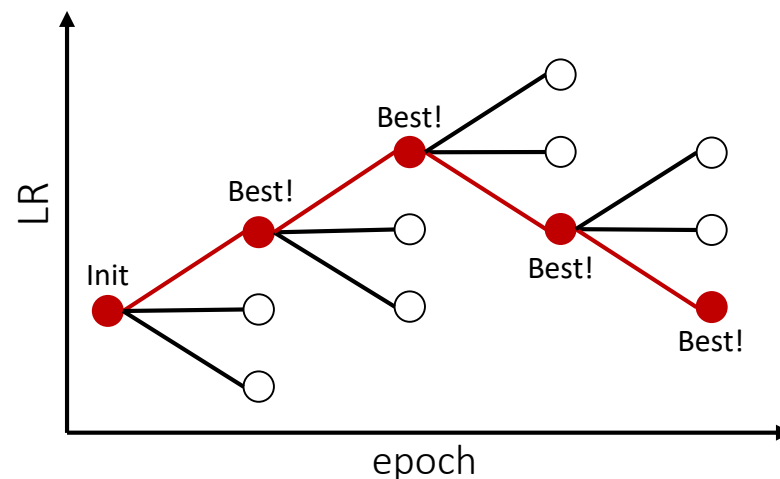
➡学習係数評価にかかるオーバーヘッドを削減(演算コスト削減)

②複数の学習係数を用いる(MLT: Multi LRs Training)

➡学習係数を事前に決定する必要がない(知識の削減)

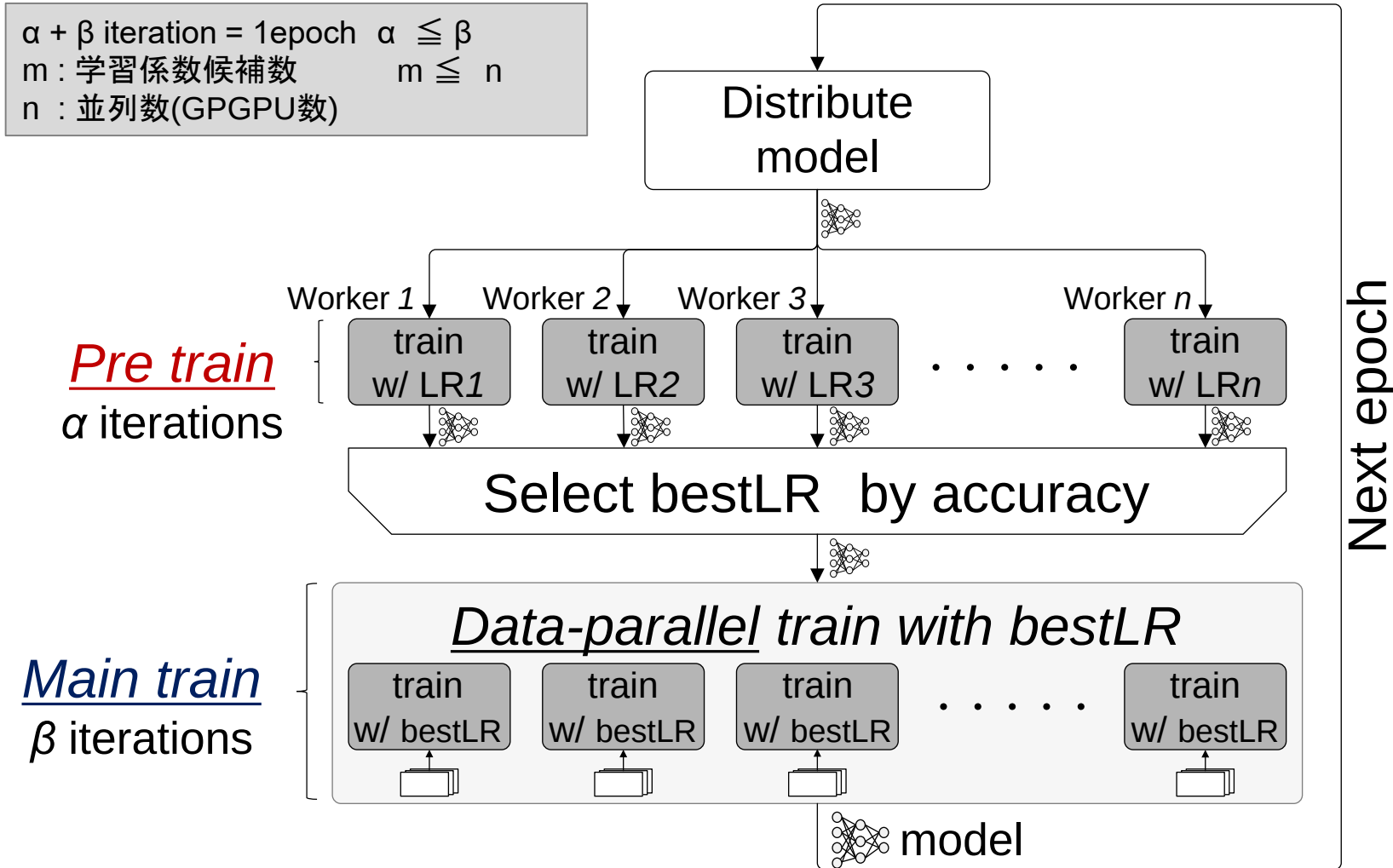


epoch毎に最適な学習係数(LR)を選択



Epoch毎に適応的なLR探索

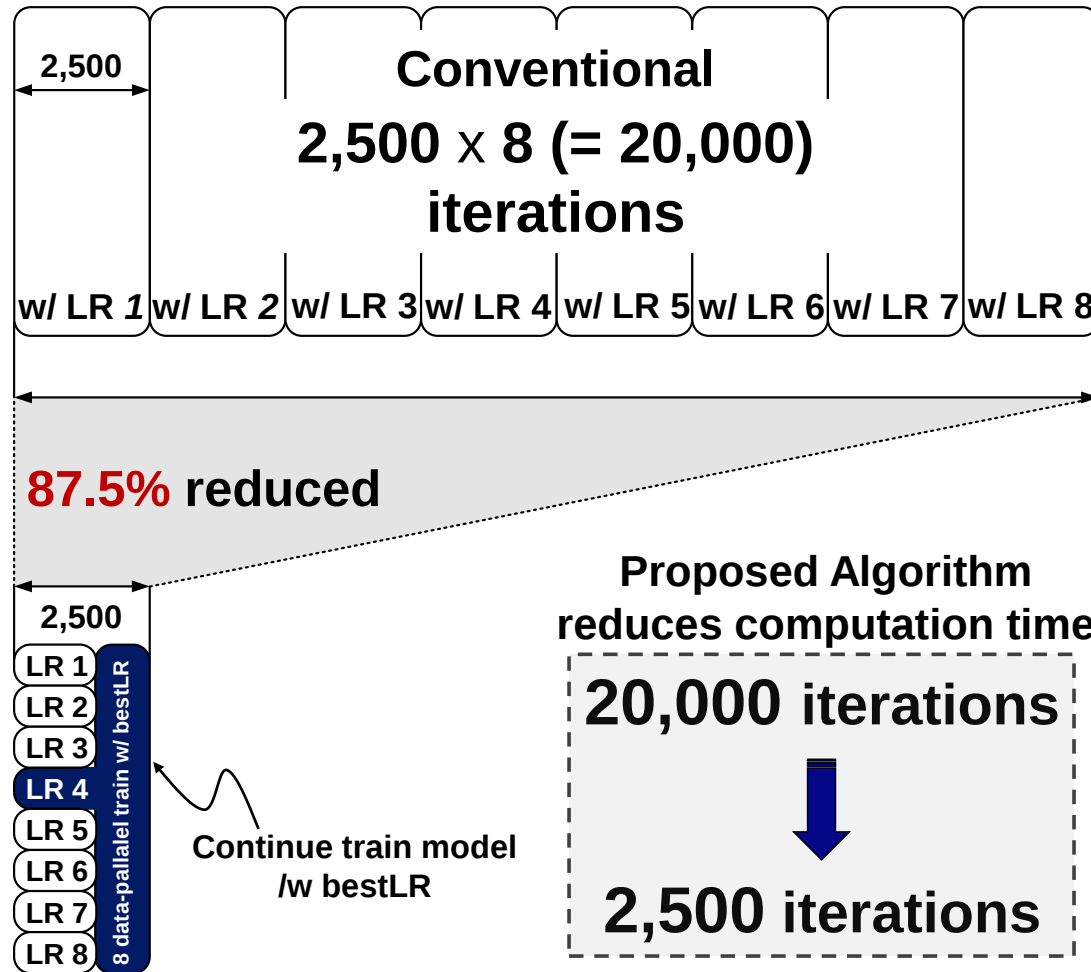
提案手法の概要図



高速かつ適応的にLRを探索可能

8並列におけるIteration数比較

・8並列 ・512 batch ・128万枚 ・学習係数8種



Iteration数を87.5%削減

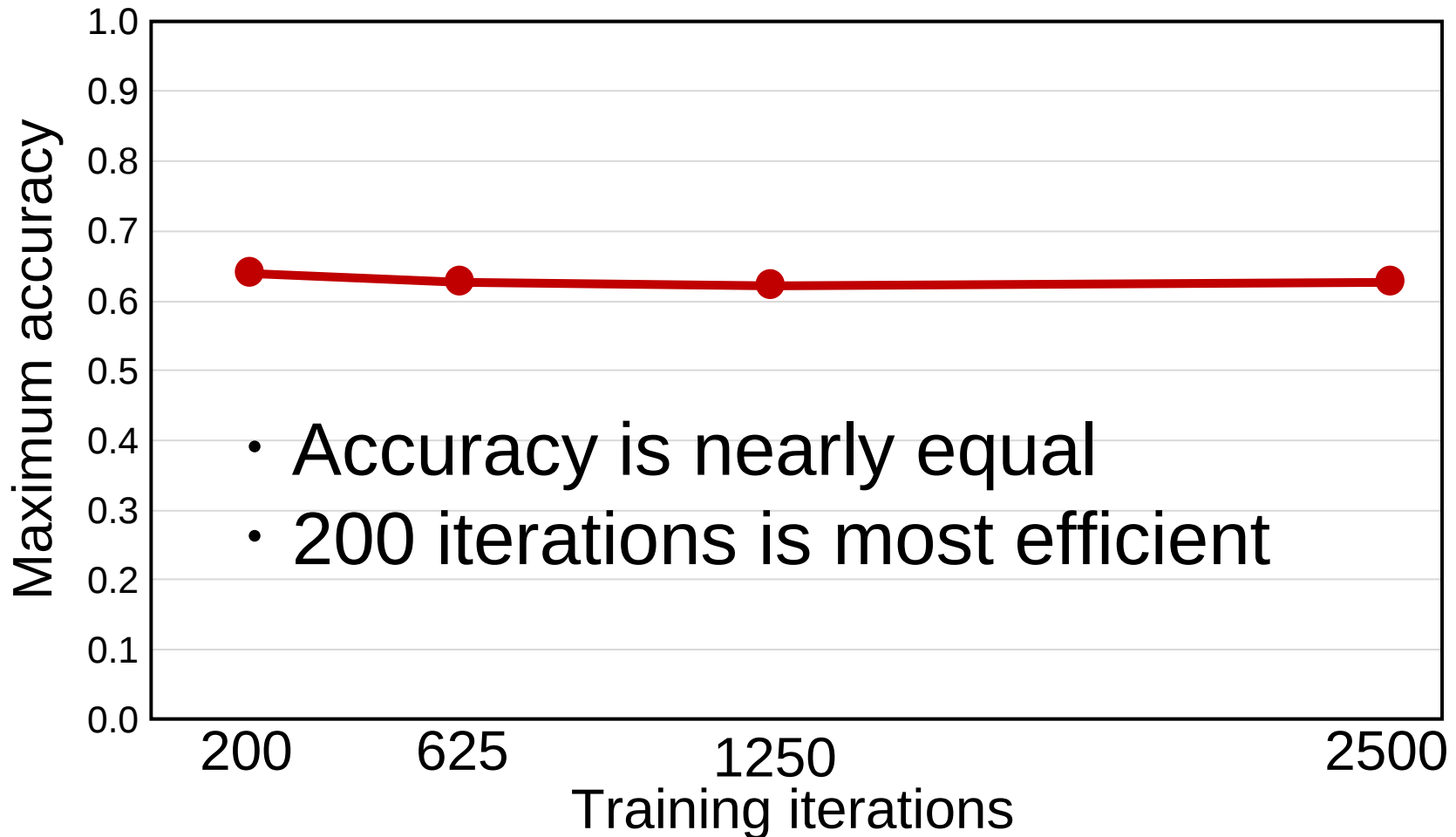
実験環境

- ImageNet-1K (Train: 1,280,000枚, Val: 50,000枚)
- Momentum SGD
- warmupを4epoch行う
- epoch毎にLR setを掛け合わせ, 学習係数候補を生成する

ネットワーク	ResNet50
GPGPUの枚数(並列度)	8
バッチサイズ (1 GPGPU)	512(64)
学習epoch数	40
Pre train iteration数(α)	200
Momentum係数	0.9
学習係数初期値	0.1
LR set	後述
Warmup epoch数	4
Warmupの学習係数	0.005

α の決定方法

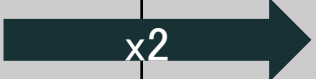
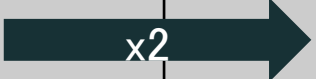
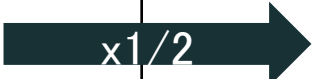
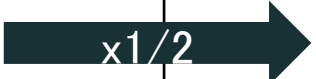
■ $\alpha = 200, 625, 1250, 2500$ で実行



α の値による影響はわずかなので200でよい

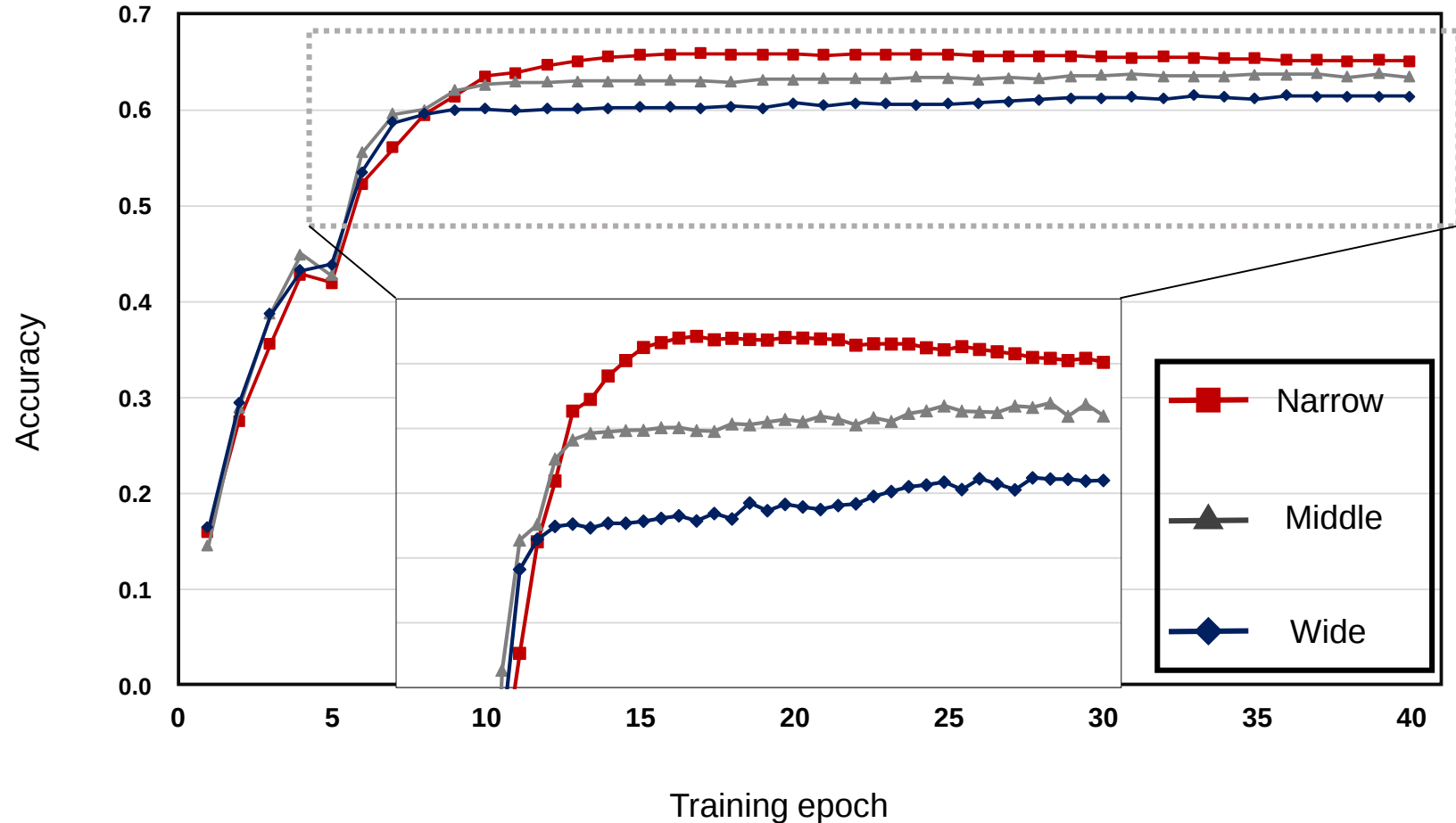
LRsetの決定方法

■ 間隔の違う3種類のLRset: Narrow, Middle, Wide

Values	(a) Narrow	(b) Middle	(c) Wide
#1	1.25 	2.50 	5.00
#2	1.17	2.00	3.67
#3	1.08	1.50	2.33
#4	<u>1.00</u>	<u>1.00</u>	<u>1.00</u>
#5	0.95	0.85	0.80
#6	0.90	0.70	0.60
#7	0.85	0.55	0.40
#8	0.80 	0.40 	0.20

LRset評価(精度推移)

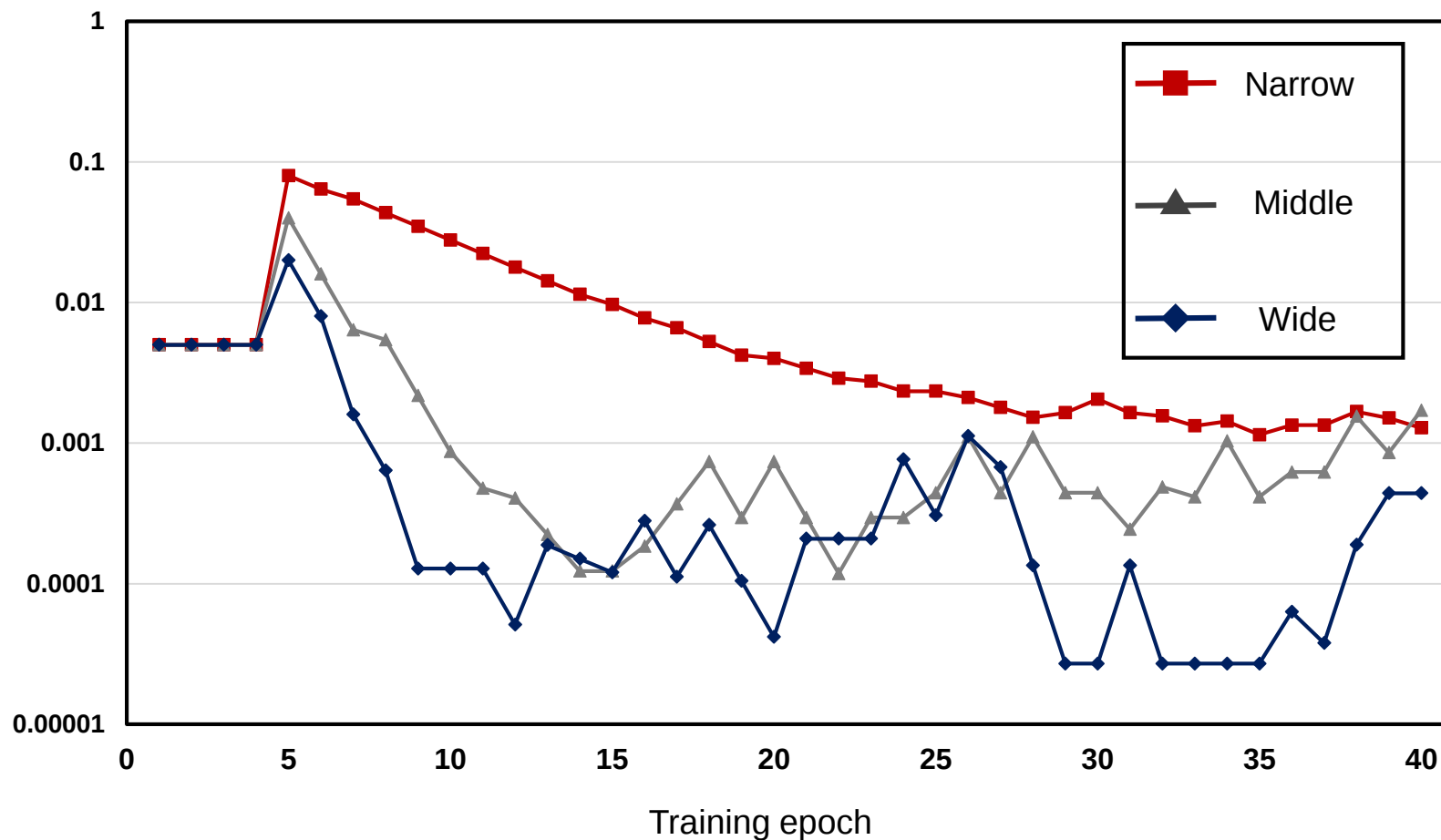
■ 各LRsetをそれぞれ用いてアルゴリズムを実行した際の精度



各LRsetに長所・短所が存在

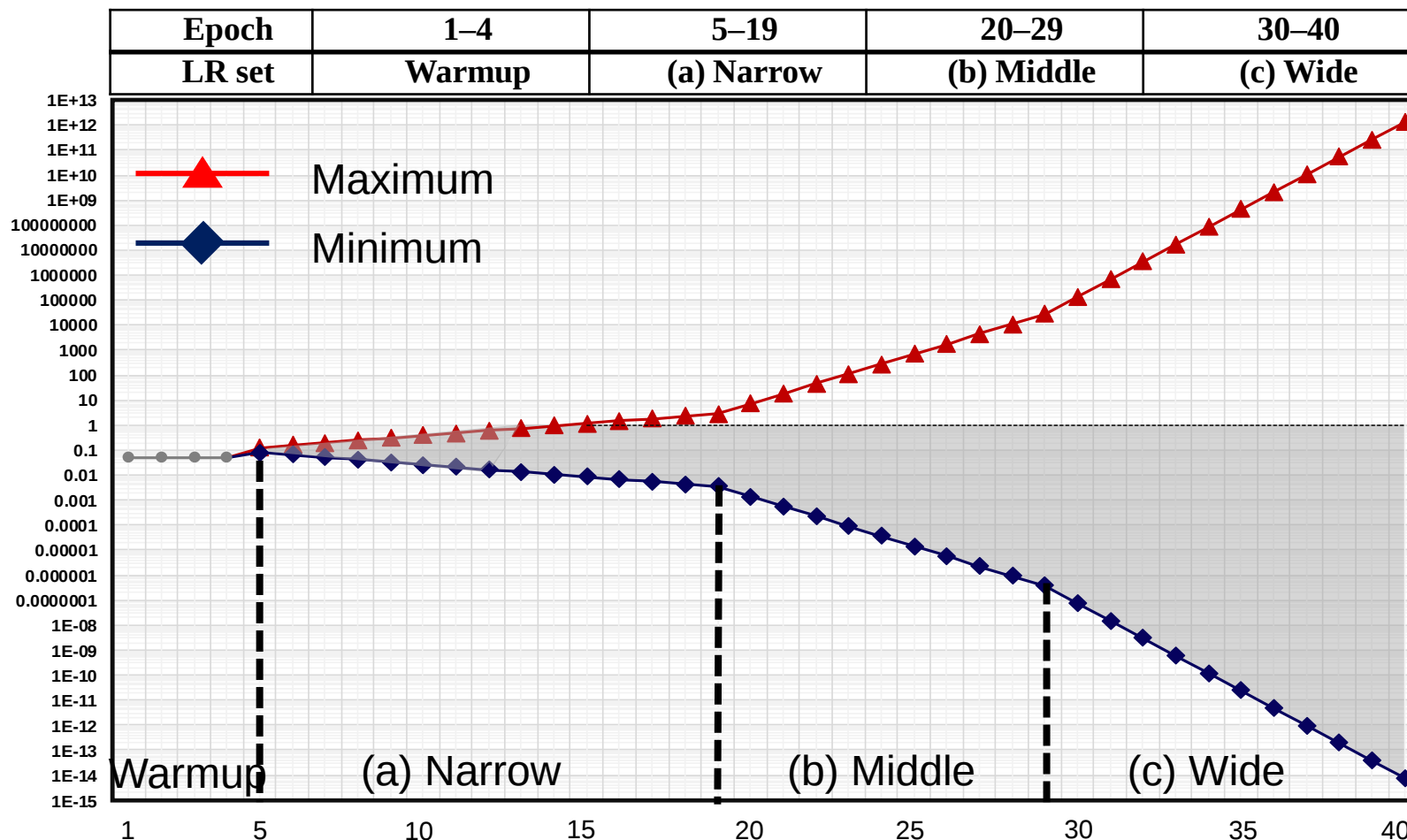
LRset評価 (LR推移)

■ 各LRsetをそれぞれ用いてアルゴリズムを実行した際のLR



Mixの提案

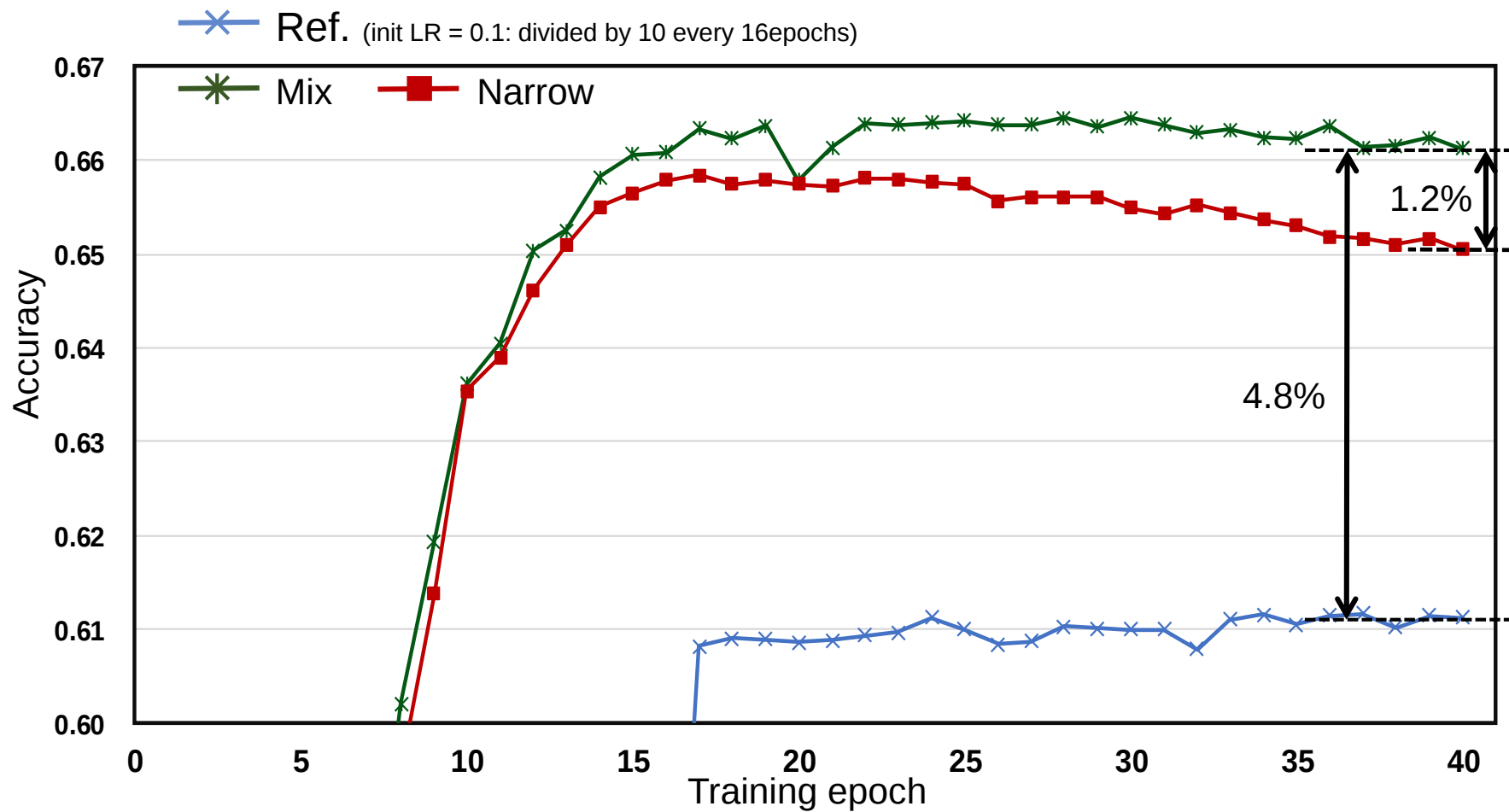
■ Mix: LRsetを切り替える



各LRsetの長所を活かす

Mixの精度推移

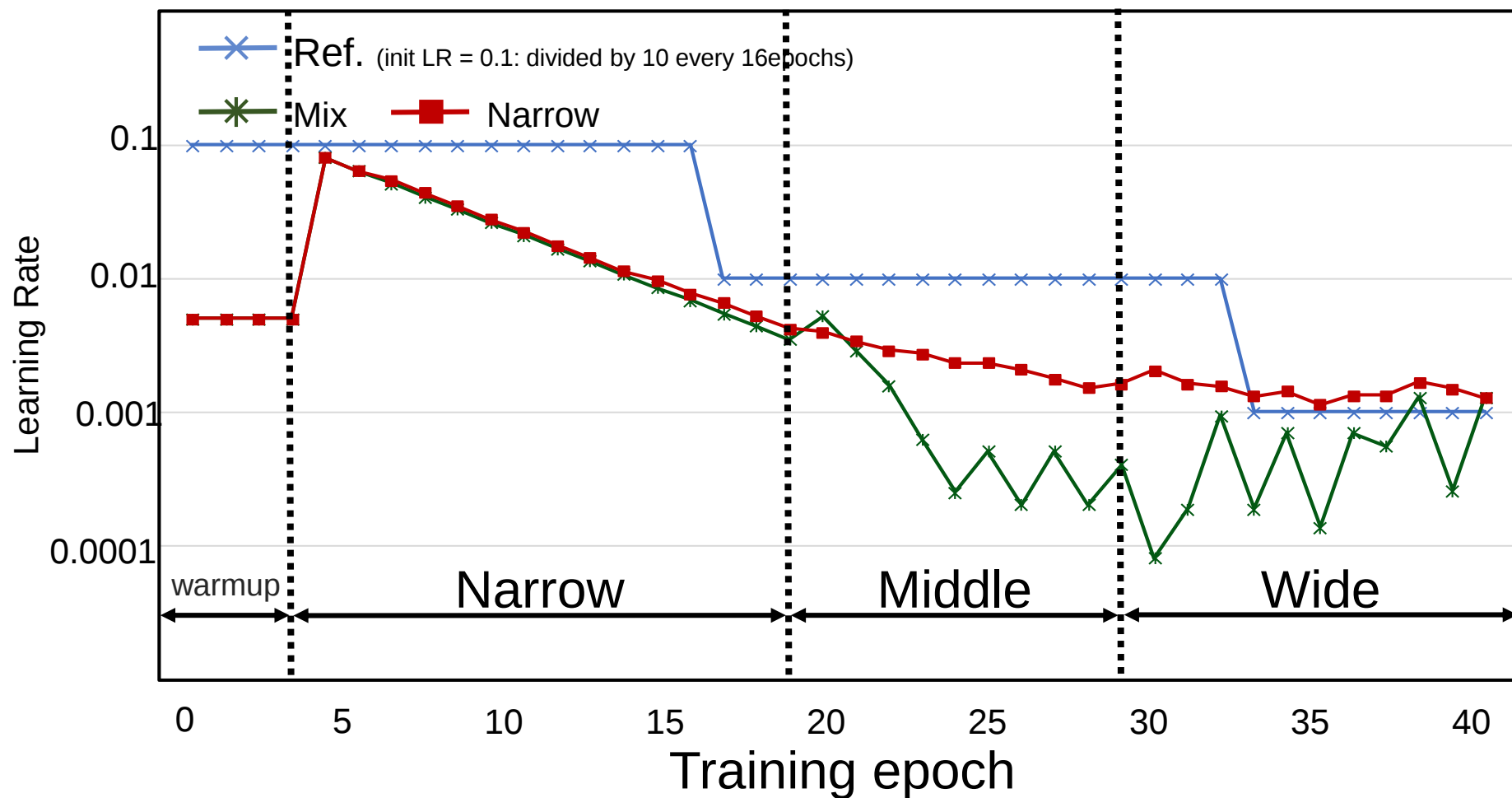
■ Mix, Narrow, Ref.の検証精度推移比較



低コストなLRの探索が劣化なく可能, 従来より4.8%高精度

MixのLR推移

■ Mix, Narrow, Ref.のLR推移比較



まとめ

①学習係数の評価を短時間で行う (STPT: Short-Term Pre-Training)

➡学習係数評価にかかるオーバーヘッドを削減
(演算コスト削減)

②複数の学習係数を用いる (MLT: Multi LRs Training)

➡学習係数を事前に決定する必要がない
(知識の削減)

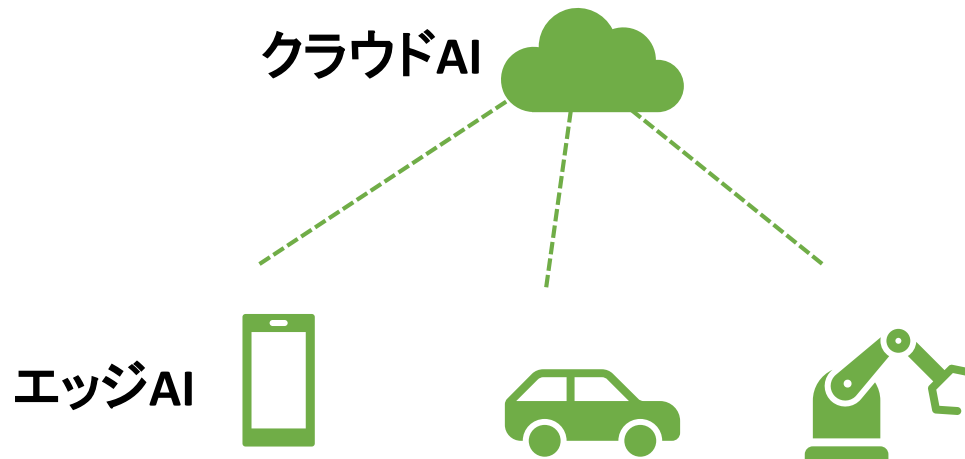
- ・上記2手法により低コストな探索でLRの設計可能
 - ・8並列, 8種の学習係数, 64batchの場合,
87.5%のiteration数を削減
 - ・段階的に減少する学習係数と比較し最大4.8%の精度向上.

想定される用途

- エッジAI: IoT機器による物体認識・セグメンテーションなどのディープラーニングアプリケーション

ex) スマートフォン、自動運転、工場の検品 etc.

- クラウドAI: 機械学習用サーバ



実用化に向けた課題

- ディープラーニング学習における学習係数の適応的調整については可能性を示したが、学習の高速化は学習係数の調整だけでなされるわけではない
→ 並列学習・データセット削減・蒸留・Neural Architecture Search (NAS)などの他技術との組み合わせで総合的コスト最適化が必要
- 学習(クラウドAI)よりマーケットの大きい推論器(エッジAI)への展開が必要。研究室では推論器の高速化・小型化や、LSI実装の知見もあり。

企業への期待

- クラウドAI関連企業は国内ではかなり限られているので、エッジAIの実用化・商品化に興味ある企業との共同研究を希望
- 学習に限らず、エッジAI推論器の高速化・小型化についても共同研究を展開したい

本技術に関する知的財産権

- 発明の名称 : 学習システム、学習方法、
およびプログラム
- 出願番号 : 特願2018-196539
- 出願人 : 神戸大学
- 発明者 : 川口博、森陽紀、山田和樹

産学連携の経歴

- 2015年-2021年 システムインテグレーターと共同研究実施
- 2016年-2019年 NEDO IoT推進のための横断技術開発プロジェクトに採択
- 2017年-2020年 自動車メーカーと共同研究実施
- 2020年-2021年 鉄鋼メーカーと共同研究実施

お問い合わせ先

神戸大学

産官学連携本部 産学連携・知財部門

TEL 078 - 803 - 5945

FAX 078 - 803 - 5389

e-mail: oacis-sodan@office.kobe-u.ac.jp