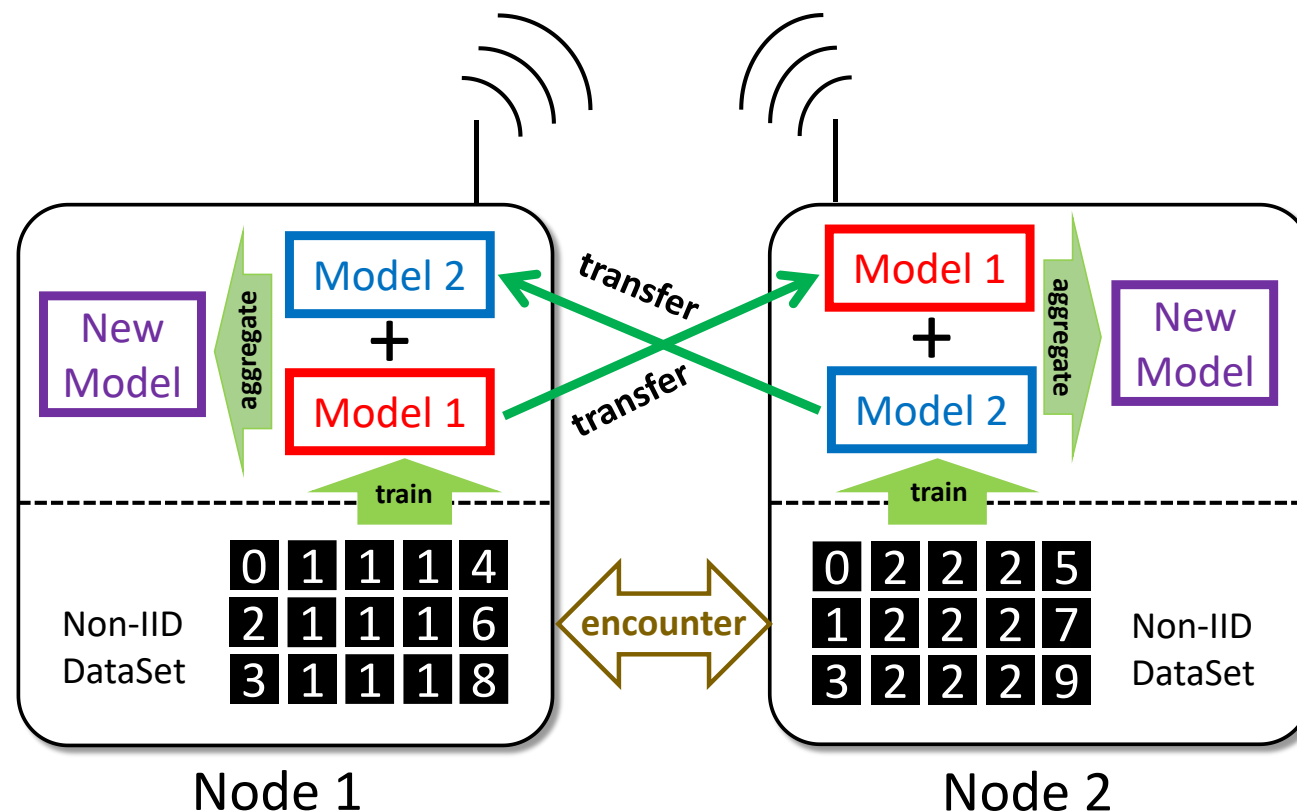


無線アドホック連合学習 ～完全分散型の協調的機械学習～



東京大学 大学院情報理工学系研究科
准教授 落合秀也, Ph.D.

Table of Contents

- 機械学習の分散化
- Wireless Ad Hoc Federated Learning (WAFL)
 - 特徴
 - 原理
 - 動作の確認(ベンチマーク評価)
 - 応用と発展
- 今後の見通し
- まとめ

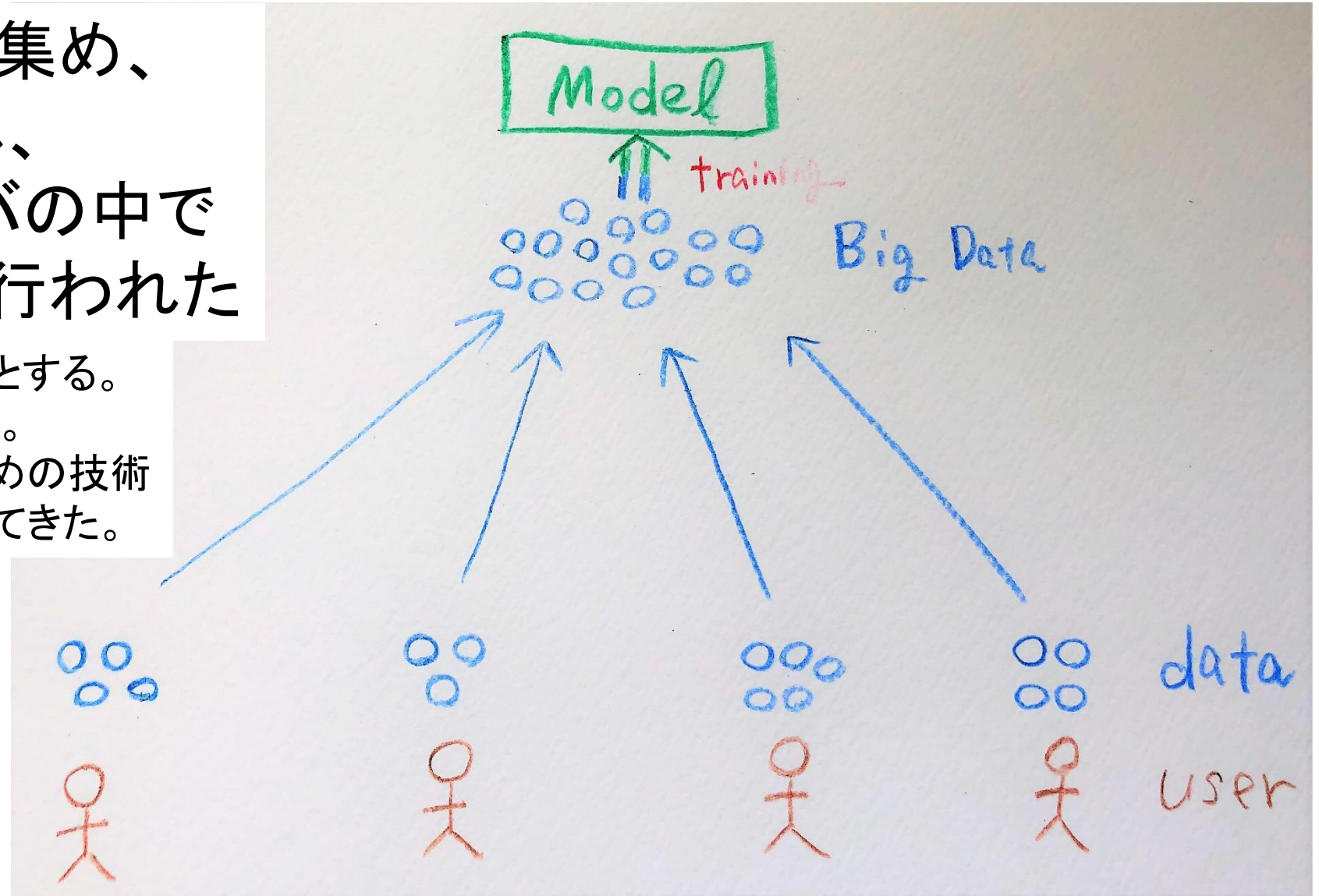
Table of Contents

- 機械学習の分散化
- Wireless Ad Hoc Federated Learning (WAFL)
 - 特徴
 - 原理
 - 動作の確認(ベンチマーク評価)
 - 応用と発展
- 今後の見通し
- まとめ

背景：機械学習アーキテクチャは中央集権的だった

1. データを1か所に集め、
2. Big Data を形成し、
3. 閉じられたサーバの中で
モデルの学習が行われた

画像処理、音声認識などを得意とする。
そこにあるデータが対象となった。
インターネットはデータを運ぶための技術
でしかなく、ほぼ無縁と考えられてきた。



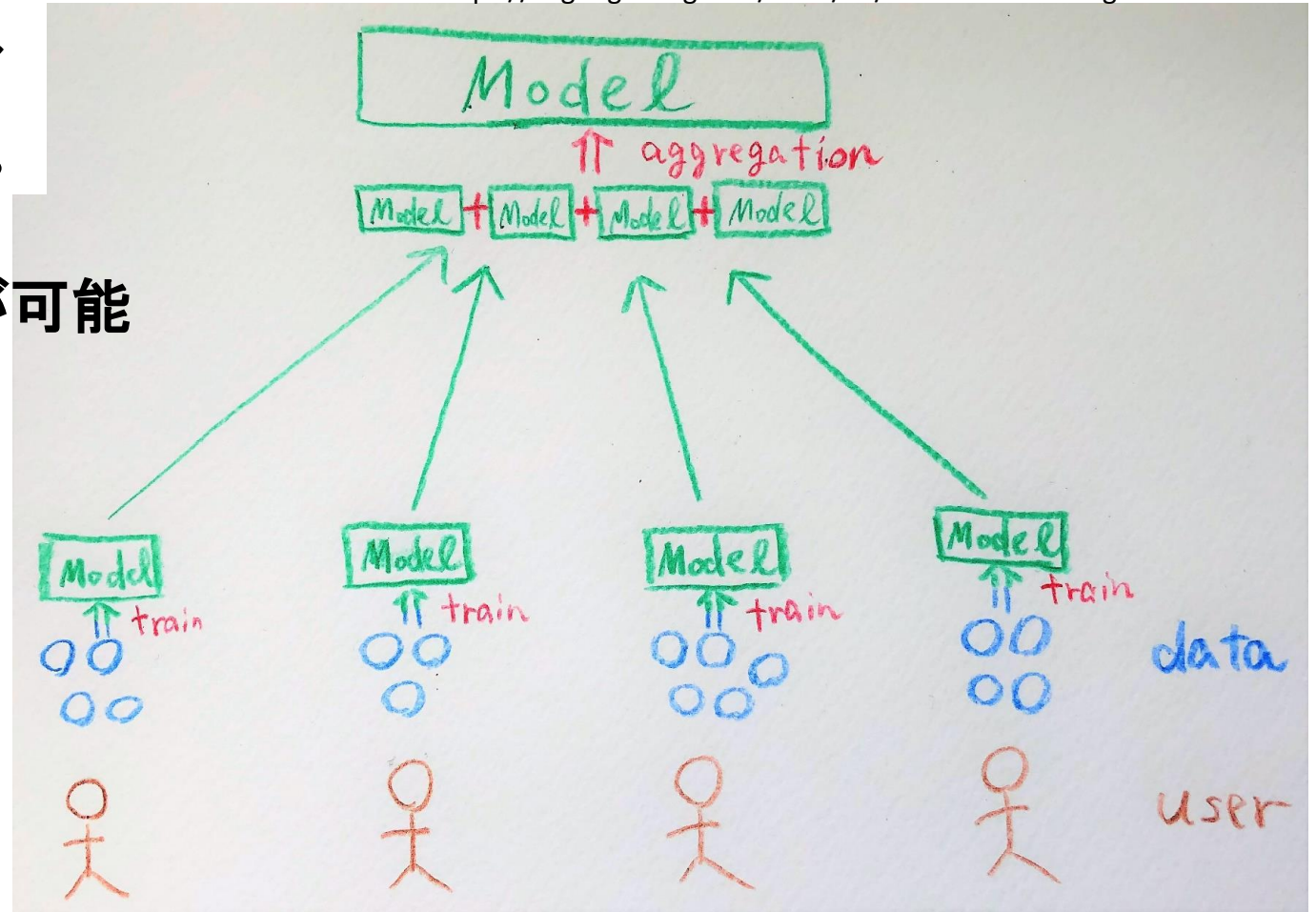
背景: Federated Learning (2017 – by Google, etc.)の登場

Federated Learning: Collaborative Machine Learning without Centralized Training Data
<https://ai.googleblog.com/2017/04/federated-learning-collaborative.html>

プライバシー規制 (欧州のGDPRなど)を背景に、
データを集めずに機械学習を可能にする
連合学習 (Federated Learning) が提唱された。

ポイント2: 学習モデルの合成演算が可能

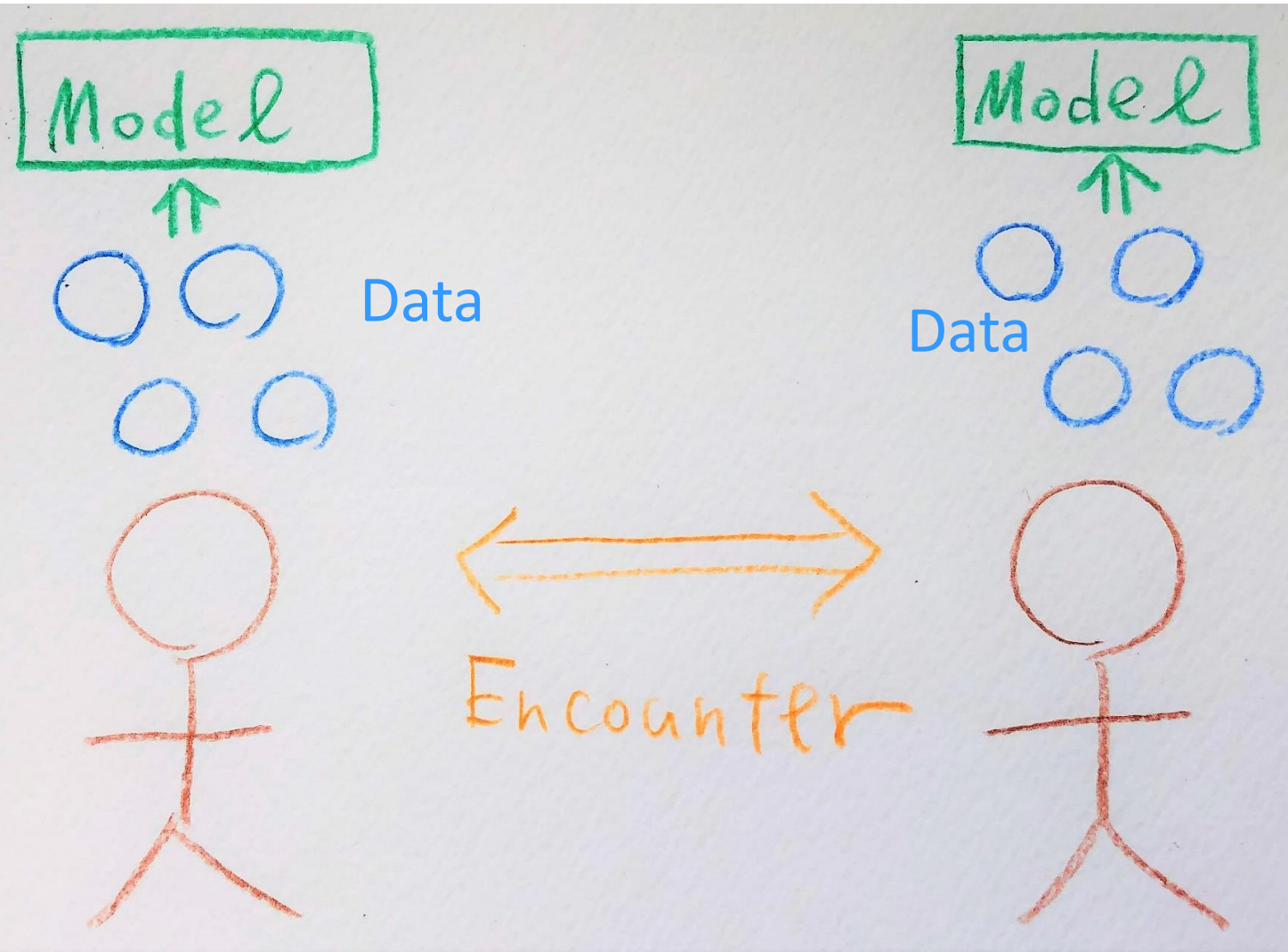
ポイント1: Data を集めずに



ただ、これはサーバ・クライアント 型だぞ ?! (いまだに中央集権的)
→ 完全な分散システムにしちやおうじゃないか !

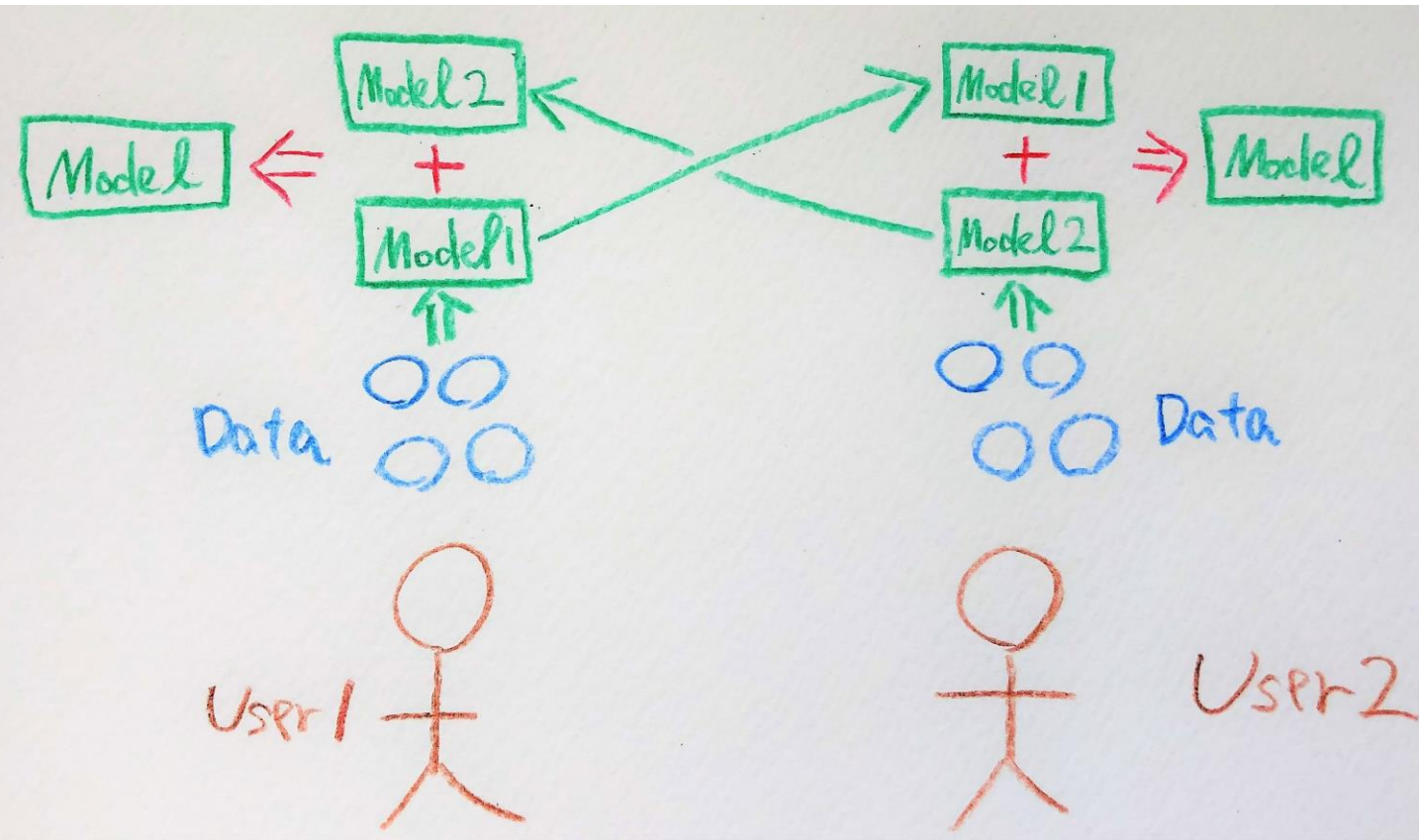
無線アドホック連合学習(WAFL) へのアプローチ

サーバ・クライアント型を Peer-to-Peer型にしてみる



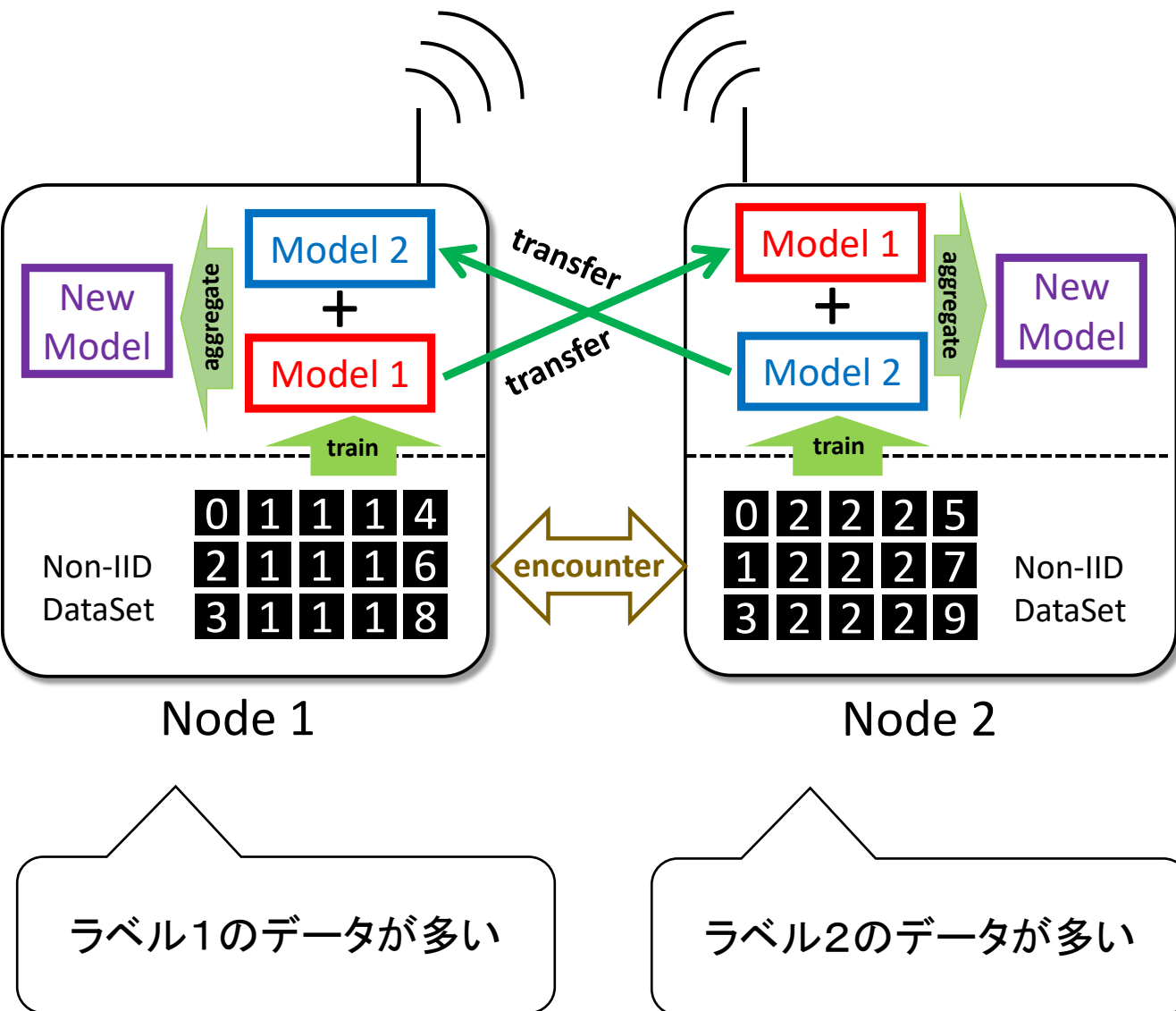
1. 各ノードは各自が持つデータを使いそれぞれモデル学習を行っている
2. それぞれのノードが出会う
3. 出会ったノード間には
Wi-Fi Ad Hocモードや
Bluetoothなど
の近隣無線技術で通信できる

無線アドホック連合学習(WAFL)へのアプローチ サーバ・クライアント型を Peer-to-Peer型にしてみる



1. 各ノードは各自が持つデータを使いそれぞれモデル学習を行っている
2. それぞれのノードが出会う
3. 出会ったノード間には
Wi-Fi Ad Hocモードや
Bluetoothなど
の近隣無線技術で通信できる
4. お互いに持っているモデルを交換し合成して、新しいモデルとする。
5. これによって、学習が協調的に
行われる。

協調的な学習がもたらす意味



1. 各ホストの持つデータにはクラスの偏りがある例)

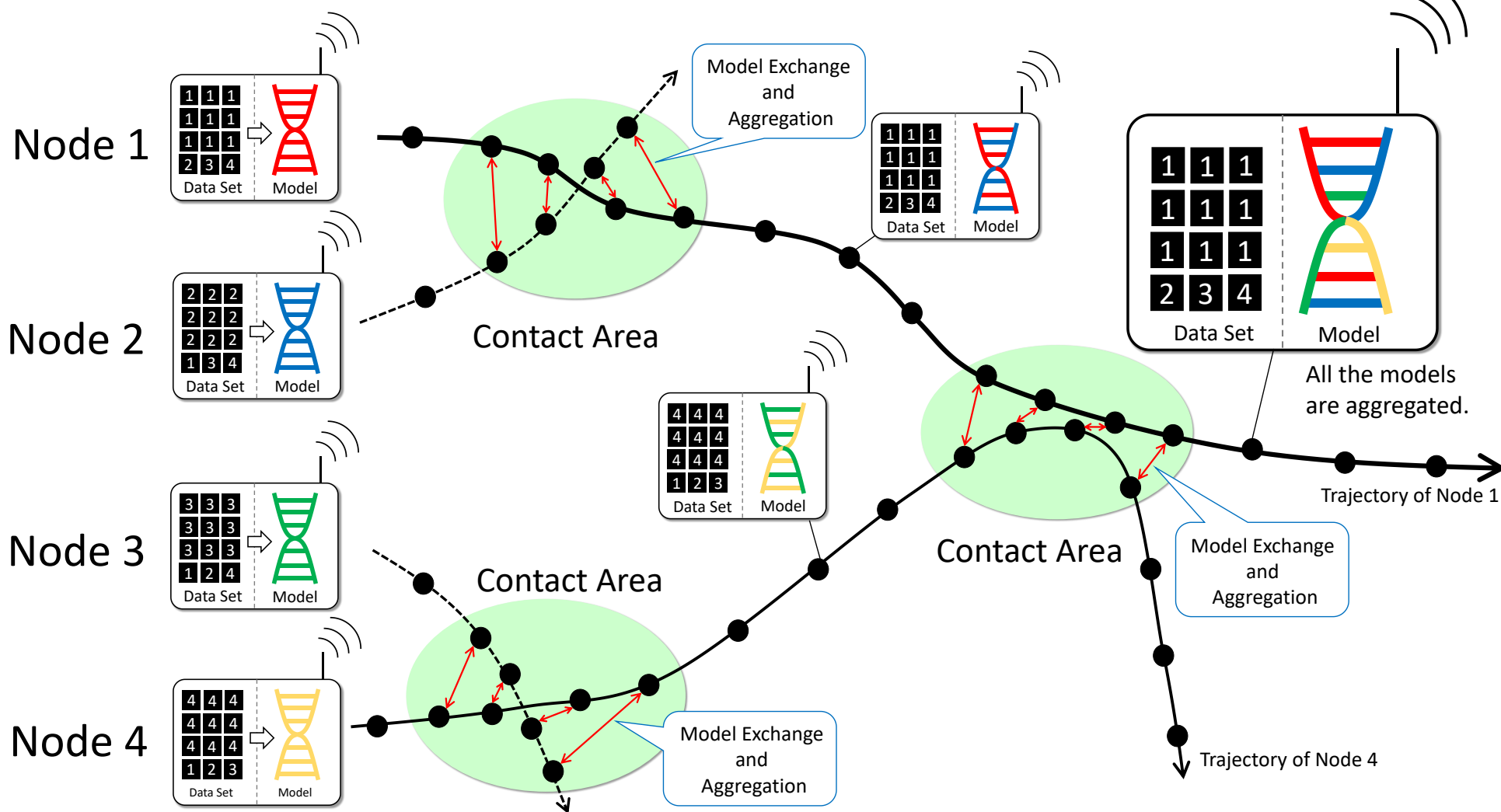
- ラーメン好きはラーメンの写真が多数
- 電車好きは電車の写真が多数

2. データを集めて Big Data を形成する場合には、クラスの偏りは中和されたが、集めないとそのような傾向が生まれやすい **Non-IID 特性** と言う。

3. 偏りがあるデータで学習を行うと汎用的な判別ができるモデルにならない例) ラーメン好きの端末では、線路の写真をラーメンと判別しやすくなる。

4. それぞれのモデルを合成することで、正しく判別できる汎化されたモデルを生成できる。

多くのノードとの出会いが汎化モデルの獲得に繋がる



偏りのある学習が行われている

出会いによるモデルの合成

汎化モデルの獲得

Table of Contents

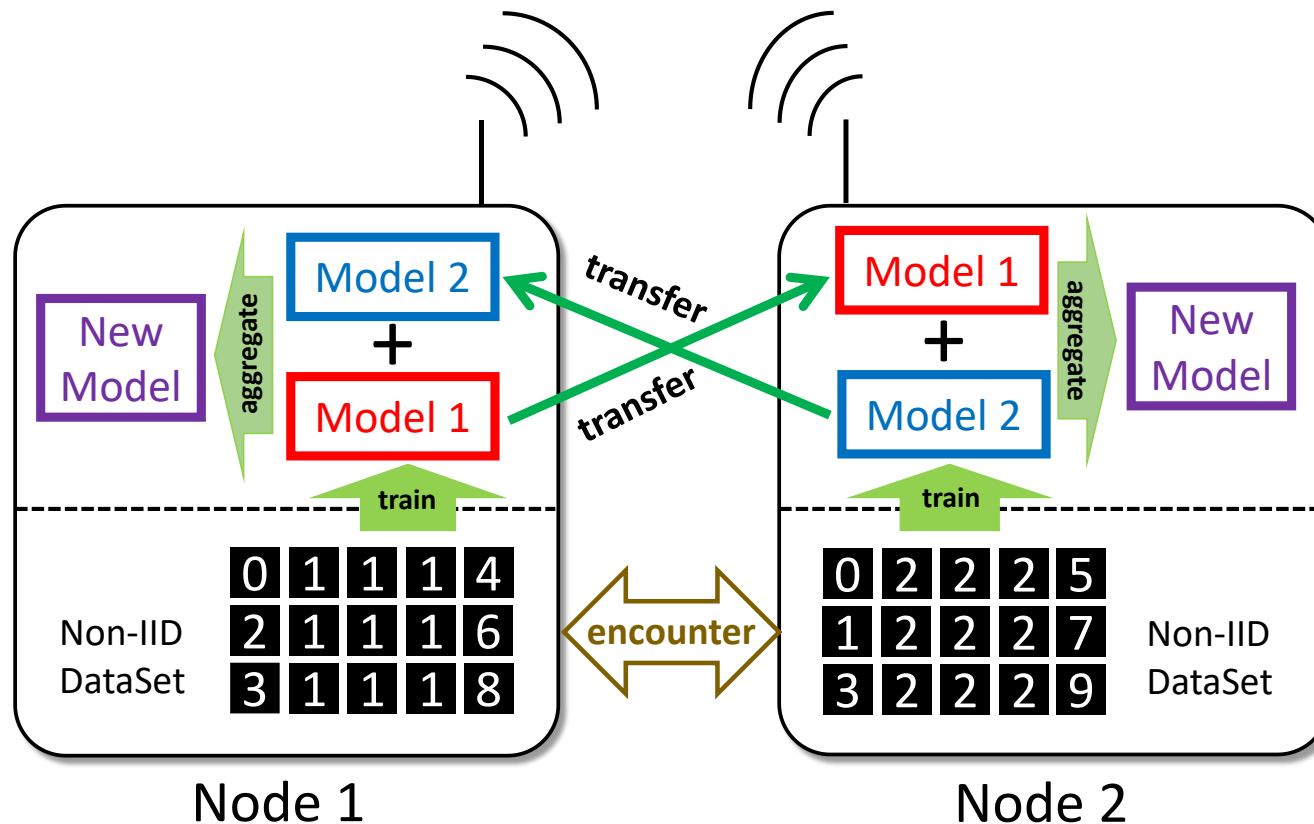
- 機械学習の分散化
- Wireless Ad Hoc Federated Learning (WAFL)
 - 特徴
 - 原理
 - 動作の確認(ベンチマーク評価)
 - 応用と発展
- 今後の見通し
- まとめ

無線アドホック連合学習(WAFL)の特徴

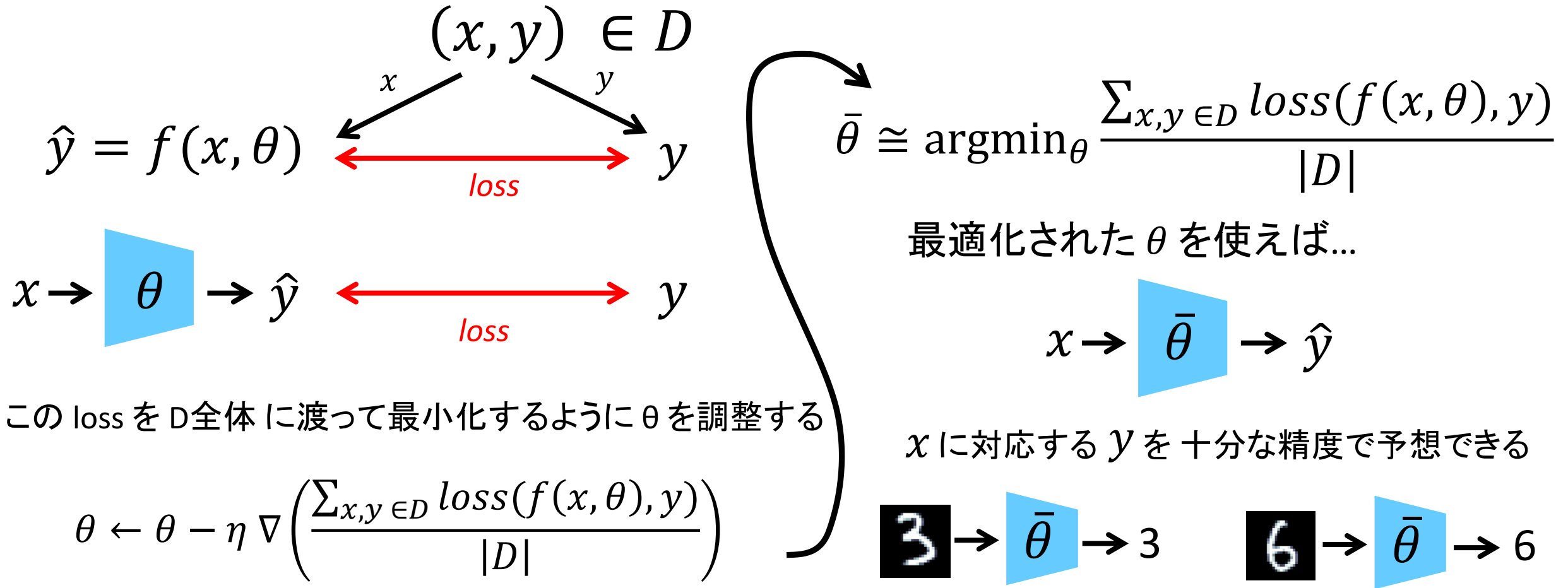
1. 第三者(仲介者)を交えずとも学習が可能
 - 仲間内の問題に第三者を介入させない
2. 権力構造が存在しない
 - Server-Client型では Server運営組織に権力が集中する
 - Server-Client型は Master-Slave型に化ける。
 - Peer-to-Peer型では対等である。
3. Multi-Vendor環境を実現可能
 - プロトコルを決めれば誰でも参入できる

無線アドホック連合学習(WAFL)の原理

WAFL: Wireless Ad Hoc Federated Learning



その前に・・・機械学習の仕組み

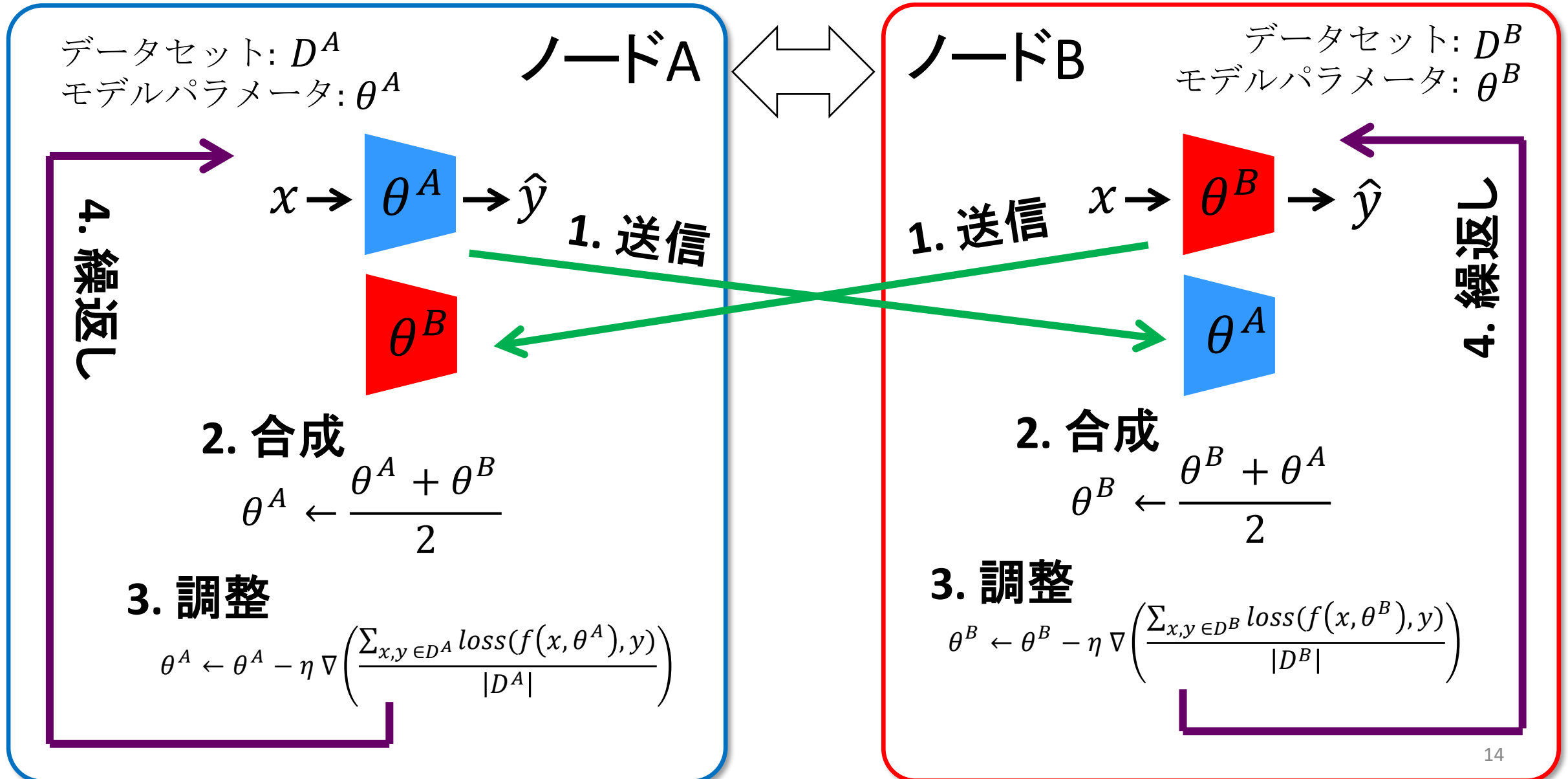


(*) θ は モデルパラメータ (1000次元など)

100発100中であれば精度(Accuracy)100%と言う
100発95中であれば精度(Accuracy)は95%

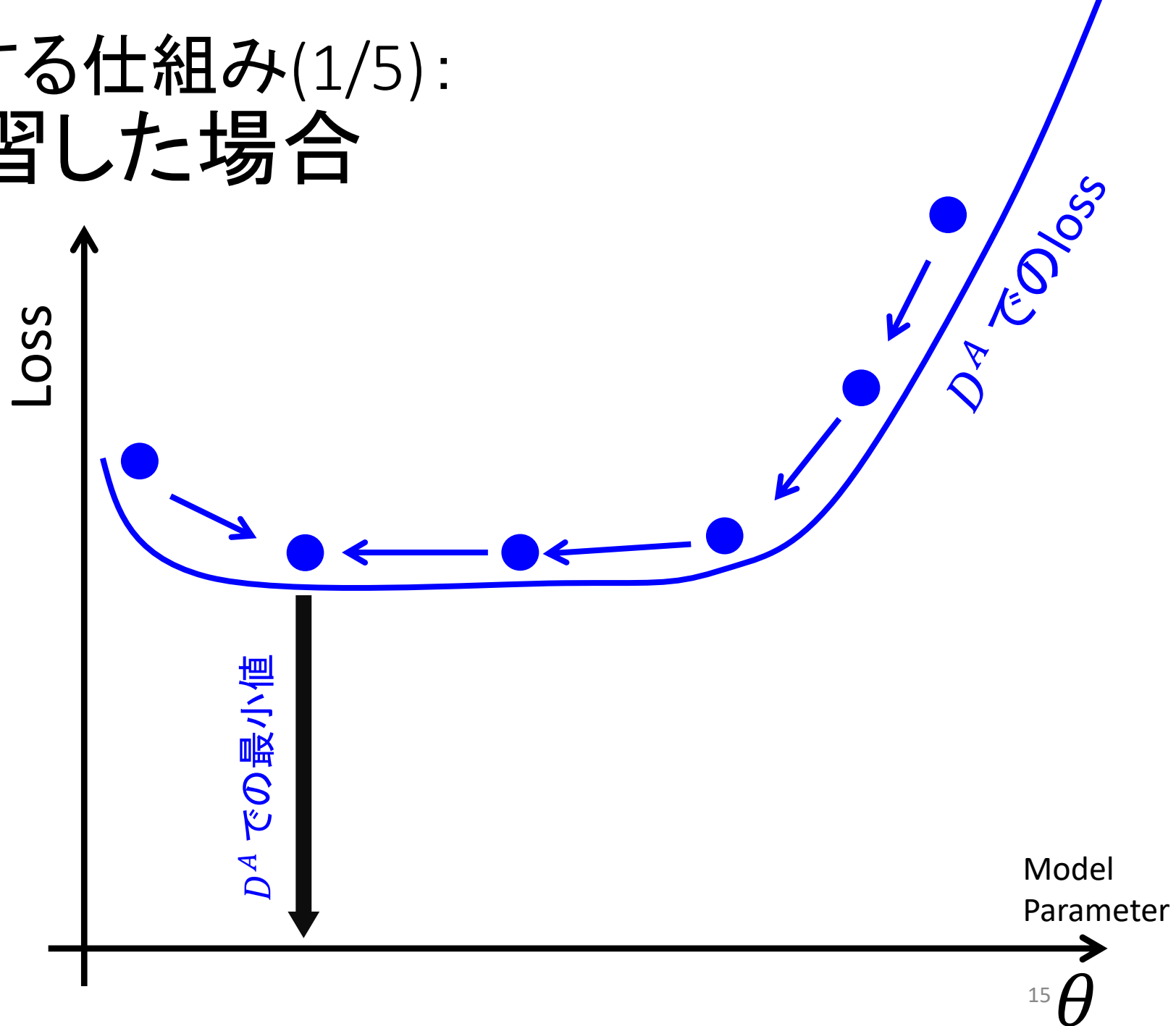
WAFLでの学習

0. 物理的な出会い



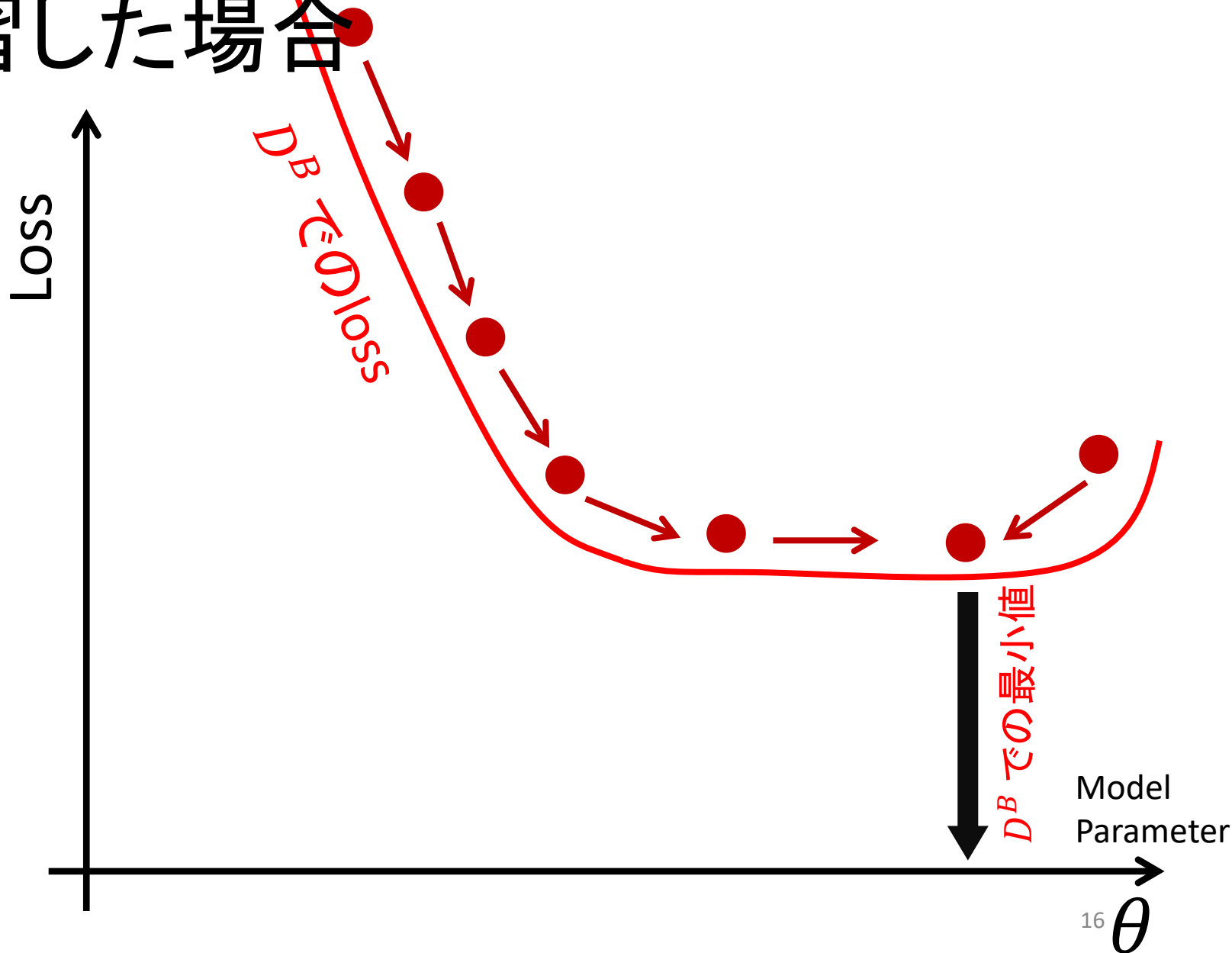
WAFLで学習が成立する仕組み(1/5): ノードA単体で学習した場合

最小化する θ にたどり着く
盆地のような構造がある



WAFLで学習が成立する仕組み(2/5): ノードB単体で学習した場合

データセットが異なれば
最小化する θ は異なる



WAFLで学習が成立する仕組み(3/5):

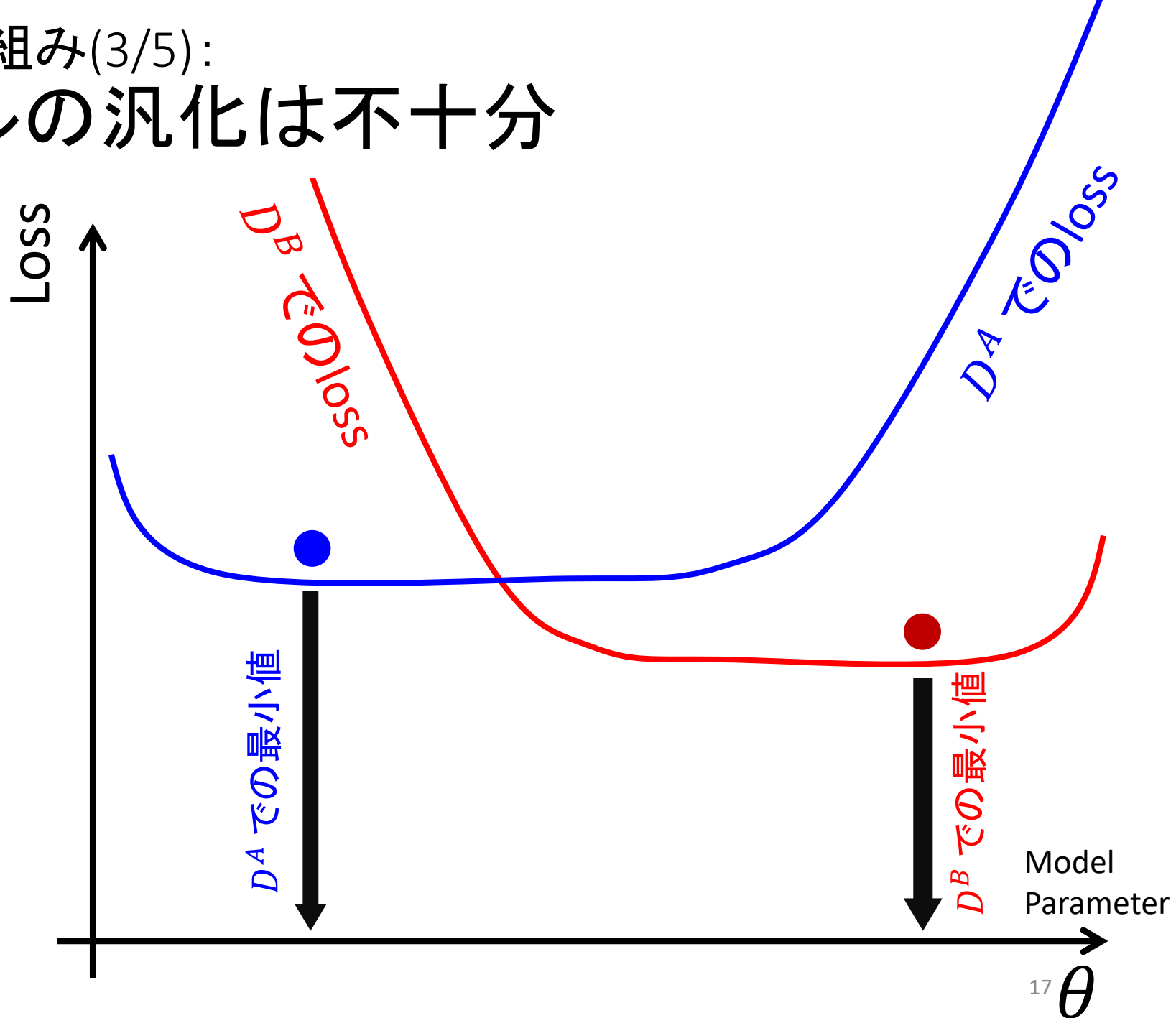
協調しないとモデルの汎化は不十分

1. D^A での最小値は
Node A にとっては最適だが、
Node B (D^B) には不適

→ モデルの汎化は不十分

2. D^B での最小値は
Node B にとっては最適だが、
Node A (D^A) には不適

→ モデルの汎化は不十分



WAFLで学習が成立する仕組み(4/5):

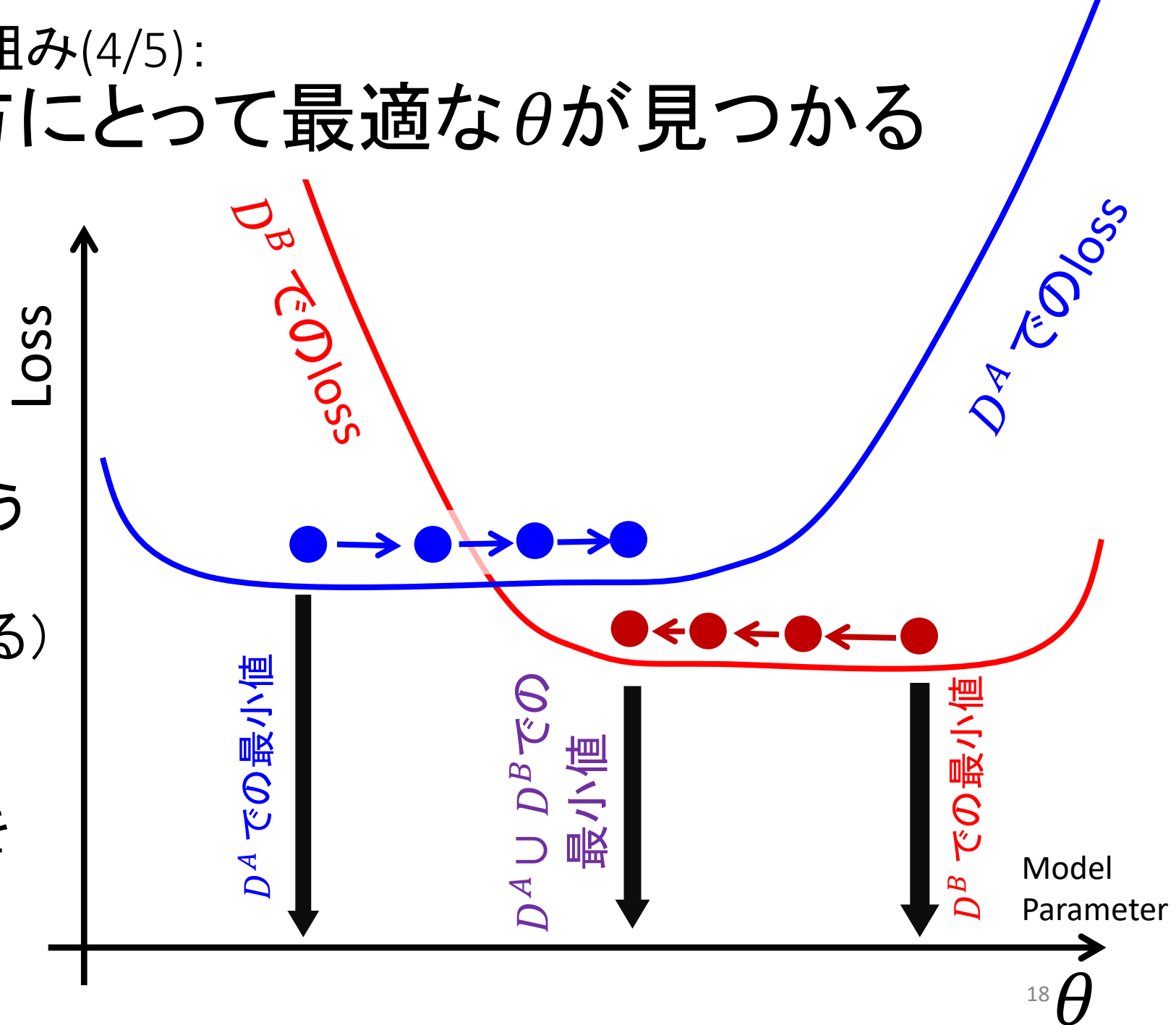
協調することで双方にとって最適な θ が見つかる

モデルの合成:

$$\theta^A \leftarrow \frac{\theta^A + \theta^B}{2}$$

は、 θ 空間上で引き寄せあう働きをする。
(上記の場合は、中点を取る)

→ 双方にとって最適な点を探索することになる



WAFLで学習が成立する仕組み(5/5):

収束の様子・・・2次元での考察

モデルの合成結果

$$\theta^A \leftarrow \frac{\theta^A + \theta^B}{2}$$

は必ずしも、盆地内にあるとは限らない

その後行われる調整、

$$\theta^A \leftarrow \theta^A - \eta \nabla \left(\frac{\sum_{x,y \in D^A} \text{loss}(f(x, \theta^A), y)}{|D^A|} \right)$$

によって、盆地内に引き戻される。

繰り返し行うことによって、
双方の共通の盆地にたどり着く

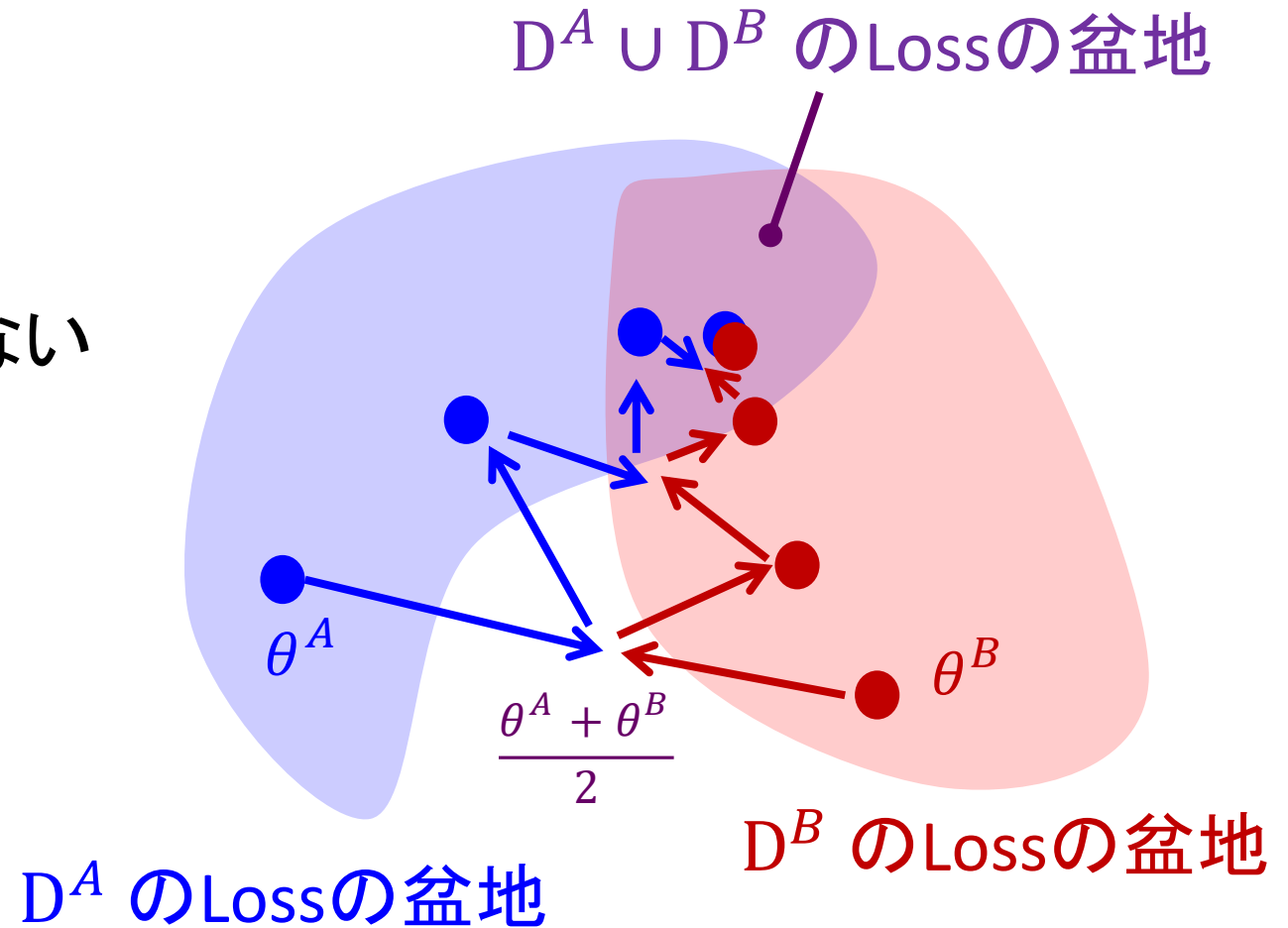
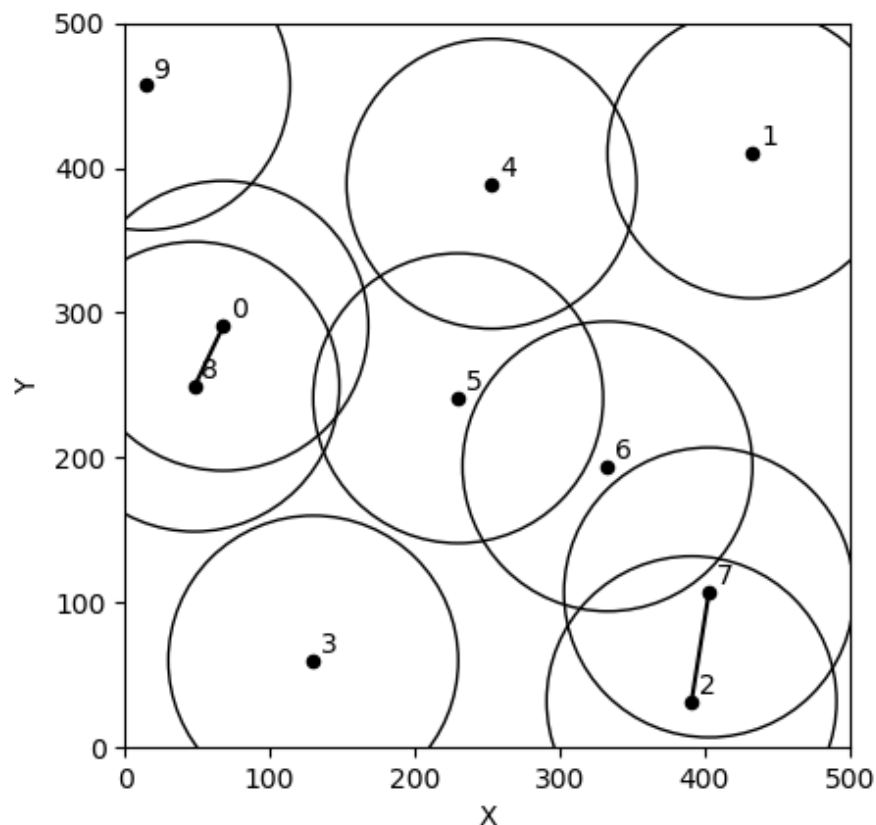


Table of Contents

- 機械学習の分散化
- Wireless Ad Hoc Federated Learning (WAFL)
 - 特徴
 - 原理
 - 動作の確認(ベンチマーク評価)
 - 応用と発展
- 今後の見通し
- まとめ

無線アドホック連合学習(WAFL)の動作確認

WAFL: Wireless Ad Hoc Federated Learning



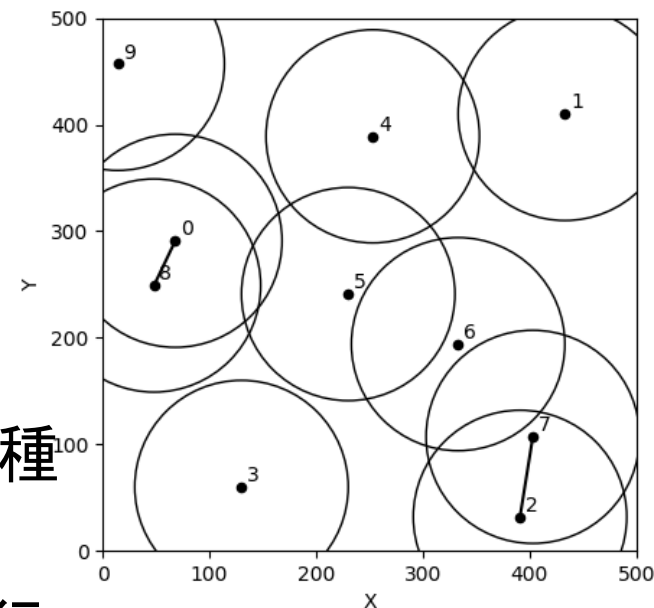
ベンチマーク評価による確認

- ベンチマークテストの意義
 - 標準的であり、再現性がある
 - 技術特性への理解が深まる
- データセット
 - MNIST を使用
- 機械学習モデル
 - Multilayer Perceptron (MLP) を使用
- モビリティ・パターン
 - 静的なトポロジー構成：4種
 - 動的なパターン Random Waypoint Mobility (RWP): 3種
 - 動的なパターン Community-Structured Environment (CSE): 3種
- シミュレーション
 - 1台のコンピュータ内にこれらを定義し、シミュレーションを実行



Yann et.al., THE **MNIST** DATABASE of handwritten digits, 1998.

<http://yann.lecun.com/exdb/mnist/>



Random Waypoint Mobility²²

ベンチマーク評価: 実験設定

90% Non-IID MNISTデータセット

■ Training用データの分布 Node n は ラベル n がノード内で90%を占めるようにサンプルを保有

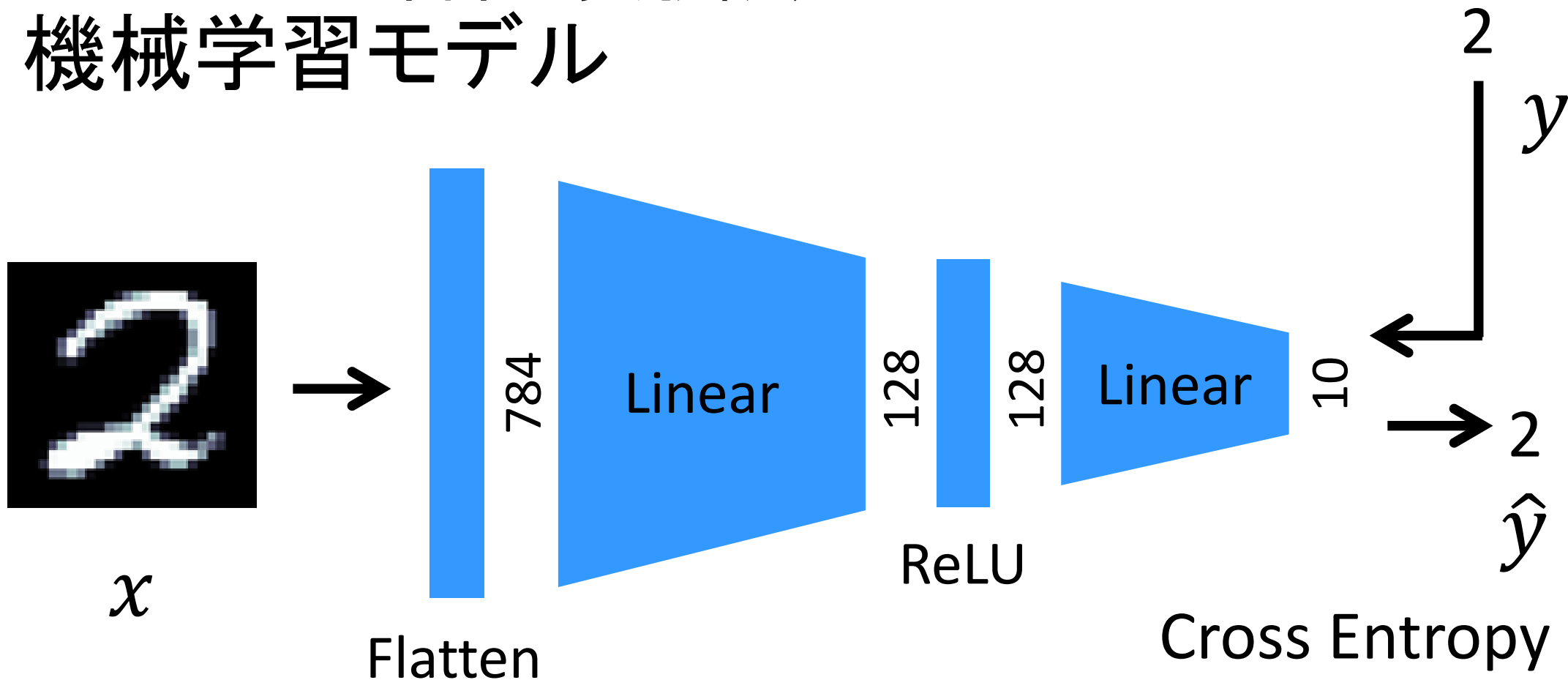
Node	L0	L1	L2	L3	L4	L5	L6	L7	L8	L9	Summary
0	5341	76	64	76	61	57	57	61	74	50	5917
1	79	6078	67	52	57	59	58	79	59	68	6656
2	58	67	5374	80	64	68	87	73	57	61	5989
3	68	74	73	5537	51	56	72	65	67	77	6140
4	73	67	80	67	5301	59	53	69	70	64	5903
5	60	66	57	74	68	4896	61	59	66	69	5476
6	52	78	53	56	58	66	5312	56	65	65	5861
7	67	90	66	74	55	61	65	5683	63	77	6301
8	59	80	59	54	63	48	88	67	5268	57	5843
9	66	66	65	61	64	51	65	53	62	5361	5914
Summary	5923	6742	5958	6131	5842	5421	5918	6265	5851	5949	60000

■ Test用データの分布

全ラベルが均等に出現するものとする (IID) – MNISTの標準テストデータセットを使用

(*) IID: *independent and identically distributed*

ベンチマーク評価: 実験設定 機械学習モデル



2層の全結合ニューラルネットワーク

Cross Entropy Loss

Adam Optimizer

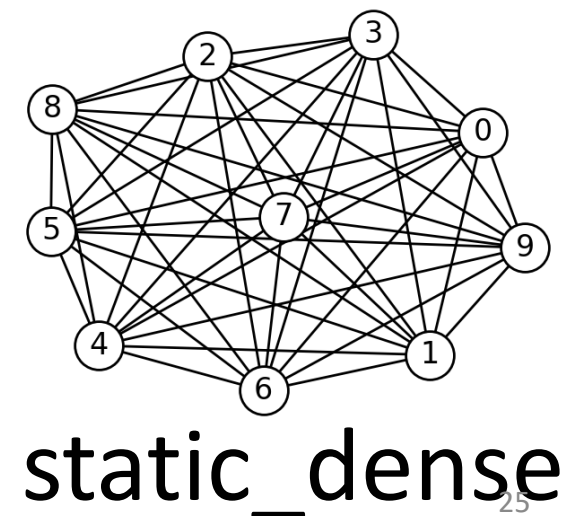
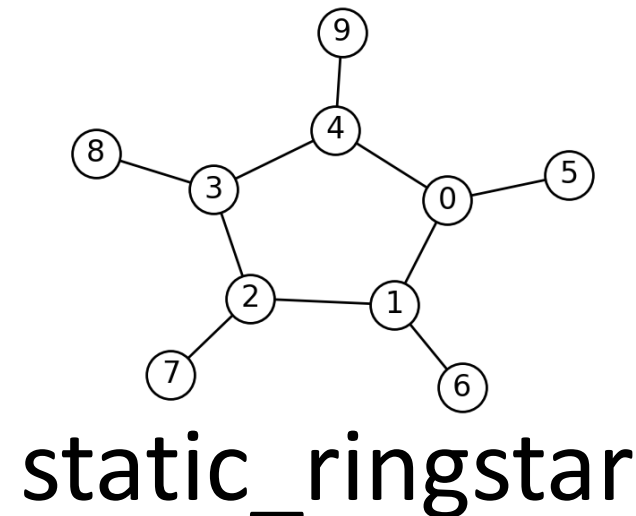
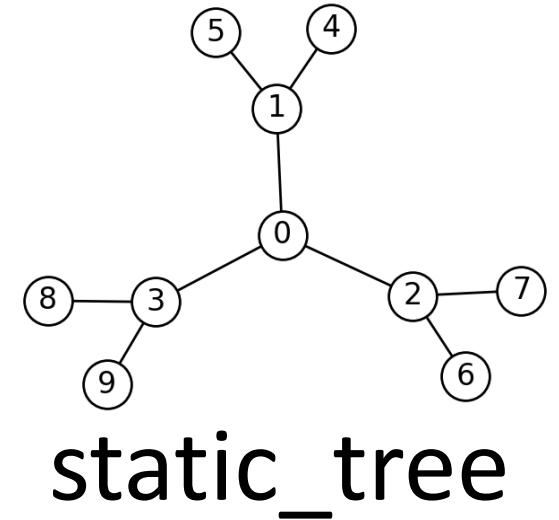
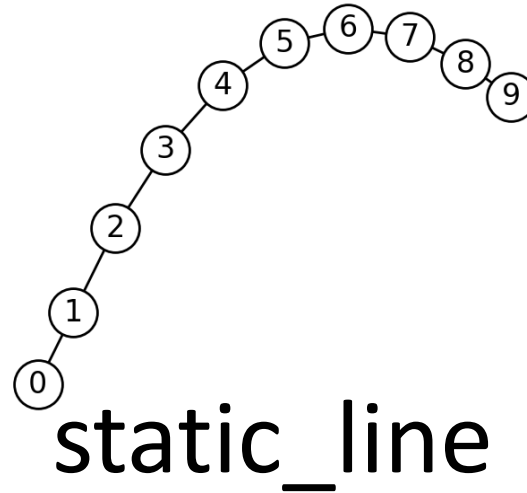
Learning rate 0.001

Coefficient of Aggregation 1.0

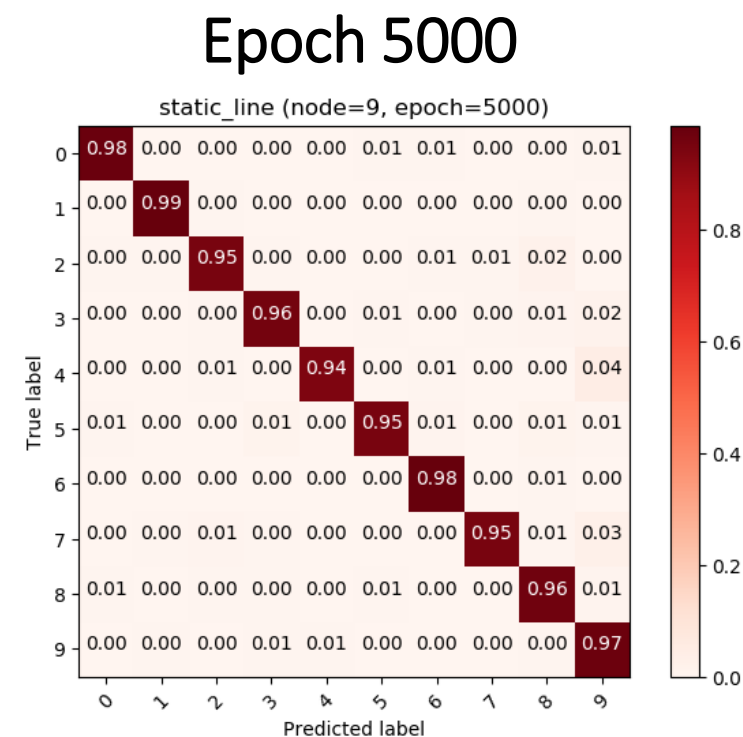
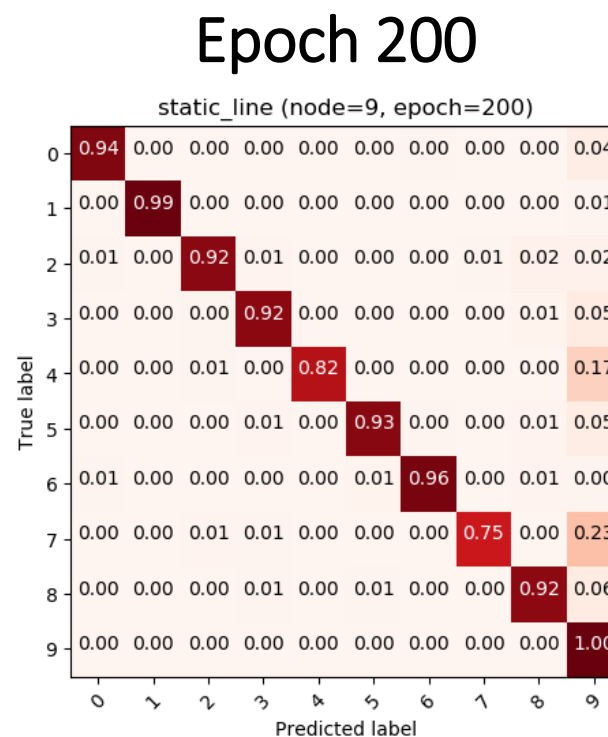
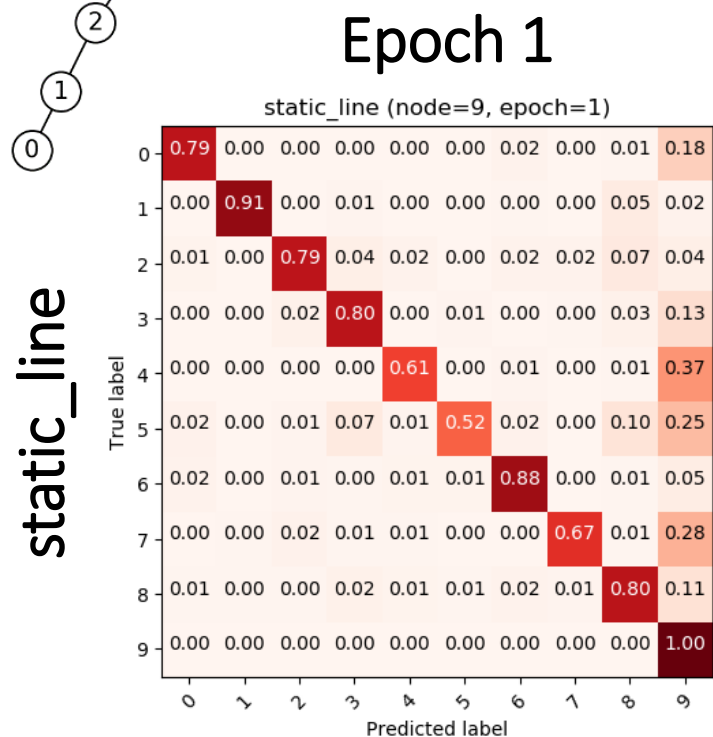
ベンチマーク評価

1. 静的なトポロジーでの検証

- センサネットワークなどで、ノードが静止しているとき、静的なトポロジーとなる
- 4種のトポロジーを想定してWAFLを実行



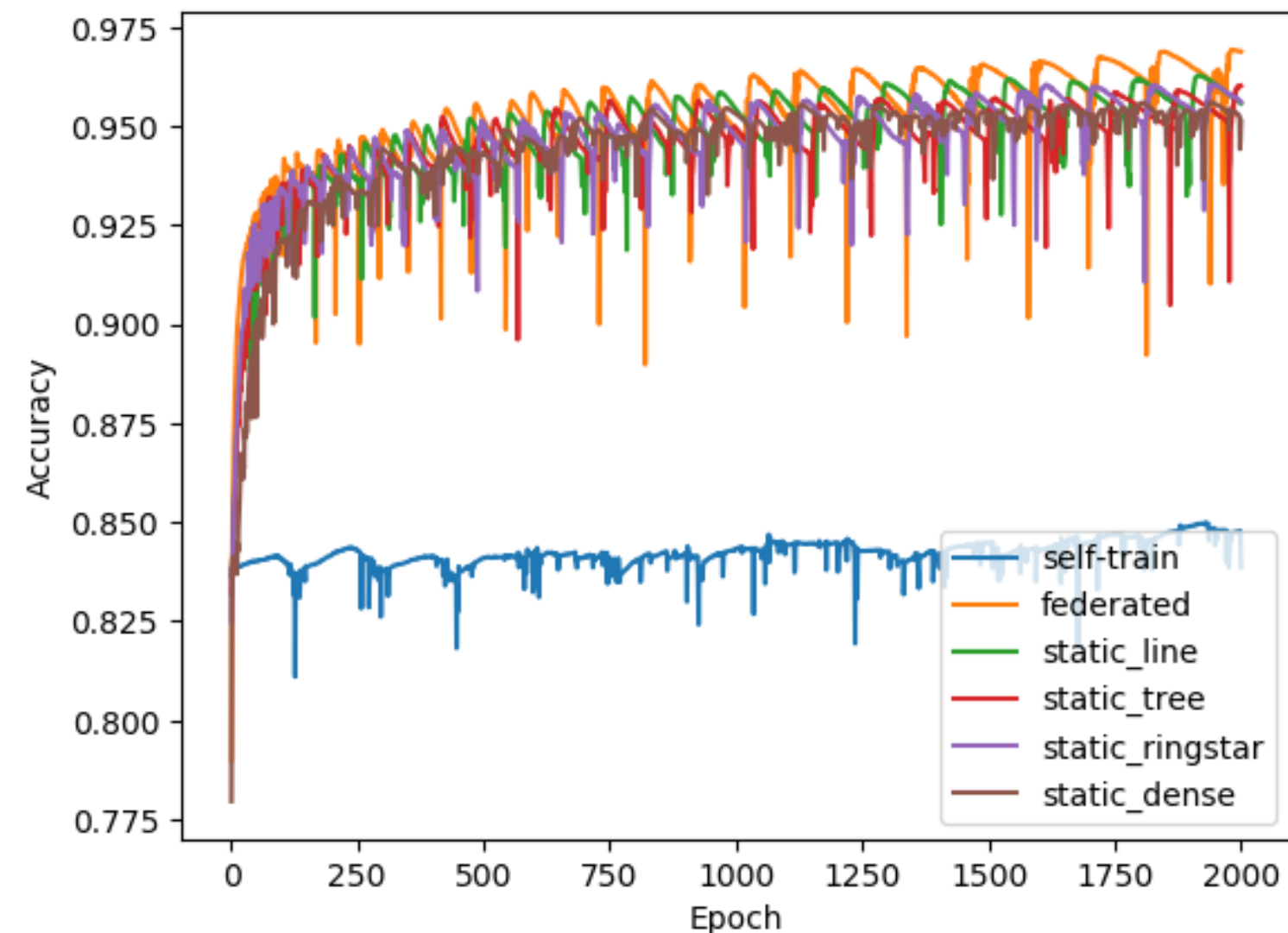
ベンチマーク評価 テストデータに対する混同行列の変化 @Node 9での様子



初め、多くのデータを9と判定してしまっている
例えば、4を9と誤判定。7を9と誤判定など

最終的に誤判定は大きく減少

ベンチマーク評価 精度の推移 (静的トポロジーの場合)



WAFLは
従来のFederated Learningと
同等の性能を出している



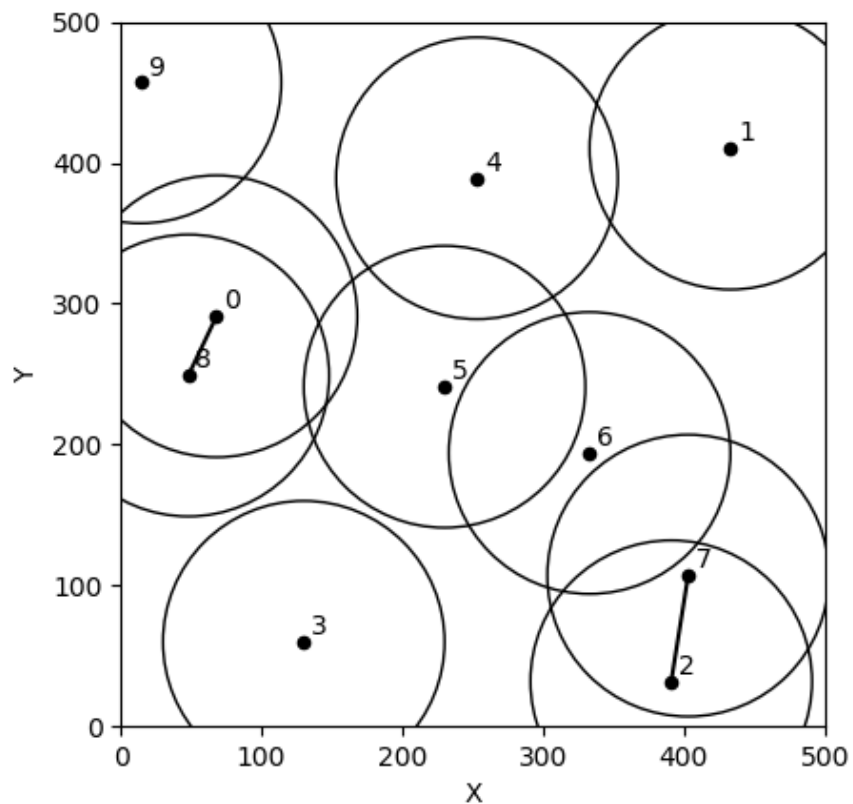
10.0-11.5%の向上

Self-Trainの精度

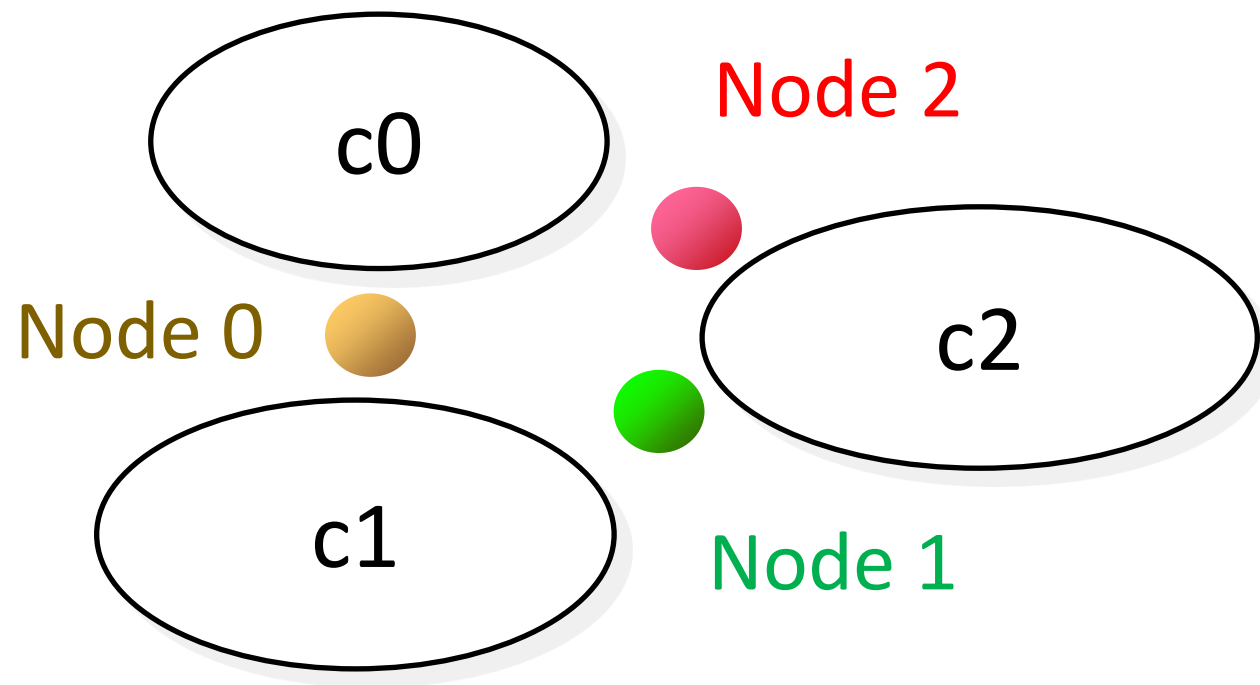
(*) 精度の振動は Optimizer Adamによるもの

ベンチマーク評価

2. 動的なパターンでの検証



Random Waypoint Mobility
(RWP)



Community-Structured Environment
(CSE)

ベンチマーク評価 テストデータに対する混同行列の変化

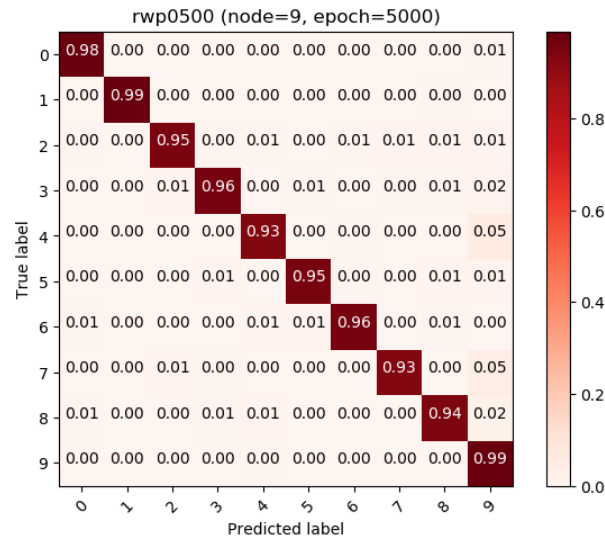
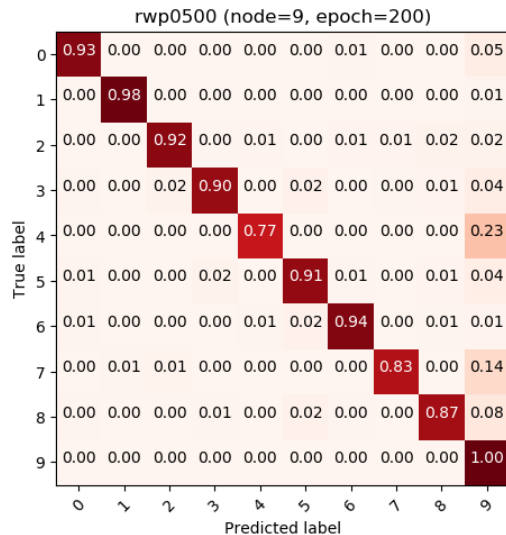
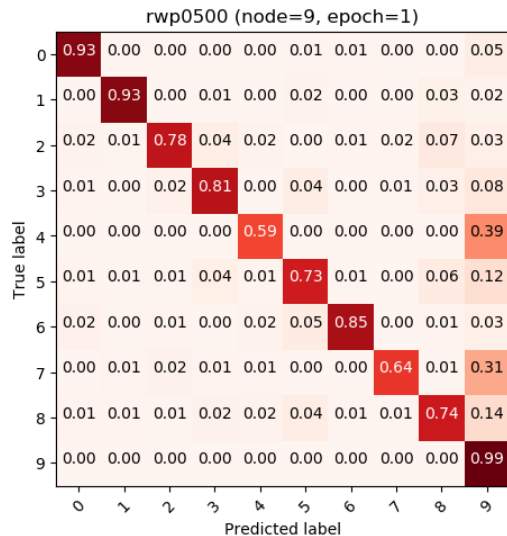
RWP, CSEの場合
Node 9 での様子

Epoch 1

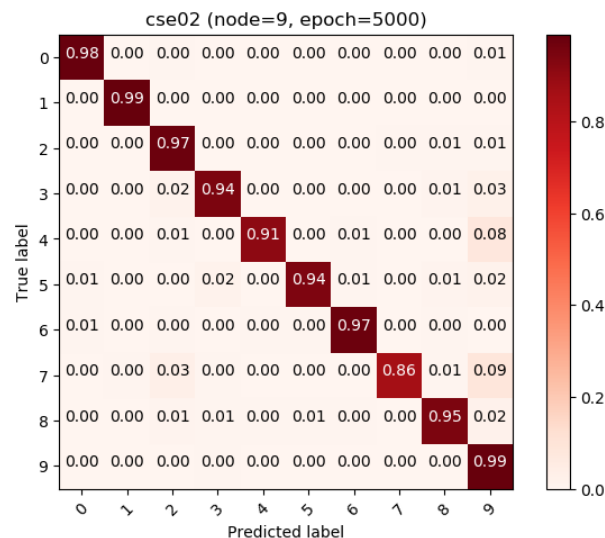
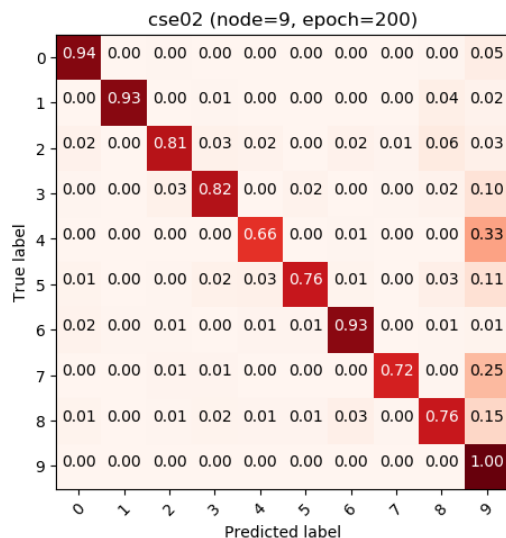
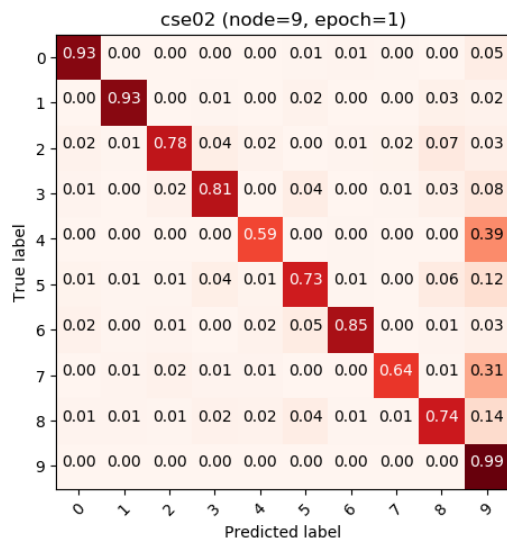
Epoch 200

Epoch 5000

rwp0500



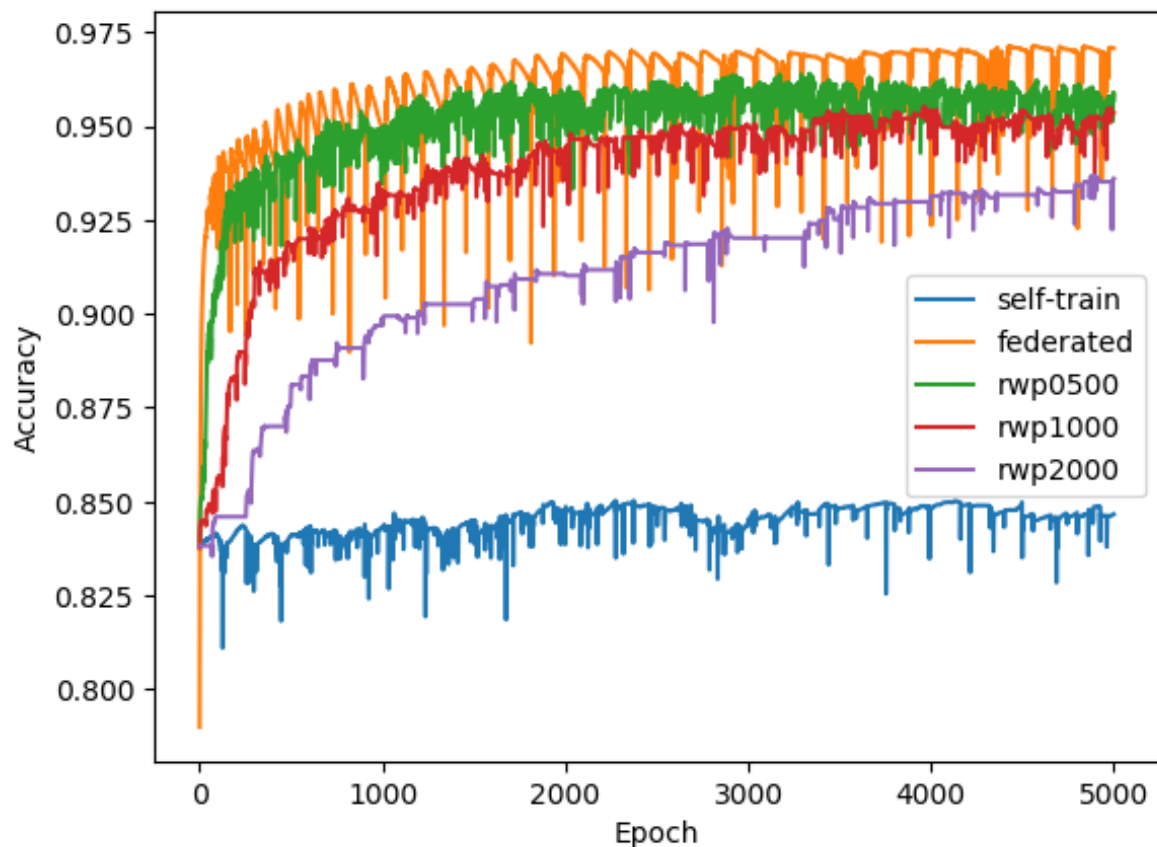
cse02



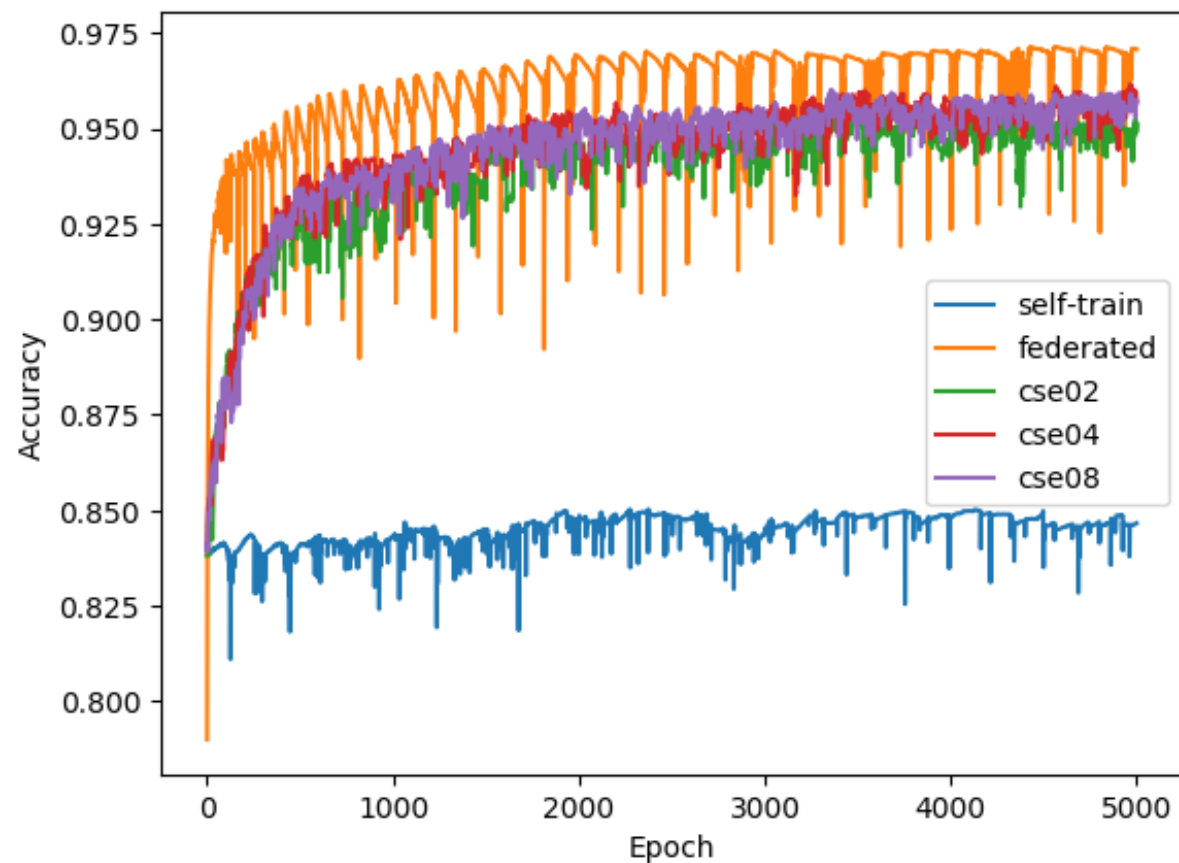
staticの場合と
同様の傾向が
読み取れる

ベンチマーク評価

精度の推移 (動的パターンの場合)



(a) RWPの場合



(b) CSEの場合

(*) 精度の振動は Optimizer Adamによるもの

ベンチマーク評価: 精度の一覧

Training Method	Accuracy [%]	Diff(FL) [%]	Diff(Self-Train) [%]
WAFL (static_line)	96.337±0.595	-0.419	11.674
WAFL (static_tree)	95.906±1.196	-0.850	11.243
WAFL (static_ringstar)	96.110±0.745	-0.646	11.447
WAFL (static_dense)	95.643±1.036	-1.113	10.98
WAFL (dynamic_rwp0500)	95.389±1.218	-1.367	10.726
WAFL (dynamic_rwp1000)	95.230±0.981	-1.526	10.567
WAFL (dynamic_rwp2000)	93.659±0.948	-3.097	8.996
WAFL (dynamic_cse02)	94.840±0.833	-1.916	10.177
WAFL (dynamic_cse04)	95.454±1.288	-1.302	10.791
WAFL (dynamic_cse08)	95.560±0.946	-1.196	10.897
Federated Learning	96.756±0.928	-	-
Self-Training	84.663±1.285	-	-

接触機会が
多い方が
若干高い

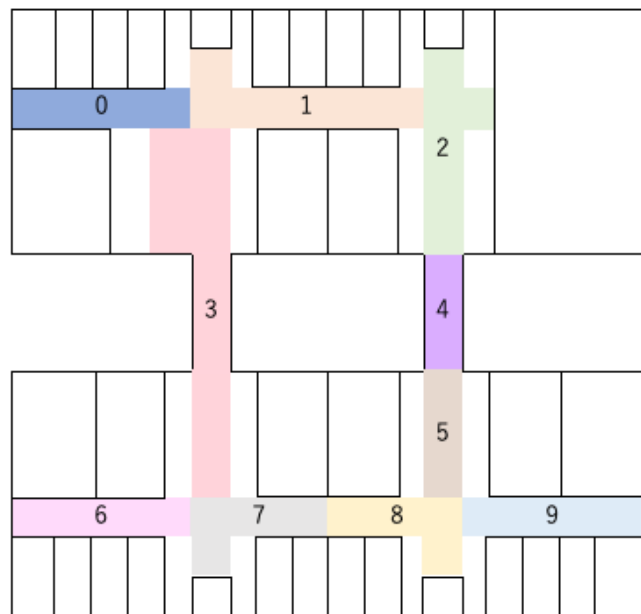
(学習は途中であり
Epochを進めれば
さらに進むことを
示唆)

Table of Contents

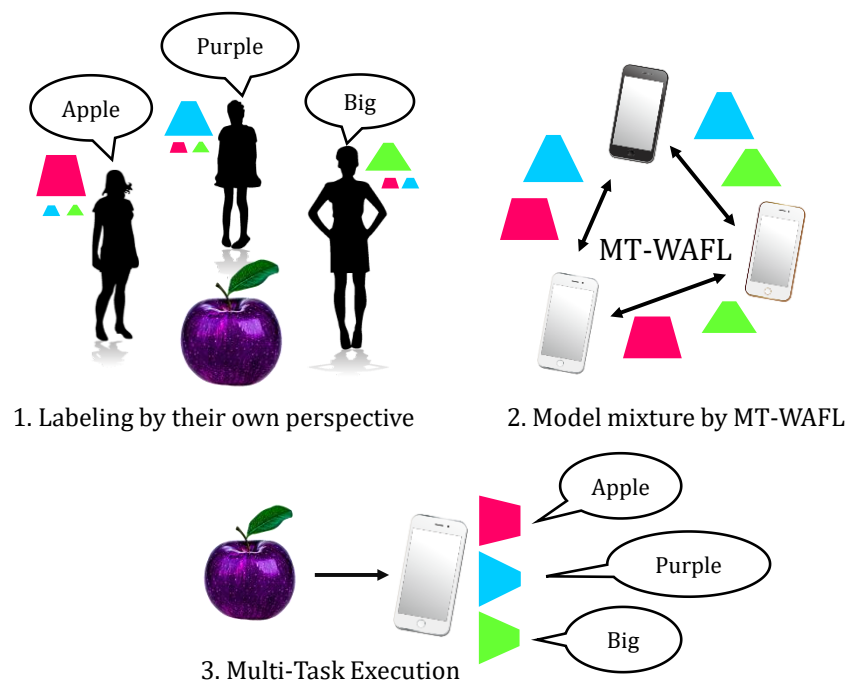
- 機械学習の分散化
- Wireless Ad Hoc Federated Learning (WAFL)
 - 特徴
 - 原理
 - 動作の確認(ベンチマーク評価)
 - 応用と発展
- 今後の見通し
- まとめ

無線アドホック連合学習(WAFL)の応用と発展

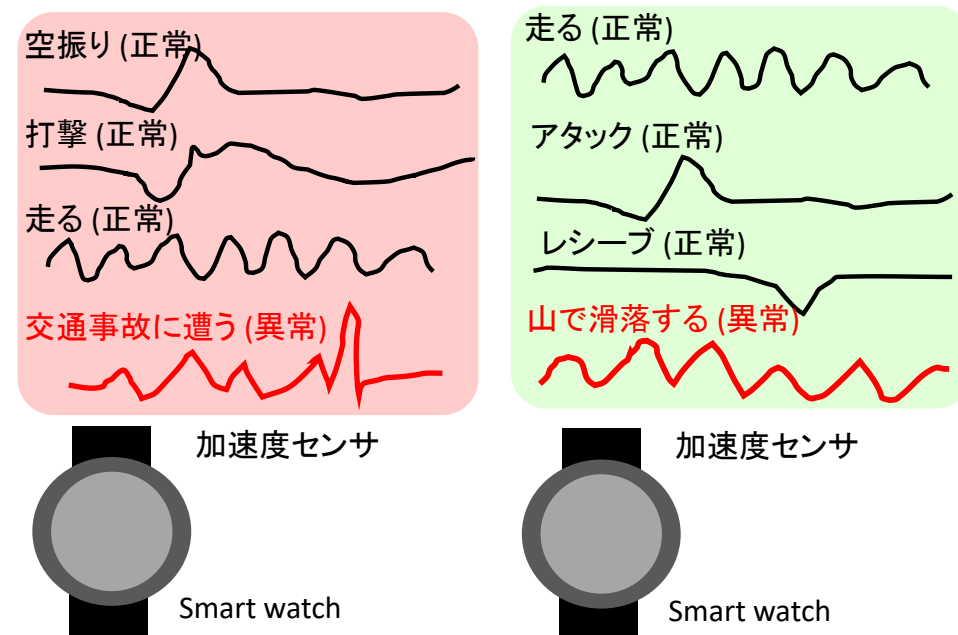
WAFL: Wireless Ad Hoc Federated Learning



(1) 自己位置推定



(2) ブレーン・ストーマ



(3) 異常検知

WAFLによる自己位置推定モデルの協調学習

1. 背景: Wi-Fi の AP の RSSI 値の表から自己位置推定が可能

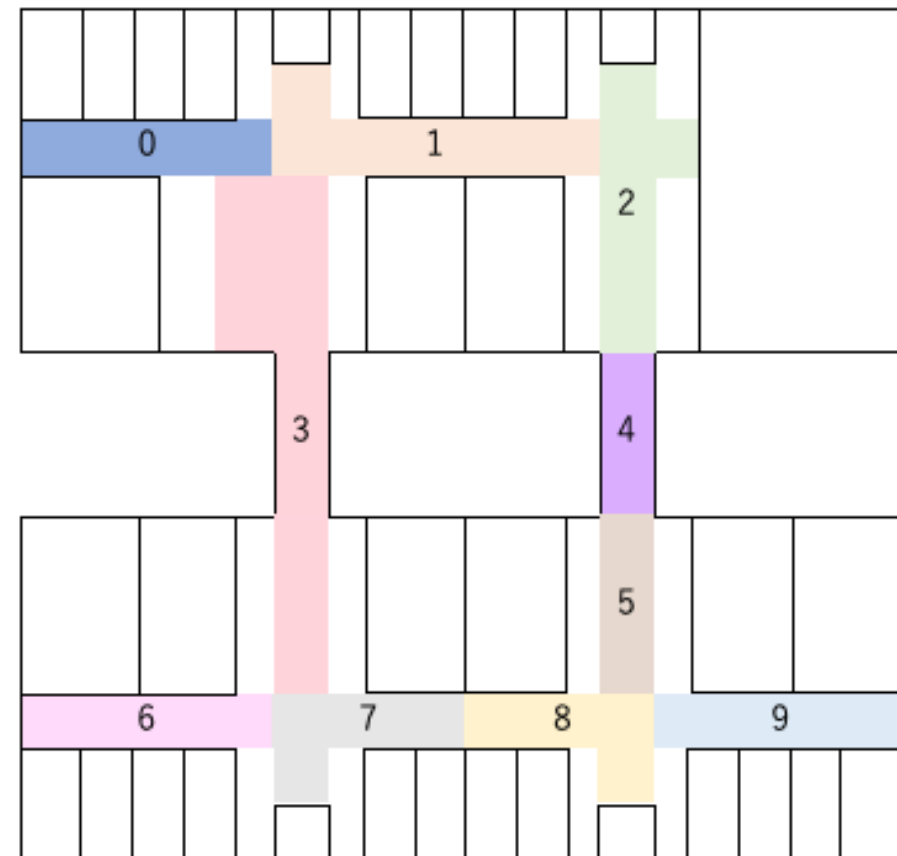
	SSID	BSSID	RSSI	CHANNEL	HT	CC	SECURITY (auth/unicast/group)
	sanshiro	74:88:bb:02:d7:c9	-90	100,+1	Y	JP	WPA2(PSK/AES/AES)
	91A2-A-WPA3	90:96:f3:c4:15:9a	-86	100	Y	JP	WPA2(SAE/AES/AES)
	91A2-A	90:96:f3:c4:15:99	-85	100	Y	JP	WPA2(PSK/AES/AES)
	elab-internal	04:fe:7f:a1:2d:43	-83	52	Y	---	WPA2(PSK/AES/AES)
	Buffalo-A-A4C0	c6:36:c0:00:a4:ca	-82	64	Y	JP	WPA2(PSK/AES/AES)

2. 従来手法の問題点:

すべての地点のRSSI表を作成して集める必要あり

3. WAFLでは、

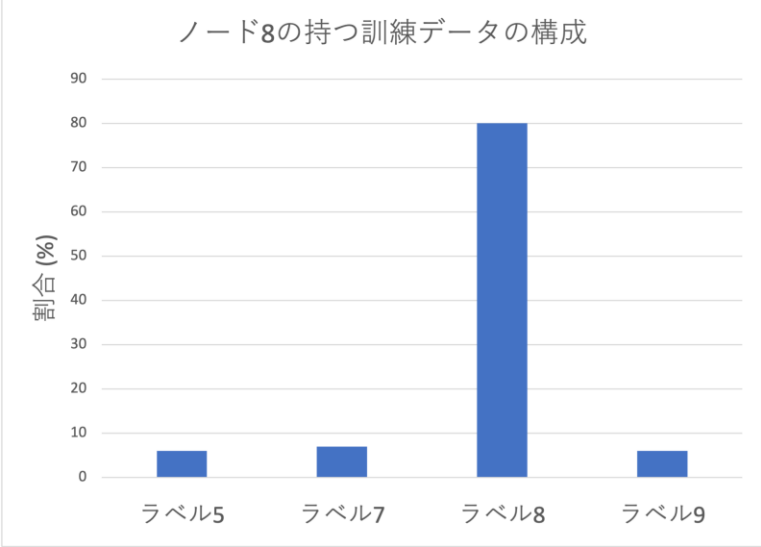
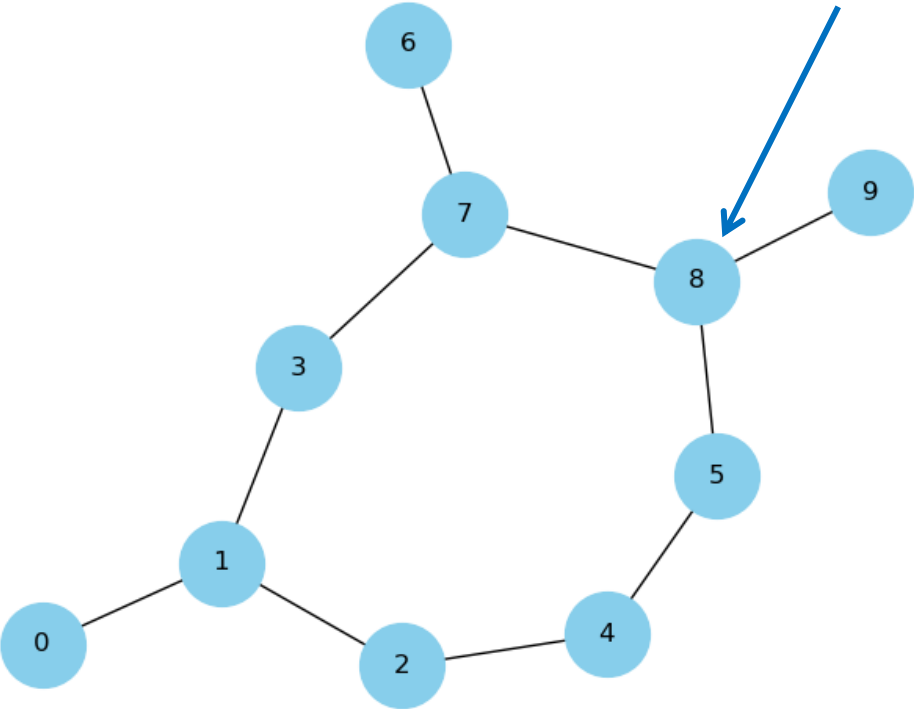
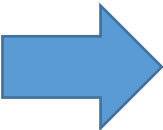
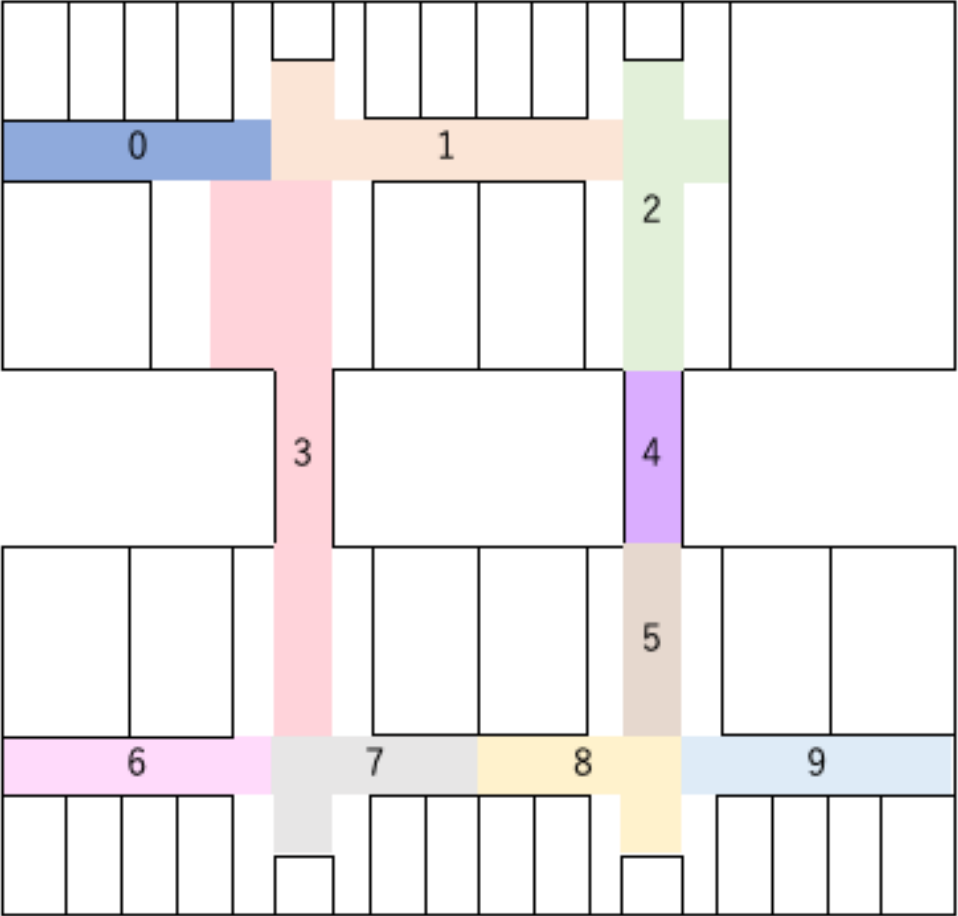
- (a) あるノードは 0 の近くのみを計測して学習
- (b) 別のノードは 9 の近くのみを計測して学習
- (c) ノード間でモデルを交換することによって、全体を計測したことになる



4. どこに行っても、自己位置を推定できるようになる

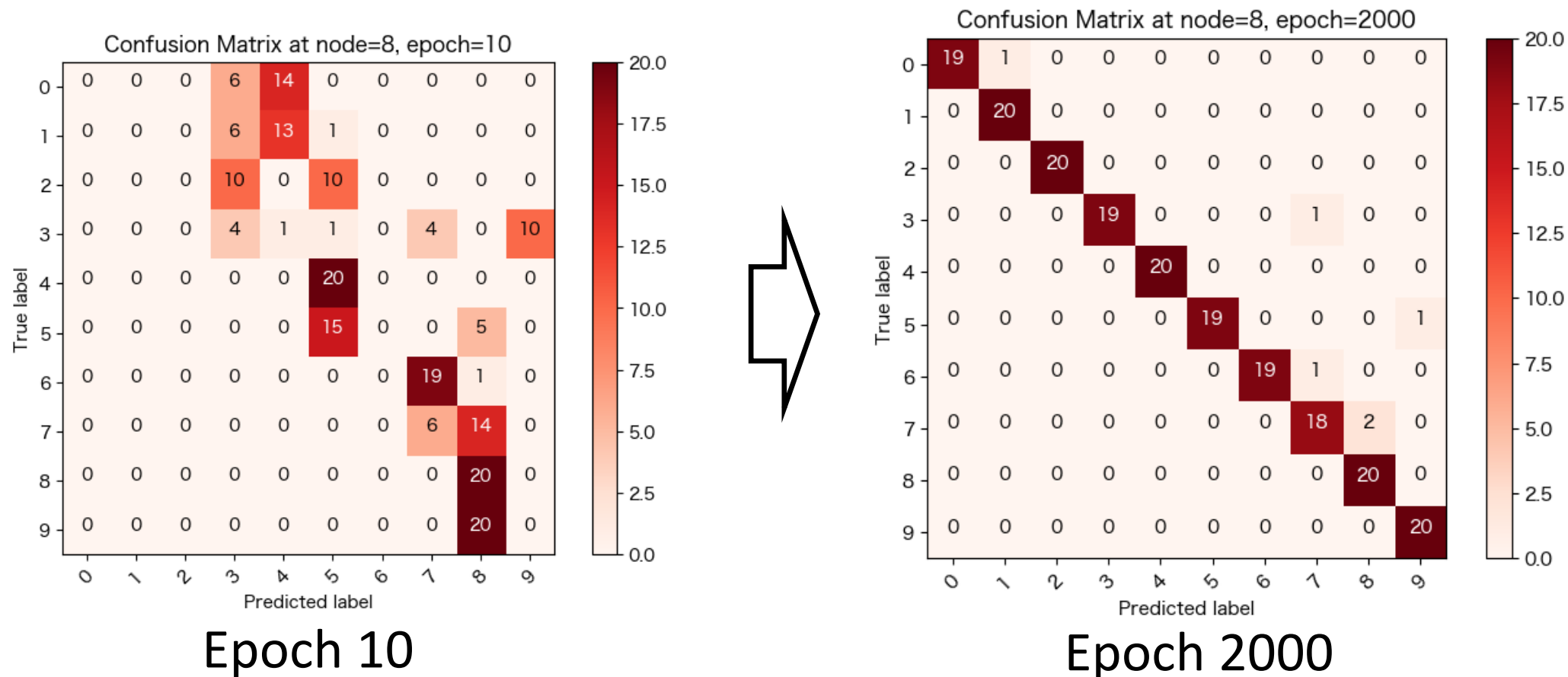
実験設定

各位置に1つずつノードが存在、
近隣ノードとの通信が行われる
状況を想定



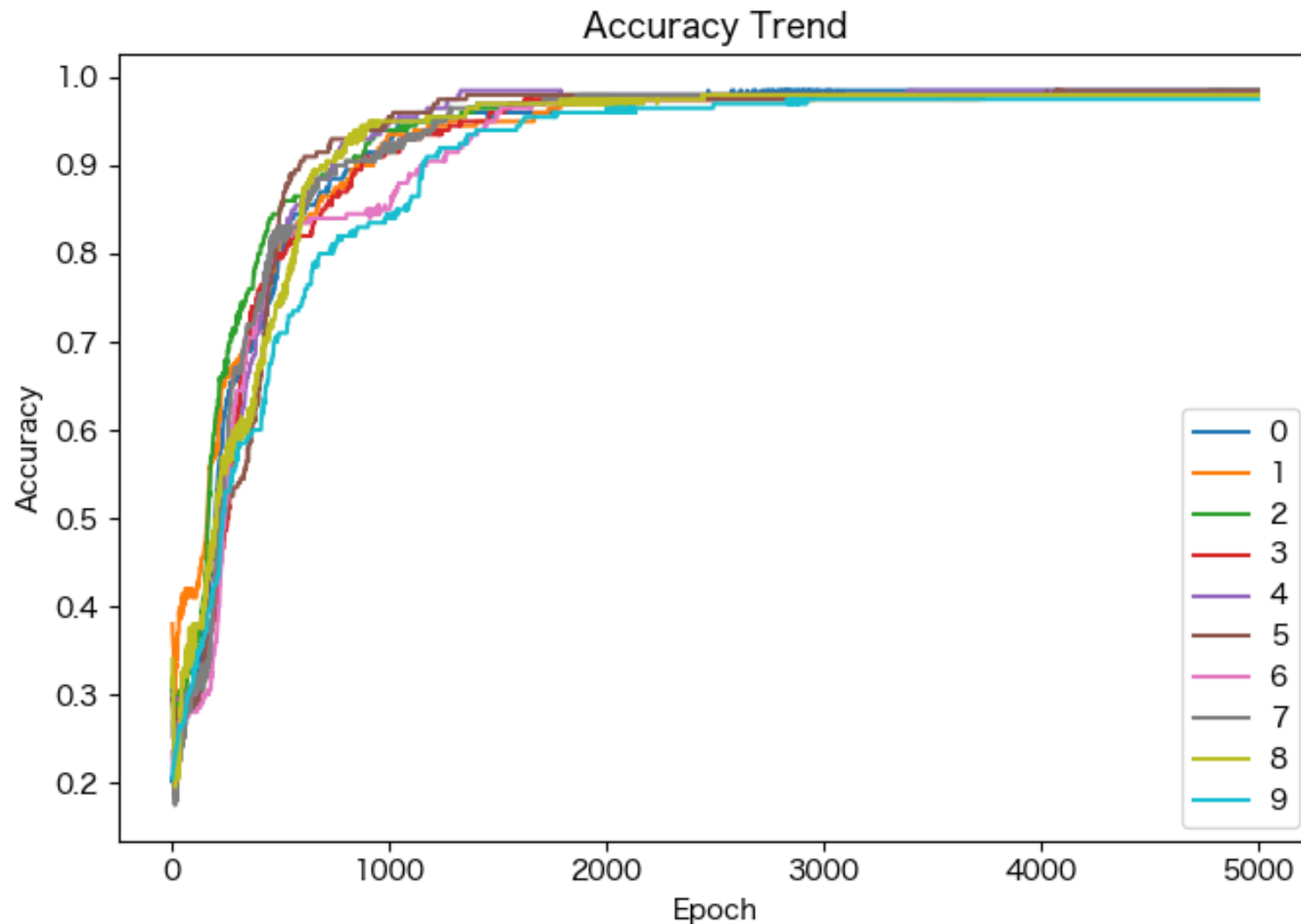
ノード間の接続関係

実験結果(1/2): ノード8 の 位置推定精度の変化



開始時には、実際の位置に対して、誤った場所を多く推定していたが、学習が進むにつれ、正しい場所を推定できるようになった。

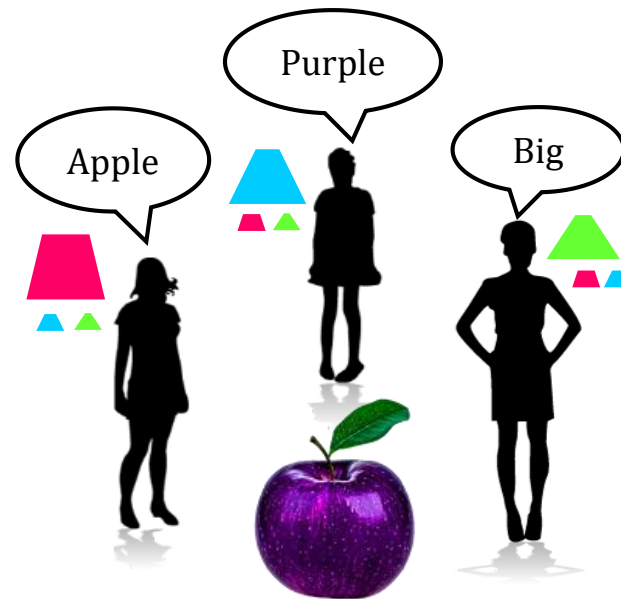
実験結果(2/2): 各ノードでの位置推定精度の変化



全てのノードで95%以上の精度が得られた

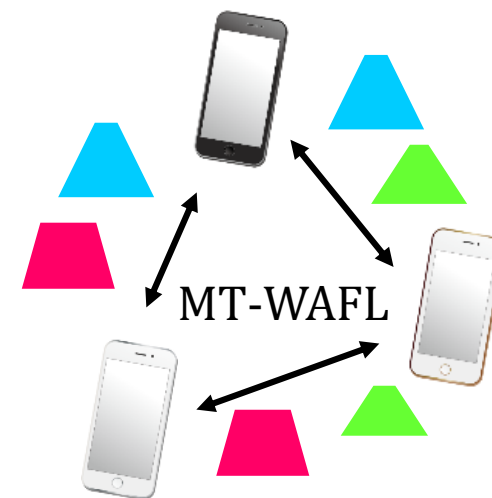
WAFLによる多角的認知の統合

1. 人々は提示されたものに対して異なる観点の反応をする

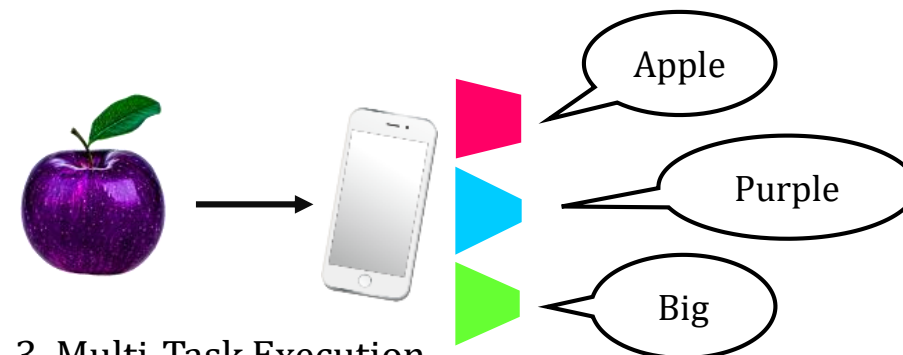


1. Labeling by their own perspective

2. これらの反応を統合すれば
そのチームの考え方に適合した
反応モデルが作られる



2. Model mixture by MT-WAFL



3. Multi-Task Execution

実験設定

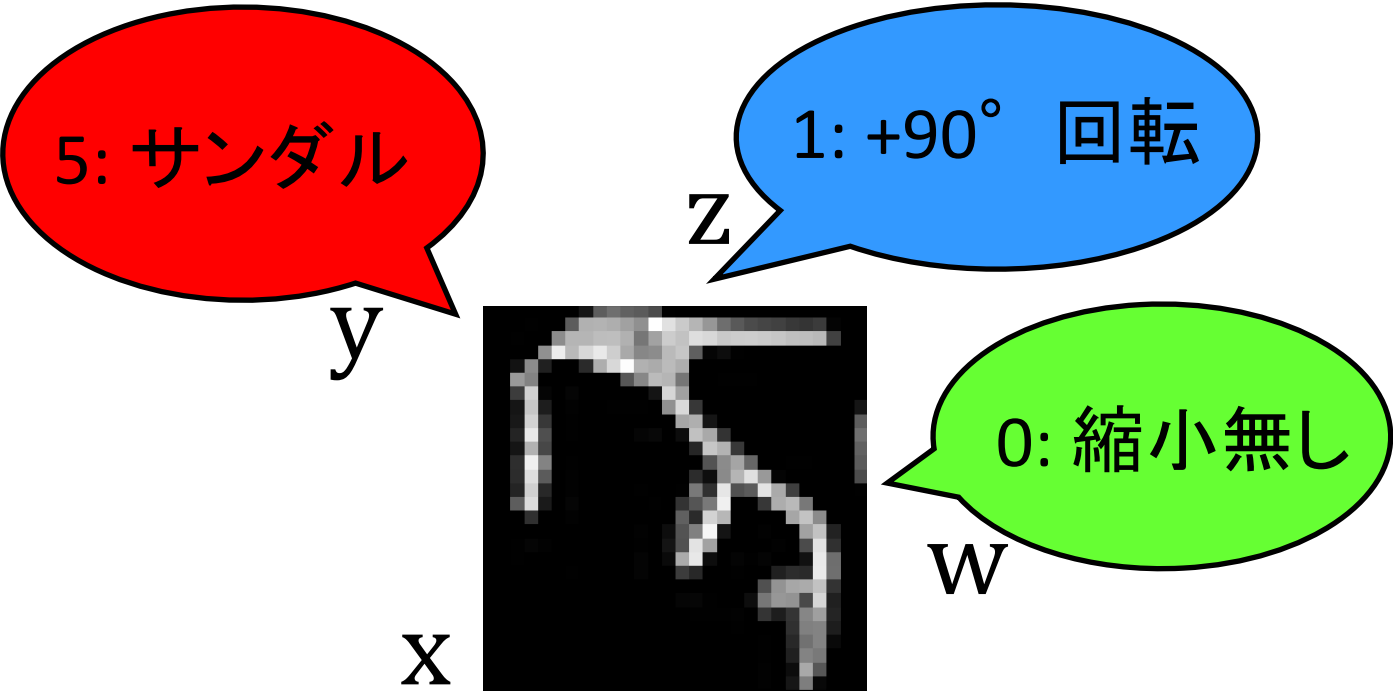
X: Fashion-MNIST画像を Rotation および Resize した画像

Y: 画像のクラス(0, 1, ..., 9)

Z: 画像のRotation (0, 1, 2, 3)

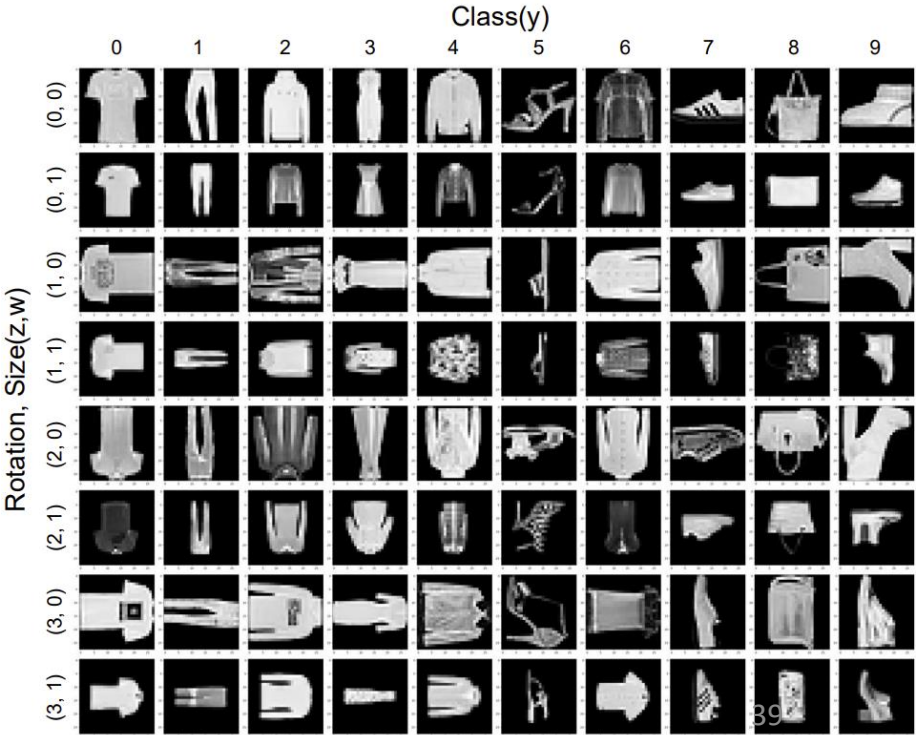
W: 画像のサイズ (0, 1)

rwp0500で実験

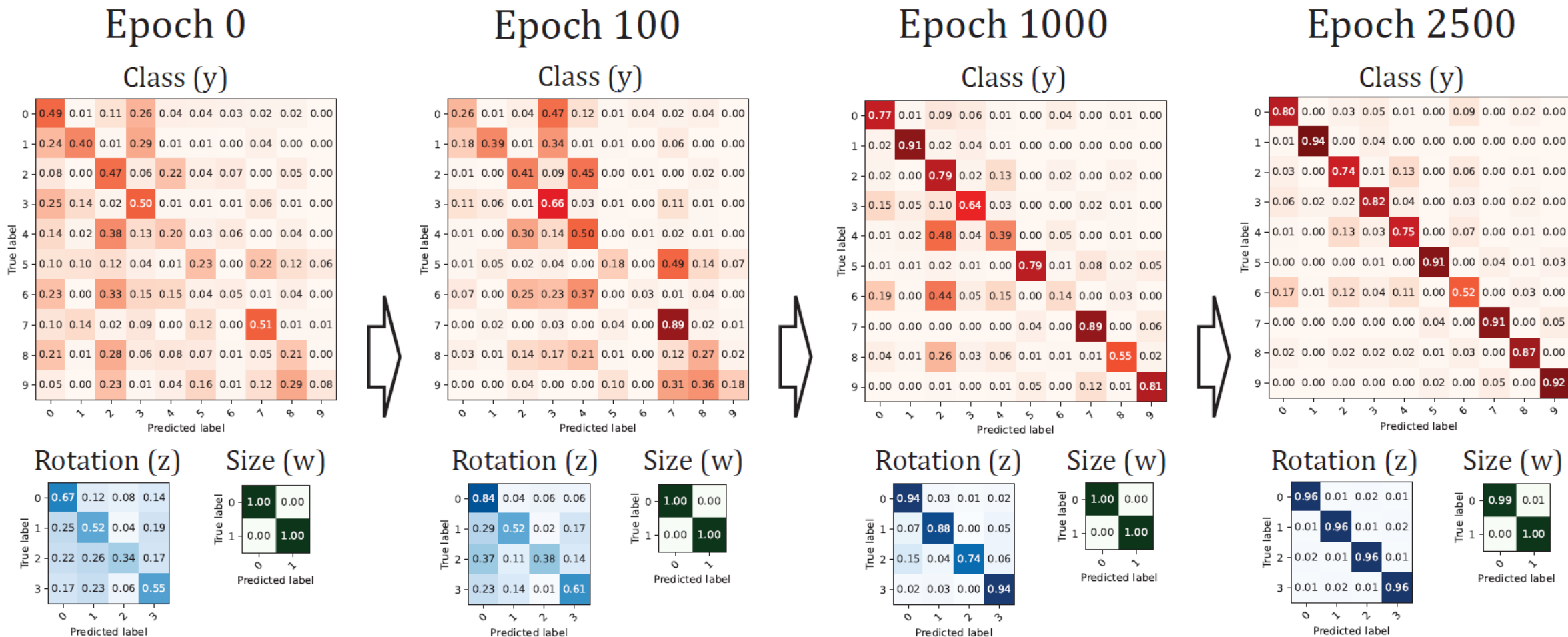


ラベル保有率の分布

	ラベルY	ラベルZ	ラベルW
Node 0-4	99%	1%	1%
Node 5-7	1%	99%	1%
Node 8-9	1%	1%	99%



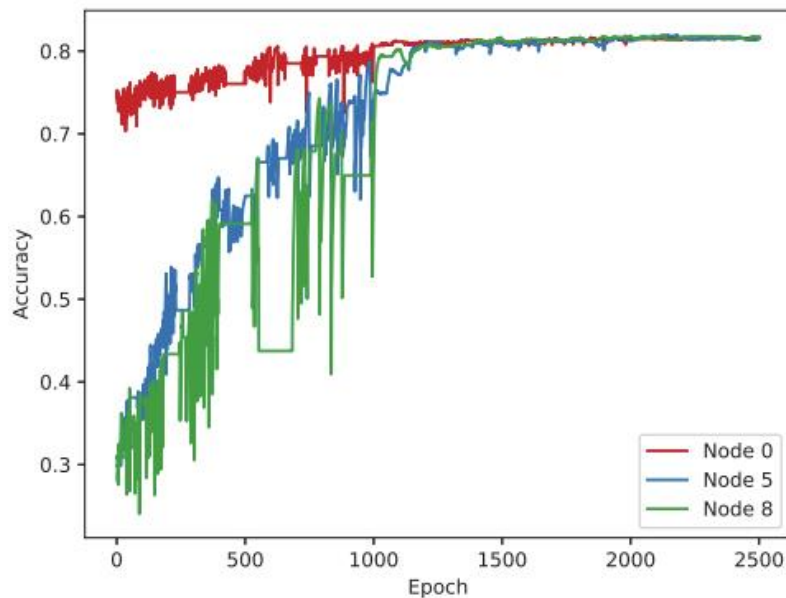
実験結果(1/2): ノード8での認知精度の変化



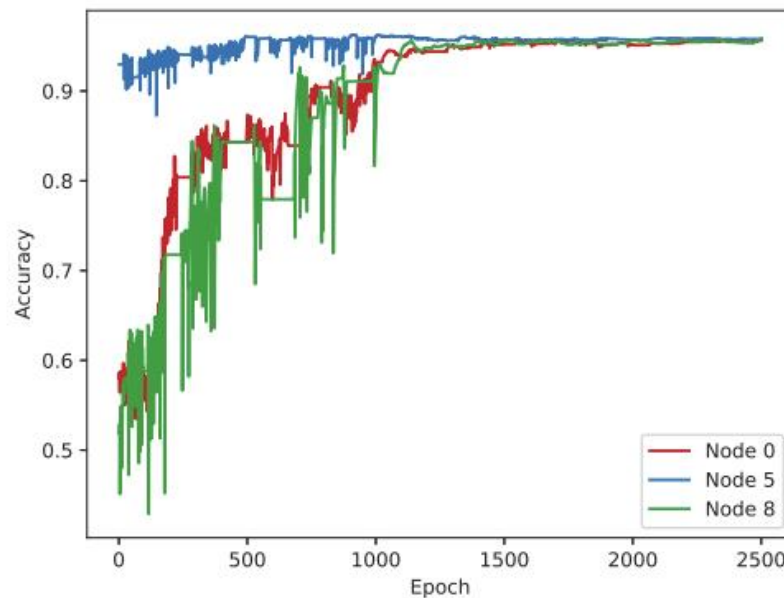
ノード8 は サイズの認知に特化。

Epochが進むにつれて、クラス(Y)や回転(Z)の認知精度も向上

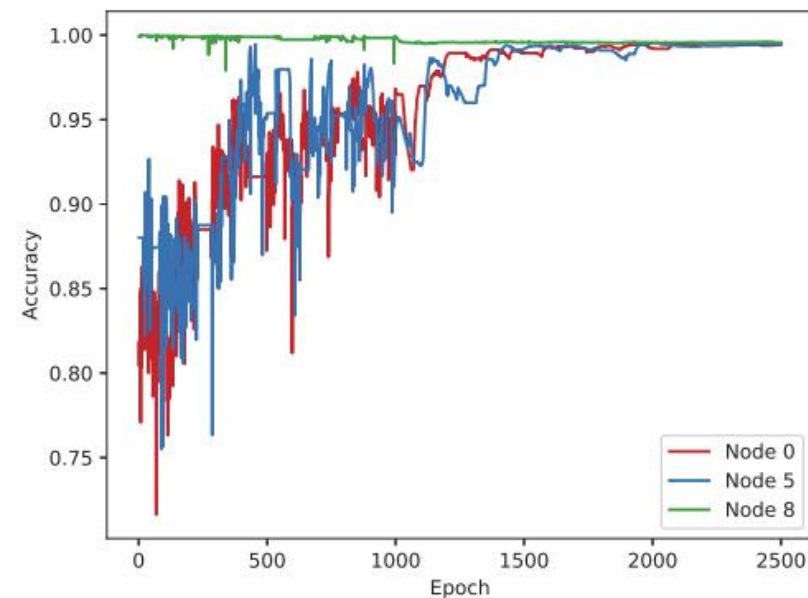
実験結果(2/2): ノード0, 5, 8の認知精度の変化



(a) Test Accuracy for Class (Y)



(b) Test Accuracy for Rotation (Z)



(c) Test Accuracy for Size (W)

ノード0はクラス(Y)の認知に特化。Node 5は回転(Z)の認知に特化。
Epochが進むにつれて、多角的に認知精度が向上。

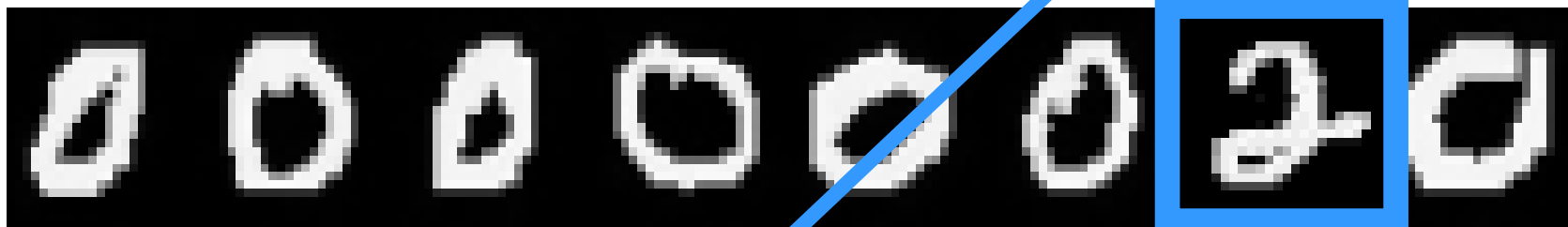
Node 8はサイズ(W)の認知に特化。

異常検知 Non-IIDの場合

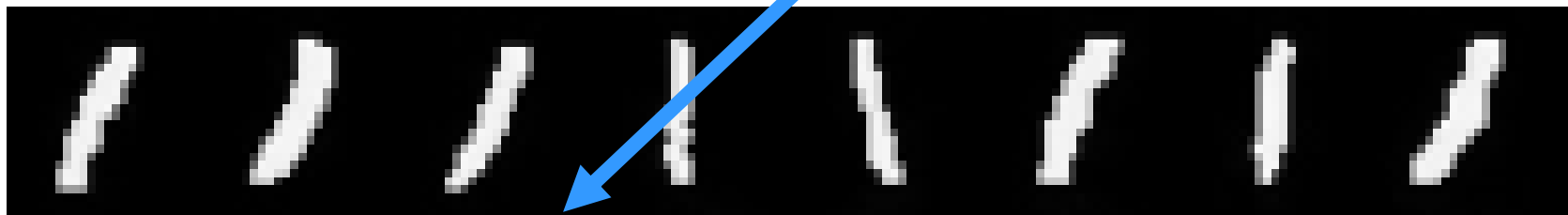
Local Anomaly

(そのノードにとっては珍しいが、全体的に見れば珍しくない)

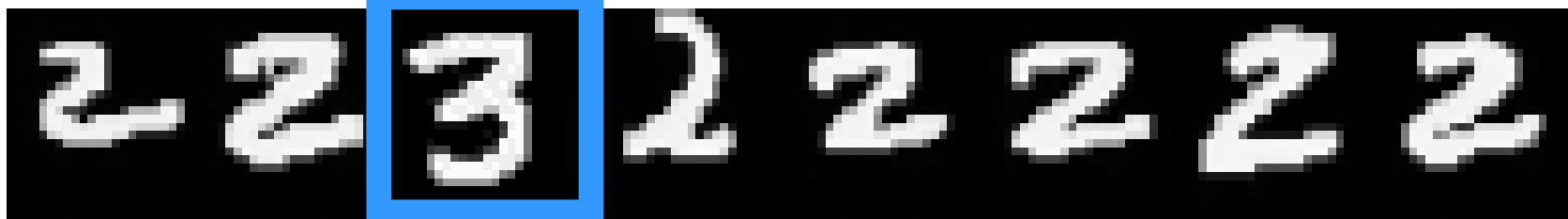
Node 0
の持つデータ



Node 1
の持つデータ



Node 2
の持つデータ



Node 3
の持つデータ

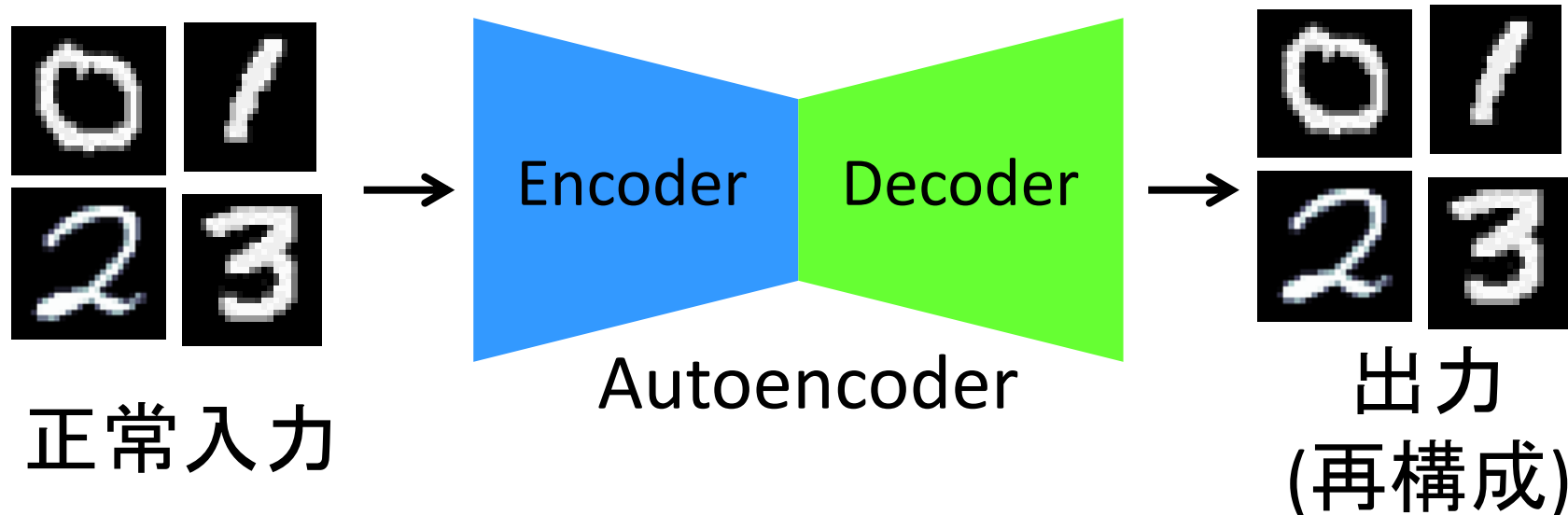


Global Anomaly

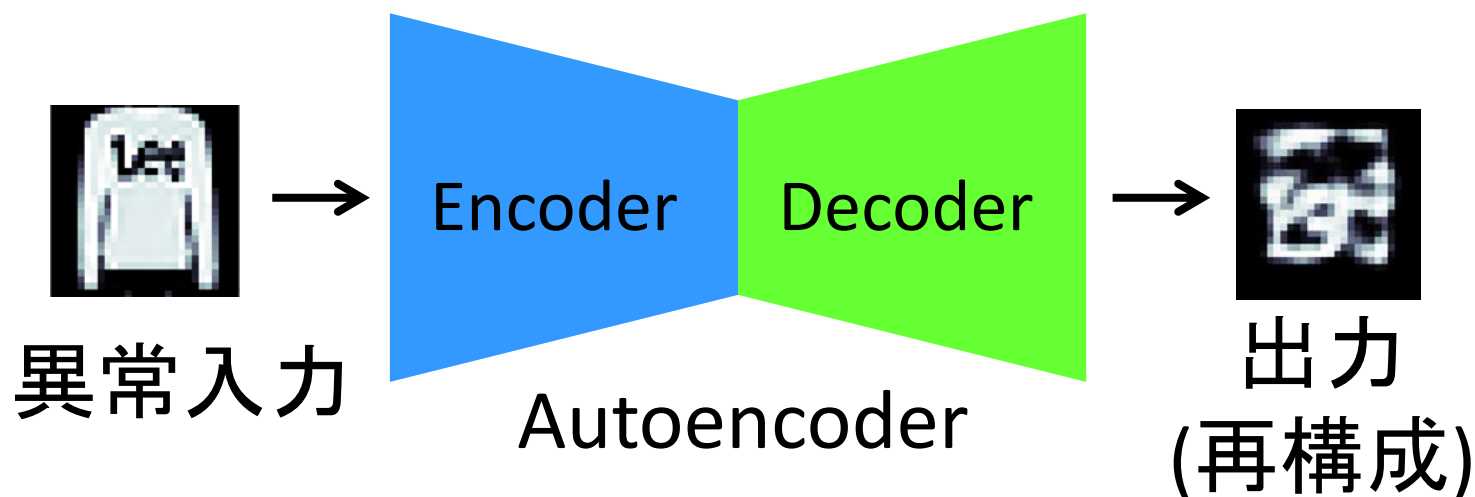
(全体的に見ても珍しいデータ)

検知したいのは Global Anomaly

Global Anomalyを検知するには



「正常データ入力では
精密に再構成する」
よう、ノードのデータ分布
の違いを吸収する学習
させる。



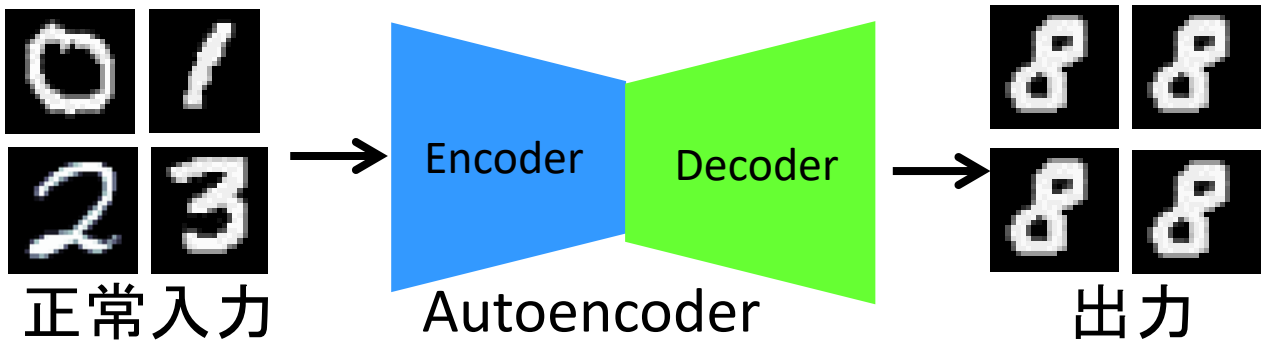
Global Anomalyの入力に
対しては、再構成に失敗する

実験設定

学習データの分布 (99.95% Non-IID)

Node	L0	L1	L2	L3	L4	L5	L6	L7	L8	L9	Total
Node 0	4736	0	0	0	0	0	0	0	0	0	4736
Node 1	0	5418	0	0	2	1	0	0	0	1	5422
Node 2	0	0	4779	0	0	0	0	0	0	0	4479
Node 3	1	0	0	4911	0	0	1	1	0	0	4914
Node 4	0	0	0	1	4733	0	0	0	0	0	4734
Node 5	0	0	0	0	0	4343	1	0	0	0	4344
Node 6	0	0	0	1	0	0	4712	0	0	0	4713
Node 7	1	0	1	0	1	0	2	5046	0	0	5051
Node 8	1	0	0	0	0	0	1	0	4714	0	4716
Node 9	0	0	0	0	0	0	0	0	1	4751	4752

例えば、Node 8単体(WAFLなし)で学習した場合のテストは:

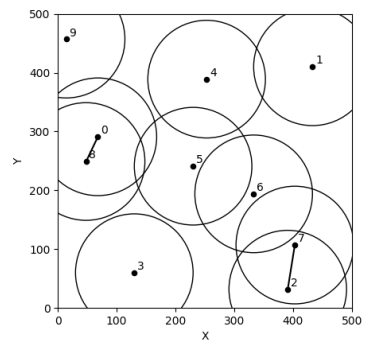


となることが期待される

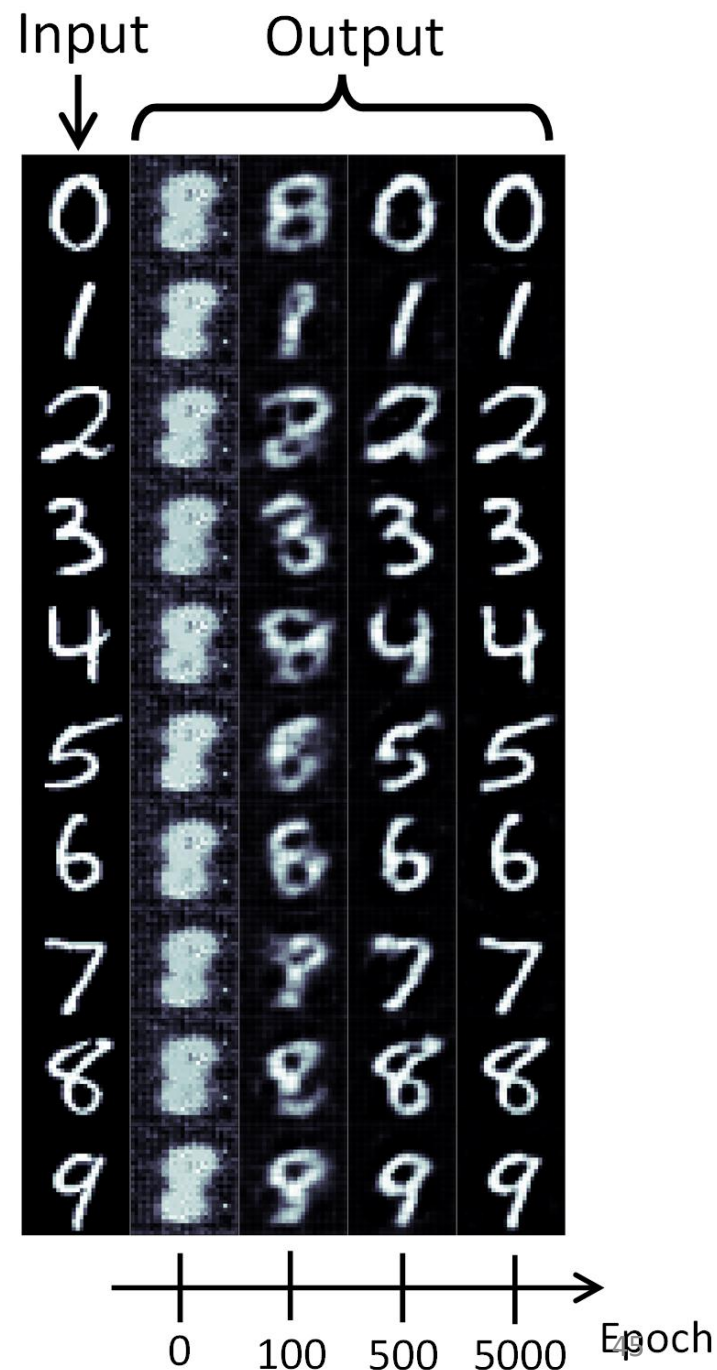
実験結果(1/2):

ノード8でのAutoencoderの再構成の変化

1. Self-Train直後の WAFL (epoch 0) では
0 ~ 9 のどの入力に対しても 8 となっている
2. WAFLの epochが進むにつれて他ノードの
モデルが合成され、
0 ~ 9 のどの入力に対しても
近い結果を出力している。



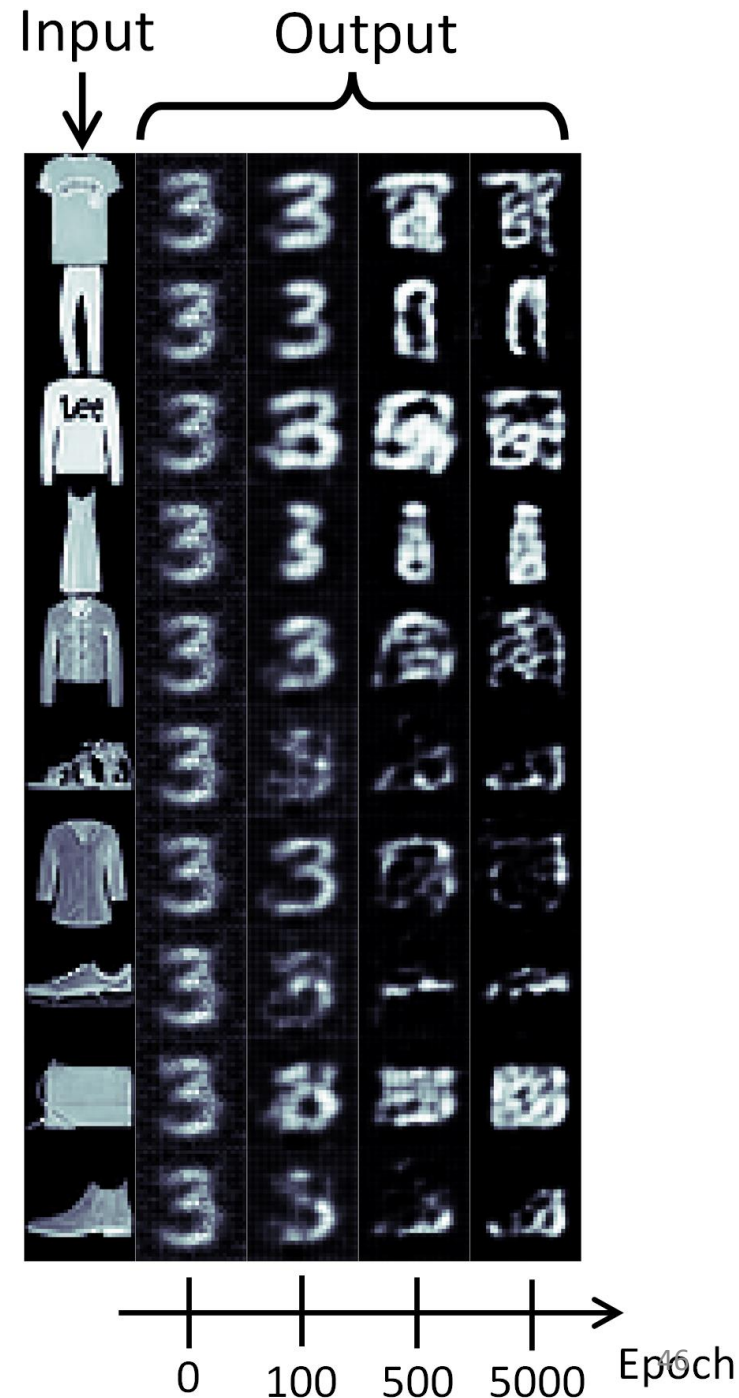
rwp0500で実験



実験結果(2/2):

Autoencoderの再構成の変化(@Node 3)

1. Self-Train直後の WAFL (epoch 0) では
どの異常入力に対しても 3 となっている
2. WAFLの epochが進むにつれて他ノードの
モデルが合成されるが、
どの異常入力に対しても不明な結果を
出力している(期待通り)。



異常検知の性能(@Node 0)

(*) 今回は解説していない Anomaly Score Threshold に関する研究を適用した結果。

1. Self-Train直後では、
0 以外のサンプルはほぼ異常と判断
2. Epoch 100では 異常データも正常と判断
3. Epoch 1000以降では正常 or 異常を
正しく捉えられるようになっている

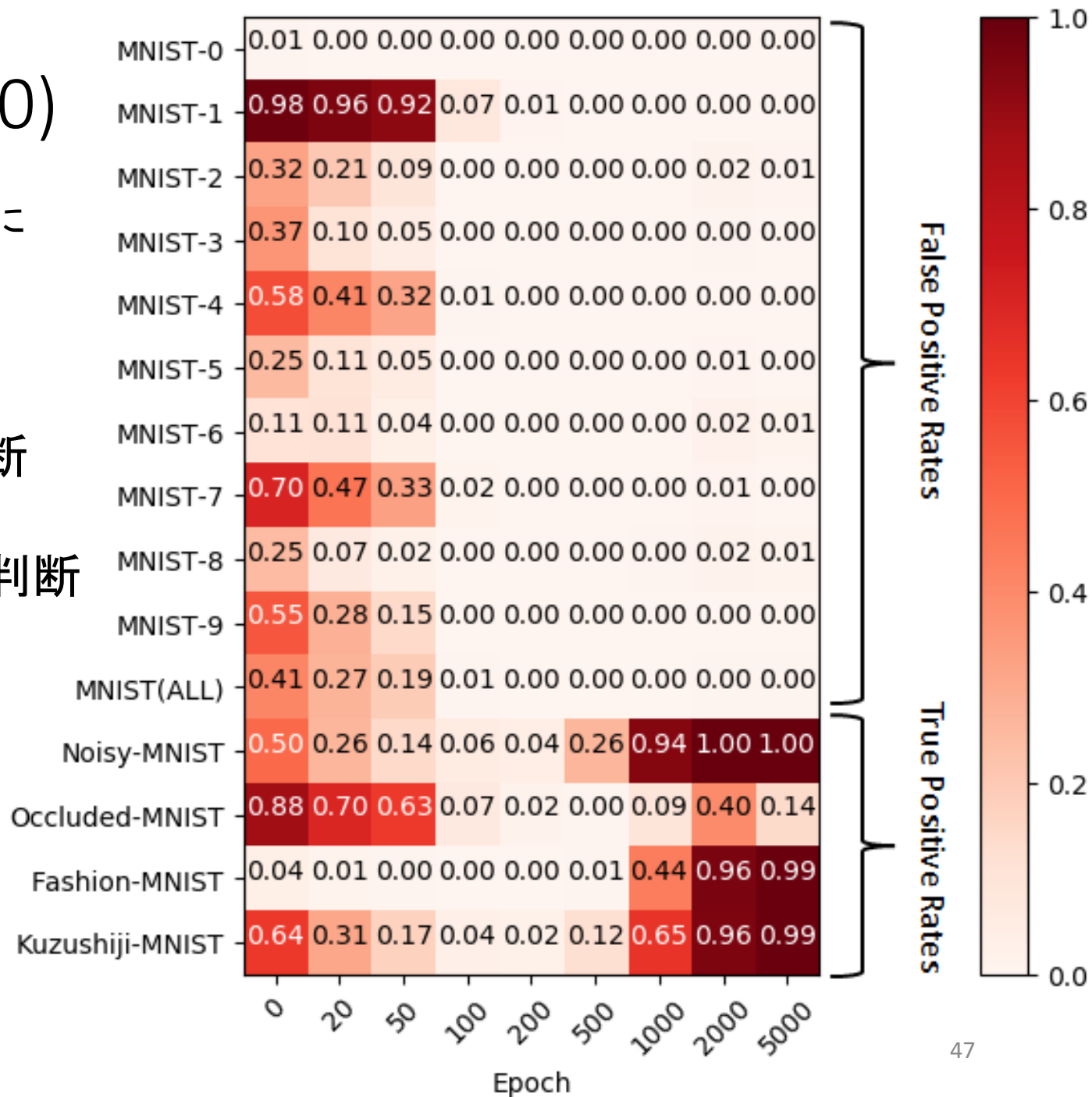


Table of Contents

- 機械学習の分散化
- Wireless Ad Hoc Federated Learning (WAFL)
 - 特徴
 - 原理
 - 動作の確認(ベンチマーク評価)
 - 応用と発展
- 今後の見通し
- まとめ

Open Research Questions (Open Issues)

- MNISTはシンプルなデータセット・・・
 - 一般的な画像/映像についてはどうなの？
 - 文章についてはどうなの？
 - 音声についてはどうなの？
 - IoT や 産業界のデータについてはどうなの？
 - ⇒ センサとセットにできないか？
 - コンピュータなどのログについてはどうなの？
 - ⇒ セキュリティ監視ができないか？
- 生成 (GANなど) についてはどうなの？
- Multi-Domain Adaptation (Out-of-Distribution) についてはどうなの？
- WAFLのセキュリティについてはどうなの？
- 実装はどうなるの？
- プロトコルはどうなるの？
- 運用はどうなるの？

まとめ

- 完全な自律分散型の協調的機械学習
 - Wireless Ad Hoc Federated Learning (WAFL) の紹介
 - WAFLの特徴
 - 中央集権的な機構が無く、すべてのノードは対等
 - プロトコルさえ決まれば、マルチベンダー・システムとして動作しうる
 - 現在の研究ステージ
 - 純粋なベンチマーク評価
 - 応用を意識したベンチマーク的な評価
 - いずれも Promising な結果を出している
 - 今後の研究課題
 - 多様なアプリケーションへの展開・評価
 - 実装、運用、プロトコルの設計、標準化
- } 一緒に活動できる仲間を探しています。