Nバージョン機械学習分類システムによる分類結果の正確性と安全性評価

町田 文雄
筑波大学 システム情報系 准教授

REAJ第30回春季信頼性シンポジウム

はじめに

- 機械学習システム

  機械学習の理論やアルゴリズムの話ではありません.

- 機械学習システムの構成法

  複数の機械学習モデルと複数の入力を組み合わせたシステムの性質について述べます.

- 発表する結果

  画像分類タスクの評価により，必ずしも冗長度の高い構成が高信頼になるわけではないという例を示します.

目次

- 背景
- Nバージョン機械学習システムの構成
- 評価指標
- 手書き文字分類による評価
- おわりに

機械学習の応用と品質管理

- 機械学習を使ったシステムの産業応用が広がる

自動運転車

ヘルスケア

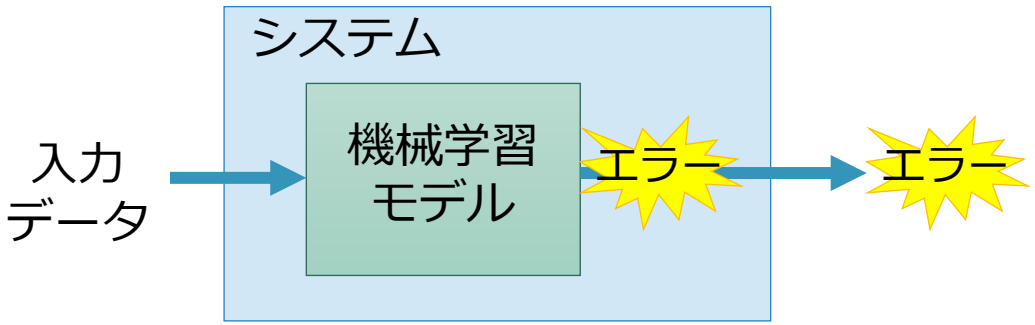
産業ロボット

- 品質管理の問題
  - 機械学習による推論結果は不確実（100%正しいということはない）

Nバージョン機械学習システム

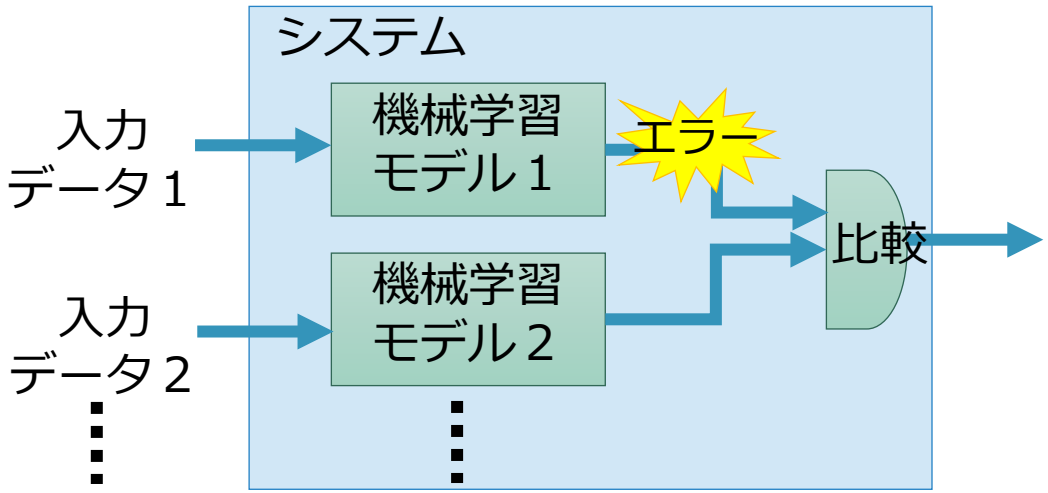
- 機械学習の推論を冗長化してエラー出力を抑える

単一の機械学習モデルを利用する場合



システムの外にエラーがそのまま出てしまう

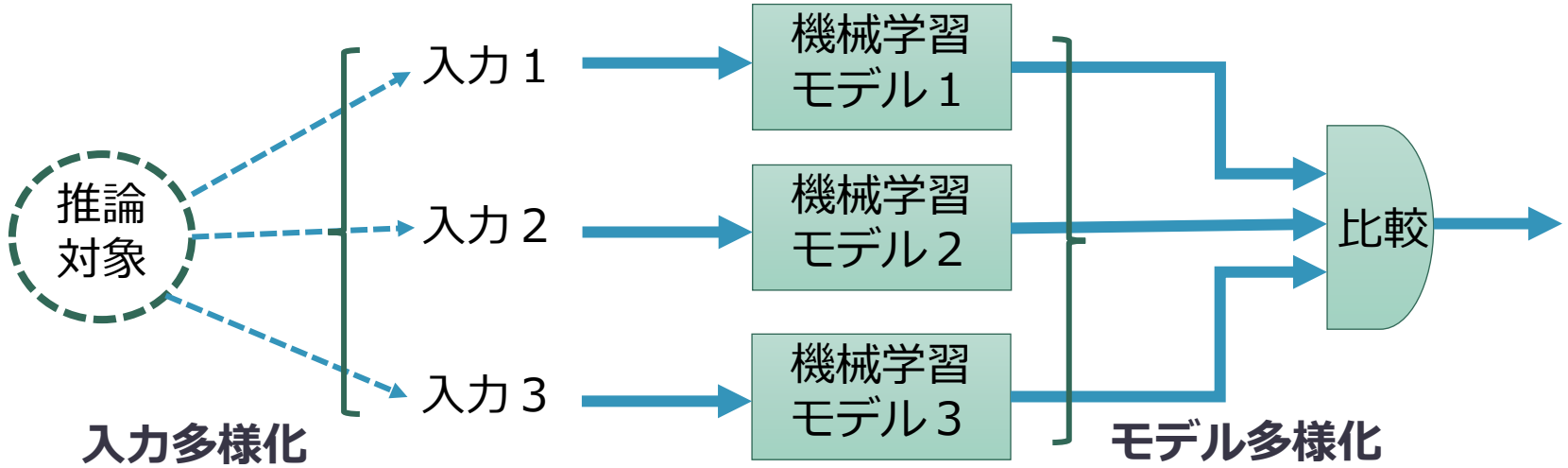
Nバージョン機械学習システムの場合



複数の結果を比較してシステムのエラー出力を抑える

モデルの多様化と入力の多様化

- 複数のモデルが同時にエラーを出力しないように
- モデル多様化
  - 異なる機械学習アルゴリズムや学習データを使ってモデルを作成する
- 入力多様化
  - 同じ推論対象に対する異なる入力データを利用する



本研究の対象

- 機械学習による分類システム
- 前提条件

入力データソース：同一の対象に対して2つの独立したセンサーから
2つの異なる入力データを得る.
機械学習モデル　：同一の分類タスクに対して独立に学習した
2つの機械学習分類モデルを利用できる.
決定器のルール　：全てのモジュールの出力結果が**一致する場合**にのみ
その結果を出力する.
一つでも不一致がある場合は出力しない.

機械学習モジュール

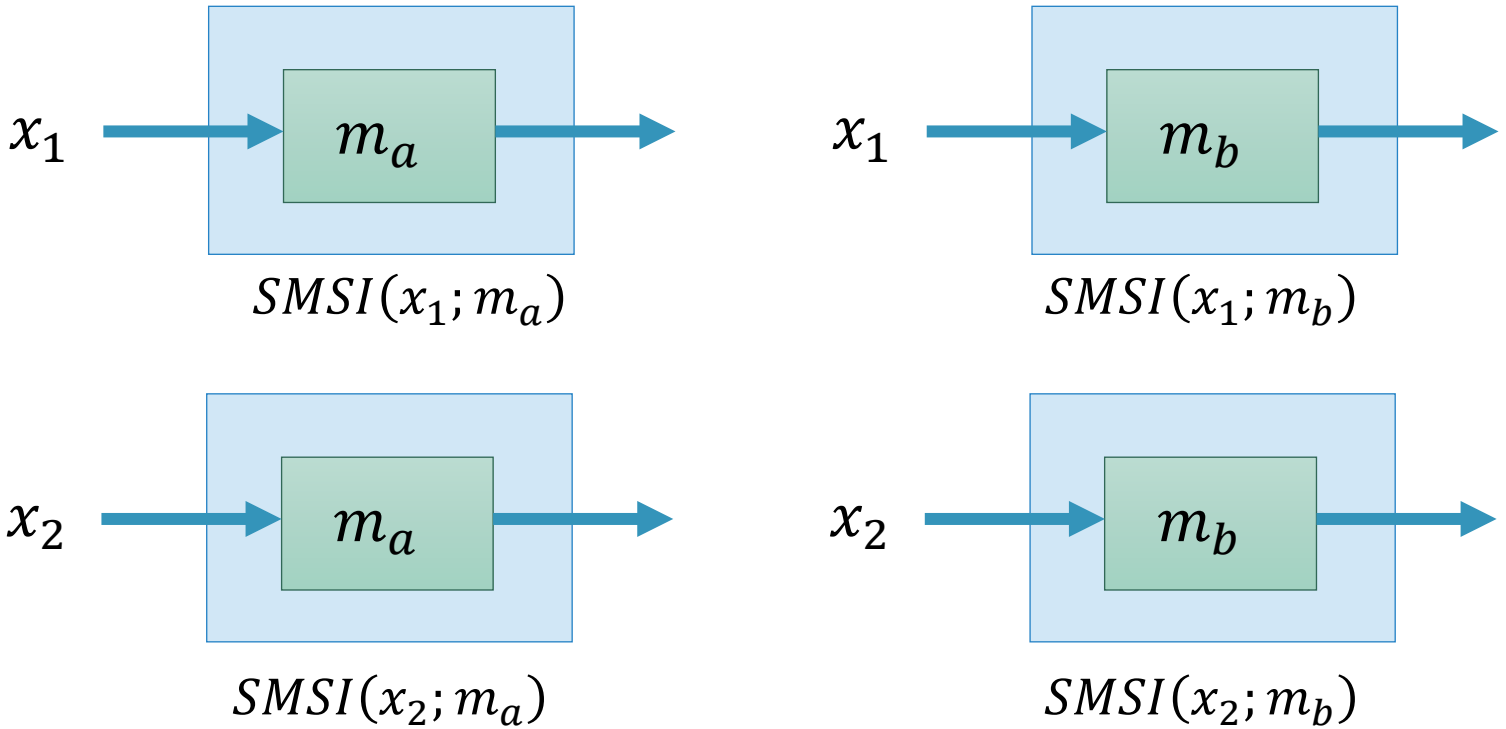
- 学習済み機械学習モデルを一つ配備し，一つの入力データソースと接続する
- 2入力2モデルで4通りのモジュール構成が取り得る
  - 入力データソース：x_1, x_2
  - 機械学習モデル：m_a, m_b
  - モジュール：

$$\begin{cases} (x_1; m_a) \\ (x_1; m_b) \\ (x_2; m_a) \\ (x_2; m_b) \end{cases}$$

これらのモジュールをどのように組み合わせるか（アーキテクチャ）

アーキテクチャ(N＝1)

- 単一モデル単一入力（Single model single input: SMSI）

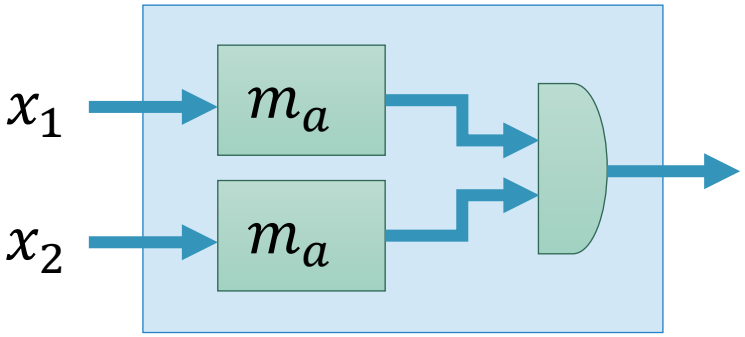


4通り

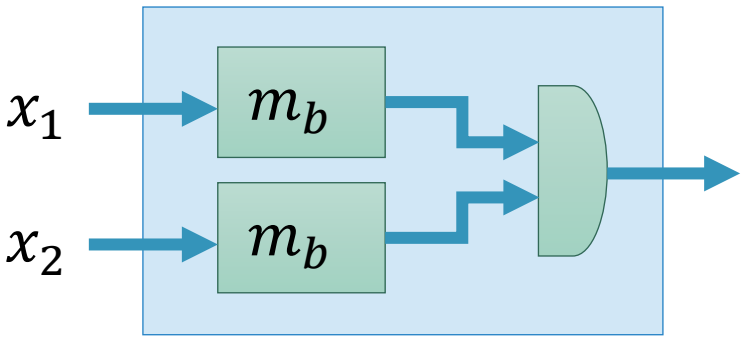
アーキテクチャ(N=2)

6通り

単一モデル二重入力
(Single model double input: SMDI)

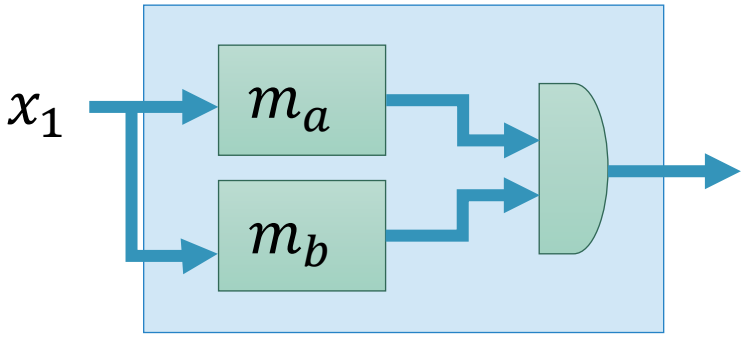


$SMDI(m_a; x_1, x_2)$

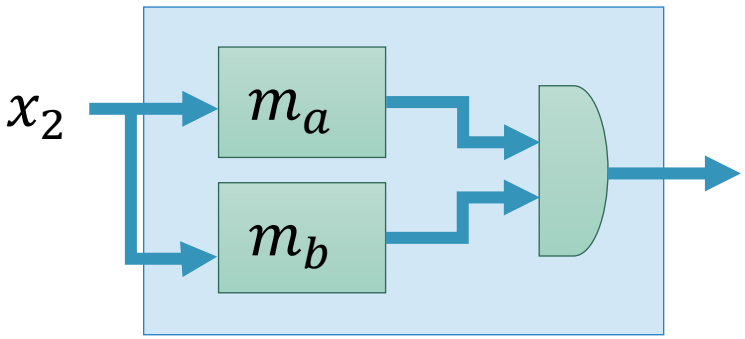


$SMDI(m_b; x_1, x_2)$

二重モデル単一入力
(Double model single input: DMSI)

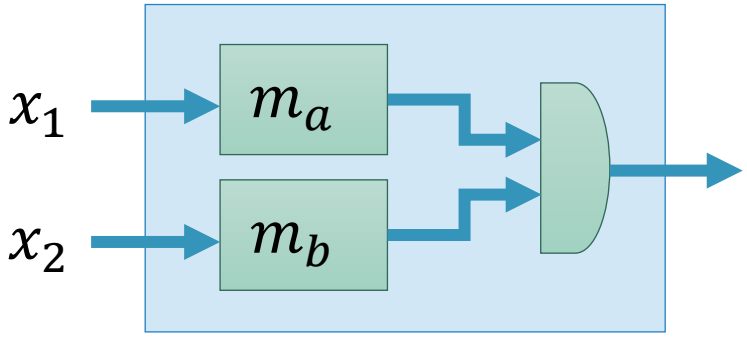


$DMSI(m_a, m_b; x_1)$

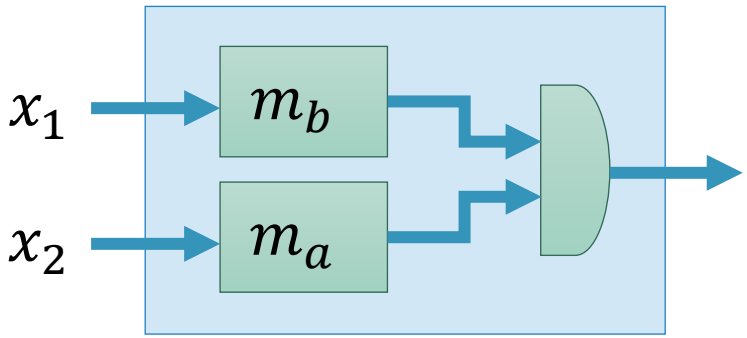


$DMSI(m_a, m_b; x_2)$

二重モデル二重入力
(Double model double input: DMDI)

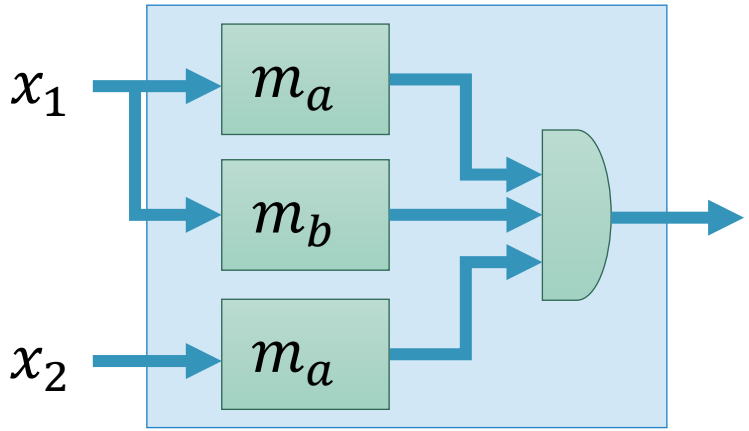


$DMDI(m_a; x_1, m_b; x_2)$

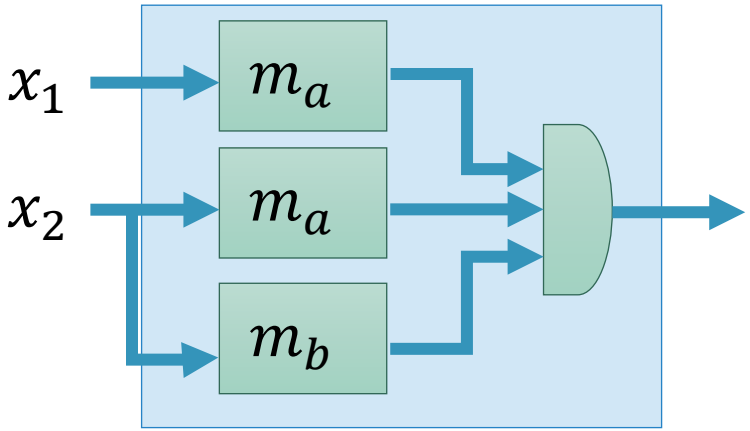


$DMDI(m_a; x_2, m_b; x_1)$

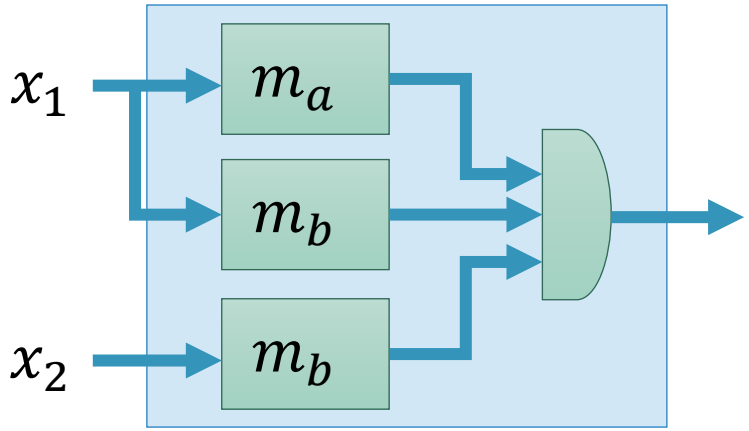
アーキテクチャ(N=3)



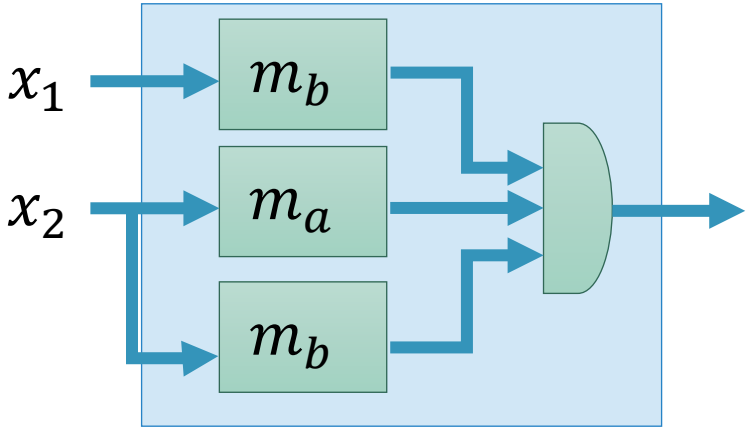
$DMDI(m_a; x_1, x_2, m_b; x_1)$



$DMDI(m_a; x_1, x_2, m_b; x_2)$



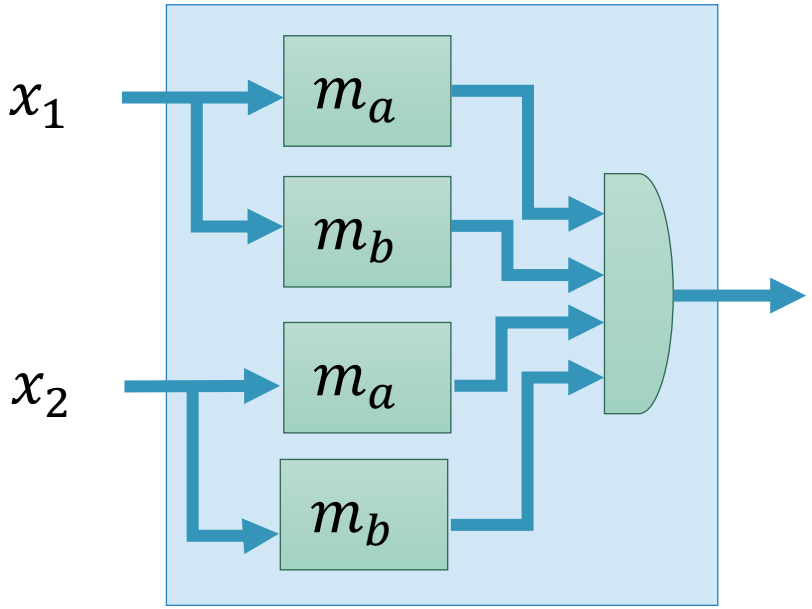
$DMDI(m_a; x_1, m_b; x_1, x_2)$



$DMDI(m_a; x_2, m_b; x_1, x_2)$

4通り

アーキテクチャ(N=4)

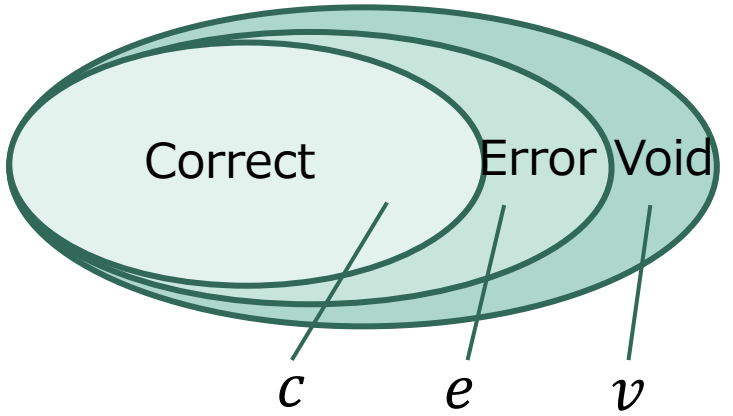


$DMDI(m_a, m_b; x_1, x_2)$

1通り

アーキテクチャ比較のための評価指標（1）

- システムの出力は３通り
  - 正しい分類結果（Correct）
  - 誤った分類結果（Error）
  - 出力なし（Void）



n：全入力数

- **正解率**
  - システムの出力のうち結果が正しい割合

$$\text{正解率} = \frac{c}{c + e}$$

アーキテクチャ比較のための評価指標 (2)

- **安全率**
  - 正しい結果を出力するか出力を棄却するかの何れかである割合

$$\text{安全率} = \frac{c+v}{n} = \frac{c+v}{c+v+e} = 1 - \frac{e}{n}$$

誤った出力を
出さなければ良い

- **応答率**
  - 全試行回数のうち出力が棄却されない割合

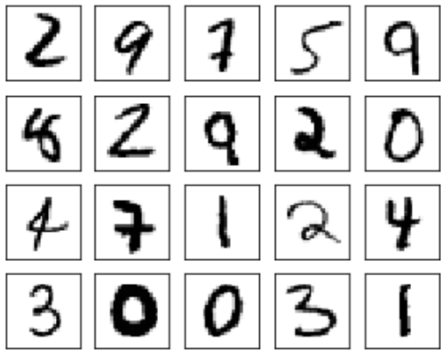
$$\text{応答率} = \frac{c+e}{n} = 1 - \frac{v}{n}$$

正誤に関わらず
出力される確率

手書き文字分類による評価

- Nバージョン機械学習分類システムのアーキテクチャの信頼性を比較評価する
- データセット：MNIST（手書き数字0～9）
- 機械学習モデル：深層ニューラルネットワーク（LeNet, AlexNet）
- 入力データ多様化：テスト画像を加工（平行移動，ノイズ，回転）

データセット



分類モデル

m_a
（LeNet）

m_b
（AlexNet）

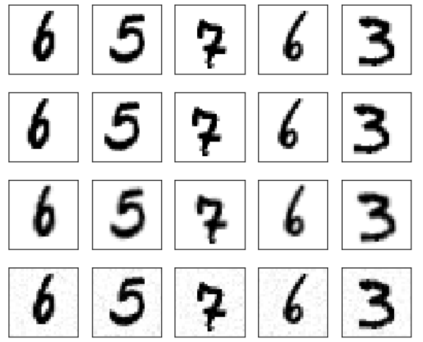
（評価用）入力データ

無加工

平行移動

回転

ノイズ



加工画像データに対する分類結果の正解率

- 加工した画像データでは多くの場合正解率は低下

データセット		LeNet	AlexNet
未加工	x_o	0.9917	0.9913
移動(1,0)	$x_{s(1,0)}$	0.9896	0.9864
移動(-1,0)	$x_{s(-1,0)}$	0.9872	0.9882
回転(1)	$x_{r(1)}$	0.9908	0.9907
回転(-1)	$x_{r(-1)}$	0.9917	0.9906
ノイズ(0.0001)	$x_{n(0.0001)}$	0.9913	0.9912
ノイズ(0.0005)	$x_{n(0.0005)}$	0.9917	0.9915

アーキテクチャ比較（移動加工）

- 左移動データx_1と右移動データx_2が与えられた場合

アーキテクチャ		正解率	安全率	応答率
N=2	$DMSI(m_a, m_b; x_1)$	0.995	0.9951	0.9868
	$DMSI(m_a, m_b; x_2)$	0.9932	0.9933	0.986
	$SMDI(m_a; x_1, x_2)$	0.9957	0.9958	0.9861
	$SMDI(m_b; x_1, x_2)$	0.9951	0.9952	0.9853
	$DMDI(m_a; x_1, ; m_b, x_2)$	0.9974	0.9974	0.9845
	$DMDI(m_a; x_2, m_b; x_1)$	0.9974	0.9975	0.9799
N=3	$DMDI(m_a; x_1, m_b; x_1, x_2)$	0.9981	0.9981	0.979
	$DMDI(m_a; x_1, x_2, m_b; x_1)$	0.9982	0.9982	0.9769
	$DMDI(m_a; x_1, x_2, m_b, x_2)$	0.9981	0.9981	0.979
	$DMDI(m_a; x_2, m_b; x_1, x_2)$	0.9963	0.9964	0.9764
N=4	$DMDI(m_a, m_b; x_1, x_2)$	**0.9988**	**0.9988**	0.9736

アーキテクチャ比較（ノイズ加工）

- 元データx_1とノイズ追加データx_2が与えられた場合

アーキテクチャ		正解率	安全率	応答率
N=2	$DMSI(m_a, m_b; x_1)$	0.996667	**0.9967**	0.99
	$DMSI(m_a, m_b; x_2)$	0.996264	0.9963	0.9903
	$SMDI(m_a; x_1, x_2)$	0.992094	0.9921	0.9992
	$SMDI(m_b; x_1, x_2)$	0.991597	0.9916	0.9996
	$DMDI(m_a; x_1, ; m_b, x_2)$	**0.996668**	**0.9967**	0.9903
	$DMDI(m_a; x_2, m_b; x_1)$	0.996667	**0.9967**	0.99
N=3	$DMDI(m_a; x_1, m_b; x_1, x_2)$	0.996667	**0.9967**	0.99
	$DMDI(m_a; x_1, x_2, m_b; x_1)$	0.996665	**0.9967**	0.9896
	$DMDI(m_a; x_1, x_2, m_b, x_2)$	0.996666	**0.9967**	0.9899
	$DMDI(m_a; x_2, m_b; x_1, x_2)$	0.996263	0.9963	0.99
N=4	$DMDI(m_a, m_b; x_1, x_2)$	0.996665	**0.9967**	0.9896

アーキテクチャ比較（回転加工）

- 左回転データx_1と右回転データx_2が与えられた場合

アーキテクチャ		正解率	安全率	応答率
N=2	$DMSI(m_a, m_b; x_1)$	0.996561	0.9966	0.9888
	$DMSI(m_a, m_b; x_2)$	0.995755	0.9958	0.9895
	$SMDI(m_a; x_1, x_2)$	0.992087	0.9921	0.9984
	$SMDI(m_b; x_1, x_2)$	0.99119	0.9912	0.9989
	$DMDI(m_a; x_1, ; m_b, x_2)$	0.996561	0.9966	0.9888
	$DMDI(m_a; x_2, m_b; x_1)$	0.996665	0.9967	0.9895
N=3	$DMDI(m_a; x_1, m_b; x_1, x_2)$	0.99656	0.9966	0.9884
	$DMDI(m_a; x_1, x_2, m_b; x_1)$	**0.996762**	**0.9968**	0.9884
	$DMDI(m_a; x_1, x_2, m_b, x_2)$	**0.996762**	**0.9968**	0.9884
	$DMDI(m_a; x_2, m_b; x_1, x_2)$	0.995754	0.9958	0.9891
N=4	$DMDI(m_a, m_b; x_1, x_2)$	0.996761	**0.9968**	0.988

評価結果のまとめ

- バージョン数Nが大きいほど正解率や安全率は向上する傾向にある.
- ただし，必ずしもNが大きい場合に正解率が最大となるわけではない.
- Nが大きいほど応答率は下がる.

望ましい性質を持つアーキテクチャを特定する手法が今後の課題

おわりに

- ２つの機械学習分類モデルと2つの入力データソースが与えられた場合のNバージョン機械学習システムのアーキテクチャを比較評価した.
- 評価指標として正解率，安全率，応答率を定義した.
- 手書き文字画像データの分類タスクでアーキテクチャの比較評価実験を行った.
- 正解率，安全率，応答率の両面で望ましいアーキテクチャを特定する方法を開発することが今後の課題.

参考文献

1. Machida, F (2019) : “N-version machine learning models for safety critical systems,” In Proc. of the DSN Workshop on Dependable and Secure Machine Learning, pp. 48-51.
2. Avizienis A. and L. Chen (1977) : “On the implementation of N-version programming for software fault tolerance during execution,” In Proc. of IEEE COMPSAC, pp. 149–155.
3. LeCun, Y., L. Bottou, Y. Bengio, and P. Haffner (1998) : “Gradient-based learning applied to document recognition.” In Proc of the IEEE, Vol.86, No.11, pp. 2278–2324.
4. Krizhevsky A., I. Sutskever, and G. E. Hinton (2017) : “ImageNet classification with deep convolutional neural networks,” Commun. ACM, Vol.60, No.6, pp.84-90.