Nバージョン構成による機械学習システムの高信頼化

町田 文雄
筑波大学 システム情報系 准教授

ソフトウェアエンジニアリングシンポジウム2024

目次

- 背景
- Nバージョン機械学習システムの概要
- 信頼性モデルとアーキテクチャ選択
- ソフトウェア開発自動化への応用

機械学習システムの応用

- 機械学習を使ったシステムの産業応用が広がる

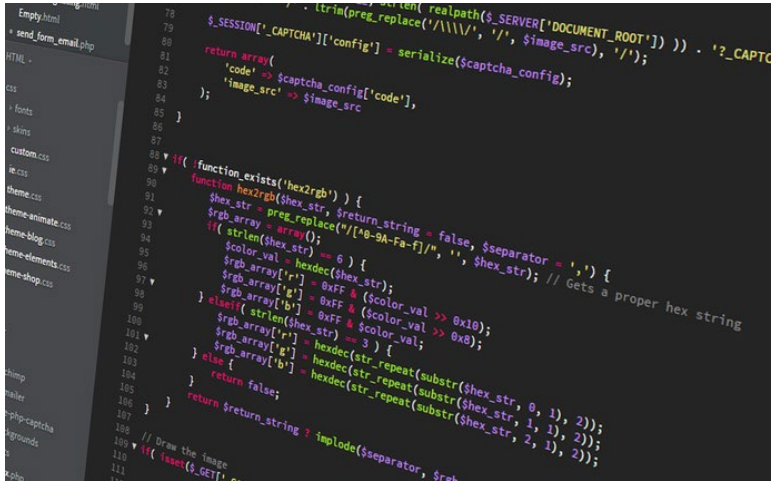
自動運転車

カメラ等のセンサから得た情報から、歩行者や自動車、信号、標識などを認識

ヘルスケア

ウェアラブル端末等から得られるバイタルデータから健康状態を推定・疾患を予測

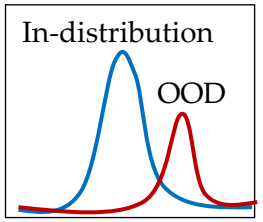
ソフトウェア開発



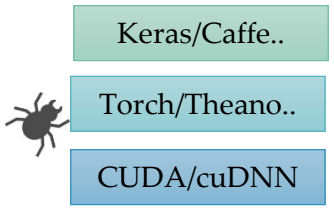
膨大なプログラムデータを学習してコードを自動生成し、プログラム開発を支援

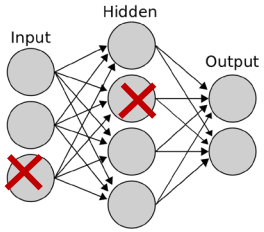
機械学習システムの信頼性リスク

- 機械学習モデルの推論エラー
  - 分布外入力（Out-Of-Distribution）
  - 敵対的サンプル（Adversarial Example）



- ソフトウェアやハードウェアの障害
  - ソフトウェアバグ
  - 一時的メモリ障害（Soft Error）

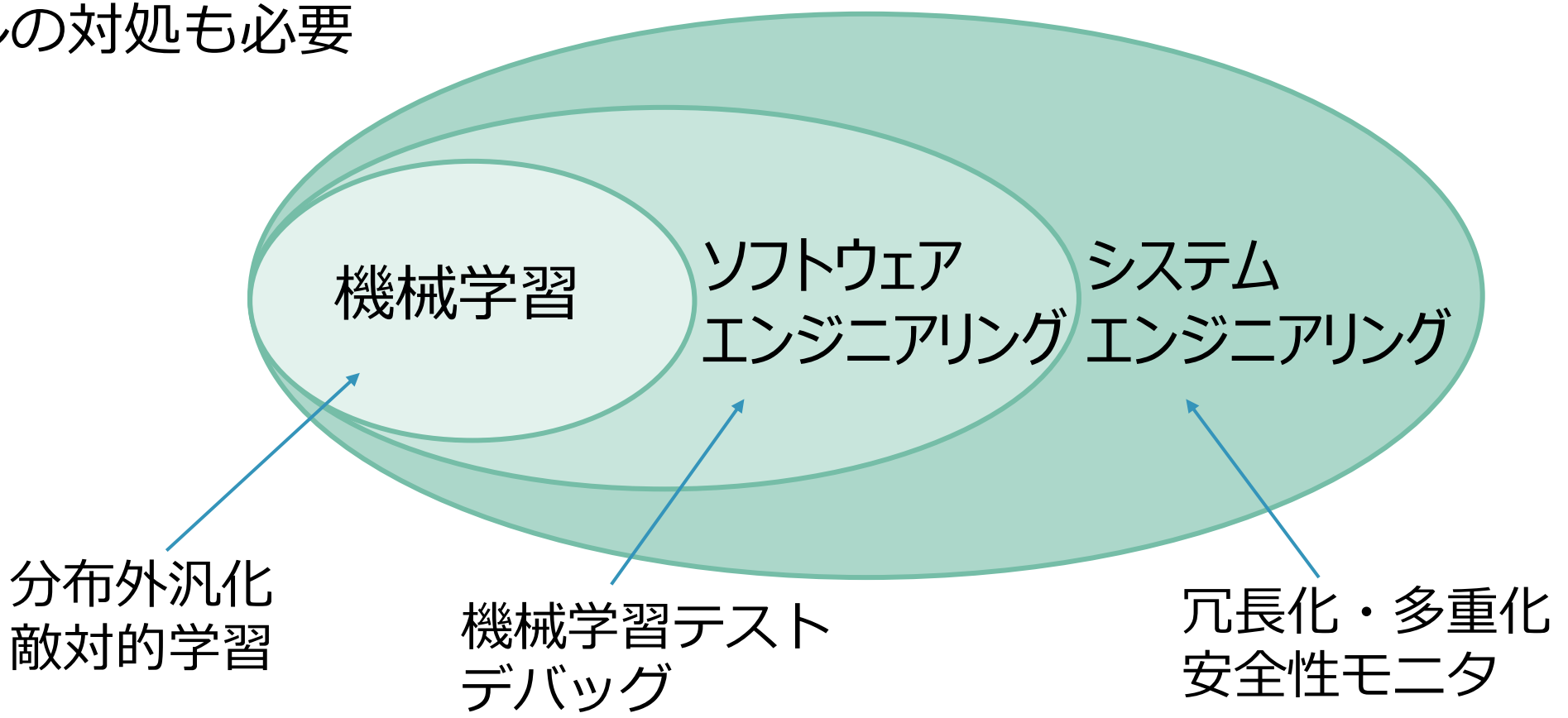




推論結果が正しいかどうか判別がつかない

機械学習システム高信頼化のアプローチ

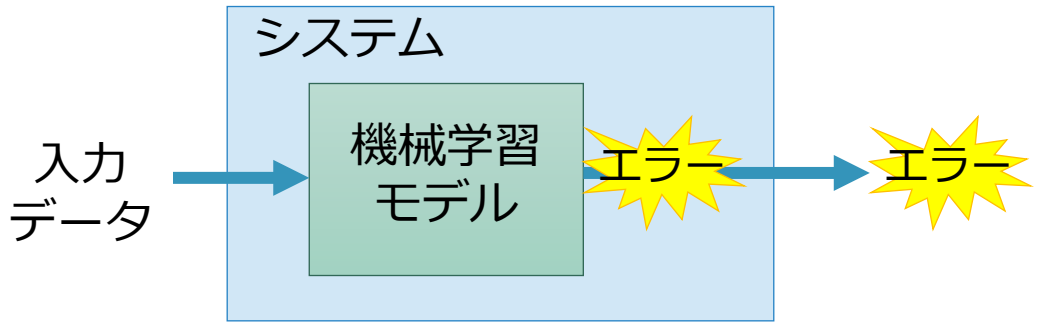
- 機械学習モデルの改良だけでなく、アプリケーションやシステムレベルの対処も必要



Nバージョン機械学習システム

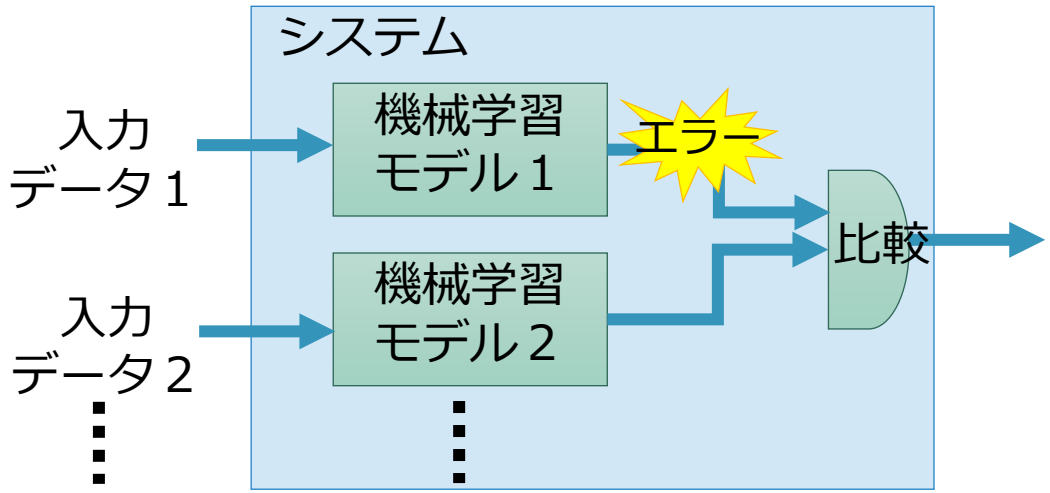
- 機械学習の推論を冗長化してエラー出力を抑える

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

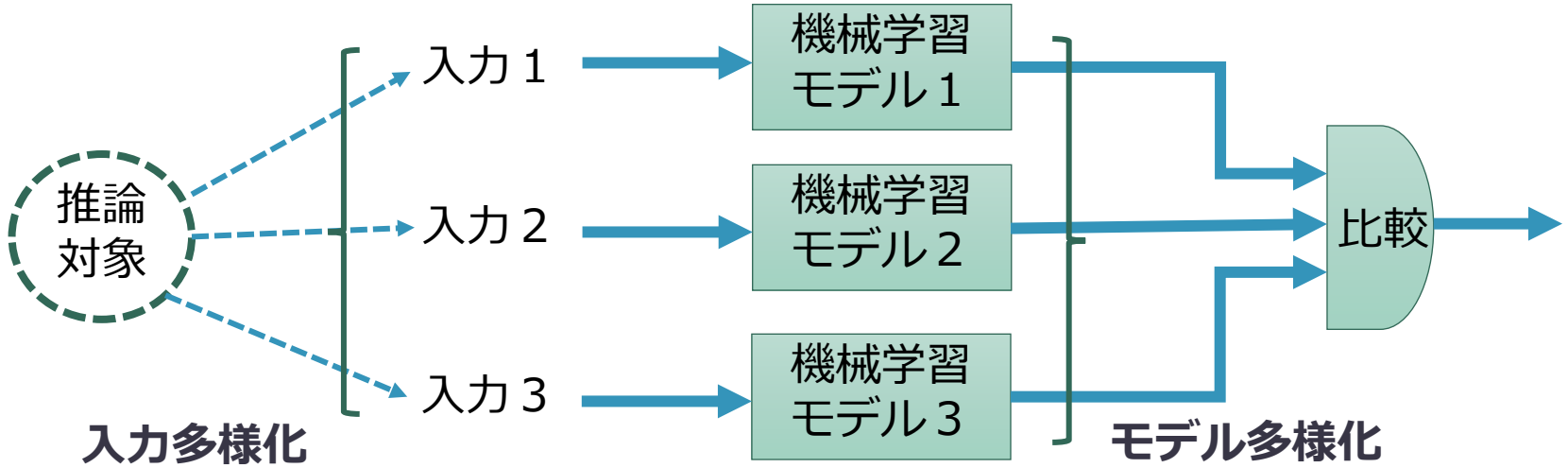
Nバージョン機械学習システムの場合



複数の結果を比較してシステムのエラー出力を抑える

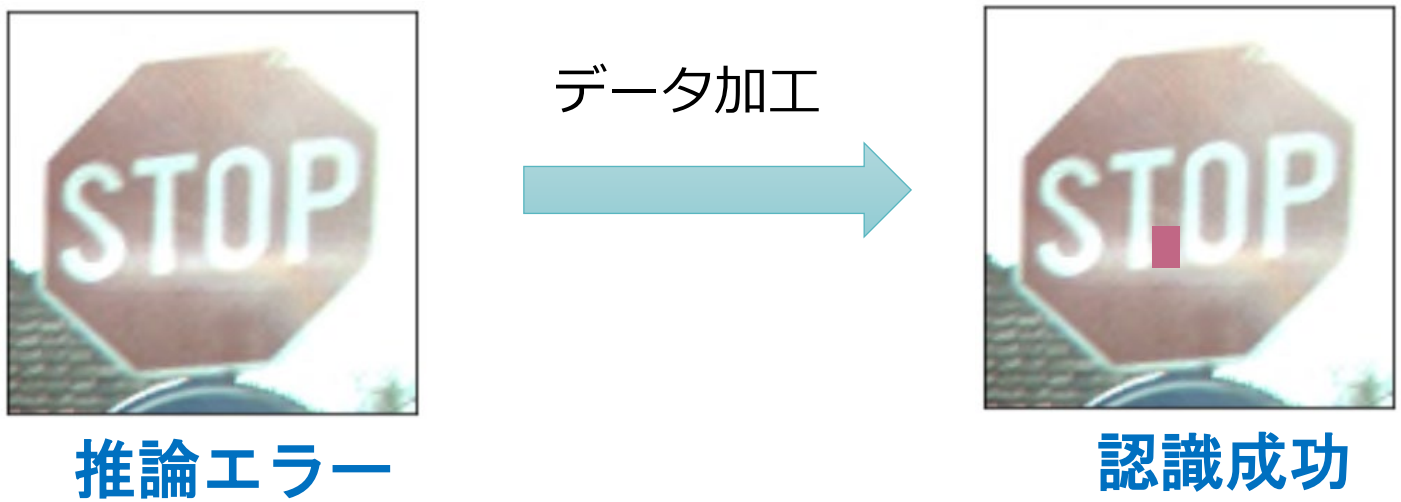
モデルの多様化と入力の多様化

- 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



入力データ多様化

- 機械学習モデルは入力データの違いに敏感
  - 入力データのわずかな加工で機械学習モデルを騙せる（敵対的サンプル）
    → 逆も起こり得る



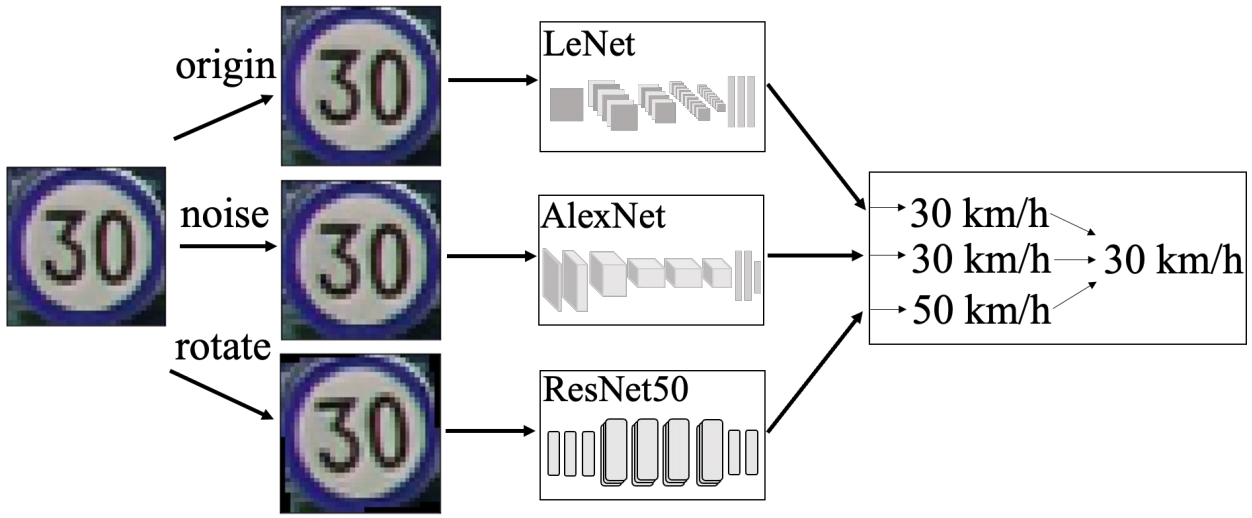
推論エラー　　**認識成功**

Nバージョンプログラミングとの違い

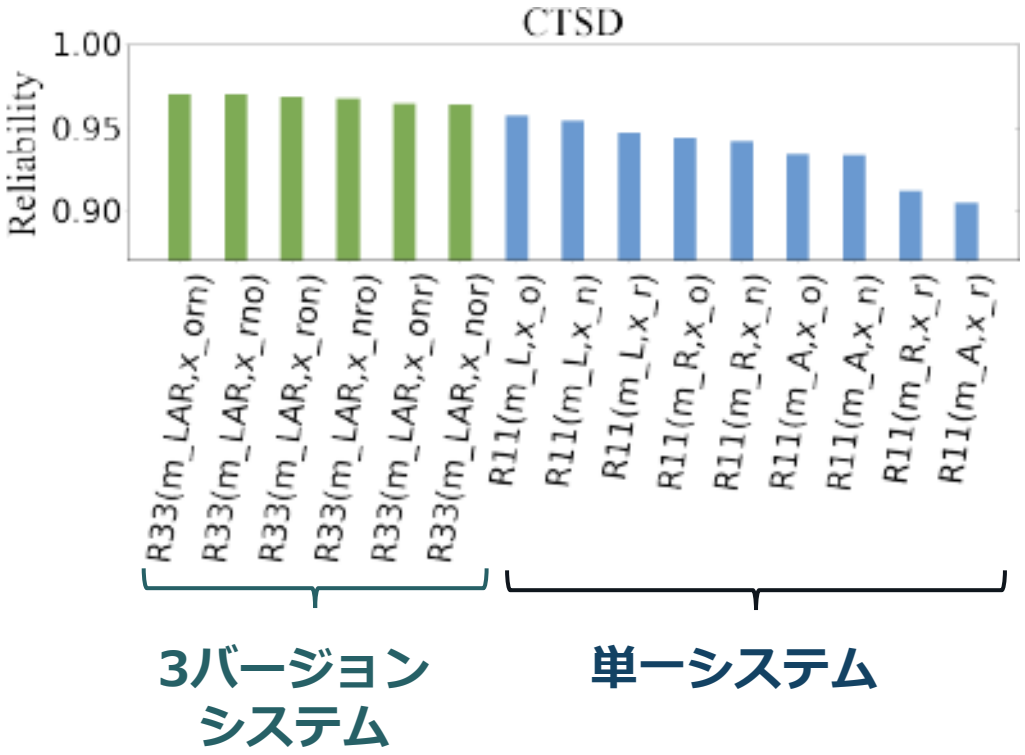
	Nバージョンプログラミング	**Nバージョン機械学習システム**
対象	プログラム（仕様に基づいて開発される）	機械学習モデル（訓練データから学習される）
対処する問題	ソフトウェアのバグ	誤判断
構成要素	2つ以上の機能的に等価なプログラム	**1つ以上**の同じタスクを実行する機械学習モデル
多様化手法	開発チーム、プログラミング言語、ライブラリ、ツール	学習アルゴリズム、ハイパーパラメータ、学習データ、**入力データ**
導入コスト	高い	**低い**

画像分類システムでの応用

- 3バージョン交通標識分類システム
  - 3つの加工入力データと3つの異なるニューラルネットワークで構成

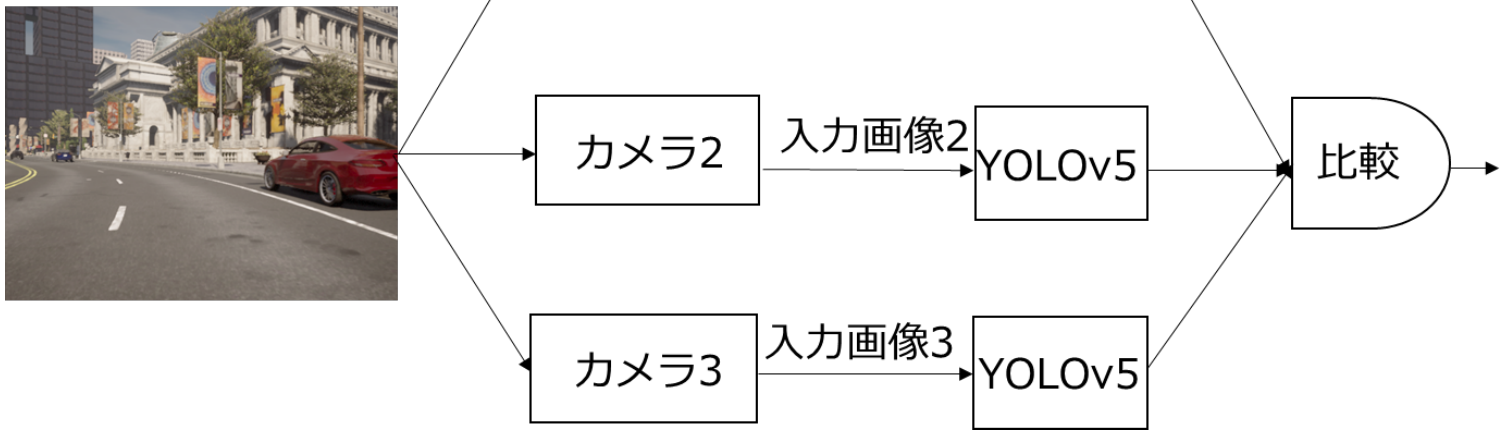


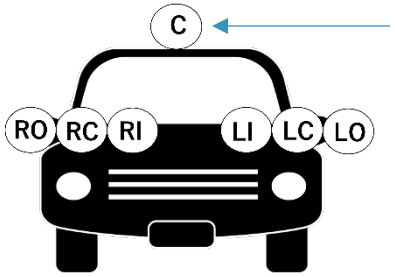
[Q. Wen, et al. ISSRE2023]



物体検出システムでの応用

- 3バージョン物体検出システム
  - 設置位置の異なる3つのカメラから撮影した画像でそれぞれ物体検出をして比較





単一カメラ画像での
正解率
0.925

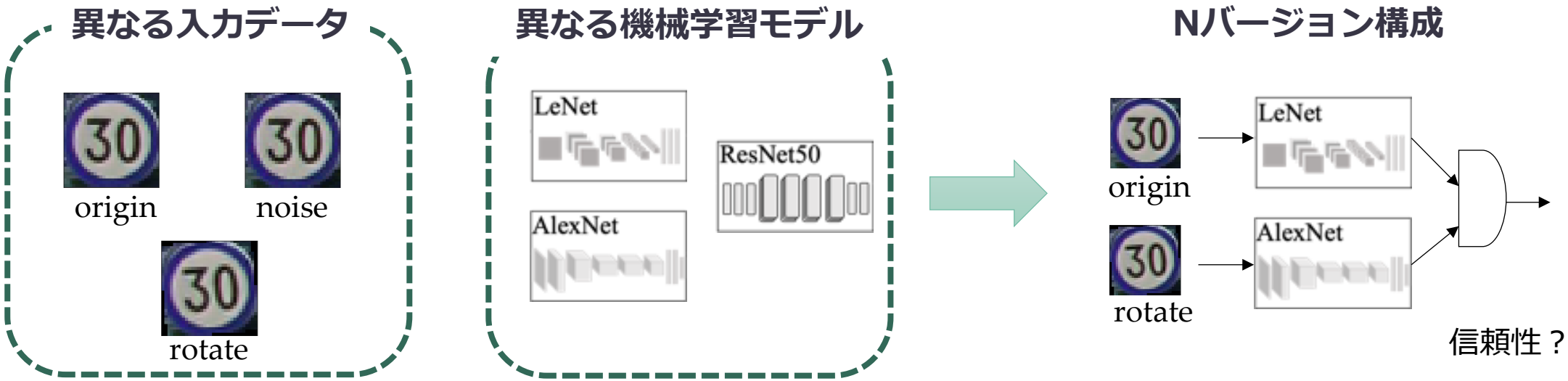
3つのカメラ画像を組み合わせた場合に
何れか一つが正解する確率

	RI	LI	RC	LC	RO
RI					
LI	0.962				
RC	0.968	0.970			
LC	0.968	0.966	0.972		
RO	0.966	0.969	0.963	0.973	
LO	0.969	0.966	0.974	0.970	0.974

信頼性モデルとアーキテクチャの選択

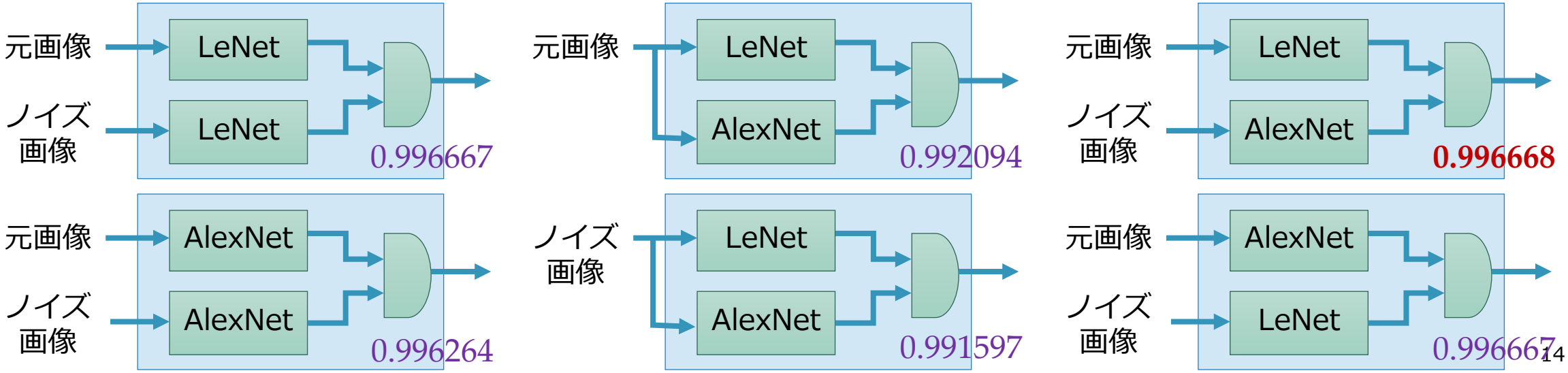
Nバージョン構成選択の問題

- 複数のモデルと異なる入力データが与えられたとき、どのような構成が最も信頼性を向上させるか？
  - どのモデルを使うか？
  - どの入力データをどのモデルに与えるか？



経験則

- 画像分類システムの信頼性はNバージョン構成によって異なる
  - 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：元画像、ノイズ追加画像
  - 比較器：不一致の場合は出力しない



Nバージョン構成の信頼性モデル

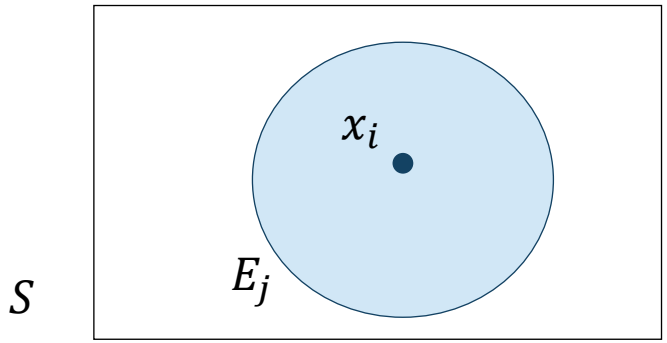
- 異なる入力データと異なるモデルの組み合わせで信頼性が異なる
  → 理論的にどこまで解析できるか？
- 分類システムを対象に信頼性モデルを考える
  - 問題設定

入力データ：同一の対象に対して**2つ**の異なる入力データを利用可能
機械学習モデル：同一の分類タスクを行う**2つ**の分類モデルを利用可能
比較器のルール：出力結果が**一致する場合**にのみその結果を出力する
信頼性：システムが誤った結果を出力しない確率

1バージョン構成の信頼性

- 記法
  - 入力データ：$x_i, i = \{1,2, \dots\}$
  - 機械学習モデル：$m_j, j = \{a, b, \dots\}$
  - 入力データの標本空間：S
  - 機械学習モデルm_jがエラーとなる入力データの集合： $E_j \subset S$
- 機械学習モデルm_jと入力データx_iを組み合わせた場合の信頼性

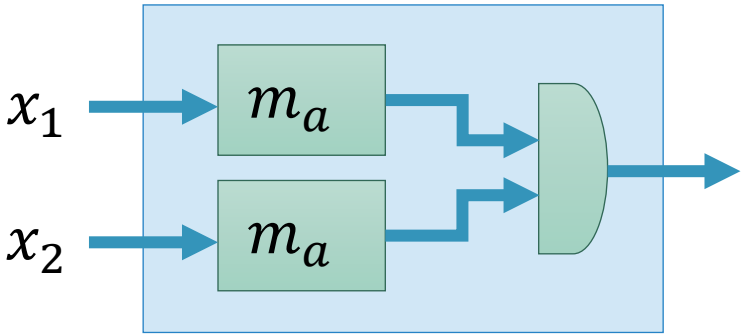
$$1 - \mathrm{P}[x_i \in E_j]$$



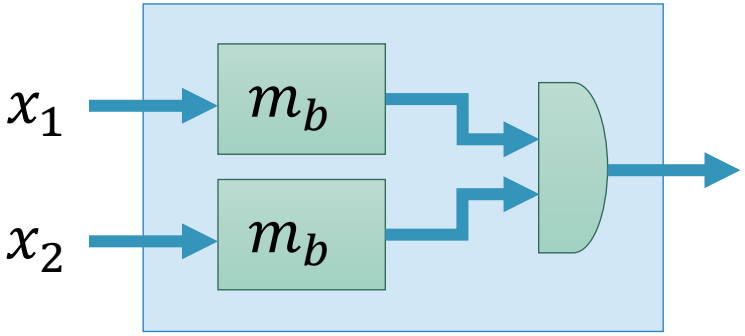
2バージョン構成のアーキテクチャ

6通り

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

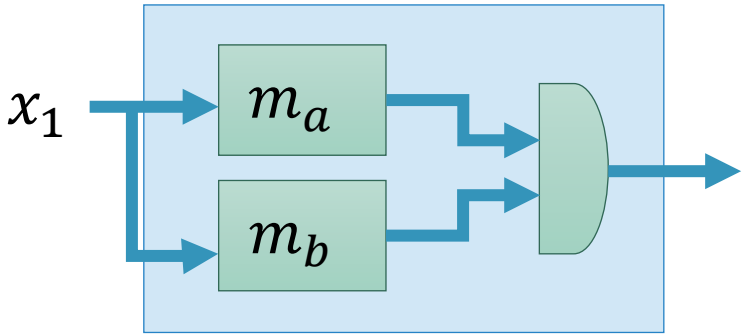


$SMDI(m_a; x_1, x_2)$

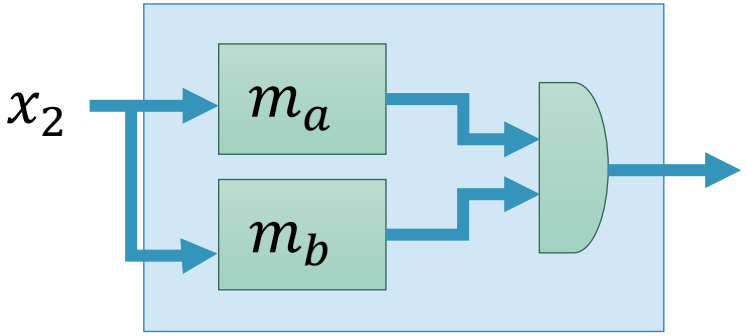


$SMDI(m_b; x_1, x_2)$

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

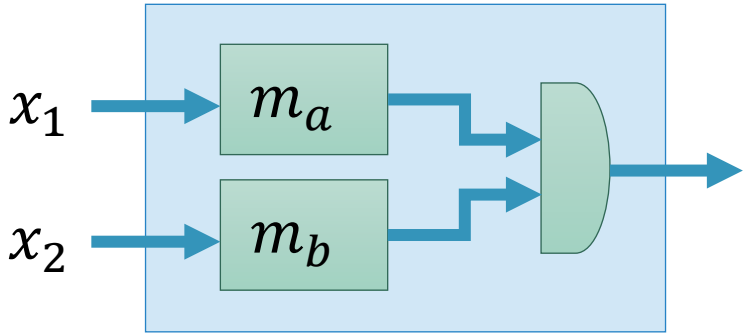


$DMSI(m_a, m_b; x_1)$

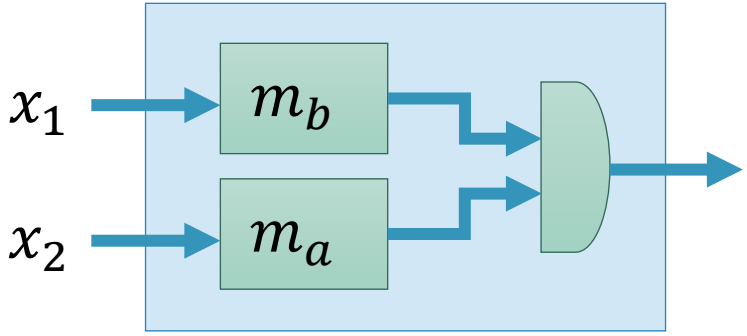


$DMSI(m_a, m_b; x_2)$

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



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

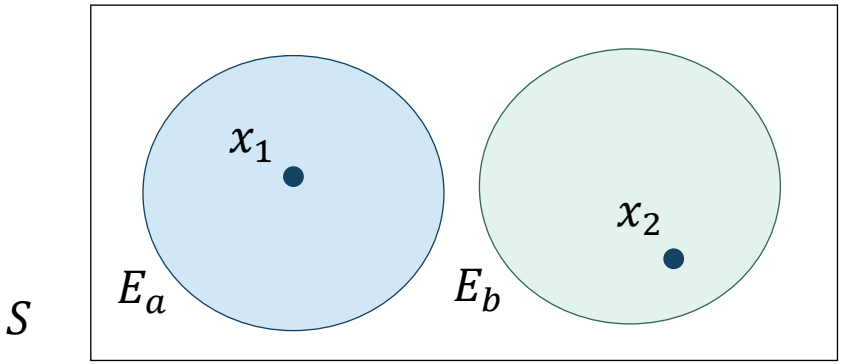


$DMDI(m_a; x_2, m_b; x_1)$

2バージョン構成の信頼性

- エラー確率$\mathrm{P}[x_i \in E_j]$が独立な場合
  - 2バージョン構成のエラー確率は1バージョン構成のエラー確率の積

    $1 - \mathrm{P}[x_1 \in E_a] \cdot \mathrm{P}[x_2 \in E_b]$



- 実際はエラー確率は独立ではない
  - エラー集合E_jは共通部分を持つ可能性がある
  - 入力データx_iの分布は同一とは限らない

多様性指標の導入

- 2つの機械学習モデルを用いる場合、エラー集合に依存関係がある

エラーの共通部分(モデル類似度)

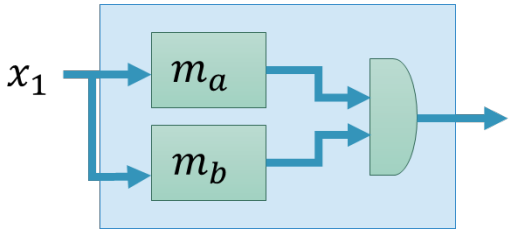
機械学習モデルm_a, m_bが誤出力となる入力データx_iの標本空間Sの部分集合をそれぞれE_a, E_bとし、エラーの共通部分$\alpha_{b|a,i} \in [0,1]$を以下で定義する。

$$\alpha_{b|a,i} = P[x_i \in E_b | x_i \in E_a] = \frac{P[x_i \in E_a \cap E_b]}{P[x_i \in E_a]}.$$

ただし$P[x_i \in E_a] > 0$とする。

- 二重モデル単一入力システム(DMSI)の信頼性

$$R_{DMSI_{a\cap b,1}} = 1 - \alpha_{b|a,1} \cdot P[x_1 \in E_a]$$



多様性指標の導入2

- 2つの入力データを用いる場合、2つのデータ分布は独立ではない

エラーの共起度(入力類似度)

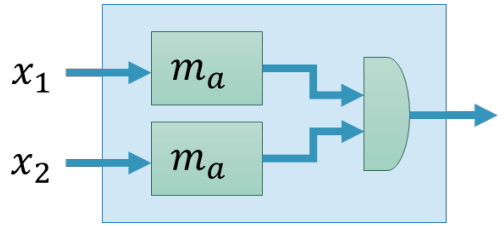
同じ標本空間Sから抽出した機械学習モデルm_j に対する入力データをx_1, x_2とし、エラーの共起度$\beta_{j,2|1} \in [0,1]$を以下のように定義する。

$$\beta_{j,2|1} = Pr\left[x_2 \in E_j | x_1 \in E_j\right] = \frac{P\left[x_1 \in E_j, x_2 \in E_j\right]}{P\left[x_1 \in E_j\right]}.$$

ただし$P\left[x_1 \in E_j\right] > 0$とする。

- 単一モデル二重入力システム(SMDI)の信頼性

$$R_{SMDI_{a,1\cap 2}} = 1 - \beta_{a,2|1} \cdot P[x_1 \in E_a]$$



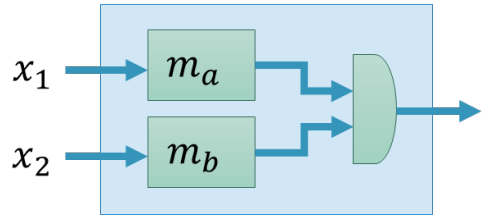
二重モデル二重入力システムの信頼性

- 二重モデル二重入力(DMDI)システムの信頼性はモデルの依存関係と入力データ分布の重なりの両方の影響を受ける
- $DMDI(m_a; x_1, m_b; x_2)$の信頼性

$$R_{DMDI_{a,1\cap b,2}} = 1 - \left[\alpha_{b,2|a,1\cap 2} \cdot \beta_{a,2|1} + \alpha_{b,2|a,1\cap \bar{2}} \cdot \left(1 - \beta_{a,2|1}\right)\right] \cdot P[x_1 \in E_a]$$

$$\alpha_{b,2|a,1\cap 2} = P[x_2 \in E_b | x_2 \in E_a, x_1 \in E_a]$$

$$\alpha_{b,2|a,1\cap \bar{2}} = P[x_2 \in E_b | x_2 \in \overline{E_a}, x_1 \in E_a]$$



入力類似度とモデル類似度に関連したパラメータで特徴づけられる

アーキテクチャの信頼性の関連付け

- 異なるアーキテクチャの信頼性は多様性指標を用いて互いに関連づけられる

信頼性行列 $\boldsymbol{R}_{a|b} = \begin{bmatrix} R_{DMSI_{a\cap b,1}} & R_{DMDI_{a,1\cap b,2}} \\ R_{DMDI_{a,2\cap b,1}} & R_{DMSI_{a\cap b,2}} \end{bmatrix} = \boldsymbol{J}_2 - \boldsymbol{A}_{b|a} \cdot \boldsymbol{B}_a^{\top} \cdot \boldsymbol{P}_a$

$\boldsymbol{J}_2 = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$

モデル類似度 $\boldsymbol{A}_{b|a} = \begin{bmatrix} \alpha_{b,1|a,1\cap 2} & \alpha_{b,1|a,1\cap \bar{2}} & \alpha_{b,1|a,\bar{1}\cap 2} \\ \alpha_{b,2|a,1\cap 2} & \alpha_{b,2|a,1\cap \bar{2}} & \alpha_{b,2|a,\bar{1}\cap 2} \end{bmatrix}$

入力類似度 $\boldsymbol{B}_a = \begin{bmatrix} \beta_{a,2|1} & 1-\beta_{a,2|1} & 0 \\ \beta_{a,1|2} & 0 & 1-\beta_{a,1|2} \end{bmatrix}$

$R_{SMDI_{a,1\cap 2}} = 1 - \beta_{a,2|1} \cdot P[x_1 \in E_a]$

$\boldsymbol{P}_a = \begin{bmatrix} P[x_1 \in E_a] & 0 \\ 0 & P[x_2 \in E_a] \end{bmatrix}$

アーキテクチャの信頼性の関連付け(続き)

- 異なるアーキテクチャの信頼性は多様性指標を用いて互いに関連づけられる

信頼性行列 $\boldsymbol{R}_{b|a} = \begin{bmatrix} R_{DMSI_{a\cap b,1}} & R_{DMDI_{a,2\cap b,1}} \\ R_{DMDI_{a,1\cap b,2}} & R_{DMSI_{a\cap b,2}} \end{bmatrix} = \boldsymbol{J}_2 - \boldsymbol{A}_{a|b} \cdot \boldsymbol{B}_b^{\top} \cdot \boldsymbol{P}_b = \boldsymbol{R}_{a|b}^{\top}$

$\boldsymbol{J}_2 = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$

モデル類似度 $\boldsymbol{A}_{a|b} = \begin{bmatrix} \alpha_{a,1|b,1\cap 2} & \alpha_{a,1|b,1\cap \bar{2}} & \alpha_{a,1|b,\bar{1}\cap 2} \\ \alpha_{a,2|b,1\cap 2} & \alpha_{a,2|b,1\cap \bar{2}} & \alpha_{a,2|b,\bar{1}\cap 2} \end{bmatrix}$

入力類似度 $\boldsymbol{B}_b = \begin{bmatrix} \beta_{b,2|1} & 1-\beta_{b,2|1} & 0 \\ \beta_{b,1|2} & 0 & 1-\beta_{b,1|2} \end{bmatrix}$

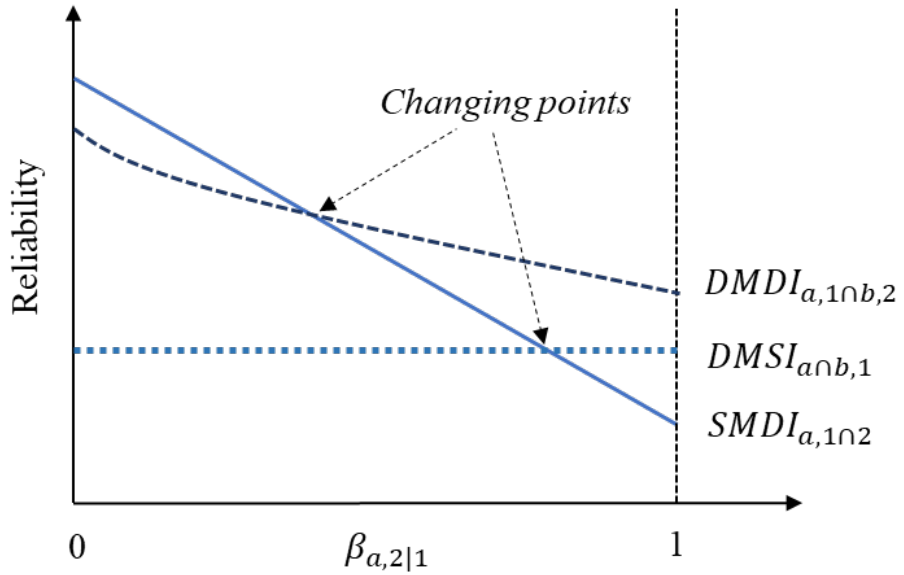
$R_{SMDI_{b,1\cap 2}} = 1 - \beta_{b,2|1} \cdot P[x_1 \in E_b]$

$\boldsymbol{P}_b = \begin{bmatrix} P[x_1 \in E_b] & 0 \\ 0 & P[x_2 \in E_b] \end{bmatrix}$

信頼性モデルから導かれる性質1

性質 1

$\boldsymbol{A}_{b|a}$が与えられたとき、信頼性の差$\boldsymbol{H}_{b|a} = \boldsymbol{R}_{b|a} - R_{SMDI_{a,1\cap2}} \cdot \boldsymbol{J}_2$の$(i,j)$要素の上界が0より大きければ、$\beta_{a,2|1}$または$\beta_{a,1|2}$の増加に伴って$R_{SMDI_{a,1\cap2}}$が$R_{DMSI_{a\cap b,1}}$、$R_{DMDI_{a,1\cap b,2}}$、$R_{DMDI_{a,1\cap b,2}}$、$R_{DMSI_{a\cap b,2}}$より小さくなる変化点が存在する。



$SMDI_{a,1\cap2}$の信頼性との差$\boldsymbol{H}_{b|a}$は
$\beta_{a,2|1}$および$\beta_{a,1|2}$に対して単調増加

$\beta_{a,2|1}$が小さいと考えられる場合は
$SMDI_{a,1\cap2}$が好ましい選択

信頼性モデルから導かれる性質2

- 優位な入力データが判明している場合

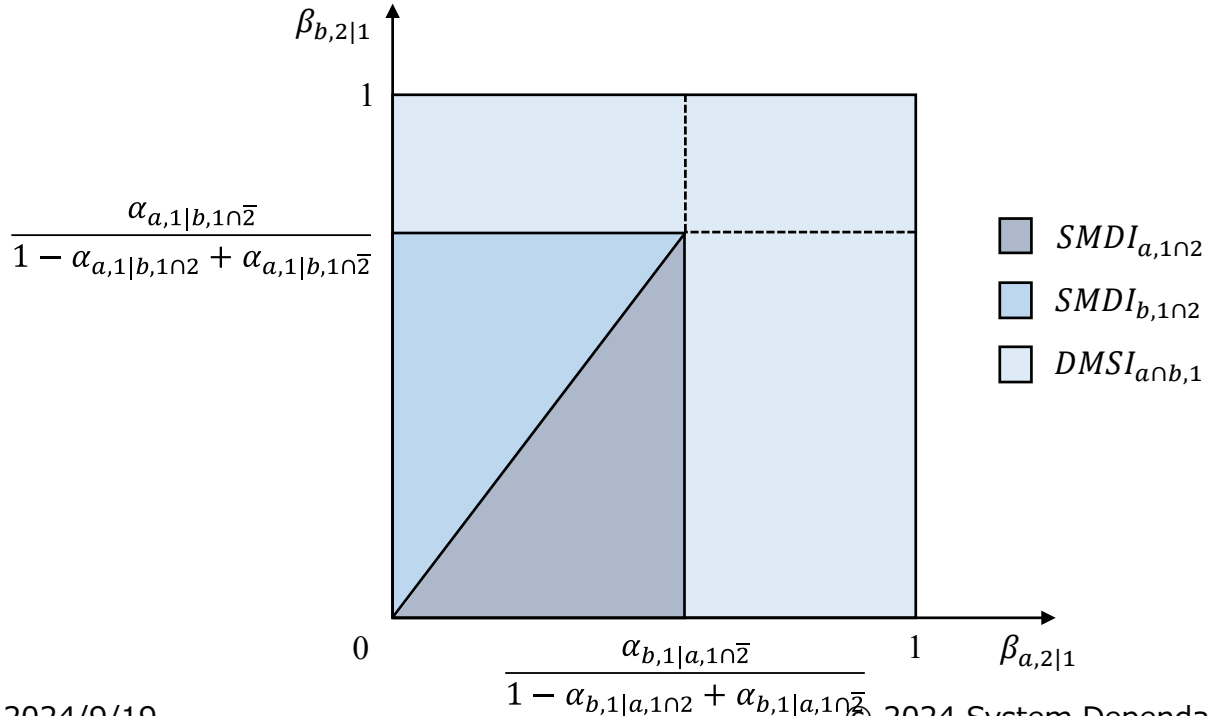
性質2

任意のエラー入力集合$E^* \subseteq E_a \cup E_b$に対して入力データの結合分布が$P[x_1 \in E^*] \leq P[x_2 \in E^*]$を満たすとき、最も信頼性の高いアーキテクチャは以下で与えられる。

$$\begin{cases} SMDI_{a,1\cap 2}, & if\ \beta_{a,2|1} \leq \dfrac{\alpha_{b,1|a1\cap\bar{2},}}{1-\alpha_{b,1|a1\cap 2,}+\alpha_{b,1|a1\cap\bar{2},}}\ and\ \ \beta_{a,2|1} \leq \beta_{b,2|1} \cdot \dfrac{P[x_1 \in E_b]}{P[x_1 \in E_a]}, \\ SMDI_{b,1\cap 2}, & if\ \beta_{b,2|1} \leq \dfrac{\alpha_{a,1|b1\cap\bar{2},}}{1-\alpha_{a,1|b1\cap 2,}+\alpha_{a,1|b1\cap\bar{2},}}\ and\ \ \beta_{a,2|1} \geq \beta_{b,2|1} \cdot \dfrac{P[x_1 \in E_b]}{P[x_1 \in E_a]}, \\ DMSI_{a\cap b,1}, & otherwise. \end{cases}$$

信頼性モデルから導かれる性質2(続き)

- 優位な入力データが判明している場合
  - 例）新しいカメラと古いカメラによる入力で新しいカメラで撮影した画像を入力した方が正解率が高い場合



$\beta_{a,2|1}$および$\beta_{b,2|1}$のバランスで最も良いアーキテクチャが変わる

入力データの類似度$\beta_{a,2|1}$、$\beta_{b,2|1}$が大きい場合は$DMSI_{a\cap b,1}$が好ましい選択

信頼性モデルから導かれる性質3

- 優位なモデルが判明している場合

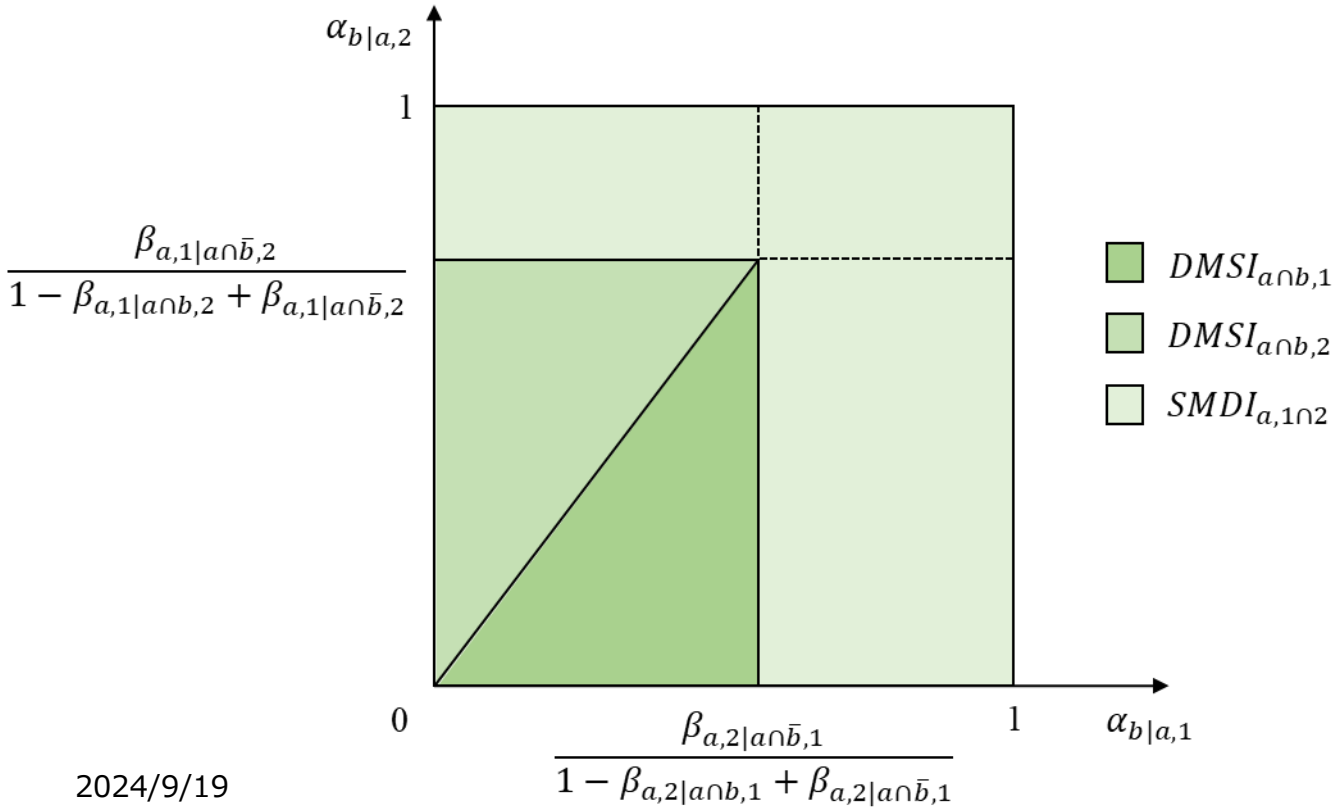
性質3

任意の入力データx_iに対して$P[x_i \in E_a] \leq P[x_i \in E_b]$が成り立つとき、最も信頼性の高いアーキテクチャは以下で与えられる。

$$\begin{cases} DMSI_{a\cap b,1}, & if\ \alpha_{b|a,1} \leq \dfrac{\beta_{a,2|a\cap\bar{b},1}}{1-\beta_{a,2|a\cap b,1}+\beta_{a,2|a\cap\bar{b},1}}\ and\quad \alpha_{b|a,1} \leq \alpha_{b|a,2}\cdot\dfrac{P[x_2 \in E_a]}{P[x_1 \in E_a]}, \\ DMSI_{a\cap b,2}, & if\ \alpha_{b|a,2} \leq \dfrac{\beta_{a,1|a\cap\bar{b},2}}{1-\beta_{a,2|a\cap b,2}+\beta_{a,1|a\cap\bar{b},2}}\ and\quad \alpha_{b|a,1} \leq \alpha_{b|a,2}\cdot\dfrac{P[x_2 \in E_a]}{P[x_1 \in E_a]}, \\ SMDI_{a,1\cap 2}, & otherwise, \end{cases}$$

信頼性モデルから導かれる性質3(続き)

- 優位なモデルが判明している場合
  - 例）新しいモデルが古いモデルより全般的に正解率が高い場合



$\alpha_{b|a,1}$および$\alpha_{b|a,2}$のバランスで最も良いアーキテクチャが変わる

入力データの類似度$\alpha_{b|a,1}$、$\alpha_{b|a,2}$が大きい場合は$SMDI_{a,1\cap 2}$が好ましい選択

信頼性モデルから導かれる性質4

- モデルの類似度と入力データの類似度が条件付き独立と仮定する場合
  - $\alpha_{b,2|a,1\cap 2} = \alpha_{b|a,2}\ and\ \alpha_{b,2|a,1\cap\overline{2}} = P[x_2 \in E_b | x_2 \in \overline{E_a}]$

性質4

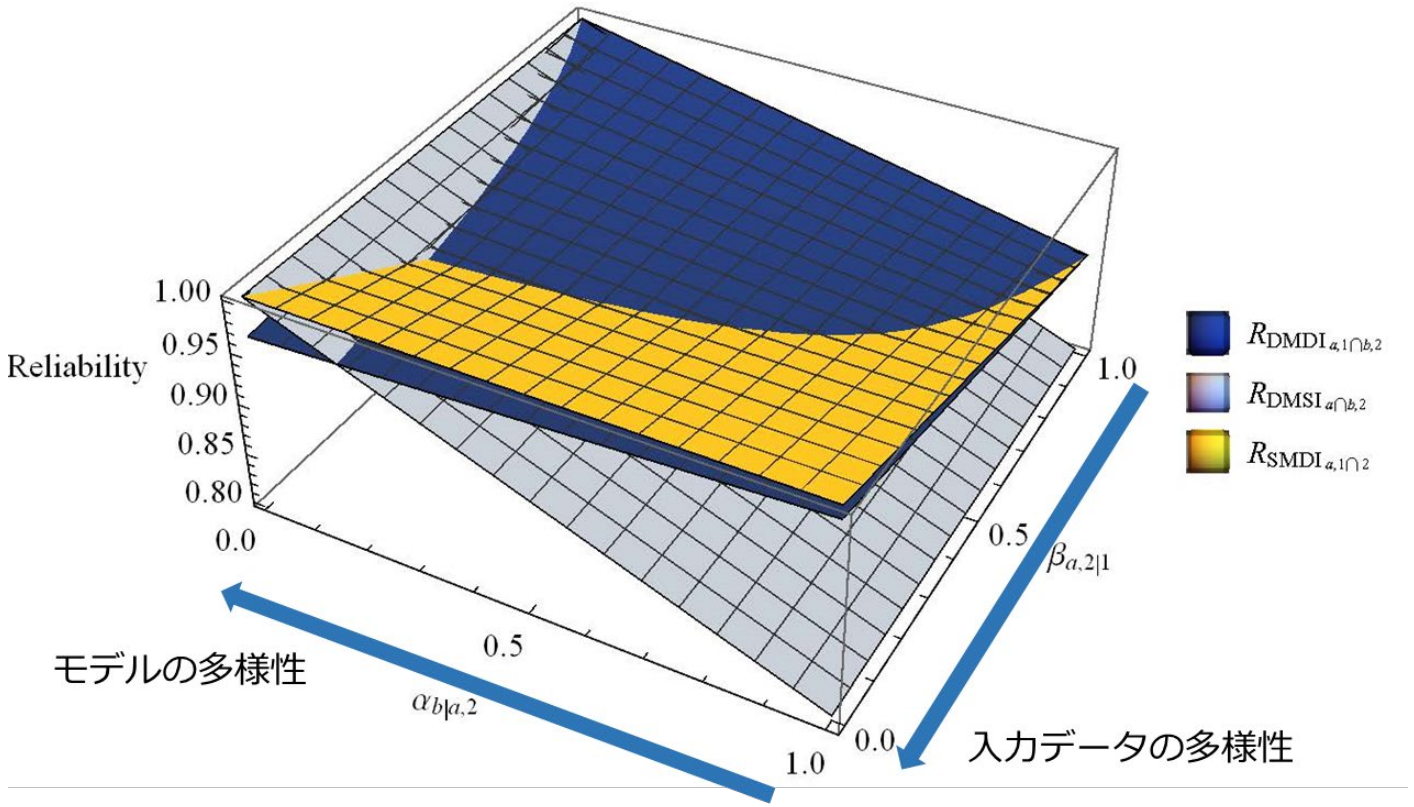
$DMDI_{a,1\cap b,2}$、$DMSI_{a\cap b,2}$ 、 $SMDI_{a,1\cap 2}$の中で最も信頼性が高いアーキテクチャは$\alpha_{b|a,2}$ と $\beta_{a,2|1}$の値から以下で求まる。

$$\begin{cases} DMSI_{a\cap b,2}, & if\ \omega(\alpha_{b|a2'},\beta_{a,2|1}) - \alpha_{b|a2'} \cdot P[x_2 \in E_a] \geq 0\ and \quad \beta_{a,2|1} \geq \alpha_{b|a2'} \cdot \dfrac{P[x_2 \in E_b]}{P[x_1 \in E_a]}, \\ SMDI_{b,1\cap 2}, & if\ \omega(\alpha_{b|a2'},\beta_{a,2|1}) - \beta_{a,2|1} \cdot P[x_1 \in E_a] \geq 0\ and \quad \beta_{a,2|1} \leq \alpha_{b|a2'} \cdot \dfrac{P[x_2 \in E_b]}{P[x_1 \in E_a]}, \\ DMDI_{a,1\cap b,2}, & otherwise. \end{cases}$$

ただし$\omega(\alpha_{b|a,2},\beta_{a,2|1}) = \frac{P[x_1\in E_a]}{1-P[x_2\in E_a]} \cdot [\alpha_{b|a,2} \cdot (\beta_{a,2|1} - P[x_2 \in E_a]) + P[x_2 \in E_b] \cdot (1-\beta_{a,2|1})]$

信頼性モデルから導かれる性質4(続き)

- モデルの類似度と入力データの類似度が条件付き独立と仮定する場合



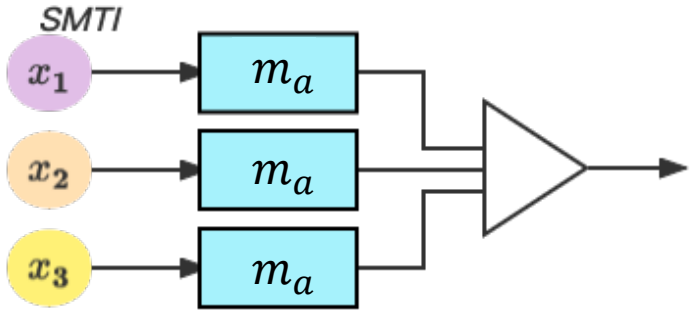
$\alpha_{b|a,2}$ と $\beta_{a,2|1}$のバランスで最も良いアーキテクチャが変わる

現実的なモデルの類似度と入力データの類似度の範囲では$DMDI_{a,1\cap b,2}$が好ましい選択

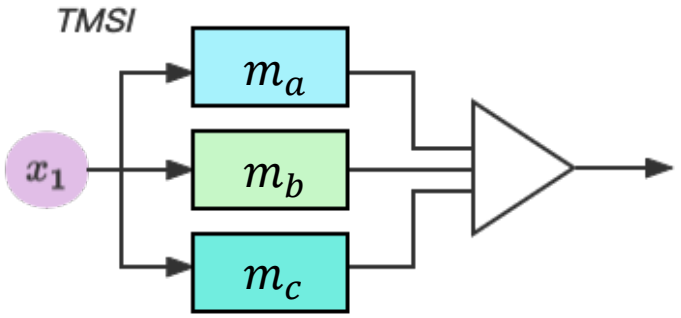
3バージョン構成のアーキテクチャ

- 3つの推論結果の多数決で最終出力を決定する

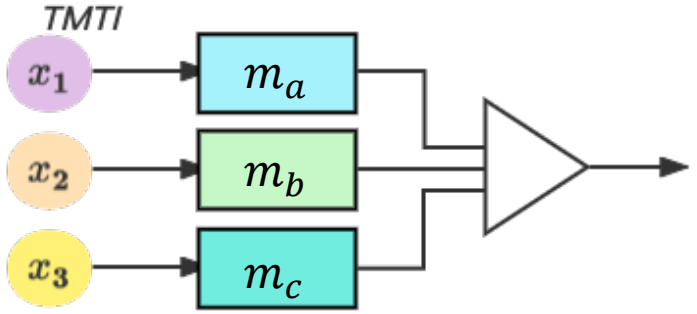
単一モデル三重入力
(Single model triple input: SMTI)



三重モデル単一入力
(Triple model single input: TMSI)



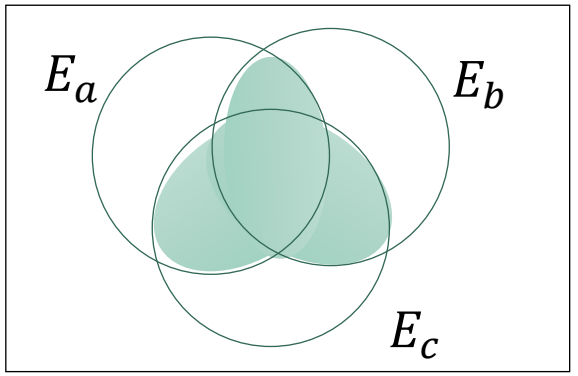
三重モデル三重入力
(Triple model triple input: TMTI)



三重モデル単一入力システムの信頼性

- エラー確率

$$f_{TMSI}(m_a, m_b, m_c; x_1)$$
$$= P[x_1 \in E_a, x_1 \in E_b] + P[x_1 \in E_a, x_1 \in E_c]$$
$$+ P[x_1 \in E_b, x_1 \in E_c] - 2\boldsymbol{P[x_1 \in E_a, x_1 \in E_b, x_1 \in E_c]}$$



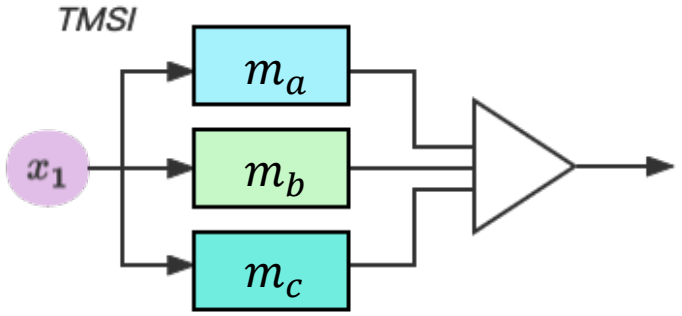
条件つき独立を仮定

$$\boldsymbol{P[x_1 \in E_a, x_1 \in E_b, x_1 \in E_c]} = P[x_1 \in E_c | x_1 \in E_a] \cdot P[x_1 \in E_b | x_1 \in E_a] \cdot P[x_1 \in E_a]$$

$$\therefore \quad f_{TMSI}(m_a, m_b, m_c; x_1)$$
$$= \alpha_{b|a,1} \cdot P[x_1 \in E_a] + \alpha_{c|a,1} \cdot P[x_1 \in E_a] + \alpha_{c|b,2} \cdot P[x_1 \in E_a] - 2\alpha_{c|a,1} \cdot \alpha_{b|a,1} \cdot P[x_1 \in E_a]$$

- 信頼性

$$R_{TMSI}(m_1, m_2, m_3; x_1) = 1 - f_{\mathrm{TMSI}}(m_a, m_b, m_c; x_1)$$



単一モデル三重入力システムの信頼性

- エラー確率

$$\boldsymbol{f_{SMTI}(m_a; x_1, x_2, x_3)}$$
$$= P[x_1 \in E_a, x_2 \in E_a] + P[x_1 \in E_a, x_3 \in E_a]$$
$$+ P[x_2 \in E_a, x_3 \in E_a] - 2\boldsymbol{P[x_1 \in E_a\ x_2 \in E_a, x_3 \in E_a]}$$

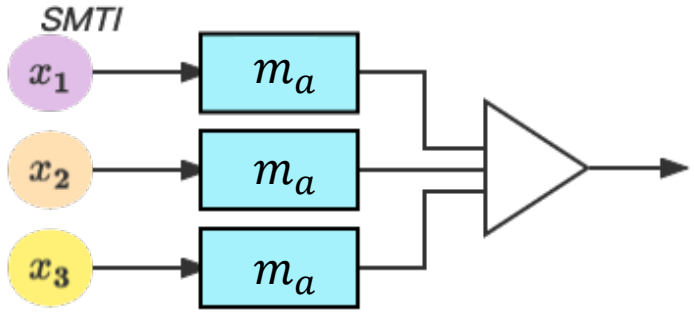
条件つき独立を仮定

$$\boldsymbol{P[x_1 \in E_a\ x_2 \in E_a, x_3 \in E_a]} = P[x_2 \in E_a | x_1 \in E_a] \cdot P[x_3 \in E_a | x_1 \in E_a] \cdot P[x_1 \in E_a]$$

$$\therefore \quad \boldsymbol{f_{SMTI}(m_a; x_1, x_2, x_3)}$$
$$= \beta_{a,2|1} \cdot P[x_1 \in E_a] + \beta_{a,3|1} \cdot P[x_1 \in E_a] + \beta_{a,3|2} \cdot P[x_2 \in E_a] - 2\beta_{a,2|1} \cdot \beta_{a,3|1} \cdot P[x_1 \in E_a]$$

- 信頼性

$$\boldsymbol{R_{SMTI}(m_a; x_1, x_2, x_3)} = 1 - \boldsymbol{f_{SMTI}(m_a; x_1, x_2, x_3)}$$



三重モデル三重入力システムの信頼性

- エラー確率

$$\boldsymbol{f_{TMTI}(m_a, m_b, m_c; x_1, x_2, x_3)}$$
$$= P[x_1 \in E_a, x_2 \in E_b] + P[x_1 \in E_a, x_3 \in E_c]$$
$$+ P[x_2 \in E_b, x_3 \in E_c] - 2\boldsymbol{P[x_1 \in E_a\ x_2 \in E_b, x_3 \in E_c]}$$

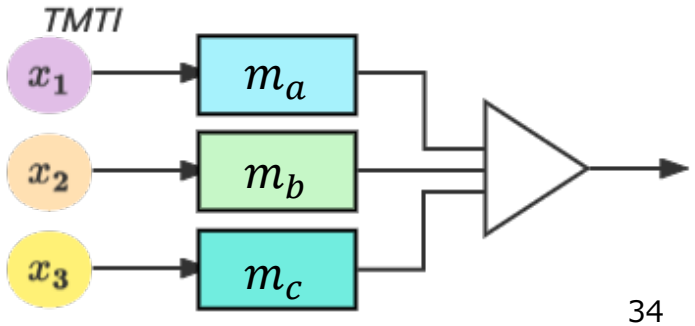
条件つき独立を仮定

$$\boldsymbol{P[x_1 \in E_a\ x_2 \in E_b, x_3 \in E_c]} = P[x_2 \in E_b | x_1 \in E_a] \cdot P[x_3 \in E_c | x_1 \in E_a] \cdot P[x_1 \in E_a]$$

$$\therefore \quad \boldsymbol{f_{TMTI}(m_a, m_b, m_c; x_1, x_2, x_3)}$$
$$= f_{DMDI}(m_a, m_b; x_1, x_2) + f_{DMDI}(m_a, m_c; x_1, x_3) + f_{DMDI}(m_b, m_c; x_2, x_3)$$
$$- 2f_{DMDI}(m_a, m_b; x_1, x_2) \cdot f_{DMDI}(m_a, m_c; x_1, x_3)/P[x_1 \in E_a]$$

$$\| $$

$$\left[\beta_{a,2|1} \cdot \alpha_{b|a,1} + \frac{(1 - \beta_{a,2|1}) \cdot (P[x_2 \in E_b] - \alpha_{b|a,1} \cdot P[x_1 \in E_a])}{1 - P[x_1 \in E_a]}\right] \cdot P[x_1 \in E_a]$$

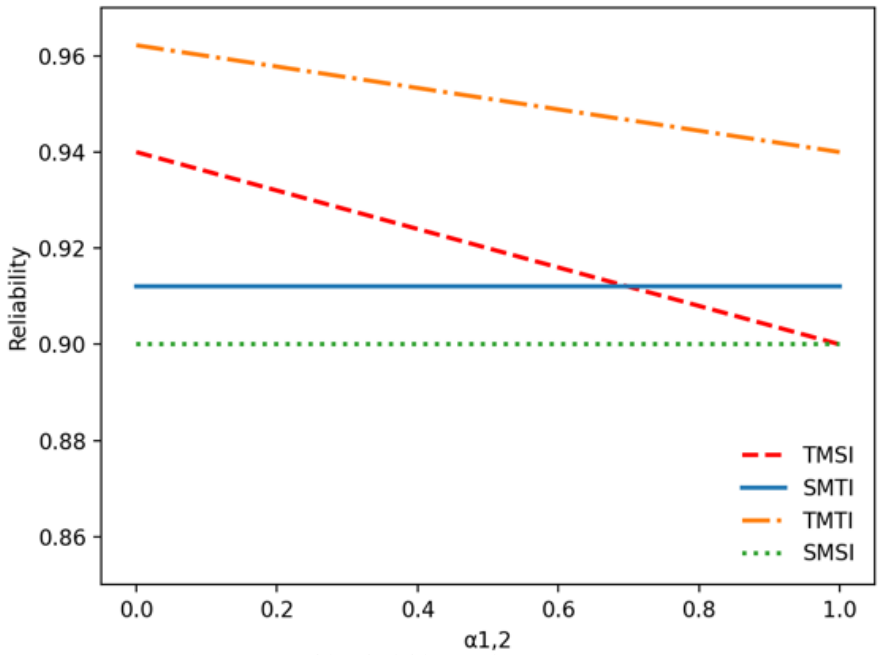


- 信頼性

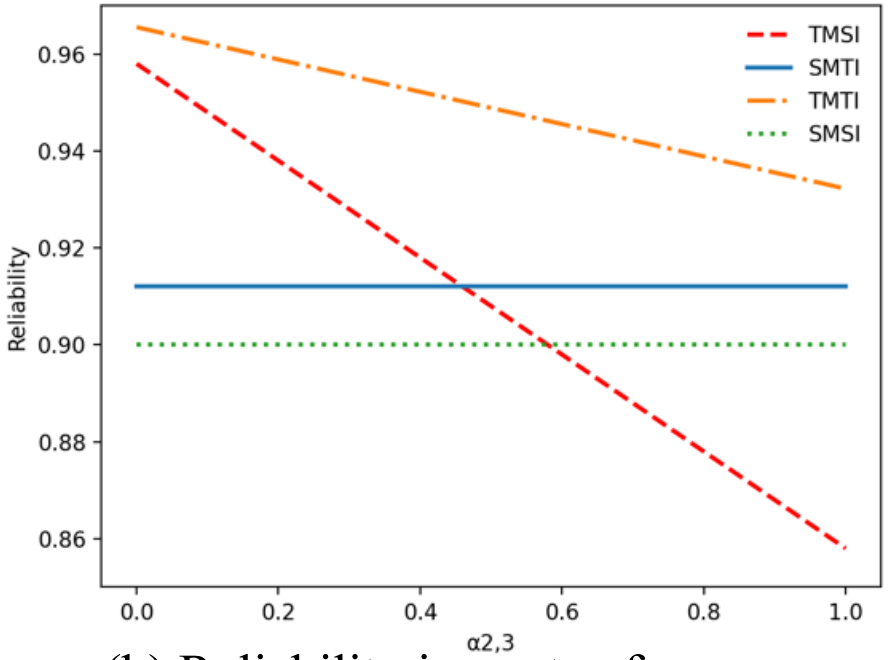
$$\boldsymbol{R_{TMTI}(m_a, m_b, m_c; x_1, x_2, x_3)} = 1 - \boldsymbol{f_{TMTI}(m_a, m_b, m_c; x_1, x_2, x_3)}$$

3バージョンアーキテクチャの信頼性比較

- モデル類似度を変化させたとき



(a) Reliability impacts of $\alpha_{b|a,1}$

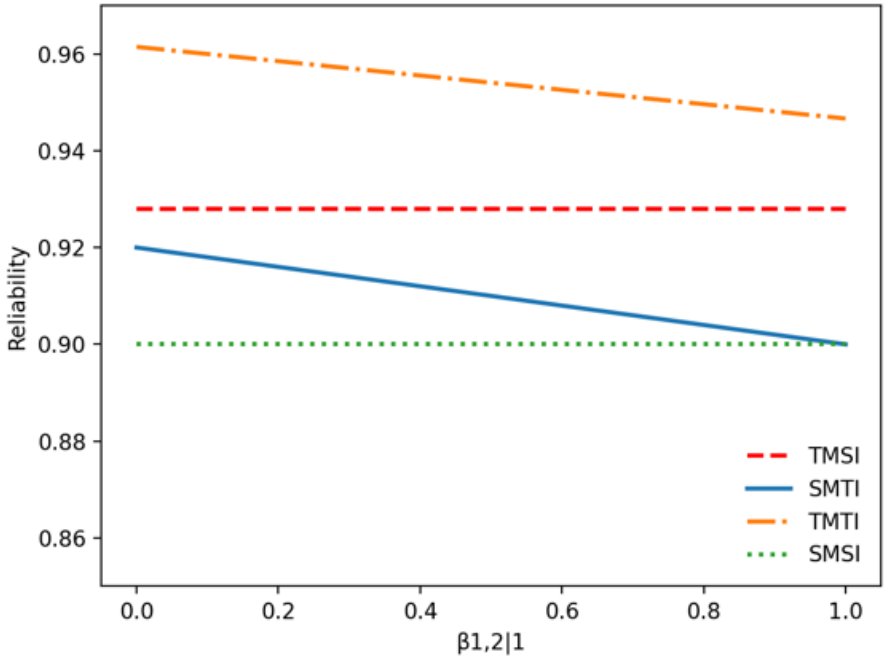


(b) Reliability impacts of $\alpha_{c|b,1}$

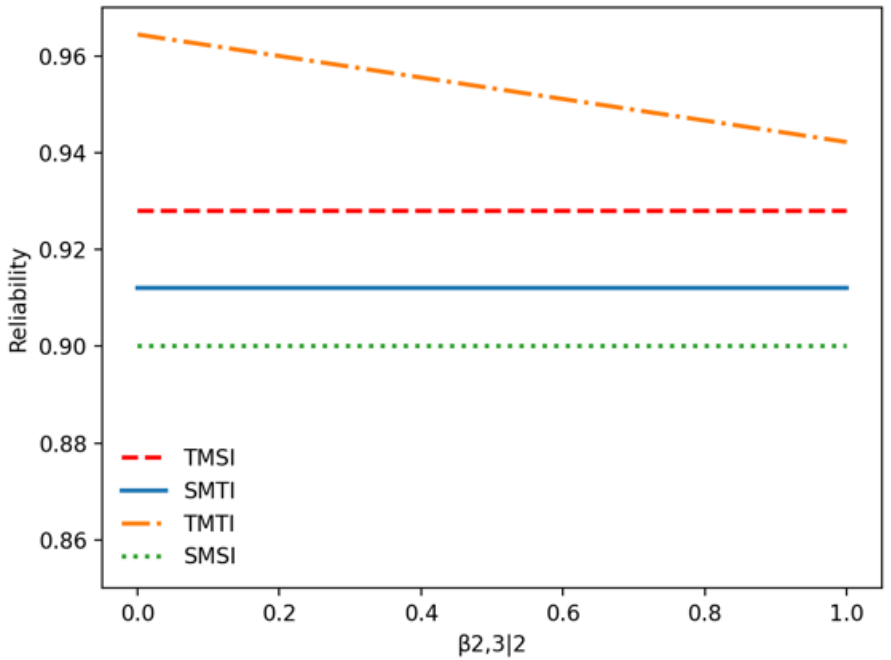
モデルの類似度$\alpha_{b|a,1}$、$\alpha_{c|b,1}$が小さい場合はTMSIとTMTIの信頼性は高まる

3バージョンアーキテクチャの信頼性比較2

- 入力データ類似度を変化させたとき



(a) Reliability impacts of $\beta_{a,2|1}$

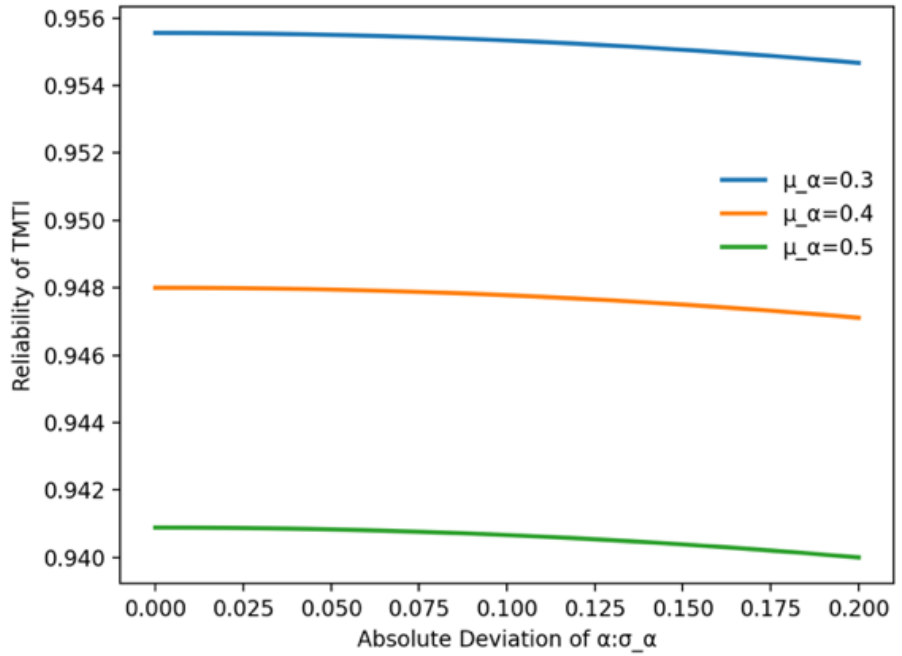


(b) Reliability impacts of $\beta_{b,3|2}$

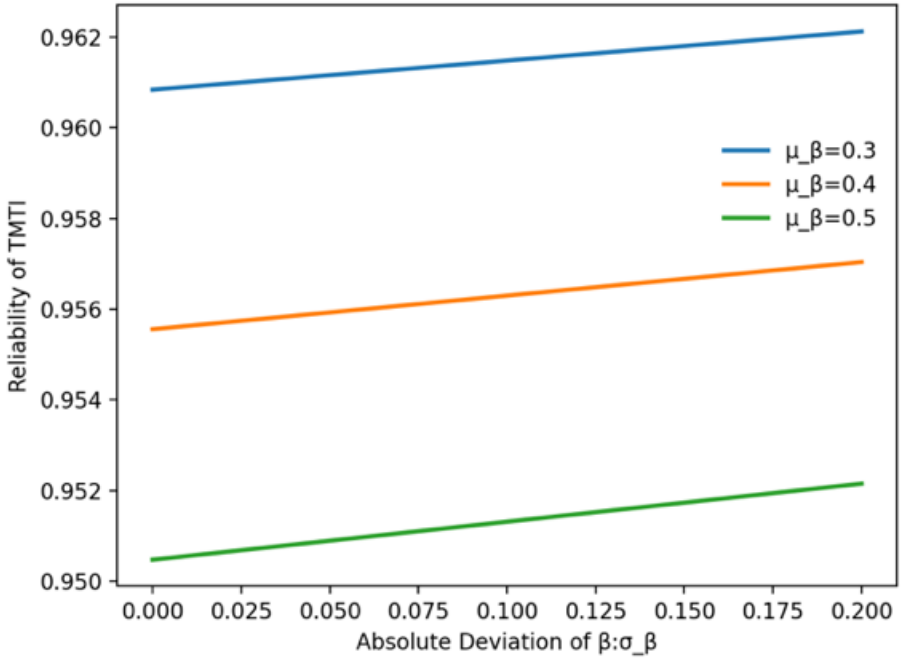
モデルの類似度$\beta_{a,2|1}$、$\beta_{b,3|2}$が小さい場合にTMTIの信頼性は高まる

多様性指標値のばらつきの影響

- 多様性パラメータのばらつきとTMTIの信頼性の関係



(a) Reliability of TMTI by varying σ_α

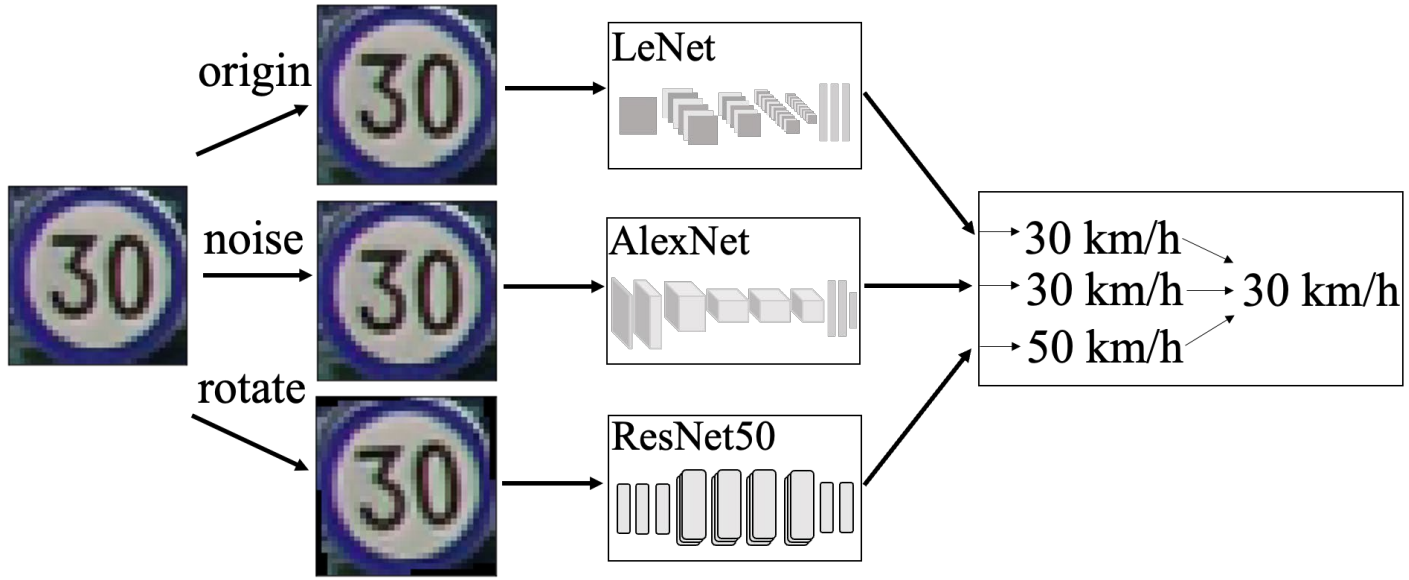


(b) Reliability of TMTI by varying σ_β

モデル類似度の分散が大きいとTMTIの信頼性は低下。
一方、入力データ類似度の分散はTMTIの信頼性向上に寄与。

交通標識分類システムでの検証

- 信頼性モデルで異なる3バージョン構成の信頼性を適切に推定できるかを検証
- 交通標識の分類タスク
  - モデル
    - LeNet
    - AlexNet
    - ResNet50
  - 入力データ
    - オリジナル
    - ノイズ追加
    - 5°回転

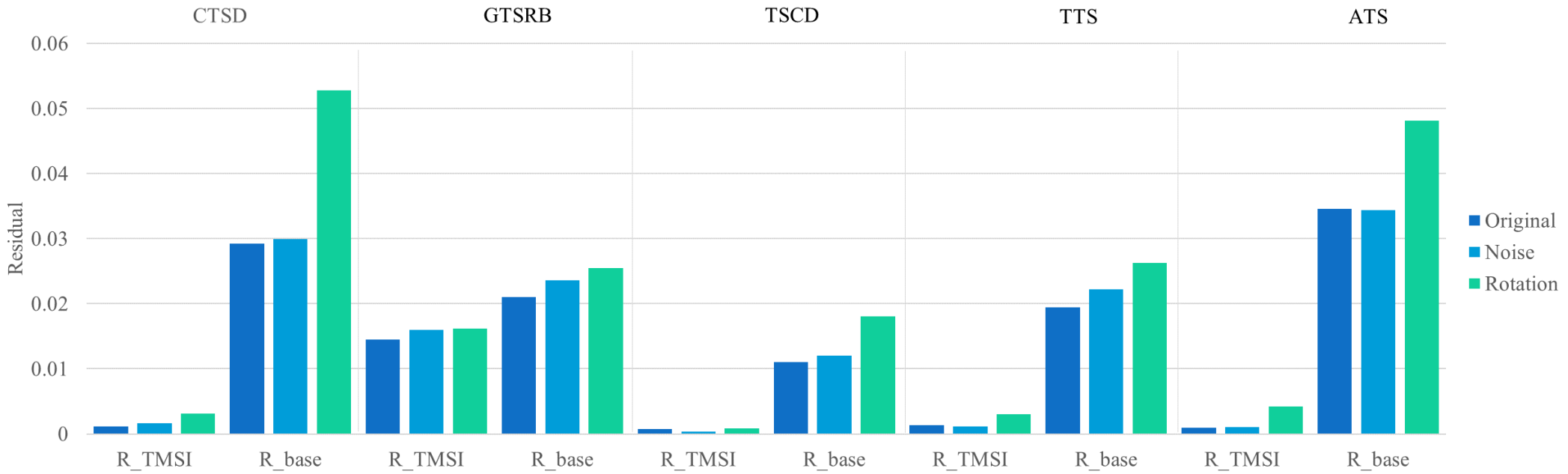


[Q. Wen, et al. ISSRE2023]

TMSIシステムの信頼性モデルの評価

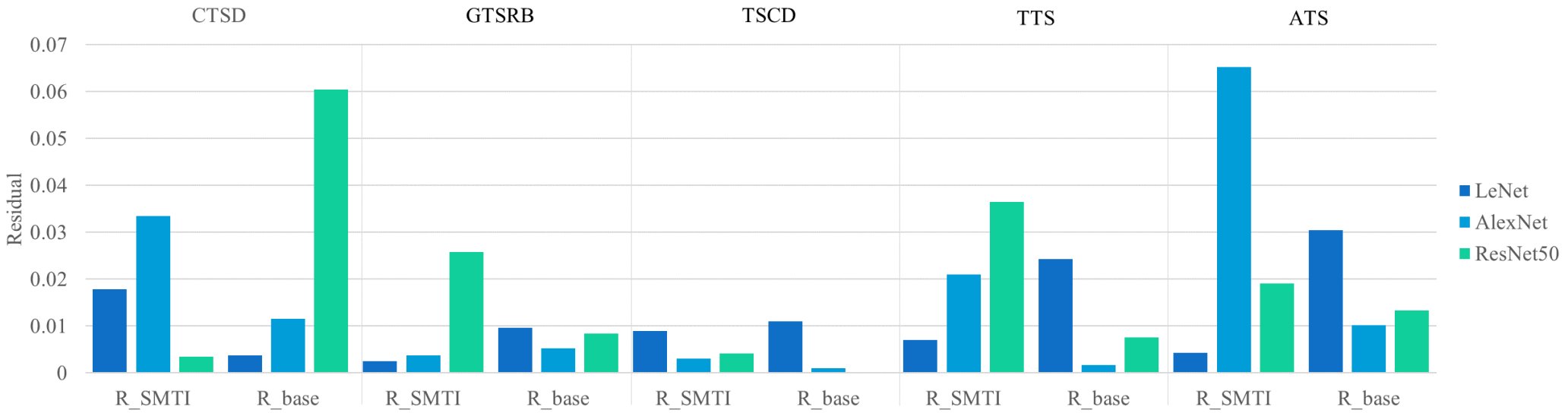
- 5つのデータセットで予測結果の誤差（residual）を評価
- 既存モデル
  - 依存関係のある3バージョンシステムの信頼性　$R_{base} = 1 - \alpha f(3 - 2\alpha)$

TMSI

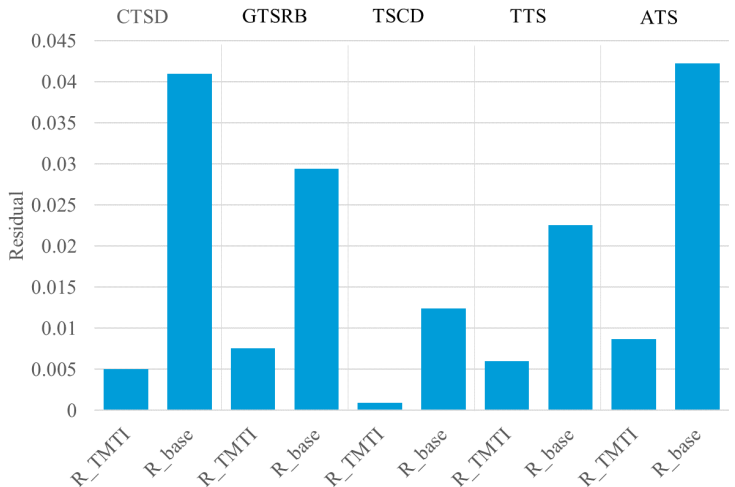


SMTI/TMTIの信頼性モデルの評価

SMTI



TMTI

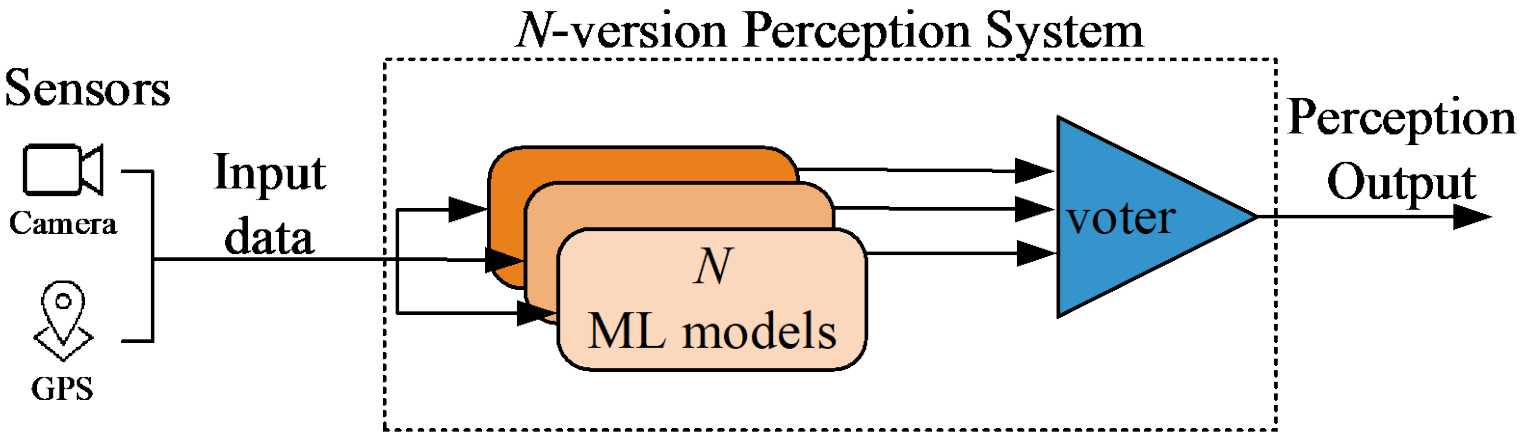


TMSIとTMTIの信頼性推定は
既存モデルよりも優れている

自動運転向け物体検出システムへの応用

- 物体検出モデルが障害や攻撃によって精度が低下すると仮定
  - 正常状態　→　劣化状態　→　故障状態
- N個の物体検出モデルを用いて信頼性を向上させる

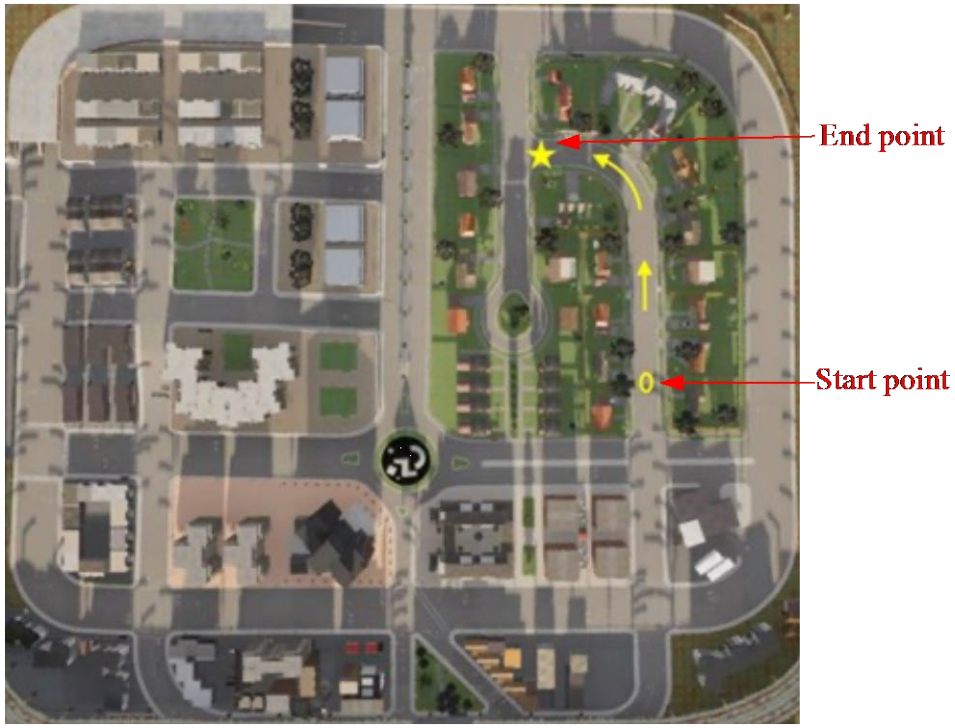
劣化状態では正しい物体検出ができない



[Q. Wen, et al. AISafety2024]

自動運転の安全性評価

- 評価環境
  - 自動運転のシミュレータCarla
  - フレームワークOpenCDA
- オブジェクト検出モデル
  - YOLOv5s6, YOLOv5m6, YOLOv5l6
- 評価指標
  - 衝突率
  - 最初の衝突までのフレーム数

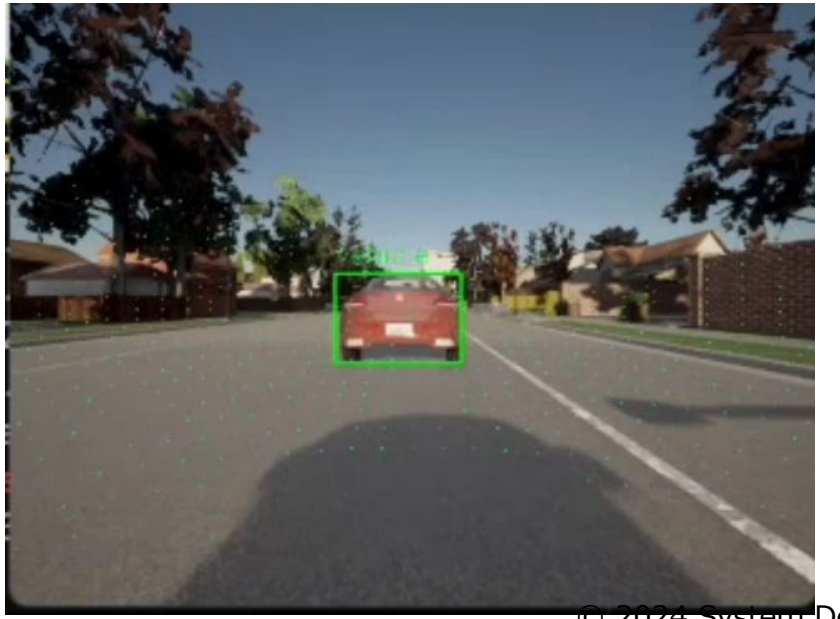


Town03 of the CARLA simulator

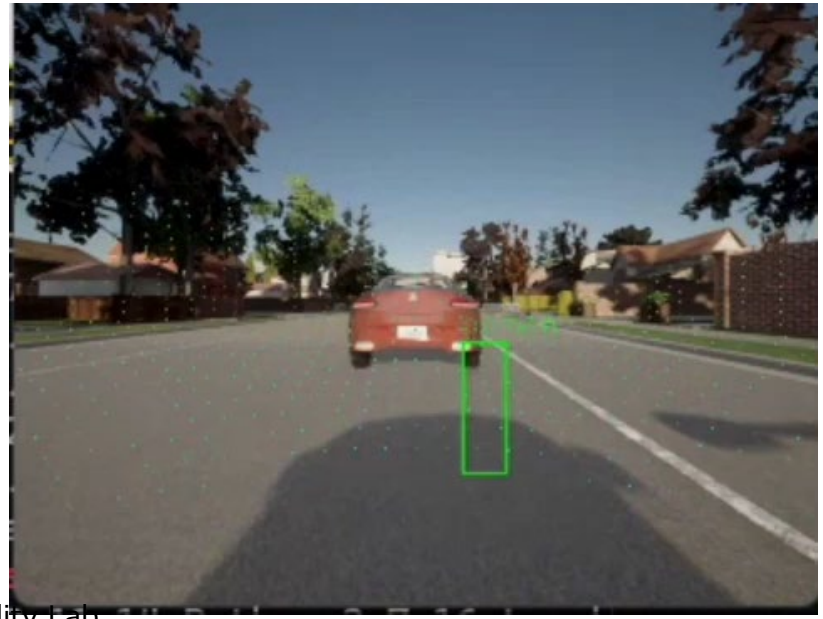
Carlaによる実験

- 故障注入により劣化モデルを生成
  - PyTorchFIでYOLOv5のパラメータをランダムに変更
- 劣化モデルによる自動運転では衝突事故が発生

正常モデルでの運転



劣化モデルでの運転



3バージョン物体検出システムの評価

- 3バージョン構成であれば、1つのモデルが劣化状態になっても自動運転の安全性を維持できる

System state	YOLO Model	1st collision frame	Total frames	Collision rate%	# Collisions
Three-version					
(3,0,0)	v5s, v5m, v5l	NA	682	0	0/10
(2,1,0)	v5s, v5m, v5m_FI	NA	693	0	0/10
(2,1,0)	v5s, v5m, v5s_FI	NA	682	0	0/10
(1,2,0)	v5s, v5s_FI, v5m_FI	272	666	28.82	5/10
(1,2,0)	v5m, v5s_FI, v5m_FI	335	654	33.08	7/10
(0,3,0)	v5s_FI, v5m_FI, v5l_FI	187	643	57.00	8/10

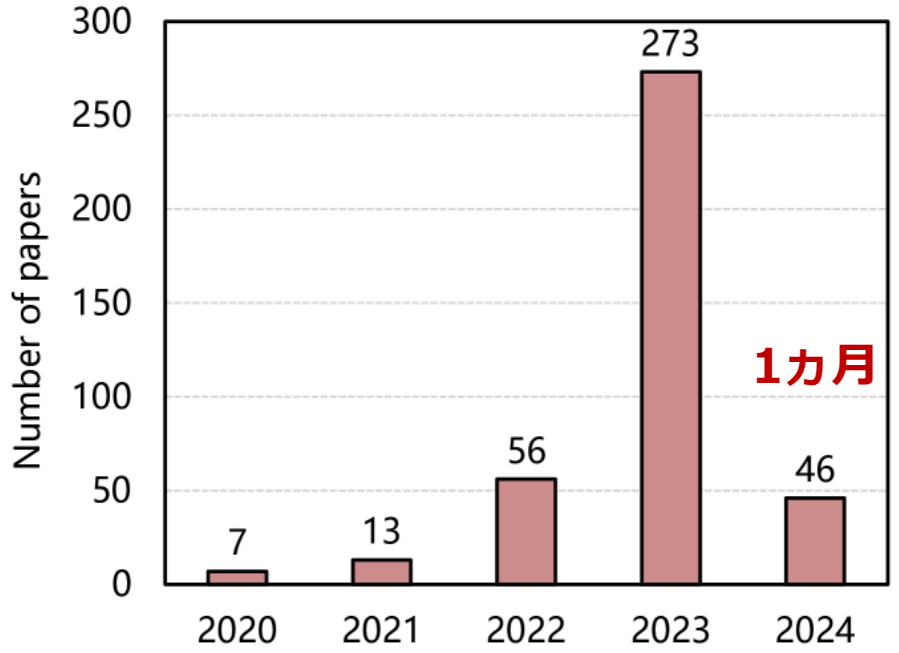
劣化したモデルの数

1つのモデルの劣化であれば衝突を防げる

ソフトウェア開発自動化への応用

LLMによるソフトウェア開発自動化

- LLMを使ったソフトウェア開発自動化の研究が近年急速に進展



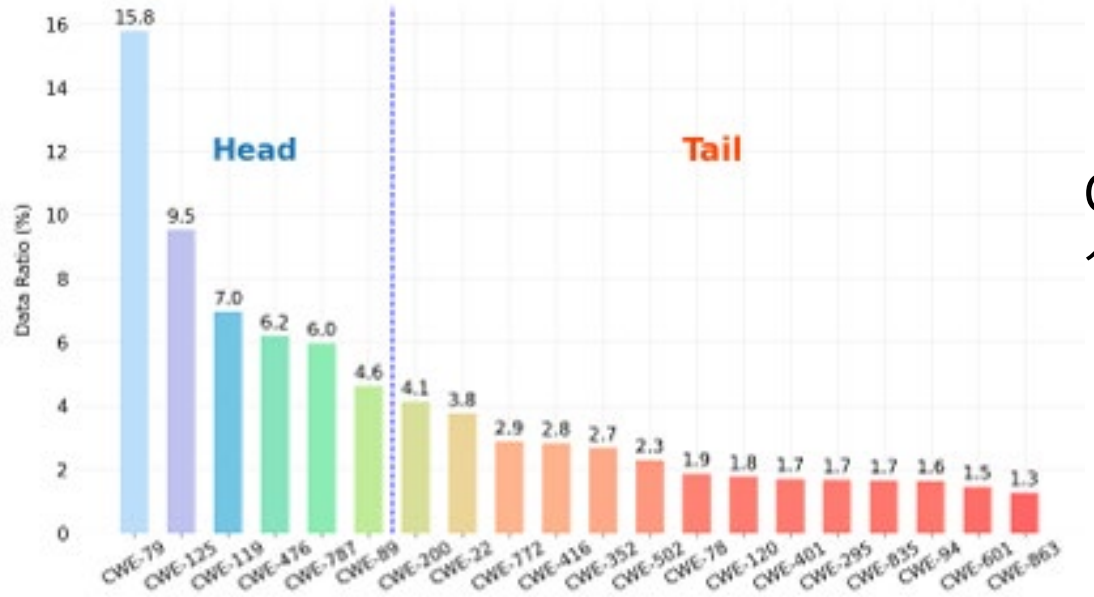
モデル	タイプ	応用
エンコーダーのみ (例：CodeBERT)	理解	コード理解 バグ箇所特定 脆弱性検出
エンコーダー& デコーダー (例：CodeT5)	理解と生成	コード要約 コード変換 バグ修復
デコーダーのみ (例：Copilot)	生成	コード生成 コード補完 テストケース生成

(source) X. Hou, et al. Large Language Models for Software Engineering: A Systematic Literature Review, 2024. https://arxiv.org/abs/2308.10620

学習データ分布の問題

- LLMを使ったコードタスクの性能は学習データに依存
- コードタスクの学習データは多くの場合ロングテール分布
  - 一部のクラスのサンプルは大量にある
  - 多くのクラスのサンプルは非常に少ない

例）



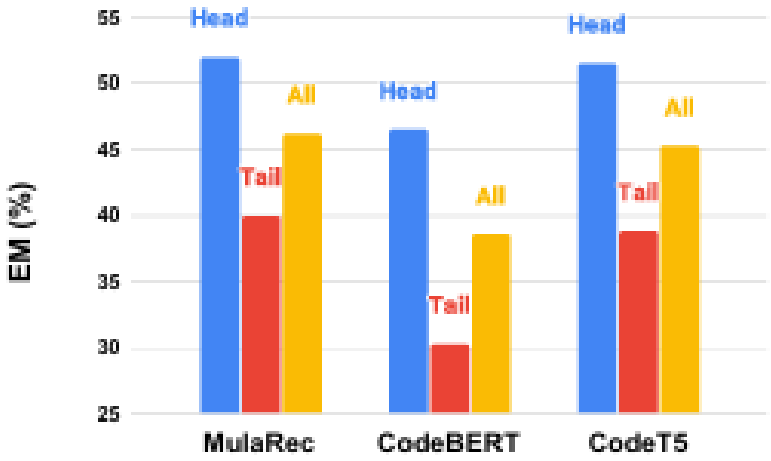
Common Weakness Enumeration (CWE)
タイプ毎のセキュリティパッチデータの分布

(source) X. Zhou, et al. The Devil is in the Tails: How Long-Tailed Code Distributions Impact Large Language Models, 2023. https://arxiv.org/pdf/2309.03567

ロングテール分布の影響

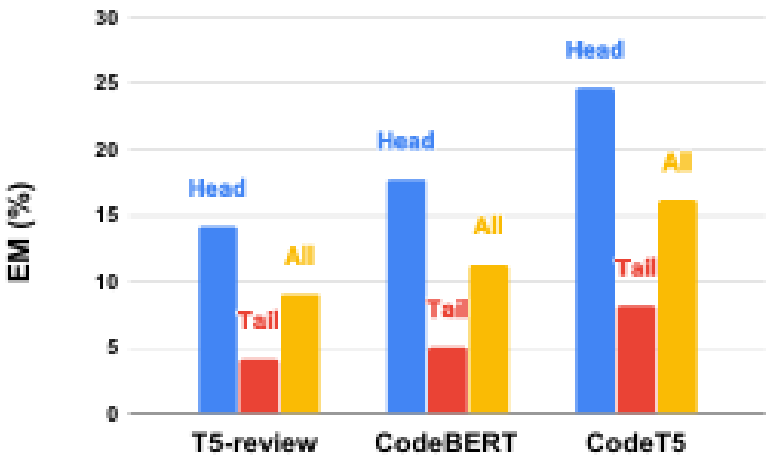
- 分布のヘッドに対する精度は良いが、分布のテールでは精度が出ない

APIシーケンス推薦



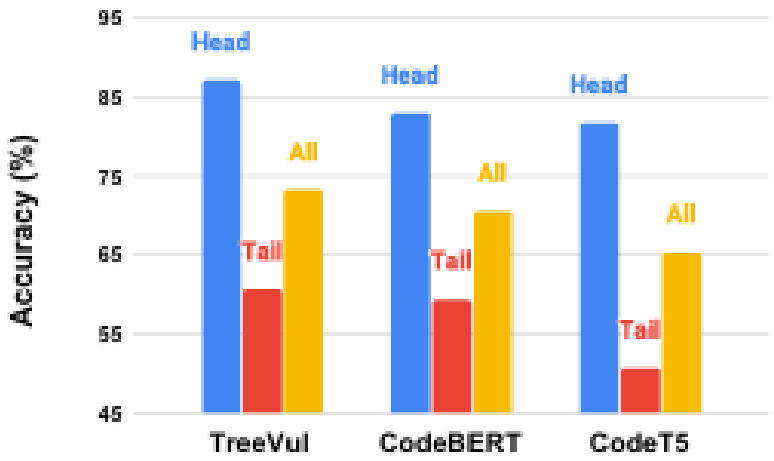
(a) API Sequence Rec.

コードリビジョン推薦



(b) Code Revision Rec.

脆弱性タイプ予測

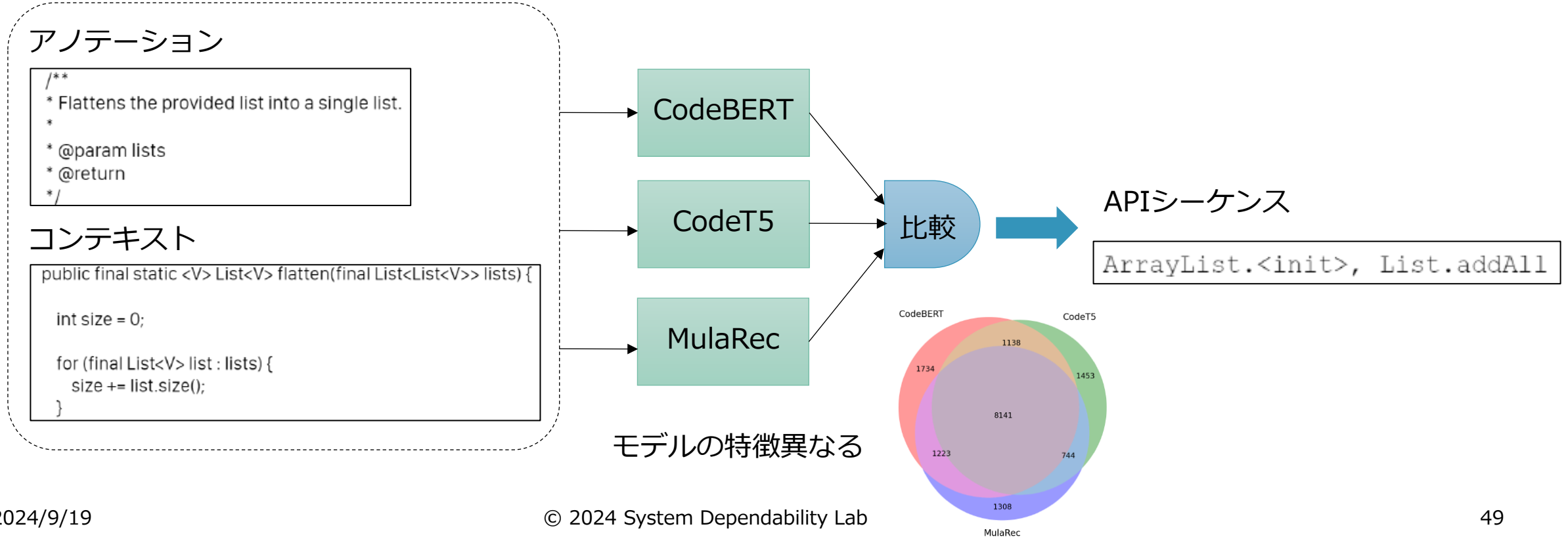


(c) Vulnerability Type Rec.

(source) X. Zhou, et al. The Devil is in the Tails: How Long-Tailed Code Distributions Impact Large Language Models, 2023. https://arxiv.org/pdf/2309.03567

複数のLLMによるテールデータへの対処

- 3バージョンAPI推薦システム
  - 3つの異なるLLMで生成されたAPIシーケンスを比較して精度を向上させる



おわりに

- Nバージョン機械学習システムの出力信頼性はモデルの多様性と入力データの多様性で特徴づけられる
- 条件付き独立の仮定でも高い精度で一部のアーキテクチャの信頼性推定が可能
  - 一般的に想定される状況においては多様なモデルと多様な入力データを組み合わせるアーキテクチャの選択が好ましい
- ソフトウェア開発自動化タスクへの応用は今後の課題

Thanks to collaborators

筑波大学
University of Tsukuba

LabSD

uni.lu
UNIVERSITÉ DU LUXEMBOURG

科研費
KAKENHI

ご清聴ありがとうございました