

1. 概要と背景
2. 提案手法
3. 実験結果
4. 結論、今後の課題

確認修正コストに基づく機械学習評価手法

**An Evaluation Method for Machine Learning Based on Verification
and Correction Costs**

尾崎 悠毘 (Haruhi Ozaki)
u320013@st.pu-toyama.ac.jp

富山県立大学 情報システム工学科

November 17, 2025

1. 概要と背景
2. 提案手法
3. 実験結果
4. 結論、今後の課題

リジェクトモデルの導入

- 機械学習システムの出力精度は 100%困難なため、リジェクトオプション/リジェクトモデルが利用される。
- リジェクトモデルでは、精度が低いと予想される結果をリジェクトし、人間が確認・修正を行う前提となる。

評価の課題

- 人間業務の代替を目的とする場合、リジェクト時の「確認コスト」と「修正コスト」を評価に反映する必要がある。

従来手法の限界

- ROC 曲線や PR 曲線は、確認修正コストを考慮できない。
- コスト曲線はコスト定数（運用形態に依存）を定める必要があり、汎用的な評価が難しい。

本論文の提案

- コスト定数に影響されず、確認修正コストを反映できる新しい評価手法（評価尺度 $W(k)$ と ARAC 曲線）を提案する。

評価尺度 $W(k)$

- コスト定数の比率 (γ, δ) を用いて、確認修正コストを最大化する（大きいほど良い）尺度を定義。



$$W(k) = \gamma X(k) + \beta - \delta R_E(k)$$

- $X(k)$ (アクセプト率): 自動化率。確認コスト削減に対応。
- β (認識率): 予測器の正解率 (閾値 k に依存しない)。
- $R_E(k)$ (エラー率): 誤りの見逃し割合。修正コストに関連。

曲線の定義

- 横軸 (X 軸): アクセプト率 $X(k)$ (自動化率)
- 縦軸 (Y 軸): 修正後正解率 $Y(k)$ (エラー率 $1 - Y(k)$)

ARAC 曲線の利点

- コスト定数なしで描画可能。
- 横軸は確認コスト、縦軸は修正コストに密接に関連し、両者を視覚的に表現。
- 右上に位置するほど良いモデルと評価できる。

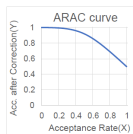


図 3: ARAC 曲線. 横軸にアクセプト率 (自動化率), 縦軸に修正後正解率を閾値 k を変化させてプロットした曲線.

ARAC-AUC と課題

- 曲線下の面積 A_{ARAC} を AUC とする。
- 認識率 β が高い場合、AUC が自動化率 $X(k)$ の改善を忠実に反映しにくい課題がある。

正規化 ARAC-AUC (A''_{ARAC})

- $W(k)$ の重み (γ) と認識率 (β) を考慮し、AUC を改良・正規化。
- A''_{ARAC} は、コスト評価尺度 $W(k)$ の変動をより正確に反映し、最大値が 1 となる。

ARAC-AUC vs 従来 AUC

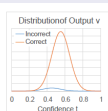
- 予測器の正予測・誤予測分布が異なる人工データ (Case 1, 2, 3) で実験。
- 確認修正コストに基づく $W(\theta)$ の順序 : $1 > 3 > 2$
- ARAC-AUC / 正規化 ARAC-AUC の順序 : $1 > 3 > 2$
- ROC-AUC / PR-AUC の順序 : $W(\theta)$ の順序とは異なる。

結論

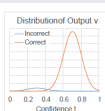
- ARAC-AUC は、ROC-AUC や PR-AUC と比較して、確認修正コストに基づく評価尺度をより忠実に反映できる。

認識率 (β) が高いケース ($\beta = 0.95$)

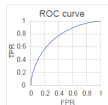
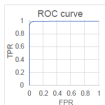
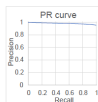
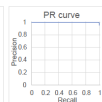
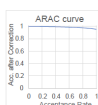
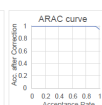
- 認識率が高いケース (Case 4, 5) では、ARAC-AUC では評価尺度 $W(\theta)$ の差がほとんど反映されない。
- **正規化 ARAC-AUC** は、認識率が高い領域でも、自動化率 $X(k)$ の差を適切に評価に反映し、実際の $W(\theta)$ の差をより良く表現できる。



(a) ケース 4 分布

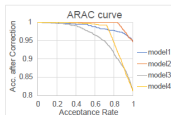


(b) ケース 5 分布

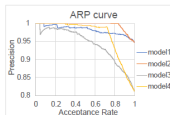

(c) ケース 4
ROC 曲線

(d) ケース 5
ROC 曲線

(e) ケース 4
PR 曲線

(f) ケース 5
PR 曲線

(g) ケース 4
ARAC 曲線

(h) ケース 5
ARAC 曲線

OCR モデル選択

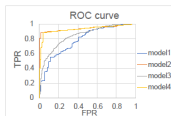
- 4 種類の OCR モデル（model 1～4）を比較。
- ARAC-AUC によるモデル選択の順序: $2 > 1 > 4 > 3$ 。
- **ROC 曲線**は認識率を無視し、**PR 曲線**はアクセプト率が低い領域で不安定となり、リジェクトモデルの評価には適さない。



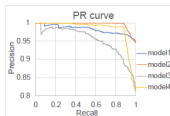
(a) ARAC 曲線



(b) ARP 曲線



(c) ROC 曲線



(d) PR 曲線

本論文の主な貢献

- 確認修正コストを反映した評価尺度 $W(k)$ の定義。
- ARAC 曲線および正規化 ARAC-AUC の提案。
- ARAC 曲線は、コスト定数が不要で描画でき、運用形態決定前のモデル評価を可能にする。
- 従来の ROC/PR 曲線に比べ、確認修正コストの観点から安定的な評価が可能であると実証。

結論

- ARAC 曲線は、確信度に依存して処理方法を選択する機械学習システムの評価において、有効な評価指標となる。

今後の課題

- ARAC 曲線の具体的な利用方法の検討。
- 確信度を出力する機械学習システムの評価方法における、ARAC 曲線の具体的な利用方法の検討