大学共同利用機関法人 情報・システム研究機構
国立情報学研究所
National Institute of Informatics

トップエスイー ソフトウェア開発実践演習

XAI の定量的活用手法としての画像分類結果の正誤判定

落合 涼太 †1 / 金田 和之 †2 / 川俣 良太 †3 / 細包 康喜 †4
†1. 富士通株式会社 / †2. 福島キヤノン株式会社 / †3. 株式会社日立製作所 / †4. 日本ユニシス株式会社

課題

・XAI の定量的な活用方法に関する先行研究がない
　・そもそも現状 XAI は活用方法が確立されていない

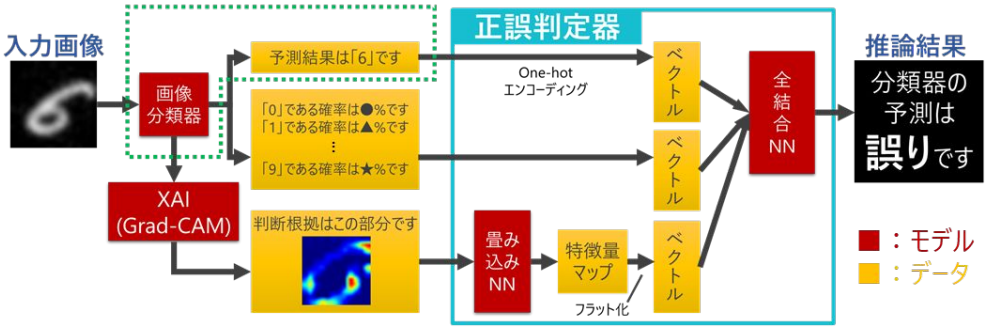
・主な活用のされ方は最終的に人間による解釈が必要
　⇒ 定性的な活用方法

取り組んだこと

・XAI の定量的活用のためのアイデア出し
　⇒ 実装することで定量的な活用ができるか評価

・以下の手法及びツールで実装，評価を実施
　・モデル構築　　：Pytorch を利用
　・XAIアルゴリズム　：Grad－CAM
　・データセット　　：MNIST
　・データ拡張　　：ホワイトノイズ，水平移動等

モデル



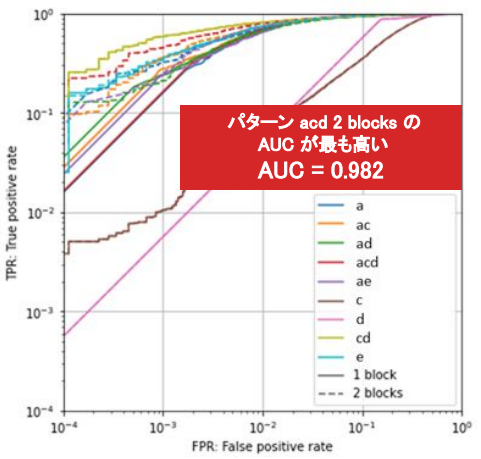
通常の分類は ⬚ の範囲.
当然、誤って判断されることもある.

画像分類器の予測結果に
XAIから抽出した追加情報 を加え,
分類の **正誤を判定するモデル** を構築.

正誤判定器の性能評価

- パターンacd 2 blocks
  - accuracy = 0.947
  - precision = 0.931
  - recall = 0.973

	誤分類と判定	正分類と判定
真の値が誤分類	6437	861
真の値が正分類	476	17227



- 各パターンの性能評価
  下記の説明変数を含むパターンの
  性能が高くなる傾向があった
  - a）Grad-CAMの出力値
  - c）分類器の予測クラスの確信度
  - d) 分類器の予測クラスのインデックス

ランダム (AUC = 0.5) よりも性能の良い判定機を構築でき XAI の結果を定量的に活用できた

判定器のユースケース

- ユースケース1: 製品検査などにおいて誤分類を減らすためのセーフティネット
  - 分類器のみ:
    誤分類 1%
  - 分類器 + 判定器:
    判定誤分類の 3.54% を人間が
    再チェックすることで、誤分類0.12% に低減
- ユースケース2: 分類機の品質保証・精度確認
  分類機の精度 99% 以上を確認するためには..
  - 分類器のみ:
    分類結果をランダムに抽出し、最低298個の分類結果のチェックが必要
    (精度 99% 以下という帰無仮説を有意水準 5% で棄却するため)
  - 分類器 + 判定器:
    判定誤分類からランダムに抽出し、最低11個の分類結果のチェックで十分

今後の課題

データセット / XAI などの一般化

- 現状
  - CIFAR-10を用いた検証でも成果が出ることを確認済み。本手法は一定の汎用性を持っている
- 課題
  - MNIST, CIFAR-10などの検証用データに留まらず、実際の業務データへの適用に関する検証が必要