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

機械学習システムにおける訓練モデル解釈による要求の明確化

NECソリューションイノベータ株式会社　川越 佑太　　株式会社デンソー　浅野 正義

機械学習の要求分析における問題点

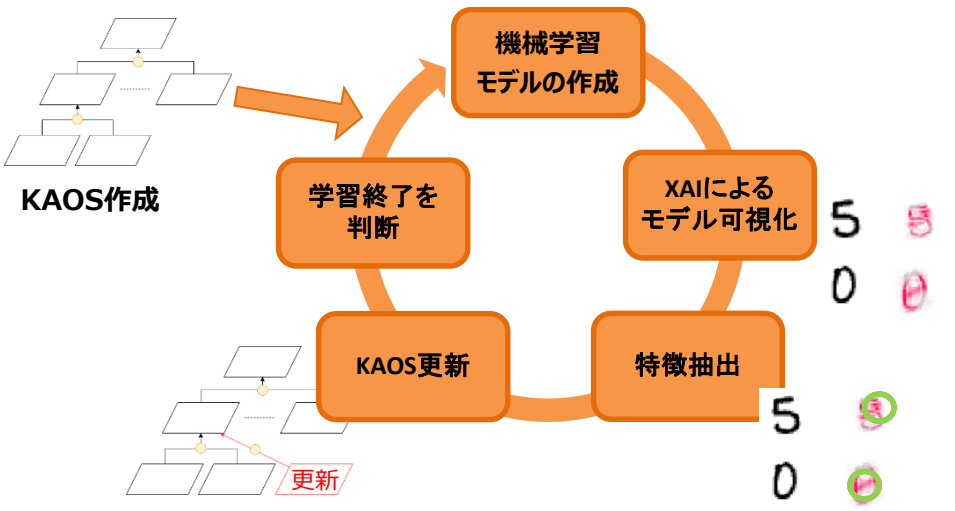
要求分析によって、とらえるべき特徴を列挙したとしても、分析結果通りに学習をさせることができないため、訓練モデルがとらえるべき特徴を要件とした開発ができない

手法・ツールの適用による解決

ゴール指向分析とXAIを組み合わせ、モデルがとらえた特徴をゴール指向分析に反映
学習時に段階的なゴール指向分析の更新を行うことにより、とらえるべき特徴に着目して学習の進捗を把握し、訓練モデルがとらえるべき特徴を要件とした開発を行う

提案手法

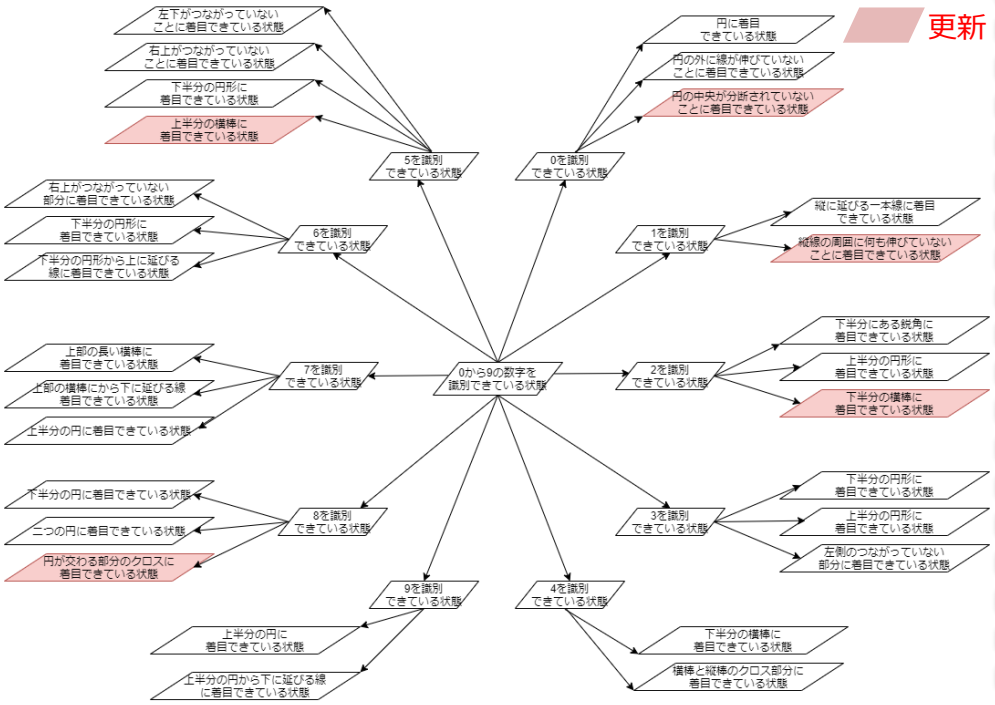
1. **KAOS法による要求分析**
   モデルが学習するべき特徴をサブゴールとして抽出
2. **機械学習モデルの作成**
   収集したデータセットから機械学習モデルを作成
3. **XAIによるモデル可視化**
   作成したモデルがとらえた識別根拠を可視化
4. **特徴抽出**
   可視化したモデルがとらえている特徴を抽出・解析
5. **KAOS更新**
   抽出した特徴をKAOS分析に追加
6. **学習終了を判断**
   サブゴールの充足度を判定し、学習終了を判断



実験結果

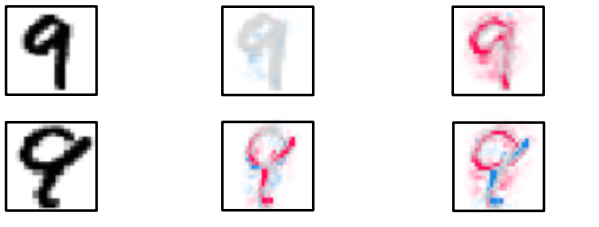
手法をMNIST識別モデルの作成に適用

学習に伴いKAOSを更新 ⇒ モデルがとらえるべき特徴を学習していることを確認



抽出できなかった特徴

誤識別結果から不足している特徴を抽出できなかった例



上記の例のように
XAIは人が理解できない特徴を抽出することもあるため、特徴を理解する為の分析を組み合わせる必要がある

今後の取り組み

今後は実システムを想定して本手法を適用することで、システム要件を満たすモデルを作成するという視点で、本手法の有効性を確認していく