機械学習をつかったサービスの品質を担保する
高信頼な機械学習応用システム

石川 冬樹、杉山 麿人、関山 太朗、吉岡 信和 他

https://qaml.jp/

どんな難しさ？

機械学習を用いて構築したシステムの振る舞いは、原則不完全で不確かなものとなります。このために、従来のソフトウェア工学におけるアプローチが通じないことがあります。

どんな活動？

産業界の課題やプラクティスを集約し議論する場づくりなどコミュニティ活動を行うとともに、JSTの未来社会創造事業のもと研究開発に取り組んでいます。

背景

訓練データ → 学習アルゴリズム → 知的機能の実装

予測，推薦，検出，意思決定等

従来は人（エンジニア）が書き下していた動作ルールを訓練データから生成する（「答えを出す真似事をする」）

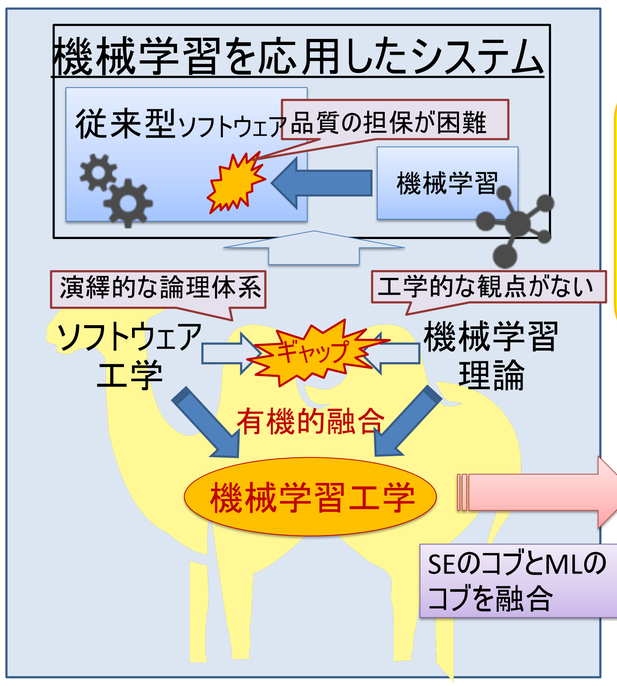
従来とは異なる性質

- 不完全で誤りをする可能性が常にある
- 論理的に挙動を予測することは難しく試してみてはじめて挙動がわかる
- どうしてその出力が得られたのか，説明することは難しい
- そもそも明確に正解を定義できない問題に適用することが多い

研究

目的： 機械学習を組み込んだソフトウェア・システムの品質を担保するための**機械学習工学を確立**しその**普及**を目指す

概要： ソフトウェア工学における論理的、演繹的アプローチと機機学習における帰納的モデル構成、統計的アプローチとを**融合**させることにより、**機械学習応用システムの品質と価値を向上**させる。



学術的チャレンジ

- 演繹と帰納のパラダイム融合
- 決定的モデルと確率・統計的理論の融合
- 機械学習とシステムのライフサイクルの融合

"きゃめる"

QAML

未来社会創造事業での研究推進

QAMLプロジェクトをJST 未来社会創造事業 探索加速型「超スマート社会の実現」領域（２０１８年度採択課題）として研究を推進中

大型研究として提案中の体制

2019年度の体制

学術：16人（6組織）
国立情報学研究所
九州大学
東京電機大学
日本工業大学
北陸先端科学技術大学院大学
早稲田大学

産業界：18人（12社）
株式会社 Udzuki
株式会社ウェザーニュース
株式会社ウルフ
株式会社エヌ・ティ・ティ・データ
NTT テクノクロス株式会社
グリー株式会社
株式会社富士通研究所
株式会社 Preferred Networks
株式会社デンソー
日本電気株式会社
株式会社日立製作所
ライフマティックス株式会社
有人宇宙システム株式会社(JAMSS)

4月の定例会にて

コミュニティ活動

研究会活動

- 産業界の参加者を中心としたワークショップを定期的に開催
- 各学会との連携イベント

ワークショップの参加・発表募集中

機械学習工学研究会 MACHINE LEARNING SYSTEMS ENGINEERING
https://sites.google.com/view/sig-mlse/
（twitter, Facebook, connpassで情報発信中）

コンソーシアム活動

AIプロダクト品質保証コンソーシアム
http://www.qa4ai.jp/

ガイドラインを5月に公開

ガイドラインや社会への提言などに向けた活動中

NII
連絡先： 吉岡 信和 （国立情報学研究所 アーキテクチャ科学研究系 准教授）
Email： nobukazu@nii.ac.jp Web：http://research.nii.ac.jp/~nobukazu/

機械学習をつかったサービスの品質を担保する

高信頼な機械学習応用システム（研究事例）

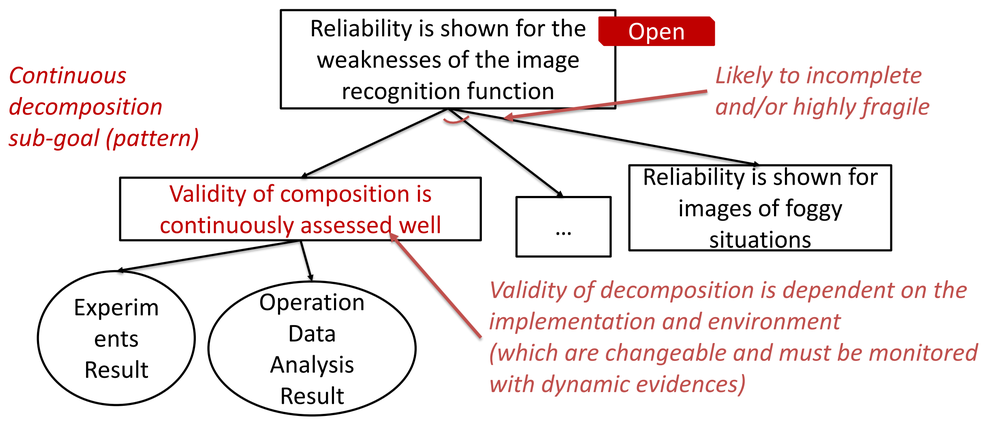
https://qaml.jp/

例：要求技術の場合

機械学習に関する不確かさ（どこまでの精度で認識できるかどうか不明など）やリスクを踏まえた要求の獲得や不明な部分の記述方法を提案

不確かさ・絶え間ない変化を踏まえたディペンダビリティ論証

[Ishikawa+, ASSURE'18]

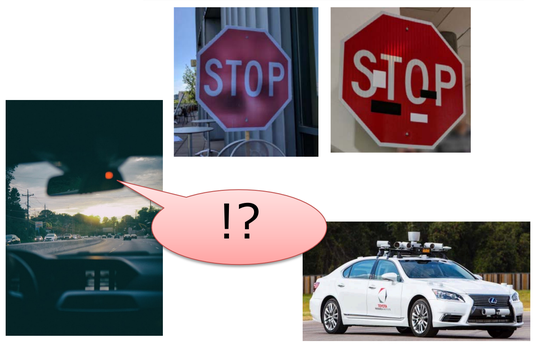


機械学習による実装による不確かさのリスクを明記し，監視機構とリンクして継続更新

例：設計技術の場合

機械学習システムのためのセキュリティ設計

誤認識する書き換え例[1]



[1]Eykholt, K., Evtimov, I., Fernandes, E., Li, B., Rahmati, A., Xiao, C., … Song, D. (2017). Robust Physical-World Attacks on Deep Learning Models.

＜背景＞
機械学習は様々なシステムで活用されているが、セキュリティの問題が発生する可能性。特に自動運転ではセキュリティは重大な問題

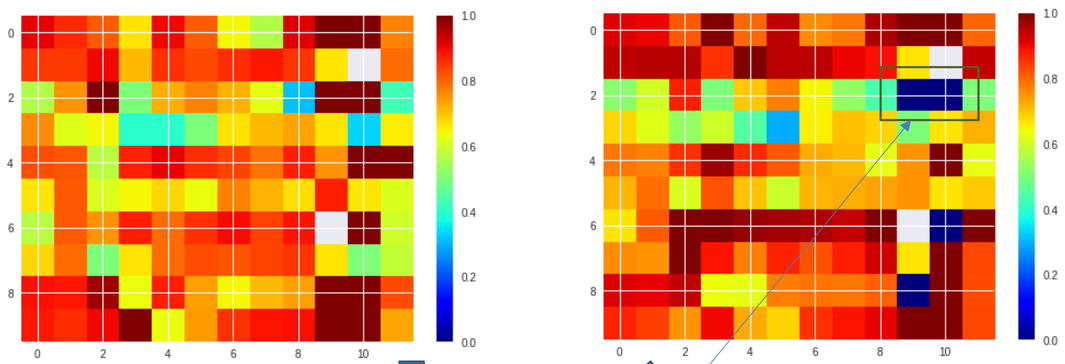
＜研究課題＞
- 訓練データ・センサーデータの悪意のある書き換えの自動検知
- 学習モデルの脆弱性の排除
- 機械学習モデルとプログラムとの組み合わせによるフェールセーフ設計

例：機械学習の成長を支える技術

工場用　敷地内用　公道用

リスクの低い学習しやすいサービスから初めて、高度な学習が必要なサービスに進化させる

要求・想定に基づいた密な精度評価・テスト



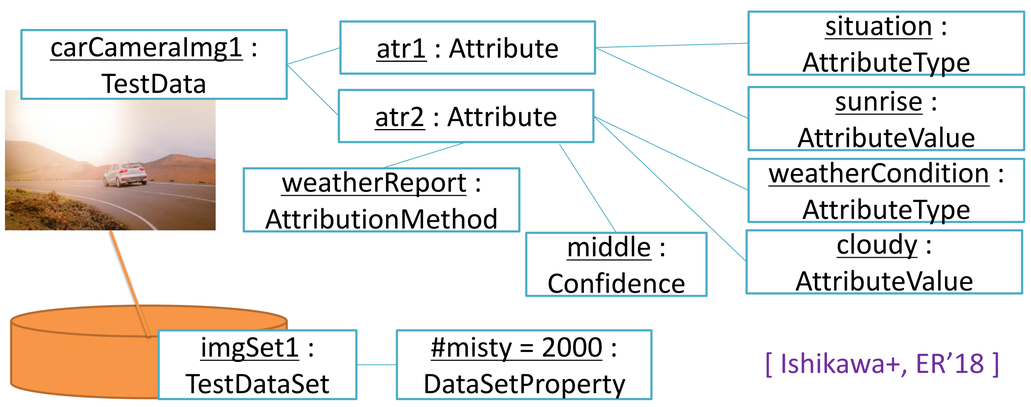
各領域は「赤い車」など属性ごとの精度

精度78%　精度82%

データの水増しで総合精度向上！（一般的な評価）
→ より密に見ると「苦手領域」が変化・増加
→ 要求との照らし合わせにより評価，再訓練

例：テスティング技術の場合

データに対する仕様・実態の記述・管理



[Ishikawa+, ER'18]

データやデータセットに対する属性を記述・管理
仕様に対して検証，変化を受けて再検証

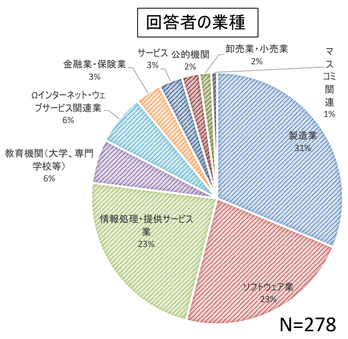
進行中：　（特にディープニューラルネットワークの）分析・説明・探索テストの手法と融合へ
「要求の観点から重要」＋「AIが見ているポイントから重要」をおさえたテスト技術へ

ニーズの調査

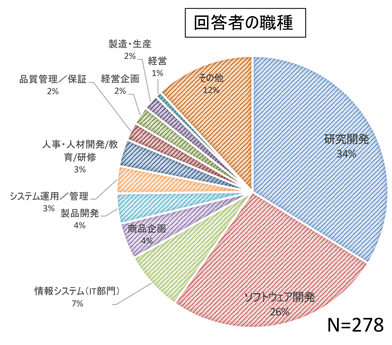
MLSE研究会でのアンケート

アンケート集計方法：Webのフォームによる記入
アンケート配布方法：MLSEコミュニティ、ソフトウェア工学関連の研究会、人工知能学会、データサイエンティスト関連のメーリングリスト、SNSで案内を通知
アンケート実施期間：2018年6月-7月
有効回答数：278回答

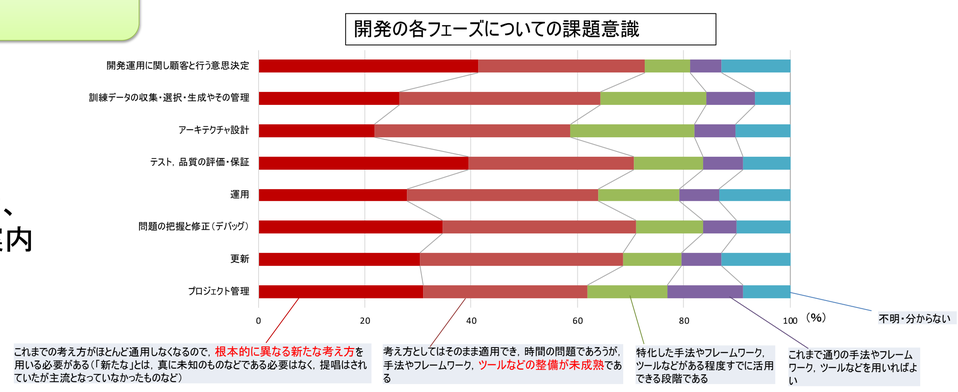
回答者の業種



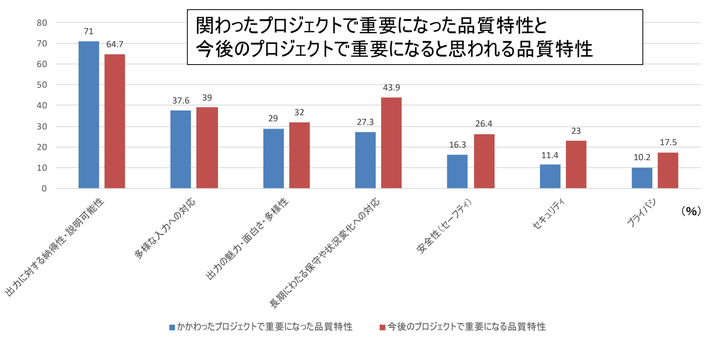
回答者の職種



開発の各フェーズについての課題意識



関わったプロジェクトで重要になった品質特性と今後のプロジェクトで重要になると思われる品質特性



インタビュー調査

機械学習応用システムを開発している企業（ベンダー企業：2社）と機械学習応用システムの開発を委託し、活用している企業（ユーザ企業：2社）にインタビュー調査（2019年2月）

機械学習応用における課題例

1. データや学習したモデルの品質保証
   現状はまだ各社とも良いデータを集めること、学習モデル、分析ロジック、学習アルゴリズムの正しさを判定することに対しては個別に対応している状況である。汎用的な基準や指針が必要
2. データやモデル、技術の引き継ぎ
   一般のシステム開発では設計データや試験データ等を決められた形で残し、後々の運用・改善時に役立てるケースが多いが、機械学習を応用したシステムでは、そもそも何をどういう形式で残せば良いかというガイドラインが十分には整備されていない
3. 説明性・解釈性
   作成した機械学習アルゴリズムが誤った答えを出す場合もある。そうした場合、原因の把握は容易ではない。多くの機械学習アルゴリズムは、教師データを使って学習し、仮想の"ブラックボックス"内で予測を行うため、その過程をたどることが難しい
4. データやモデル、技術の引き継ぎ
   一般のシステム開発では設計データや試験データ等を決められた形で残し、後々の運用・改善時に役立てるケースが多いが、機械学習を応用したシステムでは、そもそも何をどういう形式で残せば良いかというガイドラインが十分には整備されていない

NII
連絡先：　吉岡　信和　（国立情報学研究所　アーキテクチャ科学研究系　准教授）
Email：　nobukazu@nii.ac.jp　Web：http://research.nii.ac.jp/~nobukazu/