

機械学習を活用したソースコード静的解析 ツールによる解析結果の判定支援手法

日本電気株式会社

赤堀 文隆

f-akahori@bu.jp.nec.com

第三者検証サービスにおける問題

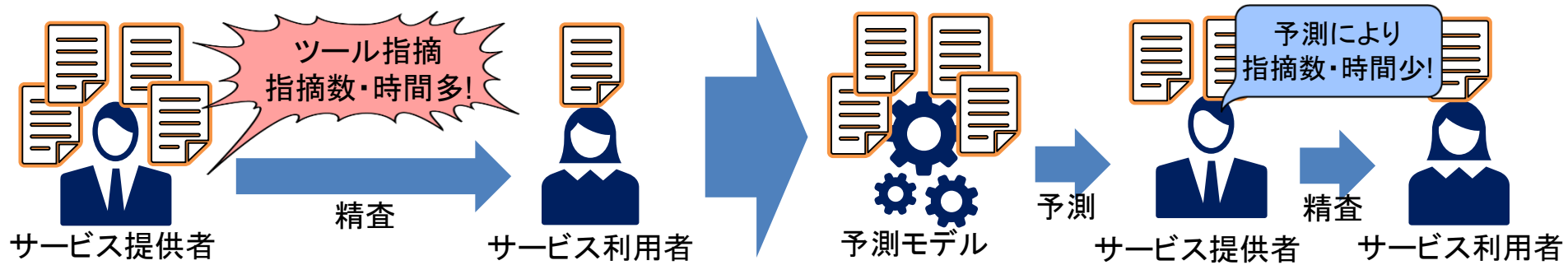
第三者検証サービスでは、ソースコード静的解析ツール等を活用してサービスを提供する。ソースコード静的解析ツールは、人手では検出が困難なソースコード上の不具合可能性を検出する。解析の特性上、誤検知となる指摘を多数含むため、サービス提供者が指摘の精査を行う場合に多大な時間がかかる問題がある。

手法・ツールの適用による解決

機械学習を用いて、ソースコード静的解析ツールの指摘が正しい検知または誤った検知であることを予測し、ツールの解析結果の確認を支援する仕組みを提案する。

研究課題とアプローチ概要

現状では、サービス提供者が多くの指摘を精査した後、サービス利用者にレポートを提供。機械学習を活用することで、サービス提供者の精査時間短縮に繋げる。



提案手法の概要

特徴量として、ソースコード静的解析ツールの解析結果から抽出できる情報を使用。学習アルゴリズムはMALSSを活用して選定し、ランダムフォレストを採用。

特徴量

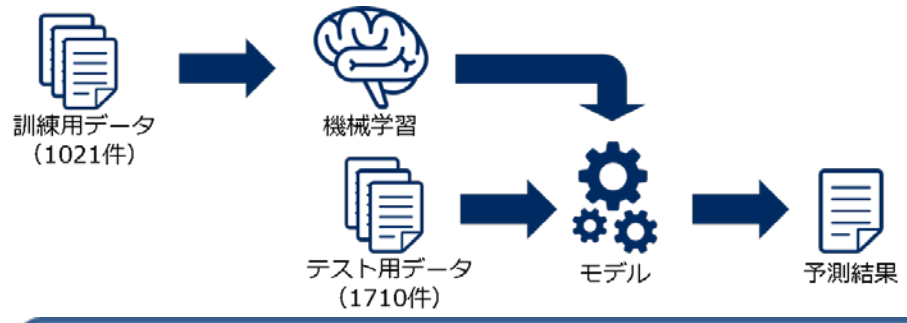
- 指摘の重要度
- 指摘のカテゴリ名
- 問題作込行で出現する関数
- 問題発生行で出現する関数
- etc...

学習アルゴリズム

- サポートベクターマシン
 - ランダムフォレスト
 - ロジスティック回帰
 - 決定木
 - k近傍法
- MALSS → ランダムフォレスト (高精度)

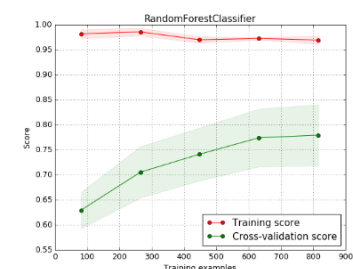
評価

訓練用データによって作成したモデルに対してテスト用データを投入した予測結果を評価



評価結果と今後の展望

- ・65%の精度でバグの可能性が高い指摘を予測
- ・学習曲線から過学習であることを検出。学習用データの追加、不要な特徴量を削除することで改善する



	精度
0(非バグ)	48%
1(バグ)	65%