

エンハンスプロジェクトでの活用を目的とした 障害予測

NRI

畠中 翔

s-hatanaka@nri.co.jp

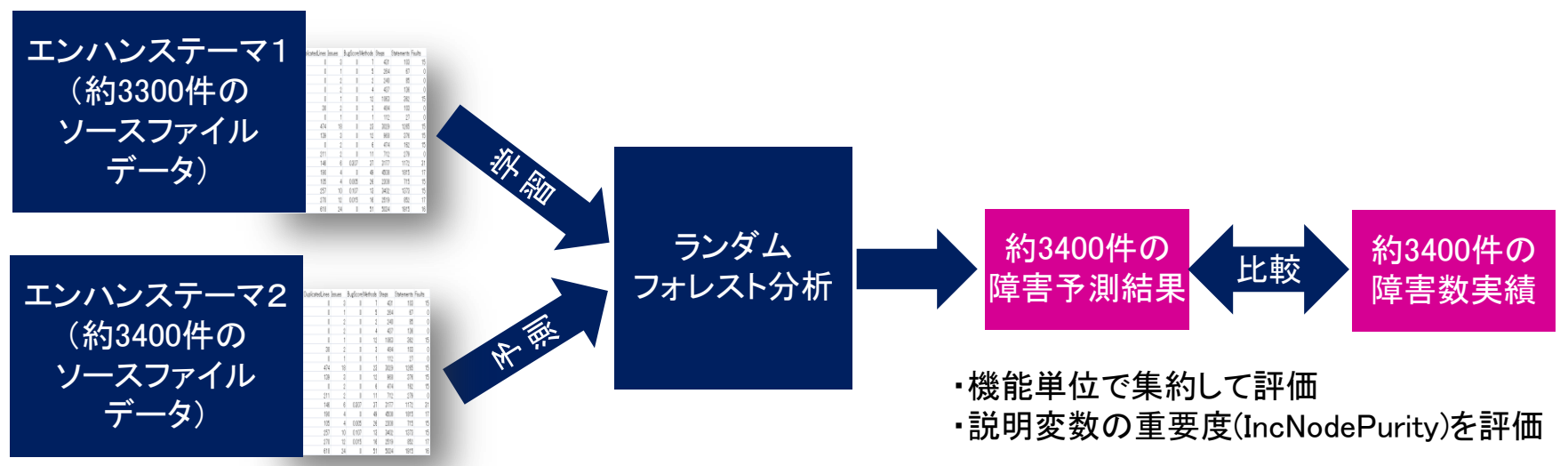
開発における問題点

ソースコード内には障害が潜んでおり,それらを効率的に予測できると,テストの優先順位の組み換え等による「障害の早期収束」といった効果等を得られることが予想される。しかし,効果的な障害予測を目的に,どのようなデータを利用すると効果的かどうか不明である。

本評価による解決

決定木の機械学習により,高精度の予測を行うアルゴリズム「ランダムフォレスト」を,社内のシステム開発プロジェクトのデータに適用し,障害予測に効果的なデータを抽出した。また,エンハンス改修時の障害予測を目的とした場合に,さらなる精度向上を見込めるデータを明確にした。

評価概要



評価時に用いた主データ

分類1	分類2	データ	
目的変数	潜在障害	障害数	
説明変数			重要コーディングルール違反数
			バグ予測ツールスコア値
	保守性	循環複雑度	
		重複行数	
規模		メソッド数	
		ステップ数	
		制御文数	

結果と今後の課題

- 結果
 - 用いたデータのうち,下記データの組み合わせが障害予測に効果的に活用できるとわかった。
 - ーデータについて(有効度順に記載)
 - バグ予測ツールスコア値,コード行数,制御文数,循環複雑度,重複行数,重要コーディングルール違反数
- 今後の課題
 - ー設計工程が起因となる障害予測
 - ⇒設計書に関するデータの活用
 - ーエンハンス改修の予測
 - ⇒改修対象,修正規模に相当するデータの活用