

開発段階における品質向上や品質保証活動への AI 適用に関する研究

アブストラクト

1. 研究の背景

システムの開発における諸作業は、ドキュメントの記述レベル、成果物に対するレビュー、および改修時の影響調査など、開発担当者の力量に左右され、品質にばらつきが生じやすい。

一方、近年進展著しい AI 技術であるが、世の中では「誰にでもすぐに『すごいこと』ができる」「AI は使い物にならない」といった両極端な誤解も依然としてされている状況である。

果たして AI は、品質確保に苦勞するシステム開発の現場を支援できるのか？ 当分科会では品質向上目的での AI 技術の適用可能性および有用性を調査研究し、活用方法を提言する。

2. 研究アプローチ

システム開発での品質向上を目的とした AI 活用は先行事例が少なく、机上では実現性と有用性の評価が難しいため、課題を明らかにした上で実証実験を実施した。準備段階として、(1) テーマの選定、(2) 解決方法の検討と AI 技術の選定、(3) 教師データの検討を行い、次に、実証実験を実施した。

(1) テーマの選定

世の中の企業が抱えている品質問題の多くは「組織」「派生開発(プロセス)」「人」に起因する。当分科会メンバーが自社で抱えている問題についてもヒアリングの結果、同様の傾向がみられた。

組織やプロセスの問題は、ルールベースでの解決が有効と考え、当分科会では AI の持つ「人の作業を代替または支援する」性質に着目し、「人の問題を AI でどの程度解決できるのか？」をテーマに研究を進めることとした。

また、システム開発においては、要件定義や設計といった上流工程、プログラミングを行う中流工程の内容のほか、その段階での品質のばらつきを是正することが重要とされる。このため、上流工程から中流工程、および品質状況の把握に利用できる AI 活用方法を重点に研究することとした。

(2) 解決方法の検討と AI 技術の選定

上流工程で用いる設計ドキュメントは自然言語で記載されているため、当初は人間に代わってドキュメント内の文章を解析する自然言語処理の適用可能性を調査した。しかし、現在の AI 技術では意味や文脈の解釈といった技術が十分に成熟しておらず、ドキュメントの分析やレビューを行うなど、正確な意味解釈を要する作業への適用は費用対効果が見込めず、現実的でないと判断した。

このため、文章内から単語を抽出する「形態素解析」および単語の前後関係から統計的に類義語検索を実現する「Word2Vec」に着目し、影響範囲の特定漏れや検討漏れを防ぐといった「開発担当者を支援する」ことの適用可能性を検討した。この結果、「ドキュメント内の類義語を開発担当者に推薦し、気づきを促す機能(類義語推薦機能)」について有用性があると判断し、実証実験で検証することとした。

また、「開発担当者に気づきを促す」という点については、他にも「開発担当者がどの場合にどのようなドキュメントを参照するか？」のノウハウが人に依存しており、品質への影響も大きいとの観点から、「協調フィルタリング」の仕組みを用いた「複数の開発担当者の評価に基づいて有用なドキュメントを推薦し、気づきを促す機能(ドキュメント推薦機能)」を考案し、その有用性を検証することとした。

さらに、中流工程以前での品質状況の把握段階において「担当者によって品質分析のレベル・観点が異なる」問題に着目した。ここでは、品質分析に用いるメトリクスが数値データであることから、「ロジスティック回帰分析」や「サポートベクターマシン」を用いた品質予測の有用性を検証することとした。

(3) 教師データの検討

AI の実証実験において大きな障壁となるのが、機械学習に必要な「教師データ」の準備である。「類義語推薦」および「ドキュメント推薦」の実験では、本来は設計ドキュメントを用いるべきであるが、複数企業で構成される当分科会の性質上、機密保護の問題や、有用と評価する基準の違いなどの問題から代替案を検討した結果、IT用語を幅広く網羅し、内容の評価が容易な過去8年間のLS研論文(133論文)を設計ドキュメントと見立てて用いることとした。また、ドキュメントの有用性を評価する開発担当者として、当分科会のメンバーを被験者として実験した。

また、品質予測の教師データは、当分科会参加企業のうち実験内容に合致した企業のソースコードから採取した品質メトリクスデータを用いた。

3. 実証実験

実証実験の結果について述べる。

(1) 類義語推薦機能

Word2Vec を用いて LS 研論文を教師データとした「IT用語モデル」を作成し、Wikipedia 日本語版を教師データとした「一般単語モデル」と類義語推薦結果の有用性を比較評価した。実験の結果、モデルによって推薦される単語が異なり、IT用語モデルのほうがIT用語として信頼性のある単語が推薦された。設計ドキュメントにはIT用語や固有の業務用語が多く含まれていることから、これを教師データに用いればシステム開発プロジェクトに有用な類義語を推薦することができるものと判断する。

(2) ドキュメント推薦機能

被験者が事前にドキュメントの有用性を5段階評価でランク付けした上で、ドキュメント推薦機能が推薦するドキュメントと比較し、上位5位が一致している割合(適合率)を評価した。機械学習の手法には、Spark MLlib の ALS(Alternative Least Squares)を採用した。

実験の結果、有用性の評価傾向が似通った被験者間で期待どおりの適合率(95%)で推薦でき、プロジェクト内の知見共有に寄与できることが確認できた。被験者が新たに加わった場合でも、評価傾向を想定した「仮想レーティング」を用意しておくことで期待したドキュメントを推薦できた。

(3) 品質予測

ソースコード単位で「品質が良い(1件以下)/品質が悪い(2件以上)」の予測がどの程度できるかを検証した。開発ステップ数、サイクロマティック複雑度、過去の障害件数などの品質メトリクスデータを教師データとして障害を予測した結果を実際の障害発生データと比較して平均正解率を確認した。

その結果、「サポートベクターマシン」での平均正解率が期待を上回る91.5%という結果を得られた。教師データ数は500件を超えたあたりから平均正解率が安定した。また、様々なメトリクスを入れ替えて予測結果を検証したが、予測精度に影響するメトリクスと影響しないメトリクスが大きく分かれており、特徴量としてどのデータを選択するかが重要であるとの知見を得ることができた。

4. 研究の提言

適用分野は限られるものの、実証実験では期待した結果が得られており、品質向上にAI技術を活用できる可能性は十分あることが確認できた。同時に教師データの内容によっては有用な結果が出ないことも明らかとなり、「AIからのアウトプットの有用性は、教師データの特徴に依存する」ことを改めて認識することとなった。現在のAIは知能そのものを持っておらず、過去のデータを統計的に判断していることから、AI導入の効果は「現時点のプロジェクト成熟度」をそのまま反映したものとなる。

しかし、これは従来のルールベースでの品質向上活動では見えなかった「プロジェクトの現状と課題」をAIが明らかにする効果があることを意味する。現状把握と対策の検討、実施にかかるコストを下げるという意味では、これも品質向上目的でのAIの活用方法であると判断できる。

AIが適切な推薦や予測を行えるよう、プロセス標準化、教育訓練、教師データ整備を進める。そして、AIが推薦や予測した結果を元に、品質向上施策を策定する。こうした「AIを育てる」そして「AIから学ぶ」という改善サイクルを導入してプロジェクト成熟度を高めることこそ、人とAIが協働する真の品質向上活動である。