

リポジトリマイニング手法の習得支援を目的としたシミュレーション環境の構築

松本 明^{1,a)} 柏 祐太郎^{1,b)} 大平 雅雄^{1,c)}

概要: ソフトウェアリポジトリマイニング (MSR) 技術は、ソフトウェア開発の生産性や品質を改善するための有力な手段の1つである。しかしながら、MSR 技術の知識やスキルがない学生にとっては、様々なリポジトリや手法を扱うまでに多くの学習コストが必要となるため、実際のプロジェクトを対象として MSR 技術を適用する事は容易ではない。そこで本研究では、学生の MSR 技術の習得支援を目的としたシミュレーション環境の構築を提案する。

1. はじめに

ソフトウェアリポジトリマイニング技術(以降では MSR 技術とする)とは、ソフトウェア開発に有用な情報や知見をソフトウェアリポジトリから導き出す技術のことで、ソフトウェア開発の生産性や品質を改善するための有力な手段の1つである。

しかしながら、MSR 技術の知識やスキルがない学生にとって、実際のプロジェクトを対象として MSR 技術を適用する事は容易ではない。主な理由として、データの収集やデータクリーニング、マイニングアルゴリズム、分析・評価に関する知識が乏しいために、様々なリポジトリや手法を扱うまでに多くの学習コストが必要となることが考えられる。それに加えて、これから MSR 技術を取得しよう

とする学生にとっては、学習意欲を維持するのも困難である [2].

そこで本研究では、MSR 技術のプロセスを体験できるシミュレーション環境を構築し、学生がシミュレーションを通してリポジトリマイニング手法の習得ができるように支援することを目的とする。また、MSR 技術のステップ毎に概略説明、課題、課題の解説といったプロセスを取り入れることで、先に述べた学習コストを軽減する。さらに、ゲーミフィケーション要素を導入することで、学生の学習意欲の維持をねらう。

2. リポジトリマイニング手法の習得支援環境

2.1 概要

本研究では、学生がシミュレーション環境上で MSR 技術のプロセスを疑似体験できるように、Xie らが文献 [1] で示した MSR 技術の一般的な手順 (図 1) に従って、シミュレーション環境を構築する。

2.2 設計方針

ソフトウェア開発に取り組んだ経験のない学生にとっては、ソフトウェア開発組織での支援ニーズがあらかじめ分かかっておらず、開発ドメインに対する知識も不足していることが多いため、最初に「どのタスクを支援すべきか」を決めることは難しい。そこで本研究では、以下の設計方針に従ってシミュレーション環境を構築する。

2.2.1 シナリオに基づく学習

本シミュレーションでは、プログラミングやテスト・品質保証などを支援対象として、MSR 技術を適用するための

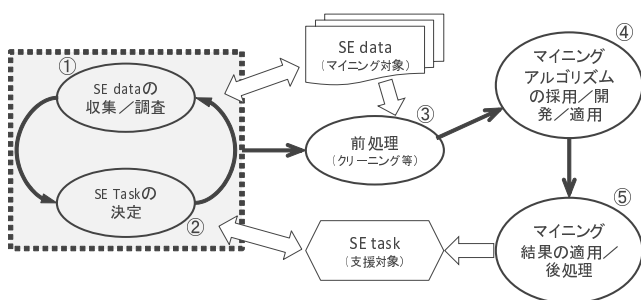


図 1 ソフトウェアリポジトリマイニングの一般的な手順 [1]

Fig. 1 A general procedure of mining software engineering data [1]

¹ 和歌山大学
Wakayama University

a) s151045@sys.wakayama-u.ac.jp

b) s141015@sys.wakayama-u.ac.jp

c) masao@sys.wakayama-u.ac.jp

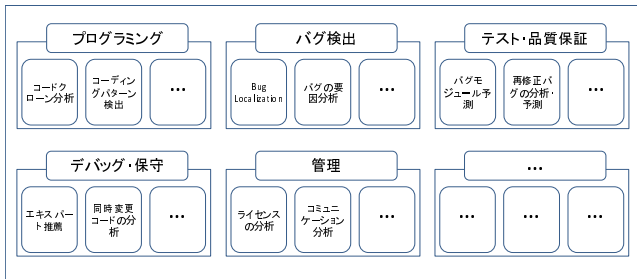


図 2 シナリオ選択画面の例
Fig. 2 An example of scenario selection

一連の流れを学習できるように、複数のシナリオを学習者に提供する。図 2 は、学習者が選択するシナリオのメニュー画面の例である*1。

2.2.2 学習要素

初学者である学生にとっては、MSR 技術の学習にあたって以下のようなことに対する知識やスキルが乏しいために、学習の障壁となる、すなわち、学習すべき要素となると考えられる。

- 収集すべきリポジトリの選択
- 調査すべきメトリクスの選択
- リポジトリの収集方法
- データクリーニングの方法
- 適当なマイニングアルゴリズムの選択
- 分析手段
- 分析結果の評価方法

本シミュレーションでは、MSR 技術のステップ毎に概略説明、課題、課題の解説を順に行うことで、上記の問題点を解決できるようになることを目的とする。

2.2.3 ゲームフィケーション

さらに、学生の学習意欲の維持を図るために、シミュレーション環境にゲームフィケーション要素を取り入れる。ゲームフィケーションとは、問題解決や学習プロセスにゲームデザインの技術やメカニズムを応用することで、近年、教育の分野にも積極的に活用されている。

ゲームフィケーションを活用したシステムの中には、課題、報酬、交流の 3 要素を組み合わせることで、ユーザが自発的にシステムの利用を継続するようにうまく工夫されているものが数多くある。本研究もこれに倣い、学習者のレベルに応じた問題（課題）、課題をクリアした際の結果に応じて与えられるポイント（報酬）、チャットを用いた学生の相互扶助やポイントランキングの表示による競争化（交流）といった要素を導入することで、学生が自発的に学習を継続できるようにする。

2.3 シミュレーションの構成

シミュレーションの構成を以下に示す。

*1 図 2 の例は、[3] を元に作成したもの。

Step1: SE task の決定

複数あるシナリオ（プログラミング、テスト・品質保証、デバッグなど）から支援すべきタスクの一つを選択する。これは、支援対象の決定にあたる。

Step2: SE data の収集/調査

データ収集の方法を理解したのち、用意されているマイニング対象データを収集する。

Step3: SE data の前処理

データクリーニングの方法を理解したのち、収集したデータをクリーニングする。

Step4: マイニングアルゴリズムの採用/開発/適用

各マイニングアルゴリズムの有用性を理解したのち、クリーニングしたデータに適当なアルゴリズムを選択し、適用する。

Step5: マイニング結果の適用/処理

マイニング結果の分析・評価手法を理解したのち、分析・評価を行い、支援対象の作業に対して、どれほどの効率向上がみられたかを確認する。

Step6: シナリオの達成

支援対象の作業に対する効率向上の度合いに応じてポイントが与えられ、Step1 に戻る。より高いポイントの獲得を目指す学生は、もう一度同じシナリオに挑戦することができる。

学生がこの一連のステップを繰り返し行うことで、MSR 技術に対する理解度の向上を図る。また、同じタスクに対して異なる手法を適用した場合の効果の比較を行えるようにすることで、より良い手法の選択方法を理解する。

3. おわりに

本研究では、初めて MSR 技術を学ぶ学生がリポジトリマイニング手法を習得できるように支援することを目的としたシミュレーション環境を提案した。今後は、シミュレーション環境の実装と評価をおこない、学習効果の検証を行う予定である。

謝辞 本研究の一部は、文部科学省科学研究補助金（基盤 (B):23300009）および（基盤 (C):24500041）による助成を受けた。

参考文献

- [1] Xie, T., Thummalapenta, S., Lo, D. and Liu, C.: Data Mining for Software Engineering, *Computer*, Vol. 42, No. 8, pp. 55-62 (2009).
- [2] 大平雅雄, 中野大輔, 松本健一: ソフトウェアリポジトリマイニング教育における探索的データ分析ツールの効用, 情報処理学会報告ソフトウェア工学研究会, Vol. 2013-SE-181, No. 5, pp. 1-8 (2013).
- [3] 門田暁人: ソフトウェアリポジトリマイニングの技術動向とその応用, ソフトウェアジャパン 2013, IT フォーラムセッション (講演資料) (2013).