

ソフトウェア開発現場での活用へ向けた ソフトウェアリポジトリマイニング手法の体系化

大 平 雅 雄^{†1}

ソフトウェアの開発過程で蓄積される膨大なデータから有用な知見を導き出すための手法として、ソフトウェアリポジトリマイニングが近年注目されている。しかしながら、新たな手法が多数提案されるものの、ソフトウェア開発の現場に適用した事例はまだ多くない。本ワークショップでは、開発現場への適用における課題について問題意識を共有するとともに、開発現場での適用を促すための手法の体系化について議論したい。

Constructing a Systematic Classification of Software Repository Mining Techniques toward Practical Use in Software Development

MASAO OHIRA ^{†1}

Mining software repository (MSR) has been attracting many researchers and practitioners as a promising approach to extract useful insights from a wealth of information sources automatically accumulated in software development. While there already exist many techniques in the field, they are not fully used for software development in practice. In this workshop, we would like to discuss the challenges in applying MSR techniques to software development. We also would like to try to construct a systematic classification of MSR techniques to helping practitioners exploit the techniques in their development environment.

1. はじめに

ソフトウェアリポジトリマイニングは、版管理システムや不具合管理システムなどのソフトウェアリポジトリに記録された膨大な開発履歴に対するデータマイニング手法を総称したものの^{*1}である。ソフトウェア開発の過程で自動的に記録されるデータを対象とするためデータ収集コストが低く、かつ、複数のリポジトリデータの組み合わせから新規性の高い有用な知見の導出を期待できるため、専門の国際会議 MSR (Mining Software Repository) が毎年 ICSE (International Conference on Software Engineering) にて併設開催されるなど、近年ソフトウェア工学の1分野として高い注目を集めている。

多数の MSR 手法が提案され、また、多くの研究者がその可能性について言及する¹⁾⁻³⁾。一方で、MSR 手法をソフトウェア開発の現場に適用した事例はまだそれほど多くないのが実情である。その理由には以下のような課題が挙げられる⁴⁾。

- (1) データ抽出の困難さ：自動的に蓄積される開発履歴データを対象とするため、データ収集のコストはほとんどかからない。しかしながら、データに含まれるノイズの除去、あるいは、異なるシステム（例えば、CVS と Subversion）からデータを抽出する際のデータ整合性の確保など、MSR 手法を適用するためには様々な前処理を施す必要がある。現時点では実務者が容易に MSR 手法を適用するための環境やシステムが十分に整備されているとは言い難い状況にある。
- (2) 手法の制約：MSR 手法の多くは、単体のリポジトリを対象とするものであり、手法により導出される知見には多くの制約条件がある。例えば、ソースコードの変更時期とその理由を正確に知るためには、版管理システムに蓄積された変更時期や変更箇所だけではなく、不具合管理システムやメーリングリストなどで議論された不具合修正理由などを総合的に解析する必要があるが、現時点では複数の種類のリポジトリデータを組み合わせて利用する手法は少ない。結果的に、開発現場のニーズを十分に満たす手法がま

^{†1} 奈良先端科学技術大学院大学
Nara Institute of Science and Technology

*1 以降では、MSR 手法と呼ぶ

だ少ないものと思われる。

- (3) Approach-driven な研究分野：MSR 分野で提案される手法の多くは、特定の種類のデータマイニング手法を特定の種類のリポジトリ（かつ、ほとんどはオープンソースプロジェクトから収集したもの）に適用することでその有用性が検証されてきた。したがって、実際の開発現場に手法を適用するためには、多数の手法の中から開発現場のニーズを満たす妥当なものを選択し、かつ、その手法が開発現場で利用しているリポジトリに適用可能かどうかを検証する必要があり、実務者の MSR 手法の利用を阻害する一因となっている。

2. タスク・目的指向な MSR 手法の体系化

先に挙げた課題のうち、(1)、(2) については様々な取り組みがある。MASU (Metrics Assessment plugin platform for Software Unit) は、プログラミング言語の違いを意識せずに各種メトリクス計測を可能にするためのソースコード解析基盤である。SO-MSR (Service Oriented Framework for Mining Software Repository)⁵⁾ は、リポジトリ間のデータ形式の違いを吸収するとともにマイニング手法の手順やノウハウを研究者間で共有するためのフレームワークである。また、EPM⁶⁾ は、CVS, GNATS, Mailman の 3 つのシステムを組み込み、種類の異なるリポジトリデータを横断的に利用して解析結果を提供するシステムである。これらの枠組みやシステムは、ソフトウェア開発現場での MSR 手法の適用を容易にするためのツールとして有用と考えられるが、(3) の課題から十分に活用されていない状況にある。特に、既存の MSR 手法が (1)、(2) の問題を十分にクリアしないまま提案されてきたため（どの手法が現場のニーズを確実に満たすのかが不透明なため）、どの手法を枠組みやツールに実装すべきかの判断を難しくしているためである。

我々は現在、ソフトウェア開発現場でのタスクあるいは目的に沿って MSR 手法を分類し体系化することを試みている。前述に挙げた枠組みやツールを開発する研究者にとって有益だけでなく、実務者が「どの工程や作業において、どのリポジトリに対するどの MSR 手法の適用が、どのような知見をもたらすのか」を一望でき、MSR 手法の適用を促すことにつながると期待している。本ワークショップでは、特に開発現場での MSR 手法の適用を重視して、MSR 手法を体系化するために必要な観点や分類軸について議論したい。

謝辞 本研究の一部は、文部科学省科学研究補助費（若手 B：課題番号 22700033）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) Godfrey, M.W., Hassan, A.E., Herbsleb, J., Murphy, G.C., Robillard, M., Devanbu, P., Mockus, A., Perry, D.E. and Notkin, D.: Future of Mining Software Archives: A Roundtable, *IEEE Software*, Vol.26, No.1, pp.67–70 (2009).
- 3) Hassan, A.E. and Xie, T.: Software intelligence: the future of mining software engineering data, *Proceedings of the FSE/SDP workshop on Future of software engineering research*, FoSER '10, New York, NY, USA, ACM, pp.161–166 (2010).
- 4) Hassan, A.E.: The road ahead for Mining Software Repositories, *Frontiers of Software Maintenance (FoSM) at the 24th IEEE International Conference on Software Maintenance (ICSM2008)*, pp.48–57 (2008).
- 5) Matsumoto, S. and Nakamura, M.: Service Oriented Framework for Mining Software Repository, *Joint Conference of the International Workshop on Software Measurement and the International Conference on Software Process and Product Measurement (IWSM/MENSURA 2011)*, pp.13–19 (2011).
- 6) 大平雅雄, 横森励士, 阪井 誠, 岩村 聡, 小野英治, 新海 平, 横川智教: ソフトウェア開発プロジェクトのリアルタイム管理を目的とした支援システム, *電子情報通信学会論文誌 D-I*, Vol.J88-D-I, No.2, pp.228–239 (2005).