

20 大規模ソフトウェアの再利用を実現するリポジトリ・マイニング技術

吉村健太郎 (株)日立製作所

計画的再利用 vs 経験的再利用

近年、組込みシステム、社会インフラシステム等の応用分野への展開を背景として、ソフトウェアの大規模化が加速している。同時に、セーフティクリティカル分野では高い信頼性が要求されるため、実績あるソフトウェア資産を用いて新規製品を開発する、再利用による大規模ソフトウェア開発の重要性が高まっている。

ソフトウェア再利用技術の歴史は古く、1968年にドイツ Garmisch で開催された世界初のソフトウェア工学会議において「ソフトウェア部品」の概念が提案されている。ソフトウェア再利用技術の多くは、計画に基づく再利用である。すなわち、将来の技術動向、市場動向に基づいて製品ロードマップを計画し、複数製品間で再利用可能なソフトウェア部品を事前に設計することによって、ソフトウェア開発を効率化する。近年、製品系列内での計画的ソフトウェア再利用技術は「ソフトウェア・プロダクトライン(SPL)」として体系化され、産業界で活用されている¹⁾。

しかしながら現実の開発現場では、計画と製品開発とが完全に一致することは稀である。たとえば、設計時点で予測していなかった機能が個別製品の開発中に場当たりに追加され、その後、複数製品で経験的に再利用されることも多い。そのため、計画的なソフトウェア再利用に加えて、製品開発実績における経験的知識を再利用資産として活用する技術が重要となる。

また、ソフトウェアの大規模化に伴って、ソフトウェア部品間の関連が複雑化している。ソフトウェア部品間の関連として、依存関係や排他関係などがある。ソフトウェア大規模化によって部品間の関連の複雑度は指数関数的に増加するため、設計段階での分析が困難となる。さらに、市場でのユーザ嗜好のように、論理的な設計が困難な関連も生じる。

上記の課題を解決して大規模ソフトウェアの再利用開発を支援するために、実際の製品開発時に追加、発見される経験的知識を抽出し、将来の製品開発に活用する技術が求められている。

リポジトリ・マイニングによる経験的再利用

そこで近年、製品開発におけるコア資産運用状況を分析し、その結果をコア資産へとフィードバックする技術「ソフトウェア・リポジトリ・マイニング」が注目されている^{2), 3)}。特に、製品リリース履歴やリポジトリ変更履歴に基づいて機能間の関連を抽出する技術は、今後のソフトウェア大規模化時代における再利用支援技術として、非常に有用性が期待される。

図-1に、リポジトリ・マイニングによる相関ルール分析例を示す。この例では、ソフトウェアリポジトリ中の製品リリース履歴における製品一機能(ソフトウェア部品)の対応関係に基づいて、機能間の関連(依存関係、排他関係)を製品リリース履歴から抽出する。

機能間関連の抽出には、統計的分析技術を活用する。図-1の例では相関ルール分析を用いている。機能間の関連を「ルール」として仮定し、そのルールが製品リリース履歴中で成立しているか否かを定量的に評価する。たとえば、「機能Aを採用した製品は機能Dも採用している」という相関ルールは、高い信頼度と支持度(ルールの出現率)を示している。このように、製品リリース履歴と照らし合わせて高い評価が得られた相関ルールを抽出し、経験的設計知識として利用する。

抽出した機能間の関連はソフトウェアリポジトリにフィードバックし、将来の再利用時の指針として活用する。活用例として、今後の製品開発時には、機能Aを選択したときの機能D選択漏れを警告することが可能になる。

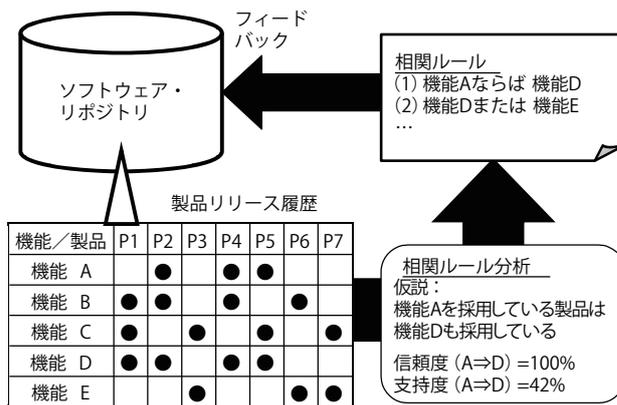


図-1 リポジトリ・マイニングによる機能間依存関係分析

実用化に向けた研究課題

大規模ソフトウェア再利用のためのリポジトリ・マイニングは、今後の実用化・発展が期待される技術であると同時に、多くの課題が残されている研究対象でもある。

第1の研究課題は、抽出対象とする経験的知識の定義である。製品開発時に追加・発見される知識は、機能間の関連のみならず、ソフトウェア構造、開発プロセス、開発者間コミュニケーション等、多くの可能性がある。また、経験的知識を抽出するためのリポジトリ・データの選択と生成、そして統計的分析を行うための前処理手法の研究もまた重要である。

第2の課題は、統計的分析手法そのものの研究である。すなわち、統計学的側面から見た開発履歴データの性質の評価手法と、それに適した分析技術の選定手法が必要である。また、大規模ソフトウェアの分析結果を効率よく実務者に提示するためには、直感的かつスケーラブルな可視化手法が重要な研究テーマとなるであろう。

第3の課題は、分析結果のフィードバック手法である。分析結果をどのように製品開発に活用していくか、特に開発環境へのマイニング技術統合が実用化に向けたキー技術となる。

本会は、情報処理技術を専門とする研究者のコミュニティである。リポジトリ・マイニング技術を発展させるため、本会における研究者間の議論および技術体系化に期待する。

参考文献

- 1) McGregor, J. D., Muthig, D., Yoshimura, K. and Jensen, P. : Successful Practices in Software Product Lines, IEEE Software, Vol.27, No.3 (May/June 2010, to appear).
- 2) Nagappan, N., Zeller, A. and Zimmermann, T. : Mining Software Archives, IEEE Software, Vol.26, No.1, pp.24-25 (Jan./Feb. 2009).
- 3) 吉村健太郎, 成沢文雄, 菊野 亨 : 製品リリース履歴における論理的結合集合に基づいた横断フィーチャ分析法, 情報処理学会論文誌, Vol.50, No.11, pp.2654-2664 (Nov. 2009).

(平成 21 年 10 月 29 日 受付)

吉村健太郎(正会員) ● kentaro.yoshimura.jr@hitachi.com

2001年早稲田大学大学院修士課程修了。同年(株)日立製作所日立研究所入社。現在、同研究所研究員。2009年大阪大学大学院博士後期課程修了。博士(情報科学)。主に組み込みソフトウェア開発手法に関する研究に従事。IEEE、日本機械学会各会員。