

大規模開発データの収集・活用とエンピリカルソフトウェア工学

大阪大学 大学院

情報科学研究科コンピュータサイエンス専攻 教授

井上 克郎

Katsuro Inoue



ソフトウェア開発プロジェクトを成功に導くためには、いろいろな技術(手法やツール、ノウハウなども含む)が利用される。しかし、たとえばUML(Unified Modeling Language:ソフトウェア分析/設計のためのモデリング言語)を用いた設計法は、構造化設計法よりも優れている、というふうなことを主張できるであろうか?

ソフトウェアの生産に関するいろいろな技術に関して、きちんとデータを揃えて科学的に評価することは今までほとんど行われなかった。これは、対象とする技術にはいろいろなバリエーションがある、技術の周りにある環境によって評価が大幅に変わりうる、評価コストが高いなどの理由による。

しかし、科学や工学の他の分野は、何か新しい提案が行われる場合、必ずそれを評価する方法が付随しており、それによって、新しい提案の良し悪しが議論される。厳密な評価を行なうことによって技術を着実に進歩させることができる。

ソフトウェア開発に関する技術に関しても、厳密な評価を行なおうという機運が盛り上がっている。環境を整えて技術を適用することによって、評価に耐えうるような実験データを収集し、技術の良し悪しを議論する。このようなアプローチをエンピリカルソフトウェア工学(empirical software engineering)と呼んでいる。

近年、エンピリカルソフトウェア工学に関する論文誌(Empirical Software Engineering)が発行され、国際会議(International Symposium on Empirical Software Engineering)が開催されるようになってきた。また、ソフトウェア工学に関する論文誌でも、新しい技術の提案に際しては、実践的なデータを用いた評価が行われているかが問われるようになってきている。

米国のSEI(Software Engineering Institute)やドイツのフランクフォラーIESE(Fraunhofer Institute for Experimental Software Engineering)などの研究機関でもエンピリカルソフトウェア工学の研究が盛んに行われている。わが国でも文部科学省の支援の下に、エンピリカルソフトウェア工学ラボが2003年4月、大阪の千里中央に設立された。このラボでは、日立GPをはじめとした企業からの出向者、専属研究者、大学の教員や学生が集まり、エンピリカルソフトウェア工学に関する研究のみならず、システムの開発、普及を行っている。

ソフトウェア開発に関する種々のデータを収集し、評価することは、すでにいろいろな現場で行われている。たとえば出来上がったモジュールの数からプロジェクト全体の進捗を判断したり、バグレポートから品質を推定したりすることは頻繁に行われている。本ラボでは、ひとつのプロジェクトのデータのみならず、多数のプロジェクトのデータを集めて分析し、次のプロジェクトに役立つ情報を抽出することをターゲットにしている。

近年、ディスク容量の巨大化や高速ネットワークの普及によって、ソフトウェア開発に関わるあらゆるデータを記録し、それらを集積して長期保存することは容易になってきている。またCPUパワーも増大しているので、保存した多数のプロジェクトの巨大なデータを目的に応じて詳しく解析したり、検索したりすることも徐々に可能になってきている。

今後、ソフトウェア開発を行う会社では、組織的に過去のプロジェクト全てのデータを残し、必要に応じて取り出し、加工して、他のプロジェクトに活用することは、会社全体の生産性、信頼性を向上させるために必須になろう。その効率によって会社全体の利益が大きく左右されるようになるのではないか。