

ソフトウェア開発プロセス設計・適応支援システム

2W-06

— 開発に必要な知識項目のモデル化 —

松下壮一 菅野亨太 野中誠 東基衛

早稲田大学大学院理工学研究科 経営システム工学専門分野

1 はじめに

開発者の能力といった人的要因は、開発工数の見積りに大きな影響を与える。COCOMO II [1]では、プログラマの能力が開発工数に与える影響が最大で1.85倍の差があるとしている。しかしCOCOMO IIでの人的要因の評価は、能力が「極めて高い」から「非常に低い」までの6段階による主観評価である。プログラマの能力による開発工数の変動範囲非常に大きいため、能力評価はより詳細かつ正確に行われる必要がある。またプロダクトの特性も考慮に入れるべきである。

筆者らはソフトウェア開発プロセス設計・適応支援システム研究の一環として、プロダクトの開発に必要なとされる知識と開発者の知識、さらにプロダクトの特性に基づいた調査活動時間の見積もり技法の開発を行っている。

本発表では既存のプログラムから開発に必要な知識項目を抽出し、調査時間を予測する手法について報告する。

2 提案システムのご概念

2.1 本報告の位置付け

筆者らは、ソフトウェアの機能要求とその他の品質要求、および資源の能力に適応したプロセス設計を支援するシステムの研究を行っている。ここで資源の能力には開発者の能力が含まれる。プロセス設計の評価を行う際には開発工数とのトレードオフを考慮する必要があり、そのため開発者の能力が正確に計測され評価される必要がある。

本報告では、個人プロセスを対象として、規模と開発対象の機能的な分類を入力とし調査時間を出力とする手法について述べる。

2.2 調査時間の定義

支援システムでは開発時間と調査時間を区別して扱う。すなわち、総開発時間を T 、開発時間を T_d 、調査時間を T_r としたとき、

$$T = T_d + T_r$$

となる。ここで調査とは、個人の資質や能力とは関係のない、知識不足を補うための知識獲得活動とする。

3 調査対象

調査すべき対象は幾つかに分けられる。東らは情報技術者に必要な知識として以下の5つを挙げている[2]。

- ① ソフトウェア工学の基礎知識
- ② 技法・方法論の知識
- ③ 開発ツール及び環境の知識
- ④ アプリケーション対象分野の知識
- ⑤ システム資源の知識

本研究では詳細設計以降の調査を扱うので、以上に挙げた②、③、⑤の知識を調査対象とする。

4 知識の表現

開発に必要な調査の時間を見積もるには、以下の項目を定量化する必要がある。

- I 開発に必要な知識
- II 開発者の知識状態

Iは、これから開発する対象に含まれることが予測される知識項目である。これは予備調査により、統計的に分類ごとの頻出API (Application Programming Interface) クラスを算出し見積りに利用する。IIは開発者の過去のプロセスデータから知識状態を予測する。これらの差異から必要と思われる調査時間を推測する。

4.1 APIクラスの利用

本研究では、知識をAPIクラスという単位でモデル化する。APIは、あるプラットフォームのソフトウェアを開発する際に使用できる命令や関数の集合であり、その機能により整然と分類されている。特に今回の対象言語であるJavaはAPIを多く用いてプログラミングされる。

そこで、ある API クラスを使用した経験があれば、その使用方法のみならず、その API を用いたアルゴリズムや関連項目の知識も同時に獲得されたものと考え、API を用いて調査時間に影響を及ぼす知識の表現を試みる。すなわち API の知識の有無で3章に述べた3つの知識領域を表現する。

4.2 開発者の知識の表現

開発者の知識状態は、API クラスを知っているかどうかの2値で表す。その際同じパッケージ内やAPI クラス間の類似度または難易度などを考慮し後のコスト計算に用いる。推定に用いる値は、開発者の過去のプロダクトのプロセス記録から抽出する。

5 プロダクトの分類

開発の対象となるプロダクトに含まれる知識を表現するためには、まずプロダクトを分類する必要がある。開発規模の見積もりを計算する手法である Function Point 法では、14 項目のプロダクト特性により見積もり値を調整しているが、項目数が多いため時間が掛かってしまう。そこで本研究では分類をなるべく少なくし、分類ごとの API クラスの含有量を調べることにより対象プロダクトの開発に必要なとされる知識を表現する。

5.1 調査例

無作為に集めた 10 個の Java プログラムを対象に、プログラムに含まれる API クラスの量を調べた(図1)。

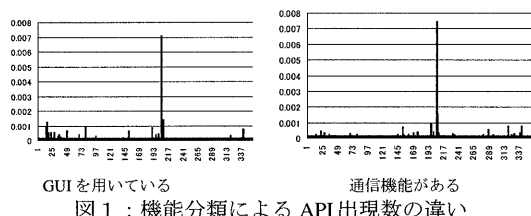


図1：機能分類による API 出現数の違い

図1の横軸は API クラス名、縦軸は API の出現回数を規模で割ったものである。このうち String や Vector といった、どのような特性のプロダクトにも出現する API クラスの出現回数は両者とも差はないが、出現率の少ない API クラスは異なっている。クラスの機能を調べると、各分類において出現回数の一定のものと異なるものでは以下のような特徴がある。

- A どのプロダクトにも同じような頻度で現れるクラス
- B 特有の機能を表すクラス

そこで、各分類において出現回数が多く、ばらつきの少ないものが高い値を示す「出現スコア」を計算する。

5.2 出現スコア

コア API のクラス $i(i=1,2,\dots,n)$ の既存プロダクト $A_k(k=1,2,\dots,m)$ 内での出現回数を A_{ki} とし、 A_{ki} のプロダクト間の平均を AV_i とする。また、

$$A_{ki} / AV_i = H_{ki}$$

である $H_{li} \sim H_{mi}$ の標準偏差を SDV_i 、 A_{ki} を規模 S で割った値 s_{ki} の平均を AVs_i とする。このとき出現スコア Y_i を、

$$Y_i = AVs_i / SDV_i$$

と定義する。

各分類における出現スコアの順位(表1)を見ると、分類の特徴をよく表していることがわかる。

表1：出現スコアの順位

順位	GUIあり	通信機能あり
1	String	String
2	System	System
3	BorderLayout	Vector
4	Dimension	Integer
5	Vector	Enumeration
6	ActionEvent	StringTokenizer
7	Button	Thread
8	WindowEvent	File
9	Thread	OutputStream
10	File	Socket

以上より、API の出現の仕方によりプロダクトを分類できることがわかった。この情報はプロダクト自身に含まれているため、分類をする際に人間の主観的な評価に依らず、プロダクト自信の特性から客観的に評価をすることができる。プロダクト特性の分類をさらに洗練し出現 API を予測することで開発に必要な知識が推測可能になると思われる。これにより、調査時間の推定が可能となる。

6 おわりに

ソフトウェア開発プロセス設計・適応支援システムにおける開発工数見積りのために、プロダクトの規模と機能的分類から開発に必要な知識を、API クラスを用いて表現する手法を提案した。今後は API クラス毎の難易度や調査時間を実験により定義していく。また、プロダクト特性の分類が急務である。

参考文献

- [1]COCOMO II Model Definition Manual Version 1.4, University of Southern California (1997).
- [2]東基衛ほか：“情報技術者自己診断システム—能力モデルと診断の枠組み”，情報処理学会第56回全国大会，1998.