

2S-6

類似度に基づく
ソフトウェア部品検索

相川正美†

上野勝彦‡

日本電気(株) † C & Cシステム研究所 ‡ 放送映像事業部

1 はじめに

ソフトウェアの生産性と信頼性を向上させるために、既存のソフトウェア資産、部品の再利用が注目されている。再利用を促進することにより、重複開発やテスト工数の削減が可能となるため、開発コストの低減、開発期間の短縮、品質の確保が期待できるが、予想したほど大きな効果をあげていないのが実状である。

本稿では、部品の再利用を支援するために、過去に開発したソフトウェア部品の中から、要求に合致する、あるいは、類似している部品を検索する方式について報告する。

2 類似ソフトウェア部品検索

2.1 従来の部品検索における問題点

一般的に考えられているソフトウェア部品の検索には、部品名称、処理機能名等を検索キーとしたキーワード検索、分野や使用頻度に基づいたブラウザ検索がある[1]。

しかし、部品検索に柔軟性を持たせて、部品の再利用の機会を増やすためには、部品検索機能として次の要件を満足する必要がある。

- 1) ユーザが部品名称、処理機能等の部品に関する情報を知らなくても検索可能であること。
- 2) 要求に100%合致した部品が存在しなくても要求に類似した部品が検索可能であること。
- 3) 類似した部品が複数存在する場合には、どれが要求に最も類似しているかを判断できる情報の提示がなされること。

2.2 類似ソフトウェア部品検索方式

2.1で述べた要件を満たすためには、要求機能と個々のソフトウェア部品の機能とを比較して、両者の類似性を判別し、この結果を定量的に表現できればよいと考えられる。

一般的に、2つ以上の対象物が似ているか似ていないかという判断は、多分に主観的要素を含み、曖昧なもので問題を定量的に扱うために類似度という量を定義することが多い[2]。

ここでは、要求機能とソフトウェア部品の持つ機能がどれだけ似ているかを表わす尺度を要求に対するソフトウェア部品の類似度と定義し、0[%]から100[%]で表現する。この類似度を導入したソフトウェアの部品検索方式を次に述べる。

ソフトウェア部品の検索において、類似の対象として扱いたいのは機能であり、量的に扱うことは難しい。そこで機能を構成するプリミティブな機能を要素機能(形式:要素機能名+特性)と定義し、この部分を個々のソフトウェア部品の該当する要素機能と照合する。各特性には、それぞれ重みを与えておき、特性の照合結果に基づいて重みを加算する。このようにして各要素機能の類似性を定量化する。尚、この照合から重みの加算までの処理を照合アルゴリズムと呼び、この結果得られる要素機能の類似性を Q_n とする。

次に、各要素機能がソフトウェア部品の中で占める割合を考慮にいれ、部品内における各要素機能単位の類似度を下に示す式①より求める。尚、ソフトウェア部品の中で占める各要素機能の割合は、要素機能に対応するソフトウェア部品のソースライン数を使用する。

最後に、これら各要素機能単位の類似度を合計して、要求に対するソフトウェア部品の類似度を下に示す式②より求める。

1) 要素機能単位の類似度 A_n

$$A_n = (Q_n / 100) * (L_n / L) \dots\dots\dots \text{①}$$

Q_n :照合アルゴリズムによって求めた要素機能の類似性

L_n :要素機能に対応するソースライン数

L :該当部品のソースライン数

2) ソフトウェア部品単位の類似度 S

$$S = (M / C) * \sum A_n * 100 \dots\dots\dots \text{②}$$

M :マッチした要素機能数

C :要求側の要素機能数

尚、要求と部品は、それぞれ定型の要求仕様、部品仕様として表現している。定型の要求仕様は、仕様原簿という形式で定義しておき、これをもとに要求内容の設定を行なう。これらの関係を図1に示す。

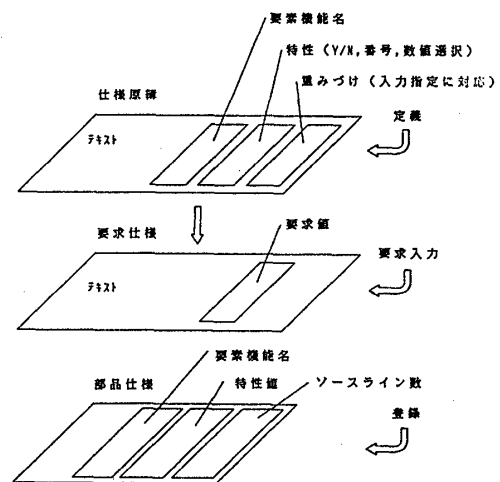


図1 仕様関連図

A method for retrieving software parts based on the degree of similarity
Masami Aikawa, Katsuhiko Ueno
NEC Corporation

3 実験システム

2で述べた検索方式に基づく類似ソフトウェア部品検索の実験システムをUTILISPを使ってEWS4800上に作成した。この実験では、放送局の自動番組制御システムのソフトウェア部品をサンプルとした。各要素機能の重みは、このシステムの開発経験者の意見を参考とした。ソースライン数の割り出しは、手作業で行なった。

実験で用いた定型の要求仕様、及び定型の部品仕様の一部を、それぞれ図2、図3に示す。

f421		特性	
		Y/N選択	番号選択 数値
1	TSコマンド使用	Y	
2	コマンドの指定	Y	
2.1	TS表示確認後、*の部分に時間、t(+、-)をとり、	Y	
1)	TS表示	Y	5.10
2)	コマンド開始時刻	Y	5.13
3)	コマンド終了時刻	Y	5.20
4)	コマンド時間	Y	5.27
5)	コマンド表示	Y	5.37
6)	コマンド終了表示	Y	5.37
7)	X指定	Y	5.40
8)	コマンド指定	Y	5.40
9)	A指定	Y	
2.2	X指定は?	Y	①
①	X指定.....指定曜日の00A1'-fのみのコマンド		
	f'コマンド指定.....0A1'-f、B1'a、D1'a、S1'aコマンド		
②	A指定.....指定曜日のLOCAL1'-f (0A1'-f) A1'a、B1'a、C1'aコマンド		
	X指定.....選択f'fのみのコマンド		
③	X指定.....指定曜日のLOCAL1'-f、A1'aコマンド		
	f'コマンド指定.....選択f'fのみのコマンド		
④	その他		

図2 定型の要求仕様

部品1				部品2			
要素機能名	Y/N選択	番号選択	数値	要素機能名	Y/N選択	番号選択	数値
実行コマンド	Y			実行コマンド	Y		
コマンドの指定	Y			コマンドの指定	Y		
TS表示	Y		5.17	TS表示	Y		5.10
コマンド開始時刻表示	Y		5.22	コマンド開始時刻表示	Y		5.13
コマンド終了時刻表示	Y		5.35	コマンド終了時刻表示	Y		5.20
コマンド時間表示	Y		5.47	コマンド時間表示	Y		5.27
コマンド表示	Y		5.57	コマンド表示	Y		5.37
コマンド終了表示	Y		5.57	コマンド終了表示	Y		5.37
X指定表示	Y		5.62	X指定表示	Y		5.40
f'コマンド指定表示	Y		5.62	f'コマンド指定表示	Y		5.40
A指定表示	Y			A指定表示	Y		5.40
X指定機能	Y		①	X指定機能	Y		②

図3 定型の部品仕様

4 実験結果

定型の要求仕様を設定した要求値から、2で述べた類似度の算出式に基づき計算した結果は、以下の通りであり、要求に対する類似性をソフトウェア部品ごとに定量的に表現することができた。

尚、比較のためにシステムの開発経験者の直感による類似度、Qn (要素機能の類似性、重み有り)、Qn' (要素機能の類似性、重み無し) もあげておく。

	類似度	直感	Qn	Qn'
部品1	54%	50-60%	71%	30%
部品2	71%	80%	84%	78%

図4 類似度との比較

5 考察

Qn' では、要求する要素機能が存在するか否かで類似性を決定してしまうため、どのくらい似ているかという違いが数字に反映されない。また、Qnは、各要素機能の持つ特性に重みを考慮したもので、直感には近いが、各要素機能の部品に占める割合が考慮されていないため、局所的に類似性が高ければ、全体の類似性が上がってしまうことになる。

今回のデータだけで結論づけることはできないが、類似ソフトウェアの部品検索においては、単に要求と部品の機能項目の共通部分の大きさだけで類似性を決めるときではないということと言える。以下に課題を列挙する。

1) 重みづけ

要素機能の特性ごとに指定する重みは、類似度に影響を及ぼすので、最初は、適当な重みを指定しておき実際の検索を通して重みを調整していくことが必要となる。

2) 機能の照合

要求仕様と部品仕様の要素機能は、同一の機能空間に定義してあるため、機能の照合は完全マッチで行なった。より汎用的な検索を考えると部分マッチ、シソーラス等の利用が必要となる。

3) ソースライン数の割り出し

要素機能に対応するソースライン数の割り出しを手作業で行なうのは、きわめて効率が悪い。したがってコーディングの際に、ソースに要素機能に該当する部分をマークしておき、これを自動的に検出してソースラインを割り出す方法等を考慮する必要がある。

6 おわりに

本稿では、ソフトウェア開発における部品の再利用という観点から、類似ソフトウェア部品の一検索方式を提案し、その有効性を確認するために実験を行い、その課題について述べた。ここで提案した検索方式の機能強化については今後の課題である。

謝辞

本研究を進めるにあたり、数々の助言をくださった日本電気(株)永井義裕部長、藤田友之課長、古関義幸主任に感謝致します。

[参考文献]

- [1] 宗近他:「部品指向の設計支援環境 50SM」
情報処理学会研究報告 Vol.89, No.11 64-14
- [2] 「数理科学」ダイヤモンド社 No.120, June 1973