

B-010

変更・閲覧履歴を利用した成果物間の依存関係予測手法に関する提案 Elicit Dependency among Existing Products Using Change and Reference Histories

川崎 将人 †
Masato Kawasaki †

高橋 洋一 †
Yoichi Takahashi †

1. 概要

近年、ソフトウェアが大規模化・複雑化する中、品質を向上するために、仕様書やソースコード等の成果物間の依存関係を把握する方法が考えられている。しかし実際には、依存関係を把握するための作業コストの大きさから、依存関係を把握せず、結果的に品質の低下を招いていることが多い。そこで、成果物間の依存関係を自動的に抽出する手法が求められている。

成果物間の依存関係を抽出する手法として、複数の成果物をほぼ同時に変更したという履歴を利用し、成果物間の依存関係を予測するものが提案されている([1])。しかし、異なる工程で作成される成果物は通常同時には変更されないため、依存関係を自動的に抽出するのは難しい。

そこで本研究では、成果物の変更履歴だけでなく、閲覧履歴も利用して、異なる工程で作成される成果物の間においても依存関係を予測できる手法を提案する。

2. 背景

近年、ソフトウェアは大規模かつ複雑になる一方、品質の向上や納期の短縮がますます求められている。このような中でも品質を向上するためには、要求、仕様書、ソースコード、テスト項目などの成果物間の依存関係を把握することが重要となる。なぜならば、依存関係を把握しなければ、ある成果物の変更に伴って変更が必要となる可能性のある他の成果物を漏れなく一覧化するのが難しく、成果物品質の低下を招くからである。例えば、要求を満たすための仕様書やソースコードが作成されているかどうかを確認できないために、実装漏れを生じる可能性がある。さらに、変更が必要となる可能性がある成果物を一覧化できなければ、変更のための工数を見積もるのが難しく、納期を守れない変更要求を受け入れた結果、品質が低下するということが考えられる。

しかし多くの場合、成果物間の依存関係は把握されていない。なぜならば、ドキュメントと他の成果物間の依存関係を自動的に抽出するのは難しく、手作業で抽出し記録することになり、その作業コストが大きいためである。また依存関係は一度抽出すればよいというものではなく、開発が進んで成果物に変更されるのに併せて、依存関係をメンテナンスする必要がある。工程遅れが顕著な場合、記録した依存関係が放置されるという傾向が強く、依存関係を活用できていない。このような経緯から、自動的に成果物間の依存関係を抽出し、なおかつ成果物の変更に伴って成果物間の依存関係も変更するための手法が求められる。

3. 従来研究

成果物間の依存関係を自動的に抽出する手法として、成果物の変更履歴を利用する手法が提案されている。

[1]では、成果物をほぼ同時に変更したという履歴を利用して成果物間の依存関係を予測する。さらに予測精度を向上させるため、ほぼ同時に変更したという履歴に加えて、成果物中の出現語句やソースコード中の構文など、種々の情報を利用するものがある([2][3])。

ここで、異なる工程の成果物について考える。異なる工程の成果物は、それぞれ作成される時期が異なるため、それぞれの変更履歴は同時とはならない。また、後工程で見つかった障害を改修するときは、本来は関連する成果物を同時に変更すべきだが、工程遅れが顕著な場合は同時に変更する方がむしろ稀である。そのため、工程遅れのプロジェクトが多い現状では、成果物改修時において同時に変更したという履歴は残りにくい。

したがって、成果物をほぼ同時に変更したという履歴を利用する従来手法では、異なる工程で作成される成果物間の依存関係を抽出することは難しいといえる。

4. 提案手法：操作履歴を利用した依存関係予測

本研究では、変更履歴と閲覧履歴を利用する手法を提案する。以降、変更操作および閲覧操作を総称して単に操作と呼ぶ。なお、操作情報には表1のような情報が記録されているものと仮定する。

表1 操作情報に記録される情報

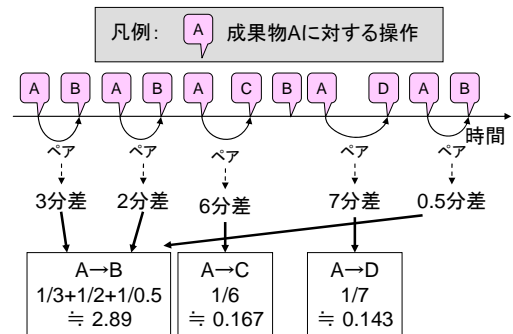
操作情報		
・ 操作対象成果物名	・ 操作者	・ 変更内容
・ 変更後の版番号	・ 操作日時	(変更時のみ)

本手法では、次のような仮定の下、成果物間の依存関係を確度として予測する。

仮定1：

2つの成果物に対するあるユーザの操作について、操作時刻が近く、かつ操作回数多いほど、依存関係が強い

成果物 x から成果物 y への依存関係の確度を、操作履歴をもとに計算する手順の概要を図1に示す。



それぞれの手順は下記の通りである。

1. 操作履歴を検索して、成果物 x の操作情報と、成果物 y の操作情報のペアを抽出する。ペアの組み方は、例えば成果物 A に対する操作の直後に行われる、成果物 B に対する操作を組む方法があり、表 2 のように組み合わせることができる。

表 2 操作情報ペアの抽出例

操作情報 ID	成果物名	操作時刻	
1	A	11:00:00	} ペア
2	B	11:03:00	
3	A	12:00:00	} ペア
4	B	12:02:00	
5	A	13:00:00	} ペア
6	C	13:06:00	
7	A	14:00:00	} ペア
8	D	14:07:00	
9	A	15:00:00	} ペア
10	B	15:00:30	

2. 表 2 のような操作情報のペアの集合から、依存関係の確度を計算する。仮定 1 を満たすような、確度を計算する関数の例として、各ペアの操作時刻差の逆数の和が挙げられる。表 2 の各ペアについて操作時刻差を計算したものを表 3 に、表 3 から依存関係の確度を計算した結果を表 4 に示す。

表 3 各ペアの操作時刻差

先に操作した成果物名	後に操作した成果物名	操作時刻差(分)
A	B	3
A	B	2
A	C	6
A	D	7
A	B	0.5

表 4 依存関係の確度

先に操作した成果物名	後に操作した成果物名	依存関係の確度
A	B	$1/3+1/2+1/0.5 \approx 2.89$
A	C	$1/6 \approx 0.167$
A	D	$1/7 \approx 0.143$

なお、変更対象の成果物名を入力し、影響を受ける可能性がある成果物を一覧化する場合には、表 4 のようなデータに対して、「先に操作した成果物名」が指定の成果物名に一致するものを検索する。さらに、それぞれの「後に操作した成果物名」と一致する「先に操作した成果物名」を再帰的に検索することで、全成果物の中で影響を受ける可能性がある成果物を抽出できる。

5. 考察

本手法は、あるユーザの操作について、操作時刻差が短いほど、かつ操作される回数が多いほど、依存関係が強い、と仮定して依存関係を予測している。このようにすることで、成果物間の依存関係を自動的に抽出し、なおかつ成果物の変更に伴って成果物間の依存関係も変更することがで

きる。さらに、変更履歴に加えて閲覧履歴を使用することで、異なる工程で作成される成果物間の依存関係も扱うことができる。本手法を利用することで、作業コストを増大させることなく、成果物の依存関係を把握することが可能となる。

本手法をさらに有用なものとするために、まず我々が置いた仮定 1 が正しいことを検証する必要がある。

また本手法は、依存関係がない成果物間に対しても依存関係があると誤って予測する可能性がある。そこで、依存関係の予測精度を向上させるために、操作時間差や操作回数等の重みを考慮した適切な確度計算関数を設定することが必要である。

さらに、確度計算関数に新たなパラメータを追加することも検討する必要がある。例えば、著しく成果物の内容を変更した場合には、それまでの依存関係が大きく崩れることが予想されるため、変更操作で行われた変更量も確度計算関数のパラメータに取り込む必要がある。他に、ユーザが手作業で依存関係を記録する場合の正確さも盛り込むために、ユーザが入力する依存関係を取り込むことも考えられる。

そして本手法では、多くの成果物間で依存関係があると予測する可能性が十分考えられる。そこで、依存関係の確度が高いものだけをフィルタリングしてユーザに提示することが必要となる。そのために、確度の高い依存関係の集合と、確度の低い依存関係の集合の間に、有意な差があるということを調べる統計的手法が求められる。

6. まとめと今後の課題

本研究では、成果物間の依存関係を予測するための手法として、変更履歴および閲覧履歴を利用する手法を提案した。本手法では、あるユーザの操作について、操作時刻差が短いほど、かつ操作される回数が多いほど、操作された成果物間の依存関係が強いと判断する。本手法を利用することで、成果物を同時に変更しない場合でも成果物間の依存関係を把握することができ、成果物の品質向上に寄与することと期待している。

今後の課題として、本手法を実開発に適用し、本手法の有効性を評価することが必要である。その後、予測する依存関係の予測精度を向上させるために、確度計算関数の改良、および有意な依存関係のみを抽出できるよう、依存関係の確度から依存関係の有無を推定する手法について、より具体的に検討していく。

7. 参考文献

- [1] 岩田英丈、阿萬裕久、山田宏之：変更履歴情報に着目した依存関係分析、社団法人電子情報通信学会技術研究報告 知能ソフトウェア工学 Vol.106, pp.7-12, 2006
- [2] 岩田英丈、阿萬裕久、山田宏之：クラスの変更履歴と依存関係に基づいた同時更新モデルの提案、社団法人電子情報通信学会技術研究報告 知能ソフトウェア工学 Vol.105, pp.43-46, 2006
- [3] 松村知子、大杉直樹、横森励士、川口真司、松下誠：ファイルの同時変更パターンと変更差分の分析による論理結合関係の自動抽出、ソフトウェアシンポジウム 2005 論文集, pp.104-112, 2005