

過去の変更仕様書を学習することにより 要求文からプログラム変更部分の候補を 推薦するシステムの提案

岩崎陽也¹ 中島毅¹

概要: 流用ソフトウェア開発において、変更要求に対するソフトウェアへの影響範囲を決定する影響分析は、その開発の品質と効率に大きく影響するため重要な位置付けをもつ作業である。この影響分析の精度と効率の向上のために、成果物間のトレーサビリティが利用されるが、その確立・維持は一般に困難である。本研究報告では、多品種小変更型開発において、大量に作成される変更設計書に注目し、変更要求とそれに対応するコンポーネントへの変更の履歴を機械学習し、変更要求を入力することで、どのコンポーネントへの変更影響がある確率が高いかを提示する機能をもつ支援システムを提案する。さらに、変更分析作業への貢献を評価するため、網羅範囲比率と網羅範囲正答率という評価指標を提案する。本提案システムを評価するために、提案支援システムを試作し実際の多品種小変更型開発に適用し、提案評価指標により評価を行なうことで提案システムの有効性を示した。

キーワード: 流用開発, トレーサビリティ, 機械学習, 変更分析

1. 研究背景

ソフトウェア開発では、各開発工程における成果物が多く、工程間にまたがる成果物（およびその記述要素）間の関連性がわかりづらいことから、変更要求に対する影響分析が容易でない。開発する機能を共有しつつ製品バリエーションに合わせて少しずつ異なる仕様を実現する複数の開発を、短期間に並行して行う、いわゆる多品種小変更型開発[1]の場合、その問題はさらに深刻となる。

トレーサビリティとは、開発工程における複数の成果物間に確立された関連性のことである。これにより、要求事項の成果物間での対応漏れを確認することや仕様変更の際の影響分析にも利用できるため、保守品質を底上げする重要な役割を果たす。特定の成果物間の関連を表した情報をトレーサビリティリンクという[2]。トレーサビリティリンクはトレーサビリティマトリクスやタグ付方式を用いて実現するが[3]、それらの確立・維持は開発者にとって作業負担が大きい。

開発者がトレーサビリティの維持管理をせずに、自動的に成果物文書や構成管理の情報から自動あるいは半自動的にトレーサビリティを抽出できれば、流用開発における変更要求の影響範囲の特定を、誤りなく効率的に進めることができる。と期待できる。

本研究報告では、多品種小変更型開発において大量に作成される変更設計書に注目し、変更要求とそれに対応するコンポーネントへの変更の履歴を機械学習し、変更要求を入力することで、どのコンポーネントへの変更影響がある確率が高いかを提示する機能をもつ支援システムを提案する。さらに、そのような機能は、変更要求に対応する実装を決定する既存（方式）設計書のレビュー範囲を限定する

ために利用されることに注目し、網羅範囲比率と網羅範囲正答率という評価指標を提案する。

本提案システムを評価するために、提案支援システムを試作し、実際の多品種小変更型開発に適用し、提案評価指標により評価を行なった。その結果、提案するシステムは全コンポーネントの20%程度にレビュー範囲を限定し、その範囲の正答率を50%以上にする効果を示し、有効性を証明することができた。

2. 関連研究

2.1 トレーサビリティリンクの回復の手法

従来トレーサビリティリンクの回復には、Retrospective（後ろ向き）な方法と Prospective（前向き）な方法がある[4]。Retrospective な方法では、ソフトウェア成果物から遡ってトレーサビリティリンクの推論を行う。例えば、テキストから候補リンクを抽出する方法や、コード解析や実行時の監視により自動的にリンクを得る方法、一部の正解付きリンクを学習して正解が未知のリンクを予測する方法などがある。

一方、Prospective な方法でモデル変換の痕跡など、解析によってトレーサビリティリンクとなるものを然るべき工程にて与える。例えば、成果物IDをソースコードに埋め込む方法や、工程ごとの成果物間の整合性を自動で維持する方法などがある。

2.2 Retrospective な手法とその問題点

本研究では、トレーサビリティリンクの確立・維持が行われていないプロジェクトを対象とするため、Retrospective な手法に注目する。

Retrospective な方法は大きく、ルールベース手法と情報検索手法の二つの手法が用いられてきている[2]。

¹ 芝浦工業大学 Shibaura Institute of Technology

■ルール

「要求仕様書の要求文と設計文書の設計文の双方に、『検索』という文字が双方に含まれている場合、それらの間にはトレーサビリティがあると識別する」

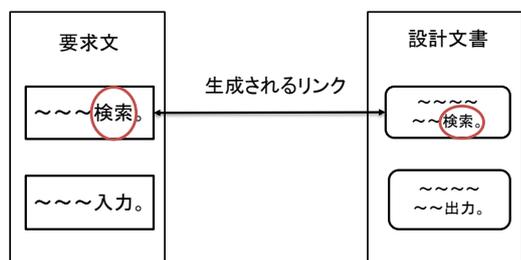


図1 ルールベース手法の例

以下に、2つの手法と適用上の問題点を示す。

(1) ルールベース手法

ルールベース手法とは、一定のルールを満たすか満たさないかによってトレーサビリティリンクがあるかないかを判断する手法である。

図1にルールベース手法の簡単な例を示す。図1において、ルールは、「要求仕様書の要求文と設計文書の設計文の双方に、『検索』という文字が双方に含まれている場合、それらの間にはトレーサビリティがあると識別する」である。このルールに基づき、要求仕様書と設計文書が解析され、ルールを満たす要求文と設計文書間にトレーサビリティリンクが生成される。

この手法の有効性は、有効なルールが抽出できるかどうかによって依存する。一般に、有効なルールは、あるソフトウェア開発にのみ存在する固有名詞の存在や、開発上の規約（プログラムへの特別なコメントの挿入）など、ドメイン依存性の高いものが多い傾向があるため、汎用的かつ有効なルールは用意できない。そのため、その適用にあたって、ルール作成のための事前の綿密な分析と効果の検証の作業が必ず必要になる問題点がある。

(2) 情報検索手法

情報検索手法とは、段落や文などの文章の単位を対象として、そのテキストとしての類似度の高さにより、トレーサビリティの有無を決定する手法である[5]。

図2に情報検索手法の簡単な例を示す。図2において、入力文は、ユーザがトレーサビリティの有無を調べたい要求文である。比較文は、設計文書から一文ずつ文章単位で抜き出したものである。情報検索手法では、入力文と比較文の意味的類似度を類似度算出アルゴリズムで求める。そして、「入力文に対してトレーサビリティの存在確立が高い比較文は意味的類似度が高い」という仮定のもと、類似度の高い順に比較文を出力する。

この手法では、事前の人手による分析を必要とせず、ルールベース手法よりも汎用性は高い。しかし、テキストの分析のために、要求仕様書や設計文書に出てくる文や単語

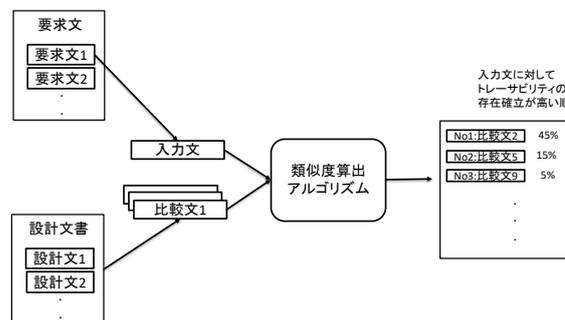


図2 情報検索手法の例

の意味のベースを与えるコーパスを必要とする。しかし、要求仕様書や設計文書は、一般文書とは異なり特殊な用語を多く用い、それがトレーサビリティのキーになるため、この手法の有効性を確保するためには、コーパスの選び方に注意が必要となる。

また、この手法は、文章単位間の類似度を用いるため、ドキュメント間に適用が限られ、例えばプログラムとの間には適用できない。

3. 支援システムとその評価指標の提案

3.1 支援システムの提案

本研究では、多品種小変更型開発における開発支援を対象とし、その変更要求の影響範囲の決定を正確かつ効率的に行うための支援機能の開発を目指した。

(1) 提案システムとその構成

本研究では、大量に作成される変更設計書に注目した。変更設計書は、派生開発 XDDP[6]で使われる変更要求一つ一つに対してその実装方法が検討されるドキュメントであるが、本研究で対象とするものは、厳密にその形式に従うものではなく、個々の変更要求とそれに対応するプログラムコンポーネントへの変更の履歴が抽出可能であるものであればよい。

提案する支援システムは、図3に示すように、変更設計書を学習しておき、入力の変更要求に対して、どのコンポーネントへの変更影響があるか、どのコンポーネントの確率が高いかを出力する機能をもつ。

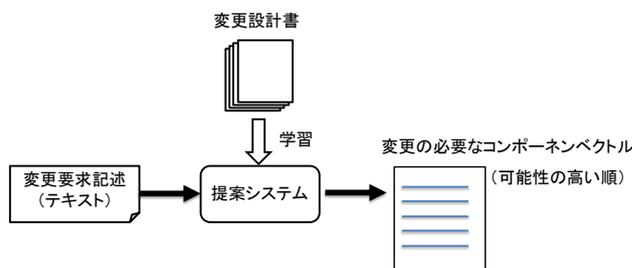


図3 提案システムの概要

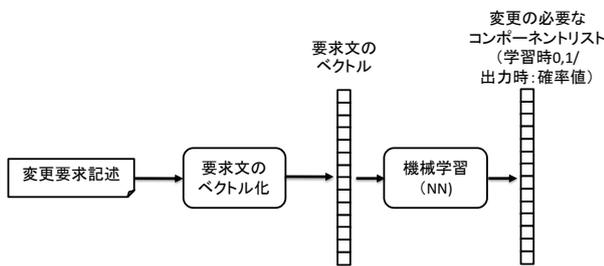


図5 提案システムの学習部の構成

図4に提案システムの構成を示す。提案システムは、変更要求記述を単語ベクトル化する部分と、その単語ベクトルを入力、コンポーネン変更ベクトル(その変更要求により、実際に変更があった場合1/ない場合0にしたコンポーネント分のベクトル) を出力とし、学習したニューラルネットワーク(NN)部分からなる。

提案システムは、ドメインに関する綿密な事前の分析を行う必要はない。また変更要求記述のベクトル化のため、意味解析にコーパスを必要とするが、これは類似度計算のためではなくベクトル化のためだけに使うため、出現する単語をカバーできればコーパスの設計の工夫を必要としない。このように、提案手法は、前章で述べた従来の Retrospective な手法の問題点をクリアしている。

(2) 変更要求記述の単語ベクトル化とその実装

変更要求記述文のベクトル化手順には、単語抽出、ベクトル化、ベクトルの結合化の3つの段階がある。その段階をどう行うかで、図6に示すようなバリエーションが考えられる。

- ① 単語抽出：変更要求変更記述から、要求の意味を代表する単語を抽出すること。全選択、名詞のみ抽出、人手による抽出などがある。
- ② 単語のベクトル化：抽出した単語を N 次元のベクトルに変換すること。Word2vec が代表的な手法であるがその中でも CBOW モデルや skip-gram のようなバリエーションがある。

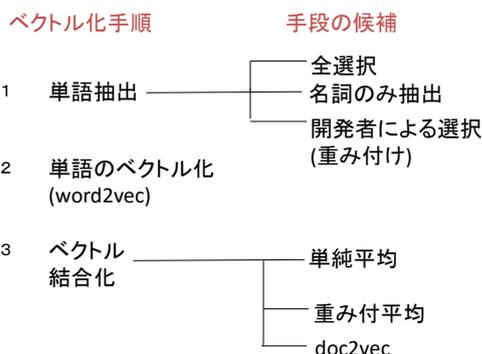


図6 要求文のベクトル化

- ③ ベクトル結合化：抽出した単語ベクトルを結合して変更要求記述全体の意味を表す一つのベクトルを生成すること。抽出した単語ベクトルの単純平均、重み付け平均、doc2vec[7]の実装方法などがある。

どの方法で実験を行うかで、学習精度が変わってくる可能性があるため、今回は、表1に示す3通りの組合せを実装した。

表1 要求文のベクトル化の実装方式

実装 No	単語抽出	単語ベクトル化	ベクトル結合化
実装 1	名詞抽出	word2vec	単純平均
実装 2	全選択	doc2vec	
実装 3	名詞抽出	doc2vec	

以下は、利用した言語解析のオープンソースである。

- 言語解析ライブラリ:genism(ver3.8.3)
- 形態素解析エンジン:mecab(mecab-Python3 ver1.0.1)

また、形態素解析および wor2vec/doc2vec による単語ベクトル化において、変更仕様書にしかない固有名詞を考慮するため、コーパスには Wikipedia のデータに加えて、4.2 節で述べる変更仕様書の要求文も用いた。

(3) 機械学習

機械学習として、図7に示すような隠れ層4層のニューラルネットワーク(NN)を利用した。入力、変更要求記述のベクトルは100次元とし、出力はあらかじめ存在がわかっているコンポーネントを31個とそれ以外のものの合計32個に固定した。

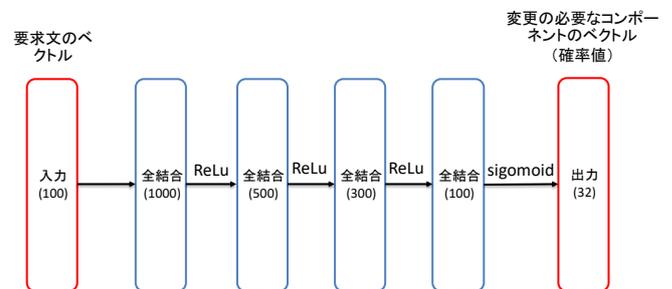


図7 採用したニューラルネットワークの構造

学習のパイパーパラメータとして以下を利用した。

- ・ 学習を行う回数：50
- ・ バッチサイズ：50
- ・ 学習率：0.1
- ・ 損失関数：バイナリークロスエントロピー誤差
- ・ 重みパラメータ更新方法: SGD (確率的勾配降下法)

(4) システムの出力方法

システムの出力は、変更の必要なコンポーネントの候補

のリストを出力する。ニューラルネットワークの出力は、sigmoid により [0, 1] の値域をもつ値をだすので、これを確率値として用い、このリストを値が高い順に並べて表示するようにする。

これにより開発者は、変更要求の必要な確率の高いコンポーネントの記述を、順に吟味していくことが可能になる。

3.2 評価方法の提案

本研究では、上記の支援システムは、変更要求に対応する実装を決定する既存（方式）設計書のレビュー範囲を限定するために利用されることに注目し、**網羅範囲比率**と**網羅範囲正答率**という2つの精度評価指標を提案する。図8は、それらの指標値を計算するための、変数の定義を与える。

(1) 網羅範囲比率 F

網羅範囲比率 F とは、ある変更要求記述に対して、変更の必要なコンポーネントは、候補のリストの上位何%までに全て含まれるかの割合である。網羅範囲比率 F は次式で表すことができる。

$$F = \frac{C}{W} \quad (\text{式1})$$

ここで、

W : 学習の際に出てくるコンポーネントの総数

C : 実際に修正が必要なコンポーネントが全て含まれている候補のリスト上の順位

網羅範囲比率 F は、変更要求に対応して成果物中においてレビューすべき範囲の割合を表し、0 に近いほど良い指標である。

(2) 網羅範囲正答率 A

網羅範囲正答率 A とは、上記網羅範囲の中に正解のコンポーネントがどれくらいの比率で含まれているかを表す。に対する評価値は A は次式で表すことができる。

$$A = \frac{T}{C} \quad (\text{式2})$$

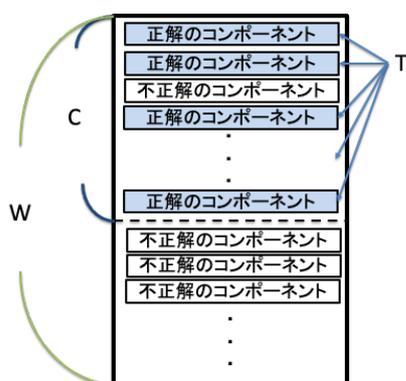


図8 指標値の各変数

ここで、

T : 対象とする変更要求記述に対して実際に変更が必要なコンポーネントの総数

網羅範囲正答率 A は、変更要求に対応して成果物中においてシステムがレビューを勧める範囲において、無駄な作業とならずにすむ割合であり、1 に近いほど良い指標である。

4. 評価実験

4.1 評価目的

この章では、3.1 で示した提案システムが、3.2 で示した提案評価方法により、望ましい精度を得ることができるかどうかを評価する。

4.2 対象プロジェクトと学習データ

対象となるプロジェクトは、多品種小変更型の開発[2]である。実際には、図9に示すように、同一プログラム母体に対して、年間30件程度の流用開発プロジェクトが走り、各プロジェクトは平均10件程度の変更要求を実装しているのが典型的な姿である。

各プロジェクトは、複数の変更要求に対して、それぞれ変更設計書を作り開発を行っている。変更影響分析は、変更要求に対して、方式設計書と詳細設計書の該当部分を開発者の経験により探索・レビューすることにより実施されている。

変更設計書は、開発成果物（要求仕様書、方式設計書、詳細設計書、およびソースコード）のどこに変更があったのかを記録した文書となっている。変更設計書の内容は、開発終了時に、開発成果物に反映される。

この変更設計書は、過去数年単位で存在しており、全てを合わせると1000個程度存在している。この資産の内405個を、学習とテストデータとして利用した。

変更設計書は、MS-EXCEL形式の一定のフォーマットとなっているファイルであり、その中の以下の情報を利用した。

- ・ 変更要求記述（20~400文字のテキストデータ）
- ・ 変更を実施したコンポーネントの名前のリスト

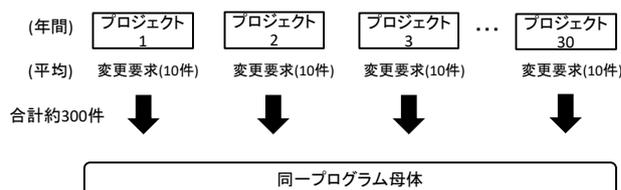


図9 対象プロジェクトにおける年間プロジェクト数と平均変更要求数

4.3 実験方法

実験では、図 10 に示すように、上記 405 個の変更設計書を用い、訓練データ 385 個として学習させ、テストデータ 80 個で評価を行なった。また、この一連の流れを他の組み合わせで 5 回行ない結果はその平均を算出した。

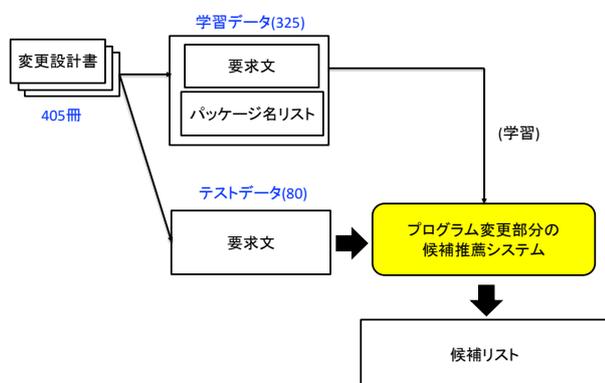


図 10 実験方法

4.4 実験結果

単語抽出に名詞のみ word2vec (実装 1) と名詞のみ doc2vec (実装 2)、全選択 doc2vec の精度比較実験の結果を示す。表 2 に網羅範囲比率について、表 3 では網羅範囲正答率の実験結果を示す。

(1) 単語抽出に「名詞のみ」を選択した精度比較実験

表 2 および表 3 において、データ 1 はランダムに取り出したテストデータ 80 個のことを表し、平均はデータ 1-5 の実験結果の平均を表している。

名詞のみ抽出した word2vec と doc2vec では網羅範囲比率、網羅範囲正答率どちらの指標でも doc2vec を用いた手法の方が良い精度を示している。

(2) Doc2vec 同士の精度比較実験

2 つの単語の抽出方法、すなわち名詞のみ抽出したものと全単語を抽出したものを利用し、doc2vec を組み合わせた 2 つの手法の比較実験を行った。

Doc2vec を用いた際にはわずかながら名詞のみ抽出して doc2vec にかけた手法の方が網羅範囲比率、網羅範囲正答率ともに良い精度を示している。

4.5 考察

上記実験結果は、変更要求文から変更の可能性があるコンポーネント候補を出す機能の精度評価になる。

訓練データとテストデータの組み合わせを変えて行った 5 回の試行に対して、どの要求文ベクトル化の手法を用いても、網羅範囲比率はほぼ 20% を切る値となっている。これはコンポーネント総数の 1/5 をレビューすれば良いことになる。また、網羅範囲正答率も 50% を超える値であるため、無駄な作業も少ないという結果を示している。この

表 2 網羅範囲比率

データ系列	word2vec	doc2vec 全結合	doc2vec 名詞のみ
データ 1	22.45%	21.22%	21.57%
データ 2	19.37%	19.21%	18.98%
データ 3	16.67%	15.32%	15.59%
データ 4	15.28%	14.08%	13.77%
データ 5	13.35%	12.77%	13.12%
平均	17.40%	16.50%	16.60%

表 3 網羅範囲正答率

データ系列	word2vec	doc2vec 全結合	doc2vec 名詞のみ
データ 1	54.77%	56.49%	57.60%
データ 2	56.95%	54.67%	58.82%
データ 3	62.37%	67.80%	67.99%
データ 4	65.62%	74.97%	72.67%
データ 5	75.05%	77.15%	78.85%
平均	62.90%	66.20%	67.10%

ように、支援機能を用いることで、影響分析の大幅な作業負担の軽減になるということがわかった。

3 通りの要求文のベクトル化について、精度比較を行なった結果は、網羅範囲比率と網羅範囲正答率のどちらも doc2vec を用いた手法の方が良い精度となったが、単語抽出に、全単語か名詞のみ選択するのかわではそんなに精度に影響を与えないことがわかった。

今後の課題として、単語抽出の方法としてまだ行なっていない開発者による選択を行なった場合に精度が良くなるかどうかを検証する必要がある。

5. まとめ

多品種小変更型開発において、大量に作成される変更設計書に注目し、変更要求とそれに対応するコンポーネントへの変更の履歴を機械学習し、変更要求を入力することで、どのコンポーネントへの変更影響がある確率が高いかを提示する機能をもつ支援システムを提案する。さらに、変更分析作業への貢献を評価するため、網羅範囲比率と網羅範囲正答率という評価指標を提案する。本提案システムを評価するために、提案支援システムを試作し実際の多品種小変更型開発に適用、提案評価指標により評価を行な、提案

システムの有効性を示した.

謝辞

本研究を進めるにあたり, データ提供と有益な議論をさせて頂いた三菱電機(株)情報技術総合研究所の徳本修一氏, 磯田誠氏, 塚本良太氏, 高橋加寿子氏に感謝致します.

参考文献

- [1] 長峯 基, 中島 毅: 多品種小変更型開発におけるコア資産保守・製品導出手法の改善と実践, 情報処理デジタルプラクティス, Vol.10, No.3 (2019).
- [2] 宇田川佳久 他: 情報システム開発におけるトレーサビリティの事例と今後. 情報処理 51.2:150-158,(2010).
- [3] 徳本 修一, 高橋 洋一, 田村 直樹, 中島 毅: 構造化タグ方式による要求トレーサビリティリンク構築ツール, 電子情報通信学会論文誌-D, Vol.J104-D, No.02, pp. 105-118, Feb. 2021.
- [4] 小方真平 他: 長期的大規模プロジェクトのドキュメントを事例とした, トレーサビリティリンク回復の実践と課題. ソフトウェアエンジニアリングシンポジウム 2019, 論文集, pp.50-59(2019).
- [5] Tomas Mikov et al.: Efficient estimation of word representations in vector space, arXiv preprint arXiv:1301:3781(2013).
- [6] 清水 吉男: 「派生開発」を成功させるプロセス改善の技術と極意, 技術評論社 (2007).
- [7] Le, Quoc, and Tomas Mikolov: Distributed representations of sentences and documents, International conference on machine learning, pp.1186-1196(2014).