

文章間の類似度を用いた要求文書内の変更の波及解析

2019SE020 木方 友香 2019SE038 森永 紫音 2019SE041 中野 樹

指導教員 佐伯 元司

1 はじめに

研究背景と研究課題

システム開発では、要求分析の段階であっても様々な理由で要求変更が発生し、要求仕様書に変更が行われることがある。変更が起きた際には、変更部分以外の箇所もその変更の影響を受ける可能性がある。これらを放置すると、矛盾などが発生し、後の工程に悪影響を与える。現在、影響の特定は人間の手によって行われている。しかし、人間による波及解析は、多くの時間や手間を必要とするという問題を抱えている。この問題を解決するために、自動で影響を受ける部分の特定をすることが必要である。

本研究では、変更文と意味的に類似した文が変更の影響を受けやすいと考え、要求文書変更により影響を受ける要求文を自動的に発見する手法を提案する。また、実際の要求文書を用いて提案手法の評価を行い、その有効性を確認する。

2 準備

本研究で使用する3つの技術、形態素解析技術、単語のベクトル化技術、類似度計算法について説明する。

1. MeCab

MeCab[1]は京都大学で開発されたオープンソース形態素解析エンジンである。言語、辞書、コーパスに依存しない汎用的な設計を基本方針としており、条件付き確率(CRF)に基づく高い解析精度が特徴にあたる。本研究では文章の分かち書きを目的として使用する。

2. Word2Vec

Word2Vecは、Googleの研究者トマス・ミコロフ氏によって提案された手法で、文章中の単語を数値ベクトルに変換してその意味を把握する自然言語処理の手法のことである[2]。Word2Vecは入力層と出力層、1層の隠れ層で構成されるニューラルネットワークであり、単語や単語間の結びつきや類似度を学習するという特徴を持つ。

3. cosine 類似度

cosine 類似度とは、2つのベクトルの類似性を表す尺度であり、2つのベクトルがなす角度の余弦のことである。具体的な式は以下に示す。

$$\text{Cos}(\vec{a} \cdot \vec{b}) = \frac{\vec{a} \cdot \vec{b}}{|\vec{a}| \cdot |\vec{b}|}$$

3 アプローチ

3.1 提案概要

変更の影響を受ける文は、変更文と意味的な類似性があると考えられる。本研究では要求文と変更文の意味的な類似度を計算し、類似度の高い要求文を波及範囲候補とする手法を提案する。また、本研究では要求分析者からみた要求文と変更文の意味的な類似性を「人間判断の類似性」、本手法で出力された意味的な類似度を「Cos類似度」と呼ぶ。図1に本手法の方式を示す。入力として要求文書に対する変更文を入力した時の処理を表している。出力となる波及範囲一覧には、波及候補として要求文が変更文との類似度順にリストアップされる。

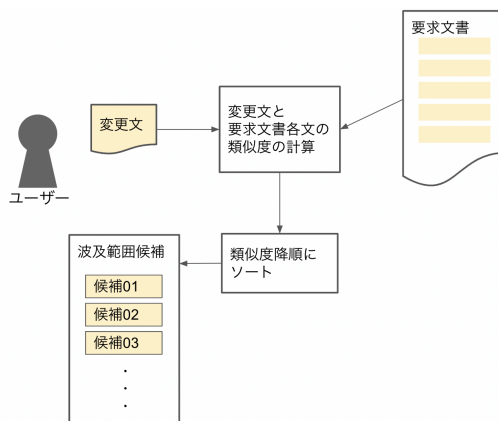


図1 提案概要

3.2 変更文と要求文書内の各文の類似度計算

3.2.1 データ処理

変更文と要求文書内各文のデータ処理を行う。MeCabを用いて形態素の情報を取得し、以下の処理を行う。

1. 文章を分かち書きにする
2. 各単語を原型へ変換
3. 名詞と動詞のみを抽出

要求文書における変更によって影響を受ける文は、変更文と同じオブジェクトを指す、もしくは関わっていることが多いと考えられる。このオブジェクトは名詞であることがほとんどのため、本研究では名詞を抽出することとした。さらに、動詞にもオブジェクトの特徴が一定程度表れると考え、動詞も抽出することとした。

4. 一部の頻出単語を除去

「する」、「こと」、「いる」などの、どの文章にも

出現しやすい頻出単語は、変更の影響を受けない文の類似度が高くなってしまいう原因となることが考えられるため除去する。

3.2.2 ベクトル化

データ処理したものを Word2Vec モデル (4.2 で説明) でベクトル化する。ベクトル化の流れは以下である。

1. 単語ごとのベクトルデータを算出
2. ベクトルデータの平均を求める

以上の手順で算出したものを本研究では文のベクトルとする。

3.3 類似度算出

要求文書各文のベクトルと変更文ベクトルから、cos 類似度を用いて類似度を計算する。

4 プロトタイプの作成

4.1 実装環境

本研究の手法を実装するプロトタイプの実装環境を以下の表 1 に示す。

表 1 プロトタイプ実装環境

コンポーネント	コンポーネント名	バージョン
OS	Ubuntu	22.04.1
形態素解析器	mecab-python	1.0.5
配列化ツール	numpy	1.23.4
ベクトル化ツール	gensim	4.2.0

4.2 ベクトル化モデル

Word2Vec モデル作成時のパラメータを表 2 に示す。

表 2 Word2Vec モデルのパラメータ

ニューラルネットワーク	CBOW
コーパス	日本語 Wikipedia ダンプデータ (データサイズ: 3.1GB)
次元数	200
単語最低出現回数	5
window サイズ	20
epoch 数	10

5 実験の実施

5.1 実験の目的

提案手法の有効性を調査し、結果の考察をするため以下の 2 点を実験で調査する。

RQ1: 変更の影響を受ける要求文は、人間判断の類似性が高いか

RQ2: 変更影響を受ける人間判断の類似性のある文が、本手法で取得した波及範囲候補の上位に現れるか

5.2 実験手順

実験の手順を図 2 に示す。

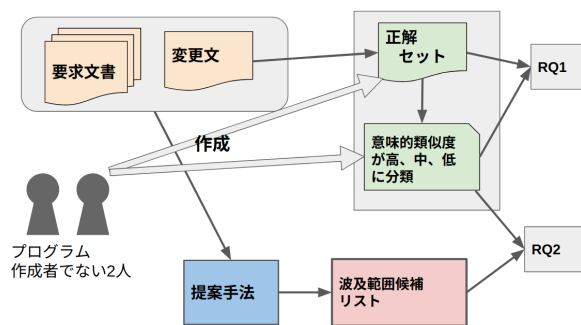


図 2 実験手順

5.2.1 変更文の作成

実験で使用する要求文書への変更内容を作成する。今回の実験では、実際の変更履歴を含んだ文書を用意できなかったため、人為的に変更箇所を作成した。

5.2.2 正解セットの作成

1 つの変更文に対して影響を受ける文を特定し、それを正解セットとした。作成には、提案手法実装者でない 2 人で作成した。正解セットの客観性を上げるため、それぞれ独立して正解セットを作成し、さらにお互いの正解セットを照らし合わせ、意見が異なる部分があれば話し合いを行い作成した。さらに、正解文の中から、変更文との人間判断の類似性「高」「中」「低」の三つに分類する。

5.2.3 評価指標

評価には平均適合率に加え、網羅範囲比率と網羅範囲正答率を使用する。図 3 はその説明図である。波及範囲候補が全体で W 個あり、上位から C 番目までに影響を受ける文がすべてリストアップされているとする。さらに C 番目までに影響を受ける文は T 個含まれる。

これらを人間判断の類似性「高」、人間判断の類似性「高・中」、人間判断の類似性「高・中・低」の 3 つのパターンで算出する。

1. 平均適合率

出力された候補のある順位まで見たときにどれぐらい正解が入っているかを示す適合率を、正解文の順位にしたがって平均したものである。

$$AP = \frac{1}{n} \sum_{i=1}^n P_i$$

P_i は出力の順位リストを i 番目まで見たときの適合率、つまり正解数を i で割ったものである。

2. 網羅範囲比率

上位何 % までに正解文が出きったかを表すものである。値が小さいほど良い結果であると言える。

$$F = \frac{C}{W}$$

3. 網羅範囲内正答率

提案手法で作った順位リストで正解となっているものの最下位順位まででの適合率である。値が大きいほど良い結果であると言える。

$$A = \frac{T}{C}$$

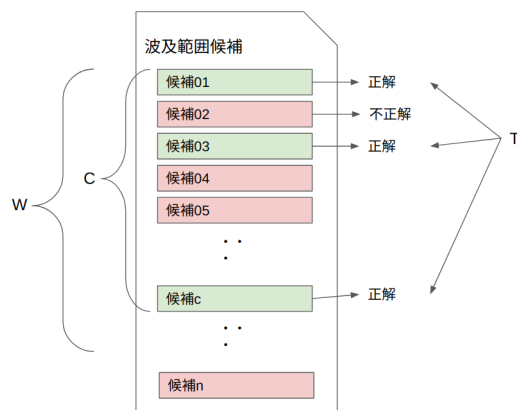


図 3 評価指標

5.3 実験の実施

実験に使用したテストデータについて、表 3, 4 に示す。

表 3 使用した要求文書

要求文書	会津大学図書管理システム
要求文数 (行数)	295

表 4 人間判断の類似性別正解セット数

	人間判断の類似性 (高)	人間判断の類似性 (中)	人間判断の類似性 (低)	全正解セット数
変更文 1	3	2	1	6
変更文 2	3	0	4	7
変更文 3	7	1	0	8

6 実験結果

実験結果として表 5, 6, 7 にそれぞれ平均適合率, 網羅範囲比率, 網羅範囲内正答率を示す。

7 結論と考察

7.1 実験目的について

(1)RQ1 への回答

全正解セット 21 文中, 人間判断の類似性「中」以上は 16 文であり, 約 75% の文が意味的類似性がある文であ

表 5 平均適合率

実験番号	人間判断の類似性 (高)	人間判断の類似性 (高・中)	人間判断の類似性 (高・中・低)
変更文 1	1	0.6466873	0.54
変更文 2	0.3166666	0.3166666	0.2355
変更文 3	0.182153	0.2258846	0.2258846
平均	0.499606533	0.3964128333	0.333794866

表 6 網羅範囲比率

	人間判断の類似性 (高)	人間判断の類似性 (高・中)	人間判断の類似性 (高・中・低)
変更文 1	0.0101	0.2837	0.75
変更文 2	0.0337	0.03367	0.4256
変更文 3	0.3277	0.3277	0.3277
平均	0.1238	0.215	0.5011

表 7 網羅範囲内正答率

	人間判断の類似性 (高)	人間判断の類似性 (高・中)	人間判断の類似性 (高・中・低)
変更文 1	1	0.0595	0.0270
変更文 2	0.3	0.3	0.0555
変更文 3	0.0824	0.0824	0.0824
平均	0.4608	0.14733	0.055

た。このことから, 変更の影響を受ける要求文は, 変更文と人間判断の類似性が高い傾向があると言える。

(2)RQ2 への回答

実験結果の, 人間判断の類似性「中」以上の結果に着目する。変更文 2 が最も良い網羅範囲比率, 網羅範囲内正答率となったが, 人間判断の類似性「中」の正解セットが 0 であるため, この結果だけでは良い結果であったと結論付けられないと考えられる。そのためこの変更文 2 を外れ値として無視し, 変更文 1, 3 の網羅範囲比率を平均すると, 上位約 30% の要求文を見れば, 意味的類似性のある影響を受ける文の特定ができるという結果となった。

このことから, 影響を受ける人間判断の類似性の高い文は, 本手法で取得した波及範囲候補の上位に現れたと言える。

7.2 結論

実験により, 変更に影響を受ける文は意味的類似性が高く, それらを本手法によって上位に表示させることができた。

7.3 誤答の原因分析

次に, 正解文でないが上位に表示された文 (偽陽性文) と, 正解文であるが低い順位に表示された文 (偽陰性文) に

ついて分析を行う。

1. 偽陽性

偽陽性の要求文に、変更文と共通の単語、もしくは類似度が非常に高い単語が含まれていることが多いことがわかった。

2. 偽陰性

共通の単語が出現しているにも関わらず類似度が低いとされた文は、出現する単語数が多く、単語ベクトルを平均した際に、文の特徴が現れなくなったことが考えられる。

これらの原因の1つとして、すべての単語が同じ重みであることが挙げられる。単語の出現頻度や、変更文中の重要な単語などを考慮し、単語の重み付けなどを行うことによりこれらの偽陽性、偽陰性を少なくすることができると思われる。

7.4 妥当性への脅威

(1) 内的妥当性

検出プログラムを作成した人は正解セット作成には関与していないため、この部分の内的妥当性への脅威は軽減されていると考えられる。しかし、変更文、正解セットの作成は著者（の一部）が関与しているため、内的妥当性への脅威はぬぐえない。

(2) 外的妥当性

使用した要求文書は1つの分野のみであったため、より多くの分野の要求文書で実験を行う必要がある。また、変更事例も3つのみであり、著者らが作り出したものであるため、多様な変更事例や現実のプロジェクトでの変更事例で行う必要がある。

8 関連研究

本研究と類似した研究として塚本ら [3] の研究では、変更による影響範囲の特定を支援するツールの提案を行っている。成果物を一定の表示単位に分割し、Doc2Vecによって表示単位内の文章のベクトルと要求文書の文章のベクトルを段落ごとに求め、全ての段落の組み合わせで \cos 類似度を計算し、類似度上位の平均をその表示単位の類似度としている。既存研究の評価において網羅範囲比率は 0.33、網羅範囲正答率は 0.14 であった。本研究においては意味的類似が「中」以上であるものの平均の網羅範囲比率は 0.215、網羅範囲内正答率は 0.147 であった。既存研究とは実験データや、文の区切り方など異なる部分があるため評価指標の比較は行えないが、本研究の評価では既存研究と同等かそれ以上の値が得られていることから、我々の提案する Word2Vec での波及範囲特定支援は特に意味的類似度の高い文に対して一定の有効性があると言える。しかし、意味的類似の低い文に関しては一定程度の有効性は見込めるものの、既存研究と比べて評価指標は低い。

9 今後の課題

本研究によって得られた知見から、今後の課題を適合率上昇と妥当性の向上の2つの面から上げていく。

9.1 適合率上昇

1. 単語の重み付

考察で、偽陽性文、偽陰性文の原因として、単語の重みが同じであることが挙げられた。これを確かめるため、tf-idf の適用などを行い、適合率の上昇を目指す。名詞、動詞以外の品詞についても考慮し、適合率の比較を行う必要がある。

2. 要求文書や特定分野に特化したモデルの作成

本研究では日本語 wikipedia をコーパスに使用し、多くの文書に対応できる汎用的なベクトル化モデルを作成した。しかし、ソフトウェア開発では特有の単語や表現が出現する。要求文書や、特定の分野に特化したモデルを作成することによって適合率上昇を目指す。

9.2 妥当性の向上

1. 実験データ数

本研究では、1つの要求文書を使用して実験を行った。これは、外的妥当性への脅威の可能性がある。今後、より多くの分野の要求文書を使用して実験を行う必要がある。

2. 実際のデータ適用による評価

本研究では、インターネットで収集した完成された要求文書に対し、我々自らが変更内容と波及範囲を決めた。そのため、考察の通り内的妥当性は低いと考えられる。このことから、実際の要求文書や変更履歴を使用した実験を行い、妥当性を担保する必要がある。

10 まとめ

本研究では Word2Vec を用いた要求文書の波及解析支援手法の評価を行った。プロトタイプを実装し、実験を行った結果、本手法の有効性を示した。その一方で、今後改善できそうな点や課題も見つかった。

参考文献

- [1] MeCab: Yet Another Part-of-Speech and Morphological Analyzer, <https://taku910.github.io/mecab/>, [アクセス日:2022.9.21]
- [2] Le, Q. M, T. , "Distributed Representations of Sentences and Documents", CoRR, abs/1405.4053.pp.1-9, 2014, [アクセス日:2022.9.21]
- [3] 塚本良太, 徳本修一, 高橋加寿子, 磯田誠, 中島毅, 「文章の類似度を用いた影響分析方式の提案」, 研究報告ソフトウェア工学 2021-SE-208 1-6 (Jul.2021)