

類似性を用いたユースケース記述変更の影響分析

2018SE083 SUNJiahong

指導教員: 佐伯 元司

1 背景と研究課題

ユースケース分析法は開発の初期段階で要求を明確化するためによく使われている。その一方で、要求変更は頻繁に起こり、ユースケース記述を変更する必要がある。しかし、ユースケース同士は関係を持っており、一つのユースケースが変更されるとほかのユースケースにも影響を及ぼすことがある。ユースケース間の関係が明示されていないような場合には、影響を及ぼすユースケースを発見することが難しい。

本研究は、ユースケース記述変更により影響を受けるユースケースを自動的に発見する方法を提案する。また、実際のユースケース記述を用い評価を行い、その有効性を確認することを本研究の課題とする。要求仕様書についての同様の手法はあるが[1]、ユースケースは構造化して書かれていること、各セクションで書くべき内容を規定していることで、精度向上が期待できる。

2 本研究で使用する技術

(1) Doc2Vec [2]

任意の長さの文書を固定長のベクトルに変換する技術で、文やテキストに対して分散表現を獲得することができる。Doc2Vec は文章の意味をベクトル空間内のベクトルで表し、意味的に近い文章のベクトルは近くなる。本研究は文章のベクトル化を目的として使用する。

(2) MeCab¹

MeCab 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いる。本研究は文章の分かち書きを目的として使用する。

3 提案手法

影響を受けるユースケースは変更要求のあるユースケースと意味的に関連があると考えられる。意味的に関連のあるユースケースを見つけるために、本研究ではユースケース記述同士の意味的な類似度を計算する。変更されるユースケースと類似度の高いユースケースを選び出し、影響範囲を特定する手法を提案する。

図 1 は提案手法のプロセスを示す。まず、変更要求のあるユースケース記述 A とそれ以外の同じシステムのユースケース記述 B1, B2, …を入力データとする。形態素解析器 MeCab を使い、入力ユースケース記述を品詞単位に分かち書きする。次に Doc2Vec を使い、分かち書きされたユースケース記述をベクトル化し、各ユースケー

ス記述のベクトルデータを取得する。ここで事前に Doc2Vec のベクトル化モデルを作成する必要がある。最後に A のベクトルデータと B1, B2, …のベクトルデータに対し、cosine 類似度を計算し、類似度の降順にソートした結果をリストとして出力する。

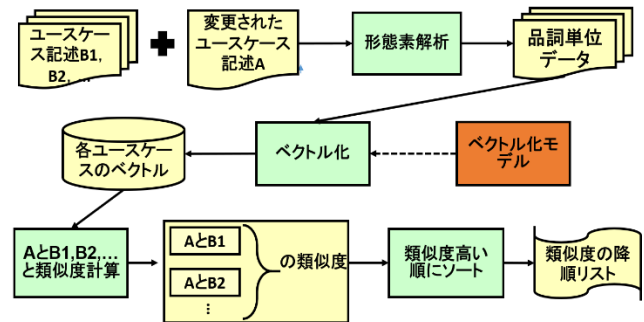


図 1 提案手法

4 使用した Doc2Vec モデル

類似度の精度を比較するために 2 種類のコーパスを使用した。以下はそれを示す。

(1) 学習済みモデル

日本語 Wikipedia の学習済 Doc2Vec モデル dmpv300d²を使用する。ここでは Doc2Vec の次元数は 300、学習回数は 20 と定義されている。

(2) 学習させたモデル

ユースケース記述は特有の文書スタイルを持つため、ユースケース記述を学習データにすると、精度があがると考えられる。そのため、実験データとは異なる 47 個のユースケース記述を学習データとした。また、データ量を増やすため、情報システムの新聞記事³1706 個を加えた。Doc2Vec の次元数は多い方が表現力が上がるが、そのため多くのデータ情報が必要である。学習データとしたユースケース記述と情報システムの新聞記事に含まれた語彙数は一般的に自然言語処理の語彙数より少ないため、ここでは Doc2Vec の次元数を 100 と 300、学習回数を 20 とした。

5 実験

5.1 実験手順

提案手法の有効性を調べるために、5 つの異なるシステムのユースケース記述を実験データとして収集し、5 回実験を行った。変更においては、ユースケース記述中の 1 文をランダムに選び、その文に変更、または削除、追加の操作を加えた。変更と追加においては、変更後、追加後のユースケース記述を入力データとする。削除においては、削除前のユースケース記述を入力データとする。

¹ MeCab: Yet Another Part-of-Speech and Morphological Analyzer. [http://taku910.github.io/mecab/#wakati\(2021.07.10\)](http://taku910.github.io/mecab/#wakati(2021.07.10))

² dmpv300d: <https://yag->

[ai.github.io/project/pretrained_doc2vec_wikipedia/](https://github.com/yag-ai/ai.github.io/project/pretrained_doc2vec_wikipedia/)

³ 情報記事: <https://www.rondhuit.com/download.html#ldcc>

実験データおよび変更操作は表 1 に示す。

表 1 実験データおよび変更操作

システム名	変更操作	ユースケース数	文字数
飛行機座席予約	変更	13	2491
ワイン注文	追加	7	1642
図書館貸出	削除	7	1854
料理店在庫管理	変更	5	1611
新幹線座席予約	追加	8	2078

次に 5 つの実験データを 4 人に渡し、変更により影響を受けるユースケース記述を指摘してもらい、4 人はいずれもユースケース法の知識を持つ人である。より多数の影響箇所を検出するため、4 人が判断したユースケース集合の和集合を正解セットとする。次に、提案手法による類似度結果と正解セットを比較する。図 2 は評価方法を示す。

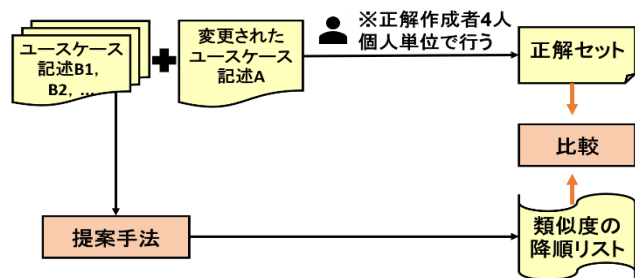


図 2 評価方法

5.2 評価基準

評価基準としては、類似度の降順リストに対するその順位までにおける正解率を調べるための適合率(Precision)と、正解の部分に限定し、それらの適合率の平均を計算し、全体的な正解率を調べるための平均適合率(Average Precision)を算出する。

平均適合率が大きいほど、影響を受けるユースケース記述を上位に検出する能力があることを示す。

5.3 実験結果と結論

平均適合率を求めた結果を表 2 に示す。ただし、モデル 1 は Wikipedia の学習済の Doc2vec モデルを指す。モデル 2 はユースケース記述と新聞記事を学習させた Doc2Vec モデルを指す。

表 2 システムごとの平均適合率

システム名	平均適合率 (Average Precision)		
	モデル 1	次元数 100	次元数 300
		モデル 2	モデル 2
飛行機座席予約	0.57	0.66	0.58
ワイン注文	0.58	1.0	0.83
図書館貸出	0.71	0.50	1.0
料理店在庫管理	0.62	0.76	0.62
新幹線座席予約	0.52	0.62	0.43

表 2 により、ユースケース記述を学習データとさせた方が平均適合率が高いと分かった。その原因はユースケース記述の特有の文書スタイルが影響したという可能性が考えられる。また、Doc2Vec の次元数が 300 より 100 とした方が平均適合率が高いものが多かった。次元数を調

整する時は記述それぞれの文章の特徴、データ量を考慮する必要があると思われる。

さらに 5 つの実験データの平均適合率の平均を取り、MAP (Mean average precision) を計算する。その結果は表 3 に示す。

表 3 MAP の結果

MAP	モデル 1	次元数 100	次元数 300
	モデル 2	モデル 2	モデル 2
	0.60	0.71	0.69

表 3 の評価結果により、最高値は 0.71 であり、提案手法により影響を受けるユースケース記述を検出することについて比較的に良い結果が出たと言える。

6 考察

(1) 平均適合率の考察

評価結果から本提案手法は一定の効果があったが、関係のないユースケースが正解のユースケースの類似度よりも高い値があった。その原因は変更要求と似たような形の単語や文が複数存在していたことが考えられる。また、正解のユースケースの類似度が低い値があった。その原因は同じ意味であった言葉が複数の表現で記されたことが考えられる。

(2) 妥当性の脅威

正解は複数人数で作成したため、正解の属人性が軽減されると考えられる。正解作成者はユースケース法の知識を持っており、他者と作業中にやりとりを行わないことで、正解セットの質による内的妥当性への脅威は軽減されると考えられる。外的妥当性については、実験したユースケース記述以外のユースケース記述でも同じ結論になるかは不明である。より多くの実験を行う必要がある。

7 今後の課題

(1) 変更箇所の考慮

本研究の対象ユースケース記述は構成が明確にされているため、どこで変更したかという変更箇所の情報を加えて類似度計算する手法も考えられる。

(2) 変更履歴の使用

本研究の変更対象は 1 文のみのため、複数の要求が時間をおいて変更される場合はその管理法も考慮する必要がある。変更を分類し、変更履歴を文書化して保存することによって、版管理の支援につながる。

(3) 多くのシステムへの適用

本研究の評価では 5 つのシステムのユースケース記述で実験を行ったため、実システムや種々の分野のシステムに同様の効果が得られるか検証する必要がある。

参考文献

- [1] 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, 情報処理学会研究報告, Vol.2021-SE-208 No.19
- [2] Le, Q. M, T. , "Distributed Representations of Sentences and Documents", CoRR, abs/1405.4053. pp. 1-9, 2014