

# 要求文のベクトル化による要求文間の包含関係検出法の提案

吉成 未菜里<sup>†</sup> 岡本 圭史<sup>†</sup>

仙台高等専門学校 情報システム工学科<sup>†</sup>

## 1 はじめに

ソフトウェア要求仕様書(以下要求仕様書)は、ソフトウェア開発プロジェクトにおいて重要なドキュメントである。要求仕様書は自然言語で記述されることが多いため、自然言語で記述された要求仕様書の問題点を計算機によって検出する手法が研究されている。そうした研究の1つとして、有賀らの研究[1]が挙げられる。有賀らは、要求文の構文と文書構造に基づいて要求仕様書の問題点を検出するチェッカーを考案・実装した。有賀らのチェッカーは、全体的に低い検出精度を持ち、被験者実験により有用性が示唆された。中でも、非あいまい性、追跡可能性に関する問題について高い検出性能を示した。一方で、変更可能性などに関する問題については、検出性能に問題があった。

変更可能性とは、要求仕様書の構造とスタイルを維持したまま、容易に、完全に、矛盾なく要求を変更できる性質である[2]。有賀らは、変更可能性に関する問題を検出できなかった理由として、ある要求文と同一の内容が他の要求文の一部に表れる場合(包含関係)を想定していなかったことを挙げている[1]。ある要求文と同一の内容が他の要求文の一部であるということは、要求が冗長であるということである。冗長な要求が多い場合、1つの要求の変更に伴い、その要求に依存する多数の要求の変更が発生し、変更が必要な箇所の見落としや変更内容のミスが起りやすくなる。要求の冗長さ自体はエラーではないが、エラーに結びつきやすい。

そこで本研究では、要求仕様書内の要求間の包含関係検出法を特に提案し、その有用性を示す。

## 2 提案手法の概要

提案手法の手順は、大きく3つに分けられる。

- ① 要求文の前処理
- ② 要求文のベクトル化
- ③ 類似度の計算

①では、②に向けての準備として、要求文に形態素解析を施し、ストップワードの削除を行う。

Checking inclusion relations between vectorized requirements

<sup>†</sup>Minari Yoshinari, <sup>†</sup>Keishi Okamoto

<sup>†</sup>Department of Information Systems,

National Institute of Technology, Sendai College

本手法におけるストップワードは、接続詞、助詞、助動詞、記号、存在動詞(「ある」「いる」)である。また、見出し語化で単語を基本形に戻し、同義の単語が複数使われていれば統一する。統一する同義語を獲得するためのソーラスとしては、日本語 WordNet[3]を利用した。

②では、要求文のベクトル化を行う。ベクトルは二値ベクトルとし、ベクトル化に用いる素性は単語とする。例として、「(1)授業に出席した学生の出席番号は出席者リストに追加される。」と「(2)講義に出席した学生の氏名を出力する。」という要求文の組のベクトル化を次に示す。

$$(1) \begin{cases} \delta("授業"), \delta("出席"), \delta("学生"), \delta("番号"), \\ \delta("者"), \delta("リスト"), \delta("追加"), \delta("する"), \\ \delta("れる"), \delta("氏名"), \delta("出力") \end{cases}$$

$$= \{1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 0\}$$

$$(2) \begin{cases} \delta("授業"), \delta("出席"), \delta("学生"), \delta("番号"), \\ \delta("者"), \delta("リスト"), \delta("追加"), \delta("する"), \\ \delta("れる"), \delta("氏名"), \delta("出力") \end{cases}$$

$$= \{1, 1, 1, 0, 0, 0, 0, 1, 0, 1, 1\}$$

ただし、 $\delta("word")$ は単語wordが当該要求文に出現するとき1、出現しないとき0としている。また、「授業」と「講義」という同義語は「授業」に統一されている。

③では、②で求めたベクトルの全ての組について類似度を計算する。類似度が閾値を超えていれば、その文の組を包含関係にある組と判断する。代表的な類似度のうち、どの類似度が包含関係の検出に適しているかについて、評価実験の結果から考察を行った。

## 3 実験

### 3-1 採用する類似度関数の決定

最初に、提案手法に採用する類似度関数の候補から最も適切な類似度関数を選定するため、予め文の包含関係が判明しているデータ集合に対して提案手法を施し、その結果を評価する実験を行った。図1に実験方法の概要を示す。

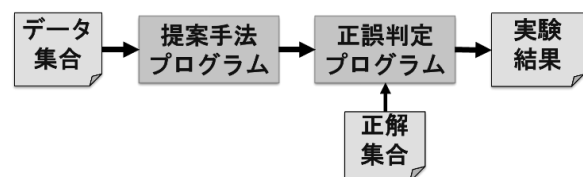


図1 実験方法の概要

初めに、データ集合を入力とし、図1中の提案手法プログラムを実行する。このプログラムは提案手法に沿ってデータ集合に含まれる文をベクトル化し、文の組に対して類似度を計算する。

次に、正誤判定プログラムでは、閾値を変更しながら正しい判定ができていくかを調べる。具体的には、「閾値以上の類似度を持ち、かつ正解である組の数」, 「閾値以上の類似度を持つが、正解ではない組の数」, 「全体の文の組の数」を、正解集合を用いて求め、F値を計算する。 $R_D$ を閾値以上の類似度を持つ組の集合、 $R$ を正解集合とすると、F値は式(3)で表される。

$$\text{適合率} = \frac{|R_D \cap R|}{|R_D|} \quad (1) \quad \text{再現度} = \frac{|R_D \cap R|}{|R|} \quad (2)$$

$$F \text{ 値} = \frac{2 \times \text{適合率} \times \text{再現度}}{\text{適合率} + \text{再現度}} \quad (3)$$

この実験を、Jaccard係数、コサイン類似度、補完類似度、 $\phi$ 相関係数、ダイス相関係数、条件付き確率の6種類の類似度に対して行い、類似度ごとに結果を比較した。その結果、F値が最高になるように各類似度の閾値を設定したときのF値の最高値は表1のようになった。条件付き確率が最も高いF値の最高値を示しており、他の類似度のF値の最高値を0.06以上上回っている。したがって、提案手法に採用する類似度は条件付き確率とした。

表1 類似度ごとのF値の最高値

	F値の最高値
Jaccard係数	0.703947
コサイン類似度	0.709220
補完類似度	0.707903
$\phi$ 相関係数	0.718861
ダイス相関係数	0.703947
条件付き確率	<b>0.786517</b>

### 3-2 人間の判断と提案手法による判定の比較

次に、本手法の有用性を示すため、実際の要求仕様書を基にしたデータ集合を用いて人間による判断と提案手法による判定の比較実験を行った。

データ集合は実際の要求仕様書[4]から同じ機能に関する要求文を抜き出し、一部の複文を単文に分割することによって作成した11文である。また、正解集合は3人の協力者の判断を基に作成した。協力者には、包含する文とされる文が逆になっている組を含めたデータ集合内の文の全ての組110組について、包含関係にあるか否かを判断させた。全110組中、2人以上の協力者が「包含関係にある」と判断した9組を正解集合とした。

このようにして作成されたデータ集合と正解

集合を用いて、図1の方法で実験を行った。ただし類似度は、採用する類似度を決定する実験でF値の最高値が最も高かった条件付き確率である。提案手法によって計算された条件付き確率の値が上位9組以内に入っているかどうかと、その組が人間の判断を基に作成された正解集合に入っているかに注目して実験結果をまとめた表が表2である。なお、条件付き確率の値の基準を上位9位以内としたのは、正解集合の組数が9組であるためである。表2より、全110組中104組、つまり約95%の組において提案手法による判定と人間の判断が一致している。

表2 人間の判断と提案手法による判定の組み合わせ別組数

条件付き 確率 正解集合	上位9組以内	上位9組以外
	入っている	6
入っていない	3	98

(全110組)

### 4 おわりに

要求文をベクトル化し、それらのベクトルの組に対して条件付き確率を計算することで、要求文間の包含関係を検出する手法を提案した。また、実際の要求仕様書と人間による判断を利用した実験により、提案手法の有用性を示した。

しかし提案手法は閾値をユーザが決めることになっており、適切な閾値を決定しなければ本来の性能を發揮できない恐れがある。閾値は機械学習等によって一定の手順で決められることが望ましい。適切な閾値の決定方法を検討することが今後の課題である。

### 参考文献

- [1] 有賀頭, 林晋平, 佐伯元司, “構文と文書構造に基づく要求仕様書の問題点発見支援”, Vol.2013-SE-179 No.4, 2013.
- [2] The Institute of Electrical and Electronics Engineers (IEEE), “IEEE Recommended Practice for Software Requirements Specifications”, IEEE Std 830-1998, 1998.
- [3] Hitoshi ISAHARA, *et al.*, “Development of the Japanese WordNet”, 2008.
- [4] 話題沸騰ポット GOMA-1015 型 要求仕様書, [http://www.sesame.jp/workinggroup/WorkingGroup2/POT\\_Specification.htm](http://www.sesame.jp/workinggroup/WorkingGroup2/POT_Specification.htm), 2016/1/6 参照.