

# 要求仕様書における曖昧語辞書作成法の研究

2020SE011 今村咲彩 2020SE012 石井里彩

指導教員：佐伯元司

## 1 はじめに

### 1.1 研究背景と研究課題

ソフトウェア開発で必要不可欠となる要求仕様書は、通常は自然言語で記述される。そのため、自然言語の持つ曖昧性により要求仕様書には多くの曖昧性が混入されることがあり、開発における障害の要因となる。この問題を解決し、要求仕様書の品質の向上を行うためにも、要求仕様書の曖昧さを解消することが重要になる。

そのためには、要求仕様書を書き上げた段階で直ちに曖昧な箇所を検出する必要がある。仕様書が膨大になると、人間で曖昧な箇所の検出を行うには多大な労力を要するだけでなく、検出誤りも起こる。本研究では要求仕様書中の曖昧な箇所の検出の自動化を目指す。要求仕様書における曖昧性は、D.Berry らが指摘しているように種々の原因によって生じる [1]。本研究では、仕様書記述者を曖昧語を使用しないように支援するために、曖昧語辞書を作成することを目的とする。

本研究では、高橋らの研究 [2] から曖昧さを引き起こすと考えられる単語に着目した。彼らの研究では、システムの振舞いを表す単語の同義語となっているものを収集し、同義語辞書を作成した。1つの要求仕様書中でシステムのある振舞いを表すために、同義語を多く使って記述すると、それらの同義語の意味が本当に同じ振舞いを表すものか、異なる意味で使われているのかが不明になり、曖昧性が増大するとしている。この手法はまず、「曖昧さを引き起こすと考えられる単語」と「その同義語」をあらかじめ特定しておく必要があり、同義語を使った判定は活用できる可能性があるが、辞書作成という本研究の目的とは異なる。

本手法での辞書作成は、まず意味がもともと曖昧な単語を収集し、辞書候補の単語リストを作成し、辞書の中に入っている語と同義な語を収集する。曖昧である単語と同義であるとみなせる語の判定は、単語のベクトル化手法と cosine 類似度を用いて行う。それらの単語を辞書に追加することによって、より広い範囲をカバーできる辞書を作成する。作成した曖昧語辞書を他の要求仕様書にも展開することにより、曖昧性を含む箇所を指摘する手法を提案する。

## 2 準備

本研究で使用する2つの技術である形態素解析技術、単語のベクトル化技術について説明する。

### 2.1 MeCab

MeCab とは、日本語の形態素解析ツールであり、AIなどが自然言語を処理するために用いるツールである。また言語や辞書、データベース化された言語資料であるコーパ

スに依存しない汎用的な設計が MeCab の特徴である。本研究では、要求仕様書におけるそれぞれの曖昧語の出現回数を調べる際に、MeCab と Python から MeCab を呼び出すためのライブラリ MeCab-Python を使用する。

### 2.2 Word2Vec

Word2Vec とは、単語を数値ベクトルに変換し、その意味を数値的に把握する自然言語処理の手法である。Word2Vec の最大の特徴は、単語間の関係性をベクトル演算により表現できることである。本研究では、曖昧語辞書作成の際、単語の類似度計算を行い辞書の拡充を行うために、Word2Vec を使用する。

## 3 アプローチ

要求仕様書における曖昧性は、仕様書中の曖昧な意味を持つ単語が1つの要因である。本研究では、単語の意味に着目し曖昧語を抽出する。

曖昧語の抽出は人手で行う。そのため、判断がどうしても主観的になってしまう。そこで抽出された曖昧語が、本当に曖昧であるかどうかを、抽出者とは別の被験者にレビューしてもらう。レビュー結果でも曖昧と判定された単語について信頼度を算出する。求めた信頼度により最終的に曖昧性を判定後、曖昧語を定める。曖昧と判定された曖昧語と、その他の判定の単語の意味的類似度を求め類似度の高い単語も曖昧語とみなして辞書を拡張する。拡張した単語が本当に仕様書中で曖昧となっているかどうかを複数人で判定し、その信頼度と偏差を求めることによって評価を行う。

## 4 提案手法

### 4.1 曖昧語の抽出

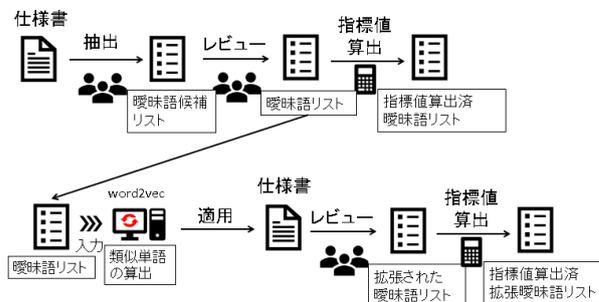


図1 提案実験概略

本研究では、n人の被験者、m個の要求仕様書を使用し曖昧語の抽出作業を行う。本研究での意味的に曖昧な単語の抽出には、何を基準に単語を曖昧とみなすかの基準が必

要である。よって曖昧語の抽出作業において、以下の3つを判定基準とした。

1. 単語によって解釈が複数存在する  
例) 効率 (得られる成果の割合, 仕事量と消費されたエネルギーとの比率), 情報 (伝えられる内容, 詳細)
2. 基準がはっきりしないものが含まれている  
例) 十分な, ほぼ
3. 代名詞の中でも何を指すか分かりにくい  
例) それ, あれ

これらの判定基準を基に, 曖昧語抽出フォームに記入してもらう。

#### 4.2 レビュー

曖昧語抽出は人手で行うため, どうしても主観的になってしまう。そのため, 抽出者とは別の被験者に単語リストを配布し, レビューを依頼する。被験者としての人的資源に制約があるため, 抽出者とレビューワの役割を交代で同じ被験者にやってもらうこともあり得るが, レビューワには自分が抽出者になったのと同じ仕様書は渡さないようにする以外, 割り当てをランダムに行う。これは, 同じ要求仕様書で抽出, レビューをやる際のバイアスがかかることを回避する目的がある。そして抽出された曖昧語が曖昧か, 前述の判定基準を基に判定してもらう。

#### 4.3 指標値算出

作成した単語リストにおける評価指標値として, 信頼度, 偏差を使用する。曖昧語抽出実験では信頼度のみ, 曖昧語拡張実験では信頼度と偏差を算出する。

##### 4.3.1 信頼度算出

信頼度とは, ある単語が文書中で曖昧だと判定したときに, その判定が信頼できるかどうかの指針である。同じ単語であっても文中での使われ方によって曖昧であったり曖昧でなかったりする。その単語だけで曖昧であると判定した場合の「正しさ」を何らかの形で数値化する必要がある。信頼度は文中に出現しているある単語が曖昧であると判定された回数をその単語の出現回数で割ったものである。単語  $w$  に対し, 以下の式とする

$$\text{信頼度}(w) = \frac{w \text{ の出現で曖昧と判断されたものの個数}}{w \text{ の出現個数}}$$

本研究では, 信頼度の算出により, 単語の曖昧度を測る。

##### 4.3.2 偏差算出

本研究では, 偏差は曖昧語辞書作成における単語データの特徴を調査するために算出する値である。被験者によって単語を曖昧と判断する基準が異なるため, 信頼度に差が生じる場合がある。そのため, 得られた単語の信頼度, と被験者がそれぞれ算出した信頼度との偏差を求める。表1に例を示す。被験者が A, B, C の3人が単語  $w$  をそれぞれの仕様書で曖昧回数, 信頼度を求める。

表1 偏差算出における例データ

被験者	出現回数	曖昧判定回数	被験者ごとの信頼度 (%)
A	3	3	100
B	4	2	50
C	8	2	25
合計	15	7	46

$w$  の信頼度の偏差は, 信頼度 ( $w$ ) と表1の4列目の値を使って,

$$\text{偏差}(w) = \sqrt{\frac{\sum_{i=1}^n (\text{信頼度}(i, w) - \text{信頼度}(w))^2}{n}}$$

信頼度 ( $i, w$ ) は被験者  $i$  の単語  $w$  の信頼度。

で求めることができ, 偏差は 33.5 となり, 単語  $w$  は比較的偏差が小さく, 人により曖昧と思う基準が同じになりやすい単語と言える。偏差を算出することで, 偏差の高い語について, 類似度を用いて曖昧語辞書の拡張を図る。

#### 4.4 類似度算出

曖昧と判断された単語と他の単語間の類似度を算出する。この類似度の算出には, Word2Vec を使用する。Word2Vec を使用することで, 単語間の類似度を求め意味が近い単語同士で集めることで, 辞書の拡充を図る。

### 5 曖昧語抽出実験

#### 5.1 曖昧語抽出実験手順

本実験には5人の被験者と25の要求仕様書を使用する。手順としては, 以下の図2に示す1人の被験者あたり5つの仕様書を配布し事前に作成した曖昧語抽出フォームと, 曖昧語抽出マニュアルに則って曖昧語の抽出を委託する。

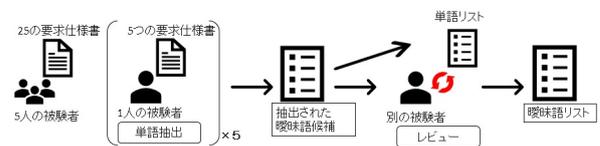


図2 曖昧語抽出実験手順の概要

1人の被験者が5つの要求仕様書から抽出した曖昧語を別の被験者に配布し, 曖昧と判断されるかどうかのレビューを委託した。

#### 5.2 実験結果

25の要求仕様書から重複した単語を除き計108単語が抽出された。

## 5.2.1 問題点

被験者によって単語の抽出数に偏りがみられた。このことからデータを分析する上での以下の問題が発生した。

1. 仕様書ごとに単語の信頼度を算出しようとしたところ、被験者により抽出された単語が1つの仕様書につき約5~7単語と著しく少ないため、有意なデータとならない可能性がある。
2. 25の仕様書を総括し、信頼度の算出を行うと単語数が108単語と増え、データにはなりそうだが、信頼度の不正確さが問題となった。

例として被験者Aは単語wを曖昧語として抽出、被験者Bは単語wを曖昧語として抽出していない場合、両者の仕様書で単語wが出現していても被験者Bでは出現回数は抽出されないために0となる。

このように被験者による曖昧さの判断の偏りにより抽出が決まるため、信頼度の正確さが不十分である。

## 5.2.2 知見

今回信頼度に不正確さが生じたのは、被験者による曖昧さの判断の偏りにより、曖昧語の抽出に差ができたことである。また、偏りがみられたことから信頼度のみだけでなく被験者それぞれの偏りのデータを数値に表すことで、新たなデータを分析する指標とした。また、単語に着目していたことから曖昧と思われる単語のレビュー時に章題に出現した単語がくることもあり、単語としては曖昧だが章題として見た場合には曖昧ではないと捉えられる場合の判断方法に関して難しいと被験者から伝えられた。

# 6 曖昧語拡張実験

## 6.1 曖昧語拡張実験手順

曖昧語抽出実験を踏まえ、被験者それぞれの判断で単語を抽出するのではなく、あらかじめ曖昧語候補リストを作成し被験者全員に同じ単語リストを配布し、被験者に曖昧であるかどうかを判断してもらい、曖昧と判断した回数を収集した。また、被験者により曖昧さの判断の偏りが生じるため、単語が被験者によって判断の偏りがあるか偏差を算出するという手法をとった。

本実験で得られた108単語をWord2Vecでベクトル化し、類似度算出によって類似した意味の単語の類似度を算出し、類似した意味の単語を抽出することで単語の拡張を狙う。手順としては、108単語から類似する4000単語をWord2Vecにより抽出した。得られた4000単語から、436単語にまで以下で述べる独自のフィルタリングにかけて抜出する。その後、新規25の要求仕様書と単語リストを被験者に配布し、レビューを実施する。そしてレビューから信頼度、偏差を算出する。図3に手順を示す。

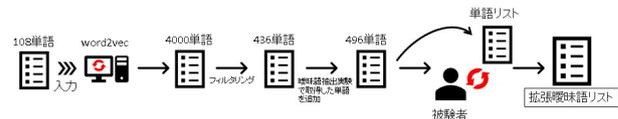


図3 曖昧語拡張実験手順の概要

## 6.2 フィルタリング手順

フィルタリングの方法として、第1に108語から抽出された4000語の類似度の平均値を算出する。それぞれの単語の類似度の上位10位としなかった理由としては、単語によって類似度が大きく異なり数値の変動が激しく正確に曖昧さを持つ単語のみを抽出する際にノイズとなる可能性があったからである。平均値0.5878を閾値として設定する。そして、フィルタリングされた結果の単語数が1602単語である。

第2に、新しく用意した25の要求仕様書の中に第1にフィルタリングされた結果の1602単語の25の仕様書での出現数調査を行いフィルタリングをかけた結果463単語にまで絞られた。

第3に、著者らが曖昧語抽出実験で単語リストの中で明らかに曖昧でない判断した単語が27単語あり、それらを除くと436単語にまで絞られる。明らかに曖昧でない判断される単語の代表格として、数字では「4週間前」といった単語が挙げられる。曖昧語拡張実験では、これらのフィルタリングで絞られた436単語と、曖昧語抽出実験で得られた108単語のうち25の仕様書内で出現する86単語を加え重複を除いた計491単語を使用する。フィルタリングの手順を図4に示す。

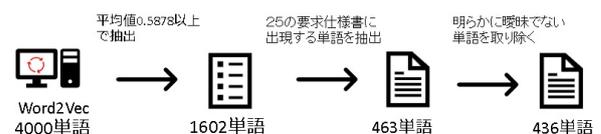


図4 フィルタリング手順

## 6.3 実験依頼

曖昧語拡張実験では、曖昧語候補リストに格納され、フィルタリング済みの491単語と新規25の仕様書を使用する。被験者に1人5つの仕様書を配布し、配布した曖昧語候補リストの曖昧な回数を記入してもらう。また、被験者には1人につき1ファイル(5つの仕様書)を配布した。それぞれのファイルで491単語のうち、どの単語が何回出現したかについてはあらかじめ調べ、それらを記入した抽出フォームを配布した。

## 6.4 実験結果

### 6.4.1 データ作成

実験の結果、491 単語中信頼度 0 の単語が 278 単語あり、曖昧語とみなす単語は 218 単語となった。得られた 218 単語を類似度で抽出された単語別にリスト化し、それぞれの信頼度、偏差をまとめた拡張済み曖昧語リストを作成した。曖昧語リストの一部を下記の図 5 に示す。図 4 の表の 2 列目が曖昧語である。

抽出単語	曖昧語	類似度	出現回数	曖昧と判断した回数	信頼度	偏差
以下	上記	0.75853	40	18	45	39.24283374
	次	0.75701	187	46	24.59893048	15.35960756
連応	依存	0.684746	4	2	50	38.7298346
	配置	0.675534	16	9	58.23	32.9862119
	列挙	0.663953	14	8	57.1439274	41.12918544
	列挙	0.534425	302	20	6.22518858	6.72122853
	変化	0.613735	7	1	14.33571429	13.964846
	変化	0.630304	1	1	100	0
	適用	0.599948	24	3	12.3	26.0998028
	影響	0.595463	22	7	31.81818182	24.99872459
	算出	0.592598	1	1	100	0
	算出	0.737012	3	2	100	0
	算出	0.735888	1	1	100	0
	算出	0.736653	1	1	100	0
	算出	0.730625	32	18	58.23	34.26783055
	算出	0.720684	4	4	100	0
	算出	0.709988	1	1	100	0
	算出	0.699165	4	4	100	0
	算出	0.69791	2	1	50	31.6227766
	算出	0.693165	1	1	100	0
	算出	0.678292	5	4	80	37.94733192
	算出	0.654359	9	5	56.3593959	25.0385052
	算出	0.599883	46	1	0.17311924	2.952781323
	算出	0.717409	6	6	100	0
	算出	0.702343	35	5	13.53551553	11.46141992
	算出	0.626846	1	1	100	0
	算出	0.594153	7	2	28.57142857	34.40455594
	算出	0.63104	47	4	8.51043676	41.44296389
	算出	0.601287	83	4	0.24020644	7.189823473
	算出	0.749454	29	21	72.4131413	37.21335152
	算出	0.62586	2	2	100	31.6227766
	算出	0.673765	119	80	67.21688772	32.51084582
	算出	0.593111	56	11	19.54285714	43.45156555

図 5 曖昧語リストの一部

## 6.5 データ分析

### 6.5.1 分類

出現回数が明らかに少ない場合、分母が小さくなるため、今回の実験で使用した 25 の仕様書のみでは信頼度に信用性があるとは言えない。図 5 の中では、今後他の仕様書でも調べてみる必要性のある分母の少ない単語の出現回数は水色でマークされている。また、信頼度が 23.28595318 % より小さい場合は緑、23.28595318 % ~ 50 % の場合は黄色、50 % より大きい場合は赤色で色分けをした。単語が被験者によってどれほど判断の偏りがあるかで、偏りの大きさによって色分けした。偏差が 50 % を超える場合ピンクでマークされている。

表 2 信頼度と偏差の分布

項目	信頼度 (%)	偏差 (%)
最小値	0.37037037	0
第一四分位数	23.2859...	0
第三四分位数	100	38.57162483
最大値	100	63.2455532

上記の表 2 より、信頼度は比較的高く、偏差は 50 より低いものが大半を占めることが分かった。よって信頼度の偏りは比較的小さく、実験 2 で得られた曖昧語は曖昧語としての正確性が高いと言える。

## 7 まとめと今後の課題

### 7.1 まとめ

本研究では、合計 50 の仕様書を使用し、本実験で得られた 108 単語の拡張実験により最終的に 218 単語を曖昧語として検出することが出来た。Word2Vec と類似度で新たに加えた 436 単語と 218 単語の共通部分は 175 単語あったため、この拡張手法の拡張部分の適合率 0.40(175/436)、拡張辞書全体の適合率 0.44(218/491) と算出された。適合率の低さの原因として Word2Vec は単語単体での意味ベクトルを算出しており、文中の使われ方によって曖昧さを初めてもつ曖昧語も存在していることから被験者はそれまで考慮してレビューを行ったためであると考えられる。そして、機械学習における単語拡張性が証明されたと思われる。今後は、218 単語の曖昧語を今後仕様書の曖昧箇所検出ツールの開発に発展させたい。

### 7.2 今後の実験課題

#### 7.2.1 辞書作成

辞書作成の目的は、曖昧語の箇所を目視で行い可視化させ要求分析者のために曖昧と思われる単語を使用している箇所の危険性を促すためのものである。そのためにこのシステムでは、算出された数値を信頼度ごとに分類して格納した単語リストファイルを 3 つ用意し、辞書を作成した。曖昧語の信頼度の低い順から赤色、黄色、緑色とハイライト化させ見やすくした。今回は python で作成したが今後は誰でも扱えるようにウェブアプリケーションでの作成を目指す。

#### 7.2.2 辞書拡張

本研究では合計 50 の要求仕様書から 218 単語の曖昧単語を抽出できたが、辞書の拡大性を考えて間違っている単語をハイライトだけでなく修正すべき補助単語までの抽出を行いたい。また、本研究の 50 の仕様書内では、出現数が少なく、信頼度の正確性が低いとみなした単語を新たな仕様書で調べ直したい。また、辞書の拡張性にはまだまだ辞書には多くの単語を格納したいので類似度からまた作成された単語リストのレビューを続けていくことでデータ数を集め今後の要求仕様書における曖昧さを軽減することが可能となる辞書ツールを作成したい。

### 参考文献

- [1] Daniel M. Berry, Erik Kamsties, and Michael Krieger. From contract drafting to software specification: Linguistic sources of ambiguity. 2003.
- [2] 高橋宏季, 井上昇, 伴凌太, 位野木万里ほか. 要求仕様をあいまいにする同義語の特性分析と同義語辞書の自動作成手法の提案. 研究報告ソフトウェア工学 (SE), Vol. 2019, No. 1, pp. 1-7, 2019.