

要求仕様書における設計要素の表記ゆれと同義語利用に着目した あいまい表現の記述パターンの分析と自動指摘手法の提案

高橋 宏季† 中島 千壽† 近藤 公久† 位野木 万里†

工学院大学†

1. はじめに

要求仕様書はシステムの開発範囲を定義する重要な成果物である[1]。要求仕様書は自然言語で記述されることが多く、あいまいな記述があれば、内容確認の手戻りの発生や、発注者の要求とは異なる解釈で開発されるリスクがある。要求仕様書中のあいまいな記述は排除することが望ましいが、どのような表現が要求仕様書をあいまいにするのかが不明確であり、複数のあいまい表現を漏れなく効率的に指摘することは困難である。著者らは、実案件の要求仕様書を対象に、要求仕様書をあいまいにする表現を分析した[2][3][4]。特定の設計要素用語に対して文字列の修飾や省略を繰り返し派生形による表記ゆれ用語や、同じ意味で異なる表現である異音同義語の用語（以下、同義語とする）が、あいまいさを引き起こす要因となることを明らかにした。著者らは、これらの表記ゆれと同義語表現の記述パターンを形式知化し、同知識を組み込んだ指摘手法によって、要求仕様書中のあいまい表現の網羅的な自動指摘を実現した。本稿では同手法の実案件への適用評価結果を報告する。

以下、2章では、あいまい表現の特性を分析し、あいまい表現を網羅的かつ効率的に指摘することへの課題を整理する。3章ならびに4章では、2章で整理した課題を解決するために、それぞれ、派生形用語指摘手法、word2vecを用いた同義語辞書自動生成手法を提案する。5章では、提案手法の実案件への適用評価を示す。6章では提案手法について、有効性、妥当性の観点から考察し、7章にて本稿のまとめを示す。

2. 要求仕様書中のあいまい表現

2.1 設計要素のあいまい表現の分類

自然言語で記述された文書のあいまい表現の分類、要因、防止策等について多くの研究がある[5][6]。要求仕様書中の設計要素に関するあいまい表現の分類を図1に示す。要求仕様書には利用者定義のような形で設計要素が定義され、定義された設計要素によるシナリオが記述される。設計要素のあいまい表現は、定義からの表記ゆれか、同義語の関係であるか、によって4種に分類できる。

例えば、管理社員が利用者定義として定義され、社員、スタッフ、マネージャ、一般社員がシナリオ中に記述されているとする。利用者とユーザーは未定義の同義語のため、図1(A)の一般的な表記ゆれである。マネージャは定義済みの管理社員の同義語のため、定義からの表記ゆれの図1(B)である。管理社員と社員は、定義されている管理社員を省略した形のため、図1(C)に該当する。社員と一般社員は両方定義と無関係であり、社員に一般を修飾した形のため派生形の図1(D)に当てはまる。

2.2 派生形アクターと振る舞い用語の同義語

著者らの要求仕様書の分析では、ある大規模システムの要求仕様書において、シナリオ中に出現しているアクター正味 91 件中 89 件は、設計要素として未定義のアクターであった。また、未定義アクターの内、73 件は、5 つの基準アクターを元に派生した、派生形アクターであった。このように、アクター名は、役職名や組織名に対し修飾または省略を繰り返して記述さ

Analysis of Patterns of Ambiguous Descriptions in Requirements Specification and Proposal of a Method for Automatic Extraction Focusing on Spelling Inconsistencies and Analogous Terms in Design Elements

†Hiroki Takahashi, Senju Nakajima, Tadahisa Kondou, Mari Inoki, Kogakuin University

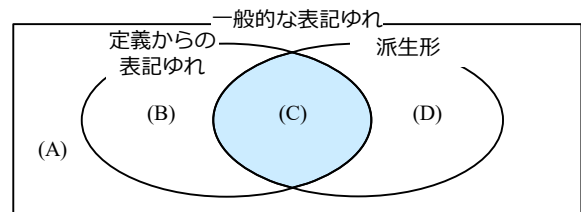


図1 設計要素のあいまい表現の分類

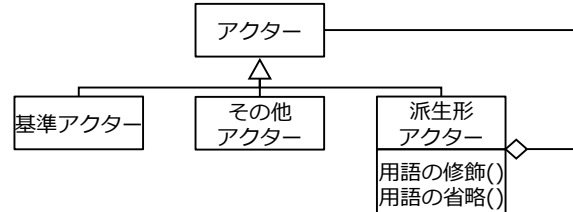


図2 派生形アクターの関係図

れると考えることができる。そこで、このようなアクター名を派生形アクターと呼ぶことにする。基準となる用語を、基準アクターと呼ぶことにする。派生形アクターの関係を図2に示す。

また、著者らは分析する過程で、複数の要求仕様書に渡り、振る舞い用語の同義語が記述されている状況に直面した。振る舞い用語の同義語が用いられた要求は、表現は異なるが同一の内容なのか、異なる内容を示すのかが不明確であり、要求仕様書の品質低下につながる。

2.3 あいまい表現を指摘する際の課題

要求仕様書中のあいまい表現の指摘作業には、高い作業コストとノウハウが必要であるという課題がある。多量の要求仕様書中に散在している膨大な件数の設計要素を、手作業で指摘するためには多大な作業コストを要する。また、同義語を自動指摘するためには事前に同義語のリストを作成する必要がある。同義語のリストの構築には、設計要素指摘の作業コストに加え、要求仕様書特有の同義語があるため、初級技術者には実施が困難と考えられる。

著者らの研究グループでは、自然言語で記述された要求仕様書の高品質化のため、要求仕様の一貫性検証支援ツール（以下、一貫性検証支援ツールとする）を開発してきた[7]。本ツールは、要求仕様書の品質特性である「一貫性」に着目し、「アクター」「データ」「画面」「振る舞い」の設計要素が、要求仕様書で一貫した定義で記述されているかを支援する。著者らは、要求仕様の一貫性検証支援ツールを拡張し、分析した表記ゆれの記述パターンを反映させた、表記ゆれの自動指摘手法を提案する。また、単語のベクトル表現化ツールである word2vec を用いて、同義語を自動指摘する手法を提案する。

3. 派生形用語指摘手法

基準アクターに基づき派生形アクターを指摘する手法を、一貫性検証支援ツールを拡張し開発した。拡張後の一貫性検証支援ツールの入出力の流れを図3に示す。本手法は、図3(B)のF30派生形用語抽出である。

本手法は自然言語で記述された要求仕様書と基準アクターを入力とする。一貫性検証支援ツールは、要求仕様書中からアクター名をアクター識別辞書によって指摘する。派生形用語指摘

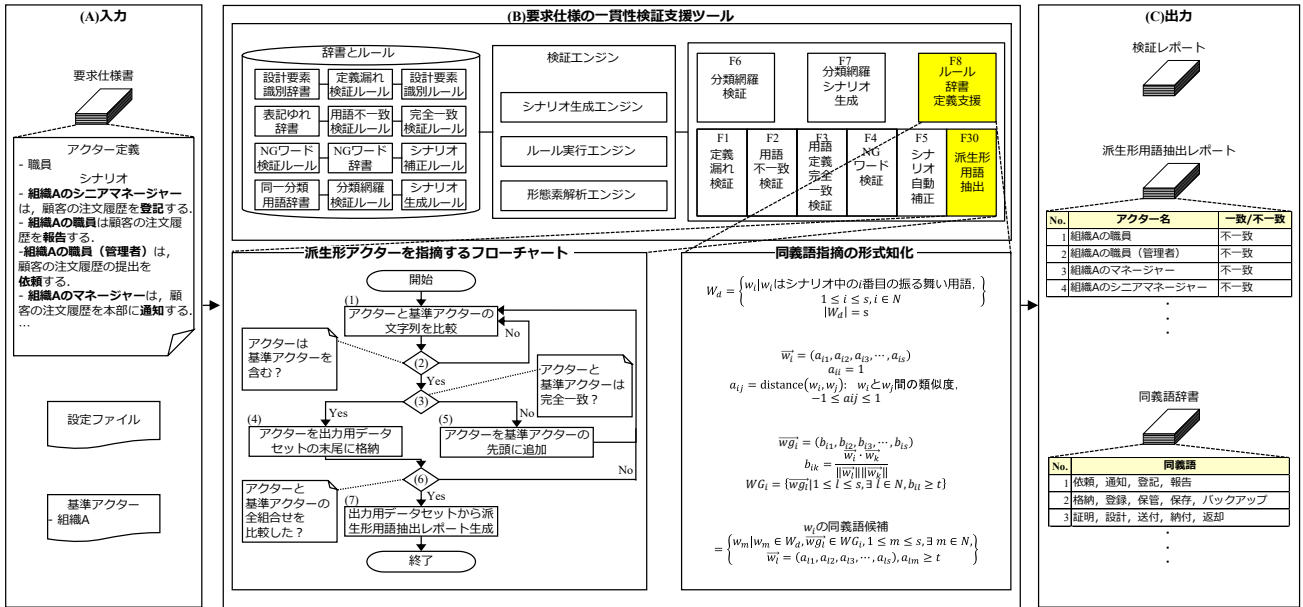


図3 提案手法拡張後の要求仕様の一貫性検証支援ツールの入出力の流れ

手法は、入力した基準アクターの派生を再帰的な繰り返しによって辿ることで、派生形アクターを指摘する。

4. word2vec を用いた同義語自動指摘手法

単語のベクトル表現化ツールである word2vec を用いて、要求仕様書内の同義語指摘作業の負荷を軽減する手法を提案する。本手法は、図 3(B) の F8 ルール辞書定義支援に相当する。本手法は自然言語で記述された要求仕様書と、同要求仕様書から一貫性検証支援ツールが指摘した振る舞い用語を入力とする。

W_d を振る舞い用語の集合、振る舞い用語数は s 個とする。 \bar{w}_i をある用語 w_i と W_d 内の全ての用語との類似度による s 次元ベクトルとみなす。同 s 次元ベクトル間の類似度を算出し、それを w_i の用語グループベクトル \bar{w}_{jk} と定義する。 WG_i を用語グループベクトルの集合とする。類似度の閾値を t と設定し、類似度が t 以上の要素を持つ用語グループベクトル \bar{w}_{jk} を構成する類似用語（ここでも閾値 t 以上）を、用語 w_i の同義語候補とする。

5. 適用評価

本章では提案手法を、設計要素毎の観点から評価する。アクターの観点からは、派生形用語指摘手法による指摘作業の効率化および、派生形アクターの要求仕様書の修正作業への効果を評価する。実案件の要求仕様書を対象に本手法により生成されたリストの結果と、人手によるリスト作成結果を比較する。作業コストは人手で 100 人時、ツールで 1 人時であった。リストの比較結果は、再現率、適合率ともに 90% 以上の精度であった。

振る舞いの観点からは、word2vec を用いた同義語自動指摘手法が指摘する同義語の、妥当性を評価する。実案件の要求仕様書から人手と提案手法によって同義語を指摘し、両者の同義語を比較する。本評価では、閾値を 0.4 に設定し、手法を適用する。本手法により作成した同義語と、手作業により特定した同義語と比較したところ、完全一致はしなかった。本手法により作成した同義語では、人手による同義語特定では見逃していた同義語を指摘した。

6. 考察

本章では提案手法を、設計要素毎の観点から考察する。アクターの観点からは、提案手法によって、90% を超える精度かつ、手作業の 1% の作業コストで、派生形アクターを自動指摘可能となった。本研究では情報システムを対象に、職員等を基準アクターとした。組込みシステムや等に対しては、操作仕様中の表記ゆれを指摘するために、センサーやデバイス等の要素を、基準アクターとすることが有効と考えられる。

振る舞いの観点からは、提案手法により、手作業による同義

語指摘の結果を含む同義語と、手作業で見落とし同義語の指摘も可能となった。本手法は、初級技術にもノウハウが必要な同義語を低コストで指摘可能となる点から、有効と考えられる。

本研究では、あいまい表現の記述パターンを明らかにしたが、基準アクターの設定のように、人手の分析が必要な部分がある。今後は、word2vec を用いた同義語自動指摘手法をベースに、自動学習による記述パターンの分析や指摘を、他分野や他の設計要素への適用に取り組む。

7. まとめ

本稿では設計に大きな影響を与える設計要素のあいまい表現である派生形アクターや同義語を明らかにし、自動指摘する手法を提案した。本手法は、要求仕様書のあいまい表現の指摘作業の効率化や負荷軽減に一定の成果を得た。今後は本研究の自動学習の成果の、他分野や他の設計要素への適用に取り組む。

謝辞

要求仕様の一貫性検証支援ツール開発に関わる研究は、独立行政法人情報処理推進機構技術本部ソフトウェア高信頼化センター (SEC: Software Reliability Enhancement Center) が実施した「2015 年度ソフトウェア工学分野の先導的研究支援事業」の支援を受けたものである。また、本研究は JSPS 科研費 JP16K00105, JP19K11907 の助成を受けた。

参考文献

- [1] 一般社団法人情報サービス産業協会 REBOK 企画 WG 編, 要求工学実践ガイド REBOK シリーズ 2, 近代科学社, 2014.
- [2] 高橋宏季, 野村典文, 近藤久久, 位野木万里, 要求仕様書における派生形アクター自動抽出手法: 組織変更による影響対応への効果, 情報処理学会, ソフトウェアエンジニアリングシンポジウム 2018 論文集, pp. 121-129, 2018.
- [3] 高橋宏季, 野村典文, 近藤久久, 位野木万里, 要求仕様書中のアクター名の定義漏れパターンと組織変更がもたらす影響対応-実案件分析と得られた教訓-, 情報処理学会, デジタルプラクティス, (採録決定 2019-10-23).
- [4] 高橋宏季, 井上昇, 伴凌太, 位野木万里, 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, 情報処理学会, 研究報告ソフトウェア工学 (SE), 2019-SE-201(1), pp. 1-7, 2019.
- [5] 阿部圭一, 情報伝達型の日本語文章に現れるあいまい表現の類型化とその改善例, 情報処理学会, 情報処理学会デジタルプラクティス Vol. 5 No. 1, pp. 70-79, 2014.
- [6] Kamsties, E, Paech, B, Taming ambiguity in natural language requirements, In Proceedings of the Thirteenth International Conference on Software and Systems Engineering and Applications, December, pp. 1-8, 2000.
- [7] 位野木万里, 近藤久久, 省略と修飾パターンを用いた用語不一致検証による要求仕様の一貫性検証支援ツールの実現と適用評価, 日本ソフトウェア科学会, コンピュータソフトウェア, Vol. 35, No. 3, pp. 109-127, 2018.