

自然言語で記述された要求仕様書からの 状態遷移抽出法の提案と評価

宮崎 大河[†] 佐伯元司[‡]

南山大学大学院理工学研究科^{†‡}

1. はじめに

要求仕様書は、システム開発工程の初期プロセスである要求定義プロセスにおいて作成され、設計プロセスや開発プロセスなどその後の開発の基準となる重要な成果物である。しかし、要求仕様書は主に自然言語で記述されているが故に、主語や機能要求の欠落・文同士や単語間の係り受けが複雑であるなどの曖昧性等を有している。これら低品質箇所を早い段階で検出することができなければ、結果として前工程への後戻りによる開発遅延やコストの増加、顧客の要求を満たさない質の悪いソフトウェアが開発されてしまう[1]。

このような背景からシステム開発の初期段階で要求仕様書の低品質箇所を特定することは、高品質なシステム開発を行うために極めて重要であると言える。本研究では、体系的な曖昧性判定の実施かつ、幅広いステークホルダーが共通の認識が得られるという観点から自然言語で記述された要求仕様書から状態遷移を抽出し、形式手法によって曖昧性の判定を実施する手法[2]に着目する。通常、状態遷移はシステム動作を記述するものであることから要求仕様文中の動詞に着目し、状態となる名詞を抽出していく。しかし、単に名詞と状態を対応づけていくだけでは、代名詞等の本来状態となり得ない名詞まで状態となってしまうため、本研究では、動詞と共起している名詞の意味的役割を解析するために格フレームを使用する。なお、状態遷移の抽出という目的から表層格の二格と深層格の対象格を、表層格のカラ格を深層格の源泉格に対応づけて研究を進める。

2. 要求仕様書からの状態遷移抽出法

2.1. 対象文抽出プロセス

要求仕様書から状態遷移を抽出するために

A Technique for Extracting State Transitions from Requirements Specifications Written in Natural Language

[†] 「Taiga Miyazaki · Nanzan University」

[‡] 「Motoshi Saeki · Nanzan University」

はじめに要求仕様書から状態遷移を抽出するための対象文を抽出・加工を行う。具体的には、次の3つの特徴を持つ文を抽出する。
(ア) 順接の接続詞を文中に保有
(イ) {条件部} + {動作部} でシステム動作を記述
(ウ) 抽出できる状態遷移が文中に1つ

これら3つの特徴を持つ文を対象文として抽出し、接続詞を含む文の前半部分を条件部、それ以降を動作部とし、原文とともに格納する。

2.2. 名詞リスト作成プロセス

次のプロセスとして、状態遷移抽出対象文を対象とした名詞ペアを作成していくプロセスについて図1に示す例文を用いてまとめる。

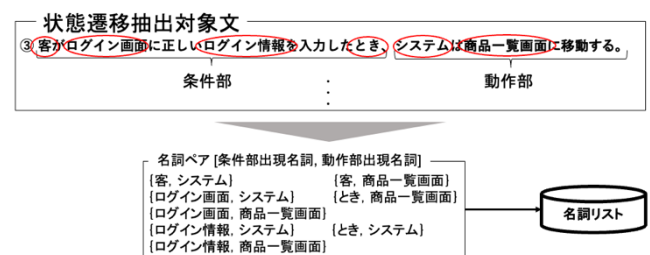


図1 名詞リスト作成例

まず、赤い楕円で囲んでいる対象文中の全名詞を抽出する。その後、条件部出現名詞と動作部出現名詞の全ての名詞の組み合わせを作成し、名詞リストに格納する。

2.3. 状態リスト作成プロセス

次のプロセスとして、状態遷移抽出対象文中の遷移を表す動詞に係る表層格：二格の抽出を行う。図2で例文を用いて示す。

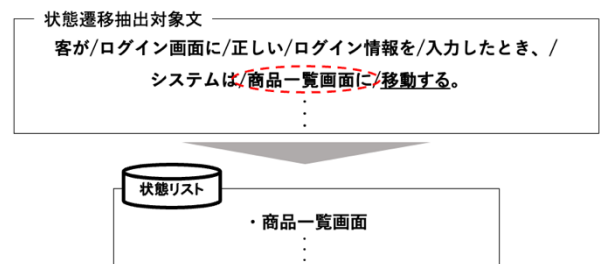


図2 状態リスト作成例

まず、対象文から特定の遷移動詞「移動する」にかかる文節を抽出する。その後、当該文節の中から【名詞】+格助詞「に」で構成されている文節の名詞のみを名詞リストに格納する。図2では、「商品一覧画面」という名詞が格納される。

2.4. 状態遷移抽出プロセス

本プロセスでは、2.1 節~2.3 節で得たデータを元に状態遷移の抽出を行う。抽出の際、遷移動詞に関する深層格の対象格と遷移後状態を、深層格の源泉格と遷移前状態を対応づける。これを踏まえて2つのケースに分けて抽出を行っていく。

2.5. 文の動作部に源泉格が存在する場合

1つ目のパターンとして、対象文の動作部に源泉格が存在する場合である。具体的には、文の動作部に【名詞】+カラ格となる文節が存在している場合である。この場合、カラ格の名詞を遷移前の状態、当該対象文の状態リストに格のされている二格の名詞を遷移後状態として、状態遷移を抽出する。

2.6. 文の動作部に源泉格が存在しない場合

抽出の2つ目のパターンとして、対象文の動作部に源泉格が存在しない場合である。この場合の状態遷移抽出プロセスを以下、図3に示す(ア)文の状態遷移抽出を例としてまとめる。

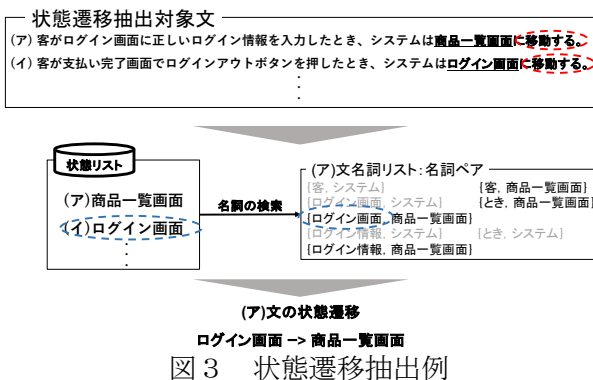


図3 状態遷移抽出例

まず初めに、(ア)文から抽出され、状態リストに格納されている遷移動詞に係る対象格「商品一覧画面」を持つ名詞ペアを状態遷移の抽出対象となる(ア)文名詞リストの名詞ペアから抽出する。その後、抽出された当該名詞ペアの中からもう一方の名詞も状態リストに含まれる名詞ペアを検索する。今回の例では、「ログイン画面」が(イ)文から抽出されているため、{ログイン画面, 商品一覧画面}という名詞ペアが該当する。以上から(ア)文における状態遷移は、「ログイン画面 → 商品一覧画面」と決定する。

3. 評価実験

3.1. 実験の目的と手順

第2章で示したプロセスを Python, MeCab, CaboCha 等を用いてプロトタイプシステムに実装した。これを用いて次の評価項目に答えるために実験を行った。具体的な評価項目は、提案手法は、1)状態を正しく・全て抽出できているか、2)状態遷移を正しく・全て抽出できているか、3)人間の作業時間を短縮できるかである。なお、「正しく」検出できることを適合率、「全て」抽出できることを再現率で評価する。

実験手順としては被験者を4名と50文程度の3つの要求仕様書を用意する。その後、被験者が人手で状態遷移を抽出し、正解セットとする。なお、被験者個人特有の影響を軽減するため、各要求仕様書に対して異なる2名の被験者で状態遷移抽出を行う。

3.2. 実験の結果

状態に関する適合率・再現率はそれぞれ0.8~1程度で推移していた。また、状態遷移に関する適合率は平均0.95であり、再現率は約0.55であった。このことからある一定の効果を得られたと言えるが、状態遷移の再現率に関しては高い数値とは言えず、課題が残る結果であった。

抽出時間については、提案手法が平均0.18分であったのに対し、正解セットは平均50分抽出に有しており、大幅に提案手法は作業時間を削減できた。

4. 考察と今後の課題

False Negative や False Positive となった状態や状態遷移について以下のような要因が考えられる。

- I. 自己遷移
- II. 複数の状態を代替する名詞
- III. 文意の考慮が必要である状態遷移
- IV. 被験者の見落とし

今後の課題として、IやIIに関しては文型や代替する名詞の使用に関する仕様書記述時の制限、特定の接頭詞に対する処理を追加することで対応できると考えられる。しかし、IIIについては、文型変更等での対応が難しく本手法の限界であると言える。

参考文献

[1] 高橋淳也・高橋 宏季・位野木 万里:「日本語で記述された情報システムの要求仕様書中のあいまい表現が開発範囲に与える影響に関する考察」. 情報処理学会第81回全国大会, 工学院大学, 東京, 2019, p2.
 [2] 中村遼太郎・林晋平・佐伯元司:「ユースケース記述の検査のための自然言語要求文の解析」, 信学技報, SS2013-76, pp.25-30, 2014.