

自然言語で書かれた要求文の規則への整合性検査手法

中村 遼太郎¹ 林 晋平¹ 佐伯 元司¹

概要: 自然言語で書かれた要求文と規則の整合性をモデル検査で検査するためには、要求文を状態遷移モデルに、規則を検査式に変換する必要がある。この際、状態遷移モデルと検査式を対応づけるには、要求文と規則の語彙マッチングを行わなければならない。本稿では、用意した格フレーム辞書と要求文とのマッチング法を開発し、これによる規則への整合性検査手法を提案する。提案手法は、類義語、同義語、上位下位語を処理し、要求文の単語と格スロットに入るべき単語の意味的制約に基づくマッチングを行う。これにより要求文を格フレーム化した後、要求文の意味を踏まえた状態遷移モデル及び該当モデルに対応する規則の検査式を生成する。要求文と規則文に提案手法を適用し有用性を評価した。

キーワード: ユースケース記述, モデル検査, 格文法

Checking the Consistency between Requirements Specification Documents and Regulations

Abstract: When developers check the consistency between requirements specification documents and regulations by Model checking, they need state transition models of the documents and logic specifications of regulations. Moreover, they have to know which words in documents correspond to words in applicable regulations in respect of meaning so that they create logic specifications. In this paper, we propose a technique to reason the meaning of words in requirements specification documents by using co-occurrence restrictions in case frames and to create state transition models based on the reasoned meaning and logic specifications of applicable regulations. These specifications are created from these reasoned meaning and logic expressions which contain case frames as factors and express regulations. Our proposal is evaluated with a case study.

Keywords: Use case description, Model checking, Case grammar

1. はじめに

システム開発では、要件定義段階で法令をはじめとする規則と要求間の整合性を確保すべきと考えられている。規則に反する要求は、開発の後戻りやコンプライアンス違反を招く。これらを防ぐには、規則に準拠しない要求を発見し、その要求に対する禁止行為の除外や、義務行為の追加などの修正を加えなければならない。

規則に準拠した要求獲得という課題のために、規則と要求間の関係のモデリングやテキストマイニングを用いた手法が提案されている [1]。その中には、ユースケースと規則間のモデル検査による整合性検査 [2] も含まれる。この手法は、ユースケースを状態遷移モデルに、規則を検査式に

変換し、両者をモデル検査に適用する。その結果、検査式が満たされていれば、規則に対して整合性があると判断する。この手法では、モデル検査によりユースケースモデルを網羅的に検査することができる。

規則から検査式を作るには、分析者は、規則がユースケース上のどの要求と関係しているか把握しなければならない。これに対し、格文法に基づき、規則に対応する要求文を検出する手法 [3] が提案されている。この手法は、要求獲得作業中に獲得した要求文の格構造を、規則文の意味を表現した格フレームと照合することで、要求文と規則の対応関係を発見する。同じく、格文法に基づき、ユースケースの状態遷移モデルを生成する手法 [4] が提案されている。両者とも格文法を用いるため、併用が容易となっている。生成される状態遷移モデルは、ユースケース文の格構造を状態変数として扱う。そのため、先述の格文法を使った照

¹ 東京工業大学
Tokyo Institute of Technology

合で発見した要求文の規則と関わる格構造を組合せた検査式は、状態遷移モデルと併せてモデル検査に適用可能な検査式となる。

しかし、この格文法による要求文と規則のマッチングは、単語の意味を考慮しないため、規則にも出現する要求文の単語が規則と異なる意味で使われる場合や、字面が異なるものの意味的な互換関係にある単語が使われる場合に、対応関係の誤検出を起こし得る。このとき、モデル検査による整合性検査が誤った結果になる。このため、要求文の単語の概念を特定した後、要求文と規則の概念間の互換関係をふまえた対応関係を発見する必要がある。

本稿では、単語の共起関係から要求文の単語の概念を特定し、特定された概念が論理式で表された規則にある格フレーム上の概念と互換関係にあるか調べることで、要求文と規則間の対応関係を発見する手法を提案する。手法は辞書を参照し、格フレームの共起制限と要求文の単語の比較から、その単語の概念を特定する。次に、概念を特定された全ての文の格構造から、論理式にある規則の格フレームの共起制限を満たすものを探す。発見された格構造と格フレームを置換し、論理式をモデル検査へ適用可能な検査式へ具体化する。この手法は、我々のこれまでのアイデア [5] を具体化したものである。

例題ユースケースに手法の実装を適用し有用性を評価した。結果、作るべき検査式 6 つのうち 4 つが出力された。また、8 つの出力のうち 4 つの検査式が不要な検査式だったものの、その原因から生じる検査式は人が容易に誤りであると判断できるものであった。これにより、モデル検査適用前に出力を確認すれば、提案手法は約半数の検査式の作成を分析者に代わり行えることが分かった。

以降の構成を以下に示す。2 章で背景を述べる。3 章で提案手法を述べ、4 章で手法の実装を紹介する。5 章で手法の評価を行う。6 章で関連研究と本研究を比較し、最後に 7 章で本稿の結論を述べる。

2. 準備

2.1 格フレーム

動詞とその動詞がとり得る格要素の構造パターンを格フレームという [6]。また、文の動詞と格要素の構造パターンを格構造という。説明のために、送るの格フレームの例を次に示す。

(送る, 動作主: 人 | 組織, 対象: 具体物, 目標: 場所)

動作主格, 対象格, 目標格は、深層格と呼ばれ、動詞に対する意味的な役割を示す。一般的に、格フレームは、動詞と共起関係にある体言を表すために、共起する格要素に対して意味的な制約 (共起制限) を設ける [7]。上の例では、目標格になり得る単語が場所の概念をもつ単語に限定されることを表している。例において、場所の下線は、場所という単語ではなく場所に相応する概念そのものを指してい

る。動詞の意味次第で共起する単語が変わるので、動詞が複数の意味を持つ場合、その動詞は複数の格フレームを持ち得る。例えば、送るの場合、ものを送るという意味の他にも、時を過ごすという意味で使われる。この場合、対象格として共起するものは具体物ではなく、時の概念をもつ単語に限定される。

格の種類は深層格の他にも、文の構文から表層的に決まる表層格がある。表層格は文の構文に規定されるので、表層格のさらに細かい分類は言語によって異なる。日本語においては、格助詞につく体言が表層格になりえ、ガ格、ヲ格、ニ格のように格助詞の読み方の後に「格」をつけた表層格に分類される。格助詞と等しい文節にある単語が、その格助詞に対応する表層格となる。

2.2 格文法による要求文と規則の対応関係の検出

1 章で挙げた、要求文と規則の対応関係の検出手法 [3] は、要求文と対応する格フレームや規則の様相に従い、整合性をとるための修正を分析者に提示する。規則が法令の場合、規則は、適用される状況を示す論理式と、その状況下における適用対象者の行為の論理式からなる。また、様相は、義務、許可、禁止、免除のいずれかを指し、行為の意味を示す。たとえば、下記の個人情報の保護に関する法律の第十八条 *1 は、

状況 (取得, 動作主: x , 対象: 個人情報, 源泉: y)

$\wedge \neg$ (公表, 動作主: x , 対象: 利用目的)

義務 (通知, 動作主: x , 対象: 利用目的, 目標: y)

\vee (公表, 動作主: x , 対象: 利用目的)

のように格フレームの論理式で表現できる。異なる深層格間で変数名を共有し、等しい単語が入ることを示す。深層格にある単語は、その単語を該当する深層格に収めた文が格フレームとマッチすることを表す。様相が義務かつ状況における格フレームにマッチする要求文が発見された場合、分析者への提示は、義務を遂行する要求文の追加や、状況を招く要求文の再考となる。なお、状況と行為を時相オペレータと含意で結合することで、様相を表現する手法が提案された [2]。義務の場合、状況を s 、行為を a とすると義務の検査式は $s \rightarrow AF a$ となる。

ここで、「システムが個人情報を顧客から取得する」という要求文を新たに獲得した場合を考える。まず、この文から格構造、(取得, 動作主: システム, 対象: 個人情報, 源泉: 顧客) を抽出する。次に、要求文の動詞「所得」を抽出し、動詞が等しい格フレームを規則の論理式から探す。動詞が等しい格フレームが発見された場合、全ての深層格において、単語が格構造と格フレーム間で等しいか、あるいは格フレームの深層格が変数であれば、格構造と格フレームが対応関

*1 個人情報取扱事業者は、個人情報を取得した場合は、あらかじめその利用目的を公表している場合を除き、速やかに、その利用目的を、本人に通知し、又は公表しなければならない。

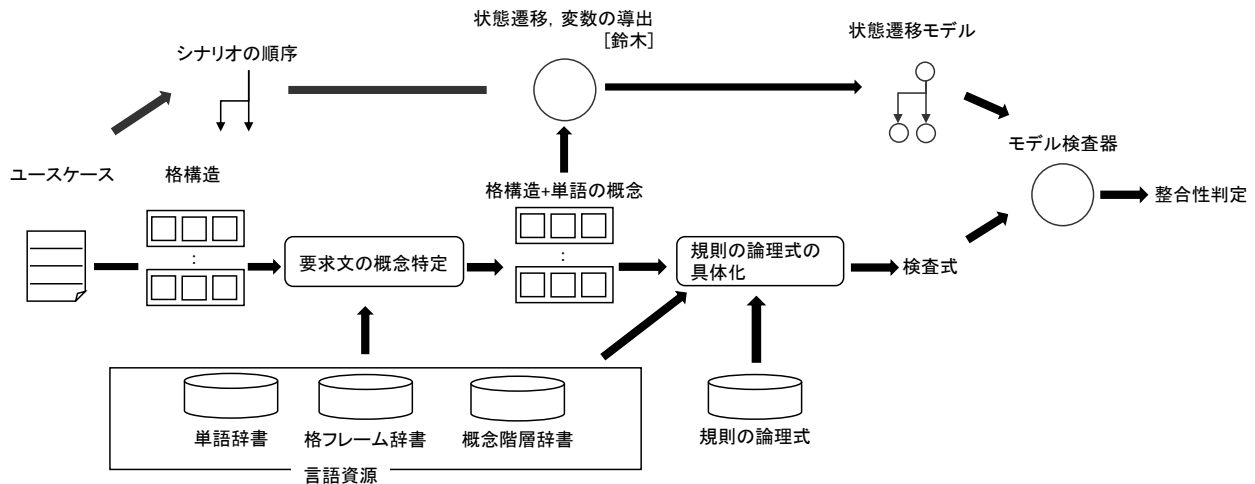


図 1 提案手法の概要

係にあると判定する．ここでは，状況に要求文の格構造と対応する格フレームが発見される．

1章で述べたように，格構造と格フレームのマッチングにおいて単語の概念が考慮されない場合，要求文と規則の対応関係の誤検出が起り得る．たとえば，先述の要求文の目的語が個人情報ではなく住所の場合，住所は個人情報にあたるものの，単語が異なるため対応関係の検出に失敗する．また，要求文においてシステムという単語は慣習的に擬人化された主体として扱われるので，ここでは個人情報取扱事業者と解釈できる．しかし，システムという単語は，文脈次第で制度という意味で使われることがある．このとき，システムを個人情報取扱事業者として解釈すると要求文と規則の対応関係の誤検出が生じかねない．したがって，要求文の文脈から単語の概念を特定し，特定された概念と規則の格フレーム上の概念間で意味的な互換関係があるか調べる必要がある．

2.3 ユースケースの状態遷移モデル生成

格文法によるユースケースの状態遷移モデルの生成手法 [8] は，ユースケース文の格構造を真偽値型の状態変数として扱い，状態遷移モデルを表現する．ユースケースでは，システムの動作はシナリオに書かれ，状態は条件に書かれるという仮定の下，この格構造は原文がシナリオ上のステップであれば動作の状態変数に，事前・事後条件であれば属性の状態変数に分類され，それぞれ異なる値の割り当て規則が割り振られる．動作の状態変数は，どのステップが実行中であることを示すために利用される．状態変数モデルは，シナリオ上のどのステップが実行中であることを，動作の状態変数のうち実行中のステップの状態変数だけに真を割り当てることで表現する．一方，属性の状態変数は，原文の条件が成立中であることを真で表現する．属性の状態変数の値は，ユースケースが主成功シナリオの最終

ステップに到達直後に変わりうる．事後条件に状態変数の格構造と等しい格構造の条件がある場合，値が真に切り替わる．他方，事後条件に状態変数の原文と対義語関係にある条件がある場合や，原文の否定文がある場合，値が偽となる．これら以外の場合，状態変数はこれまでの値を保持する．

3. 提案手法

3.1 概要

入力されたユースケースから規則の論理式上の格フレームと対応する文を発見し，その文の格構造を状態変数とする状態遷移モデルと併せてモデル検査に適用可能な検査式を出力する手法を提案する．手法の概要を図 1 に示す．手法は，図の左にあるユースケースから格構造を抽出し，辞書にある単語と概念の関係，概念間の is-a 関係，格フレームを参照しながら，格構造にある単語の概念を特定する．まず，文がとり得る全ての格フレームについて，格要素の単語がとり得る概念のうち，格フレームの共起制限を最も満たす概念を求め，概念が共起制限をどの程度満たすかを数値化する．数値に基づき，全ての格フレームのうち，共起制限が最も満たされた格フレームを文の格フレームとみなし，マッチした格フレームで求めた概念をその文における単語の概念とみなす．次に，概念を特定された全ての文の格構造から，辞書を使い，論理式にある規則の格フレームの共起制限を満たすものを探す．発見された格構造と格フレームを置換し，論理式をモデル検査へ適用可能な検査式へ具体化する．

文の格構造を状態変数とする状態遷移モデルの生成のために [8] の手法を用いる．ユースケースから抽出したシナリオの順序と文の格構造から，状態遷移モデルを生成する．ただし，[9] に従い，概念を特定した格構造を状態変数に使うことで，同義語関係を考慮した状態遷移モデルが生成さ

れるようにする。

3.2 言語資源

3.1 節で述べたように、単語と概念間の対応関係、概念間の is-a 関係、動詞の概念の格フレームを言語資源として活用する。この3種の言語資源の関係を図2に例示する。

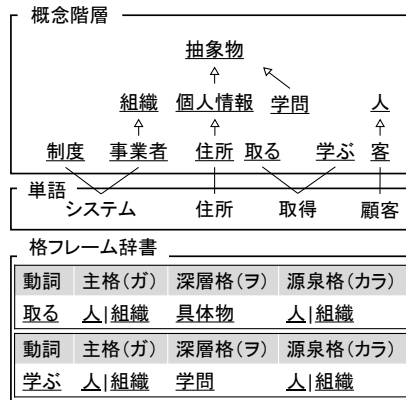


図2 言語資源

言語資源にある単語の集合を W 、概念の集合を C として、単語と概念間の意味的な多対多の関係を以下のように定める。

$$\rightarrow \subseteq W \times C$$

この関係は、文脈次第でとり得る概念が変わる単語や複数の単語で共有される概念を想定している。例えば、システムという単語は、制度という概念に言及するために使われるもの、要求文においては慣習的に擬人化された主体を指すために使われる。例えば、図2では顧客 \rightarrow 客 が成り立つ。複数の単語が関連付く、すなわち、 $|\{w \mid w \rightarrow c\}| \geq 1$ となる概念 c については、[9]にある同義語関係を踏まえた状態遷移モデル生成のために、 c を代表する単語 w_{rep} を定める。また、 $\{c \mid w_{rep} \rightarrow c\}$ から w_{rep} への関数を rep と定める。

単語と概念間の関係に続き、概念間の is-a 関係を

$$\triangleright \subseteq C \times C$$

で定義する。左の成分は右の成分と等しいか小さいものとする。例えば、図2では客 \triangleright 人が成り立つ。この is-a 関係は単語「概念」の概念である概念を最大、単語「無し」の概念である無しを最小とする半順序関係をなす。また、一つの概念は複数の親概念をとり得る。

格フレームを、動詞の概念 c_v と、深層格の種類 $case_d$ とそれに対応する共起制限 $rest$ を成分とする組の集合を組み合わせて

$$f = (c_v, {}^{case_d}rest, \dots) \quad (1)$$

と定め、またその辞書を F とおく。例えば、図2では格フ

レーム辞書は2つの格フレームを保持しており、そのひとつの f_1 は

$$f_1 = (\text{取る}, {}^{\text{主}}\text{人}|\text{組織}, {}^{\text{深層}}\text{具体物}, {}^{\text{源泉}}\text{人}|\text{組織}) \quad (2)$$

と表せる。

また、係り受け解析の結果から文節に対応する深層格を推定するために、格助詞から格助詞と同分節にある体言の対応する深層格への関数 m_f が格フレームごとに定めてあるものとする。例えば、 f_1 では主格としてガ格を取るため、 $m_{f_1}(\text{ガ})$ は主格を指す。

概念の数を考えると、共起制限を動詞と共起する概念集合として表すと表記が煩雑となる。そこで表記の簡潔化のために、共起制限に現れる概念に、その概念以下の概念が共起制限を満たすという解釈を与える。さらに論理積、論理和、否定の演算子を導入する。例えば、図2の主格の共起制限は $c \triangleright \text{人}$ または $c \triangleright \text{組織}$ となる c が取る と共起することを意味する。論理積は $\&$ 、否定は $!$ で表す。

3.3 要求文の概念特定

入力文の格解析。はじめに、深層格と動詞の概念を単語辞書に問い合わせる。入力文に係り受け解析を適用した結果から、文の動詞 w_v に加え、体言とその体言と同文節にある格助詞の組の集合 ${}^{case_s}w_n, \dots$ を求める。例えば、入力文が「システムが住所を顧客から取得する」の場合、(取得, ${}^{\text{ガ}}$ システム, ${}^{\text{ヲ}}$ 住所, ${}^{\text{カラ}}$ 顧客) が得られる。ここで、格助詞に重複はないものとする。また、修飾語のように、他の文節が体言と格助詞を含む文節に係る場合、係る文節を体言の一部として w_n に含めてもよい。

格フレーム候補の特定。文の動詞 w_v がとりうる概念と対応する格フレームの集合を求める。 w_v がとり得る概念を動詞としてもつ格フレームが候補となる：

$$F_{w_v} = \{(c_v, {}^{case_d}rest, \dots) \in F \mid w_v \rightarrow c_v\} \quad (3)$$

例えば、 $w_v =$ 取得 であれば、図2の場合、図中の格フレーム候補2つのいずれも候補となる。候補が空集合となるケースは、マッチング対象の格フレームがないことを意味する。この場合、単語の概念は特定されず、ここで終了する。

表層格と深層格の対応付け。次に、候補の要素ごとに、抽出された体言 w_n と深層格 ($case_d, rest$) の対応関係を求める。候補のそれぞれの格フレームは、表層格をどの深層格に対応付けるかを規定しているため、 $case_d = m_f(case_s)$ となる深層格 $case_d$ およびそれに関連付いた共起制限 $rest$ が、表層格 $case_s$ およびそれに関連付いた体言 w_n に対応する。例えば、先ほど求めた格解析結果からは、以下の対応付けが得られる。

$$\begin{aligned} & \{(ガ:システム \leftrightarrow 主:人|組織), \\ & (ヲ:住所 \leftrightarrow 対象:具体物), \\ & (カラ:顧客 \leftrightarrow 源泉:人|組織)\}, \end{aligned} \quad (4)$$

$$\begin{aligned} & \{(ガ:システム \leftrightarrow 主:人|組織), \\ & (ヲ:住所 \leftrightarrow 対象:具体物), \\ & (カラ:顧客 \leftrightarrow 源泉:人|組織)\} \end{aligned} \quad (5)$$

対応付けの満足度合いの測定。得られた対応付けの各要素について、体言 w_n がとり得る概念のうち、最も $rest$ を満たす概念を入力文における w_n の概念と特定する。そこで、体言の概念 c が共起制限 r を満たす程度を数値化する関数 s を導入する。共起制限の規定に合致する度合いの強弱を表現するため、ファジィ真理値関数 [10] を参考にして以下のように関数 s を定義する。

$$s(r, c) = \begin{cases} \min\{s(r_a, c), s(r_b, c)\} & (r = r_a \& r_b) \\ \max\{s(r_a, c), s(r_b, c)\} & (r = r_a | r_b) \\ 1 - s(c_r, c) & (r = '!c_r) \\ s(c_r, c) & (r = c_r) \end{cases} \quad (6)$$

この値が高いほど概念が共起制限を満たすことを意味する。ここで、 c_r は概念を表す。 $s(c_r, c)$ の値は以下のように定義される。

$$s(c_r, c) = \begin{cases} 1 & (c \triangleright c_r) \\ 1/(d_{min} + 1) & (\text{それ以外}) \end{cases} \quad (7)$$

$$d(c_c, c_p) = \begin{cases} 0 & (c_c = c_p) \\ 1 & (c_p \text{ が } c_c \text{ の親}) \\ 1 + d' & (\text{それ以外}) \end{cases} \quad (8)$$

$s(c_r, c)$ の値は、 c_r と c 間の is-a 関係で決まる。 c_r が c と is-a 関係にあるとき 1 となる。それ以外の場合、(8) の is-a 関係にある概念の世代差を返す関数 d の値で決まる。ただし、 d' は $\{d(c_c, c_{p'}) \mid c_{p'} \text{ は } c_p \text{ の子}\}$ の最小値を、 d_{min} は $\{d(c_c, c') \mid c_c \triangleright c' \wedge c_p \triangleright c'\}$ の最小値、すなわち c_c と c_p の共通の祖先と c_c の世代差を示す。例えば、図 2 の場合、 $s(\text{学問}, \text{住所}) = 1/(2 + 1) = 0.67$ となる。

この s を、得られたすべての対応付けに対して適用し、入力文の格フレームが規定するそれぞれの共起制限の満足度合いを求め、

$$slots = \{(w_n, case_d, rest, cs, score)_i\} \quad (9)$$

を得る。ここで、格フレームの深層格に対応する表層格が複数存在する場合は、値がもっとも高いもののみを採用し、 $score$ はそのときの満足度合いの最大値、 cs は、該当の満足度合いが得られた概念の集合を表す。

図 2 に従い、(4) と (5) から満足度合い (9) を求めると、そ

れぞれの対象格に対応する要素は、(住所, 対象, {住所}, 1) と (住所, 対象, {住所}, 1/3) となる。

格フレームの特定。最後に、格フレーム全体での総合的な満足度合いを求める。

$$score_f = \frac{\sum_{k=1}^{|slots|} score_k}{|slots| + 1} \quad (10)$$

(10) は、分母に深層格との対応がみられた体言の数に 1 を加えた値をとる。また、分子に、深層格ごとに求めた $score$ の合計をとる。分母に 1 を足すことで、深層格と体言のペアが多い格フレーム候補ほど、高い点数になりやすい採点方法となっている。この値が最も高い格フレーム候補を要求文の格フレームと特定し、その候補で求めた概念を要求文における単語の概念と判定する。

これまで例で用いてきた入力文「システムが住所を顧客から取得する」の計算過程より、取る と 学ぶ に対する格フレーム候補の満足度合い (10) の値は、 $3/4(0.75)$ と $2.33/4(0.58)$ となり、入力文の格フレームが 取る の格フレームと特定される。さらに、入力文において「取る」は 取る、システムは 事業者 の概念で使われていることが分かる。最終的に、 $score_f$ が最も高い格フレーム候補の一つについて (11) を出力する。

$$(w_v, c_v, score_f, slots) \quad (11)$$

3.4 規則の論理式的具体化

論理式的具体化のために、規則の論理式にある格フレームに対して互換性のある格フレームをもつ要求文を探す。

ここで、格フレーム間の互換性について定義しておく。格フレーム $f_1 = (c_1, case_i^1: rest_i^1, \dots)$, $f_2 = (c_2, case_j^2: rest_j^2, \dots)$ について、 $c_1 = c_2$ かつ任意の $case_i^1$ について $case_i^1 = case_j^2$ となる $case_j^2$ が存在し、さらに両者の共起制限間で

$$\forall c \bullet (s(rest_i^1, c) = 1 \Rightarrow s(rest_j^2, c) = 1) \quad (12)$$

が成り立つとき、かつそのときに限り f_2 は f_1 に対して互換性があるという。動詞の概念や共起制限の比較には、(11) を用いる。

検査式的具体化に用いる要求文の組み合わせを直積により作成する。規則の論理式に出現する格フレーム f_i に対して互換関係にあり、かつ体言の概念集合 cs がいずれも空集合ではない要求文の (11) の集合を CS_i とし、各 i についての直積集合 $CS_1 \times \dots \times CS_n$ を作成する。ここで、 n は論理式に出現する格フレームの総数を示す。また、 $CS_i = \emptyset$ となる場合、該当部分を偽として展開、すなわち、 $CS_i = \{FALSE\}$ とする。

概念を特定した規則の格フレーム f_i を 3.2 節で述べた関数 rep を使い、 CS_i の要素の成分からなる以下の式で置換することで論理式を具体化する。

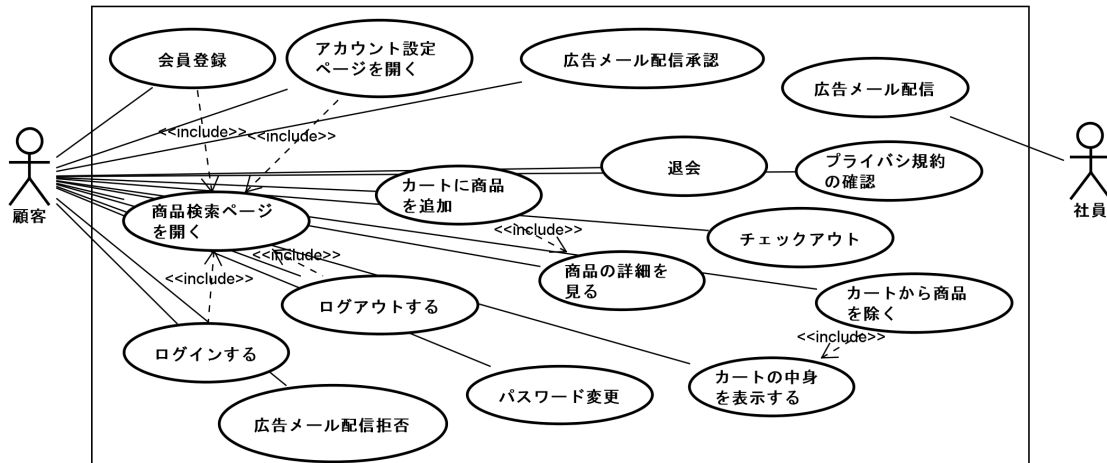


図 3 適用例題のユースケース図

$$(rep(c_v),^{case\ a:} rep(c_{rep}), \dots)) \quad (13)$$

ただし、 c_{rep} は (9) にある体言の概念集合 cs の任意の一要素とする。例えば下記の規則の論理式を具体化するケースを考える。

$$\begin{aligned} & (\text{取得}, \text{動作主: 事業者}, \text{対象: 個人情報}, \text{源泉: 人}) \\ & \rightarrow \text{AF} (\text{通知}, \text{動作主: 事業者}, \text{対象: 利用目的}, \text{目標: 人}) \end{aligned}$$

このとき、論理式の左側の取得の格フレームに対して

$$\{(\text{取得}, \text{動作主: システム}, \text{対象: 住所}, \text{源泉: 顧客})\},$$

通知の格フレームに対して

$$\{(\text{通知}, \text{動作主: システム}, \text{対象: 利用目的}, \text{目標: 顧客})\},$$

というユースケース文との対応関係が見られた場合、

$$\begin{aligned} & (\text{取得}, \text{動作主: システム}, \text{対象: 住所}, \text{源泉: 顧客}) \\ & \rightarrow \text{AF} (\text{通知}, \text{動作主: システム}, \text{対象: 利用目的}, \text{目標: 顧客}) \end{aligned}$$

という検査式が具体化される。

4. 実装

評価のために手法を実装した。実装は、ユースケースの状態遷移モデル生成と規則の論理式の具体化を支援する。状態遷移モデルは NuSMV*² のソースコードとして出力される。規則の論理式を、真偽値に加え、規則の格フレームを原子項とする CTL 式で表現できる。また、係り受け解析のために Cabocha*³ を利用する。

5. 評価

5.1 目的

例題ユースケースへの実装の適用結果に基づき、要求文と規則間の対応関係における検出精度の観点から手法を

*² <http://nusmv.fbk.eu/>

*³ <https://code.google.com/p/cabocha/>

評価する。正解として用意された作成すべき検査式に対して、実装の出力の適合率と再現率を測る。また、適合率や再現率低下の原因について考察する。

5.2 適用例題

架空のオンライン通信販売サイトのユースケースを例題とした。このユースケースは、顧客が扱える機能を主眼において機能要求を説明する。例題ユースケースのユースケース図を図 3 に示す。

例題ユースケースは我が国の 4 つの規則の適用対象となる箇所を含む。4 つの規則は、個人情報の保護に関する法律 第十八条、特定電子メールの送信の適正化等に関する法律 第三条の一、特定商取引に関する法律 第十一条、同第十三条にあたる。[2] に従い、第三条から 2 つの論理式を、残りの規則から 1 つずつ論理式を作り、計 5 つの規則の論理式を検査式の具体化のために用意した。以降の説明のために、第十八条、第十一条、第十三条の論理式を (14)、(15)、(16) に示す。右肩に「*」のない格フレームは 2.3 節で述べた動作の状態変数に、「*」のある格フレームは属性の状態変数に変換される。

$$\begin{aligned} & \text{AG}((\text{取得}, \text{動作主: 事業者}, \text{対象: 個人情報}, \text{源泉: 人}) \\ & \quad \&!(\text{公表}, \text{動作主: 事業者}, \text{対象: 利用目的}) \\ & \rightarrow \text{AF} (\text{通知}, \text{動作主: 事業者}, \text{対象: 利用目的}, \text{目標: 人}) \\ & \quad | (\text{公表}, \text{動作主: 事業者}, \text{対象: 利用目的})) \end{aligned} \quad (14)$$

$$\begin{aligned} & \text{AG} ((\text{表示}, \text{動作主: 事業者}, \text{対象: 通信販売広告}) \\ & \rightarrow \text{AF} (\text{表示}, \text{動作主: 事業者}, \text{対象: 価格})) \end{aligned} \quad (15)$$

表 1 単語と概念の関係, 概念間関係, 格フレームの規模

	→	▷	格フレーム
EDR	270000	410000	13000
EDR 以外 *	61(9)	59(17)	10(1)

* () は検査式の具体化のために参照された言語資源の数。

表 2 正解数と結果の比較

	第十八条	第三条	第十一条	第十三条
正解	2	2	1	1
出力 (正)	2	2	0	0
出力 (誤)	2	0	1	1

AG((受領, 動作主: 事業者, 対象: 代金, 源泉: 人)
→ AF((通知, 動作主: 事業者, 対象: 注文の承認, 道具: 電磁的方法)
&(承諾, 動作主: 人, 対象: 電磁的方法による通知)*
|(通知, 動作主: 事業者, 対象: 注文の承認, 道具: 書面)))
(16)

EDR 電子化辞書 [11] と著者の一人が作成した知識を併せて言語資源として使用した。辞書になく例題に出現する知識を補填するために EDR 以外の知識を追加した。言語資源の大きさを表 1 に示す。

5.3 結果

手法の適合率は 4/8(0.50), 再現率は 4/6(0.66) となった。規則ごとの正解数と正解ごとの出力結果を表 2 に示す。

5.4 考察

適合率から, 分析者は手法の出力の半数を, 修正することなくモデル検査に適用できるといえる。しかし, 適合率が 1 未満となったため, 本手法の出力をモデル検査適用前に, 分析者によって出力から不要な検査式を除外する必要があることが示された。

正解以外の検査式が出力された原因を分析する。第十八条の論理式 (14) から具体化された正解ではない検査式の一つを以下に示す。

AG(((受信, 動作主: 事業者, 対象: メールアドレス, 源泉: 顧客)& !FALSE)
→ AF((通知, 動作主: 事業者, 対象: 住所の用途, 目標: 顧客)|FALSE))
(17)

第十八条の原文に立ち返れば, 通知の対象格は取得の対象格に関する利用目的であるべきであるものの, 具体化された式にはこの対象格の対応関係がみられない。

この原因は, 格フレーム間の意味を顧みず, 論理式の具体化に使うユースケース文の組み合わせを選んだことにある。具体化された式における格フレームは, もととなる式の対応する格フレームと意味的な整合性がとれているもの

の, 同時に出現している格フレーム同士には意味的な繋がりがみられない。これは, 論理式の具体化に用いる格構造の組み合わせを意味的な整合性を考慮せず, 直積で作成したことが原因となっている。

適合率の結果から, 手法は作るべき検査式のうち, 約半分を出力できるといえる。ここで, 具体化できなかった論理式について考察する。第十三条の論理式 (16) は, 通知と受領の格フレームについてユースケース文と対応すべきであったものの, 受領の格フレームの意味に対応するユースケース文が, その格フレームにマッチしない格構造であったために, 正解通りに具体化されなかった。受領の格フレームに対応する要求が「システムが料金の支払いを承認する」というユースケース文で説明されていた。提案手法は, 格構造と格フレームの対応関係に基づき, 要求文と規則の対応関係を検出するので, 要求文の格構造が対応すべき規則の格フレームと異なる場合, 対応関係を発見できない。

第十一条の論理式 (15) は, 先頭の格フレームとユースケース文「システムがブラウザに商品のサムネイルを表示する」が対応すべきだったものの, 「サムネイル」の概念が通信販売広告ではなく画像と誤って特定されたため, 論理式が具体化されなかった。画像と広告はどちらも表示という動詞と共起するため, 概念特定が誤ったと考えられる。このように, 深層格の単語がとり得る概念のいくつか, いずれも動詞の概念と共起する場合, 概念の特定が誤り, 規則との対応関係を発見できない場合がある。

5.5 妥当性の脅威

手法を本稿の著者が作成した例題を評価した点に内的妥当性の脅威がある。ユースケースの規模が小さい場合, 要求文と規則間の対応関係の検出を誤る可能性が小さくなるので, 実験結果がよい結果となりやすい。そこで, 例題ユースケースに [4] の被験者実験で使われたユースケースと同等の複雑さと規模をもたせた。被験者実験のユースケースは, 実験結果から, 人がモデル検査器で検査するよりも, 状態遷移モデル生成手法で検査を行う方が正しい結果を得ることができる程度に, 複雑で大きなものであることが分かっている。被験者実験のユースケースは, シナリオ数 15, 総ステップ数 144, include 遷移 3 となっている。本実験の例題ユースケースは, 順に 16, 146, 5 となっている。

評価に用いられたユースケースが一つしかない点に外的妥当性の脅威がある。規則が容易に検査できるものであったために, 論理式を具体化できた可能性がある。他のユースケースに対しても, 手法が実験結果と同程度に機能することを示すには, 他のユースケースや規則に手法を適用する必要がある。しかし, 法令はあらゆるドメインで定められているだけでなく, ドメインによらず, ある程度形式化された書き方で書かれている。よって, 今回の実験結果から,

規則が法令であれば、手法が他のユースケースに対しても実験結果と同程度に機能しうることが示せたと考えられる。

6. 関連研究

黒橋らは、和文の構文・格解析器 KNP^{*4}を開発した。KNPはWebテキストから自動構築された格フレーム[12]をもとに、係り受け、格、照応関係を出力する。この格フレームは深層格のような意味的な情報を保持しないので、KNPを単語の概念特定に利用することはできない。一方、本稿の手法は、深層格の共起制限に基づき要求文の単語の概念を特定する。これにより、要求を自然言語で表現していても、[13]のような格文法に基づく仕様記述言語で要求を記述する場合と同様に、要求の機械的な検査が可能になる。

規則のモデリング手法が提案されている[14],[15]。これらの手法は規則間の関係や規則と要求間の関係をモデル化することで、要求の規則に対する適合性を形式的手法で分析する。これにより適用対象となる規則や、規則の適用対象とならない代替的な要求などを分析者に提示できる。これらの手法を利用するには、分析者自らが規則と要求の対応関係を発見し、要求を規則モデルと比較可能な形式に変換する必要がある。本手法は、格フレームの論理式で表現された規則と要求文の間の対応関係を格フレームにある共起制限を元に特定する。そして、特定された対応関係を踏まえ、整合性をとるべき規則の検査式を出力し、モデル検査により整合性を検査する。

本稿で利用したユースケースの状態遷移モデル生成手法のほかにも生成手法が提案されている。Sinnigらはユースケースの記述方式に形式的な定義を与えることで、ユースケースを状態遷移モデルに変換する手法を提案した[16]。Sinningらの手法は、ユースケース文の意味に対しては定義を与えていないので、文の意味を考慮した状態遷移モデルを生成することができない。しかし、並行実行のように本稿で利用した生成手法では対応できない振り舞いに関する変換規則に対応している。そのため、これらの規則を取り入れることで、より多くのユースケースを支援対象にすることができると考えている。

7. おわりに

モデル検査で要求文の規則への整合性を検査するための検査式を生成する手法を提案した。手法は辞書を使い、格フレームの共起制限と要求文の単語から要求文における単語の概念を特定する。そして、特定された概念が格フレームの論理結合で表された論理式上の概念と互換関係にあるか調べることで、規則と要求間の対応関係を求める。対応関係が見られた要求文の格構造で規則の論理式の格フレ

ームを置き換え、論理式をモデル検査に適用可能へ検査式を具体化する。適用事例から、モデル検査適用前に出力を確認すれば、提案手法は約半数の検査式の作成を分析者に代わり行えることを示した。適用実験を重ね、より確度の高い手法の評価を行うことを主要な今後の課題としている。

参考文献

- [1] Otto, P. N. and Antón, A. I.: Addressing Legal Requirements in Requirements Engineering, *Proc. RE*, pp. 5–14 (2007).
- [2] Saeki, M., Kaiya, H. and Hattori, S.: Detecting Regulatory Vulnerability in Functional Requirements Specifications, *Proc. ICSE*, Vol. 1, pp. 105–114 (2009).
- [3] Saeki, M. and Kaiya, H.: Supporting the Elicitation of Requirements Compliant with Regulations, *Proc. CAiSE*, pp. 228–242 (2008).
- [4] 高久陽平, 林晋平, 佐伯元司: ユースケース記述からの状態遷移モデル生成, 情報処理学会研究報告, Vol. 2010, No. 17, pp. 1–8 (2010).
- [5] 中村遼太郎, 林晋平, 佐伯元司: ユースケース記述の規則への整合性検査に向けて, ソフトウェアエンジニアリングシンポジウム 2014 論文集, pp. 76–84 (2014).
- [6] Fillmore, C. J.: Some problems for case grammar, *Monograph series on languages and linguistics*, Vol. 24, pp. 35–56 (1971).
- [7] Bird, S., Klein, E. and Loper, E.: 入門 自然言語処理, O'Reilly Japan, Inc. (2010).
- [8] 鈴木啓史: ユースケース記述の検査を目的とした状態遷移モデル生成の研究, 東京工業大学工学部情報工学科卒業論文 (2012).
- [9] 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, 電子情報通信学会技術研究報告, Vol. 113, No. 489, pp. 25–30 (2014).
- [10] Russell, S. and Norvig, P.: *Artificial Intelligence: A Modern Approach*, Prentice Hall Press, 3rd edition (2009).
- [11] 日本電子化辞書研究所: EDR 電子化辞書.
- [12] 河原大輔, 黒橋禎夫: 高性能計算環境を用いた Web からの大規模格フレーム構築, 情報処理学会研究報告, Vol. 2006, No. 1, pp. 67–73 (2006).
- [13] 大西淳, 阿草清滋, 大野豊: 要求フレームに基づいたソフトウェア要求仕様化技法, 情報処理学会論文誌, Vol. 31, No. 2, pp. 175–181 (1990).
- [14] May, M. J., Gunter, C. A. and Lee, I.: Privacy APIs: Access Control Techniques to Analyze and Verify Legal Privacy Policies, *19th IEEE Computer Security Foundations Workshop*, pp. 85–97 (2006).
- [15] Ingolfo, S., Jureta, I., Siena, A., Perini, A. and Susi, A.: Nòmos 3: Legal Compliance of Roles and Requirements, *Proc. ER*, pp. 275–288 (2014).
- [16] Sinnig, D., Chalin, P. and Khendek, F.: LTS Semantics for Use Case Models, *Proc. SAC*, pp. 365–370 (2009).

^{*4} <http://nlp.ist.i.kyoto-u.ac.jp/index.php?KNP>