

自然言語処理を利用した要求仕様からの UTP モデル抽出方法

増田聡† 松尾谷徹‡ 津田和彦††

日本アイ・ビー・エム株式会社† デバッグ工学研究所‡ 筑波大学††

1. はじめに

ソフトウェアの設計開発のため、要求仕様をモデル化する言語に Unified Modeling Language (UML) がある。同様に、ソフトウェアのテストのため、要求仕様を分析した結果をテストのモデル化する定義に UML Testing Profile (UTP) がある [3]。UTP にはテストのモデルとして、テストアーキテクチャ、テストビヘイビア、テストデータおよびタイムコンセプトが定義されている。要求仕様は自然言語で記述されており、現在は UTP モデルへの抽出はモデリング技術を有する技術者が行う。そのため、技術者のスキルに依存し UTP モデル抽出にばらつきが生じている。これらのことから、自然言語で記述された要求仕様から UTP モデルの自動抽出技術が支援装置として有効と考えられている。本論文では、英語の要求仕様を対象に、UTP のテストビヘイビアで定義されているシーケンス図によるテストケース記述に着目し、自然言語で記述された要求仕様から UTP のテストケース記述の抽出方法を提案し評価する。

2. 要求仕様から UTP モデル抽出の課題

自然言語処理の構文解析で得られる情報は、単語の品詞、係り受けとその関係などである。例えば、「The system stores the new link.」という要求仕様は、構文解析の結果、図 1 の品詞とツリー構造であることが分かる。S は文、NP は名詞句、VP は動詞句、NN は名詞、VBZ は他動詞などを表す。

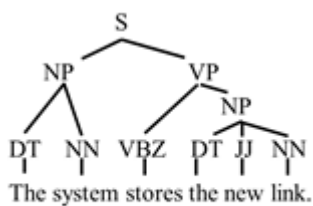


図1. 構文解析結果例

UTP のテストケースのシーケンス図の情報は、クラス、動作とその引数、動作の順番などである。例えば、図 2 のように記述される。この例では、Activity 01 では、「The user click the link

as LinkNew.jsp.」と解釈できる。

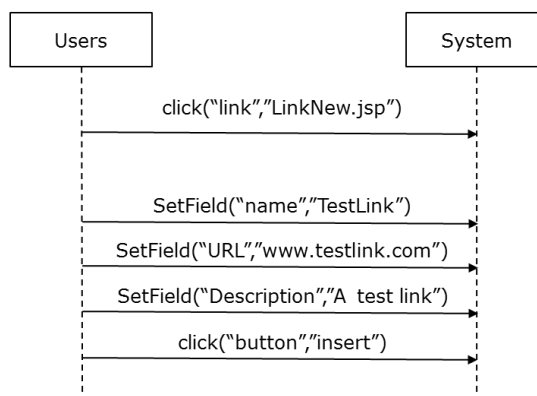


図 2. UTP テストケース記述例 ([1] Fig. 3 を編集)

ここで、自然言語で記述された要求仕様を **S** とし、UTP のシーケンス図によるテストケースのアクティビティ記述を **U** とした場合、

$$U = G \cdot U \quad (1)$$

式(1)が成り立つ抽出ルール **G** を定義することが本論文の課題である。

研究アプローチとしては、関連研究 [2] を基に UTP モデル抽出ルールを検討する。

3. 要求仕様から UTP モデル抽出方法

要求仕様から UTP モデル抽出方法の概要を図 3 に示す。要求仕様を自然言語処理し品詞や構文、係り受け情報を得る。その情報から抽出ルールに基づき UTP モデルを抽出する。

関連研究 [2] では、要求仕様から自然言語処理を利用してクラス図の情報を抽出している。その抽出ルールは、名詞はクラスとし、その名詞に係る形容詞はそのクラスの属性、動詞はそのクラスの動作としている。UTP モデルを抽出する場合にクラスの定義は同じであるのでこれらの抽出ルールは同様に適用可能と考える。しかしながら、これらの抽出ルールに加えて、シーケンス図のメッセージを記述するためには、どのクラスがどのクラスにどのような動作を行うかという情報が必要である。このクラス間のメッセージに関しては、それぞれのクラス間の動作であるので、クラス名の間には動詞が含まれる構造を抽出ルールとした。

以上のことから、抽出ルール **G** を以下のとおり

Extracting UTP Models from Requirements by Using Natural Language Processing

† Satoshi Masuda, IBM Japan Ltd.,

‡ Tohru Matsudani, Debug engineering research laboratory

†† Kazuhiko Tsuda, University of Tsukuba

とした。

- ・ルール 1: クラス抽出ルール

主語の名詞はクラス，動詞は動作，補語は引数とする．文の構造表現は(2)である．(*)は任意の品詞を表す．要求仕様の自然言語解析の結果でこの文構造を持つ仕様は NN1 をクラス，VBZ を動作，NN2 を属性とする．

S (NP ((*)(NN1) (*)) VP (VBZ)(NP ((*)(NN2)(*))) (2)

- ・ルール 2: クラス間のメッセージ抽出ルール

(2)で NN1 と NN2 の名詞が既に出現したクラス名である場合，その間の動詞 VBZ は NN1 から NN2 へのメッセージとする．

- ・ルール 3: シーケンスの順番は要求仕様に記述されている順番どおりとする．

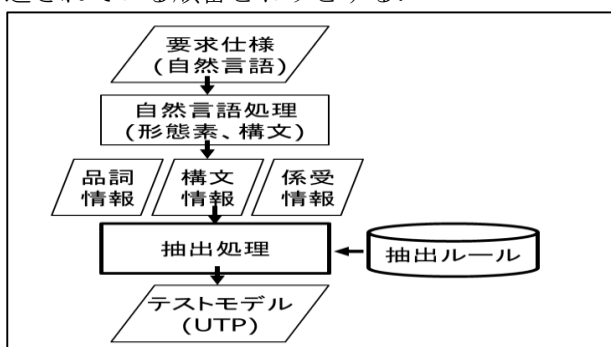


図. 3 要求仕様から UTP モデル抽出

4. UTP モデル抽出ルール適用実験と評価

提案方法を実装し適用実験と評価を行った．要求仕様の対象は参考文献の[1] (GEN とする) と [4] (CHART とする)を用いた．自然言語解析は英語用のツールを用い，抽出部分は抽出ルールをもとに自作した．評価は抽出結果が UTP テストケースのモデルとして正しいかを，UTP 有識者により判断し，精度と再現率を計算した．

表 1 に GEN の実験結果、表 2 に CHART の実験評価結果を示す．

		精度	再現率
ルール 1	クラス	1.00	0.80
	動作	0.75	0.75
	属性	0.75	0.75
ルール 2	メッセージ	1.00	1.00
対象文の数=6			

表 1. GEN に対する実験結果

		精度	再現率
ルール 1	クラス	0.81	0.81
	動作	0.75	0.75
	属性	0.56	0.56
ルール 2	メッセージ	0.75	0.86
対象文の数=15			

表 2. CHART に対する実験結果

5. 考察

GEN の評価結果が 0.75 以上と高い値になっている．もともと GEN は要求仕様から手作業で UTP を作成できるものなので，高い値になっているということは，本論文の抽出ルールが少なくとも手作業を反復することは可能であると考えられる．より一般的な要求仕である CHART への適用評価も属性を除いて，0.75 以上の高い値となった．これからも，本論文の抽出ルールが有効である可能性は確認されたと考えられる．属性が 0.56 と低い値となっている原因は，抽出ルールの構文構造に適用した結果，属性を示す NP の構造が名詞の連続など複雑な構造になり NN2 が特定できないことが原因と考えられる．また，これは英文の特徴として，主語 (NP) と動詞 (VP) は簡潔に述べて，他の情報をその後に述べるということが関係しているとも考えられる．以上のことから，複雑な文への対応などさらに対象を拡大して検証する必要はあるが，本論文で提案している UTP モデルの抽出ルールは有効である可能性は確認されたと考えられる．今後は対象を拡大し抽出ルールも追加改善し，より一般的な要求仕様へ対応可能とする．

6. おわりに

本論文では，英語の要求仕様を対象に，UTP のテストビヘイビアで定義されているシーケンス図によるテストケース記述に着目し，自然言語で記述された要求仕様から UTP のテストケース記述の抽出方法を提案し評価した．提案する抽出ルールは有効な部分があることが確認された．今後は今回の結果を踏まえ，対象を拡大し抽出ルールを改善していく．

参考文献

- [1] J. J. Gutierrez et al: "An approach to generate test cases from use cases", 6th international conference on Web engineering, pp.113-114,2006
- [2] O. Keszocze et al: "Lips: An ide for model driven engineering based on natural language processing", IEEE Natural Language Analysis in Software Engineering (NaturaLiSE), pp.31-38,2013
- [3] Object Management Group, "UML Testing Profile(UTP) Version 1.2", <http://www.omg.org/spec/UTP/1.2/>
- [4] Maryland State Highway Administration: "Coordinated Highways Action Response Team (CHART) Mapping Applications", http://www.chart.state.md.us/downloads/readingroom/ChartIntranetMapping/Detailed_System_Design_Specification_Final.doc