

パターンを活用した自然言語からのモデル生成

豊田学[†], 原口智史[†], 三上徹也[†], 坂本佳史[†]

組込みシステムの開発において、機能要求は要求文書として記述される事例が依然として多く見受けられる。しかしながら開発工程の上流で作成されるそれらの要求文書と、開発の下流工程における技術成果物であるモデルとの、技術的な観点での整合性やトレーサビリティは明確に担保されていないことが現実の開発課題となる場合がある。すなわち要件が正しくシステムに実装されることを検証の工程に至るまで確認できないことによって、開発の手戻りを発生させ、大きなスケジュール・インパクトの主原因となる。その理由の1つとして、要求文書は開発や設計における参照の対象であることが考えられる。要求文書は開発における一連の技術成果物とは考えられておらず、加えて従来の開発で用いられているモデルの構造やモデル要素との整合性も考慮されていない場合がある。この問題に対して、我々は自然言語で記述された開発文書からモデルを生成する手法を提案する。この手法においては、単一の記述文の構成を解析するための文の構文パターン、複数文の関係を解析するための複数文の関係のパターン、モデルの制御フローを解析するためのモデルのパターンの3つのパターンを用いることで、自然言語からモデルの生成を実現している。

Model Generation from Document using Pattern Matching

Manabu Toyota[†], Satoshi Haraguchi[†], Tetsuya Mikami[†], Yoshifumi Sakamoto[†]

Embedded system development has a problem of discrepancies between requirements documents written by natural language and models as technical work products in detail development phase. We propose a method to generate a model from a document written by natural language. In this method, we use 3 kinds of pattern matching, structure of a sentence, relationship of sentences, and structure of control system model.

1. はじめに

近年、組込みシステムは消費者の多様な要求を満たすために多機能化・高性能化が進んでいる。このような環境下において、組込みソフトウェアの開発を効率的に進めるための1つの手法として、モデル駆動開発に注目が集まり適用事例も多く報告されている。

モデル駆動開発においては、モデリング言語により全ての開発成果物を記述するのではなく、従来の開発手法と同様に、開発の上流工程では自然言語により記述された開発文書を作成し、段階的にモデルによる設計へと進めることが一般的である。実際の組込みソフトウェアの開発では、開発途中での要求や仕様の変更、また過去のソフトウェア資産に変更を加える差分開発の適用が何世代にもわたって繰り返される中で、ソースコードやソースコード生成の基になるモデルは更新される一方、開発文書の更新は行われず、開発文書とモデルやソースコードとの乖離が発生するケースが多い[1]。この一因として、開発プロセスにおいて要求文書に代表

される開発文書が、コミュニケーションの目的に限定され、開発者は開発文書を参照しながら別途モデルを作成しているという状況が挙げられる。

本稿では、この問題を解決する1つの手法として、開発文書の1つである要求仕様からモデルを生成する手法を提案する。自然言語で記述された要求仕様に対して3種類のパターンを適用することによって自然言語からモデル作成に必要な情報を抽出し、モデルを生成する。開発文書を次工程において再利用してモデルを生成することで、モデルやコードとの乖離を抑制し、開発中の要件の変更や、差分開発の適用において整合のとれた開発成果物群を維持することが可能となる。これらによって組込みソフトウェアの品質を高めることに貢献する事が可能であると考えられる。

2. モデル生成の手法

本稿で対象とするモデルとは、組込み機器の制御システムにおける入出力と処理内容を記述した制御フローモデルであり、具体的には **Simulink** で記述されるモデルである。また、開発文書は日本語による要求仕

[†]日本アイ・ビー・エム株式会社

様である。要求仕様は複数の文から構成され、主語や目的語などの情報欠落や係り受けの曖昧性が含まれることを前提とする。本稿では、自然言語で記述した開発文書からモデルを生成する手法を提案する(図 1)。

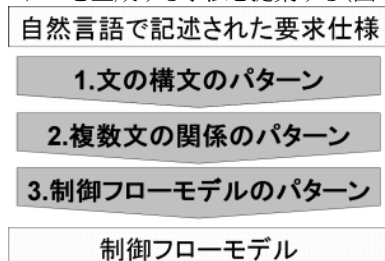


図 1. モデル生成のフロー

この手法では、自然言語により記述した要求仕様の解析に2つのパターン、モデル生成に1つのパターンを用いる。自然言語により要求仕様を記述する場合には、情報の欠落、係り受けの曖昧性などが問題になる。また、一般に一つのモデルは複数の仕様記述文と対応することから、文と文の関係を分析することが必要となる。これらの課題の解決にパターンを適用する。次にその適用のステップを示す。

ステップ1. 要求仕様からの入出力情報の抽出

制御フローモデルの生成には、自然言語で記述された要求仕様から、入力・処理・出力の情報を抽出する必要がある。これらの情報を明示的に要求仕様に記述することを目的として要求仕様を EARS テンプレートに従って記述する[2][3]。EARS テンプレートを適用するためには曖昧な記述を検出して修正する必要がある。このため、形態素解析技術を応用して、文脈から欠落情報の候補を提示することで、欠落情報を補完する。

ステップ2. 複数文の関係の解析

一つの制御フローモデルは複数文に対応することが多いことから、モデルを生成するためには文と文の関係を解析して、モデルの制御フローを明確化する必要がある。そこで複数文の関係を示す 5 つのパターンを考案し、制御フローとの関係を定義した。

1. 並列(調停)パターン
共通の処理と出力を持つ文の集合と定義。条件毎に異なる値の処理の記述等に用いる。
2. 並列(別処理)パターン
共通の入力を持つ文の集合と定義。一つの入力を複数の処理に利用する場合等に用いる。
3. 順序パターン

一方の出力と他方の入力共通の文の集合と定義。一方の処理結果を他方の処理に使用する場合等に用いる。

4. 換言パターン
共通の入力・処理・出力を持つ文の集合と定義。同一の仕様とみなす。
5. 階層パターン
仕様の概要と詳細の関係にある文と定義。概要を詳細仕様の説明文として使用。

ステップ3. システム固有のモデルパターンの適用

頻繁に利用される制御フローについて、パターン化したモデルへの適合により、モデル生成の精度を向上することが望める。そこで、対象とするシステムの既存のモデルを分析し、条件比較や最大値算出などの繰り返し使用可能な12の制御フローをシステム固有のパターンとして定義した。

上述の3ステップのパターン・マッチングにより、自然言語で記述した要求からモデルを生成する。

3. 本手法の適用結果

既存の組込みシステムの要求仕様書群に、提案する手法を適用した。対象とした要求仕様の文の数は約400である。期待する制御フローモデルを生成できた割合は 85%である。期待のモデルを生成できなかった15%を分析したところ、要求記述に情報の欠落が多く、文書の校正によって情報の補完が不可能であることが原因である。

4. 議論

本稿では、パターンを利用して自然言語により記述した開発文書からモデルを生成する手法を提案し、実際の開発文書と制御フローモデルに適用することで、高い効果が得られることを確認した。今後の課題として、より広い領域への適用や、高い抽象度表現への対応を挙げる。

参考文献

- [1] 原口智史, 坂本佳史, 三上徹也, 組込みソフトウェアにおける差分開発の問題調査とモデル駆動開発の採用による解決の提案, WWS2013.
- [2] Mavin, A., Wilkinson, P., Harwood, A., Novak, M., EARS(Easy Approach to Requirements Syntax), IEEE IREC, 2009.
- [3] 山本 修一郎, 要求工学基礎知識, 2012.