

部品合成による自動プログラミングシステム

～要求文の補完方法について～

Fang Miao[†] 清田昌宏[†] 山崎雄大[†] 上之蘭和宏[‡] 古宮誠一[†]

芝浦工業大学[†] 芝浦工業大学大学院[‡]

1. はじめに

近年、ソフトウェア開発の大規模化・複雑化が進んでいる一方で、ソフトウェアの高品質・高生産性の向上の要求が高まっている。特に、今日では短納期でのソフトウェア開発が必要な状況にあり、超短納期でソフトウェア開発を行うアジャイル等に期待が寄せられている。

本研究では部品合成による自動プログラミングシステムを、格フレーム[1]を用いた自然言語処理によって要求仕様を獲得する過程（要求獲得系）と部品合成によってプログラムを自動生成する過程（自動生成系）の2段階で実現する。そして、システムが要求を受け取り、要求通りのソフトウェアを生成することによって、より高い品質・生産性のソフトウェア開発を目指す。

本システムは要求定義者によって、「どのようなソフトウェアを作るか(=what)」の情報が書かれた要求文を入力とする。システムはこの要求文から「どのように実現するか(=how to)」の情報、即ちソースコードに置き換える。

人間が要求仕様を記述するときには、頻繁に語の省略が起こる。しかし要求定義者によって本システムに与えられる要求文は、「how to」の情報に置き換えるために必要な情報を全て持っている必要がある。そのために、要求文に欠落している情報を下記のいずれかの手段で補完する。

要求文において、文の意味要素に省略があれば、その箇所を検出し、システムが自動的に補うか、または欠落している情報の入力を要求定義者に要求する。

自動生成系がプログラム生成の際に必要な情報の中で欠落しているものを検出し、欠落している情報の入力を要求定義者に要求する。

本稿では、自然言語処理技術を用いて の補完について述べる。なお、 の補完については、自動生成系で用意する部品と、部品を組み立てる情報の知識などから要求される。

[†]"An Automatic Programming System by Composition of Reusable Program Components,"

[†]Fang Miao, Masahiro Kiyota, Yudai Yamazaki, Seiichi Komiya, Shibaura Institute Of Technology

[‡]Kazuhiro Uenosono, Graduate School Of Shibaura Institute Of Technology

2. 要求獲得処理の流れ

本システムは自然言語処理技術を用い、以下に示す流れで要求獲得系の処理を行う。

自然言語で書かれた要求文に対して cabocha[2]を用いて構文解析を行う。

構文解析で得られた係り受け情報を基に格フレームを生成し、初期状態の格フレームネットワークを構築する。

自動補完と質問による補完の2つの手法を用いて、要求文の格要素の全ての欠落を補完し、格フレームネットワークを更新する。

完全な要求文を自動生成系への中間表現に変換する。

3. 要求文の処理の事例

要求文の欠落を判断するには、文の意味を理解する必要がある。しかし、機械は自然言語をそのままの形では理解することができない。そこで格フレームを用いる。格フレームは格文法[1]に従って、文中の語について、述語に着目し、どのような意味・役割、即ち格を持つかを示すことで、文の意味を表現するものである。Fillmore によって唱えられた初期の格は8種類であるが[1]、本システムでは経験者格は不要と考え、除外する。また、本システムが補完処理を行うために6種類の格を追加する。これらの格と例を表1に示す。システムは文から格フレームを生成することで、文がどのような格構造を持っているか、またどのような格要素が不足しているかがわかる。この特徴を用いて、要求文の欠落の補完を行う。

本システムでは、係り受け情報に対して、上記のように本システム専用で格が定義された格フレーム辞書を用いて格フレームを生成し、格要素の欠落を判断する。欠落要素があれば格要素の補完を行う。補完の手段として以下の2つを用いて完全な要求文を作る。

自動補完：自動で補完できる場所を明確にし、格フレームネットワークを利用して補完を行う。

質問による補完：要求定義者に欠落箇所に埋めるべき要素を、質問し回答を得ることで補

完を行う。

なお、自動補完に用いる技術は照応省略解析手法[3]である。

以下にホテル予約システムの要求文においての例を示す。なお、サ変動詞「する」は直前のサ変接続名詞「入力」とで一つの動詞であると見なしている。

- ・ ユーザは予約入力情報を入力する。

この文に対し構文解析と格解析を行い、図1のような結果が得られた。*で始まる行は文節の情報を表し、行末に格属性が付加している。

例の場合、対象となる文の格フレームを生成する。この例では動詞「入力する」は動作主格、道具格、対象格の3つの格が必要であることがわかる。

システムはまず自動補完を行う。説明は省略するが、照応省略解析手法を使い、格フレームネットワークを利用し、「ユーザ」が「ホテル予約システムのユーザ」であることがわかり、

「ユーザ」の所有格を補完する。次に、質問による補完を行う。この例では、動作主格と対象格はそれぞれ「ユーザ」、「予約入力情報」として要求文中に現れるが、道具格は欠落している。したがって、システムは要求定義者に「ユーザはなにで予約入力情報を入力するか？」を質問し、要求定義者から「キーボードで」という回答をもらって道具格を補完し、図2になる。

* 0	2D	0/1	1.21603566	動作主格				
ユーザ	ユーザ			ユーザ	名詞-一般			0
は	ハ			は	助詞-係助詞			0
* 1	2D	2/3	0.00000000	対象格				
予約	ヨヤク			予約	名詞-サ変接続			0
入力	ニュウリョク			入力	名詞-サ変接続			0
情報	ジョウホウ			情報	名詞-一般			0
を	ヲ			を	助詞-格助詞-一般			0
* 2	-10	1/1	0.00000000	述語				
入力	ニュウリョク			入力	名詞-サ変接続			0
する	スル			する	動詞-自立	サ変・スル基本形		0
.	.			.		記号-句点		0
EOS								

図1. 構文解析と格付けの結果

入力する	
動作主格	- ユーザ
道具格	- キーボード
対象格	- 予約入力情報

図2. 「入力する」の格フレーム

表1. 要求獲得系で用いる格と例

動作主格(Agent)	動作を引き起こすもの。
例	ユーザは人数を入力する。
道具格(Instrument)	出来事の原因となるもの。
例	ユーザはキーボードで入力する。
対象格(Object)	移動や変化する対象物。
例	ユーザは人数を入力する。
源泉格(Source)	移動や変化における起点。
例	ユーザは待機状態から選択する。
目標格(Goal)	移動や変化における終点。
例	ユーザは予約システムに入力する。
場所格(Location)	出来事が起こる場所。
例	システムは予約帳データベースにあるかを確認する。
時間格(Time)	出来事が起こる時間。
例	システムは1時間内に入力を受けなかった場合、終了する。
所有格(Possessor)	所有・所属の関係を表すもの。
例	ホテル予約システムのユーザはチェックイン日時を入力する。
並列格(Parallel)	2つ以上同じ趣の言葉を並べるもの。
例	予約入力情報にはチェックイン日時、チェックアウト日時がある。
数量格(Quantity)	数量を表すもの。
例	ユーザはもう一度入力する。
条件格(Condition)	出来事の成り立つあるいは実現に必要な事柄を表すもの。
例	空室がある場合は、ユーザは顧客情報を入力する。
同格(Apposition)	同一の文法上の機能を果たし、並び置かれた2つ以上の語
例	ラーメンなど麺類が好きだ。
原因格(Cause)	ある出来事や状態を引き起こしたもとなった事・出来事。
例	在庫数が発注や返品で増減する。

4. まとめ

本稿では、自動プログラミングシステムの要求獲得の処理を、自然言語処理技術を用いて実現することを提案し、要求獲得処理の流れを示した。また、具体例を用いながら要求文の格補完処理を示した。

なお、要求獲得系の入力となる要求文には「要求の誤りがない」ことを前提とし、本研究では「要求の漏れがない」ような、自動生成系に渡す中間表現を生成することを目標としている。今後、誤りの検出について考察する。

参考文献

- [1] 田中穂積, 自然言語処理 - 基礎と応用 -, 電子情報通信学会(1999)
- [2] 工藤拓, 松本裕治, “チャンキングの段階適用による日本語係り受け解析,” 情報処理学会論文誌, vol43, no.6, pp.1834-1842, 2002
- [3] 村田真樹, 長尾真, “表層表現と用例を用いた照応省略解析手法,” 電子情報通信学会信学技報, NLC97-56, 1998-03