

大規模情報システムの要求仕様書の自動要約手法の提案

-背景や目的部分に対するマルコフ連鎖を用いた抽出型要約の適用-

石井 達樹[†] 阿部 養太[†] 大工原 正哲[†] 中島 千壽[†] 位野木 万里[†]
工学院大学[†]

1. はじめに

近年の情報システムの高度化に伴い、要求仕様書の記述も膨大になっている。膨大な量の要求仕様書を開発者が短期間のうちに理解するのは困難である。

著者らの研究グループでは、要求定義の高品質化の研究に取り組んでおり[1]、その一環としてデジタルトランスフォーメーションの実現を目的としたレガシーシステムの複数の視点からアプローチを行い、要求仕様書の内容理解を効率化する手法の開発を目指している。

文書理解の効率化手法としては、自然言語処理技術による自動要約が一般的である。しかし要求仕様書は、開発背景、ゴール、調達条件、機能要求、非機能要求などの様々な要素から構成されており、一つの要約技術の適用のみでは、妥当な要約を自動で生成することは困難である。本研究では、要求仕様書に適した要約を定義し、自動要約実現のための手法を提案する。

要求仕様書の自動要約を実現するための機能の一つとして、本研究ではシステム化の背景や目的部分の自動要約を試行する。自動要約にはマルコフ連鎖を用いた文章自動生成ツールである TextGenerator[2]を利用し、要約結果の妥当性と運用可能性について考察する。

以下、本稿は次のように構成する。2章では、要求仕様書の要約の定義と自動要約の手法を示す。3章では、背景や目的部分の自動要約を実現する手法を提案する。4章では提案する手法の適用評価を行う。5章では提案手法の考察を述べる。6章にて本稿のまとめを示す。

2. 要求仕様書の要約定義と自動要約の実現

2.1 マルコフ連鎖を用いた自動要約技術

マルコフ連鎖は文章自動生成に用いられる。元の文章を形態素単位で分解し順に並べた順序列を用意する。形態素を1つの状態と捉え次の状態への遷移を確立に基づき決定する。この遷移を文末の形態素にたどり着くまで繰り返すことによって文生成を行う。実際の文生成には直前の状態1つではなく、4個の形態素を考慮した高階マルコフ連鎖を使うことが多い[3]。元の文章の要素を抜き出す形で文章生成であるため、自動要約技術としても用いられる。

2.2 要求仕様書の自動要約の実現

著者らの研究グループでは、要求仕様書の要約を以下のように定義する。要求仕様書を、その特性を考慮して3つの構成要素に分割し、それぞれの視点から、個別に最適化

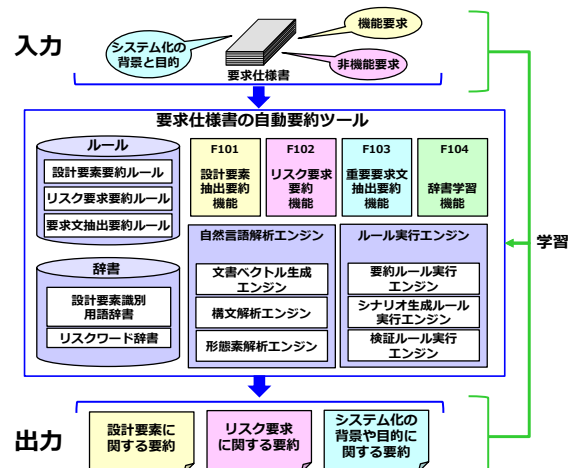


図1 要求仕様書の自動要約の実現

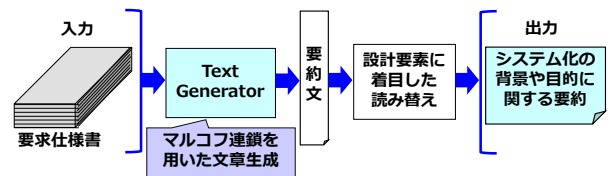


図2 TextGeneratorを用いた要約の実現

した手法を用いて要約を行い、それぞれの出力を融合させた結果を、最終的な要求仕様書の要約として定義する。要求仕様書の3つの構成要素に対する要約を、それぞれに最適化した異なる手法による自動要約ツールを適用することで、要求仕様書の自動要約を実現する。

3. 背景及び目的部分の自動要約

本研究では要求仕様書の自動要約を実現する三つの構成要素のうち、背景や目的部分の自動要約のためTextGeneratorを用いた要約を提案する。TextGeneratorは、GitHub上で公開されているオープンソースソフトウェアであり、形態素解析エンジン MeCab[4]を用いて文章の形態素解析を行い、高階マルコフ連鎖を用いて元の文章から任意の文章数の要約文を生成する。

マルコフ連鎖を用いた要約の性質上、日本語として不自然な表現や冗長な繰り返しが発生する。要約結果としての有効性を確保するために、出力された要約文を元に設計要素に着目した読み替えを行い、代表的なアクター、取り扱うデータ、運用に際し有する機能を抽出することで、要約としての最終的な出力とする。

4. 適用評価

4.1 評価概要

本研究では、前項で提案したシステム化の背景と目的部

Proposal of Automatic Summarization Method for Requirements Specification of Large-scale Information Systems -Application of Extractive Summarization Based on Markov Chain to Descriptions of Background and Objective-
[†]Tatsuki Ishii, Yota Abe, Masanori Daikuhara, Senjyu Nakajima, Mari Inoki, Kogakuin University

S1	S2	S3
<ul style="list-style-type: none"> さらに、住民票を有するすべての人に、重複することのない一意の番号を漏れなく付番し、国民にとっての利便性、行政事務の効率性・正確性、負担と給付の公平性を確保し、個人情報保護に配慮しつつ幅広い行政分野において情報連携を行う仕組みを築くことを目的とした社会保障・税に関わる番号制度の導入に伴い、厚生労働省及び機構では、公的年金業務における円滑な番号制度への対応を行うこととしている。 これらの業務及びシステムについては、「公的年金業務の業務・システム最適化計画」に基づき、最適化を実現することとしている。 厚生労働省及び日本年金機構は、社会保険オンラインシステムを利用して、国民年金、厚生年金保険等に係る事業所及び被保険者に対する適用、各種保険料の徴収、年金給付等の各種給付並びにこれらに関連する相談等の公的年金業務を行っている。 個人番号と基礎年金番号を紐付けて管理し、被保険者記録及び年金受給者記録に関する相談・照会業務において、個人番号の利用及び関係機関との情報提供の開始に向け、次の対応を行うこととしている。 個人番号と基礎年金番号を相互に変換する機能を構築し、届出、通知業務への個人番号の利用を可能とする。 	<ul style="list-style-type: none"> 厚生労働省は、「監督・安全衛生等業務及び労災保険給付業務のシステム最適化を進め、平成21年3月には第1期開発分として平成22年12月に基準システムが稼働し、平成26年1月より全国稼働している。 本仕様書は、平成28年度に予定される監督・安全衛生等業務の業務・システム最適化計画)及び「労災保険給付業務の業務・システム最適化計画)及び「労災保険給付業務に係る制度改正等に伴う基準システムのアプリケーションプログラムの改修業務の調達を目的とし、その仕様を記載するものである。 上記に加えて、労災保険給付業務に係る制度改正等に伴う基準システムのアプリケーションプログラムの改修を実施している。 また、労災業務における診療報酬明細書を電子的に受け付け、審査を行うためのシステムとして、労災セプト電算処理システムを構築し、平成26年1月より全国稼働している。 本仕様書は、平成28年度に予定される監督・安全衛生等業務に係る制度改正等に伴い、労働基準行政システムのアプリケーションプログラムの改修業務の調達を目的とし、その仕様を記載するものである。 	<ul style="list-style-type: none"> 労働保険適用徴収業務の業務・システム最適化計画)に基づき、再構築を行ったシステムが平成25年1月より稼働しており、次期の機器更改は平成29年1月に移行を予定している。 現在は、「労働保険適用徴収業務は、労働保険料の収納に係る事務等のために大量のデータを処理しており、次期の機器更改は平成29年1月に移行を予定している。 本仕様書は、高い運用能力を有するシステム運用事業者を調達し、適用徴収システムの安定的な稼働を実現することを目的とする。 現在は、「労働保険適用徴収システムは、当該業務を支援する目的から1972年より運用しており、2003年10月からは電子申請・届出に対応し、2004年1月から労働保険料の収納に係る事務等のために大量のデータを処理しており、次期の機器更改は平成29年1月を予定している。 近年、情報システムの構築について、「世界最先端IT国家創造宣言」に基づく「政府情報システム改革ロードマップ」に基づき適用徴収システムにおいても事業場公開機能の政府共通プラットフォームへの移行準備を進めており、平成29年1月に移行を予定している。

図3 背景や目的部分の自動要約結果の例

分に着目した要約を、TextGenerator を用いて実現する手法の有効性について評価する。実際の要求仕様書を用いて以下のように適用評価を行った。

- ・目的: 要求仕様書の内、システム化の背景や目的に関する部分についての要約の有効性の評価。
- ・対象: 以下S1~S3の仕様書に対して適用を行った。
S1: 年金業務システム [5]
S2: 労働基準行政情報システム [6]
S3: 労働保険適用徴収システム [7]
- ・手順: TextGenerator に対象の要求仕様書を読み込ませ、それぞれ5文の要約文を出力する。出力された文に対し、設計要素に着目し手動で読み替えを行う。

4. 2 結果

自動生成による要約文の出力内容については、図3に示す。文章としては成立しているが、意味の通らない文章や同じ情報の冗長な繰り返しが発生しており、要約結果として有効とはいえない。

そこで、要約文に対して設計要素に着目して読み替える。ここでは紙面の都合上、代表としてS1の読み替えた結果の一例を以下に挙げる。

「S1のシステムにおける代表的なアクターは厚生労働省、日本年金機構であり、取り扱うデータは個人番号、基礎年金番号及び被保険者記録、年金受給者記録である。前者に関しては紐づけて管理し相互に変換が可能であり、後者に関しては個人番号を用いて記録の照会が可能である。」

5. 考察

提案手法の適用評価を行った結果から、要求仕様書の自動要約における背景や目的部分に着目した要約の適用結果として、TextGenerator を用いた要約は対象システムの代表的なアクター、取り扱うデータ、運用に際し有する機能を把握するために有効であると考えられる。しかし、設計要素に着目した読み替えは手動で行っているため、これも含めた自動化を実現することで、自動要約としての運用がより容易になると考えられる。

6. まとめ

本稿では要求仕様書の自動要約の実現のために、構成要素のうちの一つであるシステム化の背景と目的部分についてマルコフ連鎖を用いた自動要約手法を提案し、読み替えを行うことで有効な情報を抽出できることを確認した。

今後は、本稿では手動で行った自動要約の結果に対する読み替え作業を自動化することで、自動要約としての精度を高めていく。

謝辞

要求仕様の一貫性検証支援ツール開発に関わる研究は、独立行政法人情報処理推進機構技術本部ソフトウェア高信頼化センター (SEC: Software Reliability Enhancement Center) が実施した「2015年度ソフトウェア工学分野の先導的研究支援事業」の支援を受けたものである。また、本研究はJSPS 科研費 JP16K00105, JP19K11907 の助成を受けた。

参考文献

- [1] 位野木 万里, 近藤 公久, 省略と修飾パターンを用いた用語不一致検証による要求仕様の一貫性検証支援ツールの実現と適用評価, 日本ソフトウェア科学会, コンピュータソフトウェア, Vol. 35, No. 3, pp. 109-127, 2018
- [2] TextGenerator, GitHub リポジトリ
<https://github.com/ohshige15/TextGenerator> (参照 2020-01-09)
- [3] 文章自動生成手法の比較考察-文章生成における文と文のつながりの課題と方策-, 太田 博三, 放送大学教養学部, 2017
- [4] MeCab, Yet Another Part-of-Speech and Morphological Analyzer
<https://taku910.github.io/mecab/> (参照 2020-01-09)
- [5] 年金業務システム (個人番号管理サブシステム等 (2次開発情報連携分)) に係る設計・開発等業務及びアプリケーションソフトウェア保守業務, 厚生労働省
<https://www.mhlw.go.jp/sinsei/shotatu/shotatu/shiyousho-an/160428-1.html> (参照 2020-01-09)
- [6] 労働基準行政情報システムに係るアプリケーションプログラム改修等業務一式 (平成28年度) 調達仕様書 (案), 厚生労働省
<https://www.mhlw.go.jp/sinsei/shotatu/shotatu/shiyousho-an/151204-1.html> (参照 2020-01-09)
- [7] 労働保険適用徴収システムに係るシステム運用業務一式調達仕様書 (案), 厚生労働省
<https://www.mhlw.go.jp/sinsei/shotatu/shotatu/shiyousho-an/160413-1.html> (参照 2020-01-09)