

大規模情報システムの要求仕様書における自動要約技術の適用 -自動要約結果と先入観との差異を活用した理解支援-

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

1. はじめに

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み[1]、その一環として、要求仕様書の自動要約結果を用いた理解支援を研究している。著者らは、自然言語処理技術を用いた要約技術[2]の中から、抽出型の自動要約手法に着目した。抽出型の要約手法は、入力した文章の記述を活用して要約文とするため、要求仕様書の要約には適していると考えられる。

著者らは、抽出型の自動要約技術を実際の要求仕様書に適用したところ、次のような課題に直面した。その課題とは、読み手である技術者が対象となる要求仕様書に対して先入観があると、要約文が「間違っただけ」要約と判断され、当該要約結果を要求仕様書の理解支援として活用しないリスクが存在することである。

あらかじめ先入観と要約結果にギャップがあることを技術者が自覚していれば、そのギャップの自覚を活用して、生成した要約文を対象となる要求仕様書の背景情報の理解に役立てられると期待できる。そこで、上述の課題の解決のため、技術者の先入観と生成された要約文のギャップを活用して、初級技術者の要求仕様書の早期理解を支援することを提案する。

以下、本稿は次のように構成する。2章では、要約結果と先入観の受け取り方の違いについて説明する。3章では自動要約結果を用いた実験方法を示す。4章では、3章の結果を示す。5章では、4章の結果から要約を要求仕様書の理解にいかす方法を提案する。6章で本稿をまとめる。

2. 要約文と先入観のギャップと理解支援

図1を用いて、技術者による要約文の受け取り方の違いについて説明する。図1(A)に先入観と要約文とのギャップを自覚していないケースを示す。技術者が先入観と要約文のギャップを自覚していない場合、要約結果が技術者の想定していた内容と異なった時、要約文のシナリオをそのまま解釈すると要約文が間違っていると判断し、内容理解に役立てることができない。

図1(B)に技術者がギャップがあることを自覚しているケースを示す。要約文が技術者が想定していた内容と異なった時、技術者の理解に先入観があることを自覚することで、要約文に含まれるアクターやデータに着目した要求仕様書の理解支援に役立てることができる。

3. 自動要約結果を用いた実験

本研究では技術者によるギャップがある場合とそうでない場合に、要約結果が、どのように、要求仕様書の理解に影響するのかを確認するため以下の2つの実験を行う。

An Application of Automatic Summarization Technology to Requirements Specification of Large-scale Information Systems-Support for Understanding on the Basis of Differences between Generated Summary and Engineers' Preconception-

[†]Yota Abe, Tatsuki Ishii, Masaaki Daikuhara, Senju Nakajima, Mari Inoki, Kogakuin University

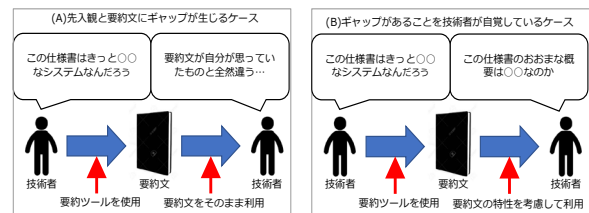


図1：要約文の受け取り方の違い

・実験 1

目的：

概要を把握している文書の要約文から、先入観と要約文にどのようなギャップがあるかを分析し、ギャップの活用方法を考察する。

対象：対象は以下の①～③とする。

①桃太郎[3] (著 芥川龍之介 4, 384 文字)

②桃太郎[4] (著 石塚浩之 14, 309 文字)

③桃太郎[5] (著 楠山正雄 5, 354 文字)

手順：

- (1) 3人の作者によるおとぎ話「桃太郎」を自動要約ツールに適用し、文書ごとに要約文を5文抽出する。要約文の抽出には、抽出型の自動要約ツール User Local テキストマイニングツール[6]を用いる。
- (2) 要約文と元の文書を入力として要約文に出現したアクター(図2(a))と元の文書に出現するアクター(図2(b))を抽出する。アクターの抽出には著者らの研究グループが開発した要求仕様の一貫性検証ツール[1]を用いる。
- (3) (1)(2)で抽出した要約文とアクター用語を分析して、先入観と要約結果にどのようなギャップがあるか明らかにし、ギャップの活用方法を考察する。

・実験 2

目的：

自動要約ツールを要求仕様書に適用し、実験1の考察を元に、先入観と要約文のギャップを活用して要求仕様書の理解を支援できることを示す。

対象：④エンタープライズ系システムの要求仕様書[7] (25, 059 文字)

手順：

- (1) 要求仕様書の本文をテキストマイニングツールを適用し、要約文を5文抽出する。
- (2) 要約文に出現したアクター(図3(a))と元の文書に出現するアクター(図3(b))を抽出する。
- (3) 実験1(3)の考察を元に要約文と先入観のギャップを用いてある一定の理解を得られることを示す。

抽出した要約文	【①の要約結果】 一万年に一度結んだ実は一千年の間は地へ落ちない。実は椰子の殻えたり、 極楽鳥 の囀ったりする、美しい天然の楽土だった。いや、 鬼 というものは元来我々人間よりも享乐的に出来上がった種族らしい。これでもまだわからないといえは、貴様たちも皆殺してしまおうぞ。」 犬も桃太郎 の洗面を見ると、口惜しそうにいつも唸ったものである。	【②の要約結果】 鬼ヶ島に行くとの噂を聞きつけ、道中で同伴を願い出たこの 鬼 は自ら 猿 と名乗り、名前を尋ねても、皆に 鬼 と呼ばれているからそれではいなどと割り切ったふうな口を利いていたが、おそらくもともと名前などないのではあるまいか。それは自覚しているものの、我ながら調子にのりすぎていたなどと今さら反省してみても後に引き返すことなどできぬところまでできてしまっており、考えれば考えるほど 勝負 などあるのかどうか不安になるばかりである。この 主 を捕まえたのは、何日続くのかもしれぬ海路で口にできるものはもともと十分ではなかったところへ、涙がこぼれるくらいありがたく心強い 猿 と 犬 の援軍があり、来るべき食糧不足の際の食料とするためなのである。けれどもそれを妄想というならばすべてが 村人 たちの妄想で、彼らが 鬼 に襲われたことなど一度もなく、実はその 鬼 すらもどこにもおらず、ただ自分たちの貧しさをありませぬ暴力への被害意識とすりかえただけなのではないか。村に帰りそのように告げてやれば、いわれのない凄然とした畏れから生まれてきた迷妄などは直ちに払拭され、ついでにおのれは 鬼 を退治した英雄として歓迎を受け、村長のあの美しいひとり 鬼 を 壺 にとり、安泰な暮らしを送れるのではない。	【③の要約結果】 と言って、ていねいに顔を下げました。「わたくしは、満き手になりましょう。こう言って、 壺 がかしに座りました。その時 犬 は門の前に立って、うちでは おじいさん と、 おばあさん が、かわるがわる、
	a 極楽鳥, 鬼, 犬, 桃太郎	男, 猿, キジ, 犬, 鬼, 村人たち, 娘, 嫁	猿, 犬, おじいさん, おばあさん
b 極楽鳥, 鬼, 犬, 桃太郎, 猿, 人間, 雉, 鬼の酋長, 家来, 鬼の子供, 他27件	男, 猿, キジ, 犬, 鬼, 村人たち, 娘, 嫁, 人間, 老夫婦, 他32件	猿, 犬, おじいさん, おばあさん, 桃太郎, 鬼, きじ, 子, 家来, 鬼の大将, 他10件	

図2: 「桃太郎」の自動要約結果と出現アクター

抽出した要約文	【④の要約結果】 日本産業規格A列3番を使用することができる。を実施する都度、書面にて報告すること。当性を検討し、報告すること。は、 職業安定局 の承認を得て実施すること。)に再委託することはできない。
	a 職業安定局
b 職業安定局, システム, 事業者, 受託者, サブシステム, 関連事業者, 求職者, 職員, 機関, 労働局, 他59件	

図3: 要求仕様書の自動要約結果と出現アクター

4. 実験結果

図2に「桃太郎」の自動要約結果と出現アクターを示す。桃太郎を「鬼退治に向かう」物語という先入観があると仮定すると、要約結果からは鬼退治について読み取ることが困難であった。これは、自動要約ツール利用者の先入観からツールに求める結果とツールがツール内のルールに基づいて出力した結果にギャップがあるからと考えられる。

ツール利用者の想定している内容と異なる要約結果が出現した時、同利用者があらかじめ先入観があることを自覚していれば、要約結果そのものは利用せず、同結果に含まれる設計要素のみに着目して元のテキストの内容理解に役立てるように誘導することが有効であると考えられる。例えば、図2①を例として、要約結果中のアクター「桃太郎」に着目し元の文章を見直すことで「桃太郎」の心情や行動について理解することが可能である。

図3に要求仕様書の自動要約結果と出現アクターを示す。図3の要約結果から全体の概要を把握するのは困難であったが、要約結果に含まれていたアクター「職業安定局」に着目して要求仕様書を確認した結果「職業安定局」が要求仕様書中の重要なアクターであることが確認できた。

5. 要約結果と先入観を活用した理解支援

実験1, 2に示したように、先入観と要約結果のギャップを活用して要求仕様書の理解支援に役立つことが期待できる。すなわち、要約結果をそのまま解釈するのではなく、設計要素別に捉えることで、要求仕様書中の中心となる要素の理解支援が可能である。一方で設計要素のみを抽

出し理解支援に役立てることも可能だが、要約文は出現頻度の高い設計要素に共起する用語も同時に特定できるため、長文の要求仕様書の場合には、コンパクトな要約結果を用いて、内容の初期の理解に有効であると考えられる。

6. まとめ

本稿では自動要約ツールを用いた要約結果と先入観のギャップを活用する方法を明らかにする実験を行った。当該実験により、自動要約ツールの利用者があらかじめ先入観があることを自覚し、要約文に含まれる設計要素に着目することで要求仕様書の内容理解に役立てることができると示した。今後は要求仕様書のより深い内容理解のために要約文の妥当性の検証方法を提案することが課題である。

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進的研究支援事業」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]奥村, 難波英嗣, テキスト自動要約に関する最近の話題向, J-STAGE, 9巻, 4号, pp. 97-116, 2002
- [3]芥川龍之介, 青空文庫, 桃太郎 (以下 URL 参照 2019-12-09)
<https://www.aozora.gr.jp/cards/000879/card100.html>
- [4]石塚浩之, 青空文庫, 桃太郎 (以下 URL 参照 2019-12-09)
<https://www.aozora.gr.jp/cards/001813/card56878html>
- [5]楠山正雄, 青空文庫, 桃太郎 (以下 URL 参照 2019-12-09)
<https://www.aozora.gr.jp/cards/000329/card18376.html>
- [6]User Local テキストマイニングツール
[https://textmining.userlocal.jp\(参照 2019-12-24\)](https://textmining.userlocal.jp(参照 2019-12-24))
- [7]厚生労働省, ハローワークシステム等に係るシステム部門支援業務一式 (令和2年度以降) 調達仕様書 (案) (参照 2020-12-03)
<https://www.mhlw.go.jp/sinsei/chotatu/chotatu/shiyousho-an/190827-1.html>