

要求仕様書における話題の記述状況の可視化 -BERTを用いた Page Trend 算出手法の提案-

中村 雄太郎[†] 長岡 武志[‡] 北川 貴之[‡] 位野木 万里[†]

工学院大学[†]

東芝デジタルソリューションズ株式会社[‡]

1. はじめに

Digital Transformation (以下, DX と呼ぶ) の促進により, DX に関連するレガシーシステム関連技術や業務の流れを理解することは重要である. これらのシステムの要求仕様書は大規模かつ複雑なものである場合が多いため, それらの開発に関わっていない技術者が, 短時間のうちに理解するのは困難である.

要求定義のタスクに自然言語処理技術を適用する研究が試みられている[1]. 自然言語処理技術の自動要約を適用することは, 要求仕様書の理解に期待できる[2].

一般的に, 要求仕様書の読者は, 冒頭の目次を参照し, 章節のタイトルからおおよその内容を理解すると考えられる. しかし, 実際の要求仕様書の中には, 章節のタイトルと記載されている内容が異なる場合がある. そのため, 主要な機能要求や非機能要求が要求仕様書のどの位置に記述されているのか特定することは容易ではない.

本稿では, 要求仕様書における重要な機能要求や非機能要求に関する話題を特定し, 仕様書中のどのページに出現しているかをわかりやすく可視化することを試みることで, 初級技術の要求仕様書の理解を支援する.

2. 研究課題と解決策へのアプローチ

要求仕様書では基本的に章節項によって内容の理解を行うが, 実際に要求仕様書では1つの章節項内でも記述された内容が異なる場合が存在する. そのため, 要求仕様書の目次に記述された各章節項のタイトルのみを見て内容を短時間のうちに理解するのは困難であり, 誤った理解になるリスクがある.

このような課題解決のため, 自然言語処理技術を用いて, 要求仕様書の記述状況を可視化する手法を提案する. 具体的には, 要求仕様書から重要な話題を特定し, 要求仕様書のページ毎に当該話題の出現率を算出し, それらを用いて記

述状況を可視化することを提案する.

3. 提案手法

3.1. 話題の可視化手法

図1に提案する要求仕様書の話題の可視化手法の概要を示す. 提案手法では, 要求仕様書を入力すると, 要求文の中から重要な要求の抽出を行い, 抽出した重用要求文と要求仕様書の全要求文に対して類似度の算出を行い, Page Trend の可視化結果を出力する.

本手法において, 要求仕様書は自然言語によって平文で書かれたものと仮定し, 1つの文を1つの要求文であるとす. 可視化する話題は, 以下のルールに従って決定する.

- ① 重要度が高い要求文から選択する
- ② すでに決定した話題と類似する要求文は選択しない
- ③ 章節項等タイトルに該当する要求文は選択しない
- ④ 用語のみの要求文は選択しない

本手法の実装は BERT を用いる. ①での各要求文の重要度算出には重要文抽出アルゴリズム

Pysummarization を利用する[3]. 選択する話題の重複をさけるために②を設定する. 要求文に該当しない, 章節項のタイトルや用語だけの記述の断片を排除するためにそれぞれ, ③と④を設定する.

3.2. Page Trend の算出

各話題のページ毎の出現率を算出するために, ページ内に含まれる話題に関連する要求文の割合を計算する. これによって得られる値を Page Trend として定義する.

式(1)にページ p における話題 t の Page Trend の算出式を示す. 話題と関連する要求文の判定には BERT Score[4]を用い, 話題とのコサイン類似度が閾値を超える文を話題に関連する文と捉える. Page Trend を用いて各ページの話題を可視化する.

$$PageTrend(t, p) = \frac{\text{ページ}p\text{において対応付けられた要求数}}{\text{ページ}p\text{における要求数}} \quad (1)$$

Visualization of description status in requirements documents
- Proposal of a page trend calculation method using BERT -

[†]Yutaro Nakamura, Mari Inoki, Kogakuin University

[‡]Takeshi Nagaoka, Takayuki Kitagawa, Toshiba Digital Solutions Corporation

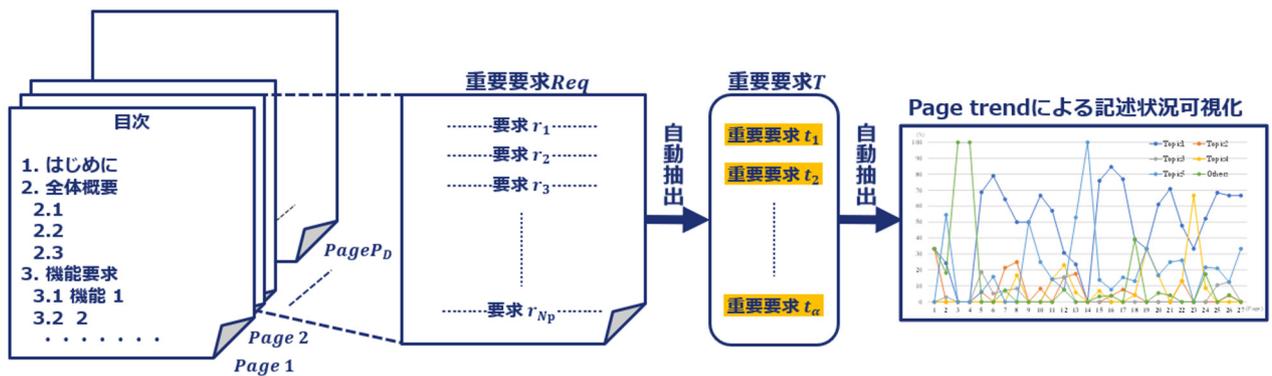


図1 話題の可視化手法の概要

4. ケーススタディ

実際の要求仕様書に対して本手法を適用した。対象は、接触確認アプリ COCOA の要求仕様書である[5]。本手法により生成した話題の可視化結果の例を図2に示す。本手法が特定した話題は、陽性者検知記録に関する本アプリの主たる機能、新型コロナウイルス感染拡大防止、利用者のプライバシー保護、新型コロナウイルス感染症対策の基本対処方針、クラウドサービス事業者が保有する通知サーバであった。加えて、コサイン類似度が話題に対応するか判断する閾値を0.65に設定した。

図2において、出現率70%以上となるページが存在する「プライバシー」に注目する。例えば、実際に仕様書の6,19ページを確認すると、確かにプライバシーに関する非機能要求について記述されており、Page Trendの値と一致していることを確認した。

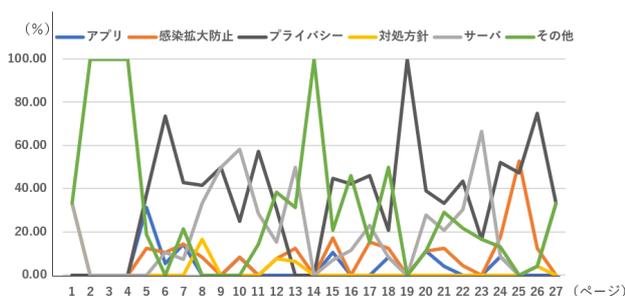


図2 Page Trend による記述状況の可視化結果

5. 考察

ケーススタディの結果から、本研究で提案した話題の出現率である Page Trend を可視化する手法は、COCOA 要求仕様書の内容に対して高い精度で記述状況が可視化されていることを確認した。このような情報は目次情報には表れていない。したがって、可視化された記述状況を活用することで、例えば非機能要求の記述箇所などを特定することが可能となり、要求仕様書の理解に有効であると考えられる。

なお、要求文の記述が少ないページに対しては、出現率が偏った結果になる可能性がある。1 ページ内の要求

文の記述量に影響を受けない記述状況の可視化が必要である。

また、今回のケーススタディでは、決定する話題の数を5つ、話題に対応する閾値を0.65に設定した。要求仕様書の対象ドメインやページ数によって、閾値の設定は変化させる必要がある。今後は、適用評価を繰り返すことで、Page Trendの精度を高めていく。

6. まとめ

本研究では、要求仕様書内の重要文と話題を定義し、ページ毎の話題の出現率である Page Trend による記述状況を可視化することによって要求仕様書の理解支援、または検証する手法を提案した。

今後は、様々な案件の要求仕様書に対する Page Trend の可視化や、可視化結果に対する評価を継続し、本手法の有効性、妥当性等を高めていく。

謝辞

本研究は JSPS 科研費 JP19K11907 の助成を受けた。

参考文献

- [1] Liping Zhao, Waad Alhoshan, Alessio Ferrari, Keletso J. Letsholo, Muideen A. Ajagbe, Erol-Valeriu Chioasca, and Riza T. Batista-Navarro, Natural Language Processing for Requirements Engineering: A Systematic Mapping Study. ACM Comput. Surv. 54, 3, Article 55, 2022.
- [2] Yamamoto S., Suzuki R., Fukusato T., Kataoka H., Morishima S. A case study on user evaluation of scientific publication summarization by Japanese students Applied Sciences (Switzerland) 2021
- [3] Automatic Summarization Library: pysummarization <https://pypi.org/project/pysummarization> (参照 2022-12-26)
- [4] BERT Score github https://github.com/Tiiiger/bert_score (参照 2022-12-20)
- [5] 新型コロナウイルス感染症対策テックチーム, 接触確認アプリ及び関連システム仕様書 <https://cio.go.jp/node/2613> (参照 2022-12-20)