

[2020R2]

要求仕様書等の技術文書の要約・検証・生成自動化手法

日立製作所
株式会社セック

照屋 絵理
平山 哲

開発における問題点

- デジタル化が加速する中、既存システムの仕様把握、作成した要求仕様書の評価・検証などの場面で、技術者が要求仕様書の読み込みや理解をしなければならない機会が増加している。
- しかしながら、要求仕様書の理解は人手に頼らざるを得ず、コストがかさむという問題がある。

手法・ツールの提案による解決

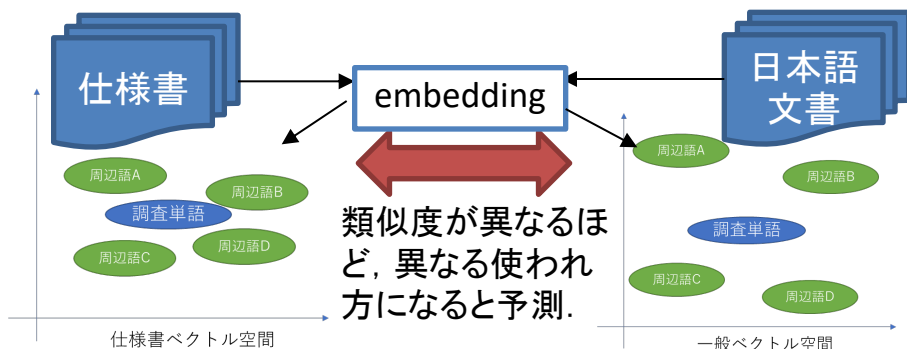
既存の自然言語処理技術を用いて、要求仕様書の重要な部分のみを自動的に抽出し、理解を支援する手法を提案・検証した。
 仕様書の重要な部分として、以下の点に着目した。

- 重要な単語
- 重要な単語に関連する重要な文

重要な単語の抽出

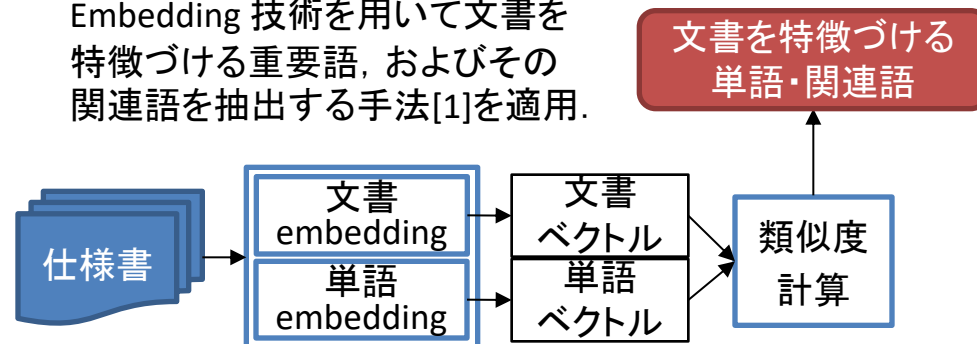
通常とは異なる使われ方をする単語の抽出

要求仕様書と一般的な日本語の文書をEmbeddingし、周辺語の類似度を比較する手法を提案。



仕様書の特徴づける単語の抽出

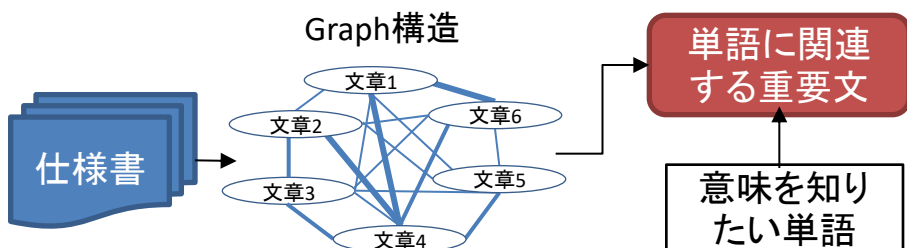
Embedding 技術を用いて文書の特徴づける重要語、およびその関連語を抽出する手法[1]を適用。



単語に関連する重要な文の抽出

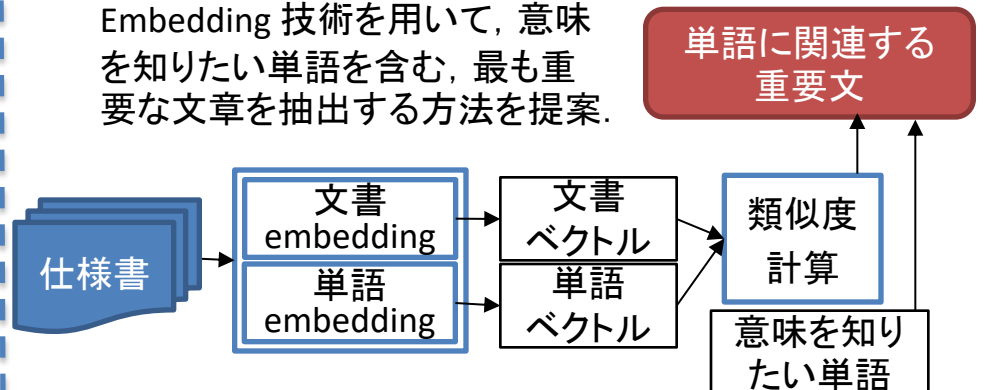
要約技術を用いた重要文抽出

文章の関連度等から構築した文章のGraph構造を用いて、要約文を抽出する手法であるLexRank[2]を適用。



Embedding手法を用いた重要文抽出

Embedding 技術を用いて、意味を知りたい単語を含む、最も重要な文章を抽出する方法を提案。



考察

- 接触確認アプリやJR券売機の仕様書に対し、提案手法を適用。その結果、提案手法・ツールにより仕様書の重要な部分を、40～70%程度の精度で抽出することができた。
- 新規着手システムの要求仕様書を読む際の、理解の効率化に期待。
- 仕様書に未記載の内容は抽出不可能。仕様書自体の品質の向上が必要。
- 今後、提案手法・ツールをPoC的に業務に導入し、課題点の洗い出し・改良をアジャイル的に進めたい。

[1] Bennani-Smires, et. al., (2018). Simple Unsupervised Keyphrase Extraction using Sentence Embeddings.

[2] Günes Erkan et.al., 2004. LexRank: graph-based lexical centrality as salience in text summarization. J. Artif. Int. Res. 22, 1 (July 2004), 457-479.