

# 画像キャプション生成を加えた文章要約手法の提案

新村 凌平<sup>†</sup> 林 実<sup>†</sup>

<sup>†</sup> 明星大学大学院理工学研究科

## 1. はじめに

近年,人間の情報処理を助けるための文章要約技術が盛んに研究されている[1][2].その研究の多くはテキストデータのみを用いた要約であるが,実際に目にする文章では画像などのビジュアル要素が,説得力を高めるために使用されていることが多い.

そこで本稿では,画像から得られる情報を要約対象のテキストに加えることで文章要約の品質を向上させる手法を試みたので報告する.

## 2. 実験

今回の実験では,“Nature”などの科学分野の英語ニュース記事を 15 種類使用し,各記事に対して以下の二つのアプローチで要約を生成した.

- ① テキストのみを使用した要約
- ② 生成キャプションありの要約

①では元の記事の文章をそのまま要約元のテキストとして使用した.②では①に加えて記事内の画像に対してキャプションの生成を行い,生成されたキャプションを本文の該当位置に挿入して,要約対象のテキストとした.

テキストの要約には T5-Base を使用した.T5-Base とは Transformer ベースのモデルである T5(Text-to-Text Transfer Transformer) [3]の中でも,約 2 億 2000 万のパラメータを持つ中間サイズのモデルである.

要約対象のテキストに対してそれぞれ要約タスクを実行したのちに,生成された要約の品質を評価するため,BLEU スコアおよび ROUGE スコア (ROUGE-1, ROUGE-2, ROUGE-L) を使用して出力結果の評価を行った.評価指標の性質は以下の通りである.

- ・ BLEU スコア: 生成された要約が,比較対象のテキストとどの程度一致しているかを示す.
- ・ ROUGE-1 スコア: 単語レベル(uni-gram)での重複を評価する.このスコアが高い要約は,比較対象のテキストの単語をより多く含んでいる.
- ・ ROUGE-2 スコア: 2 語の連続した単位(bi-gram)の重複を評価する.このスコアが高い要約は,比較対象のテキストのフレーズや文脈をより正確に捉えている.
- ・ ROUGE-L スコア: 最長共通部分列 (Longest Common Subsequence) を評価する.このスコアが高い要約は,比較対象のテキストの全体的な構造や内容を保持している.

## 3. 実験結果

テキストのみを使用した要約と画像キャプション生成

を加えた要約の両方で,BLEU スコアと ROUGE スコア (ROUGE-1, ROUGE-2, ROUGE-L) を計算した.出力結果に対する評価の平均値を図 1 に示す.

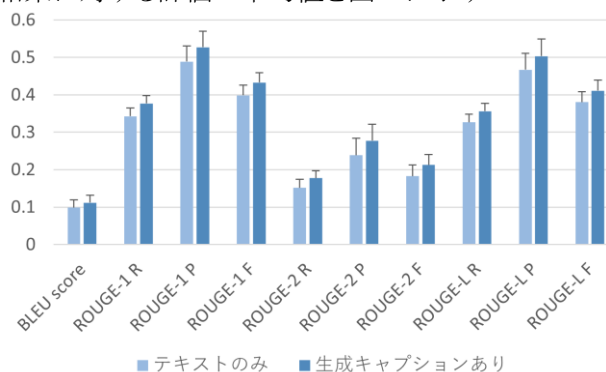


図1. 出力結果に対する評価の平均値

## 4. 考察

実験結果(図 1)より,画像キャプションを加えた要約がテキストのみの要約と比較して, BLEU スコアおよび ROUGE スコア (ROUGE-1, ROUGE-2, ROUGE-L) で一貫して高いスコアを示していることがわかる.これは生成したキャプションを要約対象の文章に加えることで,要約タスクを実行した際に要約元の文意をより正確かつ包括的に反映することができているためであると考えられる.今後は異なる種類の画像キャプションやさらに多様なデータセットを用いた実験を行うことで,これらの結果を汎化していく予定である.

## 5. まとめ

画像のキャプションを生成し要約対象のテキストに加える手法を試みた.その結果,単語,フレーズ,全体的な構造のすべてのレベルで要約の品質が向上し,提案手法が有効であることが示唆された.

## 参考文献

- [1] Alami Merrouni, Z., Frikh, B. & Ouhbi, B. EXABSUM: a new text summarization approach for generating extractive and abstractive summaries. J Big Data 10, 163 (2023). <https://doi.org/10.1186/s40537-023-00836-y>
- [2] Al-Maleh, M., Desouki, S. Arabic text summarization using deep learning approach. J Big Data 7, 109 (2020). <https://doi.org/10.1186/s40537-020-00386-7>
- [3] Raffel, C., et al. "Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer." (2023) arXiv:1910.10683, <https://doi.org/10.48550/arXiv.1910.10683v4>