

文位置を考慮した国会会議録の要約

有田智也^{1,a)} 松井くにお^{1,b)}

概要 : 政党のマニフェストと政治家の発言に齟齬がないかを確かめるためには、議会での発言を確認する必要がある。議会での発言は会議録として一般に公開されているが、文章量が多く読みづらいという問題がある。そこで本研究では、1文から12文でなりたっている議員の質問への回答文を文ベクトルに変換し、文ベクトルに文位置の情報を加算することで得られた、文位置情報が考慮された文ベクトルをニューラルネットワークで学習し、抽出型要約を作成するモデルを提案する。提案手法を評価するために、議員の発言文数に応じた文数で要約したデータを自作し、ROUGEスコアによる評価を行った。実験の結果文位置情報を付与することで、ROUGE-1において0.39から0.42に、ROUGE-2において0.26から0.29に、ROUGE-Lにおいて0.28から0.30に改善された。

キーワード : BERT, 要約, 抽出型要約, ニューラルネットワーク

1. はじめに

日本の選挙では、候補者や政党は自らの政策や方向性を示したマニフェストをネットやメディアを通じて有権者に示し、それを投票する際の参考にしてもらうという形式が取られている。しかしマニフェストには法的拘束力がないため、消費税増税などの賛否が分かれる政策をマニフェストに載せず、実権を握った後に不意打ち的に強行することができる。また実現する気がない政策を有権者の関心を集めるためだけに掲げている議員もいる。そのため投票する際にはマニフェストだけを参考にするのではなく議員の普段の発言を観察し、発言内容とマニフェストに齟齬が無いかを確認する必要がある。しかし国会会議録は文章量が多く、普段の発言のすべてに目を通すことは困難である。

そのため本研究では、入力文章の長さに応じた適切な長さで文章を要約する抽出型要約の手法を提案する。

2. 関連研究

2.1 LexRank

LexRank[1]は、Günes Erkanらが提唱した教師無し抽出型要約のアルゴリズムである。多くの文と類似している文が重要文であり、重要文と類似している文も重要文として要約を行う。LexRankでは単語 w_i の羅列からなる文 $d_i(w_1w_2 \dots w_n)$ に対して、TF-IDF値を計算し、文ベクトル $dv_i(w_1w_2 \dots w_n)$ に変換する。TFは単語の出現頻度であり、文 d_i 中の単語 w_i の割合である。IDFは逆文章頻度であり、単語 w_i の珍しさを表す。TF-IDFは式(1)によって定義される。ただし単語 w の文 d での出現回数を $n_{t,d}$ 、文 d 中の単語すべての出現回数を足した値を $\sum_{s \in d} n_{s,d}$ 、全文数を N 、単語 w

が出現する文書数を $df(t)$ と表す。

$$tfidf(t,d) = \frac{n_{t,d}}{\sum_{s \in d} n_{s,d}} \left(\log \frac{N}{df(t)} + 1 \right) \quad (1)$$

次に文ベクトル dv_i 間のcos類似度を計算し、cos類似度を閾値と比較し、大きいものを1、小さいものを0に変換することで行列 \mathbf{A} を作成する。 \mathbf{A} に対してべき乗法を行うことで関連度が計算できる。関連度が高いものから n 件を抽出することで、入力文章を n 文に要約できる。

2.2 SCDV

SCDV(sparse composite document vectors)[2]は、Dheeraj Mekalaらが提唱した、良質な文ベクトルを生成するためのモデルである。SCDVではまずSkip-Gram[3]とNegative Samplingで単語 w_i に対しての単語ベクトル V_{w_i} を計算する。Skip-Gramは単語 w_i が入力された時に周辺に現れる単語 $(w_{i-n} \dots w_{i-1}, w_{i+1} \dots w_{i+n})$ を2層のニューラルネットワークで予測するモデルであり、入力層と隠れ層の間の重みを単語ベクトルとして利用する。Negative SamplingはSkip-Gramタスクを高速に解くためのアルゴリズムである。Skip-Gramでは V 個の単語から正しい単語を選ぶ多値分類のタスクを解くため、softmaxを計算する際には語彙数に比例した計算時間がかかる。そこでNegative Samplingでは正例と少量の負例を分類する二値分類のタスクに落とし込み、ロジスティック回帰で解くことで、学習を高速化している。次に混合ガウスモデルでソフトクラスタリングし、各単語 w_i がクラス C_j に属する確率 $P(C_j(w_i))$ を計算、結合することで確率行列 p_{wi} を得る。混合ガウスモデルはクラスタリング手法の一つであり、データ点がガウス分布に属している

1 金沢工業大学
Kanazawa Institute of Technology
a) c6000578(@)planet.kanazawa-it.ac.jp
b) kmatsui(@)neptune.kanazawa-it.ac.jp

確率を推定する手法である。その後文 $d_i(w_1w_2 \dots w_n)$ に対して、 $\tilde{x} = \sum_i idf(w_i)p_{w_i} \otimes V_{w_i}$ を計算し得られる non sparse なベクトルに対して一定の値より小さい成分の値を 0 にすることで sparse な文ベクトル x_d を得る。

2.3 BERT

BERT[4]は Jacob Devlin らにより提唱された事前学習モデルであり、複数個の Transformer エンコーダーの積み重ねによって構成されており、あらゆる NLP のタスクに Fine-tuning することが可能なモデルである。BERT に文を入力するには先頭に[CLS]トークンを付与し、複数文を入力するには文と文の間に[SEP]トークンを付与する。BERT ではラベルのついてないデータから事前学習として、MLM (Masked LM)と NSP(Next Sentence Prediction)の 2 つのタスクを解いている。MLM では入力文から[CLS], [SEP]を除いたトークンを一定の割合で[MASK]トークンに置換した文を BERT に入力し、[MASK]トークンに対応する最終層の出力から置き換えられる前の単語を予測する。NSP は[SEP]トークンで接続された 2 文が連続したものかを、文頭にある[CLS]トークンの最終層の出力をもとに予測する。これら 2 つのタスクを解くことで BERT は双方向から文章を捉えることができる。

3. 提案手法

本研究で提案する手法の概要図を図 1 に示す。議員の応答文は“お伺い”、“答弁”、“でしょうか”、“いただければ”が含まれる質問の発言がなされた後の別の議員の発言を応答文として抽出した。提案手法ではモデルに入力する文ベクトルに Embedding により文位置情報を与える。具体的には、まず文 d_i を BERT に入力して得られる[CLS]トークンに対応する最終層の出力から、文ベクトル dv_i を取得する。 dv_i に対して文位置に応じたテンソルの Positional Embedding を加算することで、位置情報が考慮された文ベクトル PEd_i を取得する。 PEd_i を単層のニューラルネットワークに入力し、付与された 0 と 1 のラベルを予測できるようにニューラルネットワークの重みと、BERT の最終層の重みを更新する。

要約生成の際には、入力文長に応じた要約文数になるまで 1 のラベルが付与された文を抽出した。また 1 のラベルと予測された文数が足りない場合は 0 のラベルが予測された文から 1 のラベルの確率が高い文を抽出した。

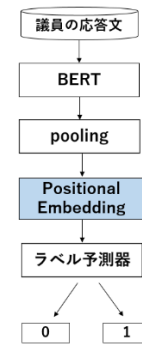


図 1 提案手法の概要図

3.1 文位置正規化 Embedding

文位置正規化 Embedding は文の位置に応じて与えられる Embedding であり、式(2)によって定義される。Pos は何番目の文章、 i はテンソル上の位置、 $\#D$ は最大の入力文数である。

$$PE = \frac{Pos_i}{\#D} \quad (2)$$

3.2 Sinusoidal Positional Embedding

Sinusoidal Positional Embedding は Transformer[5]で利用されている位置 Embedding の手法であり、学習不要で利用することができる。Transformer の位置埋め込みは単語の位置と分散表現の次元によって、一意に定まる位置情報テンソルを作成しているが、提案する埋め込み手法では文位置と分散表現の次元によって、一意に定まる位置情報テンソルを文ベクトルに加算する。Sinusoidal Positional Embedding は式(3), (4)によって定義される。ただし Pos は何番目の文章、 i はテンソル上の位置、 dim はモデルに入力する Embedding の次元数である。

$$PE_{(Pos,2i)} = \sin\left(\frac{Pos}{10000^{2i/dim}}\right) \quad (3)$$

$$PE_{(Pos,2i+1)} = \cos\left(\frac{Pos}{10000^{2i/dim}}\right) \quad (4)$$

3.3 位置埋め込み学習

BERT で利用されている構造であり、文位置に応じて一意に定まる位置情報テンソルを学習可能なパラメータとして学習させる。

4. 文章要約実験

4.1 使用するデータセット

本研究では国立国会図書館が公開している国会会議録データ[5]の内、2020 年度の予算委員会会議録から、質問者の回答者に対する質問の回答 3000 文を抽出し、要約に適切な文に 1、適切でない文に 0 のラベルを付与したデータを作成した。作成した会議録データを学習データ 8、テストデータ 2 の割合で分割し、5 分割交差検証でモデルの評価を行った。作成したデータの統計情報を表 1 に示す。また ROUGE スコアを評価するために議員の応答から 10 件をランダムに抽出し、1 から 3 文の場合は 1 文の要約を、4 か

ら 5 文の場合は 2 文の要約を、6 文以上の場合は 3 文の要約を手で作成した。

表 1 作成したデータの統計情報

文位置	ラベル	
	0	1
1	386	495
2	372	332
3	338	158
4	224	122
5	159	76
6	84	54
7	58	26
8	38	10
9	20	9
10	13	5
11	9	1
12	7	4
合計	1708	1292

4.2 実験設定

BERT の事前学習済みモデルとして、Hugging Face 社が公開している Transformers というライブラリから、東北大学が学習させた BERT のモデル(cj-tohoku/bert-base-japanese-whole-word-masking)[6]を利用し、BERT の最終層とニューラルネットワークを学習させた。バッチサイズは 16、最適化手法に Adam を利用し、BERT の最終層の学習率は $5e-5$ 、ニューラルネットワークの学習率は $1e-4$ とした。

4.3 実験結果と評価

以下に国会会議録を本手法によって要約した結果の一例を表 2 に示す。入力文は以下のとおりである。

- ①本制度は、国民がマイナンバー付きの口座を国に登録することで様々な給付金を簡単な手続きで迅速に受け取れるようにするものであって、国民の皆様は利便性を感じてもらい、希望して登録を進めてもらうということにしています。
- ②このため、口座登録も、マイナポータルのほか金融機関や行政機関等、様々なチャンネルを通じて行えるようにすることで、簡単に登録できて、また幅広い公金の受取が可能になる制度としております。
- ③それは、特定公的給付ですね、今回のやつとか年金、児童手当、失業手当、生活保護、国税の還付等々たくさんあるわけでございます。
- ④これにより、国民に口座の登録を義務とせずとも登録は広がるのではないかと考えておりますが、政府としては、口座登録のそのメリットについて国民の理解を得るべく、制度の周知、広報にもしっかりと取り組んでいきたいと思っております。
- ⑤そして、委員のその義務化という話でございますが、義務とする場合でも、その実効性を担保する措置が併せて整備されなければ登録は進まないというふうに思っておりますので、メリットをちゃんと国民にお伝えしたいと、そのように考えております。

表 2 入力例での要約結果

Model	①	②	③	④	⑤
lead3	○	○			
LexRank	○	○			
SCDV	○	○			
BERT			○	○	
BERT+シンプルEmbedding		○		○	
BERT+SinusoidalEmbedding	○				○
BERT+位置埋め込み学習	○				○

実験の結果得られた precision, recall, F-measure を表 3 に示す。Lead-n は、議員の応答文章の長さに応じて先頭から n 文を抜き出した要約の結果、LexRank は、Günes Erkan らが提案したモデルを実装し要約した結果、SCDV は Dheeraj Mekala らが提案したモデルで文ベクトルを抽出した後に、文ベクトルをニューラルネットワークに入力し要約文に相当かを分類した結果を示す。BERT の[CLS]トークンから文ベクトルを抽出し、ニューラルネットワークで分類する手法では、文章要約タスクにおいて有用なベースラインである、lead-n よりも precision は低くなったが recall と F-measure が高くなった。提案手法の中でも文位置正規化 Embedding を利用した際の評価が最良であり、[CLS]トークンの出力をそのまま利用した場合に比べて、precision で 0.07, F-measure で 0.03 改善された。また Sinusoidal Positional Embedding と位置埋め込み学習では位置 Embedding を加算しない場合に比べて、precision, recall, F-measure 共に向上は見られなかった。

表 3 提案システムの評価結果

Model	precision	recall	F-measure
lead3	0.73	0.50	0.59
LexRank	0.65	0.42	0.50
SCDV	0.59	0.55	0.58
BERT	0.63	0.61	0.62
BERT+文位置正規化Embedding	0.70	0.60	0.65
BERT+SinusoidalEmbedding	0.58	0.61	0.59
BERT+位置埋め込み学習	0.63	0.58	0.61

実験の結果得られた ROUGE スコアの F-measure を表 4 に示す。自動評価の結果と同様に、文位置正規化 Embedding を利用した際の評価が最良であり、BERT の出力を利用した場合に比べて、ROUGE-1 で 0.03, ROUGE-2 で 0.03, ROUGE-L で 0.02 の改善が見られた。

縦軸を文の位置横軸をテンソルの次元として(3), (4)式により、得られた文位置 Embedding を可視化した結果を図 2 に示す。また学習された文位置 Embedding を可視化した結果を図 3 に示す。Sinusoidal Positional Embedding によって得られる Embedding では、テンソルの先頭 20 次元程度はバラバラな値になっており、それ以降のテンソルは規則的な値になっている。学習された文位置 Embedding ではテンソルのすべての次元にバラバラな値が加算されている。実験の結果、文位置 Embedding にはテンソルに一定の値を加算する文位置正規化 Embedding が有効である。

表 4 提案システムの ROUGE の F-measure 評価結果

Model	ROUGE-1	ROUGE-2	ROUGE-L
lead3	0.38	0.25	0.28
LexRank	0.35	0.24	0.24
SCDV	0.40	0.29	0.29
BERT	0.39	0.26	0.28
BERT+文位置正規化Embedding	0.42	0.29	0.30
BERT+SinusoidalEmbedding	0.40	0.28	0.30
BERT+位置埋め込み学習	0.39	0.20	0.27

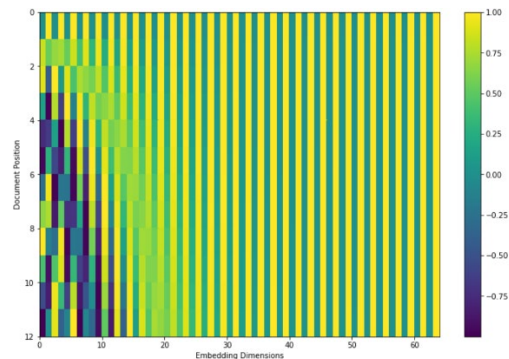


図 2 Sinusoidal Positional Embedding の可視化

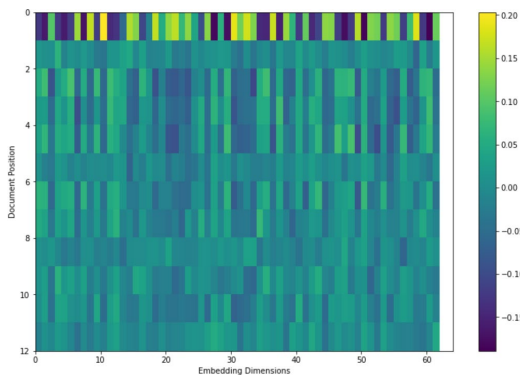


図 3 学習させた文位置埋め込みの可視化

5. おわりに

本研究では文章の抽出型要約において文位置の情報を文ベクトルに加算することによる影響を検証した。文位置を正規化した Embedding を文ベクトルに加算することで、F-measure 及び ROUGE スコアが向上することが確認された。

今後は公開されている他の文章要約データセットでも文位置情報を Embedding に付与することで、抽出型要約の精度が向上するかを検証していきたい。また CNN/Daily Mail や DUC2001~DUC2007 などの英語のデータセットでも本手法が有効であるかを検証したい。さらに新聞記事の要約や対話の要約などドメインによる最良の Embedding の方法を分析していきたい。

参考文献

[1]Günes Erkan and Dragomir R. Radev. LexRank: Graph-based Lexical Centrality as Saliency in Text Summarization.

Journal of Artificial Intelligence Research 22, pp457-479

- [2] Dheeraj Mekala, Vivek Gupta, Bhargavi Paranjape and Harish Karnick. SCDV : Sparse Composite Document Vectors using soft clustering over distributional representations. Proceedings of 2017 Conference on Empirical Methods in Natural Language Processing, pp659-669
- [3] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781, 2013
- [4] Jacob Devlin, Ming-Wei Chang, Kenton Lee and Kristina Toutanova. BERT:Pre-training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of the Association for Computational Linguistics, pp4171-4186
- [5]Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser and Illia Polosukhin. Attention Is All You Need. 31st Conference on Neural Information Processing Systems (NIPS 2017)
- [6]国立国会会議録検索システム検索 API <https://kokkai.ndl.go.jp/api.html>(引用日:2021/6/21)
- [7]Hugging face <https://huggingface.co> (引用日:2021/6/21)