

多様な地方議会議事録を対象とした トピックセグメンテーションおよび質問と回答の対応付け

東山喜輝¹ 秋葉友良²

¹ 豊橋技術科学大学大学院 ² 豊橋技術科学大学 情報・知能工学系

¹higashiyama.yoshiki.ta@tut.jp ²akiba@cs.tut.ac.jp

概要

一部の地方議会では議員がまとめて質問をし各担当者が答弁を行う形式がとられている。この形式の議事録は各質問に対する回答が離れていることや回答順が質問順ではないことにより読みづらい。本研究は議事録を一問一答形式に自動的に対応付けすることを目的とする。提案手法は機械学習によるセグメンテーションと整数線形プログラミングによるアライメントの2段階で構成されている。12都市の議会議事録を用いた評価実験の結果、質問文章に対するセグメンテーションおよびアライメントでは従来手法を上回る性能を確認した。

1 はじめに

政府や地方議会では、さまざまな議題に対して議論し、発言内容は議事録に記録され、誰でも閲覧することができる。しかし、一部の議会では、一名の議員がまとめて質問をし、各担当者が該当する質問に対して答弁を行う形式が取られている。この「一括質問・一括答弁」の形式で記録されている議会議事録は以下の要因により住民にとって読みづらい。1つ目は質問文章や回答文章が長く、各質問に対する回答が離れていることである。2つ目は回答者が答えられる質問に対してまとめて回答していることにより回答順は質問順ではないことである。これらの問題に対処するために、一つの質問に対して該当する答弁が並ぶ「一問一答」形式に変換することが望ましい。そこで、本研究では「一括質問・一括答弁」形式の議事録を「一問一答」形式に自動的に変換し、議事録の可読性の向上を目的とする。また、先行研究 [1] では東京都議会議事録を対象としていたが、本研究では東京都議会議事録のデータを用いて、多様な地方議会議事録にも対応でき、汎用的に適用可能な手法を検討する。

2 関連研究

2.1 トピックセグメンテーション

「一括質問・一括答弁」の方式の議事録は、複数の議題が並べられており、各議題の境界を見つけるトピックセグメンテーションが必要である。Hearst(1997)[2] は、テキストを単語 t とし、いくつかの単語を連結した2つのブロックを順に動かしていき、コサイン類似度を計算する TextTiling を提案した。Lukasik ら (2020)[3] は、3つの BERT ベースのアーキテクチャを提案した。1つ目は、テキストを分割する位置の前後に来るテキストのトークンの系列を用いて、事前学習済みの BERT をファインチューニングをし、学習させたモデルで分割を行う Cross-Segment BERT である。2つ目は、各テキストを BERT でトークンの系列に変換し、それを Bi-LSTM に渡し、分割を行う BERT+Bi-LSTM である。3つ目は、2つ目と同じく BERT でトークンの系列に変換し、それを BERT に渡し、分割を行う Hierarchical BERT である。大杉ら (2023)[1] は、分割する前後に含まれるキーワードを正規表現で判断し、キーワードが含まれる場合はテキストを分割する。Retkowski ら (2024)[4] は、各文をテキスト埋め込みモデルでエンベディングに変換し、機械学習モデルに対して変換したエンベディングを文章サイズでまとめて入力し、分割を行う MiniSeg を提案した。

2.2 アライメント

議事録を「一問一答」形式に変換するためには、セグメンテーションで分割した質問と答弁を対応付けるアライメントが必要である。大杉ら (2023)[1] は、分割した質問と答弁を Okapi BM25[5] の単語重みベクトルに変換し、すべての質問と回答の組み合わせでコサイン類似度を計算して、ハンガリアン・

アルゴリズム [6] を用いて最適な組み合わせを求める手法を提案した。

3 提案手法

提案手法は、図 1 に示すようにセグメンテーション (3.1 節) とアライメント (3.2 節) の 2 段階で構成されている。セグメンテーションによりテキストを分割し、アライメントにより分割した質問と答弁を対応付ける。

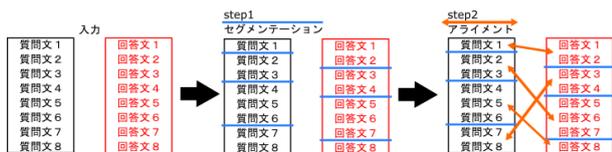


図 1 提案手法の流れ

3.1 セグメンテーション

議題 (トピック) の区切りが存在しない議会議事録に対してセグメンテーションを行い、質問や答弁の中で議題の境界を見つけてテキストを分割する。先行研究 [1] は、正規表現によるセグメンテーションを行っている。この正規表現は、東京都議会議事録に特化していることにより、多様な地方会議事録に適用できない。そこで正規表現をそれぞれの議事録に対応させることは有効であるが、議事録ごとの表記揺れや構造の違いに応じて正規表現を修正する必要がある、その都度人手による調整を要するため、コストがかかる。提案手法では、東京都議会議事録の訓練データセットを用いて学習を行った機械学習モデルによるセグメンテーションを行った。

話者が切り替わった個所でテキストを分割した後処理を行っている。話者が変われば議題も変わる可能性が高いため、話者交代は重要なセグメンテーションの手がかりとなる。

3.1.1 MiniSeg GLuCoSE Regexp

この手法は、MiniSeg[4] を採用した。この手法では、各文をテキスト埋め込みモデルでエンベディングに変換し、機械学習モデルに対して変換したエンベディングを文章サイズでまとめて入力し、セグメンテーションを行う。テキスト埋め込みモデルには、GLuCoSE¹⁾を採用した。このモデルは LUKE[7] の日本語モデルをベースにした日本語埋め込みモデルである。LUKE は単語とエンティティを独立したトークンとして扱う知識拡張型訓練済み Transformer

1) <https://huggingface.co/pkshatech/GLuCoSE-base-ja>

モデルである。このモデルを用いて、各文のエンベディングを取得する。そして、機械学習モデルは RoFormer[8] を採用した。RoFormer は、Transformer の自己注意機構に回転位置エンコーディングを導入したモデルである。MiniSeg モデルの埋め込みモデルを GLuCoSE に変更したモデルを MiniSeg GLuCoSE と呼ぶ。

この MiniSeg GLuCoSE モデルの出力と、正規表現による分割位置を組み合わせるセグメンテーションを行う。この組み合わせ手法を MiniSeg GLuCoSE Regexp と呼ぶ。学習ベース手法と明示的ルールベース手法を組み合わせることで、より高精度なセグメンテーションを実現する。正規表現は先行研究で用いられた付録 A の表 3 の正規表現を使用する。MiniSeg GLuCoSE モデルの分割確率 p_{model} と正規表現の判断結果 $r \in \{0, 1\}$ を次式で組み合わせた。モデルの確率の重み α は 0.8 に設定している。

$$p_{final} = \alpha p_{model} + (1 - \alpha)r$$

提案手法は、事前学習済みの BERT を訓練データセットでファインチューニングしたモデルを用いて、テキストを分割する。訓練データセットは、2021-22 年開催のワークショップ NTCIR-16 における QA Lab-PoliInfo タスクの QA Alignment[9] で使用された訓練データとテストデータを用いた。加えて、QA Alignment の訓練データに含まれていない東京都議会の議事録を自動分割したデータも訓練データとして用いた。自動分割の方法は、大杉ら [1] がセグメンテーションで用いた正規表現による手法である。使用した正規表現は、付録 A の表 3 に示す。また、同タスク [9] で使用されたテストデータを検証データとして使用している。加えて、付録 A の表 3 の正規表現で自動分割したデータを検証データに追加した。なお、訓練データと検証データは、質問文章と回答文章に分けて用意し、それぞれに対応する 2 つのファインチューニングモデルを用意した。

3.2 アライメント

分割した質問や回答を対応づけるアライメントを行う。

3.2.1 ベクトル変換

質問と回答を対応づけるために、テキストを比較する必要があるため、テキストをベクトルに変換する。先行研究 [1] に習い、Okapi BM25[5] の単語重み

ベクトルに変換する。加えて、異なる視点のベクトルとして、テキスト埋め込みモデルによってベクトルを取得する。本研究では、この2つのベクトルを組み合わせると比較し対応付けを行っている。

BM25では、単語の出現頻度 (TF)、単語の逆文書頻度 (IDF)、文書に含まれる総単語数 (DL) を組み合わせるとテキストごとの単語重みを計算する。文書 D の単語 Q に対するスコアが次式で表現される。

$$score(D, Q) = \sum_{i=1}^n IDF(q_i) \frac{f(q_i, D)(k_i + 1)}{f(q_i, D) + k_i(1 - b + b \cdot \frac{|D|}{avgdl})}$$

単語の逆文書頻度 (IDF) とは、テキスト集合中にある単語が出現する文書の少なさ (文書頻度の低さ) を表し、次式で表現される。ここでのテキスト集合は、分割した質問や答弁の一つを表す。

$$IDF(q_i) = \log \frac{N}{n(q_i)}$$

本研究では、テキスト埋め込みモデルとして GLuCoSE を採用している。この変換した質問と回答のベクトル間のコサイン類似度を計算し、分割した質問と答弁の対応付けを見つける。BM25 によるベクトル間のコサイン類似度と GLuCoSE によるエンベディングのコサイン類似度を加算したものを基にマッチングを行う。質問と回答のすべての組み合わせに対して、類似度を計算し、類似度行列を用意する。この類似度行列を用いてマッチングを行う。

3.2.2 一対多マッチング

先行研究では、質問や回答は一対一であることを前提に対応付けを行っている。しかし、東京都議会以外では回答された質問に対して補足するように複数の回答者が答弁する場合があるため、一対一では不十分である。そこで、一つの質問に対して複数の回答を対応付けできるように一対多のマッチングを行った。質問と回答のベクトル間のコサイン類似度の合計を最大にする最適化問題として計算する。この問題を整数線形プログラミングで解く。以下に使用した変数、目的関数、制約条件を示す。実装には PuLP²⁾ を使用した。

変数

$$x_{ij} = \begin{cases} 1 & \text{質問 } Q_i \text{ を回答 } A_j \text{ に対応付ける場合} \\ 0 & \text{それ以外} \end{cases}$$

2) <https://coin-or.github.io/pulp/>

目的関数 逆コスト行列 (類似度行列) C を用いて、合計コストを最大化する：

$$\max \sum_{i=1}^m \sum_{j=1}^n C_{ij} x_{ij}$$

制約条件

(1) 各質問は最低 1 件以上の回答を対応付ける

$$1 \leq \sum_{j=1}^n x_{ij} \quad \forall i = 1, \dots, m$$

(2) 各回答は 1 つの質問にのみ対応できる

$$\sum_{i=1}^m x_{ij} \leq 1 \quad \forall j = 1, \dots, n$$

4 評価実験

従来手法と提案手法を、新規の議会議事録を適用し、性能を比較する。

4.1 テストデータセット

12 都市 (安城市、岡崎市、横浜市、名古屋市、札幌市、川崎市、神戸市、京都市、広島市、仙台市、千葉市、北九州市) を対象にテストデータセットを作成した。1 都市に対して 1 名の議員が行った質問と対応する回答をアノテーションした。それに伴い、分割位置も同様にアノテーションした。

4.2 セグメンテーションの評価実験

4.2.1 実験方法

セグメンテーションの指標には、Pk[10]、WindowDiff[11]、F 値を採用する。Pk とは、距離 k だけ離れた 2 つのテキストが同一のかたまり内に属しているかどうかを予測と正解で比較する。エラー率であることから値が小さい方がよい。WindowDiff とは、距離 k だけ離れた 2 つのテキストの間にいくつの境界が含まれているかを比較し、その不一致数をスコアとする指標である。そして、F 値は、まず提案手法の予測データと正解データで混同行列を作成する。この行列から適合率と再現率を計算し、F 値を計算する。3 つの評価指標は、SegEval[12] という python 用の評価パッケージを使用して計算を行った。

4.2.2 実験結果と考察

セグメンテーションの評価結果を表 1 に示す。まず、質問文章では、MiniSeg GLuCoSE Regexp モデル

は F1 スコアで最も高い値を示した。予測した分割位置は正規表現よりも多く正解データと一致したことがわかる。しかし、Pk スコアと WindowDiff スコアは正規表現よりも悪化した。これは、分割数が低下したことが原因だと考えられる。次に、回答文章では、正規表現がすべての評価指標で最も高い値を示した。分割位置が正確であることがわかる。しかし、分割数が少なく、アライメントする際に対応付けられない質問が多くなる可能性がある。MiniSeg GLuCoSE Regexp モデルは、正規表現で分割できなかった分割位置を補完することで、セグメント数の増加に寄与したと考えられる。

そして、MiniSeg GLuCoSE Regexp モデルは、MiniSeg モデルよりもすべての評価指標で改善した。東京都議会に特化した正規表現でも組み合わせることで、性能が向上したと考えられる。

表 1 セグメンテーションの評価

質問文章					
手法	Pk	WindowDiff	F1 ↑	segments(423)	
正規表現	0.153	0.204	0.664		402
MiniSeg	0.212	0.281	0.628		357
提案手法	0.173	0.245	0.684		374
回答文章					
手法	Pk	WindowDiff	F1 ↑	segments(442)	
正規表現	0.133	0.159	0.801		335
MiniSeg	0.184	0.209	0.739		359
提案手法	0.185	0.223	0.730		391

5 アライメントの評価実験

5.1 実験方法

予測データと正解データと比較し、適合率、再現率、F 値を QA Alignment[9] サブタスクの評価方法で計算する。対応付けされた質問セグメントと回答セグメントの文ペアを予測データと正解データで洗い出す。予測の文ペアの内、正解の文ペアと一致する割合を適合率とする。正解の文ペアの内、予測の文ペアと一致する割合を再現率とする。F 値は、適合率と再現率の調和平均として計算する。

先行研究 [1] ではハンガリアン法を用いて一対一の対応付けを行っている。この先行研究の一対一の対応付けと、提案手法の一対多の対応付けの比較を行う。また、セグメンテーション手法の違いによるアライメント性能の差も検証する。

5.2 実験結果と考察

まず、セグメンテーション手法として正規表現を用いた場合の一対一 (1v1) と一対多 (1vM) のアライメントの評価結果を表 2 に示す。一対多の対応付けを行うことで、F1 スコアが 0.013 向上した。再現率が向上していることから、質問と回答の対応付けに漏れが少なくなったことがわかる。一方で、適合率は低下している。余分な回答を対応付けてしまう傾向があると考えられる。これは、最大化問題として定式化したため、回答がいずれかの質問に対応付けしたほうが、対応付けしない場合よりも目的関数の値が大きくなるのが原因であると考えられる。

次に、類似度行列に BM25 のみを用いた場合 (bm25) と GLuCoSE のエンベディングを組み合わせた場合 (sum-cost) の比較を行った。類似度を組み合わせることで、より類似度の差が明確になり、すべての評価指標で改善が見られた。

セグメンテーション手法別のアライメント評価結果も表 2 に示す。セグメンテーションでは正規表現が総合的に高い評価を示したが、MiniSeg GLuCoSE Regexp モデルが最も高いスコアを示した。回答文章のセグメント数が多くなり、一対多の対応付けを行うときに有利に働いたと考えられる。

表 2 アライメントの評価

手法	適合率 ↑	再現率 ↑	F1 ↑
正規表現 (bm25,1v1)	0.667	0.661	0.664
正規表現 (bm25,1vM)	0.632	0.728	0.677
正規表現 (sum-cost,1v1)	0.670	0.676	0.673
正規表現 (sum-cost,1vM)	0.645	0.760	0.698
提案手法 (sum-cost,1vM)	0.652	0.846	0.737

6 おわりに

先行研究は特定の地方議会議事録に特化していたが、多様な地方議会議事録にも対応できる手法を提案した。提案手法は東京都議会議事録を基にした訓練データのみで学習を行い、多様な地方議会の訓練データを必要としない点が特徴である。トピックセグメンテーションの手法は、従来のルールベース手法に対して、機械学習モデルを基にした手法を提案した。アライメントは、従来の 1 対 1 の対応付け手法に対して、1 対多の対応付け手法を提案した。12 件の地方議会を対象とした評価実験を行い、従来手法を上回るアライメント性能を達成した。

参考文献

- [1] 大杉了斗, 秋葉友良, 増山繁. 会議議事録を対象とした議論構造解析と議題生成による議会レポートの自動生成. 言語処理学会第 29 回年次大会 発表論文集, pp. 2803–2808, 2023.
- [2] Marti A. Hearst. Texttiling: A quantitative approach to discourse segmentation. In **Annual Meeting of the Association for Computational Linguistics**, 1993.
- [3] Michal Lukasik, Boris Dadachev, Gonçalo Simões, and Kishore Papineni. Text segmentation by cross segment attention. **CoRR**, Vol. abs/2004.14535, , 2020.
- [4] Fabian Retkowski and Alexander Waibel. From text segmentation to smart chaptering: A novel benchmark for structuring video transcriptions. In Yvette Graham and Matthew Purver, editors, **Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics (Volume 1: Long Papers)**, pp. 406–419, St. Julian’s, Malta, March 2024. Association for Computational Linguistics.
- [5] Stephen Robertson, S. Walker, S. Jones, M. M. Hancock-Beaulieu, and M. Gatford. Okapi at trec-3. In **Overview of the Third Text REtrieval Conference (TREC-3)**, pp. 109–126. Gaithersburg, MD: NIST, January 1995.
- [6] H. W. Kuhn and Bryn Yaw. The hungarian method for the assignment problem. **Naval Res. Logist. Quart.**, p. pp. 83–97, 1955.
- [7] Ikuya Yamada, Akari Asai, Hiroyuki Shindo, Hideaki Takeda, and Yuji Matsumoto. LUKE: Deep contextualized entity representations with entity-aware self-attention. In Bonnie Webber, Trevor Cohn, Yulan He, and Yang Liu, editors, **Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)**, pp. 6442–6454, Online, November 2020. Association for Computational Linguistics.
- [8] Jianlin Su, Yu Lu, Shengfeng Pan, Bo Wen, and Yunfeng Liu. Roformer: Enhanced transformer with rotary position embedding. **CoRR**, Vol. abs/2104.09864, , 2021.
- [9] Masaharu Yoshioka et al. Yasutimo Kimura, Yuzu Uchida. Overview of the ntcir-16 qa lab-poliinfo-3 task. **NTCIR 16 Conference: Proceedings of the 16th NTCIR Conference on Evaluation of Information Access Technologies**, 2022.
- [10] Doug Beeferman, Adam Berger, and John Lafferty. Statistical models for text segmentation. **Machine Learning**, Vol. 34, No. 1-3, pp. 177–210, 1999.
- [11] Lev Pevzner and Marti A. Hearst. A critique and improvement of an evaluation metric for text segmentation. **Computational Linguistics**, Vol. 28, No. 1, pp. 19–36, 2002.
- [12] Chris Fournier. Evaluating Text Segmentation using Boundary Edit Distance. In **Proceedings of 51st Annual Meeting of the Association for Computational Linguistics**, p. to appear, Stroudsburg, PA, USA, 2013. Association for Computational Linguistics.

A 先行研究のセグメンテーションで用いた正規表現

先行研究で用いられた正規表現を付表 3 に示す。

表 3 先行研究のセグメンテーションで用いた正規表現

質問	(お)?(伺い 尋ね)(を)?(いた)?し?(させて いただき)?(ます たい) (見解 答弁 所見 課題 認識 考え 説明) をお?(求 め 伺い 聞かせ 尋ね) (お)?(答え 聞かせ)(て を)?ください ありがとうご ざいました いかがですか どうですか ではありませんか るものです (ど のように どう)(考えて 認識して 取 り組む) のですか のでしょうか
答弁	(お)?(答え(を)?(いた)?(し 申し上げ)ま す 初めに、 次(いで に は)、 ま ず、 他方で、 最後に、 続きまして、 について(です であります でご ざいます) の(お話 お尋ね)(がござい ました ございます) (の に関す る)(ご)?質問で(ございま)?す (質問 指 摘 言及 お尋ね)が?ございました (質問 指摘)を?いただきました

させた場合のアライメントの評価結果を付表 6 に示す。テストデータセットとは別に、検証データセットを用いて評価を行った。安城市と岡崎市を対象に人手でアノテーションした。付表 6 の結果から、 α が 0.8 のときに最も高い F1 スコアを示した。

表 6 重み α を変化させた場合のアライメントの評価 (GLu.=GLuCoSE, Reg.=正規表現)

手法	適合率↑	再現率↑	F1 ↑
GLu.50%+Reg.50%	0.402	0.623	0.488
GLu.60%+Reg.40%	0.480	0.685	0.564
GLu.70%+Reg.30%	0.475	0.636	0.544
GLu.80%+Reg.20%	0.507	0.677	0.580
GLu.90%+Reg.10%	0.526	0.634	0.575

B 不要なセグメントの削除

セグメンテーションで分割した質問や回答の中には、質問や回答として不要なものが存在する。例えば、質問者である議員の発言には、質問以外にも要望や意見表明などの、回答を求めない発言が含まれている場合がある。そこで、不要な質問や回答を削除するために正規表現を用意し、該当するセグメントをアライメントの処理から除外する。使用した正規表現は、付表 4 に示す。

表 4 セグメント削除で用いた正規表現

質問	要望(です し(?!ました た) いたし)期待 ありがとう
答弁	お答え

C テストデータセットの統計量

テストデータセットの統計量を付表 5 に示す。

表 5 テストデータセットの統計量

質問文数	2210
質問セグメント数	423
回答分数	1826
回答セグメント数	442

D 正規表現と MiniSeg GLuCoSE の重み

MiniSeg GLuCoSE Regexp モデルで使用する MiniSeg GLuCoSE モデルの確率の重み α を変化