

# 転移学習に基づくテキスト生成モデルを用いた 多人数会話の発話分析方法の検討

福田 治輝 綱川 隆司 大島 純 大島 律子 西田 昌史 西村 雅史

静岡大学大学院 総合科学技術研究科

## 1. はじめに

近年、オンラインミーティングの普及や音声認識性能の向上により、多人数会話における書き起こしテキストデータの入手がより容易な環境が整いつつあり、さまざまな分野で多人数会話の書き起こしデータの活用が期待される。しかし多人数の発話内容が記録された発話ログを分析する際に、全ての対話ログに目を通して分析を行う作業には多大な労力が必要となり、データ活用の障害となっている。分析者が長い対話ログから必要とする情報を得るにあたり、機械的に話題転換点を抽出することができれば、全てに目を通すことなく対話の構造理解ができ、分析が容易になると考えられる。一方で、文書等におけるテキストとは異なり、会話の書き起こしテキストには砕けた言い回しや文法的に正しくない文が含まれるため、構文解析やそれを利用した談話構造解析を直接適用することが難しい[1]。そこで本研究では、事前学習済みの汎用言語モデル BERT (Bidirectional Encoder Representations from Transformers) [4]および BART[6]を用いて、多人数会話の書き起こしテキストデータの話題分割の方法を提案し、その効果を検証する。

## 2. 関連研究

本章では、議事録生成技術の先行研究と、提案手法で用いる事前学習済みの汎用言語モデルについて述べる。

### 2.1 議事録生成

会議における発言ログである会議録から、重要部分、会議内の意見や質問等をまとめた議事録を生成する研究は多く行われている[1]。特に、文書要約技術を応用した研究は多く行われており、文の関係をランクづけし、LSIを用いて重要文抽出を行うアプローチが提案されている[2]。また、議事録に特化する形で、議事録に残したい種類の情報をもとに木構造パターンを生成することでテキストフレーズ間の関係抽出を行い、短縮文を生成する手法が提案されている[3]。本研究では、複数発話の要約文の変化量から話題分割を行うことで、相槌等による発話分割など、不完全な発話の影響を受けづらいという特徴がある。

## 2.2 BERT[4]

BERT は自己注意機構を使用している Transformer[5]をベースとした汎用言語表現モデルである。大規模なテキストコーパスで事前学習を行い、各タスクに対して少量のデータでファインチューニングすることにより、多くの自然言語処理タスクにおいて有効に動作する。

## 2.3 BART[6]

BART は BERT と同様、Transformer[5]をベースとした汎用言語表現モデルである。BERT は事前学習の手法の特性により文章生成に利用できない。対して BART では、異なる事前学習の手法を用いる文章生成のタスクにも利用が可能であるという特徴をもつ。

## 3. 提案手法

図 1 に話題転換点を抽出する提案手法の概要を示す。一定発話数  $n$  のスライディングウィンドウにより発話全体を分割し、各ウィンドウをあらかじめ要約データセットでファインチューニングされた BART モデルを用いて要約を行う。このとき  $i$  番目の発話  $U_i$  から  $i+n-1$  番目の発話  $U_{i+n-1}$  の要約を  $S_j$  と表す。生成された要約テキストを BERT モデル [7]に入力し、BERT モデルの最終層の重みベクトル (768 次元)  $w(S_j)$  を取得する。連続するテキストのベクトル同士のコサイン類似度  $\cos(w(S_j), w(S_{j+1}))$  を計算し、コサイン類似度が小さい組合せは要約テキストの変化量が大きくなっていると考えられる。要約テキストのコサイン類似度  $\cos(w(S_j), w(S_{j+1}))$  が小さいとき、その要約テキストに初めて含まれた発話  $U_{i+n}$ 、または含まれなくなった発話  $U_i$  のどちらかが要約テキストに大きな影響を与えている。そこで、一つの発話  $U_i$  に対して初めて含まれる要約  $S_{j-n+1}$  のコサイン類似度  $\cos(w(S_{j-n}), w(S_{j-n+1}))$  と含まれなくなった要約  $S_{j+1}$  のコサイン類似度  $\cos(w(S_j), w(S_{j+1}))$  の最小値  $\min(\cos(w(S_{j-n}), w(S_{j-n+1})), \cos(w(S_j), w(S_{j+1})))$  を発話ごとに割り当てる。対話の最初と最後の発話において該当するコサイン類似度が一つしか存在しない場合は、そのコサイン類似度を直接割り当てることとする。この値が小さい発話を話題転換点と扱う。

## 4. 実験

日本の高校 1 年生に通常のカリキュラムの一環として行われた生物の授業におけるグループワークの書き起こしデータを用いる。この授業では生物学の

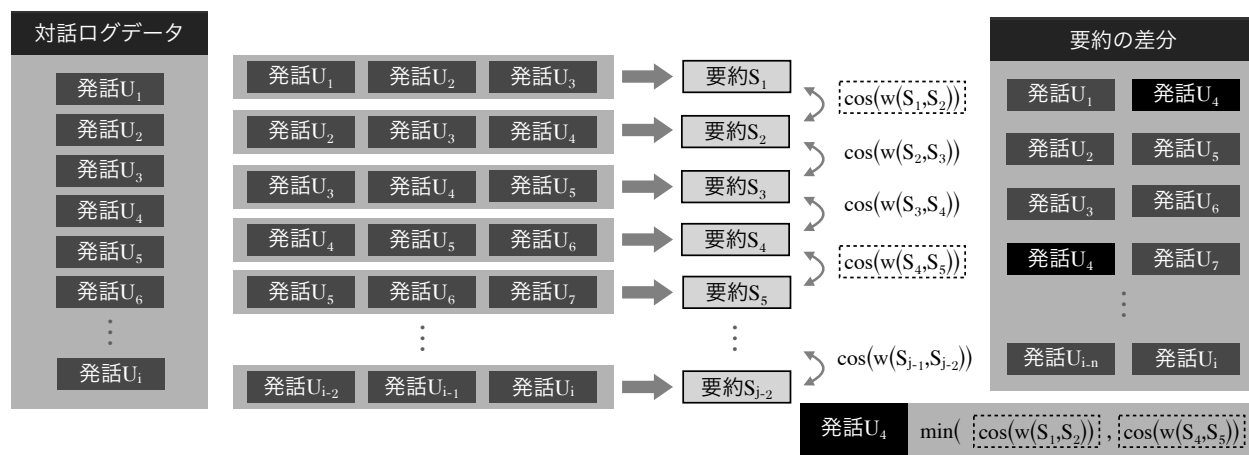


図1 提案手法の概要 (n=3としたときの発話 U<sub>4</sub>に対するスコアを求める例)

表1 発話ごとに割り当てられたコサイン類似度の平均値

	グループ番号	1	2	3	4	5	6	平均
話題転換点での コサイン類似度	平均値	0.952	0.924	0.939	0.919	0.945	0.932	0.935
	標準偏差	0.021	0.043	0.022	0.021	0.029	0.029	0.028
話題転換点以外での コサイン類似度	平均値	0.934	0.923	0.932	0.931	0.940	0.933	0.933
	標準偏差	0.026	0.046	0.028	0.025	0.042	0.032	0.033

博士号を持ち、10年以上の指導経験を持つ理科教師の指導のもと、予防接種をテーマとして3-4人ずつ12のグループでグループワークが行われた。今回はこのうち6グループの発言を人手で書き起こしたデータ(平均発話数424.8, 平均発話長16.6)を用いる。この実験対象データに対して、大学院生3名により人手で話題転換点を付与し、提案手法と比較する。

スライディングウィンドウのサイズnは20に設定した。要約にあたり、日本語には大規模な要約データセットがないことから、英語の要約データセットであるCNN/DailyMailデータセット[8]でファインチューニングされたBARTモデルを用いて要約テキストを生成する。その際、発話をDeepL翻訳[9]により日本語から英語に機械翻訳したものを用いて要約テキストを生成し、生成された要約テキストは再度DeepL翻訳により英語から日本語に機械翻訳を行い日本語の要約テキストを取得する。

得られた要約テキストを日本語Wikipedia記事によって事前学習されたBERTモデル[7]に入力し、特徴量ベクトルを得る。

### 5. 結果

発話ログ中の話題転換点でのコサイン類似度と話題転換点以外でのコサイン類似度を比較した結果を表1に示す。評定者3名によって付与された話題転換点にばらつきが見られたため、いずれかの評定者によって話題転換点が付与された発話を話題転換点と扱ったが、多くのグループにおいて平均値に大きな違いがないことから、人手による評定でばらつきが出るような特定の発話ピンポイントでの話題転換点の検出は難しいことを示している。

### 6. おわりに

多人数会話の書き起こしデータの活用を目的とし

て、事前学習済みの汎用言語モデルを用いた話題分割の方法の検討を行った。今回の検証では人手で付与された話題転換点となる特定の発話の検出は困難であることが示された。今後は話題分割の性能向上とともに、発話者ごとの分析や分割された話題ごとの要約文生成など、更なる分析支援を検討していく。

### 謝辞

本研究はJSPS科研費JP19H01714の助成を受けたものである。

### 参考文献

- [1] 三浦寛也 and 平田圭二, “議事録生成技術に関するサーベイ,” 人工知能学会 言語・音声理解と対話処理研究会, vol. 80, pp. 34-38, 2017.
- [2] Y. Gong and X. Liu, “Generic Text Summarization Using Relevance Measure and Latent Semantic Analysis,” in Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 2001, pp. 19-25.
- [3] 佐藤美沙, 黒土健三, and 柳井孝介, “係り受け構造に基づく関係抽出を利用した議事録作成のための文短縮,” 人工知能学会全国大会論文集, vol. JSAI2020, pp. 4Rin116-4Rin116, 2020.
- [4] J. Devlin et al., “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” arXiv:1810.04805, 2018.
- [5] A. Vaswani et al., “Attention Is All You Need,” in Proc. of the 31st International Conference on Neural Information Processing Systems, pp. 6000-6010, 2017.
- [6] M. Lewis et al., “BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension.” arXiv:1910.13461, 2019.
- [7] 鈴木正敏, “GitHub - cl-tohoku/bert-japanese: BERT models for Japanese text.,” Github, 2020. <https://github.com/cl-tohoku/bert-japanese> (accessed Jan. 07, 2021).
- [8] K. M. Hermann et al., “Teaching Machines to Read and Comprehend,” in Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 1, 2015, pp. 1693-1701.
- [9] “DeepL.” <https://www.deepl.com/> (accessed Jan. 07, 2021).