

ミーティング要約を目的としたプレゼンテーションスライドの 構成要素と発話の自動アラインメント

神谷 賢太郎* 東中 竜一郎† 長尾 確‡

名古屋大学 大学院情報学研究科**‡

1. はじめに

研究の進捗報告などを行うミーティングではプレゼンテーションスライド(以下, スライドとする)を用いることが多く, 発表内容について発表者と参加者が議論を展開する[2]. このようなミーティングの内容の振り返りや議事録の作成を目的に, 我々はスライドの情報と発話を紐付けて提示する手法の検討を行っている.

本研究では, スライドと発話とのアラインメントを自動で行う手法を提案する. 具体的には, まず発話列の中から同一議論に関する発話のまとまりを取得するために, 発話列のセグメンテーションを行い, 次に, 発話のセグメントとスライドの構成要素とのアラインメントを行う.

2. 提案手法

本手法のシステム構成を図1に示す. 学習フェーズでは, 発話とスライド情報が含まれたデータセットからセグメンテーションモデルの学習とアラインメントモデルの学習を行う. なお, セグメンテーションモデルの学習には既存の議事録データ(DM データと呼ぶ)を用いる.

評価フェーズでは, 第一段階として, まずセグメンテーションモデルを用いて発話列のセグメンテーションを行い, 発話セグメントを作成する. そして, 第二段階として, スライドの構成要素と複数の発話を含む発話セグメントとのアラインメントを行う. 以下, これらの処理を具体的に説明する.

2.1. 発話列のセグメンテーション

発話列のセグメンテーションを行うモデルを作成するために, 既存の BERT モデル[1]を, DM (Discussion Mining) データを用いて話題の区切れ目となる発話を判定するように追加学習を行う. DM データは, 本研究室で提案・運用しているディスカッションマイニング[4]という会議記録システムで蓄積されたデータである. このデータはプレゼンテーションを含むミーティングでの発話列が話題ごとに構造化されたテキストの議事録データである. この追加学習済みのモデルを, 後述するデータセットを用いてファインチューニングし, 発話列に含まれる議論の区切れ目となる発話を判定するセグメンテーションモデルを作成する.

2.2. 発話セグメントのアラインメント

発話列のセグメンテーションにより取得した発話セグメントとスライドの構成要素とのアラインメントを行う(図2参照). アラインメントの学習モデルは, 発話とスライドの

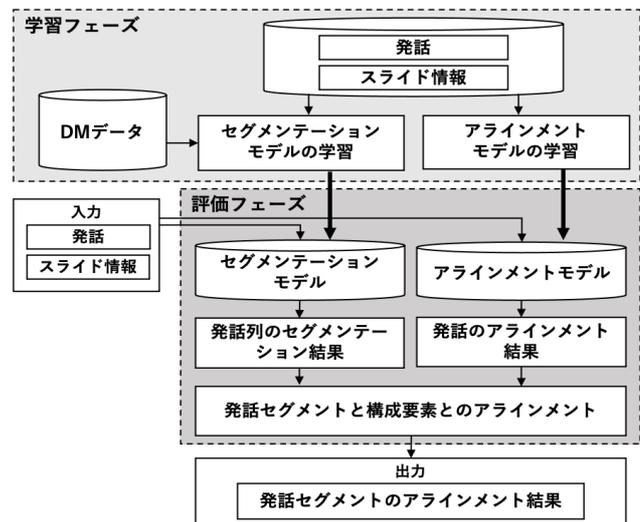


図1. プレゼンテーションスライドの構成要素と
発話の自動アラインメントシステム

構成要素を入力として学習され, 入力された発話とスライドの構成要素が対応付く尤度を出力する. 複数の発話を含む発話セグメントと構成要素とのアラインメントを行う場合は, 発話セグメントに含まれる各発話のアラインメント結果の多数決により最終的な発話セグメントと構成要素とのアラインメントを決定する. 多数決が同数の場合は, 尤度の総和が最大となる構成要素を発話セグメントとのアラインメント結果とする.

3. 評価実験

発話列のセグメンテーションを行った後, 発話セグメントとスライドの構成要素とのアラインメントを行う二段階の手法の有効性を実験により検証した.

3.1. データセット

本研究室で実施された全 17 回のミーティングにおけるプレゼンテーションでの発表者と参加者の発話列データとスライドデータを収集した[4]. 発話列データは発話音声を書き起こしたデータである. また, スライドの構成要素と各発話とのアラインメントをプレゼンテーションの参加者が手動で行い, データセットを作成した.

3.2. 比較手法

本実験では以下の手法を比較した.

(a) ベースライン

本手法は, スライドの構成要素と発話列データに含まれる各発話とのアラインメントを発話ごとに行う. アラインメントモデルの実装には, 既存の BERT モデルをベースのモデルとして使用し, このモデルを入力された発話とスライドの構成要素から, それぞ

Automatically Aligning the Components of Presentation Slides with Utterances for Meeting Summarization

* KAMIYA, Kentaro (kamiya.kentaro@b.mbox.nagoya-u.ac.jp)

† HIGASHINAKA, Ryuichiro (higashinaka@i.nagoya-u.ac.jp)

‡ NAGAO, Katashi (nagao@i.nagoya-u.ac.jp)

**‡ Graduate School of Informatics, Nagoya University

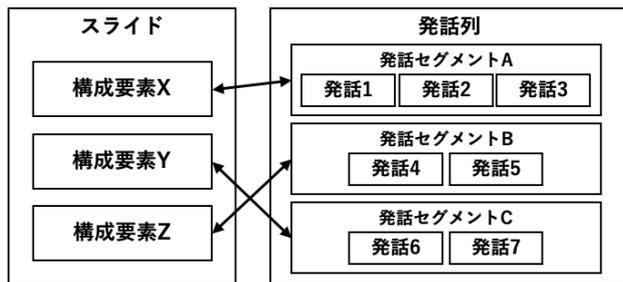


図 2. スライドの構成要素と発話セグメントとのアラインメント

れが対応付く尤度を出力するようファインチューニングした。アラインメントを行う際に当該発話の周辺の発話も考慮するため、当該発話の前後 1 ないし 2 発話を使用した。周辺発話を使用しないモデルを A0、前後 1 発話を使用するモデルを A1、前後 2 発話を使用するモデルを A2 とした。

(b) 二段階手法(提案手法)

本手法は、発話列のセグメンテーションを行った後、発話セグメントとスライドの構成要素とのアラインメントを行う。セグメンテーションモデルの実装には、ベースラインでのアラインメントモデルと同様に BERT を使い、入力された発話が議論の区切れ目となる発話であるかを示す 2 値を出力するようファインチューニングした。また、ベースラインと同様に、セグメンテーションを行う際に当該発話の周辺の発話も考慮した。周辺発話を使用しないモデルを S0、前後 1 発話を使用するモデルを S1、前後 2 発話を使用するモデルを S2 とした。セグメンテーションモデルについては、DM データで BERT モデルの追加学習を行うかどうかの 2 パターンを用意した。取得した発話セグメントと構成要素のアラインメントを行う際は、ベースラインで用いたアラインメントモデルを使用する。

3.3. 評価指標

セグメンテーションモデルの評価には、議論の区切りとなる発話をどれだけ正しく予測できたかを示す再現率を使用する。また、アラインメントモデルの評価には、予測したアラインメント結果が全体でどれだけ正しいかを示す正解率を使用する。

3.4. 評価結果と考察

17 回のミーティングデータを評価実験に用い、One-versus-Rest により提案手法を評価した。学習データのうちの 1 回分のミーティングデータを検証データとした。

ベースラインのアラインメント結果を表 1 に示す。周辺発話として前後 2 発話を使用した A2 モデルの正解率が最も良い結果となった。次に、提案手法でのセグメンテーション結果を表 2 に示す。周辺発話として前後 1 発話を用い、さらに DM データによる追加学習を行った S1 モデルの再現率が最も良い結果となった。表 3 には提案手法でのアラインメント結果を示す。DM データを用いてセグメンテーションモデルの追加学習を行った場合の S1 モデルの正解率が最も高く、ベースラインよりも約

表 1. ベースラインでのスライドの構成要素と発話とのアラインメントの正解率

A0:周辺発話不使用, A1:周辺発話として前後 1 発話使用, A2:周辺発話として前後 2 発話使用

手法	A0	A1	A2
ベースライン	0.5430	0.5649	0.5654

表 2. 提案手法での発話列のセグメンテーションの再現率
S0:周辺発話不使用, S1:周辺発話として前後 1 発話使用, S2:周辺発話として前後 2 発話使用

DM なし/あり:DM データを用いたセグメンテーションモデルの追加学習の有無

モデル	DM なし	DM あり
S0	0.7076	0.7920
S1	0.7169	0.8140
S2	0.7116	0.7814

表 3. 提案手法でのスライドの構成要素と発話とのアラインメントの正解率

発話セグメントとスライドの構成要素とのアラインメントにはベースラインの A2 モデルを使用(表 1 参照)

手法	モデル	DM なし	DM あり
二段階手法 (提案手法)	S0	0.5783	0.5781
	S1	0.5692	0.5848
	S2	0.5673	0.5721

3.4%改善した。

また、データセットの一部を用いて、人手、ベースラインでの正解率の最も高かった A2 モデルおよび提案手法で DM データによる追加学習を行った場合の S1 モデルのアラインメントの正解率を比較したところ、人手は 0.6885(アッパーバウンド)、ベースラインは 0.5198、提案手法は 0.5697 であった。

アラインメントの正解率が最も高かったベースラインの A2 モデルと提案手法の DM データで追加学習を行った場合の S1 モデルの間に差があるかどうかをマクネマー検定 [3] で確認したところ、提案手法の S1 モデルに有意水準 5%で有意差があることを確認した。

4. おわりに

本研究では、プレゼンテーションを含むミーティングにおける発話列のセグメンテーションを行った後、発話のセグメントとスライドの構成要素とのアラインメントを行う二段階によるアラインメント手法を提案した。今後は、ミーティングでの議論の内容や構造に着目し、より効果的なアラインメントアルゴリズムを検討する予定である。

参考文献

[1] J. Devlin et al., BERT: Pre-training of deep bidirectional transformers for language understanding. In Proc. of NAACL, pp. 4171–4186, 2019.
 [2] K. Kamiya, et al., Using Presentation Slides and Adjacent Utterances for Post-editing of Speech Recognition Results for Meeting Recordings. In Proc. of TSD, pp. 331-340, 2021.
 [3] Q. McNemar, Note on the sampling error of the difference between correlated proportions or percentages. Psychometrika, 12(2), pp. 153–157, 1947.
 [4] K. Nagao, Artificial Intelligence Accelerates Human Learning: Discussion Data Analytics. Springer, 2019.