

講義スライドへの書き起こしデータの対応付け

Automatic Alignment of Presentation Slides to Lecture Transcription

柿元 芳文[†], 山本 和英[†]

Yoshihumi Kakimoto, Kazuhide Yamamoto

1 はじめに

近年、インターネットの利用が教育の分野にも広がってきている。その一例として講義の動画や書き起こしデータを保存し、インターネットを通じて学生の復習に役立てようというものがある。

講義の書き起こしデータは話題の境界が明確でなく、そのままの形では学生の復習に役立てるのは困難である。そこで、学生の復習に役立つものとするためには講義の書き起こしデータの整形が不可欠である。整形とは、話し言葉から書き言葉への変換、要約処理、話題ごとの分割などである。

講義の書き起こしデータをそのまま復習に用いる場合、学生が復習したい部分に到達するまで非常に時間がかかる。これでは効率的な学習とは言えない。データが講義の内容ごとに分割されていれば学習の効率は向上すると考える。書き起こしデータに対して要約処理を行う場合にも前処理として書き起こしデータの話題ごとの分割は必要となる。

本稿ではスライドを用いた講義を対象に書き起こしデータの分割を行う。話題のまとまりをスライド一枚単位とした。そして同講義で使用されたスライドを話題の分割点を判断する情報として用いた。これは分割の単位を明確にし、また分割を行う上での情報を多くするためである。スライド一枚単位で書き起こしデータの分割を行うことにより学生は復習したい部分に素早くたどり着けることになる。よって学習の効率化が図れると考える。

そこで、我々は音声認識で得た講義の書き起こしデータを同講義で用いられたスライド単位で分割する手法を提案する。

2 関連研究

Hearst[1]は語彙的結束性を利用した分割手法である TextTiling 法を提案している。語彙的結束性とは、意味的に近い文には同一の語が出現しやすいというものである。TextTiling 法については3節で詳しく述べる。

TextTiling 法を基盤とした分割の改良手法がいくつか提案されている。福田ら [2] は TextTiling 法を話し言葉テキストへ対応させる手法を提案している。これはフィルターなどの不要語を削除し、左右の窓の大きさを可変にするというものである。システムが出力した分割点が正解分割点と一致しているかどうかの評価で、適合率 47.3%となっている。松井ら [3] は TextTiling 法で出力した境界候補を文の表層表現を用いて訂正する手法を提案している。平尾ら [4] は単語の IDF によって付与した文の重要度を用いている。重要度によって出力した境界候補と TextTiling 法で出力した境界候補の和集合をとることで精度の向上を図っている。しかし TextTiling 法には以下のような問題点がある。

1. 窓より小さな話題を検出することが困難である
2. 左右の窓が小さくなると類似度が顕著に小さくなる

3. 文章の先頭、末尾付近では前後同数の単語を含む窓を設定できない

上記の問題を解決するために本手法では TextTiling 法を用いず、同講義で使用されたスライドに現れる単語を主な情報として扱うことを提案する。北出ら [5] はスライドのキーワードや発話のポーズ長に注目し、マルコフモデルを利用してスライドと発話の対応付けを行っている。発話が適切なスライドに対応付けられているかどうかの評価で 55%の精度となっている。これに対して本手法ではすべての発話を分割点候補とし、スライドの枚数分の分割点を抽出する問題と考え動的計画法を用いる。

3 TextTiling 法

本稿では、TextTiling 法を比較手法として評価実験で用いる。この手法では、ある基準点 T から同数の語を含むように左右に窓を設け、窓同士の類似度を測ることで分割を行っている。類似度 $sim(T)$ は、 W_l, W_r を左右の窓に現れる単語の集合、 $f_l(w), f_r(w)$ を左右の窓における単語 w の頻度とすると、式 (1) で定義される。

$$sim(T) = \frac{\sum_{w \in \{W_l \cap W_r\}} f_l(w) f_r(w)}{\sqrt{\sum_{w \in \{W_l \cap W_r\}} f_l(w)^2 \sum_{w \in \{W_l \cap W_r\}} f_r(w)^2}} \quad (1)$$

式 (1) で得られた類似度が極小値となる基準点を境界とする。ただし、類似度の微小な揺れを無視するために式 (2) に示す $DepthScore(T)$ を用いる。

$$DepthScore(T) = \frac{sim(T-1) - sim(T)}{sim(T+1) - sim(T)} \quad (2)$$

$DepthScore(T)$ が式 (3) に示す閾値 d_{th} を越えた場合に分割点とする。ここで \bar{S} は類似度の平均、 σ は類似度の分散を表す。

$$d_{th} = \bar{S} - \sigma/2 \quad (3)$$

しかし今回は本手法と比較する目的のため任意の数の分割点を出力する必要がある。そこで本稿では $DepthScore(T)$ が高い基準点から順に出力することで任意の数の分割点を得ている。

4 提案手法

4.1 手法概要

本手法では講義の書き起こしデータと、講義で用いられたスライドの情報を用いている。これらの情報からスコアを算出し、動的計画法を用いて分割点を同定する。以下に処理の詳細を記す。

4.2 前処理

本稿で分割の対象として使用したデータは講義の書き起こしデータである。このデータは誤字、脱字の修正はされているが、そのほかは整形されておらずフィルターも含むデー

[†]長岡技術科学大学 電気系
Email: {kakimoto, ykaz}@nlp.nagaokaut.ac.jp

タである。また講義の途中で現れた無発話の期間 (ポーズ) で自動的に改行してある。図1に書き起こしデータの一部を示す。

発話ID	ポーズで改行された発話
111	えー知っててこれがまあ基本的には10で
112	えー情報として
113	茶室に与えられていますね じゃあこんだけの情報を
114	使ってですねえーと形態素解析実際にやってみましょう
115	あのこが一番最初っていうかここに入る時に見せたえーと全体の流れですねその次形態素解析をした後何やるか次構文解析っていうのを普通やりますね
116	で構文解析っていうのは

図1: 書き起こしデータの例

本稿ではこの改行されている単位を「発話」と呼び、改行と同時に付与してある昇順の番号を「発話ID」と呼ぶ。スライドは書き起こしデータと同じ講義で使用されたスライドを用いた。問題の簡単化のためにスライドは以下の条件を満たすものとした。

- (1) スライドに文字が使われている
- (2) スライドは1枚目から最後まで逐次的に説明される
- (3) スライドは、飛ばしたり戻ったりしない

上記の条件を満たさない部分は、発話、スライド共に人手で除外した。

4.3 スコアの設定

1 発話内では話題の変更は起こらないと考える。そこで、発話を分割の最小単位とする。

書き起こしデータの観察結果より、話題の切り替わりやすさの傾向として以下の二つがみられた。

傾向1 スライドを説明している発話群はそのスライドの単語を含みやすい

傾向2 スライドを切替える発話では文頭に接続詞やフィラーが発生し、一つ前の発話が終了形である

終了形とは、発話の最後の形態素の品詞が図2に属さない場合を指す。

名詞	/	感動詞	/	フィラー
未知語	/	接続詞	/	助詞 (終助詞は除く)

図2: 終了形とならない品詞

4.3.1 含有率スコア

i, j はそれぞれスライド番号、発話IDを表す。スライド i に含まれる単語の集合を W_i とする。ある範囲の発話に含まれる単語の集合を W_k とする。傾向1から含有率スコア $R(W_i, W_k)$ を設定した。図3を用いて説明する。

図3の $J_s(i)$ はスライド i の分割点の発話IDを表す。 $H(j)$ は発話IDが j の時の発話を表す。発話の総数を M 、スライドの枚数を N とする。スライド i の分割点を同定するためのスコアとして、書き起こしデータ中の単語に対するスライドの単語の含有率を用いる。 W_k を、式(4)に示し、含有率スコア $R'(W_i, W_k)$ を式(5)に示す。

$$W_k = \{w | w \in \bigcup_{j=J_s(i-1)}^M H(j)\} \quad (4)$$

$$R'(W_i, W_k) = \frac{|W_i \cap W_k|}{|W_k|} \quad (5)$$

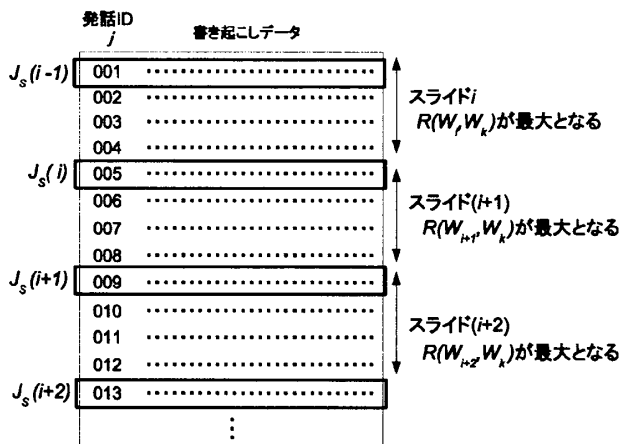


図3: 含有率スコアの特徴

単語は茶室で形態素解析した時に名詞、動詞、形容詞、未知語となったものとした。図3に示すように、含有率 $R'(W_i, W_k)$ は $J_s(i-1)$ から $J_s(i)$ の範囲の発話群を対象として算出した時に最大となると考える。ただし、式(5)では、発話群が少ないほどスコアが高くなり、傾向1に沿わないスコアを与えてしまう。短かい分割を避ける目的で式(6)に示すペナルティを与えた。 α は実験的に与えたパラメータであり、本稿では100に設定した。

$$R(W_i, W_k) = \begin{cases} R'(W_i, W_k) \times \frac{|W_k|}{\alpha} & \text{if } |W_k| \leq \alpha \\ R'(W_i, W_k) & \text{otherwise} \end{cases} \quad (6)$$

発話IDが001から003の発話群の $R(W_1, W_k)$ を計算する場合を図4を用いて説明する。アルファベットはそれぞれ単語を表している。 $R'(W_1, W_k)$ は式(5)より $R'(W_1, W_k) = 6/9 = 0.667$ となる。ただし $|W_k| \leq 100$ であるのでペナルティが加えられる。よって、式(6)より、 $R(W_1, W_k) = 0.667 * (9/100) = 0.060$ となる。

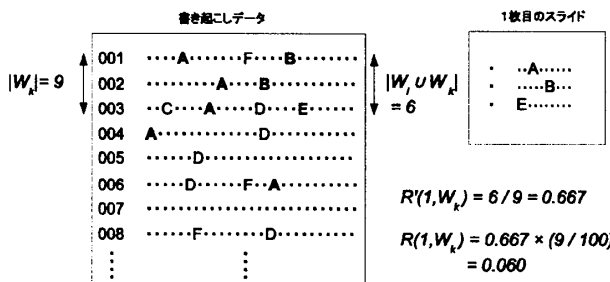


図4: 含有率スコアの計算例

4.3.2 表層表現スコア

傾向2により表層表現スコア $L(j)$ を設定する。 $L(j)$ は以下に示す場合に限って、0.01のスコアを付与した。これはスコア $R(W_i, W_k)$ に比べて十分小さな値であり、実験的に与えたものである。

- 発話の文頭が「えー」で、その前の発話が終了形の場合 (例1)
- 発話の文頭の品詞が接続詞で、その前の発話が終了形の場合 (例2)

「えー」以外のフィラーは、観察結果から分割点には出現しにくいと判断した。また例外として、接続詞「で」についてはスコアを付与していない。これは接続詞「で」が他の接続詞に比べて圧倒的に出現頻度が高く、分割点の判定としては使えないと判断したからである。ただし接続詞「で」の後に接続詞、「えー」が現れた場合はスコアを付与している(例3)。例1)

59 おー知識が必要な内容です
60 えー音声入力の利点から...

例2)

134 ... 明らかになるわけですから
135 さてそれでえーとまこういうふうに...

例3)

45 ... あまりしないことにします
46 でじゃあ機械で実現する...

4.4 動的計画法を用いた分割点の同定

今回の分割は、 M 個の分割点候補から $N - 1$ 個の分割点を抽出する問題と言える。さらに、本節で設定した2つのスコア $R(W_i, W_k)$, $L(j)$ の和が最大となる分割点で分割する問題とも言える。よって、本手法では動的計画法を用いて分割点の同定を行う。

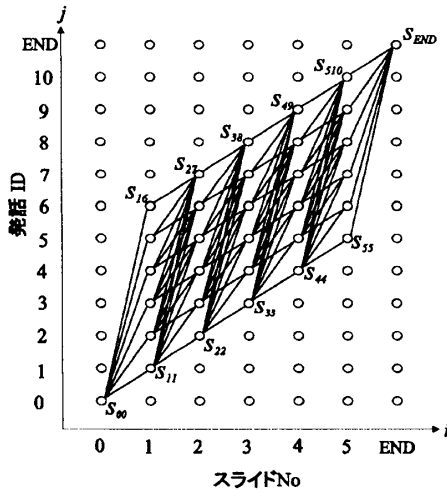


図5: 動的計画法

図5は、横軸 i がスライド番号、縦軸 j が発話 ID である。 S_{ij} は状態(分割点候補)を表す。4.3 節で示したスコアを用いて分割点の同定を行う。 $R(W_i, W_k)$ を直前の状態から接続可能な状態間(リンク)のスコア、 $L(j)$ を状態のスコアとして適用する。

S_{ij} から $S_{(i+1)j'}$ に移動する場合のスコア $Score(S_{ij}|S_{(i+1)j'})$ を式(7)に表す。ここで、 j' は直前の状態から接続可能な状態の列の発話 ID を表す。よって j' は以下の範囲となる。 $(i + 1 \leq j' \leq M - N + i + 1)$

$$Score(S_{ij}|S_{(i+1)j'}) = R(W_i, W'_k) + L(j') \quad (7)$$

$$W'_k = \{w|w \in \bigcup_{j'=J_s(i-1)}^{M-N+i+1} H(j')\}$$

上記のスコアを用いて、動的計画法により S_{00} から S_{END} へ移動するスコアが最大となる状態列を探索し、分割点を決定した。

5 評価実験

5.1 実験データ

4章で説明した提案手法を実装し、評価実験を行った。実験の対象として講義の書き起こしデータ 5 講義分と同講義で使用されたスライドを用いた。講義はすべて異なる人物によるものである。分割点の正解は人手で付与した。表1に各講義の発話数とスライド数を示す。

表1: 実験データの詳細

講義	発話数	スライド数
講義1	762	28
講義2	695	52
講義3	665	19
講義4	1487	38
講義5	480	23

5.2 評価指標

本手法との比較対象として TextTiling 法を用いた。類似度の計算に用いた単語は本手法と同じで名詞、動詞、形容詞、未知語とした。左右の窓の大きさは類似度が0とまらない最小の値とした。また窓の移動幅は10単語とした。TextTiling 法では基準点の移動を発話単位ではなく単語単位で行う。そのためその出力は発話単位にはならない。そこで、評価の際には分割点とされた基準点を含む発話をシステムの出力とした。

評価指標として式(8)に示す対応率 P_r を用いた。対応率とは発話が適切なスライドに対応付けられた割合である。

$$P_r = \frac{\text{スライドに正しく対応付けられた発話数}}{\text{全発話数}} \quad (8)$$

5.3 評価結果

本手法と TextTiling 法のそれぞれで書き起こしデータを分割した。その結果を対応率で評価した。表2に示す。

表2: 講義別にみた対応率

講義	P_r	
	本手法	TextTiling 法
講義1	0.664	0.043
講義2	0.796	0.010
講義3	0.529	0.084
講義4	0.662	0.024
講義5	0.771	0.023
全体	0.684	0.037

表2より、本手法が TextTiling 法での対応率を上回っていることが分かる。

6 考察

6.1 精度について

動的計画法を用いた本手法と語彙的結束性を用いた TextTiling 法の対応率を比較した。その結果、対応率が TextTiling 法に対して64.7ポイント向上した。このことから今回の問題に対して本手法が有効であったことがわかる。

表2より、講義によって対応率にばらつきがあることがわかる。この理由として書き起こしデータに対するスライドの単語の含有率が結果に影響を与えたと考えた。

表3に書き起こしデータに対するスライドの単語の含有率を示す。含有率は式(9)で計算した。 W_a は書き起こしデータ中の単語の集合、 W_d は全スライド中の単語の集合を表す。

$$\text{含有率} = \frac{|W_d \cap W_a|}{|W_a|} \quad (9)$$

表3: 書き起こしデータに対するスライドの単語の含有率

講義	含有率
講義1	0.434
講義2	0.613
講義3	0.416
講義4	0.478
講義5	0.408

表3より、対応率が一番高かった講義2は含有率も高い数値を示している。しかしその他の講義では含有率にそれほど差がないものの、対応率のばらつきは大きなものとなった。これは用いたスライドの単語に問題があったと考える。今回用いたスライドの単語は、名詞、動詞、形容詞、未知語とされたもの全てである。その中にはほとんどのスライドで用いられている単語もあった。よって、4.3節で述べた傾向1に沿わないスコアとなってしまう、対応率にばらつきが出たと考える。

6.2 出力分割点について

システムの出力した分割点を評価するため、式(10)に示す適合率 P_w を用いた。正解分割点の発話IDが x の時、 $x \pm n$ までを正解とした場合の適合率の推移を図6に示す。 $(0 \leq n \leq 10)$

$$P_w = \frac{\text{出力結果に含まれる正解分割点数}}{\text{出力分割点数}} \quad (10)$$

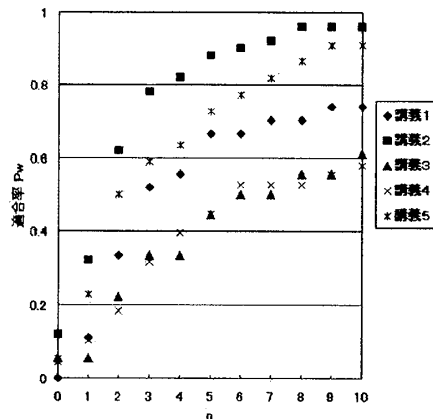


図6: 適合率の推移

図6より、適合率の変化は $n=5$ までは大きくそれ以降はなだらかになっていることがわかる。これはシステムが出力した分割点が正解の分割点付近に集まっていることを表している。よって、本稿で設定したスコアが分割点を探索するのに有効であったと言える。

6.3 リンクのスコアについて

4.3.1節で、 $R(W_i, W_k)$ は $J_s(i-1)$ から $J_s(i)$ の範囲の発話群を対象とした時に最大となると考えた。しかし実際には最大とはならなかった。書き起こしデータ中に、スライ

ドの単語以外の単語数が多かったため発話数が増えるほどスコアは減少していく傾向が見られた。これでは傾向1に沿った最適なスコアとは言えない。そこで W_k に対して前処理を行うことで最適なスコアに近付ける事が出来ると思われる。前処理としては、一般的な語を除き重要語を抽出することなどが挙げられる。

7 今後の課題

今後の課題として大きく二つ挙げることが出来る。一つは6.3節で触れたように使用する単語の選択を行う手法の考案である。単語を品詞で選択するのではなく $tf*idf$ 等を用いてスコア付けし、スコアの上位の語のみをスコアの算出に用いる。これにより4.3.1節の傾向1に沿った最適なスコアに近付けることが出来ると思われる。もう一つは表層表現スコアの信頼度を向上させることが挙げられる。分割点となる発話から特徴語を収集し、その特徴語からスコアを付与することで、より信頼できるスコアになると考える。講義の書き起こしデータからだけでは十分な量の特徴語を収集できない可能性がある。よって講義の書き起こしデータからだけでなく、話し言葉コーパス全般から特徴語を収集することも有効な手段だと考える。収集した特徴語を基にスコアを付与することにより、分割点となりやすい発話ほど高いスコアを与えることが出来ると思われる。

8 おわりに

本稿では講義の書き起こしデータと同じ講義で用いられたスライドの単語に注目した。そしてそれを基にスコアを算出し動的計画法を用いて話題の分割を行う手法を提案した。その結果、68.4%の対応率を得た。またTextTiling法と比較した結果、対応率が64.7ポイント向上した。今後の課題として、より精密な分割点の同定を行うために重要語の抽出や機械学習をスコアの算出に組み込む事が挙げられる。

謝辞

本研究の一部は、平成17-19年度総務省戦略的情報通信研究開発推進制度(SCOPE)の支援によって実施した。

使用した言語資源及びツール

- (1) 形態素解析器“ChaSen”, Ver.2.3.3, 奈良先端科学技術大学院大学 松本研究室,
<http://chasen.naist.jp/hiki/ChaSen/>

参考文献

- [1] M.A.Hearst: TextTiling:Segmenting Text into Multiparagraph Subtopic Passages: Computational Linguistics, vol.23, no.1, pp.33-64, 1997.
- [2] 福田 雅志, 延澤 志保, 太原 育夫: 語彙的結束性に基づく話し言葉のテキストセグメンテーション: 言語処理学会第11回年次大会発表論文集, pp.620-623, 2005.
- [3] 松井 祥峰, 乾 伸雄, 小谷 善行: 単語の結束度と文の表層情報を組み合わせたテキストセグメンテーション: 情報処理学会研究報告, NL-162-22, 2004.
- [4] 平尾 努, 北内 啓, 木谷 勉: 単語重要度と語彙的結束性を利用したテキストセグメンテーション: 情報処理学会研究報告, NL-130-6, 1999.
- [5] 北出 裕, 河原 達也: 講義の自動アーカイブ化のためのスライドと発話の対応付け: 情報処理学会研究報告, SLP-55-11, 2005.
- [6] 西澤 信一郎, 中川 裕志: 名詞の文書内頻度を利用したテキストセグメンテーション: 情報処理学会研究報告, NL-117-20, 2005