

係り受け情報と残存文長を考慮した講演テキストへの逐次的な改行挿入

高橋 晨成¹ 大野 誠寛¹ 松原 茂樹²

¹ 東京電機大学 大学院未来科学研究科 ² 名古屋大学 情報基盤センター

23fmi15@ms.dendai.ac.jp ohno@mail.dendai.ac.jp

matsubara.shigeki.z8@f.mail.nagoya-u.ac.jp

概要

講演を対象とした字幕生成システムにおいて、講演文特有の長い文が複数行にまたがって表示されると読みやすさが低下するため、適切な位置に改行を挿入し、読みやすい字幕を生成する必要がある。本稿では、読みやすい字幕を生成するための要素技術として、文末を未知とした問題設定において、漸進的係り受け解析結果と残存文長推定結果を考慮した逐次的な改行挿入手法を提案する。

1 はじめに

聴覚障害者や高齢者、外国人らの講演音声の聞き取りや理解を支援するために、字幕を生成することが望まれている。講演では1文が長くなる傾向があり、複数行にまたがって表示される場合、字幕を読みやすく表示するために適切な位置に改行を挿入する必要がある[1]。

これまでに、改行挿入手法がいくつか提案されている[2, 3, 4, 5, 6]。従来研究[2]は、入力された1文全体を考慮して適切な改行位置を求める手法を提案している。しかし、1文全体の発話が終わるまで改行位置の推定を行えないため、講演におけるリアルタイムでの字幕生成には必ずしも適さない。それに対し従来研究[3]は、文節が入力されるたびに、その直前の文節との間に改行を挿入するか否かを最大エントロピー法を用いて、逐次的に推定する手法を提案している。しかし、使用できる係り受け情報が限定される、また、未入力部分の情報を使えないという制約があり、1文全体の情報を用いて改行挿入する従来手法[2]と比べて、その精度は低い。

逐次的な処理における情報の制約を緩和する試みとして、従来研究[4]では、漸進的係り受け解析手法を提案するとともに、解析で得られる係り受け情

報を使用した逐次的な改行挿入手法を提案している。従来研究[4]で提案されている漸進的係り受け解析手法では、文節が入力されるごとに、既入力文節の係り先が他の既入力文節のいずれであるか、あるいは、未入力文節であるかを同定し出力することができ、この解析結果から得られる係り受け情報に関する素性を用いた逐次的な改行挿入手法（以下、従来手法[4]）が同時に提案されている。さらに、未入力部分の情報を補う試みとして、従来研究[5]は、文の残りの長さである残存文長が短いほど改行の必要性は低下すると考え、BERT[7]を用いて、残存文長を推定しつつ、それと同時に改行を挿入するか否かを逐次的に判定する手法を提案している。

両手法[4, 5]を踏まえた従来研究[6]では、逐次的な改行挿入において、事前に解析した漸進的係り受け解析結果を使用するとともに残存文長の同時推定を行う手法を提案している。しかしながら、従来手法[4, 6]では、改行挿入判定の前に別途、文単位での係り受け解析を行い、その解析結果を用いており、文末を既知として扱っている。実用上、必ずしも文末を既知として扱えるとは限らないため、文末を未知として扱うことが望ましい。

そこで本稿では、文末を未知とした問題設定において、漸進的係り受け解析結果と残存文長推定結果を考慮した逐次的な改行挿入手法を提案する。提案手法では、各文節が入力されるごとに、改行挿入判定と残存文長推定をBERTを用いて同時実行する。また、その際の改行挿入判定において、文末情報なしで別のBERTにより推定した係り受け情報を用いることにより、精度向上を試みる。

2 従来の逐次的改行挿入手法

逐次的改行挿入に関する従来手法[3, 4, 5, 6]では、共通して、形態素情報、文節まとめ上げ、節境界

表 1 提案手法と従来手法 [3, 4, 5, 6] の位置づけ

	文末	$f_{1\sim 11}$	f_{12}	$f_{1\sim 3}^A$	残存 文長	モデル
従来手法 [3]	未知	○	×	×	×	ME
従来手法 [4]	既知	○	○	×	×	ME
従来手法 [5]	未知	○	×	×	○	BERT
従来手法 [6]	既知	○	○	○	○	BERT
提案手法	未知	△	○	×	○	BERT

解析, 係り受け解析が施された文節列を入力とし, $i+1$ 番目の文節 b_{i+1} が入力されるたびに, その直前の文節との境界, すなわち b_i の直後に改行を挿入するか否かを逐次的に判定するという問題設定を用いている. また, 共通して, ディスプレイの大きさを考慮して 1 行の最長文字数を 20 字と設定し, 行頭から文節 b_{i+1} までの文字数が最長文字数を超える場合には, 判定結果に関わらず強制的に改行する.

従来手法 [3, 4, 5, 6] の相違点を以下の各節で説明する. また, 提案手法との位置づけを表 1 に示す.

2.1 基本素性に基づく逐次的改行挿入

従来手法 [3] では, 文末を未知であるとして 1 講演の文節列を入力とし, 改行を挿入するか否かを最大エントロピー法 (ME) を用いて判定している. ME の素性には, 下記 11 種の素性を用いている.

- ・素性 $f_1 \sim f_4$: 文節 b_i , 文節 b_{i+1} の形態素情報.
- ・素性 $f_5 \sim f_6$: 文節 b_i 直後の節境界情報.
- ・素性 $f_7 \sim f_9$: 文節 b_i 前後の係り受け情報.
- ・素性 f_{10} : 行頭から文節 b_i までの文字数.
- ・素性 f_{11} : 文節 b_i 直後のポーズ情報.

ここで, 素性 $f_7 \sim f_9$ は, 隣接文節に係るか否かの解析結果から得られる素性であり, 文節列間の係り受け構造を利用しているわけではない.

2.2 漸進的係り受け解析結果を用いた逐次的改行挿入

従来手法 [4] では, 漸進的係り受け解析の結果に基づく素性 f_{12} を ME の素性に追加した改行挿入手法を提案している. ただし, 従来手法 [3] では文末を未知としていたのに対し, 従来手法 [4] では文末を既知として扱い, 1 文の文節列を入力とするとともに, 学習と評価から文末となる文節を除外している. 従来手法 [4] で使用されている漸進的係り受け解析手法では, 文節が入力されるごとに既入力文節の係り先が他の既入力文節のいずれであるか, あるいは, 未入力文節であるかを同定し出力する. この出力を元にして, 素性 $f_7 \sim f_9$ のみならず, 「行頭から文節 b_i までの間で係り受けが閉じているか否か」

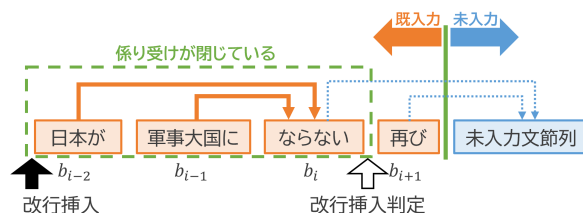


図 1 素性 f_{12} を判定する際に用いる係り受け構造

という素性 f_{12} を抽出し, 従来手法 [3] の素性に加えて用いている. 図 1 は, 講演の文節列の一部「…/日本が/軍事大国に/ならない/再び/…」中の文節 b_{i+1} 「再び」が入力され, 文節 b_i 「ならない」の直後に改行を挿入するか否かを推定するとき, 素性 f_{12} を判定する際に参照する係り受け構造である. 文節 b_{i-2} , 文節 b_{i-1} が文節 b_i に係ることから, 文節 b_{i-2} から文節 b_i までの文節列は係り受けが閉じていると判定する.

2.3 残存文長を考慮した逐次的改行挿入

従来手法 [5] では, 従来手法 [3] と同様に文末を未知であるとして 1 講演の文節列を入力とし, 残存文長推定と改行挿入判定を同一の BERT モデルで同時に実行する手法を提案している. 飯泉ら [5] は河村らの既存研究 [8] と同様に, 文 s が n_s 個の文節から成り, 文頭から i 番目の文節 b_i まで既に入力されているときの残存文長 $RL(s, i)$ を $RL(s, i) = n_s - i$ により定義している. 文節 b_{i+1} が入力されたときに, 改行挿入判定を行うのと同時に文節 b_i の残存文長 $RL(s, i)$ を推定することで, 残存文長を考慮した改行挿入判定を実現している.

2.4 漸進的係り受け解析結果と残存文長を考慮した逐次的改行挿入

従来手法 [6] では, 素性 f_{12} と音響情報に基づく 3 種類の素性 $f_{1\sim 3}^A$ を追加し, 残存文長推定と改行挿入判定を同一の BERT モデルで同時に実行する手法を提案している. 係り受けに関する素性 $f_7 \sim f_9$, 及び, f_{12} については, 従来手法 [4] と同様に漸進的係り受け解析の結果を利用して取得している. また, 文末に関しても従来手法 [4] と同様に既知として扱い, 1 文の文節列を入力とするとともに, 学習と評価から文末となる文節を除外している.

3 提案手法

提案手法では, 従来手法 [3, 5] と同様に文末を未知であるとして 1 講演の文節列を入力とし, 漸進的

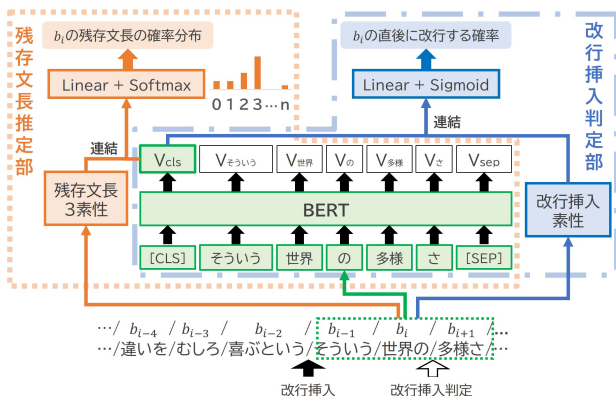


図2 提案手法の概要

係り受け解析と残存文長推定を考慮して、逐次的に改行挿入を行う。

2章で挙げた各逐次的改行挿入手法では、それぞれ以下の情報を考慮していた。

- ・素性 $F^B = \{f_1, \dots, f_6\}$: 従来手法 [3] の素性のうち、形態素情報と節境界情報の素性。
- ・素性 $f_{1-3}^D = \{f_7, f_8, f_9\}$: 従来手法 [3] の素性のうち、係り受け情報の素性。
- ・素性 $f_4^D = \{f_{12}\}$: 従来手法 [4] の漸進的係り受け解析結果に基づく素性。
- ・素性 $F^L = \{f_{10}\}$: 従来手法 [3] の素性のうち、行頭からの文字数の素性。
- ・素性 $F^P = \{f_{11}\}$: 従来手法 [3] の素性のうち、ポーズ情報の素性。
- ・素性 f_{1-3}^A : 従来手法 [6] の音響情報に基づく素性。
- ・残存文長 RL: 従来手法 [5] が改行挿入判定と同時的に推定する残存文長。

本稿では、新たに改行挿入手法を提案するに際して素性の整理を行った。具体的には、従来研究 [6] の結果から、有効性の確認できない素性 F^A を不使用とした。また、BERT における改行挿入判定において素性 F^B の情報は、BERT の事前学習済みモデルが持つ情報との重複が考えられるため、有効性を確認する事前実験を行った。その結果、素性 F^B の有効性は確認できなかったため、不使用とした。そのため、本稿では素性 F^D , F^L , F^P , 残存文長 RL の情報を考慮した手法を提案する。

提案手法の概要を図2に示す。図2は、講演の文節列の一部「…違いを/むしろ/喜ぶという/そういう/世界の/多様さ/…」中の文節 b_{i+1} 「多様さ」が入力されたとき、文節 b_i 「世界の」の直後に改行を挿入する確率と残存文長 $RL(s, i)$ の確率分布を推定する様子である。ただし、それまでの改行挿入判定結

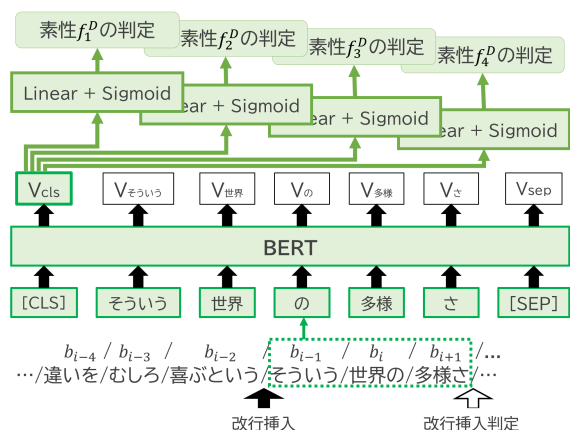


図3 BERT による係り受け素性の判定

果において、直近の改行挿入位置は文節 b_{i-2} 「喜ぶという」の直後であるとした際の様子である。

提案手法では、それまでの改行挿入判定結果において、直近の行頭文節となった文節（図2では「そういう」）から入力された文節（図2では「多様さ」）までのサブワード列を BERT への入力として用いる。また、改行挿入部では、素性 F^D , F^L , F^P からなる 6 次元のベクトルを、BERT の <CLS> の入力に対応した出力（図2では V_{cls} ）と連結したものを Linear+Sigmoid に入力することで、改行挿入の確率を出力する。残存文長推定部では、従来手法 [5] と同様に、 b_i の直後にポーズ、フィラー、言い淀みのそれぞれが出現しているか否かを表した 3 次元のベクトルを用意し、これを V_{cls} と連結したものを Linear+Softmax に入力し、その出力を残存文長の確率分布とする。

3.1 BERT による係り受け素性の判定

素性 F^D については、文末を未知として扱うため、文を入力とする漸進的係り受け解析手法 [4] を用いて取得することはできない。そのため、提案手法では、図3に示す BERT により係り受け素性の判定を事前に行い、その結果を使用する。BERT への入力は、提案手法と同様に直近の行頭文節となった文節から入力された文節までのサブワード列を用いる。係り受け素性 F_{1-4}^D は同一のモデルを用いて判定しており、その出力を Linear+Sigmoid に入力することで、各係り受け素性を判定している。

4 評価実験

提案手法の有効性を評価するために、日本語講演データを用いて改行挿入実験を行った。

表2 提案手法と比較手法の位置づけ

	文末	F^B	f_{1-3}^D	f_4^D	F^L	F^P	RL	モデル
従来手法 [5]	未知	○	○	×	○	○	○	BERT
[BERT]	未知	×	×	×	×	×	×	BERT
提案手法	未知	×	○	○	○	○	○	BERT

4.1 実験概要

実験データには、同時通訳データベース [9] に収録されている音声データと、日本語講演音声の書き起こしデータを使用した。なお、書き起こしデータの全てのデータに形態素情報、節境界情報、係り受け情報、改行位置が人手で付与されている。実験は、16 講演を用いた 16 分割交差検定により実施した。ただし、16 講演のうち 2 講演については、開発データとして使用するため評価データから取り除き、残りの 14 講演に対して評価を行った。

比較手法として以下を用意した。提案手法と比較手法の位置づけを表 2 に示す。

- ・従来手法 [5]：従来手法 [5] の改行挿入手法。
- ・[BERT]：提案手法において、素性 F^D , F^L , F^P を使用せず、残存文長の同時推定を行わない手法。

各手法では、モデルは PyTorch¹⁾ を用いて実装し、BERT の事前学習モデルは東北大学の公開モデル²⁾ を用いた。学習アルゴリズムは AdamW を採用した。また、学習率スケジューラは Hugging Face Transformers ライブラリ³⁾ の `get_linear_schedule_with_warmup` を用いた。ウォームアップステップ数は全体の 10% とした。ハイパーパラメータは、開発データを用いて予備実験を行い、F 値が最大となったものを使用した。比較手法 [BERT] に関しては、バッチサイズを 32、学習率を $4e-5$ 、エポック数を 19 とした。提案手法に関しては、バッチサイズを 32、学習率を $4.3e-5$ 、エポック数を 13 とした。

評価では、各手法について、同一実験を 10 回繰り返し、再現率、適合率、F 値の平均値を測定した。

4.2 実験結果

実験結果を表 3 に示す。提案手法は従来手法 [5] と比較して、再現率では下回ったものの、適合率、F 値では上回った。提案手法は比較手法 [BERT] と比較して、再現率、適合率、F 値のすべてにおいて上回った。提案手法と比較手法 [BERT] の F 値を比

表3 実験結果

	再現率 (%)	適合率 (%)	F 値
従来手法 [5]	83.69	73.24	78.11
[BERT]	80.69	78.45	79.56
提案手法	81.19	78.57	79.86

提案手法の出力(正解)	映像の調子はまだ終わっていない 私は不安でたまりませんでした
比較手法[BERT]の出力(不正解)	映像の調子はまだ終わっていない私は 不安でたまりませんでした

図4 提案手法が正解し、比較手法が不正解の例

較したところ、有意差が認められた ($p < 0.05$)⁴⁾。このことから、文末が未知である場合においても、漸進的係り受け解析結果と残存文長推定結果をともに考慮することの有効性を確認した。

4.3 考察

文末を未知とした問題設定において、漸進的係り受け解析結果と残存文長推定とをともに考慮することによる好影響を考察する。提案手法が正解し、比較手法 [BERT] が不正解となった例を図 4 に示す。提案手法では、「終わっていない」と「私は」の間に正しく改行を挿入できているが、比較手法 [BERT] では不正解となった。BERT による f_4^D の判定結果は、文節列「映像の調子はまだ終わっていない」は係り受けが閉じていると判定されている。残存文長推定の結果は、従来研究 [5, 8] と同様に、確率分布の期待値が「0, 1, 2~3, 4 以上」のいずれに属するかを求めると、図 4 の例では、残存文長の推定結果は「0」であり、「終わっていない」の後に続く文節が無いという情報を利用できた。そのため、入力された文節列の間に構文的なまとまりがあることや「終わっていない」が文末の文節であることを考慮でき、適切な位置で改行できたものと考えられる。

5 おわりに

本稿では、文末を未知とした問題設定において、漸進的係り受け解析結果と残存文長推定結果を考慮した逐次的な改行挿入手法を提案した。実験の結果、文末が未知である場合においても、漸進的係り受け解析結果と残存文長推定結果をともに考慮することの有効性を確認した。今後は、未入力文節の具体的な予測結果などの活用を検討し、更なる精度向上を図りたい。

1) <https://pytorch.org>

2) <https://github.com/cl-tohoku/bert-japanese>

3) <https://huggingface.co/docs/transformers/index>

4) 評価用データ 14 講演の各講演ごとに F 値を算出し、対応のある t 検定を行った。

謝辞

本研究は JSPS 科研費 JP19K12127, JP24K15076 の助成を受けたものです。

参考文献

- [1] 中野聡子, 金澤貴之, 牧原功, 黒木速人, 上田一貴, 井野秀一, 伊福部達. 聴覚障害者向け音声同時字幕システムの読みやすさに関する研究: (1)-改行効果に焦点をあてて. ヒューマンインタフェース学会論文誌, Vol. 10, No. 4, pp. 435–444, 2008.
- [2] 村田匡輝, 大野誠寛, 松原茂樹. 読みやすい字幕生成のための講演テキストへの改行挿入. 電子情報通信学会論文誌, Vol. J92-D, No. 9, pp. 1621–1631, 2009.
- [3] 大野誠寛, 村田匡輝, 松原茂樹. 講演のリアルタイム字幕生成のための逐次的な改行挿入. 電気学会論文誌, Vol. 133-C, No. 2, pp. 418–426, 2013.
- [4] 大野誠寛, 松原茂樹. 文節間の依存・非依存を同定する漸進的係り受け解析. 電子情報通信学会論文誌, Vol. J98-D, No. 4, pp. 709–718, 2015.
- [5] 飯泉智朗, 大野誠寛, 松原茂樹. 残存文長を考慮した講演テキストへの逐次的な改行挿入. 言語処理学会 第 28 回年次大会 発表論文集, Vol. 2022, No. 1, pp. 2061–2065, 2022.
- [6] 高橋晨成, 大野誠寛, 松原茂樹. 漸進的係り受け解析と残存文長推定に基づく講演文への逐次的な改行挿入. 言語処理学会 第 30 回年次大会 発表論文集, Vol. 2024, No. 1, pp. 3192–3196, 2024.
- [7] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In **Proceedings of the 2018 Annual Conference of the North American Chapter of the Association for Computational Linguistics**, pp. 4171–4186, 2018.
- [8] 河村天暉, 大野誠寛, 松原茂樹. 漸進的な言語処理のための独話文に対する残存文長の推定. 情報処理学会 第 82 回全国大会講演論文集, Vol. 2020, No. 1, pp. 447–448, 2020.
- [9] Shigeki Matsubara, Akira Takagi, Nobuo Kawaguchi, and Yasuyoshi Inagaki. Bilingual spoken monologue corpus for simultaneous machine interpretation research. In **Proceedings of the 3rd International Conference on Language Resources and Evaluation**, pp. 153–159, 2002.