

漸進的係り受け解析結果を用いた講演文への逐次的な改行挿入

高橋 晨成^{†, a)} 大野 誠寛^{†, b)} 松原 茂樹[‡]
 東京電機大学未来科学部[†] 名古屋大学情報連携推進本部[‡]

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。しかし講演では一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、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。これまでも、改行挿入手法がいくつか提案されているが[1, 2], 特にリアルタイムな改行挿入には精度向上の余地が残されている。

そこで本稿では、読みやすい字幕を生成するための要素技術として、漸進的係り受け解析結果を用い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を挿入するか否かを判定する際に、BERT[3]の出力とともに、漸進的係り受け解析から得られる構文情報を用いる。本稿の貢献は、改行挿入においてBERTを用いる際に、漸進的係り受け解析結果を併用することの有効性を示した点にある。

2 従来の逐次的改行手法

従来手法[1]では形態素情報、文節まとめ上げ、節境界解析、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の判定を最大エントロピー法(ME)を用いて逐次的に行う。MEの素性には、 b_i の主辞、語形などの形態素情報4種 $f_1 \sim f_4$ 、節境界情報2種 $f_5 \sim f_6$ 、係り受け情報3種 $f_7 \sim f_9$ 、行頭からの文字数1種 f_{10} 、ポーズ情報1種 f_{11} の合計11種を使用している。ただし、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、 b_i の直後に改行を挿入しなければ最長文字数を超える場合には、MEの判定結果に関わらず強制的に改行を挿入する。

従来研究[2]では、漸進的係り受け解析手法が提案され、その解析結果をMEの素性に利用した改行挿入手法(以下、従来手法[2])が従来手法[1]よりも高精度な改行挿入を実現することを示している。その漸進的係り受け解析手法[2]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。従来手法[2]では、従来手法[1]の11種の素性に加えて、「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」という素性 f_{12} を追加し改行を挿入するか否かを判定している。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} の抽出時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。この素性 f_{12}

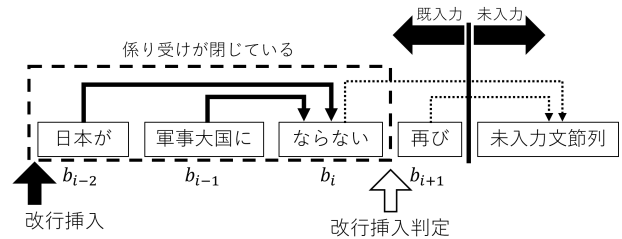


図1: 素性 f_{12} を判定する際に用いる係り受け構造

を用いることにより、構文的なまとまりの情報を利用した改行位置の推定が行える。

3 漸進的係り受け解析結果を用いた改行挿入

提案手法では、従来手法[2]と同じく、形態素情報、文節まとめ上げ、節境界解析、漸進的係り受け解析が施された文節列を入力とし、文節が入力されるごとに改行を挿入するか否かを逐次的に判定する。その判定では、BERTの出力と従来手法[2]と同じ素性(すなわち、従来手法[1]における素性 $f_1 \sim f_{11}$ と従来手法[2]独自の素性 f_{12} を加えた12種類^{*1)})を用いる。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「ご存じの/通り/わが国では/日本が/軍事大国に/ならない/再び/…」中の文節 b_{i+1} 「再び」が入力されたとき、文節 b_i 「ならない」の直後に改行を挿入する確率を推定する様子である。ただし、それまでの改行挿入結果において、改行が挿入されたのは、文節 b_{i-3} 「わが国では」の直後が直近の改行位置であったとした推定の様子である。

提案手法では、それまでの改行挿入によって直近の行頭文節となった文節(図2では「日本が」)から判定対象の文節(図2では「ならない」)までのサブワード列(図2では「[CLS] 日本が 軍事大国に ならない [SEP]」)をBERTの入力として用いる。また12種類の素性 $f_1 \sim f_{12}$ を12次元のベクトルとして用意し、BERTの[CLS]に対応した出力(図2では V_{cls})と連結したものをLinear + Sigmoidに入力し、改行を挿入する確率を出力する。この確率が0.5以上であれば改行を行う。

なお、従来研究[2]では、最ゆう推定手法と決定的手法の2種類の漸進的係り受け解析手法が提案されており、改行挿入に応用した場合、最ゆう推定手法の結果を利用した方が高精度に改行挿入できることが報告されている。そのため本研究では、 b_i の直後の改行挿入有無の判定において、文節 b_{i+1} が入力されたときに最ゆう推定手法が解析した係り受け構造を利用する。

4 改行挿入実験

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。

4.1 実験概要

実験データとして、同時通訳データベース[4]に収録されている日本語講演音声の書き起こしデータを使用した。なお、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。

*1なお、係り受け情報3種 $f_7 \sim f_9$ は、従来手法[2]と同様に、係り受け解析[2]の結果を利用して取得した。

Sequential Linefeed Insertion into Lecture Transcription using Incremental Dependency Parsing Results

Shinsei Takahashi^{†, a)}, Tomohiro Ohno^{†, b)}, Shigeki Matsubara[‡]

[†] School of Science and Technology for Future Life, Tokyo Denki University.

[‡] Information and Communications, Nagoya University

a) 19fi067@ms.dendai.ac.jp

b) ohno@mail.dendai.ac.jp

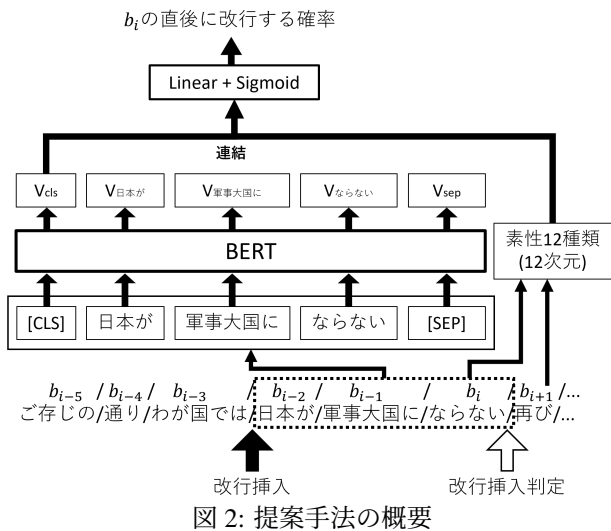


図2: 提案手法の概要

実験は、14 講演を用いた交差検定により実施した。すなわち、1 講演をテストデータとし、残りの 13 講演を学習データとする実験を 14 回繰り返した。ただし、14 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 12 講演に対して評価を行った。評価には、正解データの改行位置に対する再現率、適合率、F 値を用いた。

提案手法 [BERT+ $f_{1-11} + f_{12}$] との比較手法として以下を用意した。

- [BERT]: 提案手法において、素性 $f_1 \sim f_{12}$ を使用しない手法。
- [BERT+ f_{1-11}]: 提案手法において、素性 f_{12} を使用しない手法。

各手法では、モデルは PyTorch^{*1} を用いて実装し、BERT の事前学習モデルは東北大学が公開しているモデル^{*1} を用いた。学習アルゴリズムは SGD を採用した。パラメータの更新はミニバッチ学習 (学習率 0.002, バッチサイズ 16) により行った。エポック数は [BERT] を 26, [BERT+ f_{1-11}] を 37, [BERT+ $f_{1-11} + f_{12}$] を 51 ととした。なお、エポック数は開発データを用いて予備実験を行い、F 値が最大となる値を採用した。

4.2 実験結果

実験結果を表 1 に示す。提案手法は再現率、適合率、F 値のいずれも比較手法を上回った。[BERT] と [BERT+ f_{1-11}] を比較すると、[BERT+ f_{1-11}] は F 値において上回っているものの、有意差は認められなかった ($p \geq 0.05$)。一方、[BERT+ $f_{1-11} + f_{12}$] と [BERT+ f_{1-11}] を比較すると、提案手法は F 値が約 3 ポイント上昇しており、有意差が認められた ($p < 0.01$)。

以上より、BERT の出力だけでなく、漸進的係り受け解析結果の情報を用いることの有効性を確認した。

4.3 考察

まず、BERT の出力に加えて漸進的係り受け解析結果を用いることによる好影響を考察する。提案手法が正解し、[BERT+ f_{1-11}] が不正解となった例を図 3 に示す。提案手法では、「まきこまれる」と「ちょっと」の間に正しく改行を挿入できているが、[BERT+ f_{1-11}] では不正解となり「ちょっと」と「悲しい」の間に余分な改行が挿入されている。漸進的係り受け解析は、「また」と「ラッシュに」が「まきこまれる」に係るという係り受け構造を出力しており、提案手法は「またラッシュ

表 1: 実験結果

	再現率 (%)	適合率 (%)	F 値
[BERT]	80.04 (4,859/6,071)	68.59 (4,859/7,084)	73.87
[BERT+ f_{1-11}]	80.55 (4,890/6,071)	70.37 (4,890/6,949)	75.12
[BERT+ $f_{1-11} + f_{12}$]	83.63 (5,077/6,071)	73.28 (5,077/6,928)	78.11

提案手法 の出力(正解)	またラッシュにまきこまれるという ちょっと悲しい事態になってしまいますので
[BERT+ f_{1-11}] の出力(不正解)	またラッシュにまきこまれるというちょっと 悲しい事態になってしまいますので

図3: 提案手法が正解, [BERT+ f_{1-11}] が不正解の例

提案手法 の出力(不正解)	仲人とかに挨拶に行くことが できずに どうしようという形でした
正解の出力	仲人とかに挨拶に行くことができずに どうしようという形でした

図4: 提案手法が不正解の例

にまきこまれる」の係り受けが閉じている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を考慮でき、適切な位置で改行できたものと考えられる。

次に、悪影響を考察する。提案手法が不正解となった例を図 4 に示す。提案手法では「行くことが」と「できずに」の間に、正解にはない余分な改行が挿入されている。「行くことが」と「できずに」の間の改行挿入判定時において、漸進的係り受け解析の結果では「仲人とかに」と「挨拶に」が「行くことが」に係りとなっており、提案手法は「仲人とかに挨拶に行くことが」の係り受けが閉じているという情報を利用できたこととなる。そのため、当該文節列が構文的なまとまりがあることを考慮し、改行を挿入すると判定したものと考えられる。しかし正解の改行では、「仲人とかに」から「できずに」までの文節列の係り受けが閉じていることを考慮し、この位置に改行が挿入されていない。正解は、1 文全体での適切な改行位置が決められているのに対して、逐次的な改行挿入では、後方文脈を見ることはできない中で改行位置を判定する必要がある。このような例に対する改善方法は今後の課題である。

5 おわりに

本論文では、漸進的係り受け解析結果を用いた講演文への逐次的な改行挿入手法を提案した。実験の結果、逐次的な改行挿入において BERT に加えて漸進的係り受け解析結果を用いることの有効性を確認した。今後は、漸進的係り受け解析結果をより効果的に利用できる深層学習モデルについて検討し、更なる精度向上を図りたい。

謝辞 本研究は、一部、科学研究費補助金基盤研究 (C) No. 19K12127 により実施した。

参考文献

- [1] 大野ら, “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,” 電学論, Vol. 133-C, No.2, pp.418–426, 2013.
- [2] 大野ら, “文節間の依存・非依存を同定する漸進的係り受け解析,” 信学論, Vol. J98-D, No.3, pp.709–718, 2015.
- [3] J. Devlin et al., “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” Proc. ACL2018, pp.4171–4186, 2019.
- [4] S. Matsubara et al., “Bilingual Spoken Monologue Corpus for Simultaneous Machine Interpretation Research,” Proc. LREC2002, pp.154–159, 2002.

^{*1}<https://pytorch.org/>

^{*1}<https://github.com/cl-tohoku/bert-japanese>