

マルチタスク学習による講演テキストへの 読点と改行の自動挿入

方之成[†] 村田 匡輝[‡] 松原 茂樹[†]

[†] 名古屋大学大学院情報学研究科 [‡] 豊田工業高等専門学校

1 はじめに

講演などの音声を、音声認識技術を用いて自動でテキスト化し、字幕として提示することで、聴衆による講演の理解を支援することができる。しかし、講演では一文が長くなる傾向にある。文全体がスクリーン上で複数行にまたがって表示されると、語句の区切りがわかりにくくなることもあり、結果的に読みにくい字幕となる。このため、字幕テキストでは、適切な位置に読点と改行が挿入されていることが望ましい。

これまで、テキストへの読点挿入 [1, 2]、ならびに、改行挿入 [3] の研究がそれぞれ行われている。また、村田らは、読点と改行の用法に着目した機械学習モデルにより、読点と改行を順次的にテキストに挿入する手法を提案している [4]。しかし、適切な読点の挿入位置と改行の挿入位置は相互に影響しあうと考えられるため、読点と改行を相互作用的に挿入することが望ましい。

そこで本稿では、講演会場において読みやすい字幕を提示することを目的として、日本語講演テキストへ読点と改行を同時に挿入する方法を提案する。本手法では、読点挿入と改行挿入の2つのタスクをマルチタスクとしてBERT[5]により実行し、文中のそれぞれの挿入位置を同定する。

日本語講演データを用いた実験の結果、読点挿入、改行挿入のいずれもベースラインより性能が向上しており、本手法の有効性を確認した。

2 提案手法

本手法では文を処理の単位とする。モデルの構造を図1に示す。トークン x_i とその前後 n 個のトークンを並べた X_i がモデルへの入力となり、 x_i は文の最初のトークンから最後のトークンまで一つずつ推移する。 X_i に開始位置と終了位置を示

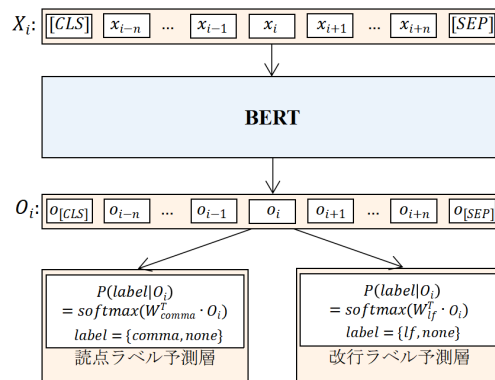


図1 モデルの構造

す [CLS] と [SEP] の2つの特徴トークンを追加したうえで、BERTによりベクトル列 O_i に変換し、それぞれ読点ラベル予測層と改行ラベル予測層へ入力する。

読点ラベル予測層では、 $softmax$ 関数を使用して、 $label = \{comma, none\}$ の確率を計算する。 $comma$ はトークン x_i の直後に読点が挿入されることを、 $none$ は挿入されないことを意味する。また、改行ラベル予測層でも同様に $label = \{lf, none\}$ の確率を計算する。

W_{comma} , W_{lf} はそれぞれ読点ラベル予測層と改行ラベル予測層の独立パラメータである。

3 実験

3.1 実験概要

同時通訳データベース (SIDB) [6] の日本語講演音声の書き起こしデータ 16 講演分を用いて読点、改行挿入実験を実施した。書き起こしテキストには一行の文字数が 20 文字以下になるように、読点と改行の正解データが人手で付与されている。2 講演を検証データ、残りの 14 講演のうち 13 講演を学習データ、1 講演をテストデータとする交差検証を実施した。

BERT には東北大学の事前学習済み日本語 BERT モデル¹、オプティマイザーは Adam、損失関数は Cross Entropy Loss を用いた。ハイパーパ

Automatic insertion of commas and linefeeds into lecture transcriptions based on multi-task learning

Zhicheng Fang[†], Masaki Murata[‡], Shigeki Matsubara[†]

[†]Graduate School of Informatics, Nagoya University

[‡]National Institute of Technology, Toyota College

¹<https://github.com/cl-tohoku/bert-japanese/tree/v1.0>

表1 読点挿入の実験結果

	再現率	適合率	F 値
提案手法	76.69% (3,103/4,046)	79.71% (3,103/3,893)	78.17
村田らの手法 [1]	71.65% (2,899/4,046)	81.07% (2,899/3,567)	76.07
ベースライン	70.17% (2,839/4,046)	76.44% (2,839/4,004)	73.17

表2 改行挿入の実験結果

	再現率	適合率	F 値
提案手法	76.08% (3,836/5,042)	74.78% (3,836/5,130)	75.42
村田らの手法 [1]	76.91% (3,878/5,042)	69.19% (3,878/5,605)	72.85
ベースライン	67.91% (3,424/5,042)	68.91% (3,424/4,969)	68.41

ラメータはバッチサイズを 16, エポック数を 5, 学習率を $1e-5$ に設定した. モデルへの入力トークン x_i と前後 $n = 8$ 個のトークンとした.

比較のためのベースラインとして BERT を用いて読点, 改行の順で挿入する手法を設定した.

評価は適合率, 再現率, および F 値により行った. 適合率, 再現率はそれぞれ,

$$\text{適合率} = \frac{\text{正しく挿入された読点数}}{\text{挿入された読点数}}$$

$$\text{再現率} = \frac{\text{正しく挿入された読点数}}{\text{正解データの読点数}}$$

を求めた. 改行挿入においても同様の式で求める.

3.2 実験結果

実験結果を表 1, 表 2 に示す. 提案手法は先行研究, およびベースラインよりも高い F 値を達成しており, 本手法の有効性を確認した.

読点挿入タスクにおける誤りを分析した結果, 正しい位置に挿入できなかった主な原因として, 並列する名詞を区切る読点が挙げられる.

- メソッド組織、ツール、それらを使って両方作ってしまおうとは考えないでしょう...

この例では, 名詞「メソッド」と名詞「組織」の間に読点が挿入されなかった. 並列された名詞を区切る読点を挿入するために, 品詞やポーズの情報を有効であると考えられる.

また, 改行挿入タスクにおける誤りを分析した結果, 正しい位置に挿入できなかったものとして, 主に行長を 20 文字以下にするための改行が挙げられる. 図 2 に例を示す. この例では一行 20 文字という行長の制限を満たしておらず, 3 文字超過している. これを解消するために, 改行ラベルの予測で, 行長制限を満たすものから選択して出力

正解:

インドの方は国際社会に訴えてですね、今回の攻撃はパキスタンの政府が支援したものだ

予測結果:

インドの方は国際社会に訴えてですね、今回の攻撃はパキスタンの政府が支援したものだ

図2 改行挿入の誤り

することが考えられる.

4 おわりに

本稿では, 講演テキストを対象に読点と改行を同時に挿入する方法を提案した. 本手法では, BERT を用いて読点挿入と改行挿入のマルチタスクを実行し, 挿入位置を決定する. 実験の結果, 本手法は読点と改行を順次的に挿入する手法よりも高い精度を達成し, その有効性を確認した.

謝辞 本研究は, 一部, 科学研究費補助金基盤研究 (C) No. 22K12122 により実施した.

参考文献

- [1] Murata, M., Ohno, T. and Matsubara, S.: Automatic Comma Insertion for Japanese Text Generation, In *Proceedings of EMNLP*, pp. 892–901 (2010).
- [2] Tilk, O. and Alumäe, T.: Bidirectional Recurrent Neural Network with Attention Mechanism for Punctuation Restoration, In *Proceedings of Interspeech*, pp. 3047–3051 (2016).
- [3] Ohno, T., Murata, M., and Matsubara, S.: Linefeed Insertion into Japanese Spoken Monologue for Captioning, In *Proceedings of ACL-IJCNLP*, pp. 531–539 (2009).
- [4] Murata, M., Ohno, T. and Matsubara, S.: Automatic Text Formatting for Social Media Based on Linefeed and Comma Insertion, In *Proceedings of IIMSS*, pp. 285–294 (2011).
- [5] Devlin, J., Chang, M., Lee, K. and Toutanova, K.: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, In *Proceedings of NAACL-HLT*, pp. 4171–4186 (2019).
- [6] Matsubara, S., Takagi, A., Kawaguchi, N. and Inagaki, Y.: Bilingual Spoken Monologue Corpus for Simultaneous Machine Interpretation Research, In *Proceedings of LREC*, pp. 153–159 (2002).