

残存文長を考慮したRNNによる逐次的な改行挿入 Sequential Linefeed Insertion based on RNN Considering Remaining Sentence Length

飯泉 智朗^{†,a)} 大野 誠寛^{†,b)} 松原 茂樹[‡]
Tomoaki Iizumi Tomohiro Ohno Shigeki Matsubara

1 はじめに

聴覚障害者や高齢者、外国人らによる講演音声の理解を支援するための技術として、字幕生成システムの開発が望まれている。字幕を生成する際、提示されたテキストが読みやすくなるように、適切な位置に改行が挿入されている必要がある。

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、残存文長を考慮したRNN[1]による逐次的な改行挿入手法を提案する。残存文長とは文の残りの長さを意味する。一般に、文がもう少しで終わる場所での改行の必要性は低下するなど、残存文長と改行位置には関連があると考えられる。評価実験の結果、従来手法[2,3]よりも、F値が大幅に向上しており、提案手法の有効性を確認した。

2 従来の逐次的改行挿入

従来手法[2]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機械学習(最大エントロピー法)を用いて逐次的に行っている。機械学習には、 b_i の主辞、語形などの形態素情報4種、節境界情報2種、係り受け情報3種、行頭からの文字数1種、ポーズ情報1種の、合計11の素性 $f_1 \sim f_{11}$ を使用している。また、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文字と設定し、 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、機械学習の判定結果に関わらず強制的に改行を挿入する。

従来手法[3]では、従来手法[2]の枠組みにおいて残存文長を考慮することにより、より適切な位置への改行挿入を実現している。残存文長とは文の残りの長さ(例えば n 個の文節から成る文において、文頭から i 番目の文節 b_i まで入力された時点での残存文長は $n-i$ となる)を意味し、河村らによって残存文長の推定手法が提案されている[4]。河村らは、文節が入力されるごとに、残存文長の確率分布をRNNにより推定し、その期待値を4クラス(0文節、1文節、2~3文節、4文節以上)に区分した結果を、その時点での残存文長としている。岩村らは、従来手法[2]の素性 $f_1 \sim f_{11}$ に、「河村ら[4]が推定した残存文長が上述の4クラスのいずれであるか」という素性 f_{12} を追加し、SVMを用いて、文節が入力されるたびに改行有無を判定している。

3 残存文長を考慮したRNNによる改行挿入

本研究では、従来研究[2,3]と同じ問題設定とし、文節 b_{i+1} が入力されるごとに、判定対象文節 b_i の直後に改行を挿入するか否かをRNNを用いて逐次的に判定する。図

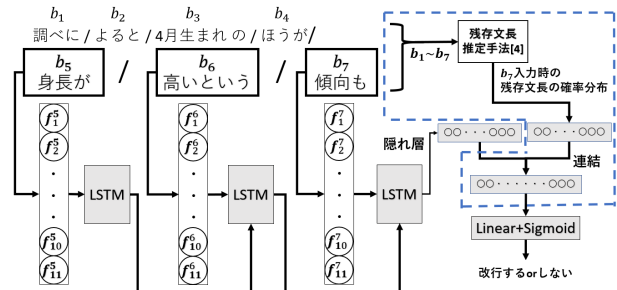


図1 提案手法の概要

1に提案手法の概要を示す。

図1では、「調べに/よると/4月生まれの/ほうが/身長が高いという/傾向も/あり」という8個の文節が入力された時点で、文節 b_7 「傾向も/あり」の直後に改行を挿入するか否かを提案手法が判定している。ただし、それまでの判定において改行が挿入されたのは、文節 b_4 「ほうが」の直後のみという状況における判定を示している。

提案手法では、行頭文節 b_k ($1 \leq k \leq i$) (図1では $k=5$) から判定対象文節 b_i (図1では $i=7$) までの各文節の素性 $f_1 \sim f_{11}$ (図1の f_j^i は文節 b_i の j 番目の素性を意味する) からなるベクトルを順にLSTM層[5]に入力する。また、残存文長を考慮するため、河村らの手法[4]が推定した b_i 入力時点の残存文長の確率分布と、LSTMの最終的な隠れ層とを連結し、Linear+Sigmoidを通すことにより、改行を挿入するか否かを判定する。なお、河村らの手法[4]は、文節 b_1 から判定対象文節 b_i までの形態素列をLSTMに入力し残存文長を推定している。

従来手法[2,3]では、判定対象文節 b_i の1文節から得られる素性のみを用いて機械学習により判定しているが、提案手法では、行頭文節から判定対象文節までの前方文脈を利用して判定できることになる。また、従来手法[3]では、残存文長を考慮する際に、残存文長の推定結果を4クラス分類した情報を素性 f_{12} として用いているが、提案手法では、残存文長推定手法が出力した確率分布をそのまま利用するため、残存文長の推定結果をより詳細に考慮して改行位置を判定できる。

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

実験データには、同時通訳データベース[6]の日本語講演音声書き起こしテキストを使用した。なお、全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、全16講演を用いた交差検定によって行った。すなわち、1講演をテストデータとし、残りの15講演を学習データとして改行位置を同定する実験を16回繰り返した。ただし、16講演のうち2講演については、開発

[†] 東京電機大学大学院未来科学研究科, Graduate School of Science and Technology for Future Life, Tokyo Denki University.

[‡] 名古屋大学情報連携推進本部, Information and Communications, Nagoya University.

a) 20fmi03@ms.dendai.ac.jp

b) ohno@mail.dendai.ac.jp

表1 実験結果

	再現率 (%)	適合率 (%)	F 値
SVM[$f_1 \sim f_{11}$]	76.27 (5,489/7,197)	70.00 (5,489/7,841)	73.00
SVM[$f_1 \sim f_{12}$]	76.56 (5,510/7,197)	71.28 (5,510/7,703)	73.83
RNN[$f_1 \sim f_{11}$]	91.40 (6,578/7,197)	71.26 (6,578/9,231)	80.08
提案手法	90.61 (6,521/7,197)	72.81 (6,521/8,956)	80.74

提案手法の出力(正解):

等距離に取り付けるなんて
何て変な考えだろうと思われるでしょう

RNN[$f_1 \sim f_{11}$]の出力:

等距離に取り付けるなんて
何て変な考えだろうと
思われるでしょう

図2 提案手法が正解, RNN[$f_1 \sim f_{11}$]が不正解の例

提案手法の出力:

いろいろまごついたり
失敗をおかしたり
しております

RNN[$f_1 \sim f_{11}$]の出力(正解):

いろいろまごついたり
失敗をおかしたり
しております

図3 RNN[$f_1 \sim f_{11}$]が正解, 提案手法が不正解の例

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. 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.

比較手法として以下の3つの手法を用意した.

・SVM[$f_1 \sim f_{11}$]: 判定対象文節 b_i から残存文長を含まない従来素性 $f_1^i \sim f_{11}^i$ を抽出し, それらを用いて SVM により判定する手法 (従来手法 [2] の最大エントロピー法を SVM に変更した手法).

・SVM[$f_1 \sim f_{12}$]: 判定対象文節 b_i から従来素性 $f_1^i \sim f_{11}^i$ と残存文長素性 f_{12}^i を抽出し, それらを用いて SVM により判定する手法 (従来手法 [3] と同一手法).

・RNN[$f_1 \sim f_{11}$]: 提案手法において, 残存文長情報を利用しない手法 (図1の青点線枠部分を除去した手法).

RNN は Pytorch を用いて実装した. 学習アルゴリズムは Adam を採用した. パラメータの更新はミニバッチ学習 (学習率 0.01, バッチサイズ 24) により行った. LSTM 層 (1層) の次元数は RNN を用いる両手法とも 25 とし, エポック数は RNN[$f_1 \sim f_{11}$] を 8, 提案手法を 10 とした. LSTM 層の次元数は 100, 75, 50, 25, 20, 10 の 6通りで変化させ, また, エポック数は 1~10 の間で 1 刻みで変化させて, 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

各手法の適合率, 再現率, F 値を表1にそれぞれ示す. 提案手法は F 値において最大値を達成しており, 提案手法の有効性が確認できる.

提案手法と SVM[$f_1 \sim f_{12}$] とを比較すると, F 値において約7ポイント上昇しており, RNN を導入した効果が大きいことがわかる. また, 提案手法と RNN[$f_1 \sim f_{11}$] とを比較すると, F 値が向上しており, 従来研究 [2] での報告と同様に, 残存文長情報の有効性を確認できた.

4.3 考察

本節では, 残存文長の推定結果を考慮することによる影響を考察する. 提案手法が正解し, RNN[$f_1 \sim f_{11}$] が不正解となった例を図2に示す. 提案手法の改行位置は正解データと完全に一致しているが, RNN[$f_1 \sim f_{11}$] では「考えだろうと」の後で余分に改行されている. 「考えだろうと」の文節が入力された際に河村らの手法 [4] によって推定された残存文長の確率分布の期待値は約1であり, 実際の残存文長と一致していた. その結果, 当該情報を

表2 残存文長が正解した場合と不正解の場合の実験結果

	再現率 (%)	適合率 (%)	F 値
正解	91.70 (4,433/4,834)	75.12 (4,433/5,901)	82.59
不正解	88.36 (2,088/2,363)	68.35 (2,088/3,055)	77.08
合計	90.61 (6,521/7,197)	72.81 (6,521/8,956)	80.74

利用している提案手法では, 文がもう少しで終わることを考慮でき, 余分な改行を抑えることに成功したものと考えられる.

次に, RNN[$f_1 \sim f_{11}$] が正解し, 提案手法が不正解となった例を図3に示す. 正解の改行位置と比べて, 提案手法の出力では, 「おかしたり」の直後に余分な改行が行われている. 「おかしたり」の文節が入力された際に河村らの手法によって推定された残存文長の確率分布の期待値は約13であったが, 実際の残存文長は1であり, 大きく外れていた. その結果, 提案手法は, まだ文が長く続くという情報を考慮することになり, この位置で過剰に改行することになったと考えられる.

また, 表2は, 残存文長の推定結果 (確率分布の期待値を4クラス分類したもの) が正解していたか否かによって場合分けし, 各場合における提案手法の再現率, 適合率, F 値を再評価した結果である. 残存文長を正しく推定できている場合は, そうでない場合と比べて, いずれの評価指標においても大幅に上回っていることがわかる.

なお, 残存文長を 100% 正しく推定できる手法を仮定し, 正解の残存文長を提案手法に入力した場合は, 再現率 90.80%, 適合率 73.85%, F 値 81.45 であった.

以上より, 逐次的な改行挿入において残存文長を考慮することは有効であることを確認した.

5 おわりに

本稿では, 講演テキストを対象に, 残存文長を考慮した RNN による逐次的な改行挿入手法を提案した. 実験の結果, 提案手法は, 従来手法を上回る再現率, 適合率, F 値を達成しており, その有効性を確認した. また, 逐次的な改行挿入において残存文長を考慮することの有効性を確認した.

今後は, RNN における残存文長情報のより効果的な導入方法等について検討し, 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] T. Mikolov et al., "Recurrent Neural Network Based Language Model," Proc. INTERSPEECH 2010, pp.1045–1048, 2010.
- [2] 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," 電気学会論文誌, Vol. 133-C, No. 2, pp.418-426, 2013.
- [3] 岩村ら, "残存文長を考慮した講演テキストへの改行挿入," 情報処理学会第82回全国大会講演論文集, No.2, pp.545-546, 2021.
- [4] 河村ら, "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, No.2, pp.447-448, 2020.
- [5] M. Sundermeyer et al., "LSTM Neural Networks for Language Modeling," Proc. INTERSPEECH 2012, pp. 194–197, 2012.
- [6] S. Matsubara et al., "Bilingual Spoken Monologue Corpus for Simultaneous Machine Interpretation Research," Proc. LREC 2002, pp.153–159, 2002.