

## 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

河村 天暉<sup>†,a)</sup> 大野 誠寛<sup>†,b)</sup> 松原 茂樹<sup>‡</sup>東京電機大学大学院未来科学研究科<sup>†</sup> 名古屋大学情報連携統括本部<sup>‡</sup>

## 1 はじめに

同時通訳や字幕生成などの音声言語処理システムでは、入力に対して漸進的に処理を行う必要があり、処理の正確さを保ちつつ、遅延時間を抑えることが求められる。このような処理を実現するにあたり、意味的なまとまりをもつ文が今後どれだけ続くかという情報は、重要な手がかりとなりうる。例えば、文がもう少しで終わることが分かれば、処理の正確さを保つため、同時通訳における訳出タイミング [1] や、読みやすい字幕とするための改行挿入タイミング [2] を遅らせるという判断が可能となる。

そこで本稿では、独話文を対象に、RNN (Recurrent Neural Network) [3] を用いて、文節が入力されるごとに残存文長を推定する手法を提案する。これまでに著者らは、新聞記事文を対象とし、文節が入力されるごとに残存文長が1文節、2~3文節、4文節以上の3クラスのいずれであるかを推定する手法を提案している [4]。本稿では、独話文を対象とするため、話し言葉では文境界が明示されない点を考慮する。具体的には、後段の音声言語処理システムに文境界の情報を提供することを念頭に、文境界を意図するクラス、すなわち残存文長が0文節というクラスを加えた計4クラスのいずれであるかを推定する。

## 2 残存文長

日本語文を構成する要素には、単語、文節、節などが存在するが、本研究では文の長さを測る単位として、文節を使用する。文  $s$  が  $n_s$  個の文節から成り、文頭から  $x$  番目の文節  $b_x$  まで既に入力されているとき (すなわち、既入力文節数が  $x$  であるとき) の残存文長  $RL(s, x)$  を  $RL(s, x) = n_s - x$  により定義する。

## 3 残存文長の推定

提案手法では、1文 ( $s = b_1 \dots b_{n_s}$ ) を構成する文節が入力されるごとに、文頭から現在入力された文節までの形態素系列 (ポーズ、フィラー、言い淀みを表す記号を含む) を RNN に入力し、そのときの残存文長 (文節数) を推定する。この推定を、文節  $b_1$  が入力されてから文節  $b_{n_s}$  が入力されるまで繰り返す。なお、話し言葉に現れるポーズ、フィラー、言い淀みについては、それらを示す記号 P, F, D でそれぞれ統一し、各入力単語と同様に扱うこととした<sup>\*1</sup>。

例えば、文  $s$  「私は P / この間 / 学校に / 行った。」<sup>\*2</sup> 中の文節  $b_2$  「この間」が入力されたときの残存文長  $RL(s, 2)$  を推定する際には、RNN への入力は、文頭から現在入力された文節までの系列「私は P この間」とし、RNN の出力は、そのときの残存文長  $RL(s, 2)$  の確率分布とする。この確率分布に基づいて、どの値を最終的な推

定結果とするかについては様々な方法が考えられるが、提案手法では、確率分布の期待値 (小数点第1位を四捨五入) を算出したうえで、その値が4クラス (0文節, 1文節, 2~3文節, 4文節以上) の中でどのクラスに属するかを求め、そのクラスを推定結果にすることとする。4クラスのうち、2~3文節のクラスは、節の平均長 (2.60文節) [6] を考慮して設けたものであり、文の残りが最終節だけであることの判別を意図したものである。

さらに、単語系列の入力方向については、文頭から順に入力していく順方向 (FW)、文末からの逆方向 (BW) の2種類を試すこととし、それぞれによる推定を行う。

なお、RNN の出力層の次元数は、最も長いと想定する文の長さ (文節数) とする。RNN の隠れ層は1層かつ LSTM (Long Short Term Memory) [7] により構成され、RNN への各入力単語、及び、ポーズ、フィラー、言い淀みの各記号は one-hot ベクトルで表現されているとする。

## 4 評価実験

提案手法の有効性を確認するために、日本語講演音声の書き起こし文に対する残存文長の推定実験を実施した。

## 4.1 実験概要

実験データには、同時通訳データベース<sup>\*3</sup>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。全てのデータに、形態素情報、係り受け情報、節境界情報が人手で付与されている。実験は全16講演、1,935文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15講演を学習データとして残存文長を推定する実験を16回繰り返した。ただし、16講演のうち2講演は開発データとして使用したため評価データから取り除き、残りの14講演 (1,714文, 20,707文節) に対する実験結果に基づいて評価した。

比較手法として、以下の2つの手法を用意した。

- **ChanceRate**: 学習データより、各入力文節位置における1文の残存文長を集計し、その割合に従ってランダムに4クラスを出力する。

- **SVM**: 学習データの文集合  $S$  における入力文節数ごとの平均残存文長 ( $ARL(S, x) = \frac{\sum_{s \in \{s | n_s > x, s \in S\}} RL(s, x)}{|\{s | n_s > x, s \in S\}|}$ ), 入力文節の語彙情報、既入力文節列の構文情報を素性とし (文献 [8] と同じ)、SVM (Support Vector Machine) により4クラスを出力する。

評価では、各クラスの適合率、再現率、F値を測定した。

RNN の学習は、Chainer V4.0<sup>\*4</sup> を介して行った。学習アルゴリズムには 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 0.001, バッチサイズ 128) により行い、更新時にユニットを 0.1 の確率でドロップアウトさせた。エポック数は 50 とした。入力層の出力ベクトルと隠れ層の次元数は共に 200 とした。この値は、200, 400, 600, 800, 1000 の5通りで学習を行い、開発データ

Estimation of Remaining Sentence Length of Spoken Monologue for Incremental Language Processing

Takaki Kawamura<sup>†,a)</sup>, Tomohito Ohno<sup>†,b)</sup>, Shigeki Matsubara<sup>‡</sup>

<sup>†</sup> Graduate School of Science and Technology for Future Life, Tokyo Denki University.

<sup>‡</sup> Information and Communications, Nagoya University.

a) 18fmi07@ms.dendai.ac.jp

b) ohno@mail.dendai.ac.jp

\*1 ポーズ、フィラー、言い淀みは事前に検出できるとする [5].

\*2 単語境界を半角スペース、文節境界をスラッシュで表す。

\*3 <http://sidb.jp/>

\*4 <https://chainer.org/>

表1 4クラス分類の実験結果 (適合率: P, 再現率: R, F値: F)

	残存文長: 0 文節			残存文長: 1 文節			残存文長: 2~3 文節			残存文長: 4 文節以上		
	P(%)	R(%)	F	P(%)	R(%)	F	P(%)	R(%)	F	P(%)	R(%)	F
Chance Rate	8.00 (137/1,676)	7.83 (137/1,714)	7.91	7.71 (131/1,661)	7.89 (131/1,700)	7.80	15.79 (513/3,249)	15.56 (513/3,296)	15.79	67.33 (9,508/14,121)	67.93 (9,508/13,997)	67.63
SVM	71.47 (1,418/1,984)	82.73 (1,418/1,714)	76.69	17.47 (1,082/6,192)	63.65 (1,082/1,700)	27.42	21.44 (947/4,416)	28.73 (947/3,296)	24.56	80.26 (6,513/8,115)	46.53 (6,513/13,997)	58.91
提案手法 (FW)	88.90 (1,201/1,351)	70.07 (1,201/1,714)	78.37	19.21 (323/1,681)	19.00 (323/1,700)	19.11	18.71 (501/2,677)	15.20 (501/3,296)	16.78	75.12 (11,267/14,998)	80.50 (11,267/13,997)	<b>77.72</b>
提案手法 (BW)	87.43 (1,516/1,734)	88.45 (1,516/1,714)	<b>87.94</b>	27.33 (540/1,976)	31.76 (540/1,700)	<b>29.38</b>	22.93 (960/4,186)	29.13 (960/3,296)	<b>25.66</b>	78.75 (10,089/12,811)	72.08 (10,089/13,997)	75.27

の推定において最も正解率<sup>\*5</sup>の高かったものを採用した。one-hot ベクトルのサイズの平均は 5610.5 であり, 未知語率のマクロ平均は 8.89%であった<sup>\*6</sup>。

#### 4.2 実験結果

表1に実験結果を示す。提案手法 (BW) と ChanceRate, SVM の F 値を比較すると, 全てのクラスにおいて, 提案手法 (BW) が上回った。また, 提案手法 (FW と BW) の間で F 値を比較すると, 4 文節以上のクラスにおいては FW が 2.45 上回っているものの, 他のクラスにおいては BW が 8.88 以上, 上回った。以上より, 残存文長の推定において, 提案手法 (BW) が有効であることを確認した。

#### 5 人間による残存文長推定との比較

4 節にて提案手法の有効性を確認したが, 残存文長の推定タスクの難しさに対して, 提案手法がどの程度の性能を示しているのかは検証できていない。そこで, 人間が残存文長を推定する能力を測定し, 残存文長の推定タスクの難しさの程度を確認するとともに, 人間による推定結果と提案手法による自動推定結果を比較評価した。

##### 5.1 人間の残存文長推定能力の測定

作業者は, 1 文が文頭から順に 1 文節ずつ提示されるという状況において, 新たな文節が提示されるごとに, これまでに提示された文節列をもとに残存文長を推定する (文献 [9] のインタフェースと同様のものを使用)。具体的には, 次の手順により推定作業を実施した。

- (1) 文 ID を選択する。(2) 既入力文節列として, 文頭文節のみが表示される。(3) 既入力文節列をもとに残存文長を推定し, 入力欄に入力する。(4) 既入力文節列に, 次の文節が足されたものが表示される。(5) (3) と (4) の推定・表示を文末まで繰り返す。

上述の作業を, 対象データ 100 文に対して大学生 8 名がそれぞれ実施した。対象データは, 4.1 節の評価用データ 14 講演よりランダムに抽出した。

評価では, 4 節と同様に比較するため, 人間による残存文長の推定結果を, それぞれ 0 文節, 1 文節, 2~3 文節, 4 文節以上の 4 クラスのいずれかに分類した上で, クラスごとの F 値を測定した。

##### 5.2 比較評価

表2に比較評価結果を示す。人間 (平均) の値は 8 名のマクロ平均値, 人間 (最高) の値は 8 名のうち, 正解率<sup>\*5</sup>が最も高い作業者の値であり, 提案手法 (BW) の値は, 4 節の実験で同手法が対象データ 100 文に対して推定した結果である。人間 (平均) と提案手法 (BW) を比較すると, 2~3 文節のクラスと 4 文節以上のクラスにおいて

<sup>\*5</sup> 全文節の各入力時における推定結果に対する正解率を指す。  
<sup>\*6</sup> 交差検定を行ったため, 評価に用いた 14 講演に対する全 14 回の実験におけるマクロ平均を求めた。

表2 人間による推定との F 値の比較 (0: 0 文節, 1: 1 文節, 2~3: 2~3 文節, 4~: 4 文節以上)

	0	1	2~3	4~
人間 (平均)	84.51	30.60	26.90	64.68
人間 (最高)	<b>88.46</b>	<b>38.84</b>	27.29	<b>77.87</b>
提案手法 (BW)	83.87	25.11	<b>31.46</b>	75.54

提案手法 (BW) が上回った。特に 4 文節以上のクラスにおいては, 10.86 の差がみられた。また, 0 文節のクラスにおいては, 人間 (平均) と提案手法 (BW) の差が 0.64 であることから, 提案手法 (BW) は人間 (平均) と同程度に文境界を判別できていると考えられる。一方, 人間 (最高) と提案手法 (BW) を比較すると, 2~3 文節のクラスを除いた全クラスにおいて人間 (最高) が上回った。このことから, 人間が捉えている特徴の中には, RNN が捉えきれていないものがあるといえる。

#### 6 おわりに

本論文では, 独話文を対象に, RNN を用いて, 1 文を構成する文節が入力されるごとに残存文長を推定する手法を提案した。比較実験の結果, 提案手法 (BW) が比較手法を全クラスの F 値において上回っており, 提案手法 (BW) の有効性が確認できた。今後は, Attention の導入などによる精度向上に取り組みたい。

謝辞 本研究は, 一部, 科学研究費補助金基盤研究 (B) No. 26280082 及び (C) No. 16K00300 により実施した。

##### 参考文献

- [1] 笠ら, “英日同時翻訳のための依存構造に基づく訳文生成手法,” 信学論, J92-D(6), pp. 921-933, 2009.
- [2] 村田ら, “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,” 信学論, J92-D(9), pp. 1621-1631, 2009.
- [3] T. Mikolov et al., “Recurrent Neural Network Based Language Model,” Proc. of INTERSPEECH 2010, pp.1045-1048, 2010.
- [4] 河村ら, “漸進的な言語処理のための RNN を用いた残存文長の推定,” 言語処理学会第 25 回年次大会発表論文集, pp. 355-358, 2019.
- [5] M. Asahara and Y. Matsumoto, “Filler and Disfluency Identification based on Morphological Analysis and Chunking,” Proc. of SSPR2003, pp.163-166, 2003.
- [6] 丸山ら, “日本語節境界検出プログラム CBAP の開発と評価,” 自然言語処理, 11(3), pp. 39-68, 2004.
- [7] M. Sundermeyer et al., “LSTM Neural Networks for Language Modeling,” Proc. of INTERSPEECH 2012, pp. 194-197, 2012.
- [8] 河村ら, “漸進的な言語処理のための残存文長の推定,” FIT2018 講演論文集, 17(2), pp. 183-184, 2018.
- [9] 河村ら, “漸進的な言語処理のための RNN を用いた残存文長の推定とその評価,” FIT2019 講演論文集, 18(2), pp. 185-186, 2019.