

## 進捗報告

## 1 進捗

- 連続  $n$  文を入力とし、正例ラベルを喜劇としたときの2クラス感情推定

## 2 実験設定

図1にネットワークの概略図を示す。  $s_i$  はそれぞれ連続する  $n$  文のセリフを BERT の id 列に変換したものである。また、組み合わせとしては同一の4コマに属し、かつ連続しているものを扱う。各4コマの序盤に現れるセリフには参照できる過去のセリフが無い場合、便宜上のセリフ [pad] を置くことで対処した。末尾のセリフ以外はオリジナルのセリフのみから抽出することで、データ数を合わせている。

よって、単語 id 列長を  $w$  とすると入力次元は  $(batch \times n \times w)$  となる。このままでは BERT の入力次元に対応していないので、まず  $(n \times batch \times w)$  へと軸を入れ替え、これを1次元目について各要素に分解し、これら  $n$  個の次元数  $(batch \times w)$  のベクトルをそれぞれ BERT への入力とし、BERT の出力から [CLS] トークンに相当するベクトルのみを抜き取り、先と逆の手順を踏むことで次元数  $(batch \times n \times 768)$  のテンソルを得ることができる。これを識別器 Bi LSTM, Self-Attention への入力とすることで末尾のセリフの感情推定を行う。

BERT は事前学習済みモデル<sup>1</sup>を用い、最終層のみをチューニングする。その他実験設定などは、前回までと同様1文のみを入力とする実験と同じである。識別器のパラメータは図1に示す。

今週は  $n = 3$  と  $n = 5$  の場合について実験を行った。

表 1: network parameters

parameter	value
lstm_in	300
lstm_hidden	128
self_attn_in	$128 \times 2$
attn_num_layers	3

## 3 実験結果

表2に実験結果を載せる。表2より B3 実験の時と同様で  $n$  を大きくすることで Accuracy が下がり、正例の Recall と F1 値が上がる傾向が見られると予想できる。(不均衡データでも過剰に正例だと判定するようになる。)

## 4 今後の研究の方向性?

- 4 コマストーリーデータセットで訓練・テストでは精度に期待できない
- 整備された他の日本語の極性評価データセットを用いて BERT の事前学習モデルを fine tuning し、4 コマ漫画の感情推定問題に転移させる?
- 仕様によっては感情ラベルの貼り直し、または人手によるラベル付けが必要になる
- manga109 のセリフのデータも用いれば半教師あり学習も可能か

## 5 今後の実験予定

- 前期発表資料作り
- 連続  $n$  文を入力とする実験 続き

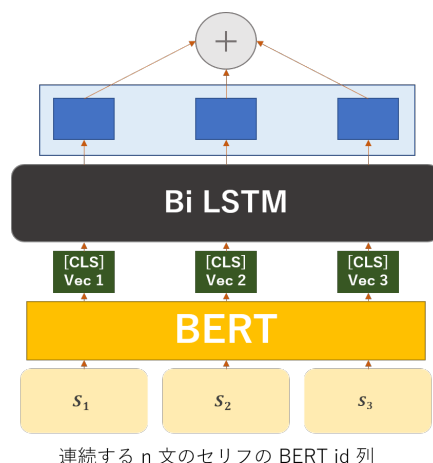


図 1: seq net

<sup>1</sup><http://nlp.ist.i.kyoto-u.ac.jp>

表 2: result

seq len	ギャグ			少女漫画			少年漫画			青年漫画			萌え系			5 タッチ平均		
	Acc	Recall	F1	Acc	Recall	F1	Acc	Recall	F1	Acc	Recall	F1	Acc	Recall	F1	Acc	Recall	F1
3	<b>0.773</b>	0.000	0.000	<b>0.627</b>	0.526	<b>0.615</b>	<b>0.766</b>	0.000	0.000	<b>0.800</b>	<b>0.286</b>	<b>0.381</b>	<b>0.625</b>	<b>0.409</b>	<b>0.429</b>	<b>0.718</b>	0.244	0.285
5	0.591	<b>0.600</b>	<b>0.308</b>	0.552	<b>0.553</b>	0.583	0.578	<b>0.333</b>	<b>0.229</b>	0.585	0.214	0.182	0.547	<b>0.409</b>	0.383	0.571	<b>0.422</b>	<b>0.337</b>
ベースライン	0.85	0	0	0.43	0	0	0.81	0	0	0.78	0	0	0.66	0	0	0.71	0	0

- コマの画像のベクトルも考慮した感情推定