発話の位置情報を考慮した対話行為認識

矢野 祐貴[1]　田村 晃裕[2]　二宮 崇[2]　大林 弘明[3]

[1] 愛媛大学 工学部情報工学科

[2] 愛媛大学 大学院理工学研究科 電子情報工学専攻

[3] トランスコスモス株式会社

{y_yano@ai., tamura@, ninomiya@}cs.ehime-u.ac.jp,
Oobayashi.Hiroaki@trans-cosmos.co.jp

1 はじめに

対話行為認識（Dialogue Act Recognition; DAR）は，対話文の各発話に対して，対話行為ラベルを識別するタスクである．対話行為ラベルは，対話文書の各発話がどのような意図を含むか，もしくはどのように機能しているかを判断する為のラベルであり，対話システムにおける話題や発話意図の推定，対話文書要約での重要文抽出などに用いられる．近年の DAR の研究では，再帰型ニューラルネットワークと条件付き確率場（Conditional Random Field; CRF）を組み合わせたモデルが高い識別性能を実現している [1, 2]．

本稿は，注意機構 [3] を用いて，階層型 BiLSTM-CRF モデル [1] に，対話文書における各発話の位置情報を組み込む新しい対話行為認識モデルを提案する．対話行為ラベルの設計は，コーパスによって異なっており，対話文書内でのラベルの種類や出現位置の分布もそれぞれ異なっている．特に，コンタクトセンター業務などで行われる問い合わせなど，対話に特有の流れがある場合には出現位置の分布に偏りがある傾向が強い．関係抽出の研究において，注意機構を介して系列の位置情報を組み込む手法 [4] が提案されているが，彼らの研究は非階層型の系列データを扱う関係抽出に対して行っており，DAR などの階層型の系列データを扱う手法とはなっていない．本研究では，階層型 BiLSTM-CRF モデルに注意機構を介して系列の位置情報を組み込む手法を適応する．

実験では，標準的な DAR のデータセットである SwDA コーパスに加え，トランスコスモス株式会社が行っているコンタクトセンター業務及びヘルプデスク業務で蓄積された問い合わせデータに，対話行為ラベルを付与した TCIDA コーパスを用いて性能評価を行う．評価実験の結果，位置情報を考慮することで，階層型 BiLSTM-CRF モデルと比べて正解率が SwDA では 0.08%，TCIDA では 0.36%向上することを示す．

2 階層型 BiLSTM-CRF

階層型 BiLSTM-CRF モデル [1] は，単語層と発話層の双方向 LSTM（BiLSTM）によって発話の表現を獲得し，その発話の表現から CRF で対話行為を識別するモデルである．このモデルでは，対話文書を $D = (u_1, u_2, \cdots, u_N)$，発話文を $\mathbf{u}_i = (w_{i1}, w_{i2}, \cdots, w_{iS_i})$，とする．$N$ は対話文書に含まれる発話の数，S_i は i 番目の発話の単語の数を表す．まず，i 番目の発話の j 番目の単語 w_{ij} を単語埋め込み層によって，埋め込みベクトル $\mathbf{e}_{ij}$ に変換する．その後，埋め込みベクトル $\mathbf{e}_{ij}$ を単語層 BiLSTM へ入力し，最後の隠れ状態を発話 $\mathbf{u}_i$ の表現ベクトル $\mathbf{v}_i$ とする．この一連の操作を次に示す：

$$\mathbf{e}_{ij} = f_{embed}(w_{ij}) \quad (1)$$

$$\mathbf{h}_{ij} = \mathrm{BiLSTM}^{word}(\mathbf{h}_{ij-1}, \mathbf{e}_{ij}) \quad (2)$$

$$\mathbf{v}_i = \mathbf{h}_{iS_i} \quad (3)$$

$\mathbf{v}_i$ は発話の前後を考慮しない i 番目の発話のみに依存した発話表現である．この発話表現 $\mathbf{v}_i$ を発話層 BiLSTM へ入力して，対話全体を考慮した発話表現 $\mathbf{z}_i$ へ更新する：

$$\mathbf{g}_i = \mathrm{BiLSTM}^{utt}(\mathbf{g}_{i-1}, \mathbf{v}_i) \quad (4)$$

$$\mathbf{z}_i = \mathbf{W}^g \mathbf{g}_i \quad (5)$$

$\mathbf{W}^g \in \mathcal{R}^{d \times |C|}$ は，重み行列であり，d は発話層 BiLSTM の次元数と $|C|$ は対話行為ラベルの種類数である．

階層型 BiLSTM によってエンコードされた発話表現 $\mathbf{z}_i$ を CRF に入力し，ラベル系列を得る．まず，発話表現 $\mathbf{Z} = (\mathbf{z}_1, \mathbf{z}_2, \cdots, \mathbf{z}_N)$ をスコア行列に変換した $P = \mathbf{Z}^T$ と，i 番目のラベルから j 番目のラベルに遷移するスコア行列 $A_{i,j}$ を用いて，ラベル系列 $\mathbf{y} = (y_1, y_2, \cdots, y_N)$ に対するスコア関数を次の様に定義する．

$$\mathrm{score}(\mathbf{Z}, \mathbf{y}) = \sum_{i=0}^{N} A_{y_i, y_{i+1}} + \sum_{i=1}^{N} P_{i, y_i} \tag{6}$$

上記のスコア関数を用いて，入力した発話表現 $\mathbf{z}$ に対して出力されるラベル系列は次式で計算される．

$$p(\mathbf{y}|\mathbf{Z}) = \frac{\exp(\mathrm{score}(\mathbf{Z}, \mathbf{y}))}{\sum_{\tilde{\mathbf{y}} \in \mathbf{Y_z}} \exp(\mathrm{score}(\mathbf{Z}, \tilde{\mathbf{y}}))} \tag{7}$$

ここで $\mathbf{Y_z}$ は，入力系列 $\mathbf{Z}$ に対するすべての可能なラベル系列である．

CRF は対数線形モデルの一種であるから，学習時には正解ラベル系列を用いて次式を最大化する．

$$\log(p(\mathbf{y}|\mathbf{Z})) = \mathrm{score}(\mathbf{Z}, \mathbf{y}) - \log(\sum_{\tilde{\mathbf{y}} \in \mathbf{Y_z}} \exp(\mathrm{score}(\mathbf{Z}, \tilde{\mathbf{y}}))) \tag{8}$$

識別を行う時は，次式の最大化問題を解くことで出力ラベル系列 $\mathbf{y}^*$ が得られる．また，計算の際にビタビアルゴリズム [5] を用いて，最適なラベル系列を得る．

$$\mathbf{y}^* = \arg\max_{\tilde{\mathbf{y}} \in \mathbf{Y_z}} \mathrm{score}(\mathbf{Z}, \tilde{\mathbf{y}}) \tag{9}$$

3 提案手法

DAR と異なる系列ラベリング問題である関係抽出の研究において，注意機構を介して系列の位置情報を組み込む手法が提案されている [4]．しかし，この研究は非階層型の系列データを扱う関係抽出に対してであって，階層型の系列データを扱う DAR での有効性は示されていない．

本稿では，注意機構 [3] を介して系列の位置情報を組み込む手法を，階層型 BiLSTM-CRF モデルに適応したモデルを提案する．提案手法を加えたモデルの全体像を図 1 に示す．

位置情報を考慮するために注意機構を介する場合，発話表現を獲得する式 (5) を次式に変更する．

$$\mathbf{z}_i = \mathbf{W}^c[\mathbf{g}_i; \mathbf{c}_i] \tag{10}$$

表 1: コーパスのラベルの種類数と構成

Dataset	$\|C\|$	Training	Dev.	Testing
SwDA	42	1003(173K)	112(22K)	19(4K)
TCIDA	38	880(129K)	147(22K)	147(22K)

$\mathbf{W}^c \in \mathcal{R}^{2d \times |C|}$ は，重み行列であり，$\mathbf{c}_i$ は位置情報を考慮した文脈ベクトルである．文脈ベクトル $\mathbf{c}_i$ は，次の手順で算出する：

$$s_{ij} = \mathbf{v}_{attn}^T \tanh(\mathbf{W}^H \mathbf{g}_N + \mathbf{W}^p \mathbf{h}_i^p + \mathbf{W}^h \mathbf{g}_i) \tag{11}$$

$$a_{ij} = \frac{\exp(s_{ij})}{\sum_{k=1}^{N} \exp(s_{ik})} \tag{12}$$

$$\mathbf{c}_i = \sum_{j=1}^{N} a_{ij} \mathbf{g}_j \tag{13}$$

$\mathbf{W}^H, \mathbf{W}^p, \mathbf{W}^h, \mathbf{v}_{attn}$ は学習可能なパラメータである．アライメントスコア a_{ij} の計算時に，発話表現 $\mathbf{g}_i$ と対話全体の表現 $\mathbf{g}_N$ および位置表現 $\mathbf{h}_i^p$ を使用している．$\mathbf{h}_i^p$ は，次式によって得られる発話の絶対的な位置を埋め込んだ発話位置のベクトル表現である．

$$\mathbf{h}_i^p = f_{embed}^p(i) \tag{14}$$

4 実験

4.1 コーパス

モデルの性能評価には，DAR のタスクで標準的に使用される SwDA コーパスに加え，トランスコスモス株式会社（TCI）が独自に設計，蓄積した TCIDA コーパスを使用する．それぞれの対話行為ラベルの種類数と，学習，開発およびテスト時に使用した対話の件数を表 1 に示す．表 1 中の $|C|$ は，コーパスの対話行為ラベルの種類数，括弧内の数字は発話の総数である．

- **SwDA**: Switchboard Dialogue Act コーパス [6] は，電話音声での会話に対話行為ラベルを付与したコーパスである．DAR タスクで広く利用され [1, 2]，タスクの標準コーパスとして，様々なアルゴリズムを比較するベンチマークデータとして用いられる．対話行為ラベルの例として，「Statement-non-opinion（意見のない陳述）」，「Backchannel（相槌）」，「Yes-No-Question（Yes/No を問う質問）」などがある．

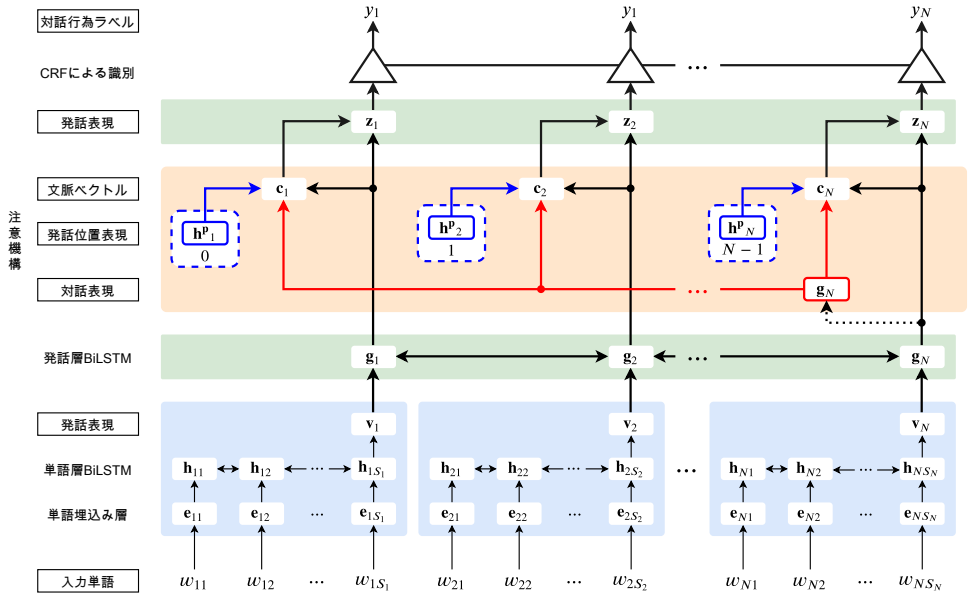

図 1: 注意機構を介して位置情報を考慮するモデル

表 2: TCIDA のオペレータ（OP）と顧客（CU）の対話サンプル

id	発話	対話行為
1	OP: お電話ありがとうございます．####ご相談センターでございます．	機能なし
2	CU: えーと修理のことで教えて欲しいんですけど	用件
3	OP: かしこまりました．どのようなご相談でございますか	機能なし
⋮	⋮	⋮
50	OP: それでは，修理料金についてご説明させていただきます．	料金案内_開始
51	CU: はい	料金案内_IN
⋮	⋮	⋮
60	OP: 詳しい料金は，訪問したときにお見積りを提示しますので，そこで確認していただきます．	料金案内_IN
61	CU: はい．	料金案内_終了
62	OP: はい．では本日の件，#日に訪問させていただきますのでどうぞよろしくおねがいします．	機能なし
63	CU: はい．失礼します．	結論
64	OP: ありがとうございました．	機能なし

- **TCIDA**: トランスコスモス株式会社では，一部のコンタクトセンター業務及びヘルプデスク業務において，応対品質の向上や VoC 分析のために，音声認識器を用いて問い合わせの会話音声をテキストデータに変換し記録している．TCI Dialogue Act コーパスは，この問い合わせ内の結論や顧客の要件にあたる発話などに，対話行為ラベルを付与したコーパスである．

  TCIDA の対話文の例を表 2 に示す．このコーパスの対話行為ラベル設計の特徴として，問い合わせの記録として機能しない相槌などの発話に対しては，「機能なし」の対話行為ラベルを付与しており，全体の約 63%を占めている．また，複数の発話にわたって，ある機能をもつ場合，_開始，_IN，_終了をラベルの末尾に付与して，複数の発話をチャンキングしている．

4.2 実験設定

SwDA と TCIDA の 2 つのコーパスを用いて，表 1 に示した学習データ，開発データ，テストデータの件数で提案手法の有効性を検証する．英語文である SwDA では NLTK tokenizer [1]を用いて単語分割し，日本語文の TCIDA では，KyTea [2] で単語分割した後，subword-nmt [3]を用いて，BPE に基づくサブワー

[1] https://www.nltk.org/
[2] http://www.phontron.com/kytea/index-ja.html
[3] https://github.com/rsennrich/subword-nmt

表 3: コーパスごとの正解率 (%) の比較

モデル	SwDA	TCI	TCI*
BiLSTM-CRF	76.73	82.25	51.76
提案手法	**76.81**	**82.61**	**52.09**

ド化を行う．なお，BPE の結合回数は 4000 回とする．いずれのモデルにおいても，バッチサイズを 16，最大エポック数は 20 とし，単語埋め込み，発話位置埋め込み，単語層 BiLSTM，発話層 BiLSTM の次元数はすべて 300 とする．

評価指標である正解率は，開発データに対して正解率が最も高い epoch のモデルパラメータを用いて，テストデータによる実験の結果で比較する．TCIDA の場合は「機能なし」ラベルを持つ発話が全体の約 63% を占めているため,「機能なし」ラベルの発話を除いた場合の正解率の評価も併せて行う．

4.3 実験結果

表 3 に従来の BiLSTM-CRF モデルと提案手法で 2 つのコーパスの対話行為ラベルを識別した結果を示す．表 3 中の TCI*の列は「機能なし」ラベルの発話を除いた場合の正解率である．提案手法によって，SwDA コーパスでは 0.08%，TCIDA コーパスでは 0.36%の正解率向上が確認できた．また，TCIDA は「機能なし」ラベルの発話を除いた場合の正解率においても，0.33%の向上が見られた．

4.4 考察

提案手法による識別性能の向上が，SwDA よりも TCIDA のほうが大きかった要因として，コーパス間で，対話行為ラベルが出現する発話の位置の傾向が異なっていることが挙げられる．

表 4 は，対話行為ラベルが出現した発話の位置を対話全体の長さで 0 から 1 の範囲に正規化し，各ラベルごとにその位置の分散をまとめた結果である．出現位置の分散が大きいラベルは，対話全体で偏りなく出現するラベルであり，反対に分散が小さいラベルは，対話中のある位置で多く出現するラベルである．SwDA と TCIDA を比べると，TCIDA の方が分散の小さいラベルが多く存在することがわかる．このことから，TCIDA は，発話の位置に依存するラベルが多く，本稿の提案手法である位置情報を考慮する仕組みが効果的に作用する設計であったことから，SwDA よりも大きな識別性能の向上が見られたと考える．

表 4: 発話位置の分散に対する対話行為ラベル数

範囲	SwDA	TCIDA
0.00 ～ 0.02	2	19
0.02 ～ 0.04	0	12
0.04 ～ 0.06	0	5
0.06 ～ 0.08	13	0
0.08 ～ 0.10	25	2
0.10 ～	2	0

5 おわりに

本稿では，発話の位置情報を注意機構によって従来の階層型 BiLSTM-CRF モデルに組み込む事で，対話行為認識の正解率を向上する手法を提案した．特に，対話に一連の流れを持ち，対話行為ラベルの出現位置にある程度偏りのあるコーパスにおいて有効性を示した．

今後は，考慮する位置情報を絶対位置から相対位置を利用することで，更なる識別性能の向上を目指す．

謝辞

本稿の研究は，トランスコスモス株式会社の助成をうけたものである．ここに謝意を表する．

参考文献

[1] Harshit Kumar, Arvind Agarwal, Riddhiman Dasgupta, and Sachindra Joshi. Dialogue act sequence labeling using hierarchical encoder with crf. In *Proc. of AAAI 2018*, pp. 3440–3447, 2018.

[2] Michael Denkowski and Alon Lavie. Dialogue act recognition via crf-attentive structured network. In *Proc. of SIGIR 2018*, pp. 225–234, 2018.

[3] Dzmitry Bahdanau, Kyung Hyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. In *Proc. of ICLR 2015*, pp. 1–15, 2015.

[4] Dai Dai, Xinyan Xiao, Yajuan Lyu, Shan Dou, Qiaoqiao She, and Haifeng Wang. Joint extraction of entities and overlapping relations using position-attentive sequence labeling. In *Proc. of AAAI 2019*, pp. 6300–6308, 2019.

[5] A. Viterbi. Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. In *IEEE Transactions on Information Theory*, pp. 260–269, 1967.

[6] Dan Jurafsky, Liz Shriberg, and Debra Biasca. Switchboard swbd-damsl shallow-discourse-function annotation coders manual. In *Proc. of WMT 2014*, pp. 376–380, 2014.