The 33rd Annual Conference of the Japanese Society for Artificial Intelligence, 2019

対話システムにおける履歴を考慮した応答の対話行為推定

Response Dialogue-Act Prediction based on Conversational History

田中 昂志　高山 隼矢　荒瀬 由紀
Koji Tanaka　Junya Takayama　Yuki Arase

大阪大学大学院情報科学研究科
Graduate School of Information Science and Technology, Osaka University

Sequence-to-sequence models are widely used to implement a chatbot. One of their advantages is that a chatbot can be trained in an end-to-end manner. On the other hand, its disadvantage is that a process of response generation is completely black-box. To solve this problem, interpretable response generation mechanism is desired. As a step forward in this direction, we focus on dialogue-acts and propose a method to predict a dialogue-act of the next response based on conversational history. Specifically, we consider both histories of utterances and their dialogue acts. Experiment results using the Switch Board Dialogue Act corpus show that our method achieves 8.6%and 1.2% higher F-score and accuracy on predicting responses' dialogue-acts, respectively, compared to a previous method that only considers the utterance history.

1. はじめに

深層学習による対話システムが盛んに研究されている [Vinyals 15]．深層学習による対話システムでは，大規模な対話データを学習することで，人手による応答ルールやパターンの設計を行わずに応答を生成できるという利点がある．一方で，応答生成のメカニズムはブラックボックスであり，ある入力発話に対する応答が生成された根拠を知ることは困難である．この問題を解決するため，Zhao ら [Zhao 18] は解釈可能な応答生成モデルの重要性を主張している．本研究では，応答生成の根拠となる有効な手がかりとして対話行為に着目し，対話行為推定に取り組む．また，応答生成においては対話行為情報が有効であることが示されており [Cervone 18]，応答の対話行為を推定することは応答生成の性能向上への寄与も期待できる．

対話は一般的に発話と応答の系列であり，ある時点で生成すべき応答の対話行為の推定には過去の発話の履歴を考慮することが有効と考えられる．既存研究 [大原 18] では発話・応答の系列を用いて応答の対話行為を推定しているが，過去の対話行為の系列は考慮していない．対話行為の系列を独立に考慮することで，過去の対話行為の系列と次の応答の対話行為との関係を直接的に学習できると期待できる．例えば，「あいづち」の発話に対しては一意に応答の対話行為の推定は困難であるが，「説明」の発話に続いて「あいづち」の発話が入力された場合には，「説明」の対話行為を持つ応答を続ける，「理解」の発話に続いて「あいづち」の発話が入力された場合には，「質問」の対話行為を持つ応答を返すなど，対話行為の系列を用いることで高精度な応答の対話行為推定が可能となると期待できる．

そこで，本研究では応答の対話行為を対話の文脈と対話行為の系列を用いて推定する手法を提案する．提案手法では，対話の文脈を捉える Recurrent Neural Network（RNN）と対話行為の系列を捉える RNN をそれぞれ構築し，それぞれの出力を用いて次の応答の対話行為を推定する．

電話での会話を書き起こし，発話毎に対話行為がアノテートされた Switch Board Dialogue Act（SwDA）コーパス *1 を用い，Precision，Recall，F 値のマクロ平均と全体の Accuracy を指標として評価実験を行った．その結果，大原らの手法と比較して F 値のマクロ平均においては 8.6%，全体の Accuracy においては 1.2% 精度が向上することが示された．

連絡先: 田中　昂志，大阪大学大学院情報科学研究科，
tanaka.koji@ist.osaka-u.ac.jp

*1 https://catalog.ldc.upenn.edu/LDC97S62

2. 関連研究

対話行為推定の研究には，発話テキストからその発話の対話行為を推定する発話の対話行為推定と，発話の系列から次の応答の対話行為を推定する応答の対話行為推定が存在する．

発話の対話行為推定の研究として，Kalchbrenner ら [Kalchbrenner 13] は発話の局所的な特徴を捉えた表現を得る Convolutional Neural Network（CNN）と発話の文脈表現を得る RNN を用いた手法を提案している．SwDA コーパス を用いて実験した結果，既存の機械学習を用いた対話行為推定の手法を上回る精度を達成している．また Khanpour ら [Khanpour 16] は発話を入力とする多段の RNN を提案している．SwDA コーパスを用いて実験した結果，80.1% の精度で発話の対話行為の推定を可能にし，既存の深層学習を用いた手法を上回る精度を達成している．

応答の対話行為推定の研究として，大原ら [大原 18] は発話の表現を得る RNN と対話の文脈を得る RNN を組み合わせ，応答の対話行為を推定する．評価実験の結果，応答の対話行為推定には対話の文脈情報が有効であることを示している．また，発話を入力としその対話行為を推定する発話の対話行為推定と比べ，次の応答の対話行為を推定する応答の対話行為推定の方が困難であることが示されている．

3. 提案手法

図 1 に提案手法の概要を示す．提案手法は発話と対話の文脈情報を保持する対話テキスト Encoder，対話行為の文脈情報を保持する対話行為 Encoder，対話行為を推定する分類器から成る．

3.1 対話テキスト Encoder

対話テキスト Encoder は，発話をベクトル化する RNN（発話 Encoder）と文脈をベクトル化する RNN（文脈 Encoder）から構成される．入力発話を単語分割したものを発話 Encoder

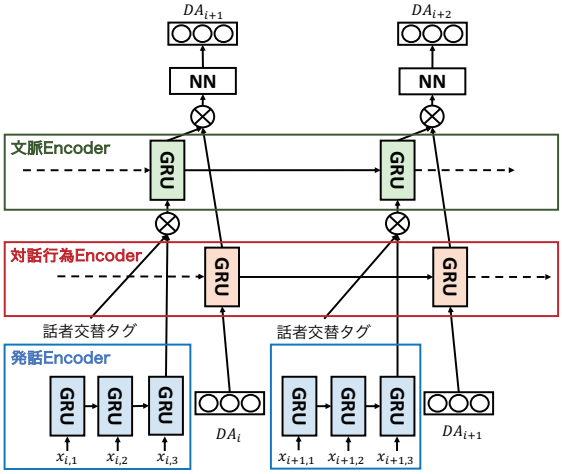

図 1: 提案手法の全体図（⊗ はベクトルの結合操作を表す）

に逐次的に入力し，発話のベクトル表現を得る．ここで，バッチ処理をするために発話の系列長を揃えるためにパディングを行うが，パディングの情報を発話ベクトルに含めると発話の情報が欠如すると考えられる．よって，発話ベクトルはパディングを行う前の RNN の状態を用いる．そして，得られた発話ベクトルを文脈 Encoder に入力し，過去の発話ベクトルの系列を考慮した文脈ベクトルを得る．

会話は常に 1 発話毎に話者交替をするとは限らず，同じ発話者が連続で発話する場合も存在する．提案手法では，大原ら [大原 18] の手法と同様に，発話ベクトルに話者交替の有無を表す埋め込みベクトルを連結し，文脈 Encoder の入力とする．

3.2 対話行為 Encoder

対話行為 Encoder は，過去の対話行為の系列のベクトル表現を RNN によって計算する．対話行為の系列を逐次的に入力することで，過去の対話行為の履歴を表現するベクトルを得る．入力の対話行為はその時刻での発話に付与されている対話行為である．

3.3 応答の対話行為推定

提案手法では，対話行為推定を多クラス分類問題として定式化する．対話テキスト Encoder と対話行為 Encoder から得られたベクトル表現を連結し，フィードフォワードニューラルネットワークを用いて対話行為の推定を行う．

4. 評価実験

4.1 実験データ

実験データは電話での会話を書き起こし，対話行為タグを付与した SwDA コーパスを用いる．SwDA コーパスに付与されている対話行為は damsl タグ [*2] に準拠しているが，付与されたタグの数が少ないものが存在し，十分に学習ができないと考えられる．そこで，簡易 damsl[磯村 09] を参考に 9 つのタグに削減したものを用いる．SwDA コーパスに含まれる対話数は 1,155，発話数は 219,297 であり，1 対話に含まれる発話数は平均 189 である．1 つの発話に含まれる発話・応答の系列が非常に長いため，本実験では 1 対話に含まれる発話・応答系列長を 5 とし，サイズ 5 のウインドウをスライドさせることで対話セットを抽出する．その結果，実験に用いるデータの対話数は 212,367，発話数は 1,061,835 となった．タグの種類とデータ中のタグの分布を表 1 に示す．訓練用，開発用，評価用にデータセットを 80%，10%，10% に対話単位でランダムに分割して使用する．

∗2 https://web.stanford.edu/~jurafsky/ws97/manual.august1.html

表 1: 実験データ内のタグの分布

タグ	タグの役割	タグ数
Statement	説明	576,005
Uninterpretable	フィラー	93,238
Understanding	理解	241,008
Agreement	同意	55,375
Directive	命令	3,685
Greeting	挨拶	6,618
Question	質問	54,498
Apology	謝罪	11,446
Other	引用，曖昧な発話	19,882

表 2: 実験結果

	Precision	Recall	F 値	Accuracy
大原ら	30.9	25.1	23.8	68.5
提案手法	**52.7**	**32.5**	**32.4**	**69.7**
DAseq only	44.7	28.7	27.9	67.1
DAseq + Utterance	45.8	29.0	29.3	68.2
DA + Utterance seq	30.1	19.5	18.9	65.1
Utterance	24.4	21.6	21.6	66.7

4.2 実験設定

本実験では，RNN として Gated Recurrent Unit（GRU）[Cho 14] を用い，単語 Embedding の次元数は 300，発話 Encoder の GRU の次元数は 512，文脈 Encoder の GRU の次元数は 513，対話行為 Embedding の次元数は 100，対話行為 Encoder の GRU の次元数は 128 とする．また，分類器の入力層の次元数は 641，中間層の次元数は 100 とする．ロス関数に交差エントロピー誤差，最適化には Adam[Kingma 14] を使用し，学習率は 0.00005 とする．学習エポック数は 30 とし，開発用データのロスが最も低い値を示した時点の重みを用いて評価する．また，評価時にも対話行為 Encoder への入力は入力発話に付与されている正解の対話行為を用いるものとする．

表 1 から分かるとおり，各対話行為の出現数は分散が大きい．そこで評価指標として，全体の Accuracy に加え，対話行為推定の Precision，Recall，F 値のマクロ平均を用いる．

4.3 比較手法

本実験では先行研究である大原らの手法と提案手法とを比較することで，対話行為系列を考慮する有効性を検証する．また，提案手法における各コンポーネントの効果を検証するため，対話行為の文脈のみから推定を行うモデル（DAseq only），および対話行為の文脈と直前の発話情報のみを用いて推定を行うモデル（DAseq + Utterance），直前の対話行為と発話系列から推定を行うモデル（DA + Utterance seq），直前の発話のみから推定を行うモデル（Utterance）を用いて精度の比較を行う．

4.4 実験結果

各モデルの評価結果を表 2 に示す．表 2 より全体の Accuracy において提案手法が最も高い値である 69.7% を示した．大原らの手法と比較して，提案手法が 1.2% 高い Accuracy を示した．このことより，対話行為系列を考慮することが応答の対

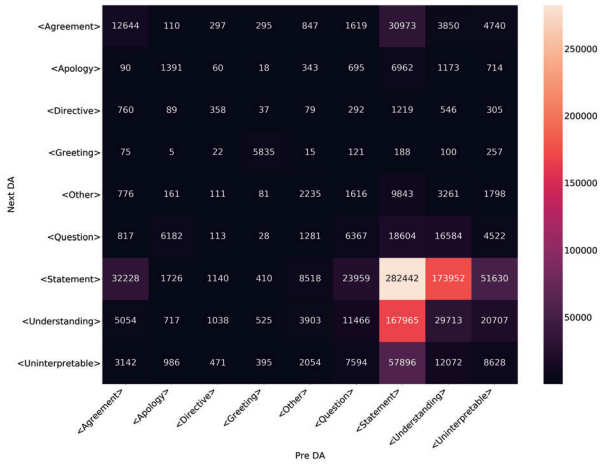

図 2: 混同行列：直前の対話行為と応答の対話行為の関係

話行為推定において有効であることが分かる．また，直前の対話行為と発話系列を用いる DA + Utterance seq と比較して，提案手法が 4.6% 高い Accuracy を示し，大原らの手法が 3.4% 高い Accuracy を示した．このことより，直前の対話行為のみを考慮する場合，推定に悪影響をもたらすことが分かる．これは，直前の対話行為と応答の対話行為の相関が薄いことが原因であると考えられる．

図 2 に直前の発話と応答の対話行為の混同行列を示す．図 2 より，対話行為として「Statement」が付与されている発話に対する応答の対話行為は「Uninterpretable」や「Statement」，「Understanding」などが続くケースが顕著に多いが，その他の対話行為の遷移については分散が大きく，対話行為遷移の目立ったパターンは観測できない．このことから，対話行為を用いる場合，系列を考慮することで推定に有益な情報を保持することが重要であることが分かる．また，対話行為の系列と直前の発話情報のみを用いる DAseq + Utterance と比較して，提案手法が 1.5% 高い Accuracy を示した．このことより，対話の系列も応答の対話行為推定において有効であることが分かる．

F 値のマクロ平均においても提案手法が最も高い値である 32.4% を示した．対話行為系列を用いているモデル全て（提案手法，DAseq only，DAseq + Utterance）において，大原らの手法と比較して高い F 値のマクロ平均を示していることから，応答の対話行為推定における対話行為の系列を考慮する有効性が明らかとなった．全体の Accuracy においては DAseq only と DAseq + Utterance は大原らの手法と比較して低いが，大原らの手法がある特定のタグに対しては高い精度で推定が可能である一方，他のタグに対しては推定精度が低いことが原因であると考えられる．

表 3 に提案手法，大原らの手法におけるタグ別の F 値を示す．表 3 より，全ての対話行為タグにおいて提案手法が大原らの手法より高い F 値を示した．特に，低頻度なタグである「Agreement」や「Greeting」,「Question」,「Apology」においては，6.1% から 34.6% の大きな改善となった．さらに，大原らの手法では「Directive」や「Other」タグについては正しく推定できていないが，提案手法では F 値は小さいが正しく推定できるケースが存在することが分かる．これは，対話行為の系列を陽に与えることによりで予測のための根拠が増えたために，出現数が少ないタグも推定できるようになったためと考えられる．

表 3: タグ別の F 値

タグ	出現数	提案手法	大原らの手法
Statement	576,005	80.8	80.4
Uninterpretable	93,238	4.7	2.6
Understanding	241,008	69.5	67.6
Agreement	55,375	23.1	15.3
Directive	3,685	2.7	0.0
Greeting	6,618	81.3	46.7
Question	54,498	8.1	2.0
Apology	11,446	22.7	11.3
Other	19,882	3.6	0.0

表 4 に提案手法と大原らの手法を用いて対話行為推定を行った例を示す．表 4 の 1 つ目の例より，表 1 においてタグ数が比較的少ない「Agreement」タグについて大原らの手法では誤判定しているが，提案手法では正しく推定できていることが分かる．表 4 の 3 つ目は対話文脈，対話行為系列双方を考慮しても推定が困難な例である．2 つ目と 3 つ目の例では，対話行為の系列は等しく，2 つ目の例では提案手法・大原らの手法ともに正しく応答の対話行為を推定できている．しかし，3 つ目の例ではどちらの手法も推定に失敗している．3 つ目の対話における 5 つ目の発話「and it's kind of dangerous.」に対する応答の対話行為は，2 つ目の発話における「aerosol」の危険性について聞き手が理解しているか，すなわち話者の知識に依存する.. この問題を解決するためには，ユーザの情報を保持する機構を用いてパーソナライズを行う必要があると考えられる．

5. まとめ

本研究では，対話文脈と対話行為の系列を考慮した応答の対話行為推定手法を提案した．

SwDA コーパスを用いて評価実験を行った結果，全体の Accuracy において 69.7%，F 値のマクロ平均において 32.4% を達成し，既存手法の性能をそれぞれ 1.2% および 8.4% 改善した．

今後の課題として，応答の対話行為推定結果を用いた応答生成を行う予定である．

参考文献

[Cervone 18] Cervone, A., Stepanov, E., and Riccardi, G.: Coherence Models for Dialogue, in *Proceedings Interspeech 2018*, pp. 1011–1015 (2018)

[Cho 14] Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., and Bengio, Y.: Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation, in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1724–1734 (2014)

[Kalchbrenner 13] Kalchbrenner, N. and Blunsom, P.: Recurrent Convolutional Neural Networks for Discourse Compositionality, in *Proceedings of the Workshop on Continuous Vector Space Models and their Compositionality*, pp. 119–126 (2013)

表 4: 推定結果例（「対話行為」はその発話の対話行為を示し,「正解タグ」は応答（次の発話）の対話行為である.）

	発話テキスト（対話行為）	正解タグ	提案手法	大原らの手法
1	what are they , (Uninterpretable)	Statement	Statement	Statement
2	the , (Statement)	Statement	Statement	Statement
3	I know , (Statement)	Statement	Statement	Statement
4	a Rabbit 's one , diesel (Statement)	Agreement	Understanding	Understanding
5	Uh-huh , (Agreement)	Agreement	Agreement	Statement
1	I hope so too .(Statement)	Statement	Statement	Statement
2	You know . Right now there 's a lot on the market for sale because of people having lost Yes .(Statement)	Understanding	Understanding	Understanding
3	Yes .(Understanding)	Statement	Statement	Statement
4	and everything(Statement)	Statement	Statement	Statement
5	so that 's , you know , that keeps prices down (Statement)	Understanding	Understanding	Understanding
1	It does n't seem like ,(Statement)	Statement	Statement	Statement
2	but I guess when you think of it everybody has some sort of aerosol in their home (Statement)	Understanding	Understanding	Understanding
3	Yeah .(Understanding)	Statement	Statement	Statement
4	you know ,(Statement)	Statement	Statement	Statement
5	and it 's kind of dangerous .(Statement)	Agreement	Understanding	Understanding

[Khanpour 16] Khanpour, H., Guntakandla, N., and Nielsen, R.: Dialogue Act Classification in Domain-Independent Conversations Using a Deep Recurrent Neural Network, in *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pp. 2012–2021 (2016)

[Kingma 14] Kingma, D. and Ba, J.: Adam: A Method for Stochastic Optimization, *International Conference on Learning Representations* (2014)

[Vinyals 15] Vinyals, O. and Le, Q. V.: A Neural Conversational Model, in *Proceedings of The 32nd International Conference on Machine Learning (ICML)* (2015)

[Zhao 18] Zhao, T., Lee, K., and Eskenazi, M.: Unsupervised Discrete Sentence Representation Learning for Interpretable Neural Dialog Generation, in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 1098–1107 (2018)

[磯村 09] 磯村 直樹, 鳥海 不二夫, 石井 健一郎: HMM による非タスク指向型対話システムの評価, 電子情報通信学会論文誌. D, 情報・システム = The IEICE transactions on information and systems (Japanese edition), Vol. 92, No. 4, pp. 542–551 (2009)

[大原 18] 大原 康平, 佐藤 翔悦, 吉永 直樹, 豊田 正史, 喜連川 優: 階層型 RNN を用いた対話における応答の対話行為予測, 言語処理学会第 24 回年次大会 (2018)