E-038

自己注意機構を用いた応答の対話行為推定
Estimating the dialogue act of the response using the self attention mechanism

宮城 孝明[1)]　遠藤聡志[2)]
Takaaki Miyagi　Satoshi Endo

1　はじめに

対話システムが適切な応答するためには, 最初にユーザの発話の意図を理解する必要がある [1]. 適切な応答を返すことの出来る対話システムを構築するために, 発話文に対する対話行為推定が行われている [2, 3, 4]. これらの研究では, 発話文の対話行為推定を行っているが, 応答文に対する対話行為推定はあまり行われていない.

大原らの研究 [5] では, 発話文から応答文の対話行為を推定する事で, 自然な対話が可能になると考えている. また, 応答文の対話行為推定の性能を上げることにより, 応答選択の精度向上が期待できる. 田中らの研究 [6] では, 大原らのモデルと組み合わせて対話行為 Encoder を構築し, 対話行為の前後関係を学習した.

本研究では, 発話文中の重要な単語に注目する事, 直近の発話を重点的に考慮する事を行う事で, 対話行為推定の精度向上を目指す. 具体的には, 田中ら [6] のモデルに対して, 単語系列を処理する箇所と, 発話系列を処理する箇所に自己注意機構 [7] を使用した拡張モデルを提案する. ここで, 自己注意機構とは 1 つの系列データの特徴を抽出する手法である. 拡張モデルを実装し, 応答の対話行為推定の精度評価を行った.

2　基礎概念

2.1　自己注意機構

注意機構 [8] とは, 入力データ中の「どの部分に注目するか」を表現する手法である. 注意機構は, 大きく 2 つの種類があり, 本研究ではその 1 つである自己注意機構を用いる.

図 1 に自己注意機構の概要を示す. 自己注意機構では入力トークンから Query, Key, Value が生成されている. Query は, 検索情報であり, Key と Value は一対一の辞書型オブジェクトである. 初めに,Query と Key の Matmul (関連度) を求め, Softmax を用いて Key 内のどこの情報に注目するかを正規化で表現する. このベクトルを Attention Weight と言う. 最後に, Value と Attention Weight の Matmul を行い, 注意すべきトークンを考慮したベクトルを出力する.

自己注意機構を利用し, 発話文の重要な単語に注目したり, 前後の発話同士の関係性を考慮する事を行う.

2.2　対話行為推定

対話行為は, 発話文の「内容」や「意味」を表しており, その「内容」や「意味」を種類分けしたものを対話行為タグと呼ぶ [9]. 対話行為推定とは, 入力された発話文の対話行為タグを予測する問題である.

大原らの研究では, 階層型 Recurrent Neural Network(以下,RNN) を用い, 応答の対話行為推定を行った. そして, 応答の対話行為推定の難しさと推定の向上は応答の選択性能の精度向上になると示した.

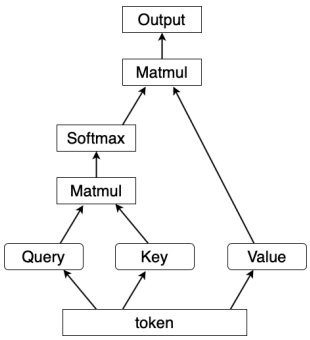

図 1　自己注意機構の概要図

また, 田中らの研究では, 対話行為系列を考慮するために対話行為 Encoder を構築した. これにより, 過去の対話行為と次の対話行為の関係性を直接学習することができ, 通常の階層型 RNN より高い精度を出す事ができた. しかし, 自然な対話を行えるようになるためには, 現在の精度をより向上させる必要がある. そのため改善案として, 発話文中の重要な単語を考慮する手法と, 直近の発話同士を重点的に考慮する手法を提案する.

3　提案手法

3.1　応答の対話行為推定

本研究では, 田中らのモデルをベースモデルとし, 自己注意機構を用いて対話行為推定の精度向上を行う. 以下の 3 つのモデルを提案し, 評価を行う.

- Utterance layer Attention (以下, Ulayer Attention)
  - 発話文中の重要な単語が注目する手法
- Context layer Attention (以下, Clayer Attention)
  - 直近の発話を重点的に考慮する手法
- Combination Attention (以下, Cmb Attention)
  - Ulayer Attention と Clayer Attention を組み合わせた手法

なお,Clayer Attention の対話, 対話行為に対して自己注意機構を用いる箇所は, Vipul ら [10] の手法を参考にした.

3.1.1　Ulayer Attention

Ulayer Attention の概要を図 2 に示す. Ulayer Attention では, まず発話の単語系列を入力とし, 位置付けエンコーダによって 1 つずつトークンに位置ベクトルを足していく. そして, 自己注意機構でそれぞれのトークンの位置関係を考慮し, 重要度の高いトークンが考慮されたベクトルを生成する. 話者を 識別するための話者ベクトルと連結させ, RNN で逐次的な処理をし, Neural Network(NN) を用いて対話行為を推定する.

1)　琉球大学大学院理工学研究科工学専攻, Graduate School of Engineering and Science, University of the Ryukyus
2)　琉球大学工学部工学科知能情報コース,Computer Science and Intelligent Systems, University of the Ryukyus

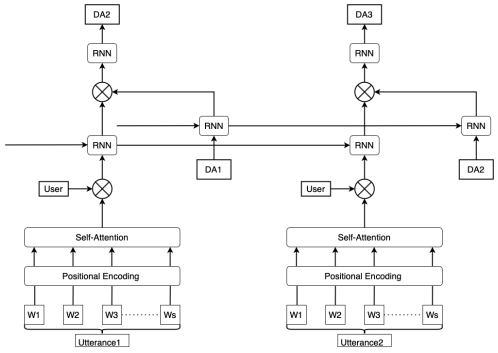

図 2　Ulayer Attention の概要図

3.1.2　Clayer Attention

Clayer Attention の概要を図 3 に示す. Clayer Attention では, RNN を用いて発話文をベクトル化し, 話者ベクトルを連結する. そして, RNN の隠れ層の出力値である前回の発話系列ベクトルと発話ベクトルを入力とし, その両方を考慮する事で, 直近の発話情報を考慮したベクトルを生成し, 再び RNN に入力する. 自己注意機構の入力から RNN に入力するまでの一連の処理部分は, Vipul らの手法を参考にした. そして,RNN の出力したベクトルを, 先ほどと同様の処理を行い生成した対話行為ベクトルと連結し,NN に入力し対話行為を推定する.

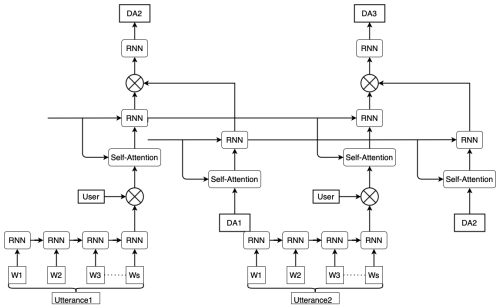

図 3　Clayer Attention の概要図

3.1.3　Cmb Attention

Cmb Attention の概要図を図 4 に示す. Cmb Attention は, 重要な単語の考慮, 直近の発話を重点的に考慮する 2 つの手法を 1 にまとめたモデルである.

Ulayer Attention と Clayer Attention の有効点である単語系列と発話文系列の処理をする箇所に自己注意機構を用いた.

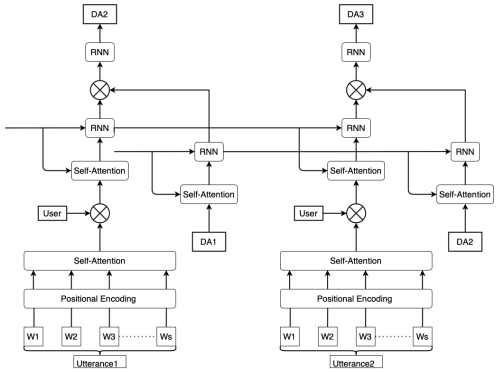

図 4　Cmb Attention の概要図

4　評価実験

4.1　実験データ

今回の実験で使用したデータセットは, Switchboard Dialog Act Corpus(以下, SwDA) [11] と, Daily Dialog Corpus(以下, Daily) [12] を使用する. SwDA では, 電話会話を文字に書き起こしているのに対して, Daily は一般的な日常会話を文字に書き起こしている. また, Daily では, 比較的にノイズが少ないという性質を持っている. 両データセットにも共通することとして, 1 つの発話文に対して, 1 つの対話行為タグが決められている. SwDA では, 本来 42 個の対話行為タグが付けられているが, 田中らの研究と同様にタグ数を 9 つに削減し, Daily では 4 個の対話行為タグで実験を行った。本研究で使用する対話行為タグの種類と割合を表 1 に示した. 1 つの対話の中に含まれている対話応答系列長は各対話によって長さが異なるため, 本研究では最大対話応答系列長 (以下, 系列長) を 5 と設定する. サイズ 5 の系列長をスライドし対話データを抽出する.

タグ	内容	割合
Statement	説明	54.3%
Understanding	理解	22.7%
Uninterpretable	フィラー	8.8%
Agreement	同意	5.2%
Question	質問	5.1%
Other	その他	1.9%
Apology	謝罪	1.1%
Greeting	挨拶	0.6%
Directive	命令	0.3%
Inform	情報	45.14%
Question	質問	30.25%
Directive	命令	15.95%
Commissive	任意の	8.66%

表 1　実験データのタグ情報　上:SwDA 下: Daily

4.2　実験設定

本実験で使用したハイパーパラメータの設定を示す.

- 単語 Embedding の次元数: 256
- 自己注意機構/RNN の隠れ層の次元数: 512
- 対話行為 Embedding/隠れ層の次元数: 128

本実験での語彙サイズを 20000 と設定し, 未知の単語に対して特殊記号”UNK”とした上で訓練を行う. また, 訓練のバッチサイズを 128 とし, ロス関数には交差エントロピー誤差, 最適化には Adam(学習率 0.00005) を用いた.

4.3　評価指標

本実験の 評価指標として Accuracy とマクロ平均 F 値を用いる. また, どの対話行為タグがよく予測されているかを調査するため, 各対話行為タグの F 値を求めた.

5　実験結果

各モデルの精度を表 2 , 各モデルの対話行為タグの F 値を表 3 にまとめた. また, 提案の有効性について考察するため, 表 4 には田中らと Ulayer Attention, Cmb Attention の対話例を示し, 図 5, 6 に可視化した結果を示す. さら

に, SwDA の F 値が最も高い Cmb Attention の混同行列を図 7 に示した.

5.1 各モデルの評価

5.1.1 各モデルの精度

表 2 では, 各モデルの Accracy と F 値を示す. SwDA では, 提案モデルの Accuracy が 1.6%, F 値が 1.8% と, Daily では Accuracy が 0.4%, F 値が 0.8% と共に, 微増だが向上した. Daily と比べ,SwDA が僅かながら向上した. SwDA が向上した理由として, 表 3 から, 「Apology」タグと「Uninterpretable」の F 値が向上したためだとわかる. しかし, 大幅な精度改善につながるとは言えない結果であった. その原因として, 応答はその時の環境や状況, 心情によって変化するため, 一貫性がみられない. これにより, 次の対話行為を予測することは非常に困難で, 自己注意機構を用いても大幅な向上につながらなかったと考えられる.

Model	SwDA(Accuracy)	Daily(Accuracy)	SwDA(F 値)	Daily(F 値)
田中ら	57.8%	60.4%	15.9%	53.3%
Ulayer Attention	58.8%	60.7%	17.5%	54.0%
Clayer Attention	58.4%	60.3%	17.4%	51.0%
Cmb Attention	**59.4**%	**60.8**%	**17.7**%	**54.1**%

表 2 各手法の評価

5.1.2 各対話行為タグの F 値

表 3 では, 各対話行為タグの F 値を示す. 表 1 から, 割合の高い「Statement」タグや「Understanding」タグの F 値が高く, 割合の低い「Directive」タグや, 「Apology」タグ, 「Other」タグなどの F 値が低い結果となった. しかし, 「Greeting」タグは 全体の 0.6 %で, 他の少数のタグと比べ F 値が高い. この理由として, 「挨拶」は会話の終わりや始めなどによく頻出するため, 他のタグと比べ比較的に予測しやすいからだと考えられる.

また, 提案モデルは「Commissive」タグ以外の対話行為タグでは F 値が低下せず, 「Understanding」タグや「Greeting」タグでは大幅な向上が見られる.

本研究では, 田中らのモデルでは全く予測が行えなかった「Apology 」タグに注目し考察を行う.

タグ	田中ら	Ulayer Attention	Clayer Attention	Cmb Attention
Statement	72.2%	**73.3**%	72.7%	**73.3**%
Understanding	50.8%	53.9%	54.4%	**55.8**%
Uninterpretable	0.4%	1.9%	0.1%	**2.5**%
Agreement	0.7%	**2.8**%	1.6%	2.3%
Question	0.0%	**0.1**%	0.0%	**0.1**%
Other	0.0%	0.0%	0.0%	0.0%
Apology	0.0%	0.3%	0.0%	**1.4**%
Greeting	19.8%	25.5%	**28.5**%	23.8%
Directive	0.0%	0.0%	0.0%	0.0%
Inform	68.7%	69.1%	65.7%	**69.9**%
Question	53.5%	**53.7**%	53.3%	53.4%
Directive	37.7%	41.5%	**43.3**%	40.3%
Commissive	**55.3**%	52.7%	53.7%	54.2%

表 3 各対話行為タグの F-Score 上:SwDA 下: Daily

5.2 実験考察

5.2.1 Apology タグの詳細分析

表 2 から, 提案モデルの推定精度の大幅な向上を示す事はできなかった. しかし, 表 3 から, 従来モデルの「Apology」タグの F 値は 0.0 %に対して, Cmb Attention では 1.4 %を出す事ができた. 「Apology」タグがなぜ予測できるようになったのか, その理由を調査するため表 4 に対話の例をまとめた.

表 4 では, 「Apology」タグを推定した対話例となる. この例文から「Apology」タグを予測するためには, 直前の発話文である「Question」タグの内容を理解する必要がある. 発話文の内容としては, 相手に対して「Yes/No」か質問を行っている. Ulayer Attention と Cmb Attention の両モデルで, 発話文「Does your dad have horses there」の可視化を行った結果を図 5, 6 に示す. 「Yes/No」質問文の内容を認識するためには, 発話文中の動詞と名詞の関係性を考慮する必要がある. 図 5, 6 から, どちらも考慮されている. さらに, 「Yes/No」質問では, 「Yes」か「No」の応答が返されるため発話同士の関係性を学習する必要がある. Cmb Attention では, 前後の発話に関して重点的に考慮しているため, Ulayer Attention よりも予測し正解できたと考えられる. 表 3 からでも, Cmb Attentoin が Ulayer Attention よりも「Apology」タグの F 値が高い事がわかる. しかし, Cmb Attention の「Apology」タグの F 値は依然として低い. 次は, その原因とその影響について示す.

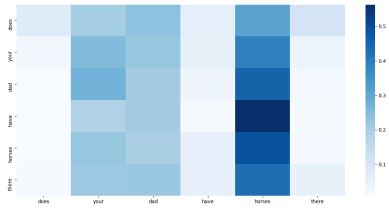

図 5 Cmb Attention の可視化

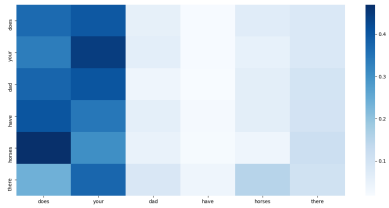

図 6 Ulayer Attention の可視化

5.2.2 SwDA の対話行為タグ

表 2 より, SwDA と Daily の F 値は最大 37.4% 差が生まれている. 表 1 から, SwDA はタグ数の不均衡性が強い事がわかる. そのため, 多数派のタグのみがよく学習されてしまう. また, SwDA と Daily の共通タグである「Question」タグと「Directive」タグも不均衡性から, F 値に大きな差が出ている. データセットの不均衡性の影響を詳細に分析するため SwDA の混同行列を示す.

SwDA の混同行列

図 7 は, Cmb Attention による SwDA の混同行列である. 横が予測タグであり, 縦が正解タグとなる. 多数派の「Statement」タグと「Understanding」タグがよく予測されている. また, それ以外のタグは全く予測できていない. これにより, SwDA の対話行為推定では, 割合の高いタグしか予測されない事がわかる. そのため, 実際の対話を

発話文 (対話行為)	正解	田中ら	Ulayer Attention	Cmb Attention
1 That's for sure the cleaning up can be a mess (Statement)	Uninterpretable	Statement	Understanding	Statement
2 Um-hum (Uninterpretable)	Question	Statement	Statemen	Statement
3 But do you have horses or anything at your dad's farm (Question)	Apology	Understanding	Statemen	Statement
4 I'm sorry (Apology)	Question	Understanding	Understanding	Understanding
5 Does your dad have horses there (Question)	Apology	Statement	Statemen	**Apology**

表 4　従来手法と提案手法の対話例

行うと, 似たような内容の発話しか行わないモデルとなり, 会話の面白みがないと考えられる. この問題を解決するために, 不均衡なデータセットにも対応した学習を行う必要がある.

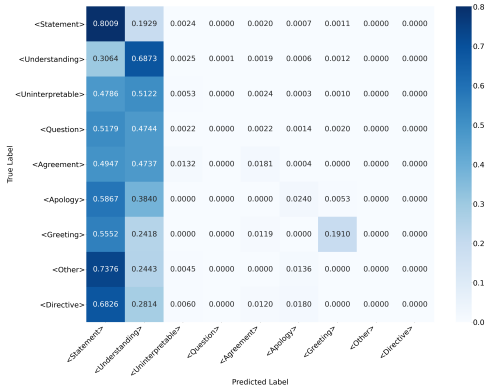

図 7　SwDA の混同行列

6　終わりに

本研究では, 自己注意機構を用いた応答の対話行為推定モデルを提案し, 評価を行った. その結果として, 各提案を組み合わせた Cmb Attention は, SwDA と Daily の Accuracy と F 値が, 従来モデルと比べ向上した. しかし, 新たな問題として SwDA の F 値が極端に低い事が判明した. その理由として, データセットの不均衡性にあるとわかった. そのため, データセットの偏りを考慮した学習を行うか, データセットの調整を行わなければならない. 後者では, 対話が持つ発話同士の前後関係を失わないように注意して行わなくてはならず, 非常に困難である. そのため, 今後は前者の取り組みを行う. コスト考慮型学習は, 多数派のタグと少数派のタグに同等の損失値を与えるのではなく, タグ数に応じた損失値を与えることで, 不均衡性をうまく学習できるようにする. 今後の研究として, 対話時系列に対応したコスト考慮型の損失関数を提案するための調査, 検討を進めていく予定である.

参考文献

[1] Ahmadvand, Ali, Jason Ingyu Choi, and Eugene Agichtein. "Contextual dialogue act classification for open-domain conversational agents." Proceedings of the 42nd international acm sigir conference on research and development in information retrieval. 2019.

[2] Li, Ruizhe, et al. "A dual-attention hierarchical recurrent neural network for dialogue act classification." arXiv preprint arXiv:1810.09154 (2018).

[3] Kumar, Harshit, et al. "Dialogue act sequence labeling using hierarchical encoder with crf." Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 32. No. 1. 2018.

[4] Chen, Zheqian, et al. "Dialogue act recognition via crf-attentive structured network." The 41st international acm sigir conference on research & development in information retrieval. 2018.

[5] 大原康平. "対話行為を考慮したニューラル雑談対話モデル." (2018).

[6] 田中昂志, 高山隼矢, and 荒瀬由紀. "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." 人工知能学会全国大会論文集 第 33 回全国大会 (2019). 一般社団法人 人工知能学会,

[7] Vaswani, Ashish, et al. "Attention is all you need." arXiv preprint arXiv:1706.03762 (2017).

[8] Bahdanau, Dzmitry, Kyunghyun Cho, and Yoshua Bengio. "Neural machine translation by jointly learning to align and translate." arXiv preprint arXiv:1409.0473 (2014).

[9] BOYER, Kristy, et al. Dialogue act modeling in a complex task-oriented domain. In: Proceedings of the SIG- DIAL 2010 Conference. 2010. p. 297-305.

[10] Raheja, Vipul, and Joel Tetreault. "Dialogue act classification with context-aware self-attention." arXiv preprint arXiv:1904.02594 (2019).

[11] Jurafsky, Daniel, and Elizabeth Shriberg. "Switchboard SWBD-DAMSL Shallow-Discourse-Function Annotation Coders Manual, Draft 13 Daniel Jurafsky*, Elizabeth Shriberg+, and Debra Biasca** University of Colorado at Boulder &+ SRI International." (1997).

[12] Li, Yanran, et al. "Dailydialog: A manually labelled multi-turn dialogue dataset." arXiv preprint arXiv:1710.03957 (2017).