傾聴態度を示す応答の生成における表出可能な応答種類の推定とその利用

田中涼雅 [1]　村田匡輝 [2]
[1] 豊田工業高等専門学校専攻科 情報科学専攻　[2] 豊田工業高等専門学校 情報工学科
j2401@toyota.kosen-ac.jp　murata@toyota-ct.ac.jp

概要

会話エージェントが人間に代わって語りの聴き手となることが期待されている．聴き手は，語り手に対して相槌などの応答をすることが重要であり，会話エージェントが聴き手として認められるためには，傾聴態度を示す応答（傾聴応答）の生成が望まれる．傾聴応答には複数の種類が存在し，語り手の発話に対する適切なものは必ずしも 1 つではない．そのため，発話に適する応答種類を複数推定したうえで，その種類に基づいて応答表現を生成することが望ましい．本論文ではある発話に対して表出可能な応答を複数生成する手法について述べる．語り手の発話に適する応答種類をマルチラベル分類により複数推定，発話と推定結果から応答表現を生成する．実験により，応答種類の推定において macro-F1 で 0.83 を得た．また応答表現の生成において，比較モデルより Distinct-N の値が向上していたことから，応答種類を利用する有効性を確認した．

1　はじめに

人間の基本的な欲求の 1 つに語ることがある．語るには聴き手が必要だが，コロナ禍におけるリモートワークの普及などによる社会の個人化および高齢化の進む現代においては，聴き手不在の状況が増加している．そこで，人の代わりに会話エージェントが聴く役割を担うことが考えられる．しかし，多くの場合，会話エージェントは人の語りを聴いている間は無反応であり，聴き手としての役割を十分に果たせていない．会話エージェントが語りの聴き手として認められるには，語りを傾聴していることを語り手に伝達する機能を備えることが必要となる．このための明示的な手段は語りに応答することであり，傾聴を示す目的で発話する応答，すなわち**傾聴応答**の表出が有力である．

傾聴応答を生成するためには，発話と応答をペアにしたデータを用いて応答生成モデルを学習することが考えられる．しかし一般に，傾聴応答にはその役割に応じて複数の種類があり，ある発話に対する適切な傾聴応答は必ずしも 1 つではない．また，そのようにして学習したモデルでは，どのような発話に対しても表出可能な汎用的な応答が生成されやすいという問題もある [1]．

そこで本論文では，ある発話に対して多様な応答を生成する手法を提案する．まず，発話に対して表出可能な応答の種類をマルチラベル分類によって推定する．次に，発話と推定した応答種類を用いて応答表現を生成する．傾聴応答を収集したデータを用いた実験を実施し，表出可能な応答種類を推定し応答生成に用いることの効果を確認した．

2　傾聴応答

2.1　語りと応答

傾聴応答とは語りを傾聴していることを語り手に伝えるための応答のことである．傾聴応答の代表例は相槌であり，それ以外にも感心，同意，繰り返しなどの種類がある [2]．語りと 2 名の聴き手による傾聴応答の例を表 1 に示す．【】で囲んだものは傾聴応答の種類を表す．「地方に行った時そこにある美術館にはなるべく行くようにしています。」という発話に対して，聴き手 A は感心を示す応答や繰り返しを，聴き手 B は評価や同意を示す応答を行っており，1 つの発話に対して複数の種類の応答が表出可能なことが分かる．

2.2　傾聴応答コーパスを用いた調査

実際に一人の語りに複数の聴き手が応答を行ったデータを使用し，ある発話に対して表出可能な応答が複数存在するか調査する．調査には Ito らが構築

表 1　傾聴応答の例

語り	傾聴応答（聴き手 A）	傾聴応答（聴き手 B）
地方に行った時	はい　【相槌】	うん　【相槌】
そこにある美術館には	えええええ　【相槌】	
なるべく行くようにしています	ああ　【感心】	あー　【感心】
	そうですかー　【感心】	いいですね　【評価】
	美術館に　【繰り返し】	行きたいです　【同意】

表 2　発話と対応付けた応答種類の例

発話	応答種類
できたひもをカラフルなひもを何に使ったのかはちょっと覚えていないし使いようがなかったのかなって	相槌，感心，同意，繰り返し，驚き，意見，想起，その他
趣味は大正琴をもう三十年近く四十代の終わりから現在までやってます。あとは	相槌，感心，評価，繰り返し，言い換え，納得，驚き

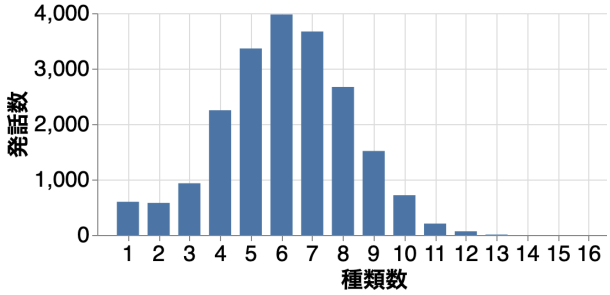

図 1　発話あたりの応答種類数

した傾聴応答コーパス [3] を用いる．傾聴応答コーパスは，語りデータとして高齢者のナラティブコーパス JELiCo[4] を利用し，録音された 30 人の高齢者の語りに対してそれぞれ 11 人の聴き手が応答を行うことで構築されている．そのため，同一の語りに対して 11 人分の多様な傾聴応答が収録されている．応答にはその役割に応じて，相槌，感心，評価，同意，不同意，繰り返し，言い換え，納得，驚き，驚きといぶかり，意見，補完，あいさつ，想起，考え中，その他の 16 種類のラベルが付与されている．

傾聴応答コーパス中の語りと収集した応答には形態素単位の発話時間情報が付与されている．この時間情報を用いて，調査のためのデータを作成する．まず，語りを文節単位に分割し，ある文節とその文節の発話時間内に発話が開始されている応答を対応付ける．そして発話の単位をある文節からその文節の N 文節前までを連結したものとし，各発話内で行われた応答の種類をリストアップする．今回は N を 10 とした．作成したデータの例を表 2 に示す．

発話あたりの応答種類数のヒストグラムを図 1 に示す．平均応答種類数は 5.09 種類であり，全発話中 2 種類以上の応答種類がリストアップされたものは 94.2 %であった．ほとんどの発話に対して複数種類の応答が表出可能なことが分かる．

2.3　関連研究

傾聴応答の生成についての研究として，代表例である相槌の生成に関するものが多く存在する [5, 6, 7]．音響情報および言語情報を用いた相槌の生成タイミングの検出が試みられている．また，相槌以外の傾聴応答に関しても，繰り返し応答や不同意応答といった特定の種類の応答を生成する手法の提案も行われている [8, 9]．さらに，それらを含めた傾聴応答全般を生成するシステムの開発が行われている [10, 11]．これらのシステムでは，ルールや優先順位を設けて発話に対する応答種類，表現が決定している．

一方，本研究では，実際の傾聴応答のデータを用い，ある発話に対する表出可能な応答の種類を複数推定し，その種類に対応する応答表現を生成する点でアプローチが異なる．

3　手法

提案手法の流れを図 2 に示す．提案手法ではまず，発話に対して表出可能な応答種類をマルチラベル分類によって推定する．次に，発話と推定した種類を入力とし，応答表現を生成する．推定手法と生成手法について，それぞれ 3.1 節と 3.2 節で述べる．

3.1　応答種類の推定手法

応答種類の推定では，発話のテキストを入力とし，16 種類の応答種類の推定結果を出力する．応答種類の推定モデルを図 3 に示す．マルチラベル分

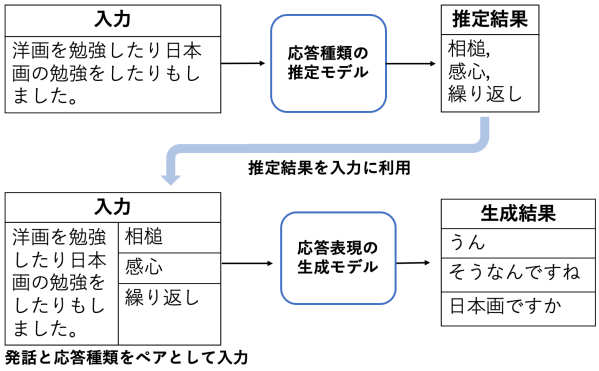

図 2　提案手法の流れ図

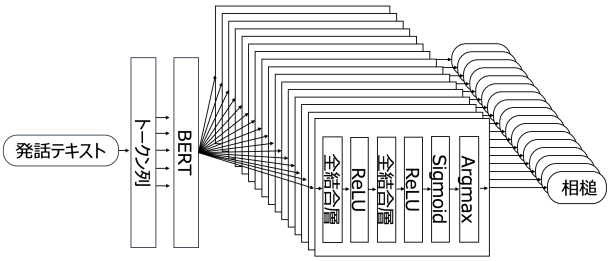

図 3　応答種類の推定モデル

類の方法として Binary Relevance Neural Network[12] を用いる．入力テキストをトークン列に分割し，BERT[13] により [CLS] トークンに対応するベクトルを得る．そのベクトルを各応答種類に対応する 16 個の二値分類器に入力し，応答種類の推定結果を得る．二値分類器は ReLU 関数を活性化関数とした 2 層の全結合層，および，Sigmoid 関数によって構成する．BERT はあるラベルに対するパラメータを他のラベルの推定時にも反映させるために各二値分類器で共有する．

事前学習済みの日本語 BERT モデルを発話と応答種類をペアにしたデータを用いて fine-tuning する．損失関数には Binary Cross Entropy Loss を用いる．以上の方法により，ある発話に対して可能な応答種類を複数推定する．

3.2　応答表現の生成手法

応答表現の生成では，発話のテキストと応答種類のテキストのペアを入力とし，応答表現のテキストを出力する．生成には BART[14] を使用する．事前学習済み日本語 BART モデルを発話と応答種類，および，応答表現をペアにしたデータを用いて fine-tuning する．応答の種類によって発話を条件付けることによって，1 つの発話に対して異なる応答の生成を可能とする．

表 3　推定実験の結果

種類	提案手法			ランダム		
	Pre.	Rec.	F1	Pre.	Rec.	F1
相槌	0.990	0.995	0.993	0.970	0.968	0.969
感心	0.984	0.988	0.986	0.925	0.922	0.923
評価	0.888	0.843	0.865	0.395	0.393	0.394
同意	0.843	0.861	0.852	0.352	0.355	0.353
不同意	0.888	0.747	0.800	0.047	0.047	0.047
繰り返し	0.900	0.928	0.917	0.672	0.665	0.668
言い換え	0.823	0.777	0.800	0.309	0.309	0.309
納得	0.843	0.799	0.820	0.344	0.350	0.347
驚き	0.866	0.783	0.823	0.217	0.212	0.214
驚きといぶかり	0.966	0.574	0.720	0.006	0.006	0.006
意見	0.861	0.708	0.777	0.141	0.139	0.140
補完	0.839	0.699	0.762	0.165	0.164	0.164
あいさつ	0.890	0.742	0.809	0.035	0.37	0.036
想起	0.880	0.758	0.815	0.028	0.029	0.029
考え中	0.863	0.690	0.767	0.081	0.080	0.080
その他	0.850	0.845	0.847	0.374	0.369	0.372

4　実験

提案手法の有効性を確認するために，傾聴応答コーパスを用いた応答種類の推定実験，応答表現の生成実験をそれぞれ行った．

4.1　応答種類の推定実験

4.1.1　推定実験の概要

傾聴応答コーパスから 2.2 節の方法で作成した，ある文節と直前 N 文節からなる発話，複数の応答種類をペアとしたものを実験データに使用した．データ数は 21,586 個であり，学習，開発，テスト用に 8:1:1 に分割して用いた．

評価指標には各応答種類の推定における適合率 (Prescision)，再現率 (Recall) および F 値を用いる．また，応答種類の推定全体に対する評価指標として macro-F1 を用いる．各応答種類の出現確率に基づいてランダムに選択する手法を比較手法とする．

事前学習済み日本語 BERT モデルには “tohoku-nlp/bert-base-japanese-char-whole-word-masking”[1] を用いた．学習率は 0.00001 とし，最適化アルゴリズムには Adam を用いた．N を 10，バッチサイズを 16，1 層目，2 層目の全結合層の次元数を 512 とし，5 エポック学習した時点のモデルが開発データにおいて最も損失が小さくなったため，このモデルをテストデータに適用した．

1)　https://huggingface.co/tohoku-nlp/bert-base-japanese-char-whole-word-masking

表 4　応答表現の生成例

発話	比較モデル	提案モデル（推定した応答種類を入力）	正解の応答
そういうことがありまして今現在でもですねそういう人達との交流があるからこそ今の私があるということで	あー	そうですか　【感心】	えー　【驚き】
		凄いですね　【評価】	凄い　【評価】
		いいですね　【同意】	いいですね　【同意】

※【】で囲んだものは応答種類を表す

表 5　生成実験の結果

応答	Distinct-1	Distinct-2
比較モデル	0.0058	0.0093
提案モデル（推定した応答種類を入力）	0.0122	0.0110
提案モデル（正解の応答種類を入力）	0.0159	0.0535
聴き手	0.0124	0.0295

4.1.2　推定実験の結果

実験結果を表 3 に示す．提案手法はいずれの応答種類においても F 値 0.7 以上を得た．提案手法の macro-F1 は 0.83，ランダムの macro-F1 は 0.35 であり，ランダムの結果を大きく上回ったことから，表出可能な応答種類の推定が高い精度で実現できていることを確認した．

4.2　応答表現の生成実験

4.2.1　生成実験の概要

ある文節と直前 N 文節からなる発話，複数の応答種類をペアとしたデータを，発話と各応答の種類，および，各応答の表現からなるデータとして作成し直し，実験データとして使用した．N は 10 とした．データ数は 145,291 であり，学習，開発，テスト用に 8:1:1 に分割して用いた．

生成した応答表現の多様性を評価するため，評価指標として Distinct-N を用いる．応答種類を用いず発話と応答表現のみで事前学習済み日本語 BART モデルを fine-tuning したものを比較モデルとする．

事前学習済み日本語 BART モデルには “ku-nlp/bart-base-japanese”[2] を用いた．学習率は 0.00001 とし，最適化アルゴリズムには RAdam を用いた．バッチサイズは 32 とし，開発データにおいて 4 エポック学習した際の損失が最小となったモデルをテストデータに適用した．

4.2.2　生成実験の結果

応答表現の生成例を表 4 に，実験結果を表 5 にそれぞれ示す．表 4 より，比較モデルは同じ発話に対して 1 つの応答を生成しているのに対し，提案モデルは応答の種類も入力することにより，異なる表現を生成できていることが分かる．しかし，正解の応答と一致しない応答表現も生成されている．表 5 には，比較モデル，推定した応答種類を入力とした提案モデルの生成結果に加え，参考として正解の応答種類を入力としたもの，聴き手による応答の値を合わせて示した．推定した応答種類を入力とした提案モデルの方が比較モデルよりも高い Distinct-1, 2 の値を得られていることから，多様な応答の生成において，応答種類を利用する有効性を確認した．

表 5 において，推定した応答種類を入力とした提案モデルの Distinct-2 の値は聴き手のそれよりも低い．これは，提案モデルでは 1 つの発話から各種類の応答表現は最大でも 1 つしか生成されないのに対し，聴き手の場合は，聴き手が異れば応答種類が同じでも異なる表現として Distinct-2 を計算しているためと考えられる．また，正解の応答種類を入力とした提案モデルによる応答の方が聴き手による応答よりも Distinct-1, 2 の値が高くなっている．正解の応答種類を入力とした提案モデルによって生成された応答を確認したところ，発話の語句を用いて応答する繰り返しなどにおいて，聴き手の応答よりも長めの表現が生成されていた．そのため，1-gram および 2-gram の異なり数が増えたと予想される．

5　まとめ

本論文では発話から多様な応答を生成する手法について述べた．応答種類をマルチラベル分類により複数推定し，発話と推定した種類から応答表現を生成する．実験では，応答種類を macro-F1 で 0.83 の精度で推定できることを確認し，また，生成においては，応答種類を用いない場合よりも生成した応答表現の Distinct-1 の値が向上することを確認した．

しかし，生成した応答表現を正解と比較すると，一致度は十分とはいえず，Distinct-1 の値が向上したとはいえ相槌など汎用的な応答の生成も少なくなかった．今後の課題として，生成における応答種類の利用方法の検討，発話に対して表出可能な応答を系列として最適化するといったことが挙げられる．

2）https://huggingface.co/ku-nlp/bart-base-japanese

謝辞

高齢者のナラティブコーパスは，奈良先端科学技術大学院大学ソーシャル・コンピューティング研究室から提供いただいた．

参考文献

[1] 村田匡輝, 大野誠寛, 松原茂樹. 系列変換モデルに基づく傾聴的な応答表現の生成. 言語処理学会第 24 回年次大会発表論文集, pp. 821–824, 2018.

[2] 日本語記述文法研究会. 現代日本語文法 7. くろしお出版, 2009.

[3] Koichiro Ito, Masaki Murata, Tomohiro Ohno, and Shigeki Matsubara. Construction of responsive utterance corpus for attentive listening response production. In **Proceedings of the 13th Language Resources and Evaluation Conference**, pp. 7244–7252, 2022.

[4] Eiji Aramaki. Japanese elder's language index corpus v2. https://figshare.com/articles/dataset/Japanese_Elder_s_Language_Index_Corpus_v2/2082706/1, 2016.

[5] Nigel Ward and Wataru Tsukahara. Prosodic features which cue back-channel responses in english and japanese. **Journal of Pragmatics**, Vol. 32, No. 8, pp. 1177–1207, 2000.

[6] 大野誠寛, 神谷優貴, 松原茂樹. 対話コーパスを用いた相づち生成タイミングの検出. 電子情報通信学会論文誌 A, Vol. J100-A, No. 1, pp. 53–65, 2017.

[7] Jin Yea Jang, San Kim, Minyoung Jung, Saim Shin, and Gahgene Gweon. BPM_MT: Enhanced backchannel prediction model using multi-task learning. In **Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing**, pp. 3447–3452, 2021.

[8] Toshiki Kawamoto, Hidetaka Kamigaito, Kotaro Funakoshi, and Manabu Okumura. Generating repetitions with appropriate repeated words. In **Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies**, pp. 852–859, 2022.

[9] 伊藤滉一朗, 村田匡輝, 大野誠寛, 松原茂樹. 語りの傾聴において不同意を示す応答の生成. 自然言語処理, Vol. 31, No. 1, pp. 212–249, 2024.

[10] 下岡和也, 徳久良子, 吉村貴克, 星野博之, 渡部生聖. 音声対話ロボットのための傾聴システムの開発. 自然言語処理, Vol. 24, No. 1, pp. 3–47, 2017.

[11] 井上昂治, ラーラーディベッシュ, 山本賢太, 中村静, 高梨克也, 河原達也. アンドロイド ERICA の傾聴対話システム –人間による傾聴との比較評価–. 人工知能学会論文誌, Vol. 36, No. 5, pp. H–L51_1–12, 2021.

[12] Huihui He and Rui Xia. Joint binary neural network for multi-label learning with applications to emotion classification. In **Proceedings of the 7th CCF International Conference on Natural Language Processing and Chinese Computing**, p. 250–259, 2018.

[13] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In **Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies**, Vol. 1, pp. 4171–4186, 2019.

[14] Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer. BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. In **Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics**, pp. 7871–7880, 2020.