

## 共感性を考慮した応答生成による傾聴対話モデルの改善

## Improvement of a Listening Dialogue Model through Empathy-Aware Response Generation

松本 奈々<sup>†</sup> 安藤 一秋<sup>‡</sup>  
Nana Matsumoto Kazuaki Ando

## 1. はじめに

近年、日本の総人口に占める高齢者人口の割合は 29.3% となり、日本における高齢化が進行するとともに、要介護者数も増加している[1, 2]。介護現場では、高齢者の発言に傾聴し、共感することが信頼関係の構築に重要であり、その実現手段の一つとして「バリデーション」が注目されている[3]。しかし、介護士の人材不足や介護負担の増大などの問題により、個人に十分な時間をかけてケアすることが困難である。

本研究では、介護環境の改善を目指して、バリデーションを活用した対話システムの構築を目的とする。著者らの先行研究[4]では、言語モデルを指示付きフィードバックを行うことにより、傾聴タイプの指示に従う応答生成が可能になったが、発話の感情傾向を考慮した応答生成や共感の深さに関して課題が残った。そこで本稿では、これらを考慮した対話応答生成を実現するために、共感の分類を新たに検討し、対話モデルの応答生成の改善を試みる。

## 2. 関連研究

## 2.1 共感の深さと応答の種類の関係

伊藤ら[5]は、傾聴応答の種類と共感度の関係に着目し、傾聴応答の種類を共感度の観点から 5 段階（Level 1（相槌）、Level 2（感心）、Level 3（同意、不同意、納得など）、Level 4（評価、繰り返し応答、言い換えなど）、Level 5（意見））に分類後、傾聴応答データを用いて分類の正当性を定量的に示した。分類の正当性を応答の汎用性の観点で評価した実験では、共感度が低いほど応答の汎用性が高く、共感度が高いほど応答の汎用性が低い結果となり、応答の汎用性と共感度はトレードオフの関係にあることを明らかにした。また、Level 3 から 4 にかけて共感度が大きく異なることを確認した。本研究では、伊藤らが提案した共感度を参考に、応答における共感の分類を定義する。

## 2.2 言語学的視点からの相槌表現

井上ら[6]は、様々な相槌表現について考察し、相槌が持つ機能に着目した上で、その機能ごとに、どのような関係性においてどの相槌表現が使用可能かを具体的に整理した。本研究では、井上らが提案した相槌表現の分類を参考に、具体的な共感表現を定義する。

## 3. 共感の分類と共感表現の定義

## 3.1 共感分類の定義

バリデーション対話技法と伊藤らの研究[5]を参考に共感

<sup>†</sup> 香川大学大学院創発科学研究科, Graduate School of Science for Creative Emergence, Kagawa University

<sup>‡</sup> 香川大学創造工学部, Faculty of Engineering and Design, Kagawa University

表 1: 共感分類の定義

共感度	分類
1	単純相槌
2	共感応答ポジティブ
	共感応答ネガティブ
3	共感応答ポジティブ+自己開示
	共感応答ネガティブ+自己開示

表 2: 共感表現の定義

分類	表現
単純相槌	「はい」、「えー」、「ええ」など
共感 P	「さすがですね」、「いいですね」など
共感 N	「お疲れ様です」、「おつらいですね」など
共感 P+自己開示	「すてきですね。私も〜」、「すばらしいですね。私も〜」など
共感 N+自己開示	「その気持ちはよく分かります。私も〜」、「それは大変ですね。私も〜」など

分類を定義する。バリデーション対話技法では、リフレーミング（会話の中で重要だと感じた言葉を反復する）というテクニックが存在することから、繰り返し応答は共感応答からは除く。また、バリデーション対話技法は、丁寧にコミュニケーションを取ることで被介護者の感情や想いを引き出すことを目的としていることから、Level 5 のような相手に自身の意見を投げかけるような発話は除外する。さらに、Level 3 から 4 にかけて共感度が異なることや、先行研究[4]での課題であった、発話の感情傾向（ポジティブ、ネガティブ）も考慮するために、本稿では、表 1 に示すように、共感を 3 つの共感度で分類して定義する。

## 3.2 共感表現の定義

共感度 1 の単純相槌は、村田ら[7]が傾聴応答データに対して応答の出現頻度を分析した結果を参考に定義する。

共感度 2 の共感応答は、井上ら[6]の研究と先行研究[4]での傾聴表現を参考に定義する。具体的には、井上らの話し手の事実報告に対する相槌、話し手の感情表出に対する相槌について、それぞれの表現を感情傾向（ポジティブ、ネガティブ）ごとに分類し、さらに、先行研究[4]で利用した表現も追加することで定義する。

共感度 3 の共感応答+自己開示は、共感度 2 の共感応答の語尾に「私も〇〇です。」という表現を追加することで定義する。これにより、発話者に対する深い共感を示すことができると考えられる。

具体的な共感表現の定義例を表 2 に示す。なお、共感応答ポジティブは共感 P、共感応答ネガティブは共感 N と表記する。ここで、共感度 1 に関しては、プロンプトに決められた表現を提示することで生成する。共感度 2 や 3 に関しては、プロンプトの工夫のみでは応答表現形式の一貫性

を保った生成が難しいことから、LLM に指示付きファインチューニングを実施し、生成を試みる。

## 4. 傾聴対話モデルの構築

### 4.1 データセットの構築

まず、日本語日常対話コーパス[8]の各対話データを発話-応答ペアの形式に整形する。次に、先行研究[4]で定義した掘り下げ質問、および、本稿で定義した共感 P、共感 N、共感 P+自己開示、共感 N+自己開示の各傾聴応答の表現に対して、掘り下げ質問は文末での一致、それ以外の応答は文頭からの一致で機械的に判定し、先ほど整形したデータに対してラベリングする。ここで、バリデーション技法のオープンクエスションでは、「なぜ～ですか?」や「どうして～ですか?」のような質問は、多くの被介護者が応答に困る傾向があるため、避けたほうがよい質問とされている。したがって、これらの表現を除外する。

ラベリングの結果、掘り下げ質問は 8,970 件、共感 P は 1,637 件、共感 N は 1,027 件、共感 P+自己開示は 91 件、共感 N+自己開示は 19 件、それ以外の応答は 23,967 件となった。ここで、共感 P、共感 N、共感 P+自己開示、共感 N+自己開示の応答件数が少ないことから、Llama-3-ELYZA-JP-8B<sup>1</sup>を用いて追加データを自動生成し、人手で選定することで、最終的にそれぞれの応答の種類が 5,000 件 (計 30,000 件) からなるデータセットを構築した。

### 4.2 モデルの構築

構築したデータセットと Llama-3-ELYZA-JP-8B を用いて QLoRA ファインチューニング<sup>2</sup>を実施することでモデルを構築する。ここで、学習プロンプトには、それぞれの発話の種類を指示して学習させる。学習は、5,000step まで実施し、loss が安定してから一定時間経過後の 3,000step を評価用モデルとして使用する。

## 5. 評価

### 5.1 評価用データセットの構築

評価用データセットでは、発話に感情傾向を持たせるために ChatGPT<sup>3</sup> (GPT-4o) を用いて、高齢者口調のポジティブな発話とネガティブな発話についてそれぞれ 500 発話を作成する。これと 4.2 節のモデルを用いて、1 発話につき 1 応答を生成させることで構築する。

### 5.2 実験設定

5.1 節のデータに対して、各傾聴応答の判定基準で、各傾聴応答の生成件数を機械的に判定する。ただし、共感応答 (P, N) & 掘り下げ質問や、共感応答 (P, N) + 自己開示 & 掘り下げ質問のような混合した応答が生成された場合、掘り下げ質問を優先してカウントする。生成される応答は毎回変化するため、3 回試行した平均値を用いて評価する。

### 5.3 評価結果

生成結果の件数の内訳を表 3 に示す。表 3 より、応答の種類

表 3 : 生成結果の件数

応答の種類	評価用発話の感情傾向	件数
掘り下げ質問	ポジティブ	462
	ネガティブ	440
共感 P	ポジティブ	472
共感 N	ネガティブ	498
共感 P+自己開示	ポジティブ	500
共感 N+自己開示	ネガティブ	499

が掘り下げ質問で、評価用発話の感情傾向がネガティブの場合以外については、9 割以上が指示に従う応答が生成できたことがわかる。掘り下げ質問に関しては、避けたほうがよい質問とされている「なぜ～ですか?」や「どうして～ですか?」のような質問が生成されていたことから、他の傾聴応答と比較して傾聴応答生成数が減少したと考えられる。共感応答に関しては、ポジティブ・ネガティブのそれぞれの発話に対して、共感 P+自己開示や共感 N+自己開示のような複雑な応答も生成できるようになり、ユーザーにより深い共感を与えられる可能性を確認した。

## 5.4 考察

本実験では、評価用データとして ChatGPT を用いて自動生成した高齢者口調の発話を用いたが、日本語日常対話コーパス以外の発話に対しても傾聴応答が生成できたことから、構築したモデルの汎用性を確認できた。掘り下げ質問に関しては、先行研究[4]と同様、避けるべき表現に対する学習を追加する必要があると考えられる。共感応答に関しては、実際に人手評価を実施するなどして、応答における共感度合いの違いを確認する必要がある。

## 6. おわりに

本稿では、新たに検討した共感分類をモデルに適用することで、応答生成の改善を試みた。先行研究と比較して、傾聴応答の種類が増加したが、傾聴応答の生成数に大きな減少は見られず、さらに、高齢者口調の発話にも応答が可能であることから、構築したモデルの汎用性を確認した。

今後は、掘り下げ質問に関しては、避けるべき応答に関する追加学習を実施し、対話制御も交えながら共感性について、人手評価を実施する。

### 参考文献

- [1] 統計からみた我が国の高齢者-「敬老の日」にちなんで-(引用日:2025年5月25日.)  
<https://www.stat.go.jp/data/topics/topi1420.html>.
- [2] 令和6年版高齢社会白書,(引用日:2025年5月25日.)  
[https://www8.cao.go.jp/kourei/whitepaper/w-2024/html/zenbun/s1\\_2\\_2.html](https://www8.cao.go.jp/kourei/whitepaper/w-2024/html/zenbun/s1_2_2.html).
- [3] 都村他, “認知症高齢者 ケアにおけるバリデーション技法に関する実践的研究”, 関西福祉大学紀要第14号, (2010).
- [4] 松本他, “傾聴タイプの指示に基づく傾聴応答生成”, 人工知能学会全国大会 (第39回), 3G6-GS-6-04, (2025).
- [5] 伊藤他, “語りの傾聴における共感の深さと応答の種類の関係”, 言語処理学会第26回年次大会発表論文集, pp.1400-1403, (2020).
- [6] 井上他, “機能による相槌の分類”, 言語処理学会第26回年次大会発表論文集, pp.1392-1395, (2020).
- [7] 村田他, “話し手への傾聴を示す応答発話の収集と分析”, 研究報告知能システム (ICS), pp.1-7, (2016).
- [8] 赤間他, “日本語日常対話コーパスの構築”, 言語処理学会第29回年次大会発表論文集, pp.108-113, (2023).

<sup>1</sup> <https://huggingface.co/elyza/Llama-3-ELYZA-JP-8B>

<sup>2</sup> [https://huggingface.co/docs/trl/sft\\_trainer](https://huggingface.co/docs/trl/sft_trainer)

<sup>3</sup> <https://chatgpt.com>