E-034

報酬表現に着目した強化学習による傾聴対話モデルの改善
Improvement of an Active Listening Dialogue Model with Reinforcement Learning Focused on Reward Expressions

松本 奈々† 安藤 一秋‡
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 29.1%と過去最高を更新し続けているともに，要介護者数も増加し続けている[1, 2]．介護現場において，高齢者の発言を傾聴することは信頼関係を築くために重要であり，一つの実現手段として「バリデーション」がある．しかし，介護士の人材不足等の問題から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[3]では，強化学習を用いた傾聴対話モデルを提案した．報酬モデル作成時の発話テンプレートとテンプレート数を改善することで，報酬推移が高くなり，対話モデルの出力における傾聴発話の割合を向上できたが，人手評価では低い評価項目がいくつか存在した．本稿では，先行研究[3]での課題に着目し，報酬モデル作成時における報酬を与える表現（以下，報酬表現）を改善する．評価実験では，尊重性や共感性の観点から効果を検証する．

2. 関連研究

2.1 報酬表現

東中ら[4]は，共感と自己開示に関する発話に着目し，ユーザとシステムにおけるこれらの発話が，ユーザの感じる親近感に対してどのような効果を持つかを対話実験により明らかにした．本研究では，東中らが提案した共感応答を参考に，報酬表現の枠組みの定義を変更する．

2.2 評価項目

細谷ら[5]は，カウンセラーの反射・バリデーション・肯定に注目し，それらとクライエントの被共感体験と心理的距離の関連を検討した．本研究では，細谷らが使用した評価尺度を参考に，提案する傾聴対話モデルに対して，尊重性や共感性を重視した評価を実施する．

3. 傾聴対話モデルの構築

強化学習による傾聴対話モデルの構築は，著者らの先行研究[3]と同様，①対話生成モデルの構築，②報酬モデルの構築，③強化学習モデルの構築の手順で実施する．①対話生成モデルには，先行研究[3]で構築したファインチューニングを 10epoch したモデル（FT10epoch）を使用する．本稿では，②報酬モデルの構築において，2 つの改善を試みる．なお，報酬モデルの構築において用いる発話テンプレート（東北大日本語日常コーパス[6]から人手で 960 発話を選定したデータ）とデータセット（発話テンプレートにおいて，1 発話につき 50 応答を生成したデータ）は，先行研究[3]で構築したものを使用する．

表 1：報酬表現の例

傾聴の種類	報酬表現の例
掘り下げ質問	？（なぜですか？，どうしてですか？ など 11 種類の表現以外）
尊重応答	「いいですね」，「良いですね」，「素晴らしい」，「すてきです」，「大変ですね」など
共感応答	「そうですね」，「よくわかります」など

3.1 報酬表現の選定

バリデーション技法のオープンクエスチョンでは，「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ことから避けたほうが良い表現とされている．したがって，掘り下げ質問に対してこれらの表現を除外する．また，評価応答と語彙的応答では，東中ら[4]や石田ら[7]の論文で提示された表現に対して，それぞれ Web 辞書[1][2]から類似表現を抽出する．その後，抽出した表現のうち，著者らの先行研究[3]で構築したデータセットに含まれている表現を新たに報酬表現に加える．ここで，東中ら[4]の研究を参考に，報酬の枠組み名について，評価応答は「尊重応答」，語彙的応答は「共感応答」と定義を変更する．報酬表現の例を表 1 に示す．

これらの報酬表現を用いて，掘り下げ質問は文末での一致，尊重応答と共感応答は文頭からの一致で機械的に判定し，条件を満たせば 1（報酬あり），条件を満たさなければ 0（報酬なし）でラベル付けする．ラベルづけの結果，正例は 19,472 件，負例は 21,505 件となり，正例負例を同比率にするために，19,400 件からなるデータを構築した．

3.2 報酬モデルの構築

報酬モデルは，構築したデータセットを用いて，BERT[3]で，2 値分類タスクとして学習することで構築する．分類判定の結果，F 値は，9 割を超える結果となった．

3.3 強化学習モデルの構築

強化学習には，Trl[4]の PPO を用いて，先行研究[3]と同様のパラメータで学習する．ただし，最大学習 step 数は 540 とする．評価対象のモデルは，報酬平均が高く，かつ平均が安定した step 数 360 で学習した傾聴対話モデルを用いる．

† 香川大学大学院創発科学研究科, Graduate School of Science for Creative Emergence, Kagawa University
‡ 香川大学創造工学部, Faculty of Engineering and Design, Kagawa University

[1] https://www.weblio.jp
[2] https://dictionary.goo.ne.jp
[3] https://huggingface.co/cl-tohoku/bert-base-japanese-v3
[4] https://huggingface.co/docs/trl

表 2: 傾聴個数の変化の結果

モデル	報酬表現	掘り下げ質問	尊重応答	共感応答	合計
ベースラインモデル	①	714	64	92	870
	②	639	65	85	790
本稿で構築したモデル	①	629	188	15	833
	②	626	**206**	26	858

4. 評価実験

本稿では，傾聴個数の変化に基づく機械的評価と，傾聴性満足度に関する人手評価の 2 つの評価を実施する．

4.1 傾聴個数の変化に基づく機械的評価

報酬表現の変更前後において，対話モデルが生成する応答の変化を確認するため，機械的評価を実施する．

先行研究[3]と同様の手順で，960 発話の発話テンプレートに対して，3.3 節で述べた提案モデルを用いて，1 発話につき 1 生成したデータに対して評価を実施する．ただし，生成される応答は毎回変化するため，3 回試行し，実験結果は 3 回の平均値を用いる．また，先行研究[3]と本稿では報酬表現が異なっていることから，先行研究[3]での報酬表現①，本稿での報酬表現②として，それぞれ算出する．

評価結果を表 2 に示す．なお，表 2 のベースラインモデルは，先行研究[3]における質問調整あり（960 発話）の FT10epoch モデルにおいて，強化学習を 360step まで実施したモデルである．表 2 より，ベースラインモデルと比較して，尊重応答が増えていることがわかる．また，掘り下げ質問や共感応答については，報酬表現の除外や変更により，生成数が減少したことが確認できる．したがって，報酬表現を変更することでモデルが生成する応答が変化したと考えられる．

4.2 傾聴性満足度に関する人手評価

本節では，提案モデルとベースラインモデルが生成した応答について，尊重性や共感性の観点から傾聴性満足度を人手で評価する．評価者は大学生 3 人である．

傾聴性満足度に関する人手評価では，4.1 節で作成したデータのうち，3 回目の生成結果を使用する．健康，運動，食事，季節・気候，趣味などに関する 30 発話を人手で選定し，1 発話に対する 1 応答レベルで評価する．評価者には，傾聴性満足度（傾聴内容に満足できる）について 4 段階（1：そう思わない〜4：とてもそう思う）で評価してもらった後，その理由を表 3 に示す項目から選択してもらう．

表 3: 評価値の選択理由の項目

理由項目	理由内容
4	相手の発話を尊重または共感しようとしていて，かつ，相手の発話に対して質問を行い，相手の発言に耳を傾けているから
3	相手の発話を尊重または共感しようとしているから
2	相手の発話に対して質問を行い，相手の発言に耳を傾けているから
1	その他（傾聴内容が発話に対して適切ではない，尊重や共感が感じられない，など）

表 2：傾聴性満足度の評価結果

モデル	傾聴性満足度
ベースラインモデル	3.055
本稿で構築したモデル	**3.511**

表 3：理由項目の選択数の 3 人の合計結果

モデル	項目 4	項目 3	項目 2	項目 1
ベースラインモデル	7	19	**39**	25
本稿で構築したモデル	**19**	**26**	35	10

傾聴性満足度の評価結果を表 4 に，評価値の選択理由の項目を表 5 に示す．ただし，傾聴性満足度の評価値は 3 人の平均値，評価値の選択理由の結果は 3 人の合計値とする．

傾聴性満足度は，ベースラインと比較すると，約 0.5 ポイント高くなった．理由項目では，項目 4 や 3 の数が多く，尊重性や共感性の個数がベースラインと比較して増加した．

4.3 考察

2 つの評価をもとに考察する．報酬表現を変更することで，掘り下げ質問や共感性に関する傾聴個数は減少したが，傾聴性満足度における評価値の選択理由では，項目 4 と 3 における選択個数が増加した．よって，報酬表現を厳密に選定することで，応答生成における内容の質が向上し，尊重性や共感性をさらに感じられるようなり，傾聴性満足度の向上につながったと考えられる．一方で，項目 1 の数より，内容に満足できない場合は，傾聴性満足度が得られないことから，発話に対する適切な応答の制御が必要であると考えられる．

5. おわりに

本稿では，報酬モデル作成時の報酬表現を改善し，評価実験により尊重性や共感性の観点からその効果を検証した．機械的評価の結果より，報酬表現を変更することで尊重応答が大幅に増加することを確認した．また，人手評価の結果より，報酬表現を変更することで，応答生成における内容の質が向上することで尊重性や共感性が増し，傾聴性満足度も向上することを確認した．

今後は，発話に対する応答の制御について，バリデーションの他のテクニックを参考にしながら検討する．

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:2024 年 6 月 5 日.) https://www.stat.go.jp/data/topics/topi1380.html.
[2] 令和 4 年版高齢社会白書, (引用日:2024 年 6 月 5 日.) https://www8.cao.go.jp/kourei/whitepaper/w-2022/html/zenbun/s1_2_2.html.
[3] 松本他, “強化学習を用いた傾聴対話モデルの構築法の改善”, 人工知能学会全国大会 (第 38 回), 2G4-GS-6-04, 2024.
[4] 東中他, “対話システムにおける共感と自己開示の効果”, 言語処理学会第 15 回年次大会発表論文集, pp.446-449, 2009.
[5] 細谷他, ” カウンセリング場面におけるカウンセラーの反射・バリデーション・肯定とクライエントの被共感体験・心理的距離との関連”, 日本女子大学大学院人間社会研究科紀要 第 22 号, pp.217-244, 2016.
[6] 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.
[7] 石田他, “共感表出と発話促進のための聞き手応答を生成する傾聴対話システム”, 人工知能学会研究会資料, SIG- SLUD-B509-02, 2018.