強化学習を用いた傾聴対話モデルの構築

松本奈々[1]　安藤一秋[1]
[1]香川大学 創造工学部
{s20t332, ando.kazuaki}@kagawa-u.ac.jp

概要

高齢者の増加とともに認知症患者も増加している．認知症患者の感情に寄り添い，意思を汲み取るためのコミュニケーション技法として「バリデーション」がある．しかし，介護業界は慢性的な人手不足であり，介護士がすべての患者に対して十分なコミュニケーション時間を確保することが難しい．本稿では，バリデーションのうち「オープンクエスチョン」に着目し，強化学習を用いて傾聴の多い応答を生成する対話モデルの構築法について検討する．評価実験では，ファインチューニングの epoch 数と強化学習の step 数の関係に着目し，応答文の多様性や傾聴性などに関して評価・考察する．

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[1]．これに伴い，認知症患者の数も今後増加することが予想されている[2]．認知症患者は，意思をうまく伝えられないため，不安やストレスを抱えやすいと考えられている．そこで，認知症患者の感情に寄り添うためのコミュニケーション法として「バリデーション」が注目されている．この方法は，6 つの基本態度（「傾聴する」や「共感する」など）と，14 の基本テクニック（「はい／いいえ」で答える質問ではなく，「いつ」，「どこで」といった自由な回答が期待できる質問を心がけることなど）を用いて認知症患者と丁寧にコミュニケーションすることで，患者が抱える悩みを軽減することを目的としている[3,4,5]．しかし，介護士がすべての患者に対して十分なコミュニケーション時間を確保することは，近年の介護士の人材不足[6]などの観点から困難である．

そこで，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．本稿では，バリデーションの基本テクニックにおける「オープンクエスチョン」に着目し，強化学習を用いて傾聴の多い応答を生成する対話モデルの構築法を検討する．特に，ファインチューニングの epoch 数と，強化学習の step 数が傾聴性のある応答生成に与える影響についても確認する．

2　関連研究

2.1　傾聴対話システム

石田ら[7]は，傾聴対話システムを構築するために，傾聴対話の聞き手応答として必要と考えられる語彙的応答，評価応答，繰り返し応答，自己開示，掘り下げ質問の 5 種類の各応答の生成方法と，各応答の言語的，文脈的妥当性から傾聴妥当性を考慮した応答選択手法を提案した．評価の結果，適切な語彙的応答の F 値は 0.85 が得られたが，評価応答と掘り下げ質問の F 値は，それぞれ 0.53 と 0.41 となり，改善の余地があるといえる．本研究では，先行研究において提示された 5 つの応答のうち，掘り下げ質問，評価応答，語彙的応答の 3 つに注目する．

2.2　強化学習

清水ら[8]は，特定のキャラクタらしさを持つ対話システムを構築するためには，対話形式のデータ収集にコストがかかるという課題に対し，強化学習を用いて特定のキャラクタらしさを持つ対話応答を生成する手法を提案した．本研究でも，傾聴の多い対話コーパスの作成にコストがかかるという同様な課題があるため，先行研究を参考に強化学習を用いて傾聴対話生成モデルの構築法を検討する．

3　傾聴対話モデルの構築

掘り下げ質問，評価応答，語彙的応答を考慮した傾聴対話モデルの構築手順を以下に示す．

1. 大規模言語モデルを対話コーパスでファインチューニングすることで，対話生成モデルを構築する．

2. 傾聴性のある応答か否かをラベリングしたデータセットを用いて，BERT で 2 値分類タスクとして学習することで報酬モデルを構築する．
3. 報酬モデルの出力を報酬として，Proximal Policy Optimization (PPO) で強化学習し，傾聴対話モデルを構築する．

3.1 対話生成モデルの構築

対話生成モデルは，東北大学が公開している日本語日常対話コーパス[9]を用いて，rinna 社が公開している GPT2-medium[i]をファインチューニングすることで構築する．特に本稿では，20 epoch まで学習し，途中保存した 5，10，15 epoch の 3 つのモデルを用いて性能を評価することで，ファインチューニングの epoch 数が強化学習に与える影響についても確認する. 以下, それぞれ 3 つのモデルを FT5, FT10, FT15 モデルと呼ぶ.

3.2 報酬モデルの構築

3.2.1 データセットの構築

報酬モデルを作成するために利用するデータセットの作成法について述べる．まず，発話テンプレートとして，日本語日常対話コーパスから人手で 320 発話を抽出し, 対話生成モデルで各発話につき 50 応答を生成した．そして，応答が重複した生成を除いてデータを構築した. データの内訳は, それぞれ FT5 モデルは 15,287 件，FT10 モデルは 13,645 件，FT15 モデルは 12,194 件である．

先行研究[7]を参考に，掘り下げ質問，評価応答，語彙的応答について，表 1 に示す表現に対して報酬を与える．

掘り下げ質問は文末での一致，評価応答と語彙的応答は文頭からの一致で機械的判定する．ファインチューニングの epoch 数の異なる 3 モデルの生成結果に対して，表 1 のいずれかの表現が一致すれば 1（報酬あり），一致しなければ 0（報酬なし）でラベル付けする．ラベル付けした結果，正例が FT5 モデルは 5,858 件，FT10 モデルは 5,364 件，FT15 モデルは 4,847 件となった．

最終的に，正例・負例の件数をそれぞれ FT5 モデルは 5,800 件，FT10 モデルは 5,300 件，FT15 モデルは 4,800 件からなるデータセットを構築した．

表 1：報酬を与える表現の例

傾聴の種類	報酬を与える表現
掘り下げ質問	？（文末のみ）
評価応答	「いいですね」，「良いですね」，「素晴らしいですね」，「頑張っていますね」など
語彙的応答	「そうですね」，「なるほど」，「そうなのですね」など

3.2.2 報酬モデル

報酬モデルは，構築したデータセットを用いて，BERT[ii]で 2 値分類タスクとして学習することで構築する. 報酬モデルを構築した結果, それぞれの FT モデルにおける分類判定の F 値は, FT5 モデルで 1.00, FT10 モデルで 0.99，FT15 モデルで 0.99 となった．

3.3 PPO による傾聴対話モデルの構築

傾聴対話モデルを構築するための強化学習には，Trl[iii]の PPO を用いる．PPO のパラメータとしては，バッチサイズは 16，エポック数は 4，学習 step 数は 200，学習率は 1e-05，seed は 42 として学習する．本稿では，FT5，FT10，FT15 モデルについて，step 数が 80 と 200 で学習した傾聴対話モデル（RL80step, RL200step）を評価用に用いる．

4 評価実験

機械的評価と人手評価により，傾聴対話モデルの性能を評価する．特に，ファインチューニングの epoch 数と，強化学習の step 数が傾聴性のある応答生成に与える影響についても確認する．

4.1 機械的評価

機械的評価には，3.2.1 項で述べた発話テンプレート（320 発話）を用いて，3.3 節で構築した傾聴対話モデルで 1 発話につき 1 応答を生成して評価する．

表 2 に評価用データの例を示す．生成される応答は毎回変化するため，3 回試行して，応答の多様性と傾聴性を評価する．なお，実験結果は 3 回の平均値を用いる．

4.1.1 Distinct-N に基づく評価

Distinct-N[10]は，n-gram ベースでテキストの多様性を測る指標である．そこで，本実験でも，強化学

[i] https://huggingface.co/rinna/japanese-gpt2-medium
[ii] https://huggingface.co/cl-tohoku/bert-base-japanese-v3
[iii] https://huggingface.co/docs/trl

習後の応答の多様性を測るために評価指標として用いる．形態素解析器（Mecab+NEologd）で生成された各応答を単語分割し，N=1, 2 について Distinct-N の値を算出して評価する．

評価結果を表3に示す. なお, 表中のFT5+RL80stepは，5 epoch のファインチューニング後，step 数 80 で強化学習したモデルを示す．強化学習前後を比較すると，FT5，FT10，FT15 モデルすべてにおいて，Distinct-1, 2 の値が共に小さくなっていることから，強化学習により多様性が減っていることがわかる．また，強化学習前は，Distinct-1 は FT10 モデル，Distinct-2 は FT5 モデルが高かったが，強化学習後は，Distinct-1, 2 ともに FT10 が高くなっていることがわかる．

表 2：機械的評価用データの例

発話	生成された応答例
ニュースやドラマをよく見ています。	どのような番組を見ているのですか?
今日は、とてもよく歩きました。	素晴らしいですね。
少しおなかがすいてきました。おやつを食べませんか？	そうですね。そうしましょう。

表 3：Distinct-N の結果

モデル	Distinct-1	Distinct2
FT5	0.2046	**0.5388**
FT5+RL80step	0.1734	0.4271
FT5+RL200step	0.1379	0.3276
FT10	**0.2053**	0.5326
FT10+RL80step	**0.1842**	**0.4459**
FT10+RL200step	**0.1688**	**0.4193**
FT15	0.1987	0.5279
FT15+RL80step	0.1767	0.4412
FT15+RL200step	0.1562	0.3734

4.1.2 傾聴個数の変化に基づく評価

強化学習後に傾聴性のある応答が増えているかどうかを測るために，強化学習前後の各モデルについて傾聴性のある応答の生成個数（傾聴個数）の変化を調べる．掘り下げ質問は文末での一致，評価応答と語彙的応答は文頭での一致でカウントする．ただし，評価応答&掘り下げ質問や，語彙的応答&掘り下げ質問のような混合した応答が生成された場合は，掘り下げ質問を優先してカウントする．

評価結果を表 4 に示す．3 つのファインチューニングモデルすべてにおいて，強化学習したモデルの傾聴個数の方が増えていることがわかる．特に，ファインチューニングの epoch 数が小さいモデルの傾聴個数の方が増えている．次に，強化学習の step 数に着目すると，step 数が大きいほど傾聴個数が増える結果が得られた．

表 4：強化学習前後における傾聴個数の変化

モデル	掘り下げ質問	評価応答	語彙的応答	合計
FT5	94	12	19	125
FT5+RL80step	177	39	42	258
FT5+RL200step	223	**43**	23	**289**
FT10	100	20	16	136
FT10+RL80step	**227**	7	9	243
FT10+RL200step	187	26	44	257
FT15	91	18	27	136
FT15+RL80step	164	23	**49**	236
FT15+RL200step	194	32	38	264

4.1.3 機械的評価全体の考察

ファインチューニングの epoch 数が小さいモデルでは，強化学習後の傾聴個数が増加したことから，強化学習の影響を受けやすい可能性が高く，短文かつ似たような応答生成が繰り返され，多様性が減る結果になったと考えられる．

4.2 人手評価

5 人の評価者に対して，以下で述べる 2 つの評価実験を依頼して, 傾聴対話モデルの性能を評価する．なお，評価結果は 5 人の平均値を用いる．

4.2.1 傾聴性満足度と日本語正確性の評価

事前に設定した発話に対して生成された応答の傾聴性満足度（応答文が傾聴してくれる，傾聴してくれる内容に満足できる), 日本語正確性を評価する．表 5 に評価データの例を示す．

まず，JPersonaChat[11]から人手で 10 発話を選定し, 各発話に対して生成された 3 つの応答に対して，傾聴性満足度と日本語正確性について 5 段階（1:全くそう思わない～5:とてもそう思う）で，5 人の評価

者に評価してもらう．ただし，傾聴性満足度について，傾聴が生成されないものは評価値を 0 とする．

評価結果を表 6 に示す．すべてのファインチューニングモデルにおいて，強化学習後のモデルの方が傾聴性満足度と日本語正確性に対する評価値が高くなっていることがわかる．また，強化学習後のモデルは，step 数が大きいほど傾聴性満足度が高くなる結果となった．具体的な生成結果と判定結果を確認すると，評価応答&掘り下げ質問，語彙的応答&掘り下げ質問は，傾聴性満足度が 4 以上で評価されており，混合した応答は，より評価値が高くなる傾向が確認できた．

表 5：発話に対して生成された応答の例

発話	生成
最近キャンプに興味あるんですよねー。行ってみようかなと！	どこか行きたいキャンプ場はありますか?
	いいですね。道具はそろっていますか?
	良いですね。場所はどこですか?

表 6：傾聴性満足度と日本語正確性の評価結果

モデル	傾聴性満足度	日本語正確性
FT5	2.140	4.233
FT5+RL80step	3.873	4.513
FT5+RL200step	**4.560**	4.700
FT10	3.013	4.433
FT10+RL80step	3.933	4.780
FT10+RL200step	4.333	4.760
FT15	2.613	4.640
FT15+RL80step	4.100	4.766
FT15+RL200step	4.113	**4.780**

4.2.2 傾聴性，継続性，満足度の評価

評価者に自由に対話してもらい，発話に対する 1 応答レベルの評価と対話全体を通じた評価を実施する．1 応答レベルの評価では，傾聴性（前項の傾聴性満足度と同様）と継続性（対話を継続しやすかった）を，また，対話全体を通じた評価では，満足度（この対話システムと話していて良かった）について 5 段階で評価してもらう．ただし，傾聴性については，傾聴がなくてもよい場合，評価値を 0 とする．ここで，対話開始の発話は 5 つの話題（趣味，旅行，スポーツ・健康，食事，週末していること）から，事前に設定した発話を選択してもらう．また，対話の際は自己開示が多い発話を心がけてもらうように評価者に指示を与えた．

評価結果を表 7 に示す．傾聴性と継続性，さらに FT5 モデル以外の対話全体評価は，強化学習後のモデルの方がより高い値を得ていることを確認した．しかし，継続性に関しては，強化学習前後での差が小さいことを確認した．

表 7：傾聴性，継続性，満足度の評価結果

モデル	傾聴性	継続性	対話全体
FT5	2.840	3.200	2.800
FT5+RL200step	4.000	3.340	2.800
FT10	2.500	3.060	2.600
FT10+RL200step	4.100	**3.640**	**3.600**
FT15	2.500	3.200	2.400
FT15+RL200step	**4.200**	3.520	3.200

4.2.3 人手評価全体の考察

強化学習することで，傾聴性満足度は向上したが，対話継続性などについては評価値に大きな差がでなかった．分析の結果，掘り下げ質問が連続した場合や，語彙的応答のみの場合に評価が低いことから，対話を継続したいと思われなかったと考えられる．今後，この結果を活かした対話制御を検討する．

5 おわりに

本稿では，強化学習を用いて傾聴の多い応答を生成する対話モデルの構築法を検討し，各モデルの性能について評価した．評価の結果，強化学習後には傾聴的な応答が増加することを確認できた．また，ファインチューニングの epoch 数が小さいモデルでは，強化学習の影響を受けやすい可能性が高く，強化学習の step 数が大きいほど人手評価の傾聴性満足度が高くなった．今回，傾聴が多い対話応答生成モデルを構築できたが，発話内容によっては傾聴的な応答が必要ない場合もあり，対話制御における課題も確認できた．

今後の課題として，報酬モデル作成時のテンプレートの改善，バリデーションをより意識した報酬表現の検討，報酬モデル構築時のテンプレート数の改善，対話制御などについて検討する．

参考文献

1. 統計からみた我が国の高齢者ー「敬老の日」にちなんでー.（引用日：2023 年 11 月 28 日.）
https://www.stat.go.jp/data/topics/topi1380.html.

2. 平成 29 年版高齢社会白書.（引用日：2023 年 11 月 28 日.）
https://www8.cao.go.jp/kourei/whitepaper/w-2017/html/gaiyou/s1_2_3.html

3.都村尚子，三田村知子，橋野建史，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.木村裕美，古賀佳代子，久木原博子，西尾美登里，行動・心理症状を表出するアルツハイマー認知症高齢者に対する身体的・情緒的介入によるストレスへの影響，日本農村医学会雑誌，71 巻，1 号，12~21 項，2022.

5. 認知症ケアのコミュニケーション方法「バリデーション」とは｜認知症のコラム.（引用日：2023 年 11 月 28 日.）
https://www.sagasix.jp/column/dementia/validation/

6. 第 8 期介護保険事業計画に基づく介護職員の必要数について.（引用日：2023 年 11 月 28 日.）
https://www.mhlw.go.jp/content/12004000/000804129.pdf

7.石田真也，井上昂治，中村静，高橋克也，河原達也，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.

8.清水健吾，上垣貴嗣，菊池英明，強化学習を用いたキャラクタらしさを持つ雑談応答の生成，言語処理学会第 28 回年次大会発表論文集，pp.1770-1774，2022.

9. 赤間怜奈，磯部順子，鈴木潤，乾健太郎，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.

10. Jiwei Li, Michel Galley, Chris Brockett, Jianfeng Gao,and Bill Dolan. A diversity-promoting objective function for neural conversation models. arXiv preprint arXiv:1510.03055, 2015.

11. Hiroaki Sugiyama and Masahiro Mizukami and Tsunehiro Arimoto and Hiromi Narimatsu and Yuya Chiba and Hideharu Nakajima and Toyomi Meguro，Empirical Analysis of Training Strategies of Transformer-based Japanese Chit-chat Systems，arXiv2109.05217，2021.