

日常生活行動ナビゲーションのための強化学習に基づく対話生成

○横田和亮（大阪工業大学） 大井翔（大阪工業大学） 佐野睦夫（大阪工業大学）

1 はじめに

近年、認知症患者や高次脳機能障がい者の数は増加の傾向にある。国内の認知症患者は厚生労働省の2015年1月の発表によると、2012年時点で約462万人とされており、65歳以上の高齢者の約7人に1人とされている。また、2025年には認知症患者数は約700万人前後に達し、65歳の5人に1人の割合になると発表されている[1]。高次脳機能障がい者数は東京都の2008年のデータによると全国で約50万人以上と推測されている[2]。高次脳機能障がい者は、脳血管障害や事故により、脳に損傷を負うことで、記憶障害、注意障害、遂行機能障害、社会的行動障害といった症状が表れ、これらの症状によって日常生活や社会生活に制約がある状態の患者のことを指す[3]。

このような認知症患者や高次脳機能障害者（以後、認知障がい者と呼ぶ）が増えるにつれて、リハビリ担当者や支える家族の負担が増大してきている。我々は、在宅での生活復帰に向けて、安心感を与える生活行動のナビゲーションと同時に、行動変容を促す振り返りを行う認知リハビリテーション支援システムの確立に取り組んできている[4][5]。このような取組の中で、本研究は日常生活行動を円滑にナビゲーションする会話生成方式に着目し、認知リハビリテーションの質的向上を図る。これまでの研究では、ユーザの体験を共有し、事実の確認・評価に対する理由質問・自己開示・共感などのコミュニケーション行動を起こすことが、ユーザの会話を促進させることがわかってきている[6]。しかしながら、ユーザの内面状態を考慮し、会話履歴から円滑なコミュニケーションの生成を学習することはできていない。阿部ら[7]は、内面状態を考慮し、子供と遊ぶロボットの行動決定モデルを提案しているが、認知障害を内包した日常生活行動の中で、ユーザの内面状態を考慮しながら、生活行動をナビゲーションするためのコミュニケーションを生成する研究は十分になされていない。

本研究では、日常生活行動として、認知障がい者が苦手意識を持つ片付け行動に着目する。片付け行動をナビゲーションする中で、ユーザ

との対話履歴からユーザに適した対話生成を行うことのできる手法の提案を行う。

2 シナリオに基づく対話生成手法

シナリオを予め設定して、それに基づき対話を進める方式である。ここで、シナリオは、ユーザに対して共通に設定し、対話生成は、シナリオに関連付けられた質問群からランダムに選択される。対話システムはこの方式に基づきデザインされることが多い。我々も、この方式に基づき、高齢者向けの健康チェックシナリオを、コミュニケーションロボット（ヴィストン株式会社 SOTA）に実装し、基礎実験を行った。高齢者向け体験デモ風景を図1に示す。



図1 デモ風景

ユーザに適合し、満足感が得られる場面もあったが、明らかにユーザに適していない質問を行う場合が多く観測された。ユーザの内面状態を考慮した対話生成方式が必要とされる。3章では、この基礎実験結果を踏まえて、ユーザの内面状態を反映した強化学習に基づく会話生成法を提案する。

3 強化学習に基づく対話生成

3.1 強化学習方式

強化学習(Reinforcement learning)とは未知の環境に対して試行錯誤を行い、環境への行動の最適化を目指す考え方である[8]。強化学習は行動する主体であるエージェント(agent)、エージェントの行動にとって働きかけられる対象である環境(environment)、エージェントが環境に働きかけを行動(action)、エージェントの行動で環境から

与えられる行動の良さを表す報酬(reward)で構成されている。これらの関係を図2に示す。

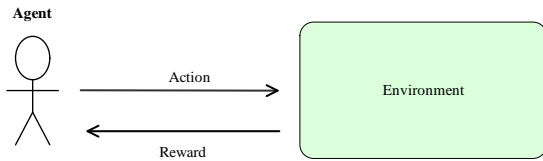


図2 強化学習関係図

3.2 提案する対話方式

本提案手法では、ユーザとコミュニケーションシステムの対話のやり取りは PC 上のコマンドプロンプトで表示されるテキストで行うものとする。ユーザは PC 上に表示されたテキストを読み、テキストの内容の回答を考え、考えがまとまれば PC の Enter キーを押すことで対話関係を築いているものとする。

ユーザと行う対話内容は片づけ行動を行う際に考えることを事前に本研究室でアンケート調査を行った結果から作成した。また、ユーザには3つの質問内容と1つの提案内容で片づける物に対する思い出や考えを想起させ、コミュニケーションシステムが処理方法の提案を行うものとする。なお、3回の質問と1回の提案で1回の対話とした。表1に質問内容と提案内容を図3に対話例を示す。

表1 質問内容と提案内容

質問群 1 ♪	最後に使ったのはいつですか? ♪
	次に使ったのはいつですか? ♪
	どれくらいの頻度で使いますか? ♪
質問群 2 ♪	何故買ったのですか? ♪
	それはまだ使えますか? ♪
	誰かからもらったものですか? ♪
質問群 3 ♪	何故迷っているのですか? ♪
	それを買ったとき、高価なものでしたか? ♪
	思い出がある物ですか? ♪
提案群 ♪	捨てましょう。 ♪
	捨てた方がいいと思います ♪
	捨てるか捨てないか決めてください ♪

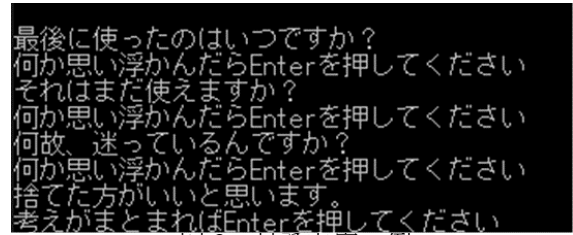


図3 対話内容一例

3.3 対話内容決定方式

本研究で提案する対話内容の決定方法には、探索アルゴリズムに1つである ϵ -greedy 法を用いて実装を行った。

ϵ -greedy 法は確率 $0 < \epsilon < 1$ ですべての選択枝からランダムに1つ選び、確率 $1 - \epsilon$ で時刻 t までの各選択枝の試行結果から得られた報酬の平均が最大な選択枝を選択する。

本研究では、コミュニケーションシステムの各状態の時刻 t における質問パターンの Q 値を ϵ -greedy 法で探索を行う。対話内容の決定までの流れを図4のフローチャートに示す。

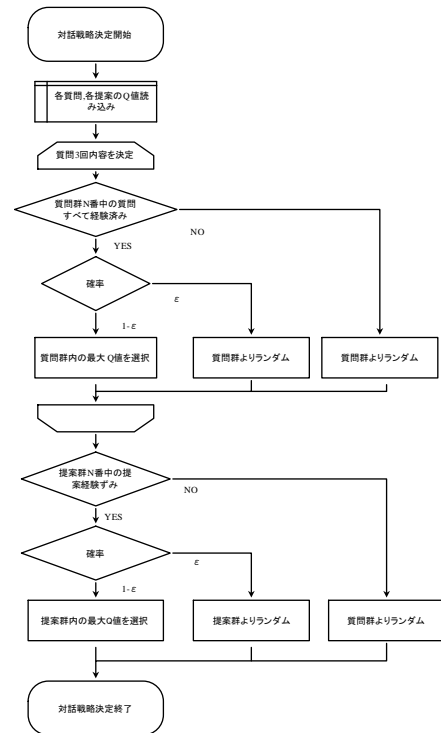


図4 対話内容決定フローチャート

3.4 対話内容の評価方式

コミュニケーションシステムがユーザに適した対話内容を行うようにするためには、行った対話内容がユーザに適していたか評価する必要がある。そこで本研究では、モンテカルロ法で対話内容の評価を行った。

モンテカルロ法は強化学習の Q 値を更新するアルゴリズムとして知られている。モンテカル

口法の特徴として、Q 値を試行途中で更新を行えないというデメリットがあるが、Q 値の更新には、実際に行動して得た結果を直接反映し、学習初期段階から学習が確かな方向性で学習することができるメリットがある。また、「試行途中で報酬を得ることができない」、もしくは「報酬の設定がしづらい」といった学習内容には対応しやすいといったメリットがある。

本研究では、1 対話終了時にユーザに対話内容の評価を行ってもらおう。例えば、「質問 A, B, C, 提案 1」の内容で、片づけ行動がスムーズに完了することができたならば、ユーザにとって適した内容での対話であったと判断した。また、コミュニケーションシステムは PC のキーボード入力以外の入力装置を備えていないため、1 対話終了後にユーザに対話内容の評価を入力してもらう必要がある。もし、対話内容が適しているなら 1 キー、そうでないなら 0 キーを入力といった方法で対話内容の評価を行う。

対話内容の評価を更新には、学習率 $\alpha=0.5$ と設定し、reward をユーザからの評価値とした。reward はユーザに適しているなら reward=1、適していないなら reward=0 とした。また、1 対話の評価値から行った各質問内容を評価するために割引報酬和を用いて、不確実性を報酬で割引く形で報酬和に反映した。なお、報酬和の割引率は $\gamma=0.8$ の固定値に設定した。対話内容の評価の更新式を式 (1)、割引報酬和の更新式を式 (2) に示す。

$$Q(s_t, a_t) = Q(s_t, a_t) + \alpha(\text{reward} + \text{totalreward} - Q(s_t, a_t)) \quad (1)$$

$$\begin{cases} \text{totalreward} = 0 & n = 0 \\ \text{totalreward} = \gamma(\text{totalreward} + \text{reward}) & n > 0 \end{cases} \quad (2)$$

4 実験概要

本実験では、実験協力者 1 人つき、500 回対話をおこなってもらいながら片づけ行動を行ってもらった。片づける物は実験協力者の身の回りにあるものとした。実験協力者には、500 回の対話終了後に表 2 の項目を記載したアンケートを実施した。また、表 3 の各質問、提案の各群から最も実験協力者に適した内容で片づけ行動ができたかをアンケート調査を行い、実験協力者の学習データとの比較を行った。

表 2 アンケート項目

質問番号	質問内容	評価方法
A1	対話を繰り返し行うことで質問内容に規則性を感じ取れたか	5 段階評価
A2	対話を繰り返し行うことで片づけがしやすい質問内容が多くなってきたと感じた	5 段階評価
A3	対話を繰り返し行うことで自分のことを理解していると感じた	5 段階評価
A4	対話を繰り返し行うことで自分のことを理解していると感じた	5 段階評価
A5	学習データ収集を行い、何か感じたことがありますか	自由記述

表 3 対話内容最適調査アンケート

質問群番号	質問番号	質問内容
B1	i	最後に使ったのはいつですか？
	ii	次に使ったのはいつですか？
	iii	どれくらいの頻度で使っていますか？
B2	i	何故、買ったのですか？
	ii	それはまだ使えますか？
	iii	誰かからもらったものですか？
B3	i	何故、迷っているのですか？
	ii	それを買ったとき、高価なものでしたか？
	iii	思い出があるものですか？
B4	i	捨てましょう
	ii	捨てた方がいいと思います
	iii	捨てるか捨てないか決めてください

5 実験結果

本実験で 500 回の対話終了後に実施したアンケート結果を表 4 に示す。

表 4 アンケート結果

質問番号	5 段階評価分布					合計人数	荷重平均
	1	2	3	4	5		
A1	1	3	0	4	1	9	3.1
A2	0	6	0	3	0	9	2.7
A3	2	3	2	2	0	9	2.4
A4	3	3	2	0	1	9	2.2

また、各実験協力者が最も片づけ行動を行いやすかったと答えた質問、提案内容と学習データと比較を行い、対話学習 100 回毎ごとの一致率を表 5 に示す。

表 5 実験協力者と学習データ一致率

学習回数	100	200	300	400	500
一致率(%)	41.67	33.33	58.33	50	50

6 考察

今回、対話を繰り返し行うことでコミュニケーションシステムが自分のことを理解していると感じたと答えた実験協力者は半数を超えた。しかし、対話を繰り返し行うことで対話に慣れ

を感じた実験協力者が多数いたことが判明し、あらかじめ用意していた質問、提案内容の数が少なかったためだと考えられる。

また、実験協力者の最も片づけ行動に適した質問内容と学習データの比較では、コミュニケーションシステムが実験協力者に適した対話方法を行えている場合とまったく行えていない場合があった。一致率が低かった要因として、ランダム要素が多く、実験協力者が最も良いと判断した対話パターンの参照が期待していたよりも低かったのではないかと考えられる。

7 おわりに

本研究では、日常生活行動の一例として、片づけ行動における対話学習について行った。現在、今回の研究結果を基に学習率向上を図れるように検討している。また、今後は高次脳機能障がい者の生活支援のナビゲーション部に組み込んでいく予定である。

本研究の一部は、JSPS KAKENHI Grant Number JP 15K00368 の支援を受けた。

参 考 文 献

- [1] 厚生労働省, “認知症とはどういうものか?,” <http://www.mhlw.go.jp/topics/kaigo/dementia/a01.html>.
- [2] 東京都高次脳機能障害者実態調査検討委員会 “東京都高次脳機能障がい者実態調査書,” 2008.
- [3] 国立障害者リハビリテーションセンター, “高次脳機能障を理解する,” http://www.rehab.go.jp/brain_fukyu/rikai/shoujou/.
- [4] 大井翔, 佐野睦夫, 渋谷咲月, 水野翔太, 大出道子, 中山佳代, “高次脳機能障害者の自立に向けた調理行動振り返り支援システムに基づく認知リハビリテーション,” 認知リハビリテーション Vol.20, No.1 , pp.51-61 (2015)
- [5] 佐野睦夫, 中川葵, 小谷凌和, 大井翔, 小山智美, 西野朋子, ” 高次脳機能障害者に対する掃除行動振り返り支援システムに基づく認知リハビリテーション, “認知リハビリテーション Vol. 22, No. 1, pp. 31-40 (2017. 10)
- [6] 佐野睦夫, 小谷凌和, 中川葵, 足立奈生, 森本亜美, 吉田祥子, 吉永千紘, “カウンセリングロボットによる生活行動振り返り対話戦略と認知リハビリテーション支援システム,” 情報処理学会研究報告ヒューマンコンピュータインタラクション (HCI) , Vol.2017-HCI-172, No.18, pp.1-7 (2017)
- [7] 阿部香澄, 岩崎安希子, 中村友昭, 長井隆行, 横山絢美, 下斗米貴之, 岡田浩之, 大森隆司, “子供と遊ぶロボット: 心的状態の推定に基づいた行動決定モデルの適用,” 日本ロボット学会誌 Vol.31 No.3,

pp.263-274 (2013)

[8] 牧野貴樹他, “これからの強化学習,” 森北出版(2016)