

ユーザ応答の言語・非言語情報に応じた 語りかけエージェントによる対話制御

Dialogue Management in Listener Agents according to Verbal and Nonverbal Information in User's Response

塚田 一史^{*1} 山内 崇資^{*2} 林 佑樹^{*1} 中野 有紀子^{*1}
Kazuhumi TSUKADA Takashi YAMAUCHI Yuki HAYASHI Yukiko NAKANO

^{*1} 成蹊大学理工学部
Faculty of Science and Technology, Seikei University

^{*2} 成蹊大学大学院理工学研究科
Graduate School of Science and Technology, Seikei University

With the goal of assessing health and mental condition of people with Dementia through interaction with a conversational agent, we have developed a listener agent that asks a sequence of questions to the user and records interaction log using a speech recognition system. However, to collect richer utterance data from the user, it is necessary to make the user more engaged in the conversation with the agent. To improve the human-agent interaction, this study aims at expanding the conversation by asking follow-up questions. We define sets of keywords by analyzing the transcription of user's utterances and expand the keywords using the EDR dictionary. If a keyword is found in the recognition result, the keyword is applied to a question generation template to produce a follow-up question.

1. はじめに

近年、高齢者の増加と共に認知症患者も急増している。しかし、介護者の数は不足しており、認知症患者を支援する環境は十分に整っていない。また、認知症患者を家族で介護する場合、認知症患者は記憶障害のため、家族に同じ質問を何度も聞くことが多々あり、家族内でコミュニケーションの問題が生じたり、介護者の大きな負担となる場合もある。このような問題を解決するために、患者の話し相手となりながら、患者の健康状態、精神状態を把握し、介護者に伝える見守りサービスが求められている。

[Pollack 2005]は認知症高齢者を支援する情報技術として、日常生活が安全に遂行されていることを確認し、問題があれば介護者に通知するシステムの重要性を述べている。[Maiden 2013]は、携帯端末用アプリケーションにより、介護記録を書き込んだり、介護者間で情報を共有したり、介護履歴を参照するなど様々な介護者支援ができると述べている。このように、認知症高齢者の見守り、介護支援環境の充実において、情報技術に大きな期待が寄せられている。本研究では、特にエージェント技術を用いた認知症高齢者の支援を目指す。[安田 2013]は、ユーザ二人とエージェントの複数人会話において、エージェントは会話の楽しさを提供できる可能性があり、複数人会話の司会を務めることが求められるとしている。また、社会から孤立した高齢者を支援するためのシステムとして、ユーザが選択肢を選ぶことでエージェントとインタラクションを行うものがある[Sidner 2013]。本研究では、会話エージェント技術を応用し、認知症高齢者との会話を通して、健康状態や精神状態を把握するシステムの実現を目指す。

2. 語りかけエージェントシステム

[比企野 2011]は認知症患者に日常の健康や食事等について質問をすることにより、患者の話の聞き役となる語りかけエージェントを作成し、認知症高齢者の方に実際にシステムと会話を行ってもらう、システムの有効性を確認した。このシステムでは、ユーザの発話音声を検出し、そのピッチとポーズ長の韻律的特徴に基づき、相槌や頷きを返す。また、ユーザの発話終了後、一定以上の時間が経過したら、ユーザが話し終えたと判断し、次の質問を行う。[酒井 2011]では、これに音声認識機能を追加し、ユーザの発話をログファイルに出力する機能を追加した。

3. システムの改善

先行研究のシステム[比企野 2011]では、予め設定したシナリオに沿って固定的な順番で質問を行うため、利用者の反応に応じて話題を展開する等の適応的な対話制御はできなかった。そのため、患者の発話も比較的短く、単調なものである場合が多く、患者の状態をより正確に把握するには、より豊富なデータから判断すべきであると考えた。

そこで本研究では、認知症患者からより多くの情報を得られるように、ユーザの反応に応じた話題の展開、切り替えを行う質問選択機構を実装することを目的とする。またこれに加え、アニメーションエンジンの改良と、相槌、頷きタイミングの変更を行う。

3.1 MMDAgentによる実装

先行研究[比企野 2011]では、アニメーション作成ソフトウェア Poser にモーションキャプチャデータを適用してエージェントアニメーションを作成したのち、これを連番画像に変換し、Adobe Flash で制御することでアニメーションを表現していた。しかしこの方法では、新しいモーションの追加やモデルの変更は簡単ではなかった。エージェントのジェスチャーや表情といった非言語情報をより豊かにすることにより、ユーザからの応答を促進す

るためには、アニメーション作成が容易にできる環境にする必要がある。そこで、エージェントのアニメーションを MMDAgent[李 2011]を用いて生成するようにした。

MMDAgent は名古屋工業大学で開発されたオープンソースの音声インタラクションシステム構築ツールキットである。MMDAgent では MMD 用に開発されてきたアニメーション技術を用いることで、簡単にエージェントのアニメーションを作成することができる。我々は、MMDAgent にソケット通信プラグインを追加することで、外部入力によるアニメーション制御を行う機構を利用した。どのような入力に対してどのような動作を行うかは、あらかじめ MMDAgent のシナリオファイルに記述しておいた。図 1 に MMDAgent を用いて実装したシステムとのインタラクションの様子を示す。



図 1 対話の様子

3.2 相槌タイミングの変更

過去の実験において、エージェントの相槌のタイミングが悪く、ユーザ発話にオーバーラップしている場面が複数回見られた。オーバーラップが起きるとユーザは発話を止めてしまい、一度止めた話を再度話始めることはなかった。また、オーバーラップではないが、発話途中のポーズでエージェントの相槌が入ると、やはり発話を中断してしまうことがあった。そこで、音声を伴わない傾きについては、以前と同じようにピッチの変化量に基づきタイミングを決定し、音声発話を伴う相槌については、発話終了時のみ生成する行動とし、3 秒以上無音区間が続いた場合に、発話が終了したとみなして相槌を返すように変更した。また、発話内容も発話終了時の相槌として適切なものに変更する必要があると考え、相槌の表現として「そうなんですか」、「なるほど」、「はい」という 3 種類を用いることにした。

4. 質問選択機構

4.1 ユーザ発話の分析

ユーザの発話に応じた応答を返すためには、システムがユーザの発話音声を認識する必要がある。しかし、発音が不明瞭であると、十分な音声認識結果が得られない場合もある。そこで、[酒井 2011]のシステムのログファイルに保存された音声認識結果と人手による書きおこしの比較を行った。その一例を表 1 に示す。多くの誤認識が含まれるものの、食べ物についての発話

表 1 音声認識結果と書き起こし例

ユーザ	音声認識結果	書き起こし
A	Ra 野菜ものとか果物は	あー野菜ものとか果物は
B	魚魚ルーアの魚グアム嫌いなものは全部刺身煮物から	魚. 魚類は, 魚類は嫌いなものはないですね. 全部, 刺身から煮物から
C	そうですね英語あまり好きではないです呉の観測が大好きです	そうですね, ご飯はあまり好き, 好きでもないです, 普通ですけどもパン

音声から、「野菜」や「刺身」といった、特定のキーワードを検出することが可能であることがわかった。

そこで、エージェントからの各質問に対するユーザの応答発話の書きおこしを分析し、どのようなキーワードが現れるのかを調査し、キーワードを選定した。しかし、書きおこしデータのみからキーワードを選定するだけでは、数が不十分であり、自動的にキーワードを追加する方法が必要となった。

4.2 辞書を利用したキーワードリストの作成

キーワードリストを拡張するために、EDR の日本語単語辞書¹を使用した。この辞書には、約 44 万個の数字や名詞、動詞の定義が収録されている。

まず、4.1 節で選定したキーワードについて辞書定義を調べ、多くの単語定義に出現する表現を抽出した。表 2 に例を示す。例えば、野菜や魚の定義には「食物」という単語が共通して含まれており、刺身や焼肉の定義には「料理」という単語が共通して含まれている。これらの単語を定義語と呼ぶ。次に、これらの定義語が単語定義に含まれている辞書項目をすべて抽出し、キーワードリストを拡張した。この方法により地名と食べ物についてのキーワードリストを作成し、さらにこれに加えて趣味のキーワードリストは、Web ページを参考に設定した。その結果、食べ物のキーワード 8330 個、地名のキーワード 3522 個、趣味のキーワード 366 個を選択した。

表 2 日本語単語辞書の例

名詞	意味
野菜	畑などに作って収穫し、副 食物 にする植物
魚	酒を飲むときに添えてつまむ 食物
刺身	刺身という、生魚をうすく切った日本 料理
焼肉	焼き肉という 料理

4.3 話題展開のための質問生成

エージェントの発話は全部で 32 個あり、そのうち質問は 16 個であるが、これらは「はい/いいえ」で答える closed-question と WH 型の open-question に分類できる。話題を展開するためには、open-question に対する応答をきっかけにするのが良いと考え、open-question に対する応答の音声認識結果中に 4.2 節で作成したキーワードが含まれている場合、これを話題展開のキーとした。次に、このキーワードを表 3 のテンプレートに適用することにより、話題展開のための質問を生成した。

表 3 質問テンプレート

キーワードリスト	テンプレート
地名	(キーワード)はどこなところですか
食べ物	(キーワード)はおいしいですね、よく食べるのですか
趣味	(キーワード)はどれくらいやっていますか

図 2 に本システムの処理の流れを示す。ユーザの発話音声は、音声認識部で Google 音声認識により処理され、音声から文字列に変換される。音声認識の結果は 10-best まで取得する。エージェントの質問は 1~32 までの ID で管理されており、システムは最後の質問 ID を現在の話題とみなし、保持している。ユ

¹ EDR : http://www2.nict.go.jp/out-promotion/techtransfer/EDR/J_index.html

ユーザ発話の音声認識結果が質問選択部に送られると、質問選択部では質問 ID を参照して最後に行った質問に対応するキーワードリストを選択する。そして、10 個の音声認識結果の文字列の中にキーワードリスト内のキーワードと一致する単語が含まれているかどうかを調べる。キーワードが複数個マッチした場合は、より上位の音声認識結果から得られたキーワードを優先する。音声認識結果にキーワードと一致する単語が含まれていた場合、音声認識結果は正しいと思われるので、音声認識とキーワードの比較を終了し、音声認識と一致したキーワードを表 3 の質問テンプレートに適用し、話題展開の質問を生成する。キーワードと一致する単語が含まれていない場合には、シナリオに従い、次の ID の質問に進む。

発話ID=22「好きな食べ物はなんですか？」

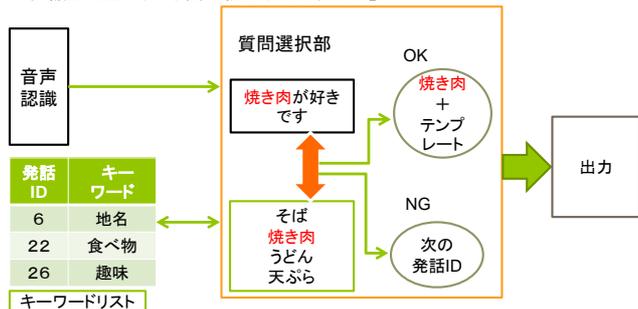


図 2 質問生成処理の流れ

5. 実行例

前節で提案した手法により、認知症高齢者の実際の発話認識結果から、どのような話題展開質問が生成されるかを、例を挙げて説明する。エージェントの質問「好きな食べ物はなんですか」に対して、表 4 の書き起こしに示すような応答が得られた。システムは質問 ID=22 から食べ物に関するキーワードリストを選択する。リスト内のキーワード「スパゲッティ、刺身、・・・焼肉、パン」が音声認識結果「you were here あざみ野 パン デイ」に含まれているか調べる。「パン」という単語が音声認識結果とキーワードリストの両方に含まれるので、キーワードが一致したと判断し、食べ物に関するテンプレート「～はおいしいですね、よく食べるのですか」と、取得したキーワード「パン」を組み合わせて「パンはおいしいですね、よく食べるのですか」という質問を生成する。

表 4 話題展開が可能な例

書き起こし	今は一、朝がパンで、えー、
認識結果	you were here あざみ野 パン デイ
抽出されたキーワード	パン
生成された質問	パンはおいしいですね、よく食べるのですか

表 5 話題展開ができない例

書き起こし	そうですね、ご飯はあまり好き、好きでもありません、普通ですけどもパン食が大好きです
認識結果	そうですね 英語 あまり好きではないです 呉の観測がだいすきです

一方、表 5 に示す例のように、同じエージェントの質問「好きな食べ物はなんですか」に対して、「そうですね、ご飯はあまり好き、好きでもありません、普通ですけどもパン食が大好きです」と

ユーザが答えているにもかかわらず、得られた音声認識結果が「そうですね 英語 あまり好きではないです 呉の観測がだいすきです」であった場合には、キーワードは含まれていないと判断されるため、テンプレートを用いた発話を行わず ID=23 の「私はお寿司が好きです」とシナリオに沿った次の発話を行う。

6. まとめ

本研究では、語りかけエージェントとの会話を通して、認知症患者の健康状態や精神状態を把握することを目指し、より豊富な発話情報を得られるような対話を実現するために、対話エージェントシステムの改良を行った。アニメーションの改善として、モーションが作成しやすい MMDAgent に変更した。エージェントの相槌がユーザ発話にオーバーラップしないよう、相槌のタイミングを再検討し、それに伴い、発話内容にも変更を加えた。さらに、シナリオに沿った発話だけでなく、ユーザの応答発話の音声認識結果から、話題を展開するための質問を生成するシステムを実装した。

今後は MMDAgent の利点を生かし、頷きのモーション以外にもお辞儀や、身振り手振りのジェスチャーを作成する予定である。また、笑顔の他に悲しみや困った顔など表情を豊かにすることも重要である。

話題展開の質問の生成については、評価実験を行い、提案手法の有効性を検証する予定である。また、現在は限定的な話題についてしか対応できず、質問の表現のバリエーションも不十分である。今後は、対応できる話題の種類を増やすとともに、質問生成のテンプレートを増やす必要がある。さらには、一つの質問に対して一度だけ問い返すのではなく、2回、3回と話題展開できる方法も検討すべきと考える。

参考文献

[比企野 2011] 比企野純太, 中野有紀子, 安田清:「会話エージェントを利用した認知症患者のためのコミュニケーション支援」, 第 73 回情報処理学会全国大会論文集, 4ZA-7, pp.195-196 (2011).

[Maiden 2013] Neil Maiden et al.: Computing Technologies for Reflective, Creative Care of People with Dementia, *Communications of the ACM*, Vol.56, No.11, pp.60-67 (2013).

[Pollack 2005] Pollack, M.: Intelligent technology for an aging population: The use of AI to assist elders with cognitive impairments, *AI Magazine*, Vol.26, No.2, pp.9-24 (2005).

[李 2011] 李晃伸, 大浦圭一郎, 徳田恵一:「魅力ある音声インタラクションシステムを構築するためのオープンソースツールキット MMDAgent」, 電子情報通信学会技術研究報告, 111(364), pp.159-164 (2011).

[酒井 2011] 酒井洋一, 野中裕子, 安田清, 中野有紀子:「認知症患者と語りかけエージェントとのインタラクション分析とシステム構築に向けた提案」, ヒューマンインタフェース学会研究報告集, Vol.13, No.8, pp.13-18 (2011).

[Sidner 2013] Candace L. Sidner, Timothy W. Bickmore, Charles Rich, Barbara Barry, Lazlo Ring, Morteza Behrooz, and Mohammad Shayganfar: Demonstration of an Always-On Companion for Isolated Older Adults, *Proc. of the SIGDIAL 2013 Conference*, pp.148-150 (2013).

[安田 2013] 安田清, 青江順一, 泓田正雄:「回想を促すアニメエージェント:認知症者 2 人とエージェントの複数人会話観察」, *Human-Agent Interaction Symposium 2013*, S-2, (2013).