

3-5 高齢者介護支援用マルチモーダル音声対話システム MICSUS

3-5 MICSUS: A Multimodal Dialogue System for Elderly Care

水野 淳太 浅尾 仁彦

MIZUNO Junta and ASAO Yoshihiko

社会の高齢化が急速に進む中で、高齢者のケア・介護分野での人手不足が深刻な問題となっており、介護従事者の作業負担軽減は急務である。MICSUS は、高齢者に対して健康状態や生活習慣などをヒアリングする介護モニタリングという作業の一部を代替し、介護従事者の負担を軽減するとともに、介護の質を向上することを目標としたマルチモーダル音声対話システムである。MICSUS の研究開発は、内閣府 SIP 第二期の支援により、国立研究開発法人情報通信研究機構 (NICT) データ駆動知能システム研究センター (DIRECT)、KDDI 株式会社、NEC ソリューションイノベータ株式会社、株式会社日本総合研究所が共同で進めている。本稿では、MICSUS の機能と実際に高齢者を対象として行った実証実験の結果について報告する。

In an increasingly aging society, the elderly care industry faces serious problems of labor shortage. To mitigate the workload of care workers and improve the quality of elderly care, we are developing MICSUS, a multimodal dialogue system that collects information on health status and lifestyle habits of elderly people who need care through interviews. MICSUS is being developed by the the Data-driven Intelligent System Research Center (DIRECT), NICT, as a collaborative work with KDDI Corporation, NEC Solution Innovators, Ltd., and the Japan Research Institute, Ltd. This paper presents the structure and functionalities of MICSUS as well as results of proof-of-concept experiments in which elderly people conversed daily with the system at their home.

1 はじめに

近年、日本においては人口減・財政難が続く中で、社会の高齢化が急速に進み、高齢者のケア・介護分野での人手不足が重大な社会課題の1つとなっている。そこで、NICT DIRECT は内閣府 SIP 第二期の支援により、KDDI 株式会社、NEC ソリューションイノベータ株式会社、株式会社日本総合研究所と共同でマルチモーダル音声対話システム MICSUS (Multimodal Interactive Care Support System) の研究開発を進めている。本研究開発の最終目的は、(1) 人手不足が社会問題となっている介護関係者(ケアマネジャー、施設の介護職員など)の業務の一部を MICSUS で代替することで業務負担の軽減や効率化を図ること、さらには(2) やはり社会問題として指摘されることが増えてきた高齢者の社会的孤立の抑制に貢献し、社会的孤立によって加速すると言われている高齢者の健康状態悪化を抑制して、最終的に将来の介護需要を低減することの2点である。

高齢者のケア・介護の分野では、今後高齢化の進展

に伴って需要が増大するものの、財政・人的資源がそれに追従できない状況が続くと予測されており、2025年には必要な人材である258万人に対して、地方都市まるまる1つの人口に相当する43万人という規模で人材不足が生じるとの予測がなされている(図1)。介護分野では日本だけでなく世界的に人手不足が懸念されており、AIの活用に期待が高まっている*1。

ケア・介護で必要な作業として、要支援・要介護の高齢者の自宅などをケアマネジャーという資格を持つ介護職が訪問し、健康状態や生活習慣などをヒアリングする介護モニタリングと呼ばれる作業がある。これによって得られた情報は介護プランの立案等で利用される。最低月1回実施される介護モニタリングは1名1回につき30分程度かかり、移動やレポート作成も合わせると、通常数十名を受け持つケアマネジャーの介護モニタリングに必要な時間は勤務時間の6割を占め

*1 日経新聞 2021.12.9 付「介護現場はDXの実験場」

<https://www.nikkei.com/article/DGKKZO78281610Y1A201C2EA1000>

3 社会知コミュニケーション技術

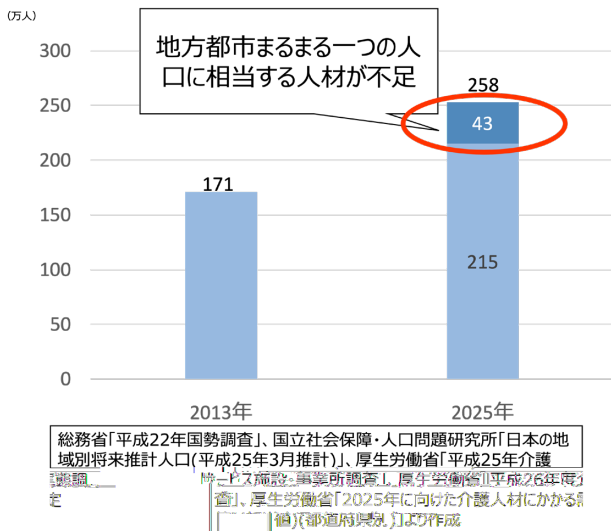


図1 介護人材の需要のギャップ(推計)

ると言われている。マルチモーダル音声対話システム MICSUS は、この介護モニタリングを一部代替する対話システムとして開発しているもので、株式会社日本総合研究所によって策定された「適切なケアマネジメント手法」*2 [1] に基づき、大型の対話端末(図2)が音声対話を介して高齢者の健康状態等のヒアリングを行う。また、端末が撮影する動画中の顔画像や音声から高齢者の感情等も認識し、より高齢者に寄り添った対話を行うために活用する。なお、大型の対話端末は高齢者の受容性を高めるために採用したが、通常のスマートフォン等のデバイス向けの移植作業も実施しており、それらのデバイスでも MICSUS の使用は可能である。

対話は、基本的に MICSUS が健康や生活習慣等に関する質問を高齢者に投げかけることで進む。高齢者の応答は最新の深層学習技術により解釈され、例えば、質問「毎日3食食べていますか?」に対して「最近、胃腸の調子が良くてね」といった遠回しな回答を行っても高精度に解釈される。また、ユーザの回答によっては質問の聞き直しや、回答の訂正などを行うこともできる。回答は音声だけでなく、^{うなず}頷きや首振りなどのジェスチャーでも可能である。ケアマネジャーは MICSUS の対話結果を「毎日3食食べている」といった簡潔な要約の形で閲覧することができ、介護プラン立案に要する時間を短縮できる。さらに、MICSUS は、飽きずに対話を継続できるように Web 等の情報を使った多様な雑談(例えば、高齢者の好きな食べ物の美味しい食べ方から、音楽鑑賞等の趣味の話まで)をする機能も備えている点が大きな特徴である。こうした種々の機能を活用した MICSUS による実際のヒアリングの様子は、実証実験時に撮影された動画*3 を参照されたい。



図2 実証実験に使用した MICSUS の犬型対話端末

実は、MICSUS の真の狙いは、介護モニタリングの一部代替を超えたところにある。MICSUS は人間とは異なり、いかなる時間、いかなる頻度でも対話に応じられるので、現在の月1回というヒアリング頻度にこだわる必要はない。場合によっては、日々健康状態が変化する高齢者と毎日対話し、月1回のヒアリングでは見落としがちな変化に関する情報等を取得することで、介護の質向上にも貢献できる。加えて、ケアマネジャーは多岐にわたる詳細な健康状態のチェックは MICSUS に任せ、人間が介在すべき重要な事案に関する相談に集中できる。また、高齢者の死亡率、要介護度の進行リスクを増大させるとして、近年、深刻な問題となっている社会的孤立 [2]、コミュニケーション不足の回避・抑制にも、Web 情報等を使った雑談の提供等で貢献できるのではないかと考えている。

本稿では、MICSUS の構成と、実証実験の結果について報告する。なお、実証実験で取得した発話は個人情報であるため、本稿に記載するユーザ発話は全て作例としている。

*2 ケアマネジャー等の介護支援専門員の知見を共通化・体系化することで利用者が必要とするケアマネジメントを一定以上の水準で提供できるようにするための手法。https://www.jri.co.jp/page.jsp?id=38679

*3 https://www.youtube.com/watch?v=bRgo31EolMg

2 MICSUS

2.1 MICSUS の構成

MICSUS の構成を図 3 に示す。ユーザは対話端末に対して、音声発話や首振りなどのジェスチャーを行うことで対話を進めることができる。音声と映像は対話端末のマイクとカメラによってそれぞれ取得され、リアルタイムに UI サーバに送られる。UI サーバ上では、感情認識エンジンが動作しており、映像及び音声からユーザの感情が推定される^{*4}。ユーザの音声は UI サーバから音声認識エンジンに送られ、認識されたテキストが対話制御モジュールに送られる。対話制御モジュールは、XML 形式で記述された対話シナリオに従い、ユーザの音声発話等をユーザ発話意味解釈モジュールで解釈した結果を使いつつ、対話を展開する。対話シナリオは 2.3 で詳述するとおり、対話シナリオオーサリングツールを用いて作成することができ、さらに対話シナリオ自動拡張モジュールによって自動的に拡張される。この拡張の必要性、内容等についても 2.3 で述べる。

MICSUS の大きな特徴の 1 つは、(1) 目的志向対話と (2) 雑談対話という 2 種類の対話をハイブリッドに実行できることである。(1) の目的志向対話は、対話シナリオに従って、ユーザに健康や生活習慣などの質問を行い、それに対するユーザ発話を解釈することで、ユーザの情報を取得する対話である。目的志向対話は、図 3 にある対話制御モジュールが制御を行い、ユーザ発話の意味解釈はユーザ発話意味解釈モジュールが深

層学習技術を用いて行う。(2) の雑談対話は、ユーザ発話の解釈結果に応じて展開される雑談である。雑談的応答は、Web 上の情報を元にした雑談応答生成モジュールである WEKDA によって生成されるものと、Web ニュースを元に雑談応答を生成する KACTUS によるものの 2 種類がある。KACTUS は対話シナリオで指定されたタイミングで雑談的応答を行うが、WEKDA による雑談的応答は、雑談のトピックとなり得る単語がユーザの発話に含まれているとユーザ発話意味解釈モジュールが判断すれば、目的志向対話の途中であってもほぼ任意のタイミングで開始され、また、高齢者が雑談的応答に興味を示さなかった等のタイミングで、健康や生活習慣を尋ねる目的志向対話に復帰する。その意味で 2 種類の対話の間の行き来はシームレスに行われることになる。

対話の結果得られる対話履歴、高齢者の動画像、回答の要約結果(介護モニタリングの報告書に相当)などはデータベースに登録され、図 4 のような対話結果確認・修正 Web アプリを使ってスマートフォン等で閲覧できる。ケアマネジャーなどの介護関係者はそれに基づいて介護プラン等を立案することができるほか、ユーザ発話意味解釈モジュールが高齢者の発話を誤って解釈したことを発見した場合には、簡単な操作で素早く修正することができる。また、今のところ実装はされていないが、介護事業者等が運用している介護関

*4 感情推定は NEC ソリューションイノベータ株式会社が開発した技術であるため、本稿では詳細は割愛する。

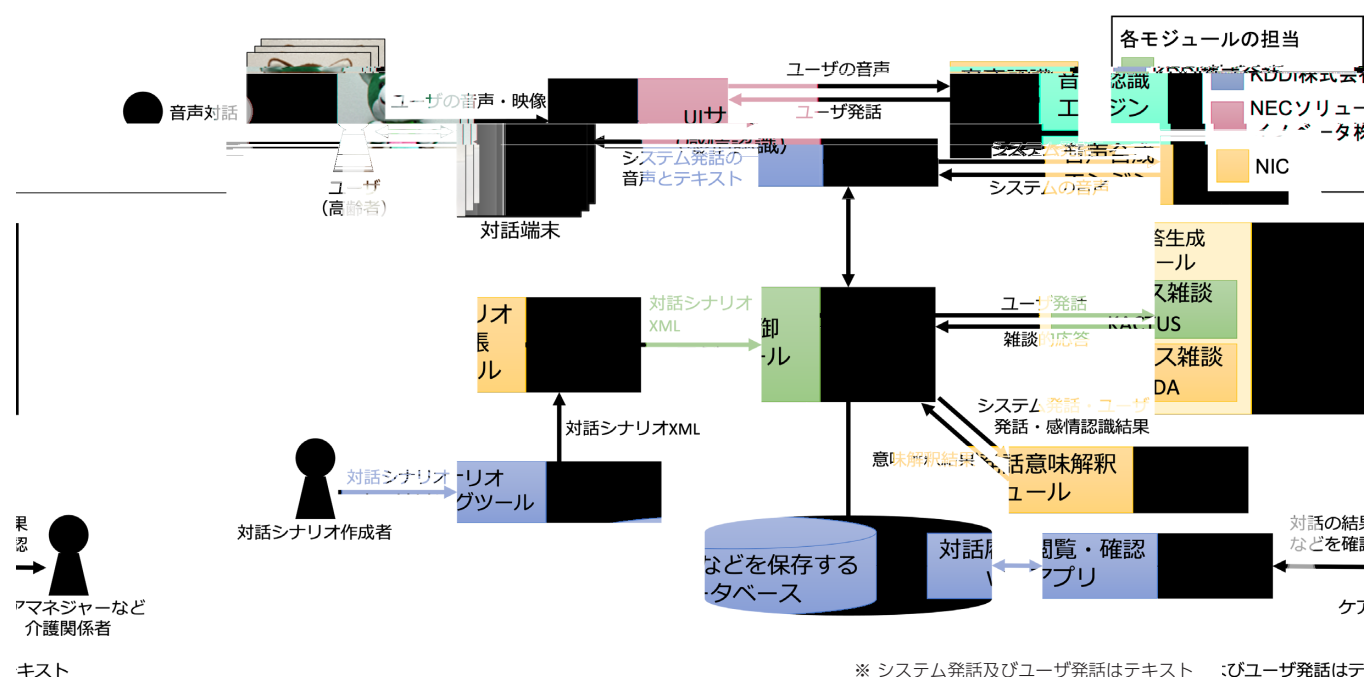


図 3 MICSUS の構成図



図4 対話結果確認・修正 Web アプリ (スマートフォン版の画面。顔写真を除き実際の動作画面のスクリーンショット)

連情報のデータベースに人手を介さず直接対話の要約結果を登録し、AIによる介護プランの立案など様々な用途で活用することも可能である。現在は、ケアマネジャーがヒアリングの結果を思い出しつつ、データベースに介護モニタリングの結果として登録する必要がある、煩雑な作業となっていたが、そうした作業負担の軽減にもつながるものと思われる。

2.2 ユーザ発話意味解釈モジュール

MICSUSの質問に対するユーザ発話の解釈は、ユーザ発話意味解釈モジュールによって行われる。このモジュールは、以下の5種類の解釈を実行し、対話はこれらの解釈結果に応じて展開される。1.と2.については質問の種類に応じて解釈を実行し、3.~5.については基本的に全てのユーザ発話に対して実行する。

1. Yes や No など回答可能な質問である Yes/No 質問 (例:「毎日3食食べていますか?」) に対する、ユーザ発話の分類 (表1)
2. 名詞や短い文で回答可能な質問である自由回答質

問 (例:「昨日の夕飯は何を食べましたか?」「水分をとるためにどのような工夫をしていますか?」) について、回答が適切になされたか否かの判定 (質問によっては高齢者が回答を拒否したり、回答を思いつかなかったりするケースがあり得る) と、ユーザ発話のうち回答を端的に表す部分 (回答キーワード) の抽出

3. 雑談を展開すべきかの判定
4. 高齢者が過去に行った回答の訂正を意図した発話をしたか否かの判定
5. 後で尋ねる予定だった質問の回答を、ユーザが先に発話したか否かの判定

これらの解釈は基本的に入力テキストの分類タスクとして定式化でき、3で説明するように深層学習技術、より具体的にはBERT [4] を使って実装されている。

上記1.のYes/No質問についてより詳細に述べると、ユーザ発話意味解釈モジュールは表1に示す6種類のラベルを出力する。分類結果が「その他」以外の場合、システムは「お薬をきちんと飲んでいないんですね。それは心配ですね。」のような解釈結果を確認する応答を発話し、次の質問に進む。ただし、上記3.の判定結果によっては雑談を行うこともある。分類結果が「その他」の場合は、回答がまだ得られていないと見なし、同じ質問をもう一度聞き直す。質問の聞き直しは指定された回数まで繰り返すことができる。後述の実証実験では最大3回とした。また、最後の聞き直しでは、「はい」「いいえ」等のシンプルな回答を求める旨、システムが発話することで、更なる聞き直しの必要性を抑制する。

上記2.の自由回答質問に対しては、回答されたかどうかを判定し、回答された場合は更にユーザ発話から回答キーワードを抽出する。対話例を図5に示す。この例では、「好きな動物は何ですか?」という質問に対して、「犬が好きです」と回答している。この場合は、回答があると判定され、回答キーワードとして「犬」を抽出する。回答キーワードは、「好きな動物は犬なのですね」のように、ユーザの回答に対するMICSUSの応答に含め、MICSUSがユーザの回答を正しく理解したこ

表1 Yes/No 質問に対する意味解釈ラベル

ラベル	説明	例 (質問「薬を水か白湯で飲んでますか?」)
Yes	肯定	「飲んでます。」
No	否定	「飲んでません。」
不明	回答が分からない	「覚えてないです。」
前提矛盾	質問の前提を否定	「薬は飲んでません。」
回答拒否	回答を拒否	「答えたくないです。」
その他	上記5種類のいずれにも分類されなかった	「薬は、えっと…。」



図5 自由回答質問と雑談の対話例(PC版の対話履歴確認・修正 Web アプリの画面)

とを伝える。さらに、この例では「犬」という単語が、雑談を開始するのにふさわしいトピックだと判断されたため、WEKDAによって生成された、犬の行動に関する雑談的応答をシステム側から発話している。

なお、後述するように、ユーザ発話意味解釈モジュールを深層学習で構築する際の学習データには、Webから抽出したのもも含め介護分野以外の質問やその回答等も含まれており、介護分野以外の対話でも高い精度で動作する。

2.3 対話シナリオ

対話シナリオは、「適切なケアマネジメント手法」に基づいて、ユーザの健康や生活習慣などを問う質問を展開していくよう作成されている。以下ではその作成手順を説明する。まず「適切なケアマネジメント手法」に基づき、人手で質問とその回答に対するシステムの応答文を作成した。例えば、「毎日3食食べていますか?」という質問と、それに対してユーザがYesと回答した際の応答文である「それは良いですね。」やNoの場合の「それは心配ですね。」といった応答文を作文した。質問や応答文はユーザである高齢者が理解できるか、気分を害するような発言をしないかなど、介護の専門家も交えて議論を繰り返しつつ整備した。この作業に加えて、質問応答の結果に応じて次にどの質問を行うかという遷移情報を決定する作業が必要である。例えば、まず福祉用具を使用しているかを問い、それ

に対してYesの回答であれば、使用している福祉用具に対して不満や要望がないかなどの質問に遷移し、Noであれば別の話題に遷移する、といった具合である。作成した質問文等と質問の遷移情報を対話シナリオオーサリングツールに入力することで、対話システムで実行可能なXML形式の対話シナリオを得ることができる。なお、オーサリングツール等のMICSUSの各種モジュールは汎用的に作られており、介護と関連のない質問等にも対応できるので、介護支援以外を目的とした対話シナリオを作成することもできる。

対話シナリオは、基本的に「適切なケアマネジメント手法」に基づいて人手で作成したが、人手による作業だけで柔軟な対話を実現することは難しい。例えば、ユーザ発話の解釈がうまくできなかった場合に再度同じ質問を聞く「回答の聞き直し」といったやりとりは全ての質問に対して発生し得るが、そのやりとりに対応する対話シナリオを逐一人手で追加するのは高コストである。そこで、シナリオ自動拡張モジュールによって、人が書いた骨格となる基本的対話シナリオを自動的に拡張し、回答の聞き直しのように対話中で広範に起こり得る一般的なやりとりを人手による記述なしで実行できるようにした。このシナリオ自動拡張モジュールによって、対話中の任意の箇所が発生し得る雑談機能の呼び出しや、一度行った回答を訂正するやり取りなどを行う対話シナリオなども実現しており、柔軟な対話が可能となっている。

2.4 雑談的応答の生成

MICSUS の雑談機能では、複数の応答生成エンジンを組み合わせることができる。現時点では Web 上の情報を基に雑談的応答を生成する WEKDA と、Web ニュースから生成する KACTUS [3] が利用可能である。本節で、NICT DIRECT で開発中の雑談対話システム WEKDA について詳しく述べる。

WEKDA は、ユーザ発話に対して、収集済みの Web 文書を元に、ユーザ発話に関係して、ユーザにとって有用と思われる情報を含む応答文を生成するように設計されている。既存の対話システムには、^{あいづち}相槌やそれに近い表現を応答として返すだけで、新規な情報は提供しないものも多いが、対話システムが新規な情報を定常的に提供しないかぎり、いずれユーザには飽きられ、使われなくなるだろうと考えており、新規な情報の提供は必須であると考えている。

ユーザにとって有用な情報についてより正確に言うと、応答文は、例えばユーザ発話の話題が「料理」であれば、「料理をすると脳が活性化する」のような、ユーザ発話の話題に関する広い意味でのチャンス（この場合は「脳が活性化する」）やリスクに関する情報を含む文が中心となるように設計されている。例えばユーザが「トマトが好きです」と発話すると、まずその話題に関するチャンスやリスクなどを問う質問を深層学習技術やパターンを用いて生成する。例えば「トマトが好きだとどうなる」「どうやってトマトを使うといい?」といった質問が生成される（「どうやって」、つまり、How-to を問う質問の回答は、何らかの目的をより適切に達成するための知識であり、目的を適切に達成するチャンスを増やすという意味でチャンスに関する知識だと考えることができる）。次に、同じく NICT DIRECT で開発し、試験公開中の大規模 Web 情報分析システム WISDOM X^{*5}（本特集号 3-3[5] 参照）を利用して質問の答えとそれを含む文章を多数取得する。その中からユーザ発話に対する応答として適したものを選択し、それに基づいて雑談的応答を生成する。先の例では、「リコピンを効率よく摂るためにはトマトを加熱することがポイント」といった情報がありましたよ。」のような応答を生成する。応答は、その情報源となった URL とともに画面に表示され、必要に応じて元のページが閲覧できるようになっている。

チャンスやリスクを記述した応答を雑談的応答の基本としているのは、雑談といえど、高齢者が将来参考にしたり、実際に活用したりできる情報を提供すべきであり、また、そうした情報とは基本的に将来におけるチャンスやリスクに関するものであるという発想、また、さらに押し進めて、雑談の本質とはそうした活用可能で有用な情報を話者間で共有することにあると

いう発想がある。

ユーザ発話に対する雑談的応答の適切さは BERT を利用したスコア付けによって判定される。この BERT の学習データ作成に際しては、作業者が書き下ろしたユーザ発話のサンプルに対して WISDOM X 等を使って雑談的応答の候補を収集し、雑談的応答候補の自然さ（例えば、日本語として理解可能か、ユーザ発話と無関係な話題になっていないかなど）、面白さ（誰でも知っている当たり前の内容になっていないかなど）といった観点に基づき、作業者がラベル付けを行った。

MICSUS で現在実際に利用している WEKDA は、これを更に簡略化、高速化したものである。高速化版を作成する準備として、Web 上の高頻度名詞 100 万件に対して、上記のアルゴリズムによって雑談的応答の候補を大量に作成し、雑談的応答の大規模なデータベースを構築した。ユーザ発話が入力されたら、まず、発話中で話題の中心となる単語（トピックワード）を BERT で選択し、先に構築したデータベースからそのトピックワードを含む雑談的応答の候補を抽出する。次に、ユーザ発話との類似性等に基づいて候補を選択し、最終的な雑談的応答として出力する。雑談的応答のデータベースを使うことにより、WISDOM X の呼び出し等の処理が不要となったため、雑談的応答の生成を大幅に高速化することができ、同時に多数のユーザが対話している状況であっても雑談を自然なタイミングで行うことが可能となった。

また、WEKDA の雑談的応答に対して、ユーザである高齢者が「教えてくれてありがとう」「面白いねえ、知らなかった」「今度、試してみます」等の応答をすることがある。こうした応答に対しても BERT を用いた意味解釈を行い、その結果に基づいて「どういたしまして」「お話しできてよかったです」「ネットには信頼の置けない情報もあるので、ほかの人にも聞くとかしてくださいね」といった応答を返すようにした。また、「料理をすると脳が活性化する」という WEKDA の情報提供に対して、「脳が活性化すると囲碁が強くなるかも」等のユーザ応答が返ってきた場合に、新たに発話に現れたキーワード「囲碁」に関して更に雑談を継続するといった機能も備えている。

3 ユーザ発話意味解釈モジュールのための 深層学習技術の開発

Yes/No 質問に対するユーザの応答を 6 種類に分類

*5 <https://wisdom-nict.jp> で試験公開中、どなたでも利用可能。

するなどといった機能を持つ、意味解釈モジュールの開発にあたっては、NICT DIRECT でアノテーション作業を実施して合計約 300 万件に上る学習データを作成し、大規模ニューラルネットワークである BERT [4]、もしくは NICT DIRECT で独自に開発した BERT の改良版である HBERT を用いて学習を行った。

BERT には 350 GB の Web テキストを用いて事前学習を行った BERT_{LARGE} を用いた (NICT DIRECT で開発した大規模ニューラルネットワークについては、本特集号 3-4[6] 参照)。ただし、通常の BERT は、仮名漢字混じり文を入力とするため、2つの単語が音声的に似ているという情報を利用することは難しい。このため、ユーザ入力に音声認識エラーがあると性能が大きく落ちる。そこで、通常の仮名漢字混じり文に加えて単語の読み情報も入力するほか、疑似的な音声認識エラーとして学習データにノイズを加えるという工夫を行うことで音声認識エラーに対する頑健性を高めた HBERT というモデルを開発し、合わせて使用している。

Yes/No 質問に対するユーザの回答を「Yes」「No」「不明」「前提矛盾」等に分類する Yes/No 質問回答分類 (表 1 参照) については、後述する実証実験時点では、1,730,875 件の学習データを用いて学習を行ったモデルを使用した。このモデルの分類性能は、アノテータがテキストベースで作成した評価データでは、平均精度のマクロ平均で BERT では 93.1 %、HBERT では 94.3 % となった。また、ノイズ (疑似的な音声認識エラー) を加えた評価データでは、BERT では 87.8 %、HBERT では 92.9 % となった。HBERT は、ノイズを加えた場合にも精度の低下が小さく、BERT に比べて約 5 % の性能向上を実現できたことがわかる。また、HBERT はノイズなしのデータに対しても性能向上を達成している。なお、データ作成時には、「はい」のような単純な応答は避け、「毎日 3 食食べていますか?」に対する「胃腸の調子がよくてね」のような、遠回しな表現を積極的に作成するようにしており、そのような

分類難易度の高いユーザ発話に対して高い精度で解釈できることがわかる。

このモデルは、介護関連の質問だけでなく、Web から取得した分野を問わない一般的な質問を学習データに含めることで、介護に限らない様々な話題について高精度で動作するよう作成している。介護分野以外の評価データで Yes/No 質問回答分類を評価した結果では、ノイズなしの評価データに対して HBERT で平均精度 98.0 % を達成した。遠回しな表現を多く含めたかどうか等、データ作成時の方針が異なるため単純な比較はできないが、介護分野以外の質問についても高い性能で動作することがわかる。

表 2 に、Yes/No 質問回答分類以外も含めたいくつかの主要なモデルの学習データの規模と性能についてまとめた (性能は 2 値分類のタスクに関しては正例の平均精度、多値分類のタスクに関しては平均精度のマクロ平均を示している)。いずれも評価データはテキストベースで作成したものであり、実際の音声対話で出現し得る音声認識エラーなどを含まないが、ノイズ (疑似的な音声認識エラー) を加えた評価データを用意して結果を比較できるようにしている (実際の音声対話を対象にした実験結果については次節を参照されたい)。表に掲載したいずれのモデルでも、ノイズありの場合は、HBERT の性能が BERT の性能を上回っていることがわかる。また、HBERT はノイズなしのデータに対しても BERT 版と同等かそれ以上の性能を示していることが読み取れる。

4 実証実験

社会実装に向けて、実際に高齢者が MICSUS と対話する実証実験を行い、技術的課題の洗い出しとその解決を実施している。実証実験は、高齢者介護施設に同居している高齢者や独居の高齢者などを対象として多数回実施している。本稿では 2022 年 2 ~ 3 月に 4 名の高齢者 (独居の 70 歳代の方 3 名と 80 歳代の方 1 名) の

表 2 意味解釈モジュールの主要なモデルの学習データの規模及び性能 (平均精度 %)

モデルの種類	学習データ件数 (うち人手作成)	ノイズなし BERT	ノイズなし HBERT	ノイズあり BERT	ノイズあり HBERT
Yes/No 質問 回答分類	1,730,875 (665,221)	93.1	94.3	87.8	92.9
自由回答質問 回答有無判定	58,522 (58,522)	98.1	98.1	97.0	97.5
内容ベース訂正対象特定	182,834 (61,342)	95.7	96.0	94.5	95.6
雑談開始判定	313,825 (139,164)	99.0	99.0	97.7	98.8

自宅に MICSUS を設置、15 日間毎日対話を行ってもらった実証実験の結果について報告する。現状の MICSUS には 15 種類の対話シナリオがあり、それぞれ異なる質問が含まれている。実証実験ではこの 15 シナリオを毎日 1 シナリオずつ使って対話を行い、健康状態や生活習慣などの情報を取得した。15 日間を通して、Yes/No 質問は 4 名の合計で 398 回実行され、その回答を解釈する機能の正解率は 93.7 % であった。6.3 % の解釈誤りのうち最多の 64.0 % は、ユーザの発話が長くて複雑なケースか、もしくは曖昧で人間でも分類しにくいケースであった。例えば、質問「いつも外へ出る時は歩くことが多いですか？」に対する「近所でしたら大体歩きですが、遠出するときは車に乗せてもらうことが多いですね」という回答のように、人間であっても Yes か No か判断しにくいような事例である。また、質問の聞き直しや回答の訂正などで、最終的にユーザの意図通りに回答が解釈されたかという基準での正解率は 95.9 % であった。Yes/No の解釈は、MICSUS の最も基本的な機能の一つであるが、かなり高精度で解釈できることが分かった。

高齢者の社会的孤立やコミュニケーション不足の抑制を目的とした雑談機能については、15 日間を通して 4 名で合計 157 個の WEKDA による雑談応答が生成された。システム開発者ではない 3 名の評価者が対話の動画を見て、雑談応答に対する高齢者の反応や応答の質を独立に評価し、多数決で評価結果を確定したところ、157 件中の 89.2 % の 140 件が雑談応答として適切であると分かった。また、157 件中の半数程度である 74 件 (47.1 %) について、高齢者が雑談応答に対して笑顔を見せるか、もしくは「やってみます」など行動することを示唆する応答をする、雑談の内容について質問をするなど積極的に興味を示していた。MICSUS の雑談応答は、高齢者が提示した話題に関する広い意味でのチャンスやリスクなど、何らかの新規情報を含む設計になっている一方、既存対話システムは相槌やそれに類する簡単な表現を多用するものもある。相槌等の表現は広範な状況で妥当な雑談応答とみなすことができるため、そうしたシステムでは雑談的応答のうち妥当なもの割合は高くなるが、MICSUS のように新規な情報を提供するタイプの対話システムで妥当な雑談応答を生成することはより難しくなる。こうした点を考慮すると、上記の数字は、極めて良好な品質が達成されていることを示していると考えている。

5 まとめ

NICT DIRECT が他社と共同で開発を進めている介護支援用マルチモーダル音声対話システム MICSUS

について紹介した。MICSUS は、人手不足が懸念されるケアマネジメント業務の一部を代替し、介護従事者の負担を軽減するとともに、より頻繁な健康状態のチェックを可能にするなど、質の高い介護サービスを実現することを目標とする。さらに、コミュニケーション不足の解消を通じて高齢者の健康状態改善を図る。本稿では、MICSUS の構成、意味解釈モジュールの詳細とその構築、また、雑談対話システム WEKDA を用いた雑談機能について詳しく紹介した。また、高齢者宅に実際に MICSUS を設置して実施した実証実験の結果についても報告した。

謝辞

本研究は総合科学技術・イノベーション会議の戦略的イノベーション創造プログラム (SIP) 第二期「Web 等に存在するビッグデータと応用分野特化型対話シナリオを用いたハイブリッド型マルチモーダル音声対話システムの研究」(管理人: NEDO) によって実施されたものである。

【参考文献】

- 1 株式会社日本総合研究所, 平成 30 年度厚生労働省老人保健事業推進費補助金 (老人保健健康増進等事業) 適切なケアマネジメント手法の策定に向けた調査研究報告書, 2019.
- 2 齊藤 雅茂, 近藤 克則, 尾島 俊之, 平井 寛, JAGES グループ, “健康指標との関連からみた高齢者の社会的孤立基準の検討: 10 年間の AGES コホートより,” 日本公衆衛生雑誌, vol.62, no.3, pp.95-105, 2015.
- 3 吳 剣明, 楊 博, 服部 元, 滝嶋 康弘, “マルチモーダル認識技術を活用した対話 AI [KACTUS],” (株) KDDI 総合研究所, 2021.
- 4 J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of deep bidirectional transformers for language understanding,” Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp.4171-4186, 2019.
- 5 吳 鍾勲, クロエツェージュリアン, “大規模 Web 情報分析システム WISDOM X 深層学習版,” 情報通信研究機構研究報告, 本特集号, 3-3, 2022.
- 6 飯田 龍, “DIRECT における深層学習を用いた大規模自然言語処理,” 情報通信研究機構研究報告, 本特集号, 3-4, 2022.



水野 淳太 (みずの じゅんた)

ユニバーサルコミュニケーション研究所
データ駆動知能システム研究センター
主任研究員
博士 (情報学)
自然言語処理

【受賞歴】

- 2019 年 文部科学省 平成 31 年度 科学技術分野の文部科学大臣表彰 科学技術賞 (研究部門)
- 2019 年 公益財団法人通信文化協会 第 61 回 前島密賞
- 2016 年 情報処理学会 2016 年度 (平成 28 年度) 山下記念研究賞



浅尾 仁彦 (あさお よしひこ)

ユニバーサルコミュニケーション研究所
データ駆動知能システム研究センター
主任研究員

Ph.D. (言語学)

言語学、自然言語処理

【受賞歴】

2022年 令和4年度情報通信研究機構成績優
秀表彰 優秀賞