

# 介護支援対話システム MICSUS のための意味解釈モジュール

浅尾 仁彦<sup>1</sup> 水野 淳太<sup>1</sup> 呉 鍾勲<sup>1</sup> Julien Kloetzer<sup>1</sup> 大竹 清敬<sup>1</sup>  
 福原 裕一<sup>1</sup> 鎌倉 まな<sup>1,2</sup> 緒形 桂<sup>1</sup> 鳥澤 健太郎<sup>1,2</sup>

<sup>1</sup> 国立研究開発法人情報通信研究機構 (NICT) <sup>2</sup> 奈良先端科学技術大学院大学  
 {asao, junta-m, rovellia, julien, kiyonori.ohtake, fukuhara, kamana, kei.ogata, torisawa}@nict.go.jp

## 概要

介護支援対話システム MICSUS でユーザ発話を解釈するためのモジュールの概要とその性能等についてまとめる。MICSUS は介護・支援の必要な高齢者の健康状態等の定期的なチェックを行うとともに、高齢者のコミュニケーション不足を補うことで、介護従事者の負担を軽減し、高齢者の健康維持を促進する目的で開発が進められている対話システムである。MICSUS は、健康状態等に関する質問に対するユーザの自由な応答を解釈し記録することができるほか、ユーザ発話によってはウェブ情報に基づいた雑談に遷移することも可能である。また、音声認識エラーによる意味解釈の失敗を防ぐため HBERT と音声認識エラー検出の2つの技術も開発した。

## 1 はじめに

介護支援対話システム MICSUS<sup>1)</sup> は、内閣府総合科学技術・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2期の支援のもと、KDDI 株式会社・情報通信研究機構 (NICT)・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の共同で開発が進められている高齢者向けの対話システムである [1, 2]。現在、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<sup>2)</sup>。MICSUS は、ケアマネジメント業務の負担軽減のため、主にあらかじめ設計した対話シナリオを元に、システムからユーザに質問する形で高齢者の健康状態、生活習慣等をモニタリングする。同時に、雑談対話システム



図1 介護用対話システム MICSUS。音声だけでなく表情やジェスチャーも認識可能

WEKDA [3] 等と接続して、ユーザ発話によってはウェブ情報を用いた雑談に遷移することで、飽きのこないシステムにするとともに、高齢者のコミュニケーション不足を補うことを狙っている。定期的な健康チェックの徹底やコミュニケーション不足の解消により要介護度の進行や認知症の発症、死亡率を抑えられることが知られており [4, 5]、MICSUS は介護の質を向上させるうえで重要な役割を果たすことが期待されている。現在、高齢者施設と連携して長期にわたる実証実験を進めている。

MICSUS の外見を図1に示した。MICSUS はジェスチャーや表情も認識可能なマルチモーダルな対話システムとして開発が進められているが、本発表では、NICT が担当して研究開発を行っている、ユーザ発話を解釈するモジュールに議論を絞る。MICSUS の主要機能には、Yes/No 質問による健康状態等のモニタリング、好みや習慣などを聞いて話題を広げるための個人属性質問応答、ユーザの回答の訂正、重複した質問を避ける含意認識機能、雑談機能がある。本発表では、このうち一部の機能について、その概要と利用例、およびその実現のため構築した、ユーザ発話を解釈するための深層学習モデルについて述べる。また、音声対話システムの課題である音

1) 紹介動画 (2020年10月公開) が <https://www.youtube.com/watch?v=gCURC3f9-Go> にある。また、けいはんな R&D フェア 2021 「高齢者介護支援マルチモーダル音声対話システム MICSUS」 <https://keihanna-fair.jp/exhibition/ai/899> では 2021年10月時点のデモ等が閲覧できる。

2) 日経新聞 2021.12.9 付 「介護現場は DX の実験場」 . [https://www.nikkei.com/nkd/industry/article/?DisplayType=1&\n\\_m\\_code=124&ng=DGKKZ078281610Y1A201C2EA1000](https://www.nikkei.com/nkd/industry/article/?DisplayType=1&\n_m_code=124&ng=DGKKZ078281610Y1A201C2EA1000)

声認識エラーへの対処について、音声認識エラーを含むテキストに対して頑健な分類を実現する言語モデル HBERT と、音声認識エラー検出の 2 種類の技術も提案する。特に HBERT は、評価データに音声認識エラーがない場合も含めて、BERT<sub>LARGE</sub> [6] からの性能向上を実現した。本発表で言及するモデルはいずれも 350GB (22 億文) のウェブテキストで事前学習を行った BERT もしくは HBERT をもとに、それぞれの機能に合わせた学習データを NICT のアノテータが作成し、自動生成した学習データと合わせてファインチューニングを行ったものである。

## 2 関連研究

少子高齢化の進むなかで高齢者を対象にした対話システムは注目を集めており、既存研究に山本ら [7]、大竹ら [8]、下岡ら [9] の対話システムなどがある。例えば下岡らは、相づち・問い返し・共感応答などによってユーザの語りを促す傾聴対話システムを構築しており、音声認識エラーへの対策として、音声認識の信頼度によって応答生成の手法を使い分けることも実現している。傾聴ないし雑談を主としたこれらの対話システムと比較すると、MICSUS は健康状態等のモニタリングという明確な目的をもつため、あらかじめシナリオを用意してシステム主導で目的指向型の対話を行うという点が異なっているが、同時に、雑談機能によってユーザの主体的な話題提供に対応することも可能となっている点にオリジナリティがある。

## 3 Yes/No 質問

「夜はよく眠れていますか？」のような Yes/No 質問による高齢者の健康状態等のモニタリングは、MICSUS の最も基本となる機能である。この機能のため、Yes/No 質問に対するユーザ応答を Yes, No, Unknown, PresuppositionFailure, Other の 5 値に分類するモデルを作成した (表 1)。Yes, No 以外のカテゴリとして、Unknown (ユーザがわからないと答えている場合)、PresuppositionFailure (ユーザが質問の前提が成り立たないと指摘している場合)、Other (その他の場合。ユーザが質問に答えず他のことを言っている場合など) が設けてある。モデルは、例えば「毎日しっかり食べていますか？」という質問に対し、「食べています」や「いいえ」のような単純な応答だけでなく、「胃は丈夫なんだよ」「それが最近すっかり食が細くなって」のような間接的な応答で

あっても適切に分類できるよう、意図的に後者のような発話事例の多い学習データを作成し学習した。

表 1 Yes/No 質問応答の 5 値分類と、「お薬を飲み忘れなく飲んでいますか？」という質問を例とした応答例

分類	ユーザ応答例
Yes	早く治したいですからね
No	実は昨日も忘れちゃって
Unknown	どうだろう、あやふやです
PresuppositionFailure	薬はもらってません
Other	なんでそんなこと聞くの？

学習・評価データには、株式会社日本総合研究所が策定を進めているケアマネジメント標準 [10] に基づいて作成された介護関連の質問のほか、介護とは関係のない一般的な質問も含めることで、幅広い目的で使用できるようにした。最新モデルは評価データ全体では平均精度 (AP) のマクロ平均 94.3%、また介護とは関係のない一般的な質問応答の評価データでは 98.1% を達成した<sup>3)</sup>。2021 年 12 月に高齢者施設で高齢者 1 名を対象に 15 日間にわたり実施した音声対話での実証実験<sup>4)</sup>では、ユーザ発話ごとに評価すると正解率 89.9% (129 件中 116 件) であった。分類結果が Other の場合は回答を確定させずに再質問する仕様になっているため、再質問のやりとりまで含めて最終的に正解できたかどうかで評価すると、正解率は 94.3% (123 件中 116 件) であった。不正解となった事例 7 件中 4 件は、ユーザ発話に Yes と判断できる部分と No と判断できる部分の両方を含むなど、人手でも分類の難しい曖昧な発話をしているケースであった。

## 4 個人属性質問

個人属性質問は「好きな食べ物は何ですか?」「日課は何ですか?」「昨日は何を食べましたか?」のように、システムがユーザの好みや習慣などを聞く質問である。ユーザの回答から雑談 (6 節) につなげるなどして、モニタリング質問の連続による単調さを避け、場を和らげることを目的としている。個人属性質問については、ユーザが (「特にない」といった回答を含め) 質問に回答したかどうかを判定する回答有無判定と、ユーザ発話から回答となるキーワードを抽出する回答部分抽出の 2 種類のモデルを用意した。回答有無判定によって、その質問を終了

3) アノテータが作文した評価データに基づくものであり、音声認識エラーなどは含まない。評価データについても、意図的に単純な応答を避け、分類が難しいものが多くなるように作成されている。

4) コロナ禍で大規模な実証実験の実施が難しい状況のため、少人数で実証実験を進めている。

してよいかどうかを判断することができ、また、回答部分抽出によって、例えばユーザが「好きな食べ物は何ですか？」に対して「最近はお魚がおいしくてね」と答えた場合に、「お魚がお好きなのですね」のように応答することが可能になっている。

回答部分抽出は、現状、上記の「お魚」の例のように回答が名詞で表せるファクトイド質問の場合にしか対応していない。例えば why, how 等を訊ねる非ファクトイド質問の扱いは今後の課題となるが、現在、NICT で公開している深層学習版 WISDOM X<sup>5)</sup> で実現している技術を応用することで解決できると考えている。

## 5 訂正機能

訂正機能は、システムの誤認識（あるいはユーザの言い間違い）のため一度誤って確定してしまった回答を訂正する機能である（表 2 に対話例）。この機能により、ユーザは「すみません、訂正があるのですが」「さっきの薬の話に戻れますか？」「一つ前の質問ですが、よく眠れていないって言ったんです」のように、任意のタイミングで自由な表現で訂正要求を行うことができる。ただし、このような訂正要求はシステムによる質問等を遮って行わなければならないため、不慣れなユーザには難しい。そこで、システムでそれまでの質問応答の結果を要約し、「ちょっとこれまでの回答を確認させてください。毎日しっかり食べている、薬を飲み忘れなく飲んでいる、ということですが、訂正があれば教えてください」のように、訂正が必要ないかどうか確認を促す機能を合わせて実装した。

表 2 訂正機能の利用例

システム：	毎日しっかり食べていますか。
ユーザ：	すみません、訂正があります。
システム：	訂正ですね。
	訂正したい質問を教えてください。
ユーザ：	薬の質問です。
システム：	それでは、薬を飲み忘れなく飲んでいるかもう一度教えてください。
ユーザ：	飲み忘れることもあります。
システム：	薬を飲み忘れることがあるのですね。

訂正機能では、「訂正したいのはその質問じゃないです」のようなユーザ発話で訂正対象の変更を行うことや、システムが訂正の意図を誤認識した場合に「訂正ではありません」のようなユーザ発話で訂正をキャンセルすることも可能になっている。

5) <https://www.nict.go.jp/press/2021/03/31-3.html>

## 6 雑談

雑談機能は、ユーザ発話の内容が雑談的であるときに、NICT で開発している雑談対話システム WEKDA [3] を使用して雑談的な応答を返す機能である。表 3 に雑談の例を掲載した。

表 3 雑談の例

システム：	毎日しっかり食べていますか。
ユーザ：	食べてるよ。鍋がおいしくてね。
システム：	毎日しっかり食べているのですね。ネットに、鍋料理も土鍋を使うと保温もしっかりとできるといった情報がありましたよ。

WEKDA は、ユーザ発話に含まれる名詞をもとにウェブ文書から雑談に適した情報を検索することで雑談応答を実現するシステムであり、深層学習版 WISDOM X の採用などにより性能が大きく向上しているが、詳細については本発表では省略する。MICSUS には、雑談に移行するのにふさわしいユーザ発話を検出する雑談開始判定と、雑談を終了するほうがよいユーザ発話を検出する雑談終了判定の 2 つのモデルを実装した（このほか 7.2 節で述べる音声認識エラー検出も利用される）。

## 7 音声認識エラーへの対応

MICSUS は音声で対話を行うシステムであるため、入力には音声認識エラーが含まれている可能性があり、音声認識エラーがあっても対話を大きく破綻させることなく動作させることが重要な課題になる。この節では、音声認識エラーの影響を抑えるため導入した HBERT と音声認識エラー検出という 2 種類の技術について述べる。HBERT は、音声認識エラーがない場合も含めて性能向上を実現した。

### 7.1 HBERT

単語列を入力とする通常の BERT では 2 つの単語等が音的に似ているという情報を利用することは難しい。そのため、ユーザ入力に音声認識エラーがあるとその性能は大きく落ちる。この問題に対処するため、HBERT と呼ばれる技術を開発した。これは、通常の漢字カナ混じり表記の入力以外に、形態素解析によって得られるカナ表記も入力に加えて事前学習を行うものである。事前学習済みのモデルをファインチューニングする際に、疑似的な音声認

表4 主なモデルでの BERT と HBERT の比較

機能	モデル	学習データ件数		評価基準	ノイズなし		ノイズあり	
		(うち人手作成)			BERT	HBERT	BERT	HBERT
Yes/No	Yes/No 質問回答分類	1,730,875	(665,221)	AP マクロ平均	89.2	<b>94.3</b>	83.6	<b>92.9</b>
個人属性	回答有無判定	58,522	(58,522)	正例 AP	97.3	<b>98.1</b>	96.3	<b>97.5</b>
訂正	内容に基づく訂正対象特定	182,834	(61,342)	正例 AP	95.3	<b>96.0</b>	94.1	<b>95.6</b>
雑談	雑談開始判定	313,825	(139,164)	AP マクロ平均	97.6	<b>99.0</b>	96.1	<b>98.8</b>

MICSUS で使用しているモデルのうち一部のみを示している。AP は平均精度 (average precision)。開発データで最も性能の良かったモデルの評価データでの性能を掲載している (ハイパーパラメータの探索範囲はモデルによって異なる)。「ノイズあり」の性能は、ノイズ率 0%、10%、20%、50% の 4 種類の評価データでの評価結果の平均。学習データ件数については括弧内で、直接アノテータがユーザ発話を人手作成したものの件数を示した。それ以外の学習データとして、既存の言語資源を利用したものや、システム発話との組み合わせの変更、複数のユーザ発話の連結などによって機械的に生成したものがあ。HBERT のためにノイズを加えて生成したものは学習データ件数に含まない。

識エラーとして学習データにノイズ<sup>6)</sup>を入れることで、音声認識エラーに対して頑健なモデルを実現することができる<sup>7)</sup>。表5 に音声認識エラーに対して頑健な動作の例を示す。

表5 音声認識エラーに頑健な Yes/No 分類の例

システム:	お水をよく飲んでますか?
ユーザ:	飲んでます (音声認識結果: 飛んでます)
システム:	お水をよく飲んでるのですね。

表4 に主要なモデルの BERT 版と HBERT 版の性能比較について要約した。ノイズなしの評価データによる性能と、ノイズありの評価データによる性能の2種類を示している。表に示したいずれのモデルでも、BERT、HBERT とともにノイズありの評価データを用いた場合には性能が落ちるが、HBERT は比較的性能の低下が小さいことがわかる。また、ノイズなしの評価データを用いた場合にも、表に示した全てのモデルで HBERT は BERT を上回る性能となっている。

## 7.2 音声認識エラー検出

個人属性質問や雑談機能などでは、HBERT を使用するかどうかにかかわらず、ユーザが用いた名詞をシステム発話で利用するため、そのキーワードが音声認識エラーの結果生じた語である場合は致命的な対話の破綻を招いてしまう(表6)。特に、雑談開始判定は、ユーザが新しい話題を出したときに雑談開始と判定することを意図しているため、音声認識エラーで予期しない語が認識されると、それが原因で誤って雑談開始してしまう可能性が高い。

6) ノイズは音素列で表現した単語間の編集距離に基づき、類似した単語にランダムに置き換えることで作成した。ノイズを入れる単語の割合は 0%、10%、20%、50% の 4 種類を試した。表4 では開発データで最高性能になったものについて報告している。評価データのノイズも同じ方法で生成している。

7) 関連研究として、通常の英単語列の他に、音声認識結果を発音記号で入力させる Phoneme-BERT [11] がある。

表6 雑談における音声認識エラーによる対話破綻の例

システム:	毎日しっかり食べていますか?
ユーザ:	はい、ちゃんと食べてます。 (音声認識結果: 愛車と食べてます。)
システム:	適切なコーティングを愛車に施すと汚れにくくなるといった情報がありましたよ。

このような対話の破綻を防ぐため、ユーザ発話に含まれる名詞について、音声認識エラーによるものかどうかを判別するモデルを作成した。このモデルを利用し、音声認識エラーがあった場合は雑談開始としない(もしくは、音声認識エラーの部分捨て、残りの部分で雑談開始するかどうかを判定する)という動作を実現できる。音声認識エラー検出モデルの学習データにあたって、人手作成したユーザ発話に含まれる名詞を他の名詞にランダムに差し替えることで疑似的な音声認識エラーとした。同じ方法で作成した疑似的な評価データでの精度は 99.9% を達成した。また、このモデルが実装されていなかった 2021 年 6~7 月の実証実験では、音声認識エラーのため誤って雑談開始してしまったケースが 13 件あったが、このモデルを実装すると、その全てのケースについて音声認識エラーが検出でき、誤った雑談開始を避けることができることがわかった。

## 8 おわりに

本発表では、介護用対話システム MICSUS でユーザ発話を解釈するために開発したモデル群について報告した。今後さらに高齢者を対象とする実証実験を重ねてその結果をシステムの改善につなげるほか、NICT で開発、公開している深層学習自動並列化ミドルウェア RaNNC [12] を用いたより大規模な言語モデルを採用することで、さらなる精度の向上を目指す。

## 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)<sup>8)</sup>「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」(管理法人：NEDO)によって実施されたものである。

## 参考文献

- [1] Yoshihiko Asao, Julien Kloetzer, Junta Mizuno, Dai Saiki, Kazuma Kadowaki, and Kentaro Torisawa. Understanding user utterances in a dialog system for caregiving. In *Proceedings of the 12th Language Resources and Evaluation Conference (LREC2020)*, pp. 653–661, 2020.
- [2] 浅尾仁彦, Julien Kloetzer, 水野淳太, 齊木大, 門脇一真, 鳥澤健太郎. 介護用対話システムのための高齢者の発話理解. 言語処理学会第26回年次大会発表論文集, pp. 125–128, 2020.
- [3] 水野淳太, クロエツェージュリアン, 田仲正弘, 飯田龍, 呉鍾勲, 石田諒, 浅尾仁彦, 福原裕一, 藤原一毅, 大西可奈子, 阿部憲幸, 大竹清敬, 鳥澤健太郎. WEKDA: Web 40億ページを知識源とする質問応答システムを用いた博学対話システム. 人工知能学会第84回言語・音声理解と対話処理研究会, pp. 135–142, 2018.
- [4] 株式会社日本総合研究所. 平成25年度老人保健事業推進費等補助金老人保健健康増進等事業居宅サービス等における適正化とサービスの質の向上および保険者機能強化のための調査研究事業報告書, 2014.
- [5] 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.
- [6] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL): Human Language Technologies, Volume 1*, pp. 4171–4186, 2019.
- [7] 山本大介, 小林優佳, 横山祥恵, 土井美和子. 高齢者対話インタフェース: 『話し相手』となって、お年寄りの生活を豊かに. 電子情報通信学会技術研究報告. HCS, ヒューマンコミュニケーション基礎, Vol. 109, No. 224, pp. 47–51, 2009.
- [8] 大竹裕也, 萩原将文. 高齢者のための発話意図を考慮した対話システム. 日本感性工学会論文誌, Vol. 11, No. 2, pp. 207–214, 2012.
- [9] 下岡和也, 徳久良子, 吉村貴克, 星野博之, 渡部生聖. 音声対話ロボットのための傾聴システムの開発. 自然言語処理, Vol. 24, No. 1, pp. 3–47, 2017.
- [10] 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.
- [11] Mukuntha Narayanan Sundararaman, Ayush Kumar, and Jithendra Vepa. Phoneme-BERT: Joint language modelling of phoneme sequence and ASR transcript. In *Interspeech 2021*, 2021.
- [12] Masahiro Tanaka, Kenjiro Taura, Toshihiro Hanawa, and Kentaro Torisawa. Automatic graph partitioning for very large-scale deep learning. In *Proceedings of 35th IEEE International Parallel and Distributed Processing Symposium (IPDPS 2021)*, pp. 1004–1013, 2021.

8) <https://www8.cao.go.jp/cstp/gaiyo/sip/>