

感情分析による体調推定とエージェント対話を通じた高齢者体調確認サービスの提案

西山 敦紀[†] 中田 匠哉[†] 陳 思楠[†] 佐伯 幸郎^{††} 増田 廣介^{†††}
露崎 雄太^{††††} 安田 清[†] 中村 匡秀[†]

[†] 神戸大学 〒657-8501 神戸市灘区六甲台町 1-1

^{††} 高知工科大学 〒782-8502 高知県香美氏土佐山田町宮ノ口 185

^{†††} 医療法人共生会 長崎友愛病院 〒851-0401 長崎県長崎市蚊焼町 2314-1

^{††††} おゆみの中央病院 〒266-0033 千葉市緑区おゆみ野南 6-49-9

E-mail: [†]2174391t@stu.kobe-u.ac.jp, ^{††}tnakata@es4.eedept.kobe-u.ac.jp, ^{†††}chensinan@gold.kobe-u.ac.jp,

^{††††}saiki.sachio@kochi-tech.ac.jp, ^{†††††}kskmsds0504m@gmail.com, ^{††††††}tsuyuzaki.yuuta@oyuminohp.or.jp,

^{†††††††}yasukiyo.12@outlook.jp, ^{††††††††}masa-n@cmds.kobe-u.ac.jp

あらまし 我々の研究グループでは、仮想エージェント（VA）を使って在宅高齢者の「こころ」の内を聞き取り、様々な支援につなげる VA 傾聴サービスを開発している。しかしながら、既存の VA 傾聴サービスでは非言語情報を十分に活用できていない課題があった。そこで本研究では、顔表情による感情認識の学習済モデルを活用し、エージェント対話中の表情から体調を推定する機械学習分類モデルを構築する手法を提案する。また、分類モデルによる体調推定の結果を VA 傾聴サービスの対話に加え、高齢者の実際の体調を聞き出し記録するサービスを実装する。

キーワード 高齢者見守り, 表情, 感情認識, 体調, 機械学習

Checking Elderly Health Condition Using Face-based Emotion Analysis and Spoken-Dialogue Agent

Atsunori NISHIYAMA[†], Takuya NAKATA[†], Sinan CHEN[†], Sachio SAIKI^{††}, Kosuke MASUDA^{†††},
Yuta TSUYUZAKI^{††††}, Kiyoshi YASUDA[†], and Masahide NAKAMURA[†]

[†] Kobe University Rokkodai-cho 1-1, Nada-ku, Kobe, Hyogo 657-8501 Japan

^{††} Kochi University of Technology 185 Miyanokuchi, Tosayamada, Kami City, Kochi 785-8502, JAPAN

^{†††} Medicaï corporation Nagasaki Yuai Hospital 2314-1 Kayaki-machi, Nagasaki City, Nagasaki 851-0401 Japan

^{††††} Oyumino Central Hospital 6-49-9 Oyumino Minami, Midori-ku, Chiba-shi, Chiba 266-0033 Japan

E-mail: [†]2174391t@stu.kobe-u.ac.jp, ^{††}tnakata@es4.eedept.kobe-u.ac.jp, ^{†††}chensinan@gold.kobe-u.ac.jp,

^{††††}saiki.sachio@kochi-tech.ac.jp, ^{†††††}kskmsds0504m@gmail.com, ^{††††††}tsuyuzaki.yuuta@oyuminohp.or.jp,

^{†††††††}yasukiyo.12@outlook.jp, ^{††††††††}masa-n@cmds.kobe-u.ac.jp

Abstract Our research group is developing a VA listening service that uses virtual agents (VAs) to listen to the inner thoughts of elderly individuals living at home and connect them to various forms of support. However, existing VA listening services have faced challenges in fully utilizing nonverbal information. Therefore, this study proposes a method to construct a machine learning classification model that estimates health condition based on facial expressions during agent conversations, utilizing a pre-trained model for emotion recognition from facial expressions. Furthermore, we implement a service that incorporates the results of health condition estimation from the classification model into the VA listening service's dialogue, thereby eliciting and recording the actual health condition of the elderly.

Key words watching over the elderly, facial expressions, facial emotion recognition, health condition, machine learning

1. はじめに

近年日本の高齢者人口は増加しており、独居高齢者も増加している [1]。高齢者は孤立死を身近に感じており、実際に孤立死と考えられる事例が多数発生している [2]。そのため高齢者が孤立せず気付かれぬ間に体調が悪化しないためにも、日々高齢者への声掛けを行い体調を把握することが大切である。先行研究として仮想エージェントを用いて、在宅高齢者の日常生活の中での傾聴を行う VA 傾聴サービスを開発しており、感情表出の場として機能している [3]。また表情から体調を推定する手法の検討を行っており、体調ごとの表情演技を撮影した動画クリップを収集し、機械学習分類モデルを構築して体調を推定できることが示された。しかし、体調推定手法では長期的なスコープで挙げられている VA 傾聴サービスとの連携部分が未実装であった。VA 傾聴サービスを通じた表情による体調を推論すること、そして推論された体調に基づいてより詳しい状況を聞く必要があり、リアルタイムに動画を収集してラベル付けする方法を構築しなければならない。そこで本研究では以下のリサーチクエスチョンを設定した。

RQ1: 日常の対話の中で表情と主観的体調の収集を行い、その結果に基づいて深掘りするシステムをいかにして構築するか

RQ2: 実際の居住環境でデータ収集を行った場合においてどのような課題が生じるか

本研究の目的は対話中の表情感情から本人の主観的な体調を推論することで体調の把握と深掘りを行うサービスを提案することである。キーアイデアは一日のうちで初めて会ったときにその時の体調を聞く会話を開始すること、会話中に表情感情を蓄積して体調推論モデルをバッチ学習で構築すること、推論された体調を対話に組み込みより詳細な状況を聞き出すことである。バッチ学習の際には収集期間としてデータを集める期間を設け、後に推論期間に移行して深掘りを始める。目的を達成するために以下の手法を提案する。

A1: 感情認識と主観的体調の収集

A2: 機械学習モデルの作成

A3: 機械学習モデルによる体調推論

A4: 体調に応じた深掘り

この手法に沿って実装を行い、A1 に関する予備的評価を行った。その結果、対話エージェントとの対話中に体に負担をかけずに自然な形で表情感情のデータが取得できることが確認された。

2. 準備

2.1 超高齢社会

近年、日本の高齢者人口は増加しており令和 6 年度版高齢社会白書 [1] によると、総人口が減少する中で 65 歳以上の者が増加することにより高齢化率は上昇を続け、令和 19 年に 33.3 % となり国民の 3 人に 1 人が 65 歳以上の者となると見込まれている。また 65 歳以上の一人暮らしの者についても 65 歳以上の男女それぞれの人口に占める割合が、令和 2 年には男性 15.0 %、女性 22.1 % となり令和 32 年には男性 26.1 %、女性 29.3 % と

なると見込まれており独居高齢者も増加している。

独居高齢者については人と話をする頻度について「毎日」と回答した人の割合がひとり暮らし以外の人の半分以下という調査もあり、独居高齢者は高齢者の中でも特に孤立しているという状況が窺える [1]。65 歳以上の者の孤立死に対する意識調査においては、孤立死を身近に感じるかという問いに対して「とても感じる」又は「まあ感じる」と回答したものは 48.7 % である [1] とあり、高齢者自身も孤立死を身近に感じている。実際に、令和 6 年度の上半期に自宅において死亡した一人暮らしの者のうち 65 歳以上は 28330 人に上り [2]、孤立死と考えられる事例が多数発生している。

2.2 注目する問題

高齢者が孤立せず、気付かれぬ間に体調が悪化しないためにも、日々高齢者への声掛けを行い体調を把握することが大切である。東京都福祉保健局の高齢者の見守りガイドブック [4] では日常生活を送る中で、住民同士が互いにさりげなく気遣い合い、見守り合う関係を緩やかな見守りと呼び、コミュニケーションを交えながら住民同士が普段との違い、異変を感じてケアをしたり、専門機関に相談するなどのさりげない見守りが、今まで以上に重要性が高まっているとある。地域住民やそれ以上に近い家族と日々のコミュニケーションがあれば、普段との違いを気にかけてくれるが、孤立した状況では行われぬため、定期的なコミュニケーションの機会を設けることや、また見守りサービスの活用で状況を把握する必要がある。

また、高齢者の社会的孤立は課題であるが、単に対話の機会を設けるだけでは不十分である [5]。特に、健康や経済的な問題などの深刻な困難に直面していても、それらを他者に打ち明けて支援を求めることは容易ではない。このような自己開示の困難さは、適切な社会的支援を受ける機会を逸することに繋がり、結果として孤立をさらに深める要因となる。これに対し、野口らは、人間同士の対話をロボットが仲介する「社会的媒介ロボット」により、喪失経験のような通常は話しにくい話題において自己開示が促進されうることを示した [5]。ロボットは人間と社会的利害を共有しない相手として捉えられやすく、対面のコミュニケーションに比べて評価懸念や心理的抵抗が緩和されうる。

2.3 先行研究

2.3.1 VA 傾聴サービス

我々の研究グループでは、在宅で生活する高齢者の日常における傾聴を目的として、仮想エージェントを活用した「VA 傾聴サービス」を開発している [6] [7] [8]。本サービスは、高齢者が日常生活で生じる不安や悩みを仮想エージェントに話すことを通じて、心理面のケアを行うことを目的とする。実証実験の結果、本サービスは感情表出を促す場として機能し、孤独感および感情状態の変化に対して有効性が示唆されている [3]。

しかし VA 傾聴サービスは対話内容のテキストを活用することが主であり、対話中の表情などの非言語的な情報が十分に活用されていない。対話中に表情を取得して分析し、VA 傾聴サービスの発話に組み込むことで見守りに生かすことができる可能性がある。

2.3.2 表情による感情認識を用いた体調推定手法

先行研究として、短時間の動画クリップから体調を推定する手法の検討と評価を行っている。長期的なゴールとして対話エージェントの声掛けからテキストと非言語的な情報である表情と声色を取得し、体調を分析してスコアとして算出して日々保存し、スコアが悪ければ詳細な内容を聞くことで支援につなげることを挙げ、その中でも表情から体調を推定できるか調べることをスコープとしている [9]。体調推定手法ではまず体調を「良い」「普通」「悪い」の三段階に分けて動画クリップを撮影する。撮影した動画クリップからフレームを抽出し、既存の感情認識の学習済みモデルを適用して七つの感情クラス (angry, disgust, fear, happy, sad, surprise, neutral) の感情値を抽出する。抽出した感情の内 happy のみを入力特徴量として機械学習分類モデルを構築し、テストデータとして用いる各フレームの体調を推定し、動画クリップごとに最頻クラスを取る。機械学習モデルの構築と体調推定は個人ごとに行い、複数人のデータを用いて機械学習モデルを構築して行わない。この手法によって体調を推定することができるかを調べるため、体調ごとの表情演技を撮影した動画クリップを収集して機械学習分類モデルを構築して体調を推定する実験を行ったところ、体調を推定できることが示されている。

その後、特徴量とモデル構築の人数を比較することで体調分類モデルの精度を上げることを目的として手法の改良と評価を行っている [10]。抽出した感情値に対して主成分分析を行い、第一、第二主成分を特徴量に三段階の体調を正解データとして機械学習分類モデルを構築する。ここで happy のみの場合と比較を行うこと、そして個人ごとにモデルを作成する必要があるかどうか、複数人のデータを用いて機械学習モデルを構築した場合と比較を行い、特徴量には七つの感情値に主成分分析を適用して第一、第二主成分を用いて、被験者ごとに個別に構築する方が高精度であることが示されている。また主成分分析により、第一、第二主成分のどちらかに happy のみが大きな正の負荷量を持つ主成分が現れ、もう片方に angry, disgust, fear, sad が大きい正の負荷量を持つ主成分が現れる特徴があることが示唆されている。

しかし、先行研究では実験環境下での動画クリップ収集と分析にとどまっており、長期的なスコープで挙げられている VA 傾聴サービスとの連携部分が未実装である。VA 傾聴サービスと連携して動画クリップを収集し、表情による体調推定を行うフローを考え実装し、実際に VA 傾聴サービスを通してデータ収集する方法を確立する必要がある。体調推定手法にある機械学習分類モデル作成のためには、リアルタイムに動画を収集してその動画に対して体調ラベルを付ける方法を構築しなければならない。そして表情から三段階の体調に分類するのみでは具体的な好不調の理由を知ることができないため、推定された体調に基づいてより詳しい状況を聞く必要がある。

そこで本研究では以下のリサーチクエスチョンを設定する。

RQ1：日常の対話の中で表情と主観的体調の収集を行い、その結果に基づいて深掘りするシステムをいかにして構築するか

RQ2：実際の居住環境でデータ収集を行った場合においてど

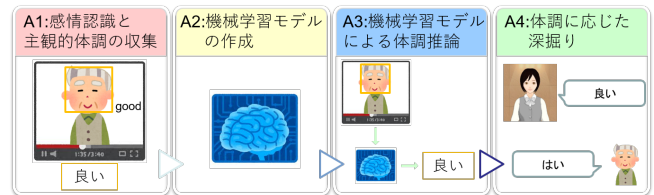


図1 アプローチの全体図

のような課題が生じるか

3. 提案手法

3.1 目的とキーアイデア

本研究の目的は対話中の表情感情から本人の主観的な体調を推論することで体調の把握と深掘りを行うサービスを提案することである。キーアイデアは一日のうちで初めて会ったときにその時の体調を聞く会話を開始すること、会話中に表情感情を蓄積して体調推論モデルをバッチ学習で構築すること、推論された体調を対話に組み込みより詳細な状況を聞き出すことである。バッチ学習の際には収集期間としてデータを集める期間を設け、後に推論期間に移行して詳細の深掘りを始める。目的を達成するためのアプローチとして以下の四つを提案する。

- A1：感情認識と主観的体調の収集
- A2：機械学習モデルの作成
- A3：機械学習モデルによる体調推論
- A4：体調に応じた深掘り

アプローチの全体図を図1に示す。A1では対話エージェントとの会話中に動画クリップを撮影して感情認識を行うこと、そして体調を何う問いかけを行うことで表情とその時の体調のデータを収集する。A2では収集した感情を特徴量に、体調を正解データとして機械学習分類モデルを作成する。A3ではA1のアプローチと同様にしてデータを収集しつつ、表情感情から機械学習分類モデルによって体調を推論する。A4では推論された体調を基に LLM が発話を生成し、体調の深掘りを行う。

本研究のアプローチでは、基本的なデータ取得・特徴抽出・フレーム単位分類と動画全体の最頻クラスによるラベル決定という流れは、筆者らの先行研究「表情による感情認識を用いた体調推定手法の開発と評価」[10]を基盤として採用する。

3.2 A1: 感情認識と主観的体調の収集

A1ではまず収集期間を設けて、体調を推論する機械学習分類モデル構築のためのデータとして表情の感情認識と主観的体調の収集を行う。先行研究であるVA傾聴サービスで非言語的な情報が十分に活用されていないという課題に対するアプローチとして表情感情を扱う。正解データとする高齢者の体調を収集するために、対話中に高齢者の主観的な体調を対話エージェントが伺い収集する。

会話の様子を図2と図3に示す。一日の内で高齢者が初めて対話エージェントと話した際に、まずは10秒間高齢者の顔をビデオ撮影する。ビデオクリップの約300フレーム(30fps×10秒)に対して感情認識モデルによる推論で七つの感情の確度を感情値ベクトルとして抽出する。その間、図2のように会話

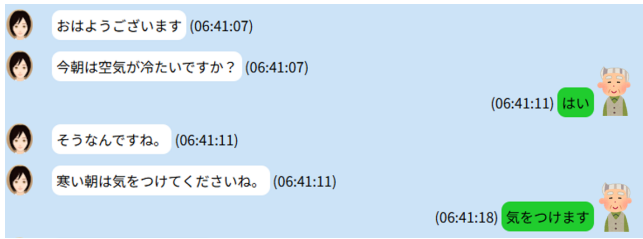


図2 感情認識中の会話の様子

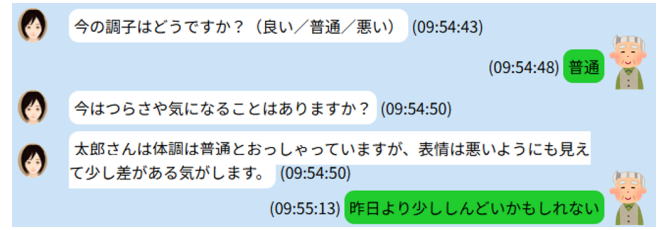


図4 体調推論後に確認する様子

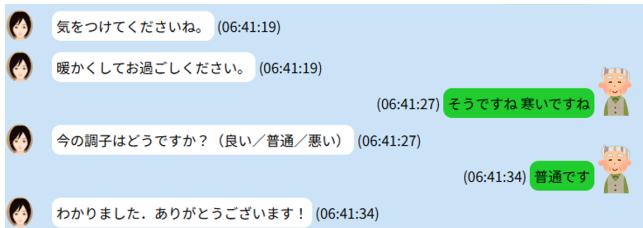


図3 高齢者が体調を申告する様子

を進行する。高齢者の表情感情が取得できた後、図3のような会話で対話エージェントが高齢者に対して体調を聞く。体調は良い、普通、悪いの三段階をとる。高齢者の申告した体調を10秒間のクリップ及びその各フレームの正解データとして、感情値ベクトルと共に保存する。瞬間的な表情の変化や認識エラーの影響を排除し、安定した体調推論を行うために、10秒を設定した。このようにして対話エージェントとの会話中に表情からデータを収集する。

3.3 A2: 機械学習モデルの作成

推論期間に高齢者の表情から体調を推論して詳細を伺う対話につなげるために、対話エージェントを通じて収集した表情感情と申告された体調から、高齢者の体調を推論するための機械学習分類モデルを作成する。感情値ベクトルに主成分分析を行いその第一、第二主成分を入力特徴量に、そして高齢者の主観的体調を正解データとしてモデル作成を行う。この際に主成分負荷量を保存することで、以降の推論プロセスでも同様の処理を適用する。良い、普通、悪いのそれぞれの体調の表情を収めた動画クリップが収集できた時点で、各フレームの体調を推論する機械学習分類モデルを作成する。モデルの構築は個人ごとに行い、他の人のデータを用いて機械学習モデルを構築しない。

主成分分析を行う理由は二つある。一つは実運用において一人から収集できるデータには限りがあるため、次元削減を行いモデルの汎化性能を高めることである。もう一つは先行研究において第一、第二主成分のどちらかがhappyのみ、もう片方にangry, disgust, fear, sadが大きい正の負荷量を持つというポジティブとネガティブの分離が見られたため、推論結果に対する説明性を高める効果があるためである。

また、機械学習分類モデルには説明可能性を重視して決定木を採用する。

3.4 A3: 機械学習モデルによる体調推論

機械学習モデル作成が完了した日以降を推論期間として、対話エージェントとの会話中に撮影した動画クリップ内の高齢者の体調を、機械学習分類モデルを使ってフレームの推論結果の

最頻クラスとして求める。発話に伴う一時的な表情変化や、瞬きや顔の角度のわずかな変化を特定の感情として評価してしまうようなノイズの影響を抑えるために、10秒間の最頻クラスを用いることで、瞬間的変動へのロバスト性を担保する。動画クリップの各フレームに感情認識を行い、それらを入力特徴量として機械学習分類モデルで体調の分類を行う。クリップ内全てのフレームの推論結果の最頻クラスを動画クリップの体調として推論する。A1と同様に図2のような一日の内で高齢者が初めて対話エージェントと話した際に、まずは10秒間高齢者の顔をビデオ撮影し、正解データの保存まで行う。体調推論の完了までの時間は対話エージェントとの会話を継続し、その裏側で主成分分析と機械学習モデルによる分類を行うことで、推論に伴う体感待ち時間を低減する。モデル作成後は体調を伺う前に高齢者の体調を推し量ることができ、エージェントの会話につなげる。

3.5 A4: 体調に応じた深掘り

推論した結果から高齢者の体調が良い、悪い理由を知るために、推論した体調を基にその体調に関する深掘りを対話エージェントが行う。表情から体調を三段階の内から推論するだけでは、具体的に高齢者がどう感じているのか詳細を読み解くことができない。そのためには対話エージェントがその詳細を伺う必要があり、推論された体調を基にした発話内容が必要である。高齢者の推論された体調に対してどんな時に感じたか、生活の中で思い当たることがあったかエピソード想起を促す質問を行うようプロンプトを設計し生成し、LLMに入力して発話を生成する。生成した発話を対話エージェントに入力する。

会話の様子を図4と図5に示す。図4のようにA3の会話中に申告された体調を復唱し、表情から推論した体調を提示して確認を取る。推論体調の提示とそれに対する確認応答を用い、LLMに入力して対話で採用する最終的な体調を決定する。これは推論結果が間違っただけで高齢者と会話を行い負担をかけることを防ぐための安全策である。それから体調を基にした深掘りを行う。深掘りを行っている様子が図5である。このようにして対話エージェントが会話と表情感情を通じて非侵襲的に体調推論とその深掘り、見守りを行う。

4. 実装

4.1 利用した技術

本研究の実装には以下の技術を用いた。

- 開発言語: Python, JavaScript (Node.js)
- バックエンドフレームワーク: FastAPI, Express

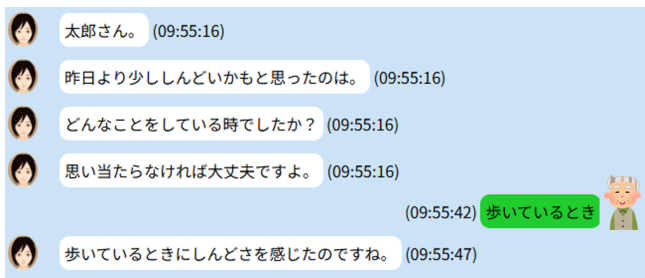


図5 深掘りの会話の様子

- 使用したライブラリやツール: scikit-learn, PyAV, OpenCV, LangChain, OpenAI API(GPT-5.2), face-api.js

- 外部データベース: MySQL
- ハードウェア: Lenovo IdeaPad Flex 5 15ITL05 (内蔵カメラ), Logicool C270 HD (Web カメラ)

システム構成はマイクロサービスアーキテクチャを採用し、動画処理と対話管理を行う Python(FastAPI) バックエンドと、顔認識処理に特化した Node.js(Express) バックエンドで構成した。動画処理、感情認識、体調推論という計算不可の高い処理の内、感情認識を分離して負荷分散を図った。感情認識モデルには軽量かつ顔の検知から感情認識までの実装が容易であることから face-api.js を採用した。具体的には、オリジナルの face-api.js [11] のフォーク版であり、最新の TensorFlow.js に対応してメンテナンスが継続されている @vladmandic/face-api [12] を使用した。機械学習分類モデルには scikit-learn の DecisionTreeClassifier を用いた。

4.2 A1の実装

10秒間の動画クリップを撮影し、WebM形式でFasiAPIを用いたバックエンドサーバに送信される。受信した動画データはPyAVライブラリを用いてデコードされる。続いて各フレームをnode.js環境の感情認識へPOSTし、face-api.jsを利用して動画クリップ中の約300フレームから七つの感情クラスの確信度を抽出する。なお、顔が検知できなかった場合に3回まではクリップを再度取り直す。

4.3 A2の実装

収集期間を通して良い、普通、悪いそれぞれの動画クリップが最低一つずつ蓄積された時点で、学習プロセスがトリガーされる。クラス間のデータ不均衡に対処するため、class_weight='balanced'を設定している。モデルの最適化にはグリッドサーチを用い、決定木の深さ(max_depth: 1-5)および分割に必要な最小サンプル数(min_samples_split: 2, 10, 20, 30)の組み合わせに対して、6分割交差検証を実施し、最も汎化性能の高いパラメータを選定する。

4.4 A3の実装

A1と同様に動画の撮影から感情特徴量の抽出までが行われる。得られた各フレームの感情値ベクトルに対し、A2で学習した主成分負荷量を適用して次元圧縮を行った後、決定木モデルに入力してフレームごとの体調クラスを推論する。最終的な動画クリップに対する体調推論は、全フレームに対する推論結果の最頻クラスを採用し、データベースに保存される。

4.5 A4の実装

体調確認以降の対話生成にはGPT-5.2モデルを使用し、LangChainフレームワークを用いて実装した。プロンプトには、対話エージェントが伺った発話内容(主観的体調)と体調推論の結果の両方を組み込み、「良い」「普通」「悪い」のいずれれとして会話を進めるかを出力するよう指示している。LangChainのPydanticOutputParserを用いることで、LLMの出力を所定の形式に厳格に従わせ、予期しない出力形式によるエラーを防いでいる。

5. 予備的試用と考察

5.1 開発者および高齢者による予備的試用

本研究ではA1からA4までを実装済みであるが、本試用では実環境における学習データの蓄積特性を把握するため、A1における運用評価に焦点を当て、開発者自身(20代男性)および高齢者1名(70代男性、以下「被験者A」とする)を対象に、1月30日から2月17日まで19日間の予備的試用を各々の居住環境において実施した。

手順として、毎日の最初の会話においてA1(表情・体調ラベル収集)を実行した。具体的には、対話エージェントとの挨拶、雑談中に10秒間の動画クリップから感情認識値を抽出するとともに、対話を通じた主観的な体調申告(良い/普通/悪い)を紐付けてデータベースに記録した。なお、被験者Aについては生活習慣上、毎朝最初での使用が困難であったため、任意のタイミングで特定の起動キーワードを発話することでシステムを起動する運用とした。

5.2 結果

期間中のデータ収集試行回数は開発者が14回、被験者Aが17回であり、そのうち顔検知および感情抽出に成功したデータ数は開発者が14件、被験者Aが6件であった。計20件のうち、システムからの問いかけに対して体調ラベルが付与された有効データは19件であり、ラベル欠損は1件のみであった。ラベル欠損は三段階の体調(良い/普通/悪い)以外の表現で体調を申告したことによるものであった。

有効データについて、処理時間を次の区間として計測した。開始時刻は、対話エージェントの初回発話時刻(または起動キーワード認識の通知発話時刻)であり、いずれもデータベースに保存されたタイムスタンプを用いた。終了時刻は、感情認識処理の完了後に体調を伺う声掛けを行った時刻とした。その結果、開発者環境では平均26.0秒(SD=4.0)、被験者A環境では平均46.2秒(SD=12.1)であった。開発者と被験者の間で時間に差が生じた要因として、ネットワーク環境の差異に加え、開発者環境で撮影された動画クリップのフレームレートが被験者A環境のおよそ半分であったことが考えられる。

試用の結果、有効データが得られたケースにおいては、VA傾聴サービスを通じたデータ収集は自然に行われ、ユーザに「待たされている」という印象を与えずに済むことが確認された。平均46.2秒という処理時間はシステム単体で見れば遅延といえるが、エージェントとの挨拶や雑談といった対話インタラクションによって吸収されることで、ユーザに監視や測定といっ

た心理的負担を与えずにデータの蓄積が可能であることが示された。

一方で被験者 A が収集を試行した内で有効データにならなかった原因の内訳は、キーワードに関するミスが 2 件、顔検知エラー件数が 5 件、実験準備のミスが 4 件あった。特に顔検知エラーが発生した際の画像を観察したところ、室内の天井照明が強く、顔領域に光が重なって撮影されている事例（ハレーション）が確認された。この場合に 10 秒間で 1 フレームも顔が検出できておらず、家庭内の照明環境によっては顔検出が失敗し、10 秒間のクリップ内で有効なフレームを確保できない可能性が示唆された。被験者 A は早朝にシステムを利用する傾向があり、その際に室内の天井照明の明るさと自然光の明るさとの差によって顔検知エラーが発生していたと考える。これは実験室環境では顕在化しなかった問題であった。

こうした顔検知エラーにより、高齢者は体調を伺う発話までたどり着かないことで途中でシステムを終了してしまう。顔検知に関する課題は A2 以降のアプローチへの移行に影響を及ぼすほか、各ラベルのフレーム数が偏ることでモデルの精度低下にもつながる可能性がある。

他には今回の試用期間中、二名の健康状態に異常がなかったため、正解ラベルとしての「悪い」が一度も報告されなかった。内訳は開発者はすべて普通、被験者 A は普通が 4 件、良いが 1 件であった。その結果、三クラスが揃うことをトリガーとする A2 の学習以降のプロセスへ自動的に遷移することが困難であった。収集期間に一度でも良い、普通、悪い体調だと感じる日がなければならず、この期間が長くなりすぎる可能性を感じた。実環境において個人適応型モデルを運用する場合、個人のデータのみ依存すると、全ラベルが揃うまでの収集期間が長期化する可能性がある。

5.3 考 察

本試用を通じて、提案手法の利点としては対話エージェントとの自然な会話の中に感情データの取得プロセスを埋め込むことで、ユーザに監視や測定といった心理的負担を与えずに継続的な見守りが可能となる点にある。有効データが得られたケースでは、処理待ち時間が対話によって吸収され、スムーズなデータ収集が実現できていた。また体調に応じて深掘りを行えること、器具の装着などの必要がなく、継続的に表情も含めた見守りが可能となることが挙げられる。

提案手法の限界としてはまず、実環境においてはカメラと顔との距離や角度、背景の光が強いと顔検知できない場合があることが挙げられる。試用において失敗する事例も確認されており、PC 内蔵カメラや Web カメラを用いる場合、設置場所やユーザの座る位置が固定されがちであり、照明環境の調整が困難な場合が多い。顔検知エラーによる対話の中断を防ぐためには、ユーザへの適切なフィードバックが必要である。試用においても対話画面上部に「お顔が見えにくいです。カメラに映るように調整してください。」と表示することで対応を図った。しかし、顔がどのように映っているか（カメラ映像の状態）が確認できた方が分かりやすい、との意見が得られた。監視の不快感を与えずに環境改善を促す UI の構築が求められる。他には、

現在は PC 環境に依存しているが、設置場所の制約を受けにくいスマートフォンやタブレット端末への拡張も考えられる。

また、先行研究 [10] では個人ごとのモデル構築が高精度であるとされたが、本試用により健康な高齢者の場合は「悪い」ラベルの収集に長期間を要することが明らかになった。この解決策として、運用初期段階においては他者のデータを用いて事前学習したモデルを利用し、データが蓄積され次第、段階的に個人モデルへ移行するハイブリッドな方法を検討するべきかもしれない。

6. ま と め

本研究では感情分析による体調推定とエージェント対話を通じた高齢者体調確認サービスの提案を行った。アプローチとしては「A1:感情認識と主観的体調の収集」、「A2:機械学習モデルの作成」、「A3:機械学習モデルによる体調推論」、「A4:体調に応じた深掘り」の四つである。この提案手法に基づき実装を行った。

予備的評価として開発者自身と高齢者一名が 2 週間サービスを使用して実環境下でのデータ収集の確立と課題抽出を行った。結果、対話エージェントとの対話中に体に負担をかけずに自然な形で表情感情のデータが取得できることが確認された。

今後の課題として顔検知に関するユーザビリティの向上と、機械学習分類モデルの作成方法を工夫すること、A2 から A4 に関する評価を行うことが挙げられる。

謝辞 本研究の一部は JSPS 科研費 JP25H01167, JP25K02946, JP25K24389, JP24K02765, JP24K02774, JP23K17006, JP23K28091, JP23K28383 の研究助成を受けて行われている。

文 献

- [1] 内閣府, “令和 6 年度版高齢社会白書,” https://www8.cao.go.jp/kourei/whitepaper/w-2024/zenbun/slash06pdf_index.html. accessed October 17, 2024.
- [2] 警察庁刑事捜査第一課, “警察取扱死体のうち、自宅において死亡した一人暮らしの者～令和 6 年上半年期 (1 6 月) 暫定値～,” https://www.npa.go.jp/publications/statistics/shitai/hitorigurashi/240827_kenshi2.pdf. accessed October 17, 2024.
- [3] 松川晃徳, K. Nahyun, 陳思楠, 片桐恵子, 中村匡秀, “音声対話エージェントが在宅高齢者の孤独に与える影響: 対話ログと孤独感尺度による分析,” 電子情報通信学会技術研究報告, 第 124 巻, pp.13–20, Nov. 2024.
- [4] 東京福祉保健局, “第 4 版住民の皆さんのための高齢者等の見守りガイドブック,” https://www.fukushi.metro.tokyo.lg.jp/kourei/koho/jyuuminnotameno.files/j_guidebook4.pdf. accessed October 17, 2024.
- [5] Y. Noguchi, H. Kamide, and F. Tanaka, “Personality traits for a social mediator robot encouraging elderly self-disclosure on loss experiences,” *J. Hum.-Robot Interact.*, vol.9, no.3, pp.●●–●●, May 2020. <https://doi.org/10.1145/3377342>
- [6] 大園隼人, 陳思楠, 佐賀雅樹, 中村匡秀, “音声対話エージェントとウェアラブルデバイスの連携による高齢者のセルフケア促進,” 電子情報通信学会技術研究報告; 信学技報, vol.121, no.336, pp.103–108, 2022.
- [7] 陳思楠, 大園隼人, 中村匡秀, “高齢者自助支援のための音声対話エージェントを用いたマルチモーダルダイアリーサービスの提案,” 電子情報通信学会技術研究報告; 信学技報, vol.121, no.437, pp.75–80, 2022.
- [8] 堀江寛, 陳思楠, 中村匡秀, 安田清, “動画を活用した在宅高齢

者のためのストレス解消サービスの研究,” 電子情報通信学会技術研究報告; 信学技報, vol.121, no.381, pp.1-6, 2022.

- [9] 西山敦紀, 中田匠哉, 陳思楠, 佐伯幸郎, 中村匡秀, “表情による感情認識を用いた体調推定手法の検討,” 電子情報通信学会技術研究報告, 第 124 巻, pp.56-64, Nov. 2024.
- [10] 西山敦紀, 中田匠哉, 陳思楠, 佐伯幸郎, 中村匡秀, “表情による感情認識を用いた体調推定手法の開発と評価,” 電子情報通信学会技術研究報告, 第 124 巻, pp.77-84, Match 2025.
- [11] V. Mühler, “face-api.js: JavaScript API for face recognition in the browser and Node.js with TensorFlow/JS,” <https://github.com/justadudewhohacks/face-api.js>, 2018.
- [12] V. Mandic, “face-api.js: JavaScript API for face recognition in the browser and Node.js with TensorFlow/JS,” <https://github.com/vladmandic/face-api>, 2024.