

表情による感情認識を用いた体調推定手法の検討

西山 敦紀[†] 中田 匠哉[†] 陳 思楠[†] 佐伯 幸郎^{††} 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

^{†††} 理化学研究所・革新知能統合センター 〒103-0027 東京都中央区日本橋 1-4-1

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, ^{†††††}masa-n@cmds.kobe-u.ac.jp

あらまし 近年、高齢者の人口は増加しており、一人暮らしの独居高齢者の人口と割合も増加傾向にある。一人暮らしの高齢者は社会的に孤立しやすく、病気やけがの発見が遅れると孤独死につながる可能性が高い。そのような高齢者の日々の体調を把握することは大切であるが、既存の安否確認や見守りサービスでは高齢者が自ら体調不良を伝えるしかないという問題点が存在する。本研究では既存の感情認識の学習済みモデルを利用し、体調別の演技をした表情を撮影した画像の感情推論結果を取得して分析する。そして感情の推論結果を特徴量に、演技した体調を教師データとして機械学習で分類することができるか実験を行う。実験より個人に適応したモデルを作成することで表情を分類できる。

キーワード 高齢者見守り, 表情, 感情認識, 体調, 機械学習

Estimating Health Condition Using Facial Emotion Recognition

Atsunori NISHIYAMA[†], Takuya NAKATA[†], Sinan CHEN[†], Sachio SAIKI^{††}, 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

^{†††} Riken AIP 1-4-1 Nihon-bashi, Chuo-ku, Tokyo 103-0027 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, ^{†††††}masa-n@cmds.kobe-u.ac.jp

Abstract In recent years, the population of the elderly has been increasing, and the population and percentage of the elderly living alone are also on the rise. Elderly people who live alone are likely to be socially isolated, and if the detection of illness or injury is delayed, it is likely to lead to a solitary death. It is important to understand the daily health condition of such elderly people, but existing safety confirmation and monitoring services have the problem that elderly people have no choice but to report their health condition by themselves. In this study, we use an existing emotion recognition model to acquire and analyze emotion inference results from images of facial expressions of people acting according to their health condition. Then, we conduct an experiment to see if machine learning can classify the images using the emotion inference results as feature values and the health condition of the person as the teacher data. Based on the experiments, it is possible to classify facial expressions by creating a model that is adapted to the individual.

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

1. はじめに

近年、高齢者の人口は増加しており、中でも一人暮らしで生活を送る独居高齢者の割合も増加している [1]。そうした中で高齢者自身が孤立死を身近に感じており、実際に自宅で死

亡した一人暮らしの高齢者は多くおり [2]、孤立死と考えられる事例が数多くある。このような背景から高齢者とコミュニケーションを取ることや見守りを行うことは大切であり、訪問による見守りやデジタルツールによる見守りサービスが行われている [3]。しかし、担い手不足や高齢者自身が体調不良に気が付く

えるしかないシステムなど課題もある。このような課題があり日々の体調について声掛けをしてくれるサービスや体調を記録しておく必要があると考える。そこで研究の長期的なゴールを対話エージェントの声掛けによる日常的な体調確認を行い支援につなげることと定め、本研究では表情から体調を推定できるかを調べることをスコープとする。

本研究では短時間での対話中の表情から体調を推定する手法を提案し、体調に合わせた声掛けにつなげることを目的とする。キーマイデアは既存の感情認識の学習済みモデルを活用して表情から感情を推論し、表情による感情認識を特徴量とする機械学習分類モデルを構築すること、そして表情の学習にはエージェントとの対話を想定した短い動画クリップを使うことである。目的を達成するために以下の手法を提案する。

A1：被験者が体調に応じた演技を行った動画クリップを収集する。

A2：動画クリップをフレームを抽出し、フレームに対して感情認識を適用することで感情値を抽出する。

A3：抽出した感情値から特徴量を作成し、フレームの正解データとセットにする。

A4：フレームの特徴量と正解データを集めて機械学習分類モデルを構築する。

A5：機械学習モデルで動画クリップ中のフレームの体調を推定し、多数決によってクリップ全体の体調を推定する。

この手法を基に実際に動画クリップを収集して機械学習モデルによりクリップ全体の体調ラベルの推定を行う。これにより被験者ごとに機械学習モデルを作成し、フレームの体調推定結果を基にした多数決を行うことで動画クリップ全体の体調ラベルを推定することができることが分かった。

2. 準備

2.1 独居高齢者の増加

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。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、気付かれない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なため、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。本研究ではこの日々のコミュニケーションと見守りを行うことに注目する。

2.3 現状の課題

高齢者の見守りに活用される見守りサービスには様々な種類の見守り方がある。総務省行政評価局の調査[3]によると訪問による見守り活動は、民生委員、社会福祉協議会、地域包括支援センター等が、対面により、生活の状況を把握する取組を行っている。課題として民生委員の担い手確保に苦慮するなど、見守り活動を行う人材の確保が難しくなっている状況がみられたとある。高齢者宅に個別に訪問するには工夫が必要であり、訪問する頻度によっては異変に気付くことができないと考えられる。

またデジタルツールの活用による見守り活動についても挙げられている。センサーを通して高齢者の異変を察知した際、センサー設置事業者へ連絡が入り、安否確認を行うものや[5]、ロボットやタブレット端末を通じて本人の姿の確認や会話を行うことができるもの[6][7]、利用者が電話で健康状態を発信し、見守りセンターがその状況を確認するもの[8]、緊急通報装置のボタンを自ら押すだけで、緊急放送装置の貸与事業者等との通話ができるものが挙げられている[9][3]。挙げられた課題としてはデジタルツールで異変を察知した際、誰が見守り対象者の自宅に向いて安否確認をするかということについて苦慮している地方公共団体がみられたとあり、またタッチパネルを操作してその日の体調を回答するケースでは、操作を面倒に感じる人もいるということがある。このことからデジタルツールを活用することにより見守る時間を増やすことができるが、詳細な体調を伝えようとすると高齢者が自ら体調不良に気づき伝えるしかないことが課題でもあると考える。

また我々の研究グループでは対話エージェントを用いた在宅高齢者の見守りを行うサービスの研究[10]が行われているが、非言語的な情報は活用できていない。このような課題から日々体調を伺う声掛けを行い、非言語的な情報も活用して体調を把握してくれるサービスが必要であり、そして体調を把握するだけでなく記録しておく必要がある。

2.4 長期的なゴールと本研究のスコープ

現状の課題を踏まえた研究の長期的なゴールを、対話エージェントの声掛けによる日常的な体調確認を行い支援につなげ

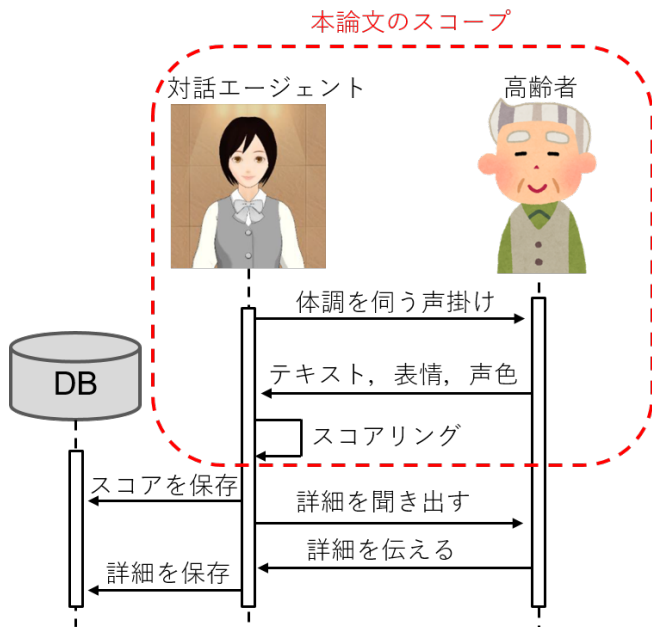


図1 長期的なゴールの全体図

ることとする。図1に長期的なゴールの全体図を示す。対話エージェントが高齢者に対して日々の声掛けを行い、それに対して高齢者が答えた際のテキスト及び非言語的な情報である表情と声色を取得する。取得した情報を基に高齢者の体調を分析し、スコアとして算出する。算出した体調スコアが悪ければ対話エージェントが詳細な不調の内容を聞き出し、支援につなげる。また算出した体調スコアは日々保存することで日を追うごとにスコアの推移が悪くなっていれば支援が必要だということがわかり、支援につなげることができる。図1の長期的なゴールの内、枠で囲まれた部分、すなわち表情から体調を推定できるか調べることを本研究のスコープとする。

2.5 関連研究

川村らは発話時の表情変化に着目した疲労推定の有効性を検証し、モーションキャプチャによる表情の測定と得られた特徴量を用いてSVMによる疲労の推定を行ない、標準顔よりも発声動作時の表情を基にしたほうが疲労の認識精度が高いことを示した[11]。Ghoshらは痛みの表情を機械学習で分類する研究を行い、SVMとCNNのスコアの積を取るシステムによって痛みの表情を分類し、競合する手法よりも提案手法が優位であることを示した[12]。ZhouらはMCIの高齢者のうつ病とアパシーをテキスト、音声、ビデオを活用して分類する研究を行い、アパシーとうつ病が共通した言語的特徴や顔の特徴を有していること、鑑別診断に使用可能な特徴も有していることや、うつ病とアパシーの併存が新たな特徴を示すことを発見した[13]。佐古らはケア受診中の対象者の表情から感情を分析し、ケアの効果を対象者の感情の変化として計測、評価するシステムの提案をしている[14]。

このように表情を活用して痛みや精神疲労など特定の状態を分類する研究は行われているが、体調について日常的に計測して記録し、支援に役立てるところまで踏み込まれておらず、結果から対話によって聞き出す研究は行われていない。非言語的

な情報も活用した日常的な体調把握と記録を行うために、本研究のスコープは表情から推論される感情から体調を推定できるかを調べることにする。

3. 提案手法

3.1 目的とキーアイデア

本研究では短時間での対話中の表情から体調を推定する手法を提案し、体調に合わせた声掛けにつなげることを目的とする。キーアイデアは既存の感情認識の学習済みモデルを活用して表情から感情を推論し、表情による感情認識を特徴量とする機械学習分類モデルを構築すること、そして表情の学習にはエージェントとの対話を想定し、体調が良い場合、普通の場合、悪い場合の3通りの短い動画クリップを使うことである。推論される感情はangry, happy, sad, fear, disgust, surprise, neutralの7つである。目的を達成するために以下の5つのアプローチを提案する。

A1: 体調に応じた動画クリップを収集

A2: フレームを抽出し、感情認識を適用して感情値を抽出

A3: 感情値から特徴量を作成

A4: 機械学習モデルを構築

A5: 多数決によってクリップ全体の体調ラベルを推論

アプローチの全体図を図2に示す。図2の左上側の枠で囲まれたA1において高齢者の表情を撮影して3段階の体調の動画クリップを1つずつ取得する。上側の枠で囲まれたA2において動画クリップから複数のフレームを抽出し、フレームごとに感情認識を適用することで感情値を抽出する。右側の枠で囲まれたA3において抽出した感情値から特徴量を作成し、フレーム抽出元の動画クリップの体調を正解データとしてセットにして学習に用いる体調データを作成する。右下側の枠で囲まれたA4において3つの動画クリップの半分の体調データを訓練データとして学習に用いて機械学習分類モデルを作成する。左下側の枠で囲まれたA5においてテストデータとして残り半分の体調データを用いてフレームごとに体調を推定し、同じ動画クリップから得られたフレームを集めて、推定したフレームの体調の多数決を取ることで動画クリップ全体の体調を推定する。アプローチの動画クリップ撮影から機械学習モデルの作成、動画クリップの体調推定までは全て個人ごとに行い、他の人の動画クリップを学習データに用いることや機械学習モデルで体調を推定することはしない。すなわち個人の動画クリップから個別の機械学習モデルを作成し、その人の機械学習モデルによって体調を推定する。

3.2 A1: 体調に応じた動画クリップを収集

体調を推定するためのデータを集めるために、まずは体調に応じた動画クリップを収集する。動画クリップを収集する際に体調は良い、普通、悪いの3段階を想定する。それぞれ良いは元気であり笑顔であること、普通は極端に笑顔であったりせずかしこまった状態、悪いは風邪や病気であるときの体調やメンタルが不安定な時を想定する。想定した表情の例は図3に示し、左から良い、普通、悪いを想定した表情である。

図2の左上側の枠で囲まれた部分のアプローチであり、体調

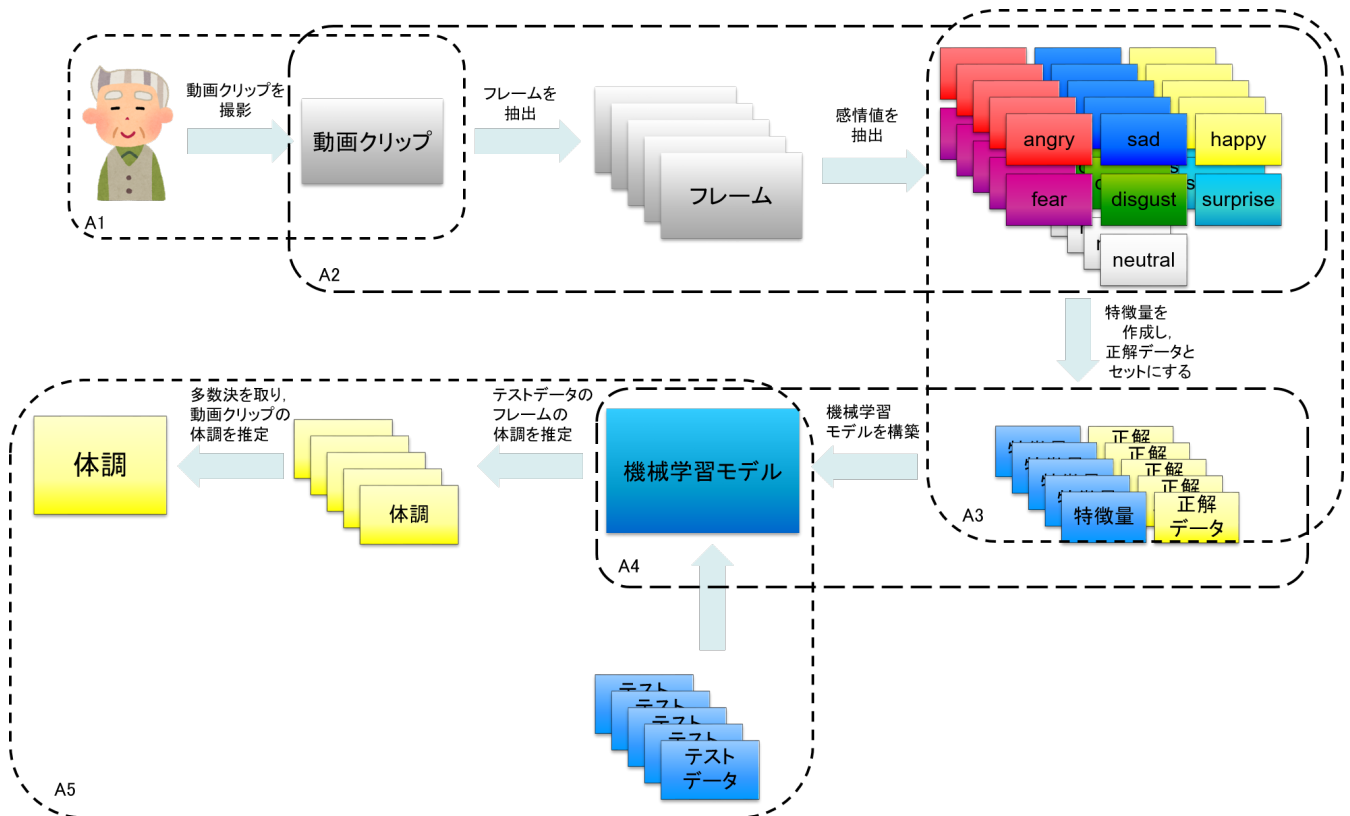


図2 アプローチの全体図

に応じた表情をした10秒間の動画クリップを収集する。長期的なゴールにおいて対話エージェントによる声掛けとそれに答えるシチュエーションを想定しているため、収集する動画クリップも会話を想定したものとし、体調を伺う声掛けから始まりそれに対して答えるまでを含める。体調には3段階を想定しているため、体調ごとに1つの動画クリップを撮影することになり、1人当たり10秒の動画を3本収集する。収集した3本の動画クリップは機械学習モデルを用いて体調を推定するための学習データに用いるという目的がある。

3.3 A2：フレームを抽出し、感情認識を適用して感情値を抽出

収集した動画クリップの表情から感情認識モデルの推論によって感情値を抽出するために、まずは動画クリップから短い時間ごとにフレームを抽出する。図2の上側の枠で囲まれた部分のアプローチであり、フレームは動画クリップを撮影した画像の集まりとした中の1つの瞬間であるため、1つの動画クリップから多数のフレームを取得することができる。表情による感情認識では一般に用いられる7種類の感情(angry, happy, sad, fear, disgust, surprise, neutral)の確度(confidence)を数値として算出し、本論文ではこれを感情値と呼ぶ。現在様々な感情認識AIが存在するが、それを有効活用する。そして得られたすべてのフレームに対して感情認識モデルで推論することで感情値を取得する。感情の推論をフレームごとに行うため、感情値はフレームごとに異なった値が得られる。

例えば図3の静止画はそれぞれの体調を想定した動画クリップから抽出したものであるが、それぞれフレームとして感情認

識モデルを適用すると静止画の下のような感情値ベクトルが得られる。括弧の中は順に angry, disgust, fear, happy, sad, surprise, neutral となっており、良い場合のフレームの happy の値は 0.961356 であると読み取れる。フレームごとに得られた感情値を動画クリップの感情値の推移としてグラフに表したものが図4である。図4では実線が happy のグラフである。感情値を取得してグラフを作成することで機械学習モデルを構築する際の特徴量の決定に役立てるという目的があり、フレームから取得した感情値は体調を推定する機械学習の特徴量に用いるという目的がある。

3.4 A3：感情値から特徴量を作成

表情から体調を推定する際の学習データを作成して機械学習モデルを構築するために、特徴量に使う感情を選択し、正解データとセットにする。図2の右側の枠で囲まれた部分のアプローチであり、抽出した感情値の中から機械学習分類モデルの学習に使う特徴量を選ぶ。本研究では図4を見て良い場合に特に happy の感情値が大きく、普通と悪い場合で小さいことから happy の値が体調に関係があると考え、特徴量として happy のみを用いる。そしてフレームから選んだ特徴量と動画クリップを収集する際に想定した3段階の体調をフレームの正解データとしてセットにしてフレームの体調データとする。

例えば良い場合の動画クリップから得られた感情値の推移が図4の良い場合のグラフであるとして、この中の各フレームの happy を正解データ(良い)とセットにする。そうすると表1の左と真ん中のセットが得られ、happy の感情値が 0.965837 の場合の正解が良いであるとわかる。これを普通の場合と悪い場合

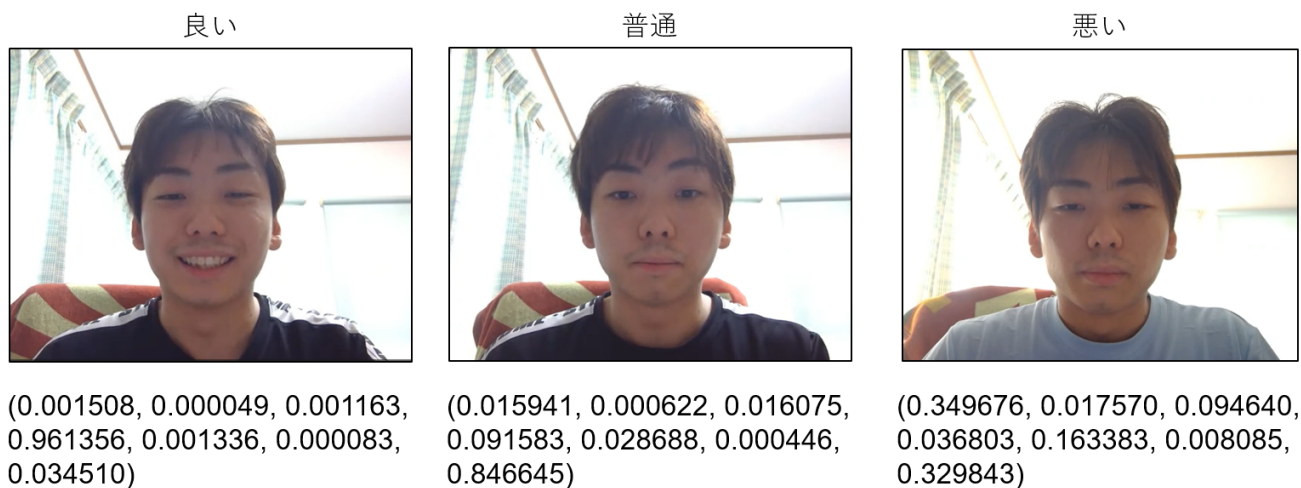


図3 想定した表情と感情値の例

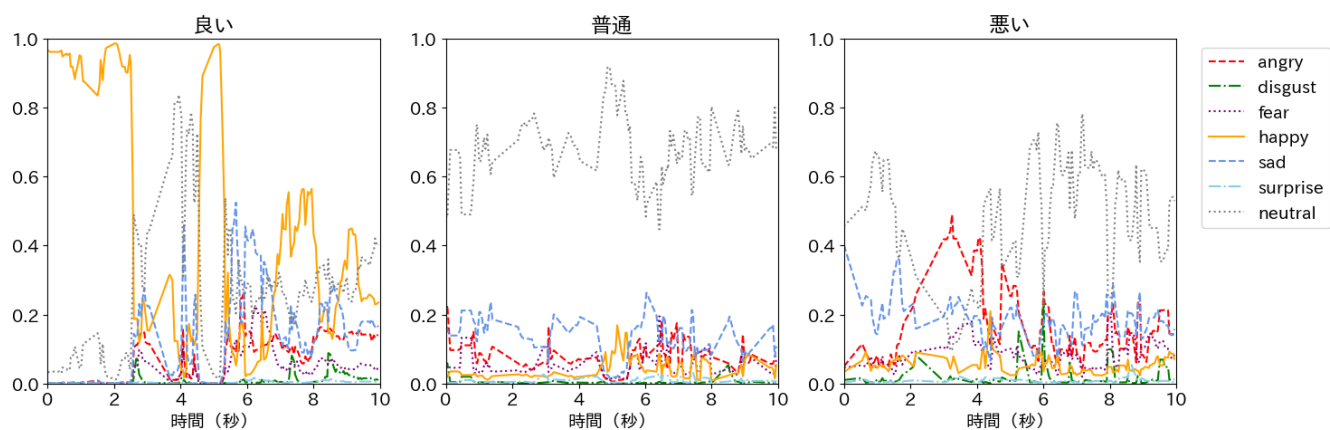


図4 想定した表情から抽出される感情値の推移

表1 良いフレームの happy の確度と推定結果の例

happy	正解データ	推定結果
0.965837	良い	良い
0.961356	良い	良い
0.961356	良い	良い
0.961356	良い	良い
0.961356	良い	良い
...	...	...
0.131833	良い	普通
0.178858	良い	普通
0.178858	良い	普通
0.142448	良い	普通
0.020623	良い	悪い

にも行い、それぞれ happy の感情値と普通、happy の感情値と悪い、という各フレームの体調データが得られる。この happy と正解データのセットは動画クリップから得られた学習データとして機械学習に用いるという目的がある。

3.5 A4: 機械学習モデルを構築

フレームごとの体調を推定するために、体調データを用いてフレームの体調を推定する機械学習モデルを構築する。図2の右下側の枠で囲まれた部分のアプローチであり、まず体調データを訓練データとテストデータに半分に分ける。分けたデータの内訓練データを用いてフレームを良い、普通、悪いの3段階の体調に分類する機械学習モデルを構築する。ここで使用する特徴量と正解データは同じ人物の動画クリップのみとして、他の人物の体調データは使用しない。すなわち機械学習モデルは体調を推定する人ごとに個別に構築する。

例えば図3の人物の機械学習モデルを構築するためには、まず3本の動画クリップから得られた図4の感情値について表1の左と真ん中のセットのような体調データを普通と悪いについても用意する。例に用いる3本の動画クリップからは441フレームが得られたので、用意した体調データを半分に分けて訓練データとテストデータにすると、正解データが良い、普通、悪いのものが訓練データにもテストデータにも含まれる。そして221フレームの訓練データを用いて機械学習モデルを構築す

ると、動画クリップではなく、その中の各フレームがどの体調に分類されるかを推定するモデルが得られる。構築した機械学習モデルは動画クリップの体調を多数決で決めるための、各フレームの体調を推定する目的で使われる。

3.6 A5：多数決によってクリップ全体の体調ラベルを推定

短時間の対話中の表情から体調を推定するために、構築した機械学習モデルを用いてテストデータの体調を推定し、多数決によって動画クリップの体調を推定する。図2の左下側の枠で囲まれた部分のアプローチであり、テストデータのフレームの体調を推定し、対話をしている動画クリップ中の体調を推定するために、機械学習モデルによって推定された体調を集めて多数決を取る。同じ動画クリップのフレームから推定した体調を集めて多数決を取ることで、動画クリップ中に最も多く表れたと推定される体調を良い、普通、悪いの中から選択し、動画クリップ全体の体調とする。

例えばこれまでのアプローチで例にした人物について、3本の動画クリップから得られた441フレームの内テストデータは220フレームである。その中から良い場合を例にとるとテストデータは88フレームあり、これを推定すると表1のような推定結果が得られる。結果として

良い：63 フレーム

普通：18 フレーム

悪い：7 フレーム

という内訳になると、多数決によりテストデータの元の動画クリップ全体の体調は良いという推定結果が得られる。このように多数決により動画クリップの体調を推定することで、短時間の対話中の表情から体調を推定するという目的がある。

4. 実験

提案手法を評価するために実験を行った。

4.1 実験の概要

まず動画クリップの収集を行うために、学習済みモデルを使用したアプリケーションのプログラムはサンプルアプリ[16]を参考にして、感情値を抽出するアプリケーションの作成と感情値をCSVファイルに保存する機能の追加にはJavaScriptを用いた。学習済み感情認識モデル[15]はFER-2013 emotion datasetで66%の精度があり、7つの感情値は足して1になるように算出される。次に感情値から体調を推定する機械学習分類モデルの構築にはPythonを使用し、教師あり学習の決定木モデルを採用した。機械学習分類モデルと提案手法のそれぞれの評価にはPythonを使用した。

実験の被験者には研究者とその家族及び我々の研究室のメンバーに協力していただき、10代女性1名、20代男性9名、50代男性1名女性1名の12名を選出した。被験者には1回10秒で良い、普通、悪い場合の3通りの動画クリップを撮影して1人あたり3本、すなわち全員で36本の動画クリップを収集した。収集した動画クリップ各3本を用いて個別に体調を推定する機械学習モデルを構築し、これによって動画クリップの体調が正しく分類できるのか実験で確認する。

4.2 実験方法

まず体調推定に使用する動画クリップの収集と感情値の取得を行った。エージェントとの会話を想定した表情を取得するために、10秒経過するまでにエージェントから体調を何う想定をした声掛けとして「調子はどうですか?」という音声を流し、被験者がそれに対して返答した。10秒経過後に動画クリップから150前後のフレームが抽出され、それぞれのフレームに対して推定された感情値が記録されたCSVファイルを得た。この動画クリップを収集してからCSVファイルを得るまでを良い、普通、悪い体調想定ごとに1回ずつ行い、3段階の体調を想定した各表情の演技の感情値を取得した。ここで感情値を取得する際の動画クリップや各フレームの静止画は保存しなかった。

次に動画クリップの体調を推定するために、訓練データの感情値から体調を推定する機械学習分類モデルを構築した。特徴量には取得した感情値の中からhappyのみを使用し、動画クリップの体調を正解データとした。モデルを構築する際の最適なハイパーパラメータはグリッドサーチを用いて、決定木の最大の深さであるmax_depthを1, 2, 3, 4, 5, 6, 7, 8, 9, 10の中から、決定木の分割に必要な最小のサンプルサイズであるmin_samples_splitを2, 10, 20, 30の中から最適な組み合わせを決定し、モデルは被験者ごとに構築した。そして構築したモデルを用いてテストデータの感情値からフレームごとの体調を推定し、同じ動画クリップのテストデータから推定した体調を集めて多数決を取り、最も多い体調を動画クリップの体調と推定した。

4.3 評価方法

構築したフレームの体調を推定するモデルを評価するために、被験者のモデルごとの訓練データの正解率とテストデータの正解率を比較し、推定した体調の中で実際の体調と一致した比率である適合率(precision)、実際の体調の中で一致した推定ができた比率である再現率(recall)、適合率と再現率の平均と解釈できる指標であるF値(f1-score)を算出する。F値は

$$F \text{ 値} = \frac{2 \times \text{適合率} \times \text{再現率}}{\text{適合率} + \text{再現率}} \quad (1)$$

で算出される。また多数決を取った動画クリップの体調推定手法について評価するために、実際の体調がどの体調であると推定されたかを混同行列に表すことと再現率、適合率、F値を算出することを行った。

4.4 実験結果

被験者ごとの機械学習分類モデルのハイパーパラメータを表2に示し、訓練データとテストデータの正解率について比較を行った結果とそれぞれの平均値を表3に示し、各体調の適合率、再現率、F値を被験者ごとに算出した結果を表4に示す。得られた決定木による分類モデルの内被験者Aの分類モデルを図5に示す。また多数決を取った結果の混同行列を図6に示し、再現率、適合率、F値を表5に示す。

初めに被験者ごとの機械学習分類モデルの結果について、表2よりハイパーパラメータは深さ1が4名、深さ2が3名、深さ3が1名とここまでで3分の2を占める結果となった。表3よりトレーニングスコアとテストスコアについては平均を見

表2 被験者ごとのハイパーパラメータ

	max_depth	min_samples_split
被験者 A	2	2
被験者 B	2	2
被験者 C	3	2
被験者 D	4	2
被験者 E	1	2
被験者 F	1	2
被験者 G	2	2
被験者 H	9	2
被験者 I	1	2
被験者 J	5	2
被験者 K	1	2
被験者 L	10	10

表3 被験者ごとのトレーニングスコアとテストスコア

	トレーニングスコア	テストスコア
被験者 A	0.725225	0.645740
被験者 B	0.915323	0.903614
被験者 C	0.828358	0.768657
被験者 D	0.939024	0.792683
被験者 E	0.707317	0.695122
被験者 F	0.805970	0.768657
被験者 G	0.883268	0.872093
被験者 H	0.882591	0.564516
被験者 I	0.812766	0.778723
被験者 J	0.819495	0.762590
被験者 K	0.707865	0.679775
被験者 L	0.706714	0.580986
平均値	0.811160	0.734430

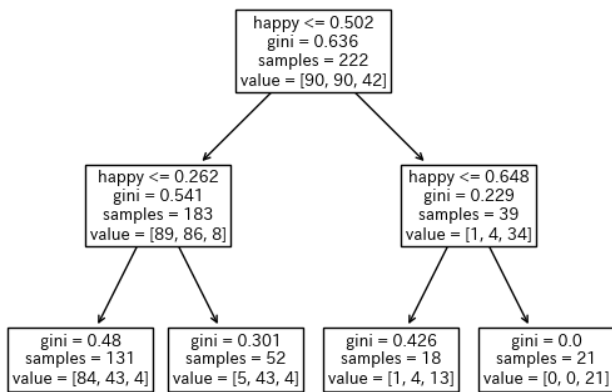


図5 決定木による被験者 A の分類モデル

るとトレーニングでは約 0.81 に近くテストでは約 0.73 という値が得られたが、max_depth が 9 と 10 の被験者 H と L に関してテストスコアが 0.6 を下回った。適合率、再現率、F 値を見ると、表 4 の体調が良い場合についてはどの被験者も高い値が得られている一方で、普通の場合や悪い場合は 0 を記録するなど分類されないケースもあった。図 5 を見ると happy の感情値を 0.502 を境に分けた後、0.502 より大きい値についてはさら

表4 被験者の体調ごとの適合率と再現率と F 値

		適合率	再現率	F 値
被験者 A	悪い	0.56	0.85	0.68
	普通	0.72	0.39	0.51
	良い	0.81	0.78	0.80
被験者 B	悪い	0.78	0.92	0.85
	普通	0.91	0.81	0.86
	良い	1.00	0.98	0.99
被験者 C	悪い	0.65	0.97	0.78
	普通	0.86	0.43	0.58
	良い	0.93	0.95	0.94
被験者 D	悪い	0.78	0.37	0.50
	普通	0.79	0.98	0.88
	良い	0.80	0.80	0.80
被験者 E	悪い	0.64	0.99	0.78
	普通	0.00	0.00	0.00
	良い	0.93	0.63	0.75
被験者 F	悪い	0.00	0.00	0.00
	普通	0.74	1.00	0.85
	良い	0.95	1.00	0.97
被験者 G	悪い	0.75	0.98	0.85
	普通	0.98	0.64	0.78
	良い	0.99	1.00	0.99
被験者 H	悪い	0.60	0.62	0.61
	普通	0.47	0.54	0.50
	良い	0.69	0.54	0.60
被験者 I	悪い	0.00	0.00	0.00
	普通	0.69	0.98	0.81
	良い	0.88	0.98	0.93
被験者 J	悪い	0.68	0.94	0.80
	普通	0.72	0.59	0.65
	良い	0.96	0.73	0.83
被験者 K	悪い	0.96	0.56	0.71
	普通	0.00	0.00	0.00
	良い	0.57	0.97	0.72
被験者 L	悪い	0.54	0.43	0.48
	普通	0.52	0.75	0.61
	良い	0.73	0.56	0.64

表5 動画クリップの体調推定の適合率と再現率と F 値

	適合率	再現率	F 値
悪い	0.73	0.67	0.70
普通	0.67	0.67	0.67
良い	0.92	1.00	0.96

に 0.648 を境に分けることでそれより大きい値はすべて良いであると分けられている。また 0.502 より小さい場合は 0.262 を境に分けることで悪いが 84 フレーム含まれるようなノードと悪いが 5 つしか含まれないノードに分けられている。

次に多数決を取った動画クリップの体調推定手法について、表 5 と図 6 より良い場合については適合率が 0.92、再現率が 1.00、F 値が 0.96 と特に良く正解できた。普通と悪いについては互いに間違えるケースもあるという結果が得られ、悪い場合は適合率が 0.73、再現率が 0.67、普通の場合は適合率が 0.67、

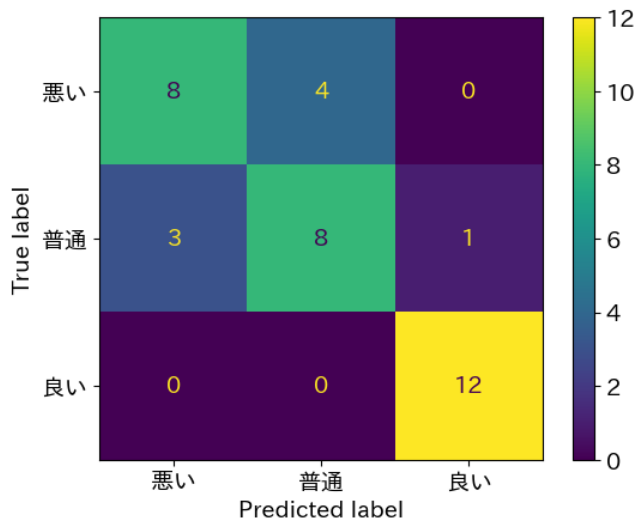


図6 動画クリップの体調推定の混同行列

再現率が0.67であるが、F値にするとそれぞれ0.70と0.67となった。

4.5 考察

初めに被験者ごとの機械学習分類モデルについて、ハイパーパラメータは特徴量がhappyのみの1つだけであることから決定木の大小比較だけではそこまで深さを必要としなかったと考えられる。トレーニングスコアとテストスコアについては深くなるほど分岐条件が多くなり分類が難しくなっているため、トレーニングスコアとテストスコアの差が大きくなっていると考える。適合率などの値については普通と悪いを互いに間違えるケースが多いため、普通と悪いが極端に低くなるという結果になり良いだけが独立して高い値が得られたと考える。また図5の分類モデルについてはhappyの感情値が大きい良い場合と小さい悪い場合の不純度(gini)が小さくなるように分類されており、結果として余ったものが普通になると考えられる。

次に多数決を取った動画クリップの体調推定手法について、図6よりhappyのみを用いたため、より笑顔が顕著なものが分類できたので良い場合が特に正解できていると考えられる。被験者ごとの機械学習分類モデルでも普通と悪いの適合率などが低くなる傾向があるため、この傾向を受けて普通と悪いについては互いに間違えやすいという結果が得られたと考えられる。このことからhappyだけでは3段階に体調を分類しきれない場合がある。これはhappyは推論される感情の内、良いという体調と幸福であるという感情が特に強く対応しているため、良いという肯定的な反応を分類できていると考え、他の感情値を使うことにより普通と悪いを区別しやすい感情によって分類精度を高めることができると考える。また表5より良い場合の再現率が高いこと、普通と悪いも7割ほどの再現率があることから、個別にモデルを作成して体調を推定する手法として用いることができるといえる。

5. まとめ

本研究では表情による感情認識を用いた体調推定手法を検討

した。本研究を行った背景としては、独居高齢者が増加しており日々のコミュニケーションと見守りと体調を把握して記録しておく必要があり、短時間での対話中の表情から体調を推定する手法を提案し、体調に合わせた声掛けにつなげることが目的である。アプローチとしては、「A1：体調に応じた動画クリップを収集」、「A2：フレームを抽出し、感情認識を適用して感情値を抽出」、「A3：感情値から特徴量を作成」、「A4：機械学習モデルを構築」、「A5：多数決によってクリップ全体の体調ラベルを推論」の5つである。この提案手法に基づきデータの収集、モデルの作成と評価を行った。

その結果、被験者ごとに個別にモデルを作成して体調を推定することができることが示された。今後の課題として、分類精度を高める特徴量を考えること、推定した体調から対話エージェントとの対話により深堀して詳細を聞き出すサービスを開発すること、音声やテキストでも体調を分類できるか実験することが挙げられる。

謝辞 本研究の一部はJSPS科研費JP20H05706, JP22H03699, JP22K19653, JP23H03401, JP23H03694, JP23K17006, および、立石科学技術振興財団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] 警察庁刑事局捜査第一課, “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年上半期(16月)暫定値～,” https://www.npa.go.jp/publications/statistics/shitai/hitorigurashi/240827_kenshi2.pdf. accessed October 17, 2024.
- [3] 総務省行政評価局, “一人暮らしの高齢者に対する見守り活動に関する調査結果報告書,” https://www.soumu.go.jp/main_content/000892187.pdf. accessed October 17, 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] いすみ市役所, “見守りあんしん電話事業,” https://www.city.isumi.lg.jp/soshikikarasagasu/fukushika/shakai-shogaifukushihan/2/1_1/1/1048.html. accessed October 17, 2024.
- [6] 厚沢部町保健福祉課地域包括支援センター係, “見守りロボット タピア 料金改定,” <https://www.town.assabu.lg.jp/uploaded/attachment/4536.pdf>. accessed October 17, 2024.
- [7] 株式会社ベネフィットジャパン, “Tapia,” <https://www.robotplanet.site/tapia/>. accessed October 17, 2024.
- [8] 社会福祉法人盛岡市社会福祉協議会, “「いわて「おげんき」みまもりシステム」のごあんない,” <https://www.morioka-shakyo.or.jp/wp/wp-content/uploads/2021/05/ogenkisystem2021.pdf>. accessed October 17, 2024.
- [9] 熊野町役場, “熊野町緊急通報事業について,” <https://www.town.kumano.hiroshima.jp/www/contents/1422406275963/index.html>. accessed October 17, 2024.
- [10] 大園隼人, 陳思楠, 中村匡秀, “音声対話エージェントを活用した在宅高齢者の支援と見守りを行う,” 電子情報通信学会技術研究報告, 第123巻, pp.18–22, Aug. 2023.
- [11] 川村亮介, 武村紀子, 佐藤宏介, “発話時の表情変化に基づいた精神疲労の推定,” 計測自動制御学会論文集, 第53巻, pp.90–98, Jan. 2017.
- [12] A. Ghosh, S. Umer, M.K. Khan, R.K. Rout, and B.C. Dhara, “Smart sentiment analysis system for pain detection using cutting edge techniques in a smart healthcare framework,” Cluster Computing, vol.26,

pp.119–135, 29 January 2022.

- [13] Y. Zhou, X. Yao, W. Han, Y. Wang, Z. Li, and Y. Li, “Distinguishing apathy and depression in older adults with mild cognitive impairment using text, audio, and video based on multiclass classification and shapely additive explanations,” *International Journal of Geriatric Psychiatry*, vol.37, pp.–, 09 October 2022.
- [14] 佐古嵐, 佐伯幸郎, 中村匡秀, 安田清, “コグニティブコンピューティングによる顔感情解析を用いたケア効果の定量的評価,” *電子情報通信学会技術研究報告*, 第 117 巻, pp.105–110, March 2018.
- [15] “face_classification,” https://github.com/oarriaga/face_classification. accessed October 17, 2024.
- [16] ポンダッド, “第 17 回 tensorflow.js で「感情」を認識してみよう,” <https://book.mynavi.jp/manatee/detail/id=99887>. accessed October 17, 2024.