

AdaBoost を用いた心の健康スコア推定システムの構築 Mental Health Score Estimation System Based on AdaBoost

中川 聡[†] 米倉 将吾[†] 金沢 星慶[†] 西川 鋭[†] 國吉 康夫[†]
Satoshi Nakagawa Shogo Yonekura Hoshinori Kanazawa Satoshi Nishikawa Yasuo Kuniyoshi

1. はじめに

我々の平均寿命は年々伸びている。しかし長寿は同時に世界人口の高齢化をも意味する。2019 年における 65 歳以上の世界人口は 7 億人であり、2050 年には 2 倍以上の 15 億人に増大すると予測されている[1]。さらに現状、男性では 10 人に 1 人が、女性では 5 人に 1 人が独居高齢者に該当する。このような世界規模での高齢化に対するアプローチとして、IoT を駆使した見守りサービスの導入が加速しており、この流れの一つに見守りロボットがある。ロボット活用に対する取り組みに関して、世界各国でロボット導入政策が発表されるなど[2]、あらゆる分野にまたがるクロスセクタなロボット開発が世界規模で加速している。さらにこの潮流が高齢者福祉分野と結びつくことで、高齢者見守りロボットへの需要もさらに高まると推測される。

高齢者福祉では、支援対象となる高齢者の生活の質 (QOL) 向上や QOL に基づく対応に努めるべきだとされる[3]。QOL は健康状態のみならず、精神状態や社会的活動なども総括した指標である。支援対象者の QOL を正確に把握することで、QOL 向上の促進や支援手法のパーソナライズが可能になる。しかし、従来の QOL 測定法には以下の 3 つの問題点が挙げられる。(1) 高齢者福祉で提案される従来の支援手法では、本来複数の尺度で構成される QOL を単一の概念として扱うことが多く、有効性の検証が適切でない。(2) QOL の定量評価はアンケートや面接の形式を採るが、これらは調査対象者に負担を強いる。また、一度の検査で測定可能なのは測定時の QOL のみであるため、QOL の継続的な測定には検査の反復実施を要し、効率性を欠く。(3) 従来の QOL 推定では、対象者に QOL 質問紙内の質問を投げかけるが、質疑応答形式は質問者と回答者の間に上下関係を築く可能性が指摘されており[4]、福祉の場では適切でない。

我々の研究では、それぞれの問題点に対する解決策を以下のように提示し、対話エージェントとのインタラクション中に得られる動画に基づく QOL 推定システムの実装および推定精度評価を通じて、提案手法の有効性を示した。

(1) 日本全国調査に基づく QOL 指標 SF-36v2 (質問紙)[5]を活用した定量評価を行った。なお、SF-36v2 で定義される QOL の構造は、8 つの尺度 (身体機能、日常役割機能 (身体)、体の痛み、全体的健康感、活力、社会生活機能、日常役割機能 (精神)、心の健康) により構成される。

(2) Human agent interaction (HAI) の過程において、QOL を推定するシステムを構築した。日常会話の中で QOL 測定を行えば、ユーザの状態を総括的に理解可能である上、推定結果は、それに基づくロボットが QOL 向上のためのコミュニケーション行動の糸口ともなる。(3) 質問回答

形式に替わり、内容に制限を設けない対話システムによる推定手法を提案した。

2. 方法

コミュニケーションエージェントと対話している際の動画とユーザの総合的な状態を表す QOL 情報が紐付いたデータベースは存在しない。したがって、本研究では自ら自然言語対話エージェントを実装し、それを用いたユーザとのコミュニケーション実験を通じて、ユーザがエージェントとインタラクションをしている際の動画、およびその時点における QOL 情報を収集した。その後、深層学習アーキテクチャを実装し、構築したデータベースに基づく QOL スコア推定実験を行い、提案手法の有効性を評価した。

2.1 対話エージェントとデータベースの構築

本研究では、発話により会話が可能な対話エージェントを PC 上に実装した。この対話エージェントはユーザの発言を収集する音声記録、音声を文章に変換する音声認識、会話を続けるためにコメントに対して返答する返答生成、返答として作成されたテキストをエージェントに発言させるための音声合成で構成される。返答生成の部分では NTT Docomo が提供する雑談対話 API[6]を一部導入した。

人が対話エージェントとコミュニケーション行動をとっている最中の動画のデータベースを作成することを目的として、対話エージェントを用いた会話実験を行った。本実験には 14 名 (男女各 7 名、年齢 26.5 (SD 12.1)) が参加した。実験の流れは図 1 に示したように、対話エージェントを起動すると、エージェントがまずは挨拶をする。それ以降は人とエージェントが交互に会話をしてゆく。実験中は、発話内容と全体的な発話時間に関しては制限を設けず、被験者の普段のコミュニケーションのとり方を優先した。人以外の実体に対して自己開示をも含む会話をするのに対して我々が抱く抵抗感の有無に関しては、対面、チャット、電話、テレビ電話、エージェントの 5 形式による半構造化面接実験を通じて検証している。実験結果から、エージェントとの対話形式はユーザに抵抗なく受け入れられる可能性があることを結論づけている[7]。

被験者はエージェントとの会話実験に先立ち、QOL 質問紙 SF-36v2 へ回答をしている。SF-36v2 は 36 の質問により構成され、3 から 6 つの選択肢の中から一つを選択する形式の質問紙である。この回答結果を既存のスコアリングアルゴリズムに入力することで、QOL を構成する 8 尺度それぞれのスコアを出力することができる。本研究ではこの従来の QOL 測定法により算出された QOL スコアを正解データとして活用した。

以上の実験により、図 2 に示したような、エージェントとの会話中の動画、被験者一人当たり QOL アンケートへの回答結果、従来手法により算出された QOL スコアにより構成されるデータベースを構築した。

[†] 東京大学大学院情報理工学系研究科 Graduate School of Information Science and Technology, the University of Tokyo

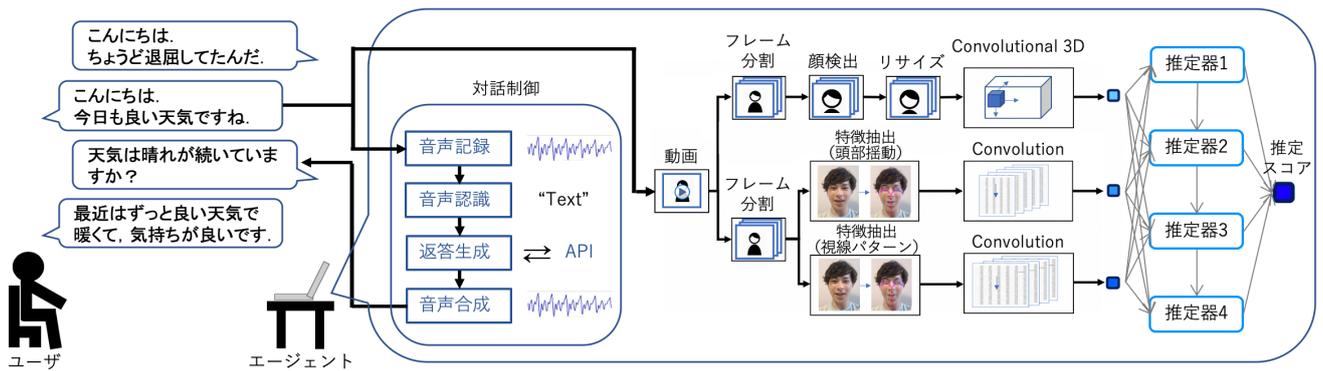


図 1 対話エージェント

2.2 深層学習アーキテクチャの実装

QOL 学習推定システムとして、図 1 右側に示した深層学習アーキテクチャを設計した。3 つの個別の特徴量に基づく推定器を並列に配し、最終段階において統合する構造である。

上段の表情に基づく推定器は、入力となる動画の三次元情報に対してフレーム分割、顔検出、サイズ調整の前処理を施したのちに Convolutional 3D (C3D) [8]を用いたニューラルネットワークへと入力される。C3D は 3 次元 (2 つの空間次元と 1 つの時間次元) で畳み込みを行うアプローチであり、これはビデオベースの感情認識において有効性が示されている [9]。この C3D を融合した表情に基づく QOL スコア推定では、QOL を構成する 8 尺度のうち、心の健康スコアの推定が困難であることを我々の先行研究 [10]において示している。そこで、さらなる QOL スコア推定精度向上を目的として、視覚情報の個別の特徴抽出に基づく学習推定を行った。

頭の揺動パターンを用いて表情認識を目指した研究 [11,12,13]や、視線パターンと感情状態に相関があることを示す研究 [14,15,16]があることから、表情の他に新たに抽出する視覚情報として、頭部揺動および視線パターンが有効であると考えられる。したがって、頭部揺動および視線パターンは人の状態を推定するのに有用な情報であり、QOL に関する情報も抽出することが可能であると仮定し、頭部揺動および視線パターンに基づく QOL 尺度の一つ「心の健康」に関するスコア推定を行った。

頭部揺動および視線パターンの抽出方法を統一することを目的として、さらに最終的にはカメラを備えたインタラクティブロボットが自律的に抽出できる手法として、OpenPose [17]を用いて両者の特徴量抽出を実行した。この抽出方法により、頭部揺動および視線パターンそれぞれの時系列データを得た。

2.3 評価実験

本研究は予備実験と本実験の二つの実験で構成される。予備実験では、頭部揺動および視線パターンの QOL 推定

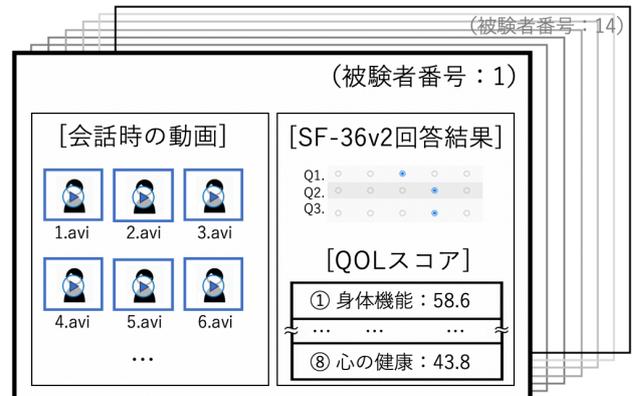


図 2 収集データ

に対する有用性を検証するために、心の健康スコアの高低の分類実験を実行した。被験者 14 名を従来のスコアリングアルゴリズムを用いて算出した QOL スコアに応じて低スコア群と高スコア群に分けた。被験者の「心の健康」尺度スコアを昇順に並べたものを図 3 に示している。縦軸が「心の健康」スコアに対応し、横軸が被験者番号に対応している。QOL を構成する 8 つの尺度にはそれぞれに国民標準値が存在するが、図中の QOL スコアは国民標準値が 50、標準偏差が 10 となるように調整している。したがって、50 をカットオフ値として、高スコア群および低スコア群の 2 群に分類することができる。続いて本実験においては、心の健康スコアを直接推定する回帰実験を実行した。提案手法の評価として、表情、頭部揺動、視線パターンのそれぞれの特徴抽出に基づく 3 種のシングルモーダル学習、および AdaBoost アルゴリズムを用いた、これら 3 種類の特徴量の統合学習の合計 4 つの学習推定を実行した。全ての学習実験において入力したデータは、対話エージェントとの会話実験の過程において収集された動画に由来するものであり、評価においては Leave one out 法 [18]を用いて交差検証を実行した。

3. 結果

3.1 頭部揺動および視線パターンに基づく推定可能性に関する検証

14 名の参加者の動画から頭部揺動および視線パターンを抽出して分類した結果をそれぞれ図 4, 図 5 に示した. 低スコアと推定された被験者を青色で, 高スコアと推定された被験者を赤色で示している. 頭部揺動に基づく分類実験の結果に関しては, スコアが 40 点未満の被験者データに関しては正しく分類ができていないが, 全体的な精度は 57% であった. この値は, 頭部揺動パターンが心の健康スコアの高低に弱い関連がある可能性を示唆している. また, 視線パターンに基づく分類実験の結果に関しては, カットオフ値に最も近い各群の 2 名の被験者の高低の逆転や低スコア者の判別誤りが確認され, 全体的な精度は 71.4% であった. この結果から, 視線パターンに関しては心の健康スコアの高低に関連があることが確認された. したがって, 頭部揺動および視線パターンの両者ともに QOL スコアに有用な特徴量の一つである可能性があることが示された.

3.2 統合学習に基づく QOL スコア推定可能性に関する検証

対話エージェントとの会話実験を通じて作成されたデータベースから抽出した表情に, 頭部揺動および視線パターンを加えた統合学習に基づく推定結果について説明する.

AdaBoost アルゴリズムを用いた統合学習に基づく推定実験結果, および視線パターン, 頭部揺動, 表情をそれぞれ個別に用いたシングルモーダル学習に基づく推定結果を比較したものを図 6 に示した. グラフは 4 種の条件下での実験による, 今回実装した学習推定システムにより推定されたスコアと, 既存のスコアリングアルゴリズムにより算出された実際のスコアとの誤差の分布をそれぞれ表している.

誤差の分布を比較すると, 今回の実験で実装した AdaBoost アルゴリズムを導入した統合学習 (median: 2.7, first quartile: 0, third quartile: 11) が, シングルモーダル学習である視線 (median: 13, first quartile: 7.6, third quartile: 16), 頭部揺動 (median: 11, first quartile: 9.6, third quartile: 17), 表情 (median: 11, first quartile: 5.6, third quartile: 26) に比べて誤差を小さく抑えて推定スコアを出力することができたことがわかる. また t 検定の結果, 表情と比較して視線パターンおよび頭部揺動の入力による誤差の減少には有意差はないものの, 統合学習と頭部揺動, 視線, 表情それぞれの比較では, AdaBoost を適用した学習推定器による推定誤差の減少に有意差 ($p < .05$) があることが確かめられ, 提案システムに QOL スコア推定機能の存在が示された.

4. 考察

頭部揺動および視線パターンに基づく推定可能性に関する検証を通じて, 被験者の心の健康スコアの高低を判別したが, 実験結果は両者から「心の健康」尺度の推定が可能であることを示唆している. したがって, 統合学習に基づく QOL スコア推定可能性に関する検証を通じて, AdaBoost を適用した統合学習と, それぞれの特徴量を個別に用いたシングルモーダル学習による結果を比較した. 頭部揺動と表情, および視線パターンと表情のそれぞれにおける t 検定では有意差は見られなかったものの, 統合学習

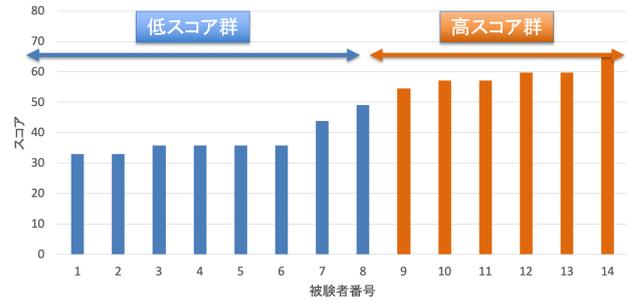


図 3 被験者 14 人の心の健康スコア

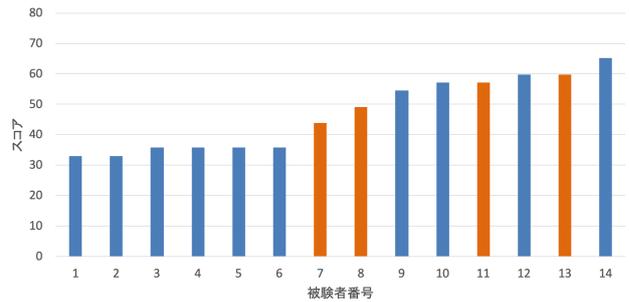


図 4 頭部揺動を用いたスコア高低判別結果

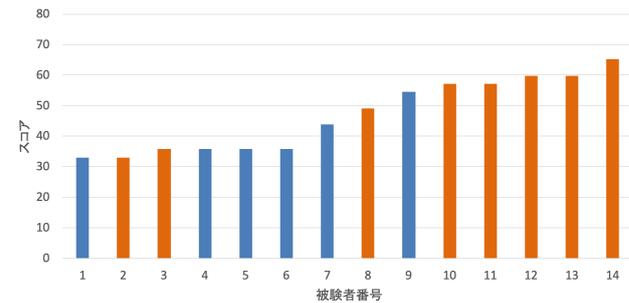


図 5 視線パターンを用いたスコア高低判別結果

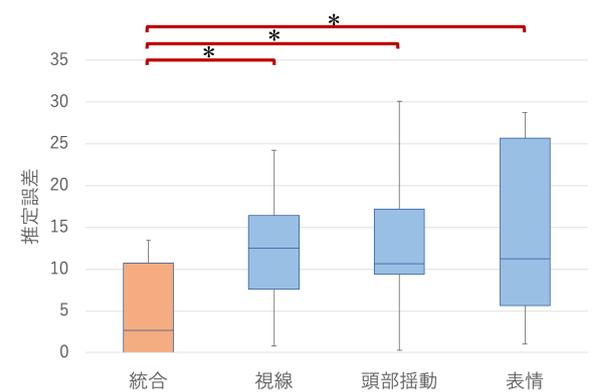


図 6 心の健康スコアの推定結果

による結果はその他のシングルモーダル学習との間に有意差が確認された. このことから, 頭部揺動, 視線パターン, 表情のそれぞれを用いた場合は弱い推定器であるものの, 統合させることで推定精度の向上を実現させることができ, 統合学習のシングルモーダル学習に対する優位性が示され

た. さらに 14 名中 9 名の被験者において誤差を一桁に抑えている上, 6 名の被験者に関しては誤差なく推定できていることから, 提案システムがユーザの QOL 推定器として十分活用可能であるといえる.

本研究では統合学習とシングルモーダル学習 3 条件の合計 4 条件で推定精度を比較したが, いずれの条件においても, 使用したデータベースは独自に実装した対話エージェントとの会話実験を通じて構築したものであった. しかし, 同一のデータベースを用いたにも拘らず, 先行研究[10]においてはシングルモーダル学習条件の一つ, 表情のみを入力とした場合の推定が困難であった. その理由としては次の 2 点を考える. まず, 学習モデルである C3D に入力するための前処理において顔を抽出した段階で, QOL 推定に有効な頭の揺動に関する情報が入力データから欠落したことである. もう 1 点としては, 表情の動画における視線パターン以外の変化量が大きく, QOL 推定に有効である視線パターンの適切な抽出, 学習が行われなかったことを考える.

インタラクションの過程において我々は, 相手の状態理解に基づいて自身の言動を生成, 選択していると考えられる. 人とコミュニケーションロボットのインタラクションの質を向上させることを目的としてロボットやエージェントにも相手の状態推定を行わせる際, 本研究の QOL スコア推定実験の結果を鑑みると, QOL はロボットによる人の状態推定の指標として十分活用可能である. したがって, ロボットに QOL 推定をさせることによって, 我々が日常的に行っているような相手の状態理解とそれに基づく行動生成の実現を目指すことが可能なのではないかと考える.

5. おわりに

本研究ではエージェントとのインタラクションの過程においてユーザの心の健康スコアを推定するシステムを設計し, 実データを用いた実験結果から, 提案手法の実用可能性を明らかにした.

一般に, HAI では人間同士のような長期的関係の構築が容易ではないとされている. これは, 人間とエージェントとの関係が経時的変化に乏しく, 人間がエージェントに関心を示さない飽きの段階に入ることが原因だと考えられている. 本研究で行った対話エージェントとのコミュニケーション実験は, 関係構築の経時的発展の中でも初期の部分抽出している. したがって今後の展望として, エージェントとの関係構築を目指す長期間の時系列データを収集し, そのデータベースに基づく学習推定を実現させることを考える. 推定したユーザの状態に基づき, QOL 最大化を図るようロボットに随時言動生成をさせることで, 高齢者とロボットに良好で長期的な関係をもたらしたいと考える.

高齢者福祉へのヒューマンロボットインタラクション導入の動きは存在するものの, 一般的な普及のためには未だ多くの課題が残されている. 例として, 従来の QOL 測定法が非効率的事であることや, 人間とロボットとの関係の画一性が原因で長期的関係の構築が困難になることなどが挙げられる. 本研究の提案手法はこれらの問題への対処を目的としている. ロボットとの日常的なインタラクションの中でユーザの QOL を推定することで, 継続的かつ効率の良い QOL 測定が可能となる. また, 今後の展望として挙げた, 推定した QOL を考慮したロボットの言動生成により, ロボットとユーザとの関係に経時的変化がもたらされ,

長期的関係構築を促進するであろう. このように, QOL 推定を軸としたマンマシンインターフェースの構築は, 知的ロボットによる状態推定と長期的な関係構築を実現し, 高齢者福祉の ICT 化における諸問題の解決策となりうる.

謝辞

本研究は, 公益財団法人 KDDI 財団 2019 年度調査研究助成, 次世代知能科学研究センター, 文部科学省の東京大学ソーシャル ICT グローバル・クリエイティブリーダー育成プログラム (GCL) の支援を受けた.

参考文献

- [1] The United Nations, "World population ageing 2019," (2019).
- [2] Prime Minister's Office, New Robot Strategy Japan's Robot Strategy - Vision, Strategy, Action Plan -, (2015).
- [3] H. Masui, M. Otaga, Y. Moriyama and T. Matsushige, "Current issues in long-term care policy and research", Journal of the National Institute of Public Health Vol. 68, No.1 (2019).
- [4] 赤間怜奈, 徳久良子, 乾健太郎, "Quality of Life 情報の伝達補助を目的とする対話応答候補生成", 自然言語処理 Vol. 26, No. 3 (2019).
- [5] S. Fukuhara, S. Bito, J. Green, A. Hsiao, and K. Kurokawa, "Translation, adaptation, and validation of the sf-36 health survey for use in japan," Journal of Clinical Epidemiology, Vol. 51, No. 11, (1998).
- [6] docomo developer support, NTT Docomo. (n.d.). <https://dev.smt.docomo.ne.jp> (accessed Jun. 16, 2020).
- [7] 中川聡, 三枝弘幸, 遠藤凌河, 成瀬加菜, Lem Wey Guan, 國吉康夫, "VR テレカウンセリングに基づくセラピスト-クライアント間インタラクションに関する評価," 第 24 回日本バーチャルリアリティ学会大会, 4D-04, (2019).
- [8] D. Tran, L. Bourdev, R. Fergus, L. Torresani, and M. Paluri, "Learning spatiotemporal features with 3d convolutional networks," in Proceedings of the IEEE International Conference on Computer Vision, (2015).
- [9] Y. Fan, X. Lu, D. Li, et al., "Video-based emotion recognition using cnn-rnn and c3d hybrid networks," in Proceedings of the 18th ACM International Conference on Multimodal Interaction, (2016).
- [10] S. Nakagawa, D. Enomoto, S. Yonekura, H. Kanazawa, and Y. Kuniyoshi, "A Telecare System that Estimates Quality of Life through Communication," 2018 5th IEEE International Conference on Cloud Computing and Intelligence Systems (CCIS), (2018).
- [11] J. F. Cohn, L. I. Reed, Z. Ambadar, et al., "Automatic analysis and recognition of brow actions and head motion in spontaneous facial behavior," in 2004 IEEE International Conference on Systems, Man and Cybernetics (IEEE Cat. No. 04CH37583), vol. 1, (2004).
- [12] Y. Zhang and Q. Ji, "Active and dynamic information fusion for facial expression understanding from image sequences," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 27, No. 5, (2005).
- [13] P. Liu and L. Yin, "Spontaneous facial expression analysis based on temperature changes and head motions," in 2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), Vol. 1, (2015).
- [14] L.M. Zhao, R. Li, W.L. Zheng, and B.L. Lu, "Classification of five emotions from EEG and eye movement signals: complementary representation properties," 2019 9th International IEEE/EMBS Conference on Neural Engineering (NER), (2019).
- [15] Y. Lu, W.L. Zheng, B. Li, and B.L. Lu, "Combining eye movements and EEG to enhance emotion recognition," 24th International Joint Conference on Artificial Intelligence, (2015).
- [16] 古川茂人, "眼から読み取る心の動き Heart-Touching-AI のキー技術", NTT 技術ジャーナル 2016, (2016).
- [17] Z. Cao, T. Simon, S.E. Wei, and Y. Sheikh, "Realtime multi-person 2d pose estimation using part affinity fields," Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit, (2017).
- [18] E. Kocagueli and T. Menzies, "Software effort models should be assessed via leave-one-out validation," Journal of Systems and Software, Vol. 86, No. 7, (2013).