

[研究助成 (A)]

対話中の非言語行動の意味・機能を推定する機械学習モデルの構築

Developing machine-learning models for estimating the meanings
and functions of nonverbal behaviors in conversations

2211004



研究代表者 横浜国立大学 大学院工学研究院 准教授 大塚 和弘

[研究の目的]

SDGs や Society 5.0, 新型感染症などの社会的課題の解決には、社会の成員による対話を介した合意形成と集合知創出が重要である。本研究では、対面・遠隔を問わず人々の対話活動を支援し、人との高度な知的協調を可能とする人工知能 (AI) の実現を長期的な目的とする。対話においては、言語だけではなく、顔表情や頭部運動、視線、身体動作・姿勢、韻律などの非言語行動が重要な役割を果たす。こうした非言語行動を介して、人は互いの意図や気持ちを読み取り、意見を調整し、相互理解や合意形成を図る。将来的にエージェントが、ユーザの心の機微や曖昧な行動の意味を読み解き（「場の空気を読み」）、また、人間のような豊かな意味を伝える行動表出を行い（「以心伝心」）、人々の議論に参画し、相互理解や合意形成、共同問題解決を支援できるような AI の構築をビジョンとする。

[研究の内容, 成果]

本研究では、人と人の対話におけるマルチモーダル非言語行動表出の機能を認識・理解するため『非言語機能スペクトラム』という新概念を提唱し [1], その計算基盤の構築を行った

[2~8]。非言語機能スペクトラムとは、対話における表出行動が持ちうる意味や機能を分布として捉える新しい表現形態である。対話においては、一つの非言語行動表出は一つの機能（や意味や意図）に対応するのではなく、本質的に多様で多対多の関係にある。例えば、一回の頷きは、単に話を聞いているとも、理解や同意のサインとも読み取れる。また、笑顔も単に幸福のサインとは限らず、興味や困惑、羞恥、嘲笑など様々な意味を持ちうる。

このように表出者が行動に込める意味も多様かつ曖昧であり、また、それを受け取る側の解釈も多様である。本研究で提唱した非言語機能スペクトラムは、このような非言語表出の多様性や解釈の曖昧性を捉えることを目的としている。

本研究では、非言語行動として、発話時のリズム取り、強調、反応確認、思考、及び、反応時の機能として、相槌、応答、思考、理解、肯定、正の感情表出などに着目した。これら各々の機能の現出の度合いを非言語行動の観測データから推測するという計算基盤の枠組みを提案した。

図1にその概略を示す。図1では、入力の特徴量として、頭部運動、及び、顔の表情を対象としている。これらに対話者の顔画像より頭部姿勢角、及び、顔面活動単位 (AU) の時

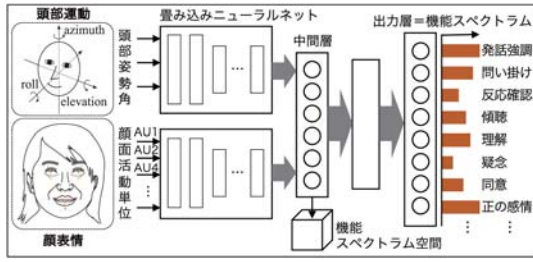


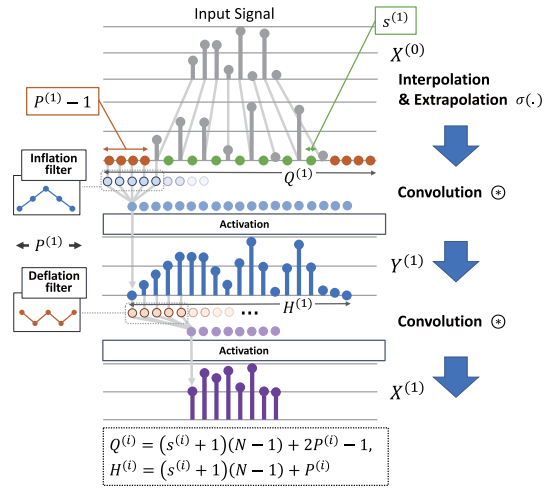
図1 非言語機能スペクトラム推定のための深層神経回路網の基本構造

系列データとして計測する。次にこれらを入力として用い、畳み込みニューラルネット (CNN) により、これら情報を統合し、各々の機能の強度を推定し、出力する。なお、入力の特徴モダリティとしては、頭部運動や顔表情に限らず他のモダリティ情報も用いることができる。また、本研究では、この枠組の有効性を検証するため、特に頭部運動に着目した。本研究では、部分課題として、以下の3テーマについて取り組み、成果を創出した。

- (1) 頭部運動機能の認識性能の向上を目的とした、新しい深層ニューラルネット (DNN) の付加機構 (I/DeF 機構と呼ぶ) の提案 [2-4]
 - (2) 話し手・聞き手間の相互作用機能を認識する DNN の構築 [5-6]
 - (3) 頭部運動の非言語機能スペクトラムから対話者の主観を予測するモデルの提案 [7-8]
- 以下、具体的に述べる。

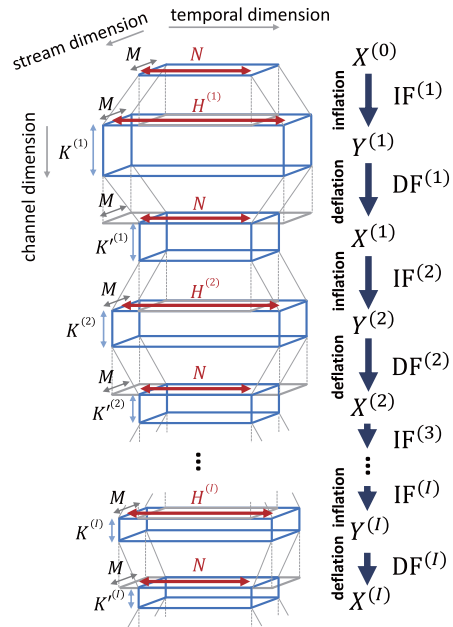
(1) 対話中の頭部運動機能を認識する特徴膨張収縮ニューラルネットワーク

従来、対話者の頭部姿勢角の時系列から頭部運動機能を認識するための畳み込みニューラルネットワーク (CNN) が提案されているが、認識性能には向上の余地があった。本研究では、この CNN の性能向上を目的とし、入力時系列の特徴学習を促進させるための付加機構として、特徴膨張収縮機構 (I/DeF 機構) を提案した。この機構は、DNN に入力される時系列データに対して、微視的スケール方向および入力時系列の時間窓の前後方向に膨張と収縮を繰り返し



([3]より引用・改変)

図2 I/DeF 機構の処理



([3]より引用・改変)

図3 I/DeF 機構の構成

行うことで、稠密かつ子細な時間構造の特徴学習の促進を狙った機構である。図2に I/DeF 機構の処理を示す。この処理は、膨張処理と収縮処理により構成される。前者では、転置畳み込みによる微視的スケール方向および時間窓の前後方向へのデータの拡大が行われる。後者では、畳み込みによる時間方向のデータの削減・復元が実施される。また、I/DeF 機構は、図2の処理を単位として、複数、連結される構成を取ることもできる (図3)。

また、この I/DeF 機構は、既存の CNN などの入力の前段に配置され、一体として学習が行われる。

提案した I/DeF 機構の有効性を検証するため、頻出する頭部運動機能 10 種を対象とした実験を行った。その結果、I/DeF 機構を統合した CNN は、従来の CNN の性能を全機能カテゴリにおいて上回り、最大で 4.5 ポイントの F 値の向上を実現した。また、I/DeF 機構を VGG, ResNet にも統合して、比較を行った結果、I/DeF 機構を統合した CNN が 8 割の機能カテゴリにおいて最も優れた性能を示した。

(2) 複数人対話における深層転移学習による頭部運動相互作用の機能認識

本研究では、聞き手と話し手の頭部運動機能の組を「相互作用機能」として定義し、既存の頭部運動機能を認識する CNN を転移させることで、これらの相互作用機能を認識するモデルを構築する戦略を提案した。この転移戦略は、

- I. 個人動作・相互作用転移 (AIAT)
- II. 多機能スペクトラムベクトル (mSFV)
- III. 多機能学習 (MFL)

から構成される。個人動作・相互作用転移は、学習済みの個人モデルを再利用し、話し手と聞き手それぞれについて組み合わせる戦略である。次に、多機能スペクトラムベクトルは、複数の個人モデルの中間層出力を取り出し、連結したベクトルであり、複数機能の出現分布を表現する形態であり、このベクトルを他の識別器に入力し、機能の認識を行う。さらに多機能学習は、

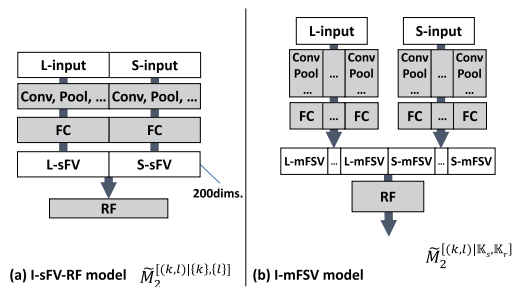
複数の機能を一つのモデルで認識するようモデルを訓練する戦略である。

図 4 にこれらの戦略を用いて構成した相互作用機能を認識するモデルの例を示す。図 4(a) は学習済みの頭部運動機能認識 CNN を話し手と聞き手それぞれについて配置し、これらの CNN の中間層出力を抽出し、ランダムフォレスト (RF) の入力し、相互作用機能を認識するモデルである。図 4(b) のモデルは、図 4(a) のモデルを拡張し、話し手、聞き手について、それぞれ複数の頭部運動機能を認識する CNN を配置し、それらの多機能スペクトラムベクトルを抽出して、RF で認識を行うモデルである。

話し手・聞き手各々の機能認識 CNN の出力について論理積をとる方策を基準として、提案した転移戦略を適用したモデルによる性能向上を検証した結果、話し手のリズム取りに対する聞き手の相槌という相互作用において、F 値にて最高 8.0 ポイントの性能向上を確認し、また、話し手のリズム取りに対する聞き手の正の感情表出という相互作用では、最高 13.9 ポイントの性能向上を確認した。このように提案した転移戦略の潜在的可能性が確認された。

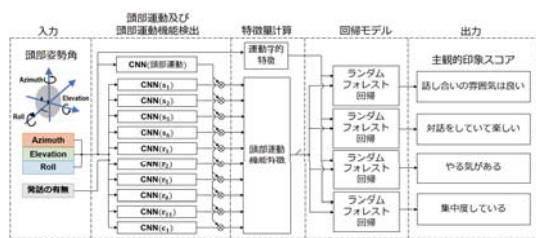
(3) 頭部運動機能特徴を用いた複数人対話における対話参加者の主観的印象の予測

対話中に表出される頭部運動の機能と対話者の主観的印象との関連性を探るため、頭部運動の機能に関する特徴量を用いた対話者の主観的印象の予測モデルを提案した。本研究では、「NTT 主観的アノテーション付きオフラインミーティングコーパス」(SA-Off コーパス)を対象として、対話者本人により事後に評定された主観の項目(「話し合いの雰囲気は良い」、「対話をしていて楽しい」、「やる気がある」、「集中している」)のスコア(1~9)を予測の対象とした。このコーパスには女性 4 名、17 グループの対話が含まれる。また、このコーパスの一部について、2 名の外部評定者により頭部運動機能のアノテーションを実施し、構築した



([6]より引用)

図 4 相互作用の認識モデルの構造



（[8]からの引用）

図5 主観的印象の予測モデルの構成

コーパスを予測モデルの検証に用いた。

図5には、提案した主観的印象の予測モデルの構成を示す。本モデルは、入力として、対話者の頭部運動と発話の有無の情報を用いる。まず、それらの情報から、頭部運動区間、並びに、10種の頭部運動機能をCNNを用いて検出する。次に、その検出結果を用いて、22種の頭部運動機能特徴、及び、6種の運動学的特徴を計算する。頭部運動機能特徴には、各々の機能の出現率や構成比などが含まれる。対話者の行動、心理的な状態がこれら特徴量に反映されていると考えられる。さらにこれらの特徴量を入力として、ランダムフォレスト回帰モデルにより、4つの印象項目それぞれについて、スコアの予測値を得る。

各グループ・項目について、従来型の運動学的特徴のみを用いた予測モデルと、これに今回提案した頭部運動機能特徴を追加したモデルとを比較した結果、後者により中程度以上の相関（相関係数0.5以上）がみられたグループ・項目の割合が25%から41%に増加することが確認され、頭部運動機能特徴は、主観的印象の予測性能の向上に寄与することが示唆された。

【今後の研究の方向、課題】

以上、本研究では、人の対話におけるマルチモーダル非言語行動表出の機能を認識・理解するため『非言語機能スペクトラム』という新概念を提唱し、その計算基盤の構築を行った。次のような今後の方向性・課題があげられる。

まず、非言語行動のモダリティの拡張が課題

としてあげられる。本研究では、頭部運動に焦点を絞ったが、他に顔表情や音声の韻律などの非言語行動も対話においても重要な役割を果たす。現在、顔表情の機能を認識するDNN、並びに、顔表情と頭部運動の連携によって発揮される相乗機能を認識するDNNについて研究を進めている。さらに今後は、韻律や視線などより多くのモダリティに対象を拡大し、それら機能について理解を深め、総合的に人の行動や心理、対話の状態を認識・理解できるような技術の研究開発を進めていく予定である。

また、今回のDNNは、機能の有無という2値情報を出力としていたが、今後は、機能の強度値を推定対象とすることで、より微細な行動の意味や機能を認識・理解する方向へと研究を進めていきたい。

以上を含め、今後も引き続き、人の対話活動を支援できるような人工知能技術の構築を進める予定である。

【成果の発表、論文等】

- [1] 大塚和弘：「対話中の非言語行動の機能スペクトラム解析に向けて～頭部運動を対象とした機能コーパスと深層学習モデルの構築事例～」，信学技報，vol. 121, no. 143, HCS2021-19, pp. 13-18, 2021
- [2] 武田一輝，大塚和弘：「対話中の頭部運動機能認識のための特徴膨張収縮深層ニューラルネットワーク」，電子情報通信学会和文論文誌 A, Vol. J105-A, No. 2, pp. 26-37, February, 2021, DOI: 10.14923/transfunj.2021JAP1016
- [3] Kazuki Takeda, Kazuhiro Otsuka, "Inflation-Deflation Networks for Recognizing Head-Movement Functions in Face-to-Face Conversations," in Proceedings of ACM International Conference on Multimodal Interaction (ICMI2021), pp. 361-369, 2021
- [4] 武田一輝，大塚和弘：「対話中の頭部運動機能を認識する特徴膨張収縮ニューラルネットワーク」，第20回情報科学技術フォーラム（FIT2021）【FIT奨励賞，FIT2021ヤングリサーチ賞】
- [5] Takashi Mori, Kazuhiro Otsuka, "Deep Transfer Learning for Recognizing Functional Interactions via Head Movements in Multiparty Conversations," in Proceedings of ACM International Conference on Multimodal Interaction (ICMI2021), pp. 370-378,

2021

- [6] 森 駿志, 大塚和弘, 「複数人対話における深層転移学習による頭部運動相互作用の機能認識」, 電子情報通信学会 HCG シンポジウム 2021 【特集テーマセッション賞】
- [7] Shumepi Otsuchi, Yoko Ishii, Momoko Nakatani, Kazuhiro Otsuka, "Prediction of Interlocutors' Subjective Impressions Based on Functional Head-

Movement Features in Group Meetings," in Proceedings of ACM International Conference on Multimodal Interaction (ICMI2021), pp. 352-360, 2021

- [8] 大土隼平, 石井陽子, 中谷桃子, 大塚和弘: 「頭部運動機能特徴を用いた複数人対話における対話参加者の主観的印象の予測」, 信学技報, vol. 121, no. 143, HCS2021-20, pp. 19-24, 2021