

対話知能学

ニュースレター (Vol. 14)



1. 研究紹介

■ 好意の返報性に基づく対話惹き込みシステム (A01 班)

A01 班の瀬島らの研究グループでは、相手から好意を寄せられると自己も相手へ好意を寄せようとする「好意の返報性」に着目した対話継続を試みています。対話相手、すなわち聞き手ロボットからの好意的な／無関心な振る舞いが対話意欲や発話継続に与える影響を解析・モデル化し、そのモデルをロボットへ組み込むことで、思わず対話へ惹き込まれてしまうコミュニケーションシステムを追求しています。とくに、好意的／無関心さが伝達されやすく、かつその解釈や意味付けが対話者へ委ねられる「瞳の動き」に着目したシステム設計を行っています。



図 1. 瞳ロボットによる傾聴



図 2. きくロボとコミュニケーションの様子

「瞳の動き」を表現する仕組みとして、直径 25cm の半球ディスプレイに瞳孔と虹彩を投影したシステムを構築しています(図 1)。このシステムは、好意的な振る舞いとしてアイコンタクトの強化と瞳孔拡大、無関心な振る舞いとして視線逸らしと瞳孔縮小を実現します。音声が入力されると、推定モデルによってこれらの振る舞いが切り替わる傾聴アルゴリズムを設計し、上記のシステムへ導入しました。コミュニケーション実験により、親近感や傾聴感に加えて対話への意欲が高まることを確認しました。この研究成果は、第 23 回計測自動制御学会システムインテグレーション部門講演会(SI2022) 優秀講演賞を受賞しました。

さらに、実証実験の試みとして 2022 年 12 月にグランフロント大阪のカフェラボにおいて「対話ロボットとコミュニケーション」というイベントを企画しました。この企画は、対話機能を

持つロボットを一般の方が体験する場を提供し、対話を通じてロボットの現状や未来のロボットに必要な機能を創造する機会を設けることを目的としました。言葉に加え、対話感を生み出す非言語情報に注目したコミュニケーション力、通称“コミュカ”を体験(バトル)し、対話ロボット技術を広く発信しました。イベントには、本領域の「きくロボ(京都大学 河原研究室)」「瞳ロボット(関西大学 瀬島研究室)」に加え、「自己推薦ロボット(サイバーエージェント)」「エモーション・クロースロボット(同志社大学 大久保研究室)」「ペコッぱ(岡山県立大学 渡辺研究室)」に参画いただきました(図 2)。幼児から大人まで参加し、対話ロボットの印象についてアンケートで回答しました。来場者からは、「楽しく会話することができた」「最近はオンラインばかりなので、このような企画があると情報をキャッチしやすい」等のコメントが寄せられ、実証実験に向けた有意義な企画となりました。

瀬島 吉裕(関西大学/A01 班 対話継続関係維持研究グループ公募班代表者)

■ 歩行映像解析による話者属性の早期認識 (A02 班)

人の歩き姿である歩容は、人物が遠方で小さく映っている場合やマスクをしていて顔がはっきりと見えないような場合であっても利用可能な生体情報(バイオメトリクス)として注目を集めている。もし、遠方にいる話者候補の属性を早期に推定できると、属性に応じた対話の呼びかけが可能といった利点が考えられる。そこで、本研究では、歩行映像解析による年齢推定・性別識別モジュールを開発した。

本モジュールは、以下の人物映像解析用の深層学習モデルを基に構成した。まず、話者候補となる人物領域は、人物のインスタンスセグメンテーションモデルの一つである PointRend を利用して抽出する。次いで、人物領域の外接矩形で切り取った画像に対して、人体モデル当てはめモデルの一つである Human Mesh Recovery (HRM) ネットワークを適用し、人体の体型や姿勢に関する標準的な 3 次元人体モデ



ルの一つである Skinned Multi-Person Linear (SMPL) モデルのパラメータを抽出する。最後に, SMPL モデルの体型パラメータもしくは姿勢パラメータを全結合層や畳み込み層に入力することで, 年齢推定や性別識別を行う。

実験では, 歩行映像の公開データベース OUMVLP(老若男女の被験者数 10,307 名を含む)による精度評価を行った。結果として, 姿勢パラメータを認識に用いた場合で, 性別の識別精度が 96.64%, 年齢推定の MAE(推定値と真値との平均絶対誤差)が5.19歳となった。顔画像による年齢推定精度と比較すると劣るものの, 大まかな年代を考慮した対話モデルの適応などには十分な精度と言える。

また, データベースによるオフライン評価に加えて, オンライン(リアルタイム)での動作検証を行った(図 3)。画像中の人物は 33 歳の男性であり, マスクを着用して顔がはっきりと



図 3. 歩容による属性推定結果 (左) と入力画像の人物領域及び頭部領域の拡大図 (右)

は見えない横向きの歩行映像に対しても良好な性別識別・年齢推定結果が得られた。

今後は対話モジュールとの連動実験を通して, 本モジュールの有効性を検証する予定である。

榎原 靖(大阪大学/A02 班 対話理解生成研究グループ計画班分担者)

■ ヒューマノイド・オルタ 3 における模倣を通した行動の概念形成(A03 班)

A03 公募班である, 東京大学の升森, 吉田, 丸山, johnsmith, 池上らの研究グループでは, ヒューマノイドロボットを用いて, 身体動作を主とした対話を通して自己を獲得していくシステムについて研究を行っている。本稿では, これまでの模倣システムに, 人の動きを模倣することで身体動作の概念を獲得していくニューラルネットを追加し, ヒューマノイドロボットに実装して実験した成果を紹介する。この模倣システムでは, ヒューマノイドロボット Alter3 が, 人の行動を真似するために, 3 つの戦略 - Mimicking, Imitation, Memory - を選択的に組み合わせて使用する。Mimicking モードでは, Alter3 は関節角度を計算することで, 人が「どのように」動かかを再現する。Imitation モードでは, 変分オートエンコーダ(VAE)の潜在空間の各クラスターの中心を身体運動の概念として位置付け, その象徴的な身体運動を生成することで, 人の行為の意味を再現しようとする。これらの模倣が成功しない場合, Alter3 は Memory モードに移行し, 場の雰囲気をもとに記憶を呼び起こして動きを生成する。目の前に被験者がいない場合, Alter3 は目を閉じて記憶から動きを取り出し, 実行することで身体動作の探索を行う(Dream)。実験の結果, Mimicking だけでも, Imitation だけで模倣するよりも, 3 つの戦略を使い分けることで, VAE の潜在空間のクラスターが安定して, かつ身体動作のクラスターと対応することが示された。また, Alter3 が Dream モードを通じて自律的に生成する動きが, Imitation を通じて生成された象徴的な動きから進化していることが示唆された。これらの



図 4. オルタ 3 を用いた模倣実験の様子

発見は, 機械的な How の模倣(Mimicking)と, 身体動作の意味的な What の模倣(Imitation)を選択的に使用することで, 身体運動の概念形成につながり, さらにそれらの記憶から新たな動きをの探索が生じるといったように発展的に概念形成をおこなっていくことが示唆された。

書誌情報: Takahide Yoshida, Atsushi Masumori, Norihiro Maruyama, John Smith, Takashi Ikegami, Development of Concept Representation of Behavior through Mimicking and Imitation in a Humanoid Robot Alter3, Proceedings of The 2023 Conference on Artificial Life (in press).

升森 敦士(東京大学/A03 班 行動決定モデル推定研究グループ公募班代表者)

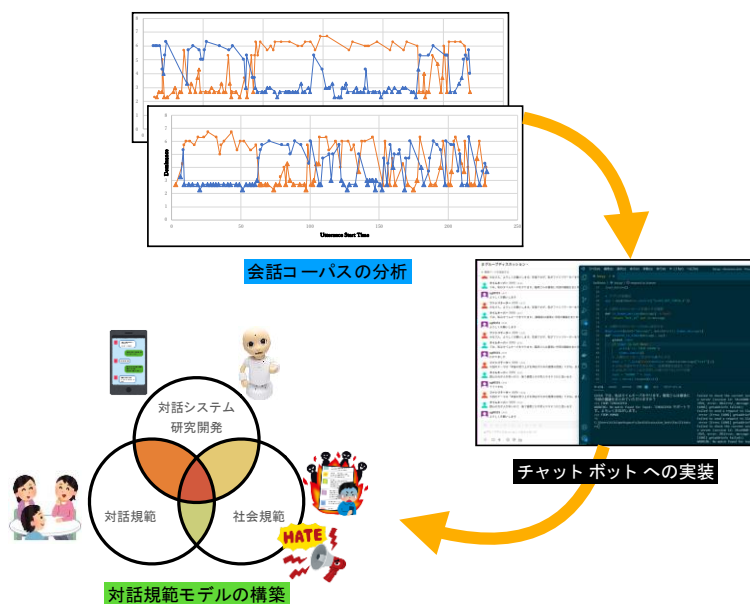
■「権利」中心的対話モデルの実装と応用に向けて(A04 班)

対話システムの研究開発や、それに関わる対話規範および社会規範についての検討には、「対話とは何か」という問いに一定の答えを与えておくことが求められる。しかし、システムの研究開発と対話規範の研究の両方にとって満足のいくような答えがどのようなものかは全く明かではない。本研究で目指すのは、対話知能システムと社会規範の関係を明らかにし、それに基づいて人間と機械が共生する社会に求められる新しい社会規範の基盤を提供することである。具体的には、人と人との対話に関する哲学、社会学、法学などの研究を元に、人とロボットの対話にも適用可能な対話モデルを構築し、その実装および応用を目指す。

そもそも対話モデルとは、対話についての見方(view)ないし捉え方(conception)を提供するものであり、例えば「人間同士の対話は、人間以外との間でも実行しうる音声によるコミュニケーションや図像によるコミュニケーションとどの点で異なるのか?」という問いに答えを与えるものでなければならない。この問いに対して本研究では、対話とそれ以外のコミュニケーションを社会規範の有無によって区別することを提案する。すなわち、対話は本質的に社会規範を伴うものであり、それゆえ、必ずしも社会規範を伴うとは限らない他の(非言語的)コミュニケーションとは区別される。具体的には、対話では参加者の間で、社会規範を伴う何らかの社会的関係(上下関係、地位の違い、それぞれどのような立場で対話に参加しているかなど)が成立していることが前提されているのに対し、対話以外のコミュニケーションは、こうした社会的関係抜きで成立しうる。

対話をこのような仕方では特別視するのは、対話が、社会規範に関わる特殊なコミュニケーションとして規定可能だからである。コミュニケーションは、身体的接触のような純粋に物理的なインタラクションと比べて社会的側面に特徴がある。すなわち、コミュニケーションの場合、参加者の間で何らかのプロトコルが共有されており、プロトコルに従っていなければインタラクションが発生しない。だが、このような社会的側

面があるからといって社会規範を伴うとは限らない。なぜなら、プロトコルに従っているかどうかは事実に関する問題であるのに対し、規範とは「すべき」と「すべきでない」に関する問題だからである。だから、例えば、プロトコルに従うべきなのか従うべきでないのかのような価値判断に関わっていないければ社会規範を伴っているとはいえない。それに対し、対話の遂行には、このような価値判断が不可欠である。なぜなら一般的なコミュニケーションとは異なり、対話では発言の適切性や対話の目的といった規範に関わる対話への移行がいつでも生じうるからである。



対話がこのように社会規範から切り離せないことは、対話の参加者は常に何らかの「権利」の主体として参加しているとみなすことにより説明できる。このような「権利」中心的な対話モデルを採用することにより、発言権のような対話規範、およびヘイトスピーチ、いじめ(無視)のような対話に関わる社会問題に適切に取り組むことが可能になると考えられる。

現在、「権利」中心的な対話モデルの実装可能性について、会話コーパスの分析をもとに、まずはチャットボットへの実装を進めている。

小山 虎(山口大学/A04 班 人間機械社会規範研究グループ公募班代表者)

2. イベント報告

■ 対話システムライブコンペティション 6 キックオフミーティング

2023 年 6 月 24 日(土)に標記のオンラインイベントを開催した。「対話システムライブコンペティション 6」のウェブページ(下記)から動画と資料を閲覧できる。 <https://sites.google.com/view/dslc6/%E3%82%AD%E3%83%83%E3%82%AF%E3%82%AA%E3%83%95>

「対話知能学」領域からは、東中竜一郎(A02 計画班研究代表者)が登壇し、当コンペティションの概要を説明するとともに、貸与するソフトウェア(新学術で開発したもの)の紹介を行った。

A02 班の活動の一環として実施している「対話システムライブコンペティション」とは、オーディエンスの前で実際に対話システムを動作させ、評価を行うイベントである。前回に

引き続き今回も、マルチモーダル対話システムを対象とする。これまでは、任意の話題についてユーザと雑談を行う能力を競うオープントラック、特定のシチュエーションの上で状況にあった対話をする能力を競うシチュエーショントラックがあった。しかしながら、昨今の大規模言語モデルの進展に鑑み、任意のトピックで雑談をするオープントラックは一定の役割を果たしたと考えたため、今年度はシチュエーショントラックのみを実施する。シチュエーショントラックは、対人関係や TPO を踏まえた高度な対話システムの実現の第一歩となることを狙いとする。また、高度な対話を実現するため、ChatGPT をはじめとする大規模言語モデルの活用を推奨している。

3. 顕著な業績

■ 論文

岡田将吾(A02 公募班研究代表者)らによる、クロスコーパス感情認識のための転移学習についての論文が Affective Computing 分野のトップジャーナルである IEEE Transactions on Affective Computing に採択された。

異なる言語圏(英語と中国語)・対話設定(人と人、人とシステムの対話)で取得された音声データコーパス間では分布が異なるため、一方のデータで訓練、他方のデータでテストした場合(クロスコーパスタスク)、音声感情認識の精度が著しく低下する問題が生じる。本論文では、従来のモデルがクロスコーパスタスクに与える影響を分析し、異なるコーパス間で共有特徴分布を学習することを目的とした ADoGT (adversarial domain generalized transformer)を提案し、有効性を検証した。将来、異なる言語圏・対話設定で取得された複数の少量データセットから学習する手法として期待できる。

書誌情報: Yuan Gao, Longbiao Wang, Jiaxing Liu, Jianwu Dang, and Shogo Okada. Adversarial Domain Generalized Transformer for Cross-Corpus Speech Emotion Recognition, IEEE Transactions on Affective Computing (in press).