

対人インタラクションのための人の 心的状態推定システムの研究

Study on Human Mental Status Estimation System for Interaction with Human

肥田 竜馬¹ 山田 徹志² 宮田 真宏¹ 大森 隆司³

Ryoma HIDA¹, Tetsuzi YAMADA², Masahiro MIYATA¹, Takashi OMORI³

¹ 玉川大学 工学研究科

¹Graduate School of Engineering, Tamagawa University.

² 玉川大学 脳科学研究所

²Tamagawa University Brain Science Institute.

³ 玉川大学 工学部

³ College School of Engineering, Tamagawa University.

Abstract: エージェントがユーザーの心的状態を推定する事が人の心に寄り添う HAI の構築に不可欠である。しかし、そのような人の心的状態推定は工学技術としては実現されていない。そこで我々は子どもを対象とし、心的状態の一つであると考えられる関心の推定を試みる。そのためにまず、子どもの関心の状態を保育経験者より評定し、それと同等の結果を AI 技術による解析から再現する方法を開発している。これより、人の関心状態を定量化し、子どもの、ひいては人の行動選択の因果関係を推定し、HAI におけるユーザー理解の高度化を目指す。

1 はじめに.

対人インタラクションを行う上で、対象人物と良好な関係を築く為に、相手の行動を予想する事は非常に重要である。近年、Pepper や Siri 等のロボットや対話型 AI 等が実用化されている。これらは、音声認識技術の精度向上や、人の発話に対する応答が比較的インテリジェントに設計されている。しかし、発話者の意図や感情を認識している訳ではない。長期的には我々は人の意図や感情を理解し、その上で発話内容や行動決定を行う事ができるエージェントの構築を目指す。そのため本研究は、対象人物の心的状態の推定を実現することで、定型的なやり取りを脱した HAI の実現の道を探る。

このようなエージェントの構築においては、いかにして対象人物の状態推定を行うかが鍵になる。現在の対人センシングは、日々新しい技術が開発されている。その代表例として、KinectV2[1]や、CMU

Panoptic Studio[2]のような人の骨格情報のセンシング技術がある。さらにその発展として Deep Learning の関数近似技術を用いた OpenPose 等が挙げられる[3,4]。これらのシステムは、人の計測を通して現実世界の出来事を幅広く捉える事ができる。しかし、これらのシステムから得られる情報は、物理量の推定でしかなく、その解釈はユーザーに委ねられている。これらの情報を用いた効果的な対人インタラクション技術は開発途上である。

一方、人同士がインタラクションを行う際は、五感から得られる物理量を元に、相手の状態推定や自身の行動決定を行っていることは間違いない。そこには、物理量を元にしたコミュニケーションの意思決定がある。このインタラクションのための情報抽出と意思決定の過程を検討、解析する事によって、この手法を HAI タスクの中に取り入れる事ができるなら、HAI の対人インタラクションの精度はさらに向上するものと考えられる。

このような背景のもとで我々は、子ども達の心的状態に基づくインタラクションの構成要素を検討、解析するために、フィールドの観察を行った。また、子ども達の行動の解釈においては、多数の保育経験

連絡先: 玉川大学工学研究科

〒194-8610 東京都町田市玉川学園 6 丁目 1-1

E-mail: hidar2is@engs.tamagawa.ac.jp

者からの聞き取り調査を元に、注目すべき観点を絞り込み、その後、3名の保育経験者（実務経験5年以上）によるアノテーションを行った。

そして、アノテーションによる動画の解析により得られたデータを元に心的状態推定に関連する物理量を抽出した。その後、子どもの心的状態推定に関連する物理量を見出す事を目指したので、その過程を報告する。

2 インタラクションにおける心的状態の有用性

2.1 心的状態としての「関心」

人は他者の心的状態（より広く内部状態）を、多くの場合は視覚や聴覚などの感覚情報から推定し、インタラクションに使用している。特に視覚情報は重要であり、我々は数メートル離れていても他者の内部状態を推定できる。より具体的に言えば、他者の内部状態の推定には、その人物の身体動作や表情、発話などの情報や、周辺にある物体・イベントや時間的文脈などの外的情報が利用されると考える。しかし、特定の状況で同様の行動をとったからといって、内部状態が同一とは限らない。これを工学技術を用いて推定する事は現状では困難であるが、人は日常的に内部状態の推定をし、インタラクションをしている。これより、観察された情報から内部状態を推定するには、解釈する側に何らかの解釈モデルが介在していると考えられる。しかし、人同士のインタラクションでは、我々はこの解釈モデルを意識する事なく使用し、推定している。

山田らは、保育士の専門性に注目し、子どもの心的状態を推定するための心的状態推定モデルを提唱している[5]。そのモデルでは、外部センサにより観測可能な子どもの行動や環境などについての物理量から子どもの心的状態を推定する事が可能であると想定している。その中で、子どもの心的状態の構成要素の1つとして「関心」が挙げられている。関心は、その対象者が何に価値を感じているかを示す、行動決定の重要な要素である。行動経済学や認知科学では価値の評価は人の行動選択に大きく影響する要素であるとされている[6,7]。この関心に関わる人（本研究では特に子ども）の行動を解明していく事ができれば、人の内部状態推定のモデル構築に大きく貢献できると考える。そのため本研究では、子どもの関心に注目した。

2.2 関心アノテーションによる評価。

内部状態（関心）の推定に入る前に、まず関心を

推定することの有用性についての検討が必要である。そこで、保育経験者による子どもの関心に関するアノテーションを行った。対象場面は、幼稚園での英語クラスのビデオ映像18分53秒間である。その映像中のすべての子ども（18名）に対して、アノテータ3名（保育実務経験者）がその一つひとつの判断について協議し、子どもの関心の対象を1秒ごとに6種類の評定項目を設け分類した。この内、特徴的なシーン2分間を選定し、子どもの内部状態（関心）についての頻度分布を作成した（図1）。

この頻度分布をそれぞれの子どもの関心の特徴を表すベクトルと考え、この集団の関心分布を表す正規分布モデルを作り、ナイーブベイズ法により、各子どもの関心ベクトルの平均ベクトルからの距離を求めた（図2）。グラフの左端がlog尤度0、すなわち平均値であり、右に行くにしたがって関心の分布が平均から離れていく。図中の点は個々の子どもの行動のlog尤度である。

その結果、1名の子どもが平均的な関心分布から大きく離れている事が発見された。この結果を元にこの幼稚園にインタビューした所、対象の子どもに配慮が必要であることを保育者（子どもの担任）が認識していることが判明した。これらより、関心についての6分類のアノテーションは、子ども達を日々観察している保育者の経験則による状態推定における有用な指標と考えられる。次に、同映像について他の保育経験者が独立にアノテーションしたところ、87.5%の一致率が得られた。これより、この関心の分類カテゴリは保育士の間で共有された特徴を捉えていることが示唆される。

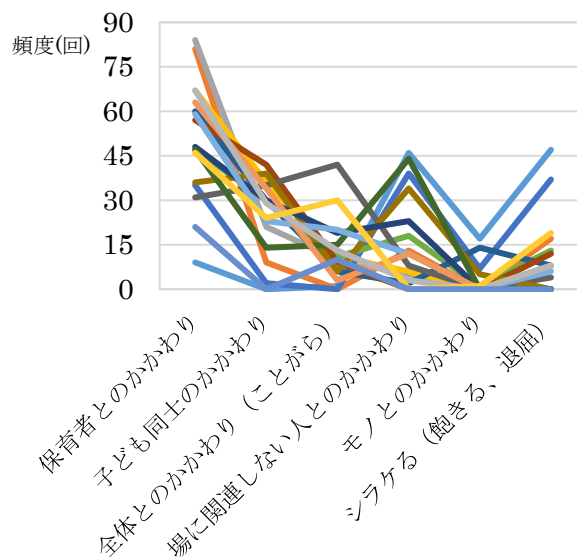


図1: 子どもの関心分布. 120秒の映像中の個々の子どもの関心を1秒ごとに分類した。個々の線が個々の子どもに対応する。

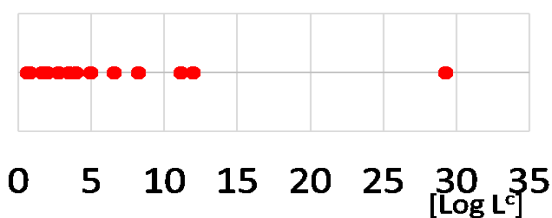


図 2: 個々の子どもの関心の log 尤度の分布.

3 心的状態推定システムの開発

子ども達の身体情報を収集するために、RGBD カメラ (KinectV2) を保育の活動が行われる部屋の四隅に設置し、カラー画像と距離画像を記録した。前処理としてカメラ間の同期処理を行い、個々の子どもの特徴抽出を試みた[8].

当初は、KinectV2 の距離画像から Point Cloud を作成し、四方向からの Point Cloud を合成してオクルージョンを回避し、個体の分離を行うことを計画していた[9]. しかし 2017 年 5 月に OpenPose が公開されたため、OpenPose による骨格の推定を行った[3,4]. OpenPose は、カラー画像から人の骨格を推定する事ができるツールであり、カメラの解像度に応じて計測可能範囲も伸び、幼稚園や保育園のような比較的広い空間での多数の人物の発見・追跡および姿勢の測定には有用である.

OpenPose は、画像中の身体像を検出し、さらにその骨格情報を画像から認識できる。(図 3 右). その出力は身体各関節のカメラ画像中の二次元座標であり、そのままでは三次元的な位置とはならない。そのため、4 方向から見ているカメラ 4 台のうち隣合わせの 2 台からの視線の向きとそのカメラの世界座標上の座標と角度から、三角法により各子どもの骨格情報の 3D 再構成を行った (図 3 左). カラー画像で観察された骨格と比較して 3D 再構成された骨格の数は約半分である。これは、2 台のカメラの方向が 90 度ずれており、オクルージョンが多く発生して 2 台から同時に見えている身体が少ないことによる。

また、OpenPose で検出された個体の識別のために OKAO Vision による顔認識を行った (図 4). これにより、骨格情報に個人を識別する情報が付与でき、オクルージョンなどのために骨格情報の時系列が途切れても連続的な追跡が容易になる。

現在までの計測では、KinectV2 のカラー画像センサを使用しているが、子どもの顔の認識可能距離が 5m 程度と保育室に対して短いため、今後はさら



図 3: OpenPose による骨格の検出と(右), 三角法による 3D 空間への投射(左).

に高解像度で観察範囲の広い 4K カメラの使用を検討している。これにより、顔の検出件数の増加と個人識別の精度向上を目指す。しかし、認識可能距離と視野の広さはトレードオフの関係にあり、目的とするフィールドに合わせての視野の広さの選択が必要である。

OpenPose の骨格推定では頭の向きの推定は不安定である。そこで 3D 再構成された骨格情報から子どもの体の向きを推定した。その精度については検証中であるが、骨格推定の精度に大きく依存することは言うまでもない。



図 4: OKAO Vision による表情検出と個人識別.



図 5: OpenPose の骨格座標の 3D 再構成による子どもの体の向きの推定.

4 今後の課題

対人インタラクション場面における他者の心的状態推定のための予備的実験と、推定のためのセンシングシステムについて述べた。心的状態のうち「関心」を示す行動は、人の意思決定にかかわる価値の知覚の表出とも考えられ、他者の心的状態を表す重要な指標と考えることができる。

また、先行研究として集団行動の中で関心行動の分布は、個々人の特性を表すものであることが発見されている[10]。今回は英語教室という場面での全体平均からのずれのみを報告したが、個々の活動場面でのずれを評価すると、個々人のある種の特性を観察できるのではないかと考えている。その方法については今後の課題である。

子どもの位置と向きへの推定については、現在、アノテーションを人手でも並行して行っており、この結果と自動推定の結果を比較する事で、推定精度の検証を行う。この推定に多少の誤りはあるとしても人による修正を追加すれば利用可能という状態になれば、子どもに限らず人の行動研究の労力が大幅に削減されると期待される。本手法の当面の適用領域はそのような人間行動観察研究であろう。

センシングで用いた RGB センサは単眼カメラを4方向から囲むように設置しているため、オクルージョンには強い一方、3D再構成の際には同一人物が2台以上のセンサで撮影されている必要があり、人数が多い場合には骨格情報の検出件数が頭打ちになっている。これについても今後の課題とする。

以上より、心的状態推定に関連する情報を収集するシステムを安定的に稼働させ、その場にいるすべての人物の骨格情報、身体情報、表情、個人識別の情報収集を目指す。これらの技術を統合することで匿名化されたデータの収集が可能となり、それに基づいた工学的解析をより簡易に行えるようになると期待する。

謝辞

本研究の一部は産業技術総合研究所人工知能研究センターからの委託研究により実施された。

参考文献

- [1] Microsoft: Kinect Sensor for Xbox One, https://www.microsoftstore.com/store/msusa/en_US/pdp/Kinect-Sensor-for-Xbox-One/productID.2267482500.
- [2] CMU Panoptic Dataset: CMU Panoptic Studio <http://domedb.perception.cs.cmu.edu/>.
- [3] Shih-En Wei, Varun Ramakrishna, Takeo Kanade, Yaser Sheikh: Convolutional pose machines, wei2016cpm, (2016)
- [4] Zhe Cao, Tomas Simon, Shih-En Wei, Yaser Sheikh: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields, cao2017realtime, (2017).
- [5] 山田徹志, アッタミミ・ムハンマド, ジャン・ビン, 宮田真宏, 中村友昭, 大森隆司, 長井隆行, 岡夏樹, 西村拓一: 保育の質の定量化に向けた子どもとロボットの関わりー子どもの心的状態推定へのアプローチ, 2016年度日本認知科学会第33回大会, OS13-4, pp.358-362, (2016).
- [6] マッテオ・モッテルニーニ, 2008, 『経済は感情で動く はじめての行動経済学』(泉 典子訳), 紀伊国屋書店.
- [7] エヤル・ヴィンター, 2017, 『愛と怒りの行動経済学 賢い人は感情で決める』(青木 創訳), 早川書房
- [8] Ryoma Hida, Tetsuji Yamada, Masahiro Miyata, Takashi Omori: Development of Interest estimation Tool for effective HAI, HAI 2017 - Human-Agent Interaction, pp. 483-486, (2017)
- [9] PCL: Point Cloud Library, <http://pointclouds.org/>.
- [10] 山田徹志, 宮田真宏, 肥田竜馬, 大森隆司: 子どもの主体的な行動を通じた保育の質の客観化手法の検討ーAIを用いた子どもの行動計測と心的状態推定ー, 2017年日本発達心理学会第28回大会, P4-4, (2017).