

コミュニケーションロボットを用いた適応的人-XAI インタラクションに向けて

福地庸介^{1*} 山田誠二^{2,3}

¹ 東京都立大学

¹ Tokyo Metropolitan University

² 国立情報学研究所

² National Institute of Informatics

³ 総合研究大学院大学

³ The Graduate University for Advanced Studies (SOKENDAI)

Abstract: コミュニケーションロボットによって、人と説明可能 AI (XAI) のインタラクションを促進できる可能性がある。しかし、動的に進行する人と AI の協調において、文脈に合わせてロボットの身体表現を適応的に活用する方法は明らかではない。そこで本研究は、コミュニケーションロボットが身体表現を用いて、説明のどの部分を強調すべきかを適応的に決定する手法を提案する。この手法では、特定の部分を強調した際のユーザーの判断への影響を予測することで、ユーザの判断と AI が提案する判断の差を減らそうとする。ユーザスタディの結果、AI の提案のパフォーマンスが高い場合に、本手法がユーザーをより良い意思決定へと導く可能性が示唆された。

1 はじめに

説明可能 AI (XAI) は、AI の推論過程がブラックボックス化し、人間が理解できなくなるという問題 [1] に対する解決策として、盛んに研究されている。大規模言語モデル (LLM) によって、AI の推論結果をサポートする説明を大量に生成可能することができるようになった [2]。説明によって AI の推論の過程や根拠をユーザーが理解できるようになれば、AI の性能を過信して失敗を招いたり、逆に不信を抱いて AI 利用を忌避することを防ぐことができ、AI の適正な利用につながると期待される [3]。

本研究では、人と XAI のインタラクションにおいてコミュニケーションロボット (以下、ロボット) を活用することを考える。ロボットは、人間と XAI の効果的なインタラクションを実現する上で有望なインタフェースとなり得る [4, 5]。例えば、単に説明を提示するだけでなく、身体や音声的な表現によってニュアンスを付加することで、より豊かで効果的なコミュニケーションが実現可能だと期待される。

XAI のインターフェースとしてロボットを活用する上で、ロボットの表現を状況に応じていかに選択するかが問題になる。ジェスチャーや声色といった表現には解釈の幅が大きく、状況や相手次第で、むしろミス



図 1: コミュニケーションによる説明の強調

リードを引き起こす可能性もある。そのため、AI の予測における重要度などといった AI 側の要因だけでなく、ロボットの表現がユーザにどう受け取られるか、相手や文脈を考慮した上で、適切な表現を選択する必要があると考えられる。

そこで本稿では、著者らが提案する、ロボットが身体表現を用いて説明のどの部分を強調すべきかを適応的に決定する手法の検証結果を報告する。検証では、株式取引支援を例に、ロボットが今後の株価の値動きを説明する場面を考える。提案手法は、説明の強調がユーザの株取引に与える影響を、機械学習によって予測する。そして、予測されるユーザの売買の数量が AI が推奨するものに近づくように、説明において強調す

*連絡先: 東京都立大学システムデザイン学部情報科学科
〒191-0065 東京都日野市旭が丘6丁目6
E-mail: fukuchi@tmu.ac.jp

るポイントを決定する [6, 7, 8]。ここでいう強調表現は、ロボットの腕振りと声色の変化として実装されている (図 1)。

2 検証結果

検証では以下の 3 条件を比較した：

1. 強調なし (FLAT)
2. 常に正確な予測を行う AI が推奨株数を決定していると仮定し、それにユーザの売買を合わせようとする条件 (ORACLE)
3. 強化学習 (RL) で得られた方策をもとに推奨株数を決定する条件 (RL)

結果、ORACLE では参加者の収益の平均値が大きく伸び、タスクパフォーマンスを向上させる効果が示唆された。ORACLE の結果は、中央値では FLAT と大きく変わらないものの、上位 25% の参加者の成績が伸び、また下位 25% の損失額を圧縮する効果があった。

一方、RL は、FLAT と比べた顕著な改善を確認できなかった。結果をさらに分析したところ、参加者の約 31% は AI の推奨する取引量と実際の取引量が強く相関した一方で、約 27% は AI が売りを推奨している際に即時的に対応できず、約 22% はむしろ AI の推奨に反する取引を行うことで、パフォーマンスを低下させたことがわかった。今後は、このようなユーザの特性を考慮したロボットの表現方法の選択が重要であると考えられる。

謝辞

本研究は、JST CREST (JPMJCR21D4)、JSPS 科研費 (JP24K20846) の支援を受けた。

参考文献

- [1] Raja Parasuraman and Victor Riley. Humans and automation: Use, misuse, disuse, abuse. *Human Factors*, 39(2):230–253, 1997.
- [2] Sarah Wiegrefe, Jack Hessel, Swabha Swayamdipta, Mark Riedl, and Yejin Choi. Reframing human-AI collaboration for generating free-text explanations. In Marine Carpuat, Marie-Catherine de Marneffe, and Ivan Vladimir Meza Ruiz, editors, *Proc. of the 2022 Conf. of NAACL*, pages 632–658, Seattle, United States, July 2022. ACL.
- [3] Yosuke Fukuchi and Seiji Yamada. Dynamic selection of reliance calibration cues with ai reliance model. *IEEE Access*, 11:138870–138881, 2023.
- [4] Tim Miller. Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*, 267:1–38, 2019.
- [5] Marco Matarese, Francesca Cocchella, Francesco Rea, and Alessandra Sciutti. Ex(plainable) machina: how social-implicit xai affects complex human-robot teaming tasks. In *IEEE ICRA 2023*, pages 11986–11993, 2023.
- [6] Yosuke Fukuchi and Seiji Yamada. User decision guidance with selective explanation presentation from explainable-ai. In *2024 33rd IEEE International Conference on Robot and Human Interactive Communication (ROMAN)*, pages 338–343, 2024.
- [7] Yosuke Fukuchi and Seiji Yamada. Towards adaptive explanation with social robot in human-xai interaction. In *Proceedings of the 12th International Conference on Human-Agent Interaction, HAI '24*, page 353–355, New York, NY, USA, 2024. Association for Computing Machinery.
- [8] Yosuke Fukuchi and Seiji Yamada. Should xai nudge human decisions with explanation biasing? In *The 9th issue of the International Conference Series on Robot Ethics and Standards (ICRES)*, pages 199–207, 2024.