

ロボットの行動可能性に基づいた 自然言語指示の目標状態推定

○佐藤夏樹（慶大院） 川崎陽祐（慶大院） 高橋正樹（慶應大）

1. 緒言

近年の生活支援ロボットの非専門家ユーザに対する需要の増加に伴い、生活支援ロボットに対する自然言語でのコミュニケーションの要求が高まっている。自然言語は強力なインターフェースである[1]ことから、非専門家ユーザであっても生活支援ロボットと容易にコミュニケーションが取れる可能性があるためである。本稿では特に、自然言語を使用したロボットへの指示である、自然言語指示に着目する。

自然言語指示からロボットが行動を遂行するためには、実空間における行動計画に必要な目標状態を推定する必要がある。目標状態とは、ある指示がなされた際に、その指示を達成したとみなせる環境の状態のことを示す。しかし、自然言語指示には明確な情報は含まれず曖昧である[2]ことから、自然言語指示のみから目標状態を推定することは困難である。

従来、自然言語指示を想定したロボットの研究[3]においては、環境におけるロボットの行動可能性が考慮されていないため、実空間を考慮した目標状態や、行動ができない際の対応が困難であった。そのため、自然言語指示をより理解するためには、ロボットの行動可能性を検討する必要がある。

また、目標状態推定において課題となる自然言語指示の曖昧性として、指示が意味する実空間の動作の曖昧性がある。加えて、人間は実空間の状態に対して曖昧に指示している可能性があることから、自然言語指示が意図する目標状態が実現不可能な場合が存在する。つまり、環境におけるロボットの行動可能性を考慮することで、自然言語指示を実空間において解釈し、実現不可能な目標状態が指示された際にも行動できるシステムが必要とされている。

本研究では、環境におけるロボットの行動可能性に基づき、人間の自然言語指示から目標状態を推定するシステムを提案する。また、想定実験を通じて提案システムを検証する。

2. 提案システム

人間の生活環境においてロボットに自然言語指示を行うことを想定する。環境センシングによって、環境中の物体の位置と名称が既知であり、自然言語指示から対象物、場所、行動が既知であるとする。

本提案システムで用いる環境表現を図1に示す。図1左上の Scene graph とは、物体と物体同士の関係を環境情報で表現した有向グラフのことである。ノードが物体 (object)、エッジが関係性 (relation) を表現している。人間の指示の言語的な理解や、目標状態の推定のために使用した。図1右上の Action graph とは、時空間的な行動の可能性の表現と理解を表す

Scene Graph G^S :
物体と物体同士の関係の表現

Action Graph G^A :
時空間的な行動の理解と表現

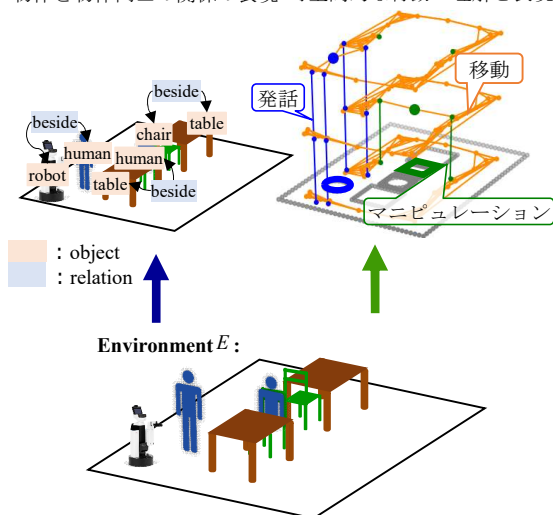


図1 環境表現

グラフのことである[4]。ノードが行動の実行場所を、エッジが行動の実行を表現している。多層化を行うことによって、環境変化を含む行動も表現することが可能になっている。本研究においては、ロボットの行動可能性の表現のために使用した。

システムへの入力は、自然言語指示から得られた対象物、場所、行動に加え、環境情報である。出力は、自然言語指示の意図した目標状態となる尤度付き Scene graph である。

システムの流れとして、はじめに、自然言語指示の対象物、場所、行動から、対象物と場所間に存在する relation のランキングを得る。続いて、Action graph で環境におけるロボットの行動可能性を示し、行動可能な環境変化後の環境を取得する。得られた環境から Scene graph を作成し、その際に、物体間の relation らしさを算出する。最後に、行動可能性の Scene graph 内の対象物と場所の間にある relation を抽出し、自然言語指示から得た relation のランキング内でのスコアと、relation らしさを積算し、各行動可能な Scene Graph の尤度を得る。

3. 実験

3.1. 実験目的

指示された対象物に対する行動が制限されている、もしくは到達することが困難である環境で提案システムにより推論を行う。言語的な推論と環境におけるロボットの行動可能性の両方を考慮していることの効果を検証する。

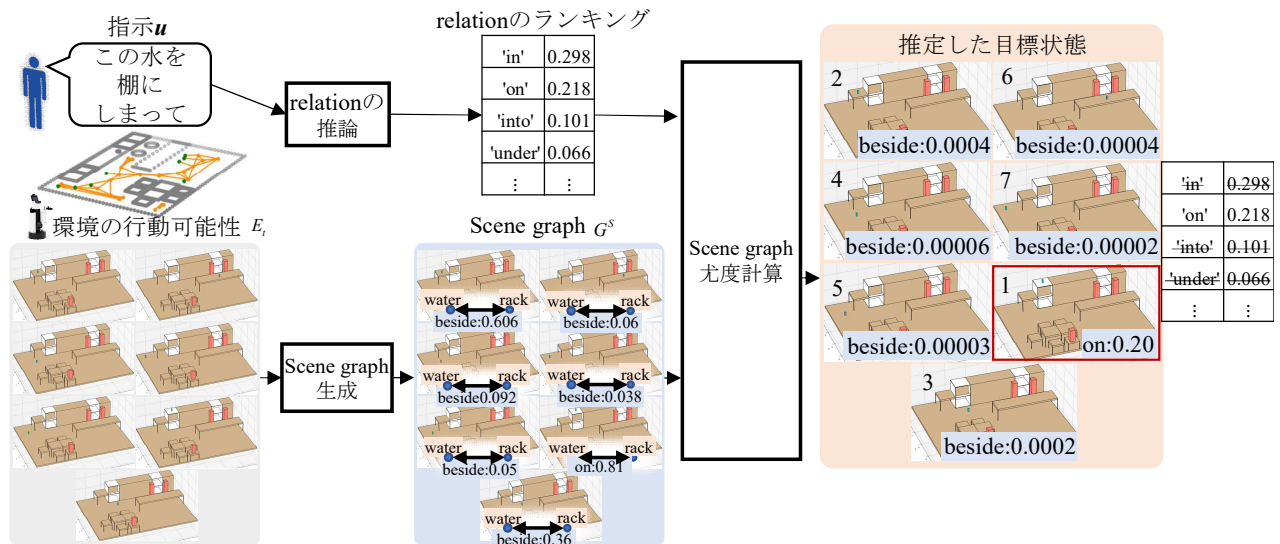


図 2 実験結果

3.2. 実験条件

自然言語指示として「この水を棚にしまっ」という指示がロボットに与えられ、対象物を **water**, 場所を **rack**, 行動を **put** として既知であるとする。

HSR[5]などに代表される、マニピュレーションにより **on**, **in**, **beside** が実現可能なロボットを想定した。

環境中の物体として、棚、テーブル、水、箱、人を想定し、アフォーダンスとして、棚は **placeable_on**, **placeable_in**, **placeable_beside** を持ち、テーブルは **placeable_on**, **placeable_beside** を持つ。水、箱、人は本実験において関係するアフォーダンスを持たないものとした。

3.3. 結果と考察

実験結果を図 2 に示す。上部の表に言語的な **relation** のランキングの結果を、下部に各行動可能な環境における **relation** らしさの結果を、右に提案システムの出力した尤度と順位の結果を記載する。

言語的な推論にてランキングが高い **in** が、行動可能な環境として存在しないため出力されず、次点で高い **on** が行動可能であり、推論結果として尤度が最大で出力されていることが確認できた。また、**relation** らしさを判断することによって、同じ **beside** という **relation** の中でも自然言語指示の解釈に近い順番に目標状態の仮説を持てることが確認できた。

4. 結論

環境におけるロボットの行動可能性に基づき、自然言語指示から目標状態を推定するシステムを提案した。想定実験を通して、言語的な推論と環境におけるロボットの行動可能性を活用し、目標状態の推論が可能になっていることを確認した。

今後の課題として、人間の意図に沿った行動ができていないか感性評価を行い、本提案システムの推定

結果を検証する必要があると考えられる。また、対象物の特性と周辺にある場所との関連を活用し仮説の尤度を算出することで、システムが持つ仮説を改良することが可能であると考えられる。そして、言語的な自然言語指示の解釈にて上位でありながら実行できない **relation** を対話へ活用することで、新たなインタラクションとして、ユーザの満足度へ影響を与えられると考える。

謝辞

本研究の一部は、科学技術振興機構 (JST) 戦略的創造研究推進事業 (CREST) 「文脈と解釈の同時推定に基づく相互理解コンピューテーションの実現」によるものであることを記し、謝意を表す。

参考文献

- [1] M. Shridhar, D. Mittal, and D. Hsu, "INGRESS: Interactive visual grounding of referring expressions," *Int. J. Rob. Res.*, vol. 39, no. 2–3, pp. 217–232, 2020.
- [2] M. A. V. J. Muthugala and A. G. B. P. Jayasekara, "A review of service robots coping with uncertain information in natural language instructions," *IEEE Access*, vol. 6, pp. 12913–12928, 2018.
- [3] D. K. Misra, J. Sung, K. Lee, and A. Saxena, "Tell me Dave: Context-sensitive grounding of natural language to manipulation instructions," *Int. J. Rob. Res.*, vol. 35, no. 1–3, pp. 281–300, 2016.
- [4] 川崎陽祐, 望月竣介, 高橋正樹, "多層アクショングラフを用いた行動計画", 第38回 日本ロボット学会学術講演会, 2020.
- [5] T. Yamamoto, K. Terada, A. Ochiai, F. Saito, Y. Asahara, and K. Murase, "Development of Human Support Robot as the research platform of a domestic mobile manipulator," *ROBOMECH J.*, vol. 6, no. 1, 2019.