

## AI活用でロボットも日常動作が可能に

### ◆Google DeepMind、視覚と言語を行動に変換できる視覚言語行動モデルを発表

2023年7月28日、グーグル傘下のGoogle DeepMindは、視覚と言語を行動に変換できる世界初の視覚言語行動（VLA:vision-language-action）モデル「RT-2 (Robotics Transformer 2)」を発表した。現在、さまざまな分野でロボットは活用されているが、基本的にはあらかじめプログラムされた動作しかできない。高速で自動車を組み立てる能力を持つ高性能ロボットでも、「目の前にある三つの物体からリンゴを選び、それを青い箱に移動させる」といった一見単純な動作ができない。RT-2は、オフィスのキッチンで働く13台のロボットが17ヵ月かけて収集したデータで訓練された前身モデル「RT-1」をベースに開発された。ChatGPTを動かしている「GPT-4」などの大規模言語モデル（LLM）と同様に、ウェブ上のテキストと組み合わせることでロボットの汎化能力<sup>1</sup>を向上させ、初歩的な推論機能を持つようになったといわれている。

すなわち、ロボットがそのタスクをこなすように特別に訓練されていなくても、ウェブからの知識を利用してパターンを認識し、アクションが実行できる。例えば、従来のロボットにゴミを捨てさせたい場合、ゴミを認識し、拾い上げ、所定の場所に捨てるように訓練する必要があったが、RT-2はウェブデータからの学習でゴミが何であるかをすでに理解しており、訓練なしにゴミを識別し、捨てることができる。さらにRT-2は、バナナは食べ物だが、食べた後の皮はゴミになるといったことも理解している。今回RT-2モデルに対し、「バナナを3に移動させて」「テー

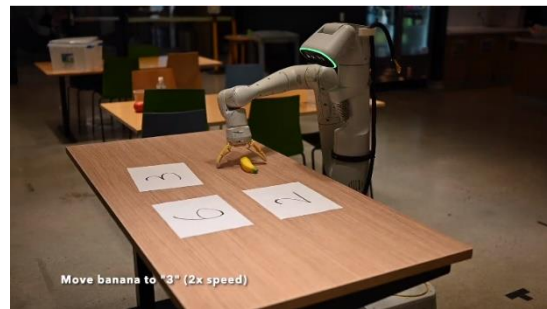


図. 1 RT-2実験の様子

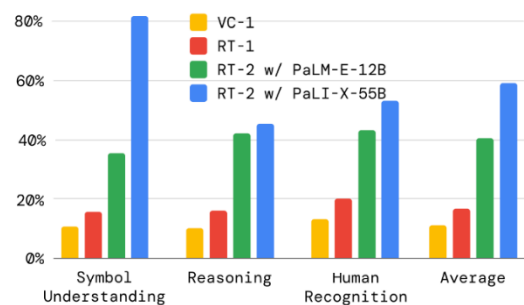


図. 2 RT-2定性的・定量的実験結果

出典：<https://robotics-transformer2.github.io/>

<sup>1</sup> 汎化能力：機械学習の訓練データにない未知のデータに対する推論の精度

ブルから落ちそうなバッグを拾って」といった6,000回を超える一連の定性的・定量的実験を行って、以前のRT-1モデルや、Visual Cortex (VC-1)のような大規模な視覚データセットで事前に訓練されたモデルなどと比較して汎化能力が3倍以上向上することを確認した（図. 1、2）。

RT-2は、AIの進歩がロボット工学に急速に接近していることを示すだけでなく、より汎用的なロボット開発にも大きな可能性を示している。

#### ◆AIロボットシステムが曲線も含む柔軟物のファスナー開閉作業を実現

23年8月23日エクサウィザーズは、JAXAと共同で実施している有人宇宙拠点内クルー作業の自動化・自律化に向けた技術検討の取り組みにおいて、さまざまな用途に合わせて学習できる、汎用性の高い模倣学習プラットフォームを開発したと発表した。従来のロボット技術では、資源輸送用バッグやケーブル類といった形状が変化する対象を通信の遅延がある環境で操作することは困難であるほか、操作対象のサイズや質量が作業の精度に影響を与えやすいという課題があった。また、ロボットが柔軟物のファスナーを開閉する作業は高度な模倣学習・予想学習技術を必要とするため、実現することは困難だった。

今回、同プラットフォームを活用し、エクサウィザーズ顧問である早稲田大学理工学術院・尾形哲也教授が論文発表した、「視覚と触覚の情報を併用する」基本アーキテクチャをもとに開発を行った。従来の技術ではロボットによる自律操作が困難だった作業のうち、ファスナー開閉のような柔軟物の作業の実用性を高めるため、触覚に代わり力覚（関節モータ電流から推定したトルク値）を用いて、柔軟物におけるファスナーの開閉作業を実現した。これに、直線だけでなく曲線も含む柔軟物のファスナー開閉作業を、対象物の位置を固定した環境下において学習させ、その固定位置における操作を100%精度で実現した。さらに、上記条件に加え対象物の位置を左右にずらして学習させ、その後、学習条件とは違う位置に動かした条件下においても、曲線を含むファスナー開閉作業を80%以上の精度で実現した。

今後は宇宙での適用に向けて、対象物の場所や形だけでなく、重力が変化する環境下においても高い精度で作業が実施できるよう開発を進めていく。

ロボット進化はAIの活用でより汎用性・柔軟性が高まっている。 【成田誠】