

AI 特許紹介(93)
AI 特許を学ぶ！究める！
～TOWARDS A UNIFIED AGENT WITH
FOUNDATION MODELS 特許～

2026 年 10 月 9 日
河野特許事務所
所長弁理士 河野英仁

「AI 特許紹介」シリーズは、注目すべき AI 特許のポイントを紹介します。熾烈な競争となっている第 4 次産業革命下では AI 技術がキーとなり、この AI 技術・ソリューションを特許として適切に権利化しておくことが重要であることは言うまでもありません。

AI 技術は Google, Microsoft, Amazon を始めとした IT プラットフォーマ、研究機関及び大学から毎週のように新たな手法が提案されており、また AI 技術を活用した新たなソリューションも次々とリリースされています。

本稿では米国先進 IT 企業を中心に、これらの企業から出願された AI 特許に記載された AI テクノロジー・ソリューションのポイントをわかりやすく解説致します。

1.概要

特許出願人 DeepMind

出願日 2024 年 2 月 15 日

公開日 2024 年 9 月 19 日

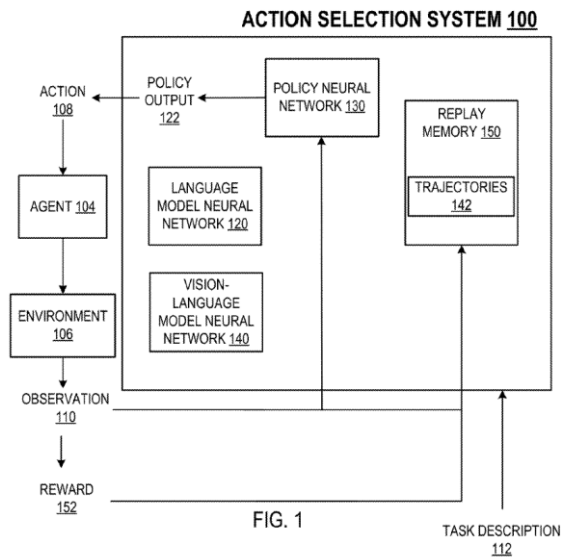
公開番号 US20240311617

発明の名称 言語モデルニューラルネットワークによって生成されたサブゴールを使用したエージェントの制御

617 特許はロボットに対するタスクを LLM によりサブタスクに分解し、サブタスクのポリシーに従ったアクション後に VLM から得られる観察結果がサブタスクを達成したか否かを判断することにより、順次サブタスクを実行する TOWARDS A UNIFIED AGENT WITH FOUNDATION MODELS 技術に関する。

2.特許内容の説明

下記図はアクション選択システム 100 のネットワーク構成図である。



アクション選択システム 100 は、言語モデルニューラルネットワーク 120 および視覚言語モデル (VLM) ニューラルネットワーク 140 を使用して、(i) 環境 106 内でタスクを実行するために環境 106 と相互作用するエージェント 104 (ロボット) の制御を改善する、または (ii) 環境 106 内でタスクを実行するために環境 106 と相互作用するエージェント 104 を支援する。システム 100 は、エージェント 104 を制御する際、当該タスクのエピソードの実行中における複数のタイムステップにおいて、エージェント 104 によって実行されるべきアクション 108 を選択することにより、タスクを遂行するようにエージェント 104 を制御する。

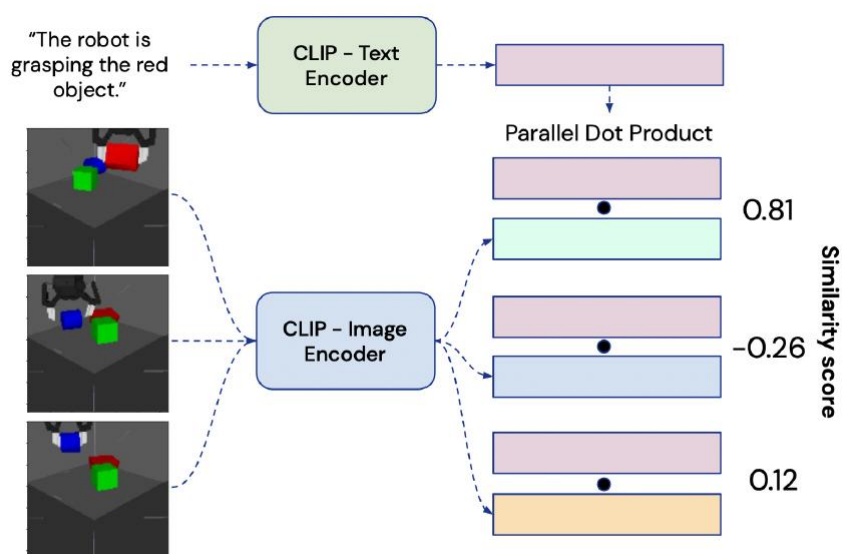
タスクエピソード中の各タイムステップにおいて、システム 100 は、そのタイムステップにおける環境 106 の現在の状態を特徴付ける観測 110 を受け取り、それに応答して、そのタイムステップにおいてエージェント 104 によって実行されるべきアクション 108 を選択する。エージェント 104 がアクション 108 を実行した後、環境 106 は新しい状態へと遷移する。システム 100 は、エージェントがアクションを実行したことに応じて、環境から外部報酬 152 を受け取る。

所定のタイムステップにおいてエージェント 104 を制御または支援するために、システム 100 は、例えばユーザから、タスクエピソード中にエージェント 104 によって実行されるタスクを記述した自然言語タスク記述 112 を受け取る。システム 100 がエージェント 104 を制御する際、システム 100 は、言語条件付きポリシーニューラルネットワーク 130 を用いて、現在の観測に応答してエージェントが実行すべきアクション 108 を選択する。

ポリシー出力 122 は、複数のアクションからなる固定セット内の各アクションについて、それぞれの Q 値を含む。システム 100 は、それらの Q 値を処理して（例えば、ソフトマックス関数を用いて）、各アクションについてそれぞれの確率値を生成する。この確率値は、アクションを選択するために使用されるか、あるいは、システム 100 は、最も高い Q 値を有するアクションを選択する。

システム 100 は、言語モデルニューラルネットワーク 120 によって生成された記述と、視覚言語モデル（VLM）ニューラルネットワーク 140 によって生成されたサブゴールおよび観測画像の埋め込みとを使用して、ポリシーニューラルネットワーク 130 の学習を改善する。特に、ポリシーニューラルネットワーク 130 を使用してエージェントを制御しタスクを実行させる前に、システム 100 は、リプレイメモリ 150 に格納された学習データを用いてポリシーニューラルネットワークを学習させる。

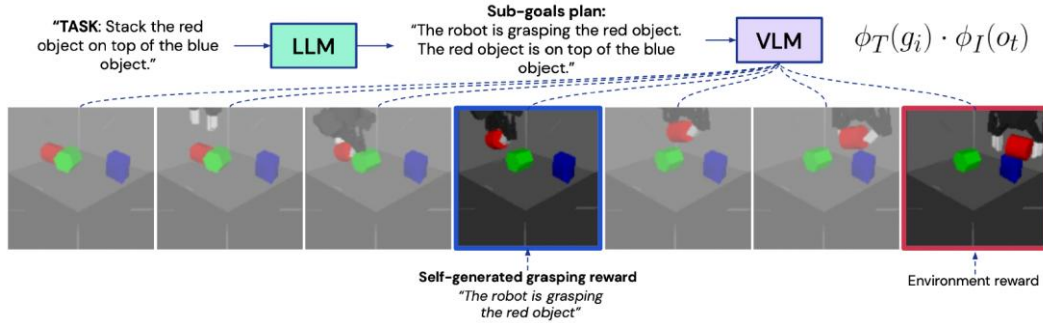
下記図は、VLM を用いて、観測画像とサブゴールとの間の類似度スコアを決定する例を示している。



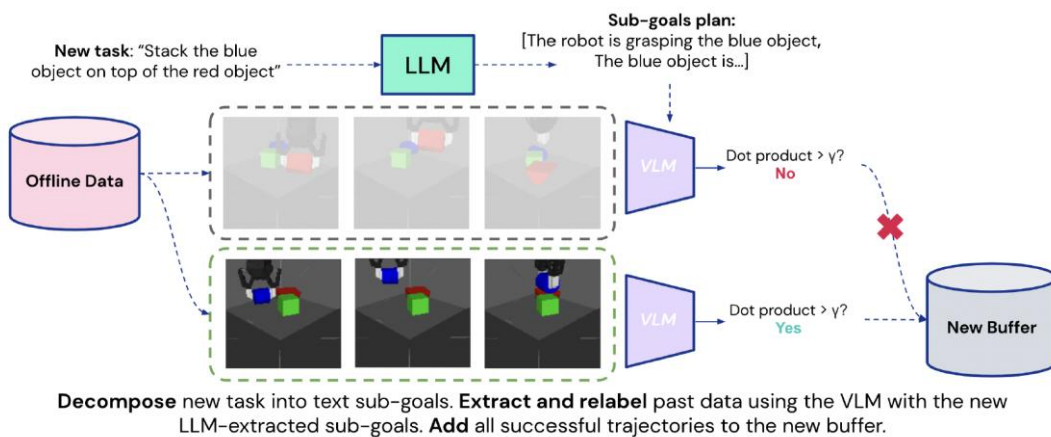
図の例では、システムは、VLM のテキストエンコーダを用いてサブゴールを処理し、サブゴールのテキスト埋め込みを生成する。また、システムは、VLM の画像エンコーダを用いて各観測画像を処理し、各観測画像の観測埋め込みを生成する。そして、システムは、各観測画像について、その観測画像の観測埋め込みとサブゴールのテキスト埋め込みとの間の類似度スコアを算出する。

その場合、システムは、類似度スコアが閾値を満たすならば、所与の観測に対応する

タイムステップにおいてエージェントがサブゴールを正常に達成したと判定する。上記図の例では、類似度スコアはドット積であり、そのドット積が閾値を超えたときに閾値が満たされる。下記図は、VLM および言語モデルニューラルネットワークを使用して学習データを生成する例を示す。

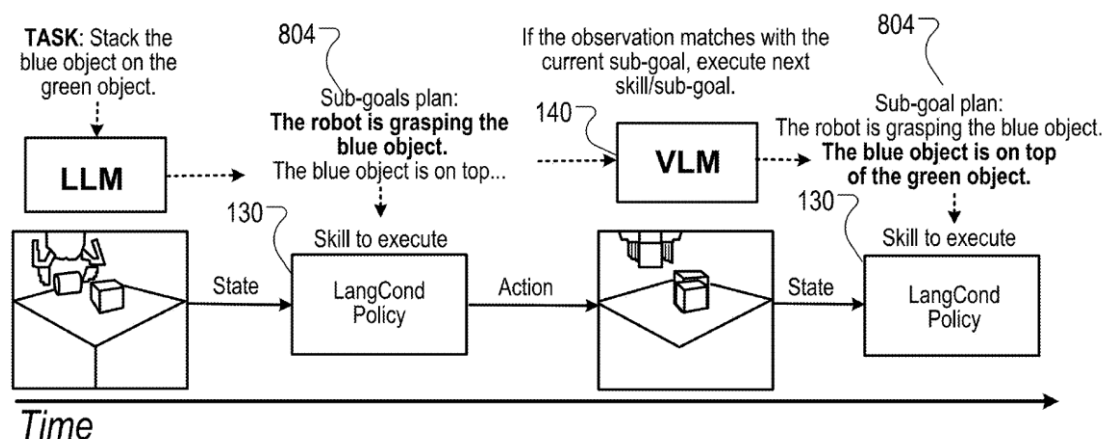


図の例において、システムは、軌跡に対する自然言語のタスク記述を受け取る。システムは、LLM を使用して、タスク記述から、一連のサブゴールの自然言語記述を含む計画を生成する。軌跡の各タイムステップについて、システムは、VLM を使用して、いずれかのサブゴールが達成されたかどうかを判定する。図の例では、システムは、青枠のタイムステップにおいて「ロボットが赤い物体を把持している」というサブゴールが達成されたと判定する。その結果、システムは、青枠タイムステップに先行するタイムステップを含む軌跡をリプレイメモリに追加する。また、システムは、軌跡の終了時に環境報酬を受け取り、その環境報酬に基づいて、軌跡全体をリプレイメモリに追加するかどうかを判定する。下記図は、新しいタスクのためにオフラインデータを再利用する例（例えば、異なるタスクのために生成された軌跡）を示す。



図に示すように、システムはオフラインデータから 2 つの軌跡を抽出する。次いでシステムは、上側の軌跡について、その軌跡内のいずれのタイムステップにおいても新しいタスクのサブゴールがいずれも達成されなかったと判定する。その結果、システムは、

新しいタスクの学習に使用するための新しいバッファにその軌跡を追加しない。しかしながら、システムは、下側の軌跡の最後のタイムステップにおいてサブゴールの 1 つが完了したと判定する。その結果、システムは下側の軌跡を新しいバッファに追加する。下記図はサブゴールを用いてタスクエピソードを実行するようにロボットを制御する例を示す。



図からわかるように、任意のタイムステップにおいて、システムは、LLM を用いて生成されたサブゴールシーケンス 804 のうちの 1 つのサブゴールに基づいて、言語条件付きポリシーニューラルネットワーク 130 を条件付ける。特に、任意のタイムステップでニューラルネットワーク 130 を条件付けるために使用されるサブゴールの記述は、図において太字で強調表示されている。システムは、VLM 140 を使用して、任意のタイムステップにおいて、次のサブゴールの「実行」に切り替えるか否か、すなわち、シーケンス内の次のサブゴールに基づいてポリシーニューラルネットワーク 130 を条件付けるように切り替えるか否かを判定する。

3.クレーム

617 特許のクレーム 1 は以下の通りである。

1. 環境と相互作用してタスクを実行するエージェントのための命令を生成する方法において、

前記エージェントによって実行されるタスクの自然言語による記述を受け取り、

言語モデルニューラルネットワークを用いて、前記タスクの実行中に前記エージェントが達成すべき一連のサブゴールの各々についての自然言語による記述を含む出力テキストシーケンスを生成するために、前記タスクの自然言語による記述から導出される入力シーケンスを処理し、

1 つまたは複数のタイムステップの各々において、

当該タイムステップにおける環境の現在の状態を特徴付ける現在の観測画像を受け取り、

当該タイムステップにおいて前記エージェントによって実行されている前記一連のサブゴールの中の現在のサブゴールを特定し、

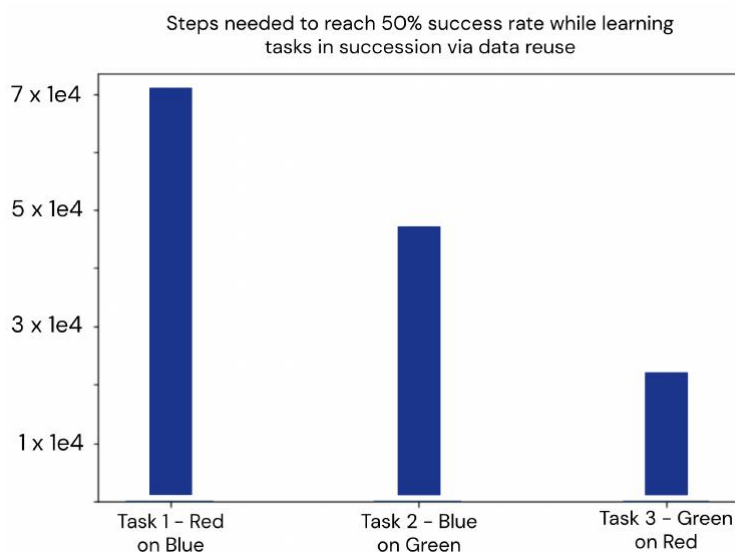
視覚言語モデル (VLM) ニューラルネットワークを用い、かつ前記現在の観測画像から、前記現在の観測の観測埋め込みを生成し、

当該タイムステップの時点で、前記エージェントが前記シーケンス内の現在のサブゴールを正常に達成したかどうかを判定し、

当該タイムステップの時点で前記エージェントが前記シーケンス内の現在のサブゴールを正常に達成したと判定したことに応じて、前記シーケンスにおいて前記現在のサブゴールに続く次のサブゴールを実行するように前記エージェントに指示する。

4. 本特許に関連する論文

本特許に関する論文 “TOWARDS A UNIFIED AGENT WITH FOUNDATION MODELS”¹が、DeepMind の Arunkumar Byravan 氏らにより公表されている。論文には本手法による有用性が示されている。下記図は性能を示すグラフである。

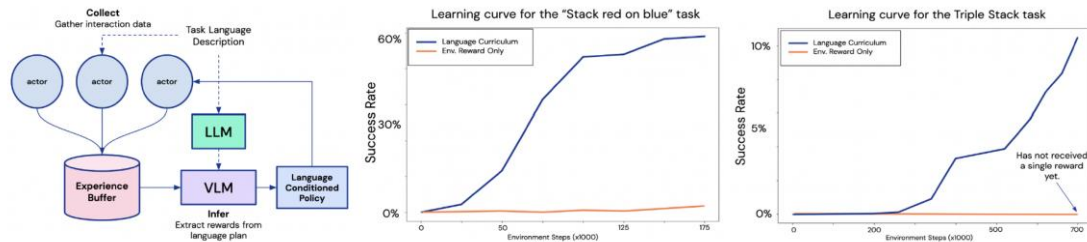


図では、3つのロボットによる積み上げタスクにおいて、新しいタスクのために学習データを再利用することが、いかに学習を加速させるかを示している。図から分かるように、最初の2つのタスクの学習データを再利用することで、ロボットは3つ目のタスクにおいて、最初のタスクと同程度の複雑さであるにもかかわらず、最初のタスクに要

¹ Arunkumar Byravan, et al. “TOWARDS A UNIFIED AGENT WITH FOUNDATION MODELS” rXiv:2307.09668v1 [cs.RO] 18 Jul 2023

した制御ステップ数の半分未満で 50%の成功率を達成している。

下記図左は、本論文における「収集・推論 (Collect & Infer)」パイプラインの説明図であり、中央及び右はそれぞれ「Stack Red on Blue」および「Triple Stack」タスクにおける、本論文のフレームワークとベースラインの学習曲線を示している。



図から分かるように、ポリシーニューラルネットワークを使用することで、より高品質な学習データが得られる結果、ベースライン技術よりも著しく高い成功率が達成されている。実際、「トリプルスタック (3 段積み)」タスクにおいては、タスクの複雑さゆえにベースライン技術では一度も成功を収めることができないのに対し、上述の技術では学習の進行に伴って性能が向上し続けている。

著者紹介

河野英仁

河野特許事務所、所長弁理士。立命館大学情報システム学博士前期課程修了、米国フランクリンピアースローセンター知的財産権法修士修了、中国清華大学法学院知的財産夏季セミナー修了、MIT(マサチューセッツ工科大学)コンピュータ科学・AI 研究所 AI コース、生成 AI ビジネスコース修了。MIT スローン経営大学院 Agentic AI コース修了。

[生成 AI・エージェント型 AI 特許コンサルティング](#)、[フィジカル AI 特許コンサルティング](#)、[医療 AI 特許コンサルティング](#)の他、米国・中国特許の権利化・侵害訴訟を専門としている。著書に「世界のソフトウェア特許(共著)」、「FinTech 特許入門」、「[AI/IoT 特許入門 4](#)」、「[ブロックチェーン 3.0](#)(共著)」がある。