

2024年4月23日
K u d a n株式会社

Kudan の視点～Kudan CTO が読み解く 人工知能と人工知覚の結びつき～を公開

高度な人工知覚（AP）技術で世界をリードするK u d a n株式会社（本社：東京都渋谷区、代表取締役 CEO：項 大雨、以下 Kudan）は、この度、『Kudan の視点～Kudan CTO が読み解く人工知能と人工知覚の結びつき～』と題した記事を別紙にて公開しましたのでお知らせいたします。

この記事では、Kudan の事業環境や将来予測、その中における経営戦略などの解説を行ってまいります。今回は、世界中で注目が高まっている半導体産業において、パートナーとKudan が取り組む人工知覚と人工知能の融合など、今後の展望について解説しています。

【K u d a n株式会社について】

Kudan は、人工知覚（AP）のアルゴリズムを専門とする Deep Tech（ディープテック）の研究開発企業です。人工知覚（AP）は、人工知能（AI）と相互補完する技術として、機械を自律的に機能する方向に進化させるものです。現在、Kudan は高度な技術イノベーションによって幅広い産業にインパクトを与える Deep Tech に特化した独自のマイルストーンモデルに基づいた事業展開を推進しています。

詳細な情報は、Kudan のウェブサイト（<https://www.kudan.io/jp/>）をご参照ください。

■ 会社概要

会社名 : K u d a n株式会社
証券コード : 4425（東証グロース）
代表者 : 代表取締役 CEO 項 大雨

■ お問い合わせ先は[こちら](#)

Kudan の視点～Kudan CTO が読み解く人工知能と人工知覚の結びつき～

Written by Anthony Glynn, Kudan CTO

ロボット工学とエッジ・コンピューティングに関する NVIDIA の視点

先月、Kudan はサンノゼで開催された NVIDIA の GPU Technology Conference (GTC) と Jetson Partner Day に参加しました。ジェネレーティブ AI や広範なロボティクス分野の最新動向に直接触れ、市場動向に対する NVIDIA の視点について理解を深め、NVIDIA のロボティクスチームやその他の潜在的な協力者と協力を進める絶好の機会となりました。

NVIDIA の成長の大部分は現在データセンターに集中しており、これは計算集約的なモデルが主流である生成 AI の現在の状況を反映しています。大規模言語モデル (LLM) のパラメータ数は通常数十億であり、GPT-4 のような最近の進歩は 1 兆パラメータの大台に乗ると推定されています。

しかし、ロボット工学やエッジ・コンピューティングにおけるアクセラレーテッド・コンピューティングとなると、まだ日が浅いです。小型言語モデル (SLM) とミニ視覚言語モデル (VLM) は、Orin Nano を含む NVIDIA Jetson デバイス上で実行可能です。しかし、エッジ・コンピューターは単一のタスク以上の処理を期待され、ロボットやエッジ・デバイスが実行するタスクの重要な性質は、エラーに対する許容度をはるかに低くする必要があります。チャットボットは何度かミスを行ってもユーザーに価値を提供できるが、ロボットがミスを行えば、その代償は破滅的なものになりかねません。

エッジ・コンピューティングとロボティクスの将来に対する NVIDIA 社の強気な見通しは、これらの分野への戦略的投資の指針となっています。この初期段階とその先に横たわる課題にもかかわらず、エッジ AI の市場機会は紛れもなく大きいです。この市場拡大の時期はまだ不確定だが、それが意味する機会の大きさは明らかです。

ロボット工学における AI の役割

人工知能 (AI) は、ロボット工学の分野に計り知れない影響を与えています。ディープ強化学習技術は、脚式ロボットによる凹凸のある地形の移動や、高速で移動するドローンの操縦など、予測不可能で複雑な環境におけるロボットの制御に不可欠なものとなっています。大規模

言語モデル（LLM）もまた、タスク・プランニングや人間とロボットのインタラクション強化に直接応用されています。

多くの議論において繰り返されたテーマは、ロボット工学における学習済み AI モデルの将来的な役割でした。この分野は、エンド・ツー・エンドの学習済みモデルのみを使用する方向に進むのだろうか、それともモデル予測制御（MPC）のような伝統的な手法の居場所はまだあるのだろうか。最終的には学習ベースのアプローチが優位に立つと主張する者もいたが、一般的な意見としては、学習済み手法と古典的手法のどちらにもメリットがあり、相乗的に共存できるというものでした。

学習ベースの AI モデルをロボット工学に応用する際の主な課題は、質の高い学習データの取得でした。言語モデルや視覚モデルを学習するために容易に入手できるデータとは異なり、ロボット工学モデルには、具体的な実世界の相互作用から得られたデータが必要です。シミュレーションは、この問題を軽減するための重要な戦略となっており、NVIDIA の高度なロボティクス・シミュレーション・ツールへの投資を支えています。とはいえ、実世界のデータの必要性は依然として残っており、シミュレーション環境と実世界のギャップを埋めるという継続的な課題が浮き彫りになっています。

言語モデルが SLAM に与える影響

大規模言語モデル（LLM）と視覚言語モデル（VLM）が初めて提示した能力は、人工知能（AI）と人工知覚（AP）の融合による、真のセマンティック SLAM への実現可能な道を示唆しています。歴史的に、SLAM は主にメトリック SLAM に焦点が当てられてきました。メトリック SLAM は、環境内の特徴の幾何学的配置をマッピングし、正確な位置と向き の推定値を生成することに関係します。一方、セマンティック SLAM は、環境からより高いレベルの意味を抽出することを目的としており、例えば、システムが、テーブル、椅子、カトラリー、グラス、食事をしている人々を識別することで、レストランのセッティングを認識できるようにします。

知覚ネットワークが、入力データや地図データから、人や車などいくつかの基本的なオブジェクトのカテゴリを識別するために使われてきたことは事実です。しかし、このアプローチは意味理解への初期段階に過ぎず、拡張性に欠けていました。言語モデルは 2 つの利点を提供します。それは、以前に見たことのないカテゴリに対するオブジェクト検出器の記述と一般化を可能にし、新しいオブジェクトクラスに対して常にモデルを再学習する必要性をなくします。言語モデルはまた、環境のより高度な記述を構築するための構造化された出力の作成を容易にします。

真のセマンティック SLAM は多くの利点を提供します。SLAM によって生成されたマップを搭載した AI エージェントと対話する個人にとって、ユーザビリティの向上は即座にもたらされます。例えば、「階段を上って 2 階に行き、右に曲がって廊下を進み、青い両開きのドアを開く

ぐると、両開きのドアの先、左側の 3 つ目のドアがある」といった、システムから提供される指示は、格段に直感的で意味のあるものになります。

真のセマンティック SLAM のもう一つの利点は、システムが何に注意を払う価値があり、何を無視してもよいかを理解できることです。例えば、家具や駐車中の車のような一過性のオブジェクトは、位置が変わる可能性があるため、再訪時に場所を認識するための信頼性が低くなります。より高度な環境記述を活用することで、システムのロバスト性は大幅に向上します。このアプローチでは、環境の外観だけでなく、環境の構造やレイアウトに注目することで、照明の違いなどさまざまな条件下でも、同じ場所を確実に認識することができます。

Kudan の方針

ロボット工学におけるエンド・ツー・エンドの学習済みモデルと従来のアプローチの採用に関する議論は、ディープラーニングが脚光を浴び始めた SLAM コミュニティで生まれた過去の同様の議論を非常に思い起こさせるものでした。その際にコンセンサスとなった見解は、私たち自身の結論とも一致しています。たとえば、三角測量法のように、あるプロセスが方程式で簡潔に記述できる場合、その方程式を直接実装するのが最も効率的であることが多いです。しかし、ディープラーニングの強みは、異なる照明条件下で特徴がどのように見えるかを記述するような、簡単な数学的定式化が困難な課題に対処することにある。このことが、例えばビジュアル SLAM システムでディープラーニングされた特徴量を使用することを探求する動機となり、私たちは人工知能 (AI) を組み込んだ SLAM の開発に成功するだけでなく、ディープラーニングモデルを SLAM に統合する新しい方法を検討し続けています。

NVIDIA 社をはじめとするパートナー企業との議論を通して、特にマッピング、ローカリゼーション、ナビゲーションに言語モデルの最近の開発をどのように活用できるかということに関連して、いくつかの新鮮な新しいアイデアを呼び起こしました。私たちは、これらのアイデアを試し、さらに探求することを楽しみにしていますし、これがモバイル・マシンのための空間知能の未来にどのような意味を持つのか、非常に期待が高まるとともに、今後公開を進めていきたいと考えています。