

報告書

1 今週の進捗

- MCP-Zero: Active Tool Discovery for Autonomous LLM Agents の調査
- Process vs. Outcome Reward: Which is Better for Agentic RAG Reinforcement Learning の調査
- 最終目標 “ドメイン特化型 LLM” の確認

2 論文 Process vs. Outcome Reward: Agentic RAG の調査

本論文は、検索拡張生成 (Retrieval-Augmented Generation; RAG) を自律的に行うエージェント (Agentic RAG) の学習方法に関する研究である。従来の強化学習 (Reinforcement Learning; RL) を用いた手法では、最終的な回答の正しさのみを報酬としていた。しかし、このアプローチは探索効率の低さ、報酬の疎性、学習の不安定性といった課題を抱え、大量の学習データと計算コストを必要としている。

そこで本論文では思考の途中経過を細かく評価して Agentic RAG に適用する新たな手法 ReasonRAG [1] を提案している。ReasonRAG は、モンテカルロ木探索 (Monte Carlo Tree Search; MCTS) と独自の報酬推定アルゴリズム (Shortest Path Reward Estimation; SPRE) を組み合わせ、高品質なプロセスレベルの選好データセット RAG-ProGuide を自動で構築する。このデータセットを用いて LLM エージェントを学習させることで、従来手法よりも大幅に少ないデータ量と計算コストでそれを上回る性能を達成できることを 5 つのベンチマークデータセットで実証している。

2.1 提案手法の仕組み

提案手法 ReasonRAG は大きく分けて “プロセス評価データの生成” と “選好最適化による学習” の 2 つのフェーズから構成される。

- プロセス評価データセット RAG-ProGuide の自動構築:
 1. 多様なプロセスの探索: モンテカルロ木探索 (MCTS) を用いて LLM エージェントに検索クエリ生成、証拠抽出、回答生成などの多様な思考プロセスを探索させる。これによりやみくもな探索ではなく有望な思考プロセスを効率的に見つけ出す。
 2. プロセス品質の評価: 探索された各プロセスの途中経過の良し悪しを、独自に提案した最短経路報酬推定 (SPRE) で評価する。SPRE はそのプロセスを続行した場合に “最終的に正解にたどり着ける確率” と “完了までにかかるステップ数” の両方を考慮して報酬を算出する。
 3. 選好ペアの作成: SPRE によって算出された報酬に基づき、良いプロセスと悪いプロセスのペアを作成し、データセット化する。
- 選好最適化によるエージェントの学習: 生成したデータセット RAG-ProGuide を用いて直接選好最適化 (Direct Preference Optimization: DPO) により LLM エージェントをファインチューニングする。

2.2 利点

本手法は結果のみを評価する従来手法と比較して以下の点で利点を持つ。

- 高いデータ効率と学習効率: 最終結果だけでなく途中の各ステップで報酬を与えながら学習するため効率が良い。結果として従来手法の約 5% (90,000 件に対し 5,000 件) という圧倒的に少ないデータ量で同等以上の性能を達成していた。
- 優れたタスク遂行性能: 特に複数回の検索と情報統合が必要となる複雑な“複数ホップ QA タスク”において、既存手法を上回る高い性能を示していた。これはプロセスの学習により、戦略的な思考能力が身についたと考えられる。
- 高い汎化能力: 学習データに含まれていない未知のドメインのタスクに対しても高い性能を維持していた。これはモデルが表面的な解法パターンを暗記するのではなく、より本質的で応用可能な問題解決能力を学習していることを示している。
- アノテーションコストの削減: 高品質なプロセスレベルの報酬データを人間の手作業に頼ることなく MCTS と SPRE を用いて完全に自動で構築できる。

2.3 欠点

本論文の内容から以下の欠点が挙げられる。

- データ構築時の初期モデルへの依存: お手本データセット RAG-ProGuide の品質はデータ構築フェーズで探索役を担う LLM (本研究では GPT-4o) の性能に依存する。もし初期モデルの能力が低い場合、そもそも有望なプロセスを探索できず高品質なデータセットが生成できない可能性がある。
- SPRE の計算コスト: 途中経過を評価するために複数回シミュレーションする SPRE は、原理的に計算コストが高い。最終的なエージェントの学習コストは低いが、その前段階であるデータ構築フェーズには相応の計算リソースが必要となる。
- 報酬設計の一般性: 本研究では質問応答タスクにおける F1 スコアを正解らしさの指標としていた。しかし、要約や対話、創造的な文章生成といった正解が 1 つに定まらないタスクにおいてプロセスの良し悪しを自動で評価する報酬関数を設計することは依然として困難な課題である。

3 論文 MCP-Zero の調査

本論文では、LLM エージェントが外部ツールを利用する際の根本的な問題を解決する新しいフレームワーク MCP-Zero [2] を提案している。

従来手法では利用可能な全ツールの説明書をプロンプトに含めるため、コストが高く非効率であり、また、LLM は与えられたリストから選ぶだけの受動的な存在であった。一方、MCP-Zero は、LLM 自身に必要なツールを考えさせ、それをその都度自ら要求させる能動的なツール発見という新しいパラダイムを提唱する。これにより、エージェントは自律的になり、システム全体も劇的に効率化されることを実証していた。

3.1 MCP-Zero の仕組み

MCP-Zero は主に以下の 3 つのシンプルなメカニズムで動作する。

- 能動的なツール要求 (Active Tool Request)
LLM はタスクを解決する上で能力が不足していると判断すると, “どのサーバのどんな機能のツールが欲しいか” という構造化された要求書 (リクエスト) を自ら生成する.
- 階層的セマンティックルーティング (Hierarchical Semantic Routing)
LLM からの要求書に基づき, システムは 2 段階検索をする.
 - 第 1 段階: サーバで大まかに候補を絞り込む.
 - 第 2 段階: 候補の中からツールの意味が最も近いものを特定する.これにより数千あるツールの中からでも高速かつ高精度に目的のツールを発見できる.
- 反復的な能力拡張 (Iterative Capability Extension)
一度の検索でツールが見つからない場合, LLM は要求書を修正して何度も検索をやり直すことができる. これにより, 複雑なタスクをこなしながら段階的に必要なツールを揃えていくことが可能である.

3.2 利点

- 圧倒的な効率性かつ低コスト
必要なツール情報のみをプロンプトに含めるため, 実験ではトークン消費量を最大 98% 削減しており, API の利用コストと処理時間を大幅に削減できる.
- 高精度なツール選択
LLM 自身が検索キーワード (リクエスト) を生成するため, ユーザの曖昧な指示で検索するより遥かに精度が高く, 的確なツールを見つけ出せる.
- 自律性の実現
LLM が受け身で選ぶ存在から自ら考えて行動する能動的な主体へと変わり, より高度なエージェントシステムの構築が可能になる.

3.3 欠点

- モデルの性能依存性
そもそも高い推論能力を持たない LLM では適切なツール要求を生成できず, 性能が上がらない可能性がある. 元々高性能な GPT-4.1 では改善効果が限定的であった.
- リクエスト生成の失敗リスク
LLM が最初のツール要求の生成に失敗したり, 見当違いな要求をしたりするとタスクが停滞する可能性がある.
- 検索対象の品質
検索対象となるツールの説明文が不十分な場合, 正しいツールを発見できない.
- ツールの不在
解決策となるツール自体が存在しない場合, このフレームワークだけではタスクを解決できない. これはあらゆるツール利用エージェントの共通の課題でもある.

4 最終目標 “ドメイン特化型 LLM” の確認

アニメーション制作の知識に特化した LLM のシステムの概要である。本システムでは、RAG とファインチューニングの技術を連携させることで専門的かつ高精度な応答が可能なローカル LLM システムの実現を目標としている。以下では、システムの全体像、採用する技術スタック、および開発ロードマップを定義する。

4.1 目的

本システムの主目的は、アニメーション制作現場における知識検索、ノウハウ共有、および新人教育を支援するためのドメイン特化型 LLM を構築することである。ユーザ (アニメ制作者) からの専門的な質問に対し、正確かつ文脈に即した回答を生成することで制作プロセスの効率化と品質向上に貢献する。

4.2 コアコンセプト

本システムは以下の 2 つの技術を統合することで実現される。

1. RAG: 外部ナレッジベースから関連情報を動的に検索し、それを基に回答を生成することで情報の正確性と最新性を担保する。
2. ファインチューニング: ベースとなる LLM に対し、アニメ制作ドメインの応答スタイルや専門用語、思考プロセスを学習させる。これにより、RAG から得た知識をより専門家らしく解釈し、質の高い応答を生成する能力を獲得する。

この 2 つの組み合わせにより単一でありながらも専門性の高い LLM システムを構築する。

4.3 採用技術スタック

本システムで採用する主要なツール、モデル、およびプラットフォームは以下のものを考えている。

4.3.1 開発および実行環境

- ハードウェア: 研究室もしくは研究科のサーバ
- ソフトウェア: Python, Docker, LangChain

4.3.2 主要モデル

- ベース LLM: `elyza/Llama-3-ELYZA-JP-8B`
Meta 社の高評価モデル Llama3 をベースに ELYZA 社が日本語性能を強化したモデル。RAG の回答生成およびファインチューニングの対象として利用する。
- エンベディングモデル: `intfloat/multilingual-e5-large-instruct`
多言語対応であり、特に日本語のセマンティック検索において高い性能評価を得ているモデル。RAG におけるナレッジベースのベクトル化に用いる。

4.3.3 データ管理

- フィードバック UI: Gradio
ユーザからの質問, LLM の回答, およびその評価を収集するための簡易的な Web インターフェースを構築する.
- 一時データストア: MySQL
Gradio から送信されたフィードバックデータをリアルタイムに保存する. 開発初期はローカル PC 上の Docker コンテナで運用し, 外部公開時にはクラウドサービスである PlanetScale への移行を計画している.
- データセット原本管理: Hugging Face Hub
MySQL に蓄積されたデータを定期的に集約し, バージョン管理された公式なデータセットとして非公開リポジトリに保存する. ファインチューニングやナレッジベースの更新はこのリポジトリを基にする. GitHub でも管理は可能だが容量制限があるため, 容量制限のない Hugging Face Hub を使用するほうが望ましい.

4.3.4 RAG システム構成

- フレームワーク: LangChain
RAG のパイプライン (ドキュメント読み込み, 分割, 検索, 生成) を構築するために利用する.
- ベクトルデータベース: ChromaDB
エンベディングモデルによってベクトル化されたナレッジデータを格納し, 高速な類似度検索を実現する. ローカル環境で容易にセットアップ可能である点を評価する.

4.4 開発ロードマップ

本システムは以下のフェーズに分けて段階的に推進する.

フェーズ 1: 基盤構築とナレッジ収集

アニメ制作に関するテキストデータ (技法書, Web 記事, インタビュー等) を収集整理する. 同時にローカル PC 上に開発環境を構築する.

フェーズ 2: RAG プロトタイプ開発

選定したベース LLM, エンベディングモデル, ChromaDB を用いて, 基本的な RAG システムを構築する. この段階で検索精度や応答品質に関する初期課題を洗い出す.

フェーズ 3: フィードバックシステムの構築とデータセット作成

Gradio と MySQL を用いてユーザーからのフィードバック収集システムを構築する. 収集したデータおよび既存のナレッジを基にファインチューニング用の QA データセットを作成し, Hugging Face Hub で管理する.

フェーズ 4: ファインチューニングによるモデルの専門化

作成したデータセットを用いて LoRA 学習にてベース LLM をファインチューニングする. これにより, モデルにアニメ制作ドメイン特有の応答スタイルを学習させる.

フェーズ 5: 統合システムの完成と評価

フェーズ 2 で構築した RAG システムの LLM をフェーズ 4 でファインチューニングしたモデルに置き換える. これにより, 知識の正確性 (RAG) と応答品質 (ファインチューニング) を両立した統合システムを完成させる.

フェーズ 6: 継続的な改善サイクル

完成したシステムを実際のユーザに試用してもらい、継続的にフィードバックを収集する。収集したデータは次回のファインチューニングやナレッジベースの拡充に活用し、システムの精度を継続的に向上させるサイクルを確立する。

フェーズ 7: 追記

Mixture of Agents (MoA) や Mixture of Experts (MoE) を組み合わせることで精度の向上が考えられる。また、OCR タスクのような質問応答以外のタスクも扱えるようになる。

4.5 elyza/Llama-3-ELYZA-JP-8B の検証

RAG なしで LLM を試した。

4.5.1 実験環境

- **モデル:** elyza/Llama-3-ELYZA-JP-8B
- **モデル種別:** 日本語に対応した事前学習済みモデル。指示チューニングは施されていない。

表 1: テキスト生成パイプラインのパラメータ設定

パラメータ	値
max_new_tokens	512
temperature	0.7
repetition_penalty	1.1
do_sample	True

4.5.2 実験結果と考察

3 つの異なる質問に対するモデルの応答を評価した。

質問 1: アニメ制作におけるレイアウトの役割について、詳しく教えてください。

[モデルの回答]

レイアウト！アニメ制作では非常に重要な工程の 1 つです。

レイアウトは、原画を基にした作業で、動きや構図を決めるために行われます。具体的には、以下のような役割があります。

1. ****シーンの全体像を把握する**:** レイアウトでは、カメラ位置、キャラクターの配置、背景などを含むシーンの全体像を決定します。これにより、演出家や監督が意図する世界観や雰囲気を視覚化することができます。
2. ****キャラクターの動きと距離感を調整する**:** レイアウトでは、キャラクターのポーズや動きを微調整し、動線や空間を考慮して最適な配置を行います。この工程で、キャラクター同士の関係性や感情表現をコントロールします。

3. ****フレームの構成要素を設計する****: レイアウトでは、フレーム内での各要素（キャラクター、背景、道具など）のバランスを調整し、見栄えの良さや情報の伝わりやすさを向上させます。
4. ****絵コンテや原画の指示を反映させる****: レイアウトでは、絵コンテや原画に描かれた指示や意図を忠実に再現し、必要に応じて修正や提案を加えます。
5. ****次工程への橋渡しとなる****: レイアウトが完成すると、次工程である清書やペイントの作業が始まります。レイアウトは、次の工程にスムーズに移るための重要な資料となります。

アニメ制作のレイアウトは、作品のクオリティや世界観を形作る上で非常に重要な工程です。担当者は、創造的な判断力、技術力、コミュニケーション能力などを兼ね備える必要があります。

質問 2: 3DCG の制作工程で「リギング」とは、具体的にどのような作業ですか？

[モデルの回答]

リギング (Rigging) とは、3DCG キャラクターの体や骨格にスケルトンを設定し、動きや物理演算を可能にする準備作業のことを指します。

具体的には、以下のような作業が含まれます：

1. スケルトンの作成：3D モデルにスケルトンと呼ばれる内蔵骨格を設置します。これにより、モデルの内部に空間を作り、後工程での動きや物理演算が可能になります。
2. ジョイントの設定：スケルトンにジョイントと呼ばれる接続点を設定し、モデルの各部位を関節でつなぎます。これにより、モデルの可動域や動きを定義します。
3. ウェイトの調整：モデルの各パーツにウェイトと呼ばれる重みを設定し、スケルトンとの関係を調整します。これにより、モデルの動きや物理演算の精度を高めます。
4. リグの構築：上記の作業を基に、実際に動かすためのリグ (Skeletone) を構築します。リグは、スケルトンとモデルの結合、動きの制御、衝突検出などを一元管理する重要な要素です。

リギングは、3DCG キャラクターやロボット、生物などの動きやアクションシーンを再現するために不可欠な工程です。

質問 3: アニメ業界の専門用語「デジタル中割り自動生成 AI」について、現状の課題は何ですか？

[モデルの回答]

デジタル中割り自動生成 AI とは、手描きのアニメーターが行っていた中割り (cut-in と cut-out を繋ぐ動画素材) を、AI が自動で生成する技術です。現在の課題は主に以下のような点が挙げられます。

1. 中割りの定義やルールの難しさ：中割りには様々なパターンや例外があり、AI に正しく学習させるためには、膨大な量のデータと高精度のラベル付けが必要です。
2. 中割りの品質向上：現在の AI は、多少の不自然さや違和感が生じることがあります。中割りは、微妙な間や動作のつながりが重要な要素であるため、AI がこれらを正確に捉え、かつ高品質な中割りを生成することは非常に困難です。

3. 作画の個性や表現力の再現性: AI が生成した中割りは、機械的に処理されるため、人間の作画のように個性や表現力が反映されません。作品の世界観やキャラクターの特徴を損なう可能性があります。
4. 制作フローの変化に伴う新たな問題: 自動生成された中割りを活用する場合、制作フローも変化します。新しいツールやソフトウェアを導入し、スタッフが適応する必要がある一方で、新たなスキルや知識が求められることになります。
5. 知的財産や著作権に関する問題: AI が生成した中割りは、誰が創作者として認められるのか、また所有権や使用許諾など、法律的な問題が未整備です。

以上の課題を解決するために、研究者や企業は日夜努力しています。AI が持つ可能性を最大限に活用して、更に高度なアニメーション制作を実現することが期待されています。

質問 1, 2 の結果は、モデルが事前学習の段階でアニメ制作に関する豊富な知識を獲得していることを示している。しかし、これらの情報がどの文献やデータに基づいているかは不明であり、その正確性を検証する手段はない。

質問 3 は、現実には確立されていない架空の技術用語を含んでいる。この結果は、自身が知らない架空の用語に対して“知らない”と回答せず、あたかもそれが実在する技術であるかのように振る舞い、もっともらしい課題を 5 項目も創作した。典型的なハルシネーションが生じている。

5 今後の課題

- アニメ制作に関するデータの収集
- PlanetScale を用いた RAG の導入

参考文献

- [1] Wenlin Zhang, Xiangyang Li, Kuicai Dong, Yichao Wang, Pengyue Jia, Xiaopeng Li, Yingyi Zhang, Derong Xu, Zhaocheng Du, Huifeng Guo, Ruiming Tang, and Xiangyu Zhao. Process vs. outcome reward: Which is better for agentic rag reinforcement learning. 2025.
- [2] Xiang Fei, Xiawu Zheng, and Hao Feng. Mcp-zero: Active tool discovery for autonomous llm agents. 2025.