

概要

マルチホップ推論を伴うAgentic RAGを実行可能な日本語SLMの学習

- 1. Chain of Abstractionに基づくAgentic RAG能力をSarashina 7BにSFT
- 2. Agentic RAG評価ベンチマーク 虚構マルチホップQA の構築

1. 背景

- マルチホップ推論を伴う検索タスク：単発検索を前提とする従来RAGでは困難
- Agentic RAG: 自律的に検索計画を立て、検索・思考の反復を行う (例: ReAct^[1], CoA^[2])
- 日本語LLMにおけるAgentic RAG の学習手順に関する報告は限られる
- PromptingのみでAgentic RAGを実行可能なLLMの使用を前提 → 当該能力未獲得の日本語SLMの学習

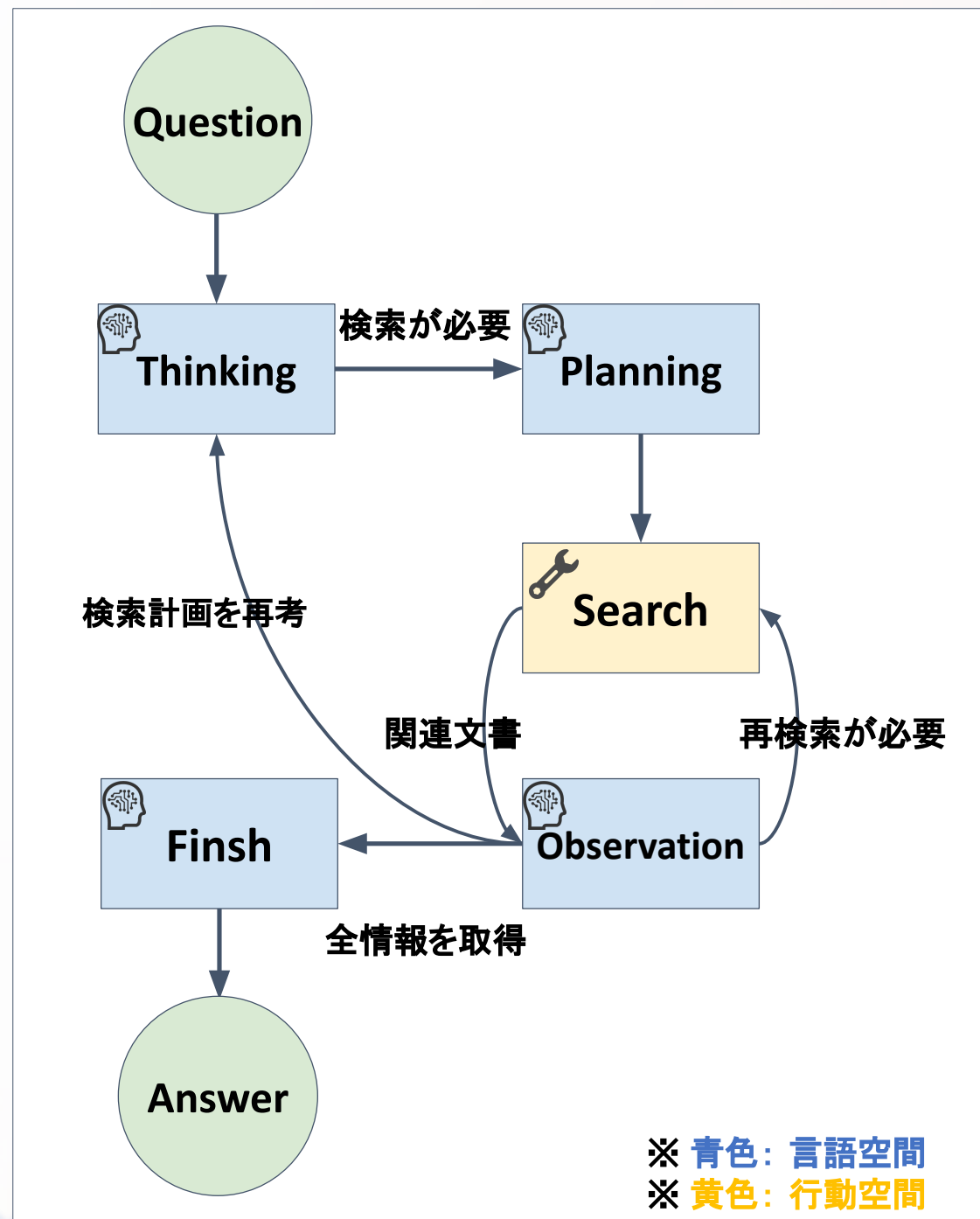


図1. 本研究におけるAgentic RAGの状態遷移図

2.1. Agentic RAG 学習

Chain of Abstraction (CoA)^[2]

- 抽象的なプレースホルダを用いた検索計画
- 構成質問に対して優れる
- 本研究では検索計画で以下の要素を管理
 - step: 特定すべき対象のインデックス
 - target_entity: stepで特定すべき対象
 - query: 検索クエリ
 - fact: 知識・プレースホルダを含む

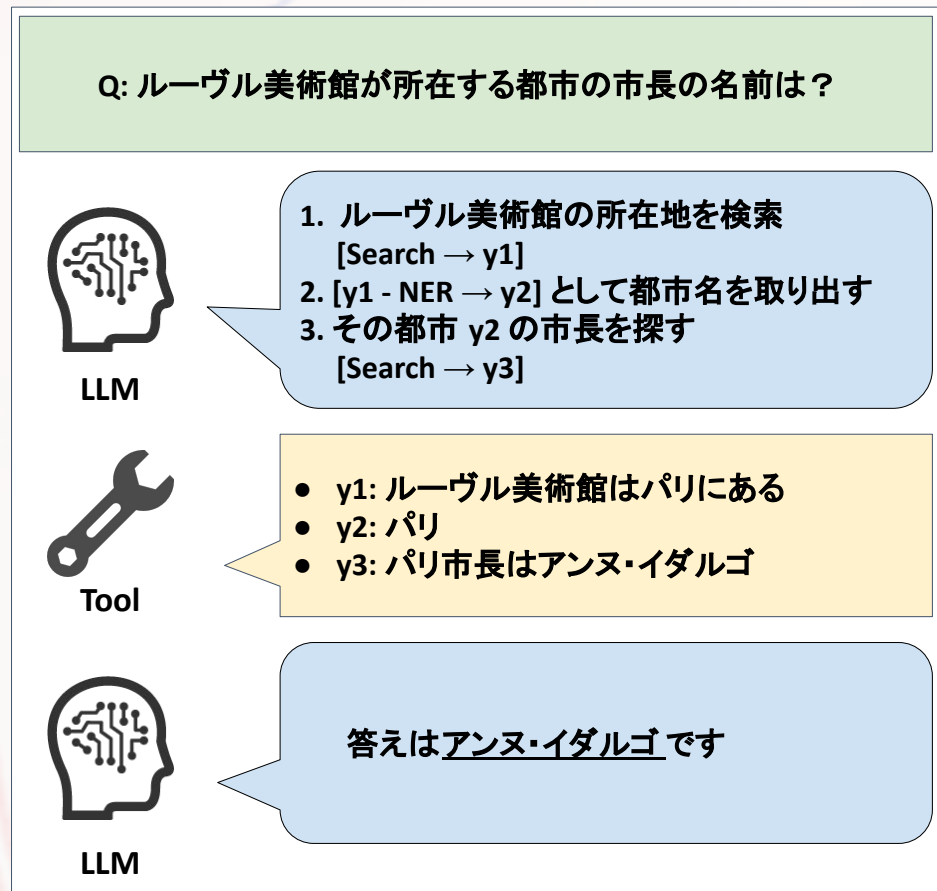


図2. CoAに基づく推論の例

行動軌跡データの合成

- Multi-Agent Distill: 複数LLMによるAgentic RAGの軌跡を生成
 - PathSynth: 状態遷移を網羅する行動軌跡を合成
- 以下3つの再検索を含む行動パターンを定義

- $\sigma((S(e_1)O(e_1))(S(e_2)O(e_2))\text{fin})$
 - $\sigma((S(e_1)O(e_1))^2(S(e_2)O(e_2))\text{fin})$
 - $\sigma((S(e_1)O(e_1))(S(e_2)O(e_2))^2\text{fin})$
- σ : 状態遷移関数
 - e : エンティティ
 - S : Search
 - O : Observation

以下の合成行動軌跡データを用いて 教師あり指示学習 (SFT)

合成手法	QAデータ	合成器	件数
Multi-Agent Distil	JEMHopQA	GPT 4o mini ¹	248
PathSynth	JQaRA ^[5] (1-hop QA)	Qwen3 235B ²	500
PathSynth	JEMHopQA	Qwen3 235B	1,123

表1. 学習に用いた合成行動軌跡データ

- GPT 4o mini: gpt-4o-mini-2024-07-18 (OpenAI API)
- Qwen3 235B: Qwen/Qwen3-235B-A22B-Instruct-2507

2.2. 虚構マルチホップQAベンチマーク

既存研究

- JEMHopQA^[3]: 日本語マルチホップQAベンチマーク

- 比較質問: 同じ関係の2エンティティを比較

例: 「夏目漱石と太宰治はどちらが早く生まれた？」

- 構成質問: 隠れた橋渡しエンティティを特定

例: 「実写映画 魔女の宅急便の主演俳優の生年月日は？」

⇒ LLMが事前学習で得た知識のみで回答が可能

- 虚構RAGベンチマーク^[4]: 実在しないエンティティや出来事に関する合成文書に基づくQAによるRAG評価

提案ベンチマーク：虚構マルチホップQA

- 架空文書集合・マルチホップQAをLLMを用いて合成
- 例: 「青風文化社と光下音楽社はどちらが早く創業した？」
- 80件の評価セット (比較質問: 40問, 構成質問: 40問)
- 評価: LLM as a Judgeによる2値判定 → Accuracy
- 結果: 関連文書なし → 回答不可, Agentic RAG → 性能向上

生成手法	合計	比較質問	構成質問
関連文書なし	0.063	0.063	0.000
RAG(単発検索)	0.300	0.410	0.175
Agentic RAG:CoA	0.738	0.769	0.700

表2. 参考: Qwen3 235B A22B Instructの評価結果

3. 実験

学習設定

- ベースモデル: Sarshina2 7B (指示学習済みモデル)
- ベンチマーク: 虚構マルチホップQA
- LLM as a Judge: Qwen3 235B A22B Instruct
- 検索器: BM25 (kuromojiを用いて単語分割)
- 検索対象: 架空文書集合, Wikipedia文書集合^[6]
- ツール呼び出し
 - search: クエリ入力としてindex検索を実行(最大5回まで呼び出し可能)
 - finish: 最終回答をユーザに提示する

4. 評価結果

Qwen3 235B Agentic RAG ≥ 学習モデル > Qwen3 235B RAG

モデル	生成手法	合計	比較質問	構成質問
Qwen3 235B	RAG(単発検索)	0.300	0.410	0.175
Qwen3 235B	Agentic RAG: CoA	0.738	0.769	0.700
Sarshina2 7B	RAG(単発検索)	0.175	0.282	0.075
学習モデル	-	0.700	0.775	0.625

表3. 虚構マルチホップQAベンチマークによる評価結果

5. 分析

Agentic RAG能力の獲得

複数ターンのTool呼び出し・検索結果を踏まえた再検索判断を実現

課題

- ツール呼び出し時のフォーマットエラー
- 本来2stepが望ましい場合に1stepの検索計画を立案
- 関連文書不在時のObservationでのハルシネーション

Question	瑞野りあねが所属するグループを運営している事務所は？	Question	MERRY-TUTORIALの連載雑誌を発行していた出版社はどこか？
Thinking	(省略) まず、瑞野りあねが現在所属しているグループの名前を特定し、その後、そのグループを運営する事務所を調べる必要がある。	Planning	Step1. 「MERRY-TUTORIAL 連載 雑誌」 → ルミナスフレーム Step2. 「ルミナスフレーム 出版社」 → {fact_step2}
Planning	Step1. 瑞野りあね 所属グループ Step2. {fact_step1} 事務所	Tool	【出版社】(省略: 出版社に関する概要文章) 角川書店のように出版社が他業種へと進出する場合...
		Observation	角川書店のように出版社が...という記載があり、ルミナスフレームの発行出版社が角川書店であることを示唆している。(省略)

- (左) 図4. 検索計画の成功例と失敗例: 正しくは2stepで検索計画を立てる
(右) 図5. Observationの失敗例: 無関係な角川書店を関連事実として判断

- ⇒ SFTのみで
- Agentic RAGの一連の流れを獲得可能
 - 柔軟な推論を伴うAgentic RAGの実現は困難

今後の展望：強化学習等による柔軟な思考能力の獲得

PathSynthの具体例: 行動パターン2の例

- Question: 「ルーヴル美術館が所在する都市の市長の名前は？」(fact1: パリ, fact2: アンヌ・イダルゴ)
- Planning: SearchPlan(step=1, query='ルーヴル美術館 所在地', fact='{fact1}'), SearchPlan(step=2, query='{fact1} 市長', fact='{fact2}')
 - Search: 「ルーヴル美術館 所在地」
 - Observation: 「見つからなかった」◦ 根拠情報 未発見
 - Search: 「ルーヴル美術館」◦ クエリ再修正
 - Observation: 「パリに所在することがわかった」◦ 根拠情報 発見
 - Search: 「パリ 市長」
 - Observation: アンヌ・イダルゴであることがわかった◦ 根拠情報 発見
 - Answer: 答えはアンヌ・イダルゴです。

参考文献

- ReAct: Synergizing Reasoning and Acting in Language Models, ICLR 2023
- Efficient Tool Use with Chain-of-Abstraction Reasoning, COLING 2024
- JEMHopQA, <https://github.com/aiishii/JEMHopQA>
- 実在しないエンティティや出来事に関する合成文書を用いた RAGベンチマークの構築, IPSJ NL 第263回
- JQaRA, <https://github.com/hotchpotch/JQaRA>
- Wikipedia 2021, 森羅プロジェクト, <http://shinra-project.info/>