

# MCMC を用いた前提検索による LLM の仮説推論能力の強化

WANG YUANYI<sup>1</sup> 小林一郎<sup>1</sup>

<sup>1</sup> お茶の水女子大学

{wang.yuanyi, koba}@is.ocha.ac.jp

## 概要

本稿では、大規模言語モデル (LLMs) の推論能力、特に仮説推論能力を向上させるため、モンテカルロ・マルコフ連鎖 (MCMC) を活用した新しいアルゴリズムパイプラインを提案する。観察を最も適切に説明する仮説と、それを支持する関連前提を効率的かつ計算コストを抑えて探索するため、完全教師なしの MCMC アルゴリズムを導入した。本手法は、関連性の高い前提の順序を優先することで、LLMs における正当な仮説を効果的に構築し、生成文の精度と前提知識の再現率の向上を確認し、仮説推論能力を向上させることを実証している。

## 1 はじめに

近年、大規模言語モデル (LLMs) は、Chain of Thought (CoT) プロンプティングによって推論能力が大幅に向上していることが示されている [1]。しかし、最も高い尤度 (Plausibility) を持つ仮説を選ぶ仮説推論は、LLM がもつ推論能力が正当な推論として成り立っているものなのかを評価するのが難しい推論の一つである。人間の推論は、経験や文脈に基づく因果関係の理解によって形成されるが、LLM は統計的パターンに依存するため、因果関係を深く理解しているかを判断することが難しい場合が多い。また、LLM は観察を説明するために尤もらしい仮説を生成することはできても、その仮説を支える全ての前提を特定するのは困難である。これが、環境から前提を使いながら仮説を生成する仮説推論の完全に再現する際の大きな障壁となっている。前提の順序が推論プロセスに与える影響は、DeepMind の研究 [2] により明らかになっており、前提の順序次第で精度が最大 30% 変動することが示されている。この知見は、関連性の高い前提を選択するだけでなく、それらを適切な順序で提示することが、

LLM の推論能力を向上させるために重要であることを示唆している。

本研究では、LLM に仮説推論をさせる際の問題や現象に対する解釈と推論能力を向上させるために、仮説を支持する前提条件となる命題を立てることにおいてマルコフ連鎖モンテカルロ法 (MCMC) を導入し、確率的なモデリングを通じて不確実性を管理するだけでなく、前提空間における探索と活用のバランスを取る方法を提案する。

## 2 仮説推論

仮説推論は、最良の説明への推論 [3] 観察された現象に対してもっともらしい説明を作るための基礎的な認知プロセスである。一般的には以下のように定義される：

$$H \cup P \implies O,$$

ここで、前提 *Premises* の集合  $P = \{p_1, p_2, \dots, p_n\}$  と観察 *observation*  $O$  が与えられたとき、仮説推論とは観察  $O$  を最も簡潔かつ合理的に説明する仮説 *Hypothesis*  $H$  を見つけること。しかし、このプロセスをモデル化するのは一般的に難しい。仮説の尤度 (確からしさ) を測る標準的な方法が確立されていない。この難しさは、観察の原因となる環境からすべての情報や前提知識を取得することが基本的に不可能であることに起因している。そのため、本研究では以下のタスクを設定し、その問題解決に取り組む：

- 仮説生成:** 観察と前提の集合が与えられたとき、その観察を説明する仮説を生成する。
- 前提選択:** 与えられた前提の集合から、生成された仮説を支持する関連する前提をすべて特定する。

## 2.1 提案手法

仮説推論では、観察  $O$  を最もよく説明する前提集合  $P^*$  と仮説  $H^*$  を探索する必要がある。しかし、前提空間が大きい場合、全ての前提の組み合わせを試すのは非現実的なことである。この問題を解決するために、モンテカルロ・マルコフ連鎖 (MCMC) を使って効率的な探索を行う。そして、MCMC によって得られた前提を元の前提リストの先頭に再配置し、それをういて LLM にプロンプトを生成させ、最適な説明を再構築する。

### モンテカルロ・マルコフ連鎖の利点

- **確率的枠組み**：MCMC を使うことで、前提空間を効率的に探索し、計算資源を節約できる。
- **探索と活用のバランス**：新しい前提を試す「探索」と、既存の良い前提を利用する「活用」のバランスをとりながら最適化が可能である。本手法では完全教師なし (unsupervised) である点特徴的である。
- **一般化可能**：予測エントロピーや尤度を使って、不確実性を定量化しながら前提選択を改善できる。また、アルゴリズム内でペナルティ項を導入することで柔軟性を確保している。さらに、教師なし手法であるため、幅広い応用に対応可能である。

MCMC の反復的なサンプリングにより、観察  $O$  を最もよく説明する前提集合と仮説を効率的に探索できる。その具体的なアルゴリズム 1 は、Metropolis–Hastings アルゴリズム [4] を参考し、以下の手順で最適な前提と仮説を探索する。

**入力と初期化** 観察  $O$  と全ての前提集合  $P = \{p_1, p_2, \dots, p_n\}$  を含むプロンプトを作成し、LLM の出力をもとに初期仮説  $H^{(0)}$  と初期前提  $P^{(0)}$  を生成する。これらを最初の最良仮説として記録する。また、初期スコア  $S^*$  は含意モデル (例：RoBERTa[5]) をういて対数尤度  $\log P(H^{(0)} | O)$  を計算し、MCMC の進行状況を評価する指標となる。

**提案ステップ** 現在の前提集合  $P'$  に対してランダムに選んだ前提  $p$  を追加する。追加後の  $P'$  と観察  $O$  をういて新たなプロンプトを生成し、LLM を利用して新しい仮説  $H'$  を生成する。

**スコア計算ステップ** 含意モデルをういて、新たな仮説  $H'$  のスコア  $\log P(H' | O)$  を計算する。さらに、LLM の生成確率  $\log P$  を基に、現在の前提  $p$  を

---

### Algorithm 1 MCMC に基づく仮説推論

---

入力:

観察  $O$ , 前提集合  $P = \{p_1, p_2, \dots, p_n\}$ ,  $P' = [P^{(0)}]$

初期仮説  $H^* = H^{(0)}$

entail モデルで  $S^* = \log P(H^{(0)} | O)$

出力:

最適な前提と仮説  $E^* = (P^*, H^*)$

for  $t = 1$  から  $T$  まで do

  Step1: 提案

  現在の前提集合  $P'$  に random  $p$  を追加

  prompt  $\leftarrow$  template( $P', O$ )

$H' \leftarrow$  LLM(prompt)

  Step2: スコア計算

$\ell \leftarrow \log P(H' | P) - \log P(H^{(t-1)} | P^{(t-1)})$

$S^t \leftarrow \log P(H' | O)$

  受け入れ確率を計算する:

$$\alpha \leftarrow \min(1, \exp(S^t - S^* + \ell - \text{Penalty}))$$

  Step3: 採用基準

  乱数  $u \sim$  Uniform(0, 1)

  if  $u < \alpha$  then

$P^* \leftarrow P'$

  end if

end for

  prompt  $\leftarrow$  template( $P^* + P, O$ )

$H' \leftarrow$  LLM(prompt)

  if  $\log P(H^* | O) > S^*$  then

$H^* \leftarrow H'$

  end if

  最適説明  $E^*$  を返す

---

追加することによるスコア変化 (残差)

$$\ell = \log P(H' | P') - \log P(H^{(t-1)} | P^{(t-1)})$$

を算出する。そして、この値をういて受け入れ基準を評価する。

**受け入れステップ** 乱数  $u \sim$  Uniform(0, 1) をサンプリングし、 $u < \alpha$  の条件で前提  $p$  を採用または却下する。受け入れ基準は、現在の確率分布とターゲット分布の差に基づいて計算される。元の式は以下の通り:

$$\alpha = \min\left(1, \frac{\pi(x') q(x | x')}{\pi(x) q(x' | x)}\right),$$

ここで:

- $\pi(x)$  はターゲット分布 (目的の確率分布).
- $q(x' | x)$  は現在の状態  $x$  から提案された状態  $x'$  への遷移確率 (提案分布).

本手法では,  $\pi(x)$  を  $\log P(H^* | O)$  と定義し, 遷移確率  $q(x' | x)$  を提案された前提  $p$  の対数確率に基づいて計算する.

**ペナルティ** 全ての前提を無制限に  $P^*$  に追加するのを防ぎ, 生成仮説  $H'$  の品質を重視するため, 2種類のペナルティスコアを  $\alpha$  に加える.  $\beta$  と  $\gamma$  は重み関数のことである.

- **長さペナルティ (Length Penalty)**  $P^*$  の長さに比例したペナルティを課す:

$$LP = -\beta \times \text{len}(P^*)$$

- **仮説ペナルティ (Hypothesis Penalty)** 仮説  $H'$  の品質変化を含意モデルで計算し, その差に基づいてペナルティを課す:

$$HP = -\gamma \times (\log P(H' | O) - \log P(H'^{-1} | O))$$

これらのペナルティにより, 不要な前提の追加を抑えつつ, 高品質な仮説の生成を促進する.

**出力** 最終的に, 最適な前提の優先順位を高め, 全前提集合の先頭に配置した上で, LLM を用いて最良仮説  $H^*$  を生成する.

Q: Stars are organized into patterns called constellations. One constellation is named Leo. Which statement best explains why Leo appears in different areas of the sky throughout the year?

H: the earth revolving around the sun causes leo to appear in different areas in the sky at different times of year

sent1: leo is a kind of constellation

...

sent6: a constellation contains stars

...

sent17: the earth revolving around the sun causes stars to appear in different areas in the sky at different times of year

...

sent25: the earth revolves around the sun

図1 Entailment Bank Task3 Data

## 3 実験

### 3.1 実験設定

**Entailment Bank Dataset** Entailment Bank [6] は, 多段階の推論と自然言語における含意生成を研究するために設計されたベンチマークデータセットである. このデータセットは, 複数の前提をもとに目標仮説を導出するタスクにおいて包括的な枠組みを提供する. 本研究では, Entailment Bank のタスク3を基に実験を行い, 各データアイテムは1つの質問文 (観察  $O$  として扱う), 1つの正解仮説 (説明), および25の前提から構成されている. **実験の主なタスク**は, 25の前提集合から最も支持的な前提を検索し, 正解仮説に近い仮説を生成することである. 図1には, Entailment Bank タスク3のデータ構造が示されている.

### 3.2 評価指標

**Premises の評価** 選択された前提の品質は, 真陽性の前提が前提集合全体に占める割合を基に精度 (Precision) と再現率 (Recall) で評価される. 正解前提データは, Entailment Bank のタスク1データから取得しており, このデータはタスク3と同じ質問-仮説ペアを共有している. さらに, 人間が作成した洗練された証明プロセスを含み, 重要な正解前提が明示されている. 図2にはタスク1のデータで, このデータを図1のデータの ground truth として採用し, 選択された前提の評価を行った例を示す.

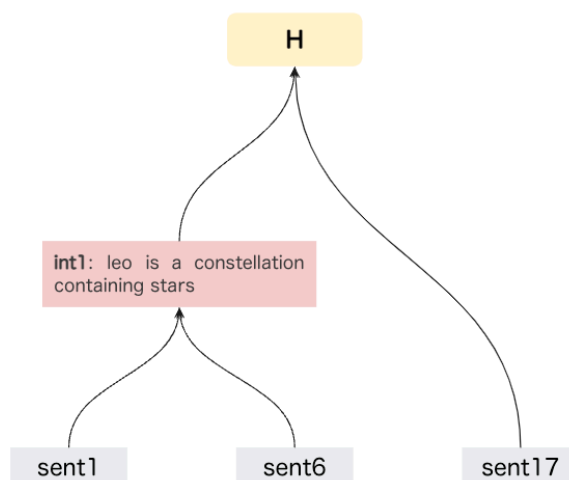


図2 Entailment Bank Task1 Data

表1 異なる GPT モデルにおける MCMC ベースの前提検索の性能比較

Model	Precision(P)	Recall(P)	F1 Score(P)	Semantic Overlap(H)
GPT-4o-mini	0.1593	0.2085	0.1700	0.9775
GPT-4o-mini (MCMC)	<b>0.2381</b>	<b>0.2881</b>	<b>0.2450</b>	0.9777
GPT-4o	0.2809	0.3742	0.3017	0.9772
GPT-4o (MCMC)	0.2490	0.3663	0.2800	<b>0.9787</b>

**Hypothesis の評価** 生成された仮説を正解の答えと比較評価するために、2つのアプローチを採用した。一つ目のアプローチは「意味的オーバーラップ (semantic overlap)」と呼ばれ、表1に示されている。この手法では、2つの仮説の文モデル埋め込み間のコサイン類似度を直接計算する。

2つ目の評価方法は、生成された仮説と正解の仮説間の意味的不確実性を、元の質問文を基に計算するものである。このアプローチは Lorenz Kuhn ら [7] によって提案された。この手法では、まず文対 [質問文, H [SEP] 質問文, GT] と逆文対 [質問文, GT [SEP] 質問文, H] のロジット値を計算し、それぞれについて DeBERTa MNLI モデル [8] を使用して分類を行う。分類結果は、それぞれの文が「Entailment (類似)」、「Neutral (異なるが矛盾しない)」、または「Contradiction (矛盾)」のいずれに該当するかを示す。もし入力ペアのいずれかが 1 (Neutral) または 2 (Contradiction) と予測された場合、これらの仮説は意味的に正解と異なると見なされる。この手法を採用する理由は、セマンティック不確実性に基づく類似性チェックが入力テキストの長さといった要因に影響されず、人間の評価により近い結果が得られるためである。

### 3.3 実験結果とその考察

表1では、直接プロンプトを使用した GPT-4o-mini および GPT-4o の結果と、MCMC 前提検索による強化結果を比較している。ランダム性の影響を抑えるため、データセット (1,007 サンプル) 全体で5回の試行を行い、その平均を算出した。GPT-4o-mini (小型 LLM) では、MCMC 手法によってより多くの支持的前提が見つかり、精度 0.2381 および再現率 0.2881 を達成した。一方、GPT-4o では、MCMC アルゴリズムにより仮説形成がわずかに改善された。

生成された仮説がアルゴリズムによって改善されたかを判断するのは依然困難である。この課題に

対応するため、本研究ではセマンティック不確実性 (3.2 参照) を利用し、生成仮説が質問文に対する因果の意味で正解仮説とどの程度類似するかを評価した。セマンティック不確実性チェックを通過しなかったデータは、MCMC パイプラインで改善され、最終的にチェックを通過した。改良された仮説の生成例はいくつかのケースを付録表 A に示している。特に、元々間違った答えが強化後に正解となるケースでは、選択された前提が正解仮説に必要な因果的情報を含むことが確認された。

**Hyper-parameter** について、Entailment Bank データセットを分析し、各データポイントに **25 の前提セットから平均 4.4** の重要な前提を特定する必要があることがわかった。このデータは探索規模に制限があるものの、MCMC が多くの反復回数で良い性能を示す特性を考慮した。そのため、反復回数を 25 とより少ない 10 に設定し、計算コストを抑えつつ十分な探索性能を確保した。この設定は、現実世界の大規模データベースにおけるヒット率を模倣することを目的としている。

## 4 おわりに

本研究では、完全に教師なしの統計的 MCMC 探索アルゴリズムを導入し、LLM の推論結果を洗練する手法を提案した。このアルゴリズムは、前提知識を伴う推論精度と信頼性の向上を示し、GPT シリーズの性能評価に基づいてその有効性が確認された。本研究の成果は、効果的な前提検索を通じて LLM の推論能力を強化する一助となることが期待される。しかし、現時点では、データセット規模などの制約が存在する。今後は、より複雑なタスクへの適用が期待されており、例えば、11941 の前提を含む WorldTree からの前提検索に取り組むことが課題となる。また、より良い仮説的推論を実現するための報酬プロセスの構築も今後の重要な課題である。

## 謝辞

本研究は異種専門知識の融合と価値観の多様性に基づく次世代 AI 開発人材の育成プログラムの助成を受けたものです。

## 参考文献

- [1] Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Brian Ichter, Fei Xia, Ed H. Chi, Quoc V. Le, and Denny Zhou. Chain-of-thought prompting elicits reasoning in large language models. In **Proceedings of the 36th International Conference on Neural Information Processing Systems**, NIPS '22, Red Hook, NY, USA, 2024. Curran Associates Inc.
- [2] Xinyun Chen, Ryan Andrew Chi, Xuezhi Wang, and Denny Zhou. Premise order matters in reasoning with large language models. In Ruslan Salakhutdinov, Zico Kolter, Katherine Heller, Adrian Weller, Nuria Oliver, Jonathan Scarlett, and Felix Berkenkamp, editors, **Proceedings of the 41st International Conference on Machine Learning**, Vol. 235 of **Proceedings of Machine Learning Research**, pp. 6596–6620. PMLR, 21–27 Jul 2024.
- [3] Gilbert H. Harman. The inference to the best explanation. **Philosophical Review**, Vol. 74, No. 1, pp. 88–95, 1965.
- [4] W. K. Hastings. Monte carlo sampling methods using markov chains and their applications. **Biometrika**, Vol. 57, No. 1, pp. 97–109, 04 1970.
- [5] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. Roberta: A robustly optimized bert pretraining approach. **ArXiv**, Vol. abs/1907.11692, , 2019.
- [6] Bhavana Dalvi, Peter Jansen, Oyvind Tafjord, Zhengnan Xie, Hannah Smith, Leighanna Pipatanangkura, and Peter Clark. Explaining answers with entailment trees. **EMNLP**, 2021.
- [7] Lorenz Kuhn, Yarin Gal, and Sebastian Farquhar. Semantic uncertainty: Linguistic invariances for uncertainty estimation in natural language generation. In **The Eleventh International Conference on Learning Representations**, 2023.
- [8] Pengcheng He, Xiaodong Liu, Jianfeng Gao, and Weizhu Chen. DeBERTa: Decoding-enhanced bert with disentangled attention. In **International Conference on Learning Representations**, 2021.

## A 生成結果例

質問文 1	Drew knows that Earth is tilted on its axis. He also knows this tilt is responsible for the season that a region on Earth will experience. When the South Pole is tilted toward the Sun, what season will it be in Florida?
直接出力	When the South Pole is tilted toward the Sun, <b>Florida will be experiencing summer.</b>
強化後出力	When the South Pole is tilted toward the Sun, it will be summer in the Southern Hemisphere and winter in the Northern Hemisphere, which means <b>Florida, located in the Northern Hemisphere, will be experiencing winter.</b>
正解文	It is winter in Florida.
質問文 2	When the Northern Hemisphere is tilted toward the Sun, what season is occurring in Australia?
直接出力	When the Northern Hemisphere is tilted toward the Sun, <b>it is summer in Australia.</b>
強化後出力	When the Northern Hemisphere is tilted toward the Sun, it is summer in the Northern Hemisphere, which means that <b>Australia, located in the Southern Hemisphere, is experiencing winter.</b>
正解文	australia is in winter
質問文 3	Sheep Challenge Evolution A species of sheep lives on the Scottish island of Hirta. Hirta has plenty of grass for the sheep to eat, and no natural predators of the sheep live on the island. Some sheep have dark wool, and some have light wool. The sheep with dark wool tend to be much larger and stronger than the sheep with light wool. However, researchers have found that in the last 20 years, the sheep with light wool have increased in number. Usually, strength and size help animals survive and breed. On Hirta, sheep with light wool outnumber sheep with dark wool because sheep with light wool
直接出力	The increase in the population of sheep with light wool on Hirta is likely due to their better camouflage in the grasslands of the island, which provides them with a survival advantage despite the larger size and strength of the dark wool sheep.
強化後出力	Sheep with light wool have a better ability to survive and reproduce on Hirta due to their adaptation to the environment, <b>which may include factors such as improved thermoregulation, camouflage against the grass, or other advantages that outweigh their size and strength compared to dark wool sheep.</b>
正解文	sheep with light wool have a better chance of surviving and reproducing than sheep with black wool on hirta