

大規模言語モデルによる仮説推論のための 対話的プロンプティング

Interactive Prompting for Hypothesis Exploration with Large Language Models

中根慎哉¹ 亀谷由隆^{2*}
Shinya Nakane¹ Yoshitaka Kameya²

^{1,2} 名城大学 理工学部 情報工学科

^{1,2} Dept. of Information Engineering, Faculty of Science and Technology, Meijo University

Abstract: In this study, we propose a prompting method that encourages large language models (LLMs) to explore hypotheses. Specifically, we consider C. S. Peirce's logic theory of hypothetical reasoning (the logic of inquiry) as a model of the thought process of a human reasoner when reasoning about a hypothesis, and attempt to encourage LLMs to perform hypothetical reasoning by providing them with a prompt template based on this theory. This method is inspired by the Chain-of-Thought Prompting approach, which encourages LLMs to reason by providing them with the thought processes of human reasoners as prompts. The results of our experiments using GPT-4 suggested that the proposed prompting method increases the testability of the hypotheses generated by the LLMs and helps continuous and appropriate hypothesis reasoning.

1 はじめに

近年、大規模言語モデル (large language models, 以下 LLM) の性能を高めるプロンプティング手法が研究されている。Chain-of-Thought (CoT) プロンプティングはそのような中で開発された手法の一つであり、モデルの推論能力の増強を目的として提案された。タスクの最終的な答えを導く中間推論ステップの生成をモデルに促す CoT プロンプティングは、多くの推論タスクにおいて、標準的なプロンプティングに比して、規模の増加に伴う正確性向上の度合いをより著しいものにする事が明らかになった [1]。このことから、人間が推論する際の思考過程の模倣をモデルに促すようなプロンプティング手法は、モデルが正常にこなすタスクの幅を拡大するとの見方が広がっている。そのような手法を包括的に調査した Huang & Chang [2] は、経験的結果から、そのような手法が、「LLM の算術タスクや記号推論タスクや常識推論タスクに対する正確性を (ときとして顕著に) 高めることができる」としている。

本研究では、推論の一つである仮説推論のためのプロンプティング手法を提案する。この手法は、科学的仮説探究はアブダクション (仮説の発想) と演繹 (仮説の分析) と帰納 (仮説の実証) の三つの推論段階から

なるとした C. S. Peirce の理論 (the logic of inquiry) [3] を、人間が仮説推論する際の思考過程のモデルとみなし、その模倣を言語モデルに促すものである。これによって適切な仮説を生成することができるか検証する。また、生成された仮説から導かれる予測を正しいと行うために観察されるべきとモデルが主張する事柄が、実際に観察されたか否かを追加のプロンプティングとして対話的に与えたとき、それらの事実を受けたモデルが適切に仮説推論を継続することができるか検証する。

2 提案手法

提案するプロンプティング手法は、Peirce による科学的仮説探究についての理論 [3] を人間が仮説推論する際の思考過程のモデルであると仮定して、その模倣を促すことによる効果を期待している。

Peirce の理論によると、科学的探究は、説明を要する事実の観察から始まり、三つの推論段階へと続く。第一段階のアブダクションは、観察されたある事実が何故起こったかについての何らかの可能な説明を与える仮説を考え出す。第二段階の演繹は、仮説が真であるとしたとき、その仮説からどのような予測が必然的にまたは高い確率で導かれるかを示す。第三段階の帰納は、仮説から導出された諸予測がどれだけ経験的諸事実と一致す

*連絡先: 名城大学 理工学部 情報工学科
〒468-8502 名古屋市天白区塩釜口一丁目 501 番地
E-mail: ykameya@meijo-u.ac.jp

Model Input

以下の問いに答えてください。

事実「陸地のずっと内側で魚の化石が見つかった」

問1:この事実を説明する現実的でもっともらしい仮説を考えてください。

問2:もしあなたの仮説が真であるとすると、あなたの仮説からどんな予測が必然的に(あるいは非常に高い確率で)導かれますか。列挙してください。

問3:あなたの仮説から導かれた予測を正しいと言うためには、それぞれどのような事実が調査・観察されればよいでしょうか。考えてください。

図 1: 提案するプロンプティングの例。

Additional Input

調査の結果、以下の事実が観察されました。

- ・その地域がかつて水域だったことを示す堆積岩層や、水底環境の特徴を示す化石層は発見されなかった
- ・過去に水域が存在していた時期は特定されなかった
- ・同じ地層から他の水生生物の化石が発見されなかった

このことを受けて、あなたは以下の三つの選択肢から一つを選ぶことができます。

- 1:仮説は現時点では受容されるため、引き続きこの仮説の検証を続ける。
- 2:仮説は本質的でない修正を必要とするため、修正を施し、引き続きこの仮説の検証を続ける。
- 3:仮説は拒否されるため、新しい仮説の検証を始める。

図 2: 仮説推論の継続を促す追加プロンプティングの例。

るか確かめる。

以上を踏まえ、提案するプロンプティング手法の例を図 1 に示す。まず説明を要するある事実を与え、その暫定的な答えである仮説の生成をモデルに促す。そして仮説から導出される予測の生成をモデルに促し、その予測を確証する為に必要な事実の生成をモデルに促す。また Peirce によると、帰納は繰り返されることで正当化される推論であるため、提案手法では、モデルからの出力に対し、図 2 に示すような追加プロンプティングを繰り返し対話的に与えることで帰納の正当性を保証することを目指した。

追加プロンプティングは、モデルに追加の事実を与える部分と、それを受けて仮説をどうするかという選択を促す部分からなる。モデルに与える選択肢は受容・修正・拒否の三つである。

3 実験 1：演繹と帰納を促すことによる影響の検証

3.1 実験方法

実験 1 では、演繹と帰納の段階をモデルに促すことが、モデルの生成する仮説にどのような影響を与えるか検証する。具体的には、図 3 に示す、提案手法から演繹

Model Input

以下の問いに答えてください。

事実「陸地のずっと内側で魚の化石が見つかった」

問1:この事実を説明する現実的でもっともらしい仮説を考えてください。

図 3: 標準的なプロンプティングの例。

Model Output

仮説:哺乳類において大きい種ほど顔が長く、小さい種ほど顔が短くなる傾向は、体のサイズと代謝率、および環境への適応の違いに関連している可能性があります。大きな体を持つ哺乳類は一般に低い代謝率を持ち、それによって長い顔が効率的な呼吸や冷却システムを提供することができます。一方で、小さい哺乳類は高い代謝率を持ち、よりコンパクトな顔の構造が必要で、これは熱の保持や迅速な顔の摂取に有利です。さらに、大きな顔はセンサー機能(視覚、嗅覚、聴覚)を強化し、捕食者や獲物を探知するのに有効であり、これは大型の哺乳類に特に重要な特徴です。一方、小型の哺乳類では、このようなセンサー機能の拡張は生存戦略としてそれほど重要ではないことがあります。

図 4: 標準的な手法による仮説

Model Output

仮説:哺乳類において、大きい種ほど顔が長く、小さい種ほど顔が短くなる傾向は、体のサイズと感覚器官の配分との関係によるものです。大きな体を持つ種では、鼻や耳などの感覚器官がより広範囲に分布し、その結果、顔の長さが増す傾向があります。これは、体の大きさに比例して感覚範囲を広げる必要があるためです。一方、小さい種では、感覚器官がより集中して配置され、結果的に顔が短くなります。

仮説から導かれる予測:

1. 大型の哺乳類(例:ゾウ、馬)は、感覚器官が顔全体に広範囲に分布している。
2. 小型の哺乳類(例:ネズミ、リス)は、感覚器官が顔の限定された部分に集中している。
3. 体のサイズが異なる哺乳類間で、感覚器官の相対的な位置やサイズに顕著な違いが見られる。

観察されるべき事実:

1. 大型の哺乳類の顔の構造を詳細に分析し、感覚器官が広範囲に分布していることを確認する。
2. 小型の哺乳類の顔の構造を分析し、感覚器官がより限定された領域に集中していることを確認する。
3. 異なるサイズの哺乳類を比較し、顔の長さや感覚器官の位置やサイズとの関係を統計的に分析する。

図 5: 提案手法による仮説

と帰納を促す問 2, 3 を除いたプロンプティングを標準的な手法として、両手法によって生成される仮説にどのような差が見られるか検証する。

Manning [4] は LLM のアプローチについて、「この手法では、システムは極めて大量のテキストに触れるだけで、言語や世界に関する膨大な知識を学習することができる」と述べ、LLM が世の中の一般的な知識を獲得しているという見方を示している。このような見方に示されているように、モデルに常識と呼ぶべきものが蓄えられているとすれば、それをバイパス的に用いることで、たとえ演繹と帰納を生成しない一度きりの仮説生成でも、尤もらしい仮説を生成することは可能であるように思われる。

そのため、説明を求める事実として、「哺乳類は大きい種ほど顔が長く、小さい種ほど顔が短くなる傾向がある」という最近説明がなされた事象 [5] を与え、提案手法と標準的な手法のそれぞれによって、GPT-4 を用いて仮説生成を 8 回行った。

3.2 実験結果

それぞれの手法によって生成した仮説の一例を図 4, 5 に示す。提案手法によって生成された出力は仮説だけでなく、そこから導かれる予測と、その予測を正しいと言うために観察されるべきとする事柄も含むため、仮説の部分を緑色のマーカーで示した。

Mitchell ら [5] は、このような傾向が見られる理由について、哺乳類は一般に、四足歩行を常とし、手先を器用には用いることができないため、顔を長くして口先の

機能性を高める必要があるが、大型種に比して顔まわりの筋肉量に限りがある小型種においては、顔を短くして噛む力を高める必要の方が優先されるためである、と説明付けている。それに対し、図4の標準的な手法による仮説では、大型種について「代謝率が低いため、顔を長くすることで表面積を増し冷却の効率を高める」「顔を長くすることで感覚機能（視覚や聴覚や嗅覚）が強化される」「呼吸効率を高める」、小型種について「感覚機能の拡張が必要ない」「代謝率が高いため顔を短くすることで熱を保持する」という観点から説明している。一方、図5の提案手法による仮説では、「身体の大きさ」の観点から説明している。

両手法によって生成された仮説は、試行ごとに内容のばらつきが多く見られたが、それぞれ共通する特徴も見られた。標準的な手法によって生成された仮説は一見尤もらしいが、視点が一貫しておらず、全体としてどのようなことを主張しているのかわかりにくい。その一方で、提案手法によって生成された仮説は、内容の尤もらしさにはさほど変化が見られないが、主張していることが明確でわかりやすいものであった。この結果から、GPT-4の仮説推論性能は限られたものであることや、提案手法はモデルの仮説推論性能自体は高めないものの、生成する仮説の検証可能性を高めることなどが示唆された。

4 実験2：仮説推論を適切に継続することが可能かどうかの検証

4.1 実験方法

実験1の結果を受けて、実験2では、提案手法によってGPT-4が仮説推論を適切に推し進めることができるか検証する。具体的には、まずGPT-4が適切な仮説を安定的に生成できることが分かっている「陸地のずっと内側で魚の化石が見つかった」という仮想的な事実について提案手法によって仮説を生成させる。このとき、仮説と共に生成される、仮説から導かれる予測を正しいと言うために観察されるべきとモデルが主張する事柄が、実際に観察されたか否かを図2に示したような追加のプロンプティングとして与えたとき、それを受けたモデルが自身の仮説を適切に受容・修正・拒否することができるかということを検証する。

また、追加プロンプティングによる事柄の与え方は、モデルによる予測と反する事柄だけを観察された事実として与えた場合、モデルによる予測と一致する事柄だけを観察された事実として与えた場合、そして、モデルによる予測と一致する事柄と反する事柄の両方を観察された事実として与えた場合に分けて実験する。

Model Output

仮説:この地域はかつて海や湖といった水域だったが、地質学的な変化(例えば、地盤の隆起や気候変動による乾燥化)によって現在は内陸となった。この過程で、かつての水域に生息していた魚の化石が埋もれ、今になって発見された。

仮説から導かれる予測:

- その地域の地質学的構造には、かつて水域であったことを示す証拠(例えば、堆積岩や古代の海底環境を示す化石など)が存在する。
- 地域の気候史や海水準の変化に関するデータが、かつて水域が存在していた時期を示している。
- 他の水生生物(貝殻、水草、他の魚類等)の化石も同様に発見される可能性が高い。

観察されるべき事実:

- 地質調査により、その地域がかつて水域だったことを示す堆積岩層や、海底環境の特徴を示す化石層が発見される。
- 気候史や海水準の変化に関する研究が、過去に水域が存在していた時期を特定し、その時期が現在の内陸地形に変わる前であることを示す。
- 同じ地層から他の水生生物の化石が発見され、それらが同じ時期に生息していたことが確認される。

Additional Input

調査の結果、以下の事実が観察されました。

- その地域がかつて水域だったことを示す堆積岩層や、海底環境の特徴を示す化石層が発見されなかった
- 過去に水域が存在していた時期は特定されなかった
- 同じ地層から他の水生生物の化石が発見されなかった

このことを受けて、あなたは以下の三つの選択肢から一つを選ぶことができます。

- 仮説は現時点では受容されるため、引き続きこの仮説の検証を続ける。
- 仮説は本質的に修正を必要とするため、修正を施し、引き続きこの仮説の検証を続ける。
- 仮説は拒否されるため、新しい仮説の検証を始める。

選択肢1を選んだ場合の回答のフォーマットを以下に示します。
 [選んだ選択肢]
 仮説から導かれる予測:[ここに問2の回答]
 観察されるべき事実:[ここに問3の回答]

選択肢2および3を選んだ場合の回答のフォーマットを以下に示します。
 [選んだ選択肢]
 新しい仮説:[ここに問1の回答]
 仮説から導かれる予測:[ここに問2の回答]
 観察されるべき事実:[ここに問3の回答]

Model Output

[3:仮説は拒否されるため、新しい仮説の検証を始める]

新しい仮説:この魚の化石は、他の地域から運ばれてきた可能性がある。地質学的な変化(例えば洪水や土砂崩れなど)や、人為的な活動(例えば古代の人々が魚を狩猟し、運搬してきたなど)によって、魚の化石が内陸部に移動した。

仮説から導かれる予測:

- 化石が見つかった地層は、遠方の水域の地層と類似しているか、あるいは異なる地層の間に挟まれている。
- 魚の化石の保存状態が、運搬による特徴(破損や摩耗など)を示している。
- この地域の周辺に、同じ時期の水域が存在していた証拠が見つかる。

観察されるべき事実:

- 化石が発見された地層の地質学的な分析により、化石が他の場所から運ばれてきたことを示唆する証拠(異なる地層の組み合わせ、遠方の水域の特徴を示す砂や岩石など)が見つかる。
- 化石の詳細な分析により、運搬の過程で生じた可能性のある特徴(摩耗や破損)が確認される。
- 地質学的、地理学的な調査により、この地域の近くに同じ時期の水域が存在していた可能性を示す証拠が見つかる。

図6: 予測と反する事柄のみを与えた場合の推論の進行。

GPT-4とのやり取りの例においては、モデルによる選択を赤色のマーカー、生成された仮説を緑色のマーカー、そこから導かれた予測を正しいと言うために観察されるべきとする事柄と、それらが観察されたか否かを追加プロンプティングで与えた部分を黄色のマーカーで示す。

4.2 実験結果

図6に、モデルの予測を反する事柄のみを観察された事実として追加プロンプティングで与えた場合の推論の進行を示す。この場合、モデルは与えられた事柄を受けて自身の仮説を拒否し、そのたびに新しい仮説を生成して仮説推論を継続した。推論が進むにつれて仮説の尤もらしさは減じられていくような印象を受けた。

Model Output

仮説: 内陸で魚の化石が見つかった事実は、過去にその地域が水域（海や湖など）であったという証拠です。数百万年前、その地域は海または大きな湖の一部で、魚が豊富に生息していた可能性があります。地質学的プロセスにより、その水域は徐々に陸地になり、現在の内陸地形になったと考えられます。

仮説から導かれる予測:

- 同じ地域で他の海洋生物や湖の生物の化石が見つかる可能性が高い。
- 地層の分析から、その地域が過去に水域であったことを示す証拠（例えば、水生植物の化石、水中の堆積物など）が見つかるはず。
- 地質学的調査により、過去にその地域が海や湖に接していたことを示す地形学的特徴（例えば、古代の海岸線、堆積物の層）が見つかる可能性がある。

観察されるべき事実:

- 同じ地層や近隣地域で他の海洋生物や湖の生物の化石が発見される。
- 地層の化石記録や堆積物の分析から、その地域が過去に水域であったことを示す明確な証拠が得られる。
- 地質学的な証拠（例えば、地形学的特徴、堆積物の層の分析）から、その地域がかつて水域であったことを裏付ける情報が得られる。

Additional Input

調査の結果、以下の事実が観察されました。

- 同じ地層や近隣地域で他の海洋生物や湖の生物の化石が発見された。
- 地層の化石記録や堆積物の分析から、その地域が過去に水域であったことを示す明確な証拠が得られた。
- 地質学的な証拠（例えば、地形学的特徴、堆積物の層の分析）から、その地域がかつて水域であったことを裏付ける情報が得られた。

このことを受けて、あなたは以下の三つの選択肢から一つを選ぶことができます。

- 仮説は現時点では受容されるため、引き続きこの仮説の検証を続ける。
- 仮説は本質的でない修正を必要とするため、修正を施し、引き続きこの仮説の検証を続ける。
- 仮説は拒否されるため、新しい仮説の検証を始める。

選択肢1を選んだ場合の回答のフォーマットを以下に示します。
[選んだ選択肢]
仮説から導かれる予測:{ここに問2の回答}
観察されるべき事実:{ここに問3の回答}

選択肢2および3を選んだ場合の回答のフォーマットを以下に示します。
[選んだ選択肢]
新しい仮説:{ここに問1の回答}
仮説から導かれる予測:{ここに問2の回答}
観察されるべき事実:{ここに問3の回答}

Model Output

[選んだ選択肢]
1: 仮説は現時点では受容されるため、引き続きこの仮説の検証を続ける。

仮説から導かれる予測:

- 追加の地層や地域で、さらに多様な海洋生物や湖の生物の化石が見つかる可能性がある。
- 地層のさらなる分析から、その地域が特定の地質時代に水域であったことを示す、より詳細な証拠が発見される。
- 地質学的調査が進むにつれ、過去の水域の大きさや形状、生態系の特徴に関するより具体的な情報が明らかになる。

観察されるべき事実:

- 同様の地層や他の近隣地域での追加的な化石の発見。
- 地層の詳細な分析から、その地域が特定の時期にどのような水域であったか（例えば、深海、沿岸、湖沼等）を示す証拠の発見。
- 地質学的調査から、過去の水域の具体的な特徴や生態系に関する情報の明確化。

図 7: 予測と一致する事柄のみを与えた場合の推論の進行。

次に、図 7 に、モデルの予測と一致する事柄のみを観察された事実として追加プロンプティングで与えた場合の推論の進行を示す。この場合、モデルは与えられた事柄を受けて自身の仮説を受容して、仮説推論を継続した。始めの出力で演繹した予測をより深く検証するような方向に推論を進めており、一見仮説の尤もらしきは向上するようだったが、自身の仮説が怪しくなることを避けているような印象を受けた。

最後に、図 8 に、モデルの予測と一致する事柄と反する事柄の両方を観察された事実として追加プロンプティングで与えた場合の推論の進行を示す。与えられた事柄を受けて自身の仮説を修正したり拒否したりして、仮説推論を継続した。また、どの場合でも、演繹と帰納の段階では誤りらしい誤りは見られなかった。

この結果から、GPT-4 には自身の仮説が怪しくなる

Model Output

仮説: この地域が以前は海や大きな湖の一部であったが、長い地質学的期間を経て陸地になった。過去の海洋や湖の存在は、堆積層の分析や地層の構造によって裏付けられる可能性がある。

仮説から導かれる予測:

- 魚の化石が見つかった地層には、他の海洋生物や湖の生物の化石も含まれている可能性がある。
- この地域の地質学的調査により、古い海洋や湖の底を示す堆積物や特徴的な地層が見つかる可能性がある。
- 地域の地形や地質から、かつての海岸線や湖岸線の跡が観察できるかもしれない。

観察されるべき事実:

- 同じ地層から海藻や貝殻、魚以外の水生生物の化石が見つかること。
- 地質学的調査によって、海洋や湖の堆積物や、その他の関連する地質学的特徴（例えば、波の作用による堆積物の層）が確認されること。
- 地形分析や地質学的証拠によって、かつての海岸線や湖岸線が推定され、その推定が現在の地形と整合性を持つこと。

Additional Input

調査の結果、以下の事実が観察されました。

- 同じ地層から海藻や貝殻、魚以外の水生生物の化石が見つかった。
- 地質学的調査によって、海洋や湖の堆積物や、その他の関連する地質学的特徴（例えば、波の作用による堆積物の層）は確認されなかった。
- 地形分析や地質学的証拠によって、かつての海岸線や湖岸線が推定されなかった。

このことを受けて、あなたは以下の三つの選択肢から一つを選ぶことができます。

- 仮説は現時点では受容されるため、引き続きこの仮説の検証を続ける。
- 仮説は本質的でない修正を必要とするため、修正を施し、引き続きこの仮説の検証を続ける。
- 仮説は拒否されるため、新しい仮説の検証を始める。

選択肢1を選んだ場合の回答のフォーマットを以下に示します。
[選んだ選択肢]
仮説から導かれる予測:{ここに問2の回答}
観察されるべき事実:{ここに問3の回答}

選択肢2および3を選んだ場合の回答のフォーマットを以下に示します。
[選んだ選択肢]
新しい仮説:{ここに問1の回答}
仮説から導かれる予測:{ここに問2の回答}
観察されるべき事実:{ここに問3の回答}

Model Output

[選んだ選択肢]: 2: 仮説は本質的でない修正を必要とするため、修正を施し、引き続きこの仮説の検証を続ける。

新しい仮説: 地域が以前に海洋や湖であったことを示す地質学的証拠が見つからなかったため、別の可能性として、この地域にはかつて淡水の河川や湖が存在し、それが現在の地形に変化したことを仮定する。これは河川や湖に由来する堆積物の特徴、例えば粗大な砂や砂利、河川の流れによる堆積層のパターンによって検証される。

仮説から導かれる予測:

- 魚の化石が発見された地層には、淡水生物の化石も含まれている可能性がある（例: 淡水魚、水生昆虫、淡水性の貝類など）。
- 地質学的調査では、河川や湖由来の特徴的な堆積物（例: 粗大な砂や砂利、河川由来の層状構造）が確認される可能性がある。
- 地形分析や地質学的証拠から、古い河川の流れや湖の存在を示す特徴が見つかるかもしれない。

観察されるべき事実:

- 同じ地層から淡水性生物の化石が発見されること。
- 地質学的調査により、河川や湖由来の堆積物の特徴（例: 粗大な砂や砂利、河川由来の層状構造）が確認されること。
- 地形分析や地質学的証拠により、古い河川の流れや湖の存在が推定されること。

図 8: 予測と一致する事柄と反する事柄の両方を与えた場合の推論の進行。

ような検証を避けようとする消極的な傾向があることが示唆された。また、事実として与えられた事柄が総体として説明することが難しい事柄になっていくとき、生成される仮説の尤もらしさが減じられていくことから、GPT-4 の仮説推論性能は限られたものであるということや、提案手法はモデルの仮説推論性能自体は高めないということが再び示唆された。しかし、演繹と帰納の段階では誤りらしい誤りは見られないことから、GPT-4 は仮説推論を適切に継続できることが示唆された。

5 おわりに

本研究では, Peirce の理論に基づいたプロンプティング手法を提案した. 実験の結果, GPT-4 の仮説推論性能は限られたものであることや, 提案するプロンプティング手法はモデルの仮説推論性能自体は高めないものの, 生成する仮説の検証可能性を高め, 適切な仮説推論の継続を助けることが分かった.

今後の課題として, 仮説推論性能を高めるプロンプティング手法の開発や, モデルの仮説検証に対する消極的な姿勢を是正するプロンプティング手法の開発などが挙げられる.

参考文献

- [1] J. Wei et al.: Chain of Thought Prompting Elicits Reasoning in Large Language Models, *Proc. of NeurIPS*, 2022.
- [2] J. Huang and K. C. Chang: Towards Reasoning in Large Language Models: A Survey, *Proc. of ACL*, 2023.
- [3] 米盛祐二: アブダクション, 勁草書房, 2007.
- [4] C. D. Manning: Human Language Understanding & Reasoning, *Daedalus*, vol. 151, No. 2, pp. 127-138, 2022.
- [5] D. R. Mitchell et al.: Facing the Facts: Adaptive Trade-offs Alongbody Size Ranges Determine Mammaliancraniofacial Scaling, *Biological Reviews*, 2023.