

2023 年度（令和 5 年度） 卒業論文

多段推論を用いた社会問題に対する発想支援

指導教員

村田 真樹

村上 仁一

鳥取大学工学部 電気情報系学科

自然言語処理研究室

B20T2068B 中田 鈴太郎

概要

本研究では ChatGPT を用いて、社会問題に対して多段推論を行った。多段推論は、複雑な問題に対するアプローチを段階的に進め、異なる観点や情報を組み合わせながら推論を深める手法である。本研究では ChatGPT-3.5 を用いて、提案手法の性能を確かめるためにベースラインを用いて実験を行った。

提案手法で用いたプロンプトは Tree of Thoughts であり、ベースラインで用いたプロンプトは Input-Output である。実験では 10 個の異なる社会問題を対象にし、ChatGPT と人手による評価を通じて提案手法とベースラインの各設問の正解率を F 値を用いて求め、その平均を算出した。

評価においては、提案手法の出力の”画期的さ”とベースラインの出力の”画期的さ”を比較し、どちらが人から見たときに画期的と感じるかという検証を行った。実験は被験者 4 人で行い、被験者ごとにベースラインと提案手法の F 値とその平均値を求めた。結果として、ベースラインでの F 値は 0.21、提案手法での F 値は 0.15 となり提案手法よりもベースラインの方が性能が高いという結果が得られた。

これらの結果から、提案手法のプロンプトの改善や調整が必要である可能性が示唆され、今後の研究方針や手法の検討が求められる。

目次

第1章	はじめに	5
第2章	関連研究	6
	2.1 大規模データを用いた社会構造ネットワークモデル	
	2.2 Tree of Thoughts: Deliberate Problem Solving with Large Language Models.	
	2.2.1 Tree of Thoughts (思考の木)	
	2.2.2 Chain of Thought (思考の連鎖)	
	2.2.3 Input-Output (入力と出力)	
第2章	提案手法	9
	3.1 ベースライン	
	3.1.1 プロンプト	
	3.1.2 入出力例	
	3.2 提案手法	
	3.2.1 プロンプト	
	3.2.2 入出力例	
	3.3 ChatGPT	
第4章	実験	21
	4.1 実験データ	

4.2 実験方法

4.3 実験結果

4.3.1 Tree of Thoughts を用いた実験結果

4.3.2 Input-Output を用いた実験結果

4.3.3 性能比較

第5章 考察.....29

第6章 おわりに.....30

謝辞.....31

参考文献.....32

第1章 はじめに

多段推論は、複雑な問題に対するアプローチを段階的に進め、異なる観点や情報を組み合わせながら推論を深める手法であり、本研究では ChatGPT を用いて社会問題においてその有用性を確かめる。ChatGPT を用いて多段推論を行い、社会問題に対する新たな洞察や解決策を得ることを目指す。以下は本研究の主張点である。

- ・ 提案手法では、ChatGPT を用いて社会問題に対する多段推論を行う。
ChatGPT は自然言語処理の分野で高い性能を発揮し、複雑な問題に対しても柔軟に対応することができる。
- ・ 被験者実験において、提案手法とベースラインを比較した結果、ベースラインの方が F 値が高い結果が得られた。この結果から、提案手法がベースラインよりも性能が低い可能性が示唆される。これは、提案手法の改善や調整が必要であることを示唆している。
- ・ 提案手法において、プロンプトの改善が必要であることが示唆される。現在のプロンプトでは出力が被験者に対して適切な情報やガイダンスが提供されていない可能性があると考えられる。プロンプトの改善を通じて、被験者が問題に対する適切な回答を提供することができるようにする必要がある。

論文の構成は以下の通りである。第2章では、本研究に関連する研究を説明する。第3章では提案手法の説明を行う。第4章では実施した実験とその結果について記述する。第5章では得られた実験結果の考察を行う。第6章では本研究のまとめを記述する。

第2章 関連研究

本章では、先行研究について記述する。2.1節では大竹が行った単語ネットワーク[1]について記述する。2.2節では Tree of Thoughts について記述する。

2.1 大規模テキストデータを用いた社会構造ネットワークモデルの自動抽出

大竹ら[1]は構築したい単語ネットワークの主となる概念をキーワードとして設定し、それに関係した電子テキストを抽出する。そのテキストにおいて関係性が強い単語を抽出し、キーワードと抽出した単語をノードとしたネットワークを構築した。構築した単語ネットワークを図1に示す。

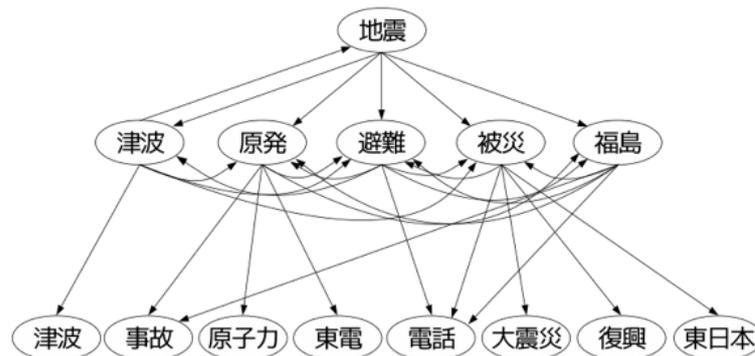


図1：抽出された社会構造モデルの一部

2.2 Tree of Thoughts: Deliberate Problem Solving with Large Language Models.

本節では, Tree of Thoughts[2]と, 従来の手法である Chain of Thought, Input-Output の説明を記述する. また, 図 2 は Tree of Thoughts, Chain of Thought, Input-Output のフレームワーク[3]であり ChatGPT を用いたとき, (a)Input-Output は 質問と回答が一对一の手法である. (b)Chain of Thought は結論に行き着くためにステップバイステップで質問を行い回答を得る手法である. (c)Self Consistency with CoT は(b)Chain of Thought を複数回行い, 多数派の回答を得る手法である. (d) Tree of Thoughts は複数のアイデアを回答してもらい, それに対して評価と深掘りを行い, 最終的にベストなものを選ぶ手法である.

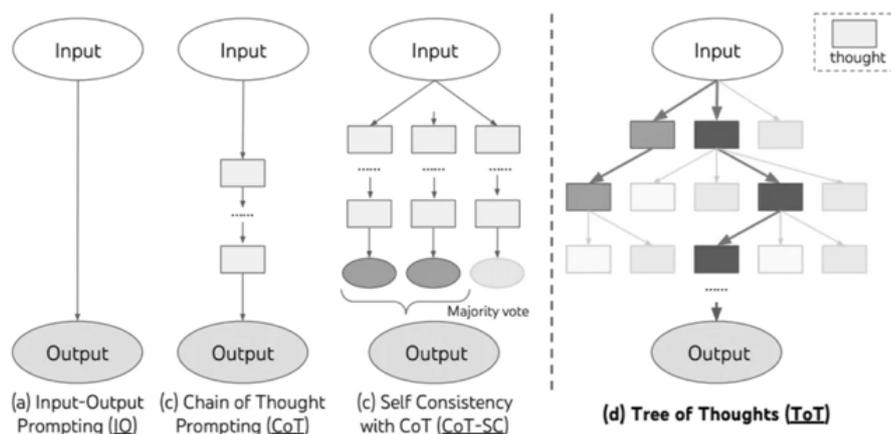


図 2 : ToT, CoT, Input-Output の概要図

2.2.1 Tree of Thoughts (思考の木)

Tree of Thoughts は, 樹木のように, 一つの中心的なアイデアやトピックから派生したり, 関連するアイデアが連鎖的に広がったりする構造を指す. この手法では, アイデアが根から芽を出し, それが枝分かれして新しい派生アイデアを生み出す様子をイメージできる. 例えば「環境問題」という中心のアイデアから「プラスチックごみ削減」という枝が生え, さらにその枝から「リサイクルプログラムの拡充」「バイオデグレード可能な代替材料の研究」といった具体的なサブアイデアが派生する構造である. このような思考の構造を用いることで, 問題解決やアイデアの発展において, 従来の一直線的なアプローチに比べて柔軟性があり, より多様かつ豊かな解決策やアイデアを生み出すことが期待される.

2.2.2 Chain of Thought (思考の連鎖)

Chain of Thought は、一つのアイデアが次のアイデアに連なり、それがまたその次のアイデアに続くような連続的で順次的な思考のプロセスを指す。これは、一つの考えや概念が連鎖的に発展し、次から次へと続いていく構造を表現している。例えば「新しい商品開発」という中心の考えから、「市場調査を実施する」「顧客のフィードバックを得る」「競合商品の分析を行う」といった具体的なステップやサブアイデアが順次展開されていく構造である。この手法では、一つのアイデアが次の段階に発展していく連鎖的なプロセスが、問題解決や新しいアイデアの生成において非常に有益であるとされている。

2.2.3 Input-Output (入力と出力)

Input-Output は、ユーザが対話システムに質問をする際に、その質問に対してシステムが正確な応答を生成する仕組みを指す。ユーザの質問が入力として機能し、それに基づいてシステムが適切な出力を生成する。質問と答えは一对一の関係を持ち、ユーザが特定の情報を尋ねると、対話システムがそれに適した回答を提供する構造となっている。

第3章 提案手法

本研究では ChatGPT を使用して社会問題に対する多段推論を行う。本章ではベースラインと提案手法の説明を記述する。3.1 節はベースラインについて、3.2 節では提案手法についての大まかな流れを記述する。3.3 節では本研究で使用する ChatGPT の解説を記述する。

3.1 ベースライン

本研究では、提案手法の有効性を検証するために、ベースラインとして Input-Output を用いる。

3.1.1 プロンプト

ベースラインとして Input-Output を用いることで、提案手法の相対的な優位性を明確に示すことができる。提案手法と比較して、Input-Output がどの程度の画期的な対策を生成し、その対策の画期的さをどの程度評価するかを比較することで、提案手法の効果や改善点を把握することができる。また、Input-Output をベースラインとして用いることで、提案手法の改善点を特定することができる。提案手法がどのような点で Input-Output よりも優れているか、またどのような点で改善の余地があるかを明確にすることで、提案手法の品質向上に向けた具体的な方針を立てることができる。

表1のベースラインのプロンプトは、参加者に対して20個の画期的な対策を挙げ、その”画期的さ”を評価するように求めている。

表1：ベースラインのプロンプト

(1)	(設問)において、今までにはない画期的な対策を20個挙げてください
(2)	(設問)において、以下の対策の画期的さを評価し、10点満点で評価してください {出力(1)}

3.1.2 入出力例

ベースラインにおいて表2に表1の設問を「海洋プラスチック問題」としたときの入出力例を示す。

表2：ベースラインの入出力例

(3)	<p>User : 海洋プラスチック問題において、今までにはない画期的な対策を20個挙げてください</p> <hr/> <p>ChatGPT : 海洋プラスチック問題に対する画期的な対策を以下に挙げます。</p> <ol style="list-style-type: none"> 1. プラスチック分解酵素の利用: <ul style="list-style-type: none"> • 海洋に流れ込んだプラスチックを分解する特殊な酵素の導入。 2. バイオプラスチックの推進: <ul style="list-style-type: none"> • 海洋プラスチックの代替となるバイオプラスチックの研究・普及。 3. 海洋プラスチック回収ドローン: <ul style="list-style-type: none"> • ドローンを用いて海洋上のプラスチックを効率的に検知・回収。 • • (省略) • 20. プラスチック規制強化: <ul style="list-style-type: none"> • 未処理のプラスチックの排出を厳しく規制し、法的な制裁を導入。
(4)	<p>User : 海洋プラスチック問題において、以下の対策の画期的さを評価し、10点満点で評価してください</p> <ol style="list-style-type: none"> 1. プラスチック分解酵素の利用: • 海洋に流れ込んだプラスチックを分解する特殊な酵素の導入。 2. バイオプラスチックの推進: • 海洋プラスチックの代替となるバイオプラスチックの研究・普及。 3. 海洋プラスチック回収ドローン: • ドローンを用いて海洋上のプラスチックを効率的に検知・回収。

	<p style="text-align: center;">.</p> <p style="text-align: center;">.</p> <p style="text-align: center;">.</p> <p>20. プラスチック規制強化: • 未処理のプラスチックの排出を厳しく規制し、法的な制裁を導入。</p> <hr/> <p>ChatGPT :</p> <p>1. プラスチック分解酵素の利用 (9/10)</p> <ul style="list-style-type: none"> • 海洋のプラスチックを分解する酵素の利用は、プラスチックの除去において革新的であり、環境に対する影響が最小限。 <p>2. バイオプラスチックの推進 (8/10)</p> <ul style="list-style-type: none"> • バイオプラスチックの研究・普及はプラスチック代替の一環として有望であるが、市場への普及までには時間がかかる可能性。 <p>3. 海洋プラスチック回収ドローン (9/10)</p> <ul style="list-style-type: none"> • ドローンによる海洋プラスチックの検知・回収は効率的であり、プラスチックの迅速な除去に寄与。 <p style="text-align: center;">.</p> <p style="text-align: center;">.</p> <p style="text-align: center;">.</p> <p>20. プラスチック規制強化 (9/10)</p> <ul style="list-style-type: none"> • 未処理のプラスチック排出への厳しい規制は、責任の強化と環境保護に効果的。
--	--

3.2 提案手法

本研究では, ChatGPT を利用し, Tree of Thoughts を用いることで社会問題に対する多段推論を行い, その問題に対する対策方法や解決策を得る.

ChatGPT は ChatGPT3.5 を利用する. 以下に提案手法の手順を記載する.

- (1) ニーズに対する課題を提案させる.
- (2) 課題に対するアプローチを提案させる.
- (3) 提案されたアプローチの中から画期的なものを ChatGPT と人手で抽出する.
- (4) 抽出されたアプローチを ChatGPT を用いて点数化する.

3.1.1 プロンプト

Tree of Thoughts は、問題を階層的な構造で表現する手法である。これにより、複数の観点やアプローチを包括的に捉えることができる。表3の提案手法で用いるプロンプトでは、(1)から(3)の段階的なプロンプトを通じて、問題のニーズから課題、そしてアプローチへと段階的に推論を進める。このような階層的な構造により、多角的な視点からのアプローチを促進し、問題解決の網羅性を高めることが期待される。提案手法のプロンプトには、特定の出力やアプローチに対する追加の情報や詳細な分析を求める部分が含まれている。例えば、(2)では特定の課題に対するアプローチを求める際に、前段階の出力を参照することで、そのアプローチの正確性や妥当性を詳細に検討することができる。これにより、問題をより深く理解し、より効果的なアプローチを見つけるための手がかりを得ることが期待される。また、(4)、(5)の ChatGPT による評価で“画期的さ”を定めている。

表3：提案手法のプロンプト

(1)	以下の問題に対するニーズを教えてください (設問)
(2)	以下のニーズに対する課題を教えてください {(1)の出力の X 個目}
(3)	以下の課題に対するアプローチを教えてください {(2)の出力の X 個目}
(4)	(設問)において、以下の中から今までにない画期的なアプローチだと思われるものを説明と共に 20 個選んでください {(3)の出力}
(5)	(設問)において、以下のアプローチの画期的さを評価し、10 点満点で評価してください

3.1.2 入出力例

提案手法において表4に表3の設問を「海洋プラスチック問題」としたときの入出力例を示す。

表4：提案手法の入出力例

<p>User: :</p> <p>以下の問題に対するニーズを教えてください</p> <p>海洋プラスチック問題</p> <p>ChatGPT :</p> <p>海洋プラスチック問題に対処するためのニーズは多岐にわたります。これは、海洋環境におけるプラスチック汚染が増加しており、生態系、人間の健康、経済に深刻な影響を及ぼしていることから、緊急性が高い課題です。以下は、この問題に対処するための主要なニーズです：</p> <ol style="list-style-type: none">1. プラスチックごみの削減: プラスチック製品の使用を削減し、環境に配慮した代替材料の普及を促進するための取り組みが必要です。このためには、政府規制、産業界の協力、消費者意識の向上が不可欠です。(1) 2. プラスチックの適切な廃棄物管理: プラスチック廃棄物の適切な管理と処理システムの整備が必要です。これには、リサイクル施設の強化、廃棄物処理施設の整備、プラスチックごみの違法投棄の取り締まりが含まれます。3. プラスチックごみの除去: 海洋中のプラスチックごみを効果的に除去するための技術とプロジェクトへの投資が必要です。浮遊するプラスチックを回収し、浜辺や海底からのプラスチックのクリーンアップ活動を行うための資金と技術が必要です。4. 国際協力: 海洋プラスチック問題は国境を越えた問題であるため、国際的な協力と規制が重要です。プラスチックの流通や廃棄物管理に関する国際的な規制を整備し、国際的な協力体制を構築する必要があります。5. 環境教育と意識向上: 人々に対して、プラスチック汚染が環境や生態系に与える影響を教育し、環境に対する責任感を高めるためのプログラムが必要です。
--

	<p>6. プラスチックのリサイクル技術の改善: プラスチックのリサイクル技術を改善し、より効率的かつ持続可能な方法でプラスチック廃棄物を再利用するための研究と開発が必要です。</p> <p>7. 緊急資金の提供: 海洋プラスチック問題に対処するための緊急の資金が必要です。これによってクリーンアッププロジェクト、環境保護団体、研究機関が支援されます。</p> <p>これらのニーズへの取り組みが進むことで、海洋プラスチック問題の緩和と将来のプラスチック汚染の防止に向けた進展が期待されます。</p>
(2)	<p>User :</p> <p>以下のニーズに対する課題を教えてください</p> <p>1. プラスチックごみの削減: プラスチック製品の使用を削減し、環境に配慮した代替材料の普及を促進するための取り組みが必要です。このためには、政府規制、産業界の協力、消費者意識の向上が不可欠です。</p> <hr/> <p>ChatGPT :</p> <p>プラスチックごみの削減に関する課題は多岐にわたり、以下にいくつかのポイントを挙げてみましょう。</p> <p>1. 政府規制の強化と実施:</p> <ul style="list-style-type: none"> • 政府はプラスチックごみ削減のための規制を強化し、プラスチック使用量を減少させる法律や政策を制定する必要があります。例えば、プラスチック製品の制限、リサイクル割合の向上、リサイクル料金の導入などが考えられます。 <p>2. 産業界の協力:</p> <ul style="list-style-type: none"> • 産業界は、環境に配慮した代替材料の研究開発や導入に取り組む必要があります。また、プラスチックの生産から廃棄物処理までのサプライチェーン全体で持続可能なプラクティスを採用し、循環型経済を促進する役割を果たすべきです。 <p>3. 消費者意識の向上:</p> <ul style="list-style-type: none"> • 消費者に対して、プラスチックごみの削減やリサイクルの重要性を啓発する情報キャンペーンが必要です。持続可能な選択肢へのアクセスを向上させ、消費者がプラスチックを避けたり、リサイクルを積極的に支持するための教育プログラムも重要です。 <p>4. 代替材料の研究と普及:</p>

	<ul style="list-style-type: none"> • 環境に優しい代替材料の研究と開発を支援し、これらの材料を市場に普及させるための支援が必要です。代替材料のコスト効率性や性能向上に対する投資が必要です。 <p>5. プラスチック廃棄物の適切な処理:</p> <ul style="list-style-type: none"> • プラスチック廃棄物の適切な処理とリサイクルインフラの整備が重要です。リサイクル施設の拡充、プラスチックごみの適切な分別とコレクションの促進、プラスチックごみの海洋投棄の防止が必要です。 <p>これらの課題に対処するためには、政府、産業界、消費者が協力し、持続可能な未来に向けた具体的な行動を共に取ることが不可欠です。</p>
(3)	<p>User :</p> <p>以下の課題に対するアプローチを教えてください</p> <p>1. 政府規制の強化と実施: 政府はプラスチックごみ削減のための規制を強化し、プラスチック使用量を減少させる法律や政策を制定する必要があります。例えば、プラスチック製品の制限、リサイクル割合の向上、リサイクル料金の導入などが考えられます。</p> <hr/> <p>ChatGPT :</p> <p>政府規制の強化と実施に関するアプローチは、プラスチックごみ削減のための重要なステップです。以下は、政府がこの目標を達成するために取るべき主要なアプローチです。</p> <p>1. プラスチック製品の制限:</p> <ul style="list-style-type: none"> • 一回限りの使い捨てプラスチック製品の制限を導入します。これにはプラスチックストロー、プラスチックバッグ、プラスチックカトラリーなどが含まれます。 • 従来のプラスチック製品に代わる環境にやさしい代替品を促進し、普及させます。 <p>2. リサイクルの促進と割合の向上:</p> <ul style="list-style-type: none"> • プラスチック製品のリサイクル率を向上させるための施策を導入します。これには、リサイクル施設の拡充、リサイクルプログラムの強化、市民への啓発キャンペーンなどが含まれます。 • プラスチックの新たなリサイクル技術の研究と開発を奨励し、循環型経済を構築します。 <p>3. リサイクル料金の導入:</p>

	<ul style="list-style-type: none"> • プラスチック製品のリサイクル料金制度を導入し、プラスチック製品の使用者に責任を持たせます。これにより、リサイクルの促進と廃棄物の減少を実現します。 • リサイクル料金から得られた収益をリサイクルインフラの整備や環境保護プロジェクトに再投資することを検討します。 <p>4. プラスチック削減目標の設定:</p> <ul style="list-style-type: none"> • 政府はプラスチックごみ削減の具体的な目標を設定し、定期的に進捗を監視・評価します。 • 達成可能なターゲットを設け、関連する産業や事業者に対して協力を要請します。 <p>5. 国際協力と規制の強化:</p> <ul style="list-style-type: none"> • 国際的な協力を通じて、プラスチックごみの輸出・輸入を制御し、海洋ごみ問題への対処を強化します。 • 国内外のプラスチック関連規制を調整し、国際的なベストプラクティスに合致させます。 <p>6. 啓発キャンペーンと市民教育:</p> <ul style="list-style-type: none"> • 政府は市民に対して、プラスチックごみの問題と持続可能な消費に関する教育と啓発キャンペーンを実施します。 • プラスチック削減への市民の積極的な参加を奨励します。 <p>7. 監視と執行:</p> <ul style="list-style-type: none"> • 規制を監視し、違反者に対する厳格な罰則を実施します。違反行為を取り締まるための効果的な監視体制を構築します。 <p>これらのアプローチを組み合わせることで、政府はプラスチックごみの削減と環境への影響の軽減に向けた重要なステップを踏むことができます。同時に、産業や市民の協力を得ながら、より持続可能な未来を築く手助けをすることができます。</p>
(4)	<p>User :</p> <p>海洋プラスチック問題において、以下の中から今までにない画期的なアプローチだと思われるものを説明と共に 20 個選んでください プラスチック製品の制限 リサイクルの促進と割合の向上 リサイクル料金の導入 プラスチック削減目標の設定 国際協力と規制の強化 啓発キャンペーンと市民教育 監視と執行 代替材料の研究と開発 サプライチェーンの持続可能性 再利用とリサイクルの促進 持続可能な認証と規制遵守 消費者教育と共感 産業連携と共同研究 情報キャンペーンの実施 学校教育プログラム</p>

	<p style="text-align: center;">・ ・ ・</p> <p>国際的な会議とイニシアティブ 国際的な基金設立 資金供与の促進 科学的根拠とデータ共有 教育と啓発 フェアなリーダーシップ</p> <hr/> <p>ChatGPT :</p> <p>以下は、海洋プラスチック問題において今までにない画期的なアプローチと思われる 20 の提案です。</p> <ol style="list-style-type: none"> 1. プラスチックの種類に応じた選択 <ul style="list-style-type: none"> • プラスチックの種類ごとに最適な代替品やリサイクル技術を開発。それにより、特定のプラスチック種の使用や処理を最適化し、効果的な解決策を提供。 2. 科学的根拠に基づくアプローチ <ul style="list-style-type: none"> • プラスチックの環境への影響について科学的な根拠に基づき、それに応じた対策を講じる。環境への悪影響の正確な理解が、より効果的な対策を生む。 3. 持続可能な代替品の奨励 <ul style="list-style-type: none"> • 持続可能かつ生分解性の高い代替品の研究と奨励。プラスチックの使用を減らすために、代替品の市場普及を促進。 <p style="text-align: center;">・ ・ ・</p> <ol style="list-style-type: none"> 20. プロジェクトの計画と組織 <ul style="list-style-type: none"> • 継続的かつ綿密なプロジェクト計画と組織の構築。目標を達成するための明確な計画と効果的な組織体制が、画期的なアプローチの成功に不可欠。
(5)	<p>User :</p> <p>海洋プラスチック問題において、以下のアプローチの画期的さを評価し、10点満点で評価してください</p> <p>1. 持続可能な代替材料の研究と開発: • 新たな環境にやさしい材料の研究と開発を奨励し、生分解性プラスチック、植物由来のバイオプラスチック、再生可能な素材などの代替材料を普及させることが、プラスチック使用を大幅に削減する可能性があります。これにより、プラスチックの生産から廃棄物までの影響を劇的に減少させることができます。</p>

2. サプライチェーンの持続可能性: • サプライチェーン全体で持続可能なプラクティスを採用し、省資源型・低炭素経済を実現します。これにより、原材料調達から製造、包装、流通、廃棄物処理までのプロセスで環境への影響を最小限に抑えることができます。

3. 再利用とリサイクルの促進: • 製品設計段階から、製品の再利用やリサイクルを容易にする工夫を行うことで、廃棄物削減を実現できます。また、産業界が自社でリサイクルプログラムを実施し、リサイクル製品の供給を増やす取り組みを強化すべきです。

•
•
•

20. プロジェクトの計画と組織 • 継続的かつ綿密なプロジェクト計画と組織の構築。目標を達成するための明確な計画と効果的な組織体制が、画期的なアプローチの成功に不可欠。

ChatGPT :

以下は各アプローチの画期的さに対する評価です。

1. プラスチックの種類に応じた選択 (9/10)

• プラスチックの種類ごとに対策を講じるアプローチは環境への影響を最小限に抑え、解決策を特定する上で画期的。

2. 科学的根拠に基づくアプローチ (9/10)

• 科学的根拠に基づいたアプローチは、問題を正確に理解し、有効な対策を講じる上で重要。

3. 持続可能な代替品の奨励 (8/10)

• 持続可能な代替品の奨励はプラスチックの使用削減に寄与しますが、市場普及には時間がかかる可能性がある。

•
•
•

20. プロジェクトの計画と組織 (9/10)

• 綿密なプロジェクト計画と組織 • 継続的かつ綿密なプロジェクト計画と組織の構築。目標を達成するための明確な計画と効果的な組織体制が、画期的なアプローチの成功に不可欠。

3.3 ChatGPT

ChatGPT[4]は, OpenAI が開発した自然言語処理モデルの一種である. 具体的には, GPT (Generative Pre-trained Transformer) アーキテクチャを基にしたモデルで, 大規模かつ多様なデータセットで事前に学習された後, 特定のタスクに対して fine-tuning が行われることがある.

GPT モデルはトランスフォーマー (Transformer) と呼ばれるニューラルネットワークアーキテクチャを採用している. このアーキテクチャは, 文脈を理解し, それを元に文章を生成するのに優れており, 自然言語処理のタスクにおいて高い性能を発揮する.

ChatGPT は, 対話型のアプリケーションに適した形で調整された GPT モデルである. ユーザーとの対話に応じて柔軟にテキストを生成することができ, 質問への回答や文章の生成など, 対話的なコンテキストに基づいたタスクに利用される. ChatGPT は, 大量のオンラインのテキストデータを学習しており, そのため一般的な言語理解や生成において高い精度を持っている.

第4章 実験

本章では、本研究が行った実験の入力として用いたデータを4.1節で説明し、本研究が行った実験方法を4.2節で説明する。

4.1 実験データ

本研究では、ベースライン及び提案手法の入力として10個の設問を考え、用いる。表5に実験で用いる設問を示す。

表5：実験データ

設問1	いじめ問題
設問2	フードロス問題
設問3	介護離職問題
設問4	海洋プラスチック問題
設問5	限界集落問題
設問6	少子高齢化問題
設問7	人身売買問題
設問8	男女格差問題
設問9	地球温暖化問題
設問10	労働問題

4.2 実験方法

本研究では、設問1から設問10の10個の社会問題に対して、Tree of Thoughts と Input-Output の2つの異なる手法を用いて実験を行う。

4.3 評価方法

本研究では ChatGPT と人手での評価を比較し、一致率を確かめるために F 値を使用する。具体的には, ChatGPT が 9 点以上と評価した回答と, 人手で"画期的"と感じられた回答が一致した場合を正解として扱う。表 6 は混同行列 (Confusion Matrix) に基づく評価指標の定義[5]である

表 6 : 混同行列

		予測	
		Positive(正と予測)	Negative(負と予測)
実際	Positive(正)	TP(True Positive)	FN(False Positive)
	Negative(負)	FP(False Positive)	TN(True Negative)

- TP (True Positive, 真陽性) : 予測値を正として, その予測が正しい場合.
- TN (True Negative, 真陰性) : 予測値を負として, その予測が正しい場合.
- FP (False Positive, 偽陽性) : 予測値を正として, その予測が誤りの場合.
- FN (False Negative, 偽陰性) : 予測値を負として, その予測が誤りの場合.

これらの指標を用いて, 適合率 (Precision), 再現率 (Recall), および F 値 (F-measure) を以下の式(1), (2), (3)で計算する.

$$\text{適合率(Precision)} = \frac{TP}{TP+FP} \quad (1)$$

$$\text{再現率(Recall)} = \frac{TN}{FP+TN} \quad (2)$$

$$\text{F 値(F - measure)} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (3)$$

4.4 実験結果

本節では、ベースライン(Input-Output)と提案手法(Tree of Thoughts)を用いたときの各設問に対しての適合率, 再現率, F 値の結果を示す.

4.4.1 Input-Output を用いた実験結果

設問 1 から設問 10 の 10 個の社会問題に対して, Input-Output を用いたときの正解率を F 値の平均値として, 表 7 から表 10 に示す.

表 7 : 被験者 1 の正解率(Input-Output)

	適合率	再現率	F 値
いじめ問題	0.67(2/3)	0.20(2/10)	0.31
フードロス問題	0.67(2/3)	0.22(2/9)	0.33
介護離職問題	0.33(1/3)	0.17(1/6)	0.22
海洋プラスチック問題	1.00(4/4)	0.40(4/10)	0.57
限界集落問題	0.00(0/1)	0.00(0/5)	0.00
少子高齢化問題	0.00(0/2)	0.00(0/8)	0.00
人身売買問題	0.25(1/4)	0.08(1/13)	0.11
男女格差問題	1.00(1/1)	0.11(1/9)	0.20
地球温暖化問題	0.00(0/1)	0.00(0/5)	0.00
労働問題	0.33(1/3)	0.11(1/9)	0.17
平均値			0.19

表 8 : 被験者 2 の正解率(Input-Output)

	適合率	再現率	F 値
いじめ問題	1.00(2/2)	0.17(2/12)	0.29
フードロス問題	1.00(3/3)	0.30(3/10)	0.46
介護離職問題	1.00(1/1)	0.20(1/5)	0.33
海洋プラスチック問題	0.00(0/2)	0.00(0/6)	0.00
限界集落問題	1.00(1/1)	0.11(1/9)	0.20
少子高齢化問題	—	0.00(0/7)	0.00
人身売買問題	0.33(1/3)	0.11(1/9)	0.17
男女格差問題	0.50(1/2)	0.13(1/8)	0.20
地球温暖化問題	0.50(1/2)	0.13(1/8)	0.20
労働問題	0.00(0/2)	0.00(0/6)	0.00
平均値			0.18

表 9 : 被験者 3 の正解率(Input-Output)

	適合率	再現率	F 値
いじめ問題	1.00(5/5)	0.50(5/10)	0.67
フードロス問題	1.00(3/3)	0.33(3/9)	0.50
介護離職問題	0.67(2/3)	0.33(2/6)	0.44
海洋プラスチック問題	0.25(1/4)	0.10(1/10)	0.14
限界集落問題	0.33(1/3)	0.20(1/5)	0.25
少子高齢化問題	—	0.00(0/8)	0.00
人身売買問題	0.50(1/2)	0.08(1/13)	0.13
男女格差問題	0.00(0/1)	0.00(0/9)	0.00
地球温暖化問題	0.33(1/3)	0.20(1/5)	0.25
労働問題	1.00(2/2)	0.22(2/9)	0.36
平均値			0.28

表 1 0 : 被験者 4 の正解率(Input-Output)

	適合率	再現率	F 値
いじめ問題	0.50(1/2)	0.10(1/10)	0.17
フードロス問題	0.10(2/2)	0.22(2/9)	0.36
介護離職問題	0.00(0/2)	0.00(0/6)	0.00
海洋プラスチック問題	0.50(1/2)	0.10(1/10)	0.17
限界集落問題	0.50(1/2)	0.20(1/5)	0.29
少子高齢化問題	0.50(1/2)	0.13(1/8)	0.30
人身売買問題	0.50(1/2)	0.08(1/13)	0.13
男女格差問題	0.50(1/2)	0.11(1/9)	0.18
地球温暖化問題	0.00(0/5)	0.00(0/5)	0.00
労働問題	0.33(1/3)	0.11(1/9)	0.17
平均値			0.17

4.4.2 Tree of Thoughts を用いた実験結果

設問 1 から設問 10 の 10 個の社会問題に対して, Tree of Thoughts を用いたときの正解率を F 値の平均値として, 表 1 1 から表 1 4 に示す.

表 1 1 : 被験者 1 の正解率(Tree of Thoughts)

	適合率	再現率	F 値
いじめ問題	0.00(0/1)	0.00(0/12)	0.00
フードロス問題	1.00(1/1)	0.10(1/10)	0.18
介護離職問題	0.67(2/3)	0.40(2/5)	0.50
海洋プラスチック問題	0.00(0/2)	0.00(0/6)	0.00
限界集落問題	0.00(0/1)	0.00(0/9)	0.00
少子高齢化問題	0.00(0/2)	0.00(0/7)	0.00
人身売買問題	0.00(0/5)	0.00(0/9)	0.00
男女格差問題	0.50(1/2)	0.13(1/8)	0.20
地球温暖化問題	1.00(1/1)	0.13(1/8)	0.22
労働問題	1.00(2/2)	0.33(2/6)	0.50
平均値			0.16

表 1 2 : 被験者 2 の正解率(Tree of Thoughts)

	適合率	再現率	F 値
いじめ問題	1.00(2/2)	0.17(2/12)	0.29
フードロス問題	1.00(3/3)	0.30(3/10)	0.46
介護離職問題	1.00(1/1)	0.20(1/5)	0.33
海洋プラスチック問題	0.00(0/2)	0.00(0/6)	0.00
限界集落問題	1.00(1/1)	0.11(1/9)	0.20
少子高齢化問題	—	0.00(0/7)	0.00
人身売買問題	0.33(1/3)	0.11(1/9)	0.17
男女格差問題	0.50(1/2)	0.13(1/8)	0.20
地球温暖化問題	0.50(1/2)	0.13(1/8)	0.20
労働問題	0.00(0/2)	0.00(0/6)	0.00
平均値			0.18

表 1 3 : 被験者 3 の正解率(Tree of Thoughts)

	適合率	再現率	F 値
いじめ問題	0.67(2/3)	0.17(2/12)	0.27
フードロス問題	0.60(3/5)	0.30(3/10)	0.40
介護離職問題	—	0.00(0/5)	0.00
海洋プラスチック問題	0.00(0/3)	0.00(0/6)	0.00
限界集落問題	0.33(1/3)	0.11(1/9)	0.17
少子高齢化問題	—	0.00(0/7)	0.00
人身売買問題	0.20(1/5)	0.11(1/9)	0.14
男女格差問題	0.50(2/4)	0.25(2/8)	0.33
地球温暖化問題	—	0.00(0/8)	0.00
労働問題	0.00(0/3)	0.00(0/6)	0.00
平均値			0.13

表 1 4 : 被験者 4 の正解率(Tree of Thoughts)

	適合率	再現率	F 値
いじめ問題	1.00(2/2)	0.17(2/12)	0.29
フードロス問題	0.50(1/2)	0.10(1/10)	0.17
介護離職問題	0.00(0/1)	0.00(0/5)	0.00
海洋プラスチック問題	0.00(0/2)	0.00(0/6)	0.00
限界集落問題	0.60(3/5)	0.33(3/9)	0.43
少子高齢化問題	0.00(0/3)	0.00(0/7)	0.00
人身売買問題	0.00(0/1)	0.00(0/9)	0.00
男女格差問題	0.00(0/1)	0.00(0/8)	0.00
地球温暖化問題	0.50(1/2)	0.13(1/8)	0.20
労働問題	0.00(0/3)	0.00(0/6)	0.00
平均値			0.11

4.4.3 性能比較

被験者 1 から被験者 4 までの提案手法とベースラインの F 値とその平均値を表 1 5 に示す.

表 1 5 : 提案手法とベースラインの比較

F 値	被験者 1	被験者 2	被験者 3	被験者 4	平均
Tree of Thoughts	0.16	0.18	0.13	0.11	0.15
Input-Output	0.19	0.18	0.28	0.17	0.21

第5章 考察

本章では、得られた実験結果から Input-Output と Tree of Thoughts の有効性について考察を記述する。

被験者実験において、提案手法よりもベースラインの方が F 値が高い結果が得られたことから、提案手法がベースラインよりも性能が低い可能性が示唆される。また、この結果から、Tree of Thoughts よりも Input-Output の方が性能が良いという結果が得られた。この結果に基づく考察をより具体的に行うと、以下のような理由が考えられる。

- ・推論の階層構造の有効性

ベースラインの Input-Output が提案手法よりも性能が高かったことから、推論の階層構造よりも入力と出力の直接的な対応が有効である可能性が考えられる。Input-Output では、入力と出力の対応が直接的であり、課題や解決策の関連性が明確になる。一方、Tree of Thoughts では、問題を階層的な構造で表現することで複数の観点やアプローチを捉えることができるが、この階層構造が必ずしも性能向上につながらなかった可能性が考えられる。

- ・問題の特性

使用された問題や課題によっては、Tree of Thoughts のような階層構造が適切でない場合があると考えられる。問題の特性や複雑さによっては、直接的な入力と出力の関連性がより重要であり、その方が解決策の発見や評価に効果的である可能性があると考えられる。

以上の考察から、今後の課題としては提案手法のプロンプトの改善を行うことで、性能向上や有効性の確認が行われると考えられる。

第6章 おわりに

本研究では ChatGPT を用いて多段推論を行い，提案手法の性能を検証を行った．提案手法では Tree of Thoughts を用いたプロンプトを用い，ベースラインでは Input-Output のプロンプトを用いた．評価では提案手法とベースラインの出力の”画期的さ”を比較し，実験は4人の被験者で行なった．被験者1，被験者2，被験者3，被験者4のベースラインのF値はそれぞれ0.19, 0.18, 0.28, 0.17であり，提案手法のF値はそれぞれ0.16, 0.18, 0.13, 0.11であった．結果として，どの被験者においてもベースラインの性能が提案手法を上回る傾向が見られた．これらの結果から，提案手法の改善や調整が必要であると考えられ，今後の研究の方向性や手法の検討が求められる．

謝辞

鳥取大学工学部電気情報系学科自然言語処理研究室の村田真樹教授には、終始に渡り研究の進め方や本論文の書き方など、細部にわたる御指導を頂きました。ここに御礼申し上げます。また、御助言を頂きました村上仁一助教授に心から御礼申し上げます。そして、被験者実験にご協力いただいた自然言語処理研究室の皆様には感謝申し上げます。

参考文献

- [1] 大竹竜太, 村田真樹, 徳久雅人. 大規模テキストデータを用いた社会構造ネットワークモデルの自動抽出. 言語処理学会第 19 回年次大会発表論文集, pp. 798-801, 2013.
- [2] Shunyu Yao, Dian Yu, Jeffrey Zhao, Izhak Shafran, Thomas L. Griffiths, Yuan Cao, Karthik Narasimhan Tree of Thoughts: Deliberate Problem Solving with Large Language Models. arXiv:2305.10601v1 [cs.CL] 17 May 2023.
- [3] 那須 太陽, 師田 賢人, 居戸 和由貴. “【ChatGPT をフル活用】高精度なプロンプト「Tree-of-Thoughts」で AI からより良い回答を引き出そう”. ハッシュタグ. 2024/1/5. https://hashtag-jp.com/technology/chatgpt_tree-of-thoughts/#index_id0
- [4] Hitachi, “ChatGPT とは？ 始め方や賢い活用方法などわかりやすく解説”. 日立ソリューションズ・クリエイト. 2023/9/24. <https://www.hitachi-solutions-create.co.jp/column/technology/chatgpt.html>
- [5] FukuharaYohei. “【入門者向け】機械学習の分類問題評価指標解説(正解率・適合率・再現率など)”. Qiita. 2023/1/21. <https://qiita.com/FukuharaYohei/items/be89a99c53586fa4e2e4>