

GRPO強化学習に基づき構成要素 対応を考慮した特許請求項翻訳

浅見 遥斗, 宇津呂武仁(筑波大), 永田 昌明 (NTT)

- **背景**
- **実験設定**
- **提案手法**
- **実験結果**
- **まとめ**

- **背景**
- **実験設定**
- **提案手法**
- **実験結果**
- **まとめ**

特許文中で特に重要性の高い特許請求項は以下の特徴から翻訳が難しい

1. 極めて長い 1 文から構成される

2. 特殊な文構造を持っている

前記走行経路生成部が生成した前記曲線の情報を取得する走行経路情報取得部と、前記走行経路情報取得部が取得した前記曲線の情報に基づいて、前記車両が前記進行すべき経路に従って走行するように前記車両の操舵制御を行う前記車両制御部と、を備え、前記車両制御部は、前記制御参照点における前記曲線の曲率、前記制御参照点と前記曲線との距離、および前記制御参照点における前記曲線のベクトルと前記車両の速度ベクトルとがなす角度のうち少なくとも1つを用いて前記操舵制御を行うことを特徴とする、請求項5または6に記載の車両制御装置。

訳抜け：入力文にあった内容が翻訳時に抜け落ちる誤訳

例： 原文：…少なくともアルミニウムおよびニッケルと少なくとも…

翻訳文：… at least aluminum and at least…

繰り返し：入力文の一部が翻訳時に不必要に繰り返される

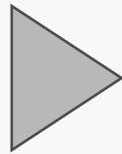
例： 原文：…少なくともアルミニウムおよびニッケルと少なくとも…

翻訳文：… at least aluminum and aluminum and…

LLMの訓練手法

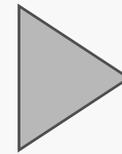
① 事前訓練

大量の自然言語データから、
自然言語のルール
(文法、品詞など)や、
知識を獲得させる



② 指示学習

指示、応答形式データからそれぞれの指示についての応答の形式を学習
→ タスク（翻訳・要約など）に正しく従えるようになる



③ アライメント

人間の好みや指定した指標などに合わせて回答を調整
→ より望ましい・評価の高い出力を選ぶようになる

アライメントによる性能向上の実証

- OpenAI：モデル規模よりも「アライメント」が重要 [Ouyang et al., 2022]
 - 比較対象: 1,750億パラメータのGPT-3（事前訓練のみ） vs 13億パラメータの InstructGPT（+SFT/RLHF）
 - 結果: 1/100のパラメータサイズでも、アライメント済みモデルが指示追従性・安全性で勝利
- Meta：SFTの性能は早期に頭打ちになる傾向 [Touvron et al., 2023]
 - 検証内容: Llama 2 開発における、SFTモデルと+RLHFモデルのスコア推移の比較
 - 結果: SFTによる性能改善は比較的早期に飽和する傾向が確認された一方、SFT後に適用されたRLHF（アライメント）は、人間評価に基づく有用性を継続的に向上させた。

事前訓練

⇒ 特許データを用いた継続事前訓練

⇒ 特許データを用いたSFTの組み合わせ

により、従来のNMT(LLM以前の手法)に対し

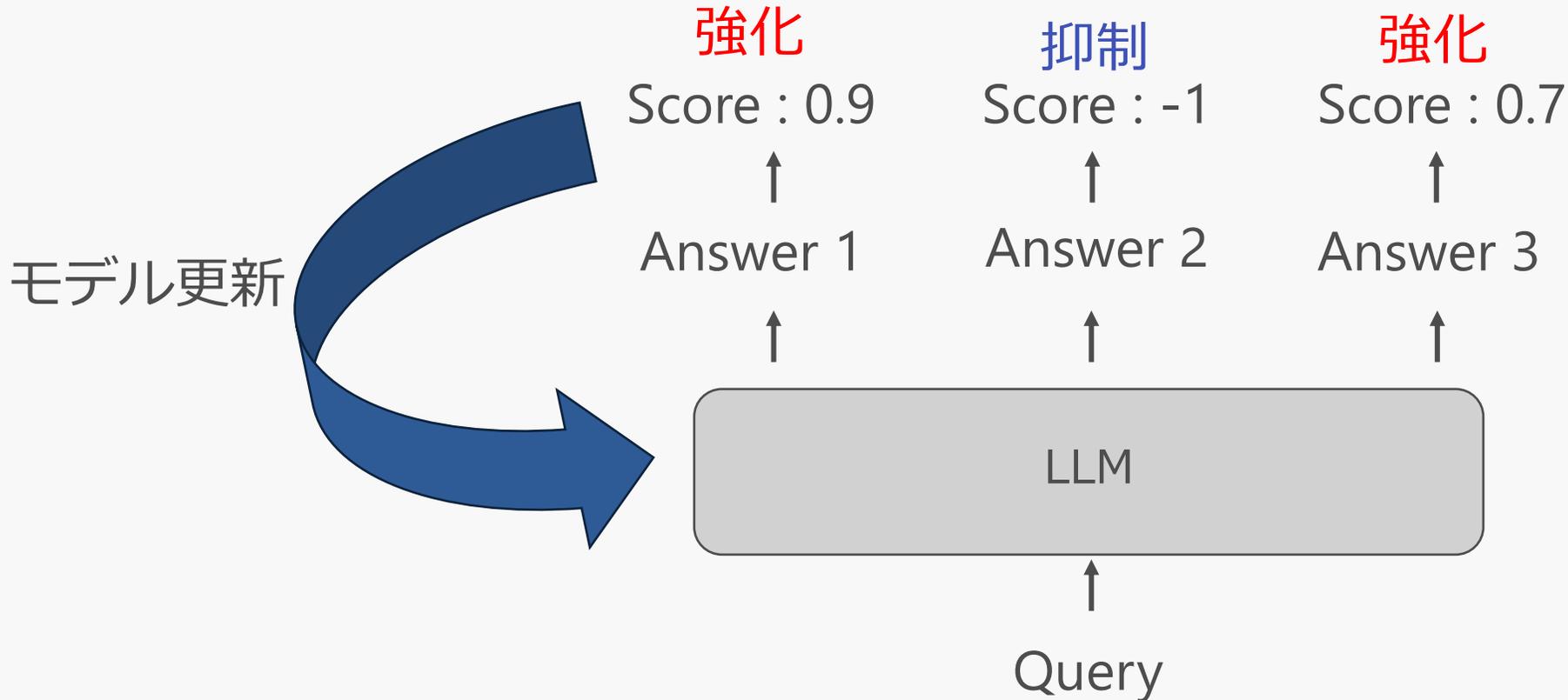
翻訳精度向上を確認[Azami et al. MT Summit2025]

アライメントについてまだ未検討

⇒ 当然やるべき

- DeepSeek社により開発されたアルゴリズム
- GRPOにより訓練されたDeepSeekMath-7B
 - ● ● **数学の問題においてGPT-4に匹敵する推論能力**を獲得
- 安全性向上、LLMエージェント、ロボティクスなど幅広い分野で活用
- **先行研究[Feng+, EMNLP2025]によりSFTをせずGRPOのみで、翻訳性能の向上が確認された(一般分野のみ)**

- 一つの指示ごとに複数候補生成しそれぞれ評価
- Group内のスコア平均との差に基づき，相対的に良い応答を強化、悪い応答を抑制



現状の課題:

- 特許請求項翻訳においてGRPOはどのような効果を発揮するのか
- 特許請求項翻訳でおきる**訳抜け、繰り返し**に対して
効果のある報酬の検討

本研究の立ち位置:

- 「**特許分野外で翻訳精度の検証がされた手法組**」を検証し、
「**特許翻訳において最も有用な訓練手法**」を発見
- GRPOで用いる「**特許請求項に特化した報酬**」を提案

- 背景
- 実験設定
- 提案手法
- 実験結果
- まとめ

使用したデータセット：JaParaPat

- データセットの概要:2016~2020までの日英特許文が対訳形式になったデータ
- 内訳：
 - 継続事前訓練用：約100万文対(2016~2020年)
 - SFT用：3万文対(2020年特許請求項)
 - GRPO用：1万文対(2020年特許請求項)

使用したベースモデル：Qwen/Qwen3-8B

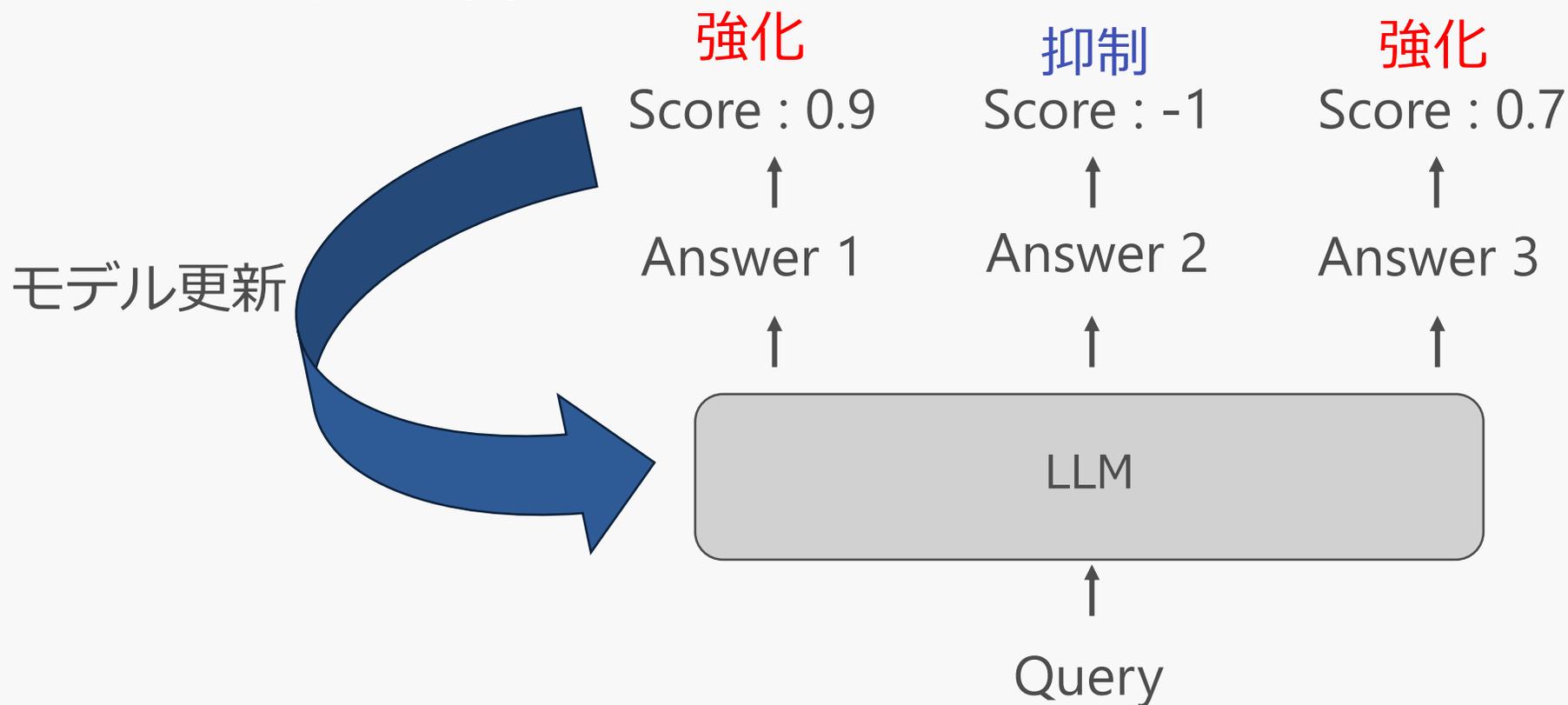
- 多言語対応モデル、翻訳についての指示学習済み

実験：訓練手法の比較

- テストデータ：WAT2025の特許請求項翻訳タスクにおける開発データ(英日、日英)
 - 評価尺度
 - XCOMET：
 - 評価にはUnbabel/XCOMET-XLを使用
 - モデルの翻訳文と参照訳文との間で、意味がどの程度一致しているかで評価
 - MetricX：
 - 人手評価をもとに学習したエラー分析指標
 - エラーが多いほどスコアが高い
- ⇒低いほど精度の高い翻訳

- **背景**
- **実験設定**
- **提案手法**
- **実験結果**
- **まとめ**

- 一つの指示ごとに複数候補生成しそれぞれ評価
- Group内のスコア平均との差に基づき、相対的に良い応答を強化、悪い応答を抑制



- フォーマットスコア(データの損傷がないことを規定):

- Qwenモデルは推論時思考を行うことで最終出力の質を高めている
- 思考部分を<think></think>、翻訳部分を<translate></translate>で囲むことを強制させることでその後の報酬計算の際に扱いやすくする

¥n<think>¥nOkay, let's tackle this translation. The user wants to translate a Japanese patent claim into English. First, I'll read ... Finally, check for any grammatical errors or awkward phrasing to ensure clarity and precision.¥n</think>¥n¥n<translate>A vacuum-assisted force multiplier device comprising: a housing forming at least one pressure chamber inside; ... position by a reduction of the input to the input member during activation of the actuator.</translate>

- フォーマットリワードの与え方

$$R_{\text{format}} = \begin{cases} 1, & \text{if format is correct} \\ -1, & \text{if format is incorrect} \end{cases}$$

- **メトリック(評価尺度)スコア(翻訳精度を報酬としてGRPOで訓練)** :
 - `<translate></translate>`タグで囲まれた部分のみ抽出し、以下のスコアを計算し報酬を与える
 - **BLEU** : 正解翻訳との文字的類似度 ([0,1]の範囲)
 - **Cometkiwi** : 入力文との意味的類似度([0,1]の範囲)

- **先行研究における最終的なスコア計算**

- 以下のように、フォーマットが正しい場合全てを足し合わせる

$$R_i = \begin{cases} 1 + R_{BLEU} + R_{Cometkiwi}, & R_{format} = 1 \\ -3, & R_{format} = -1 \end{cases}$$

・・・ 翻訳精度を報酬とするGRPOに追加

● 課題：自動評価の限界

- 特許請求項の長文性から、一部の構成要素が翻訳されずに抜けても、全体スコアへの影響が小さい場合がある

● 提案手法：構成要素ごとの長さ比をスコア化

- 発想：訳抜けや繰り返しが発生⇒文の長さに変化が生じるはず
- 特許請求項に含まれる「改行」を利用し、文を構成要素単位に分割、原言語文と翻訳文における各要素の長さを比較
- 評価：期待される長さ比からの乖離が大きければ0に近く、小さければ1に近づくようスコアリング

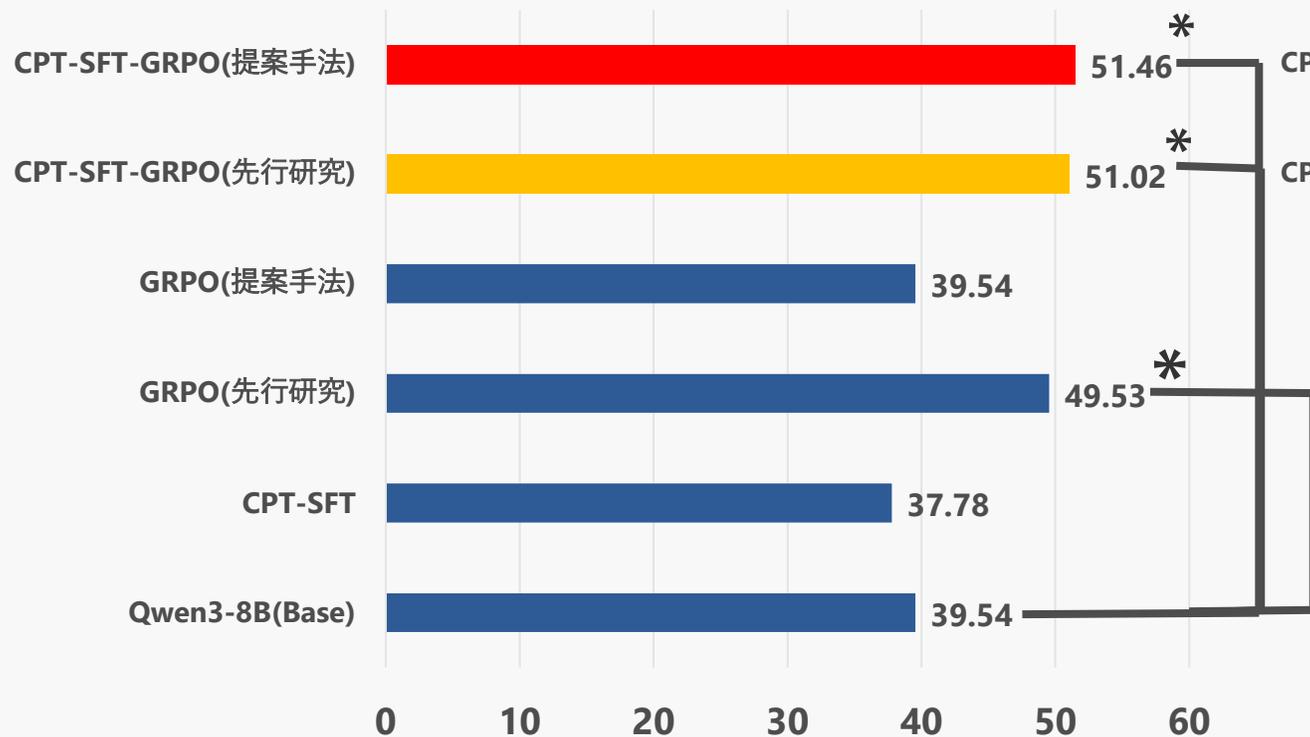
● 最終的なスコア計算(赤字は先行研究との違い)

$$R_i = \begin{cases} 1 + R_{BLEU} + R_{Cometkiwi} + R_{length} & , R_{format} = 1 \\ -4, & R_{format} = -1 \end{cases}$$

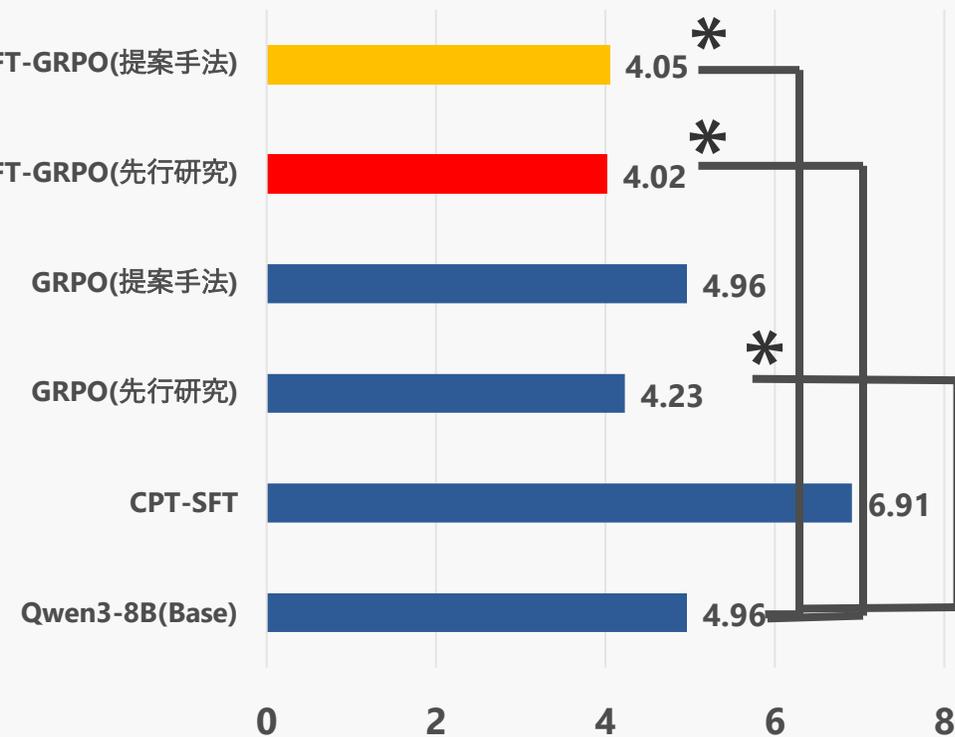
- 背景
- 実験設定
- 提案手法
- **実験結果**
- まとめ

GRPOによりXCOMET, MetricXスコアが有意に向上

XCOMET(↑)



MetricX(↓)

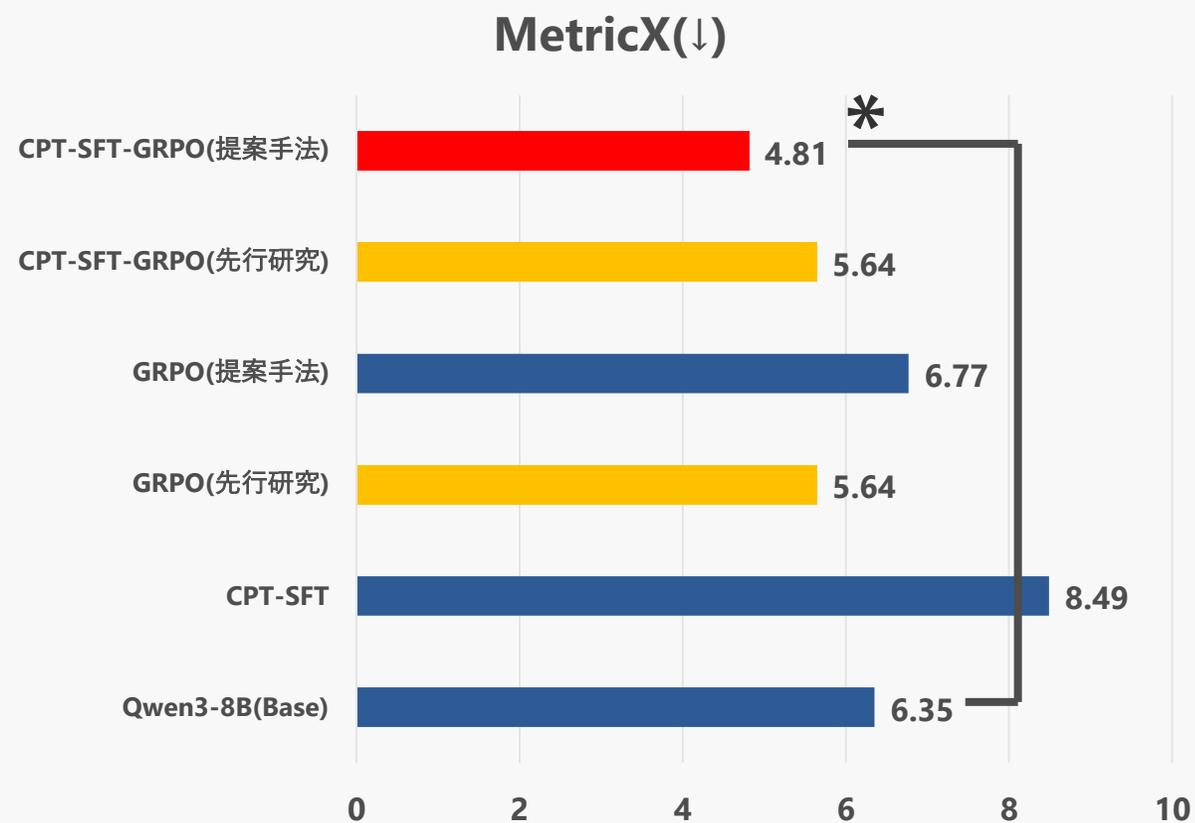
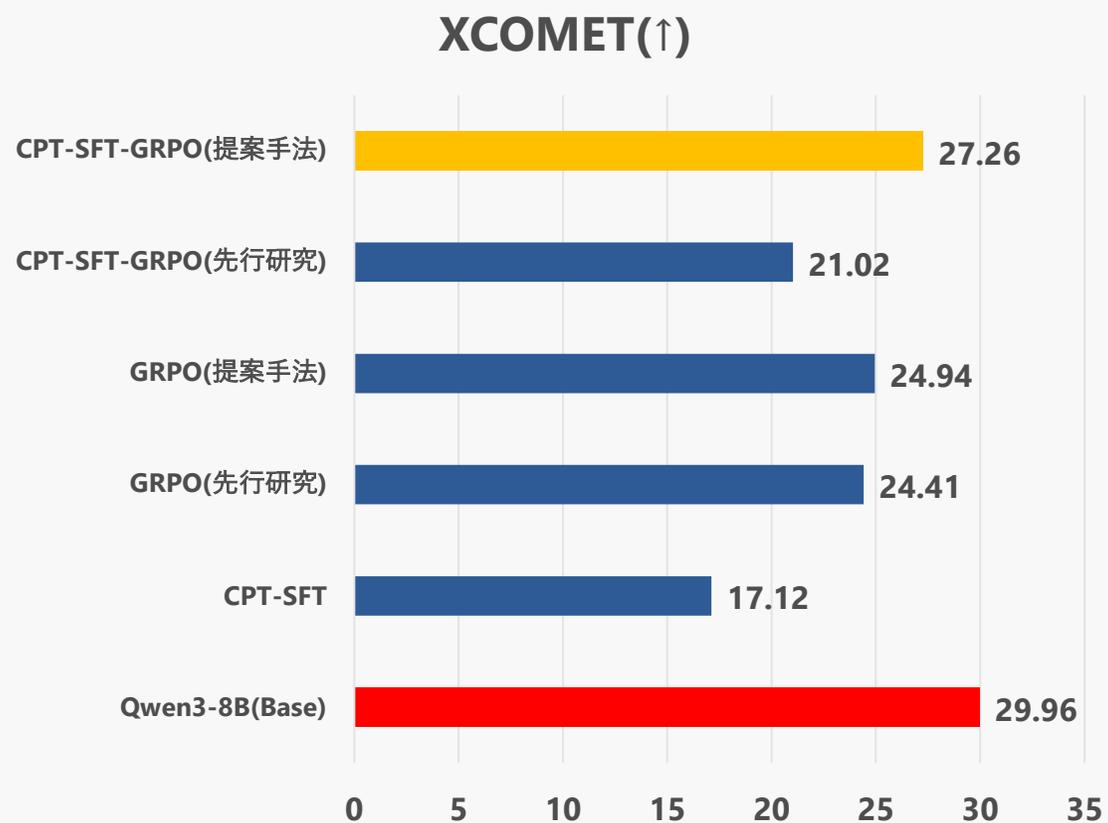


*:p<0.05

訳抜けがCPT-SFT-GRPOで改善

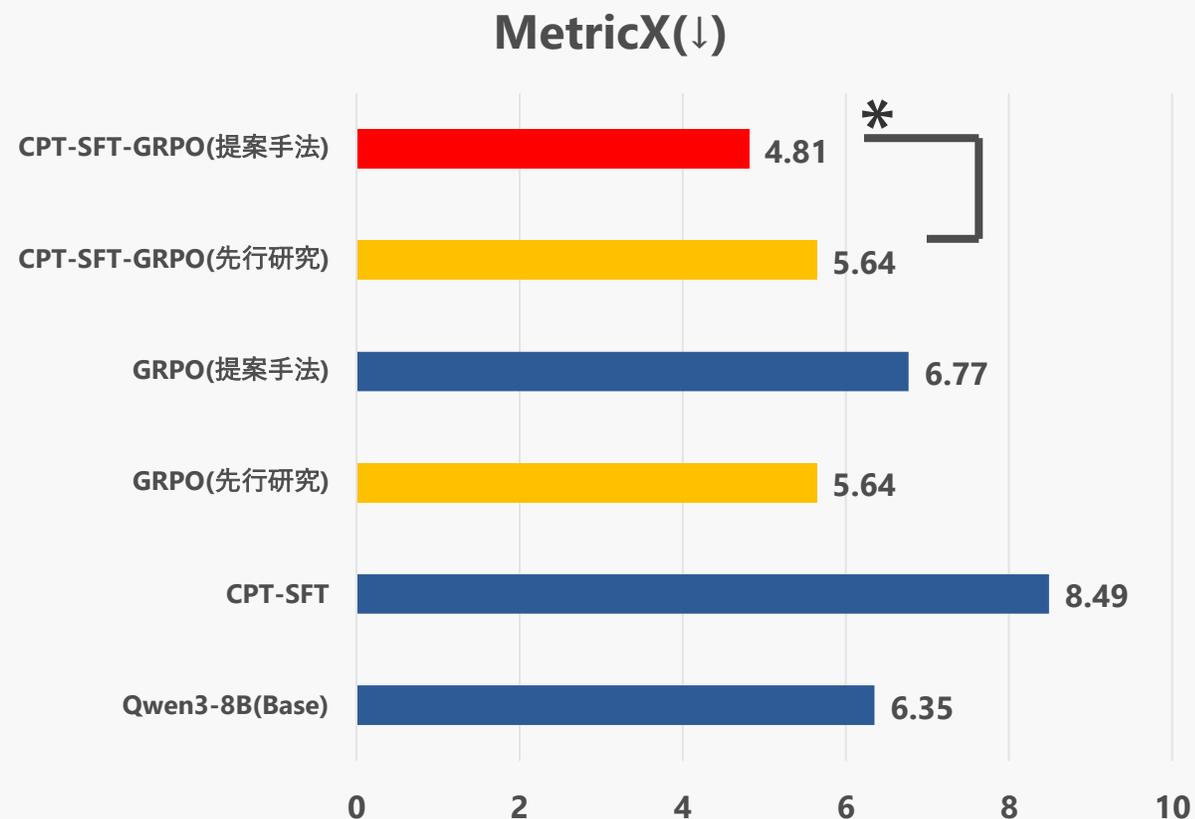
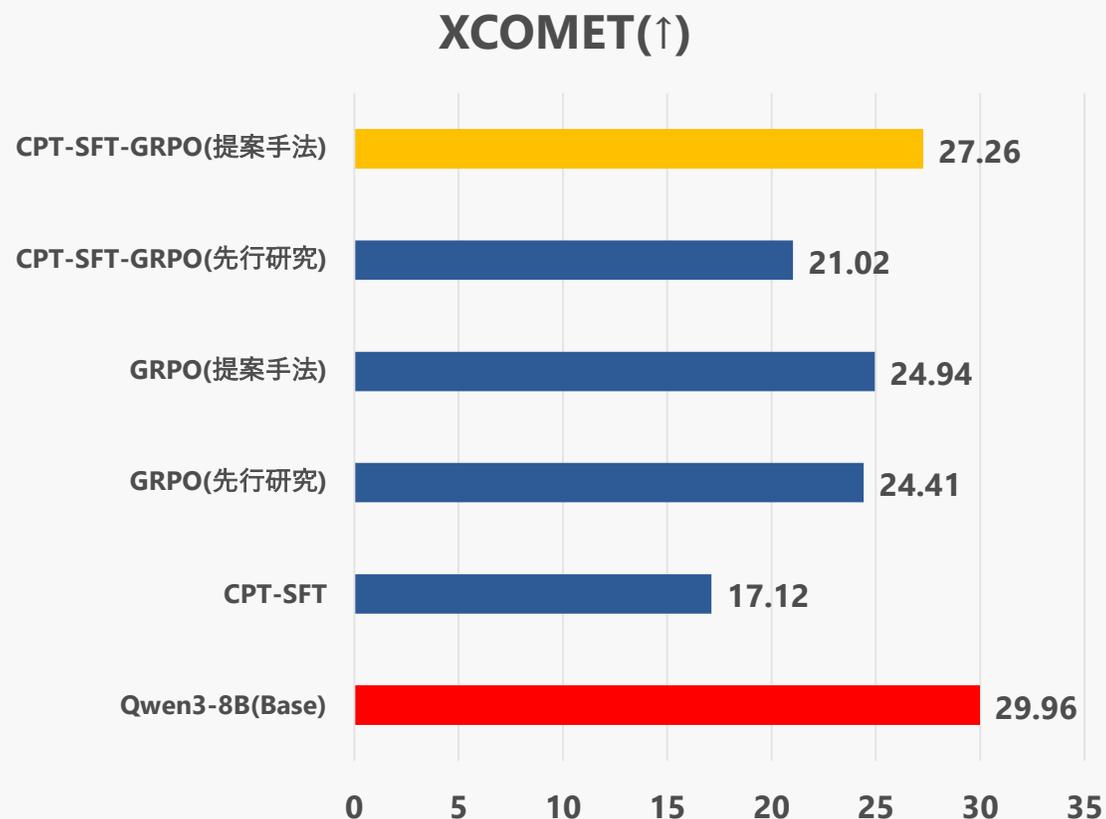
原言語文	参照文	Qwen3-8B	Qwen3-8B-CPT-SFT-GRPO(先行研究)	Qwen3-8B-CPT-SFT-GRPO(提案手法)
<p>ポリエーテルポリオール（a 1）と有機ポリイソシアネート（a 2）を反応させて得られる、... 制電剤（b 3）が炭素数 10～20 の脂肪族系不飽和炭化水素基及び 1 個の... 質量%であることを特徴とする半導電性ウレタンエラストマー形成性組成物。</p>	<p>A semiconductive urethane... antistatic agent (b3) is an ionic salt composed of a quaternary ammonium cation having an aliphatic unsaturated hydrocarbon group having 10 to 20 carbon atoms and one hydroxyalkyl group...</p>	<p>A semiconductive urethane... antistatic agent (b3) is an ionic salt consisting of a quaternary ammonium cation having a hydroxyalkyl group...</p>	<p>A semiconductive urethane... antistatic agent (b3) is an ionic salt consisting of a quaternary ammonium cation having an aliphatic unsaturated hydrocarbon group of 10 to 20 carbon atoms and one hydroxyalkyl group...</p>	<p>A semiconductive urethane... antistatic agent (b3) is an ionic salt composed of a quaternary ammonium cation having an aliphatic unsaturated hydrocarbon group of 10 to 20 carbon atoms and one hydroxyalkyl group...</p>

提案手法によりMetricXスコアが向上



*:p<0.05

提案手法によりMetricXスコアが向上



*:p<0.05

提案手法により翻訳の質が改善

原言語文	参照文	Qwen3-8B	Qwen3-8B-CPT-SFT-GRPO(先行研究)	Qwen3-8B-CPT-SFT-GRPO(提案手法)
<p>A brain machine interface decoding method based on a spiking neural network, wherein ... the trace is the accumulation of the decaying pulse signal, and the formula is:</p>	<p>スパイキングニューラルネットワークに基づくブレインマシンインタフェースデコーディング方法であって、 ... 前記トレースは、 衰退パルス信号の 累積であり、前記 式は：</p>	<p>脳機械インターフェースの解码方法は、スパイキングニューラルネットワークに基づいた液体状態機械モデルを構築することを含み、 ... 、トレースは減衰パルス信号の累積であり、その式は：</p>	<p>スパイキングニューラルネットワークに基づく脳機械インターフェース復号方法であって、... は、前後のニューロンシナプスのトレースの時間微分の時間微分であり、τは、トレースの減衰時間定数であり、 は、前後のニューロンシナプス...(以降繰り返し)</p>	<p>スパイキングニューラルネットワークに基づく脳機械インターフェース復号方法であって、 ... 前記トレースは減衰パルス信号の累積であり、その式は：</p>

- 背景
- 実験設定
- 提案手法
- 実験結果
- まとめ

● まとめ

- 特許請求項翻訳に特化したGRPOに用いる報酬を設計した
- 特許請求項翻訳において、訓練手法と報酬設計の有効な組み合わせを示した

● 今後の展望

- 各翻訳結果の更なる誤訳分析
- 検討中の報酬について実装し精度確認