

Explanation-Guided Prediction of Multi-Dimensional Feedback Scores in Psychological Counseling Dialogues

背景

- 世界では約8人に1人が精神疾患を抱えている [WHO, 2022]
- しかし、カウンセラーは深刻に不足しており、養成が急務
- 現実のカウンセラーはフィードバックを得にくい
 - 相談がうまくいかない → クライアントが再訪しない
 - 相談がうまくいく → 問題が解決し再訪しない
- よって、**自動的・体系的なフィードバック支援の仕組みが必要**
- 先行研究では、評価付きカウンセリング対話 KokoroChat を収集し、スコア予測モデルを構築 [Qi+2025]
- 課題**
 - スコアだけでは、「なぜその評価になったのか」が分からず、改善が難しい
 - 精度自体も十分ではない

提案アプローチ：説明誘導型スコア予測

- 評価理由の説明文を生成した上でスコアを予測する手法
- 特徴・利点**
 - スコアの根拠を明示し、カウンセラーが結果を理解・改善しやすくなる
 - CoT (Chain-of-Thought) 形式の推論により、スコア予測の精度向上が期待できる

KokoroChat

- 訓練を受けたカウンセラーによるロールプレイを通じて収集された**日本語カウンセリング対話データセット**

カウンセリング対話

- 対話数: 6,589, 平均発話数: 91.2
- 日本語母語話者480名構成 (約1/3がプロのカウンセラー、他も6~12か月の養成課程を修了)
- 家族問題、職場の悩み、メンタルヘルス、進路・キャリアなど、多様なトピック

クライアントによるフィードバック

- 公認心理師が監修した20評価項目 (カウンセリングスキルの評価、会話全体の印象)
- 0から5の6段階でクライアント役の話者が評価

ステップ 1

データ構築

- モチベーション**
 - カウンセラーが**評価スコアの根拠を理解**できるように、理由説明文が必要
 - 説明文付きの評価データが既存ではなく、モデル学習のために新たにデータを構築
- KokoroChatデータセットを基盤**に構築
- GPT-5**に対し、カウンセリング対話履歴と20項目のクライアント評価スコアを入力
- 各スコアに対応する**理由説明文**を生成

ステップ 2

説明文付きの点数予測モデルの学習・推論

- 先行研究は、対話履歴から直接スコアを予測
- 本研究では、Step1で構築した説明文付きデータを用い、**CoT形式でLLMを学習**
- 入力
 - 対話履歴
- 出力
 - 「**理由 -> スコア**」の順で、20項目について予測

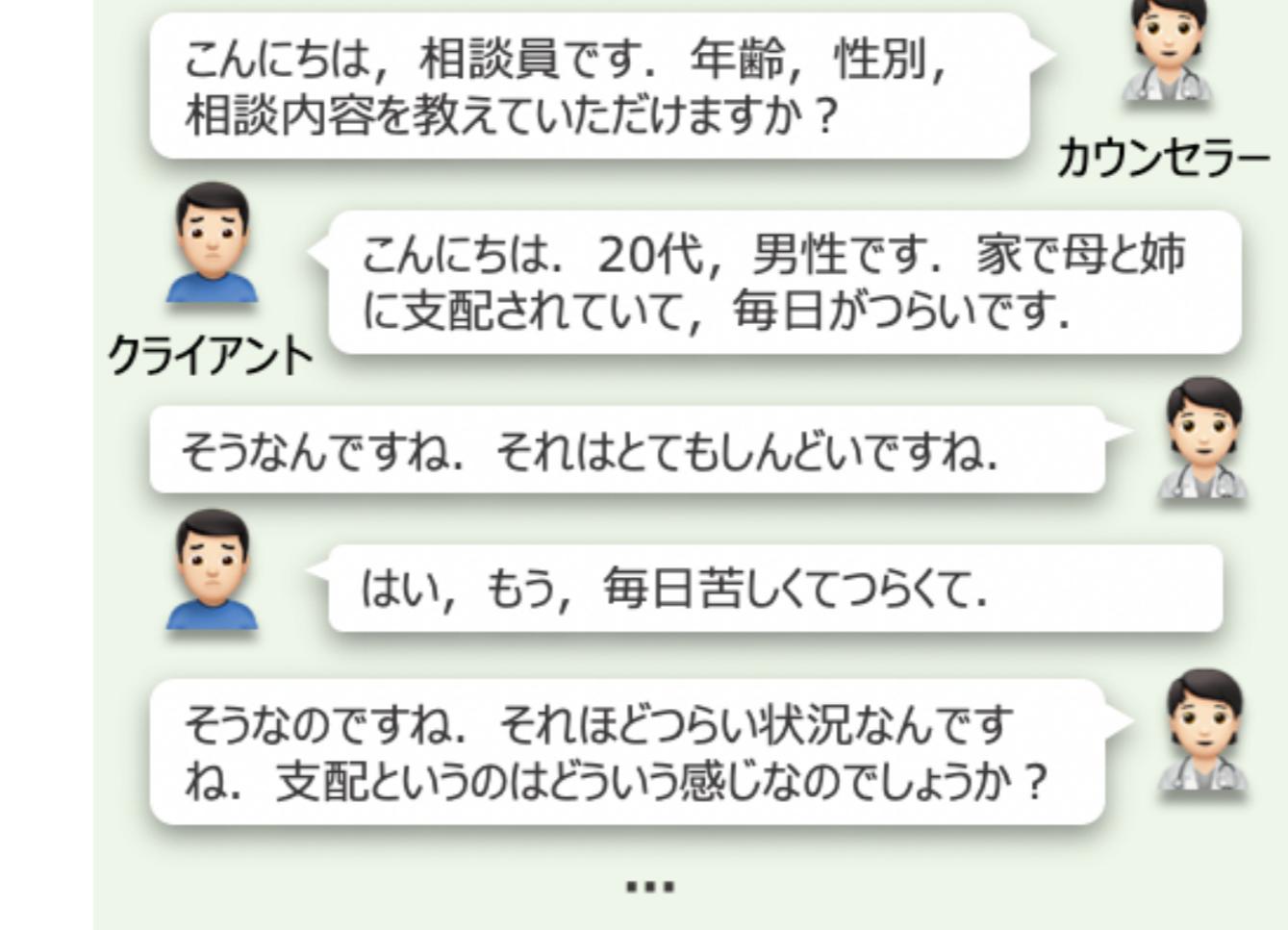
目的

- 理由説明文により、**スコアの根拠を明確化** → カウンセラーが具体的な改善点を理解
- 思考過程を促すことで、**スコア予測の精度向上**を実現

提案手法

Step1: データ構築 (点数説明文の生成)

カウンセリング対話



クライアントによるフィードバック

全体的な印象

- | | |
|-------------------------|-------|
| 1. 聴いてもらえた、わかつてもらえたと感じた | 5 / 5 |
| 2. 尊重されたと感じた | 3 / 5 |
| 3. 新しい気づきや体験があった | 4 / 5 |
| 4. 希望や期待を感じられた | 4 / 5 |

カウンセリングスキル

- | | |
|-------------------|-------|
| 11. 受容・共感 | 4 / 5 |
| 12. 肯定・承認 | 4 / 5 |
| 13. 的確な質問による会話の促進 | 4 / 5 |
| 14. 問題の明確化 | 5 / 5 |

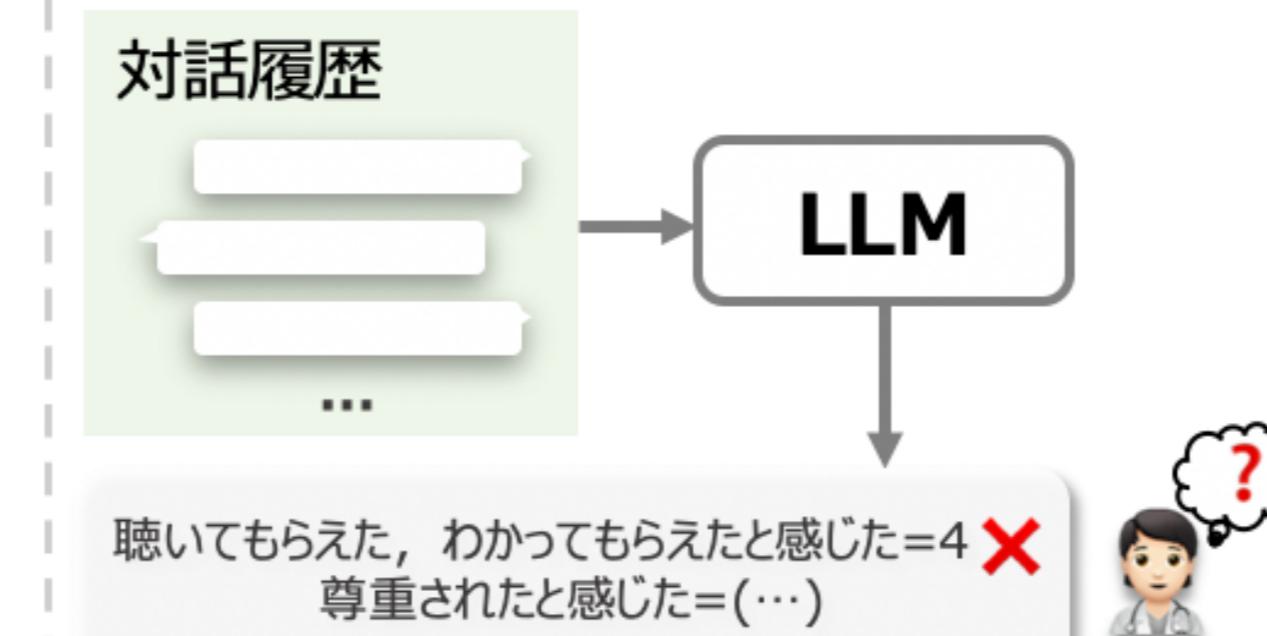
聴いてもらえた、わかつてもらえたと感じた=5

理由: この対話では、クライアントが抱える問題についてカウンセラーが一定の理解を示し、共感を表す場面が見られますが、その表現がやや表面的であるため、クライアントが「聴いてもらえた、わかつてもらえた」と強く感じるには至っていない印象があります。 (...)

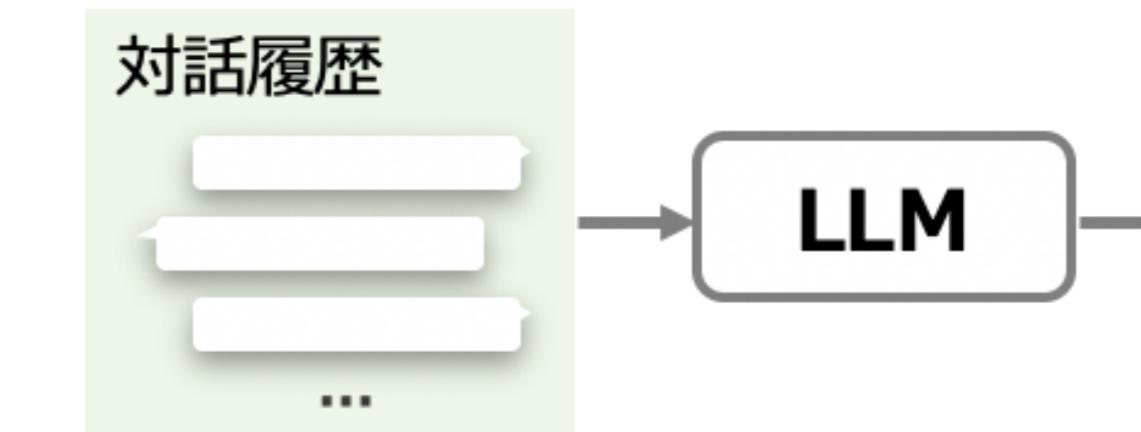
KokoroChat

点数予測

先行研究：直接予測するように学習



Step2: 説明文付きの点数予測モデルの学習・推論



[聴いてもらえた、わかつてもらえたと感じた]
理由: この対話では、クライアントが抱える問題についてカウンセラーが一定の理解を示し、共感を表す場面が見られます (...)

[尊重されたと感じた]
理由: (...)
点数: (...)



実験計画

- 使用モデル: Llama-3.1-Swallow-8B-Instruct-v0.3
- 実験1: 自動評価実験**
 - 目的:** 説明文がスコア予測精度の向上に寄与するかを検証
 - 評価指標:** KokoroChatと同一の指標を使用 [Qi+2025]
 - ACC 正解スコアとの一致率
 - ACC_{soft} ±1点差までを許容する柔軟な一致率
 - MAE 平均絶対誤差
- 実験2: 人間評価実験**
 - 目的:** 説明文がカウンセラーの理解促進・改善支援に役立つかを検証
 - 評価条件**
 - スコアのみ提示
 - スコア + 理由説明文を提示
 - 被験者は2条件で主観的評価を実施

今後の展望

モデルの学習・推論および評価実験の実施