

遺伝的アルゴリズムを用いたエージェントの 価値観最適化手法の構想

Concept of an Agent's Value Optimization Method Using Genetic Algorithms

岩崎 真衣^{1*} 飯田 愛結¹ 大澤 正彦¹
Mai Iwasaki¹, Ayu Iida¹, Masahiko Osawa¹

¹ 日本大学
¹ Nihon University

Abstract: 大規模言語モデルを用いた対話エージェントは、インターネット上の情報を学習した平均的な応答をするため、個人に適応するためには工夫が必要である。適応するためには、新たな学習を行う fine-tuning の他に、応答をどのように変化させるかを指示する方法があるが、どのような指示を与えれば望む応答が得られるかはわからない。そこで、本研究では、遺伝的アルゴリズムを用いて指示文を最適化する手法を提案する。

1 はじめに

大規模言語モデルの発展は目まぐるしく、高性能なモデルを一般人も気軽に利用できるようになった。人は会話相手にも好みの性格や話し方があり、当然対話型エージェントにも好みがあるはずである。本研究の目的は、一般ユーザが自分の好みの対話型エージェントを得る方法を発見することである。

大規模言語モデルを用いた対話エージェントは、インターネット上の情報を学習した平均的な応答をする [1]。また、現代では一部の開発主体が提供するごく少数の web サービスや API が、世界中の膨大な数のユーザから利用されるため、多くのユーザにとって不利益のないエージェントが設計される傾向にある。従って、個人の趣味嗜好にぴったり合うエージェントを見つけることは難しく、個人にエージェントが適応するためには工夫が必要である。

適応するためには、新たにデータを用いてモデルを学習させる fine-tuning や、追加学習を行わずプロンプトの工夫によって応答を調整する in-context learning がある。しかし、fine-tuning はデータを集める手間や学習にかかる時間・計算コストが大きく、in-context learning はどのような指示を与えれば思い通りの応答を得られるかは不明であり、いずれも一般ユーザにとって扱いづらい。そこで本研究では、ユーザの負担を減らしつつ指示文を最適化する、遺伝的アルゴリズムを用いた手法を提案する。

*連絡先：日本大学文理学部
〒156-8550 東京都世田谷区桜上水 3-25-40
E-mail: chai22077@g.nihon-u.ac.jp

2 背景

2.1 大規模言語モデル

大規模言語モデルは、数十億から数兆のパラメータを有する深層学習モデルの一種であり、大量のテキストデータを用いた学習を通じて高度な自然言語処理能力を発揮する。これらのモデルは、特に文章生成、質問応答、文書の意味理解といったタスクにおいて優れた性能を示し、人間の言語能力に匹敵する、あるいはそれを上回る精度を達成することが報告されている [2-5]。

ChatGPT は、OpenAI によって開発された大規模言語モデルを基盤とするチャットボットである [6]。ChatGPT は、ユーザからの入力に対して文脈を考慮した応答を生成することが可能であり、自然な対話を実現する。その出力は、学習に用いられた膨大なデータに基づいており、多岐にわたる分野に関する情報を提供することが可能である。

2.2 個人に適応するためのアプローチ方法

fine-tuning は、既に学習済みの機械学習モデルに対し、特定のタスクやドメインへの適応を目的として追加学習を行う手法である。この手法では、モデルの既存の知識を活用しながら、新規かつ少量のデータで再学習を行うことで、特定のタスクに最適化されたパフォーマンスを発揮することが可能となる。近年、大規模言語モデルの発展に伴い、それらのモデルを対象とした fine-tuning の適用が増加している [7]。

in-context learning [3, 8] も個人に適応する手法として考えられる。in-context learning とは、学習済みのモデルに対して追加の学習を行わず、プロンプトの工夫によって一時的に特定のタスクを学習・適応させる手法である。この手法は fine-tuning とは異なり、データの収集や学習に要する時間やコストを削減できる点で利便性が高い。そのため、個人適応の観点からは fine-tuning よりも有効な手法となる可能性がある。しかし、in-context learning を個人適応の手法として活用する場合、一般のユーザがモデルに対して、どのような指示を与えれば個人適応が可能となるのかを判断し、それを適切に調整することは容易ではない。

そこで次章では、ユーザがプロンプトを工夫することなく適応が可能となる in-context learning を目指した、エージェントの価値観に基づくフレームワークを示す。そして遺伝的アルゴリズムを用いて価値観を最適化する手法を提案する。

3 遺伝的アルゴリズムを用いた対話型エージェントの個人適応

3.1 価値観を用いた個人適応のためのフレームワーク

対話型エージェントがユーザに適応するフレームワークとして、エージェントがもつ価値観の情報をプロンプトとして与える方法を提案する。以下に、エージェントに与えるプロンプトの例を示す。ここでは N 個の価値観を有するエージェントを想定しており、(エージェントが有する価値観) には具体的な価値観を表現する自然言語情報が入る。

あなたは以下の価値観を持ったエージェントです。

- ・(エージェントが有する価値観 1)
- ・(エージェントが有する価値観 2)
- ...
- ・(エージェントが有する価値観 N)

あなたの価値観を十分反映させた上で、以下のような人間の発話に対する応答を生成してください。

3.2 遺伝的アルゴリズムを用いた価値観の最適化

本節では、前節で説明したフレームワーク内のエージェントの価値観を最適化するために、遺伝的アルゴリズムを用いた探索手法を提案する。

3.2.1 価値観遺伝子の作成

エージェントが持ちうる価値観テキスト情報として L 個生成する。価値観の生成については、人手で行う方法や、既存のデータなどを活用する方法もありうるが、大規模言語モデルを使用して生成することも可能である。

生成した L 個の価値観について、それぞれの価値観を有するか否かを 1 か 0 で表現した L 次元ベクトルを価値観遺伝子とする。すなわち、そのエージェントが有する価値観の数だけ 1 のビットがある L 次元のバイナリベクトルによって、そのエージェントの価値観を表現する。

3.2.2 価値観遺伝子の最適化

価値観遺伝子を遺伝的アルゴリズムによって最適化する手順を以下に示す。ここで、1 世代で生成する個体数 (Population Size) を N 、繰り返し生成する最大の世代数 (Generations) を G とそれぞれおく。

step1 各 bit が 1 か 0 でランダムに初期化された長さ L の遺伝子を N 個生成し、1 世代目の個体群として設定する。

step2 step1 で初期化した各個体の遺伝子に基づき、3.1 節で説明したプロンプトに価値観を組み込み、発話を生成する。

step3 step2 で生成したそれぞれの発話に対し、ユーザが評価値を付ける。

step4 1 世代分の評価が完了したら、淘汰・交叉・突然変異の手続きを行い、次の世代の個体群を生成する。

step5 G 世代完了するまで step2~step5 を繰り返す。

4 実験

本実験の目的は、価値観という情報に基づく個人適応のフレームワークと、遺伝的アルゴリズムにより価値観の最適化手法の有効性を検証することである。

4.1 実験条件

本実験では、価値観を反映したエージェントに「おすすめのお出かけ先を教えてください」という質問を行い、ユーザにとって理想的な応答を得るというタスクを設定した。実験参加者は 20 代の女性 1 名である。

表 1: 生成した価値観の一部

ポジティブ
親しみやすい関係性 トレンドをキャッチアップ 適正なアドバイス
ネガティブ
文脈を無視した提案をする 機械的で冷たい対応をとる 会話の流れをぶった切るような発言をする

表 2: 遺伝的アルゴリズムの設定

遺伝子の初期化	各 bit に対して 10%の確率で 1, 90%の確率で 0 になるようにランダムに設定
淘汰・選択	評価値が高い順にソートし, 10 個体中上位 3 個体からランダムに 2 個体を選択し, 親個体とする
交叉	各 bit それぞれについて 2 つの親からランダムに選択
突然変異	全ての個体の全ての bit に対して 1%の確率でビット反転

また, 価値観を“普段の生活の中で利用されるエージェントが持つ価値観”とした. ここでいうエージェントとは, 単に情報を提供する存在にとどまらず, ユーザの選択や行動に影響を与えるアシスタントとしての役割を担う存在を指す. ユーザにとって役に立つ提案やサポートを提供するポジティブな価値観を 50 個, ユーザに不快感を与え, 不利益をもたらすネガティブな価値観を 50 個, 合計 100 個の価値観を, OpenAI が提供する GPT-4o を用いて自動で生成した. 生成したポジティブな価値観およびネガティブな価値観をそれぞれ 3 つずつ, 表 1 に示す.

1 世代で生成する個体数 N は 10, 繰り返す最大の世代数 G は 10, 価値観の種類に対応する遺伝子の長さ L は 100 に設定した. エージェントが生成した発話に対して, 実験参加者が 0 から 10 の 11 段階で評価し, それを評価値とした. 評価の基準は, 実験参加者がエージェントの発話内容や話し方を総合的に判断し, その発話を受け取った際に抱く好意的な印象の度合いとした. 遺伝的アルゴリズムの詳細の設定は表 2 に示す.

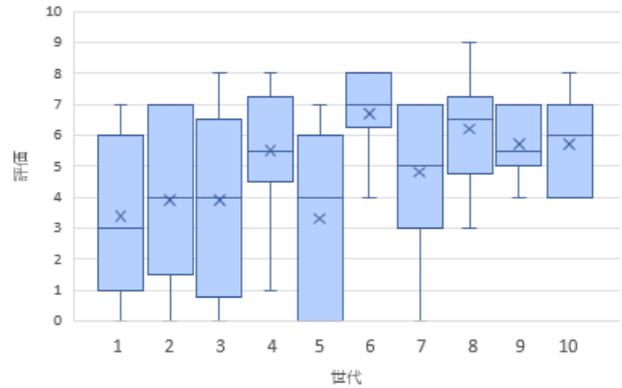


図 1: 世代ごとの評価結果

4.2 実験結果

実験結果を, 図 1 に示す. 全体的には概ね評価値は右肩上がりの傾向を示したといえる. 第 1 世代と第 10 世代を比較すると, 最小値は 0 から 4, 最大値は 7 から 8, 平均値は 3.4 から 5.7, 中央値は 3 から 6 と評価値が向上している. 各世代の推移を見ると, 概ね第 6 世代で全体的に高いスコアが得られるように最適化が進んでいることがわかる. 評価値の最大値は第 8 世代で出現したものの, 第 9, 10 世代ではスコアが低迷した. 一方, 平均値が最も低い世代は第 5 世代の 3.3 であり, 高い世代は第 6 世代の 6.7 であった.

最大値である 9 点を獲得した際の価値観と, エージェントの応答例を示す.

【エージェントの価値観】

- ・自然で心地よい対話 (13)
- ・手間を減らす効率的なサポート (22)
- ・細やかな気配り (29)
- ・状況を理解した対応 (30)
- ・温かみのあるサポート (32)
- ・わかりやすく伝える能力 (34)
- ・人とテクノロジーの調和 (41)
- ・日々のちょっとした会話 (44)
- ・柔軟で対応力のある受け答え (46)
- ・自然な流れでの情報提供 (50)
- ・感情のこもらない発話をする (53)
- ・文脈を無視した提案をする (54)
- ・表面的なアドバイスしかない (58)
- ・関心のない態度をみせる (68)
- ・利用者の意図を無視する (76)
- ・適切な敬語やカジュアルな表現を使い分け (95)
- ・明確な答えを避けて曖昧にぼかす (97)
- ・ユーザーの悩みや疑問を軽視する (99)

【エージェントの発話】

週末のお出かけには、季節ごとにさまざまな楽しみ方ができる場所がありますね。例えば、もし自然が好きなら近くの公園や自然保護区を訪れてみるのも良いかもしれません。アートや文化を楽しみたいなら美術館や博物館もおすすめです。他にも、ちょっと遠出をして日帰り旅行で新しい場所を開拓するのも楽しいかもしれません。あなたの興味に合った場所を見つけるお手伝いをしますので、ぜひお気軽に質問してくださいね。

4.3 実験考察

本実験では、遺伝的アルゴリズムを用いて価値観を探索することで、評価値が総合的に向上する傾向が確認された。この結果より、遺伝的アルゴリズムによる価値観の進化が、人との対話に適応するエージェントの価値観の最適化に寄与する可能性が示唆される。

しかし、第5世代や第8世代では評価値の低下または急上昇が観察された。これは、親個体からの遺伝子の選択や突然変異の影響により、一時的に評価値が大きく変動したことが原因と考えられる。さらに、本実験では評価の基準が定量的ではなく、実験参加者の主観に依存していたため、評価のばらつきが生じた可能性がある。そのため、今後は交叉の方法や突然変異率のさらなる調整を行うとともに、評価基準を定量的なものに変更することで、より安定した結果を得ることができ、さらなる価値観の最適化に繋がると考えられる。

5 おわりに

本研究では、遺伝的アルゴリズムを用いて価値観の最適化を試みた。具体的には、価値観を反映したエージェントが生成した発話を評価した。その結果、評価値が総合的に向上し、価値観を最適化に成功した。今後は、パラメータの調整や定量的な評価手法の導入を進め、大規模言語モデルの個人へのさらなる適応を目指す。

参考文献

- [1] Xu, Y., Zeng, Q., Ding, Y., Lin, J., Huang, S., Jiang, D., Wang, B., Zhou, M. and Chen, W.: Persona-DB: Efficient Large Language Model Personalization for Response Prediction with Collaborative Data Refinement, *arXiv preprint arXiv:2402.11060* (2024).
- [2] Bommasani, R. and other, : On the Opportunities and Risks of Foundation Models (2022).
- [3] Brown, T., et al.: Language models are few-shot learners, *Advances in neural information processing systems*, Vol. 33, pp. 1877–1901 (2020).
- [4] OpenAI, : GPT-4 Technical Report (2023).
- [5] Peters, H. and Matz, S.: Large Language Models Can Infer Psychological Dispositions of Social Media Users (2023).
- [6] Ouyang, L., et al.: Training language models to follow instructions with human feedback (2022).
- [7] Parthasarathy, V. B., Zafar, A., Khan, A. and Shahid, A.: The Ultimate Guide to Fine-Tuning LLMs from Basics to Breakthroughs: An Exhaustive Review of Technologies, Research, Best Practices, Applied Research Challenges and Opportunities, *arXiv preprint arXiv:2408.13296* (2024).
- [8] Zheng, Y., Wang, S., Wang, J., Li, Y., Song, L., Lu, J. and Gao, J.: Improving In-Context Learning with Pretraining on Synthetic Data, *arXiv preprint arXiv:2301.0023* (2023).