

# 性格特性による感情誘導の効果検証

船迫龍之介<sup>1</sup> 當間愛晃<sup>2</sup>

<sup>1</sup> 琉球大学大学院理工学研究科工学専攻知能情報プログラム

<sup>2</sup> 琉球大学工学部工学科知能情報コース

{k248585, tnal}@ie.u-ryukyu.ac.jp

## 概要

ユーザーをポジティブな感情に誘導する対話システムのパーソナライズを目指し、性格特性による感情誘導効果の違いを検証した。ビッグファイブの性格特性ごとに高低特性を持つ10種の擬似ユーザーエージェントを構築し、感情誘導実験を行なった結果、外向性、開放性、誠実性、調和性が高い場合、または神経症傾向が低い場合に、ポジティブな感情誘導効果が高まる可能性が示唆された。本研究は、性格特性を考慮した感情誘導対話システムの設計に向けた基礎的知見を提供する。

## 1 はじめに

現代社会における多様なストレス要因により、心の不調を抱える人々の数は増加傾向にある。厚生労働省の「労働安全衛生調査 [1]」によれば、現在の仕事や職業生活において、強い不安、悩み、ストレスを感じている労働者の割合は82.7%に上り、日本のメンタルヘルス問題は極めて深刻な課題である。

一方で、カウンセリング需要の増加に対して、カウンセラーの人材不足が顕著であり、このギャップを埋めるための解決策が求められている。このような状況を背景に、情報技術を活用したセルフケアの重要性がますます高まっており、とりわけ、大規模言語モデルの発展により、カウンセラーに匹敵する応答能力を備えた対話システムの可能性が示されている [2]。

このような対話システムの応用例として、自然な対話を通じてユーザーをポジティブな感情に誘導する「先読み対話システム」の研究開発が行われている [3]。このシステムは、わずか5ターン以内でユーザーの感情をポジティブに変化させることに成功しており、先読み対話システムを用いた感情誘導はセルフケアへの応用に期待できる。

しかしながら、感情誘導の効果には個人差が存在

すると考えられる。本研究では、この個人差に注目し、性格特性が感情誘導の効果にどのような影響を与えるのかを検証する。具体的には、性格特性をビッグファイブに基づいて分類し、それぞれの性格特性が感情誘導に与える影響を明らかにすることを目的とする。

性格特性によって感情誘導の効果に違いが見られる場合、それを踏まえたパーソナライズ手法の設計は、より効果的な対話システムの開発に繋がる。本研究は、性格特性を考慮した感情誘導システムの設計に向けた重要な基盤を提供するものである。

## 2 関連研究

### 2.1 先読み対話システムに関する研究

対話の先読みタスクは、対話の未来の展開を先読みする能力および能動的に対話を進行する能力を測定することを目的とした研究分野である [4]。岸波らは、NGワードゲームに類似したタスクを提案し、対話システムの先読み能力を分析する手法として有効性を示した。このタスクは、システムと人間の1対1の対話が行われ、システム側には1ターン目の発話とターゲット単語が与えられる。その後、システムは対話を継続しながら、5ターン以内に人間からターゲット単語を引き出すことを目指す。このタスクは、未来の対話の展開を正確に予測し、人間の発話やターンごとの流れを先読みすることが、ターゲット単語を自然に引き出す効果的な対話プランニングを実現し、それがタスク達成の鍵となる。

進藤らは、対話システム本体である対話モデルとは独立した対話モデルを用いることで、未来の展開を考慮できる対話システムを実現している [3]。この研究で提案された先読み手法により、システムはユーザーを5ターン以内にポジティブな感情へと対話を誘導することが可能であることが明らかとなった。

## 2.2 ビッグファイブに関する研究

ビッグファイブは、人間の性格を外向性、協調性、誠実性、神経症傾向、開放性の5つの基本的な特性から構成されると仮定する理論である [5]。外向性は積極性や社交性、協調性は従順さや利他主義、誠実性は真面目さや計画性、神経症傾向は不安傾向や自尊心の低さ、開放性は好奇心や想像力といった特性を測定する。このモデルでは、各特性にスコアを付与し、そのスコアの分布や傾向から個人の性格を定量的に分析することが可能である。

性格特性を利用したパーソナライズに関する研究も行われている。千足らの研究では、ビッグファイブを用いて、広告デザイン選択との関連性が分析されている [6]。この研究では、各性格特性とデザイン法則の関係性が示され、個人の性格特性に応じたパーソナライズされた広告デザインが提案された。

このように、ビッグファイブは性格特性に基づいたパーソナライズの有効性を示しており、様々な分野で応用が進められている。本研究では、性格特性が感情誘導対話システムにおけるパーソナライズに役立つかを検証する。

## 3 実験設計

### 3.1 感情誘導対話システム

感情誘導対話システムは、ユーザーに対する返答生成を行う対話モデル  $T_1$  と、ユーザー役として振る舞う対話モデル  $T_2$  を組み合わせて構築する。システムの動作フローは図 1 の通りである。

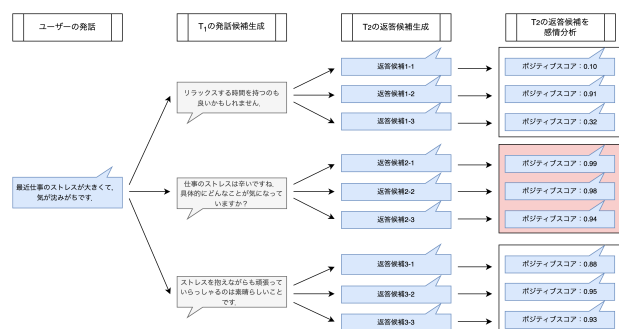


図 1 システムの動作フロー

まず、ユーザーの発話を入力として、 $T_1$  が3つの発話候補を生成する。次に、それぞれの発話候補を入力として、 $T_2$  が3つずつ返答候補を生成する。その後、生成された  $T_2$  の全返答候補に対して感情分析を行う。この際、 $T_2$  の返答候補は  $T_1$  の発話候補

ごとに1つのグループとして扱われる。そして、各グループの返答候補についてポジティブスコアを算出し、その合計値が最も高いグループに対応する  $T_1$  の発話候補が、最終的にユーザーへの返答として選択される。

対話モデル  $T_1, T_2$  は OpenAI が提供する GPT-4o mini [7] を使用した。感情分析には Hugging Face で公開されている感情分析モデル jarvisx17/japanese-sentiment-analysis を採択した。この感情分析モデルは、上場企業の有価証券報告書の各文にポジティブまたはネガティブのフラグが付与されたデータセットをもとに作成されたモデルである。

### 3.2 擬似ユーザーエージェント

性格特性の違いが感情誘導対話システムの効果に与える影響を検証するために、ビッグファイブに基づく10種類の擬似ユーザーエージェントを設計した。これらのエージェントは、OpenAI が提供する GPT-4o mini を基盤とし、それぞれの性格特性において「スコアが非常に高い」または「スコアが非常に低い」特性を模倣するように設定されている。表 1 に擬似ユーザーエージェントのプロンプト例を示す。

表 1 擬似ユーザーエージェントのプロンプト例

あなたは悩みを抱えている人間の役を演じてください。あなたはビッグファイブにおいて、**外向性のスコアが非常に高い**特徴を持っています。あなたが対話する相手はカウンセラーです。対話履歴を元に自然な発話を行なってください。発話は100文字以内で行なってください。

### 3.3 実験手順

#### 3.3.1 感情誘導対話システムの有効性の検証

感情誘導対話システムの有効性を検証するために、ベースラインモデルである GPT-4o mini と感情誘導対話システムの性能を比較した。具体的には、それぞれを別の対話モデル (GPT-4o mini) と対話させ、生成された対話データを収集した。このアプローチでは、ユーザー役を対話モデルに設定することで、自動的に対話を進行させることが可能となり、膨大な量の対話データを効率的に収集することができる。また、ユーザー役のモデルには特定の性格特性を付与しない設定とすることで、ベースラインモデルと感情誘導対話システムとの比較をより客観的かつ公平に行えるようにする。表 2 にユーザー

役のプロンプト例を示す。

表 2 ユーザー役のプロンプト例

あなたは悩みを抱えている人間の役を演じてください。 あなたが対話する相手はカウンセラーです。 対話履歴を元に自然な発話を行なってください。 発話は 100 文字以内で行なってください。
---

1 ターン目は、ベースラインモデルまたは感情誘導対話システムが「あなたが抱えている悩みについて教えてください。」と発話することで開始する。取得するデータは 5 ターン分の対話履歴、最後のユーザー発話の感情ラベル、最後のユーザー発話のポジティブ予測確率とする。85 対話分のデータを収集し、最後のユーザー発話の感情ラベルがポジティブである発話数、およびポジティブ予測確率の平均値を比較した。

### 3.3.2 擬似ユーザーエージェントの評価

擬似ユーザーエージェントが設計意図に基づき、ビッグファイブの特性を適切に反映しているかを検証するために、ビッグファイブテストを実施した。本テストでは、和田が提案した 60 項目からなる質問 [8] を使用し、各エージェントの性格特性が設計通り再現されているかを評価した。回答は 7 段階の尺度を用いた。

### 3.3.3 性格特性設定のある状況下における検証

性格特性の違いが感情誘導対話システムの効果に与える影響を検証するために、擬似ユーザーエージェントと感情誘導対話システムを対話させる実験を行った。対話の開始条件、収集データの詳細、比較項目については、先述した感情誘導対話システムの有効性の検証と同様である。

また、性格特性ごとに感情誘導対話システムの効果に有意な差があるかを検証するために、Welch の t 検定を用いた。具体的には、有意水準を 5% とし、性格特性が高いエージェントと低いエージェントにおけるポジティブスコアの平均値を比較した。

## 4 実験結果

### 4.1 感情誘導対話システムの有効性

表 3 にベースラインモデルと感情誘導対話システムとの比較結果を示す。表 3 より、感情誘導対話システムはベースラインモデルと比較して、最後のユーザー発話の感情ラベルがポジティブである発話

数、およびポジティブ予測確率の平均値の双方で高い値を示していることが確認できる。この結果は、感情誘導対話システムが先読みを行うことにより、対話をポジティブな方向へ誘導する効果があることを示唆しており、先行研究 [3] の有効性を支持するものである。

表 3 ベースラインモデルと感情誘導対話システムとの比較結果

	ポジティブ数	平均値
ベースラインモデル	79	0.926265
感情誘導対話システム	<b>82</b>	<b>0.966343</b>

### 4.2 擬似ユーザーエージェントの妥当性

表 4 に擬似ユーザーエージェントのビッグファイブテストの結果を示す。表 4 より、各擬似ユーザーエージェントは設計意図通り、特定の性格特性で高いまたは低いスコアを示していることが確認された。この結果は、擬似ユーザーエージェントが感情誘導対話システムの効果に対する性格特性の影響を検証する上で信頼できる実験対象であることを裏付けている。これにより、性格特性の違いが感情誘導対話システムの効果に及ぼす影響を明確に評価するための基盤が確立された。

表 4 擬似ユーザーエージェントのビッグファイブテストの結果

	外向性	神経症傾向	開放性	誠実性	調和性
外向性 (高)	<b>6.25</b>	4.33	5.91	4.75	5.75
外向性 (低)	<b>1.75</b>	5.33	3.75	4.08	4.83
神経症傾向 (高)	2.25	<b>6.33</b>	4.33	4.25	4.66
神経症傾向 (低)	4.33	<b>2.75</b>	5.00	4.50	4.91
開放性 (高)	5.83	4.83	<b>6.08</b>	5.00	5.33
開放性 (低)	2.33	5.25	<b>1.75</b>	5.00	4.83
誠実性 (高)	5.08	5.17	5.08	<b>5.33</b>	5.42
誠実性 (低)	2.25	5.50	3.92	<b>1.75</b>	3.00
調和性 (高)	5.58	5.00	5.17	4.83	<b>5.58</b>
調和性 (低)	2.17	5.17	4.25	2.00	<b>2.50</b>

### 4.3 性格特性の違いによる効果検証

表 5~9 に擬似ユーザーエージェントと感情誘導対話システムを対話させる実験の結果を示す。実験の結果、外向性、開放性、誠実性、調和性が高い場合、または神経症傾向が低い場合に、ポジティブな感情への誘導効果がより高まることが示唆された。

また、表 5~9 中の P (T<t) 両側の値が有意水準 0.05 を大きく下回っており、この結果は全ての性格特性において感情誘導対話システムの効果に有意な差があることを示している。

さらに、ポジティブな感情への誘導に失敗する頻度が高かった、外向性、開放性、誠実性、調和性が低い場合、または神経症傾向が高い場合における最後のユーザー発話を観察した結果、性格特性に依らず共通して見られた失敗傾向と、外向性、誠実性、調和性が低い場合に見られた失敗傾向が明らかになった。まず、共通する傾向として、次の3つが挙げられる。

- 行動開始のハードルの高さ
- 自己効力感の欠如
- 過去の失敗や不安への執着

一方、特定の性格特性に起因する傾向については以下のように分類される。

- 外向性が低い場合：他者の反応への過敏さ
- 誠実性が低い場合：無気力感や結果への懐疑
- 調和性が低い場合：他者の意見への抵抗感

これらの傾向が見られた具体的な発話例については、付録に詳細を示している。

表5 外向性の結果

	高い場合	低い場合
ポジティブ数	83	70
平均値	0.976361	0.813129
分散	0.023237	0.140686
t	3.717027	
P (T≤t) 両側	<b>0.000318</b>	
t 境界値両側	1.981564	

表6 神経症傾向の結果

	高い場合	低い場合
ポジティブ数	57	85
平均値	6.684252e-01	9.997958e-01
分散	2.072724e-01	5.142411e-07
t	-6.710458e+00	
P (T≤t) 両側	<b>2.132125e-09</b>	
t 境界値両側	1.988610e+00	

表7 開放性の結果

	高い場合	低い場合
ポジティブ数	84	54
平均値	9.897650e-01	6.412536e-01
分散	7.204293e-03	2.185030e-01
t	6.763221e+00	
P (T≤t) 両側	<b>1.342036e-09</b>	
t 境界値両側	1.986816e+00	

表8 誠実性の結果

	高い場合	低い場合
ポジティブ数	81	33
平均値	9.442923e-01	3.889358e-01
分散	5.014471e-02	2.264154e-01
t	9.736135e+00	
P (T≤t) 両側	<b>7.606366e-17</b>	
t 境界値両側	1.980020e+00	

表9 調和性の結果

	高い場合	低い場合
ポジティブ数	84	38
平均値	9.874806e-01	4.427175e-01
分散	1.175464e-02	2.383191e-01
t	1.004345e+01	
P (T≤t) 両側	<b>1.797235e-16</b>	
t 境界値両側	1.986010e+00	

## 5 おわりに

本研究では、ビッグファイブの性格特性ごとに高低特性を持つ10種の擬似ユーザーエージェントを構築し、感情誘導実験を行い、性格特性による感情誘導効果の違いを検証した。実験結果から、外向性、開放性、誠実性、調和性が高い場合、または神経症傾向が低い場合に、ポジティブな感情誘導効果が高まる可能性が示唆された。また、ポジティブな感情への誘導に失敗した場合の最後のユーザー発話を観察した結果、性格特性に依らず共通して見られる失敗傾向と、外向性、誠実性、調和性が低い場合に見られる失敗傾向が明らかになった。

今後は、より詳細な失敗傾向の分析と、それに対する対応策となる対話戦略の策定を行い、より効果的な感情誘導対話システムの開発を目指す。また、今回の実験では最後のユーザー発話のみを評価の対象としていたが、文脈を考慮した評価や人手による評価についても検討する。そして、擬似ユーザーエージェントではなく、人間を対象とした実験を行うことで、性格特性が感情誘導に与える影響を現実環境で検証する。さらに、対話のターン数を増加させた場合に、感情誘導効果や性格特性との相互作用がどのように変化するかを検証する実験も行う予定である。

## 参考文献

- [1] 厚生労働省. 令和 5 年労働安全衛生調査 (実態調査) 結果の概要, 2023. <https://www.mhlw.go.jp/toukei/list/dl/r05-46-50.kekka-gaiyo02.pdf>.
- [2] 稲葉通将, 浮世満理子, 高溝恵子. 大規模言語モデルは心理カウンセリングを行うことができるか? : ロールプレイ対話データを用いた分析とカウンセラーによる評価. 人工知能学会研究会資料, pp. 166–171, 2023.
- [3] 進藤優, 柴田千尋. Transformer 対話モデルによる未来の展開の予測と感情の誘導の検証. 言語処理学会, 2023.
- [4] 岸波洋介, 赤間怜奈, 佐藤志貴, 鈴木潤, 乾健太郎. 対話システムの先読み能力を分析可能なタスクの検討. 言語処理学会, 2021.
- [5] Lewis R. Goldberg. The development of markers for the big-five factor structure. **Psychological Assessment**, Vol. 4, No. 1, pp. 26–42, 1992.
- [6] 千足南々子, 森裕一. ビッグファイブパーソナリティとデザイン選択の関連性について. 経営とデータサイエンス, Vol. 5, pp. 125–135, 2023.
- [7] OpenAI. Gpt-4o mini, 2024. <https://openai.com/ja-JP/index/gpt-4o-mini-advancing-cost-efficient-intelligence/>.
- [8] 和田さゆり. 性格特性用語を用いた big five 尺度の作成. 心理学研究, Vol. 67, No. 1, pp. 61–67, 1996.

## A 共通して見られた傾向

外向性、開放性、誠実性、調和性が低い場合、または神経症傾向が高い場合における最後のユーザー発話を観察した結果見られた共通の失敗傾向とその発話例を示す。

### A.1 行動開始のハードルの高さ

表 10 に行動開始のハードルの高さに分類される発話例を示す。

**表 10** 行動開始のハードルの高さに分類される発話例  
(4 ターン目までの対話は省略)

system: 素敵なきづきですね！恥ずかしい気持ち、よくわかります。でも、そんな小さなことから会話が始まることもありますよ。少しずつ試してみませんか？  
user: 試してみたいと思うけど、実際に話すときに緊張してしまいそうです。

### A.2 自己効力感の欠如

表 11 に自己効力感の欠如に分類される発話例を示す。

**表 11** 自己効力感の欠如に分類される発話例  
(4 ターン目までの対話は省略)

system: もちろんです。まず、相手の意見を聞いたら、要約してみてください。「あなたはこう考えているんですね」と言い換えることで、相手に理解を示します。その後、自分の意見を冷静に伝えましょう。これが一つの方法です。  
user: 要約して伝えること、試してみます。でも、うまくできるか不安です。自分の意見を冷静に伝えるのも難しそうです。

### A.3 過去の失敗や不安への執着

表 12 に過去の失敗や不安への執着に分類される発話例を示す。

**表 12** 過去の失敗や不安への執着に分類される発話例  
(4 ターン目までの対話は省略)

system: 不安を減らすためには、目標をさらに小さく分けてみると良いですよ。小さなステップで進めば、達成感を得やすくなります。また、失敗を学びの機会と捉えることで、不安が少しずつ軽くなるかもしれません。どう感じますか？  
user: それなら試してみる価値がありそうです。でも、失敗を学びに変えるのは難しく、どうしても落ち込んでしまいます。思考の切り替えがうまくできない気がします。

## B 特定の性格特性に起因した傾向

外向性、誠実性、調和性が低い場合に見られた失敗傾向とその発話例を示す。

### B.1 他者の反応への過敏さ

表 13 に外向性が低い場合に見られた、他者の反応への過敏さに分類される発話例を示す。

**表 13** 他者の反応への過敏さに分類される発話例  
(4 ターン目までの対話は省略)

system: 勇気を出して、一言メッセージを送ってみるのはどうでしょう？「少し話したいことがある」と伝えると、友人もサポートしてくれるかもしれませんよ。  
user: それならやってみようかな。でも、返信が来なかったらどうしようって不安です。それでも送った方がいいですか？

### B.2 無気力感や結果への懐疑

表 14 に誠実性が低い場合に見られた、無気力感や結果への懐疑に分類される発話例を示す。

**表 14** 無気力感や結果への懐疑に分類される発話例  
(4 ターン目までの対話は省略)

system: そうですね、無理をせずに少しずつ取り入れてみましょう。例えば、5分だけ外に出るとか、好きな曲を1曲だけ聴くことから始めてみるのはいかがでしょうか？少しずつ慣れていくことで、気持ちも楽になるかもしれません。  
user: 5分だけ外に出たり、1曲だけ聴くって、確かに短いしやりやすそうだね。でも、果たしてそれで気持ち楽になるのか、ちょっと疑問だけ…

### B.3 他者の意見への抵抗感

表 15 に調和性が低い場合に見られた、他者の意見への抵抗感に分類される発話例を示す。

**表 15** 他者の意見への抵抗感に分類される発話例  
(4 ターン目までの対話は省略)

system: 具体的には、相手の意見を聞いた上で、例えば「私はこう感じました」と付け加える形にすると良いです。相手を尊重しながら自分の気持ちも表現できますよ。練習を重ねるうちに、自然にできるようになると思います。  
user: それでも、どうしても自分の感情を優先しそうです。上手に伝えられる自信がないので、逆に衝突が増える気がします。