

人狼ゲームにおける対話相手の性格特性に基づく説得戦略の検討

A study of Persuasion Strategies based on the Interlocutor's Personality Traits in Werewolf Games

小濱 賢史¹⁾ 南條 浩輝¹⁾
Satoshi KOHAMA Hiroaki NANJO

1 はじめに

不完全情報コミュニケーションゲーム「汝は人狼なりや?」(以下, 人狼ゲームと表記)を人工知能に高度にプレイさせる「人狼知能プロジェクト」[1]が進められている。人狼ゲームは対話を通じて相手を説得することが重要なカギとなるゲームである。

人狼ゲームにおいても現実社会においても, 相手を説得する際に重要と考えられる点として, 説得対象の相手に適した説得発話を行うことが挙げられる。すなわち, 同じ内容を説得する場合でも, 相手の年齢や性格などを考慮して最も伝わりやすい, 受け入れられやすい話し方をすることが重要である。

このような背景に基づき, 本研究では, 人狼ゲームにおいて対話相手の性格特性に基づいて説得を行う方法を研究する。具体的には, 対話相手の性格特性を推測し, 相手の性格特性に適した説得発話を行う戦略について検討する。本研究では, 性格特性ごとに好まれる説得発話を調査したので, その結果について述べる。

2 性格特性

本研究では, 対話相手の性格がわかっていた場合に, その性格の人の説得に適した発言を生成 AI が生成可能かを検証する。これが可能となる場合, 対話相手を説得しやすい文を生成することで, ゲーム中の議論を自身に有利になるように進めることができる。

本研究では性格の指標として「ビッグファイブ性格特性」を使用する。ビッグファイブ性格特性とは, 「An Alternative "Description of Personality": The Big-Five Factor Structure」[2]で紹介された, 人間の性格構造を表すためには五つの基本的な因子で説明できる, とされる考え方である。この引用した論文では, これらの因子が特定の語彙や尺度に依存せず, 異なる方法によっても再現されることを複数のデータセットを用いて実証している。この特性により, 性格を安定して測定できる枠組みとして信頼されており, 本研究でも性格の分析における基盤としてビッグファイブ性格特性を採用する。

本研究では, ビッグファイブ特性診断サイト [3] で提供される 5 つの特性を利用する。具体的には, 「外向-内向」, 「協調-排他」, 「勤勉-怠惰」, 「論理-情動」, 「創造-保守」の 5 特性を利用する。

3 実験内容

本研究では, 対話相手の性格特性がわかっている状況のもとで, その対話相手を説得するのに適した発話を人工知能が生成できるかを調べる。初めに, ChatGPT に対して, 説得対象となる対話相手の性格特性を教え, 対話相手を説得するために適していると考えられる発言を生成させる。例えば, 「外向-内向」から「外向的な性格を

持つ対話相手の説得発話」と「内向的な性格を持つ対話相手の説得発話」のペアを生成させる。次に, 生成された発話のペアを人間の被験者に提示し, どちらを支持するかをアンケート形式で聞く。人狼ゲームに関する説得発話に加え, 一般の説得場面を設定した説得発話も用いる。被験者には性格診断も受けてもらい, その結果とアンケートの答え (どちらを支持したか) を比較する。これにより, ChatGPT の発言が対話相手の性格に適しているかを確認する。

アンケートは, 滋賀大学の関係者に対して行い, 31 名分の回答を得た。収集したデータの中には, 回答時間が 1 分や 2 分と異常に短いものが含まれており, これらは適切な回答ではない可能性が高いため, データセットから除外した。データの整理後, 最終的に使用したデータは 29 人分である。また, アンケートは途中で辞めることが可能な形式であったため, 人狼ゲームに関する説得文と人狼ゲームに関係しない説得文では総回答数に差異が生じた。

これらの収集したデータを用いてアンケートで得られた回答と性格診断結果を照らし合わせ, クロス集計表を作成した。クロス集計表の数字の入れ方について説明する。表 1 の外向性についてのクロス集計表を例とする。ビッグファイブ性格診断を受けると, 5 つの性格特性の強弱が出力される。例えば, 外向性について外向的特性が 80% で内向的特性が 20%, と出力される。ここで出力されたパーセンテージのうちの片方, 今回は外向的のパーセンテージを入力してもらい, それを 5 クラスに分類する。具体的には 20 ずつ区切って, 100 から 81 を外向的, 80 から 61 をやや外向的, 60 から 41 を中立, 40 から 21 をやや内向的, 20 から 0 を内向的とする。

これを表 1 のようにクロス集計表に割り振った。ビッグファイブ性格診断を受けて外向的特性 90% の人が外向的性格者向けの説得文を支持した場合表 1 の左上隅 (5 と書かれている部分) の数字が一つ増える。この分析により, ChatGPT の考えた性格特性に応じた発言がその性格特性を持つ人に支持されるのかを検証した。

4 実験結果

表 1, 表 2, 表 3 にアンケートの回答から得られたクロス集計表を示す。表に書かれた数字は人数を表している。回答者の性格は 5 段階のスケールで評価されており, 列の端に行くほど上部に書かれている性格特性を強く持つことを表す。LLM の説得文は, 各性格特性ごとに対立する 2 つの性格特性に適した説得文を ChatGPT が生成したもののうち, どちらを支持するか回答者に選んでもらったものである。これらのクロス集計表は人狼ゲームに関する説得文, 人狼ゲームに関わらない説得文, すべての説得文の 3 種類の集計方法で分けられており, それぞれの説得文に外向性, 協調性, 勤勉性, 情動性, 創造性の 1 つずつの性格特性に合わせた説得文が用

1) 滋賀大学

表1 人狼ゲームに関する説得文

LLMの説得文	回答者の性格					Total
	外向的	←	中立	→	内向的	
外向的性格者向け	5	4	2	3	2	16
内向的性格者向け	4	4	10	7	2	27
Total	9	8	12	10	4	43

LLMの説得文	回答者の性格					Total
	論理的	←	中立	→	情動的	
論理的性格者向け	10	9	4	2	12	37
情動的性格者向け	1	1	2	0	2	6
Total	11	10	6	2	14	43

表2 人狼ゲームに関わらない説得文

LLMの説得文	回答者の性格					Total
	外向的	←	中立	→	内向的	
外向的性格者向け	7	1	7	6	0	21
内向的性格者向け	3	7	5	4	4	23
Total	10	8	12	10	4	44

LLMの説得文	回答者の性格					Total
	論理的	←	中立	→	情動的	
論理的性格者向け	5	4	2	0	10	21
情動的性格者向け	7	6	4	2	4	23
Total	12	10	6	2	14	44

意されている。

本稿では紙幅の制約により、外向性および情動性の2指標に関するグラフのみを掲載する。

5 考察

得られた結果を基に、LLMを人狼ゲームに応用できるか確認する。表3を確認すると、情動的な人が論理的な性格特性の人向けの説得文を多く支持していたり、説得文の支持数が同数になるパターンが見られるなど、相手の性格に合わせた発言が必ずしも成功していない結果が示された。特に、表1の「人狼ゲームに関する説得文」においては、回答者の性格特性に関係なく、論理的な人を説得する文を好む傾向が明らかとなった。また、保守的な性格の人を説得する文、勤勉な性格の人を説得する文も一定の支持を得る傾向が確認された。一方で、表2の「人狼ゲームに関わらない説得文」においては、このような傾向は見られなかった。この結果から、人狼ゲームという限定的な状況においては、性格特性を考慮することよりも、これらの傾向を踏まえた発言を意識することが有用である可能性がある。これらの結果は、人

表3 すべての説得文

LLMの説得文	回答者の性格						Total
	外向的	←	中立	→	内向的	無回答	
外向的性格者向け	12	5	9	9	2	9	46
内向的性格者向け	7	11	15	11	6	17	67
Total	19	16	24	20	8	26	113

LLMの説得文	回答者の性格						Total
	論理的	←	中立	→	情動的	無回答	
論理的性格者向け	15	13	6	2	22	14	72
情動的性格者向け	8	7	6	2	6	12	41
Total	23	20	12	4	28	26	113

狼ゲームにおける性格特性を用いたコミュニケーション戦略を再評価する必要性を示していると考えられる。対戦相手の性格特性に基づいたアプローチが効果を発揮しない場合でも、人狼ゲームにおける一般的な傾向を把握し、発言内容を調整することで、より効果的なコミュニケーションが可能になると考えられる。

6 まとめ

本研究では、人狼ゲームにおける対話相手の説得戦略を検討した。具体的には性格特性がわかっている対話相手に対して ChatGPT が適切な説得ができるかを調査した。ChatGPT が各性格の人の説得に適していると考えた説得発言を被験者に提示し、アンケート形式で選択させたところ、被験者が自身の性格特性に合致する説得発言を受け入れるという傾向はみられなかった。一方で、人狼ゲームに関する説得文を使用したアンケートに限定した場合は、論理的な説得文が特に好まれる傾向が確認できた。このことは、人狼ゲームにおいては相手の性格特性を考慮して説得を切り替えるよりも、常に論理的性格を持つ相手を仮定した説得戦略をとる方が説得成功率が高くなることを示唆している。なお、今回のアンケート対象は大学関係者であった。今後は様々な属性の人を対象として性格特性に基づく説得戦略の有効性を調査したい。さらに、相手の性格特性によらない、説得成功率が高い発言の生成戦略についても調査したい。

謝辞

本論文の作成にあたり、アンケートにご協力いただいた皆様に感謝する。

参考文献

- [1] Artificial intelligence based werewolf 人狼知能プロジェクト. <https://aiwolf.org/> (2025.01.10 アクセス) .
- [2] Lewis R. Goldberg. An alternative "description of personality": The big-five factor structure. In *Journal of Personality and Social Psychology*, pp. 1216–1229, 1990.
- [3] ビッグファイブ性格診断 big5-basic. <https://big5-basic.com/front/index.php?route=common/home> (2025.01.10 アクセス) .