

ChatGPT を用いた教育的ノベルゲームの評価実験の拡張

赤田直弥¹ 村田真樹²

^{1,2} 鳥取大学大学院 持続性社会創生科学研究科

² 鳥取大学工学部附属クロス情報科学研究センター

m25j4002a@edu.tottori-u.ac.jp murata@tottori-u.ac.jp

概要

我々の先行研究 [1] では、村田による ChatGPT を用いた教育的小説の生成 [2] や、宮本らによる ChatGPT によるノベルゲーム自動生成システム [3] を参考にして、教育的ノベルゲームを作成した。作成したノベルゲームの教育効果を小テストによって評価した結果、事前テストに対する事後テストの正解数が向上し、教育効果の可能性が見られたが、被験者数が少なく、評価のばらつきやノベルゲームのもつ課題について十分に分析できなかったという問題があった。そこで、本研究では、被験者数を増やし、2つの教育的なノベルゲームに対して各6人ずつの計12人による評価実験を行い、小テストによる教育効果の確認や、アンケートによる物語の質の評価を行った。実験の結果、いずれのノベルゲームにおいてもすべての被験者において事後テストの正解数が事前テスト以上となっており、全体平均では事前6.3問から事後8.6問に向上した。そのため、評価者数を増やした場合においても、本研究で作成したノベルゲームには教育効果があることが確認された。一方で、教育的内容と物語的内容のバランスに課題があり、教育的内容がノベルゲーム内で十分に提示されない場合や、物語の内容が薄くなってしまいう問題があった。そこで、本研究では学問要素と物語要素を段階的に追加する添加形式でのノベルゲームの生成を検討した。

1 はじめに

近年、大規模言語モデルの進化がめざましい。特に、ChatGPT を用いて生成された文章は、多岐の分野に渡って活用されている。ChatGPT を用いた小説や、教育的な小説 [2]、またノベルゲーム [3] を半自動的に作る研究が確認されている。一方ではデジタルゲームによる教育の研究もなされており、Manuel の研究 [4] では幼稚園生を対象としたビジュアルノ

ベルゲームによる教育効果が確認された。

そこで、我々の先行研究 [1] では、これを受けて ChatGPT を用いた教育的なノベルゲームの生成を行い、3人の被験者によって事前、事後に行うテストによる教育効果の検証や、アンケートによる物語内容の評価を行った。その結果、事後テストの点数が事前テストに比べて向上した。

また、ChatGPT によるノベルゲーム生成システムとしては、DeepGame [5] が存在する。DeepGame は、ChatGPT の質問に対してプレイヤーが逐次何をするかを入力して物語を進行させていく形式である。このシステムの問題点として、ChatGPT 側の質問と、プレイヤー側の回答をひたすら繰り返すため、ノベルゲームの終わりが曖昧になる問題がある。そのため、我々の研究では事前に物語の始まりから終わりまでの流れを決めるフローチャートを生成し、後から細かく文章化する形式を採用している。

我々の先行研究では、評価数が少ないため、個人差が結果に強く反映された可能性が否定できず、その傾向を十分に分析できなかった。そのため、今回の研究では、個人差の傾向やノベルゲームの課題をより詳細に分析することを目的として、被験者数を増やした実験を行う。具体的には、物語要素の異なる2つの教育的なノベルゲームに対して各6人の計12人による被験者実験を行う。

また、実験で得られたノベルゲームにおける物語要素や学問要素の不足という課題に対して、それらを物語に段階的に付加する添加形式による生成方法を検討する。本研究の内容の一部は文献 [6] にも記載している。

2 関連研究

2.1 What Is the Impact of ChatGPT on Education? A Rapid Review of the Literature

Chung[7]は、ChatGPTのさまざまな分野における機能や、教育での使用方法、またリリース初期の2022年12月から2023年2月までに研究者によって提起された潜在的な問題についての理解を深めることを目的にして、Google Scholarを検索して見つかった50件の論文を参照してレビューを行った。

ChatGPT-3.5の学力的パフォーマンスはテストや試験、面接を用いた21の研究で評価された。その結果は分野ごとに異なり、批判的思考や高次思考、また経済学で優れた結果を示したが、法律や医学、数学などの他の科目ではその結果はあまり良いとはいえなかった。例えば、経済学においてはChatGPTは平均的な学生よりも優れた成績を取めたが、他の科目では平均的な学生より100点満点で8~40点低い結果となった。

2.2 ChatGPTを用いた教育的ノベルゲーム

赤田ら[1]は、ChatGPTを用いて教育的なノベルゲームの半自動生成を行った。筆頭著者自らによる複数のノベルゲームの評価や、3人の被験者による実験の結果、この研究で生成したノベルゲームには教育効果があることがわかった。しかし、1つのみのノベルゲームについて、3人という少ない人数で実験を行っており、評価の信頼性が薄いという問題がある。

3 提案手法

3.1 ノベルゲームの生成方法

本研究では、先行研究[1]と同様にChatGPT-4o-miniを用いてノベルゲームを生成する。

本研究で生成するノベルゲームでは、ある1つのシーンから始まり、状況や情景を示す文章が表示された後、提示された選択肢を選ぶことで他のシーンに分岐していき、様々なエンディングに到達する方式とする。以下に示す手順でノベルゲームを生成する。

- (1)特定の学問に関する講義資料を生成させる。
- (2)講義資料をもとに、小テストを2つ生成させる。(小テストは被験者実験に用いる)

(3)物語の全体構想を生成する。没入性の高いノベルゲームを作るために、物語の要素(海外旅行、SF、料理など)を1つ入力する。これを物語のテーマとする。

(4)全体構想をもとに、シーン間の分岐を表すフローチャートを生成する。

(5)シーン毎に、詳しい状況や情景を描写する文章を生成する。必要ならばプロンプトで細かい指示をする。

(6)エンディングの内容となる文章を生成する。

実際の生成例は、付録A節に記載する。

3.2 各シーンの生成方法について

各シーンを生成する際の生成方法は、筆頭著者が研究しているものでは以下に示す2通りが存在する。

- ゲームブック型生成
- 文脈型生成

ゲームブック型生成は、すべてのシーンを1から順番に生成し、それらのシーンを1つのゲームブックとして纏めてから読むという方式である。ゲームブックとして生成を終えればそこでノベルゲームとして完成する。1から順にシーンを作るため、シーンの数が多い場合それだけ作るのに時間がかかる問題や、複数のシーンから1つのシーンへと分岐が発生している場合、そこで矛盾が生じるという問題がある。

文脈型生成は、まず生成したフローチャートをChatGPTに入力し、1つのシーンからスタートして、1つの選択肢を選んだと仮定して選んだシーンのみをそれまでの文脈も用いて生成していく方式である。ゲームブック型生成と比べ文脈の矛盾が少なくなり、また、作る時間も格段に短く済むが、フローチャートをChatGPTに入力してからその都度生成するため、被験者の手元にChatGPT環境が必要である問題や、ChatGPT自体のランダム性によって安定性が低く、シーンの内容やシーン番号などが誤って変化してしまう問題がある。

本研究では、一度生成を終わらせてしまえば被験者実験が容易にできるゲームブック型生成でノベルゲームを生成している。

4 実験

4.1 評価方法

ノベルゲームの有効性について評価するために、学習者に実際に使用してもらい、ノベルゲームとしての面白さや学習効果をアンケートや事前テスト、事後テストによって判定する。

以下の手順で評価を行う。

- (1)事前テストを解く。
- (2)ノベルゲームを読み、アンケートに回答する。
- (3)事後テストを解く。

アンケートによる評価では、村田 [2] の研究を参考にし、興味(興味を持ってプレイできたかどうか)、個人的な有用性(学問について勉強になったか)、他人への有用性(他の人の教育にも良いと思うか)のそれぞれについて-3 から+3 までの7段階で評価する。

また、ノベルゲームの生成時に生じる学問的な正確性や物語の一貫性の問題を評価できるように、正確性(物語の内容が正しいか)や一貫性(物語に筋が通っているか)についても-3 から+3 までの7段階で評価する。

4.2 実験結果

海外旅行をテーマとしたノベルゲームである「自己情報量の発見」について、著者以外の複数の被験者で実験を行った。我々の先行研究では被験者はA,B,Cの3人のみであったが、今回は新たにD,E,Fの3人を追加し計6人で実験を行っている。その結果を表1に示す。

アンケート項目の平均を見ると、正確性と一貫性では+2.3点と高い数値を示したが、興味や有用性(個)では相対的に低い数値を示した。すべての被験者において事後テストの正解数が事前テスト以上となり、平均では6.8問から9.0問に上昇した。被験者ごとに読んだ時間にばらつきが見られ、9分から27分までと幅広い。

また、新たにSFをテーマとしたノベルゲームである「情報の宇宙」についても、著者以外の6人の被験者で実験を行った。その結果を表2に示す。

アンケート項目の平均を見ると、正確性や一貫性については+2.5点、+2.3点と、表1と同様の高い数値を示した。また、興味や有用性(個)、有用性(他)についても+2点程度と高い数値を示した。表1と

同様に、すべての被験者において事後テストの正解数が事前テスト以上となり、平均では5.8問から8.2問に上昇した。

さらに、海外旅行とSFそれぞれの評価をまとめたものを表3に示す。全体平均から見ると、事前テストの正解数6.3問に対して事後テストが8.6問と上がっている。また、アンケート項目では全体的に+1.3点以上と比較的高い評価で、正確性や一貫性については+2.3点以上と高い数値を示している。海外旅行とSFの平均を比べると、興味において海外旅行+0.8点に対しSF+1.7点、有用性(個)は海外旅行+0.3点に対しSF+2.2点と大きな差が見られる。

5 考察

5.1 被験者による結果の差

表1や表2の各被験者の結果を見比べると、一人ひとりの結果に差があることがわかる。特に、読み終えた時間や、事前事後テストでの点数の上がり方に差が出ている。

事前テストにおいて点数の高かった被験者に質問したところ、既に自己情報量の内容はある程度理解しているとのことだった。そのため、教育効果をより詳しく知るためには実験前の調査で被験者の持つ知識を把握しておく必要があると考えられる。

5.2 テーマにおける違い

表3において、同じ学問で海外旅行というテーマとSFというテーマについて見比べると、興味の項目では海外旅行が+0.8点に対してSFが+1.7点、有用性(個)では海外旅行が+0.3点に対してSFが+2.2点と大きな差が現れていることがわかる。この点に関しては、自己情報量という学問が、SFというテーマに対して相性が良く、また、海外旅行のノベルゲームの流れが「海外のセミナーに行く」という現実的なものだったのに対してSFのノベルゲームでは「データセンターを管理するために自己情報量を学ぶ」という非現実的なものだったため、被験者の興味をより惹きつけた可能性がある。

5.3 学問とテーマの内容のバランス

本研究における実験では、被験者からの感想として、ノベルゲームの内容に学問の内容が少ないという指摘や、物語としてあまり面白くなく冗長であるという指摘を頂いた。本研究でのノベルゲームの生

表1 被験者実験の結果 (海外旅行)

被験者	興味	有用性 (個)	有用性 (他)	正確性	一貫性	事前テスト	事後テスト	かかった時間
A	-2	+2	+2	+3	+3	4	10	10分
B	+2	+1	+1	+3	+3	5	9	20分
C	+1	+0	+2	+2	+3	9	9	27分
D	+2	+2	+1	+3	+3	5	8	22分
E	+1	-2	-1	+0	-1	9	9	10分
F	+1	-1	+2	+3	+3	9	9	9分
平均	+0.8	+0.3	+1.2	+2.3	+2.3	6.8	9.0	16.3分

表2 被験者実験の結果 (SF)

被験者	興味	有用性 (個)	有用性 (他)	正確性	一貫性	事前テスト	事後テスト	かかった時間
G	+2	+1	+2	+3	+2	8	8	3分50秒
H	+3	+3	+2	+3	+3	7	10	7分18秒
I	+2	+3	+1	+3	+1	6	8	5分7秒
J	+2	+3	+2	+2	+3	5	7	7分41秒
K	+1	+2	+2	+2	+2	3	8	7分22秒
L	+0	+1	+2	+2	+3	6	8	3分36秒
平均	+1.7	+2.2	+1.8	+2.5	+2.3	5.8	8.2	5分49秒

表3 被験者実験の結果 (全体)

テーマ	興味	有用性 (個)	有用性 (他)	正確性	一貫性	事前テスト	事後テスト
平均 (海外旅行)	+0.8	+0.3	+1.2	+2.3	+2.3	6.8	9.0
平均 (SF)	+1.7	+2.2	+1.8	+2.5	+2.3	5.8	8.2
平均 (全体)	+1.3	+1.3	+1.4	+2.4	+2.3	6.3	8.6

成方法だと、事前に生成した講義資料から、学問の内容とテーマの内容を含めて一度で物語を生成するため、命令が複雑化してしまいバランスが悪くなったり両方の内容が薄くなったりしてしまうのだと考えられる。そこで、学問要素とテーマ要素の両方を強化するために、現在は単純化した命令を2つ与えて物語の内容を補強する添加形式を試みている。

添加形式については次節で説明する。

6 添加形式

教育的なノベルゲームに学問の内容が少ない問題や、テーマの内容に沿っていない問題が見られたため、物語のそれぞれの要素を強化することを目的に、物語に要素を段階的に添加する方式を検討する。

以下にその生成手順を示す。ChatGPT-4o-mini を用いて生成する。一例として、「SF」の物語に「自己情報量」の学問を添加する際の手順を示す。

- (1) SF要素のある物語を生成する。
- (2) 事前生成した学問の講義資料や小テストを入力する。
- (3) 物語に講義資料や小テストの内容を添加するように指示する。

具体的な生成例は付録Bに示している。

7 おわりに

本研究では、先行研究 [1] で生成した教育的なノベルゲームの実験を拡張し、2つのノベルゲームに対して各6人の計12人の被験者による評価実験を行った。その結果、いずれにおいても全被験者で事後テストの正解数が事前テスト以上となり、平均でも事前6.3問から事後8.6問へ向上した。このことから、評価者数を増やした場合においても、本研究で生成したノベルゲームに学習効果があることがわかった。また、アンケートでは正確性と一貫性が相対的に高い評価となったが、興味や有用性はテーマによって差が見られた。

一方で、被験者の感想から、学問的な内容や物語的な内容が薄くなる場合があることが明らかになった。これは、双方の内容を同時に生成するように指示を行うと、複雑な指示により学問要素と物語要素の両立が難しくなるためだと考えられる。そこで本研究では、物語要素と学問要素を段階的に付加する「添加形式」による生成を検討した。

今後は、従来方式との比較実験を行いたい。

8 謝辞

本研究はJSPS 科研費JP25K06544の助成を受けたものである。

参考文献

- [1] 赤田直弥, 村田真樹. ChatGPT を用いた教育的ノベルゲーム. 言語処理学会第 31 回年次大会発表論文集, pp. 1844–1849, 2025.
- [2] 村田真樹. ChatGPT を用いた小説関連研究：教育的小説の生成、マダミスの生成、推理小説での犯人推定. 言語処理学会第 30 回年次大会, pp. 3176–3181, 2024.
- [3] 宮本茂則, 高橋力斗, 馬淵浩希, 山田暉, 松原卓二, 森川幸人. Red Ram: 生成モデルを用いたオンデマンドなミステリーアドベンチャーゲーム生成システム. 人工知能学会第 38 回全国大会, pp. 1–4, 2024.
- [4] Manuel B. Garcia. Kinder Learns: An Educational Visual Novel Game as Knowledge Enhancement Tool for Early Childhood Education. *International Journal of Technologies in Learning*, Vol. 27, No. 1, pp. 13–34, 2020.
- [5] DeepGame - ChatGPT. <https://chatgpt.com/g/g-TzI2BlJPT-deepgame>.
- [6] Naoya Akada and Masaki Murata. Generating Educational Interactive Fiction Games Using ChatGPT. In *The 21st International Conference on Frontiers in Education: Computer Science and Computer Engineering (FECS'25) in The 2025 World Congress in Computer Science, Computer Engineering, and Applied Computing*, pp. 1–15, 2025.
- [7] Chung Kwan Lo. What is the impact of ChatGPT on education? A rapid review of the literature. *Education Sciences*, Vol. 13, No. 4, pp. 410:1–410:15, 2023.

付録 A ノベルゲームの生成例

本節の例では、自己情報量と海外旅行をテーマにしたノベルゲームを作る。

以下は、自己情報量について講義資料を生成させる手順1の入出力例である。

入力：自己情報量についての詳しい講義資料を作成してください。

出力：自己情報量についての講義資料 1. 自己情報量とは、自己情報量 (Self-information) とは、特定の出来事が発生したときに得られる情報の量を定量化したものです。(略)

以下は、物語の構想を生成させる手順3の入出力例である。入力には被験者プロフィールとして「海外旅行が好きな22歳の大学生」であるという情報を含めている。

入力：以上の講義資料や小テストの内容をすべて含むような、自己情報量初学者向けの物語を作ってください。(中略) まず、何章かに分けて全体の話の流れを書いてください。

出力：物語の流れ：『情報の旅—自己情報量を求めて』

第1章：旅立ちの朝 22歳の翔太は、長い間夢見ていたバックパッカーの旅に出る準備をしていた。出発前に、友人から「情報理論についての本」をプレゼントされる。(中略)

第3章：確率の森 翔太はリュカと共に、フランスの美しい森林を探索する。途中、リュカが自己情報量の公式を説明し始める。(略)

この後、章をさらに細かい部に分ける。

以下は、物語をフローチャート形式に変換する手順4の入出力例である。入力では few-shot としてフローチャートの例を与えている。

入力：次に、今生成した物語をノベルゲーム形式で読めるようにします。(中略) まず、3章1部のフローチャートを作ってください。(略) 以下にフローチャートの例を示します。

[シーン 1101] 学校のスタート 新しい学校での最初の日。授業を真剣に受ける → [シーン 1102: 授業での取り組み] 友達と遊ぶ → [シーン 1103: 友達との関係] サボって休む → [パッドエンド 1: 成績不振](略)

出力：フローチャート：『確率の森』第3章・第1部 [シーン 3101] 森の入口

翔太とリュカは、森林の入口に立っている。リュカは自然の中で確率の概念について話し始める。

1. 事象の確率について質問する → [シーン 3102: 確率の学び] 2. 自分の観察をシェアする → [シーン 3103: 自己観察] 3. 確率に興味がないと答える → [パッドエンド 1: 知識の遅れ](略)

以下は、シーン毎に詳しい内容を生成させる手順5の入出力例である。

入力：3章1部の、以下のシーンの内容を詳しく書いてください。

[シーン 3101] 森の入口

翔太とリュカは、森林の入口に立っている。リュカ

は自然の中で確率の概念について話し始める。(略)

出力：[シーン 3101] 森の入口

翔太は、リュカと共にフランスの広大な森林の入口に立っていた。「翔太、自然を観察することで、確率の概念を理解することができるんだ。」リュカは、周囲の木々や草花を指し示しながら言った。彼は、リュカにもっと詳しく聞きたいと思った。

選択肢 1. 事象の確率について質問する → [シーン 3102: 確率の学び] 2. 自分の観察をシェアする → [シーン 3103: 自己観察] 3. 確率に興味がないと答える → [パッドエンド 1: 知識の遅れ](略)

付録 B 添加形式による物語生成例

本節の例では、「自己情報量」で作成した物語に対し、「SF」の内容を添加することを目的とする。

まず、SF要素のある物語の生成を行う。

入力：SF要素のある物語を作ってください。何章かに分けて全体の話の流れを書いてください。矛盾がないようにしてください。

出力：物語タイトル：『時の狭間』第一章：異変

近未来の地球。人類は高度なテクノロジーを駆使し、日常生活を便利に過ごしていた。しかし、ある日、時空の歪みが発生し、世界各地で異常現象が起り始める。人々は時間が止まったり、逆行したりする現象に戸惑う中、主人公の若き科学者アヤは、異常の原因を探るために研究を始める。

次に、講義資料と小テストの入力を行い、学問の要素を読み込ませる。

入力：以下を読み込んで、はいと回答してください。

自己情報量に関する講義資料自己情報量 (Self-Information) とは、ある事象が発生することによって得られる情報の量を定量化したものです。(略)

小テスト1問題

1. 自己情報量とは何ですか？

A) 事象が発生する確率を表す B) 事象が発生することによって得られる情報の量 C) エントロピーの平均値 解答:B (略)

出力：はい。

最後に、学問の内容を物語に添加する。

入力：今生成した物語に、講義資料や小テストの内容を添加し、自己情報量初学者向けの物語にしてください。

出力：物語タイトル：『時の狭間 - 自己情報量の冒険』第一章：異変

近未来の地球。人類は高度なテクノロジーで生活を便利にしていたが、ある日、時空の歪みが発生し、世界各地で異常現象が起り始める。主人公の若き科学者アヤは、これらの異常が自己情報量に関連していることを直感する。自己情報量とは、ある事象が起ることによって得られる情報の量を測るものであり、エントロピーの概念とも深くつながっていることを思い出す。

学問の内容の添加命令を行うことで、元のSFの物語に自己情報量についての説明文章が追加されていることがわかる。少しではあるが教育的な内容が物語内に付加されていることが確認でき、学問要素と物語要素を同時に生成するよりも、要素を両立して物語の生成を行えることがわかった。