

ChatGPT を用いた教育的ノベルゲーム

赤田直弥¹ 村田真樹^{1,2}

¹ 鳥取大学 工学部 電気情報系学科 ² 鳥取大学工学部附属クロス情報科学研究センター
b21t2002c@edu.tottori-u.ac.jp
murata@tottori-u.ac.jp

概要

本研究は、村田 [1] による ChatGPT を用いた教育的小説の生成や、宮本ら [2] による ChatGPT によるノベルゲーム自動生成システムを参考にして、教育的ノベルゲームを作成した。ChatGPT を用いたノベルゲームには DeepGame[3] があるが、DeepGame はユーザ入力による逐次生成であるため、物語の全体像が見えにくい。そのため、初めに物語の全体像をフローチャート形式で生成してから細かい生成を行う。生成の際は、被験者が興味を持つ内容になるように、海外旅行や料理などの物語のテーマを設定する。

ノベルゲームの面白さや教育的効果についてアンケートと事前、事後の小テストを用いて被験者実験を行った結果、筆頭著者による被験者実験では興味や一貫性などについて概ね良いという評価になったが、小テストの結果はほぼ変わらなかった。また、物語のテーマによって評価が異なり、海外旅行や料理をテーマにしたものよりも、SFをテーマとしたもののほうが評価が高かった。著者以外の被験者による被験者実験では、被験者によって評価が大きく異なり、特に評価の低かった被験者 A は小テストの点数も低かった。被験者 A は短時間で物語の途中にあるバッドエンドに辿り着いており、内容を十分に理解できなかったからではないかと考えられる。読むのにかかった時間やシーンの選択で評価が変化すると考えられる。

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 [1]、またノベルゲーム [2] を半自動的に作る研究が確認されている。

他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[4] の研究では幼稚園生にビジュアルノベルゲームを用いた教育の効果が確認された。

本研究では、それらの手法を組み合わせることで教育的なノベルゲームを作ることを目的とする。

ChatGPT によるノベルゲーム生成システムとして、DeepGame[3] 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。このノベルゲームの問題点として、ChatGPT 側の質問と、プレイヤー側の回答をひたすら繰り返していく関係上、同じような内容が繰り返され、ノベルゲームとしての物語が進展しにくいという問題がある。その問題を、事前にフローチャートを生成させるというアプローチで解決する。

また、宮本ら [2] によるノベルゲーム生成システムの研究では、人手で事前に作成したプロットやミステリー執筆ガイドを用いており、これらを事前に作成せずにノベルゲームを作ることで、コストの削減も目的にしている。

2 関連研究

2.1 ChatGPT を用いた小説関連研究

村田 [1] は、ChatGPT を用いて教育的な小説の半自動生成を行った。はじめに、10 章ほどの簡潔なプロットを出力させ、その後それぞれの章について詳細な執筆を行わせた。それぞれの小説について、以下の 3 つの観点から人手評価を行った。

- 興味 (興味を持って読めたかどうか)
- 勉強 (各小説での学問について勉強になったか)
- 有用性 (教育に役立つ小説と思うか)

特に評価の高かった小説”心の選択”では小説生成前に学問に関する知見を生成させており、この操作によってより良い小説が生成できる可能性が示唆された。

2.2 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

宮本ら [2] は、殺人事件を推理するノベルゲームを生成するシステムである Red Ram を構築した。生成されるゲームの内容は、ユーザが殺人事件を担当する刑事として事件の解明を目指すというものである。

Red Ram は、テキストデータの生成に GPT-3.5 および GPT-4 を用いている。テキストデータの生成は、以下の4段階に分けて行っている。

- (1)プロットの生成
- (2)ストーリーデータの生成
- (3)詳細な設定の生成
- (4)会話シナリオの生成

プロットの生成を行う際、事件の設定としてユーザが「凶器の名称」「殺害現場の名称」「被害者の職業」「犯行動機」を入力する。これに加えて、「ミステリー執筆ガイド」と呼ばれるミステリー創作の手順書と、事前に作られた7種類の進行フローをランダムに1つ用いることにより、プロットが生成される。その後、ストーリーデータや詳細な設定などについても ChatGPT で生成する。

2.3 ChatGPT の教育における影響

Chung[5] 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。

2.4 地球温暖化に関するビジュアルノベルの生成及び評価

Mustafa ら [6] は、ChatGPT を利用して、ビジュアルノベルを通じて地球温暖化を認識するという文脈で、ビジュアルノベルゲームを評価及び生成する方法を探った。この研究は、ChatGPT を利用して2種

類のストーリーを生成した。1つには地球温暖化関連のキーワードや結末を指定し、もう1つはそれらを指定しなかった。それらのノベルゲームに対し、ChatGPT を用いて以下の5つの指標で10点満点で評価を行った。

1. 一貫性
2. インスピレーション
3. 話の流暢さ
4. 可読性
5. 単語の複雑さ

評価の結果、どちらのノベルゲームについてもほぼ評価が変わらなかったが、単語の複雑さという指標においてのみ、キーワードを指定したノベルゲームのほうが評価が高くなった。どちらのノベルゲームの評価についても、標準偏差は0.2以下と非常に小さかった。

3 本研究の考え方

村田 [1] の研究では、事前に知見を生成させることで教育的小説の評価が上がっている。また、宮本ら [2] の研究では、ミステリーの内容をユーザの入力により決めている。そのため、本研究にもこれらの手法を取り入れて教育的なノベルゲームの生成を行う。また、宮本らの研究では事前に人手で進行フローを作成するが、本研究ではより簡単にノベルゲームを生成するために、その手順を省いて ChatGPT で進行フローを生成する。

4 提案手法

4.1 ノベルゲームの生成方法

この研究では、ChatGPT-4o-mini を用いてノベルゲームを生成する。

この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、被験者プロフィールを入力し、被験者の好みに合わせた物語にする。
- (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、A 節に記載する。

4.2 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。詳細を B 節に示す。

- 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。

今回は、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。

5 実験

5.1 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 [1] の研究に倣い、以下のそれぞれについて-3 から+3 までの 7 段階で評価する。

- 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)

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、以下についても-3 から+3 までの 7 段階で評価する。

- 正確性
- 一貫性

5.2 実験結果

筆頭著者が 3 つのノベルゲームについて被験者実験を行った。学問は、どのノベルゲームでも自己情報量としている。その結果を表 1 に示す。

「自己情報量の発見」の実験の際、筆頭著者が誤って小テスト 20 問をノベルゲーム生成後に解いたため、「自己情報量の発見」の小テストの結果は平均の計算には入れていない。

興味や有用性については平均+1、正確性や一貫性は平均+1.3 を示した。海外旅行をテーマにした「自己情報量の発見」については正確性や一貫性の評価が低く、SF をテーマにした「情報の宇宙」では高くなった。小テストの結果は事前事後でほとんど変わらず、1 点下がっていた。

また、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。その結果を表 2 に示す。

小テストの点数は事前テストで 6 点だったものが事後テストで 7.7 点となり、上がっている。平均として普通より良い結果となったが、被験者ごとに結果が異なり、被験者 A の評価は正確性以外-2 や-3 と低かった。

読むのにかかった所要時間は、被験者 A は 10 分、B は 20 分、C は 27 分だった。

6 考察

6.1 筆頭著者による実験

表 1 の筆頭著者による実験では、テーマによって結果に差が出ており、海外旅行や料理をテーマにしたノベルゲームよりも SF をテーマにしたノベルゲームのほうが評価が高くなり、興味や正確性、一貫性の評価が他よりも 1.2 点高くなった。海外旅行や料理をテーマとしたノベルゲームでは、海外に行ったり料理をするという物語の流れはあるが、その後それらとは関係なく自己情報量について学び続けるという流れになっていたのに対し、SF のノベルゲームでは電子世界のデータセンターを管理するために自己情報量を身に付けるという内容で、上手

表 1 筆頭著者による実験の結果

タイトル	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
自己情報量の発見	海外旅行	+1	+1	+1	-2	-1	N/A	(17)
料理の確率とユニークな創造	料理	0	+1	+1	+3	+2	9	8
情報の宇宙	SF	+2	+1	+1	+3	+3	10	9
平均		+1	+1	+1	+1.3	+1.3	9.5	8.5

表 2 被験者実験の結果

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
A	-2	-2	-2	+3	-3	4	5
B	+2	+1	+1	+3	+3	5	9
C	+1	0	+2	+2	+3	9	9
平均	+0.3	-0.3	+0.3	+2.7	+1.0	6	7.7

く内容が合わさっていたため、評価が高くなった。

また、小テストの結果は事前テストよりも事後テストのほうが点数が1点下がっているが、これは筆頭著者に自己情報量の知識が既にあっただけで、事後テストでノベルゲームの内容として含まれていない知識が出題されたためであると考えられる。

6.2 被験者による結果の差

表2の被験者実験では、被験者によって評価に差が出た。特に、被験者Aの結果は正確性以外-2や-3となっており低かった。

その理由として、被験者によって所要時間や、ノベルゲームで選んだ選択肢が異なることがあげられる。実験後、ノベルゲームでどの選択肢を選んだかを聞いた結果、被験者Aはノベルゲーム中盤でバッドエンドへ向かう選択肢を選び、途中で物語が終わってしまったとのことだった。そのため、十分に物語を読み進めることができず、評価が悪くなったと思われる。

また、全ての被験者について事後テストの点数が事前テストの点数以上になっている。その理由について、特に点数が上昇している被験者Bに質問したところ、ノベルゲーム内で計算方法を学べたため、事後テストを解きやすくなったとしている。また、事前テストでも9点を取得した被験者Cは、元々その分野についての知識を持っていたとのことだった。

6.3 内容の正確性

ノベルゲームの中に、学問として間違っていると考えられる箇所があった。サイコロの目から自己情報量を求めるというシーンで、サイコロの目の出る確率はすべて等しいのに、主人公は「出た目が1~3の自己情報量を計算すべきか、4~6の自己情報量を計算すべきか」と考え始めている。また、ここで出た目が1~3であるときの自己情報量が「低確率の自己情報量」、4~6のときの自己情報量が「高確率

の自己情報量」と表現されている。

[シーン2202]サイコロの結果を観察
リョウはサイコロを振り、その結果を観察する瞬間が訪れた。結果の目は「3」だった。(略)
まず、リョウは出た目が1~3の場合の自己情報量を計算するべきか、4~6の場合の自己情報量を計算するべきかを考えた。出た目が3であるため、彼は1~3の範囲に含まれていることを確認した。彼は次の計算式を心の中で反芻する。

$$I(x) = -\log_2(P(x))$$

リョウはこの計算を通じて、自己情報量の概念を深く理解しようとしていた。
出た目が1~3の確率を計算する → [シーン2204: 低確率の自己情報量]
出た目が4~6の確率を計算する → [シーン2205: 高確率の自己情報量]
計算をしない → [バッドエンド2: 理論の理解不足]

6.4 似たような内容

ノベルゲーム中で似た内容が繰り返されるシーンがあった。その例をB節に示す。

7 おわりに

本研究では、教育的なノベルゲームを生成した。

これまでのノベルゲームの生成手法としては、Deep Game という手法や、Red Ram という手法があったが、逐次生成のため終わることができないという問題や、人手で事前にフローチャートを生成しなければならぬという問題があった。そこで、事前にフローチャートを ChatGPT で生成し、そのフローチャートを用いてノベルゲームを生成することで、先行研究よりも簡単にノベルゲームの作成ができるようになった。

また人手評価では、興味や有用性は低い結果となったが、どの被験者も事後テストの点数が事前テストより上がっており、平均として1.6点ほど上がっているため、ノベルゲームの学習効果はあると言える。

今後は文脈型ノベルゲームの生成や被験者実験を行いたい。また、興味や有用性の向上のためのアプローチも考えたい。

参考文献

- [1] 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.
- [2] 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.
- [3] DeepGame - ChatGPT. <https://chatgpt.com/g/g-TzI2BLJPT-deepgame>.
- [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, pp. 13–34, 10 2020.
- [5] Chung Kwan Lo. What is the impact of ChatGPT on education? A rapid review of the literature. **Education Sciences**, Vol. 13, p. 410, 4 2023.
- [6] Mustafa Can Gursesli, Pittawat Taveekitworachai, Febri Abdullah, Mury F Dewantoro, Antonio Lanata, Andrea Guazzini, Van Khôi Lê, Adrien Villars, and Ruck Thawonmas. The Chronicles of ChatGPT: Generating and Evaluating Visual Novel Narratives on Climate Change Through ChatGPT. pp. 181–194, 11 2023.

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。物語の1つの章が長い物語になると、フローチャートのシーン数が多くなり、矛盾のない生成が難しい。そのため、1つの章全体を部分けしたものを ChatGPT の新しいセッションに入力として与えた後、フローチャートにしている。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。

教育に役立つノベルゲームのためのフローチャート

こちらは教育に役立つノベルゲームのためのフローチャートです。

[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 → [バッドエンド 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 → [バッドエンド 1: 知識の遅れ]

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 → [バッドエンド 1: 知識の遅れ]

B 似たような内容

以下の例は、リョウが実生活に自己情報量を役立てようとしているシーンである。

自己情報量の例として、「スポーツの試合で、勝つ確率が低いチームが勝ったら価値が高いが、勝つ確率が高いチームが勝っても価値は低い」という例が繰り返されている。

[シーン 2212] 実生活への応用彼はまず、スポーツの試合を例に挙げた。勝つ確率が低いチームが勝った場合、それは驚きの出来事であり、得られる情報の価値が非常に高い。逆に、勝つ確率が高いチームが勝つ場合は、ほとんどの人が予想できるため、情報の価値は低い。このように、日常の出来事を通じて自己情報量の概念を理解できると、リョウは感じた。

[シーン 2216] 知識の実践

まず、彼は身近な例として、友人とのスポーツの試合を思い出した。試合の結果を予測する際、リョウは各チームの勝利確率を考え、勝つ確率が低いチームが逆に勝った場合に得られる情報の価値を評価することを考えた。

[シーン 2223] 実生活への応用

まず、リョウは友人との会話の中で、自己情報量を活用できるシチュエーションを見つけることにした。「例えば、スポーツの試合での予測だ。勝つ確率が低いチームが勝った場合、私たちはそれに対して強い驚きを感じる。これは、自己情報量が高い出来事だと言えます」と、リョウは友人たちに話し始めた。