

VR-FlowGen : 読者をフロー状態に導くための拡散モデルを用いた VR 空間生成システムの開発

瀧澤 琴萌[†] 中村 亮太[†]

武蔵野大学データサイエンス学部[†]

1. 背景

近年、スマートフォンやタブレットの普及に伴い動画視聴が主流となったことで若者の読書離れが急速に進んでいる。実際に全国大学生協連[1]が行った1日の読書時間に関するアンケートに対して、0分と回答する大学生の割合は2014年から2017年にかけて急激に増加していた。東京都教育会[2]は、読書離れが進むことは読解力の低下を招き、長い文章を書く力を低下させ、意味の通じない文章しか書けなくなっていることを提言している。意味の通じない文章しか書けないことは、自分の伝えたい内容を適切に伝えることができなくなり、受け取り側に間違った認識をさせる恐れがある。そのような問題を防ぐためにも読書習慣は重要であると考ええる。

そこで、著者らは読書のフロー状態を擬似的に体験させ読書の成功体験を繰り返すことが、読書に対するハードルを下げ、容易に取り組みやすくなることにつながるのではないかと考え、画像生成モデルと大規模言語モデルを用いて小説の文章からVR空間の背景画像を生成し読者をフロー状態に導く手法を提案した[3]。ここで述べるフロー状態とは[3]で定義したように「物語への注意の集中」、「自己の意識や周囲に対する注意の減退または消失」、「物語世界の鮮明なイメージ化」、「登場人物の感情をそのまま体験する共感」、「登場人物の状況に対する自分自身の感情としての感情移入」、「上記の体験の総体として感じられる物語を現実のように感じる体験」の6つの要素を読書中に体験している状態のことを指す。提案手法では情景描写用テキスト抽出ならびにVR空間の背景画像生成を実現した。本稿ではこれらの手法をシステム「VR-FlowGen」として実装することを行った。また、提案手法の実装に向けた課題として、VR空間の背景画像の切り替わりによるVR酔いの対策や文章の提示方法の確立が挙げられたため

VR-FlowGen : A Diffusion Model-based VR Reading Space Generation System for Inducing Flow State

Komo Takizawa[†], Ryota Nakamura[†]

[†] Faculty of Data Science, Musashino University

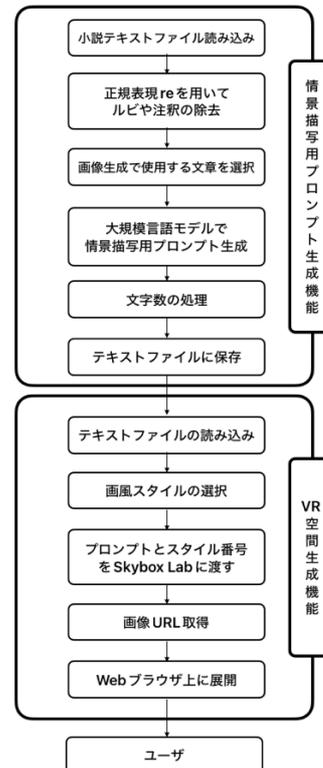


図1 VR-FlowGenの全体図

VR-FlowGenのシステム概要ならびに適切なVR空間の背景画像の生成・切り替えのタイミングについても述べる。

2. VR-FlowGen

VR-FlowGenの全体図を図1に示す。本システムは、情景描写用プロンプト生成機能とVR空間生成機能の二つで構成されている。ユーザは二つの機能を用いて生成されたVR空間の背景画像をVRHMD (Virtual Reality Head Mounted Display) を用いて閲覧し、読書をする。

2.1. 情景描写用プロンプト生成機能

本機能は、小説の文章から画像生成のためのプロンプトを生成する機能である。読み込んだ小説のテキストファイルから、ルビや注釈を除いた本文のみを抜き出す。この時、形式段落または意味段落の単位で本文を分割する。本文を

そのまま画像生成に用いると、比喩表現や感情描写がある文章の場合は適切な VR 空間の背景画像が出力されないため、GPT-4 を用いて画像生成に適したプロンプトを生成する必要がある。本機能で生成されたプロンプトは VR 空間生成機能への受け渡しのため、標準化された命名規則でテキストファイルとして保存される。

2.2. VR 空間生成機能

VR 空間生成機能は、2.1 節で生成されたプロンプトを使用して VR 空間の背景画像を自動生成ならびに Web ブラウザ上に表示するための機能である。なお、本システムで使用している画像生成モデルの Skybox Lab は、39 種類の異なる画風から好みの画風を選択することが可能である。本機能では、ユーザによって選択された画風の識別番号と情景描写用テキスト抽出機能で生成されたプロンプトを、画像生成モデルが動作するサーバに送信し、生成された画像を受け取る仕組みになっている。受け取った画像は A-Frame ライブラリを用いて作成した WebXR コンテンツのテンプレートに自動的に配置され、VRHMD「Meta Quest2, 3」に標準で搭載されている Web ブラウザで閲覧することが可能である。なお、小説のテキストについては空間に表示せず、Web Speech API の音声読み上げ機能を用いて文章内容をユーザに提供することを想定している。

3. VR 読書に適した VR 空間の背景画像の生成・切り替えタイミング

本システムをサービスとして提供する場合、文章ごとに複数の画像を適切なタイミングで切り替えることが必要となる。そこで、著者らは形式段落を画像切り替えタイミングと仮定し、改行コードが事前に含まれていることを利用して、自動的に形式段落ごとに分割、VR 空間の背景画像の生成を行っていた。しかし、形式段落で分割すると小説によっては文章が短すぎるものや、台詞のみの場合がある。これは、画像生成の精度が下がる可能性があるとともに頻繁に画像を切り替えることはユーザに不快感を与える可能性がある。この問題を防ぐため、複数の形式段落を含んだ意味の変化によって分けられる段落である意味段落を画像切り替えタイミングとする方法を考えた。

意味段落を切り替えタイミングにするにあたり、意味段落の分割がされていない小説でも自動的に認識できる必要がある。そのため、意味

段落の認識を大規模言語モデルで行えるか調査を行った。

調査方法として事前に意味段落が判明している小説「蜘蛛の糸（芥川龍之介著）」を、大規模言語モデル「Microsoft Copilot」に意味段落ごとに分割させ適切な形として出力されるかを確認した。原則としてプロンプトには、結果のみ日本語で出力すること、分割した場所で改行を行い全文表示することを記載しているものとする。

調査の結果、プロンプトによって精度が不安定になることがわかった。「下記文章を意味段落ごとに分けてください」というプロンプトの時は形式段落と同じ形で分割が行われてしまったが、「下記文章を情景描写が切り替わるところで大まかに分けてください」というプロンプトの時は、実際の意味段落の分割より多いが複数の形式段落を含んだ分割が行われていた。

4. 今後の予定

本稿では画像生成モデルと大規模言語モデルを用いて小説の文章から VR 空間の背景画像を生成し、読者をフロー状態に導くシステムである VR-FlowGen の実装方法について述べた。また、VR 空間の背景画像を切り替えるタイミングについて検討した結果、小説のテキストを意味段落に基づき分割し、その単位で切り替えることが適切であることが分かった。しかし、意味段落の認識を大規模言語モデルを使用して自動的に処理ができるか調査を行ったところプロンプト文によって精度が変化することが確認された。

今後の展望としては、意味段落の認識方法を検討するとともに、本システムを読書のフロー状態に近づけるための文章の提示方法や VR 酔いを緩和させる画像の切り替え方法の確立を予定している。

参考文献

- [1] 第 54 回学生生活実態調査の概要報告：Available : <https://www.univcoop.or.jp/press/life/report54.html>.
- [2] 定言 126 日本の若者の読解力か危ない：Available : <https://www.t-kyoikukai.org/teigen/teigen126/teigen126.pdf>.
- [3] 瀧澤琴萌, 阿部 広夢, 中村 亮太：読者をフロー状態に導くための拡散モデルを用いた VR 空間生成手法, マルチメディア, 分散, 協調とモバイルシンポジウム 2023 論文集, Vol. 2023, pp. 1078-1083, 2023.