

ニューラル言語モデルを利用した映画鑑賞支援

谷 悠翔† 牛尼 剛聡††

†九州大学 芸術工学部 芸術情報設計科 〒815-8540 福岡県福岡市南区塩原 4-9-1

††九州大学大学院 芸術工学研究院 〒815-8540 福岡県福岡市南区塩原 4-9-1

E-mail: †tani.haruto.025@s.kyushu-u.ac.jp, ††ushiyama@design.kyushu-u.ac.jp

あらまし 映画の発展に伴い観客側の作品を鑑賞する姿勢も多様化している。アクション、ホラーといった直接的な刺激を楽しむために娯楽作品として映画を見る場合や、映画の意味を深く考え、自己省察を行う映画に対する洞察が深いという姿勢も存在する。映画の意味を深く考える場合、さらにその考えを深化させる方法として、映画の感想を言語化することが挙げられる。言語化により映画で感じたことを可視化し、映画の俯瞰的な理解が促される可能性がある。しかし、映画の感想を言語化することは容易ではない場合がある。本論文では既に映画の感想として言語化され、ウェブ上に投稿された映画レビューを利用して映画鑑賞者の言語化を支援し、映画理解の支援を行うことを目的とする。具体的に、映画レビューを巨大言語モデルに学習させ、その言語モデルは映画鑑賞者の感想を基に、鑑賞者の感想を詳細に言語化したものを出力する。

キーワード 映画レビュー, ニューラル言語モデル, テキスト生成, 鑑賞支援

1 はじめに

映画は上映が初めて行われて100年以上もの間発展し続け、モノクロの無声で単に情景を映し出すだけだったものが、様々な作品のジャンルや撮影技術の向上などに伴い、観客に高いクオリティの作品を数多く提供されている。

映画の発展に伴い、観客側の作品を享受する姿勢も多様化している。Oliverら[1]によると、観客の中には映画を単なる娯楽作品と捉えアクションやコメディなどの直接的で単純な楽しさを得ることによる気晴らしを最終目的とする人がいる。その一方で映画を見る際の観客自身の感情的・認知的反応を重視し、悲劇的であったり、不条理的な映画を見ることで自己省察、人生の意味の探究などを行う、映画に対しての思慮が深い観客も存在する。Oliverら[2]はこの観客のように映画の意味を深く考え、その映画体験に触発された自身の感情・思考をさらに深めようとするような状態になることを「鑑賞」と定義している。以降、本論文では鑑賞をする観客のことを鑑賞者と呼ぶ。

鑑賞者が映画の意味や自身との関連を深く考える際に、パンフレットなどに掲載されている制作者側の制作意図等を読む以外にも、さまざまな方法が考えられる。

一つに対話型鑑賞[3][4]がある。この手法では複数の鑑賞者が進行役に促されながら作品を見て気づいたこと、感じたこと、連想されることを発話していき、鑑賞者同士の発見を深めていくものである。一般的に進行役を務める人が鑑賞者とともに鑑賞者と作品の関係を対話形式によって創造する。つまり対話に参加した鑑賞者全員で作品の意味や価値を創出していくことも考えられる。この鑑賞法では、作品に抱いた印象を改めて言葉にすることで、鑑賞者の作品の捉え方を明らかにし、他の鑑賞者の新しい視点や思考から鑑賞を深めることができる。

また、鑑賞者が自身の映画にまつわる感想を詳細に言語化す

ることは鑑賞体験を俯瞰的に理解することに役立つ。竹内ら[5]は口頭や記述による言語的表現には認識や感覚が明確に意識される働きがあると述べている。例えば、ユーザが明確に内容を理解できない作品に関して、ユーザが感じ取った作品のメッセージ性や描写ごとの解釈が言語化されれば、映画に対する理解が明確になり、わかったつもりになっていた部分を解消できる可能性がある。

これら二つの方法に共通しているのは、鑑賞者が自身の感想を発話なり記述する形で言語化していることにある。このように鑑賞者が言語化を行おうとする際には、同じ映画を見たことのある身近な人物とその内容について語り合う他、レビューサイトを利用する方法が考えられる。

しかし、実際には映画の閲覧者数とその映画へのレビュー投稿者数には差がある。例えば映画情報サービス Filmmarks のサイト内において、映画『ForrestGump』の視聴者数は約18万人であるのに対して、レビュー投稿数は2万件ほどとなっており、レビュー数の方が少ない。レビューは多くのサイトユーザにとって見るために存在しているものであり、それをユーザ自らが書くものではないと認識しているユーザが多いと考えられる。

そこで、本論文では文書生成技術を利用して、感想の言語化を支援するシステムを提案する。ニューラル言語モデルを利用した自然言語処理の技術の発達は目覚ましく、文書を生成するシステムとして GPT-3 [6] や BERT [7] などが主流となっている。本論文で提案する手法もこうした言語モデルを用いて文書生成を行う。

本論文では鑑賞者が映画感想を詳細に言語化して映画に対する理解をさらに深めることを支援するためにユーザが簡単に感想を言語化できるシステムとして、鑑賞者の短い感想文の入力に基づいて、詳細な感想文を生成する手法を開発することを目的とする。提案手法では代表的な言語モデルの一つである

GPT-2 [8] を映画レビューを用いてファインチューニングしたのち、鑑賞者の感想をモデルに入力し、出力としてその感想文から連想される次の文を生成する。鑑賞者は自身の感想から生成されたいくつもの文からその感想に関連するものを選択し、ブロックを積み上げていく要領で文を作る。これにより言語化する際の負担を軽減する。さらに、被験者による主観評価の評価実験によって提案手法の有効性を示す。

本論文の構成は次のとおりである。2章で関連研究について述べる。3章では提案手法についての説明を行う。4章では実際にレビューを生成した結果とその評価について述べる。5章では本論文のまとめと今後の課題について述べる。

2 関連研究

これまでウェブレビューからニューラル言語モデルを用いた文章生成を応用する研究が行われてきた。ここではそれらの研究のうち、提案手法に関連するものについて述べる。

橋爪ら [9] は宿泊施設サイトに投稿されたレビュー文書に対する返信文を生成するためにレビュー文書を自動変換する手法を提案している。この手法では事前学習済みの BERT [7] を教師データとして実際にサイトに投稿された宿泊施設レビュー文書とそれに対するサイト側の応答文を用いて学習させた。学習を行った BERT により変換の対象となるレビュー文書の分散表現を取得し、レビュー文書をサイト側の応答文へと変換している。評価実験ではどのようなレビュー文書にも当てはまる応答文書ではなく、レビュー文書の内容を汲み取りその話題について言及した応答文が生成されたとしている。

この手法では言語モデルにレビュー文を入力した際にその文書の内容に言及している文書を生成するという点で本研究と類似している。しかし、本研究ではレビューを受け取ったサイト側のために返信文を生成するのではなく、レビューを書くユーザー自身が自己の感情等を整理するため入力した文書の内容を詳細に言語化する補助を行うことを目的とする。

3 提案手法

3.1 提案手法の概要

本論文で提案する手法の概念図を 1 に示す。本論文では映画作品の理解を深めるためにユーザーが抱いた映画に対する印象の言語化を支援する。手法は以下の順で行い、これらはウェブアプリとして実装する。

1. 映画レビューを学習した言語モデルにユーザーがシンプルな感想文を入力する。
2. システムが感想の続きとなる複数の文章を生成する。
3. システムは生成した文章をそれぞれ句点ごとに区切ったものをユーザーに提示する。
4. ユーザーは句点で区切った文の中から、自身の感想に近いものを選択し、文ごとのブロックを組み替えながら文章を作成する。

この手法では、生成文章を句点ごとに区切った文をブロック

のように扱う。ユーザーはウェブアプリの画面上でそのブロックを組み替えて自身の趣向に沿った文章を作成する。これによりユーザーはある程度意味のまとまった文から文章を作成することができ、ユーザーが文章を書く際の負担を減らすことができる。

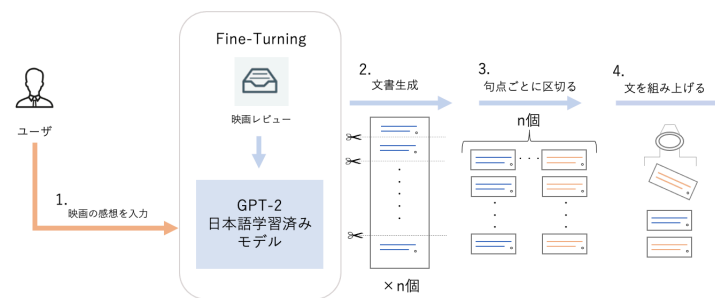


図 1 提案手法の概要

文書生成モデルとしては rinna 社が提供する GPT-2 日本語学習済みモデル [10] を使用する。

このモデルの次に登場した GPT-3 は GPT-2 の 100 倍のパラメータを持つが日本語によって学習されたモデルが公開されていないため GPT-2 日本語学習済みモデルを使用する。

3.2 GPT-2 モデル言語モデル

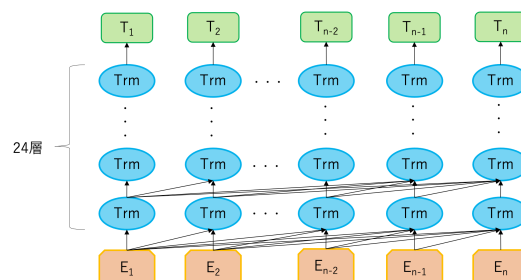


図 2 GPT-2 の構造

GPT-2(Generative Pretrained Transformer 2) は Transformer とよばれるニューラルネットワークアーキテクチャーを基にした機械学習モデルである [8]。GPT-2 では 800 万文書もの大規模な言語コーパスを用いて事前学習を行い、特定のタスクに特化した教師あり学習は行わずに様々なタスクに応用できる汎用的なモデルである。

GPT-2 は Transformer の decoder 層のみを使用し文章を生成する。その仕組みを説明するために、単語列 (x_1, x_2, \dots, x_n) で表される文章 (X_1, X_2, \dots, X_n) を仮定する。GPT-2 の言語モデルは式 (1) で表される。

$$p(X) = \prod_{t=2}^n p(x_t | x_1, \dots, x_{t-1}) \quad (1)$$

式 (1) が示すように GPT-2 はある時刻 t の単語を t 以前の単語からのみ推定するように学習する。以上の内容を図で表すと図 2 のようになる。これは Masked-Self-Attention 機構と呼

ばれるものである。これにより注目している単語以降は計算に参考せず、今まで見ていた単語だけを使って、確率的に次の単語を予測し、自己回帰的に入力トークンを学習しもっともらしい文章を生成することができる。

このモデルを用いて映画レビューを学習させる。学習させる映画レビューの種類は一つの作品に限定する。これにより生成される文章にユーザの期待していない作品情報が含まれることを防ぐ。

このモデルを通してユーザが入力した感想から、ユーザの趣向に関連すると考えられる文章を生成する。

3.3 文章作成支援

図 3.3 に具体的にユーザが文章を構成していく過程を示す。2.3 の言語モデルにより生成した複数の文章を句点で区切り、文のブロックをユーザに提示する。この文のブロックを使ってユーザは一つの文章を作成する。

例えば図 3 のようにユーザがユーザはそうした、自身の感情や感情の起因を詳しく説明したり、感想の内容に関連するような生成文のブロックを選択していく。

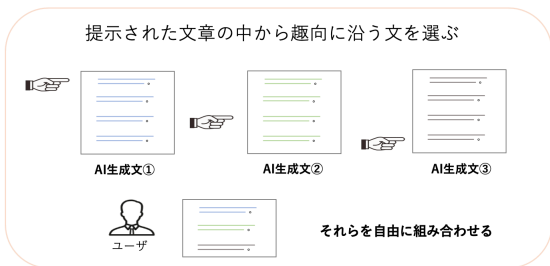


図 3 ユーザの文章作成過程

一般的に文章を書く際に行うこととしては図 4 のような過程を辿ることになる。まず言語化したい内容にふさわしい単語を選択し、選んだ複数の単語を用いて文を形成し、それらがいくつか集まったものが文章ができる。本論文では文を形成する工程は言語モデルが行うため、ユーザはすでに出来上がった文を基に文章を作る。そのためこの提案手法によりユーザが映画に対する印象を言語化する際の負担を減らすことが期待できる。



図 4 文章を書く過程

4 プロトタイプ

提案手法に基づいたシステムプロトタイプをウェブアプリとして実装した。ウェブアプリの実装には Python のフレームワークである Flask [11] を使用した。Flask はサーバーとウェブアプリケーションを繋ぐ共通のインターフェイスを Python で定義したものである。これにより Python で実装した GPT-2 言語モデルを使用したユーザインターフェイスを持つアプリを構築する。

4.1 インターフェイスの説明

図 5 は開発したアプリを起動した際に基にユーザに提示される画面の例である。ユーザは画面中央にある「感想を入力してください」と書かれ、グレーの枠で囲われている入力フォームに感想を入力する。その後、ウェブアプリ画面は図 6 に示す画面に遷移する。遷移した画面左には GPT-2 より生成された複数の文章が生成され、それらが句点区切りで一文ずつ縦に連なってユーザに提示される。一つの文が一つの文を形成している。

ユーザはこれらの文の中から自身の趣向に沿ったものを選択し、図 7 のようにブロックを画面右側の空欄の部分にドラッグアンドドロップする。ユーザはこの右の欄で文を組換えながら自身の文章を組み上げていく。

文を組み立ていく中で接続詞が必要になった場合、対象の文をクリックすると図 8 のようにその文を編集することができる。そこで接続詞を付け足すことができる。

4.2 バックエンドの処理

バックエンドの概要について説明する。まず、図 5 のフォームに入力された内容をサーバーは受け取る。サーバは受け取った内容を GPT-2 モデルに入力する。サーバーが GPT-2 が生成した文章を受け取ったら、サーバはウェブ画面を図 6 に遷移させる。図 7 で行うブロックのドラッグアンドドロップや、接続詞の付与の処理は JQuery, および JQueryUI を用いて実装した。



図 5 スタート画面

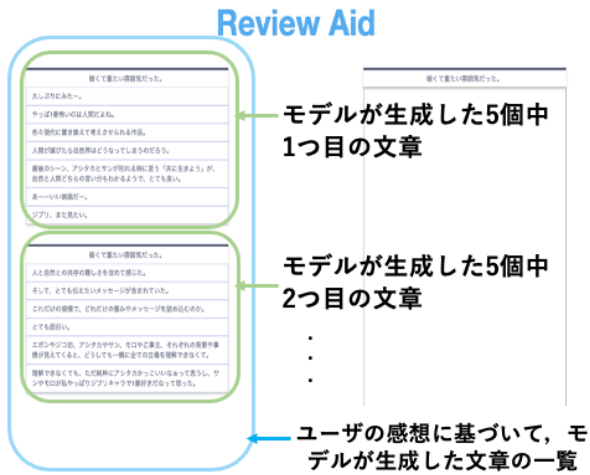


図6 遷移画面



図7 操作説明

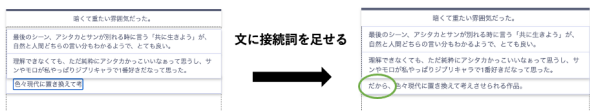


図8 操作説明

5 評価実験

5.1 実験方法

評価実験として、提案手法の内容をウェブアプリとして実装しそれを実際に被験者に試してもらい、それにより被験者の映画に対する印象を言語化できていたか、映画理解は深まったかを目的として行った。

5.2 生成結果

実験に使用する映画は「もののけ姫」とした。この映画は2020年に映画館での再上映も行われており、世間の認知度が高く、投稿されているレビューも数多くあるため実験に最適だと判断した。

映画レビューは Filmmarks 映画よりスクレイピングしたものを使用した。・記号の削除・不要な改行の削除・25文字以下のレビュー文書の削除・「鑑賞」および「観賞」の記載のあるレビュー文の削除収集したレビューは以上のクリーニングを行っ

た。収集したレビュー文書の中には「面白かった」や「好みでなかった」だけで記されたメモ書きのような感想が含まれている。短い文章では映画自体の内容に多くは言及できていない。それらを学習させた際、生成文が単調なものになり本研究の目的から外れてしまうため文字数が25文字未満のものを削除した。またレビュー文書中に書き込まれた「2020年〇〇映画館にて鑑賞」などの作品について一切言及していない「鑑賞」を含む文章は削除し生成文にそれらの冗長的な文が含まれないようにした。結果、使用する映画レビュー文書は9506件となった。

表1 文書生成時のハイパーパラメータ

top-k	0.95
top-p	40
num	5

文書生成時の GPT-2 の生成時の設定を表1に示す。top-k はモデルが次の単語を生成するとき確率の高い上位 k 個の単語からランダムにサンプリングするものである。また top-p は top-k で候補にした k 個の単語の中からそれらの確率の合計が p になるように動的に決めたものである。よって本研究では次単語を予測するとき確率の高い上位 40 個をサンプリングし、その中で次単語の確率の高いものから順に確率の合計が 0.95 になるまでサンプリングする。その中からランダムに選んだものを次の単語とする。そのため入力文が同じであっても生成文は毎回異なるものが出力された。num は生成する文書の数である。文書は 100 200 文字数の間で生成した。

「雰囲気为重めでテーマが暗い。」と入力したときに実際に生成した文書を表2に示す。候補1は入力文の内容から派生した文書になっている。作品全体のテーマから話を広げ、作品の解釈を述べる事ができていた。また「テーマ」という入力を受け、他の文書に比べて映画の内容についての具体的な表現が多く使われていた。候補2は「共生」というテーマがテーマとはいえ…」から矛盾した内容について述べていた。また日本語もやや不自然なものであった。候補3は「怖くて」や「音楽も切なかったり」の部分が入力文から派生した内容を表していた。また「…歳をとると見え方が違うもんだなあ。」の部分は前半の内容から逸脱しすぎない範囲で情緒的な文を生成していた。候補4は文章全体としてまとまりがなく、不自然な点が散見された。文章に前後のつながりがないためかなり突飛な文章になってしまっていた。候補5は文章のつながりも薄く、「…ずっとアシタカはサンを思おうって思えた」など日本語として不自然な部分があった。

また GPT-2 で生成した文章が被験者の映画印象に関する言語化に寄与できているかを検証するために、提案手法による実験と同じ被験者に「もののけ姫」以外の映画についても比較手法による実験を行ってもらった。比較手法では言語モデルによる文章生成を行わず、映画レビューサイトから収集したレビューの原文を被験者に提示した。これにより、被験者にとって言語モデルで生成された文章とレビューサイトの原文のどちらが映画印象の言語化の際に有益であるかを検証した。比較対

表 2 生成文

入力文	
候補 1	昔の映画だけれども今の環境問題に当てはめるとどう感じるのかと考えてしまった。人間の欲のせいで自然が壊されていく。それを嘆くアシタカとサンに涙が零れた。共に生きよう。そう言いたかった。昔の映画でありながら現代にも通ずる強いメッセージを感じた。アシタカはエボシ御前を守るために動く。サンも森と人間の共存を守るために動く。その姿に自然と人間は歩み寄ることはできないのかと苦しんだが、最後にはお互いの存在を認め合い共存への道を模索する二人を見て少し明るい気持ちになった。
候補 2	エボシが悪者というわけではなく、すべてエボシに原因がある。それがまた複雑。アシタカはエボシも自分と同じように、森や自然を守る存在だと考えていて、ただ単に人間嫌いというわけではない。でも、森を人間が侵していくのを目にして、なんとも言えない気持ちになった。「共生」というテーマがテーマとはいえ、人間側から見れば、いくら自然が正義であるといえども、森を破壊するのは悪なのだ。それを善悪の線引きなしで描いてしまうのが、なんだか悪者がいない、すごく難しいということを思わされた。’
候補 3	小さい頃は見れなかった、怖くて。でも今なら見れる。映像が繊細だし、音楽も切なかったり、ハッピーだったりすると同時に、人間の怖さもあると思う。昔は、ああ、アシタカかっこいい、いい男だなあ、くらいにしか思ってたなかっただろうな。歳をとると見え方が違うもんだなあ。ヤックルかわいい。それにしても久石譲は良いなあ。
候補 4	映像が綺麗で迫力があつた。メッセージ性が強すぎてメッセージと一緒にマイナスイオンをもらえた。山犬が可愛かった。人間と自然の共存について考えさせられた。音楽がよかった。久石譲の音楽が素晴らしすぎてびっくりした。主題歌も頭から離れなかった。自然を大切にしたいと思った。1 番のメッセージはヤックルの”だいだらぼっち”になりたい。誰にもなれず、誰も救えず、それでも生きていく。理想像が現実になるのは辛いなあ。
候補 5	アシタカのやりきれない気持ちがなんとなくわかった気がする。「共に生きる」ってそういうことなのかな。最後、和解したしこれからもずっとアシタカはサンを思おうって思えた。エンドロールが短く感じたけど、これは意図的で、実際は 5 分くらいなんだと思う。みんな気づいてないことだけど、もののけ姫は、こんなにも長大だったんだな、と圧巻。

象に用いる映画レビューは 2019 年に公開された「Jocker」を用いた。またレビューは映画レビューサイト「Filmarks」より、「Jocker」のページのトップに掲載されているものから 5 件、無作為に選んだ。

被験者はいずれの映画も視聴したことのある大学生 3 名である。

5.3 結果

図 9 に「このアプリが映画内容の印象の言語化に役立ったか」という質問に対する回答を示す。5 段階評価になっており、数字が大きくなるほど被験者にとってアプリが言語化する際に役に立ったことを表す。「5」が非常に役立ったことを、「1」が全く役に立たなかったことを意味する。図 10 には「アプリを通して映画の理解度が深まったか」という質問に対する回答を示す。10 段階評価になっており、数字が大きくなるほど被験者がアプリを通して映画の理解を深められたことを表す。「10」が非常に深まったことを、「1」は実験前と実験後で映画に対する理解度が全く変わらなかったことを意味する。

5.4 考察

「アプリが映画内容の印象の言語化に役立ったか」の質問について、提案手法、比較手法ともに評価が高く、提案システムが被験者の言語化に貢献することが確認できた。しかし、手法ごとに見比べてみると、3 人中 2 人は比較手法の方が評価が高い。この理由としては、提案手法において生成する文章の中には登場人物同士の相関が事実と異なるものが生成されるといった、被験者が文の意味を汲み取りづらいものがユーザに提示さ

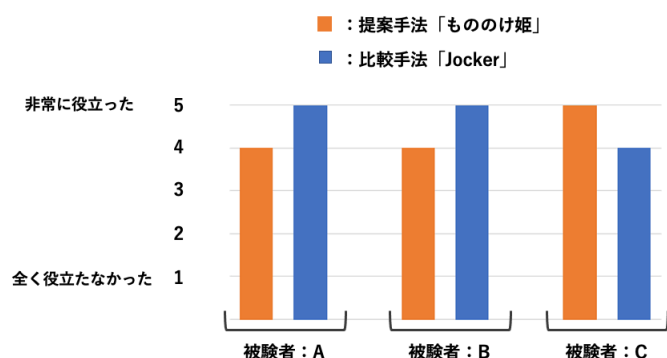


図 9 「提案システムが映画に対する印象の言語化に役立ったか」の回答

れていたことが挙げられる。

また、「映画への理解度がどのくらい深まったか」の質問について、提案手法、比較手法ともに評価が高かった。こちらは提案手法の評価が若干高いものの、サイトに投稿されたレビューよりも優れているとはいえない結果だった。

二つの質問の結果から、提案システムのインターフェイス自体はユーザの感想の言語化に有益だったと考えられる。しかし、提案手法により生成した文章はユーザの感想に関連したことを述べることはできても、そこからユーザの感想が発展していかず、ユーザに提示される生成文が全て似通った内容になってしまう。もしくは、それとは反対に、全く脈絡のない文章が生成

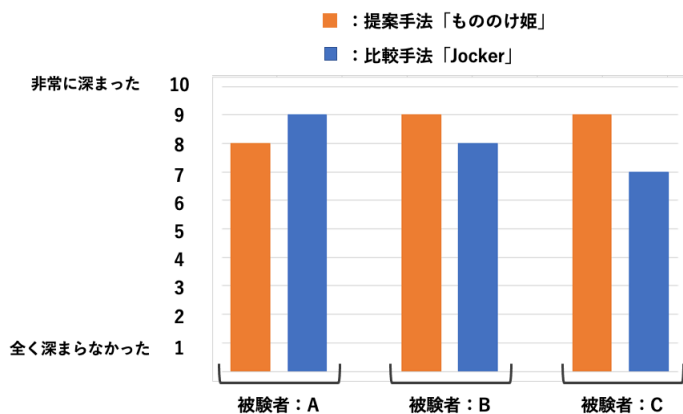


図 10 「映画への理解度がどれくらい深まったか」の回答

されてしまうことがあった。よって、システム上でユーザに提示する文章は、提案手法の言語モデルにより生成した文章である必要性はあまりないと考えられる。

言語モデルが生成する文章に問題があったため、本論文では提案手法の有効性を確認できなかった。そのため改善点として、言語モデルを GPT-2 ではなく、GPT-3 を用いることができれば、提案手法が有効性を示したと考えられる。

6 ま と め

本論文では、映画鑑賞者が映画の理解を深めるために、鑑賞者の感想に基づいて文章を生成し、その章の中から鑑賞者の趣向に沿った文を選んでもらうことで簡易的に感想を言語化するシステムを提案した。評価実験を行った結果、言語モデルが生成する文章は鑑賞者の感想を詳細に言語化することはできなかったが、プロトタイプとして実装したウェブアプリケーションは鑑賞者が言語化を行う補助として効果的であったと考えられる。

本論文の課題点としては、評価実験の際の被験者数が少なかったことが挙げられる。被験者を増やし、主観評価の正確性をより確かなものにする必要がある。

今後の展望としては一方的にユーザに文章を提示するのではなく、ユーザと言語モデルが対話的に映画レビューのやり取りを行いより映画理解を深めるシステムを検討していきたいと考えている。

7 謝 辞

本研究は JSPS 科研費 19H04219 の助成を受けたものです。

文 献

- [1] Mary Beth Oliver and Arthur A Raney. Entertainment as pleasurable and meaningful: Identifying hedonic and eudaimonic motivations for entertainment consumption. *Journal of Communication*, Vol. 61, No. 5, pp. 984–1004, 2011.
- [2] Mary Beth Oliver and Anne Bartsch. Appreciation as audience response: Exploring entertainment gratifications beyond hedonism. *Human communication research*, Vol. 36,

- No. 1, pp. 53–81, 2010.
- [3] 松岡宏明. 対話型鑑賞と対象作品についての再考. *美術教育*, Vol. 2012, No. 296, pp. 26–32, 2012.
- [4] 長井理佐. 対話型鑑賞の再構築. *美術教育学: 美術科教育学会誌*, Vol. 30, pp. 265–275, 2009.
- [5] 竹内晋平, 橋本侑佳. 鑑賞的体験の言語化を通じた美術の俯瞰的理解—中学校美術科学習におけるアクティブ・ラーニングの視点導入に基づく試み—. *美術教育学研究*, Vol. 49, No. 1, pp. 209–216, 2017.
- [6] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language models are few-shot learners. *Advances in neural information processing systems*, Vol. 33, pp. 1877–1901, 2020.
- [7] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.
- [8] Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever, et al. Language models are unsupervised multitask learners. *OpenAI blog*, Vol. 1, No. 8, p. 9, 2019.
- [9] 橋爪友莉子, 山本幹雄. 階層的 encoder-decoder モデルによる宿泊施設レビュー文書に対する応答文書生成. *人工知能学会全国大会論文集 第 34 回 (2020)*, pp. 2H6GS903–2H6GS903. 一般社団法人 人工知能学会, 2020.
- [10] rinna. japanese-gpt2-medium. <https://huggingface.co/rinna/japanese-gpt2-medium>, 2021.
- [11] Flask documentation. <https://msiz07-flask-docs-ja.readthedocs.io/ja/latest/notes>, 2010.