生成型言語モデルによる仮想映画レビューを介した映画検索

宮下　天祥†　　荘司　慶行††　　藤田　澄男†††　　大原　剛三†

† 青山学院大学大学院 理工学研究科　〒 252 – 5258 神奈川県 相模原市 中央区 淵野辺
†† 静岡大学 情報学部　〒 430 – 0807 静岡県 浜松市 中区 城北 3 - 5 - 1
††† LINE ヤフー株式会社　〒 102 – 8282 東京都 千代田区 紀尾井町 1 - 3
E-mail: †miyashita@sw.it.aoyama.ac.jp, ††shojiy@inf.shizuoka.ac.jp, †††sufujita@lycorp.co.jp,
††††ohara@it.aoyama.ac.jp

あらまし　本論文では，映画への感想のような抽象的な検索要求に対し，それに近い映画を発見する手法を提案する．近年主流の生成型言語モデルは，文の続きを予測することに長けている．そこで，映画レビューを用いて追加学習し，書き出しを与えると続きのレビューを生成するモデルを構築した．このモデルに「北野武の撮ったスターウォーズみたい」のような抽象的な要求を与えると，その文に続く，具体的で現実的なレビュー文が生成される．生成されたレビューと類似したレビューをもつ映画をランキングすることで，抽象的な検索要求に合致する映画が検索できる．実レビューサイトのデータを用いた被験者実験における定量的，定性的評価を通して，生成されたレビューに具体的かつ検索に有用な語が多く含まれ，実際に映画を検索可能であることが明らかになった．

キーワード　情報検索，映画レビュー，GPT-2，生成型言語モデル

1　はじめに

近年，ストリーミングによる映画配信サービス，サブスクリプションなどの流行を受けて，映画の視聴方法が多様化している．従来であれば，映画を観る際の選択肢として，映画館で鑑賞するか，DVD レンタルショップで DVD をレンタルするしかなかった．いずれの場合も，あらかじめ観たい映画を決めるか，数少ない候補の中から一番見たい映画を選択するのが一般的であった．しかし，インターネット上には昔の映画から最新の映画まで数多くの映画が存在し，ストリーミングといったサービスを利用することでこれらを自由に選んで視聴することが可能である．このように現代のユーザは，日々膨大な数の映画の中から次に観たい映画を選んで視聴している．そのため，より柔軟でユーザの細かい要求を反映可能な検索アルゴリズムへの要求が高まっている．

このようなアイテム検索においては，メタデータ，アイテムの画像，ユーザの購買行動などが利用されている [1–3]．しかし，映画は映像作品であるため，こうした情報は多くない．そのため，現状の映画検索アプローチは，限られたメタデータからの検索，協調性フィルタリング [4] などに限られる．ここで，まだ活用しきれていない外部リソースとして，レビューが挙げられる．レビューには，ユーザの映画に対する感想，どのような内容だったかなどの情報が含まれるため，映画検索に用いる情報として活用可能であることが期待される．杉木ら [5] の研究では，レビュー文からアイテムの意見情報を抽出することで，メタデータのみからでは検索できないアイテムを検索可能にしている．Yang ら [6] は，Online Customer Review（OCR）を用いて検索した意見文から製品特徴を抽出する手法を提案している．このように，レビューを情報検索に活用することの有効性はすでに示されている．そのため，こうしたレビューの中から，ユーザの検索要求に近い記述が含まれるレビューを検索できれば，ユーザの求める映画を発見できる可能性がある．しかし，レビューは多数のユーザによる自由記述の文章であり，文体，語彙なども書き手によってさまざまである．したがって，ユーザの入力する検索クエリと実際のレビューでは語彙空間が異なるため，直接クエリからレビューを検索することは困難である．

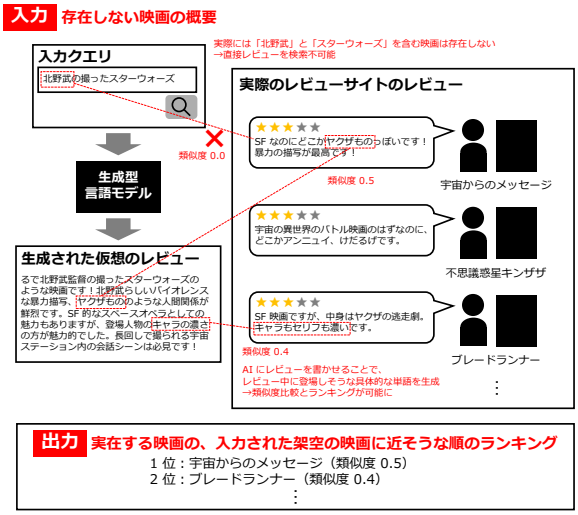

図 1　提案する検索モデルの入出力例と生成レビューの使用方法

そこで本研究では，生成型言語モデルを用いて仮想レビューを生成し，生成されたレビューから映画を検索する手法を提案する．提案手法では，生成型言語モデルによって映画に対する仮想レビューを生成させることで，実際のレビューと検索クエリとの語彙空間のギャップを軽減する．図 1 に示すように，生

成型言語モデルにレビューの書き出しを与えることで，その続きに来そうなレビュー文を生成させる．そして，その続きのレビュー文を拡張クエリとして扱うことで，映画を検索可能にする．

このような，書き出しを与えると具体的な続きのレビュー文を生成する言語モデルを作成するために，実際の映画レビューを教師データとして用いる．レビュー文の続きのレビューを生成するというタスクで言語モデルを学習させる．この際，実際のレビューには映画の内容に言及しない文が多く含まれる．たとえば,「友達と観ました.」や「期待外れだった.」など，映画の内容とは関係ない文が多く含まれる．また実際のレビューの書き出しは,「主人公が自身を鼓舞するために叫ぶシーンがかっこよかった.」といった具体的すぎる記述や,「ヒロインのひたむきな姿がかわいかった.」といった映画全体を評さない書き出しを持つものが多く含まれる．

そこで本研究では，学習時の工夫として，

（1） 映画内容に関連する記述の抽出，

（2） 重要部分の書き出しへの移動

という，2 つの処理を施した学習用のレビューデータを作成し，言語モデルを学習させた．こうすることで，抽象的な映画の説明文が入力された際に，その続きとして，具体的なレビューを生成可能にした．

提案手法が実際に映画の発見に役立つかを検証するために，実際の映画レビューを用いて学習した生成型言語モデルの生成レビューの質を定量的に評価し，仮想レビューを利用して検索した映画を定性的に評価した．これにより，生成型言語モデルを用いることでクエリとレビューという異なる言語空間をまたいだクエリ拡張が可能であるかとレビューの重要部分を書き出しに移動することの効果を明らかにした．

本論文は本章を含めた 6 章から構成される．第 2 章では，本研究と関連のあるアイテム検索，クエリ拡張，大規模言語モデルを用いた情報検索に関する先行研究について述べる．第 3 章では，提案手法の詳細について述べる．第 4 章では，提案手法の生成した仮想レビューに対する評価実験の結果について述べる．第 5 章では，実験結果に対する考察ついて述べる．最後に，第 6 章で本研究から得られた結論を述べる．

2 関 連 研 究

本研究は，映画というアイテムを検索することを目的としている．これを実現するために，生成型言語モデルを用いて仮想レビューを生成し，この生成レビューを拡張的なクエリと見なして映画を検索するという手法を取る．このように生成型言語モデルから出力された文書を用いた情報検索に関する研究は近年盛んにおこなわれている．そのため本章では，提案手法と関連する，アイテム検索，クエリ拡張，大規模言語モデルを用いた情報検索について述べる．

2.1 アイテム検索

オンラインショッピングサイトや特定のサービスに関する情報サイトにおいては，無数に存在する商品やサービスの中からユーザの求めるアイテムを精度よく検索することが求められる．本節では，こうしたアイテム検索に対する先行研究について述べる．

Karmaker ら [1] は，ランキング学習を用いて E コマース上のアイテム検索を最適化する手法を提案している．実際の E コマースデータセットを用いた実験から，ランキング学習手法のうち， LambdaMART [7] を適用すると最も検索の精度が高くなることを明らかにしている．Gao ら [2] は，ファッション分野の商品のテキスト-画像ペアで BERT をファインチューニングすることで，オンラインショッピング上のアイテムをテキストと画像の両方から検索可能にする手法を提案している．ファッション製品の画像とプロのスタイリストによる説明が含まれる Fashion-Gen データセットを用いたアイテム検索タスクから，Image-to-Text， Text-To-Image の両方で提案手法が最も高精度な検索を可能にしたことを明らかにしている．Lu ら [3] は，Transformer ベースの多言語モデルとグラフニューラルネットワークアーキテクチャを組み合わせた，E コマース検索のためのフレームワークを提案している．この手法では，アイテムを特徴量ベクトルにエンコードする際に，ユーザの購買行動に基づくグラフを用いて類似アイテムを推定し，類似アイテムの特徴量も同時に元のアイテムの特徴量に含めることで，クエリとアイテム間の語彙のギャップを軽減している．複数の国で展開している E コマースプラットフォームの検索ログからサンプリングされたデータセットを用いたアイテム検索タスクに対して，提案手法がアイテムの検索精度を大幅に向上させたことを明らかにしている．

しかし，映画は映像作品であるためメタデータ，購買行動などといったアイテムの特徴を表す情報が多くない．そこで，本研究ではユーザ投稿のレビューをアイテム検索のための情報として用いる．このようにアイテム検索にレビューを活用している研究について述べる．杉木ら [5] は，レビュー文の係り受け解析によってそれぞれのアイテムの意見情報を抽出し，検索クエリの要求を満たすアイテムを検索可能にする手法を提案している．実際の宿泊予約サイトに投稿されたレビューを用いた検索タスクから，宿泊施設の提供する情報のみでは検索が困難なクエリからの検索が可能となったことを明らかにしている．Yang ら [6] は，OCR（Online Customer Review）をセンチメント分析することで，クエリに関連する製品特徴に言及したレビューの検索精度を向上させる手法を提案している．実際のオンラインショッピングサイトのアイテムデータを用いたレビュー検索タスクから，提案手法が関連レビューの検索精度を大幅に向上させたことを明らかにしている．そのため，本研究では映画検索のための情報としてレビューを対象にする．

2.2 クエリ拡張

情報検索分野において，インターネット上に存在する情報量の指数関数的な増加とクエリに含まれる情報量の少なさによる検索文書のミスマッチが長らく問題視されている．このような問題に対処するためにユーザの入力したクエリを拡張するとい

うアプローチを選択している手法が数多く存在する．

例として，Voorhees ら [8] は WordNet を用いて入力クエリに含まれる単語の同義語を検索し，拡張クエリとして用いる手法を提案している．TREC [9] と呼ばれるテキスト検索用データセットを用いた実験から，提案手法によるクエリ拡張が特に短いクエリからの検索に有効であることを明らかにしている．Carpineto ら [10] は，クエリ拡張のための外部コーパスから単語の共起に基づいてクエリを拡張する単語を選択する手法を提案している．TREC を用いた情報検索タスクを通して，単語の共起に基づく拡張クエリ選択手法が，いくつかのタスクで他のクエリ拡張手法より効率の良い検索を可能にしたことを明らかにしている．Cui ら [11] は，ユーザの過去の検索クエリログからクエリに含まれる単語と検索に文書含まれる単語の相関関係を抽出し，拡張クエリの選択に活用する手法を提案している．TREC を用いた情報検索タスクから，ユーザの検索ログを用いたクエリ拡張によって情報検索の精度を向上させることを明らかにしている．Salton ら [12] は，適合性フィードバック手法を用いたクエリ拡張による情報検索の精度向上への効果を検証している．

また，クエリから検索された上位 k 件の文書をクエリと関連した文書として使用する疑似適合性フィードバックを用いたクエリ拡張手法も存在する．Xu ら [13] は，擬似適合性フィードバックと単語の共起を組み合わせたクエリ拡張手法を提案している．TREC を用いた情報検索タスクから，提案手法によるクエリ拡張がグローバル分析手法と比較して高精度な情報検索を可能にしたことを明らかにしている．そのため本研究では，クエリとユーザ投稿のレビュー間のギャップ軽減のためにクエリを拡張するというアプローチをとる．

2.3 大規模言語モデルを用いた情報検索

大規模言語モデルは Transformer [14] をベースとした，大規模なパラメータ，データ量が特徴の自然言語処理モデルである．大規模言語モデルは様々なタスクに対して，モデルのパラメータの更新なしに応用可能なことから様々な分野で用いられている．本節では，大規模言語モデルを情報検索に応用した既存研究について述べる．

Zhu ら [15] の論文によれば，大規模言語モデルを用いた情報検索は大きく Rewriter，Retriever，Reranker および，Reader に分けられる．本研究は映画レビューを検索するためのクエリを大規模言語モデルに生成させる研究であるため，大規模言語モデルを用いた Rewriter の研究にフォーカスする．Dai ら [16] は，大規模言語モデルに文書に対するクエリを生成させるタスクを few-shot で学習させることで，ラベルなし文書に対する正解クエリを生成し，検索器を訓練するためのデータセットを作成する手法を提案している．BEIR ベンチマークを用いた検索性能評価の結果から，提案手法によって訓練された検索器が人手によってラベル付けされたデータから訓練された検索器と同等の性能を発揮することを明らかにしている．

Gao ら [17] は，大規模モデルが生成した回答から実際の文書を検索することで信憑性のある回答を提示する手法を提案

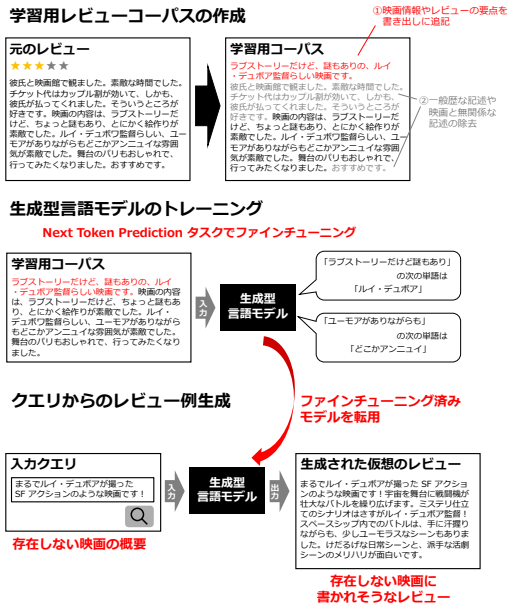

図 2　提案手法の概要

している．TREC DL19，DL20，BEIR データセットを用いた情報検索タスクから，この手法が zero-shot によるものであるにも関わらず，豊富な学習用データを用いた手法と比較して遜色ない性能を発揮したことを明らかにしている．Wang ら [18] は，大規模言語モデルにクエリとクエリに対する回答の文書の例をいくつか与えることで，未知のクエリに対する回答の文書を生成させる手法を提案している．こうして大規模言語モデルによって生成されたクエリと文書のペアを用いてリトリーバを学習させることで情報検索タスクに応用させることが可能である．TREC DL19，DL20，BEIR データセットを用いた情報検索タスクから，提案手法が従来の検索手法による検索精度を大幅に改善したことを明らかにしている．このように，大規模言語モデルを用いた情報検索手法は数多く提案されている．そこで本研究では，これを映画分野に応用して，映画検索の精度を向上させるための手法を提案する．

3 生成レビューを介した映画検索手法

本研究では，生成型言語モデルを用いて仮想レビューを生成し，生成レビューから類似する実際のレビューの投稿された映画を検索する手法を提案する．

3.1 手法の流れ

図 2 に，本手法で提案する生成型連語モデルを用いて仮想レビューを生成する一連の流れを示す．はじめに，レビューデータから，学習用レビューコーパスを作成する．ここでは，レビューデータに含まれるレビューに対して単語登場頻度による映画内容に関連する文の抽出（クレンジング）と，レビュー中の重要部分の書き出しへの移動を行う．次に，作成した学習用コーパ

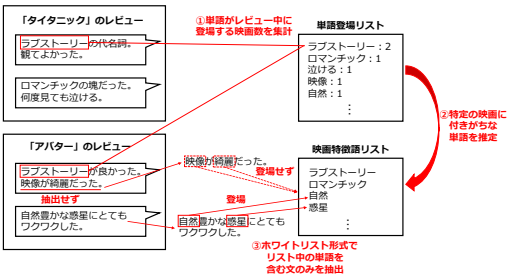

図 3　映画内容に関連する記述の抽出の流れ．それぞれのレビューから多くの映画に登場しすぎない特徴的な表現を含むレビューを抽出

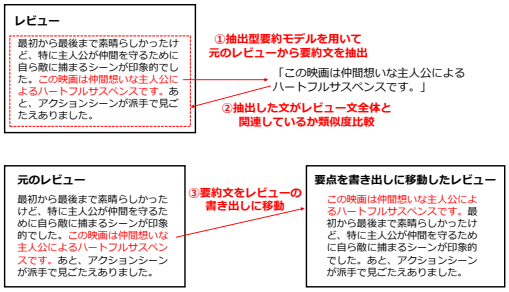

図 4　抽象型要約モデルを用いた重要部分の書き出しへの移動の流れ

スを用いて，生成型言語モデルを Next Token Prediction タスクでファインチューニングする．こうすることで，与えられた書き出しからレビューを生成する言語モデルを作成する．最後に，生成されたレビューと類似したレビューを持つ映画を，実際に検索する．

3.2　学習データ用の映画内容に関連する記述の抽出

本研究では，生成型言語モデルに仮想レビューを生成させるために，実際のレビューデータを用いてモデルを学習する．しかし，実際のレビューデータはユーザの自由記述による文章であるため，一般的な記述，映画の内容と無関係な記述などのノイズが多く含まれる．そのため，映画に関連する内容のレビューを書くようにモデルを学習可能にするため，クレンジングが必要である．そのため，学習用のレビューデータセットから映画内容に関連する記述だけを抽出する．

映画内容に関連する記述の抽出の手順は，

（1）　レビューデータに含まれるすべての単語の登場映画数を計算し，
（2）　登場映画数により映画特徴語を抽出し，
（3）　映画特徴語を含むレビュー文を抽出する

という 3 工程からなる．

図 3 に，具体的な映画内容に関連する記述の抽出の流れを示す．はじめに，単語がレビュー中に登場する映画数を集計し，単語登場リストを作成する．次に，これらの単語登場リストの中から登場映画数が一定以下の単語を映画特徴語として抽出する．そして，それぞれのレビュー文の中で映画特徴語を含む文を抽出する．普遍的な表現を含む文を除去するという方法を取らなかった理由としては，具体的な映画の内容に言及している文であっても普遍的な表現を含む場合は多く存在するからである．たとえば，「面白い」や「すごい」といった表現は普遍的であるが，「二人の恋の駆け引きが面白い」「銃撃戦の迫力がすごい」などは具体的な映画の内容に言及した文と言える．

映画特徴語を抽出する際に，登場映画数の上限と下限を設けてクレンジングを行った．あまりにも少ない映画にしか登場しない単語は，固有名詞や作中に登場する人物名ばかりになる．逆に，あまりに多くの映画に登場する単語は，普遍的な単語である可能性が高い．そのため，予備実験を通して，複数の映画への感想に含まれ，かつ 1 %以上の映画へのレビューに含まれる語だけを残した．

最後にレビューデータから映画特徴語を含む文を抽出する．それぞれのレビューを文単位で分割し，映画特徴語リストの単語を 1 つでも含む文を抽出し，映画特徴語リストの単語を 1 つも含まない文は除去する．こうして映画内容に言及している文を抽出する．

3.3　学習データ用の重要部分の書き出しへの追加

実際のレビューは，必ずしも書き出しがレビュー全体の内容を表した一文であるとは限らない．具体的には，「主人公が自身を鼓舞するために叫ぶシーンがかっこよかった.」などの具体的すぎる記述，「ヒロインのひたむきな姿がかわいかった.」などの映画全体を評さない記述からレビューが書き出されている場合が多い．こういった書き出しを持つデータで言語モデルを学習した場合，入力された文の内容に対応する続きのレビューを予測するというタスクを言語モデルが正しく学習できない恐れがある．そこで，本節ではレビュー中の重要部分を書き出しへ移動する処理を施す．図 4 に，重要部分の書き出しへの移動の流れを示す．

はじめに，抽出型要約モデルを用いて，元のレビューからもっとも重要そうな部分を，要約文として抽出する．そして，この抽出された要約文が，レビュー全体と関連しているかを判定する．要約文と本文について，類似度がある一定以上の場合，この要約文はレビュー全体を表した文であるとみなせる．つまり，検索システム全体の入力である架空の映画の概要と近いものであると考えられる．そのため，抽出した要約文をレビューの書き出しに移動し，学習データとして用いる．

この際，要約文とレビュー全体の類似度が高いレビューのみを使用したのは，レビュー中にレビュー全体を表した文が存在しない場合を考慮したためである．たとえば，1 文目で映画のキャスティングについて言及し，2 文目で印象的なワンシーンについて言及し，3 文目で映画を観た後の感想を述べているようなレビューの場合，レビューで説明しているすべてのことをまとめた 1 文は存在しない．このようなレビューの要約文を書き出しへ移動しても，映画の概要からその映画に付きそうなレビューを生成するというタスクを言語モデルが学習できない可能性が高いため，このようなレビューは除いた．

このような処理を実現するために，はじめに，レビューを抽出型要約モデルに入力し，要約文を得た．具体的には，近年主流の大規模言語モデルを用いた抽象型要約器である BERTSUM [19]

を用いた．次に，要約文がレビュー文全体と意味的に類似するかを判定する．本手法の学習では，書き出しがレビュー文全体を表すような，概要的な文から始まるレビューだけを学習に用いたいと考えた．そのために，要約文が，レビュー文全体と意味的に類似するかを判定した．

意味的類似度の判定には，近年主流の類似度判定用モデルである Sentence-BERT [20] を用いた．Sentence-BERT では，2 つの文章を入力すると，その意味的類似度を 0 から 1 で計算できる．類似度が閾値（実験の際には 0.6 とした）以下だった場合は，そのレビューは学習用データから除外した．一方，類似度が閾値より高かった場合，学習用に，レビュー文から要約文と同じ文章を除去したうえで，先頭に移動した．これは，元のレビューから抽出された文を除かないと，1 つのレビュー文中に全く同じ文が 2 回登場すると，言語モデルがレビュー内で同じことを繰り返し記述する場合があるためである．こうして，重要部分を書き出しに移動したレビューからなる，学習用データセットが作成できた．

3.4 書き出しから続きのレビューを予測させるタスクによるモデルの学習

入力された書き出しに対する続きのレビューを生成可能にするために生成型言語モデルを追加学習する．学習には Next Token Prediction タスクを用いる．このタスクは，入力されたトークン列の次に来る確率が高いトークンを予測する機械学習タスクである．

言語モデルの事前学習では，Next Token Prediction を繰り返すことで文章を生成する．このタスクはラベルを必要としない，自己教師あり学習である．このタスクでは，入力トークン列に対して次に来る確率が最も高いと予測されたトークンを選択する．この際利用される損失関数 $L(U)$ は，言語モデルの語彙を $U = u_1, u_2, ...u_N$，言語モデルのパラメータを θ，言語モデルの語彙中に含まれる i 番目のトークンが，入力されたトークン列 $u_k, ...u_{-1}$ の次に来る確率を $p(u_i|u_k, ...u_{-1}, \theta)$ とした際に，

$$L(U) = \sum_{i=1}^{N} \log(p(u_i|u_k, ...u_{-1}, \theta)) \quad (1)$$

と定義される．

4 評　　価

提案手法の生成したレビューが映画検索における有用性を向上させているかを検証するために，生成レビューの定量的評価と仮想レビューを利用して検索した映画の定性的評価の 2 つの評価実験を実施した．

4.1 データセット

本実験における生成型言語モデルの学習および仮想レビューを利用した検索した映画の定性的評価には，IMDb Review Dataset - ebD[1]を利用した．このデータセットは 453,528 件の映画等に対して投稿された 5,571,499 件のレビューデータから構成されている．この中から，232,622 件の映画に対するレビューを使用した．また，投稿されたレビューの数が 10 件を超える映画に関してはその映画に投稿されたレビューの中から 10 件をランダムサンプリングして使用した．サンプリング後のレビューの総数は 1,007,151 件である．

4.2 比 較 手 法

提案手法の生成レビューに対する効果を検証するために，

- **ベースライン**：生成型言語モデルに対してレビューデータを用いた追加学習をせず，そのままのモデルでレビューを生成する手法，
- **追加学習のみ**：処理を施していない生のレビューデータで追加学習した生成型言語モデルでレビューを生成する手法，
- **映画に関連する文の抽出＋追加学習**：映画内容に関連する記述の抽出を施したレビューデータで追加学習した生成型言語モデルでレビューを生成する手法，
- **提案手法**（書き出し移動＋映画に関連する文の抽出＋追加学習）：重要部分の書き出しへの移動と映画内容に関連する記述の抽出を施したレビューデータで追加学習した生成型言語モデルでレビューを生成する手法

の 4 つの方法でモデルをトレーニングし，それぞれの生成レビューを人手で評価した．

4.3 実　　装

実験のために，実際に生成型言語モデルにレビューデータを学習させ，任意の書き出しの続きを生成できるようにした．生成型言語モデルとして，近年主流である GPT-2 [21] を用いた．学習データ用の重要部分の推定に，抽出型要約モデルである BERTSUM を用いた．また，文の類似度比較のために，SentenceBERT の paraphrase-MiniLM-L6-v2[2] モデルを用いた．

追加学習のみの手法の学習には，1,007,151 件のレビューから 10 万件のレビューをランダムにサンプリングして使用した．映画に関連する文の抽出＋追加学習の手法の学習には，10 万件のサンプリングされたレビューから，映画内容に関連する記述のみを抽出して用いた．提案手法の学習は，映画内容に関連する記述を抽出した 10 万件のレビューで追加学習をしてから，さらに重要部分の書き出しへの移動した 10 万件のレビューで追加学習を行った．

GPT-2 の追加学習時はすべての場合で，エポック数を 3，ブロック数を 512，訓練時のバッチサイズを 1 とした．fast tokenizer の使用は False に設定し，その他のハイパーパラメータはデフォルト値のままで学習と生成を行った．

4.4 生成レビューの定量的評価

本節では，提案手法が生成した仮想レビューの映画検索にお

1：「IMDb Review Dataset - ebD」：
https://www.kaggle.com/datasets/ebiswas/imdb-review-dataset
2：haggingface「paraphrase-MiniLM-L6-v2」：
https://huggingface.co/sentence-transformers/paraphrase-MiniLM-L6-v2

ける有用性検証のための被験者実験について述べる．本実験では，あらかじめ設定したクエリから提案手法および比較手法の生成したレビューの映画検索における有用性を 4 つの評価項目で被験者に評価させた．

仮想のレビューを生成するための入力クエリを 48 件，あらかじめ人手で作成した．これらのクエリは，すべて，「This is a movie like STAR WARS taken by Takeshi Kitano」のような，存在しない映画の概要である．1 件のクエリに対して 4 つの比較手法からそれぞれ 3 件ずつ，計 12 件のレビューを生成した．そのため，生成レビューの総数は 576 件である．

今回はこれらの生成レビューを，12 人の被験者に評価させた．1 件のレビューにつき 3 人の被験者で評価し，その平均値を評価値として用いた．1 人の被験者には 12 件のクエリに対する 144 件の生成レビューについて評価させた．なお，レビューが英語であるのに対し，被験者は全員，日本人である．そのため，実際の実験では元の英語のレビューと Google 翻訳によるレビューの日本語訳を同時に提示した．提案手法および比較手法の生成したレビューを，

- 映画内容言及率,
- クエリ説明性,
- 視聴促進度,
- クエリ説明単語量

の 4 項目で被験者に評価させた．

映画内容言及率は生成レビューを構成する文のうち，具体的な映画の内容に言及している文の割合がどのくらいであるかという評価項目である．被験者には，レビュー各文に対して，この文が具体的な映画の内容に言及しているかどうかを，言及しているなら 1，していないなら 0 で評価させた．そして，レビューを構成する文の数に対する映画内容に言及する文の数の割合で評価した．

クエリ説明性は生成レビューがクエリをどのくらい説明できているかという評価項目である．被験者には，生成レビューがクエリをどのくらい説明できているかを 1 を最低評価，5 を最高評価とした 1 から 5 の 5 段階で評価させた．

視聴促進度は，仮にクエリで表されるような映画が観たいと思っているとき，この生成レビューが投稿された映画をどれくらい観たくなるかという評価項目である．被験者には，生成レビューがどのくらい映画の視聴を促進するかを 1 を最低評価，5 を最高評価とした 1 から 5 の 5 段階で評価させた．

クエリ説明単語量は，生成レビューを構成する単語のうち，クエリを説明している単語の割合がどのくらいであるかという評価項目である．被験者には，生成レビューの中からクエリを説明していると思う表現をすべて抽出させた．なお，抽出する表現は 1 単語に限定しない．このようにして 1 件の生成レビューから抽出された表現すべてを構成する単語数を求めることで，生成レビューを構成する単語数に対する割合を算出し，クエリ説明単語量を評価した．

4.5 生成レビューの定量的評価の結果

各手法の生成したレビューに対する被験者実験の結果を表 1 に示す．表 1 から，提案手法はすべての評価項目において，ベースライン，追加学習のみの 2 つの手法と比較して有意な精度向上が見られた．しかし，提案手法の精度と映画に関連する文の抽出+追加学習の精度の間にはすべての評価項目で有意差が見られなかった．映画内容言及率，視聴促進度といった評価項目では，映画に関連する文の抽出+追加学習が提案手法と比較して高精度という結果となった．クエリ説明単語量においては有意差はなかったものの，提案手法が映画に関連する文の抽出+追加学習を上回る結果となった．

表 1　提案手法（書き出し移動＋映画に関連する文の抽出+追加学習）と比較手法の生成したレビューに対する被験者評価の結果（Student t 検定, **: $p < 0.01$, *: $p < 0.05$）

手法	映画内容言及率	クエリ説明性	視聴促進度	クエリ説明単語量
提案手法	0.29	0.24	0.21	**3.28%**
映画に関連する文の抽出+追加学習	**0.30**	**0.24**	**0.22**	2.74%
追加学習のみ	0.20**	0.16**	0.16**	2.18%*
ベースライン	0.21*	0.09**	0.09**	1.41%**

表 2　提案手法のクエリ説明性が最も高かったクエリと最も低かったクエリから生成されたレビューの例

	クエリ	生成レビュー	平均クエリ説明性
クエリ説明性の最も高かった生成レビューの例	This is a movie of western, but samurai	This movie has been made a decade before samurai movies which were made for the Japanese audience. The samurai movies and movies from that era did have the use of a lot of stylized weapons in their games, which is a nice feature that you could expect from samurai movies. Although there are some flaws in the movie, I think it is good and watch it for the style of the movies. My only criticism is the movie lacks realism in its depiction of the samurai at first. Also, I would have liked to see the movie take place in some kind of western town, but it would have been better if the movie has a more Western setting and that was the case. If you like samurai movies and you like the movies more than samurai movies, it is definitely worth to watch this one. I am not kidding: If you liked samurai movies, you will like this one.	0.75
クエリ説明性の最も低かった生成レビューの例	This is a movie that makes you unbelievable everything	The movie shows you the world of the main character's life as he travels around the world with a different purpose. At the beginning it is obvious that the plot is extremely complicated due to the fact that the main character is not at all human. But then you understand that he is also a normal person because when he goes to the movies he is all dressed in black and a strange woman comes across him. His acting is extremely believable especially in the scenes where he tries to escape from the strange woman by using a wheelbarrow and then through a window.The story was amazing and it was very interesting to think about it, even if you don't know much about Thailand. The way of seeing the movie is a combination of watching the movie with your typical Thai movie experience, as if you are going through a similar movie experience. I know that there are Thai movies that does not use the same genre as the movie	0.0

次に，提案手法から生成されたレビューが被験者から高い評価を得たクエリと低い評価を得たクエリについて議論する．提案手法のクエリ説明性が最も高かったクエリと最も低かったクエリから生成されたレビューの例を表 2 に示す．クエリ説明性が最も高かった「This is a movie of western, but samurai」というクエリから生成されたレビューには，「samurai」や「Western」といった表現は生成されているが，クエリに含まれる以上の情報はほとんど存在しなかった．クエリ説明性が最も低かった「This is a movie that makes you unbelievable everything」というクエリから生成されたレビューには，映画の内容に言及した文が含まれているがクエリと関連するような表現は存在しなかった．

4.6 仮想レビューを利用して検索した映画の定性的評価

実際に生成されたレビューを用いることで，どのような映画が検索可能になるかを検証するために，被験者実験を実施した．本実験では，提案手法から生成された仮想レビューを用い

表 3　提案手法が生成した仮想レビューによる映画検索が有効に働いた例（クエリ：日本版の騎士道物語）

クエリ	This is a movie like a Japanese version of Tales of Chivalry
クエリを説明している表現	‘some interesting side stories about two men who are both friends’, ‘a movie about two lovers who are fighting over Chivalry in the Japanese colonial war.’, ‘The story of The Sword and Sandal is a very modern one, as it concerns the life and death of a knight from a noble family.’, ‘One day a girl comes to meet him and she starts looking for him, but a samurai comes after her and takes her away.’, ‘samurai’
クエリを説明している表現から検索された映画	Shikonmado- Dai tatsumaki(1964) Junko’s Shamisen (2010) Bushi no ichibun(2006) Otsuyu:Kaidanbotan-dôrô(1998) Za samurai (1987) Kakushi-torideno san-akunin(1958) Bushidô zankokumonogatari(1963) The Sea Is Watching (2002) Miyamoto Musashi (1954) The Twilight Samurai (2002)

表 4　提案手法が生成した仮想レビューによる映画検索が有効に働かなかった例（クエリ：マトリックスのようなホラー）

クエリ	This is a movie of Matrix-like horror
クエリを説明している表現	‘attacking humans’, ‘The movie isvery violent’, ‘raped and murdered hundreds of women’, ‘enemy’, ‘using thelatest technology’, ‘fear’, ‘kill’, ‘horror’, ‘Matrix’, ‘The film is full of a very real sense of fear and horror.’, ‘murdered’
クエリを説明している表現から検索された映画	The Houses October Built (2014) 咒怨 2 (2003) Voodoo Dolls (1991) Un uomo, un cavallo, una pistola (1967) Take My Eyes (2003) 駭客任務完結篇：最後戰役 (2003) O fovos (1966) Jûsan-nin renzoku bôkôma (1978) Napasta (1982) Village of the Damned (1995)

て映画を検索し，検索された映画とクエリの関連度を主観で評価した．被験者は 1 名で，検索結果の映画を知らなかった場合，ウェブでその映画について調べ，その映画がクエリと適合するかを判断した．また，主観的に，検索結果の映画や，抽出されたキーワードが正しかったかを分析した．

4.4 節の実験で被験者が抽出した「クエリを説明している表現」を，元のクエリを拡張するクエリとして使用する．BM25 を用いて，被験者が抽出した「クエリを説明している表現」を入力し，関連度の高いレビューをランキングする．検索対象は，4.1 節で紹介した 232,622 件の映画に対する 1,007,151 件のレビューである．ランキングされたレビュー上位 10 件の投稿先の映画を検索結果として，それぞれの映画の内容を確認した．実際の検索には，提案手法が生成した 3 件レビューに対して 3 人の被験者が抽出した「クエリを説明している表現」すべてを用いた．また，検索モデルとして，Python 用ライブラリである Rank BM25[3] の Okapi BM25 を用いた．

4.7　仮想レビューを利用して検索した映画の定性的評価の結果

提案手法が生成した仮想レビューによる映画検索が有効に働いた例を表 3 に示す．この結果では，「This is a movie like a Japanese version of Tales of Chivalry」というクエリに対して，武士や侍が主役の映画が多数登場している．武士や侍は，クエリ中の「Japanese」「Tales of Chivalry」といった要素と関連がある上に，これらを直接的に表す表現はクエリに登場していないため，仮想レビュー生成が映画検索に有効に働いた例と言える．また，実際に生成されたレビューから被験者が抽出した「クエリを説明している表現」には，「samurai」という単語が登場している．

被験者からの評価が高かったにもかかわらず提案手法が生成した仮想レビューによる映画検索が有効に働かなかった例を表 4 に示す．この結果では，「This is a movie of Matrix-like horror」というクエリに対して，「Matrix」らしさのないホラー映画ばかりが登場している．実際に生成されたレビューから被験者が抽出した「クエリを説明している表現」は，ほぼすべて「horror」に関係する表現であった．

5　考　　察

生成レビューの定量的評価の結果から，映画内容に言及した文の抽出は生成レビューの映画検索における有用性向上に大きく貢献することが明らかとなった．このことから提案手法では，映画内容に言及した文を精度よく抽出できていたため，言語モデルが与えられた書き出しの内容を表す続きの文を生成するというタスクを正しく学習できたと考えられる．

一方，重要部分の書き出しへの移動は生成レビューの映画検索における有用性向上にあまり効果がないことが明らかとなった．このことについて考察するために，実際に提案手法で使用した抽出型要約モデルが抽出したレビュー要約の例を提示する．抽出型要約モデルが抽出したレビュー要約の中に「Skye insists that it was some sort of evil some kind of “darkness” which was the cause of her boyfriend’s death.」というものがあった．これは，映画に登場する人物の主張を要約したものである．本研究では，映画の内容を端的に表した概要からのレビュー生成を目的としているため，映画に登場する人物の主張を要約したこの文は目的と齟齬がある．このように，学習に使用したレビューの書き出しすべてが映画の内容を端的に表した概要となっていなかったため，生成レビューの映画検索における有用性向上にあまり効果がなかったことが考えられる．

仮想レビューを利用して検索した映画の定性的評価の結果から，クエリによっては映画検索に仮想レビューを利用することが有効であることが明らかとなった．このことから，提案手法で学習した生成型言語モデルが，クエリに含まれる要素に関連する表現を学習できていたこと，入力された書き出しの内容を説明するレビューを生成するというタスクを理解できていたことが考えられる．しかし，多くのクエリからは，クエリを網羅的に説明できる表現が生成されないという結果になった．このことから，生成型言語モデルはあくまで入力文に対する続きを予測するモデルであるため，入力文に要素が複数存在する場合，特定の一部の要素に対する続きを予測しやすいと考えられる．

3：「Rank BM25」：
https://pypi.org/project/rank-bm25/

6 まとめと今後の課題

本研究では，生成型言語モデルを用いて仮想レビューを生成し，生成されたレビューと類似する実際のレビューの投稿された映画を検索する手法を提案した．生成型言語モデルに，実際のレビューを用いて，書き出しの続きとなる文を予測させることで仮想レビューを生成可能にした．この際，生成レビューが与えられた書き出しの内容を具体的に説明したものになるように，学習用レビューから映画内容に関連する記述を抽出し，重要部分を書き出しへ移動した．提案手法の生成したレビューが映画検索においてどれだけ有用であるかを検証するために，生成レビューの定量的評価と仮想レビューを利用して検索した映画の定性的評価を実施した．

生成レビューの定量的評価の結果として，映画内容に関連する記述の抽出は，生成レビューの映画内容言及率，クエリ説明性，視聴促進度の 3 つの評価値を有意に向上させた．一方，重要部分の書き出しへの移動は，今回の評価において有意な精度向上をもたらさなかった．これらのことから，映画内容に関連する記述の抽出が生成レビューの映画検索における有用性向上に大きな影響があることが明らかになった．仮想レビューを利用して検索した映画の定性的評価の結果として，一部のクエリでは映画検索に生成レビューを利用することで，クエリの情報を拡張可能であった．一方，多くのクエリでは，クエリの全ての要素を網羅的に説明する表現が生成されなかった．これらのことから，まだ精度に改善の余地があるものの，生成型言語モデルを用いた仮想レビューが映画検索において有用であることを示した．

今後の課題として，生成レビューに含まれる情報量の少なさの改善，生成レビューを有効に活用するための検索アルゴリズムの提案が挙げられる．提案手法が生成したレビューには，クエリに含まれる表現がそのまま登場してしまうことが多く，それらを具体的に説明するような表現はあまり登場しなかった．これに対しては，映画内容に言及するレビューの抽出精度の改善，レビュー本文に対する書き出しを説明する表現の追記などの対策が考えられる．また，今回の実験では，映画を検索するための拡張クエリを人手で抽出した．生成レビューからの映画検索手法について，今後検討する必要がある．

謝　　辞

本研究の一部は JSPS 科研費 21H03775，21H03774，22H03905 による助成を受けたものです．ここに記して謝意を表します．

文　　献

[1] Shubhra Kanti Karmaker Santu, Parikshit Sondhi, and ChengXiang Zhai. On application of learning to rank for e-commerce search. In *Proc. ofthe 40th international ACM SIGIR conference on research and development in information retrieval*, pp. 475–484, 2017.

[2] Dehong Gao, Linbo Jin, Ben Chen, Minghui Qiu, Peng Li, Yi Wei, Yi Hu, and Hao Wang. Fashionbert: Text and image matching with adaptive loss for cross-modal retrieval. In *Proc. ofthe 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 2251–2260, 2020.

[3] Hanqing Lu, Youna Hu, Tong Zhao, Tony Wu, Yiwei Song, and Bing Yin. Graph-based multilingual product retrieval in E-commerce search. In *Proc. ofthe 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies: Industry Papers*, pp. 146–153, 2021.

[4] Taghi M. Khoshgoftaar Xiaoyuan Su. A survey of collaborative filtering techniques. In *Advances in Artificial Intelligence*, Vol. 2009, p. 19, 2009.

[5] 杉木健二, 松原茂樹ほか. 消費者の意消費者の意見に基づく商品検索見に基づく商品検索. 情報処理学会論文誌, Vol. 49, No. 7, pp. 2598–2603, 2008.

[6] Chin-Sheng Yang, Chih-Ping Wei, and Christopher C. Yang. Extracting customer knowledge from online consumer reviews: a collaborative-filtering-based opinion sentence identification approach. In *Proc. ofthe 11th International Conference on Electronic Commerce*, pp. 64–71, 2009.

[7] Christopher JC Burges. From ranknet to lambdarank to lambdamart: An overview. *Learning*, Vol. 11, No. 23-581, p. 81, 2010.

[8] Ellen M. Voorhees. Query expansion using lexical-semantic relations. In *Proc. ofthe 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, p. 61–69, 1994.

[9] Donna K Harman. *The first text retrieval conference (TREC-1)*, Vol. 500. US Department of Commerce, National Institute of Standards and Technology, 1993.

[10] Claudio Carpineto, Renato de Mori, Giovanni Romano, and Brigitte Bigi. An information-theoretic approach to automatic query expansion. *ACM Transactions on Information Systems*, Vol. 19, No. 1, p. 1–27, 2001.

[11] Hang Cui, Ji-Rong Wen, Jian-Yun Nie, and Wei-Ying Ma. Probabilistic query expansion using query logs. In *Proc. ofthe 11th International Conference on World Wide Web*, p. 325–332, 2002.

[12] Gerard Salton and Chris Buckley. Improving retrieval performance by relevance feedback. *Journal of the American society for information science*, Vol. 41, No. 4, pp. 288–297, 1990.

[13] Jinxi Xu and W. Bruce Croft. Query expansion using local and global document analysis. In *Proc. ofthe 19th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 4–11, 1996.

[14] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Proc. ofthe 31st International Conference on Neural Information Processing Systems*, p. 6000–6010, 2017.

[15] Yutao Zhu, Huaying Yuan, Shuting Wang, Jiongnan Liu, Wenhan Liu, Chenlong Deng, Zhicheng Dou, and Ji-Rong Wen. Large language models for information retrieval: A survey. *arXiv preprint arXiv:2308.07107*, 2023.

[16] Zhuyun Dai, Vincent Y Zhao, Ji Ma, Yi Luan, Jianmo Ni, Jing Lu, Anton Bakalov, Kelvin Guu, Keith B Hall, and Ming-Wei Chang. Promptagator: Few-shot dense retrieval from 8 examples. In *The Eleventh International Conference on Learning Representations*, 2023.

[17] Luyu Gao, Xueguang Ma, Jimmy Lin, and Jamie Callan. Precise zero-shot dense retrieval without relevance labels. In *Proc. ofthe 61st Annual Meeting of the Association for Computational Linguistics*, pp. 1762–1777, 2023.

[18] Liang Wang, Nan Yang, and Furu Wei. Query2doc: Query

expansion with large language models. In *Proc. ofthe 2023 Conference on Empirical Methods in Natural Language Processing*, pp. 9414–9423, 2023.

[19] Yang Liu. Fine-tune bert for extractive summarization. *arXiv*, 2019.

[20] Nils Reimers and Iryna Gurevych. Sentence-BERT: Sentence embeddings using Siamese BERT-networks. In *Proc. ofthe 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*, pp. 3982–3992, 2019.

[21] Alec Radford, Jeff Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. Language models are unsupervised multitask learners. 2019.