

商品のユーザレビューを利用した体験的な記述の生成とその根拠の推定

沈 辰夫[†] 莊司 慶行[†] 山本 岳洋^{††} 田中 克己^{†††} Martin J. Dürst[†]

[†] 青山学院大学 理工学部 〒252-5258 神奈川県 相模原市 中央区 淵野辺

^{††} 兵庫県立大学 社会情報科学部 〒651-2197 兵庫県 神戸市 西区 学園西町

^{†††} 福知山公立大学 情報学部 〒620-0886 京都府 福知山市 字堀

E-mail: [†]shen@sw.it.aoyama.ac.jp, ^{††}{shoji,duerst}@it.aoyama.ac.jp, ^{†††}t.yamamoto@sis.u-hyogo.ac.jp,
^{††††}tanaka-katsumi@fukuchiyama.ac.jp

あらまし 本稿では、理解に前提知識を要するような商品説明を、理解容易な体験的な記述に変換し、その記述が実際に商品説明のどの部分に関連するかを提示する手法を提案する。具体的には、「ISO 感度 51200」のようなカメラの知識を要する記述を、「夜でもくっきり人の顔が写ります」などの身近で体験的な記述に置き換える。そのために、まず、GPT-2に商品情報とレビューからなるデータを学習させ、商品情報を入力すると、その商品につきそうなレビューを生成する言語モデルを作成した。次に、商品説明の一部を隠したデータを言語モデルに与え、レビューを生成させた。こうすることで、「ISO 感度」に関する商品説明を除去すると、「夜」、「くっきり」などの語が、生成されたレビューに表れにくくなるなどの傾向を、生成されたレビュー中の単語の出現確率の増減として表せるようにした。楽天市場の商品レビューデータセットを用いた被験者実験を通じて、実際に提案手法が正しく商品特徴をとらえた体験的な記述を生成できることを明らかにし、また、体験的な記述が、商品説明に記されたスペック情報よりも理解しやすいことが示された。

キーワード レビュー, 推薦, 機械学習, GPT-2, 説明可能性

1 はじめに

近年、オンラインショッピングサイトの隆盛や、メルカリに代表される個人間通信販売アプリの普及を受けて、誰もが気軽にインターネットで商品を選び、購入する機会が増えてきている。総務省による調査では、現在の日本人の73.4パーセントは、インターネット上で買い物をしたことがあるとされる¹。このような状況下で、通販サイトに掲載されている膨大な商品の中から、用途に適した商品を探し出すことの重要性が増してきている。一般的に、用途に適した商品を探し出す際には、商品の詳細な特長が述べられた商品説明文を読む場合が多い。一方で、実際に通販サイトに記載されている商品説明文は、無駄な記述が多く読みづらかったり、商品のドメインの専門的な知識を持っていないと理解することが難しい内容が記載されている場合がある。

具体的な例として、カメラについて詳しく知らない初心者が、野鳥を綺麗に撮影できるようなカメラを、通販サイトで購入したい場合を考える。一般的に、インターネットショッピングサイトの販売ページに記載されたカメラの商品説明文には、「撮像素子：aps-c」や「ISO 25600」などのカメラの詳細なスペックが書かれている。これらのスペック情報は、カメラに詳しい人であれば商品選択の参考にはできるが、初心者の場合、読んでそれが何を意味しているのか理解できない。そのため、このカ

メラが「野鳥を撮影する」という目的に適したものか、判断できない。

こういった際に、商品説明を補う形で、購入する商品の選択のために使われる情報源として、利用者投稿の商品レビューがある。例えば、ある利用者の書いたレビュー内に、「鳥を綺麗に撮影することができる」といった旨が記述されていれば、初心者でも「どうやら遠くにある小さくて動くものを撮影できるようだ」と知ることができる。このような購入者による体験的な記述は、多くの初心者にとって、その商品が何に使えるかを判断し商品を選ぶ助けとなる。

一方で、すべての鳥を撮影できるカメラに「鳥を綺麗に撮影することができる」とレビューが書いてあるとは限らない。また、複数のカメラに、「鳥を綺麗に撮影することができる」とレビューが書かれていた場合、どちらを選んでいいか判断することが難しい。こういった場合に正しく商品を選ぶためには、体験的な記述を、商品説明と合わせて理解することが必要である。具体的には、購入者はたくさんのレビューと商品説明を見比べて、「『鳥を綺麗に撮影することができる』とレビューに書かれる商品の多くは、ズーム性能が高くシャッタースピードが速い傾向がある」という関係性を自分で見つけ出さないと、正しい商品を選択できない。このような商品の特徴と、それがもたらす結果の関係を理解することは、商品選びをする初心者にとって、大きな負担である。

そこで本研究では、商品説明文を入力すると、その商品を使った場合に得られるであろう体験的な記述と、その記述の根拠となる商品説明文の箇所を出力するアルゴリズムを提案する。実際のアプリケーションの動作例を1に示す。

1: インターネットショッピング及びオークションフリマの利用状況：令和3年版情報通信白書、第1部、第1章、第1節:

<https://www.soumu.go.jp/johotsusintokei/whitepaper/ja/r03/pdf/n1100000.pdf>

入力

商品説明文

このカメラは、ISO 204800で、焦点距離最大800mm。シャッター速度は8~1/4000。連写は秒間10フレーム、バッテリー性能1100mAh。撮像素子はフルサイズ！

出力

体験的な記述

遠くの鳥、撮ってます！
根拠 (理由：焦点距離最大800mm)

夜でも顔がくっきり！
(理由：ISO 204800)

サッカー観戦に最適。
(理由：焦点距離最大800mm)

商品説明文を入力として 体験的な記述と根拠を出力

図1 アプリケーションの動作例

このアルゴリズムでは、入力として、任意の商品の商品説明文を受け付ける。すると、アルゴリズムは、その商品特徴から、その商品につきそうなレビューを複数、自動で生成する。そして、生成されたそれぞれのレビュー文について、「この文が生成されたのは、商品説明にこの単語が含まれていたためである」というように、根拠推定を行う。こうすることで、「ISO感度409600」という専門知識を要する記述を、「夜間でも顔がくっきり撮れる。」という体験的な記述に変換し、「根拠：ISO感度409600」と根拠を提示できる。このように、体験的な記述とその根拠を並べて提示することで、初心者でも商品を選びやすくなるのが期待される。

このように商品説明を体験的な記述に変換するにあたって、本研究では、事前に大規模コーパスで学習された生成的言語モデルを用いた。これは、

- 一般利用者の書いた曖昧で多様なレビュー表現を理解可能にする、
- 自然言語的な推論を可能にする、
- レビューのついていない商品にも対応可能にする

ためである。投稿レビューは、メーカーや販売サイトの専門家が書いた商品説明文とは異なり、一般的な利用者が書いたものである。そのため、レビュー文は曖昧で、質の低いものもある。加えて、レビュー中の表現は、人によって大きく異なる。例えば、野鳥の撮影に適したカメラへのレビューで、「遠くの鳥をばっちり撮れます」と書く人もいれば、「ピーちゃんをパシャリ」と書く人もいる。こういった際に、それぞれの表現を正しく理解するためには、大規模コーパスで学習された言語モデルを使う必要がある。

また、レビュー中に含まれる体験的な記述には、自然言語的推論が適用可能な場合もある。例えば、「メジロを撮るのに使っています」と「鶯の撮影に最適です」というレビューは、「野鳥の撮影に使える」とまとめることができる。加えて、「野鳥の撮影」と「サッカー場での撮影」は、遠くの動くものを撮影するという意味で、類似している。大規模言語モデルを用いることで、このような自然言語処理に基づく推論が可能になると考え

られる。

最後に、実際の商品にはそれぞれ必ずレビューがついているわけではなく、購入者が現れて初めてレビューが書かれる。通販サイトに掲載して間もない商品や、ニーズが比較的低い商品等にはレビューはつきにくい。生成的なアプローチを用いることで、新しい商品や、マイナーな商品でも、体験的な記述を生成可能である。

そこで、本研究では、このような体験的な記述の生成にあたって Radford ら [1] によって提案されている機械学習モデルである GPT-2 を利用する。現在一般的に配布されている学習済み GPT-2 モデルを用いて、商品説明文から、体験的な記述を生成する。このモデルは、40 ギガバイト程度の Web 上のテキストデータを用いて、15 億のパラメータを持つ大規模な Transformer を用いて学習したモデルであり、機械翻訳や要約といった文章生成の研究に多く使われている。

汎用言語モデルである GPT-2 を今回の体験的な記述の生成に用いるために、商品とレビューからなるデータセットでファインチューニングを施す。具体的には、商品説明とレビューを連結した文章を GPT-2 に与え、文の続きを書くように学習させる。こうすることで、商品説明文を与えると、その続きとしてレビュー文を生成するようになる。

こうして生成したレビュー文について、それぞれのレビュー文が、商品説明のどの部分に対応して生成されたのかを推定し、根拠として提示する。そのために、生成した体験的な記述の根拠となる商品説明文の箇所を推定するために、アプリケーションに基づく手法を用いる。具体的には、商品説明文中の一文を取り除いた状態で体験的な記述を生成し、その生成された文章に現れる単語の変化を見ることで、根拠となる商品説明文箇所を推定する。

このような提案アルゴリズムについて、実際に GPT-2 を用いて実装し、有用性について検証する。実際に正しく商品特徴を捉えた体験的な記述を生成できるか、正しく根拠を推定できているか、そして実際に商品選択の意思決定に使えるかについて、被験者実験を通して評価した。

本論文の構成は以下の通りである。本論文は、本章を含めて全6章からなる。本章では、本研究を行うにあたってどのような経緯や背景があったのかを述べた。第2章では、本研究に関連した研究について紹介し、本研究の位置づけを示す。第3章では、本研究が提案する手法を具体的に述べる。第4章では、本研究の提案した手法の評価について述べる。第5章では、本研究の実験結果の考察を行う。

2 関連研究

本研究は、ユーザーレビューを利用した研究、推薦フレーズの生成に関する研究、推薦理由の提示に関する研究の3つの分野に関連する。そのため、2.1節ではユーザーレビューを利用した研究について、2.2節では推薦フレーズの生成に関する研究について、2.3節では推薦時の根拠の提示に関する研究について、それぞれ関連研究を紹介し、論じる。

2.1 ユーザレビューを利用した研究

近年では、ユーザレビュー情報を用いてアイテムを推薦する手法が一般的になりつつある。ユーザレビューを用いた研究の例として、Zheng ら [2] は、ユーザレビューから商品やサービスの特性とユーザの行動を共同で学習する深層学習モデルを用いて、推薦を行う手法を提案している。

とくに本研究に近い例として、江田ら [3] は、書籍レビューサイトにおいて対象とするレビュー文に「いいね」を押すユーザを予測する機械学習モデルを構築し、機械学習モデルを解釈する代表的なアルゴリズムの1つである LIME を用いて、ユーザに適した推薦フレーズを自動的に抽出する手法を提案している。この研究では、あるユーザがどの書籍に対して「いいね」を押しているか、というデータが必要不可欠であり、このデータが不必要である本研究と条件が異なる。また、本研究では、生成的なアプローチで推薦フレーズを提示しているのでレビューのついていない新規の商品についても推薦することが可能である。

2.2 推薦フレーズの生成に関する研究

本研究では商品説明から体験的な記述を生成するが、情報推薦の分野などで、推薦フレーズを自動生成する研究も行われてきている。Zhang ら [4] は、実際に JD.com の商品推薦プラットフォームにおいて、機械学習モデルを用いた商品コピーライティング自動生成システム (APCG) を導入し、7ヶ月で 253 万件の商品説明文を生成した結果を報告している。この研究では、機械学習モデルを用いた推薦フレーズが実用化できるものであり、その有効性があることを示している。

ほかに Zhang ら [5] は、条件付き変分オートエンコーダー (CVAE) を基にした Self labeling CVAE (SLCVAE) を提案し、商品説明文を入力に推薦フレーズを生成している。Li ら [6] は、生成的なアプローチを用いた推薦システムにおいて、文章の品質管理と表現力を両立させる試みとして Neural Template (NETE) 説明文生成フレームワークを提案している。Chan ら [7] は、複数商品にまたがる商品説明文や広告を作成する、S-MG Net と呼ばれる機械学習モデルを提案している。Deng ら [8] は、消費者のニーズに合わせて広告コンテンツを自動的にパーソナライズすることができる広告文の生成システムである SGS-PAC を提案している。これらの研究では商品広告やキャッチコピーを生成することを目的としているが、本研究では、フレーズと根拠を合わせて提示することで、初心者でも商品を選びやすくすることを目的としている。

2.3 推薦時の根拠の提示に関する研究

本節では、推薦する際の根拠の提示、また機械学習モデルの説明可能性に関する研究について説明する。近年では、機械学習モデルの判断根拠を説明する「説明可能な AI」と呼ばれる研究分野が急速に発展している [9]。特に、本研究のような商品説明などの分野においても、推薦理由などを明らかにし、信頼性を確保する研究は活発になってきている [10]。

zhang ら [11] は、商品の特徴やユーザの傾向から、ユーザの興味に応じて推薦と非推薦の両方の商品を提示し、さらになぜ

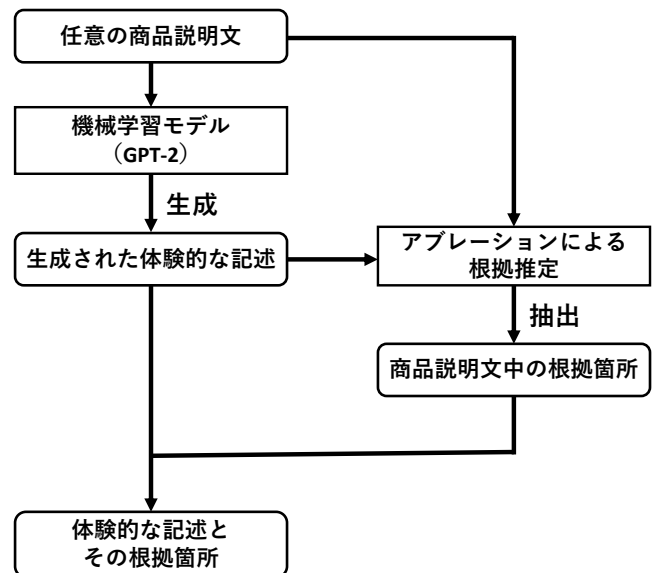


図 2 提案手法の全体図

その商品が推薦されるのか、あるいは推薦されないのかという根拠を提示する Explicit Factor Model (EFM) を提案している。Sinha ら [6] は、5つの音楽推薦システムに対してユーザの認識について調査し、ユーザはどのような理由で推薦されたのかという透明性が高いと感じる推薦に対して好感を持つことが示されている。この研究は、推薦時の根拠を提示することの有効性を示している。

3 提案手法

本手法では、任意の商品の商品説明文を入力すると、その商品の長短を表す体験的な記述を複数出力すると同時に、その体験的な記述の根拠となる商品説明文の箇所を提示する。

このような処理を、商品説明文と商品レビュー文を学習データとして、学習済み言語モデルである GPT-2 [1] をファインチューニングして体験的な記述を生成、そしてアブレーションに基づく手法を用いて根拠となる商品説明文の箇所を抽出し、提示することで実現する。

3.1 商品・レビューデータのクレンジングと前処理

本手法では、商品説明文とその商品のユーザレビューをデータセットとして用いる (図 3)。

本手法では、GPT-2 のファインチューニングをより効率的で効果的なものにするために、データのクレンジングを行った。データのクレンジングについては商品説明文と商品レビューについてそれぞれ行い、基本的にはルールベースで不必要であると判断した単語が含まれる文章を削除する処理を行った。具体的な処理の内容は下記の通りである。

商品説明文には、キャンペーン、送料等のように商品そのものについて直接関係のない記述が含まれているため、このような単語が含まれる文章は取り除く処理をした。また、実際に目

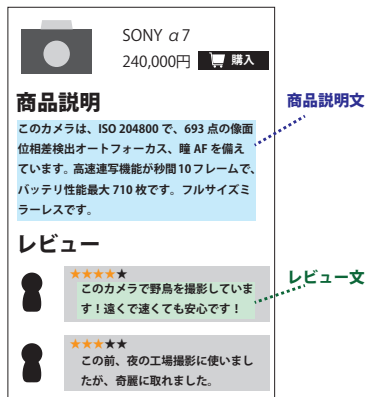


図3 学習で用いる実際の商品説明文とレビュー文からなるデータセットの例



生成されたレビュー中の単語の登場確率を分析「レンズ」という単語は「Fマウント搭載機です」という記述を除去すると、登場せず

図5 アブレーションによる商品特徴ごとの体験的記述推定の例。商品説明文から一部の文を除去してレビューを生成し、除去した単語と生成されたレビュー中の単語の登場頻度を比較する。この例では、「Fマウント搭載機です」という記述を除去すると、「レンズ」という単語の出現頻度が下がっている。そのため、「Fマウント搭載機です」という商品説明に対する体験的記述は「いろんなレンズをつけて遊んでいます」などだと推定できる。

学習させたファインチューニング済み GPT-2 に、推論時には任意の商品説明文と [SEP] を与えることでその続きの文章であるその商品説明文にふさわしいようなレビュー文を生成させることができる。そのようにして生成されたレビュー文は、1行、1文程度のものから、数行に渡るものまでまばらである。そこで生成されたレビューを1文ずつに分割し、15文字以上70文字以下のものを抽出したものを今回目的とする体験的な記述とする。

3.4 ユーザレビューのみを用いたファインチューニング

上述の通り、本提案手法の本筋としては、商品説明文とその商品のレビューを [SEP] トークンで繋いだ文章を用いて、GPT-2 をファインチューニングする手法である。一方で異なるアプローチとして、商品説明文とその商品のレビュー文を学習させるのではなく、商品のレビュー文のみを GPT-2 をファインチューニングする手法も本研究では試みた。この手法では、推論時には任意の商品説明文を入力として、その続きの文章を GPT-2 に生成させる。

3.5 根拠となる商品説明文箇所の推定

体験的な記述の生成について説得力を持たせるために、体験的な記述の根拠となる箇所の商品説明文中の一文を抽出する。それを実現するためのアプローチとして、アブレーションに基づく手法を用いる（図5の左側）。

本手法で用いるアブレーションに基づく手法とは、まず初めに複数の文で構成されている商品説明文中のある1文を取り除いて、その取り除いたあとの残った商品説明文を入力として、ファインチューニング済み GPT-2 に体験的な記述を20個を50回で、合計1,000個生成させる。次に、そのようにして生成された1,000個の体験的な記述群に登場する単語の名詞のみを

学習時

入力：商品説明文とレビュー文をセパレータでつないだもの

このカメラは、ISO 204800で、693点の像面位相検出オートフォーカス、瞳AFを備えています。高速連写機能が秒間10フレームで、バッテリー性能最大710枚です。フルサイズミラーレスです。[SEP] このカメラで野鳥を撮影しています！遠くで速くても安心です！

GPT-2

このカメラは、ISO 204800で、の次に来る単語は 693 点
ミラーレスです [SEP] の次に来る単語は この

次トークンを予測する言語モデリングで学習

生成時

入力：未知の商品説明文とセパレータの後に空白

D60はISO 1600で、Fマウント搭載機です。画質は1075万像素です。セルフタイマー搭載です。[SEP]

空白

GPT-2

D60はISO 1600で、Fマウント搭載機です。画質は1075万像素です。セルフタイマー搭載です。[SEP] 昼はいいけど、夜景には不向きなカメラです。多彩なレンズが使えて楽しいです！細かいものはぼやけがちかも。集合写真を撮るのに使っています！

[SEP] までの内容を考慮して、以降に続きそうな文を補完

図4 商品説明文からレビュー文を生成するためのファインチューニングタスクでの学習と、実際の生成の流れ。

的とする体験的な記述として、商品の性能をよく表した、具体的な使い道や使い心地を示す記述がふさわしい。そのため、使用する商品のユーザーレビューのデータについて、発送や梱包、プレゼント等についてのものを除く処理をした。

3.2 商品説明文とユーザーレビューを用いたファインチューニング

本手法では、商品説明文とその商品のレビュー文を [SEP] トークンで繋いだ文章を学習データとして GPT-2 のファインチューニングに利用する（図4の左側）。

3.3 ファインチューニング済み GPT-2 を用いた体験的な記述の生成

次に、ファインチューニングした GPT-2 言語モデルを用いて、実際に体験的な記述を生成する（図4の右側）上述のように

抽出し、それぞれ異なる入力の商品説明文中で生成した体験的な記述の中で単語の出現確率を計算する。以上のような作業を商品説明文中の文の全てに行う。

このとき、名詞を抽出する際には MeCab² という形態素解析ツールを利用した。また、解析に使用する辞書は更新頻度が高く、新語・専門用語を多く含んでいる MeCab-ipadic-NEologd³ を使用した。

そして、それぞれ異なる入力の商品説明文中で生成した体験的な記述群間で単語の出現確率の差を計算してランキングする。一連の手順で、商品説明文中に含まれる文の数だけのランキングされた、単語とその単語の異なる入力の商品説明文中で生成された体験的な記述群間との出現確率の差のデータが得られる。このデータの中から、異なる入力の商品説明文中で生成された体験的な記述群間を問わず、最も出現確率の差が大きい単語上位 5 つのものを抽出する。このとき、抽出した単語がそれぞれの箇所の商品説明文中の一文を取り除いたときのものであるかを控えておく。そして、どの文も取り除いていない商品説明文中を入力として生成された体験的な記述の中から、それぞれそのような単語が含まれているものを抽出する。最終的に、抽出した体験的な記述群は、控えておいた商品説明文中の一文が根拠となっている箇所として得ることができる。本手法の、最後に抽出した体験的な記述が、控えておいた商品説明文中の一文が根拠となっている箇所であると考えられる背景を説明する。

例えば、ある商品説明文中から、その商品説明文中の 1 文 l_i だけを取り除いた文章から体験的な記述を生成する。その結果、商品説明文中の異なる 1 文 l_j を取り除いた文章から体験的な記述を生成したときに比べて、ある単語 w の出現確率が低くなったとする。それはつまり、ある単語 w が含まれる体験的な記述は、商品説明文中の 1 文 l_i と相関関係があり、つまり根拠となっている箇所であると考えられる。

3.6 体験的な記述と体験的な記述の根拠となる商品説明文中の箇所の提示

3.5 節で得られた、体験的な記述とその根拠となる商品説明文箇所を提示する。3.5 節で得られた体験的な記述群に含まれる体験的な記述の数は不定である。そこで、本研究ではその体験的な記述群の中から無作為に 1 つ抽出し、それを根拠とする商品説明文箇所と共に提示することとする。つまり、最も出現確率の差が大きかった単語と、根拠となる商品説明文のセットが 5 つあり、それに対応する体験的な記述を 1 つ提示するので合計して 5 つの体験的な記述とその根拠となる商品説明文箇所を提示する。

4 評価実験

本章では、提案手法の有用性を示すために用いたデータセッ

トと、そのデータセットを用いて行った評価実験の詳細、そしてその結果について述べる。本研究では、提案手法の有効性を評価するために被験者実験を行った。

4.1 データセット

GPT-2 のファインチューニングのための学習データを作成するために、楽天グループ株式会社が国立情報学研究所 情報学研究データリポジトリを通じて提供している楽天市場データを利用した。

本研究では、提案手法の評価にあたってイヤホン、カメラ、T シャツ、パンの 4 つの商品分野について評価を行った。

学習データの詳細な内容としては、まずレビューが少なくとも 1 つ付いている商品の商品説明文及びレビュー文を抽出する。そして商品説明文とレビュー文を合わせて 512 トークンという制限の下で、商品説明文は 320 トークンまで、レビュー文は 192 トークンまでで切り上げて、これを [SEP] トークンで繋いだもので学習データを作成した。実際に用いたデータセットの規模としては、イヤホンでは商品説明文とレビュー文のセットが 13,000 件程度、カメラでは同様のセットが 3,500 件程度である。

4.2 実装

本研究で用いる GPT-2 モデルとして、rinna 社がオープンソースとして公開している日本語に特化した GPT-2 の大規模言語モデル⁴を用いる。この GPT-2 モデルは、訓練データとして、CC-100⁵ のオープンソースデータに含まれる 70 ギガバイトの日本語テキストを約 1 ヶ月の長期間にわたってトレーニングされ、汎用性があるものである。この GPT-2 モデルに対して、4.1 節で述べたデータセットを用いて学習させた。

4.3 比較手法

本節では、提案手法の有効性を調べるために提案手法と比較する手法について述べる。本提案手法を評価するために、ファインチューニング時に異なるアプローチを用いた手法、共起頻度に基づく抽出的なアプローチを用いた手法、根拠となる商品説明文中の箇所を提示しない手法、そして参考として元の商品説明文のみを提示する場合と比較する。具体的には、

- **提案手法**：本研究で提案している、商品説明文とレビュー文を学習データとしてファインチューニングを施した GPT-2 で体験的な記述を生成する手法、
- **レビュー文のみ学習**：商品のレビュー文のみを学習データとしてファインチューニングを施した GPT-2 を用いた手法、
- **根拠箇所提示なし**：GPT-2 で生成した体験的な記述の根拠となる商品説明文中の箇所を推定せず、生成された体験的な記述群の中で最も頻出した単語上位 5 つを含む体験的な記述を提示する手法、

2 : MeCab :

<https://taku910.github.io/mecab/>

3 : MeCab-ipadic-NEologd :

<https://github.com/neologd/mecab-ipadic-neologd>

4 : rinna/japanese-gpt2-medium :

<https://huggingface.co/rinna/japanese-gpt2-medium>

5 : CC-100 :

<http://data.statmt.org/cc-100/>

- **共起頻度（ベースライン）**：元のデータセット中の、商品説明文とそのレビュー文との間で共起して現れる単語のセットを計算し、共起する回数の多い順番にランキングする。そして、入力された商品説明文中に現れる単語に対して共起する単語を含むデータセット中の実際のレビュー文を抽出する。その際に商品説明文中に現れる単語が含まれる商品説明文中の1文が抽出したレビュー文の根拠箇所となり、これを提示する手法、
- **商品説明文のみ提示**：体験的な記述を提示せず、そのままの商品説明文のみを提示する手法

の5つの手法である。

4.4 実験方法

本提案手法が出力した商品の体験的な記述とその根拠である商品説明文中の箇所を、被験者実験によって評価する。以下の項目について、3名の被験者が評価する。被験者には実際の商品説明文と、その商品説明文から生成した体験的な記述が提示され、それぞれの項目について評価する。実際に評価する項目として、

- (1) 体験的な記述は自然な日本語の文章であるか、
- (2) 体験的な記述は体験的な記述らしい文章か、
- (3) 体験的な記述の内容は正しいか、
- (4) 体験的な記述とその根拠である商品説明文中の箇所を見て商品に興味を持ったか、
- (5) 体験的な記述とその根拠である商品説明文中の箇所を見て商品の理解が深まったか、
- (6) 体験的な記述の根拠として正しい商品説明文中の箇所が提示されているか、

の6つについて、それぞれ検証した。このうち、「6. 体験的な記述の根拠として正しい商品説明文中の箇所が提示されているか」という項目に関しては、そのドメインの商品について深い知識を有した有識者でないと判断が困難である。そのため、この項目については追加で各ドメイン1名ずつ、合計2名の被験者を追加で用意し、評価した。

実際に行った実験の詳細としては、以下の通りである。まず、それぞれの商品分野について商品説明文を15件、計60件の商品説明文を用意した。また、その1件の商品説明文に対して各手法5件ずつの体験的な記述及びその根拠である商品説明文中の箇所のセットを提示する。なお、商品説明文のみ提示では、体験的な記述及びその根拠である商品説明文中の箇所の提示はなし、根拠箇所提示なしでは、根拠である商品説明文中の箇所を提示せず体験的な記述のみを提示する。この体験的な記述及びその根拠である商品説明文中の箇所のセットそれぞれ1つ1つに対して上述の6つの項目について評価する。なお、商品説明文のみ提示では、商品に興味を持ったか、そして商品の理解が深まったかの2項目のみ、根拠箇所提示なしでは、体験的な記述の根拠として正しい商品説明文中の箇所が提示されているか、の項目を除いた5つの項目について評価する。

表1 イヤホン、カメラの両商品分野の平均評点（共起と比べて** $p < 0.01$, * $p < 0.05$ ）

評価項目	提案手法	レビュー学習	共起	根拠なし	商品説明文
自然な日本語の文章であるか	**4.36	4.25	4.18	**4.01	
体験的な記述らしい文章か	4.20	**4.04	4.32	**3.96	
体験的な記述の内容は正しいか	**3.66	**3.86	3.45	3.40	
双方を見て商品に興味を持ったか	**3.90	**3.87	3.58	**3.23	3.80
双方を見て商品の理解が深まったか	**3.80	**3.86	3.61	**3.16	3.77
正しい根拠箇所が提示されているか	**3.39	**3.69	3.00		

表2 イヤホンの商品分野のみの平均評点（共起と比べて** $p < 0.01$, * $p < 0.05$ ）

評価項目	提案手法	レビュー学習	共起	根拠なし	商品説明文
自然な日本語の文章であるか	*4.35	4.18	4.16	**3.81	
体験的な記述らしい文章か	4.35	*4.12	4.34	**3.82	
体験的な記述の内容は正しいか	*3.59	**3.83	3.40	3.33	
双方を見て商品に興味を持ったか	**3.93	**3.88	3.52	**3.11	3.78
双方を見て商品の理解が深まったか	*3.77	**3.88	3.56	**3.09	3.73
正しい根拠箇所が提示されているか	3.23	**3.70	3.04		

表3 カメラの商品分野のみの平均評点（共起と比べて** $p < 0.01$, * $p < 0.05$ ）

評価項目	提案手法	レビュー学習	共起	根拠なし	商品説明文
自然な日本語の文章であるか	4.36	4.32	4.20	4.20	
体験的な記述らしい文章か	*4.06	**3.96	4.30	4.11	
体験的な記述の内容は正しいか	*3.73	**3.88	3.51	3.47	
双方を見て商品に興味を持ったか	*3.88	*3.87	3.64	*3.36	3.82
双方を見て商品の理解が深まったか	3.82	3.84	3.65	**3.24	3.80
正しい根拠箇所が提示されているか	**3.54	**3.67	2.96		

表4 Tシャツの商品分野のみの平均評点（共起と比べて** $p < 0.01$, * $p < 0.05$ ）

評価項目	提案手法	レビュー学習	共起	根拠なし	商品説明文
自然な日本語の文章であるか	4.56	**4.32	4.67	**4.46	
体験的な記述らしい文章か	4.29	**4.13	4.44	*4.24	
体験的な記述の内容は正しいか	**3.44	**3.43	2.91	2.95	
双方を見て商品に興味を持ったか	**3.72	*3.64	3.40	3.31	*3.80
双方を見て商品の理解が深まったか	3.50	3.62	3.62	**3.28	3.87
正しい根拠箇所が提示されているか	**3.42	**3.54	2.71		

4.5 実験結果

本節では、4.4節で述べた被験者実験の結果を述べる。

実験結果から、提案手法等と共起による手法とを t 検定を行い比較し、有意差が認められた項目については数値には*をつけて表記する。

イヤホン、カメラの両商品分野をまとめて比較したものを表1に示す。イヤホンの商品分野のみで比較したものを表2に示す。カメラの商品分野のみで比較したものを表3に示す。Tシャツの商品分野のみで比較したものを表4に示す。パンの商品分野のみで比較したものを表5に示す。

実験の結果、提案手法とベースライン手法と比較して、すべての商品分野で商品に興味を持ったか等の複数の評価項目において有意差が認められた。また、レビュー文のみを学習したGPT-2を用いた手法においても、同様にすべての商品分野で複数の評価項目において有意差が認められた。

5 考察

本章では、評価実験によって得られた結果を考察する。

結果から、すべての商品分野においてベースライン手法と比べて、商品に興味を持ったか等の複数の評価項目において有意

表 5 パンの商品分野のみの平均評点（共起と比べて ** $p < 0.01$, * $p < 0.05$ ）

評価項目	提案手法	レビュー学習	共起	根拠なし	商品説明文
自然な日本語の文章であるか	4.28	**4.47	4.27	4.27	
体験的な記述らしい文章か	**3.65	3.91	4.00	**3.45	
体験的な記述の内容は正しいか	**3.34	**3.51	2.81	2.66	
双方を見て商品に興味を持ったか	**3.54	**3.77	3.22	*2.91	**3.82
双方を見て商品の理解が深まったか	**3.48	**3.64	3.05	*2.82	**3.84
正しい根拠箇所が提示されているか	**3.29	**3.66	2.84		

表 6 提案手法におけるイヤホン分野の評点の高い例

商品説明文	繊細でリッチなサウンド（独自ドライバー×ハイレゾ再生）さらに進化した Anker 独自ドライバー「A.C.A.A 3.0」搭載。2つのダイナミックドライバーにより、細かなサウンドも明瞭に聞こえる繊細でリッチな音質を実現。高音質コーデック・LDAC に対応し、通常のコーデック（※ Bluetooth A2DP の SBC, 328kbps, 44.1kHz 時）と比較して 3 倍の情報量を伝送できるため、原音に忠実に音楽を再現することができます。3D オーディオによる 360° の音響体験。Anker 独自のアルゴリズムでリアルタイムに音源を処理し、ライブ会場や映画館にいるような音響体験を提供します。また、ジャイロセンサーが頭の動きを検知し、つねに音楽に囲まれた感覚を体験できます。音楽モードとムービーモードの選択が可能です。ウルトラノイズキャンセリング 2.0 Anker 独自技術のウルトラノイズキャンセリング 2.0 により、周囲の騒音レベルに応じてノイズキャンセリングの強さを自動で調節し、環境に左右されることなく音楽への没入感を極限まで高めます。ヘルスマニタリング 心拍モニタリング、ストレッチチェック、姿勢リマインダー、ワークアウト機能等をご利用いただけます。
体験的な記述	普段の通勤では地下鉄、飛行機で使っていますが、満員電車でもノイズキャンセリングは切れず、不満はありません
根拠箇所	ウルトラノイズキャンセリング 2.0 Anker 独自技術のウルトラノイズキャンセリング 2.0 により、周囲の騒音レベルに応じてノイズキャンセリングの強さを自動で調節し、環境に左右されることなく音楽への没入感を極限まで高めます

表 7 提案手法におけるカメラ分野の評点の高い例

商品説明文	フルサイズカメラで培った技術により、瞳を検出する精度と速さ、追従性が、α 6300 の瞳 AF から大幅に向上。動きのあるポートレート撮影でも簡単・確実に、瞳にピントを合わせ続けることが可能になりました。また、シャッターを半押しするだけで瞳 AF が作動。ピントを合わせたい瞳の左右を切り換えることも可能になりました。さらに、「リアルタイム瞳 AF」は一部の動物にも対応。これにより、ペットや野生動物の瞳も高速・高精度に検出し、追従可能です。撮像エリアの約 84 % をカバーする 425 点の位相差 AF センサーを高密度に配置。加えて、コントラスト AF 枠を α 6300 の 169 点から 425 点に多分割化しました。さらに従来のロックオン AF を一新した「リアルタイムトラッキング」機能を搭載。狙いたい被写体を指定し、シャッターボタンを半押し、もしくはモニター上でタッチするだけで、カメラまかせて精度高く自動追尾させることができます。高速連写時にもサイレント撮影の使用が可能になりました。静かに撮影したいシーンでもシャッター音を気にせずに撮影でき、決定的瞬間を逃しません。さらに、AF-C やリアルタイム瞳 AF、リアルタイムトラッキングとの併用も可能です。
体験的な記述	私は動物を撮影しますので、接写で撮ったときの、動いている対象の表情がそのまま表情に出ます
根拠箇所	静かに撮影したいシーンでもシャッター音を気にせずに撮影でき、決定的瞬間を逃しません

差が認められた。これより、生成的アプローチを用いた手法は、簡単な抽出アルゴリズムと比べてより人を引きつける体験的な記述を提示することができることがわかった。

本章では、いくつかの観点に注目して提案手法の有効性について考察する。

はじめに、各商品分野に現れる特徴について考察する。

提案手法と商品説明文のみを提示する手法について、イヤホンとカメラの 2 つの商品分野では、商品に興味を持ったか、商品の理解が深まったかの 2 つの項目において同等の評点であった。一方で、T シャツ、パンの 2 つの商品分野では、これらの 2 つの項目において商品説明文のみを提示する手法が高評価であった。

イヤホンとカメラの 2 つの商品分野では、これらの商品は複数のスペック項目を有していて、それらに対応する様々な体験的な記述の提示できると考えられる。例えば、イヤホンでは有線かワイヤレスか、ノイズキャンセリングがついているか、使っているドライバは何か、等といった複数のスペック項目がある。そして「ノイズキャンセリング機能搭載」という商品説明文ならば、「電車の騒音が気になりません」という体験的な記述が対応するといった具合である。

一方で、T シャツ、パンの商品分野ではスペック項目が少なく、見た目や味といった商品説明文からは推定することが難しい要素を含む分野である。また、ユーザが商品説明文を見てスペックを理解できなかったり、何に使えるのかわからないといったようなケースが少ないと考えられる。そのため、この 2 つの商品分野は本手法を用いるにはあまり適切ではない分野であると考えられる。

そのため、以降はイヤホンとカメラの 2 つの商品分野について考察することとする。

次に、体験的な記述を提示する手法の有効性について考察する。

提案手法と商品説明文のみ提示する手法を比較すると、商品に興味を持ったか、商品の理解が深まったかの 2 項目のどちらにおいても有意差が認められず同等の評価であった。

提案手法によって生成された体験的な記述の中には明らかに元の商品説明文の内容に矛盾するものが存在し、これによって提案手法の平均評点が下がった点を考えると、生成する体験的な記述の精度を向上させることで体験的な記述を提示する手法の有効性を示すことが可能であると考えられる。

また、提案手法における生成した体験的な記述は、商品説明文中の 1 文のみに即したものであり、商品の 1 スペックのみについて述べられているものである。一方で、商品説明文全体では当然その商品のすべてのスペックが記されている。そのため、体験的な記述と商品説明文では、そもそもの情報量の違いがありこれが評価に影響したことが考えられる。

次に、根拠となる商品説明文箇所を提示する手法の有効性について考察する。提案手法と根拠箇所を提示しない手法を比較すると、5 つの項目すべてにおいて有意差が認められる結果になった。この結果から、根拠箇所を提示する手法は、根拠箇所を提示しない手法よりも有効であるとわかった。

また、根拠箇所を推定するアルゴリズムについて考察する。本提案手法で行った推定方法では、商品説明文中の 1 文ずつに対して推定を行うような手法であったため、他に関連のある商品説明文が存在したとき等に適切に推定することができなかったと考えられる。具体的には、このメーカーの出している製品は前提としてこのような機能が付属している、ある機能が備わっているときには前提として他のある機能も備わっている、そもそも同様な機能が複数の文で商品説明文に登場しているといったような例である。そのため、本提案手法よりも細かい粒度で根拠を推定したい際には、単語単位で推定する必要があると考えられる。

最後に、GPT-2 の異なるファインチューニングタスクを用

いた手法について考察する。

レビュー文しか含まれない学習データを用いてファインチューニングを施した GPT-2 を用いた手法は、学習データに含まれない商品説明文を入力として受け取っているにもかかわらず、提案手法と遜色がない評価であった。

これは、データセットのレビュー文にはそもそも「このスペックだからこの用途にいい」というように本来商品説明文に含まれるスペック項目が含まれるものがあつたり、逆に、入力される商品説明文の中に「この用途に最適のものです」というように本来レビュー文に含まれる用途や使い心地等が含まれていることがあり、それらが一つの要因となっていると考えられる。

6 まとめと今後の課題

本研究では、商品説明文を入力として与えると、その商品説明文に適した体験的な記述と、その根拠となる商品説明文中の箇所を出力するアルゴリズムを提案した。提案手法では、商品説明文とレビュー文を学習データとして用いて GPT-2 をファインチューニングし、体験的な記述を生成した。そして、商品説明文の一部を隠した状態でレビュー文を生成することで、商品説明文に何が書かれていると、どういうレビュー文が書かれやすいかを明らかにした。

また、提案手法と他比較手法の出力について6つの評価項目において被験者実験によって評価することで、手法の有用性について議論した。評価実験によって、提案手法のような生成的アプローチや根拠箇所を提示する手法の有効性は示された。一方で、実際に提示した商品説明文中の根拠箇所は間違っていたものも多く存在した。現在の手法では、3.5節において、単純な出現確率の差で上位5つの単語を抽出し、さらにその単語が含まれる体験的な記述群の中から無作為に抽出して提示している。この2点において、さらに工夫を施すことで提示する体験的な記述と、商品説明文中の根拠箇所の質の向上が見込めると考えられる。これらについて、本研究における今後の課題であると考えている。

謝 辞

本研究の一部は JSPS 科研費 21H03775, 21H03774, 22H03905 による助成、ならびに 2022 年度国立情報学研究所共同研究 22S1001 の助成を受けたものです。本研究では、楽天グループ株式会社が国立情報学研究所 情報学研究データリポジトリを通じて提供している楽天市場データ [12] を使用しています。ここに記して謝意を表します。

文 献

- [1] 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.
- [2] Lei Zheng, Vahid Noroozi, and Philip S Yu. Joint deep modeling of users and items using reviews for recommendation. In *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*, pp. 425–434, 2017.
- [3] 怜央江田, 剛聡牛尼. ユーザレビューを用いた機械学習モデルの説明可能性に基づく推薦フレーズの抽出. *情報処理学会論文誌データベース (TOD)*, Vol. 15, No. 3, pp. 50–62, 2022.
- [4] Xueying Zhang, Yanyan Zou, Hainan Zhang, Jing Zhou, Shiliang Diao, Jiajia Chen, et al. Automatic product copywriting for e-commerce. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 36, pp. 12423–12431, 2022.
- [5] Yuchi Zhang, Yongliang Wang, Liping Zhang, Zhiqiang Zhang, and Kun Gai. Improve diverse text generation by self labeling conditional variational auto encoder. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2767–2771. IEEE, 2019.
- [6] Rashmi Sinha and Kirsten Swearingen. The role of transparency in recommender systems. In *CHI '02 Extended Abstracts on Human Factors in Computing Systems*, pp. 830–831, New York, NY, USA, 2002. Association for Computing Machinery.
- [7] Zhangming Chan, Yuchi Zhang, Xiuying Chen, Shen Gao, Zhiqiang Zhang, Dongyan Zhao, and Rui Yan. Selection and generation: Learning towards multi-product advertisement post generation. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 3818–3829, Online, November 2020. Association for Computational Linguistics.
- [8] Shasha Deng, Chee-Wee Tan, Weijun Wang, and Yu Pan. Smart generation system of personalized advertising copy and its application to advertising practice and research. *Journal of Advertising*, Vol. 48, No. 4, pp. 356–365, 2019.
- [9] David Gunning and David Aha. DARPA's explainable artificial intelligence (XAI) program. *AI magazine*, Vol. 40, No. 2, pp. 44–58, 2019.
- [10] Alexandra Vultureanu-Albiși and Costin Bădică. Recommender systems: An explainable AI perspective. In *2021 International Conference on INnovations in Intelligent Systems and Applications (INISTA)*, pp. 1–6, 2021.
- [11] Yongfeng Zhang, Guokun Lai, Min Zhang, Yi Zhang, Yiqun Liu, and Shaoping Ma. Explicit factor models for explainable recommendation based on phrase-level sentiment analysis. In *Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval, SIGIR '14*, p. 83–92, New York, NY, USA, 2014. Association for Computing Machinery.
- [12] 楽天グループ株式会社. 楽天市場データ, 2020. <https://doi.org/10.32130/idr.2.1>
<https://rit.rakuten.com/data.release/>.

- [1] 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.
- [2] Lei Zheng, Vahid Noroozi, and Philip S Yu. Joint deep modeling of users and items using reviews for recommendation.