

生成的言語モデルによるレビュー文のキャッチフレーズ文体への変換

沈 辰夫^{†a)} 莊司 慶行^{††b)} 山本 岳洋^{†††c)} 大原 剛三^{†d)}
田中 克己^{††††e)}

Transforming Review Texts into Catchy Sales Slogans with Generative Language Models

Shen CHENFU^{†a)}, Yoshiyuki SHOJI^{††b)}, Takehiro YAMAMOTO^{†††c)}, Kozo OHARA^{†d)}, and
Katsumi TANAKA^{††††e)}

あらまし これは電子情報通信学会論文誌に掲載された論文の最終版投稿前の著者稿です。正式版は公式の
ポジトリを参照してください (<https://doi.org/10.14923/transinfj.2024dep0001>)。引用時には以下の版を
ご引用ください:

“生成的言語モデルによるレビュー文のキャッチフレーズ文体への変換”, 沈辰夫, 莊司慶行, 山本岳洋, 大原剛三,
田中克己, 電子情報通信学会論文誌 D, Vol. 18, No.5, pp.332-343, 2025.

本稿では、レビュー文をキャッチコピーのような文体に変換する方法を提案する。例えば、カメラのレビュー中
の「ポケットに入れていても邪魔にならない」というような単純なレビュー中の一文を入力すると、「ポケットに
インして、気軽にスナップ。」のように、キャッチーな文体に変換可能にする。このような変換をローカルで動作
可能な小規模な LLM (大規模言語モデル) で実現するために、1) プロンプトを考慮したファインチューニング、
2) In-context Learning を伴う変換候補の生成、3) 生成された候補の再ランキング、という 3 種類の工夫を施した。
実際に、書籍「キャッチコピー大百科」から人手で作成した 1,300 件程度のコーパスを用いて Open-Calm をファ
インチューニングし、実レビューサイトから収集したレビュー文を用いて被験者実験を行った。実験結果から、そ
れぞれの工夫が想定通りに効果を発揮し、すべてを組み合わせることで最も魅力的な変換を行えることを示した。

キーワード LLM, テキストマイニング, 商品レビュー, キャッチコピー

1. はじめに

近年、amazon や楽天といったオンラインショッ
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
ネットの商品を選び、購入するようになってきている。
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^(注1)。

こうした状況下で、ある商品カテゴリの中から、自
分の求めている性質をもつ商品を探すことは一般的
なタスクになりつつある。この際、商品の性質を知るた
めに、商品のスペックだけでなく、商品レビューを読
むことが一般的である。一方で、レビューの数は膨大
で、それぞれのレビューは長いので、すべてのレビュー
を読むことは現実的ではない。そこで、多くの場合、
レビュー文を検索し、最も商品特徴を表していたり、
個人の求める性質に関与するレビュー文を読む。この

[†] 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

Aoyama Gakuin University
Fuchinobe, Sagami-hara-shi, Kanagawa 252-8603, Japan

^{††} 静岡大学 情報学部 〒430-0807 静岡県 浜松市 中区 城北
Shizuoka University

Johoku, Hamamatsu-shi, Shizuoka 432-8011, Japan

^{†††} 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

Gakuennishi-machi, Nishi-ku, Kobe, Hyogo, 651-2197, Japan

^{††††} 関西学院大学 特定プロジェクト研究センター 〒669-1330 兵庫
県 三田市 学園上ヶ原

Gakuen-uegahara, Sanda-shi, Hyogo 669-0013, Japan

a) E-mail: shen@sw.it.aoyama.ac.jp

b) E-mail: shojiy@inf.shizuoka.ac.jp

c) E-mail: t.yamamoto@sis.u-hyogo.ac.jp

d) E-mail: ohara@it.aoyama.ac.jp

e) E-mail: tanaka-katsumi@fukuchiyama.ac.jp

DOI:10.14923/transinfj.????????

(注1)：インターネットショッピングおよ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>

際、単純に切り出してきたレビュー文は、あくまで長い文章の一部なので、読みづらかったり、商品の魅力が伝わらなかったりする。

そこで、本研究では商品の特長や用途が記述されている任意のレビュー文中の1文を入力として与えると、そのレビュー文の内容を反映したキャッチコピー文体に変換する手法を提案する。具体的には、「運動会やアスレチックなどで激しく動く子どものような動体でもピントが合います」というような記述を、「動きの速い被写体も、確実にキャッチ!」という、短くてキャッチーな文に変換する。

このような変換を実現するために、本研究では、近年主流になりつつある生成的大規模言語モデル (Generative Large Language Model, Generative LLM) を利用する。一般的な言語モデルは、主に文章の続きを書くことに長けており、要約や翻訳などに広く用いられ始めている。こうした言語モデルで効率よく、高精度なキャッチコピー文体への変換を実現するために、本研究では、

- (1) キャッチコピーの文体と変換タスクを学習するための、大規模なコーパスを用いた追加学習 (Fine Tuning),
 - (2) レビュー文と少量のキャッチコピーの例示による変換候補の生成 (Few-Shot Example による In-context Learning), および
 - (3) 生成された変換候補のキャッチー度合いと内容の正確さに基づく再ランキング
- という3段階の工夫を施した。

実際の手法の動作例を、図1に示す。はじめに、この手法ではファインチューニングとして、大規模なキャッチコピーコーパスを用いて、言語モデルにカテゴリ名からレビューを生成するタスクを解かせる。具体的には、

“商品：「デジカメ」 -> キャッチコピー：「

という文の続きとして、

“笑顔に、ズーム”

を推論するタスクでモデルをファインチューニングする。こうすることで、モデルは、キャッチコピー全体の大まかな文体と、ある商品分野でどのような表現が使われがちかを学習できる。

次に、こうして学習したモデルに、いくつかのレビュー文と、それらを正しく変換した文のペアを例として与えたうえで、実際に任意の文を変換させる。具

体的には、

“商品：「デジカメ」

レビュー文：「子供の写真を撮るのに最適で、特に顔がはっきり写ります。」

-> キャッチコピー：「笑顔に、ズーム」…

“商品：「デジカメ」

レビュー文：「【任意のレビュー文】」

-> キャッチコピー：「

という文の続きとして、実際に与えられたレビュー文をキャッチコピー文体に言い換えた文を生成させる。

この際、与える例を変えながら変換を繰り返すと、複数の変換候補を得られる。これらの変換候補に対して、その変換候補がキャッチーな文体であるか、元のレビューと内容が同一であるかを計算し、両方の値の高い順に並び替える。このような手順を踏むことで、元のレビュー文から、キャッチコピーらしい文体で、なおかつ元のレビュー文の意味を保った変換が可能になる。

このような手法の有用性と、3つの工夫それぞれの有効性について検証するために、被験者実験を行った。実験のために、優れたキャッチコピーを網羅的に掲載している書籍[1]から、人手で1,300件程度のコーパスを作成した。このコーパスを用いて日本語LLMであるOpen-CALMをファインチューニングしキャッチコピー変換機を実装した。また、LLMであるBERTをファインチューニングして、ある文がキャッチコピーであるかを分類する2値分類器、ある文とある文が意味が同一かを判定する分類器をそれぞれ実装した。これらを用いた被験者実験を通して、

- ファインチューニングを行うことでその分野のキャッチコピーらしい文を生成可能にできること、
- レビュー文とキャッチコピーのペアを例として与えることで、レビュー文の内容を保った生成が可能にできること、および
- 2つの分類器を用いることで、キャッチコピーらしさ、内容の元レビュー文との近さを判定できること

が明らかになり、それぞれの工夫が正しく動作し、すべてを組み合わせたときに最も商品の魅力を伝えやすい変換が行えることが分かった。

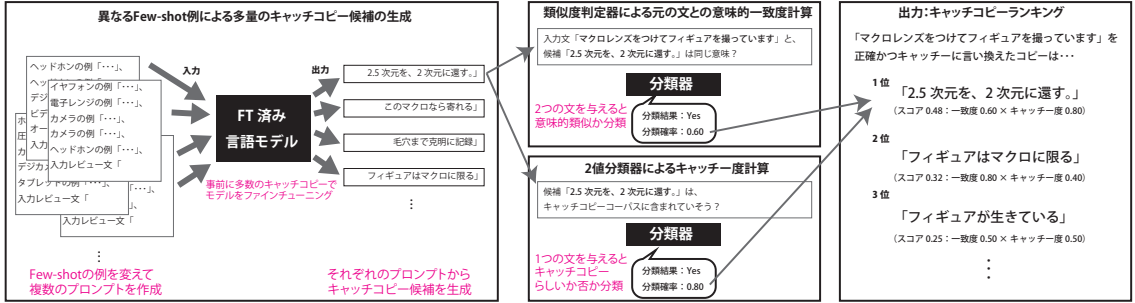


図1 アルゴリズム全体の概要。In-Context Learning に基づいて生成したキャッチコピー候補のランキング。入力レビュー文と内容の一致度・キャッチ度から算出される最終スコアを基にランキングされる。

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、各分野の先行研究を紹介し、本研究で用いた技術と位置づけについて論じる。

2.1 ユーザーレビューを利用した研究

近年ではユーザ投稿レビュー数が増加してきており、これらを有効活用する研究が多く行われている。例えば Hu ら [2] は、多数の商品レビューから商品の特長について述べたものをマイニングし、要約を作成する手法を提案している。

ユーザーレビュー情報を用いてアイテムを推薦する手法も一般的になりつつある。Zheng ら [3] は、ユーザーレビューから商品やサービスの特性とユーザー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。江田ら [4] 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。これらの研究は、商品購入の意思決定や商品推薦システムの構築にユーザーレビューが有用であることを示している。

2.2 推薦フレーズの生成に関する研究

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 [5] 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 [6] 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 [7] 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 [8] 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 [9] 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。大曾根ら [10] は、既存の広告文から抽出した複数のキーワードやカテゴリを入力として、それらに適するような広告文を生成するように、GPT-2 を転移学習する手法を提案している。

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、生成されるキャッチコピーが商品の特長を反映しているか、という点に重きを置いている。

2.3 テキストのスタイル変換に関する研究

本研究では、レビュー文をキャッチコピーに変換するが、これは広義にテキストのスタイル変換である。大規模言語モデルを用いたテキストのスタイル変換や、言い換え文の生成の研究は、近年、盛んに行われている。Liu ら [11] は、元の文に含まれる任意の単語を、言語モデルによって予測された別の単語に置換する手法によって言い換え文を生成し、言い換え文生成

タスクを最適化問題として解くことで最適な言い換え文を得る手法を提案している。Reif ら [12] は、プロンプティング手法である Augmented zero-shot learning prompt を提案し、センチメントやフォーマルかインフォーマルかであったり、コミカルであるか、ドラマチックであるか等の様々なスタイル変換について評価している。本研究はこれらの研究と似ているが、レビュー文からキャッチコピーという、大きく書き換えてほしいが内容の正確性は保ちたいという、特殊な状況下でのテキスト変換を目指している。

3. 提案手法

本手法では、ある商品に対する 1 つのレビュー文とその商品のカテゴリを入力すると、元のレビューと意味的に類似し、なおかつキャッチーな順にランキングされた複数のキャッチコピーを出力する。

このような処理を実現するために、書籍から商品カテゴリとキャッチコピーからなるコーパスを作成した。そして LLM をファインチューニングし、モデルにキャッチコピーの文体や、カテゴリごとの語彙や文法の違いを学習させた。こうして学習されたモデルにキャッチコピー化したいレビュー文を与え、キャッチコピーの文体へと変換させた。この際、いくつかの変換例をプロンプトとして与えることで、よりレビュー文の特徴をつかんだキャッチコピーが生成されるようにした。最後に、複数生成したキャッチコピーについて、別の言語モデルを用いて、キャッチーさと内容の正しさを計算し、ランキングした。

3.1 書籍からのキャッチコピーコーパスの作成

はじめに、ファインチューニング用のデータとして、商品カテゴリごとのキャッチコピーを列挙したキャッチコピーコーパスを作成した。本研究では、書籍「カタログ・チラシ キャッチコピー大百科 [1]」の内容を、人手でコンピュータに打ち込んで、コーパスとした。カタログ・チラシ キャッチコピー大百科は、「食品」「生活用品」「ファッション」など、12 個のカテゴリに分かれている。今回の実験では、その中から「デジタル関連製品」および「家庭用電化製品」のカテゴリの 1,322 個のキャッチコピーをファインチューニングに用いた。さらに、それらのキャッチコピーに対して、人手で「デジタルカメラ」「イヤホン」といったようなより細粒度の商品カテゴリのラベル付けを行った。

キャッチコピーコーパス

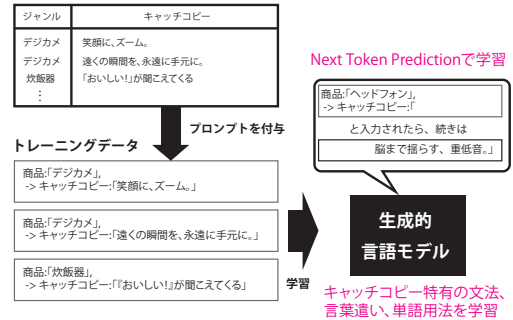


図2 プロンプトを考慮したファインチューニングの例。書籍をもとに人手でキャッチコピーコーパスを構築し、商品ジャンルとキャッチコピーとプロンプトからなる学習データを作成した。Next Token Prediction タスクで言語モデルに続きの文を予測させることで、通常の日本語の文とは異なる、キャッチコピー特有の表現方法を学習させる。

3.2 キャッチコピーコーパスを用いたプロンプトを考慮したファインチューニング

本手法では、キャッチコピーコーパスを用いて生成的言語モデルのファインチューニングを行う。プロンプトを考慮したファインチューニングの例を、図2に示す。

ファインチューニングを行う際に、学習データを、推論時のプロンプトを考慮して、テンプレートに埋め込んだ形式にした。具体的には、“商品:「[商品カテゴリ名]」-> キャッチコピー:「[キャッチコピー本文]」”という形式のテンプレートを用いた。たとえば商品が「デジタルカメラ」、キャッチコピーが「笑顔に、ズーム。」であった際には、学習用の1件のデータは、“商品:「デジタルカメラ」-> キャッチコピー:「笑顔に、ズーム。」”となる。このようなテンプレートに当てはめることで、言語モデルは“キャッチコピー:「」”から始まるカッコの中には、日本語と似ているが文法や語彙の異なる、キャッチコピーが入ることを学習できると考えられる。

このような学習データでファインチューニングすることで、生成的言語モデルの推論時に、文の続きを推定するタスクで、任意の文をキャッチコピーに変換可能になる。たとえば、“商品:「ヘッドホン」-> キャッチコピー:「」”のように、学習データを途中で打ち切ったような形式のプロンプトを入力する。すると、“商品:「ヘッドホン」-> キャッチコピー:「脳まで揺らす、重低音。」”のように、与えられたプロンプトの続きとして、キャッチコピー部分が生成される。

3.3 In-context Learning に基づくプロンプトを用いたキャッチコピー候補の生成

本手法では、In-context Learning に基づいたプロンプティング手法を用いてキャッチコピー候補を生成する。テキストのスタイル変換を行う際に、いきなり変換したい文をモデルに与えずに、先にいくつかの変換例をプロンプトとして与えることで、変換精度が向上することが知られている [12]。このような手法は Few-Shot プロンプティングと呼ばれており、このような生成時に例を与えてその場限りの学習をさせることは In-context Learning と呼ばれる。

ファインチューニングに加えてこのような手法をとった理由は、ファインチューニングで解くタスクと、実際に解きたいタスクが異なるためである。書籍から構築したコーパスでは、商品カテゴリとキャッチコピーが列挙されている。そのため、こうしたコーパスで学習すると、カテゴリごとのキャッチコピーで使われがちな表現や、キャッチコピー特有の文法などは学習できる。一方で、ある文を別の表現に変換するには、変換前と変換後の文のペアが必要である。今回の場合、このようなペアを大規模に用意することは、困難である。一般的なキャッチコピーはコピーライターがゼロから考えており、変換元の文が存在しないためである。

そこで本研究では、コーパス内のキャッチコピーからいくつかのキャッチコピーを抜き出し、それらの変換元を人手で独自に付与した。具体的には、ある商品カテゴリのキャッチコピーについて、それに関連しそうなレビュー文を、実レビューサイトから人手で探し、紐づけした。このような、キャッチコピーの変換元になりそうなレビュー文を人手で探すことは、ハイコストである。そこで、ファインチューニングではなく、生成時の Few-Shot 例として、モデルに与えることにした。

具体的な Few-Shot プロンプティングによるキャッチコピー候補の生成の例を図 3 に示す。この手法では、カテゴリ、レビュー文、キャッチコピーの三つ組を 1 つの例としてモデルに与える。例は、異なる商品カテゴリのものを、複数個与える。そして最後に、変換したい商品について、商品カテゴリと、レビュー文だけを与え、キャッチコピー候補を生成させる。

この際、キャッチコピーへの変換を 1 回で終わらせず、与える例を変えて、何度もキャッチコピー候補を生成するようにした。このような Few-Shot で与える

レビュー文

「マクロレンズをつけてフィギュアを撮っています」

のキャッチコピー候補への変換

実レビュー文とコーパス中の文のペアを Few-shot の例としてランダムに 5 個列挙

生成用プロンプト

商品：「ヘッドホン」、商品レビュー：「ノイキャンがすごいので、毎朝の通勤で使っています。」-> キャッチコピー：「満員電車が、コンサートホールに。」
商品：「ヘッドホン」、商品レビュー：「軽いののでどこにでも持ち運んでいます。」-> キャッチコピー：「ポケットに楽団を持ちこよう」
商品：「デジタルカメラ」、商品レビュー：「ハイキングで野鳥を取るのに使っています。」-> キャッチコピー：「自然の中の一瞬を切り出す」
商品：「ビデオカメラ」、商品レビュー：「簡単操作なので私でも子供の成長記録がとれます。」-> キャッチコピー：「泣き顔も、笑顔も。今日の君を、永遠に残す」
商品：「オープンレンジ」、商品レビュー：「ダイエット中だけど、ヘルシー料理が簡単に作れるから安心です」-> キャッチコピー：「カラダに優しいレシピ」
商品：「デジタルカメラ」、商品レビュー：「マクロレンズをつけてフィギュアを撮っています」-> キャッチコピー：「

... (いくつかの具体例) ...
商品：「デジタルカメラ」、
商品レビュー：「マクロレンズをつけてフィギュアを撮っています」
-> キャッチコピー：「

と入力されたので、続きは

2.5 次元を、2 次元に運ぶ。」

キャッチコピー化

生成的
言語モデル

実際にキャッチコピー化したい文

図 3 In-Context Learning に基づくキャッチコピー候補の生成。人手で作成した「商品カテゴリ、レビュー文、キャッチコピー」の組からなるデータセットから 5 件の例をランダムに抽出し、Few-Shot 学習の例としてプロンプトに含める。最後に、実際にキャッチコピーに変換したいレビュー文を入力し、その後の空白に来る文を生成させる。

例を変えて複数の出力を得るやり方はアンサンブルと呼ばれ、高い精度のテキスト生成を可能にする [13]。今回の実験時には、人手で作成した 56 個のデータから、5 個取り出したものを 1 回分の例として扱った（つまり 5-shot）。5 個の例からなるプロンプトを、ランダムに 10 個作り、10 回、キャッチコピー候補を生成させた。

3.4 キャッチコピー候補のランキング

本手法では、別に用意した 2 種類の言語モデルを用いて、キャッチコピー候補をランキングする。生成したキャッチコピー候補のランキングの例を図 1 に示す。ランキングのために、それぞれのキャッチコピー候補に対して、入力レビュー文との内容一致度と、キャッチ度をそれぞれ計算する。そして、それらを合算したスコアで候補を並び替え、最終的なキャッチコピーランキングとして出力する。

はじめに、キャッチコピー候補と入力レビュー文の内容一致度を計算する。本手法では、内容一致度を計算するために、学習済み日本語言語モデルである Sentence-LUKE [14] モデルを用いた。これは、Sentence-BERT [15] と同じタスクで学習したモデルである。具体的には、ある文章とそれに類似する文章のペアを学習データとし、類似した文章から生成される

文章ベクトルが似たベクトルになるように学習されている。このモデルに、元のレビュー文とキャッチコピー候補をそれぞれ入力し、それぞれの文章ベクトルのコサイン類似度を計算する。このようにして計算した値を、キャッチコピー候補と入力レビュー文の内容一致度のスコアとする。また、このモデルは、ほとんど元の商品レビュー文と同様な文章で生成されたキャッチコピー候補のフィルタリングにも利用する。

次に、それぞれのキャッチコピー候補のキャッチ度を計算する。本手法では、キャッチコピー候補のキャッチ度を計算するために、汎用言語モデルである BERT [16] をファインチューニングして用いた。具体的なファインチューニングの手順として、キャッチコピーと商品レビュー文が同量含まれる、ラベル付きの学習データを作成した。このデータで、ある文がキャッチコピーか商品レビュー文であるかを見分けるタスクで BERT をファインチューニングした。このようにしてファインチューニングした BERT は、任意の文章を入力すると、その文をキャッチコピー、もしくは商品レビュー文のどちらかに分類する。このモデルにキャッチコピー候補を入力文章として入力し、キャッチコピー、もしくは商品レビュー文のどちらかに分類する。このときに計算されるキャッチコピーへの分類確率をキャッチ度のスコアとする。

最後に、内容一致度とキャッチ度の両方を考慮した、最終スコアを計算する。最終スコアは、内容一致度とキャッチ度の重み付きの和として定義した。すなわち、入力されたレビュー文を r 、あるキャッチコピー候補を c_i として、内容一致度を $s_{\text{content}}(r, c_i)$ 、キャッチ度を $s_{\text{catchy}}(c_i)$ と定義した際、最終スコア $s_{\text{final}}(r, c_i)$ は、

$$s_{\text{final}}(r, c_i) = s_{\text{content}}(r, c_i)^\alpha s_{\text{catchy}}(c_i)^{1-\alpha} \quad (1)$$

である。ただし、この際、重み α は内容一致度を重視する度合いを表し、0 から 1 の値を任意で設定できる。この最終スコア $s_{\text{final}}(r, c_i)$ の大きい順で候補を並び替えることで、キャッチコピーのランキングを作成する。

4. 評価

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

本研究では、大規模なキャッチコピーコーパスと、実サービスのレビューデータを用いて実験を行った。キャッチコピーコーパスは、書籍「カタログ・チラシキャッチコピー大百科 [1]」をもとに人手で作成した。この本には、商品の大きなカテゴリごとに実際に社会的に評価の高かったキャッチコピーが辞典形式で掲載されている。これらのうち、1,322 個のキャッチコピーを、人手で入力した。この際、各キャッチコピーについて、実際にそのキャッチコピーの対象となった商品名から、細粒度の商品カテゴリを、人手で付与した。

実サービスのレビューデータとして、価格.com^(注2) からレビュー文を収集した。これらのレビューのうち一部は、キャッチコピーと人手で紐づけた。具体的には、キャッチコピーコーパスの中から、商品の特長や用途をアピールしているようなキャッチコピーを 56 個、人手で選出した。そして、そのキャッチコピーが付く商品にふさわしいような商品レビュー文と人手で紐づけ、レビュー文とキャッチコピーが対応する少数のデータを作成した。

4.2 実装

本研究で用いる大規模言語モデルの詳細や実装について述べる。キャッチコピー候補を生成する生成的言語モデルとして、サイバーエージェント社による OpenCALM^(注3) を用いた。このモデルについて、キャッチコピーコーパスを用いてプロンプトを考慮したファインチューニングを施し、Few-Shot のプロンプトをアンサンブルで用いてキャッチコピー候補を生成した。Few-Shot プロンプトのアンサンブルについて、人手で作成したレビュー文とキャッチコピーが対応する 56 個のデータの中から 5 つ選択して 1 つのプロンプトを作成した。また、そのような組み合わせのプロンプトを 10 個作成し、これら 10 個のプロンプトのアンサンブルでキャッチコピー候補を生成した。このとき、選んだ 5 つの例の中で商品カテゴリが 3 つ以上重複しないようにプロンプトを作成し、プロンプト内の商品カテゴリの比率が偏らないようにした。

実際の学習時には、使用機材のスペックに合わせ、バッチサイズを 4、勾配累積を 16、エポック数を 20

(注2)：価格.com - 「買ってよかった」をすべてのひとに。：

<https://kakaku.com/>

(注3)：cyberagent/open-calm-3b：

<https://huggingface.co/cyberagent/open-calm-3b>

に設定した。この際、学習率は $2e-4$ とした。

キャッチコピー候補と入力レビュー文の内容一致度を計算するために用いる言語モデルとして、Studio Ousia 社の日本語 LUKE モデルを元に学習された Sentence-LUKE モデル^(注4)を用いた。このモデルは、内容一致度の計算だけでなく、キャッチコピー候補のクレンジングにも用いる。生成的言語モデルが生成したキャッチコピー候補の中には、元の商品レビュー文とほぼ同じ文章が含まれる場合がある。こういった候補を取り除くために、このモデルで内容一致度 $s_{\text{content}}(r, c_i)$ が 0.85 より大きかった場合、キャッチコピー候補から除外した。

キャッチコピー候補のキャッチ度を計算するために用いる言語モデルとして、東北大学が公開している BERT モデル^(注5)を用いた。このモデルを、キャッチコピーコーパスから抽出したキャッチコピーと、価格.com から抽出した商品レビュー文を同じ割合で含むラベル付きの学習データで、2 値分類タスクでファインチューニングした。

実際の学習には、キャッチコピー 500 件、レビュー 500 件に対して、学習率 $2e-5$ 、エポック数 3 で、cls トークンを用いて分類を行うタスクを用いた。

実験の際に、内容一致度とキャッチ度の両方を考慮した最終スコアの計算時に、内容一致度を重視する度合いである重み α を 0.2 とした。この値は、キャッチコピー候補内での内容一致度のスコアのばらつきが、キャッチ度のスコアのばらつきよりも比較的大きかったため、予備実験を通して経験的に設定したものである。

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。本実験では、ファインチューニングの効果と、Few-Shot プロンプティングの効果とをそれぞれ確認するために、提案手法と比較手法をそれぞれ作成した。具体的には、

- **提案手法**：本研究で提案している、生成的言語モデルにファインチューニングを施し、Few-Shot プロンプティングでキャッチコピー候補を生成、入力レビュー文との内容一致度・キャッチ度でランキングする手法、

- **ファインチューニングなし**：生成的言語モデルはそのまま用いて、Few-Shot プロンプティングでキャッチコピー候補を生成、入力レビュー文との内容一致度・キャッチ度でランキングする手法、
- **Zero-Shot プロンプティング**：生成的言語モデルにファインチューニングを施し、プロンプトでは例を与えずにシード値を変えて複数回生成することで複数のキャッチコピー候補を生成、入力レビュー文との内容一致度・キャッチ度でランキングする手法、
- **ランキングなし**：生成的言語モデルにファインチューニングを施し、Few-Shot プロンプティングでキャッチコピーを生成、別途ランキングは行わない手法、
- **ベースライン**：生成的言語モデルをそのまま用いて、Zero-Shot プロンプティングでキャッチコピーを生成する手法、
- **S_{content} のみ**：生成的言語モデルにファインチューニングを施し、Few-Shot プロンプティングでキャッチコピー候補を生成、入力レビュー文との内容一致度のみでランキングする手法、
- **S_{catchy} のみ**：生成的言語モデルにファインチューニングを施し、Few-Shot プロンプティングでキャッチコピー候補を生成、キャッチ度のみでランキングする手法、

の 7 つの手法である。

ここで使われている Zero-Shot プロンプティングを用いた手法は、キャッチコピー候補の生成時に、プロンプト中で例を与えない。つまり、「商品：「デジタルカメラ」、商品レビュー：「マクロレンズをつけてフィギュアを撮っています」→キャッチコピー：」のような短いプロンプトだけから、続きとしてキャッチコピーを生成する手法である。シード値を変えて複数回生成することで、複数の候補を生成した。

4.4 実験方法

本提案手法によって生成されたキャッチコピーを、被験者実験によって評価する。具体的には、30 個のキャッチコピーの元となる商品レビュー文を用意し、それぞれ 1 つの商品レビュー文に対して生成された、各比較手法につき 3 つのキャッチコピーを評価する。このとき実験に用いる商品レビュー文については、そのレビュー文が確かにその商品の特長や用途を記述しているか、を基準に我々が選定した。実際に評価する

(注4)：sonoisa/sentence-luke-japanese-base-lite :

<https://huggingface.co/sonoisa/sentence-luke-japanese-base-lite>

(注5)：cl-tohoku/bert-base-japanese-v3 :

<https://huggingface.co/tohoku-nlp/bert-base-japanese-v3>

表 1 各比較手法のそれぞれの評価項目に対する平均評点 (1 から 5, ベースラインと比べて ** $p < 0.01$, * $p < 0.05$. 魅力向上度はキャッチコピーの魅力とレビューの魅力の差, 一致係数は Krippendorff の α)

評価項目	キャッチー度	内容反映度	変換度	魅力向上度
提案手法	**3.65	** 3.80	3.79	** 0.02
Zero-Shot	** 3.86	**3.23	** 4.15	** -0.29
FT なし	**3.77	**3.68	3.90	**0.01
ランキングなし	**2.68	3.09	4.01	** -0.59
ベースライン	1.93	2.83	3.82	-1.40
S_{content} のみ	**2.97	**4.15	3.23	** -0.13
S_{catchy} のみ	**4.06	3.03	4.18	** -0.23
(一致係数)	0.55	0.55	0.32	0.39

項目は,

- (1) キャッチー度: キャッチコピーは確かにキャッチーであるか,
- (2) 内容反映度: キャッチコピーは元の商品レビュー文の内容を反映しているか,
- (3) 変換度: レビュー文とキャッチコピーを比較して, どれくらい変換されているか,
- (4) レビュー魅力アピール度: レビュー文が商品の魅力をどれくらいアピールできているか,
- (5) キャッチコピー魅力アピール度: キャッチコピーが商品の魅力をどれくらいアピールできているか,

の 5 項目である. それぞれの項目について, 5 段階のリッカート尺度で評価した.

被験者は 3 名で, 全員が日本人の情報系学部の大学生である.

4.5 実験結果

本節では, 4.4 節における被験者実験の結果を述べる. 実験結果から, 提案手法等と生成的言語モデルをそのまま用いたベースラインの手法とを 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. 各比較手法のそれぞれの評価項目に対する平均評点をまとめて比較したものを表 1 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, 変換度を除くすべての評価項目において有意差が認められた.

被験者間の意見一致係数として, クリッペンドルフの α を算出した. 本タスクの内容は主観的に左右されがちなものであるが, すべての評価項目で一致傾向が見られ ($\alpha > 0$), キャッチー度と内容反映度の評価では, 過半数のキャッチコピーにおいて評点が一致した.

5. 考 察

本章では, 評価実験によって得られた結果について考察し, 提案アプローチの妥当性と限界について議論する. 実験結果から, 提案手法はベースライン手法と比較して, すべての評価項目において有意差が認められた. このことから, ファインチューニング, プロンプティング, ランキングによる後処理を行うことで, 単純に生成的言語モデルをそのまま用いるよりも, キャッチーかつ, レビュー文の内容を反映し, 商品の魅力をアピールできるキャッチコピーを生成できることがわかった. さらに, いくつかの観点に注目して考察を述べる.

はじめに, キャッチコピーコーパスを用いたプロンプトを考慮したファインチューニング手法について考察する. まず, 提案手法と, ファインチューニングなしの手法を比較すると, どの評価項目においても有意差は認められず, 評点はほとんど同様であった.

ファインチューニングの効果がなかった原因のひとつに, ファインチューニングのタスクが実際の変換タスクと異なることの影響が考えられる. 学習時に, モデルはカテゴリ名だけをプロンプトとして受け取り, キャッチコピーを生成するタスクで学習を行った. 一方で, 実際の生成時には, カテゴリ名とレビュー文がプロンプトとして与えられる. 実際の生成時のプロンプトが学習時のプロンプトと異なったため, モデルはその場で与えられたプロンプトと, モデルが元から持っていた言語的知識を用いてタスクを解いた可能性がある. 学習時と異なる, 長いプロンプトが与えられた場合, ファインチューニングの効果は限定的になることが示唆されている.

一方で, Zero-Shot プロンプティングと, ファインチューニングなしの手法を比較すると, キャッチー度と内容反映度で有意差が認められた. このことから, プロンプティング手法による影響よりは小さいが, プロンプトを考慮したファインチューニング手法は生成文をよりキャッチーするように影響していると考えられる.

次に, In-context Learning に基づく Few-Shot プロンプティング手法について考察する. 提案手法と, Zero-Shot プロンプティングの手法を比較すると, キャッチー度と内容反映度で有意差が認められた. このことから, Few-Shot プロンプティングによって, 生成文がよりレビューを反映するように影響していると考えら

表2 各手法で被験者からの評点が相対的に高かった生成キャッチコピーの例

カテゴリ	レビュー文	手法	生成されたキャッチコピー
デジタルカメラ	ポケットに入れていても邪魔にならず快適です!	提案手法 Zero-Shot FT なし ランキングなし ベースライン	ポケットにインして、気軽にスナップ。 ずっと付き合いたくなるポケットサイズ。 いつでも、どこでも、どんなシーンでも。 旅行やレジャーにもぴったり。 コンパクトなのに、高画質。
デジタルカメラ	運動会やアスレチックなどで激しく動く子どものような動体でもピントが合います。	提案手法 Zero-Shot FT なし ランキングなし ベースライン	動きの速い被写体も、確実にキャッチ! 動いているものを、ピンポイントで。 スポーツシーンでも、動く被写体でも、ピントが合う。 「動くもの」を撮るなら、このカメラ。 動く被写体でも、ピントが合う。
イヤホン	かなりの重低音で、ライブ会場さながらの迫力で思わず体を揺らしてしまうようなノリの良いサウンドが聞けます。	提案手法 Zero-Shot FT なし ランキングなし ベースライン	まるで、ライブ会場にいるような臨場感。 もう、ボータブルでライブ会場です。 それは、ライブ会場さながらのサウンド。 音楽好きにはたまらない、ライブ会場さながらのサウンド。 重低音が魅力のイヤホン
炊飯器	しっとり且つお米の一粒一粒が立っており、炊きあがりの香りや艶も問題なしでした。	提案手法 Zero-Shot FT なし ランキングなし ベースライン	ふっくらと、つやつやと。お米の味わい方、無限大。 ひと粒ひと粒に感動を。 ふっくら、つやつや、おいしいお米を炊く。 炊飯器なのに、まるで土鍋で炊いたような艶やかさ。 ふっくら美味しいご飯が炊ける
エアコン	パワーも十分、音も静かで、大変満足しています。	提案手法 Zero-Shot FT なし ランキングなし ベースライン	静かで、パワフルで、快適。 静かな寝室にも、パワーあり。 省エネで、快適で、静か。 静かだから、寝室にもぴったり。 快適空間を創造する、エアコン

れる。

この際、Zero-Shot プロンプティングではよりキャッチーな文章が生成され、Few-Shot Example を用いるとよりレビューを色濃く反映した文章が生成される傾向が示された。ファインチューニングでは、カテゴリ名という限られた内容から、独創的なキャッチコピーらしい文を生成するようにモデルが学習されている。そのため、モデルは抽象的でキャッチコピーらしい文体の文を、多数生成した。一方で、Few-Shot プロンプティングによってレビュー文とキャッチコピーのペアを与えられると、モデルはキャッチーさを犠牲にして、変換元レビューと内容的に近い文を、例を参考に生成するようになった。キャッチーさとレビュー内容反映度は、ある程度、トレードオフの関係にある可能性がある。

直観的には、Few-Shot 例にレビュー文と対応するキャッチコピーを多く含めば、よりキャッチーな文が多く生成されると予想されたが、実際には例を追加するとキャッチー度が下がる傾向がみられた。これは、Few-Shot 例に用いたキャッチコピーが、データセット全体から見ると特殊なものだったためである可能性がある。書籍に掲載されるような優れたキャッチコピーの多くは、抽象的なものであった。一方で、人手で選んだキャッチコピーはデータセット全体から見ると少数の、レビューに関連しそうなものに限られた。その

ため、これらを例に含めることで、逆にレビューと対応しないような候補が生成されづらくなったと考えられる。実際にキャッチコピーを生成するという実運用においては、このトレードオフを考慮して、後処理でのキャッチコピーらしさを重視する重み α の値を調整したり、Few-Shot 例の数を調整するなどの対応が考えられる。

次に、レビューとの内容一致度およびキャッチー度でのランキング手法について考察する。提案手法と、ランキングなしの手法を比較すると、キャッチーな文章であるか、およびレビューを反映しているかの2項目において有意差が認められた。このことから、別に用意した2種類の言語モデルによる、レビューとの内容一致度、およびキャッチー度でのランキング手法は有効であると考えられる。さらに、 S_{content} のみの手法とランキングなしの手法を比較すると、内容反映度の評点が高く、レビューとの内容一致度によるランキングが機能していることがわかる。また、 S_{catchy} のみの手法とランキングなしの手法を比較すると、キャッチー度の評点が高く、キャッチー度によるランキングが機能していることがわかる。

次に、プロンプトを考慮したファインチューニングや Few-Shot プロンプティング手法による生成文への影響について考察する。ベースライン手法の生成する文章の中には、商品カテゴリのようなものが含まれた

りしていたが、これはプロンプトの形式を学習できていないからであると考えられる。他の手法では、そのようなことがほとんどなかったことから、プロンプトを考慮したファインチューニングや Few-Shot プロンプティング手法は、生成文を正しくキャッチコピーらしいものにすることに對して有効であったと考えられる。

次に、キャッチコピーの魅力アピール度について考察する。キャッチコピーの魅力アピール度とレビューの魅力アピール度の差である魅力向上度では、提案手法が一番評点が高かった。しかし、提案手法のキャッチコピー魅力アピール度と、レビュー魅力アピール度の評点はほぼ同様であった。実験に用いたレビュー文は商品の長を具体的に記述しているものが多かったが、生成されたキャッチコピーの中には抽象的すぎる表現や、焦点がやや逸れているように見受けられるものが含まれていた。このことが一因であると考えられる。

各手法で被験者からの評点が相対的に高かった生成キャッチコピーの例を表 2 に示す。提案手法、Zero-Shot プロンプティング、ファインチューニングなしの各手法については、キャッチかつ、レビュー文の内容を反映したキャッチコピーが生成できていることがわかる。

最後に、本研究で採用したキャッチコピー生成における、ファインチューニングと In-context Learning を組み合わせた変換方式について、妥当性と限界について議論する。本研究で採用した手法では、ファインチューニング時と実際のキャッチコピー変換時に、異なるタスクを用いている。ファインチューニング時は任意のカテゴリのキャッチコピーを生成するタスクで学習し、キャッチコピー変換時にはカテゴリ名と入力レビュー文からキャッチコピーを生成している。本研究では採用しなかったが、これらの行程を分けずに、直接、入力レビュー文からキャッチコピーへの変換を行うタスクでファインチューニングを行うアプローチも考えられる。

それぞれのキャッチコピーに対して、そのキャッチコピーと関連しそうなレビュー文を機械的に選択し、キャッチコピーとレビュー文のペアを作成することができる。たとえば、コーパス中のそれぞれのキャッチコピーから、そのキャッチコピーを特徴づける単語を抽出し、その単語を含むレビューをそのキャッチコピーの変換元レビューと見なせる。そうすると、レ

ビュー文からキャッチコピーへの変換そのものをファインチューニングタスクとしてモデルを学習させることができる。

一方で、機械的に発見した関連レビューを正解として学習するアプローチをとった場合、出力の多様性の低下が懸念される。多くの場合、キャッチコピーは比喩表現や婉曲表現を用いており、直接的にレビュー文に表れるような観点名を含まないことが多い。あるキャッチコピーが製品のどの長を言い表しているかは、人間がキャッチコピーの内容を吟味し、意味を理解してはじめて判断がつく場合が多い。機械的にキャッチコピーに含まれる単語を含むレビューとペア付けして学習を行った場合、生成されるキャッチコピーは、直接的にレビュー中の単語を含むものばかりになる。今回のような限られたデータから、ファインチューニングだけで、多様で比喩的な、元のレビュー文中の単語に縛られないような変換を学習することは困難である。

将来的な課題として、十分に質の高いレビュー文とキャッチコピーのペアを機械的に作成することが考えられる。任意のレビュー文と任意のキャッチコピーのペアに対して、「直接的には類似しないが、よく読むと関連している」かを判定する分類機が作成できれば、教師データを自動で作成できる。多くのキャッチコピーについて関連するレビューを手で紐づけし、対照学習などで学習すれば、十分な精度の分類機が作成できる可能性がある。今回のアプローチでは、人手で 56 個しかペアを作成できなかったため、Few-Shot プロンプトとして用いるアプローチを採用した。今後、より多くの正解ペアが作成できれば、直接的にレビュー文をキャッチコピーに変換するタスクでファインチューニングしたモデルで、より高度なキャッチコピー変換が実現できる可能性がある。

6. ま と め

本稿では、レビュー文を魅力的なキャッチコピーに変換する方法を提案した。具体的には、「ポケットに入れていると邪魔にならない」のようなレビュー文を入力すると、「ポケットにインして、気軽にスナック。」のような創造的なキャッチコピーを出力する。そのために、大規模言語モデルをプロンプト付きのデータセットでファインチューニングし、Few-Shot プロンプティングでキャッチコピーの候補を多数生成した。生成された候補を、入力レビューとの内容の一致度・キャッ

チ一度を考慮して並び替え、キャッチコピーをランキング化した。実際に生成されたキャッチコピーを被験者に評価させることで、生成されたキャッチコピーがキャッチで、内容を反映しており、商品選びの参考になることが示された。

今回の研究では、任意のレビューを入力するとその文をキャッチコピーに言い換える部分だけを可能にした。実際に商品を比較検討する際には、その商品の特徴や魅力がよく表れたレビュー文を特定する必要がある。商品特徴を考慮したキャッチコピー生成などの、継続的な研究を行う予定である。

謝 辞

本研究の一部は JSPS 科研費 24K03228, 22H03905, JP21H03554, 21H03775 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] 久野寧子 (編), カタログ・チラシ キャッチコピー大百科, ビエ・ブックス, 2008.
- [2] M. Hu and B. Liu, “Mining and summarizing customer reviews,” Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, p.168–177, KDD '04, Association for Computing Machinery, New York, NY, USA, 2004. <https://doi.org/10.1145/1014052.1014073>
- [3] L. Zheng, V. Noroozi, and P.S. Yu, “Joint deep modeling of users and items using reviews for recommendation,” Proceedings of the Tenth ACM International Conference on Web Search and Data Mining, pp.425–434, 2017.
- [4] 江田怜央, 牛尾剛聡, “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.
- [5] X. Zhang, Y. Zou, H. Zhang, J. Zhou, S. Diao, J. Chen, Z. Ding, Z. He, X. He, Y. Xiao, B. Long, H. Yu, and L. Wu, “Automatic product copywriting for e-commerce,” Proceedings of the AAAI Conference on Artificial Intelligence, vol.36, pp.12423–12431, 2022.
- [6] Y. Zhang, Y. Wang, L. Zhang, Z. Zhang, and K. Gai, “Improve diverse text generation by self labeling conditional variational auto encoder,” ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)IEEE, pp.2767–2771 2019.
- [7] R. Sinha and K. Swearingen, “The role of transparency in recommender systems,” CHI '02 Extended Abstracts on Human Factors in Computing Systems, pp.830–831, Association for Computing Machinery, New York, NY, USA, 2002. <https://doi.org/10.1145/506443.506619>
- [8] Z. Chan, Y. Zhang, X. Chen, S. Gao, Z. Zhang, D. Zhao, and R. Yan, “Selection and generation: Learning towards multi-product advertisement post generation,” Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp.3818–3829, Association for Computational Linguistics, Online, Nov. 2020. <https://aclanthology.org/2020.emnlp-main.313>
- [9] S. Deng, C.-W. Tan, W. Wang, and Y. 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.
- [10] 大曾根宏幸, 張 培楠, “Gpt-2 の転移学習によるキーワードを考慮した広告文生成,” 人工知能学会全国大会論文集, vol.JSAI2021, pp.2D4OS7b03–2D4OS7b03, 2021.
- [11] X. Liu, L. Mou, F. Meng, H. Zhou, J. Zhou, and S. Song, “Unsupervised paraphrasing by simulated annealing,” Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, eds. by D. Jurafsky, J. Chai, N. Schluter, and J. Tetreault, pp.302–312, Association for Computational Linguistics, Online, July 2020. <https://aclanthology.org/2020.acl-main.28>
- [12] E. Reif, D. Ippolito, A. Yuan, A. Coenen, C. Callison-Burch, and J. Wei, “A recipe for arbitrary text style transfer with large language models,” Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), eds. by S. Muresan, P. Nakov, and A. Villavicencio, pp.837–848, Association for Computational Linguistics, Dublin, Ireland, May 2022. <https://aclanthology.org/2022.acl-short.94>
- [13] Z. Jiang, J. Araki, H. Ding, and G. Neubig, “How can we know when language models know? on the calibration of language models for question answering,” Transactions of the Association for Computational Linguistics, vol.9, pp.962–977, 2021. <https://aclanthology.org/2021.tacl-1.57>
- [14] I. Yamada, A. Asai, H. Shindo, H. Takeda, and Y. Matsumoto, “LUKE: Deep contextualized entity representations with entity-aware self-attention,” Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), eds. by B. Webber, T. Cohn, Y. He, and Y. Liu, pp.6442–6454, Association for Computational Linguistics, Online, Nov. 2020. <https://aclanthology.org/2020.emnlp-main.523>
- [15] N. Reimers and I. Gurevych, “Sentence-BERT: Sentence embeddings using Siamese BERT-networks,” Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), eds. by K. Inui, J. Jiang, V. Ng, and X. Wan, pp.3982–3992, Association for Computational Linguistics, Hong Kong, China, Nov. 2019. <https://aclanthology.org/D19-1410>
- [16] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of deep bidirectional transformers for language understanding,” Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), eds. by J. Burstein, C. Doran, and T. Solorio, pp.4171–4186, Association for Computational Linguistics, Minneapolis, Minnesota, June 2019. <https://aclanthology.org/N19-1423>

Abstract This paper proposes a method for transforming a review text into a catchy tagline-like phrase. For instance, given a simple review statement such as “It does not get in the way when in my pocket,” the proposed method outputs catchy phrases like “Slides into your pocket, snap with ease.” To achieve such transformations using large-scale language models (LLM), the method employs three techniques: 1) fine-tuning with prompts, 2) generating transformed candidates with in-context learning, and 3) re-ranking of generated candidates. The Open-Calm model is fine-tuned using a Catchy-Copy Corpus created manually from a book. Subject experiments are conducted using the model and review data from actual review sites. The experimental results demonstrate that each technique effectively contributes as expected, and their combination enables the most appealing transformations.

Key words LLM, text mining, product review, catchphrase