

AdParaphrase: 魅力的な広告表現の分析を目的とした 広告文言い換えデータセット

村上 聡一郎¹ 張 培楠¹ 上垣外 英剛^{2,3} 高村 大也³ 奥村 学³

¹ 株式会社サイバーエージェント ² 奈良先端科学技術大学院大学 ³ 東京科学大学

{murakami_soichiro,zhang_peinan}@cyberagent.co.jp

kamigaito.h@is.naist.jp {takamura,oku}@pi.titech.ac.jp

概要

広告の成功には人々を惹きつける効果的な言葉選びが欠かせない。本研究は広告文の言語表現に焦点を当て、どのような言語的特徴を持つ広告文が好まれるか明らかにすることを目的とし、選好評価データ付きの広告文言い換えデータセット AdParaphrase を提案する。AdParaphrase は広告文の言い換えペアから構成され、選好評価データを含む。これにより人々が魅力的に感じる広告表現の分析が可能となる。実験では広告文の言語的特徴量と選好評価データの関係を分析し、魅力的な広告文の特徴を明らかにした。またこれらの知見や提案データセットを活用し、魅力的な広告文を生成する手法を探求した。

1 はじめに

広告の目的は人々の注意を引き付け、クリックや購入等の行動を促すことである。そのためには人々の興味関心を引く内容を書くことが重要だが、それだけではない。その内容をどう表現して伝えるか、すなわち、広告の言語表現も広告の成功には欠かせない。本研究では広告文を魅力的にする言語表現に焦点を当て、どのような表現を持つ広告文が好まれ魅力的と感じるか明らかにすることを目的とする。

しかし言語表現の魅力度を分析する上で2つの課題に直面する。1つ目は広告文の内容と表現の切り分けの難しさである。例えば、広告文「アディダス50%OFF」と「ナイキ半額」の比較で後者が好まれた場合、ブランド名(ナイキ)という内容が要因か「50%OFF」や「半額」等の表現が要因かを特定することは容易ではない。2つの広告文の表現の差異に着目するためには内容を統一した上で比較する必要がある。2つ目は広告文の魅力度を分析するためのオープンデータセットが不足している点である。一

表1 AdParaphrase の例。()内は各文を選好した人数。

初回購入で最大50%割引(0) ↔ 初回購入最大50%オフ(9)
【公式】マイナビバイト(8) ↔ マイナビバイト公式(0)
業界一の安さ(3) ↔ 業界トップクラスの低価格(7)

般に広告文の評価ではクリック率等の実績値や選好評価が用いられる。しかし広告文の選好評価データはこれまで公開されておらず、広告魅力度に寄与する要因を分析する上で障壁となっている¹⁾[1, 2]。

そこで本研究では、これら2つの課題を解決するために選好評価データ付きの広告文言い換えデータセット AdParaphrase を提案する²⁾。表1に例を示す。AdParaphrase は表現が異なるが同じ内容の広告文ペア、すなわち、言い換えペアから構成され、各ペアに対して10名の選好評価データが付与されている。これにより人々が魅力的に感じる広告表現の分析が可能となる。実験では AdParaphrase を用いて各文に含まれる言語的特徴量と選好評価データの関係を分析し、魅力的と感じる広告文の特徴を明らかにした (§3.1)。さらに分析で明らかになった知見や選好評価データを活用し、魅力的な広告文を生成する手法を探求した (§3.2)。その結果、分析で得られた知見が広告文の魅力度改善に寄与することを示した。

2 言い換えデータセットの構築

データセット構築手順は言い換え候補の収集 (§2.1)、言い換え判定 (§2.2)、選好評価データの収集 (§2.3) からなる。本研究では言い換え元の広告文として、広告文データセット CAMERA [3] を用いる。

2.1 言い換え候補の収集

言い換え候補の収集は、(1) 広告ライターによる言い換え例の作成、(2) 大規模言語モデル (LLM)

1) 特にクリック率などの実績値は多くの企業にとって秘匿情報に当たり、それが障壁の要因の一つとなっている。

2) 構築したデータセットは公開予定である。

表2 言い換えペアの収集結果 (件数)

モデル	生成結果	フィルタ後	言い換え
CALM2-7B [4]	16,365	2,107	1,173
CALM3-22B [5]	16,365	6,287	4,551
Swallow-8B [6, 7]	16,365	4,942	3,623
Swallow-70B [8, 7]	16,365	5,226	4,174
Crowdworker	5,000	3,775	2,939
合計	70,460	22,337	16,460

やクラウドワーカーによる言い換え候補の作成の2段階で実施した。言い換え作成は理想的には広告制作の専門家である広告ライターに依頼することが考えられるがリソースの限界により大規模には難しい。そこで広告ライターには広告文の魅力的な言い換え例の作成を依頼し、それらを参考事例としてLLMやクラウドワーカーにより大量の言い換え候補を作成する方法を採用した。

広告ライターによる言い換え例の作成 言い換え例の作成を広告ライター2名に依頼した。広告ライターには、原文をより魅力的な表現に言い換えること、全角15文字以内の文長制約³⁾を遵守することを指示した。100件程度の言い換え例を作成するように依頼し、133件の言い換え例が得られた。言い換え元の文は、CAMERAの開発セットから抽出した。

LLM/クラウドワーカーによる言い換え候補作成 LLMではIn-context learning (ICL) [9]により言い換え候補を作成する。具体的には広告ライターの言い換え例をFew-shot事例として与え、CAMERAの全広告文の言い換えを生成した。なお、LLMとして学習データやモデルサイズが異なる複数のモデルを使用した。クラウドワーカーによる作成ではクラウドソーシングを活用し、LLMと同じ指示文や言い換え例をガイドラインとして提示した。ワーカーの多くは広告制作の経験が乏しいと予想されるため、広告制作の知見をガイドライン⁴⁾に加えた。例えば、「行動を促す言葉を使う」や「重要な情報を先頭に書く」である。言い換え元にはCAMERAから無作為抽出した5,000件を使用した。表2にLLMとクラウドワーカーが作成した言い換え候補数を示す。

2.2 言い換え判定

言い換え候補70,460件が言い換えかを判定するアノテーションを実施した。まず作業効率化とデー

3) 一般的にウェブ広告では入稿する広告文に文長制約が設けられている。本研究ではGoogle広告等の検索連動型広告における見出し文の文長制約である全角15文字を採用した。

4) 言い換え作成のガイドラインを付録Aに示す。

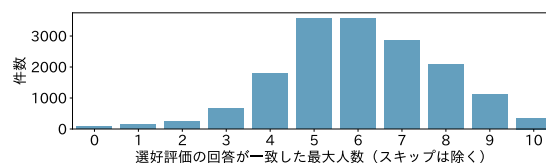


図1 言い換えペアに対する選好評価の分布

タ品質の向上のためルールによるフィルタリングを実施した。これには(1)明らかに言い換えではない事例(例:日付や金額が異なる)と(2)文長制約を超過する事例を予め除外する目的がある。ウェブ広告では文長制約を超える文は入稿できないため、文長制約を満たす広告文を作成することが求められる。そのため実用上の観点から(2)を分析対象外にすることが妥当と判断し、データセットから除外した。

言い換え判定はクラウドソーシングにより各文ペアに対して5名で実施した。評価者には文ペアを提示し、言い換えとして成立するか二値で判定するよう依頼した。各事例の最終的なラベルは多数決で決定した。表2にルールによるフィルタリングと言い換え判定の結果を示す。16,460件が言い換えペアとして判定された。Inter-Annotator Agreement (IAA)は0.442 (Fleiss' Kappa [10])と中程度の一致である。

2.3 選好評価データの収集

言い換えペア16,460件に対して選好評価を実施した。評価はクラウドソーシングにより各文ペアに対して10名で実施した。評価者には各文ペアを提示し、より魅力的と感じる広告文を選択するよう依頼した。2つの広告文の魅力度が同じ場合は「スキップ」を選択するように指示した。また選好評価の主観的な性質に配慮し、Wangらのガイドライン [11]を参考に評価観点を複数例示した。例えば「クリックしたいか」「理解しやすいか」等の観点がある。

図1に評価結果のヒストグラムを示す。X軸は10名中で回答が一致した最大人数を示す(スキップは除く)。例えば、6は10名中6名が同じ広告文を選好し、0は全員がスキップしたことを表す。この結果から広告文の言い換えペアに対する選好評価は全体的には回答が一致しづらいことが分かった。評価データ全体のIAAは0.167 (Fleiss' Kappa)である。一方、評価データ全体の約22%を占める3,570件では10名中8名以上の選好が一致した。これらのIAAは0.480 (Fleiss' Kappa)である。また、この結果は広告文ペアの表現の差異が評価者の選好に影響を与えた可能性があることを示唆している。

表3 χ^2 二乗検定の結果

特徴量		df	N	χ^2	p 値
基本特徴量	文長	文字	1 2,925	721.25	< 0.01
		単語	1 2,725	678.43	< 0.01
語彙的特徴量	内容語	名詞	1 1,406	326.61	< 0.01
		動詞	1 535	6.94	< 0.01
		形容詞	1 99	0.88	0.35
		副詞	1 127	0.68	0.41
		単語頻度	1 2,657	70.54	< 0.01
統語的特徴量	語彙選択	一般名詞	1 1,397	288.12	< 0.01
		固有名詞	1 152	7.58	< 0.01
		ひらがな	1 2,047	23.24	< 0.01
文体的特徴量	文字種	カタカナ	1 601	42.57	< 0.01
		漢字	1 1,503	257.72	< 0.01
		記号	1 2,332	795.93	< 0.01
統語的特徴量	係り受け	深さ(最大)	1 1,914	16.93	< 0.01
		長さ(最大)	1 2,349	1.89	0.17
	その他	Perplxity	1 3,570	223.26	< 0.01
文体的特徴量	感情	Joy	1 693	70.18	< 0.01
		Anticipation	1 683	89.29	< 0.01
	その他	具体性	1 186	116.44	< 0.01
		括弧	1 1,667	1372.57	< 0.01

3 実験

選好評価 (§2.3) により 3,570 件の事例で 10 名中 8 名以上の選好が一致することが分かった。そこで実験ではこれらの事例に焦点を当て、2 文のどのような表現の差異が選好に影響を与えたか分析する (§3.1)。さらに分析で明らかになった知見を踏まえ、魅力的な広告文の生成手法を探求する (§3.2)。

3.1 選好に影響を与える広告表現の分析

実験の目的は言い換えペアの選好評価に影響を与えた要因を明らかにすることである。特に言い換えペアの表現の差異に着目し、どのような言語的特徴を持つ広告文が好まれる傾向があるか分析する。

3.1.1 特徴量

広告の役割は人々の注意を引き付け、商品やサービスに興味を向けることである。そのため広告文の視認性や可読性、情報量は魅力度改善において重要な観点である [1, 12]。実験では広告文の表現やスタイルに関する様々な特徴量を定義し、選好評価との関係を分析した。特徴量は基本特徴量、語彙的特徴量、統語的特徴量、文体的特徴量に大別される。

基本特徴量 可読性や情報量に関連する基本的な特徴量として文長（文字数、単語数）を使用した。

語彙的特徴量 語彙的特徴量は内容語の数、語彙選択、文字種である。内容語を多く含む文は情報量が高く魅力的と仮説を立てた。語彙選択についてはより一般的な単語を含む広告文の方が好まれると仮説を立て、BCCWJ [13] に基づく平均単語頻度を算出した。加えて一般名詞と固有名詞の数も算出した。また各文に含まれる文字種の数も算出した。

統語的特徴量 統語的特徴量は文全体やその一部の構造に関する特徴量である。具体的には係り受け木の深さや依存リンクの長さ、Perplexity を含む。

文体的特徴量 文体的特徴量は広告文の言い回しやスタイルに関する特徴量である。本研究ではよりポジティブまたは具体的な広告文は好まれると仮定し、各文の感情と具体性に関するラベルを導入した。これらのラベルは感情と具体性を判別する独自の分類器を構築し判定した。分類器の詳細は付録 B を参照されたい。さらに、広告文で広く用いられる括弧記号 (【, 「」) の有無も特徴量として用いる。これらの括弧記号は「【公式サイト】ABC 保険」のように視認性向上や重要情報の強調に使用される。

3.1.2 分析方法

各特徴量と選好評価の関係を分析するために、独立性の χ^2 二乗検定を使用した。本手法は、2 つのカテゴリ変数の独立性を検証するものであり、本研究では (1) 多数の評価者に好まれた広告文と (2) 各特徴量の大小関係の関連性を検証する。例えば Perplexity の場合、好まれた広告文とスコアの関係进行分析する。また分析では広告文ペアの表現の差異に着目するために各特徴量のスコアが異なる文ペアを分析対象とした。よって特徴量ごとに事例数が異なる。例えば文字数が異なる文ペアは 2,925 件である。

3.1.3 実験結果

表 3 に χ^2 二乗検定の結果を示す。分析の結果、複数の特徴量が選好評価と有意な関係を持つことが明らかとなった ($p < 0.01$)。例えば、文字数や名詞の数、係り受けの深さ、Perplexity、括弧の有無などが選好評価と有意な関係を持つことが分かった。一方で形容詞や副詞の数、係り受け木の依存リンクの長さについては有意な関係は見られなかった。

また各特徴量と選好評価のクロス集計によると、例えば次のような特徴を持つ広告文が好まれる傾

表 4 広告文生成実験の評価結果 (%)

モデル	言い換え	魅力度	魅力度&文長
CALM3-22B-zeroshot	74.0	23.0	12.8
CALM3-22B-zeroshot-findings	74.0	42.6	23.0
CALM3-22B-fewshot-findings	85.0	38.8	31.2
CALM3-22B-instruct-zeroshot	90.5	31.5	29.3
CALM3-22B-dpo-zeroshot	70.5	84.4	8.5
Swallow70B-zeroshot	90.5	15.5	8.3
Swallow70B-zeroshot-findings	80.0	44.4	17.5
Swallow70B-fewshot-findings	86.5	40.5	26.0
Swallow70B-instruct-zeroshot	94.0	18.6	17.6
Swallow70B-dpo-zeroshot	62.5	71.2	8.0
GPT-4o-zeroshot	86.0	12.8	12.8
GPT-4o-zeroshot-findings	95.5	39.3	34.6
GPT-4o-fewshot-findings	92.5	37.8	33.5
Crowdworker	89.1	23.9	22.3

向があることが分かった: 文字数が多い, 名詞の数が多し, 係り受けの深さが小さい, Perplexity が低い (流暢性が高い), 括弧記号を含む. 以上の結果から, 魅力的な広告文を作成するためにはこれらの特徴量を考慮することが重要であると示唆される.

3.2 広告文自動生成

各特徴量と選好評価の関係分析 (§3.1) により明らかになった知見を踏まえ, 魅力的な広告文を生成する手法を探求する. 本実験では与えられた広告文を情報の追加削除なしで魅力的な表現に言い換える広告文生成タスクに焦点を当てる [14, 15].

3.2.1 実験設定

生成手法 LLM を用いた生成手法を探求する. 分析で得られた知見や評価者の選好を LLM に導入する方法は複数考えられるが, 広告文生成においてどの手法が有用かは明らかではない. そこで実験では ICL, Instruction tuning (IT) [16], Direct preference optimization (DPO) [17] を例として, 各学習手法に基づく生成手法を比較する. ICL では入力文を魅力的な表現に言い換えるように基本的な指示のみを与える zeroshot, 分析で明らかになった知見⁵⁾を加えた zeroshot-findings, 言い換え例を複数与えた fewshot-findings の 3 パターンのプロンプトを検証した. IT では選好評価で少数の評価者に選好された広告文を入力, 多数に選好された広告文を出力として追加学習した. また IT 済みのモデルに対して DPO

5) 本実験では文字数の多さ, 流暢性の高さ (Perplexity の低さ), 括弧記号の利用の 3 点を魅力的な広告文を作成するための知見としてプロンプトに記載した (付録 C を参照).

で選好チューニングを実施した. 付録 D.1 に詳細を記載する. モデルは CALM3-22B [5], Swallow70B [7], GPT-4o [18] を用いた. また zeroshot と同じ指示でクラウドワーカーが作成した言い換えも評価する.

選好データセット 実験では AdParaphrase を選好チューニング向けに再構成したデータを使用する. 具体的には, まず言い換え原文 x と 2 つのモデルが生成した言い換え文 y_1, y_2 の三つ組 (x, y_1, y_2) を作成する. その後 y_1 と y_2 に対して新たに 10 名の選好評価データを収集した. これにより原文 x と選好評価付きの生成文 y_1^{pref}, y_2^{pref} からなる三つ組データ $(x, y_1^{pref}, y_2^{pref})$, 合計 8,721 件を構築した. なお, 学習, 開発, 評価用に 9:0.5:0.5 の比率で分割した.

評価方法 生成文が (1) 言い換えか, (2) 魅力的か, (3) 魅力的かつ文長制約を満たすかの観点で評価した. (1) と (2) は §2.2 及び §2.3 と同手順で人手評価し, 過半数が言い換え, 魅力的と判定した割合を報告する. (3) は過半数が魅力的と判定かつ文長制約の全角 15 文字を満たす割合を算出した. なお, 評価には評価データから抽出した 200 件を用いた.

3.2.2 実験結果

表 4 に実験結果を示す. 各評価観点について考察する. 生成文が言い換えと判定された割合は IT により追加学習した手法が高い傾向があった. ICL では fewshot-findings が zeroshot に比べて改善する傾向を確認した. 生成文が魅力的と判定された割合については DPO で選好チューニングしたモデルが高い傾向にあることを確認した. ICL では zeroshot に対して知見を加えた zeroshot-findings が生成文の魅力度が高い傾向だった. これにより言語的特徴量と選好評価の関係性分析 (§3.1) で明らかになった知見が生成文の魅力度改善に寄与することが示唆された. 一方で生成文が文長制約を満たしかつ魅力的と判定された割合については zeroshot-findings が最も高い傾向だった. DPO では文長制約を超えた文を生成する傾向があり, 生成文の情報量が増えたことで魅力的と判定される割合が高くなったと考えられる. 付録 D.2 に生成例を示す.

4 おわりに

本研究では魅力的な広告表現の分析を目的とした広告文言い換えデータセット AdParaphrase を提案した. 今後の方向性として評価者の属性情報を考慮した分析や広告文生成手法の改善などが考えられる.

参考文献

- [1] Soichiro Murakami, Sho Hoshino, and Peinan Zhang. Natural language generation for advertising: A survey, 2023.
- [2] Reid Pryzant, Sugato Basu, and Kazoo Sone. Interpretable neural architectures for attributing an ad’s performance to its writing style. In Tal Linzen, Grzegorz Chrupala, and Afra Alishahi, editors, **Proceedings of the 2018 EMNLP Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP**, pp. 125–135, November 2018.
- [3] Masato Mita, Soichiro Murakami, Akihiko Kato, and Peinan Zhang. Striking gold in advertising: Standardization and exploration of ad text generation, 2024.
- [4] Ryosuke Ishigami. cyberagent/calm2-7b-chat, 2023. Hugging Face.
- [5] Ryosuke Ishigami. cyberagent/calm3-22b-chat, 2024. Hugging Face.
- [6] Naoaki Okazaki, Sakae Mizuki, Youmi Ma, Koki Maeda, Kakeru Hattori, Masanari Ohi, Taihei Shiotani, Koshiro Saito, Rio Yokota, Kazuki Fujii, Taishi Nakamura, Takumi Okamoto, Ishida Shigeki, and Hiroya Takamura and. tokyotech-llm/llama-3.1-swallow-8b-instruct-v0.1, 2024. Hugging Face.
- [7] Kazuki Fujii, Taishi Nakamura, Mengsay Loem, Hiroki Iida, Masanari Ohi, Kakeru Hattori, Hirai Shota, Sakae Mizuki, Rio Yokota, and Naoaki Okazaki. Continual pre-training for cross-lingual LLM adaptation: Enhancing japanese language capabilities. In **First Conference on Language Modeling**, 2024.
- [8] Naoaki Okazaki, Sakae Mizuki, Youmi Ma, Koki Maeda, Kakeru Hattori, Masanari Ohi, Taihei Shiotani, Koshiro Saito, Rio Yokota, Kazuki Fujii, Taishi Nakamura, Takumi Okamoto, Ishida Shigeki, and Hiroya Takamura and. tokyotech-llm/llama-3.1-swallow-70b-instruct-v0.1, 2024. Hugging Face.
- [9] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel Ziegler, Jeffrey Wu, Clemens Winter, Chris Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. Language models are few-shot learners. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, editors, **Advances in Neural Information Processing Systems 33**, Vol. 33, pp. 1877–1901, 2020.
- [10] J.L. Fleiss, et al. Measuring nominal scale agreement among many raters. **Psychological Bulletin**, Vol. 76, No. 5, pp. 378–382, 1971.
- [11] Xiting Wang, Xinwei Gu, Jie Cao, Zihua Zhao, Yulan Yan, Bhuvan Middha, and Xing Xie. Reinforcing pre-trained models for generating attractive text advertisements. In **Proceedings of the 27th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining**, pp. 3697–3707, 2021.
- [12] Hsueh-Cheng Wang and Marc Pomplun. The attraction of visual attention to texts in real-world scenes. **Journal of Vision**, Vol. 12, No. 6, pp. 26–26, 06 2012.
- [13] Kikuo Maekawa, Makoto Yamazaki, Takehiko Maruyama, Masaya Yamaguchi, Hideki Ogura, Wakako Kashino, Toshinobu Ogiso, Hanae Koiso, and Yasuharu Den. Design, compilation, and preliminary analyses of Balanced Corpus of Contemporary Written Japanese. In Nicoletta Calzolari, Khalid Choukri, Bente Maegaard, Joseph Mariani, Jan Odijk, Stelios Piperidis, Mike Rosner, and Daniel Tapias, editors, **Proceedings of the Seventh International Conference on Language Resources and Evaluation**, pp. 1483–1486, 2010.
- [14] Brit Youngmann, Elad Yom-Tov, Ran Gilad-Bachrach, and Danny Karmon. The automated copywriter: Algorithmic rephrasing of health-related advertisements to improve their performance. In **Proceedings of The Web Conference 2020**, pp. 1366–1377, 2020.
- [15] Shaunak Mishra, Manisha Verma, Yichao Zhou, Kapil Thadani, and Wei Wang. Learning to create better ads: Generation and ranking approaches for ad creative refinement. In **Proceedings of the 29th ACM International Conference on Information and Knowledge Management**, pp. 2653–2660, 2020.
- [16] Jason Wei, Maarten Paul Bosma, Vincent Zhao, Kelvin Guu, Adams Wei Yu, Brian Lester, Nan Du, Andrew Mingbo Dai, and Quoc V. Le. Finetuned language models are zero-shot learners. In **The Tenth International Conference on Learning Representations**, 2022.
- [17] Rafael Rafailov, Archit Sharma, Eric Mitchell, Stefano Ermon, Christopher D. Manning, and Chelsea Finn. Direct preference optimization: your language model is secretly a reward model. In **Advances in Neural Information Processing Systems 36**, 2023.
- [18] OpenAI. Hello gpt-4o, 2024. Accessed: 2025-01-03.
- [19] Tomoyuki Kajiwara, Chenhui Chu, Noriko Takemura, Yuta Nakashima, and Hajime Nagahara. WRIME: A new dataset for emotional intensity estimation with subjective and objective annotations. In Kristina Toutanova, Anna Rumshisky, Luke Zettlemoyer, Dilek Hakkani-Tur, Iz Beltagy, Steven Bethard, Ryan Cotterell, Tanmoy Chakraborty, and Yichao Zhou, editors, **Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies**, pp. 2095–2104, 2021.
- [20] Ikuya Yamada, Akari Asai, Hiroyuki Shindo, Hideaki Takeda, and Yuji Matsumoto. LUKE: Deep contextualized entity representations with entity-aware self-attention. In Bonnie Webber, Trevor Cohn, Yulan He, and Yang Liu, editors, **Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)**, pp. 6442–6454, 2020.
- [21] OpenAI. GPT-4 technical report, 2024.
- [22] Tim Dettmers, Artidoro Pagnoni, Ari Holtzman, and Luke Zettlemoyer. Qlora: efficient finetuning of quantized llms. In **Advances in Neural Information Processing Systems 36**, 2023.

提示した広告文がさらに魅力的になるよう「言い換え」を作成してください。言い換えとは「言い回しは異なるものの、意味的に同じ文」のことを指します。

【言い換える条件】

- 以下の条件に従うよう言い換えを作成して下さい。
- ・必ず、全角15文字以内で書いてください
- ・広告文に新たな情報を追加したり、含まれる情報を削除しないでください
- ・感嘆符（！）は利用しないでください

【言い換えるコツ】

- 言い換えるコツを以下に示します。ただし、これらのコツに必ず従う必要はありません。
- ・より分かりやすくなるよう語順を変えてみる
 - ・簡単な言葉を使う
 - ・よりキャッチな言葉を使う
 - ・同じ意味を持つ言葉を使う（例：おすすめ → 人気）
 - ・記号で装飾する（例：公式 → 【公式】）
 - ・かな漢字アルファベット表記を変える（例：TOP3 → トップ3）
 - ・重要な情報を前に持ってくる（例：今日中にお金借りる → お金を今日中に借りる）
 - ・より抽象的な表現に変える（例：10%OFF → お得）
 - ・より具体的な表現に変える（例：若者に人気 → 20代に人気）
 - ・行動を促す言葉を使う（例：お得な情報 → お得な情報をチェック）
 - ・カジュアルな言葉を使う（例：お金が必要 → お金が欲しい）
 - ・疑問文に変える（例：お金が必要 → お金が必要？）

【作成例】

広告文：ネット申込みで最大21,000円割引
 言い換え：ネット申込【最大¥21,000割引】

図2 言い換え作成のガイドライン

表5 広告文生成実験の生成例。

モデル	生成例
Swallow-70B	
zeroshot	2022年版おすすめクレカ5選
zeroshot-findings	【2022年版】おすすめクレカ5選を紹介
fewshot-findings	【2022年版】おすすめクレカ5選を紹介
instruct-zeroshot	【2022年版】おすすめクレカ5社
dpo-zeroshot	【2022年版】おすすめクレジットカード5選比較
Crowdworker	2022年人気のクレカTOP5
入力文	【2022年版】おすすめクレカ5選

A 言い換え作成のガイドライン

言い換え候補の作成 (§2.1) でクラウドワーカーに提示したガイドラインを図2に示す。

B 言語的特徴量の詳細

広告文の感情及び具体性ラベルを判定する独自の分類器を構築した。各分類器の詳細を以下に示す。

感情 各広告文の感情ラベルを判定するために WRIME [19] で学習された LUKE⁶⁾ [20] を用いた。本モデルは8つの感情ラベル (joy, sadness, anticipation, surprise, anger, fear, disgust, trust) の中から最も相応しいラベルを予測する8クラス分類器である。本研究ではポジティブな表現を持つ広告文は好まれやすいと仮説を立て、joy及びanticipationと判定された広告文を分析に使用した。WRIMEの評価データに対する正解率は68.6%である。

具体性 本研究ではより具体的に書かれた広告文は好まれやすいと仮説を立て、広告文の具体性を判

6) <https://huggingface.co/Mizuiro-sakura/luke-japanese-large-sentiment-analysis-wrime>

あなたはプロの広告ライターです。検索連動型広告の制作を担当しています。以下の条件に従って提示した広告文の表現をさらに魅力的に言い換えてください

条件

- 全角15文字以内で書いてください
- 広告文に新たな情報を追加したり、含まれる情報を削除しないでください
- 感嘆符（！）は利用しないでください
- 以下に魅力度を改善できた事例を示します。これらの例を参考に、魅力度を改善できるように広告文を言い換えて下さい。

魅力的な広告文を作るコツ

- 「**【**」や「**」**」などの括弧記号を活用する
- 文字数制限（全角15文字）いっぱいを書く
- より流暢な文章にする

分析で明らかになった知見

例

魅力度を改善できた事例 (20件)

入力: 中学受験に強い個別指導塾
 出力: 中学受験に特化した個別指導塾

入力: 高収入を得る吉祥寺の副業で稼ぐ
 出力: 吉祥寺で高収入の副業を始めよう

入力: 法人向け【即戦力採用】

出力: 即戦力採用【法人向け】

回答

入力: 【言い換え元の広告文】

出力: 【生成文】

図3 広告文生成実験で使用したプロンプト

定する分類器を GPT-4 [21] により構築した。本モデルは与えられた2つの広告文（言い換えペア）のうち具体性の高い広告文（広告1または広告2）を選択する。具体性が同じ場合は「equal」を選択する。したがって、3クラス分類器である。予測結果の中からランダムに選択した100件を人手評価したところ、正解率は88.0%であった。

C プロンプト例

図3に広告文生成実験 (§3.2) で使用したプロンプト (fewshot-findings) を示す。

D 広告文生成実験の詳細

D.1 実験設定

fewshot-findings に与えた言い換え例は学習データから20件サンプリングした。このうち、少数の評価者が選好した広告文を入力、多数が選好した広告文を出力とした (図3を参照)。またITおよびDPOではQLoRA [22] を用い、1 epoch 学習した。実装は本コード⁷⁾を使用した。

D.2 広告文の生成例

表5にSwallow-70B [7, 8] 及びクラウドワーカーによる広告文の生成例を示す。これらの生成文のうち、評価者の過半数が魅力的と判定した文はSwallow70B-dpo-zeroshotだった。それ以外の生成文では言い換え元の入力文が魅力的と判定された。

7) <https://github.com/ghmagazine/llm-book>