

AdTEx：大規模言語モデルを活用した多様な広告テンプレート拡張フレームワーク

大竹 啓永¹ 張 培楠² 坂井 優介¹ 大内 啓樹^{1,2} 渡辺 太郎¹
¹ 奈良先端科学技術大学院大学 ² 株式会社サイバーエージェント
 otake.hiroto.od2@naist.ac.jp
 zhang_peinan@cyberagent.co.jp
 {sakai.yusuke.sr9, hiroki.ouchi, taro}@is.naist.jp

概要

表現や内容に関する複数の制約を満たした広告文の人手による作成は、多くの時間を要する。そのため、あらかじめテンプレートを作成しておき、商材ごとに当てはめることで、似た商材に対する広告生成を効率的に行う広告配信手法がしばしば採用される。しかし、テンプレートを用いた手法では、広告文が似た表現となるため、多様性低下が問題となる。そのため、多様なテンプレートを作成することで、表現の多様性を向上させることが重要である。本研究では、大規模言語モデルを用いて既存の広告テンプレートを拡張する手法、AdTExを提案する。実験結果より、本手法では内容の整合性や類似性、広告配信に向けた形式の遵守に関してさらなる向上が必要となる一方で、文字数制限を満たしつつ、表層の多様性を向上させたテンプレート拡張が可能であることがわかった。

1 はじめに

広告配信プラットフォームの発展により、オンライン広告はデジタルマーケティングにおける中核的な役割をなしている。中でも検索型連動広告は、検索キーワードと関連した広告が表示される [1] ため、広告主は宣伝対象の商品と関連性を持ち、訴求力の高い広告文を作成する必要がある。さらに、文字数制限や構文的制約、内容の類似性といった条件を満たす必要があり、人手による作成は時間を要する。

この課題に対処するため、広告テキスト生成 (ATG) の研究が広く行われてきた。先行研究では、テンプレートベースの手法 [2, 3] や文書から重要情報を抽出する手法 [4, 5]、要約生成や文書書き換えとして定式化した手法 [6] が提案され、大規模言

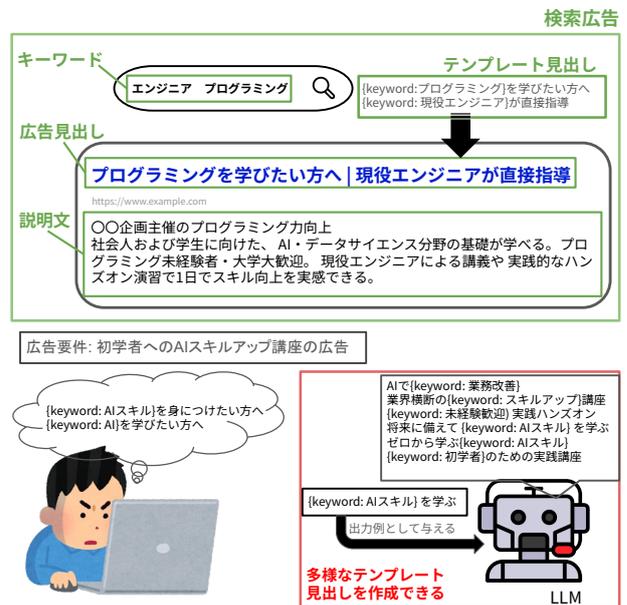


図1 検索連動型広告におけるテンプレート作成プロセスの自動化: 検索連動型広告では検索キーワードとの関連性が高い広告見出しと説明文が表示される。LLMにより制約を守った多様なテンプレートを効率的に作成できる。

語モデル (LLM) を含む自然言語生成 (NLG) 技術が ATG に適用されている [7, 8, 9]。

LLM は高い生成能力を有する一方で、文字数制限を満たせない場合があり [10]、制御性の低さが指摘されている。広告以外の分野では制御性の問題に対し、テンプレートベースの手法が検討されている [11]。しかし、広告分野におけるテンプレートベースの手法では、人手による作成に多くの時間を要するうえ、自動生成においても、事前に定義された構造や語彙的制約の影響により、表現の多様性が制限されやすい。類似した広告文が繰り返し表示された時に、広告に対して否定的な印象を抱く広告疲労 [12] が生じる。そのため、広告への関心を維持する観点から、多様性の確保が重要となる。

図1に示すように、本研究では、多様な広告テンプレートの作成を促進するために、テンプレート拡張フレームワーク **AdTE_x** を提案する。NLGにおける多様性は、表層的多様性と意味的多様性に大別されるが[1]、広告分野では表層的多様性が注目されている[13]。しかし、表層的多様性のみを追求したテンプレートでは、内容の一貫しない広告を繰り返し作成することにつながる。そのため、AdTE_xでは表層的多様性の向上を目指しつつ、広告内容の類似性および広告配信上の形式的制約を維持することを目的とする。

実験の結果、AdTE_xは内容の類似性が低下する一方で、高い表層的多様性を達成することを示した。広告分野の専門家による評価においては、広告文としてのほかの指標と比べ、内容の整合性が最も低い評価を受けた一方で、生成されたテンプレートが概ね文法的に正しく、内容的にも妥当であると評価された。そのため、提案手法により、内容の整合性や類似性、広告配信に向けた形式の遵守に関してさらなる改善が必要である一方で、文字数制限を満たしつつ、表層的多様性を向上させたテンプレート拡張が可能であることが示された。

2 検索連動型広告

検索連動型広告では、検索エンジンに入力されたキーワードに関連する広告見出しや説明文が検索結果の一部として表示される。検索連動型広告の中でも、レスポンス検索広告¹⁾は、ユーザの検索キーワードに最も関連した広告見出しや説明文が広告配信時に自動的に組み合わせられる。そのため、広告主は複数の広告見出しおよび説明文を事前に登録する必要がある。

事前に登録する広告見出しには、キーワード挿入機能²⁾を利用することができ、1つの広告見出しから、検索キーワードに対応した複数の広告を表示できる。キーワード挿入機能を用いた広告見出しは、検索キーワードにより変化しない固定された部分をもち、広告見出しのテンプレートとして機能する。似た広告見出しが繰り返し表示されることを防ぐために、広告見出しのテンプレートの多様性が重要となる。テンプレートが類似している場合、検索キーワードが異なっても似た広告見出しが繰り返

し表示されるため、テンプレートの多様性が重要となる。

広告見出しの自動生成手法は、これまでも提案されている[14, 8]が、既存手法ではテンプレートの多様性を明示的に考慮していない。

3 提案手法

本研究では、広告配信の制約を満たしながらテンプレートの多様性を高めるために、LLMを利用したフレームワーク **AdTE_x** を提案する。まず、テンプレート拡張の基盤となる広告データを収集し、各業界ごとに分類した。さらに、業界ごとに広告をいくつかのグループに分け、それぞれのグループにテンプレートを割り当てた。そして、既存のテンプレートを出力の例として用い、いくつかの広告から共通部分を抽出することで、テンプレートの拡張を行った。

広告データ収集と業界分類 収集した広告データに対して前処理を行い、重複データおよび省略記号(...)で終わるものを除去した。キーワード挿入機能で用いられる波括弧 {} を含むものを**テンプレート**、含まないものを**広告見出し**とした。これらの処理の結果、151,103件のテンプレートと12,076件の広告見出しを得た。さらに、テンプレートおよび広告見出しに対して、26のIABカテゴリ³⁾に基づく業界ラベルを付与した。分類の詳細については、付録Aに記載する。

広告データのグループ化とテンプレートの割り当て 共通部分の抽出を行うために、表層的な類似度に基づいて各業界内の広告見出しをまとめた。Pairwise-BLEU (pBLEU) [15]を算出し、相互に高い類似度を示す5つの広告見出しを、1つのグループとした。すべての広告見出しとテンプレートの組み合わせに対してBLEU[16]を計算し、各グループ内における平均が最大となるテンプレートを割り当てた。広告見出し間のpBLEU、および広告見出しとテンプレート間のBLEUの算出には、4-gram BLEUを用いた。この結果、合計132,863グループを得た。

テンプレート拡張 本研究では、CAMERA [17]およびCAMERA³ [10]に倣い、生成結果が満たすべき制約をあらかじめ定義した。その上で、共通部分を抽出するための広告見出しと、出力形式を示すためのテンプレートを入力として与え、広告見出しの生成を行った。テンプレート拡張に用いたプロンプト

1) <https://support.google.com/google-ads/answer/7684791?hl=ja>

2) <https://support.google.com/google-ads/answer/2454041?hl=ja>

3) <https://www.iab.com/>

の詳細については、付録 B に記載する

4 実験設定

評価指標 AdTE_x の有効性を検証するため、自動評価および人手評価を実施した。自動評価では、元データに対する増加率、多様性、広告配信上の制約の遵守率の三つの観点から評価を行い、人手評価ではテンプレートとしての品質を評価した。増加率 (AUG) では、元のデータと比較して、テンプレートの数がどの程度増加したかを評価する。多様性の評価には、pBLEU および pBLEURT [18] を用い、それぞれ表層的な多様性と意味的な多様性を評価した⁴⁾。さらに、広告配信における制約に対する評価として、文字数制限の遵守率 (REG) [17] と、キーワード挿入形式への制限の遵守率 (KREG) を導入した。KREG は、広告配信において求められるテンプレート形式を満たしている割合を表し、REG は広告見出しの文字数制限 (15 文字以内) を満たしている割合を示す。pBLEURT および REG の評価では、生成されたテンプレートから作成した広告見出しを対象として評価を行った。自動評価におけるベースラインとしては、各評価指標について元のテンプレートに対するスコアを採用した。

テンプレートとして適切であると判断した 2,619 件のテンプレートに対して、人手評価を実施した。文法的正確さ (Lexicality)、内容の整合性 (Naturalness)、およびテンプレートとしての妥当性 (Rationality) の 3 つの基準に基づき、生成結果を広告分野の専門家が評価した。収集したデータはすでに広告配信されているため、人手評価の 3 つの基準すべてのベースラインは 100% を用いた。

評価に使用した LLM 本研究では、Llama-3-ELYZA-JP-8B [20]、Llama-2-ELYZA-JP-7B および、13B [21, 22]、CALM3-22B [23]、CALM2-7B⁵⁾、Gemini-2.0-Flash [24]、GPT-4o-mini [25] の計 7 種類の LLM を用いて AdTE_x の評価を行った。各 LLM の詳細な設定については、付録 C で報告している。

5 実験結果

表 1 は全体の結果を示す。自動評価において、AUG は全ての業界で評価した一方、その他の指標では各モデルが 100 件以上のテンプレートを生成した 12 業界に限定して評価した。

4) 本研究における BLEU および BLEURT の計算には、mbrs [19] ライブラリを使用した。

5) <https://huggingface.co/cyberagent/calm2-7b>

自動評価 表 1 中の AUG の結果に着目すると、すべてのモデルで 100 以上のスコアとなっており、元データよりも多くのテンプレートを生成できている。この結果は、AdTE_x によりテンプレートを拡張できる一方で、拡張度合いにはモデル間で差があることを示している。表 1 中の pBLEU および pBLEURT の結果に着目すると、いずれの指標においてもベースラインより低い値を示している。この結果から、AdTE_x は表層的多様性を向上させる一方で、内容の類似性を低下させる傾向であることが示唆される。表 1 中の REG および KREG の結果に着目すると、REG ではほとんどのモデルにおいて半数以上の生成結果が文字数制限を守る一方で、すべてのモデルにおいて KREG は半数以上が制約を遵守していない。この結果から、AdTE_x は文字数制約を満たす生成は比較的容易である一方、形式的制約を満たした生成は困難であることが示唆される。

人手評価 表 1 に示す人手評価の結果を見ると、半数以上のモデルにおいて、生成されたテンプレートのうち Naturalness で適格と評価された割合は約 60% にとどまっている一方で、Lexicality および Rationality では、適格と評価された割合が 80% 以上となっている。この結果は、広告文としての内容の整合性には改善の余地が残るものの、生成されたテンプレートが概ね文法的に正しく、内容的にも妥当であることを示唆している。

6 分析

テンプレートの種類による制約遵守率への影響の分析 収集されたテンプレートの多くが KREG を満たしていなかったため、KREG を満たすテンプレートを理想的なテンプレートとし、テンプレートの種類が制約遵守率に与える影響を分析する。表 2 は、2 種類のテンプレートをそれぞれ用いた場合における、2 つの制約遵守率の結果を示している。表 2 の KREG の結果に着目すると、理想的なテンプレートを与えた場合に、全体結果から KREG のスコアが増加している。この結果は、理想的なテンプレートを用いることで、形式的制約に沿ったテンプレートが生成されやすくなることを示唆している。

業界ごとの多様性評価結果の分析 モデルの性能が業界ごとに異なることが報告されている [17] ため、各業界の 2 つの多様性への影響を検証する。図 2 には、12 の業界における pBLEURT および pBLEU の結果を示す。pBLEU の結果に着目すると、

表 1 全体の実験結果。太字は最高値を示す。青は文字数遵守率が 50 %以下のモデルを示す。± の値は標準偏差を示す。

	自動評価					人手評価		
	AUG(↑)	pBLEU (↓)	pBLEURT (↑)	REG (↑)	KREG (↑)	Lexicality (↑)	Naturalness (↑)	Rationality (↑)
Baseline	100.00	10.75(± 5.34)	15.48(± 7.36)	98.72	31.91	100.00	100.00	100.00
GPT-4o-mini	591.79	6.64(± 1.99)	11.07(± 4.47)	61.85	38.81	91.15	50.08	85.33
Gemini-2.0-Flash	789.96	5.32(± 1.26)	13.42(± 4.35)	59.14	10.57	90.18	69.03	83.94
CALM2-7B	1303.07	2.33 (± 1.44)	8.86(± 4.23)	80.36	3.60	93.76	73.24	84.01
CALM3-22B	642.17	5.45(± 2.32)	9.4(± 4.05)	65.00	38.31	91.98	61.35	83.61
Llama-3-ELYZA-JP-8B	890.23	5.06(± 1.79)	13.88 (± 5.04)	62.14	6.06	77.79	44.93	79.02
Llama-2-ELYZA-JP-7B	187.78	3.72(± 2.34)	6.66(± 3.43)	82.93	26.69	56.81	31.00	41.95
Llama-2-ELYZA-JP-13B	105.59	2.97(± 2.16)	9.81(± 5.66)	41.70	26.42	80.50	57.33	71.65

表 2 テンプレートの種類による 2 つの制約遵守率の結果。左列および右列はそれぞれ理想的なテンプレートを使用した場合と使用しない場合の結果を示す。Llama-2-ELYZA-JP-7B については、有効な生成結果がすべて理想的なテンプレートを使用していたため、理想的なテンプレートを使用しない場合の結果を省いている。太字は各指標における最良の値を示す。括弧内の値は、分割前の結果からの差分を表す。

	REG (↑)		KREG (↑)	
GPT-4o-mini	86.13(+24.28)	48.26(-13.59)	88.19 (+49.38)	8.84(-29.97)
Gemini-2.0-Flash	63.51(+4.37)	56.94(-2.2)	16.90(+6.83)	5.19(-5.38)
CALM2-7B	82.18(+1.82)	79.45 (-0.91)	8.40(+4.80)	1.18(-2.42)
CALM3-22B	88.88 (+23.88)	48.41(-16.59)	74.93(+36.62)	10.16 (-28.15)
Llama-3-ELYZA-JP-8B	74.55(+12.41)	57.40(-4.74)	14.29(+8.23)	1.22(-4.84)
Llama-2-ELYZA-JP-7B	82.93(+0.0)	-	26.69(+0.0)	-
Llama-2-ELYZA-JP-13B	52.72(+11.02)	36.33(-5.37)	51.06(+24.64)	3.75(-22.67)

表 3 自動評価と人手評価間のピアソン (r) とスピアマン (ρ) による相関の結果

	Lexicality		Naturalness		Rationality	
	r	ρ	r	ρ	r	ρ
pBLEU	0.06	0.06	0.00	0.00	0.10	0.10
pBLEURT	-0.13	-0.13	-0.08	-0.08	-0.04	-0.04
REG	0.04	0.04	0.07	0.07	0.20	0.20
KREG	0.09	0.09	0.06	0.06	0.11	0.11

人手評価と自動評価の相関による分析 自動評価と人手評価の相関を算出し、自動評価指標の有効性を検証する。表 3 には、Pearson(r) および Spearman(ρ) による各指標間の相関の結果を示す。その結果、多くの指標の組み合わせにおいて、自動評価と人手評価の間に明確な相関は確認されなかった。一方で、REG と Rationality の間には弱い相関が見られた。この結果は、文字数制限を遵守することで、テンプレートとしての妥当性が向上することを示唆している。

7 まとめ

本研究では、広告作成の自動化支援のために広告配信における制約を満たした多様な広告テンプレートの拡張を行うフレームワーク **AdTEx** を提案した。実験結果は、AdTEx は内容の整合性や類似性、広告配信に向けた形式の遵守について改善が必要とする一方、文字数制限を遵守し、表層的多様性を向上させた上での拡張が可能であることを示した。分析結果から、自動評価と人手評価との相関が弱いことが示された一方で、形式的制約を満たしたテンプレートを利用することによる形式的な制約の改善や内容の類似した業界においては、内容の類似性を維持することが示唆された。

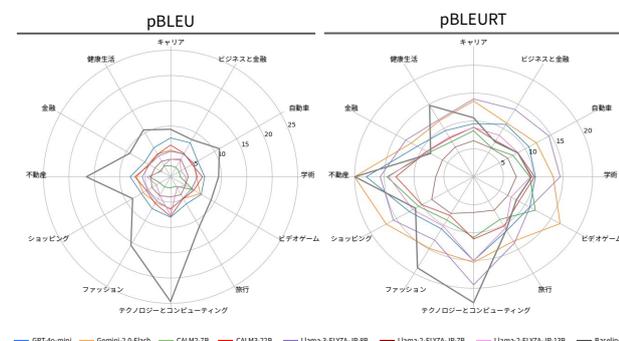


図 2 業界ごとの多様性評価の結果

すべてのモデルにおいてベースラインと比較してスコアが低下しており、業界間のばらつきも小さいことから、全ての業界において表層的多様性が向上している。一方で、pBLEURT の結果では、元の広告見出しでのスコアが高い業界では、各モデルのスコアも相対的に高くなっている。この結果は、広告見出しの内容が類似している業界においては、生成される広告見出し間の内容の類似性が高くなることを示唆している。

参考文献

- [1] Soichiro Murakami, et al. Natural language generation for advertising: A survey, 2023.
- [2] Kevin Bartz, et al. Natural language generation for sponsored-search advertisements. In **Proceedings of the 9th ACM Conference on Electronic Commerce**, EC '08, p. 1–9, New York, NY, USA, 2008. Association for Computing Machinery.
- [3] Khalid Alnajjar and Hannu Toivonen. Computational generation of slogans. **Natural Language Engineering**, Vol. 27, No. 5, p. 575–607, 2021.
- [4] Stamatina Thomaidou, Konstantinos Leymonis, and Michalis Vazirgiannis. Grammads: Keyword and ad creative generator for online advertising campaigns. In Pierre-Jean Benghozi, Daniel Krob, and Frantz Rowe, editors, **Digital Enterprise Design and Management 2013**, pp. 33–44, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg.
- [5] Konstantin Golobokov, Junyi Chai, Victor Ye Dong, Mandy Gu, Bingyu Chi, Jie Cao, Yulan Yan, and Yi Liu. DeepGen: Diverse search ad generation and real-time customization. In Wanxiang Che and Ekaterina Shutova, editors, **Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing: System Demonstrations**, pp. 191–199, Abu Dhabi, UAE, December 2022. Association for Computational Linguistics.
- [6] Atsushi Fujita, Katsuhiko Ikushima, Satoshi Sato, Ryo Kamite, Ko Ishiyama, and Osamu Tamachi. Automatic generation of listing ads by reusing promotional texts. In **Proceedings of the 12th International Conference on Electronic Commerce: Roadmap for the Future of Electronic Business**, ICEC '10, p. 179–188, New York, NY, USA, 2010. Association for Computing Machinery.
- [7] Hidetaka Kamigaito, Peinan Zhang, Hiroya Takamura, and Manabu Okumura. An empirical study of generating texts for search engine advertising. In Youngbum Kim, Yunyao Li, and Owen Rambow, editors, **Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies: Industry Papers**, pp. 255–262, Online, June 2021. Association for Computational Linguistics.
- [8] Chang Wang, Siyu Yan, Depeng Yuan, Yuqi Chen, Yanhua Huang, Yuanhang Zheng, Shuhao Li, Yinqi Zhang, Kedi Chen, Mingrui Zhu, and Ruiwen Xu. Beyond quality: Unlocking diversity in ad headline generation with large language models, 2025.
- [9] Yanda Chen, Zihui Ren, Qixiang Gao, Jiale Chen, Si Chen, Xubin Li, Tiezheng Ge, and Bo Zheng. Ctr-driven ad text generation via online feedback preference optimization, 2025.
- [10] Go Inoue, Akihiko Kato, Masato Mita, Ukyo Honda, and Peinan Zhang. CAMERA³: An evaluation dataset for controllable ad text generation in Japanese. In Nicoletta Calzolari, Min-Yen Kan, Veronique Hoste, Alessandro Lenci, Sakriani Sakti, and Nianwen Xue, editors, **Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)**, pp. 2702–2707, Torino, Italia, May 2024. ELRA and ICCL.
- [11] Yookyung Lee, Soonwon Ka, Bokyung Son, Pilsung Kang, and Jaewook Kang. Navigating the path of writing: Outline-guided text generation with large language models. In Weizhu Chen, Yi Yang, Mohammad Kachuee, and Xue-Yong Fu, editors, **Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 3: Industry Track)**, pp. 233–250, Albuquerque, New Mexico, April 2025. Association for Computational Linguistics.
- [12] Cornelia Pechmann and David W. Stewart. Advertising repetition: A critical review of wearin and wearout. **Current Issues and Research in Advertising**, Vol. 11, No. 1-2, pp. 285–329, 1988.
- [13] Yoichi Aoki, et al. Exploring the relationship between diversity and quality in ad text generation, 2025.
- [14] Yashal Shakti Kanungo, Sumit Negi, and Aruna Rajan. Ad headline generation using self-critical masked language model. In Youngbum Kim, Yunyao Li, and Owen Rambow, editors, **Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies: Industry Papers**, pp. 263–271, Online, June 2021. Association for Computational Linguistics.
- [15] Tianxiao Shen, Myle Ott, Michael Auli, and Marc'Aurelio Ranzato. Mixture models for diverse machine translation: Tricks of the trade. In Kamalika Chaudhuri and Ruslan Salakhutdinov, editors, **Proceedings of the 36th International Conference on Machine Learning**, Vol. 97 of **Proceedings of Machine Learning Research**, pp. 5719–5728. PMLR, 09–15 Jun 2019.
- [16] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. Bleu: a method for automatic evaluation of machine translation. In Pierre Isabelle, Eugene Charniak, and Dekang Lin, editors, **Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics**, pp. 311–318, Philadelphia, Pennsylvania, USA, July 2002. Association for Computational Linguistics.
- [17] Masato Mita, et al. Striking gold in advertising: Standardization and exploration of ad text generation. In Lun-Wei Ku, Andre Martins, and Vivek Srikumar, editors, **Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)**, pp. 955–972, Bangkok, Thailand, August 2024. Association for Computational Linguistics.
- [18] Yusuke Sakai, Hidetaka Kamigaito, and Taro Watanabe. Revisiting compositional generalization capability of large language models considering instruction following ability. In Wanxiang Che, Joyce Nabende, Ekaterina Shutova, and Mohammad Taher Pilehvar, editors, **Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)**, pp. 31219–31238, Vienna, Austria, July 2025. Association for Computational Linguistics.
- [19] Hiroyuki Deguchi, Yusuke Sakai, Hidetaka Kamigaito, and Taro Watanabe. mbrs: A library for minimum Bayes risk decoding. In Delia Irazu Hernandez Farias, Tom Hope, and Manling Li, editors, **Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing: System Demonstrations**, pp. 351–362, Miami, Florida, USA, November 2024. Association for Computational Linguistics.
- [20] Masato Hirakawa, Shintaro Horie, Tomoaki Nakamura, Daisuke Oba, Sam Passaglia, and Akira Sasaki. elyza/llama-3-elyza-jp-8b, 2024.
- [21] Akira Sasaki, Masato Hirakawa, Shintaro Horie, and Tomoaki Nakamura. Elyza-japanese-llama-2-7b, 2023.
- [22] Hugo Touvron, et al. Llama: Open and efficient foundation language models. **arXiv preprint arXiv:2302.13971**, 2023.
- [23] Ryosuke Ishigami. cyberagent/calm3-22b-chat, 2024. Hugging Face.
- [24] Google Deepmind. Gemini 2.0 flash. <https://deepmind.google/technologies/gemini/flash/>, 2025.
- [25] OpenAI, :, Aaron Hurst, et al. Gpt-4o system card, 2024.
- [26] AI@Meta. Llama 3 model card. 2024.
- [27] Yusuke Sakai, Hidetaka Kamigaito, and Taro Watanabe. mC-SQA: Multilingual commonsense reasoning dataset with unified creation strategy by language models and humans. In Lun-Wei Ku, Andre Martins, and Vivek Srikumar, editors, **Findings of the Association for Computational Linguistics: ACL 2024**, pp. 14182–14214, Bangkok, Thailand, August 2024. Association for Computational Linguistics.

A 分類データセットの詳細

表 4 ファインチューニングにおけるハイパーパラメータ設定。

Hyperparameter	Value
Batch Size	128
Learning Rate	2×10^{-5}
Max Epochs	5
Weight Decay	0.01
Early Stopping Patience	3
Random Seed	0

本研究では、IAB カテゴリ分類データセットを用いてファインチューニングした事前学習済み日本語 BERT モデルを用いて、広告見出しの分類を行う。事前学習済み日本語 BERT モデル (tohoku-nlp/bert-base-japanese-v3)⁶⁾ のファインチューニングには、871,909 記事とそれぞれの記事に対する IAB ラベルが割り当てられている shishir-dwi/News-Article-Categorization_IAB⁷⁾ を利用した。データセットは、訓練・検証・テストの各セットに対して 8:1:1 の比率で分割した。モデルの学習には、表 4 に示したハイパーパラメータを用いた。過学習を防止するため、検証損失を監視して最適な停止点を決定する early stopping を採用し、patience は 3 エポックに設定した。

B プロンプトの詳細

図 3 に示すようなプロンプトを構築し、5 つの具体的な広告見出しから共通する構造を抽出するようモデルに指示した。

C LLM の詳細

本研究では、Llama-3 [26] を基盤とする Llama-3-ELYZA-JP-8B [20]、Llama-2-ELYZA-JP-7B および 13B [21, 22]、ならびに日本語コーパスを用いたスクラッチで学習された CALM3-22B [23] と CALM2-7B⁸⁾ を使用した。商用モデルとしては、Gemini-2.0-Flash [24] および GPT-4o-mini [25] を用いた。表 5 は、各 LLM のソース情報を示している。オープンな LLM の推論には、A6000、6000、および 3090 GPU を使用した。また、商用モデルである Gemini-2.0-Flash

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

7) https://huggingface.co/datasets/shishir-dwi/News-Article-Categorization_IAB

8) <https://huggingface.co/cyberagent/calm2-7b>

複数の具体的な広告文をもとに、共通構造を抽出した広告文テンプレートをいくつか作成してください。テンプレートは以下の条件を厳守してください。

- テンプレートは広告文の構造のみを表し、実際の社名・サービス名・地名など具体的情報は含めないこと。
- 置換できない固定部分は合計 10 文字以内とすること。
- 置換可能な部分には必ず {} を使用し、{} 内に例として入る単語を 1 つ記載してください。
- 説明文や補足は一切記載せず、テンプレート本文のみ返すこと。
- 広告文が複数例与えられる場合、最も汎用性の高い共通の構造を 1 つにまとめてテンプレート化すること。

【出力形式

期待する出力形式を参考にして、テンプレート文のみ返してください。

広告見出しの例

子育て中の保護者の方へ【フルタイム求人】

【協同互助】妊婦の方へ

【求職中の方へ】

床ワックスがけをご検討中の方へ

働く保護者に適した求人

期待される出力形式

{keyword: 技術サポート}をお探しの方へ

図 3 テンプレート拡張に用いたプロンプト。

LLMs	HuggingFace ID / API 名称
Gemini-2.0-Flash	Gemini/gemini-2.0-flash
GPT-4o-mini	OpenAI/gpt-4o-mini-2024-07-18
CALM3-22B	cyberagent/calm3-22b-chat
CALM2-7B	cyberagent/calm2-7b-chat
Llama-3-ELYZA-JP-8B	elyza/Llama-3-ELYZA-JP-8B
Llama-2-ELYZA-JP-7B	elyza/ELYZA-japanese-Llama-2-7b-instruct
Llama-2-ELYZA-JP-13B	elyza/ELYZA-japanese-Llama-2-13b-instruct

表 5 実験に使用した LLM の詳細。

および GPT-4o-mini については、可能な限り決定論的な出力を得るために temperature を 0 に設定した [27]。その他のオープンな LLM についても、固定のシード値を用い greedy decoding により出力を生成した。