

# カテゴリ説明文を用いた大規模言語モデルによる 日本語文書分類

小野寺優<sup>1</sup> 新納浩幸<sup>2</sup>

<sup>1</sup> 茨城大学大学院理工学研究科情報工学専攻 <sup>2</sup> 茨城大学大学院理工学研究科情報科学領域  
{24nm714r, hiroyuki.shinnou.0828}@vc.ibaraki.ac.jp

## 概要

大規模言語モデル (LLM) を用いた文書分類は、従来の Fine-tuning を利用する手法と比較して大量の訓練データを必要としない利点がある。一方で、分類精度においては Fine-tuning した BERT 系モデルと同等以上の性能を示した事例は限られている。本研究では、日本語ニュースコーパスを対象に、Fine-tuning を行った RoBERTa と LLM (GPT-4o mini, Gemini 2.5 flash) を用いたカテゴリ分類の性能比較を行った。特に、実際の訓練データを基にカテゴリの特徴を言語化した「カテゴリ説明文」を LLM で自動生成し、プロンプトに付与する手法を提案した。実験の結果、提案手法は従来の Zero-shot 分類と比較して精度が向上することを確認した。一方で、Fine-tuning した RoBERTa との比較では、ラベル境界が曖昧なデータにおいて LLM が誤分類する傾向が高いことが明らかになった。

## 1 はじめに

文書分類は、膨大な文書データを特定の基準で定義されたカテゴリごとに整理する自然言語処理の重要なタスクである。従来、BERT などの事前学習済みモデルを特定のタスクに適応させる Fine-tuning というアプローチが主流であり、多くの分類タスクにおいて高い精度を達成してきた。しかし、これらのモデルが高い性能を発揮するためには大量のラベル付き訓練データを必要とし、データ収集やアノテーションに多大なコストを要することが課題であった。これに対し、近年の LLM では、Zero-shot や Few-shot のような少数データ環境における高度な推論を可能にする手法が提案された。しかし、ラベルの境界が曖昧な多クラス分類において、LLM が Fine-tuning を施した従来モデルに匹敵するほどの成果を示した事例は未だ限定的である。

本研究では、少数データ環境における LLM の分類精度を向上させるため、各カテゴリの特徴を言語化した「カテゴリ説明文」を自動生成し、日本語ニュース記事分類のプロンプトに利用する手法を提案する。本手法では、各カテゴリから抽出した少数のニュース記事を要約し、それらを LLM に分析させることで、詳細なカテゴリ特有の説明文を生成する。この説明文をプロンプトに付与することで、LLM による文書分類を試みた。評価実験の結果、提案手法は従来の Zero-shot 手法を上回る精度を記録した。一方で、Fine-tuning を施した RoBERTa の性能には及ばず、現状の LLM におけるラベルの曖昧性への識別能力に課題があることを明らかにした。

## 2 関連研究

文書分類は自然言語処理における代表的なタスクであり、従来は BERT や RoBERTa などの事前学習済みモデルを Fine-tuning する手法が主流であった。しかし、このアプローチでは高品質なラベル付きデータを大量に必要とするため、データ作成コストやドメイン適応の難しさが課題であった。

この課題に対し、LLM を用いてデータセットを構築あるいは拡張する試みがある。Fujii ら [1] は、LLM による日本語データセットの生成が少量データ環境での性能向上に影響する可能性を示した。Onodera ら [2] は、LLM にキーワードと簡潔な特徴文を与えることでデータを生成し、そのデータで訓練データを拡張することで分類器の精度が改善するかを試みたが、拡張したデータがノイズとなり、精度の改善に至らない可能性を指摘した。

近年、LLM を直接分類器として利用する Zero-shot や Few-shot 分類の研究が盛んに行われている。Sun ら [3] は、LLM を分類器として直接利用する枠組みを整理し、プロンプト設計が性能に大きな影響を与えることを示した。また、金融分野や規制文書など

の専門領域においても、Few-shot 分類の有効性が報告されている [4, 5, 6]。しかし、LLM による分類性能はタスクやデータセット、プロンプトの質に大きく依存し、従来の Fine-tuning モデルを常に上回るわけではないことも指摘されている [7, 8, 9]。

このような背景から、LLM を用いた文書分類においては、プロンプト設計やラベルの詳細な情報の与え方を工夫する研究が進められている。高野ら [10] は、分類タスクにおける「定義文」を LLM の知識から生成し、それを活用することで金融テキスト分類の精度向上を報告している。Yoshimura ら [11] は、階層構造をもつラベルに対して、LLM により分類理由も含めて構造化した推論をさせる分類手法を提案している。しかし、これらの研究では、ニュース記事のようにカテゴリが多岐にわたり、ラベル境界が曖昧なタスクにおいて、分類器としての有効性があるかについては十分に検討されていない。

本研究では、カテゴリごとの要約文を基に LLM 自身にカテゴリ特徴を分析させて生成した「カテゴリ説明文」をプロンプトに組み込むことで、カテゴリ間の意味的な差異を明示的にプロンプトへ反映させ、分類性能の向上を目指す。

## 3 プロンプト構築手法

本研究では、訓練データに含まれる文書の傾向を反映した「カテゴリ説明文」を自動生成し、それをプロンプトに組み込む手法を提案する。Livedoor ニュースコーパス<sup>1)</sup>に含まれる 9 カテゴリの日本語ニュース記事を対象とし、LLM を分類器として利用するためのプロンプトを構築した。

### 3.1 ニュース記事の要約文生成

まず、各カテゴリからランダムに抽出した 50 件の記事 (合計 450 件) を対象に、カテゴリの特徴を抽出するための要約文を生成した。大量のニュース記事の原文を直接 LLM に与えると、コンテキスト長の制限や推論時間の増大といった課題がある。そのため、各記事を 200 文字程度へと要約し、カテゴリ特有のトピックを凝縮した形式へと変換した。要約文を生成する際に使用したプロンプトは以下の表 1 の通りである。

表 1 要約文を生成するためのプロンプト

```
あなたは優秀なニュース記事要約システムです。与えられた日本語ニュース記事本文を、意味と文体を正確に保ったまま、約 200 文字で自然な日本語に要約してください。固有名詞はできるだけ維持し、主語と述語の対応を正確に保ってください。
```

```
ニュース記事: {text}
```

## 3.2 カテゴリ説明文生成

3.1 節で得られた 50 件の要約文をすべて入力し、LLM にそのカテゴリの包括的な特徴を分析させた。ニュースカテゴリにどのような特徴があるのかを説明させるプロンプトを以下の表 2 に示す。単なる辞書的な定義ではなく、実際のニュース記事の傾向に基づいた詳細な説明文を生成するために出力する文字数は 500 文字程度と設定する (付録 A 節参照)。

表 2 カテゴリ説明文を生成するためのプロンプト

```
以下の 50 個の文章はあるカテゴリに属するブログ記事を要約したものです。この 50 個の文章の内容から、このカテゴリにはどのようなブログ記事が掲載されているかを 500 文字程度で詳細に説明して下さい。
```

```
1 番目の記事の要約: {summary_01}  
2 番目の記事の要約: {summary_02}  
...  
50 番目の記事の要約: {summary_50}
```

## 3.3 文書分類プロンプトの作成

3.2 節で得られた 9 カテゴリ分の説明文を用いて、ニュース記事を分類させるためのプロンプトを構築する。推論時には、入力記事に対して最も適切なカテゴリを選択させると同時に、分類の根拠 (理由) を併せて出力させることで、LLM の分類の際の思考プロセスを可視化した。作成した文書分類のためのプロンプトの例を表 3 に示す。

表 3 文書分類のためのプロンプト

```
あなたは、与えられた 9 つのカテゴリの説明に基づき、記事を厳密に分類するアシスタントです。
```

```
第 0 カテゴリ (ラベル名: dokujo-tsushin) の説明: {description_0}  
第 1 カテゴリ (ラベル名: it-life-hack) の説明: {description_1}  
...  
第 9 カテゴリ (ラベル名: topic-news) の説明: {description_9}
```

```
以下の記事が、上記で示した 9 つのカテゴリのどれに属するかを判定してください。  
入力記事: "{article_text}"
```

## 4 実験

### 4.1 実験設定

本研究では、日本語ニュース記事分類タスクにおいて、Fine-tuning を行った RoBERTa と LLM による分類手法による性能の変化を比較する。

RoBERTa には、日本語事前学習済みモデルの

1) <https://www.rondhuit.com/download.html>

**表 4** ニュース記事カテゴリ分類の Top-1/Top-3 正解率 (%) と Macro-F1(%) の比較 (RoBERTa では 5 シードの平均精度と標準偏差 (SD) を記す。)

モデル	手法	Top-1 Accuracy (SD)	Macro F1 (SD)	Top-3 Accuracy
RoBERTa	Fine-tuning	88.62 (1.12)	88.65 (1.14)	-
GPT-4o mini	Zero-shot	64.67	62.66	86.67
	提案手法	65.11	64.53	88.67
Gemini 2.5 flash	Zero-shot	66.44	64.54	90.22
	提案手法	74.44	74.21	94.67

rinna/japanese-roberta-base<sup>2)</sup> を利用し、Livedoor ニュースコーパスから構築した訓練データを用いて教師あり学習による Fine-tuning を行った。LLM は、gpt-4o-mini<sup>3)</sup> と gemini-2.5-flash<sup>4)</sup> を使用した。LLM には、Fine-tuning は行わず、推論時に与えるプロンプトのみを用いて分類を行う。ただし、カテゴリ説明文の構築とそれを用いた分類には、一貫性を持たせるため同一のモデルを使用することとする<sup>5)</sup>。プロンプトは以下の 2 種類を用意した。

1. 簡潔なカテゴリ定義を付与した従来の Zero-shot プロンプト (以降、Zero-shot と記す。)
2. 3 節で作成した詳細なカテゴリ説明文が付与されたプロンプト (以降、提案手法と記す。)

実験に用いたニュース記事は、訓練/検証/テストでそれぞれ 450 件 (各カテゴリ 50 件ずつ)、重複なしで分離した。訓練データは、Fine-tuning と提案手法で同一のものを使用し、テストデータについてもすべての手法の分類の際に同一のものを使用した。パラメータ設定については付録 C 節に記載する。

## 4.2 評価方法

評価指標には、分類精度の全体的な傾向を示す Accuracy(正解率) および各カテゴリの性能を平均した Macro-F1 を用いる。RoBERTa は 5 つのシード値による平均値で評価し、LLM は Top-1 および Top-3 の正解率を算出する。

## 4.3 実験結果

4.1 節にて、設定した各手法について、分類精度をまとめたものを表 4 に示す (具体的な分類例は付録 B 節参照)。実験の結果、Top-1 の精度において、両 LLM ともに提案手法の方が従来の Zero-shot 手

法の精度を上回った。一方で、RoBERTa の精度には依然として及ばないことが確認された。しかし、Top-3 正解率に着目すると、提案手法を施した LLM は RoBERTa と同等以上の正解率を記録していることが分かった。

ニュースカテゴリ別の F1 スコアを算出した結果を表 5 に示す。これを見ると、全カテゴリにおいて、RoBERTa が最も高いスコアを記録していることが分かる。LLM の傾向を見ると、Zero-shot では、「家電チャンネル」や「Livedoor HOMME」のスコアが著しく低い。一方で、提案手法を適用した Gemini 2.5 flash では、「IT ライフハック」を除くすべてのカテゴリで Zero-shot を上回っている。特に、「Livedoor HOMME」や「トピックニュース」では約 20 % 伸びていることが分かる。GPT-4o mini については、「Movie Enter」のように精度が上昇している例もあれば、「独女通信」のように精度が低下している例も見られた。

## 5 考察

4 節で示した実験結果に基づき、提案手法の有効性と限界について考察する。まず、全体的な分類性能の傾向を分析し、次に低スコアのカテゴリにおける誤分類理由を定性的に分析、およびモデル間での誤分類傾向の比較を行う。

### 5.1 カテゴリ説明文の効果について

カテゴリ説明文をプロンプトに付与することで、従来の zero-shot 手法を上回る精度が示された。これは、LLM が詳細なカテゴリ定義を参照することで、より文脈に沿った推論が可能になったためと考えられる。一方で、依然として Fine-tuning を施した RoBERTa の精度には及ばなかった。しかし、Top-3 正解率に着目すると、両 LLM ともに RoBERTa と同等以上の性能を記録している。これは、LLM が正解の可能性の高いカテゴリ候補を絞り込む能力を持

2) <https://huggingface.co/rinna/japanese-roberta-base>  
 3) <https://platform.openai.com/docs/models/gpt-4o-mini>  
 4) <https://ai.google.dev/gemini-api/docs/models?hl=ja>  
 5) GPT-4o mini によって構築されたプロンプトは GPT-4o mini の分類にのみ適用する。モデル間の転用は行わない。

表5 ニュースカテゴリごとの F1 スコア (%) の比較 (RoBERTa は 5 シードの平均精度と標準偏差 (SD) を記す.)

カテゴリ	RoBERTa	GPT-4o mini		Gemini-2.5-flash	
	Fine-tuning	Zero-shot	提案手法	Zero-shot	提案手法
独女通信	88.72 (2.10)	77.67	75.47	71.91	80.90
IT ライフハック	88.52 (3.47)	57.78	61.54	65.98	63.53
家電チャンネル	86.63 (2.98)	34.48	30.77	39.47	48.00
Livedoor HOMME	79.57 (1.23)	41.27	47.76	49.28	72.50
Movie Enter	90.69 (1.60)	73.44	85.44	66.67	84.35
Peachy	83.29 (2.34)	62.00	60.76	66.67	76.36
SMAX	95.53 (1.05)	76.92	74.16	80.00	80.34
Sports Watch	92.74 (2.82)	83.50	80.77	85.98	86.60
トピックニュース	92.11 (2.66)	56.88	64.12	54.95	74.80

ちながらも、最終的に 1 つに決定する段階で、類似カテゴリ間の境界を識別しきれていないことを示唆している。

## 5.2 誤分類理由の定性分析

分類の際に LLM が生成した理由 (3.3 節) から誤分類の原因を分析した。表 5 より、特に提案手法の精度が低かった「IT ライフハック」、「家電チャンネル」、「Livedoor HOMME」の 3 カテゴリから誤分類した例を 10 件ずつランダムに抽出 (合計 30 件) し、その原因を以下の基準で分析した。

**基準 A** 複数のカテゴリにまたがる共通トピックが含まれ、判断が困難なもの

**基準 B** 特定の強いキーワードに引っ張られ、全体の文脈を考慮していないもの

分析結果を表 6 に示す。両モデルともに「基準 A」による誤分類が 8 割以上を占めた。これは、LLM の読解力不足というよりも、RoBERTa のようにデータセット固有のラベル境界を学習していない LLM が、より広義な意味で関連性の高い他カテゴリを選んでしまう傾向を示唆している。

表6 LLM が誤分類した回答の理由の分析

モデル	基準 A	基準 B	合計
GPT-4o mini	27	3	30
Gemini 2.5 flash	24	6	30

## 5.3 モデル間における誤分類の共通性

各モデルの誤分類データの重複状況を分析した結果 (表 7)、3 モデル全てが誤った記事は 23 件に留まった。これは、人間が見ても特に判断するのが難しい高難易度なラベルであると考えられる。また、RoBERTa のみが正解した例が 65 件存在している。Fine-tuning はデータセット全体の特徴を直接学習で

きる一方、LLM は一般的な言語理解に基づいて学習を行う。そのため、Livedoor ニュースコーパスはデータセット独自の判断基準があることから、入力プロンプトのみに依存する Zero-shot 手法では分類が難しいと考えられる。一方、モデル間で誤分類の傾向が異なる (GPT のみ、Gemini のみの誤り) 事例は、各モデルの学習データやパラメータサイズによる得手不得手が反映された結果と考えられる。

表7 モデル別誤分類数の比較 (○が正解, ×が不正解)

RoBERTa	GPT-4o	Gemini 2.5	誤分類数
42 件	157 件	115 件	
×	○	○	9
○	×	○	84
○	○	×	48
×	×	○	8
×	○	×	2
○	×	×	65
×	×	×	23

## 6 おわりに

本研究では、実際の訓練データの要約を基に LLM が生成したカテゴリ説明文をプロンプトに付与する文書分類手法を提案した。結果として、提案手法は従来の Zero-shot 手法より優れた精度を記録したが、Fine-tuning した RoBERTa の精度には劣ることが明らかになった。一方で、Top-3 正解率の分析から、LLM が高い分類能力を有していることが確認された。したがって、最終的なカテゴリ決定の精度向上には類似カテゴリ間における境界の識別能力を高めることが必要だと考えられる。今後の課題として、各カテゴリの独立した説明文だけでなく、他カテゴリとの相違点や境界を明示したプロンプトを設計することにより、カテゴリ間の曖昧さを低減させる手法を検討していきたい。

## 謝辞

本研究は JSPS 科研費 23K11212 の助成を受けています。

## 参考文献

- [1] Takuro Fujii and Satoru Katsumata. Japagen: Efficient few/zero-shot learning via japanese training dataset generation with llm, 2024.
- [2] Yu Onodera and Hiroyuki Shinnou. Exploring data augmentation for keyword-based document classification using large language models. In **Neural Information Processing**, pp. 273–286, Singapore, 2026. Springer Nature Singapore.
- [3] Xiaofei Sun, Xiaoya Li, Jiwei Li, Fei Wu, Shangwei Guo, Tianwei Zhang, and Guoyin Wang. Text classification via large language models. In **Findings of the Association for Computational Linguistics: EMNLP 2023**, pp. 8990–9005, Singapore, December 2023. Association for Computational Linguistics.
- [4] Lefteris Loukas, Ilias Stogiannidis, Prodromos Malakasiotis, and Stavros Vassos. Breaking the bank with chatgpt: Few-shot text classification for finance, 2023.
- [5] Hamed Meshkin, Joel Zirkle, Ghazal Arabidarrehdor, Anik Chaturbedi, Shilpa Chakravartula, John Mann, Bradlee Thrasher, and Zhihua Li. Harnessing large language models’ zero-shot and few-shot learning capabilities for regulatory research. **Briefings in Bioinformatics**, Vol. 25, No. 5, p. bbae354, August 2024.
- [6] Rómulo Nogueira, Hugo Mentzingen, and Nuno Garcia. Leveraging large language models for document classification in the banking sector, 2025. Preprint, Version 1, Research Square.
- [7] Sowmya Vajjala and Shweta Shimangaud. Text classification in the llm era – where do we stand?, 2025.
- [8] Arina Kostina, Marios D. Dikaiakos, Dimosthenis Stefanidis, and George Pallis. Large language models for text classification: Case study and comprehensive review, 2025.
- [9] Lan Jiang, Mengfei Lan, Joe D. Menke, Colby J. Vorland, and Halil Kilicoglu. Text classification models for assessing the completeness of randomized controlled trial publications based on CONSORT reporting guidelines. **Scientific Reports**, Vol. 14, No. 1, p. 21721, 2024.
- [10] 高野海斗, 中川慧, 藤本悠吾. 大規模言語モデルを用いた金融テキスト二値分類タスクの定義文生成とチューニング手法の提案. 人工知能学会 第二種研究会資料 金融情報学研究会 SIG-FIN-033-24, pp. 155–162, 2024. Research report (preprint).
- [11] Kosuke Yoshimura and Hisashi Kashima. Hierarchical text classification using black box large language models, 2025.

## A カテゴリ説明文の例

以下の表 8 に 3.2 節にて生成したカテゴリ説明文および Zero-shot で使用した説明文のうち「Livedoor HOMME」の例を示す。

表 8 「Livedoor HOMME」カテゴリ説明文の例

Zero-shot
ファッションやビジネス、ライフスタイルなど、男性向けの話題。
提案手法
<p><b>GPT-4o mini</b></p> <p>この 50 個の文章から推測されるブログ記事のカテゴリは、主に「ビジネス」「ライフスタイル」「テクノロジー」「スポーツ」に関連する内容です。特に、転職やキャリア形成に関する記事が多く見受けられ、採用担当者の心理や求められるスキル、面接のポイントなど、転職希望者が知りたい情報が豊富に取り上げられています。転職に関連する記事では、ポテンシャルやヒューマンスキルの重要性が強調されており、応募者に求められる姿勢や考え方が具体的に示されています。</p> <p>また、ライフスタイルに関する記事も多く、特にゴルフや自動車、メンズファッションに関する情報が豊富です。ゴルフに関してはクラブの選び方や練習方法、パッティングの技術向上に関する内容があり、愛好者にとって有益な情報が提供されています。自動車関連の記事では、新車の紹介やカスタムカーの展示、環境問題への企業の取り組みが取り上げられ、読者の興味を引く内容が展開されています。</p> <p>さらに、テクノロジーや SNS の活用に関する記事もあり、特に Twitter やスマホの普及状況、セキュリティ対策に関する情報が含まれています。これにより、ビジネスパーソンや若者に向けた情報提供が行われています。</p> <p>全体を通じて、ビジネスシーンでの実用的なアドバイスや、ライフスタイルを充実させるためのヒントが散りばめられており、読者が自分のキャリアや生活を見つめ直すきっかけとなる内容が多く含まれています。</p>
<p><b>Gemini 2.5 flash</b></p> <p>このカテゴリのブログ記事は、主にビジネスパーソン、特に男性のキャリア形成とライフスタイルを多角的に支援する内容が特徴です。</p> <p>キャリア関連では、転職の具体的なノウハウや採用担当者の視点、年収アップの秘訣、リーダーシップや部下育成、コミュニケーション術といったビジネススキル向上のためのアドバイスが豊富に提供されています。若手社会人の悩み相談から、年収 1000 万円を超える成功者の事例、新報事業立ち上げの課題まで、幅広いキャリアステージに対応。自己分析や目標設定の重要性も繰り返し強調されています。</p> <p>ライフスタイル面では、自動車（新車、高級車、カスタムカー、メーカー戦略）、ゴルフ（最新ギア、練習法、上達のコツ）、メンズファッション（オーダーメイド、トレンド、着こなし）、美容（スキンケア、身だしなみ）、健康（リカバリーウェア、イットネス）といった、男性が関心を持つであろう趣味やライフスタイルに関する情報が充実しています。</p> <p>さらに、スマートフォンや SNS (Twitter、はてなブックマーク、スポーツ SNS) の活用法、デジタルコンテンツ（ケータイ専用放送局）の紹介など、最新の IT トrend やデジタルライフに関する実用的な情報も扱っています。</p> <p>記事は、採用担当者や専門家、著名人による解説や対談形式が多く、読者が実践できる具体的なノウハウやトレンド解説が中心。向上心を持ち、仕事もプライベートも充実させたいと考えるビジネスパーソンを主なターゲットとしていることが伺えます。</p>

## B 提案手法による LLM の文書分類の結果と分類理由の出力例

提案手法によりニュース記事を分類した結果を以下の表 9 および表 10 に示す。ここでは、特に精度の低かった「家電チャンネル」カテゴリの例を記す。

表 9 GPT-4o mini の分類の例

正解例
<b>True label:</b> kaden_channel
<p><b>Original Text</b></p> <p>2012 年春の花粉飛散量予測は全国的に少なく、全国平均で 2011 年の約 3 割になるといふ。花粉症の人にとっては嬉しいニュースだ。とは言い、実は花粉症は春だけのものではなく、秋にも雑草が引き起こす花粉症がある。また、秋はダニなどハウスダストによるアレルギーが増える季節でもあるので、1 年中部屋の掃除は怠れない。シャープは、これからの季節にぴったりな窓を開けずに掃除ができる EC-WX300 を含むプラスマクラスターサイクロン掃除機 3 機種を 10 月 20 日から発売する。…</p>
<p><b>Output</b></p> <p>predict label: kaden_channel</p> <p>reason: "記事は掃除機の新製品に関する詳細な情報を提供しており、家電製品の紹介が中心となっているため、家電関連のカテゴリに適しています。"</p>
不正解例
<b>True label:</b> kaden_channel
<p><b>Original Text</b></p> <p>ハンマーでたたいても、釘でひっかいても iPhone の液晶に傷がつかないという、超強力な保護フィルムが登場した。その名も「ウルトラ衝撃吸収プロテクター」だ。どれくらい強力か、という実験動画が強烈だ。マグカップ、ハンマーそしてブロックでガンガン液晶画面に衝撃をくわえるが、iPhone はなんともない。…</p>
<p><b>Output</b></p> <p>predict label: it_life_hack</p> <p>reason: "記事は新しいスマートフォン用の保護フィルムに関する情報を提供しており、特にその機能や価格について詳しく述べています。テクノロジーやガジェットに関連する内容であるため、このカテゴリが最も適しています。"</p>

表 10 Gemini 2.5 flash の分類の例

正解例
<b>True label:</b> kaden_channel
<p><b>Original Text</b></p> <p>撮影した写真や動画をネットで共有するのは既におなじみだ。もし、イベントやパーティ会場、ビジネスのプレゼンテーションの場などで撮影した画像をすぐに大画面で見ることができれば、大勢の人と即座に共有できてもっと便利だろう。住友スリーエムは、プロジェクター機能を搭載したビデオカメラ、「3M™ ビデオカメラプロジェクター CP45」を…</p>
<p><b>Output</b></p> <p>分類結果: kaden_channel</p> <p>理由: 記事は「3M ビデオカメラプロジェクター CP45」というデジタルガジェットの最新製品情報と詳細なレビューであり、カテゴリ説明の「スマートフォン、デジタルカメラ、テレビ、ゲーム機、PC 周辺機器、生活家電などの新製品情報、詳細なレビュー、技術解説」に完全に合致します。関連記事も IT・Web 関連やエンターテインメントコンテンツに触れており、カテゴリの範囲内です。</p>
不正解例
<b>True label:</b> kaden_channel
<p><b>Original Text</b></p> <p>ツイッターや Facebook などで話題になっているのが「WEB 業界ろくろ回しすぎ」というもの。いったいどういうことだろうか。これは、ネットなどで多く掲載されている、インタビュー記事の感想だ。インタビューに答えている人の写真が掲載されていることが多いが、…</p>
<p><b>Output</b></p> <p>分類結果: topic_news</p> <p>理由: 記事は Twitter や Facebook などのソーシャルメディアにおける「痛い人」の行動とその周囲の反応に焦点を当てており、これは「topic_news」カテゴリが扱う「現代社会で注目を集める多岐にわたる話題」や「インターネット上の反応」に合致します。特に、SNS 上での人間関係の問題や世論の動向を伝える内容が特徴です。</p>

## C パラメータ設定

4.1 節にて、設定したモデルの詳細なパラメータについて、以下の表 11 および表 12 に示す。説明文生成の際には、テキストの多様性と忠実性のバランスを考慮し Temperature を 0.5 に設定した。一方、記事分類の際には、結果の再現性（決定論的生成）を確保するため 0.0 とした。

表 11 RoBERTa の Fine-tuning におけるハイパーパラメータ設定

項目	設定値
最適化関数	AdamW
学習率	1e-5
損失関数	交差エントロピー誤差
バッチサイズ	8
エポック数	最大 30
Early Stopping	適用 (patience=2)
最大系列長	510

表 12 LLM の推論パラメータ設定

項目	説明文生成 (提案手法)	文書分類 (推論)
Temperature	0.5	0.0
Max Tokens	-	4000