Adobe Behaviour Simulation Challenge

Ebad Sayed & Aaditya Jain
Indian Institute of Technology, Dhanbad
https://github.com/ES7/Adobe-Behaviour-Simulation-Challenge

Abstract

In the modern era of e-commerce and marketing, digital marketing has surpassed its competitors in terms of user engagement and product sales. With the rise of generative AI and AI-driven image creation, digital marketing has reached unprecedented levels. However, when analyzing large-scale data, it's evident that certain images on social media tend to attract more consumer attention and engagement compared to others, even when conveying the same message or featuring the same subject. Therefore, it is crucial to understand these user behavior patterns to refine the image generation process. The goal is to enable AI to produce not only visually appealing images but also those that resonate with users and drive significant engagement, fulfilling their intended purpose.

1 Introduction

We utilized a comprehensive dataset from Twitter enterprise accounts, performing exploratory analysis and data cleaning. By processing both text and images with pretrained models, we built a multi-modal dataset. An ensemble of regressors was employed to predict user engagement, while media content was converted to text for task two. Additionally, prompt engineering and the fine-tuning of a Large Language Model using LoRA and PEFT strengthened our approach. As a further enhancement, we developed an image generation model to complement the content, optimizing for user engagement and driving key performance indicators (KPIs).

2 Exploratory Data Analysis

Analyzing user engagement across a large corpus of tweets uncovers diverse patterns. The average tweet garners 773 likes, but a considerable standard deviation of 4931 suggests a long-tail distribution. Prominent usernames like CNN and EuroLeague, along with businesses such as IndependentNGR and AMCTheatres, lead in engagement. The inferred companies include news organizations (CNN, CBC), tech firms (Cisco), and financial institutions. Daily patterns suggest a possible weekly engagement cycle, with likes peaking on Sundays (889) and declining on Thursdays (675).

3 Data Processing

We implemented a comprehensive text preprocessing pipeline that included the removal of punctuation, stop words, and URLs, as well as the correction of spelling and grammar to enhance the overall quality of the text. Additionally, temporal features such as day of the week, time of day, and the interval since the company's previous tweet were extracted to account for time-sensitive factors influencing user engagement. To enable brand-specific insights, the inferred company name was appended to each cleaned tweet, facilitating the analysis of engagement patterns at the brand level. Furthermore, we incorporated tweet images into a multi-modal analysis, enriching our understanding of engagement dynamics by combining both visual and textual elements.

Step	Description
Text Cleaning	Remove noise, standardize formatting
Date-time Feature	Extract temporal attributes for analysis
Image Captioning	Downloaded the media files and applied image captioning model on those files
Reducing Data	Reduced the dataset length from 300K samples to 100K samples
Tokenization	Tokenized the content column of the given dataset
Splitting the Data	Combined captioned and tokenized data and split into train and test dataset

Table 1: Data Preparation Steps

Recognizing the critical role of visual content, we utilized **Salesforce's BLIP (large)**, an advanced image captioning model, to automatically generate textual descriptions for the downloaded images. This provided a rich set of additional features that captured the visual context of the tweets, allowing us to integrate them into the language model for generating more contextually relevant and engaging content. Finally we applied text cleaning techniques to correct any spelling mistakes occured due to BLIP.

4 Task 1: Behaviour Analysis

Given the content of a tweet (text, company, username, media URLs, timestamp), the task is to predict its user engagement, measured by likes.

4.1 Training Classifier with BERTweet Embeddings

Leveraging the power of BERTweet embeddings, we trained a classifier to categorize each tweet into predefined classes based on the anticipated number of likes. The established classes represent a logarithmic scale of engagement:

Bucket No.	Likes
Bucket 1	0-10
Bucket 2	11-100
Bucket 3	101-300
Bucket 4	301-500
Bucket 5	501-1000
Bucket 6	1001-2500
Bucket 7	2500

Table 2: Bucket and Likes

Using the DistilBERT model, we trained a classifier to categorize metadata prompts generated in the earlier stage into buckets based on their ability to predict likes. DistilBERT, a highly efficient pre-trained transformer model, was employed to generate rich contextual representations of tweet text, effectively capturing subtle nuances in meaning and sentiment.

4.2 Future Work

Regression on Buckets

We plan to further refine our approach by applying a regression model to each bucket to predict likes within the specified range of that bucket. Models such as **XGBoost** or **CatBoost** can be employed for this task. Following this, our predictive framework will be capable of estimating the number of likes for new, unseen data by first classifying the data into a bucket and then applying the corresponding regression model.

Results

Upon completion of training, the predictive model can be applied to the test dataset, with results evaluated using **RMSE** to measure the accuracy of predicting likes. Although there may be some loss reflected in the RMSE, the classification of tweets into buckets ensures that we can accurately predict the relative success of a tweet. By leveraging bucket classification, we are able to assess a tweet's potential even when there is some loss in RMSE.

5 Task 2: Content Analysis

Given the tweet metadata (company, username, media URL, timestamp), generate the tweet text.

5.1 Model Selection and Training

For fine-tuning both the **Qwen 1.5B** and **Bloom 7B** models, we used two advanced techniques to optimize performance. First, we implemented **PEFT** (Prefix-Tuning with Early Fine-tuning), focusing on fine-tuning the early layers of the models with domain-specific data while keeping the later layers more generalizable. This allowed us to leverage the models' pre-trained knowledge while adapting them specifically for social media content generation. Second, we utilized **LoRA** (Low-Rank Adaptation), which improved efficiency by reducing the number of parameters that needed updating during fine-tuning. This technique boosted training efficiency, leading to faster model adaptation to new data.

Configuration	Value
r (Attention heads)	16
lora_alpha	32
lora_dropout	0.05
bias	none
task_type	CAUSAL_LM

Table 3: Model Configuration

Arguments	Value
per_device_train_batch_size	4
gradient_accumulation_steps	4
warmup_steps	100
max_steps	400
learning_rate	2×10^{-4}
fp16	True
logging_steps	1
output_dir	outputs

Table 4: Training Arguments

6 Results

The fine-tuned Qwen 1.5B model achieved a loss of 1.808 after 400 iterations on the training data, highlighting its proficiency in text prediction. During evaluation on the test dataset, this model processed each iteration in approximately 2.61 seconds. In comparison, the fine-tuned Bloom 7B model recorded a loss of 1.9436 after

400 iterations, showcasing its effectiveness in predicting text. This model took about 11.42 seconds per iteration on the test dataset.

6.1 Evaluation Metrics

The evaluation was conducted using BLEU (1-4), ROUGE (1, 2, L), and CIDEr metrics on the first 500 predicted outputs from both models. For each metric, we calculated the average values to provide a comprehensive comparison of model performance.

Metrics	Value
BLEU	0.1364, 0.1123, 0.1033, 0.09844
ROUGE	0.2192, 0.1494, 0.2106
CIDEr	0.0000

Table 5: Evaluation of Qwen 1.5B

Metrics	Value
BLEU	0.4935, 0.4935, 0.4935, 0.4935
ROUGE	0.7317, 0.7176, 0.7317
CIDEr	0.0000

Table 6: Evaluation of Bloom 7B

6.2 Model Comparision

Based on the evaluation metrics, the Bloom 7B model consistently outperformed the Qwen 1.5B model in generating text that closely resembles the reference data. Bloom demonstrated superior ability in capturing both individual words and longer sequences, making it more effective for the text generation task. Additionally, in the context of social media content creation, Bloom's outputs were analyzed for quality and coherence. The model successfully incorporated appropriate emojis, enhancing the visual appeal and emotional resonance of the message while maintaining textual coherence and alignment with the intended narrative. In contrast, although the Qwen model conveyed the same message, it encountered character encoding issues that resulted in garbled emoji representations, diminishing both the visual appeal and potentially confusing readers. This highlights Bloom's overall effectiveness in generating contextually relevant and visually appealing content compared to Qwen.

References

[Sanh, Debut, Chaumond, Wolf,] Victor SANH, Lysandre DEBUT, Julien CHAUMOND and Thomas WOLF *DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter.* arXiv preprint arXiv:1910.01108, 2019. https://arxiv.org/abs/1910.01108.

[Li, Li, Xiong, *et al.*] Junnan Li, Dongxu Li, Caiming Xiong, Steven Hoi and Salesforce Research *BLIP: Bootstrapping Language-Image Pre-training for Unified Vision-Language Understanding and Generation.* arXiv preprint arXiv:2201.12086, 2022. https://arxiv.org/pdf/2201.12086.

[Hui, Yang, Cui, *et al.*] Binyuan Hui, Jian Yang, Zeyu Cui, Jiaxi Yang, Dayiheng Liu, Lei Zhang, Tianyu Liu, Jiajun Zhang, Bowen Yu, Kai Dang, An Yang, Rui Men, Fei Huang, Xingzhang Ren, Xuancheng Ren, Jingren Zhou, Junyang Lin†, Qwen Team and Alibaba Group *Qwen2.5-Coder Technical Report.* arXiv preprint arXiv:2409.12186, 2024. https://arxiv.org/pdf/2409.12186.

[Scao, Fan, Akiki, *et al.*] Teven Le Scao, Angela Fan, Christopher Akiki, Ellie Pavlick, Suzana Ili ´c, Daniel Hesslow, Roman Castagné, Alexandra Sasha Luccioni, Fran¸cois Yvon, Matthias Gallé, Jonathan Tow, Alexander M. Rush, Stella Biderman, Albert Webson, Pawan Sasanka Ammanamanchi, Thomas Wang, Benoit Sagot, Niklas Muennighoff, Albert Villanova del Moral, Olatunji Ruwase, Rachel Bawden, Stas Bekman, Angelina McMillan-Major, Thomas Wolf, Iz Beltagy, Huu Nguyen, Lucile Saulnier, Samson Tan, Pedro Ortiz Suarez, Victor Sanh, Hugo Lauren¸con, Yacine Jernite, Julien Launay, Margaret Mitchell, Colin Raffel *BLOOM: A 176B-Parameter Open-Access Multilingual Language Model.* arXiv preprint arXiv:2211.05100, 2023. https://arxiv.org/pdf/2211.05100.

[Kishore, Salim, Todd and Wei-Jing,] Papineni Kishore, Roukos Salim, Ward Todd and Zhu Wei-Jing *[BLEU: A Method for Automatic Evaluation of Machine Translation.* Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL), 2002. https://aclanthology.org/P02-1040.

[Lin,] Lin, Chin-Yew Lin *[ROUGE: A Package for Automatic Evaluation of Summaries.* Proceedings of the Workshop on Text Summarization Branches Out (W04), 2004. https://aclanthology.org/W04-1013.

[Vedantam, Zitnick and Parikh,] Ramakrishna Vedantam, C. Lawrence Zitnick and Devi Parikh *CIDEr: Consensus-based Image Description Evaluation.* arXiv preprint arXiv:1411.5726, 2015. https://arxiv.org/pdf/1411.5726.