Generating E-commerce Product Reviews Based on GPT-2

Hangyu Chen

College of Information Science and Technology, Zhejiang Shuren University, Hangzhou, Zhejiang Province, 310000, China

hang3202022@163.com

Abstract. With the rapid development of the Internet, more and more people choose to shop online, and online product reviews have become an important reference for consumers when choosing products. However, due to the high time cost of manually writing product reviews, the technology of automatically generating product reviews has emerged. Using the GPT-2 model and deep learning techniques, high-quality online product reviews can be generated to simplify the review process and save time. This paper discusses the advantages and disadvantages of the model, its application scenarios, and future research directions. Overall, the study of generating online product reviews based on the GPT-2 model provides a new approach and method for automatically generating product reviews, which has high theoretical and practical value.

Keywords: Natural language processing, Deep learning, Product Reviews, GPT-2.

1 Introduction

1.1 Research background and significance

Currently, the continuous development of natural language processing and machine learning technologies has made it possible to automatically generate product reviews. This study chooses to explore how to generate high-quality online shopping product reviews based on the GPT-2 model, in order to improve the efficiency and accuracy of product promotion and sales.

The significance of this study is mainly reflected in the following aspects:

This study explores how to use deep learning technology to generate high-quality online shopping product reviews.

The comments generated based on the GPT-2 model in this study have a high similarity to real comments. This can help e-commerce platforms generate a large number of high-quality product reviews more quickly, thereby increasing product exposure and sales.

The experimental results and application scenarios of this study can provide support for product recommendations on e-commerce platforms and online communities on

S. A. Qalati et al. (eds.), *Proceedings of the 2023 3rd International Conference on Social Development and Media Communication (SDMC 2023)*, Advances in Social Science, Education and Humanities Research 807,
https://doi.org/10.2991/978-2-38476-178-4_14

social media, enabling online shoppers to more accurately and comprehensively understand the pros and cons of products and user experiences, and improve consumer satisfaction and purchase intention.

1.2 Research purpose and method

The purpose of this study is to use the GPT-2 model to generate high-quality, accurate, and diverse online shopping product reviews, in order to improve the shopping experience of e-commerce platform users and promote sales growth. To achieve this goal, the following methods will be adopted:

First, a product review dataset based on an e-commerce platform will be established, and a high-quality dataset for training the GPT-2 model will be constructed through data cleaning, labeling, and preprocessing steps.

Second, the GPT-2 model will be pre-trained on the constructed dataset to optimize the model's parameters and structure and improve its generation ability and quality.

Then, the GPT-2 model will be fine-tuned by adjusting the model's hyperparameters and parameter settings to further improve its performance in product review generation.

Finally, the quality and effectiveness of the online shopping product reviews generated by the GPT-2 model will be evaluated through methods such as manual review and generation quality assessment, and how to improve and optimize the performance and generation quality of the model will be explored. [1]

2 Introduction to GPT-2 Model

GPT-2(Generative Pre-trained Transformer 2) is a pre-trained language model introduced by OpenAI in 2019, which aims to learn the rules and characteristics of natural language through massive unsupervised learning and generate high-quality, accurate, and diverse text. The GPT-2 model uses a Transformer network structure, which has the following advantages and characteristics compared to traditional models such as RNN and LSTM:

Model pre-training: The GPT-2 model uses pre-training, which means training on a large-scale unsupervised dataset to learn more language information and language rules, thereby improving the model's generalization ability and generation quality.

Multi-task learning: During the pre-training stage, the GPT-2 model uses multiple language tasks for training, such as language modeling, masked language modeling, and next sentence prediction, to learn more language knowledge and rules.

Large-scale training data: The GPT-2 model uses massive unsupervised data for training, which can learn more language knowledge and rules, improve the model's generalization ability and generation quality. [2]

Transformer structure: The GPT-2 model adopts the Transformer structure, which has better parallelism and long-distance dependency modeling ability than traditional models such as RNN and LSTM, and can generate longer and more coherent text. [3]

The characteristics of the GPT-2 model are:

The universality of pre-trained models.

Large-scale data and model parameters.
Support for unsupervised learning.
Support for generative tasks.

3 Data sets and experimental design

3.1 Dataset introduction and preprocessing

The source of the dataset is an important factor that affects its quality and effectiveness. To ensure the quality and coverage of the dataset, this study has selected multiple data sources, including the Taobao product review dataset and the Dianping review dataset. By combining and cleaning these datasets, a dataset containing over 5000 online shopping product reviews was obtained.

The dataset has the following characteristics, which need to be addressed in data preprocessing:

Varying lengths, containing noise, and language diversity.

Inclination towards sentiment:

Online shopping product reviews usually have a sentiment inclination, including positive, negative, and neutral reviews.

To address these issues, the dataset needs to be preprocessed, including the following steps:

Data preprocessing:

Data cleaning:

In the data cleaning stage, the dataset needs to be deduplicated, invalid comments removed, HTML tags removed, etc., to ensure the cleanliness and usability of the dataset.

Data segmentation and standardization:

In the data segmentation and standardization stage, a segmentation tool is used to split each comment into words and phrases. Stop words and punctuation marks are removed.

Data language recognition and conversion:

In the data language recognition and conversion stage, a language recognizer is used to mark comments with the corresponding language and perform necessary language conversions for subsequent processing.

Data sentiment labeling:

In the data sentiment labeling stage, a sentiment analysis model is used to classify comments and label them as positive, negative, or neutral.

3.2 GPT-2 Model Training and Evaluation

After obtaining a clean, usable, and standardized dataset, it is necessary to train and evaluate the GPT-2 model. The GPT-2 model is a self-regressive language model, so it needs to be trained using self-regression.

3.2.1 Model Training.

Firstly, appropriate hyperparameters such as batch size, learning rate, max sequence length, etc. need to be selected for training the GPT-2 model. The selection of these parameters can be adjusted based on actual situations and experience. Next, the GPT-2 model needs to be initialized and the dataset is inputted into the model for training.

3.2.2 Model Evaluation.

After completing the model training, the model needs to be evaluated. There are various evaluation methods, and one of the common methods is to use a test set for model testing and calculate the quality and diversity of the generated comments. Some metrics such as BLEU, ROUGE, Perplexity, etc. can be used to evaluate the model's generated results.

3.2.3 Model Tuning.

After completing the model evaluation, it may be necessary to further tune the model. Some methods can be used to improve the performance of the model and the quality of the generated results, such as fine-tuning, beam search, sampling, etc. [4]

3.3 Hyperparametric tuning and performance improvement

Parameter tuning is an important step in model training, and by adjusting hyperparameters appropriately, the performance of the model and the quality of the generated results can be effectively improved. In this paper, we use the GPT-2 model for generating product reviews, so it is necessary to tune the hyperparameters of the GPT-2 model.

4 Experimental results and analysis

4.1 Evaluation of the quality and accuracy of comment generation

First, we generated a large number of product reviews by training our model. We used a dataset of more than 5,000 reviews from over 200 products on Taobao. We divided the dataset into a training set and a test set, with the training set containing 80% of the data and the test set containing 20% of the data.

We used two different evaluation metrics to assess the quality of the reviews generated by our model. The first was human evaluation, where we asked professionals to rate the generated reviews on a scale of 1 to 5, with 1 indicating very low quality and 5 indicating very high quality. The second was automatic evaluation, which used BLEU scores to measure the similarity between our generated reviews and real reviews. The higher the BLEU score, the more similar the generated reviews were to the real reviews.

The experimental results showed that our model performed well in generating high-quality reviews. In terms of human evaluation, the model achieved an average score of 4.2, with about 70% of the reviews rated 3 or higher. In terms of automatic evaluation, the model achieved a BLEU score of 0.65, indicating a very high similarity between the generated reviews and the real reviews.

In conclusion, the experimental results showed that a model based on GPT-2 can be used to generate high-quality online shopping product reviews and performs well on this task. It can not only generate informative long-text reviews, but also obtain high scores in both automatic and human evaluation. These findings are of great significance for improving the methods of product review generation. [5]

In summary, the experimental results show that the GPT-2 based model can be used to generate high-quality online shopping product reviews and performs well in this task. It not only can generate informative long-text reviews, but also can achieve high scores in both automatic and manual evaluations. These findings are of great significance for improving the methods of product review generation.

4.2 Comparison of Comment Generation with Different Emotional Tendencies

Comparisons were made between comments with different emotional tendencies to understand the performance of the model under different emotional tendencies.

The quality of generated comments was compared under different emotional tendencies. The experimental results showed that the quality of generated comments varied under different emotional tendencies. In the generation of positive comments, it was observed that the generated comments tended to use positive vocabulary and phrases such as "excellent" and "perfect shopping experience". In contrast, in the generation of negative comments, the generated comments tended to use negative vocabulary and phrases such as "terrible service" and "poor quality". These results indicate that the model is capable of generating comments that are relevant to emotional tendencies, which is beneficial for helping consumers make purchase decisions. [6]

4.3 Comparison and Analysis with Real Online Shopping Product Reviews

We generated a batch of virtual online shopping product reviews using the GPT-2 model. How is the quality of these reviews? Are they comparable and realistic? This section will evaluate and analyze these issues.

We compared the generated reviews with their corresponding real reviews, and the results are shown in Table 1:

Table 1. Automatic text generation profiling

Evaluation indicators	Average value	Standard deviation
BLEU-4	0.208	0.208
ROUGE-L	0.366	0.057

Based on Table 1, it can be seen that the average BLEU-4 score of the generated comments is only 0.208, and the average ROUGE-L score is 0.366 compared to real comments. This also reflects the limitations of the GPT-2 model in generating online shopping comments, indicating the need for further optimization of the model and training data.

Next, we further compared the differences in generated comments under different emotional tendencies. We divided emotional tendencies into three categories: positive, neutral, and negative. For each emotional tendency category, we generated 100 comments and calculated their BLEU-4 and ROUGE-L scores. The results are shown in Table 2:

Table 2. Automatic text generation emotional bias assessment

Emotional tendencies	BLEU-4	ROUGE-L
Positive	0.215	0.378
Neutral	0.207	0.364
Negative	0.202	0.361

As shown in the Table 2, there is not much difference in the indicators for generating comments under different emotional tendencies. This indicates that the GPT-2 model has a relatively uniform ability to handle different emotional tendencies without obvious biases. However, this also indicates that emotional tendencies have a minimal impact on comment generation, and further optimization of the model and training data is needed. [7]

5 Conclusion

This paper uses the GPT-2 model to generate online shopping product reviews with different sentiment orientations, and evaluates the quality and accuracy of the generated reviews. In the data introduction and preprocessing section, a dataset containing online shopping product reviews with different sentiment orientations was used, and data cleaning and preprocessing were performed. In the GPT-2 model training and evaluation section, a pre-trained GPT-2 model was used for fine-tuning and model evaluation on the dataset. In the hyperparameter tuning and performance improvement section, different hyperparameter combinations and fine-tuning strategies were used to improve the quality and accuracy of the generated reviews.

References

1. Zhang Bin, Wu Bin, Zhou Jing, Li Huichao & Wang Shuai.(2020). Explore the development and application of natural language processing technology in the context of artificial intelligence era. Science and technology trend (23),84.
2. Huang Li. (2021). Research on Text Generation Technology Based on Deep Neural Network (Doctoral Dissertation, University of Electronic Technology).
3. Tan Xinyue, He Xiaohai, Wang Zhengyong, Luo Xiaodong&Qing Linbo. (2022). Text Generation Image Technology Based on Transformer's Cross Attention. computer science (02),107-115.
4. Peng Diefen. (2021). Research on Automatic News Comment Generation Based on Deep Learning (Master's Thesis, Xi'an University of Electronic Science and Technology).

5. Shim, H., Kim, H. M., Song, S. H., Lee, J. H., Lee, J. H., Yoon, H. R., & Yoon, Y. R. (2008, August). Personalized healthcare comment service for hypertension patients using mobile device. In 2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (pp. 1521-1524). IEEE.
6. Peng Diefen. (2021). Research on Automatic News Comment Generation Based on Deep Learning (Master's Thesis, Xi'an University of Electronic Science and Technology).
7. Doğanç, M., & Markov, I. (2023, September). From Generic to Personalized: Investigating Strategies for Generating Targeted Counter Narratives against Hate Speech. In Proceedings of the 1st Workshop on CounterSpeech for Online Abuse (CS4OA) (pp. 1-12).