Journal of Contemporary Educational Research, 2024, Volume 8, Issue 11
http://ojs.bbwpublisher.com/index.php/JCER
ISSN Online: 2208-8474
ISSN Print: 2208-8466

Training and Implementation of Subjective Questions Scoring System Based on the Baidu Qianfan Model Platform

Xiaoyun Zhu*

Yunnan Open University, Kunming 650500, Yunnan Province, China

***Corresponding author:** Xiaoyun Zhu, 15812040598@139.com

Abstract: Leveraging the Baidu Qianfan model platform, this paper designs and implements a highly efficient and accurate scoring system for subjective questions, focusing primarily on questions in the field of computer network technology. The system enhances the foundational model by utilizing Qianfan's training tools and integrating advanced techniques, such as supervised fine-tuning. In the data preparation phase, a comprehensive collection of subjective data related to computer network technology is gathered, cleaned, and labeled. During model training and evaluation, optimal hyperparameters and tuning strategies are applied, resulting in a model capable of scoring with high accuracy. Evaluation results demonstrate that the proposed model performs well across multiple dimensions—content, expression, and development scores—yielding results comparable to those of manual scoring.

Keywords: Subjective score; Natural language processing; Deep learning; Baidu Qianfan large model platform; Supervised fine-tuning; Model training and evaluation

Online publication: November 27, 2024

1. Introduction

Subjective assessment, as a key part of educational assessment, has long been trapped by high labor costs, low efficiency, and subjective deviation. In order to solve this problem, a subjective questions scoring system based on natural language processing and deep learning technology has emerged [1]. The Baidu Qianfan model platform brings together the essence of artificial intelligence (AI) and provides solid technical support for building an efficient and accurate subjective scoring system.

Based on the Baidu Qianfan model, this paper studies the training and implementation of the subjective questions scoring system [2]. Through accurate demand analysis and integration of platform resources, an automatic scoring and evaluation system for computer network technology is built [3].

In the training process, we make full use of the preset data set and training tools provided by the Qianfan platform and optimize the basic model through advanced methods such as supervised fine-tuning

(SFT) [4]. At the same time, we also design a complete set of evaluation systems to ensure that the trained model can accurately and stably output scoring results.

2. Building the development environment of the Qianfan model platform

To use the Baidu Intelligent Cloud Qianfan platform to train a large model for subjective question scoring, the following steps are followed:

(1) Registering as a Baidu Intelligent Cloud user: Registration is done for non-users of Baidu Intelligent Cloud. On the official website of Baidu Intelligent Cloud, clicking the "Register" button on the home page, and filling in the relevant information to complete the registration process. During the registration process, please ensure that the information provided is accurate, so that one can log in and use the services smoothly in the future.

(2) Logging in to Baidu Intelligent Cloud account: After completing the registration, the account information is used to log in to Baidu Intelligent Cloud. The username and password are entered on the login page, or logging in through other login methods (such as mobile verification code, WeChat login, etc.). After successfully logging in, one will enter the console interface of Baidu Intelligent Cloud.

(3) Entering Qianfan AI Native App Store: In the console interface of Baidu Intelligent Cloud, looking for and clicking the entrance of "Qianfan AI Native App Store." This app store brings together various AI services and solutions provided by Baidu Intelligent Cloud, including the Big Model Platform [5].

(4) Visiting the Big Model Platform: Browsing or searching for keywords such as "Big Model Platform" in the Qianfan AI Native App Store. After finding the corresponding service, clicking to enter the detail page of the Big Model Platform. Here, one can find information about Big Model's features, documentation, and sample code.

Install Qianfan Python SDK via pip, and set the related environment variables:

```
! pip install -U "qianfan[dataset_base]"
import os
# The following environment variables are used by the Qianfan OpenAI Adapter
os.environ["QIANFAN_ACCESS_KEY"] = your_qianfan_console_access_key"
os.environ["QIANFAN_SECRET_KEY"] = your_qianfan_console_secret_key"
```

3. Problem description

A teaching assistant in a computer network technology course is required to mark a number of subjective questions, using the Baidu Qianfan model platform to build a subjective questions scoring system. The score for directly calling the model is 56, but the accuracy and completeness of the answers are not enough to reach the level of high scores. To analyze the reasons, we need to solve them from the following aspects:

(1) Insufficient domain adaptability of the model: Although the general model can deal with a wide range of natural language tasks, it lacks a deep understanding and professional knowledge of the specific field of network technology, which makes it difficult for the model to accurately identify and evaluate the professional terms, concepts, and the logical relationships between them when evaluating students' work, thus affecting the accuracy of the scoring [6].

(2) The lack of training data and low annotation quality: Computer network technology is highly professional, the relevant training data is small, and high-quality annotations are needed to ensure that the model can learn the correct knowledge. Due to insufficient training data or inaccurate labeling, it is difficult for the model to form a comprehensive and accurate understanding of network technology in the training process, which will affect the scoring effect [7].

(3) Standardization and consistency of scoring criteria: The scoring criteria of subjective questions are often complex and diverse, so it is necessary to ensure that the model can accurately understand and apply these criteria in the scoring process. However, the general model may not be able to fully understand and adapt to these specific scoring criteria, resulting in biased scoring results. Therefore, we need to build a model that can accurately understand and apply the scoring criteria of subjective questions of computer network technology to ensure the accuracy and consistency of the scoring [8].

According to the above problem, we need to optimize from several aspects: collect and obtain high-quality field data set, select or construct the model of task in the field of architecture, design targeted training strategies and evaluation mechanism, and build a model of continuous optimization and update mechanism.

4. Data preparation

4.1. Preparing the dataset

Firstly, it is necessary to widely collect all kinds of subjective questions in the field of computer network technology and their reference answers, student answers, and corresponding ratings. These data can come from test questions over the years, online learning platforms, teacher resources, etc. In the process of collection, it is necessary to ensure the diversity and coverage of the data to fully reflect the characteristics of subjective questions in the field. Subsequently, data cleaning was carried out to eliminate repetition, error, or information unrelated to students' answer scoring to ensure the accuracy and validity of the data. Finally, according to the scoring standards, each question and answer were analyzed in detail, and the key scoring points, important keywords, highlights, and errors in students' answers were marked, which provided detailed guidance for subsequent model training [9].

After finishing the data set, it is necessary to convert the data into the format supported by Baidu Qianfan large model platform. This usually involves the data presented in a structured way, such as JSON or CSV file format. During the conversion process, a reasonable field structure needs to be designed for each subjective question and its related information, including question number, question content, reference answer, student answer, grading criteria, scoring points, etc. At the same time, the basic information of the data set, such as the name of the data set, the data type, and the possible additional information, such as the description of the data set, the creator, and the creation time, should be filled in according to the requirements of the platform [10]. It is ensured that the data format is strictly consistent with the platform requirements, so that the platform can smoothly load and parse the data, laying the foundation for subsequent large model training.

Using the Qianfan Python SDK, we can easily load the dataset:

```
import json
from qianfan.dataset import Dataset
# load the preset dataset for training
```

```
train_ds = Dataset.load(qianfan_dataset_id="ds-553hczysf3um4cc9")
# load the preset dataset for evaluation
eval_ds = Dataset.load(qianfan_dataset_id="ds-6ubasnsry5pa4azi")
```

4.2. Preparing the training parameters

Before we are ready to train the model, we must specify a number of key hyperparameters, such as learning rate and training rounds. The setting of these parameters has a significant impact on the final performance of the model on the test set, and different settings can guide the model to optimize in different directions [11]. Based on the accumulation of past experience, we specifically provide a set of SFT hyperparameter configuration schemes for ERNIE Speed, a lightweight high-performance large language model developed by Baidu. ERNIE Speed, as an innovation launched by Baidu in 2024, not only shows excellent general ability but also is particularly suitable as a cornerstone model for fine-tuning to better adapt to and solve complex problems in specific scenarios. At the same time, the model also has outstanding performance of reasoning and laid a solid foundation for efficient handling of various tasks. The SFT technique further enhances the adaptability of the model, so that it can continuously optimize its performance in the supervised learning environment.

5. Model training and evaluation

5.1. Initiating training

In the model training stage, a suitable pre-trained model should be carefully selected as a starting point in the large model training interface of the Baidu Qianfan large model platform. The platform supports a variety of mainstream large models including the Wenxin large model. Then, according to the specific needs of the subjective questions of computer network technology, the training parameters, such as iteration rounds and learning rate, are carefully configured, and special attention is paid to improving the semantic understanding and similarity calculation ability of the model. Next, an appropriate tuning strategy was selected, such as ERNIE Speed (suitable for general scenarios with limited data and little difference from the base model) or full update (suitable for professional fields with rich data and significant differences from the base model). After that, the data set was scientifically split into training set, validation set, and test set to ensure the multi-dimensional verification of model performance [12]. Finally, the carefully labeled data set was imported to start the model training process, and the running status and performance indicators of the model were closely monitored during training to ensure that the training effect met expectations.

5.2. Model evaluation

After completing model training, a crucial step is a thorough evaluation of the fine-tuned model, aiming to verify whether it has stably converged and achieved the desired performance goal. Using the powerful functions of the Qianfan Python SDK, users can not only easily use the built-in evaluation tools of the platform, but also customize the evaluation logic according to their personalized needs. We decided to build a special evaluator, which not only checks whether the model output conforms to the established format but also deeply analyzes the subtle differences between the generated results of the large model and the expected output, as well as various key evaluation indicators, so as to comprehensively measure the efficiency and applicability of the model.

It is verified that the model obtained by the evaluation has improved the stability of the answers

compared with the previous basic model, and the scoring results are closer to the results of manual scoring [13].

6. Model optimization

Model optimization is critical in the training and implementation of the subjective questions scoring system based on the Baidu Qianfan model platform. In order to ensure that the system can accurately and efficiently evaluate the answers of students or testers, we adopt a series of carefully designed optimization strategies [14]. Firstly, we use the powerful computing power and rich data resources of the Baidu Qianfan model platform to train the model, so that it can deeply understand the professional knowledge of different subject fields, as well as the subtle differences in language expression. Secondly, by introducing advanced natural language processing technology and deep learning algorithms, the model is fine-tuned to enhance its ability to identify key points of the text, evaluate the logical structure, and understand the emotional color, so as to more accurately capture the core information and value in the answer. Optimization techniques such as attention mechanism and multi-task learning are also used to further improve the generalization ability and robustness of the model, so as to ensure that it can maintain stable scoring performance in different scenarios. Through these comprehensive optimization measures, the system significantly improves the accuracy and efficiency of subjective question scoring.

7. Conclusion

Based on the Baidu Qianfan model platform, this study successfully constructed and trained a scoring system for subjective questions of computer network technology. Through the carefully designed data preparation, model training, and evaluation process, the system demonstrated efficient and accurate scoring ability. In the data preparation stage, we extensively collected and cleaned a large amount of subjective data to ensure the diversity and accuracy of the data set [15]. In the model training phase, we adopted the ERNIE-Speed-8K model and fine-tuned it with supervision to adapt to the specific needs of subjective question scoring. Through evaluation, we found that the model achieved high accuracy in multiple dimensions such as content score, expression score, and development score, which was similar to the results of manual scoring. With the accumulation of more data and the optimization of the algorithm, the accuracy and robustness of the scoring system will be further improved, bringing more innovation and value to the field of education.

Disclosure statement

The author declares no conflict of interest.

References

[1] Xiao L, Liu J, 2021, Combination of Subjective Item Grading Method Based on Similarity Research. Journal of Guizhou University: Natural Science Edition, 38(05): 64–68.

[2] Li Y, 2023, Big Model Reshapes Digital World. New Security, (06): 24–26.

[3] Zhang L, Zhang S, 2012, The Subjective Topic Grading Algorithm Based on Semantic Similarity Research. Journal of Hebei University of Science and Technology, 33(03): 263–265.

[4] Xu Z, 2021, Research on Short Text Subjective Question Scoring Review Algorithm and System Implementation,

dissertation, Zhongran University of Economics and Law.

[5] Zhou Z, 2023, From Baidu AI Painting, the Future Reform Behind the Tiangong Feast Screen Video. Big Data Era, 2023(2): 54–59.

[6] Wang S, Gong J, Wang Y, et al., 2023, Key Points Matching Based Scoring Method for Liberal Arts Subjective Questions. Journal of Chinese Information Processing, 37(6): 165–178.

[7] Yang J, 2018, Application of Machine Learning Algorithm in Data Mining. Electronic Technology and Software Engineering, (04): 191.

[8] Nan X, 2016, Automatic Scoring System of Subjective Questions Based on Sentence Similarity. Silk Road Vision, 2016(28): 54–57.

[9] Chen F, Yang Q, 2015, On the Related Technology of Data Integration. Science and Technology Information, 13(08): 30.

[10] Huang Q, 2018, Jiangxi Examination Prose Reading Subjective Item Grading Standard Research. Journal of Literature Education, 2018(3): 3.

[11] Wang X, Fu X, 2023, The Application Trend of Artificial Intelligence Technology in the Era of Big Model—Taking the Big Model of Cloud Slave Technology as an Example. China Security and Defense, 2023(12): 53–58.

[12] Huang C, Li H, Qi X, 2022, Design of a Small Program for Subjective Question Scoring Based on Natural Language Processing. Wen Yuan (High School Edition), 2022(2): 199–201.

[13] Yao X, 2021, An Improved KMP Algorithm and Its Application in the Subjective Topic Grading. Journal of Network Security Technology and Application, (06): 27–29.

[14] Hu Y, 2024, Big Models Accelerate the Release of AI Potential. Outlook, 2024(3): 60–61.

[15] Zhang F, Han L, 2024, Research on Application of Generative Artificial Intelligence Based on Large Model in Audit Practice. Internal Auditing in China, (08): 42–48.