User Rating Prediction Method Based on Fine-tuning of Large Language Models

Qi Zhang
School of Information Technology and Engineering
Guangzhou College of Commerce
Guangzhou, China
Email: qzhang [AT] gcc.edu.cn

Hao Zhong*
School of Computer Science
South China Normal University
Guangzhou, China
Corresponding author's email: scnuzhonghao [AT] foxmail.com

Abstract—**Online reviews in social networks reflect users' preferences for specific attributes of products. Accurate predictions of user ratings based on these reviews can help businesses better understand genuine user feedback. The purpose of this study is to fine-tune large language models using online reviews and corresponding user rating data, generating a large model for predicting user ratings based on reviews. An attention mechanism is introduced to calculate attention weights for fine-grained review texts, reflecting the contribution of different text features to user rating prediction. By visualizing these weights, the process of calculating the predicted rating values can be explained. Experimental results show that the proposed interpretable user rating prediction method can effectively visualize the attention weights of important text features in the decision-making process of the large rating prediction model. Compared to the baseline model, the mean absolute error is reduced by 1.96, and the root mean square error is reduced by 1.73.**

Keywords-large language model; fine-tuning; attention mechanism; interpretable prediction

I. INTRODUCTION

Online reviews are a typical form of user-generated content, usually written and published by end-users based on their personal experiences, opinions, or knowledge, reflecting the subjective views of users towards products, services, or content [1]. The user ratings associated with online reviews represent the overall evaluation of users towards products, services, or content. Online reviews and user ratings [2] provide valuable feedback to businesses and content creators, helping optimize products and services, thereby enhancing user satisfaction.

User rating prediction [3] is a task aimed at predicting the rating or preference of users for unassessed products, services, or content, playing an important role in recommendation systems. It reflects the overall preferences of users for unassessed products, services, or content, providing data support for businesses to adjust operational decisions. On Netflix (a streaming platform), user rating prediction helps recommend movies and TV shows. For example, if a user has consistently rated sci-fi movies (e.g., Interstellar) highly but given low ratings to romantic films, the system may predict a 4.5/5 rating for a new sci-fi release (Dune) and only 2/5 for a romantic movie (The Holiday). As a result, Netflix prioritizes Dune in the user's recommendations to maximize engagement and satisfaction. On Amazon (an e-commerce platform), rating prediction assists in product recommendations. If a user has previously purchased and highly rated smart home devices (e.g., smart bulbs, smart speakers), the system may predict a 4.8/5 rating for a smart plug and recommend it, increasing the likelihood of purchase.

Specifically, predicting user ratings based on online reviews [4][5] makes the prediction results interpretable, because online reviews contain rich semantic information, reflecting the degree of user preference for specific attributes of products, services, or content. Emerging large language models such as BERT (Bidirectional Encoder Representations from Transformers) [6], GPT (Generative Pre-trained Transformer)-3 [7], ESM (Evolutionary Scale Modeling) [8], ROBERTA (A Robustly Optimized BERT Pretraining Approach) [9], and XLNET (eXtreme Learning Machine Network) [10] are capable of semantic understanding of the input review text, extracting relevant features from the input reviews, and generating feature vectors containing the semantic information of the input text, which are then used for rating prediction tasks. However, general-purpose large language models are designed to perform well across a wide range of tasks and domains but often lack optimization for specific domains. For example, in the context of user rating prediction, user rating behavior and preferences may be closely tied to the context of a specific domain (e.g., categories like movies, books, or e-commerce products), and general models may fail to fully capture the domain-specific information.

The research presented in this paper builds a rating prediction model based on emerging large language models and online review data, aiming to improve the accuracy of rating predictions and visualize the prediction process. The main content of the research includes two parts. The first part involves using existing large language models such as BERT, GPT, ESM, ROBERTA, and XLNET to understand the semantics of online reviews. Based on the correspondence

between online reviews and user ratings, the existing large language models are fine-tuned to construct a user rating prediction model based on online reviews. The second part introduces an attention mechanism to compute the fine-grained attention weights of review texts, which represent the contribution of different textual features to the user rating prediction. By visualizing these weights, the calculation process of the predicted rating value can be explained.

The goal of this research is to propose an interpretable rating prediction method based on online reviews, with the following main contributions: (*i*) defining a feature representation method for fine-grained review text based on large language models; (*ii*) constructing a supervised dataset based on the correspondence between online reviews and user ratings, fine-tuning the large language model, and generating the user rating prediction model; (*iii*) calculating and visualizing the fine-grained attention weights based on the attention mechanism, achieving interpretable user rating predictions.

II. RELATED WORK

A. User Rating Prediction

User ratings can generally be predicted by extracting features from users' online reviews, behavioral data, or other user-generated content, using methods such as matrix factorization, machine learning, or deep learning [11]. For example, factors such as personal interests, interpersonal interest similarity, interpersonal rating behavior similarity, and interpersonal rating behavior diffusion can be integrated into a unified matrix factorization framework for rating prediction [12]. By considering the viewpoints and sentiment features of online reviews, as well as users' social behaviors, it is possible to construct user preferences and business characteristics, and apply matrix factorization to predict users' ratings for businesses [13]. Using LDA (Latent Dirichlet Allocation) to extract user-topic features from user labels, and integrating factors such as user sentiment and interpersonal influence into a probabilistic matrix factorization-based recommendation system can also predict user ratings [14]. By exploring the impact of internal factors, such as user reliability and popularity, which come from explicit behavioral data (ratings and trust), a better understanding of users can be achieved, leading to more accurate user modeling. Incorporating these internal factors into singular value decomposition models can enable user rating prediction [15]. Deep feedforward networks can uncover the true nature of user factors and item features, as well as their complex hierarchical relationships, influencing the rating prediction process. This can lead to the construction of user profiles and item characteristics, which are used to predict users' ratings of items [16].

Existing rating prediction methods can construct features of users and rating targets from different perspectives, effectively completing the user rating prediction task. However, these methods suffer from poor interpretability, especially for deep learning-based rating prediction approaches. How to interpret and visualize the prediction results requires further research, which is one of the objectives of this study.

B. Large Language Models and Fine-Tuning Methods

Large language models (LLMs) are complex models trained on massive text data, capable of performing various natural language processing tasks. Fine-tuning large language models [17] refers to further training the model on a specific task or domain dataset to improve performance for that task. For example, fine-tuning a large language model on a biomedical domain-specific labeled dataset can adapt the model to biomedical natural language processing tasks [18]. Fine-tuning a large language model within a federated learning framework, which provides privacy and legal protections, enables the model to assess tasks, architectures, customer numbers, and other factors in the healthcare domain [19]. By collecting written feedback from human participants on thousands of questions involving ethical and political issues, and evaluating the consistency and quality of candidate consensus statements generated by the large language model, a reward model can be trained to predict individual preferences. This allows the large language model to quantify and rank the attractiveness of group consensus statements [20]. A comprehensive collection of 140 biomedical text mining datasets, covering more than 10 task types (102 English datasets and 38 Chinese datasets), can be used as a corpus for fine-tuning LLMs on biomedical question answering and dialogue tasks [21].

Compared to training a large language model from scratch, fine-tuning requires less data and time. Fine-tuning allows the large language model to quickly adapt to different tasks and domains, thereby expanding its applicability. By applying online review and corresponding user rating data to fine-tune a large language model, the model can predict user ratings based on online reviews, reducing prediction errors. This is the main goal of this study.

C. Attention Mechanism-Based Explainable Methods

The attention mechanism [22] is a process that mimics human visual attention, allowing the model to focus on more important parts of the information while ignoring less important ones. This is particularly useful in natural language processing. Introducing attention mechanisms in machine learning and deep learning models not only improves model performance but also enhances interpretability [23]. For example, in recommendation systems, the attention mechanism can be used to capture user preferences and item attributes, enabling the system to explain its recommendations through dual attention-based recommendations [24]. In the field of medical image segmentation, incorporating attention mechanisms into convolutional neural networks allows for more accurate and interpretable medical image segmentation, helping to understand the most important spatial locations, channels, and scales in medical images [25]. For network intrusion detection tasks, combining attention mechanisms with multi-output deep learning strategies helps focus on accurate, interpretable multi-class classifications of network traffic data, thereby better distinguishing intrusion categories [26]. Friend link prediction is a key research problem in recommendation systems. For each user, their direct similarity relationships can be computed by merging user embeddings. Indirect similarity

relationships can be calculated using attention mechanisms based on intermediate neighbors, aiming to explain how neighbors influence the social relationships of the target user [27].

The attention mechanism has been widely used to explain the results of various tasks. A key challenge is the ability to visualize attention weights to enhance model interpretability. This is significant in practical applications, as it helps understand the decision-making process of the model and improves its transparency. As user-generated content, online reviews already possess a certain level of interpretability, but the challenge lies in the fact that online reviews are unstructured data containing substantial redundancy and irrelevant information. Another key objective of this study is to employ the attention mechanism, enabling large language models to automatically identify and focus on the most important parts of reviews for rating predictions, thereby providing more accurate user ratings and making the decision process of the model interpretable to users.

D. Challenge

The computational complexity of attention mechanisms is typically O(n2) is the sequence length,where n is the sequence length. This leads to excessive resource consumption for long text processing, especially during training and inference. Although optimizations like sparse attention and linear attention exist, computational efficiency remains a significant challenge.

Although attention mechanisms can capture long-range dependencies, as sequence length increases, attention weights for distant words may gradually weaken, leading to insufficient information propagation. Furthermore, some improvements, such as windowed attention, may restrict global information access, potentially affecting model performance.

Attention weights can provide some level of interpretability, but they do not always accurately reflect the model' s decision-making process. For example, high attention weights do not necessarily mean that a specific part of the input is decisive for the output, making attention-based interpretability controversial.

The attention mechanism amplifies biases in training data, potentially leading to unfairness or incorrect reasoning in real-world applications. For instance, if training data are biased toward certain topics or groups, the model may reinforce such biases, and the attention mechanism itself cannot effectively mitigate this issue.

While attention mechanisms can learn patterns in data, their generalization ability remains limited when dealing with different domains or unseen data. This is particularly true for low-resource languages or specialized domain texts, where the model may struggle to leverage attention effectively to capture sufficient contextual information, affecting output quality.

Since attention computation requires storing all query, key, and value matrices, memory requirements grow exponentially with sequence length. This not only affects large-scale model deployment but also poses challenges for mobile devices or low-resource environments.

Although attention mechanisms have significantly improved large language models, their computational cost, long-range dependency issues, interpretability, training data biases, generalization ability, and storage requirements remain key challenges. Future directions include more efficient attention variants (e.g., sparse attention), stronger bias correction methods, and task-specific optimizations.

III. Interpretable User Rating Prediction Method

The objective of this study is to propose an interpretable user rating prediction method based on online reviews from social networks. The main work consists of two parts. The first part involves constructing a supervised dataset using online reviews and their corresponding user ratings for fine-tuning large language models. This enables the large model to better apply to user rating tasks, thereby constructing the corresponding user rating prediction model. The second part introduces an attention mechanism during the fine-tuning process to calculate the weights of fine-grained review texts. By visualizing these fine-grained text weights, we aim to explain the contribution of different parts of the text to the rating prediction. The technical approach to be adopted is shown in Figure 1.

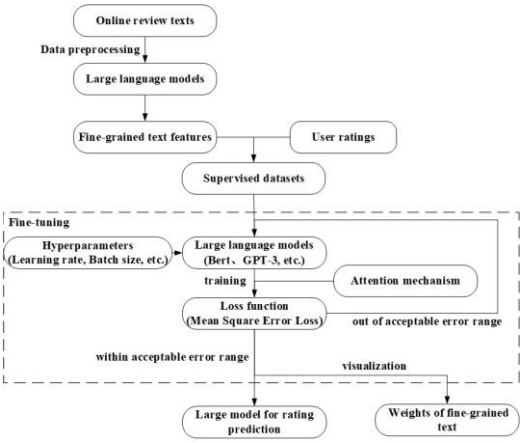


Figure 1. Research Technology Roadmap

A. Construction of the User Rating Prediction Model

The goal of this study is to leverage the generalization ability of large language models by fine-tuning them on a supervised dataset consisting of online reviews and user ratings, aiming to improve the model's performance on the rating prediction task.

After collecting data from various online platforms that includes online reviews and user ratings, data cleaning is performed. This includes removing noise such as labels, special characters, etc., correcting spelling errors, and eliminating duplicate content, ensuring the cleanliness and consistency of the supervised dataset. The input format for the large language model requires tokenization and segmentation, where the tokenizer built into the large language model is used to convert the supervised dataset into a form that the model can process.

This ensures that the input data is consistent with the processing method used during the model' s pre-training phase, reducing information loss caused by format conversion.

Once the supervised dataset is prepared, it is fed into the large language model, and appropriate training parameters (such as learning rate, batch size, and number of epochs) are selected to optimize the model. Optimizers and learning rate schedulers are set up to dynamically adjust the learning rate during training, improving both the training efficiency and effectiveness of the large language model. During training, the loss function is monitored to evaluate the model' s learning progress and generalization ability. The loss function is defined as the Mean Squared Error Loss (MSELoss), which is computed as shown in formula (1):

$$\text{MSELoss} = \frac{1}{n}\sum_{i=1}^{n}(y_i^{'} - y_i)^2 \ . \tag{1}$$

Where n is the size of the training set, $y_i^{'}$ is the predicted rating value from the large language model, and y_i is the true rating value. Suppose the input online review text x_i is passed through the large language model to obtain the hidden state h_i, and a linear layer is added for the rating prediction task. This linear layer maps the hidden state to the predicted rating $y_i^{'}$, where

$$y_i^{'} = W \times h_i + b \ . \tag{2}$$

W is the weight matrix of the linear layer, and b is the bias vector of the linear layer. Based on formulas (1) and (2), the gradient for the predicted rating $y_i^{'}$ can be calculated as follows:

$$\frac{\partial \text{MSELoss}}{\partial y_i^{'}} = \frac{2}{N}(y_i^{'} - y_i) \tag{3}$$

Then, the gradient of the loss with respect to the output h_i of the large language model can be calculated as:

$$\frac{\partial \text{MSELoss}}{\partial h_i} = \frac{\partial \text{MSELoss}}{\partial y_i^{'}} \times W = \frac{2}{N}(y_i^{'} - y_i) \times W \tag{4}$$

Based on the gradient computed from formula (4) and the selection of an optimizer (such as Adaptive Moment Estimation [27] or Stochastic Gradient Descent [28]), the parameters of the large language model are updated.

B. *Interpretable Prediction Based on the Attention Mechanism*

To make the predicted values output by the constructed rating prediction model interpretable, the attention mechanism is integrated into the fine-tuning process of the large language model. This improves the model's interpretability by generating a set of attention weights. These weights represent the contribution of each word in the input user text to the model's prediction. By visualizing these weights, we can clearly see which words the model "focused on" during prediction, making the prediction results easier to understand.

An attention mechanism is introduced between the output hidden state h_i of the large language model and the linear layer. The definition of the attention layer is as follows:

$$\text{Attention}(Q,K,V) = \text{softmax}(\frac{QK^T}{\sqrt{d_k}}) \times V \ . \tag{5}$$

Where Q,K,V is the parameter obtained by applying a linear transformation to the hidden state h_i, and d_k is the dimension of K. The attention weights can be calculated using the following formula:

$$\begin{aligned} Q &= W_Q \times h_i \\ K &= W_K \times h_i \\ V &= W_V \times h_i \end{aligned} \ , \tag{6}$$

where W_Q, W_K, and W_V are trainable weight matrices. The output of the attention scores is then used as the input to the linear layer:

$$y_i^{'} = \text{Attention}(Q,K,V) + b \tag{7}$$

The interpretability of the attention mechanism is reflected in the attention weights. By analyzing the attention weights, we can understand which parts of the input online review the model focused on during prediction. Specifically, the attention weights can reveal which words or sentences the large language model places more emphasis on when predicting user ratings based on the online review.

IV. EXPERIMENT

A. *Dataset and Evaluation Metrics*

The goal of the experiment is to evaluate whether the rating prediction model obtained through fine-tuning can reduce the error in rating prediction, and to explain the prediction results using visualization methods. The datasets used include two benchmark datasets [29], with descriptions of the datasets shown in Table 1.

Table 1 Online reviews dataset in social networks

Datasets	Reviews	Users	Products
Fine foods	568454	256059	74258
Movies	7911684	889176	253059

The error in rating prediction is evaluated using two metrics: Mean Absolute Error (MAE) and Root Mean Square

Error (RMSE). The formulas for calculating MAE and RMSE are shown in formula (8) and formula (9), respectively:

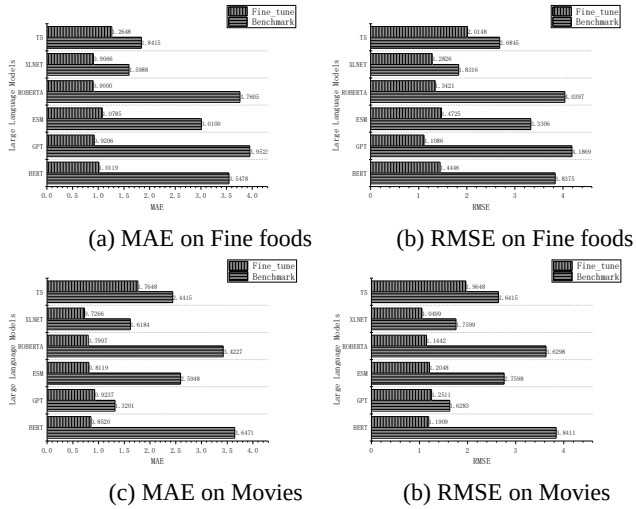
$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|y_i^{'} - y_i\right| \tag{8}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i^{'} - y_i)^2} \tag{9}$$

Where n is the size of the training set, $y_i^{'}$ is the predicted rating value from the large language model, and y_i is the true rating value.

B. Experimental Results and Analysis

The first part of the experiment compares the MAE and RMSE values of the ratings predicted by the large language model before and after fine-tuning. The large language models used for testing include the benchmark versions of BERT [6], GPT-3 [7], ESM [8], ROBERTA [9], XLNET [10] and T5[30]. The experiment compares the rating prediction performance of these five large language models before and after fine-tuning. The experimental results on the Fine Foods and Movies datasets are shown in Figure 2, respectively:



(a) MAE on Fine foods (b) RMSE on Fine foods

(c) MAE on Movies (b) RMSE on Movies

Figure 2. Comparison of Rating Prediction Results

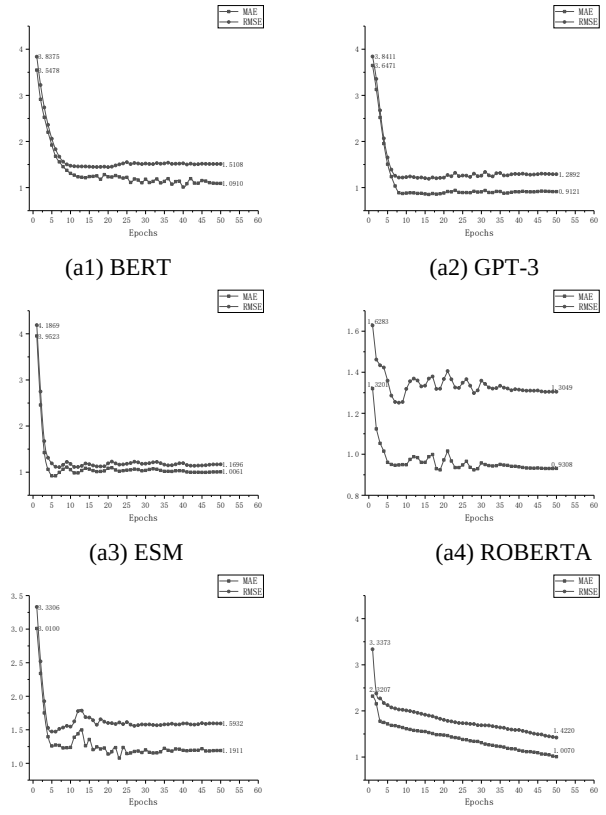
As shown in Figure 2, a supervised dataset was constructed using online reviews and user ratings from social networks, which was then used to fine-tune the large language model to build the user rating prediction model. Compared to the general-purpose large models, this method effectively reduced the error in rating prediction. The experimental results not only validate the effectiveness of the fine-tuning approach but also demonstrate that the fine-tuned models are better at capturing domain-specific linguistic features and sentiment tendencies, thereby improving their performance in rating prediction tasks. This innovation lies in optimizing the model with task-relevant data, making it more adaptable and accurate for specific tasks.

The second part of the experiment tests the trend of rating prediction error changes during the fine-tuning process of the large language model. A supervised dataset constructed from online reviews and user ratings in social networks was used to fine-tune the large language models BERT [6], GPT-3 [7], ESM [8], ROBERTA [9], XLNET [10] and T5[30]. The initial hyperparameter settings are listed in Table 2.

Table 2 The initialization hyperparameters of large language models

Learning rate	Batch size	Optimizer	Epochs
2*10^(-5)	16	AdamW	50

The changes in the rating prediction error are shown in Figure 3, respectively.

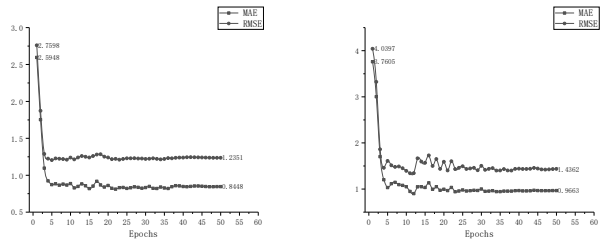


(a1) BERT (a2) GPT-3

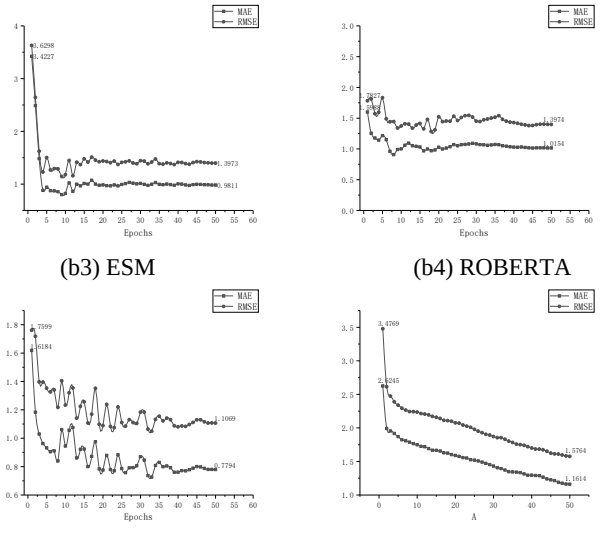
(a3) ESM (a4) ROBERTA

(a5) XLNET (a6) T5

(a) Prediction results of ratings using different large language models on Fine foods



(b1) BERT (b2) GPT-3



(b3) ESM (b4) ROBERTA

(b5) XLNET (b6) T5

(b) Prediction results of ratings using different large language models on Movies

Figure 3. Rating prediction results using different large language models on two publicly available datasets

As shown in Figure 3, as the fine-tuning process progresses, the performance of the rating prediction model gradually improves until it reaches a stable state. Specifically, the evaluation metrics MAE and RMSE decrease gradually in the early stages of training and stabilize after approximately 50 epochs. This is because, in the initial stages of training, the parameters of the large language model are randomly initialized, and its ability to predict user ratings is weak. Through continuous adjustment of parameters, the model gradually learns the features and patterns in the online reviews, leading to a significant decrease in error metrics. As the training continues, the model's parameters gradually approach an optimal point, and at this point, the rating prediction model's ability to predict user ratings reaches its best performance, and the error metrics stabilize. By this time, the model has captured the main patterns and features of the online review data, and further training only brings minimal performance improvements. If the number of training epochs is too high, the rating prediction model may begin to overfit the training data, leading to a decline in performance on the test data. The observed stabilization of the MAE and RMSE values indicates that the number of epochs at this point is appropriate, avoiding overfitting while ensuring the accuracy of the rating prediction model.

The third part of the experiment tests whether the integration of the attention mechanism into the constructed rating prediction model can explain the computation process of the predicted rating values. Taking the Fine foods dataset and the fine-tuning of the BERT model as an example, after incorporating the attention mechanism, the attention weights for certain parts of the Fine foods review text (ranked by the size of the attention weights) are shown in Table 3.

Table 3 Attention weight of some texts in Fine Foods reviews

en-US	Attention weight
Often	6.24×10^(-2)
Mild	7.81×10^(-3)
Hesitant	4.36×10^(-3)
Due	4.11×10^(-3)
First	3.54×10^(-3)
…	…

Based on the attention weights of the above text, when a user review is input, the rating prediction large model will predict the rating based on the attention weights corresponding to the text in the review. The combination of attention weights assigned to different tokens in the review determines the prediction. Tokens with higher attention weights contribute more to the prediction, indicating their significance in determining the sentiment or meaning of the review.

V. CONCLUSION

The purpose of this research is to use a supervised dataset constructed from online reviews and user ratings for fine-tuning a large language model, generating a user rating prediction model. Compared to the benchmark large language models, the user rating prediction model obtained through fine-tuning is more effective in rating prediction tasks, reducing both Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE). In particular, incorporating the attention mechanism during the fine-tuning process makes the predicted rating values of the rating prediction model more interpretable.

In future work, we will focus on exploring how to set more initialization hyperparameters to better fine-tune the large language model for more accurate rating predictions. Additionally, we will investigate how to fine-tune the large language model for application in more vertical domains.

ACKNOWLEDGMENT

This research was funded by the National Key Research and Development Program of China (Research and Demonstration Application of Key Technologies for Personalized Learning Driven by Educational Big Data) under Grant 2023YFC3341200, the National Natural Science Foundation of China under Grant 62377015, the Collaborative Innovation Center for Intelligent Educational Technology of Guangzhou under Grant 2023B04J0002, Tertiary Education Scientific research project of Guangzhou Municipal Education Bureau 2024312300, the National Natural Science Foundation of China under Grant 62407016, and the Research Cultivation Fund for The Youth Teachers of South China Normal University under Grant 23KJ29.

REFERENCES

[1] Bahtar A Z, Muda M. The impact of User–Generated Content (UGC) on product reviews towards online purchasing – A conceptual framework[J]. Procedia Economics and Finance, 2016, 37: 337-342.

[2] Cai Y, Ke W, Cui E, et al. A deep recommendation model of cross-grained sentiments of user reviews and ratings[J]. Information Processing & Management, 2022, 59(2): 102842.

[3] Lei X, Qian X, Zhao G. Rating prediction based on social sentiment from textual reviews[J]. IEEE transactions on multimedia, 2016, 18(9): 1910-1921.

[4] Cheng Z, Ding Y, Zhu L, et al. Aspect-aware latent factor model: Rating prediction with ratings and reviews[C]// Proceedings of the 2018 World Wide Web Conference on World Wide Web, Lyon, Apr 23-27, 2018. France: ACM, 2018: 639-648.

[5] Sadiq S, Umer M, Ullah S, et al. Discrepancy detection between actual user reviews and numeric ratings of Google App store using deep learning[J]. Expert Systems with Applications, 2021, 181: 115111.

[6] Devlin J , Chang M W , Lee K , et al. BERT: Pre-training of Deep Bidirectional Transformers for Language Under-standing[C] //Proceedings of the 2019 Conference of the North American Chapter of the Association for Computa-tional Linguistics: Human Language Technologies, Min-neapolis, Jun 2-7, 2019. USA: Association for Computa-tional Linguistics, 2009: 4171-4186.

[7] Ekgren A, Gyllensten A C, Gogoulou E, et al. Lessons Learned from GPT-SW3: Building the First Large-Scale Generative Language Model for Swedish[C]//Proceedings of the 13th Language Resources and Evaluation Confer-ence, Marseille, Jun 20-25, 2022. France: European Language Resources Association, 2022: 3509-3518.

[8] Rives A, Meier J, Sercu T, et al. Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences[J]. Proceedings of the National Academy of Sciences of the United States of America, 2021, 118(15):e2016239118.

[9] Liu Y, Ott M, Goyal N, et al. RoBERTa: A Robustly Optimized BERT Pretraining Approach[J]. arXiv preprint arXiv.1907.11692, 2019.

[10] Yang Z , Dai Z , Yang Y ,et al.XLNet: Generalized Autoregressive Pretraining for Language Understanding[C] // Proceedings of the 33rd International Conference on Neural Information Processing Systems, Vancouver, Dec 8-14, 2019. Canada: ACM, 2019: 5753–5763.

[11] Khan Z Y, Niu Z, Sandiwarno S, et al. Deep learning techniques for rating prediction: a survey of the state-of-the-art[J]. Artificial Intelligence Review, 2021, 54: 95-135.

[12] Zhao G, Qian X, Xie X. User-service rating prediction by exploring social users' rating behaviors[J]. IEEE transactions on multimedia, 2016, 18(3): 496-506.

[13] Lai C H, Hsu C Y. Rating prediction based on combination of review mining and user preference analysis[J]. Information Systems, 2021, 99: 101742.

[14] Ma X, Lei X, Zhao G, et al. Rating prediction by exploring user′s preference and sentiment[J]. Multimedia Tools and Applications, 2018, 77: 6425-6444.

[15] Shi W, Wang L, Qin J. Extracting user influence from ratings and trust for rating prediction in recommendations[J]. Scientific reports, 2020, 10(1): 13592.

[16] Purkaystha B, Datta T, Islam M S. Rating prediction for recommendation: Constructing user profiles and item characteristics using backpropagation[J]. Applied Soft Computing, 2019, 75: 310-322.

[17] Ding N, Qin Y, Yang G, et al. Parameter-efficient fine-tuning of large-scale pre-trained language models[J]. Nature Machine Intelligence, 2023, 5(3): 220-235.

[18] Tinn R, Cheng H, Gu Y, et al. Fine-tuning large neural language models for biomedical natural language processing[J]. Patterns, 2023, 4(4): 100729.

[19] Hilmkil A, Callh S, Barbieri M, et al. Scaling federated learning for fine-tuning of large language models[C]// Proceedings of International Conference on Applications of Natural Language to Information Systems, Saarbrucken, Jun 23-25, 2021. Germany: Springer, 2021: 15-23.

[20] Bakker M, Chadwick M, Sheahan H, et al. Fine-tuning language models to find agreement among humans with diverse preferences[J]. Advances in Neural Information Processing Systems, 2022, 35: 38176-38189.

[21] Luo L, Ning J, Zhao Y, et al. Taiyi: a bilingual fine-tuned large language model for diverse biomedical tasks[J]. Journal of the American Medical Informatics Association, 2024: ocae037.

[22] Niu Z, Zhong G, Yu H. A review on the attention mechanism of deep learning[J]. Neurocomputing, 2021, 452: 48-62.

[23] Tokuoka Y, Yamada T G, Mashiko D, et al. An explainable deep learning-based algorithm with an attention mechanism for predicting the live birth potential of mouse embryos[J]. Artificial Intelligence in Medicine, 2022, 134: 102432.

[24] Tal O, Liu Y, Huang J, et al. Neural attention frameworks for explainable recommendation[J]. IEEE Transactions on Knowledge and Data Engineering, 2019, 33(5): 2137-2150.

[25] Gu R, Wang G, Song T, et al. CA-Net: Comprehensive attention convolutional neural networks for explainable medical image segmentation[J]. IEEE transactions on medical imaging, 2020, 40(2): 699-711.

[26] Andresini G, Appice A, Caforio F P, et al. ROULETTE: A neural attention multi-output model for explainable network intrusion detection[J]. Expert Systems with Applications, 2022, 201: 117144.

[27] Liu X, Pan Z, Yang H, et al. An Adaptive Moment estimation method for online AUC maximization[J]. PloS one, 2019, 14(4): e0215426.

[28] Chen H, Zheng L, Al Kontar R, et al. Stochastic gradient descent in correlated settings: A study on gaussian processes[J]. Advances in neural information processing systems, 2020, 33: 2722-2733.

[29] McAuley J J, Leskovec J. From amateurs to connoisseurs: modeling the evolution of user expertise through online reviews[C]//Proceedings of the 22nd international conference on World Wide Web, Rio de Janeiro, May 13-17, 2013. Brazil: ACM, 2013: 897-908.

[30] Raffel C, Shazeer N, Roberts A, et al. Exploring the limits of transfer learning with a unified text-to-text transformer[J]. Journal of machine learning research, 2020, 21(140): 1-67.