https://doi.org/10.62836/iaet.v3i1.213

Global Science Publishing

IAET

Innovations in Applied Engineering and Technology

https://ojs.sgsci.org/journals/iaet

Article

The Research on Intelligent News Advertisement Recommendation Algorithm Based on Prompt Learning in End-to-End Large Language Model Architecture

Yunxiang Gan [1] and Diwei Zhu [2,*]

1 Moloco, CA, USA; yg281@scarletmail.rutgers.edu

2 New York University, New York City, USA

Abstract: With the explosive growth of information on the internet, users are increasingly facing the problem of information overload, making precise news and ad recommendations an important area of research. While traditional recommendation algorithms can meet user needs to some extent, they still have limitations in dealing with complex and changing user behaviors and dynamic content environments. This paper addresses the shortcomings of existing news and ad recommendation systems by proposing an intelligent recommendation algorithm based on an end-to-end large language model architecture. Firstly, we utilize the BERT model as the foundation, leveraging its powerful text representation capabilities to achieve deep semantic understanding of news and ad content, thereby capturing more detailed content features. Secondly, we apply prompt learning to fine-tune the BERT model, designing specific prompts for the model to better understand the implicit needs and preferences of users. Finally, we integrate these steps into an end-to-end architecture, enabling the model to achieve automated learning and optimization throughout the entire process from input to output, thus improving the precision and efficiency of recommendations. Experimental results demonstrate that the proposed method significantly outperforms traditional methods in the task of news and ad recommendation, not only enhancing the accuracy and relevance of recommendations but also effectively improving the model's interpretability and flexibility. This research explores new possibilities for the application of large language models in recommendation systems.

Keywords: Intelligent Recommendation Algorithm; BERT model; Prompt Learning

1. Introduction

With the rapid development of internet technology, the global volume of data has been experiencing explosive growth, resulting in users facing an increasing amount of information daily. In this environment of information overload, providing users with precise and personalized content recommendations has become a key focus for both academia and industry [1]. As an effective tool to address information overload, recommendation systems have been widely applied in fields such as e-commerce, social media, and news portals [2]. In recent years, deep learning models have achieved significant progress across various fields [3–5]. Machine learning-based fault prediction and diagnostics have been integrated into recommendation systems to enhance stability and accuracy, particularly in handling data anomalies and detecting user behavior biases [6, 7]. However,

Received: 1 August 2024; Accepted: 21 August 2024.

* Corresponding: Diwei Zhu (dz1397@nyu.edu)

traditional recommendation systems, which mainly rely on collaborative filtering and content-based algorithms, although achieving some early success, often show significant shortcomings when faced with the complex and variable needs of users and the diversification of content. These shortcomings include issues such as the cold start problem, data sparsity, and recommendation content homogeneity [8].

To address these issues, researchers have proposed various improved algorithms and strategies in the fields of image classification, news classification, and semi-supervised learning [9, 10]. For instance, matrix factorization-based recommendation methods, which uncover latent preference patterns by decomposing the user-item rating matrix, have made some progress. Deep Neural Networks (DNN) [11] and Recurrent Neural Networks (RNN) [12] have brought new possibilities to recommendation systems, particularly excelling in capturing nonlinear relationships and temporal sequence patterns. Graph Neural Networks (GNN) [13], by constructing graph structures between users and content, can better capture complex interaction patterns. Additionally, multi-model integration strategies have been widely applied in malware detection, enhancing system robustness through machine learning algorithms and offering improved paths for optimizing recommendation systems [14, 15]. Attention mechanism-based recommendation models, such as Transformer [16], have greatly enhanced the model's ability to understand user behavior sequences. Although these recommendation systems can provide users with personalized news and ad recommendations to some extent, they still face numerous challenges in practical applications due to the diversity of user needs and the complexity of content forms. Firstly, some recommendation algorithms show significant limitations in dealing with issues like cold start, data sparsity, and content homogeneity, resulting in recommendation outcomes that cannot fully meet users' personalized needs [17, 18]. Secondly, existing models often struggle to capture deep semantic information when processing complex natural language texts, making it difficult to improve the accuracy and relevance of recommendations. Moreover, many recommendation systems lack a deep understanding of user intent, leading to lower relevance of the recommended content and poor user experience [19]. In response, some studies have applied attention mechanisms combining DCGAN and autoencoders, significantly enhancing model classification and system robustness, offering new insights for optimizing recommendation systems [20,21].

With the rapid advancement of Natural Language Processing (NLP) technology, the emergence of pre-trained models like BERT [22] offers new possibilities for enhancing recommendation system performance. In this context, large-scale dataset analysis of economic stability has expanded the application of machine learning algorithms and provided new data insights for user segmentation in recommendation systems [23, 24]. The BERT model, through its bidirectional encoder mechanism, can deeply understand textual semantics and capture complex relationships between contexts, providing strong support for content understanding and user profile construction in recommendation systems. However, relying solely on the BERT model for feature extraction still struggles to fully address the diversity of user needs and the personalization of recommended content. On this basis, Prompt Learning [25,26], as an emerging learning paradigm, has shown great potential. By introducing appropriate prompts in the input, Prompt Learning can guide large language models to generate outputs more aligned with user expectations, thereby enhancing the model's ability to understand user intent. This method offers a novel optimization pathway for recommendation systems, particularly in responding to dynamically changing user needs and content environments, significantly improving the relevance and accuracy of recommendation results. To fully leverage these advanced technologies, this study aims to construct an end-to-end intelligent news and ad recommendation model based on Prompt Learning within a large language model framework. Through this approach, we expect not only to solve many issues of traditional recommendation systems but also to enhance the model's capabilities in handling complex natural language texts and understanding user intent, thereby providing users with more accurate and personalized recommendation services.

The structure of this paper is as follows: The first section introduces the background, motivation, objectives, overall structure of the paper, and main contributions, clearly outlining the research direction and goals, laying the foundation for the subsequent content. The second section reviews the major research progress in the field of recommendation systems, analyzing the current application status of these technologies in news and ad

recommendation and their limitations. The third section provides a detailed description of the design and implementation of the intelligent news and ad recommendation algorithm proposed in this paper. First, it introduces the application of the BERT model in text representation, then discusses the introduction of Prompt Learning and its specific implementation in the recommendation system, and finally describes the integration and optimization process of the end-to-end architecture. The fourth section introduces the experimental environment, dataset selection, and evaluation metrics, and presents the experimental results. Through comparative analysis, it verifies the accuracy of the proposed algorithm in recommendations and discusses the experimental results in detail. The fifth section summarizes the main contributions and outcomes of the research, reviews the effectiveness of the proposed method, and suggests further optimization and expansion possibilities in the news and ad recommendation system.

The main contributions of this paper include:

1. Introducing the BERT model into news and ad recommendation systems, fully utilizing its bidirectional encoder mechanism to deeply understand the semantics of news and ad texts. BERT can capture the complex semantic relationships between contexts, thereby generating more accurate text feature representations. This allows better adaptation to the diverse content needs of users, overcoming the limitations of traditional recommendation systems in dealing with complex texts.

2. Introducing the Prompt Learning method into the recommendation system. By designing specific prompts in the model input, the model is guided to generate outputs that better meet user expectations. This not only enhances the model's ability to understand users' implicit needs and preferences but also allows for dynamic adaptation to changes in user behavior, improving the relevance and personalization of recommendation results. By integrating accelerated attention mechanisms and implicit contrastive learning, the system's performance in handling heterogeneous data is further enhanced, leading to more accurate outcomes across diverse user groups [27,28].

3. Integrating BERT's text representation, Prompt Learning's fine-tuning, and the overall process of the recommendation system. Through this end-to-end architecture design, the model can achieve automated learning and optimization throughout the entire process from data input to recommendation output, avoiding the error accumulation issues that may arise in traditional step-by-step methods, thereby improving system efficiency and performance. Additionally, the model's design incorporates successful experiences from deep neural networks in image recommendations, particularly in social networks, further improving the system's recommendation performance [29,30].

2. Related Work

As a core technology for addressing the problem of information overload, recommendation systems have been widely adopted in various internet applications and have undergone rapid evolution from simple rule-based algorithms to complex deep learning models. The development of recommendation system technology has gone through several major stages, starting from early methods based on collaborative filtering and content recommendation, gradually evolving to more complex algorithms based on matrix factorization and deep learning [31,32]. Although traditional recommendation methods achieved significant success in the early stages, they often fall short when faced with the diversity of user needs and the complexity of content. In recent years, the rise of deep learning technologies, including Neural Collaborative Filtering (NCF), Graph Neural Networks (GNN), and attention mechanism-based models [33], has greatly advanced the progress of recommendation systems. Moreover, by integrating heterogeneous information streams from various fields and incorporating a time-adaptive sentiment tracking mechanism, the recommendation model's adaptability to complex content environments is further enhanced [34,35]. However, these methods still face challenges in dealing with dynamic user needs and complex content environments.

In recommendation systems, the BERT model has become an important tool in the field of natural language processing. The BERT model, with its powerful semantic understanding capabilities, significantly improves the representation of textual content, enabling recommendation systems to better understand user interests and content features. Yi Yang et al [36]. proposed an enhanced Video BERT (EVB) method, which combines the

advantages of Tree-based Deep Model (TDM) to optimize the efficiency of Video BERT in large-scale video retrieval. EVB achieves practical application in video advertising platforms by integrating local features and dynamic structure optimization, significantly improving conversion rates (CVR) and click-through rates (CTR). Kezhi Lu et al. [37] proposed a neural personalized recommendation system called BERT-RS. This method uses BERT to extract semantic representations from textual reviews and user-item interactions and generates latent representations of users and items through three different deep architectures, ultimately used for personalized recommendation tasks. Ikram Karabila et al [38]. proposed a personalized e-commerce recommendation system enhanced by BERT-based sentiment analysis. This method is implemented in three steps: first, a fine-tuned BERT model is developed for accurate sentiment classification; second, a hybrid recommendation model based on collaborative filtering is created; finally, the sentiment analysis results from BERT are used to improve the product selection process. This method significantly enhances the accuracy and personalization of the recommendation system, performing excellently in the e-commerce domain. However, the application of the BERT model in recommendation systems also faces challenges such as high computational complexity and substantial training resource requirements, which somewhat limit its widespread application in recommendation systems. Prompt Learning, an emerging fine-tuning method, can guide large language models to better understand task requirements and generate expected results by designing reasonable prompts. Lei Li et al. [39] proposed two prompt learning-based methods to improve the interpretability of recommendation systems: discrete prompt learning and continuous prompt learning. Discrete prompt learning integrates user and item IDs into the pre-trained model through substitute words, while continuous prompt learning directly inputs ID vectors and proposes sequential fine-tuning and recommendation regularization training strategies. Results show that the continuous prompt learning method outperforms traditional baseline methods across multiple datasets. Ye Jiang et al. [40] proposed the SAMPLE framework for fake news detection. SAMPLE applies prompt learning to multimodal fake news detection, using three prompt templates and a soft verbalizer to identify fake news and introducing a similarity-aware fusion method to adaptively merge the strengths of multimodal representations, reducing noise caused by irrelevant cross-modal features. Tao Guo et al. [41] proposed the pFedPrompt method, which enhances personalization in federated learning by learning personalized prompt words in vision-language models. Leveraging the advantages of multimodality, this method learns user consensus in the language space and adapts to user characteristics in the visual space, achieving comprehensive personalization of prompt words. However, research on prompt learning in the field of ad recommendation is still in its early stages, and how to design effective prompts to optimize the performance of recommendation systems remains a topic worthy of in-depth research.

In this study, we combine the powerful semantic representation capabilities of the BERT model with the flexible fine-tuning characteristics of prompt learning to propose an end-to-end news and ad recommendation system. The end-to-end recommendation system architecture integrates the entire recommendation process into a unified model, achieving automated learning and global optimization, simplifying the design and implementation of the recommendation system. Compared to traditional modular recommendation systems, the end-to-end architecture can automatically extract features, optimize the model, and reduce information loss during data transmission. This enables better capture of user interests and needs, effectively matching suitable ad content, thereby significantly enhancing the effectiveness of the recommendation system. This architecture has been widely applied in fields such as e-commerce, video recommendation, and social media, utilizing deep neural networks, graph neural networks, and Transformer models to improve recommendation outcomes.

3. Method

Figure 1 shows the overall architecture of the end-to-end large language model recommendation algorithm based on prompt learning. The figure includes several key components: First, the original input is processed by the prompt encoder to generate fixed and adjustable tags. Then, these tags are processed together with the original input through the embedding layer to form a complete embedding vector. Next, the embedding vector is encoded through the Transformer encoder, the label word embedding is added, and finally the output prediction is generated. This process shows how the model combines prompt learning with the BERT architecture to

achieve an automated recommendation process from input to output.

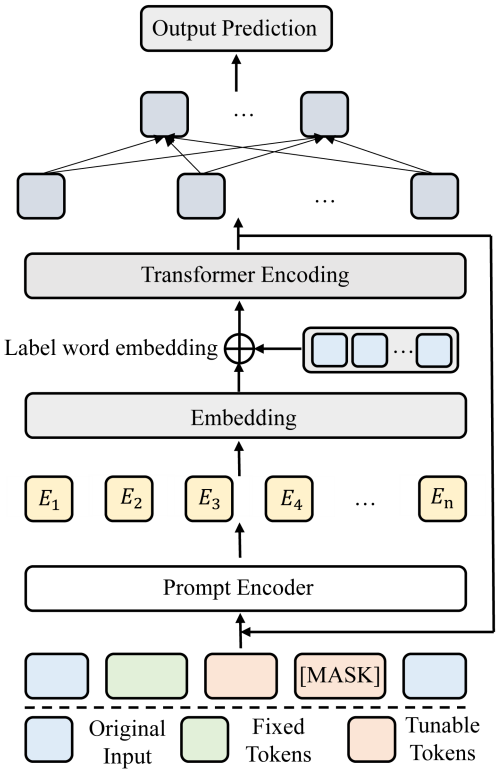

Figure 1. Overall algorithm architecture.

3.1. BERT Network Architecture

BERT considers the relationship between each word and other words in a sentence through a bidirectional attention mechanism, thereby capturing richer contextual information. This enables it to perform exceptionally well in various natural language processing tasks. BERT is based on the Transformer model, which consists of multiple self-attention layers and feed-forward neural network layers. The core components of the Transformer include the multi-head self-attention mechanism, residual connections, and layer normalization. The algorithm architecture diagram is shown in Figure 2.

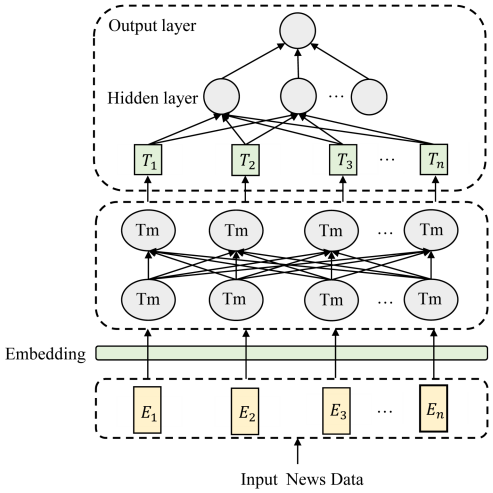

Figure 2. BERT network architecture diagram.

The input representation of BERT is composed of three parts: token embeddings, position embeddings, and segment embeddings. Token embeddings convert the input words into dense vector representations. For

example, if the input sentence is "[CLS] This is a book [SEP]", each word will be mapped to a vector of fixed dimensions. Position embeddings provide positional information for each word in the sentence through positional encoding. For the *i*th word in the input sequence, its position embedding can be represented as:

$$PE_{(i,2j)} = \sin\left(\frac{i}{10000^{2j/d_{\text{model}}}}\right) \tag{1}$$

$$PE_{(i,2j+1)} = \cos\left(\frac{i}{10000^{2j/d_{\text{model}}}}\right) \tag{2}$$

Where i is the word position, j is the dimension, and $d_{\{model\}}$ is the dimension of the embedding.

Paragraph embedding is used to distinguish different sentence fragments. In BERT, two sentence marker symbols [SEP] are usually used to represent different paragraphs. The final input representation is the vector sum of these three parts:

$$\text{Input Embedding} = \text{Token Embedding} + \text{Position Embedding} + \text{Segment Embedding} \tag{3}$$

The self-attention mechanism is a key component of BERT. It captures contextual information by calculating the relevance of each word in the input sequence to all other words. For a given input sequence X, the self-attention mechanism is computed as follows:

$$\text{Attention}(Q,K,V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{4}$$

Here, Q, K, and V are the Query, Key, and Value matrices, respectively, obtained from the input X through different linear transformations. d_k is the dimension of the Key, used to scale the dot product result.

BERT is pre-trained using two main tasks: the Masked Language Model (MLM) randomly masks some words in the input sequence (replacing them with [MASK]) and then lets the model predict these masked words. For example, if the input sequence is "[CLS] The [MASK] is on the table [SEP]", the model needs to predict that the word corresponding to [MASK] is "book". The objective of the MLM is to maximize the prediction probability of the masked words:

$$L_{\text{MLM}} = \sum_{t \in \text{MASK}} logP(X_t \mid X_{mashed}) \tag{5}$$

Where, x_t represents the ttt-th masked word, X_{masked} is the input sequence after some words have been masked. The model predicts the masked words based on the remaining context.

The Next Sentence Prediction (NSP) task predicts whether two sentences are consecutive to capture relationships at the sentence level. Given a pair of sentences, the model outputs a binary classification result. The objective of the NSP is to maximize the correct next sentence prediction probability:

$$L_{\text{NSP}} = -\log P(\text{IsNext}|X) \tag{6}$$

Where, X is the input sentence pair, which is the concatenation of two sentences, IsNext is a binary label that represents the relationship between the two sentences. If the second sentence follows the first, the label is 1; otherwise, the label is 0.

3.2. Prompt Learning

Prompt Learning is a technique for applying pre-trained language models to various downstream tasks by designing specific prompts within the input text. By leveraging the language generation capabilities of pre-trained models, it can significantly improve model performance without the need for large amounts of labeled data. The core idea is to transform downstream tasks into something resembling a "fill-in-the-blank" or "complete-the-sentence" task, utilizing the pre-trained model's natural language understanding abilities to generate the desired output for the task. The prompt learning architecture diagram is shown in Figure 3.

Given an input text x , the traditional approach is to directly input this text into the model and obtain the output y. However, in Prompt Learning, we first convert x into an input $\tilde{x}$ that includes a prompt.

$$\tilde{x} = \text{template}(x) \tag{7}$$

Here, $\tilde{x}$ is a template function used to construct the prompt. This template function usually replaces some words or phrases in the input text with a "blank" position (e.g., [MASK]), or adds some prompt phrases before

or after the text. For example, for the sentence "This product is great," the template might be "This product is great, it is [MASK]," where [MASK] represents the word to be predicted by the model.

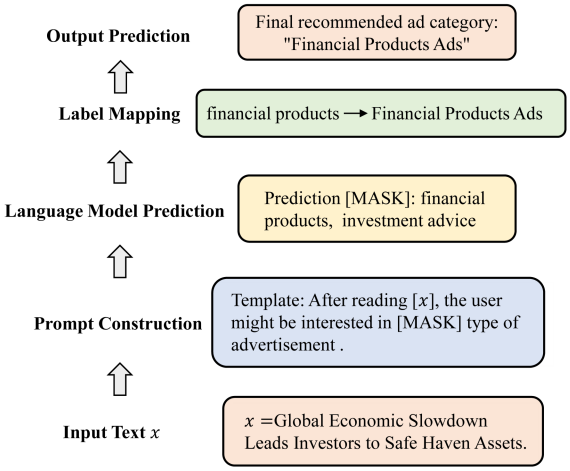

Figure 3. Prompt learning architecture diagram.

Next, the input $\tilde{x}$ with the prompt is fed into the pre-trained language model $f(\bullet)$, and the model predicts the word or phrase for the [MASK] position based on contextual information. This process is carried out by calculating the hidden state vector of the input.

$$P(y|\tilde{x})=f(\tilde{x}) \tag{8}$$

Here, $P(y \mid \tilde{x})$ represents the probability distribution of the output y given the input $\tilde{x}$ by the language model.

Once the model provides a prediction, the next step is to map the predicted result back to the label space of the target task. Let the mapping function be $g(\bullet)$, then the final output after mapping is:

$$y=g(\text{argmax}(P(y|\tilde{x}))) \tag{9}$$

Finally, the performance of the model can be further optimized by tuning the prompts. This includes adjusting the length of the prompt, choice of words, and the position of the prompt.

3.3. End-to-End Architecture

End-to-end architecture refers to a unified model that directly maps input data to the desired output result without requiring manually designed intermediate steps. Compared to traditional multi-stage processing pipelines, end-to-end architecture is simpler and more efficient, reducing the accumulation of errors in the intermediate processes and improving the overall performance of the system. In end-to-end architecture, the entire system can be seen as a complex function mapping. Given an input x and a target output y, the traditional segmented approach might break down this mapping process into multiple steps, such as $x \to h_1 \to h_2 \to y$, where h_1 and h_2 are intermediate features or outputs of subtasks. However, in end-to-end architecture, we directly generate the output y from the input x through a function $f(\bullet)$:

$$y=f(x;\theta) \tag{10}$$

Where θ represents the parameters of the model, which are learned and optimized through training data.

Let's assume the input x is a complex data structure; for instance, in natural language processing tasks, the input might be a sentence or a piece of text. First, we represent the input, usually by converting it into a dense vector representation x through an embedding method:

$$\boldsymbol{x}=\text{Embed}(x) \tag{11}$$

Here, the embedding function {Embed}(•) could be the output of a pre-trained model like BERT.

Next, the input is mapped to the output through a complex nonlinear function $f(\bullet;\theta)$. This function is typically implemented using a deep neural network, consisting of multiple layers of neurons and nonlinear activation functions. For a neural network with L layers, the mapping process can be expressed as:

$$\boldsymbol{h}^{(l)} = \sigma(W^{(l)}\boldsymbol{h}^{(l-1)} + b^{(l)}), \quad l = 1, 2, \ldots, L \tag{12}$$

Where, $h^{(0)} = x$ is the embedded representation of the input. $h^{(l)}$ is the output of the l-th layer. $W^{(l)}$ and b^{l} are the weight matrix and bias vector of the l-th layer, respectively. $\sigma(\bullet)$ is the activation function, such as ReLU or Sigmoid.

Finally, the network's output $h^{(L)}$ is mapped to the target space, generating the final output y:

$$y = \text{Output}(\mathbf{h}^{(L)}) \tag{13}$$

4. Experiment

4.1. Experimental Environment

In this experiment, we used high-performance hardware configurations, including an Intel Core i9-10900K processor and an NVIDIA GeForce RTX 3090 graphics card, combined with 64GB of memory and a 1TB NVMe SSD, to ensure the speed and efficiency of model training. In terms of the software environment, the experiment was conducted on the Ubuntu 20.04 operating system, with Python 3.8 chosen as the programming language. For the deep learning framework, we used PyTorch 1.9. To fully leverage the computational power of the RTX 3090, we utilized CUDA 11.1 and cuDNN 8.0, ensuring efficient GPU acceleration. Data processing and development were primarily carried out using Pandas, NumPy, and Jupyter Notebook, while Git and GitHub were used for version control and collaboration.

4.2. Experimental Data

● OpenNewsArchive Dataset

OpenNewsArchive [42] is a high-quality news dataset that compiles 8.8 million news reports from mainstream media, covering multiple domains such as finance, health, military, and more. The dataset has undergone rigorous cleaning and deduplication to ensure purity and diversity, primarily consisting of news published in 2023. OpenNewsArchive is suitable for tasks like news recommendation, natural language processing, text classification, and sentiment analysis, making it an essential resource for developing and optimizing large language models. The dataset is available under the open CC BY 4.0 license, supporting non-commercial research and development applications.

● MIND Dataset

MIND Dataset [43] is a large-scale news recommendation dataset released by Microsoft, designed specifically for researching personalized news recommendation systems. The dataset includes click behavior data from millions of users and detailed news content, including titles, summaries, categories, and entities. The MIND dataset is available in two versions, MIND-small and MIND-large, to cater to different research needs. The news content in the dataset is sourced from the Microsoft News platform, covering multiple domains such as politics, technology, entertainment, and more, making it suitable for various news recommendation and natural language processing tasks.

● AG News Corpus Dataset

AG News Corpus [44] is a widely-used news text dataset containing approximately 120,000 news articles, divided into four main categories: World News, Sports, Business, and Science & Technology. The news articles under each category have been preprocessed, making them suitable for text classification, natural language processing (NLP) tasks, topic modeling, and more. Due to its clear structure and diverse content, AG News Corpus is often used to train and evaluate machine learning and deep learning models, particularly in tasks like news classification and information retrieval.

● NewsQA Dataset

NewsQA Dataset [45] is specifically designed for research in machine reading comprehension and question-answering systems. It contains over 100, 000 news articles from CNN, along with corresponding question-answer pairs. What makes this dataset unique is that the questions are generated by human annotators rather than being directly extracted from the articles, resulting in a more complex relationship between the questions and answers. This makes NewsQA an important benchmark for evaluating machine reading comprehension abilities,

especially in tasks that require reasoning and semantic understanding.

4.3. Evaluation Metrics

● Accuracy

Accuracy measures the overall correctness of the model's predictions, indicating the proportion of correct predictions out of all predictions made.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{14}$$

where TP is True Positive, TN is True Negative, FP is False Positive, and FN is False Negative. In a news advertisement recommendation system, accuracy indicates how many of the recommended advertisements are correct, including correctly recommended ads and correctly not recommending wrong ads.

● Precision

Precision measures the proportion of true positives among all the predictions classified as positive.

$$\text{Precision} = \frac{TP}{TP + FP} \tag{15}$$

Precision indicates how many of the recommended advertisements are actually of interest to the user. It is a measure of the "precision" of the recommendation system.

● Recall

Recall measures the proportion of true positives that were correctly identified out of all actual positive samples.

$$\text{Recall} = \frac{TP}{TP + FN} \tag{16}$$

Recall indicates the proportion of advertisements that the system successfully recommended to the user out of all ads that the user is actually interested in. It measures the "coverage" of the system.

● NDCG

NDCG is a metric that measures the quality of the ranking in the recommendation results, taking into account both the relevance of the recommended content and its position in the list. The higher the NDCG, the better the recommendation system is at not only recommending relevant ads but also ranking them according to the user's interests.

$$\text{NDCG} = \frac{DCG}{IDCG} \tag{17}$$

Where, $\text{DCG} = \sum_{i=1}^{p} \frac{2^{rel_i} - 1}{\log_2(i+1)}$, i represents the position of the item in the ranked list,rel_i is the relevance score of the item at position I, p is the number of positions considered in the list., IDCG = DCG for the ideal ranking. NDCG measures the overall quality of the recommendation list, considering both the position of the advertisements in the list and their relevance to the user.

4.4. Task Description

In this experiment, our goal is to predict the ad categories that users are most likely to engage with. Table 1 shows the task description and data description of this article. By analyzing the user's historical ad interaction data (including ad categories, interaction time, engagement, etc.) and recent news reading behavior, we can identify the user's interest trends and engagement patterns. The experiment will combine time information and the user's interaction behavior in different time periods and different dates to infer the ad categories that users are most likely to be interested in at a given point in time. Ultimately, the model will output a set of ad category prediction results sorted by probability to help optimize the accuracy and user experience of the ad recommendation system.

Table 1. Task description and data description.

Section	Details
Specify the task	Your task is to predict the most relevant categories of advertisements (e.g., finance, sports, technology) for a user based on their news reading habits and content consumption history.
Describe the data	You will be provided with `<history>` which is a list containing this user's historical interactions with different categories of advertisements (e.g., finance, sports, health), then `<context>` which provides contextual information about the user's recent activities, including the news articles they have read. Interactions in both `<history>` and `<context>` are in chronological order. Each interaction takes on such form as (interaction_time, day_of_week, duration, ad_category, ad_id, engagement_level). The detailed explanation of each element is as follows: interaction_time: the time of the interaction in 12h clock format. day_of_week: indicating the day of the week. duration: an integer indicating the duration (in minutes) of each interaction. ad_category: a string representing the category of the advertisement (e.g., finance, sports, technology). ad_id: an integer representing the unique advertisement ID. engagement_level: a score representing the user's level of engagement with the advertisement (e.g., click, view, ignore). Then you need to predict the next likely category of advertisement that the user is most likely to engage with, denoted as `<next_ad_category>`. The prediction target will include an unknown advertisement category denoted as `<next_ad_category>` and an unknown engagement level denoted as None, while time information is provided.
Specify the number of output ad categories	Please infer what the `<next_ad_category>` might be (the {k} most likely ad categories which are ranked in descending order in terms of probability).
Guide the model to "think"	Please consider the following aspects: 1. The engagement pattern of this user that you learned from `<history>`, e.g., repeated interactions with certain types of ad categories during certain times; 2. The context interactions in `<context>`, which provide more recent activities of this user, particularly in terms of the types of news articles consumed; 3. The temporal information (i.e., interaction_time and day_of_week) of the target advertisement, which is important because people's ad engagement with different categories may vary during different times (e.g., nighttime versus daytime) and on different days (e.g., weekday versus weekend).
Format the output and ask for explanations	Please organize your answer in a JSON object containing following keys: "prediction" (the {k} most probable ad categories in descending order of probability) and "reason" (a concise explanation that supports your prediction). Do not include line breaks in your output.
Provide the data	The data are as follows: `<history>`: {historical_interactions} `<context>`: {context_interactions} `<target>`: {target_interaction}

4.5. Experimental Comparison and Analysis

Table 2. Comparison of relevant indicators of this method with other methods on OpenNewsArchive and MIND Dataset.

Model	OpenNewsArchive Dataset				MIND Dataset			
	Accuracy	Precision	Recall	NDCG	Accuracy	Precision	Recall	NDCG
Azizi et al. [46]	87.95	89.53	88.37	28.22	85.04	86.19	89.71	27.24
Vo et al. [47]	86.72	91.49	89.71	29.84	86.04	88.28	84.44	27.76

Cont.

Model	OpenNewsArchive Dataset				MIND Dataset			
	Accuracy	Precision	Recall	NDCG	Accuracy	Precision	Recall	NDCG
Suhartono et al. [48]	86.12	90.58	87.55	29.87	85.14	90.38	87.15	27.83
Wang et al. [49]	88.42	91.28	88.34	31.09	87.65	86.04	89.97	30.30
Dang et al. [50]	86.55	87.34	86.89	28.56	88.75	88.09	84.63	28.71
Li et al. [51]	87.45	89.26	89.93	32.85	89.60	90.85	87.27	30.69
Ours	91.76	92.57	90.48	34.85	90.42	92.74	91.54	32.24

Table 2 presents a comparison of our proposed method with several other methods on the OpenNewsArchive and MIND datasets. The comparison shows that on the OpenNewsArchive dataset, our method achieved an accuracy of 91.76%, a precision of 92.57%, a recall of 90.48%, and an NDCG of 34.85%. This is significantly better than the other methods; for example, the method by Azizi et al. achieved only 28.22% NDCG, while ours reached 34.85%. Notably, our method outperformed the next best method (Wang et al.) by approximately 3.34% in accuracy and 1.29% in precision. On the MIND dataset, our method achieved an accuracy of 90.42%, a precision of 92.74%, a recall of 91.54%, and an NDCG of 32.24%. Compared to the best results achieved by Li et al. (with a precision of 90.85% and a recall of 87.27%), our method shows significant improvements, particularly in recall, where we outperform them by 4.27%. Additionally, Figure 4 provides a visual comparison, making it easier to observe the differences between the models.

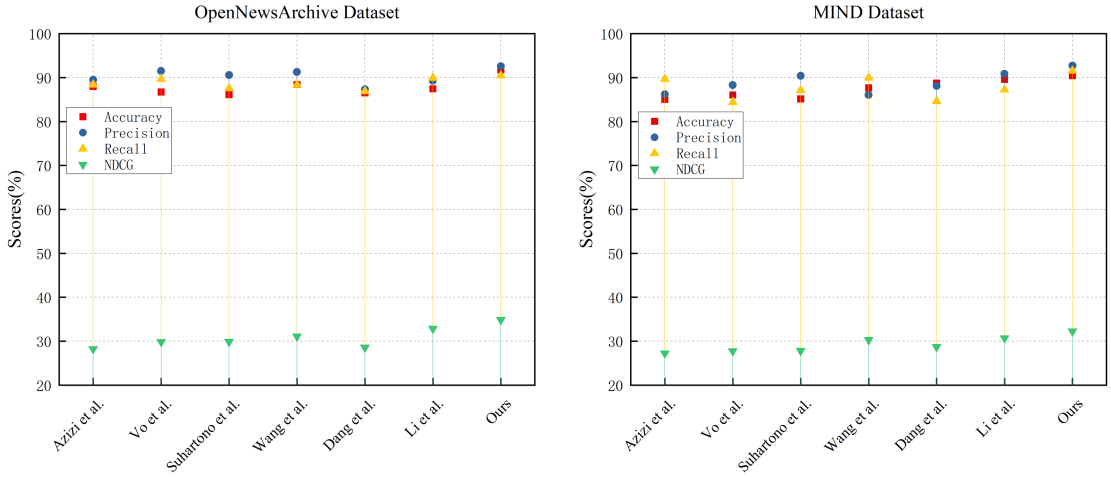

Figure 4. Visual comparison of relevant indicators on OpenNewsArchive and MIND Dataset.

Table 3. Comparison of relevant indicators of this method with other methods on AG News Corpus and NewsQA Dataset.

Model	AG News Corpus Dataset				NewsQA Dataset			
	Accuracy	Precision	Recall	NDCG	Accuracy	Precision	Recall	NDCG
Azizi et al. [46]	89.36	90.82	86.52	30.82	89.77	90.88	90.13	29.43
Vo et al. [47]	89.28	89.53	90.54	31.68	89.95	89.99	86.24	29.48
Suhartono et al. [48]	87.53	90.75	88.66	28.48	89.89	90.36	90.23	29.20
Wang et al. [49]	88.35	90.39	86.41	30.90	87.57	87.34	87.17	28.83
Dang et al. [50]	87.77	89.29	86.26	30.54	90.54	90.59	91.13	30.14
Li et al. [51]	90.21	89.79	88.42	31.45	90.31	89.13	91.53	28.75

Cont.

Model	AG News Corpus Dataset				NewsQA Dataset			
	Accuracy	Precision	Recall	NDCG	Accuracy	Precision	Recall	NDCG
Ours	91.53	92.83	91.24	33.72	93.47	91.82	93.04	34.17

Table 3 presents a comparison of our proposed method with several other methods on the AG News Corpus and NewsQA datasets. On the AG News Corpus dataset, our results also outperform all other methods listed in the table, especially in precision and recall, where our method exceeds the next best method (Li et al.) by approximately 3.04% and 2.82%, respectively. Additionally, our method shows a significant advantage in NDCG, leading other methods by at least 2.27%. On the NewsQA dataset, our method also performs well, particularly in recall, where it surpasses the best result from Dang et al. by about 1.91%. Notably, our NDCG score of 34.17% is significantly higher than other methods, indicating a substantial advantage in the ranking quality of the recommended results. Similarly, Figure 5 provides a visual comparison, making it easier to observe the differences between the models.

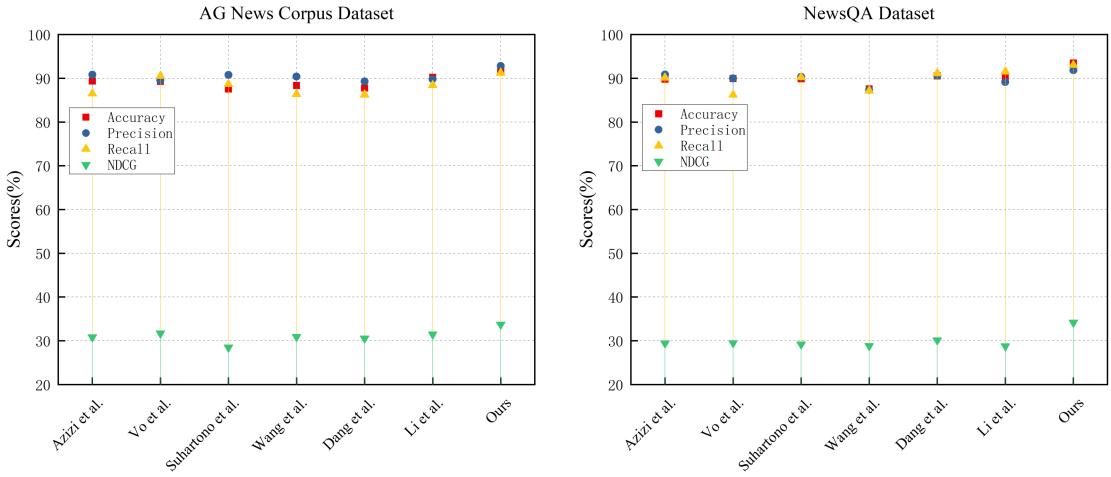

Figure 5. Visual comparison of relevant indicators on AG News Corpus and NewsQA Dataset.

Table 4. Comparison of training indicators on four datasets.

Model	OpenNewsArchive Dataset			MIND Dataset		
	Parameters(M)	Inference Time (ms)	Trainning Time(s)	Parameters(M)	Inference Time (ms)	Trainning Time(s)
Azizi et al. [46]	362.80	349.71	261.50	350.94	360.68	289.66
Vo et al. [47]	395.80	365.33	203.55	371.97	335.69	207.41
Suhartono et al. [48]	394.41	380.07	215.37	391.41	331.02	226.95
Wang et al. [49]	360.76	378.51	261.34	389.54	366.63	284.64
Dang et al. [50]	398.91	378.13	210.18	363.67	330.75	292.27
Li et al. [51]	374.20	333.04	215.42	363.60	385.37	267.50
Ours	342.41	312.74	168.52	334.37	319.62	186.04
Model	AG News Corpus Dataset			NewsQA Dataset		
	Parameters(M)	Inference Time (ms)	Trainning Time(s)	Parameters(M)	Inference Time (ms)	Trainning Time(s)
Azizi et al. [46]	381.44	366.30	264.79	358.43	360.10	248.35
Vo et al. [47]	351.57	372.17	235.17	356.27	297.00	237.84

Cont.

Model	OpenNewsArchive Dataset			MIND Dataset		
	Parameters(M)	Inference Time (ms)	Trainning Time(s)	Parameters(M)	Inference Time (ms)	Trainning Time(s)
Suhartono et al. [48]	393.46	348.52	206.94	360.79	323.56	230.34
Wang et al. [49]	380.44	345.92	234.95	383.07	339.19	292.53
Dang et al. [50]	372.88	342.13	287.47	352.67	356.20	272.88
Li et al. [51]	397.33	364.52	202.40	367.17	313.45	297.28
Ours	337.57	322.59	181.95	340.74	278.35	212.06

In Table 4, we compared the training indicators of our method with other models across four datasets, focusing on parameters, inference time, and training time. Compared to other methods, our method demonstrates significant advantages in several key metrics. Firstly, in terms of parameter count, our method has the lowest across all datasets. For example, in the AG News Corpus dataset, our method's parameter count is 337.57M, whereas the model with the highest parameter count (Li et al.) reaches 397.33M. Fewer parameters mean our model is more lightweight, offering better scalability and resource efficiency. Secondly, in terms of inference time and training time, our method also performs exceptionally well. On the NewsQA dataset, our method's inference time is 278.35ms, significantly outperforming the longest inference time of 372.17ms by Vo et al. Additionally, our training time is 212.06s, which is substantially lower than other methods; notably, compared to the longest training time by Dang et al., our method saves nearly 60 seconds. Compared to other methods, our approach not only significantly reduces the computational resources required but also accelerates the inference and training speeds of the model, providing a more efficient and cost-effective solution for practical applications. Figure 6 presents a visual comparison of these training indicators.

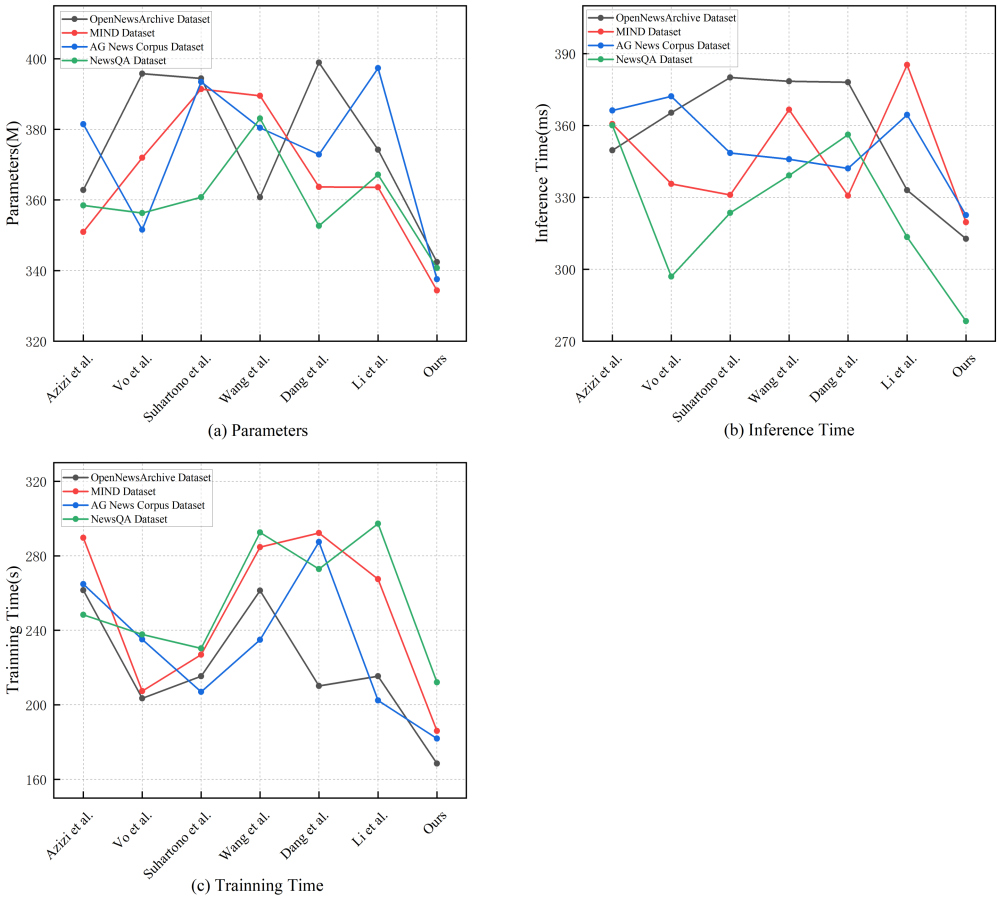

Figure 6. Visual comparison of training indicators.

Table 5. Ablation experiments on OpenNewsArchive and MIND Dataset.

Model	OpenNewsArchive Dataset			MIND Dataset		
	Accuracy	Precision	Recall	Accuracy	Precision	Recall
baseline	76.35	75.04	73.21	74.51	77.42	76.32
+BERT	85.07	85.64	86.42	82.64	84.52	83.21
+Prompt	88.16	89.32	88.39	86.75	87.21	87.47
+BERT-Prompt	91.76	92.57	90.48	90.42	92.74	91.54

Table 6. Ablation experiments on AG News Corpus and NewsQA Dataset.

Model	AG News Corpus Dataset			NewsQA Dataset		
	Accuracy	Precision	Recall	Accuracy	Precision	Recall
baseline	77.26	75.42	76.33	78.47	76.46	77.26
+BERT	83.73	84.34	82.61	84.28	85.16	86.43
+Prompt	86.21	88.16	87.24	88.76	87.06	89.24
+BERT-Prompt	91.53	92.83	91.24	93.47	91.82	93.04

In Tables 5 and 6, we conducted ablation experiments to assess the impact of BERT and Prompt Learning on model performance. In the OpenNewsArchive dataset, the baseline model's performance was relatively weak, with an accuracy of 76.35%, precision of 75.04%, and recall of 73.21%. When we introduced the BERT model, all metrics improved significantly, with accuracy rising to 85.07%. This indicates that BERT plays a crucial role in enhancing text representation and understanding. Further adding Prompt Learning improved the model's performance even more, with accuracy reaching 88.16%. Ultimately, when BERT and Prompt Learning were combined, the model achieved optimal performance, with an accuracy of 91.76%, precision of 92.57%, and recall of 90.48%. This demonstrates that the combination of BERT and Prompt Learning better captures user needs and content features, resulting in more accurate recommendations. A similar trend was observed in the MIND dataset. The baseline model's performance was basic, but after introducing BERT and Prompt Learning, the model's performance improved significantly, with the BERT-Prompt combination ultimately achieving 90.42% accuracy and 92.74% precision, showing outstanding performance. In the AG News Corpus and NewsQA datasets, we observed similar patterns. Notably, in the NewsQA dataset, the combination of BERT and Prompt Learning led the model to achieve an accuracy of 93.47% and a recall of 93.04%, the highest among all experiments. Additionally, Figures 7 and 8 provide a visual comparison of the ablation experiments.

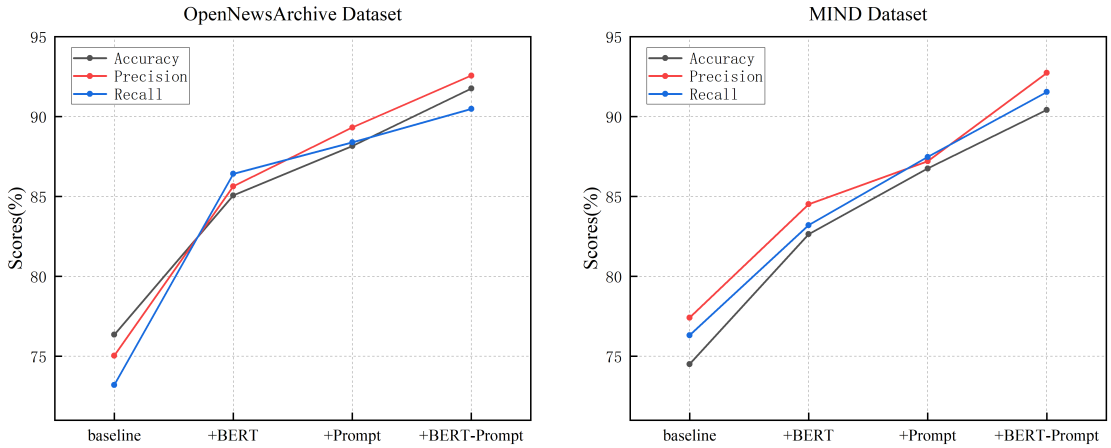

Figure 7. Visual comparison of ablation experiments on OpenNewsArchive and MIND Dataset.

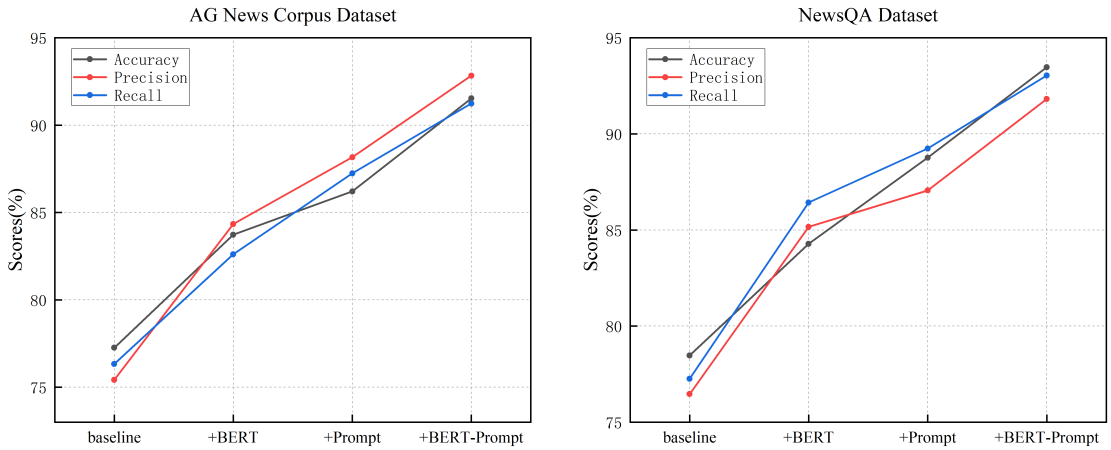

Figure 8. Visual comparison of ablation experiments on AG News Corpus and NewsQA Dataset.

5. Conclusion

In this paper, we propose an intelligent recommendation algorithm that combines BERT and Prompt Learning to address the complex challenges faced by news and advertisement recommendation systems. We introduce the BERT model, leveraging its powerful text representation and deep semantic understanding capabilities to significantly enhance the system's ability to handle complex textual content. By incorporating Prompt Learning, we design specific prompts to guide the model in better understanding users' implicit needs and preferences, thereby improving the personalization and accuracy of recommendations. Finally, we construct an end-to-end recommendation system architecture, which significantly improves the model's performance across multiple datasets. Through ablation experiments, we validated the effectiveness of this approach on several datasets. The experimental results show that the BERT model significantly enhances the system's semantic understanding of text, while Prompt Learning further improves the model's ability to identify users' implicit needs and the accuracy of recommendation results. When these two techniques are combined, the model achieves optimal performance in terms of accuracy, precision, and recall, particularly achieving the highest accuracy (93.47%) and recall (93.04%) on the NewsQA dataset. This demonstrates that the synergy between BERT and Prompt Learning can effectively handle different types of datasets and tasks, making the recommendation system more adaptable and precise in processing complex texts and user behavior patterns. In the future, we may consider further validating the applicability of this method on larger datasets and exploring how to integrate other advanced technologies to further enhance the model's performance. Additionally, given the diversity of practical application scenarios, maintaining efficiency and accuracy without significantly increasing computational costs will be an important direction for future research [52–65].

Funding

Not applicable.

Author Contributions

Conceptualization, Y. G. and D. Z.; writing—original draft preparation, Y. G. and D. Z.; writing—review and editing, Y.G. and D.Z.; All of the authors read and agreed to the published the final manuscript.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Data Availability Statement

Not applicable.

Conflicts of Interest

The authors declare no conflict of interest.

Reference

1 Wu C, Wu F, Huang Y, Xie X. Personalized News Recommendation: Methods and Challenges. *ACM Transactions on Information Systems* 2023; **41**(**1**): 1–50.

2 Chen J, *et al*. When Large Language Models Meet Personalization: Perspectives of Challenges and Opportunities. *World Wide Web* 2024; **27**(**4**): 42.

3 Wang M, Zhang H, Zhou N. Star Map Recognition and Matching Based on Deep Triangle Model. *Journal of Information, Technology and Policy* 2024; **1**(**1**): 1–18.

4 Ye X, Luo K, Wang H, Zhao Y, Zhang J, Liu A. An Advanced AI – Based Lightweight Two – Stage Underwater Structural Damage Detection Model. *Advanced Engineering Informatics* 2024; **62**: 102553.

5 Jihu L. Green Supply Chain Management Optimization Based on Chemical Industrial Clusters. *arXiv preprint* 2024; arXiv:2406.00478.

6 Chen X, Wang M, Zhang H. Machine Learning–based Fault Prediction and Diagnosis of Brushless Motors. *Engineering Advances* 2024; **4**(**3**): 130–142.

7 Wang X, Zhao Y, Wang Z, Hu N. An Ultrafast and Robust Structural Damage Identification Framework Enabled by an Optimized Extreme Learning Machine. *Mechanical Systems and Signal Processing* 2024; **216**: 111509.

8 Li X, Sun L, Ling M, Peng Y. A Survey of Graph Neural Network Based Recommendation in Social Networks. *Neurocomputing* 2023; **549**: 126441.

9 Li S, Kou P, Ma M, Yang H, Huang S, Yang Z. Application of Semi – Supervised Learning in Image Classification: Research on Fusion of Labeled and Unlabeled Data. *IEEE Access* 2024; **12**: 27331–27343.

10 Zhao F, Yu F. Enhancing Multi–Class News Classification through Bert–Augmented Prompt Engineering in Large Language Models: A Novel Approach.In Proceedings of the 10th International Scientific and Practical Conference "Problems and Prospects of Modern Science and Education", Stockholm, Sweden, 12 – 15 March 2024.

11 Wang H, *et al*. A Dnn–Based Cross–Domain Recommender System for Alleviating Cold–Start Problem in E –Commerce. *IEEE Open Journal of the Industrial Electronics Society* 2020; **1**: 194–206.

12 Jiang Z, Gao S. An Intelligent Recommendation Approach for Online Advertising Based on Hybrid Deep Neural Network and Parallel Computing. *Cluster Computing* 2020; **23**(**3**): 1987–2000.

13 Gao Y, Guo H, Lin D, Zhang Y, Tang R, He X. Content Filtering Enriched GNN Framework for News Recommendation. *arXiv preprint* 2021; arXiv:2110.12681.

14 Xiong S, Zhang H. A Multi – model Fusion Strategy for Android Malware Detection Based on Machine Learning Algorithms. *Journal of Computer Science Research* 2024; **6**(**2**): 1–11.

15 Xiong S, Chen X, Zhang H, Wang M. Domain Adaptation – Based Deep Learning Framework for Android Malware Detection Across Diverse Distributions. *Artificial Intelligence Advances* 2024; **6**(**1**): 13–24.

16 Yuan E, Guo W, He Z, Guo H, Liu C, Tang R. Multi – Behavior Sequential Transformer Recommender. In Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, Madrid, Spain, 11–15 July 2022.

17 Hao Y, Chen Z, Jin J, Sun X. Joint Operation Planning of Drivers and Trucks for Semi–Autonomous Truck Platooning. *Transportmetrica A: Transport Science* 2023; 1–37. DOI: 10.1080/23249935.2023.2266041.

18 Hao Y, Chen Z, Sun X, Tong L. Planning of Truck Platooning for Road – Network Capacitated Vehicle Routing Problem. *arXiv preprint* 2024; arXiv:2404.13512.

19 Ford J, Jain V, Wadhwani K, Gupta DG. AI Advertising: An Overview and Guidelines. *Journal of Business Research* 2023; **166**: 114124.

20 Xiong S, Zhang H, Wang M. Ensemble Model of Attention Mechanism–Based DCGAN and Autoencoder for Noised OCR Classification. *Journal of Electronic & Information Systems* 2022; **4**(**1**): 33–41.

21 Li L, Li Z, Guo F, Yang H, Wei J, Yang Z. Prototype Comparison Convolutional Networks for One–Shot Segmentation. *IEEE Access* 2024; **12**: 54978–54990.

22 Wang J, *et al*. Utilizing BERT for Information Retrieval: Survey, Applications, Resources, and Challenges. *ACM Computing Surveys* 2024; **56**(**7**): 1–33.

23 Qiu Y, Wang J. A Machine Learning Approach to Credit Card Customer Segmentation for Economic Stability. In Proceedings of the 4th International Conference on Economic Management and Big Data Applications, Tianjin, China, 27–29 October 2023.

24 Qiu Y. Financial Deepening and Economic Growth in Select Emerging Markets with Currency Board Systems: Theory and Evidence. *arXiv preprint* 2024; arXiv:2406.00472.

25 Xin X, Pimentel T, Karatzoglou A, Ren P, Christakopoulou K, Ren Z. Rethinking Reinforcement Learning for Recommendation: A Prompt Perspective. In Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, Madrid, Spain, 11–15 July 2022.

26 Ye M, Zhou H, Yang H, Hu B, Wang X. Multi–Strategy Improved Dung Beetle Optimization Algorithm and Its Applications. *Biomimetics* 2024; **9**(**5**): 291.

27 Chen Z, Fu C, Wu R, Wang Y, Tang X, Liang X. LGFat–RGCN: Faster Attention with Heterogeneous RGCN for Medical ICD Coding Generation. In Proceedings of the 31st ACM International Conference on Multimedia, Ottawa, ON, Canada, 29 October–3 November 2023.

28 Xu H, Shi C, Fan W, Chen Z. Improving Diversity and Discriminability Based Implicit Contrastive Learning for Unsupervised Domain Adaptation. *Applied Intelligence* 2024; **54**: 10007–10017.

29 Du S, Chen Z, Wu H, Tang Y, Li Y. Image Recommendation Algorithm Combined with Deep Neural Network Designed for Social Networks. *Complexity* 2021; **2021**(**1**): 5196190.

30 Chen Z, Fu C, Tang X. Multi–domain Fake News Detection with Fuzzy Labels. In Proceedings of the DASFAA 2023: The 28th International Conference on Database Systems for Advanced Applications, Tianjin, China, 17–20 April 2023.

31 Li Y, Liu K, Satapathy R, Wang S, Cambria E. Recent Developments in Recommender Systems: A Survey. *IEEE Computational Intelligence Magazine* 2024; **19**(**2**): 78–95.

32 Wang Z, Zhao Y, Song C, Wang X, Li Y. A New Interpretation on Structural Reliability Updating with Adaptive Batch Sampling–Based Subset Simulation. *Structural and Multidisciplinary Optimization*, 2024; **67**(**1**): 7.

33 Sharma K, et al. A Survey of Graph Neural Networks for Social Recommender Systems. *ACM Computing Surveys* 2024; **56**(**10**): 1–34.

34 Wang Y, Chen Z, Fu C. Synergy Masks of Domain Attribute Model DaBERT: Emotional Tracking on Time–Varying Virtual Space Communication. *Sensors* 2022; **22**(**21**): 450.

35 Li B, Ma Y, Liu Y, Gu H, Chen Z, Huang X. Federated Learning on Distributed Graphs Considering Multiple Heterogeneities. In Proceedings of the ICASSP 2024–2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Seoul, Korea, 14–19 April 2024.

36 Yang Y, *et al*. Enhanced Video BERT for Fast Video Advertisement Retrieval. In Proceedings of the 2022 IEEE International Conference on Big Data (Big Data), Osaka, Japan, 17–20 December 2022.

37 Lu K, Zhang Q, Zhang G, Lu J. BERT–RS: A Neural Personalized Recommender System with BERT. In Proceedings of the Machine Learning, Multi Agent and Cyber Physical Systems: Proceedings of the 15th International FLINS Conference (FLINS 2022), Tianjin, China, 26–28 August 2022.

38 Karabila I, Darraz N, EL–Ansari A, Alami N, Mallahi M EL. BERT–Enhanced Sentiment Analysis for Personalized E–Commerce Recommendations. *Multimedia Tools and Applications* 2024; **83**(**19**): 56463–56488.

39 Li L, Zhang Y, Chen L. Personalized Prompt Learning for Explainable Recommendation. *ACM Transactions on Information Systems* 2023; **41**(**4**): 1–26.

40 Jiang Y, Yu X, Wang Y, Xu X, Song X, Maynard D. Similarity–Aware Multimodal Prompt Learning for

Fake News Detection. *Information Sciences* 2023; **647**: 119446.

41 Guo T, Guo S, Wang J. Pfedprompt: Learning Personalized Prompt for Vision – Language Models in Federated Learning. In Proceedings of the ACM Web Conference 2023, Austin, TX, USA, 30 April–4 May 2023.

42 He C, Li W, Jin Z, Xu C, Wang B, Lin D. Opendatalab: Empowering General Artificial Intelligence with Open Datasets. *arXiv preprint* 2024; arXiv:2407.13773.

43 Wu F, et al. Mind: A Large–Scale Dataset for News Recommendation. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Online Meeting, 5–10July 2020.

44 Zhang X, Zhao J, LeCun Y. Character–Level Convolutional Networks for Text Classification. *Advances in Neural Information Processing Systems* 2015; **28**: 649–657.

45 Trischler A, et al. Newsqa: A Machine Comprehension Dataset. *arXiv preprint* 2016; arXiv:1611.09830.

46 Azizi A, Momtazi S. SNRBERT: Session – Based News Recommender Using BERT. *User Modeling and User–Adapted Interaction* 2024; 1–15. DOI: 10.1007/s11257-024-09409-x.

47 Vo T. An Integrated Topic Modeling and Auto–Encoder for Semantic–Rich Network Embedding and News Recommendation. *Neural Computing and Applications* 2023; **35**(**25**): 18681–18696.

48 Suhartono D, Subalie A. Book Recommendation Using Double – Stack BERT: Utilizing BERT to Extract Sentence Relation Feature for a Content – Based Filtering System. In Proceedings of the International Conference on Multi–Disciplinary Trends in Artificial Intelligence, Hyderabad, India, 21–22 July 2023.

49 Wang S, Guo S, Wang L, Liu T, Xu H. Multi–Interest Extraction Joint with Contrastive Learning for News Recommendation. In Proceedings of the Joint European Conference on Machine Learning and Knowledge Discovery in Databases 2022, Grenoble, France, 19–23 September 2022.

50 Dang TK, Nguyen QP, Nguyen VS. A Study of Deep Learning – Based Approaches for Session – Based Recommendation Systems. *SN Computer Science* 2020; **1**(**4**): 216.

51 Li Z, Shen Z. Deep Semantic Mining of Big Multimedia Data Advertisements Based on Needs Ontology Construction. *Multimedia Tools and Applications* 2022; **81**(**20**): 28079–28102.

52 Y Gu, K Chen. GAN–Based Domain Inference Attack. In Proceedings of the AAAI Conference on Artificial Intelligence, Washington, DC, USA, 13–14 February 2023.

53 Gu Y, Sharma S, Chen K. November. Image Disguising for Scalable GPU – accelerated Confidential Deep Learning. In Proceedings of the 2023 ACM SIGSAC Conference on Computer and Communications Security, Copenhagen, Denmark, 26-30 November 2023.

54 Gu Y, Yan D, Yan S, Jiang Z. Price Forecast with High – Frequency Finance Data: An Autoregressive Recurrent Neural Network Model with Technical Indicators. In Proceedings of the 29th ACM International Conference on Information & Knowledge Management, Virtual Event, 19–23 October 2020.

55 Chen Z, Fu C, Wu R, Wang Y, Tang X, Liang X. LGFat-RGCN: Faster Attention with Heterogeneous RGCN for Medical ICD Coding Generation. In Proceedings of the 31st ACM International Conference on Multimedia, Ottawa, ON, Canada, 29 October–3 November 2023.

56 Chen Z, Fu C, Tang X. Multi-domain Fake News Detection with Fuzzy Labels. In Proceedings of the Database Systems for Advanced Applications. DASFAA 2023 International Workshops: BDMS 2023, BDQM 2023, GDMA 2023, BundleRS 2023, Tianjin, China, 17–20 April 2023.

57 Yin N, Wang M, Chen Z, De Masi G, Xiong H, Gu B. Dynamic Spiking Graph Neural Networks. In Proceedings of the Thirty-Eighth AAAI Conference on Artificial Intelligence, Vancouver, Canada, 20 – 27 February 2024.

58 Su J, *et al*. GSENet: Global Semantic Enhancement Network for Lane Detection. In Proceedings of the Thirty-Eighth AAAI Conference on Artificial Intelligence, Vancouver, Canada, 20–27 February 2024.

59 Su H, *et al*. Sharpness-Aware Model-Agnostic Long-Tailed Domain Generalization. In Proceedings of the Thirty-Eighth AAAI Conference on Artificial Intelligence, Vancouver, Canada, 20–27 February 2024.

60 Lu L, Chen Z, Lu X, Rao Y, Li L, Pang S. Uniads: Universal Architecture-Distiller Search for Distillation Gap. In Proceedings of the Thirty-Eighth AAAI Conference on Artificial Intelligence, Vancouver, Canada, 20–27 February 2024.

61 Chen J, *et al*. Sparse Enhanced Network: An Adversarial Generation Method for Robust Augmentation in Sequential Recommendation. In Proceedings of the Thirty-Eighth AAAI Conference on Artificial Intelligence, Vancouver, Canada, 20–27 February 2024.

62 Wang Y, *et al*. A Closer Look at Classifier in Adversarial Domain Generalization. In Proceedings of the 31st ACM International Conference on Multimedia, Ottawa, ON, Canada, 29 October–3 November 2023.

63 Wang M, *et al*. Joint Adversarial Domain Adaptation with Structural Graph Alignment. *IEEE Transactions on Network Science and Engineering* 2024; **11**(**1**): 604–612.

64 Fu C, *et al*. HAG: Hierarchical Attention with Graph Network for Dialogue Act Classification in Conversation. In Proceedings of the ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Rhodes Island, Greece, 4–10 June 2023.

65 Gu Y, Sharma S, Chen K. Image Disguising for Scalable GPU-accelerated Confidential Deep Learning. In Proceedings of the 2023 ACM SIGSAC Conference on Computer and Communications Security, Copenhagen, Denmark, 26–30 November 2023.