IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 20 July 2023, accepted 9 August 2023, date of publication 17 August 2023, date of current version 23 August 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3305954

RESEARCH ARTICLE

An Enhanced Recommendation Model Based on Review Text Graph and Interaction Graph

SHUANG YANG[1,2] **AND XUESONG CAI**[3]
[1]Shanghai Institute of AI for Education, East China Normal University, Shanghai 200062, China
[2]School of Software Engineering, East China Normal University, Shanghai 200062, China
[3]Shanghai Synergy Digital Technology Innovation Institute, Shanghai 200940, China

Corresponding author: Shuang Yang (52205902021@stu.ecnu.edu.cn)

ABSTRACT User's review text data and rating data, as two major information sources of the recommender system, reflect user's preferences and item characteristics from two different perspectives. Many existing methods rely on one or the other data to make recommendations, ignoring the potential collaborative effects between the two. In terms of processing review data, the existing neural combination models mainly capture the local and continuous dependencies between adjacent words in the review text but have limited ability to capture the global and discontinuous dependencies. Therefore, we propose an enhanced recommendation model based on both the review text graph and the interaction graph. On the one hand, the model represents the review text of each user and item as a graph and uses the graph structure to capture the long-term, global, and discontinuous dependencies between words in the review text. A graph attention network based on connection relationships is used to aggregate the adjacency information of each node while taking word order relationships into account. On the other hand, the model builds an interaction graph based on user-item ratings for feature mining. The results of the two parts are combined to complete the prediction. We conduct experiments on three datasets and the results show that the proposed method can improve the recommendation performance.

INDEX TERMS Recommender system, sentiment analysis, review graph, interaction graph.

I. INTRODUCTION

Recommender system (RS) has become one of the most effective ways to solve information overload. The application of RS in various fields can not only help users quickly dig out the content they are interested in from the massive resources but also help the corresponding service providers to make their products attract the attention of target user groups.

Early recommendation methods mainly rely on rating data to make recommendations [1]. This kind of method is mainly applied to scenarios where users score or mark items, such as movie recommendations, music recommendations, and commodity recommendations. The limitation of rating-based methods is that they may be affected by users' rating preferences, and the sparsity of rating data will also affect the accuracy of recommendation results.

The associate editor coordinating the review of this manuscript and approving it for publication was Valentina E. Balas.

To deal with the sparse data in the RS and improve the recommendation performance, researchers introduce various auxiliary information into the RS, which mainly includes user/item attributes, social network information, image or text, and other multimedia information. Among them, the review-based recommendation methods have gained more and more attention and development. Compared with the basic feedback such as click and purchase, the user's reviews on the product not only contain the characteristic information of the product, but also contain rich personal sentiments, which has important reference significance for recommendation.

Initially, review-based recommendation methods are mainly topic-based modeling methods [2], [3], [4]. With the development of natural language processing, deep learning, and other technologies, methods based on deep learning (such as convolutional neural networks and recurrent neural networks) [5], [6] have been used for sentiment analysis and content analysis of review information to extract users' preferences and needs. The review-based methods can help to more accurately understand the needs and feedback of users and provide more personalized recommendation services.

Despite the success of existing review-based recommendations, there are still some challenges that need to be addressed. First, different users focus on different information points in reviews. If the review information is separated from the user's characteristic attributes, the accuracy of the reflected characteristics will be greatly affected. Therefore, review-based models should not learn user and item characteristics in an independent manner. Second, existing neural combination models mainly capture local and continuous dependencies between adjacent words in the review text, but have limited ability to capture global and discontinuous dependencies, and face challenges in dealing with complex and heterogeneous graph data. Third, the existing text-based methods only consider the local co-occurrence relationship of the words in the text and usually ignore the word order and the global context at the document level.

To solve the above problems, we propose an **E**nhanced **R**ecommendation model based on **R**eview text graph and **I**nteraction graph (**ERRI**). The model integrates the user's rating data and review information, and learns user preferences and item characteristics by constructing graph structure data and applying the graph neural network method.

Specifically, on the one hand, the model builds a specific review graph for each user/item. The topology of the graph is used to capture the long-term, global, and discontinuous dependencies between words in the review text. A graph attention network based on connection relationships is used to aggregate the adjacency information of each node while taking word order relationships into account. On the other hand, the model constructs an interaction graph based on the user's rating information to mine user preferences and item features. Finally, the two parts of features are integrated to make the final prediction.

To verify the performance of the proposed method and the influence of the review part and the rating part on recommendations, we conduct a comprehensive experimental study on three public datasets from different fields.

II. RELATED WORK

A. REVIEW-BASED RECOMMENDATION

As an important branch of RS, review-based recommendation methods have been widely concerned by domestic and foreign research institutions and researchers in recent years.

The traditional review-based recommendation methods are based on topic modeling. These methods combine topic modeling techniques such as Latent Dirichlet Allocation (LDA) and word embedding models with matrix decomposition to learn the underlying feature distribution of users and items. Twil et al. [2] combined topic modeling and sentiment analysis, as well as manual validation techniques for topic tags to extract valuable insights from reviews, utilizing aspect-based sentiment analysis to explore the strengths and weaknesses of each travel location. Xie et al. [7] used LDA topic modeling and sentiment analysis techniques to mine and analyze microblog information to understand the public's emotional and mental state in response to the COVID-19 epidemic. Wang et al. [8] used the LDA model to identify the topic after preprocessing online videos and then used the multi-attention bidirectional long short-term memory (LSTM) to identify the emotional polarity corresponding to each topic of each speaker in the video. The topic-based modeling method can only capture the semantic information of the text at the global level, ignoring the important word order and word context information in the text.

In recent years, the rapid development of deep learning techniques has prompted different neural combination models, such as convolutional neural networks (CNN) and recurrent neural networks (RNN), to be applied to the analysis of review text. Zohair et al. proposed a deep learning model based on bidirectional LSTM and embedded CNN for aspect-level sentiment analysis to solve the problem that the set of input words cannot contain contextual knowledge [9]. Yang et al. proposed a hybrid recommendation model, which combined the user-movie rating matrix and the dimension reduction method to extract user preferences by analyzing user reviews at the sentence level [10]. Arno Breitfuss et al. extracted sentiment from existing movie reviews and constructed a sentimental knowledge graph, combining user sentiment extracted from robot chat messages to recommend movies [11].

In addition, attention-based methods have emerged in recent years to assign importance weights to each word, sentence, or review in the review document. Chen et al. [12] used an attention mechanism to select useful reviews and used a CNN to extract key features from the review set to predict users' ratings simultaneously. Du et al. [13] proposed a hierarchical attention cooperative neural network (HACN) model for recommendation, using two parallel networks based on review text to model users and items respectively.

Although the method based on deep learning can effectively capture the context information of adjacent words, it has some limitations in capturing the long-term, global, and discontinuous dependency between words.

B. GRAPH-BASED RECOMMENDATION

The recommendation problem is naturally a graph problem, because the interaction between users and items in a RS can be represented by a bipartite graph [14], [15]. Therefore, graph representation learning and graph-based recommendation algorithms [16] have become research hotspots in recent years, gradually replacing traditional recommendation methods [17] for modeling user and product features.

Early studies on the graph mainly focus on path-based models [18], which recommended items by modeling the path between users and items. Subsequent studies are mostly based on embedding methods, encoding the graph as low-rank embedding for recommendation [19]. In addition, many hybrid models combining the advantages of the two are constructed.

With the continuous in-depth study of graph-based recommendation algorithms, the graph convolutional networks (GCN) methods have attracted attention because of their ability to process graph structure data [20], [21]. Wu et al. proposed a session-based graph neural network recommendation algorithm, which used GCN to perform sequence prediction tasks [22]. Song et al. used social graphs to propagate information among users and proposed a social network based on an attention network [23]. Fan et al. proposed a graph neural network model combining a user social graph and a user-item graph for recommendation [24]. Liang et al. proposed a dynamic heterogeneous graph convolutional network (DHGCN) for item recommendations that considered not only user-to-item interactions but also user-to-user and item-to-item interactions during graph evolution [25]. La Gatta et al. used powerful representations of hypergraph data structures, combined with modern graph machine learning techniques, to study in the context of music recommendation [26].

Various applications of graph models make the research of text sentiment have more positive progress [27]. Many scholars have adopted graph-based methods to deal with problems related to text data. For example, Xiao B. et al proposed a graph neural network recommendation model, RTN-GNNR, which fused the feature of review text and node features [28]. Yang et al proposed an interactive recommendation model based on an enhanced graph convolution network and fused review attribute [29]. Zhang et al. proposed a multi-aspect enhanced graph neural networks (MA-GNNs) model that learned aspect-based sentiments from reviews and used them to build multi-aspect-aware user-item graphs for item recommendations [30]. Using the advantages of graph structure, the graph model can not only directly correlate words in the text, effectively solve the problem of contextual correlation, but also effectively solve the problem of text out of the semantic environment by using the specific semantic content contained by nodes in the graph.

Although some good results have been achieved, the existing methods of using neural networks to analyze text sentiment still have drawbacks, such as not considering the importance of word order on understanding text semantics.

III. PROBLEM STATEMENT

Let $U = \{u_1, u_2, \cdots, u_{|U|}$ and $I = \{i_1, i_2, \cdots, i_{|I|}$ represent user set and item set respectively, where |U| is the number of users and |I| is the number of items. R $\in \mathrm{R}^{|\mathrm{U}|\times|\mathrm{I}|}$ is the user-item rating matrix. The rating score of user u for item i is denoted as $r_{u,i}$, and the rating scores for the items that the user did not interact with are denoted as 0.

Let $S = \{S_1, S_2, \cdots, S_u, \cdots, S_{|U|}$ represent users' review set, where $S_u = \{s_u^1, \cdots, s_u^t, \cdots, s_u^{N_u}$ is user u's review set and N_u is the number of reviews of user u. Similarly, $\mathrm{T} = \{\mathrm{T}_1, \mathrm{T}_2, \cdots, \mathrm{T}_i, \cdots, \mathrm{T}_{|\mathrm{I}|}$ represents the set of reviews for items, where $T_i = \{T_i^1, \cdots, T_i^t, \cdots, T_i^{N_i}$ represents the set of reviews for item i, and N_i is the number of reviews of item i.

To put it simply, in the case of input of user-item rating matrix and review texts, the task of the recommendation model is to predict the user u's rating $\hat{y}_{u,i}$ of the item i that has not been interacted with, which reflects the degree of user u's preference for item i.

IV. THE PROPOSED MODEL

A. ARCHITECTURE OF THE MODEL

The ERRI model consists of three main parts, the review graph-based modeling module (RG module), the interaction graph-based modeling module (IG module), and the prediction module, as shown in Fig.1.

The RG module is responsible for capturing the sentimental representation of users and items from their review texts. The RG module first constructs the reviews into review graphs. Nodes in the review graph are keywords in reviews, and edges in the review graph represent co-occurrence relations between words in a certain sliding window [31]. Then, the GCN method, which combines node information and retains structural features, is used to obtain the structural context information and learn potential semantic information of the text. On this basis, we use a relational type attention network to distinguish the importance of different types of relationships in the process of node aggregation. Finally, user sentimental embedding and item sentimental embedding are obtained.

The IG module is responsible for mining the user's potential preferences and the item characteristics from the user's historical interaction with the items. The IG module builds the interaction graph according to the user-item ratings. Users and items are the nodes in the interaction graph, and the edge represents the rating relationship between the user and the item. The GCN method is applied to capture the higher-order features in the graph, and the interaction graph-based embedding of the user and the item are obtained.

The prediction module is responsible for integrating the outputs of the RG and IG modules to complete the final prediction task. By integrating the sentimental representation based on reviews and the feature representation based on historical interactions, the module obtains the final user representation and item representation and sends them into a Factorization Machine (FM) [32] for the final prediction.

B. THE RG MODULE

1) REVIEW GRAPH CONSTRUCTION

We use the method in reference [33] to construct the review graph. For user u's review set S_u, we first complete pre-processing operations such as sentence segmentation and removal of stop words on the review text, and extract the keywords of each review. We build a directed graph where the nodes represent the keywords of the text, and the edges

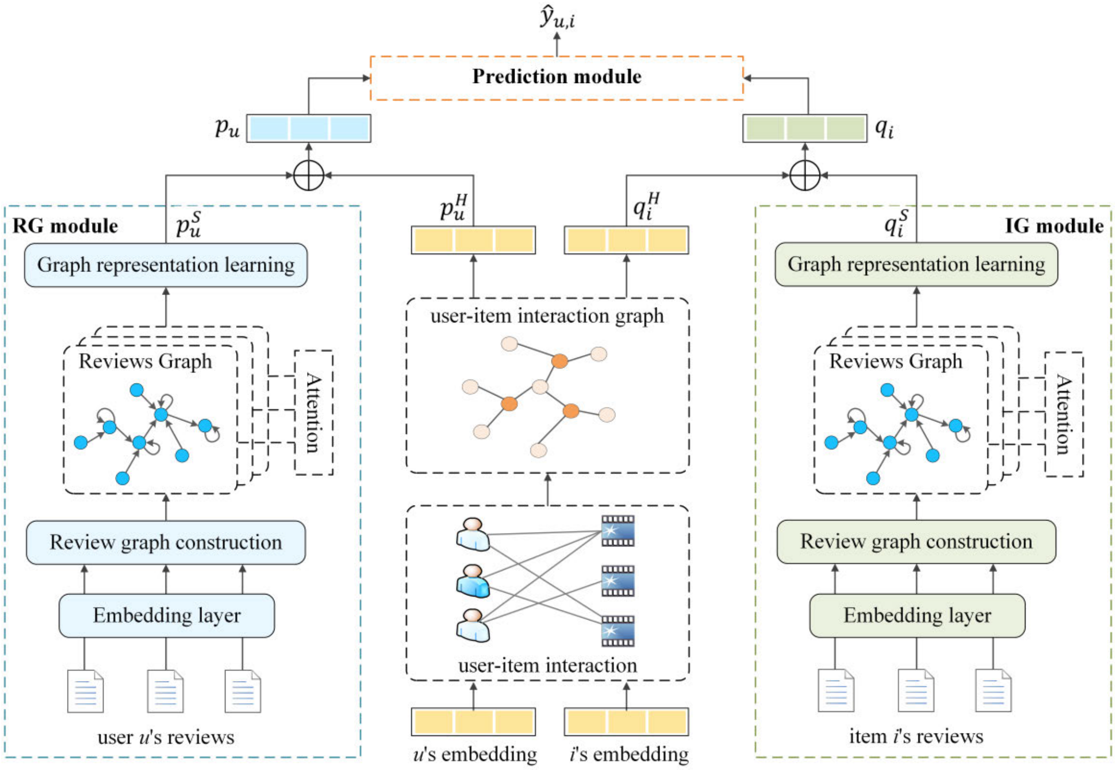


FIGURE 1. Architecture diagram of the ERRI model.

in the graph describe the co-occurrence relationship between words in a fixed-size sliding window. A review graph constructed in such a way can clearly represent the associations between words while preserving the semantic information of the review.

The order of the words in the text has an important influence on the analysis of sentimental characteristics [34]. For example, the sentimental expression "*not really good*" and "*really not good*" have different word orders and express different sentimental information. To record the order relationship between words in a review, we distinguish this order relationship into three types, r_f, r_b, and r_s. r_f represents a forward relationship type, and r_b represents a backward relationship type. For example, if two keywords w_1 and w_2 appear together in a sliding window of size w, and w_1 appears before w_2, there is an edge of type r_f that points from w_1 to w_2. At the same time, there is an edge of type r_b from w_2 to w_1.r_s is a relationship type where a node points to itself, that is, we add an edge to each word in the review graph that points to itself. The main reason for setting this type is to add semantic information about the word itself to the aggregation operation, which helps simplify the computation.

Applying the above approach, a concrete example of constructing the graph with the review "*I like this interesting video game*" is shown in Fig.2. Keywords "*I*", "*like*", "*interesting*", "*video*", and "*game*" are nodes in the graph, and the window size w is set to 3.

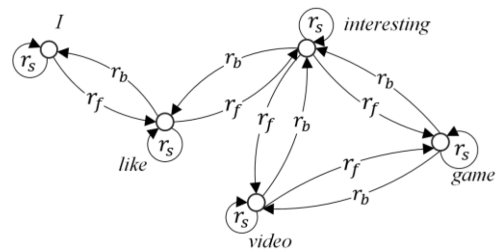


FIGURE 2. The graph constructed with the review "I like this interesting video game".

For each review of user u in the review set S_u, we first use phrase preprocessing technology to clean the words and sentences in the reviews and then select the keywords. On this basis, an un-weighted and directed review graph $G_u^S = \{V_u, T_u$ is constructed, where $\mathrm{V_u}$ represents the set of keywords nodes, T_u represents the set of the triples v_h, r_v, v_t. v_h and v_t represent the two nodes connected by an edge. r_v is one of the three relationship types, representing the type of the edge from node v_t to node v_h. Similarly, we can construct the review graph $G_i^S = \{V_i, T_i$ for the item reviews.

2) GRAPH REPRESENTATION LEARNING

After constructing the review graph, we used the GCN-based method to capture feature information from the graph. For a review graph $G_u^S = \{V_u, T_u$ of a given user u, pre-trained word embeddings are used to make initial representations of nodes in the graph.

For node v_h in the graph, we represent the set of its neighborhood nodes as $N(v_h) = \{v_t|(v_h, r_v, v_t) \in T_u$. Let z represent the semantic representation of node in the review graph, then the representation $z_{v_h}^{(l+1)}$ of node v_h in layer $l+1$ can be calculated by the aggregation of the representations of all its neighbor nodes in layer l, as

$$z_{v_h}^{(l+1)} = AGG_v(z_{v_t}^{(l)}|v_t \in N(v_h)) \tag{1}$$

where, $z_{v_t}^{(l)}$ is the embedding of node v_h's neighbor v_t in layer l. Since each node has an edge that points to itself, the node v_h itself is also included in $N(v_h)$. AGG_v is an aggregation function, which aggregates the semantic representation of all the neighbor nodes at layer l of node v_h to obtain its embedding in layer $l+1$.

Through the above operations, the central node can aggregate the features of neighbor nodes to further enrich its semantic representation.

3) GRAPH ATTENTION NETWORK

In general, the mean aggregator is often used as the aggregation function. Considering the importance of word order relationship to semantic understanding, referring to [35], a relational type attention network is used to aggregate the information of neighbor nodes more emphatically. For $v_t \in N(v_h)$, we define the attention weight as

$$a^{(l)}(v_h, r_v, v_t) = \frac{exp[\pi^{(l)}(v_h, r_v, v_t)]}{\sum_{v_t \in N(v_h)} exp[\pi^{(l)}(v_h, r_v, v_t)]} \tag{2}$$

where $\pi^{(l)}(v_h, r_v, v_t)$ is calculated as

$$\begin{aligned}&\pi^{(l)}(v_h, r_v, v_t)\\ &\quad = LeakyReLU[(v_h^{(l)} W_1^{(l)})(v_t^{(l)} W_2^{(l)} + r_v^{(l)} W_3^{(l)})^T] \end{aligned}\tag{3}$$

where $v_h^{(l)}$, $v_t^{(l)}$, and $r_v^{(l)}$ denote the vector of v_h, v_t and r_v in layer l, respectively. $W_1^{(l)}$, $W_2^{(l)}$, and $W_3^{(l)}$ are the weight matrices, and the weight represents the contribution of each neighbor node to the central node when aggregating the neighborhood information. The method of aggregating neighbors under the relational type attention network can be expressed as

$$z_{v_h}^{(l+1)} = \sum\nolimits_{v_t \in N(v_h)} a^{(l)}(v_h, r_v, v_t) v_t^{(l)} W_4^{(l)} \tag{4}$$

where $W_4^{(l)}$ is the weight matrix, $z_{v_h}^{(l+1)}$ is the output vector of v_h in layer $l+1$.

Compared with the mean aggregator, the aggregation based on the relational type attention network can obtain global context information from reviews and assign different importance to different contexts, which can more accurately mine the hidden sentimental features of the texts.

After the graph convolution operation of L layers, we get L embeddings about the node v_h. We assume that the output of each layer contributes equally to the final embedding, so the embeddings from L layers are averaged as the final representation of node v_h, which can be calculated as

$$z_{v_h} = \frac{1}{(L+1)} \sum\nolimits_{l=0}^{L} z_{v_h}^{(l)} \tag{5}$$

where L represents the number of GCN propagation layers.

Finally, we average the embedding z_{v_h} of all nodes in the review graph G_u^S, and obtain the sentimental representation p_u^S of user u as

$$p_u^S = \frac{1}{|V_u|} \sum z_{v_h}, v_h \in V_u \tag{6}$$

where $|V_u|$ is the number of nodes in V_u.

Similarly, the sentimental representation q_i^S of item i can be calculated as

$$q_i^S = \frac{1}{|V_i|} \sum Z_{v_h}, v_h \in V_i \tag{7}$$

where $|V_i|$ is the number of nodes in V_i.

C. THE IG MODULE

1) INTERACTION GRAPH CONSTRUCTION

We construct the user-item interaction graph G_u^H based on the user's ratings of the items. The nodes in the interaction graph represent users and items. If user u's rating of item i is greater than the threshold we set, an edge is added between user u and item i node. All the historical interaction items associated with the user in the interaction graph are used to model the user's preferences.

2) INTERACTION GRAPH-BASED MODELING

We use an approach similar to the RG module to propagate information between the users and the items and model each node hierarchically. Specifically, the whole operation process is divided into two steps: propagation and aggregation. In the propagation process, the characteristics of nodes are iteratively propagated to adjacent nodes along the connections between nodes. In the aggregation process, the information of the neighbor nodes is aggregated first, and then the aggregated information is aggregated with the information of the central node.

Let h represent the semantic representation of node in the interaction graph, then use the user's historical interaction items to model the user's implicit preferences, which can be expressed as

$$h_{N(u)}^{(k+1)} = AGG_u(h_j^{(k)}|j \in N(u)) \tag{8}$$

where $h_j^{(k)}$ is the embedding of historical item j in the kth layer, $N(u)$ is the set of historical items that user u has interacted with, AGG_u represents the aggregation method used to aggregate the neighborhood information of user u in the interaction graph.

To distinguish the different contributions of each interactive history item to user preference modeling, we designed an attention network to calculate the weight of different historical interactions to user implicit preference modeling, as

$$a_{ij}^* = \sigma(h_i^T h_j + b_1) \tag{9}$$

where h_i is the candidate item, h_j is the historical item, and b_1 is the bias. To make the weights of different interactions easier to compare, the attention score is standardized as

$$a_{ij} = \frac{exp(a_{ij}^{*})}{\sum_{j \in N(u)} exp(a_{ij}^{*})} \tag{10}$$

The normalized attention score is used to calculate the weighted sum of all historical interactions as an implicit preference of the user, as

$$h_{N(u)}^{(k+1)} = \sum_{j \in N(u)} a_{ij} h_j^{(k)} W_5^{(k)} \tag{11}$$

where a_{ij} is the attention of interaction $h_j^{(k)}$ and $W_5^{(k)}$ is the parameter matrix.

Finally, the representation of aggregated neighbor nodes and the representation of central node u in layer k are reaggregated to obtain the representation of user u in layer $k+1$, as

$$h_u^{(k+1)} = AGG_u(h_{N(u)}^{(k+1)}, h_u^{(k)}) \tag{12}$$

Through the above operations, we get the high-order structural features between the nodes in the interaction graph through propagation, and get the user's potential representation p_u^H (i.e. $h_u^{(K)}$) based on the interaction graph finally. Similarly, we can obtain the item feature representation q_i^H based on the interaction graph.

D. PREDICTION MODULE

Through the above two modules, we get the sentimental vectors p_u^S of user u and q_i^S of item i from reviews, and the feature vectors p_u^H of user u and q_i^H of item i from historical interactions. To predict the user's rating of the items, we first integrate the representation from the reviews with that from historical interactions to obtain the representation p_u of user u and q_i of item i, which can be expressed as (13) and (14)

$$p_u = \sigma\left(W_u \times \left[p_u^S, p_u^H\right]\right) \tag{13}$$

$$q_i = \sigma\left(W_i \times \left[q_i^S, q_i^H\right]\right) \tag{14}$$

where W_u and W_i are the parameter matrices and σ is a nonlinear function.

Subsequently, p_u and q_i are concatenated into $[p_u, q_v]$ and then fed into a FM for prediction, as

$$\hat{y}_{u,i} = w_0 + \sum_{i=1}^{n} w_i x_i + \sum_{i=1}^{n} \sum_{j=i+1}^{n} \langle v_i, v_j \rangle x_i x_j \tag{15}$$

where x represents the input vector, w_0 is the global bias, and $\langle \cdot, \cdot \rangle$ denotes the dot product operation. v_i and v_j are parameters for modeling pairwise interaction (x_i, x_j).

The mean square error (MSE) is adopted as the objective function of optimization, and the calculation formula is as

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^{N} (\hat{y}_{u,i} - y_{u,i})^2 \tag{16}$$

where N is the number of user-item pairs, $\hat{y}_{u,i}$ and $y_{u,i}$ represent the predicted and actual ratings of user u on item i, respectively.

TABLE 1. Basic statistics of three datasets.

Dataset	#users	#items	#ratings & reviews	density
Toys	19,412	11,924	167,596	0.072%
Video	24,303	10,672	231,778	0.089%
Yelp2017	35,232	32,636	1,293,357	0.112%

TABLE 2. Summary of compared methods.

Method	Ratings	Reviews	GCN
NeuMF	√		
NARRE	√	√	
DAML	√	√	
MRCP	√	√	
DAM4R	√	√	
AGCN	√	√	√
ERRI	√	√	√

V. EXPERIMENTS

A. DATASETS

We use three publicly available datasets from different domains to evaluate the performance of the model: Toys and Games (Toys for short) and Video Games (Video) from the Amazon 5-core[1] dataset, and the Yelp 2017[2] dataset. All three datasets include user information, product information, and user-written reviews for products. For the Yelp2017 dataset, we filtered users who scored less than 10 times. Table 1 shows the basic statistics for the three datasets.

B. COMPARATIVE MODELS

To evaluate the performance of the ERRI model, we select representative methods from multiple perspectives for comparative experiments, including the rating-based method, the review-based method, and the method considering both interaction and review. NeuMF [36] makes recommendations based on ratings. NARRE, MRCP [6], DAML [37], and DAM4R [38] are review-based methods, of which the first three are models based on the CNN or its deformation, and the last one is a model based on transformer architecture. AGCN [39] considers both review and rating information and applies GCN to capture features in the graph. Table 2 lists the summary information for each model.

C. EXPERIMENT SETUP

In the ERRI model, 80% of the data is randomly selected as the training set, and the rest is equally divided into the verification set and test set. Each experiment is repeated 5 times and the average performance is calculated. We optimize the model hyperparameters on the verification set and conduct experiments on the test set to evaluate the performance of the model. Word vectors are initialized using the Glove [40] pre-trained embeddings. The embedding dimension of the nodes in the review graph is set to 100, and the embedding dimension in the interaction graph is 64. When building the

[1] https://nijianmo.github.io/amazon/index.html
[2] https://www.yelp.com/

TABLE 3. Comparison results of different models.

Model	Toys		Video		Yelp 2017	
	MSE	Impv.	MSE	Impv.	MSE	Impv.
NeuMF	0.8462 ± 0.008	8.86%	1.2510 ± 0.006	12.81%	1.0526 ± 0.017	6.31%
NARRE	0.8273 ± 0.002	6.78%	1.1452 ± 0.002	4.76%	1.0228 ± 0.005	3.58%
DAML	0.8223 ± 0.003	6.21%	1.1347 ± 0.003	3.88%	1.0183 ± 0.004	3.15%
MRCP	0.8163 ± 0.004	5.52%	1.1283 ± 0.004	3.33%	1.0148 ± 0.005	2.82%
DAM4R	0.8077 ± 0.002	4.52%	1.1168 ± 0.003	2.34%	1.0041 ± 0.003	1.78%
AGCN	0.7882 ± 0.003	2.16%	1.1021 ± 0.005	1.03%	0.9972 ± 0.004	1.10%
ERRI	**0.7712** ± 0.001	-	**1.0907** ± 0.002	-	**0.9862** ± 0.001	-

review graph, the size of the sliding window w is 3. We use MSE as the evaluation index.

D. EXPERIMENTAL RESULTS

1) COMPARISON RESULTS OF DIFFERENT MODELS

The comparison results of different models are shown in Table 3. We can draw the following key observations from the results.

NARRE, DAML, MRCP, and DAM4R apply review information to help with recommendations, showing significant improvements compared to models that apply ratings alone. This is mainly because these models can take advantage of the rich semantic features contained in reviews to capture users' sentimental preferences and item characteristics at a deeper level. The experimental results demonstrate the value of review data in mining sentimental preferences and improving recommendation performance.

Based on the comprehensive consideration of ratings and reviews, the AGCN model uses graph structure to represent review data and uses the GCN method to capture features in graph structure data. Its performance is not only better than CNN-based models (NARRE, MRCP, DAML) but also better than the transformer-based model (DAM4R). We analyze that this is mainly due to the greater flexibility of the graph structure, which can directly pass the features of each node to other nodes, better capture the correlation between nodes, and thus better understand the semantic and sentimental information in the reviews. In contrast, transformer-based architectures can only consider the correlation between nodes through the attentional mechanism and may be affected by the distance and dependencies between nodes.

Our ERRI model has the best performance on all three datasets, with 2.16%, 1.03% and 1.10% improvement over the suboptimal method, respectively. As for the reasons, in addition to making full use of review text information and user-item interaction information, our model uses a directed review graph that retains word order to describe review information in the processing of review text, which can not only process complex and heterogeneous data, but also effectively transfer information between nodes, and capture the global discontinuous topological structure and dependency information between words in reviews. Therefore, our model can better understand the semantic and sentimental information in the text.

In addition, compared to the best-performing AGCN model, our model not only has a slightly higher MSE, but also a lower time complexity. We compared the average training time of our model on three datasets with the AGCN model, and the results show that our model is more computationally efficient (i.e. 1h 41m 13s vs 2h 23m 8s).

2) COMPARISON RESULTS OF DIFFERENT VARIANTS

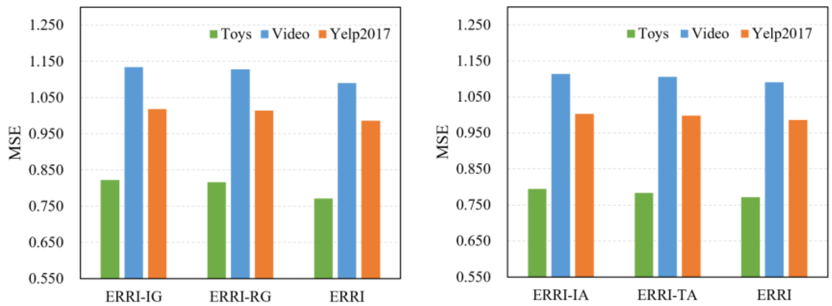
To explore the influence of each component of the model on the overall performance, we design two groups of variants for experiments.

The first group of variants checks the impact of a certain component by removing it from the model. The variant ERRI-RG removes the RG module from the complete model, that is, the model becomes dependent only on the interaction graph for the prediction. The variant ERRI-IG removes the IG module, that is, the model only relies on review texts for the prediction.

Another group of variants tests the effect of each attention network on recommendation performance. ERRI-IA removes the attention network in user preference modeling and item feature modeling of the IG module. ERRI-TA removes the relational type attention network in the RG module. The experimental results of the two groups of variables are shown in Fig.3.

As can be seen from Fig.3(a), in the three datasets, the MSE of ERRI is 5.51%, 3.32%, and 2.80% higher than that of ERRI-RG, and 6.18%, 3.85%, and 3.13% higher than that of ERRI-IG, respectively. This result suggests that a combination of methods that consider rating and review information between users and items generally yields better MSE performance than either the text information model or the rating information model alone. This shows that explicitly modeling the association of rating and review information between users and items can further enhance the model's ability to learn user preferences and item attributes and bring more benefits to the model.

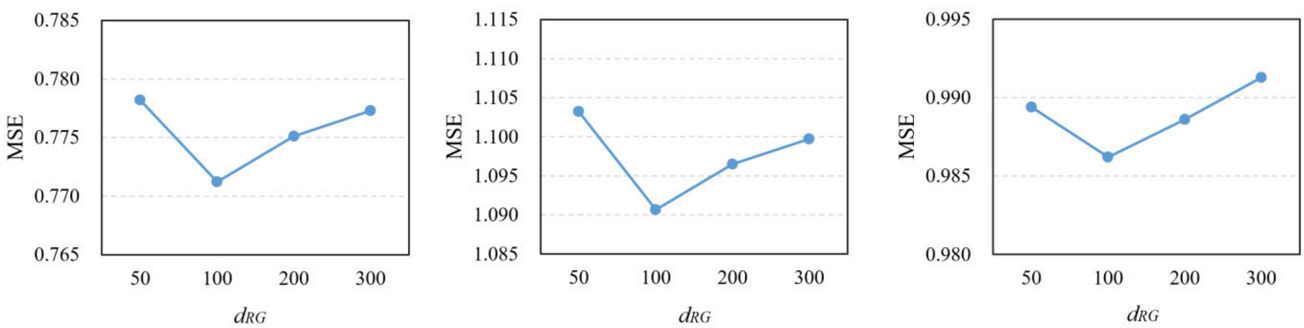
The results in Fig.3(b) demonstrate the effectiveness of introducing the attention networks into IG and RG modules. ERRI performed better than ERRI-IA, with MSE improvements of 2.91%, 2.06% and 1.63% on the three datasets, respectively. Compared with ERRI-TA, the MSE of ERRI increased by 1.56%, 1.31% and 1.21%, respectively. This result suggests that by learning the importance of relevant



(a) Variants on components

(b) Variants on attention networks

FIGURE 3. Comparison of variants on three datasets.



(a) Toys

(b) Video

(c) Yelp2017

FIGURE 4. The impact of embedding dimensions in the RG module.

parts, the relational attention networks can more effectively capture higher-level semantic associations between words, leading to a better understanding of user's preferences, and therefore have better performance than approaches that do not apply the attention mechanism.

E. PARAMETER ANALYSIS

1) EMBEDDING DIMENSIONS IN THE RG MODULE

We analyze the influence of different embedding dimensions on the experimental results by changing the dimension of the word vector in the RG module. The embedding dimensions we selected are 50,100,200 and 300, and the results are shown in Fig.4. As can be seen from the figure, the model performs best when d_{RG} is set to 100. Continuing to increase dimensions does not result in further performance improvements as expected. We infer that this is mainly because when the dimension is too small, it is not enough to represent the rich semantic features of the texts, and when the dimension is large, it is easy to cause overfitting.

2) EMBEDDING DIMENSIONS IN THE IG MODULE

By changing the dimension of the node vector in the IG module, we analyze the influence of different dimensions on the experimental results. We select 16, 32, 64, and 128 embedding dimensions respectively, and the results are shown in Fig.5. As we can see from the figure, the model performs best when d_{IG} is set to 64. We infer that this is mainly because the too-small dimension is not enough to fully express the rich semantic features of nodes, and too large dimension is easy to cause overfitting.

3) NUMBER OF GCN LAYERS

To explore the influence of the number of GCN propagation layers on the model performance, the values of layers L in the RG module and K in the IG module are studied experimentally. The value of K is fixed as 2, and different values from 1 to 4 of L are selected for the experiments. The results are shown in Fig.6. With the same method, the value of L is fixed as 2, and the experimental results of different K values are shown in Fig.7.

As can be seen from the two figures, the experimental results of the two modules are similar, that is, when the number of layers is set to 2, the performance of the model reaches the best. We infer that the main reason for this result is that when L is small, the high-order connectivity between nodes cannot be captured in the review graph, and with the increase of L, the data noise becomes larger, which may be mixed with more irrelevant node information. Therefore, the nodes at $L = 2$ are enough for the model to learn useful relevant information. The same is true for K.

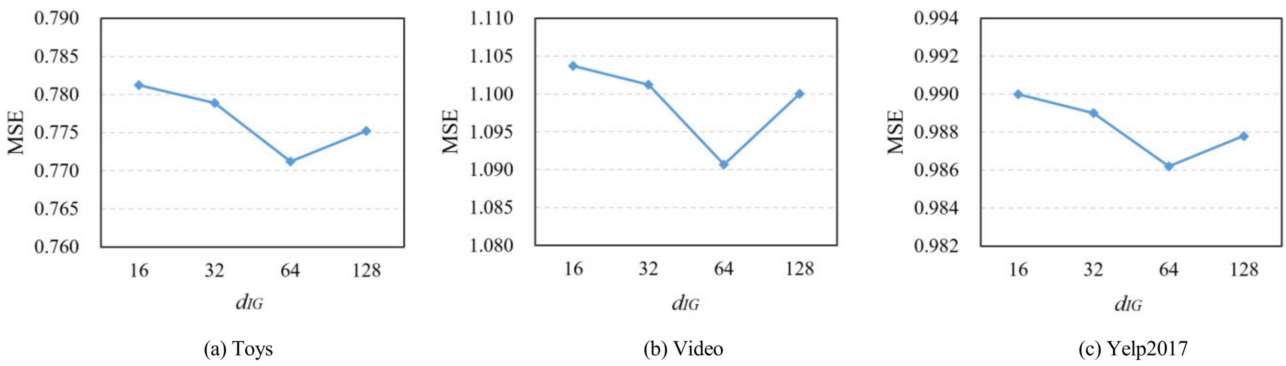


FIGURE 5. The impact of embedding dimensions in the IG module.

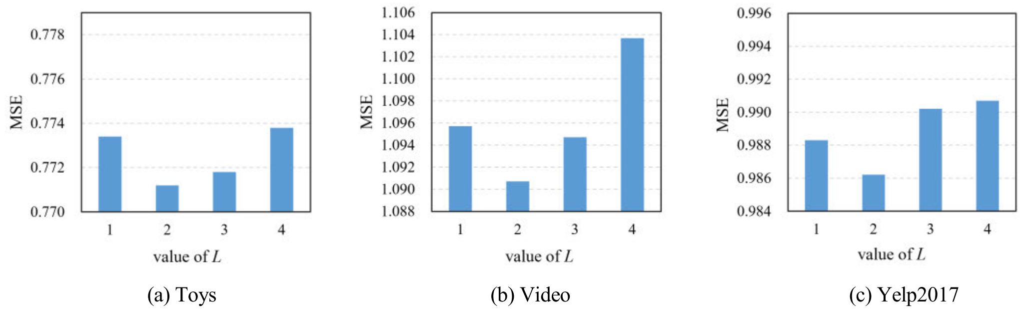


FIGURE 6. The impact of the value of L in the RG module.

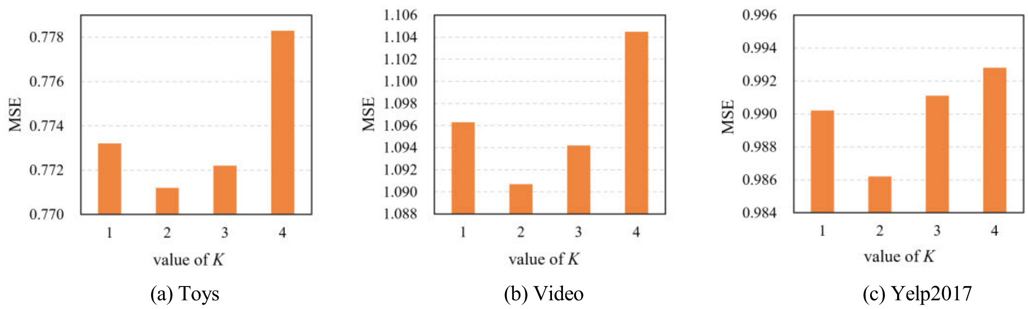


FIGURE 7. The impact of the value of K in the IG module.

VI. DISCUSSION

In summary, the experimental results show that compared with other relevant models, our ERRI model achieves better performance on Video and Yelp2017 datasets, with the greatest performance improvement on Toys dataset. This means that integrating the two types of data for recommendation is an effective design. Furthermore, compared with the traditional methods based on CNN or transformer-based architecture, our ERRI model adopts a graph-based approach to represent review and interaction information, especially the key information of word order is retained in the process of constructing the review graph, which improves the model's ability to understand the semantic information of review text. In addition, the relational attention mechanism employed in the RG and IG modules proved to be valuable in the targeted understanding of user preferences. By providing users' ratings and reviews on items, our model can provide users with personalized recommendation services more effectively.

Even so, our model still has some limitations. The performance improvement of our model is not significant, especially on relatively large datasets. This may be because there may be more noise and bad data in large datasets, which can negatively impact model training and thus subsequent recommendation performance. Our approach addresses noise in user reviews by building a global review graph and adopting a relational attention mechanism to eliminate noise from unrelated reviews. The mutual attention between user reviews and item reviews is not included in the model, which will affect the performance of the model to some extent.

VII. CONCLUSION AND FUTURE WORK

Interaction-based feature modeling and review-based sentiment analysis are two of the research hotspots in recent years. The focus of this paper is to integrate review-based sentiment analysis and interaction-based feature mining, as the double information of recommendation model, to help users quickly dig out the items they are interested in from the massive resources. In the review-based sentiment analysis part, this paper explores the implicit high-order semantic correlation between words by constructing the review graphs and introduces a relational type attention network to distinguish the importance of words with different position relationships. In the experimental part, three datasets are used to conduct relevant experiments, and it is verified that the recommendation model based on review and interaction can effectively improve the performance of item recommendation.

In future work, we can carry out further research in two aspects. First, to address the problem that user reviews may contain noise unrelated to item recommendations, we intend to add a calculation of mutual attention between user reviews and item reviews to help the model focus more on useful review information, thereby improving predictive performance. In addition, this paper only uses user-item interaction data for modeling, whereas in specific application scenarios, there is also a lot of other item-related attribute information, such as item category, keywords, etc., which are also important for item modeling. In the follow-up research, the content of item modeling can be further enriched to obtain better recommendation effects.

REFERENCES

[1] Y. Pan, D. Wu, C. Luo, and A. Dolgui, "User activity measurement in rating-based online-to-offline (O2O) service recommendation," *Inf. Sci.*, vol. 479, pp. 180–196, Apr. 2019.

[2] T. Ali, B. Omar, and K. Soulaimane, "Analyzing tourism reviews using an LDA topic-based sentiment analysis approach," *MethodsX*, vol. 9, Jan. 2022, Art. no. 101894.

[3] J. J. Kim, H. Jang, and S. Roh, "A systematic literature review on humanitarian logistics using network analysis and topic modeling," *Asian J. Shipping Logistics*, vol. 38, no. 4, pp. 263–278, Dec. 2022.

[4] I. Vayansky and S. A. P. Kumar, "A review of topic modeling methods," *Inf. Syst.*, vol. 94, Dec. 2020, Art. no. 101582.

[5] W. Li, L. Zhu, Y. Shi, K. Guo, and E. Cambria, "User reviews: Sentiment analysis using lexicon integrated two-channel CNN–LSTM family models," *Appl. Soft Comput.*, vol. 94, Sep. 2020, Art. no. 106435.

[6] H. Liu, W. Wang, Q. Peng, N. Wu, F. Wu, and P. Jiao, "Toward comprehensive user and item representations via three-tier attention network," *ACM Trans. Inf. Syst.*, vol. 39, no. 3, pp. 1–22, Jul. 2021.

[7] R. Xie, S. K. W. Chu, D. K. W. Chiu, and Y. Wang, "Exploring public response to COVID-19 on Weibo with LDA topic modeling and sentiment analysis," *Data Inf. Manage.*, vol. 5, no. 1, pp. 86–99, Jan. 2021.

[8] Z. Wang, P. Gao, and X. Chu, "Sentiment analysis from customer-generated online videos on product review using topic modeling and multi-attention BLSTM," *Adv. Eng. Informat.*, vol. 52, Apr. 2022, Art. no. 101588.

[9] Z. Ahmed and J. Wang, "A fine-grained deep learning model using embedded-CNN with BiLSTM for exploiting product sentiments," *Alexandria Eng. J.*, vol. 65, pp. 731–747, Feb. 2023.

[10] C. Yang, X. Chen, L. Liu, and P. Sweetser, "Leveraging semantic features for recommendation: Sentence-level emotion analysis," *Inf. Process. Manage.*, vol. 58, no. 3, May 2021, Art. no. 102543, doi: 10.1016/j.ipm.2021.102543.

[11] A. Breitfuss, K. Errou, A. Kurteva, and A. Fensel, "Representing emotions with knowledge graphs for movie recommendations," *Future Gener. Comput. Syst.*, vol. 125, pp. 715–725, Dec. 2021, doi: 10.1016/j.future.2021.06.001.

[12] C. Chen, M. Zhang, Y. Liu, and S. Ma, "Neural attentional rating regression with review-level explanations," in *Proc. World Wide Web Conf. World Wide Web*, 2018, pp. 1583–1592.

[13] Y. Du, L. Wang, Z. Peng, and W. Guo, "Review-based hierarchical attention cooperative neural networks for recommendation," *Neurocomputing*, vol. 447, pp. 38–47, Aug. 2021.

[14] J. Wu, X. Wang, F. Feng, X. He, L. Chen, J. Lian, and X. Xie, "Self-supervised graph learning for recommendation," in *Proc. 44th Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, 2021, pp. 726–735.

[15] J. Liao, W. Zhou, F. Luo, J. Wen, M. Gao, X. Li, and J. Zeng, "SocialLGN: Light graph convolution network for social recommendation," *Inf. Sci.*, vol. 589, pp. 595–607, Apr. 2022.

[16] W. Guo, Y. Yang, Y. Hu, C. Wang, H. Guo, Y. Zhang, R. Tang, W. Zhang, and X. He, "Deep graph convolutional networks with hybrid normalization for accurate and diverse recommendation," in *Proc. 3rd Workshop Deep Learn. Pract. High-Dimensional Sparse Data KDD*, 2021, pp. 1–9.

[17] X. Wang, X. He, M. Wang, F. Feng, and T.-S. Chua, "Neural graph collaborative filtering," in *Proc. 42nd Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, 2019, pp. 165–174.

[18] Z. Sun, J. Yang, J. Zhang, A. Bozzon, L. Huang, and C. Xu, "Recurrent knowledge graph embedding for effective recommendation," in *Proc. 12th ACM Conf. Recommender Syst.*, 2018, pp. 297–305.

[19] H. Wang, F. Zhang, X. Xie, and M. Guo, "DKN: Deep knowledge-aware network for news recommendation," in *Proc. World Wide Web Conf. World Wide Web*, 2018, pp. 1835–1844.

[20] A. Van Den Oord, S. Dieleman, and B. Schrauwen, "Deep content-based music recommendation," in *Proc. Adv. Neural Inf. Process. Syst.*, 2013, pp. 1–9.

[21] Z. Liu, C. Chen, L. Li, J. Zhou, X. Li, L. Song, and Y. Qi, "Geniepath: Graph neural networks with adaptive receptive paths," in *Proc. AAAI Conf. Artif. Intell.*, 2019, pp. 4424–4431.

[22] W. Shu, Y. Tang, Y. Zhu, L. Wang, X. Xie, and T. Tan, "Session-based recommendation with graph neural networks," in *Proc. AAAI Conf. Artif. Intell.*, vol. 33, 2019, pp. 346–353.

[23] W. Song, Z. Xiao, Y. Wang, L. Charlin, M. Zhang, and J. Tang, "Session-based social recommendation via dynamic graph attention networks," in *Proc. 12th ACM Int. Conf. Web Search Data Mining*, Jan. 2019, pp. 555–563.

[24] W. Fan, Y. Ma, Q. Li, Y. He, E. Zhao, J. Tang, and D. Yin, "Graph neural networks for social recommendation," in *Proc. World Wide Web Conf.*, May 2019, pp. 417–426.

[25] T. Liang, L. Ma, W. Zhang, H. Xu, C. Xia, and Y. Yin, "Content-aware recommendation via dynamic heterogeneous graph convolutional network," *Knowl.-Based Syst.*, vol. 251, Sep. 2022, Art. no. 109185.

[26] V. La Gatta, V. Moscato, M. Pennone, M. Postiglione, and G. Sperlí, "Music recommendation via hypergraph embedding," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Oct. 10, 2022, doi: 10.1109/TNNLS.2022.3146968.

[27] P. Pham, L. T. T. Nguyen, W. Pedrycz, and B. Vo, "Deep learning, graph-based text representation and classification: A survey, perspectives and challenges," *Artif. Intell. Rev.*, vol. 56, no. 6, pp. 4893–4927, Jun. 2023, doi: 10.1007/s10462-022-10265-7.

[28] B. Xiao, X. Xie, C. Yang, and Y. Wang, "RTN-GNNR: Fusing review text features and node features for graph neural network recommendation," *IEEE Access*, vol. 10, pp. 114165–114177, 2022, doi: 10.1109/ACCESS.2022.3218882.

[29] Z. Yang, Y. Wang, Y. Cheng, T. Zhang, and X. Wang, "Recommendation model based on enhanced graph convolution that fuses review properties," *IEEE Trans. Computat. Social Syst.*, early access, Jun. 14, 2022, doi: 10.1109/TCSS.2022.3181065.

[30] C. Zhang, S. Xue, J. Li, J. Wu, B. Du, D. Liu, and J. Chang, "Multi-aspect enhanced graph neural networks for recommendation," *Neural Netw.*, vol. 157, pp. 90–102, Jan. 2023.

[31] J. Gu, Z. Wang, J. Kuen, L. Ma, A. Shahroudy, B. Shuai, T. Liu, X. Wang, G. Wang, J. Cai, and T. Chen, "Recent advances in convolutional neural networks," *Pattern Recognit.*, vol. 77, pp. 354–377, May 2018.

[32] S. Rendle, "Factorization machines," in *Proc. IEEE Int. Conf. Data Mining*, Dec. 2010, pp. 995–1000.

[33] R. Mihalcea and P. Tarau, "Textrank: Bringing order into text," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, Jul. 2004, pp. 404–411.

[34] D. Shen, G. Wang, W. Wang, M. R. Min, Q. Su, Y. Zhang, C. Li, R. Henao, and L. Carin, "Baseline needs more love: On simple word-embedding-based models and associated pooling mechanisms," in *Proc. 56th Annu. Meeting Assoc. Comput. Linguistics*, 2018, pp. 440–450.

[35] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio, "Graph attention networks," 2017, *arXiv:1710.10903*.

[36] X. He, L. Liao, H. Zhang, L. Nie, X. Hu, and T.-S. Chua, "Neural collaborative filtering," in *Proc. 26th Int. Conf. World Wide Web*, 2017, pp. 173–182.

[37] D. Liu, J. Li, B. Du, J. Chang, and R. Gao, "DAML: Dual attention mutual learning between ratings and reviews for item recommendation," in *Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Jul. 2019, pp. 344–352.

[38] R. Yu, K. Niu, X. Li, and X. Yang, "Dual attention modeling between users and items using review for recommendation," in *Proc. 3rd Int. Academic Exchange Conf. Sci. Technol. Innov. (IAECST)*, Dec. 2021, pp. 675–679.

[39] Y. Zhang, W. Zuo, Z. Shi, and B. K. Adhikari, "Integrating reviews and ratings into graph neural networks for rating prediction," *J. Ambient Intell. Hum. Comput.*, vol. 14, no. 7, pp. 8703–8723, Jul. 2023.

[40] J. Pennington, R. Socher, and C. Manning, "Glove: Global vectors for word representation," in *Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP)*, 2014, pp. 1532–1543.

SHUANG YANG received the M.S. degree in computer science from Tongji University, China, in 2007. She is currently pursuing the Ph.D. degree with East China Normal University, China. She has been involved in studying machine/deep learning for intelligent education recommendations. Her major research interests include recommender systems, knowledge graph, and graph representation learning.

XUESONG CAI received the Ph.D. degree in software engineering from East China Normal University, Shanghai, China. He is currently with the Shanghai Synergy Digital Technology Innovation Institute. His major research interests include intelligent tutoring systems and recommender systems.