

<b>ITC 2/53</b> <b>Information Technology and Control</b> <b>Vol. 53 / No. 2 / 2024</b> <b>pp. 554-569</b> <b>DOI 10.5755/j01.itc.53.2.35290</b>	<b>Design and Implementation of an English Learning System Based on Intelligent Recommendation</b>	
	Received 2023/10/08	Accepted after revision 2024/03/20
	<b>HOW TO CITE:</b> Yuan, J. (2024). Design and Implementation of an English Learning System Based on Intelligent Recommendation. <i>Information Technology and Control</i> , 53(2), 554-569. <a href="https://doi.org/10.5755/j01.itc.53.2.35290">https://doi.org/10.5755/j01.itc.53.2.35290</a>	

# Design and Implementation of an English Learning System Based on Intelligent Recommendation

**Jinli Yuan**

School of Foreign Languages, Weinan Normal University, Weinan 714000, China

Corresponding author: [teachermary@163.com](mailto:teachermary@163.com)

The Internet has led to the rapid development of online education, but it has also caused redundancy in educational information. How to choose appropriate courses from a large number of online education resources has become a major problem for current learners. Therefore, the study proposes an English learning system based on efficient and deep Matrix decomposition. The results of the experiments showed that, in practical teaching applications, about 57.5% of students with good grades have improved their grades due to the use of the English learning system proposed by the research institute, with only about 17.5% of their grades decreasing. 67.5% of students with average grades have improved their grades after using the system, with only 10% decreasing. Among the students with poor grades, about 50% of them improved their academic performance through the system, while about 27.5% of them experienced a decrease. Meanwhile, the experiment also tested the efficient deep Matrix decomposition model in the learning system: the minimum absolute average errors of the model on different data sets are about 0.61, 0.69, 0.77 and 0.82, respectively. The minimum Root-mean-square deviation is about 0.91, 0.98, 1.06 and 1.1, which is far lower than other recommended models. The above results show that the system constructed in this paper can recommend courses according to students' actual learning level, and can effectively improve students' academic performance in the actual teaching process.

**KEYWORDS:** Online education, intelligent recommendation, matrix decomposition, learning system.

---

## 1. Introduction

With the growth of science and technology, learning methods are also gradually changing, from the original single offline learning mode to Paperless office E-learning. E-learning has the advantages of personalization, informatization and rich resources, which makes it convenient for students to choose the courses they need, and at the same time, learning is not limited by time and place, making Lifelong learning possible. For example, the digital escape rooms proposed by Sidekerskiene and Damasevicius provide online alternatives to traditional physical escape rooms, enabling students to comfortably solve problems and challenges in their own homes or in the classroom [12]. However, due to the continuous increase in online education resources, it is easy for learners to fall into an information maze, making it difficult to find suitable or interesting learning resources. Therefore, how to quickly find the necessary learning resources has become a major challenge for current learners. The intelligent recommendation system is undoubtedly a very effective method to solve this problem. For example, JiZ and his team have implemented personalized e-commerce information recommendation through a mixed recommendation model based on user ratings, comments, and social data [5]. It provides corresponding suggestions based on the personalized needs of users, and its working principle is to connect product attributes with user needs for precise matching. Due to its ability to effectively overcome the problem of information overload, it is widely used in fields such as news, e-commerce, and entertainment. The current intelligent recommendation algorithms are mainly segmented into three categories, namely, Collaborative filtering algorithm (CFA), content-based recommendation algorithm and hybrid recommendation algorithm. The second one has limited recommendation performance due to its difficulty in mining the relevant characteristics of objects and user preferences. Finally, due to the mixing of multiple algorithms, often has too many parameters and is difficult to update and adjust. The first algorithm not only has good personalized recommendation ability, but also has fewer parameters compared to hybrid algorithms, making it widely used [10, 18]. Common methods of CFA include association rules, clustering, Principal component analysis

(PCA), Matrix factorization (MF), among which MF is the most commonly used method. MF establishes implicit factors between users and objects through a user rating matrix to achieve personalized recommendations [11]. However, due to the sparsity of rating data, traditional MF is difficult to accurately construct implicit factors between users and objects, and the training time is long. Therefore, this study proposes an Efficient deep Matrix factorization (EDMF) model based on L0 norm and Convolutional neural network (CNN) to overcome the above problems. Moreover, the particularity of educational scenarios and target populations makes personalized recommendations not entirely dependent on user interests. Personalized recommendations should be completed by analyzing learners' historical learning data to fully understand users' needs and preferences. Based on the above issues, the study proposes to establish a user preference model by analyzing students' past learning data, and combines it with EDMF to achieve accurate recommendation of the courses required by users and effectively improve students' academic performance. The innovation of the research is to use L0 norm to constrain the sparsity of comments, and to construct a Loss function through the maximum a posteriori theory. At the same time, the alternating minimization algorithm is introduced to optimize the Loss function.

This article is divided into five chapters. Section 1 is the introduction. Section 2 will briefly introduce the research status of intelligent recommendation and MF algorithm. Section 3 will study the algorithm of an English learning system based on the EDMF model. Section 4 will analyze the test results of the proposed English learning system. Section 5 will summarize the research findings of the entire article.

---

## 2. Related Work

Currently, the Internet has penetrated into every aspect of people's lives, and has also led to the explosive growth of various types of data. How to quickly and accurately find the necessary information in the massive data has become a major challenge. Li and Yang proposed an intelligent recommendation algorithm

based on topic aware network embedding for resource recommendation on social networks. This algorithm extracts features from the context of user published content through the Topic model, and then uses the topic awareness network embedding framework for intelligent recommendation. After testing, the algorithm has good recommendation performance [8]. Zhang and Dong put forward a recommendation method - LockRec based on Random forest algorithm to solve the problem of how to choose an appropriate locking mechanism in parallel programming. This method first analyzes the feature attributes of the program through program static analysis technology, and then uses forest algorithm to recommend the lock mechanism. The recommendation accuracy of this model is as high as 95.1% [22]. Duan et al. proposed a model based on probability latent semantic index and integrated probability MF to address the issue of how to achieve a better recommendation experience through service rating. It first extracts user and service features by integrating probability MF, and then uses an index model to train user access records to achieve accurate recommendation of services. This model has high accuracy and recall [3]. Cao and his team proposed an intelligent algorithm based on large-scale multi-objective optimization algorithm and MF to address the optimization of intelligent IoT systems. Through testing, this algorithm can effectively optimize the model objectives. Compared with the inflection point driven Evolutionary algorithm, its F1 measure is increased by 7.78% [1]. Scholars such as Zhao proposed a recommendation algorithm based on DP-CRNN to address the issue of how to provide automated clinical guidance for online medical services. It generates a recommendation list based on the characteristics of the patient's inquiry statement and accurately diagnoses the range based on clustering mechanism. This algorithm has certain effectiveness in intelligent pre diagnosis [24]. Jang et al. raised an algorithm grounded on edge computing support to solve the problem of lack of trainers in large-scale network defense exercises, which can realize large-scale network defense exercises without trainers [4]. The MF algorithm is widely used in various industries due to its advantages of simple programming, good scalability, and the ability to deeply explore the connections between users and objects. Kumar and his team proposed a recommendation system that

combines MF and average cumulative score for travel plans and route recommendations. This system can fully consider users' interests and life experiences, and recommend travel plans that users may be interested in based on this [6]. Luo et al. put forward a feature extraction method built on MF for the problem of feature extraction from a large amount of data. This method extracts important features of data through clustering labels of MF prediction data, and its precision and normalized Mutual information are superior to other methods [9]. Xie and other scholars have proposed data processing techniques based on sparse non negative MF to address the issue of data clustering. This technology can effectively extract low dimensional features of data and has excellent performance in data clustering [17]. Zhang et al. proposed a distributed MF algorithm to address the issue of ensuring privacy and efficiency in the intelligent recommendation process. It provides privacy protection by integrating the local Differential privacy paradigm into DS-ADMM, and introduces a random quantization function to reduce the transmission overhead of ADMM to further improve efficiency. After testing, the system can effectively protect user privacy and improve recommendation efficiency, but at the same time, it loses some accuracy [20]. Lee et al. proposed a non-negative MF based detection system for weakly supervised sound events. The system trains the frequency basis matrix through heterogeneous databases, calculates the time basis using the NMF method, and uses it as a feature of the classifier to detect sound events. It has F1 mean fraction performance comparable to the Mel spectrum and  $\gamma$  spectrum, and its performance is 3-5% better than the log Mel spectrum and constant Q transformation [7]. Du et al. proposed a multi view clustering framework based on deep multiple non negative MF to address the issue of integrating heterogeneous information in multi view data. This framework captures hierarchical information by automatically decomposing the input data of the view using an encoder, and utilizes graphic regularization to maintain the local geometric information of the view. Compared with other baseline algorithms, this framework has better clustering performance [2, 13]. In summary, there have been many achievements in academic research on intelligent recommendation algorithms, but there is relatively little research on intelligent recommendation algorithms in the field of

education. MF is widely used in the field of intelligent recommendation due to its ability to fully explore the connections between users and objects. Due to the fact that basic education places more emphasis on students' academic performance rather than just their interest in learning. Therefore, this study proposes an intelligent recommendation algorithm based on EDMF and L0 norm, and determines the weak links of students' knowledge through their historical learning data to sort the recommended courses and improve the accuracy of course recommendations.

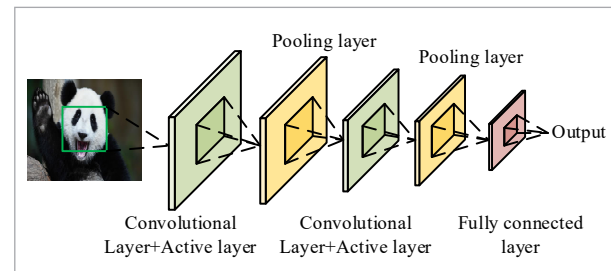
### 3. English Intelligent Learning Recommendation System Based on EDMF

With the increasing availability of text, video, and audio resources on online learning platforms, users are spending more and more time choosing suitable learning resources. At the same time, due to the different learning situations of users and their different needs for learning resources, the traditional learning recommendation system is difficult to meet the personalized needs of users. Based on the above issues, a deep MF based intelligent recommendation algorithm was proposed to achieve personalized recommendations for user learning needs through user comments. [14, 19]

#### 3.1. Research on EDMF Algorithm Based on L0 Norm

In the evaluation of courses by users, the characteristics and applicability of the courses are often hidden. By combining them with the user's past learning data, learning recommendations can be achieved for users. The MF algorithm can deeply explore the connection between users and products, achieving accurate product recommendations. However, traditional MF has the drawbacks of poor visible interpretability of recommendation results and long training time. Therefore, research has proposed using CNN to improve the performance of MF. Figure 1 shows the CNN structure. In Figure 1, CNN is divided into four parts, namely convolution, activation, pooling, and fully connected layers. The function of convolutional layer is to extract the features of input information. The function

**Figure 1**  
Schematic diagram of CNN structure



of pooling layer is to extract main features and reduce the number of features. The role of the activation layer is to simulate the transmission excitation activation mechanism of human brain neurons, so as to control transmission [18-20]. Throughout the EDMF algorithm, the input of the CNN is the evaluation of the course, and the output is the feature vector of the comments. The calculation formula of Activation function is Formula (1).

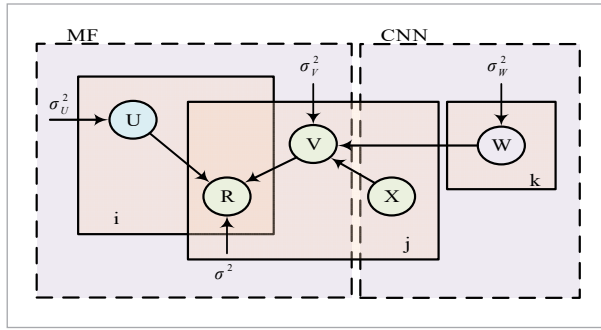
$$y = \sigma(\langle x, w \rangle) = \sigma\left(\sum_{i=0}^D x_i w_i\right) \quad (1)$$

In Equation (1),  $x$  represents the input vector,  $x \in \mathbb{R}^D$ ;  $w$  represents weight;  $x_i$  represents the  $i$ -th input signal;  $w_i$  represents the weight of  $x_i$ . The common Activation function are Sigmoid, Tanh, ReLU and Elu functions. The selection of Activation function should be specific to specific problems. At the same time, in order to solve the problem of overfitting, the dropout technique was chosen to improve the generalization ability of the model. By combining CNN with MF, the ConvMF algorithm is obtained, which obtains predictive scores by inner product of the implicit factors of users' objects. In the ConvMF algorithm, CNN first converts the original information into a digital matrix, extracts its features, and finally obtains the final hidden factor representation through the fully connected layer [15, 21, 25]. The calculation formula for the final output vector is Equation (2).

$$s_j = CNN_w(X_j) \quad (2)$$

In Equation (2),  $X_j$  represents the original information;  $CNN_w$  represents the entire convolution operation;  $W$  represents the set of all parameters in a DJ convolutional network. Figure 2 shows the ConvMF probability MF model.

**Figure 2**  
ConvMF Probability Matrix decomposition Model



From Figure 2, the final output vector obtained through CNN will serve as the input of MF. By processing it, implicit factors related to users and objects can be obtained, thereby achieving predictive scoring of unknown objects. The formula for calculating the probability distribution of observation conditions is Equation (3).

$$p(R|U, V, \sigma^2) = \prod_i^N \prod_j^M \Psi(r_{ij} | u_i^T v_j, \sigma^2)^{I_{ij}} \quad (3)$$

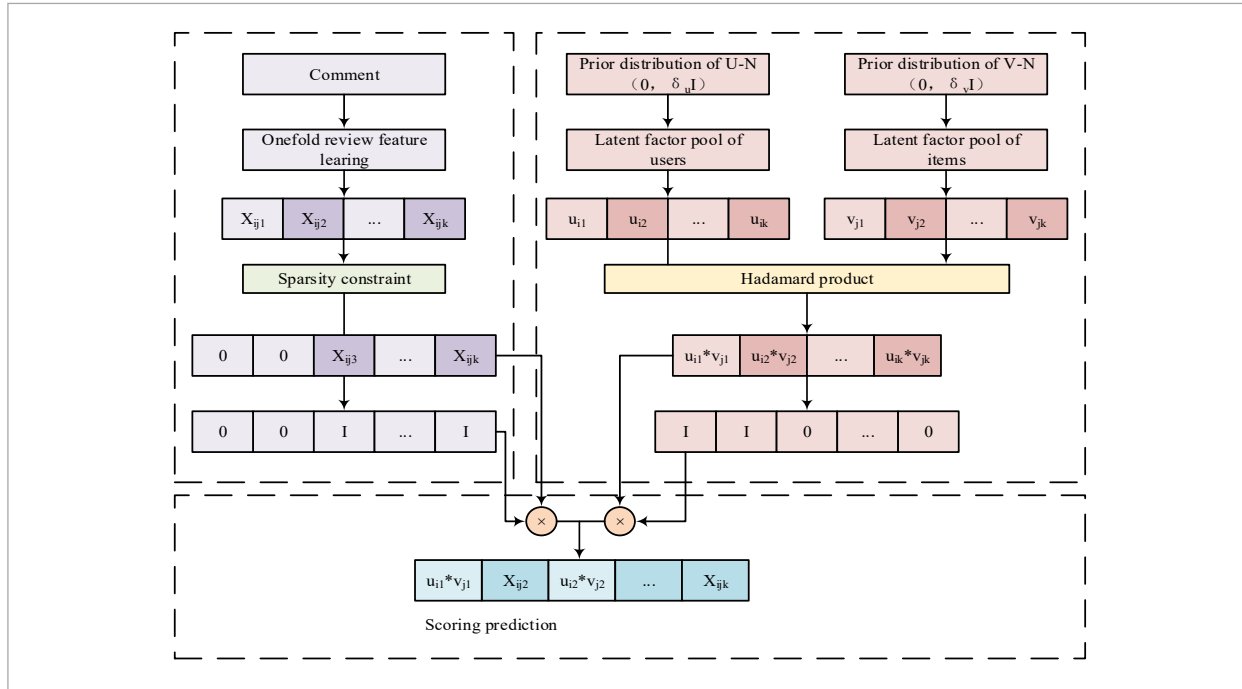
In Equation (3),  $U$  represents the implicit factor matrix of the user.  $V$  represents the implicit factor matrix of the object.  $\mu$  represents the mean.  $\Psi(x|\mu, \sigma^2)$  represents a Gaussian distribution with a mean variance of  $\sigma^2$ .  $I_{ij}$  stands for Indicator function.  $u_i$  represents the implicit factor of user  $i$ .  $v_j$  represents the implicit factor of object  $j$ . The formula for calculating the implicit factor of an object is Equation (4).

$$v_i = CNN_w(X_j) + \varepsilon_j, \text{ where } \varepsilon_j \sim N(0, \sigma_v^2 I) \quad (4)$$

In Equation (4),  $X_j$  represents the original data;  $W$  represents the internal parameters of the CNN;  $\varepsilon_i$  represents Gaussian noise. Due to the sparsity and interactivity of comments, to accurately process the implicit information of comments, it is needed to align the features of comments with the implicit factors of users and objects in the same space. An EDMF based on L0 regularization was proposed for this study. Figure 3 shows the structure of the EDMF algorithm.

In Figure 3, the EDMF algorithm first extracts the contextual features of comments using CNN. Then constrain the sparsity of comments and align them with the implicit factors of users and objects through

**Figure 3**  
EDMF Algorithm Structure



feature alignment; Finally, an optimization algorithm incorporating sparsity constraints is used to obtain feature representations of users and objects. The probability density function of the observation score is Equation (5).

$$p(R|U, V, \Gamma_U, \Gamma_V, \delta) = \prod_{j=2}^M \prod_{i=1}^N \left[ \left( R_{ij} | \hat{R}_{ij}, \delta^2 \right) \right]^{I_{ij}} \quad (5)$$

In formula (5),  $I_{ij}$  represents the Indicator function, and when it is 1, it indicates that the score exists;  $\Gamma_U$  represents the deviation term of the user;  $\Gamma_A$  represents the deviation term of the object. The formula for calculating the Prior probability of users and objects is Formula (6).

$$\begin{cases} p(U|\delta_U) = \prod_{i=1}^N \Psi(U_i | 0, \delta_U^2 I) \\ p(V|\delta_V) = \prod_{i=1}^M \Psi(V_j | 0, \delta_V^2 I) \end{cases} \quad (6)$$

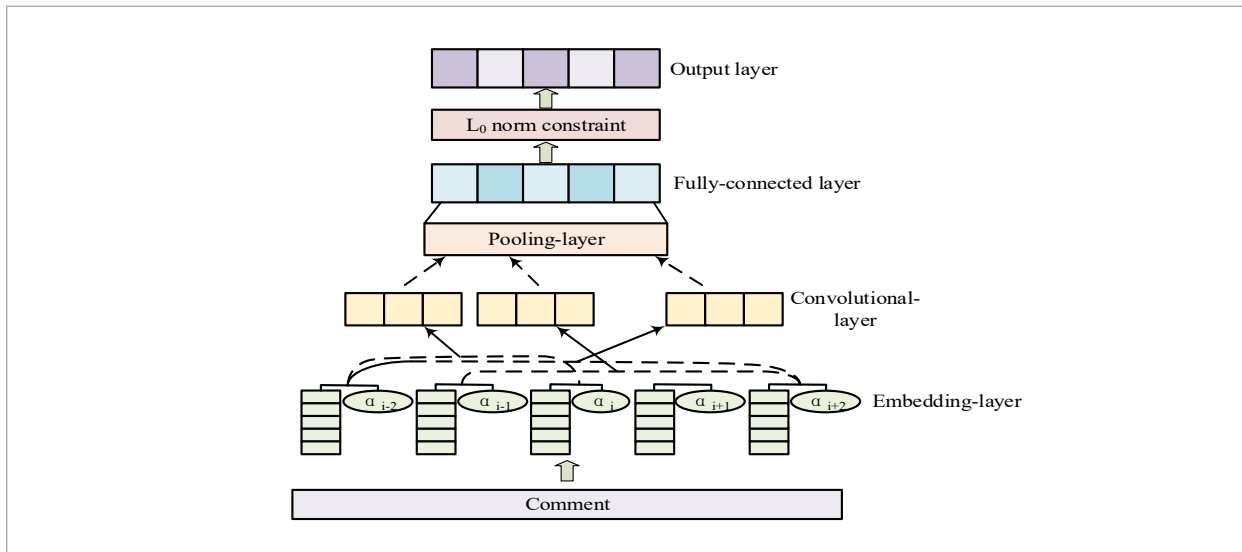
In formula (6),  $\delta_U^2$  represents the Covariance matrix of the user implicit factor matrix.  $\delta_V^2$  represents the Covariance matrix of the hidden factor unit of the object. In the EDMF algorithm, feature learning for a single comment is performed through a CNN based attention mechanism. Figure 4 shows the structure of the comment feature learning algorithm.

In Figure 4, the comment feature learning algorithm first converts the original text information of the comment into a word vector matrix. Then feature extraction is performed through shared convolutional filters; Then sample the extracted features and extract global features. Finally, norm constraints are applied to the output data of the preceding step to gain the ultimate comment features [16, 23]. The contribution of words to comment feature vectors and the formula for calculating contextual features are Equation (7).

$$\begin{cases} D \circ \Phi = \begin{bmatrix} | & | & | \\ \varphi_{i-1} d_{i-1} & \varphi_i d_i & \varphi_{i+1} d_{i+1} \\ | & | & | \end{bmatrix} \\ c_i^j = W^j * (D \circ \Phi)_{(:,i(i+t-1))} + b^j \end{cases} \quad (7)$$

In Equation (7),  $D$  represents the word vector matrix;  $\Phi$  represents the word attention matrix.  $c_i^j$  represents contextual features;  $W^j$  is the  $j$ -th convolutional filter.  $*$  represents a convolution operator;  $b_j$  is the bias of the convolutional filter.  $t$  represents the length of the filter sliding window.  $i$  means the index value in the comment. Due to the varying length of comments, generally speaking, the longer the comment, the more information it contains and the higher its credibility. Therefore, the calculation formula of Loss function considering the influence of comment length is Formula (8).

**Figure 4**  
Structure of Comment Feature Learning Algorithm



$$\begin{aligned}
L_{EDMF+} = & \sum_j^M \sum_i^N \frac{I_{ij}}{2} (R_{ij} - \hat{R}_{ij})^2 + \frac{\lambda_U}{2} |U|_F^2 + \frac{\lambda_V}{2} |V|_F^2 \\
& + \frac{\lambda_R}{2} \sum_j^M \sum_i^N I_{ij} \left\| (U_{i^*} \circ V_{j^*} - s_{ij}) \circ \text{mask}(s_{ij}) \right\|_F^2 \\
& \cdot \log(\text{len}(\Omega_{ij})) + \lambda_s \sum_j^M \sum_i^N I_{ij} C(s_{ij})
\end{aligned} \quad (8)$$

In Equation (8),  $\lambda_s, \lambda_U, \lambda_V, \lambda_R$  represents the parameters of the equilibrium regularization term and the fidelity term.  $| \cdot |_F^2$  is the square of the Frobenius norm.  $s_{ij}$  is the dimension vector  $k$  after norm constraint.  $L_0(s_{ij})$  is the amount of non-0 elements in  $s_{ij}$ .  $\text{mask}(s_{ij})$  represents the position of vector's non-0 elements.  $\text{len}(\Omega_{ij})$  is the length of the comment. At the same time, due to the high difficulty in solving the L0 norm, research has been conducted to find approximate solutions through alternating minimum algorithms to reduce computational difficulty.

### 3.2. Design of English Learning System Based on EDMF

In the learning process, learners often find it difficult to objectively evaluate their learning effectiveness, so course recommendations achieved solely through comments are not accurate. It is necessary to objectively assess students' learning situation based on data such as learning duration, exercise duration, and question accuracy, in order to achieve accurate course recommendations. Figure 5 shows the process of the course recommendation model.

In Figure 5, the model first narrows the search range based on students' learning objectives. Then, based on

data such as students' learning duration, exercise duration, and accuracy, determine their learning situation and analyze their knowledge points and weak links. Then select the corresponding courses based on their similarity, learning situation, and similar users, and generate a course recommendation list. In the calculation of learning situation, the accuracy of exercises can accurately reflect the students' mastery of knowledge points. The formula for calculating the answering situation of unanswered exercises is Equation (9).

$$\rho_{uv} = \prod_{N=1}^N m_{un} q_{vn} = f(m_u, q_v) \quad (9)$$

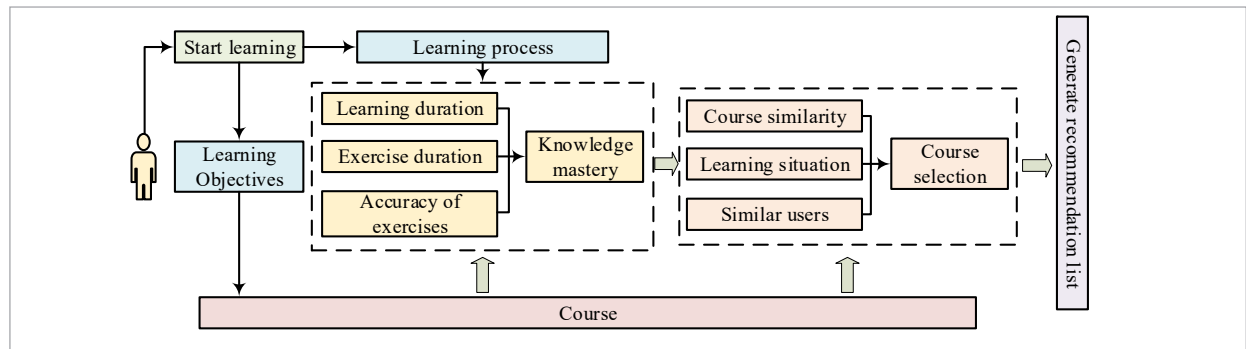
In Equation (9),  $m_u$  represents the knowledge point mastery vector;  $q$  represents the knowledge point matrix;  $f$  represents a power operation. Due to the fact that there are often guesses in correctly answered exercises, in order to accurately measure students' mastery of knowledge points, it is necessary to remove the guessed exercises. The formula for calculating the probability of guessing exercises is Equation (10).

$$P(R_{uv} = 1 | \rho_{uv} = 0) = G_v \quad (10)$$

In Equation (10),  $P$  represents the student;  $R_{uv} = 1$  represents the set of exercises that were done correctly.  $R_{uv} = 0$  is a set of exercises that were done incorrectly.  $G_v$  is the probability of guessing the exercise correctly. At the same time, due to the difficulty of the exercises having a significant impact on the accuracy rate, it is necessary to divide the difficulty level of the exercises. The pass rate of an exercise can effectively reflect the difficulty level of the exercise. The lower the pass rate, the greater the difficulty of the exercise. The formula for calculating the pass rate is Equation (11).

Figure 5

Course Recommendation Model Process



$$A_i = 1 - \frac{\bar{X}_i}{S_i} \tag{11}$$

In Equation (11),  $A_i$  represents the pass rate;  $\bar{X}_i$  represents the average score of the question;  $S_i$  The score of this question. The study defines exercises with a pass rate below 0.6 as difficult; Exercises with a pass rate higher than 0.8 are defined as simple; Exercises with a pass rate between 0.6 and 0.8 are defined as having moderate difficulty. The calculation of the mastery of user knowledge points is Equation (12).

$$E(U, K) = \alpha R_m + \beta R_n + \gamma R_s - G_v \tag{12}$$

In Equation (12),  $E(U, K)$  represents the mastery of knowledge point  $K$  by student  $U$ ;  $\alpha$ ,  $\beta$  and  $\gamma$  represent the weight of simple exercises, exercises with moderate difficulty and the difficult system;  $R$  represents the accuracy of the exercise. When  $0 < E(U, K) < 0.6$  is reached, the user has a poor grasp of the corresponding knowledge points and needs to learn about them; When  $0.6 \leq E(U, K) < 0.8$ , the situation is qualified, but still needs to be consolidated; When  $E(U, K) \geq 0.8$  is reached, it indicates that the user has mastered and should study higher difficulty courses. According to the user's answering pattern, users with the same wrong questions often have more similar knowledge points and weak areas. By searching for similar users, courses that are suitable for the target user but have not been encountered by the target user can be recommended to the target user. The formula for calculating user similarity is Equation (13).

$$W_{ij} = \frac{|N_i \cap N_j|}{\sqrt{|N_i||N_j|}} \tag{13}$$

In Equation (13),  $W_{ij}$  is user similarity.  $N_i$  represents the set of incorrect questions for user  $i$ .  $N_j$  is the set of incorrect questions using  $j$ .  $|N_i \cap N_j|$  is a common set of incorrect questions.  $|N_i N_j|$  is the total number of incorrect question sets. The formula for calculating the user's interest in incorrect questions is Equation (14).

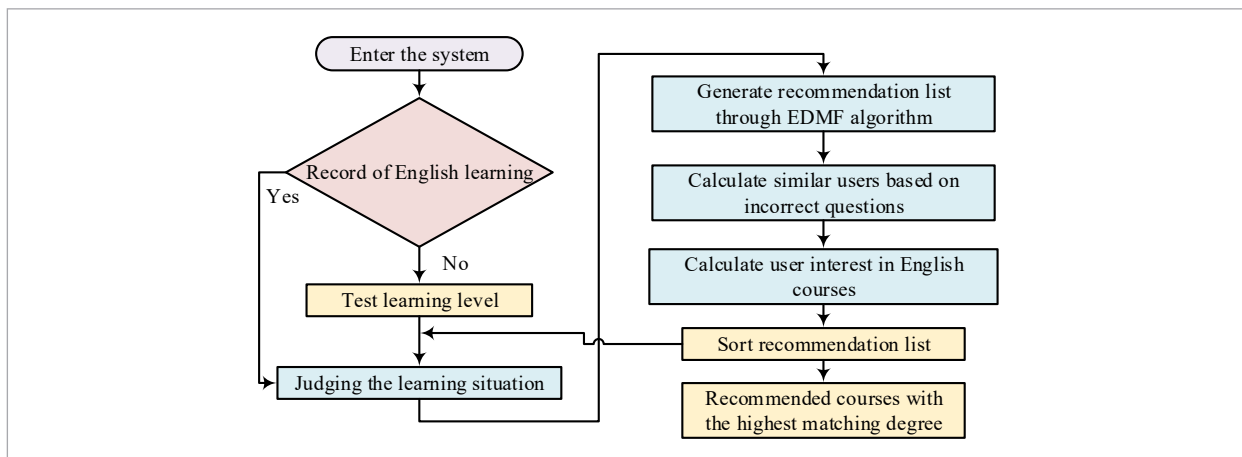
$$P(u, i) = \sum_{v \in S(u, k) \cap N_i} W_{uv} r_{vi} \tag{14}$$

In Equation (14),  $P(u, i)$  represents the degree of interest in incorrect questions.  $S(u, k)$  is the  $k$  users with the highest similarity in target users.  $N(i)$  is a collection of users who have acted on Exercise  $i$ .  $r_{vi}$  represents user  $v$ 's interest in the exercises. Due to the fact that this data is a single behavior implicit feedback data, the formula for calculating the interest of the target user in the course can be simplified as Equation (15).

$$p(A, e) = W_{AB} * R_{Be} + W_{AC} * R_{Ce} + W_{AD} * R_{De} \tag{15}$$

In Equation (15),  $p(A, e)$ ,  $R_{Be}$ ,  $R_{Ce}$  and  $R_{De}$  represent the interest of target users  $A$ ,  $B$ ,  $C$ , and  $D$  in courses  $e$ . By combining the above algorithm with EDMF, an intelligent learning recommendation system can be obtained, and its overall process is Figure 6.

**Figure 6**  
Process of Intelligent Learning Recommendation System

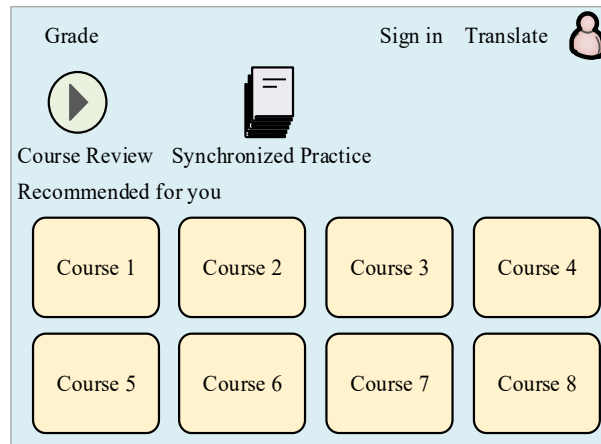




From Figure 6, after the user enters the interface, the system will proceed with the next step based on whether the user has a record of answering questions. If there is no record of the answer question, the user's learning level will be initially measured through a small test. If there is a record of answering questions, the student's learning situation will be judged, and the EDMF algorithm will be used to calculate the attributes of the course and the relationship between the course and users, generating a recommendation list. Then calculate similar users through incorrect questions and calculate the user's interest in the course. Then sort the recommendation list based on interest level. The above recommendation list will be ranked according to the matching degree, and students can study the TOPN course with the highest matching degree to make up for the weak links in learning and improve the effect of learning. Figure 7 is the recommendation interface of the system.

**Figure 7**

Recommended interface of the system



In Figure 7, the “Recommend for You” section in the system will automatically generate a recommendation list. The higher the position of a course in the list, the higher its compatibility with users. After the course ends, there will be corresponding exercises for students to practice. If the accuracy rate meets the standard, the corresponding course will no longer be recommended. Otherwise, continue to recommend corresponding courses to deepen the learning of knowledge points.

## 4. Experimental Results and Analysis

To verify the performance of the EDMF model, simulation experiments were conducted and compared with PMF, HFT, DeepCoNN, and NARRE models. All the above four models were used in course recommendation in online education. The PMF model uses Gaussian distribution to construct the object implicit factor; the HFT model learns the implicit factor; DeepCoNN extracts the implicit factor by parallel CNN; and the NARRE model uses the neural network with attention mechanism to extract the feature vectors. This experiment mainly focuses on the problem of score prediction, to determine the matching degree of the course recommendations. The experiment was conducted on four online education datasets of login.csv (Including login information for online education users, tianchi. aliyun. com/dataset), new\_users.csv (including registration information for new users of online education, tianchi. aliyun. com/dataset), study\_information.csv (Including learning information for online education users, tianchi. aliyun. com/dataset) and MOOC (Including course information for online education, obtained through scraping), respectively. The input of the system is the user characteristics and course characteristics, and the output is the course recommendation results, which specifically shows the user's matching degree for a certain course. At the same time, the experiment also tested the interface pressure and compatibility of the English learning system based on the EDMF model, and conducted an experiment on the effectiveness of the English learning system using a certain grade of students as a sample. The study grouped students based on exam scores. The study will use test scores as the basis for student classification, with scores below 60 for underperforming students, scores between 60-80 for average students, and scores above 80 for top students. In the experiment, 40 students with good, average, and poor grades each took the exam, with a total of 120 students participating in the experiment. The experiment will compare the English grades of students before and after using the English learning system, including their mastery of grammar and spelling of words, to confirm the effectiveness of the system. The parameters of the CNN are shown in Table 1.

**Table 1**  
Parameter settings of the CNN

Name	Value
Kernel Size	2*2/3*3/4*4
Filter	9
The hyperparameter of the Dropout	0.5
The dimension of the implied factor	10
Learning rate	0.001
Train batch size	256

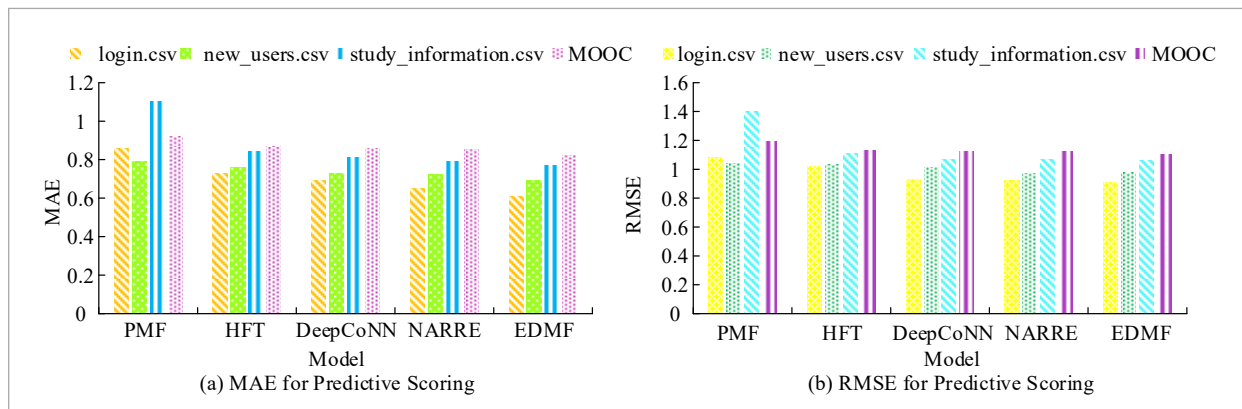
According to Table 1, CNN has three different convolution cores, respectively  $2 * 2, 3 * 3, 4 * 4$ ; the number of filters is 9. The hyperparameter and implied factor dimension of Dropout are 0.5 and 10 respectively; the learning rate and training batch are 0.001 and 256, respectively. Figure 8 shows the MAE and RMSE of the scoring prediction results for the five models.

From Figure 8(a), the MAE of the rating prediction results of PFM on the four datasets of login.csv, new\_users.csv, study\_information.csv and MOOC are approximately 0.86, 0.79, 1.1, and 0.92, respectively. HFT is approximately 0.73, 0.76, 0.84, and 0.87. DeepCoNN is 0.69, 0.73, 0.81, and 0.86; NARRE is 0.65, 0.72, 0.79, and 0.85. EDMF is 0.61, 0.69, 0.77, and 0.82. The

MAE of DEMF on all four datasets is smaller than that of other models. In Figure 8(b), the RMSE of PMF on different datasets is approximately 1.08, 1.04, 1.4, and 1.19, respectively; HFT is 1.02, 1.03, 1.11, and 1.13; DeepCoNN is 0.93, 1.01, 1.07, and 1.12; NARRE is 0.92, 0.97, 1.07, and 1.12; EDMF is 0.91, 0.98, 1.06, and 1.1. Except for the fact that the RMSE of EDMF on the new\_users.csv datasets is not the minimum, the RMSE on other datasets is smaller than that of other models. Figure 9 shows the training efficiency of five models on study\_information.csv and MOOC.

From Figure 9(a), on the study\_information.csv dataset, the PMF, HFT, DeepCoNN, and NARRE models begin to converge after about 25, 18, 17, and 17 training rounds, respectively. At this time, the RMSE of the predicted scores is about 0.98, 0.94, 0.95, and 0.93; EDMF begins to converge after about 13 rounds, with an RMSE of approximately 0.9. From Figure 9(b), on the MOOC dataset, the PMF, HFT, DeepCoNN, and NARRE models begin to converge after approximately 32, 25, 30, and 23 training rounds, respectively. The predicted RMSE scores are approximately 1.15, 1.13, 1.12, and 1.12, respectively. EDMF begins to converge at approximately 22, at which point RMSE is approximately 1.08. Thus, compared with the other four models, EDMF has higher training efficiency, faster Rate of convergence, and lower RMSE value. Figure 10 shows the parameter sensitivity of EDMF to  $\lambda_U, \lambda_V$  and  $\lambda_R$ .

**Figure 8**  
MAE and RMSE of the scoring prediction results for five models



**Note:** Automotive=A; Movies and TV=M/T; Video Game=VG; Yelp-2018=Y.

Figure 9

Training efficiency of different models on study\_information.csv and MOOC

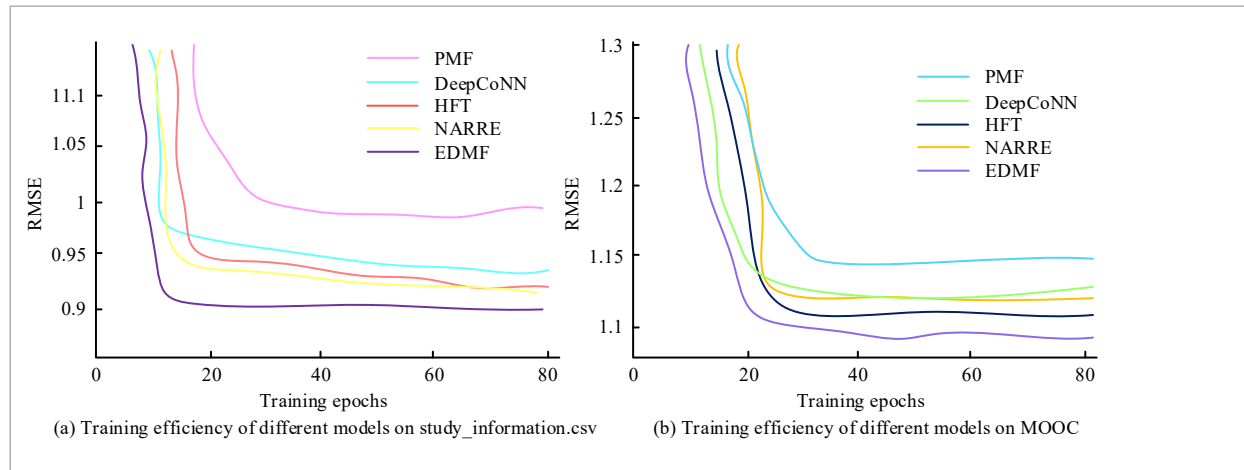
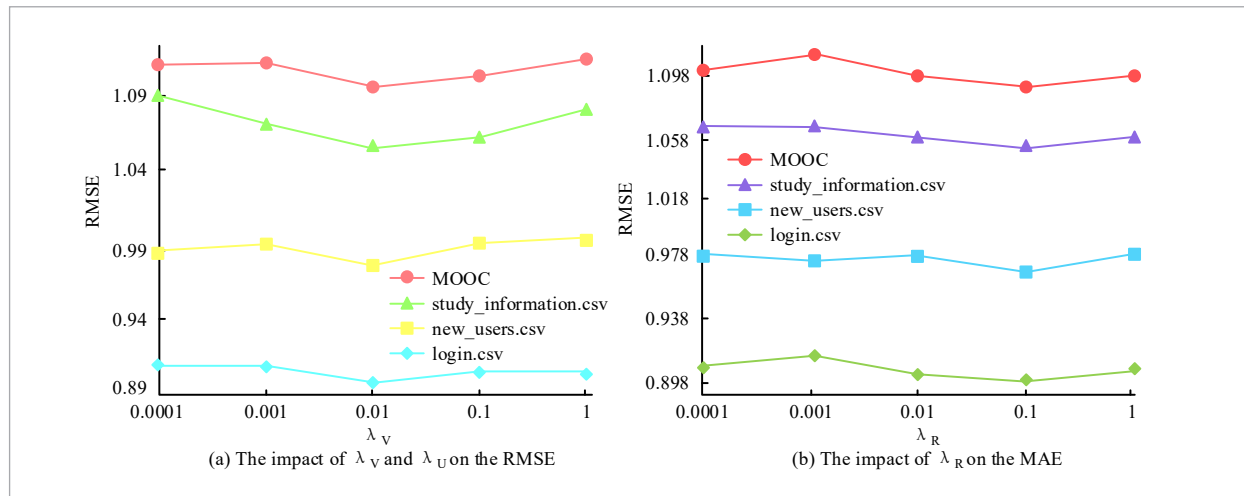


Figure 10

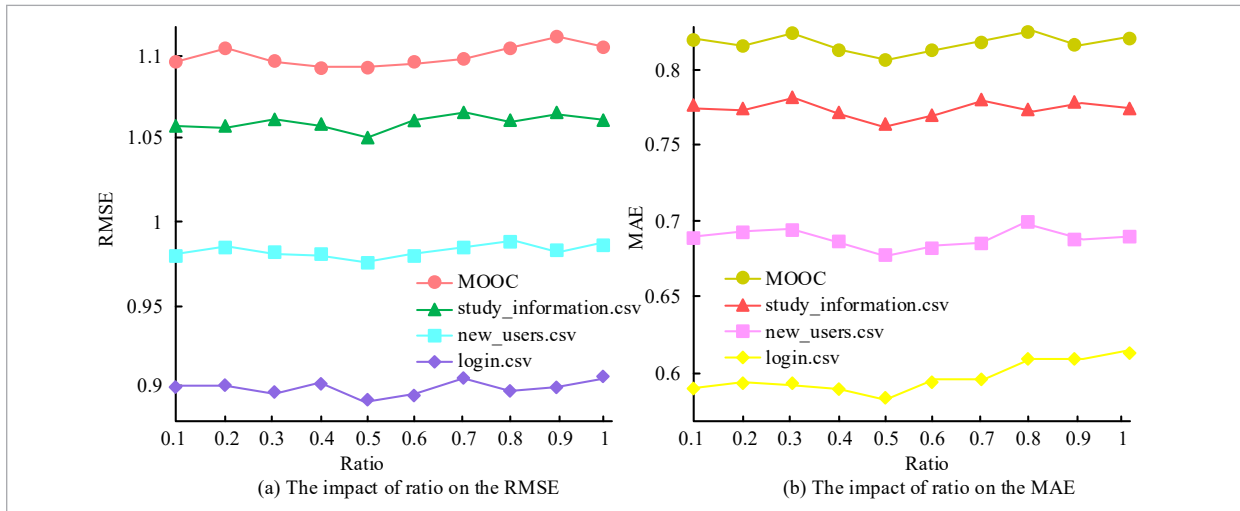
Parameter Sensitivity of EDMF to  $\lambda_U$ ,  $\lambda_V$  and  $\lambda_R$ 

In Figure 10(a), the RMSE values of EDMF on all four datasets vary with changes in  $\lambda_U$  and  $\lambda_V$ . When both the two approach 0, their RMSE is approximately 1.13, 1.09, 0.99, and 0.92, respectively. When they are both 0.01, they are approximately 1.1, 1.05, 0.97, and 0.9; When  $\lambda_U$  and  $\lambda_V$  are both 1, they are 1.12, 1.08, 1.01, and 0.91. This indicates that when both  $\lambda_U$  and  $\lambda_V$  are 0.01, the RMSE values of all four datasets are the smallest. In Figure 10(b), when  $\lambda_U=0.01$ , the RMSE is 1.1, 1.05, 0.98, and 0.9, respectively. When  $\lambda_R=0.1$ , RMSE is 1.09, 1.04, 0.96, and 0.9. When  $\lambda_R=1$ , RMSE is 1.09, 1.05, 0.99, and 0.92. Therefore, when  $\lambda_R=0.1$ , the RMSE value is

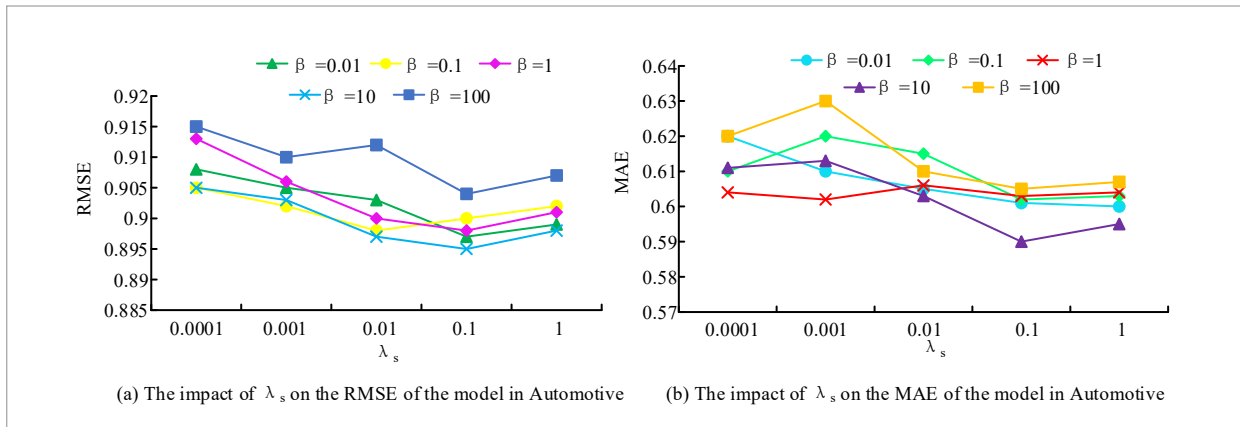
the smallest. To verify the impact of Dropout ratio  $\phi$  on the generalization performance of the EDMF model, the performance was tested on four datasets under different ratios, as displayed in Figure 11.

In Figure 11(a), without the use of Dropout technology, the minimum RMSE values for the four datasets are around 0.89, 0.97, 1.06, and 1.09, respectively, with a ratio of 0.5. In 11(b), when using the Dropout technique, the minimum RMSE values are around 0.55, 0.67, 0.77, and 0.86, respectively. Therefore, it can be concluded that although the size of the ratio value will have an impact on the RMSE value, when the Dropout

**Figure 11**  
Impact of Dropout Ratio on EDM Performance



**Figure 12**  
Experimental results of the EDMF model on MOOC under different  $\lambda_s$  and  $\beta$  values



technique is used, the RMSE value of the model is significantly reduced and the performance is improved, and when the ratio is 0.5, the RMSE value is the smallest. In EDMF,  $\lambda_s$ , as a regularization parameter, has a significant impact on the performance of the model. To verify the effects of regularization parameter  $\lambda_s$  and hyperparameter  $\beta$  on model performance, experiments were conducted on models with different values of  $\lambda_s$  and  $\beta$ . Figure 12 shows the experimental results of the model on the dataset MOOC.

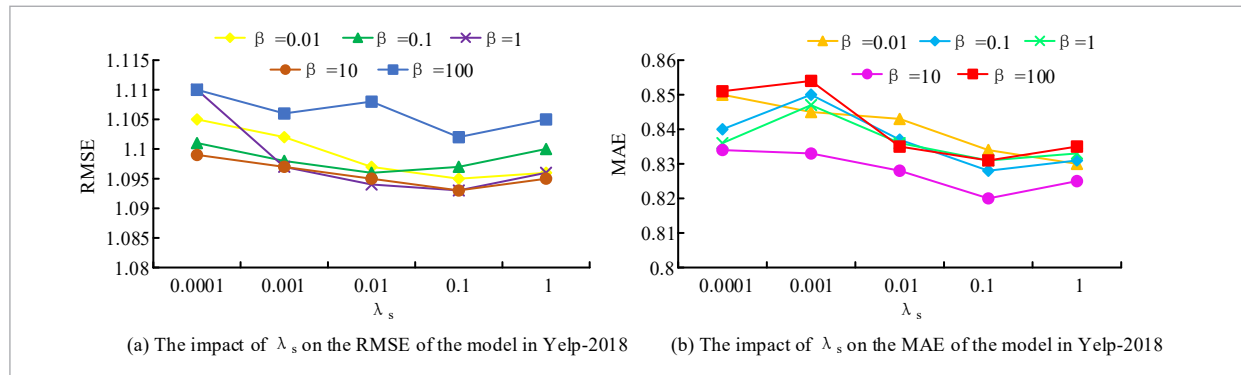
As Figure 12(a), in MOOC, the RMSE value of the model decreases first and then increases with the in-

crease of  $\lambda_s$  and  $\beta$ . When  $\lambda_s = 0.1$  and  $\beta = 10$ , the RMSE value is the smallest, approximately 0.895. In 12(b), as  $\lambda_s$  increases, the MAE value of the model changes irregularly, but as  $\beta$  increases, the MAE value still decreases first and then increases. When  $\lambda_s = 0.1$  and  $\beta = 10$ , the MAE value is the smallest, approximately 0.59. The performance impact of  $\lambda_s$  and  $\beta$  on the EDMF model on the Yelp-2018 dataset is exhibited in Figure 13.

From Figure 13(a), in study\_information.csv, the RMSE value of the model does not vary significantly with  $\lambda_s$  and  $\beta$ . However, when  $\lambda_s = 0.1$  and  $\beta = 10$ , the

**Figure 13**

Performance impact of  $\lambda_s$  and  $\beta$  on the EDMF model on study\_information.csv

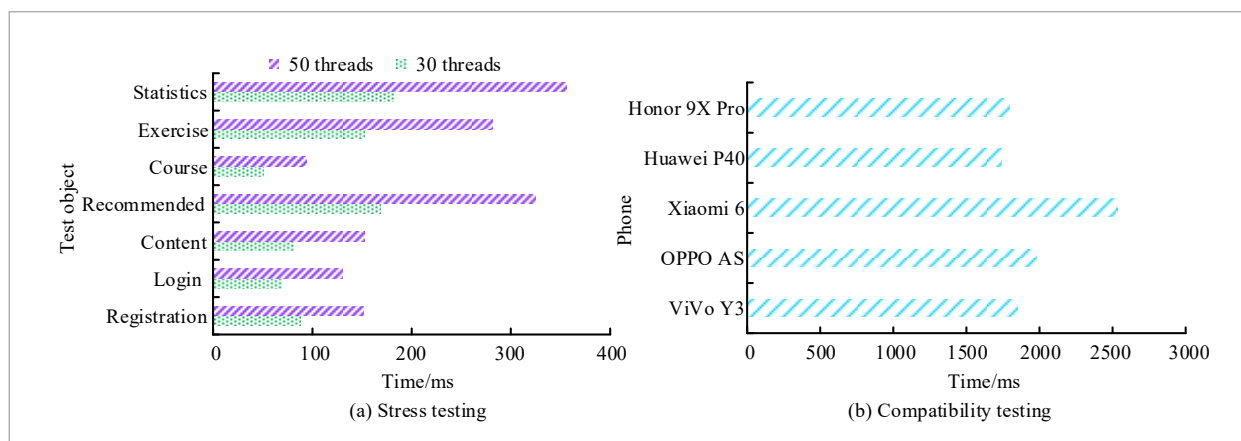


RMSE value is still the smallest, at about 1.093. In 13(b), MAE basically decreases and then increases with the increase of  $\lambda_s$  and  $\beta$ . The minimum MAE is approximately 0.82, at which point  $\lambda_s = 0.1$  and  $\beta = 10$ . From this, when  $\lambda_s$  and  $\beta$  values are large, the comment feature vectors of the model have strong sparsity constraints. To verify the response performance of the interface with high concurrency requests for docking port tuning, the study conducted a stress test. In addition, in order to test the performance of the English learning system in different models of mobile phones and the operating environment, the study also conducted the compatibility test. Figure 14 shows the stress and compatibility test results of an English learning system based on the EDMF model.

As Figure 14(a), when the number of threads is 30, the average time spent on registration, login, content, recommendations, courses, exercises, and statistical interfaces is 88 ms, 69 ms, 82 ms, 169 ms, 51 ms, 153 ms, and 182 ms, respectively. When the number of threads is 50, it is 152 ms, 130 ms, 153 ms, 325 ms, 94 ms, 281 ms, and 356 ms, respectively. From 14(b), the system can start smoothly on different models and brands of mobile phones. The startup times for Vivo Y3, OPPO AS, Xiaomi 6, Huawei P40, and Honor 9X Pro were 1847 ms, 1977 ms, 2536 ms, 1743 ms, and 1796 ms, respectively. Therefore, the interface performance, response speed, and compatibility of the English learning system based on the EDMF model are all qualified. Figure 15 shows the actual application effect of the learning system.

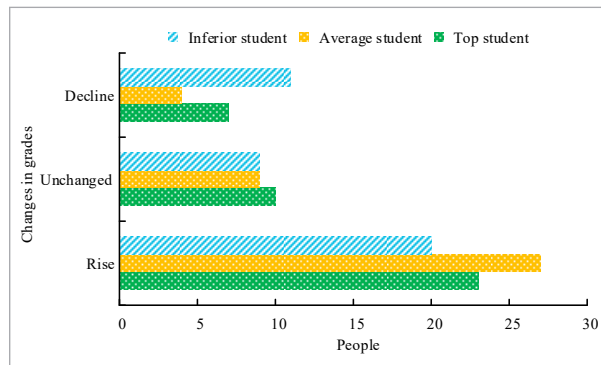
**Figure 14**

Stress and Compatibility Test Results of the English Learning System



**Figure 15**

Practical application effect of the learning system



In Figure 15, the scores of 23 students with good grades have increased, accounting for 57.5% of the total number of students with good grades. The scores of seven students with good grades have decreased, accounting for 17.5%, while the scores of the remaining students with good grades remain unchanged. Twenty-seven students with average grades have improved their grades, accounting for 67.5% of the total number of students in this group. A decrease of 10% in the performance of four students. The remaining grades remain unchanged. The scores of 20 students with poor grades have increased, accounting for 50% of the total group. Eleven places decreased, accounting for 27.5%, while the rest remained unchanged. From this, the English learning system based on EDMF has a probability of improving students' grades by at least 50%. In this experiment, there is no difference in teaching methods and modes of students, so it can be seen that EDMF model can effectively improve students' academic performance.

## 5. Conclusion

Due to the current learning system being filled with a large number of educational resources of varying quality and cumbersome information, students not only need to spend a lot of time choosing suitable courses, but also need to choose the appropriate course among a large number of courses. Based on the above issues, this study proposes an English learning system based on the EDMF model. The system uses EDMF to find connections between us-

ers and courses, and ranks recommended courses based on users' learning situation to improve their adaptability. After testing, the minimum MAE and RMSE values of EDMF prediction scores on the four datasets of Automotive, Video Games, Movies and TV, and Yelp-2018 were 0.61 and 0.91, respectively, which were lower than those of PMF, HFT, Deep-CoNN, and NARRE. Moreover, EDMF converges after a maximum of 22 training rounds on different datasets, with an RMSE value of approximately 1.08. The training efficiency and convergence are better than other models. For the numerous parameters of EDMF, the performance of the model is optimal when  $\lambda_U=0.01$  and  $\lambda_V=0.01$ ,  $\lambda_R=0.1$ ,  $\varphi=0.5$ ,  $\lambda_s=0.1$ , and  $\beta=10$ , with MAE-min of approximately 0.58, 0.69, 0.77, and 0.87 on different datasets, respectively. In the stress testing of an EDMF based English learning system, when the number of threads is 30, the average time required for registration, login, content, recommendations, courses, exercises, and statistical interfaces is at least 51 ms. When the number of threads is 50, the maximum time consumption for each interface is 182 ms. In compatibility testing, the systems on different models of mobile phones can operate normally, with the longest startup time on Xiaomi 6, approximately 2536 ms. In practical experiments, at least 50% of students from different student groups showed an improvement in their grades after using the system. The above results indicate that the recommendation performance of an English learning system based on EDMF is superior to other models, and can effectively improve the learning performance of users. However, although this study has made certain improvements to recommendation algorithms, there is still a lack of consideration in integrating multi resource information and unstructured data.

## Funding

The research is supported by A distinguished scientific research project sponsored by Department of Education, Shaanxi Province, "A Study on Chinese to English Translation of Ancient Chinese Books and Records from the Perspective of Meme Theory: A Case Study of The Historical Records" (No. 20JZ050).

## References

1. Cao, B., Zhang, Y., Zhao, J., Liu, X., Lv, Z. Recommendation Based on Large-Scale Many-Objective Optimization for the Intelligent Internet of Things System. *IEEE Internet of Things Journal*, 2021, 9, 16, 15030-15038. <https://doi.org/10.1109/JIOT.2021.3104661>
2. Du, G., Zhou, L., Kevin Lü., Ding, H. Deep Multiple Non-Negative Matrix Factorization for Multi-View Clustering. *Intelligent Data Analysis*, 2021, 25(2), 339-357. <https://doi.org/10.3233/IDA-195075>
3. Duan, L., Gao, T., Ni, W., Wang, W. A Hybrid Intelligent Service Recommendation by Latent Semantics and Explicit Ratings. *International Journal of Intelligent Systems*, 2021, 36(12), 7867-7894. <https://doi.org/10.1002/int.22612>
4. Jang, M., Kim, D., Seo, D., Ju, Y., Yoon, H. An Intelligent Recommendation Algorithm for Red Team Strategy in Edge Computing Powered Massive Cyber Defense Exercise. *Computer Communications*, 2021, 165(1), 141-148. <https://doi.org/10.1016/j.comcom.2020.10.00>
5. Ji, Z., Pi, H., Wei, W., Xiong, B., Woźniak, M., Damaševičius, R. Recommendation Based on Review Texts and Social Communities: A Hybrid Model. *IEEE Access*, 2019, 7, 40416-40427. <https://doi.org/10.1109/ACCESS.2019.2897586>
6. Kumar, N., Pallavi K V., Hanji, B R. Personalized Travel Recommendation System Using Average Cumulative Rating Matrix Factorization Technique: Concept and Framework. *Vietnam Journal of Computer Science*, 2023, 10(2), 159-195. <https://doi.org/10.1142/S2196888822500361>
7. Lee, S., Kim, M., Shin, S., Park, S., Jeong, Y. Data-Dependent Feature Extraction Method Based on Non-Negative Matrix Factorization for Weakly Supervised Domestic Sound Event Detection. *Applied Sciences*, 2021, 11(3), 1040-1056. <https://doi.org/10.3390/app11031040>
8. Li, J., Yang, G. Network Embedding Enhanced Intelligent Recommendation for Online Social Networks. *Future Generation Computer Systems*, 2021, 119(3), 68-76. <https://doi.org/10.1016/j.future.2021.01.017>
9. Luo, C., Zheng, J., Li, T., Chen, H., Huang, Y., Peng, X. Orthogonally Constrained Matrix Factorization for Robust Unsupervised Feature Selection with Local Preserving. *Information Sciences: An International Journal*, 2022, 586(68), 662-675. <https://doi.org/10.1016/j.ins.2021.11.068>
10. Masood, F., Masood, J., Zahir, H., Driss, K., Mehmood, N., Farooq, H. Novel Approach to Evaluate Classification Algorithms and Feature Selection Filter Algorithms Using Medical Data. *Journal of Computational and Cognitive Engineering*, 2023, 2(1), 57-67. <https://doi.org/10.47852/bonviewJCCE2202238>
11. Ouyang, T. Feature Learning for Stacked ELM via Low-Rank Matrix Factorization. *Neurocomputing*, 2021, 448(11), 82-93. <https://doi.org/10.1016/j.neucom.2021.03.110>
12. Sidekerskienė, T., Damaševičius, R. Out-of-the-Box Learning: Digital Escape Rooms as a Metaphor for Breaking Down Barriers in STEM Education. *Sustainability*, 2023, 15(9), 7393. <https://doi.org/10.3390/su15097393>
13. Simin, Z., Haitao, Z. Depth Estimation of Monocular Infrared Images Based on Attention Mechanism and Graph Convolutional Neural Network. *Journal of Applied Optics*, 2021, 42(1), 49-56. <https://doi.org/10.5768/JAO202142.0102001>
14. Tencer, J., Potter, K. A Tailored Convolutional Neural Network for Nonlinear Manifold Learning of Computational Physics Data Using Unstructured Spatial Discretizations. *SIAM Journal on Scientific Computing*, 2021, 43(4), 2581-2613. <https://doi.org/10.1137/20M1344263>
15. Wang, H., Zhang, C., Zhang, S. Robust Bayesian Matrix Decomposition with Mixture of Gaussian Noise. *Neurocomputing*, 2021, 449(18), 108-116. <https://doi.org/10.1016/j.neucom.2021.04.004>
16. Wang, X., Yu, P., Wu, Y., Huang, Y. K. ALBERT-BPF: A Book Purchase Forecast Model for University Library by Using ALBERT for Text Feature Extraction. *Aslib Journal of Information Management*, 2022, 74(4), 673-687. <https://doi.org/10.1108/AJIM-04-2021-0114>
17. Xie, T., Zhang, H., Liu, R., Xiao, H. Accelerated Sparse Nonnegative Matrix Factorization for Unsupervised Feature Learning. *Pattern Recognition Letters*, 2022, 156(4), 46-52. <https://doi.org/10.1016/j.patrec.2022.01.020>
18. Yang, Y., Song, X. Research on Face Intelligent Perception Technology Integrating Deep Learning under Different Illumination Intensities. *Journal of Computational and Cognitive Engineering*, 2022, 1(1), 32-36. <https://doi.org/10.47852/bonviewJCCE19919>
19. Zhang, C., Okafuji, Y., Wada, T. Reliability Evaluation of Visualization Performance of Convolutional Neural Network Models for Automated Driving. *International Journal of Automotive Engineering*, 2021, 12(2), 41-47. [https://doi.org/10.20485/jsaeijae.12.2\\_41](https://doi.org/10.20485/jsaeijae.12.2_41)

20. Zhang, F., Xue, E., Guo, R., Qu, G., Zhao, G., Zomaya, A. Y. DS-ADMM++: A Novel Distributed Quantized ADMM to Speed Up Differentially Private Matrix Factorization. *IEEE Transactions on Parallel and Distributed Systems: A Publication of the IEEE Computer Society*, 2022, 33(6), 1289-1302. <https://doi.org/10.1109/TPDS.2021.3110104>
21. Zhang, T. W., Li, W. P., Wang, L., Yang, J. Social Recommendation Algorithm Based on Stochastic Gradient Matrix Decomposition in Social Network. *Journal of Ambient Intelligence and Humanized Computing*, 2020, 11(2), 601-608. <https://doi.org/10.1007/s12652-018-1167-7>
22. Zhang, Y., Dong, S. Intelligent Recommendation Method for Lock Mechanism in Concurrent Program. *Journal of Computer Applications*, 2021, 41(6), 1597-1603. DOI: 10.11772/j.issn.1001-9081.2020121929.
23. Zheng, X., Tao, Y., Zhang, R., Yang, W., Liao, Q. TimNet: A Text-Image Matching Network Integrating Multi-Stage Feature Extraction with Multi-Scale Metrics. *Neurocomputing*, 2021, 465(20), 540-548. <https://doi.org/10.1016/j.neucom.2021.09.001>
24. Zhou, X., Li, Y., Liang, W. CNN-RNN Based Intelligent Recommendation for Online Medical Pre-Diagnosis Support. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2020, 18(3), 912-921. <https://doi.org/10.1109/TCBB.2020.2994780>
25. Zhu, D., Renaut, R. A., Li, H., Liu, T. Fast Non-Convex Low-Rank Matrix Decomposition for Separation of Potential Field Data Using Minimal Memory. *Inverse Problems and Imaging*, 2020, 15(1), 159-183. <https://doi.org/10.3934/ipi.2020076>



This article is an Open Access article distributed under the terms and conditions of the Creative Commons Attribution 4.0 (CC BY 4.0) License (<http://creativecommons.org/licenses/by/4.0/>).