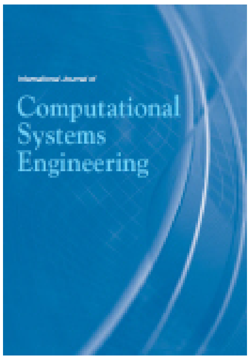
INDERSCIENCE PUBLISHERS
Linking academia, business and industry through research




**International Journal of Computational Systems Engineering**

ISSN online: 2046-3405 - ISSN print: 2046-3391
https://www.inderscience.com/ijcsyse

# A review of hybrid collaborative filtering algorithms for ELT resources under cognitive diagnosis price

Xianghong Tang

**DOI:** 10.1504/IJCSYSE.2023.10055410

**Article History:**

| | |
|---|---|
| Received: | 08 August 2022 |
| Last revised: | 08 September 2022 |
| Accepted: | 13 February 2023 |
| Published online: | 19 March 2024 |

# A review of hybrid collaborative filtering algorithms for ELT resources under cognitive diagnosis price

Xianghong Tang

School of Intelligent Manufacturing and Architectural Engineering,
Yongzhou Vocational Technical College,
Yongzhou, 425000, China
Email: 201523020305@stu.zuel.edu.cn

**Abstract:** The study takes English exercises in English teaching resources as the starting point, and combines cognitive diagnosis theory to assess students' knowledge and ability levels. On the basis of integrating the traditional collaborative filtering algorithm, the sorting learning method is introduced, and the combination of the two becomes a hybrid collaborative filtering algorithm. The results show that the accuracy of the proposed hybrid collaborative filtering algorithm under cognitive diagnosis is as high as 98%, with stable performance in accuracy, F1 value and recall rate. The results outperformed the collaborative filtering algorithm, providing learners with English teaching resources that are more in line with their cognitive ability. The English exercises recommended by the algorithm have better learning effects than those recommended by the collaborative filtering algorithm, effectively providing learners with personalised English teaching resources. In practice, the English exercises recommended by the algorithm were more effective than those recommended by the collaborative filtering algorithm.

**Keywords:** cognitive diagnosis; sequencing learning methods; hybrid collaborative filtering algorithms; English language teaching.

**Reference** to this paper should be made as follows: Tang, X. (2024) 'A review of hybrid collaborative filtering algorithms for ELT resources under cognitive diagnosis price', *Int. J. Computational Systems Engineering*, Vol. 8, Nos. 1/2, pp.56–65.

**Biographical notes:** Xianghong Tang obtained her BA in English Education from Hunan College of Science and Technology in 2006. She obtained her MA in Education from Hunan University of Science and Technology in 2021. Presently, she is working as an Associate Professor in Yongzhou Vocational and Technical College. Her areas of interest are English education, English culture and international cultural exchange and popularisation.

## 1 Introduction

With the rapid development of information technology in education, online learning of English has become an important learning tool and students have access to online teaching resources in a variety of ways. However, the sheer volume of teaching resources on the internet often leads to 'information overload', and learners have to spend a lot of time searching for them. To solve this problem, education services have introduced recommendation system technology, which analyses learners' cognitive ability and knowledge level and sends them teaching resources that meet their individual needs. It is the equivalent of information retrieval, and it is easy to use and meets the individual needs of the user, reducing the time spent by learners searching for resources to a certain extent. Moreover, educational resource recommendation services make the learning system change from the traditional model of people finding resources to a new interactive model of people finding resources and resources finding people, which has received more and more attention from related professionals. Ortega et al. (2018) developed a hybrid recommendation algorithm with a multi-class classification algorithm to improve the quality of recommendations, and implemented it based on user rating behaviour, which improved the prediction and recommendation quality. However, current recommendation services for ELT resources do not sufficiently take into account students' cognitive abilities, i.e., there is a lack of research that uses cognitive diagnostic theory to analyse learners' English knowledge and abilities. In addition, current collaborative filtering algorithms suffer from difficulties in fusing multiple algorithms, sparse data, and cold starts, and they only consider user behaviour and have low recommendation accuracy. Therefore, the research improves the collaborative filtering algorithm by combining the cognitive diagnosis theory, in order to recommend more personalised English teaching resources.

## 2 Related work

In recent years, the application of hybrid collaborative filtering algorithms in ELT resources has received a lot of attention from many domestic and foreign researchers related professionals, and related professionals have also conducted in-depth research on it. Milli and Bulut (2017) proposed a hybrid algorithm combining content collaboration-based methods and collaborative filtering techniques in order to determine the appropriate k value for users, and the results showed that the algorithm improved the average prediction accuracy and Liu et al. (2018) built a deep hybrid recommendation system framework based on autoencoder (DHA-RS ) to address the problem that implicit feedback does not directly reflect user preferences, and built a hybrid recommendation system by integrating item-side information and users, which was shown to have higher performance. Xiong R et al. proposed a hybrid Web service recommendation method based on deep learning by combining textual content and collaborative filtering, and the experiments showed that the method has better recommendation performance (Xiong et al., 2018). Dang et al. (2019) proposed a hybrid optimisation algorithm for flexible multi-objective and based on the proven model using teaching and learning based optimisation to ensure the effectiveness of solving complex multi-objective optimisation problems. Heeryong et al. (2017) proposed a user relationship algorithm that includes four types of user relationships: direct trust, indirect trust, direct distrust and indirect distrust, which not only incorporates the user numerical rating model but also considers the user's relationship from trust and distrust Wang X's team introduced an information retention period based on the information half-value period and proposed a time-weighted function to incorporate the temporal factor into the collaborative filtering algorithm, which was shown to be sufficient to solve the sentiment analysis and temporal effects in recommendation (Wang et al., 2017).

Tm et al. (2019) improved the collaborative filtering algorithm by using a gravity simulation local search algorithm, and the results showed that the proposed hybrid metaheuristic algorithm requires a higher running time and can improve the root mean square error, mean absolute error and coverage criterion. Li and Han (2020) proposed a hybrid collaborative filtering and content-based recommendation scheme and built a time-aware of a user preference model, and the results showed that the model can solve the information overload problem. Baskaran and Murugan (2019) predicted new user interests based on content and using a strategy of supportive filtering and embedded them in consumer expertise, and the results showed that the algorithm improved the accuracy of recommendations. Hattori S et al. proposed a collaborative filtering hybrid recommendation system based on personal values, and the experimental The results showed that hybrid recommendation improved coverage and accuracy (Hattori et al., 2017). Rojo et al. (2018) proposed a hybrid recommendation method based on a classification algorithm in order to improve the quality of recommendations, and the experimental results showed that the algorithm improved the quality of both prediction and recommendation. Li and Li (2019) proposed an improved collaborative filtering algorithm for effectively mining communities in the network, and the results showed that the algorithm reduced the computation time and improved the performance of the recommender system.

In summary, most hybrid collaborative filtering algorithms combine other algorithms with collaborative filtering algorithms and use them to build relevant models, but there is less research on hybrid collaborative filtering algorithms under cognitive diagnosis. Therefore, the study investigates cognitive diagnosis, analyses students' cognitive ability, and explores the establishment of a recommendation system for teaching resources under hybrid collaborative filtering algorithms in order to provide students with high-quality English teaching resources.

## 3 Hybrid collaborative filtering algorithm under cognitive diagnosis for ELT resources

### *3.1 Model design based on cognitive diagnosis*

There is a large amount of student learning information hidden in educational data. The application of big data analysis technology can effectively collect educational data and mine students' learning situation, that is, realise personalised recommendation in the field of education through learning analysis. (Azmi, 2020). As the main recipients of online education, students are the core of learning analysis. However, if we only rely on cognitive diagnosis to analyse students' knowledge level, we will fall into the dilemma of single data and inaccurate evaluation results. Therefore, in order to adequately assess students' knowledge levels and match them with accurate ELT resource services, the existing cognitive diagnostics need to be optimally designed. English exercises, as an important part of ELT resources, can better reflect students' English knowledge mastery level (Rao and Mounika, 2018). The general cognitive diagnosis process, which uses students' practice records to model and thus diagnose their proficiency in knowledge points, is shown in Figure 1.

Factors for conducting cognitive diagnostics can usually be categorised as both test questions and students (Yang, 2021). General cognitive diagnostics usually use manually designed functions to tap into students' learning processes, but these functions are relatively simple and do not capture the complex relationships between test questions and students well. Moreover, when a question has hidden multiple knowledge requirements, this has a direct impact on the observed parameters representing the student, which in turn affects the outcome of performance prediction. The student-side knowledge proficiency vector is shown in equation (1).

$$a = sigmoid(x \cdot A) \tag{1}$$

In equation (1), $a$ represents the knowledge point proficiency vector, $x$ is the student reading hot code vector, and $A$ is the training matrix. Taking four students and five knowledge points as an example, the process of calculating the knowledge point proficiency vector is shown in Figure 2.

**Figure 1** Simple diagram of cognitive diagnosis process (see online version for colours)

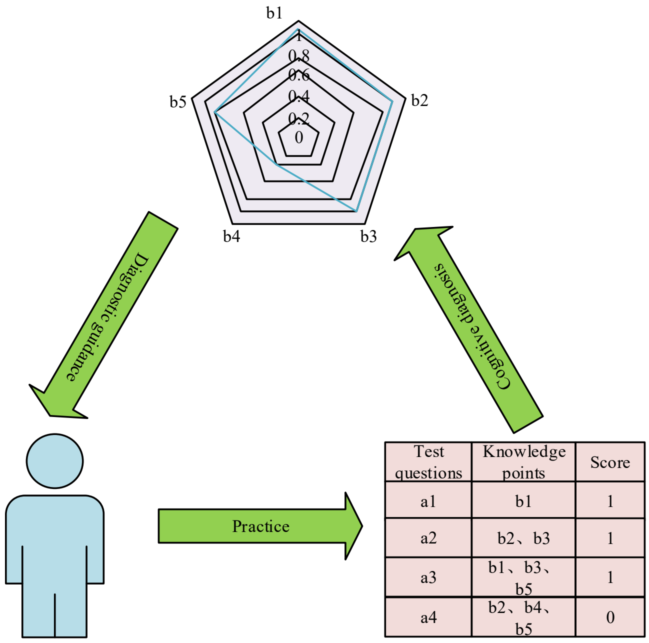
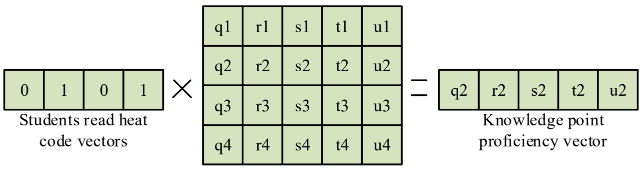



**Figure 2** Proficiency calculation process of knowledge points




The test aspect knowledge point association vector comes from a pre-established matrix as shown in equation (2).

$$S = z \cdot Q \tag{2}$$

In equation (2), $S$ is the knowledge point association vector, $z$ represents the test question reading hot code vector, and $Q$ is the established matrix. It is generally accepted that the difficulty of a knowledge point is related to the importance of the knowledge point, and the importance of the knowledge point is also inextricably linked to the number of relevant test questions, i.e., the more test questions on the knowledge point in the test bank, the more important it is. The ratio of the number of times a knowledge point is examined in a test to the number of test questions determines the frequency information of the knowledge point, as shown in equation (3).

$$Df_i = \frac{1}{V} \cdot \sum_{v=0}^{V} Q_{vi} \tag{3}$$

In equation (3), $D_{ft}$ is the frequency of test preparation, $d_i$ represents the knowledge points and $V$ represents the number of questions. The knowledge point difficulty vector and the knowledge point frequency vector are stitched together in columns to give the matrix of knowledge points tested in all questions in equation (4).

$$\pi = \begin{bmatrix} df_1 & df_2 & \dots & df_i \\ \beta_1 & \beta_2 & \dots & \beta_i \end{bmatrix} \tag{4}$$

In equation (4), $\beta$ is the difficulty vector, $d_f$ is the frequency vector and $\pi$ is the matrixing. If the vectorisation of the $n$ knowledge point is denoted as $\pi_i$ = [$k_{fi}$, $\beta_i$], then the knowledge point tested in all questions is $\pi$ = [$\pi$1, $\pi_2$, …, $\pi_I$]$^T$. Therefore, a filter of size 1 is used to convolve the knowledge points tested as shown in equation (5).

$$c_{ni} = f\left(w_n \cdot \pi_i + b\right) \tag{5}$$

In equation (5), $b$ represents the bias, $f$ is the activation function, $c_{ni}$ is the result of the convolution operation of a single filter, and $w_n$ represents the filter. In order to accurately obtain the importance of the knowledge points, the feature vectors are extracted by multiple filters and then normalised to keep them within a reasonable range, as shown in equation (6).

$$\gamma = soft\max(Cw)^T \tag{6}$$

In equation (6), $C$ represents the matrix and $\gamma$ is the normalised importance vector. In order to achieve better integration of the diagnostic factor vectors and to improve the fit of the interaction function between the test questions and the students, information on the importance of knowledge difficulty and proficiency in predicting students' performance was obtained through a self-attentive mechanism. Firstly, $\beta$ and $a$ were spliced to produce $\Omega = [a, \beta]^T$, which was then fed into the attention network to obtain the assigned weights, as shown in equation (7).

$$\begin{cases} a_n^{'} = sigmoid\left(W_\theta \Omega_n + b_\theta\right)^T \Gamma_\theta \\ a_n = \dfrac{\exp(a_n^{'})}{\sum\limits_{n=1} \exp(a_n^{'})} \end{cases} \tag{7}$$

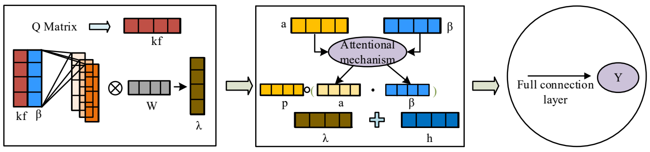
In equation (7), $W_e$ is the weight of the perceptron, $b_e$ is the bias of the perceptron, $a_n$ is the importance of the predicted outcome, and $\Gamma_e$ represents the initialisation vector of the context vector. After obtaining the diagnostic factors for the test questions and students, the knowledge point relevance vector, the knowledge point difficulty vector. Knowledge point proficiency vector and test question differentiation were fused as in equation (8).

$$h = p \circ \left(a \cdot \alpha_1 - \beta \cdot a_2\right) \cdot \mu \tag{8}$$

In equation (8), $\beta$ is the difficulty vector, $a$ is the proficiency vector, $p$ is the relevance vector, $\mu$ is the test differentiation, and $h$ represents the fused diagnostic factors. Therefore, the improved cognitive diagnostic model is shown in Figure 3.

As can be seen from Figure 3, the improved cognitive diagnostic model mainly includes: interaction functions, calculation of knowledge point importance vectors, vectorised representation of diagnostic factors and diagnostic vector fusion. The model first inputs the reading heat coding vector corresponding to the test questions and students, and then multiplies it with the trainable matrix to carry out correlation vectorisation processing; Then it analyses the difficulty of the knowledge points that students should master and the examination frequency, describes the importance of the relevant knowledge points with convolution neural network, and obtains the weight vector; Then calculate the proficiency and difficulty of the test questions through the attention mechanism, and predict the importance information of students' performance based on this; Finally, all the diagnostic vectors are fused, and then directly input into the interactive function, so as to diagnose the mastery of students' knowledge points.

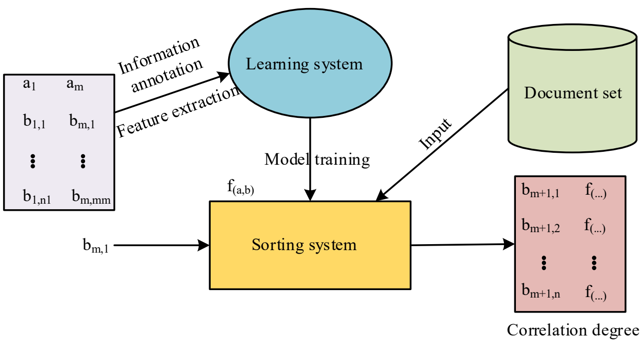
**Figure 3** Improved diagnostic model (see online version for colours)




### 3.2 *A hybrid collaborative filtering algorithm-based recommendation system for English teaching resources*

A ranking learning method is introduced to incorporate multiple collaborative filtering algorithms to form a hybrid collaborative filtering algorithm. The sorting to rank (LTR) method is a technique for further processing sorting results through machine learning methods, which mainly consists of a training part and a testing part (Liu et al., 2018). The training part is based on manually annotated query and sorting results, and uses machine learning related techniques to train an optimal sorting model, which then receives feedback from users and performs sorting recommendations on items recalled in multiplexes; the testing part is based on the former and processes the content of the test set accordingly such as sorting or classifying (Sharma et al., 2019). The general framework of the sorting learning model is shown in Figure 4.

**Figure 4** General framework of sequencing learning model (see online version for colours)




In the ranking learning model shown in Figure 4, the corresponding information features are first extracted from the query documents to obtain the corresponding feature vectors and feed them into the learning system, then a suitable ranking learning algorithm is selected to train the optimal ranking model, and finally the optimal ranking model will give the relevant document list scores for the given query. Supervised and semi-supervised machine learning using ranking learning methods can then produce ordered ranked lists based on the training data information for the purpose of personalised recommendation systems providing users with an ordered list of preferred items (Dong, 2020). For a specific user, the ranking system will display the list of items in descending order based on relevance to provide personalised recommendations for them, and the user's portrait information is characterised as shown in equation (9).

$$\begin{cases} p(u_i) = \sum_{k \in M_i} con(i_k) = \left(g_1(u_i), g_2(u_i), \cdots, g_N(u_i)\right) \\ g_n(u_i) = \dfrac{\sum_{k \in M_i} f_n(i_k)}{M_i}, 1 \le n \le N \end{cases} \tag{9}$$

In equation (9), $i_k$ represents a certain item, $u_i$ represents a user, $N$ is the dimension of the feature space, $f(\cdot)$ is the corresponding feature function, $M_i$ represents the set of historical behavioural items that the user $u_i$ has had, and $con(i_k)$ represents the feature vector. For each new item, the recommender system needs to predict whether the target user will behave accordingly to it. The potential connection between user profile information and items is shown in equation (10).

$$\begin{aligned} sim(i_k, u_k) &= \frac{con(i_k) \times p(u_k)}{\|con(i_k)\| \times \|p(u_k)\|} \\ &= \frac{\sum_n^N f_n(i_k) \times g_n(u_k)}{\sqrt{\sum_{n=1}^N f_n^2(i_k)} \times \sqrt{\sum_{n=1}^N g_n^2(u_k)}} \end{aligned} \tag{10}$$

In equation (10), $u_k$ represents the metric user and $i_k$ represents the new item. Based on the calculation of the cosine similarity between items, the items are sorted in descending order by similarity and the sorted list is output to the user. After obtaining the similarity between items, vector space models, such as the topic distribution model and word frequency-inverse document frequency (TF-IDF), are applied to extract the item features respectively. The feature vector of indexed items is used to represent the text in the vector space model, where each item corresponds to an element in the feature vector, and the weight value of the feature word in the text is expressed in terms of word frequency-inverse document frequency, and the item content feature information is shown in equation (11).

$$f_n(i_k) = TF - IDF(w_n, i_k) = TF(w_n, i_k) \times IDF(w_n) \tag{11}$$

In equation (11), $f_n(i_k)$ is the score of the first $n$ word, $N$ is the number of occurrences of different words, $W = \{w_1, w_2, \ldots, w_n\}$ is the set of individual $n$ words, and $IDF(w_n)$ represents the inverse document rating of the word $w_n$, as shown in equation (12).

$$IDF(w_n) = \log\left(|I| / count(w_n)\right) \tag{12}$$

In equation (12), $count(w_n)$ represents the number of items included in $w_n$ and $|I|$ represents the number of items. When fewer keywords appear in the item-related features or the item content is long, the similarity of the item content is difficult to calculate through the vector space model, so a topic distribution model is introduced for further modelling of the item content, as shown in equation (13).

$$f_n(i_k) = p\left(t_n / i_k\right) \tag{13}$$

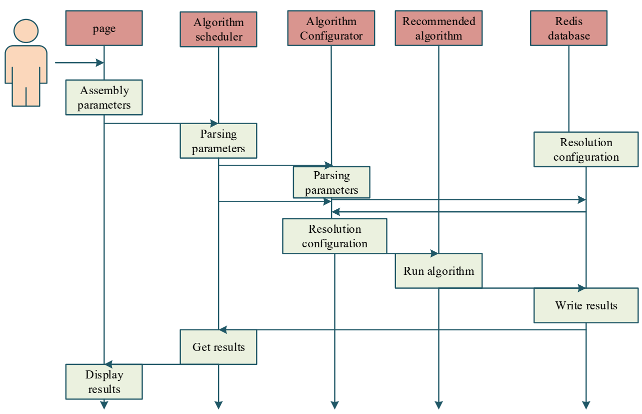
In equation (13), $f_n(i_k)$ represents the calculation function for the generated probabilities and $N$ represents the set of topics. To reduce computational complexity, a conditional probability is used to express the user's preference for an item at this time when known information about the user's behaviour is determined, as shown in equation (14).

$$p(c_{in} = 1 \,|\, c_{kn} = 1) = \frac{\sum_{n \in T} r_{in} \cdot r_{kn}}{\sum_{n \in T} r_{kn}} \tag{14}$$

In equation (14), $T$ represents the set of known items in the training set, is the set of items with behavioural information generated by the user $\sum_{n \in T} r_{in} \times k_n$ $u_i$ in combination with $u_k$, and $\sum_{n \in T} r_{kn}$ is the set of items with behavioural information generated by the user. It is further subdivided to predict the user's preference for the item based on the determination of the known behavioural information, and the final evaluation score is calculated by three strategies: averaging, maximising and summing, as shown in equation (15).

$$\begin{cases} Score_{sum}(i,n) = \sum_{k \in I_n} p\left(c_{in} = 1 \,|\, c_{kn} = 1\right) \\ Score_{\max}(i,n) = Max_{k \in I_n} p\left(c_{in} = 1 \,|\, c_{kn} = 1\right) \\ Score_{avg}(i,n) = \sum_{j \in I_n} p\left(c_{in} = 1 \,|\, c_{kn} = 1\right) / \left|I_n\right| \end{cases} \tag{15}$$

**Figure 5** Recommended algorithm collaboration diagram (see online version for colours)




In equation (15), $u_i$ is the predicted user and $i_k$ represents the item. Once the item content features have been transformed, it is possible to merge the feature scores of the items in the personalised hybrid recommendation model through the relevant ranking learning methods and convert the personalised hybrid recommendation problem into a pairwise ranking learning problem. The optimal fusion weights between the different algorithms are calculated automatically from the training data, to a certain extent avoiding the relatively complex manual tuning of references. The hybrid recommendation model has strong applicability, that is, when different recommendation algorithms can meet the requirements of the model for the established project feature engineering, it can provide personalised recommendation services to the user. The recommendation algorithm framework is based on a configurable mechanism. The general process is as follows: the learner user logs in to the recommendation function interface, and the navigation information and user parameters are transferred here through the page; the system retrieves the page location and the user's recommended algorithm configuration table to obtain dependent data, result output location and recommended algorithm type; Then get the recommendation dependency data, call the recommendation algorithm process, and transfer user information and recommendation dependency data; recommend the execution of the algorithm and output the results at the given position, and notify the caller of the completion; the recommendation results obtained are automatically checked by the page; finally, a list of recommendations is obtained according to the results, with the relevant process collaboration shown in Figure 5.

## 4 Analysis of results

Two datasets were chosen for the experiment. The first dataset contains two parts of data: students' scores on English exercises and the association of knowledge point exercises. 366 learners' scores on 30 English exercises were collected. Incorrect answers to the exercises are indicated by 0, and correct answers to the exercises are indicated by one. The association between the exercises and the knowledge points includes 30 exercises and 10 knowledge points, where the exercise examines the knowledge point is indicated by one, otherwise it is indicated by 0. The second dataset was taken from an English online learning platform, which is the real data of English learning in the second year of a junior high school. The dataset mainly contains 16 English knowledge points, 1,026 English exercises and 657 learners in total. The time period is from 18 February 2022 to 19 June 2022, and the data truly reflects their learning situation. The English exercises were recommended to the target learners by fusing the learners' feedback on their assignments and learning objectives to obtain their personal profiles. The dataset is randomly divided into a test set and a training set, and the recommendation effectiveness of this recommendation algorithm is evaluated using the metrics of recall, accuracy and F1 of the recommendation system. Recall rate (also called recall rate) is the ratio of the number of relevant documents retrieved to the number of all relevant documents in the document library, which measures the recall rate of the retrieval system; Precision is the ratio of the number of relevant documents retrieved to the total number of documents retrieved, which measures the accuracy of the retrieval system. F1 value is the harmonic average of precision and recall. If only accuracy or recall is considered, it can not be used as an indicator to evaluate the quality of a model. Therefore, F1 value is used to reconcile the two, which is compatible with accuracy and recall. Firstly, the maximum number of recommended exercises for each knowledge point should be selected and set as the parameter $\beta$, and the test set should be selected for the experiment with the proportions of 10%, 20 and 40% respectively, and the experimental results are shown in Figure 6.

As can be seen from Figure 6, as the parameter $\beta$ set for the experiment was continuously increased, the accuracy of all three proportional test sets gradually decreased, but the recall rate showed a steady increase. This is because in the English exercise recommendation process, when the number of exercises under each English knowledge point is small, the learners have a higher probability of answering the recommended exercises correctly, and therefore the accuracy rate is high; at the same time, the recall rate is low due to the small number of recommended exercises, but when the value of $\beta$ gradually increases, the recall rate then gradually increases, which is consistent with the actual situation and proves the effectiveness of the exercise recommendation method. The algorithm was then compared with 80% of the randomly selected learners in the training set and 20% of the remaining learners in the test set, and the exercises were tested separately in the training set. The results of the test set are shown in Figure 7.

As can be seen from Figure 7, in terms of accuracy, the hybrid collaborative filtering algorithm was above 90% and performed steadily, with a maximum of 98%, while the collaborative filtering algorithm's accuracy was stable at around 70%, with a maximum of 72% only; in terms of recall, the hybrid collaborative filtering algorithm was between 40% and 50%, with a maximum of 50%, while the collaborative filtering algorithm had a maximum of 40%; in terms of F1 value, the hybrid collaborative filtering In terms of F1 value, the hybrid collaborative filtering algorithm ranged from 60% to 70%, with the highest being 70%, while the collaborative filtering algorithm ranged from 40% to 60%, with the highest being 54%, indicating that the hybrid collaborative filtering algorithm outperformed the collaborative filtering algorithm in terms of F1 value, recall rate and accuracy rate, and the recommendation results were more in line with the cognitive ability of the learners and better met their requirements. The results of the training set are shown in Figure 8.

As can be seen from Figure 8, the F1 value, accuracy rate and recall rate of the hybrid collaborative filtering algorithm are stable at 60%~70%, 80%~95% and 50%~65% respectively, all of which are more advantageous than the collaborative filtering algorithm and can increase the recommendation accuracy and improve the recommendation effect.At the same time, the hybrid collaborative filtering algorithm proposed in the study is compared with the collaborative filtering algorithm based on deep learning:

a the adaptive collaborative filtering algorithm

b the hybrid collaborative filtering algorithm based on feature clustering

c The selected English exercises are still 1026, and the results are shown in Table 1.

**Table 1** Comparison of accuracy, recall and F1 value results of four algorithms

| *Category* | *Accuracy* | *Recall* | *F1* |
|---|---|---|---|
| A | 89.1% | 57.2% | 72.3% |
| B | 83.2% | 53.6% | 77.9% |
| C | 90.2% | 55.4% | 78.0% |
| HCF | 92.3% | 65.3% | 78.1% |

It can be seen from Table 1 that compared with the collaborative filtering algorithm based on deep learning, the adaptive collaborative filtering algorithm and the hybrid collaborative filtering algorithm based on feature clustering, the HCF algorithm proposed in the study shows better performance in terms of accuracy, recall and F1 value. To further test the effectiveness of the practical application, i.e., to apply the algorithm for exercise recommendation in a real-world situation to verify whether it can improve and enhance learning outcomes and whether the recommended English exercise resources can satisfy learners, a second experiment was conducted in a real environment with 58 students randomly selected from four classes in the second year of a junior high school. In order to reduce the effect of non-experimental confounding factors, i.e., the level differences that emerged between the two groups of subjects during the experiment, the experimental sample was selected by random sampling, with a total of 20 students randomly selected from the three achievement levels of low, medium and high. The respective randomly selected 20 students were then disorganised and then 9 students were selected as a group each by using a random number table and applying the simple random sampling method, making a total of 29 students to form group A1, which served as the control group, and the rest as the experimental group A2. The experiment was conducted in the second year of junior high school, and both groups of students had completed the part before the experiment. The whole experimental process is: before the experiment to introduce the use of the learning platform, precautions, the purpose of the experiment, etc., while requiring all experimental subjects to carefully carry out this experiment, and pre-experimental achievement test; then A2 group and A1 group in two hours to review the required content, after the end of the performance test again; finally through the system to analyse the results of the experimental subjects' performance test, that is, the pre-test results and post-test The results were compared and analysed using SPSS 20.0. The results of the intra-group t-test for the A1 and A2 groups are shown in Table 2. It should be noted that t-value refers to t-test, which is divided into single population test and double population test; The single population t-test tests whether the difference between a sample average and a known population average is significant. When the population distribution is normal, such as the population standard deviation is unknown and the sample size is less than 30, the statistics of the deviation between the sample average and the population average are t-distributed; Double population t-test is to test whether the difference between the average of two samples and the population they represent is significant.

From Table 2, it can be seen that the mean score of the A1 group was 71.81 on the pre-test and 75.71 on the post-test, and the mean score of the post-test was 3.9 points higher than that of the pre-test, with a significance value of 0.000 (< 0.001) and a correlation coefficient of 0.958 (> 0.05), and the difference between the pre-test and post-test was significant at 0.000 (< 0.001), indicating that the mean score of the A1 group on the post-test and the mean score of the pre-test were significantly linearly correlated. The mean value of the post-test in group A2 was 83.00 and the mean value of the pre-test was 71.69, the post-test was 11.31 higher than the pre-test, the significance value was 0.000 (< 0.001), the correlation coefficient was 0.999 (> 0.05), the significance of the difference between the pre-test and the post-test was 0.000 (& lt; 0.001) indicating that the mean post-test scores of the A2 group were significantly and linearly correlated with the pre-test and that the mean post-test scores of the A2 group were significantly different from the pre-test.

**Table 2** Intra group t-test results of groups A1 and A2

| *Category* | | *Mean value* | *N* | *Standard deviation* | *Mean standard error* | *Correlation coefficient* | *Significance* | *T* | *Significance of difference* |
|---|---|---|---|---|---|---|---|---|---|
| A1 | Pretest | 71.81 | 29 | 11.738 | 2.181 | 0.958 | 0.000 | –5.663 | 0.000 |
| | Post test | 75.71 | 29 | 9.891 | 1.838 | | | | |
| A2 | Pretest | 72.71 | 29 | 12.702 | 2.360 | 0.999 | 0.000 | –11.848 | 0.000 |
| | Post test | 83.02 | 29 | 7.585 | 1.500 | | | | |

**Table 3** T-test results between groups A1 and A2

| *Category* | | *N* | *Mean value* | *Standard deviation* | *Mean standard error* | *T* | *Significance* |
|---|---|---|---|---|---|---|---|
| Pre-test | A1 | 29 | 71.81 | 11.738 | 2.181 | 0.034 | 0.977 |
| | A2 | 29 | 71.71 | 12.702 | 2.360 | | |
| Post test | A1 | 29 | 75.71 | 9.891 | 1.838 | –3.159 | 0.003 |
| | A2 | 29 | 83.02 | 7.585 | 1.500 | | |

**Figure 6** Training set result graph (a) 100% (b) 20% and (c) 40% (see online version for colours)

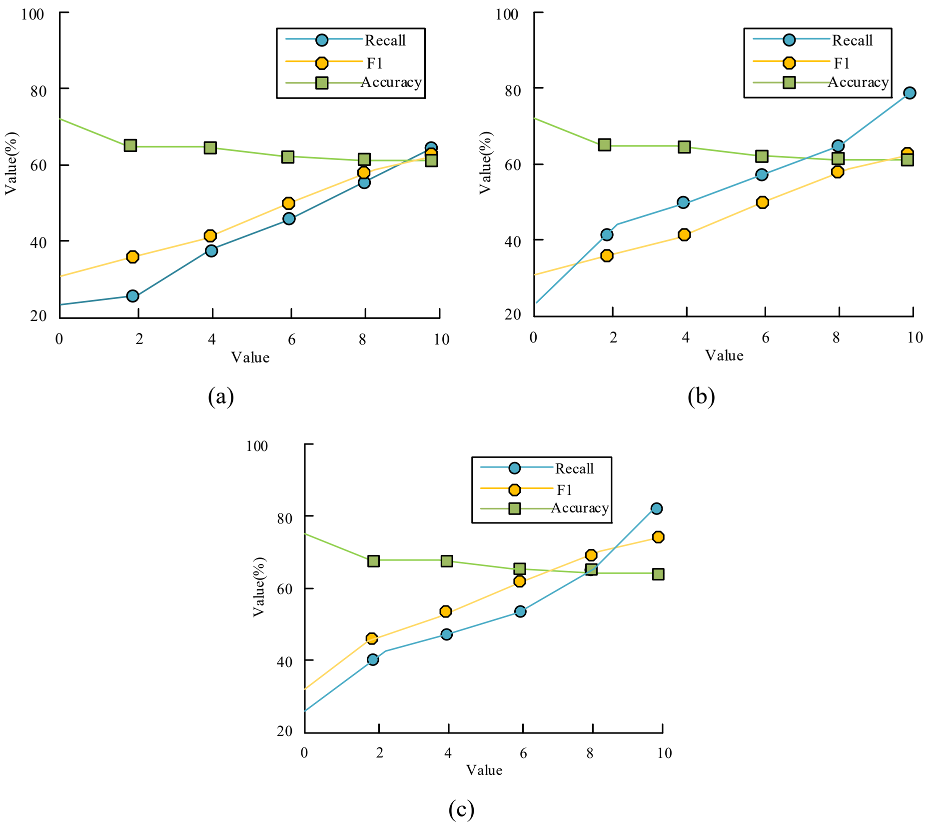


**Figure 7** Test set results (a) accuracy (b) recall (c) F1 (see online version for colours)

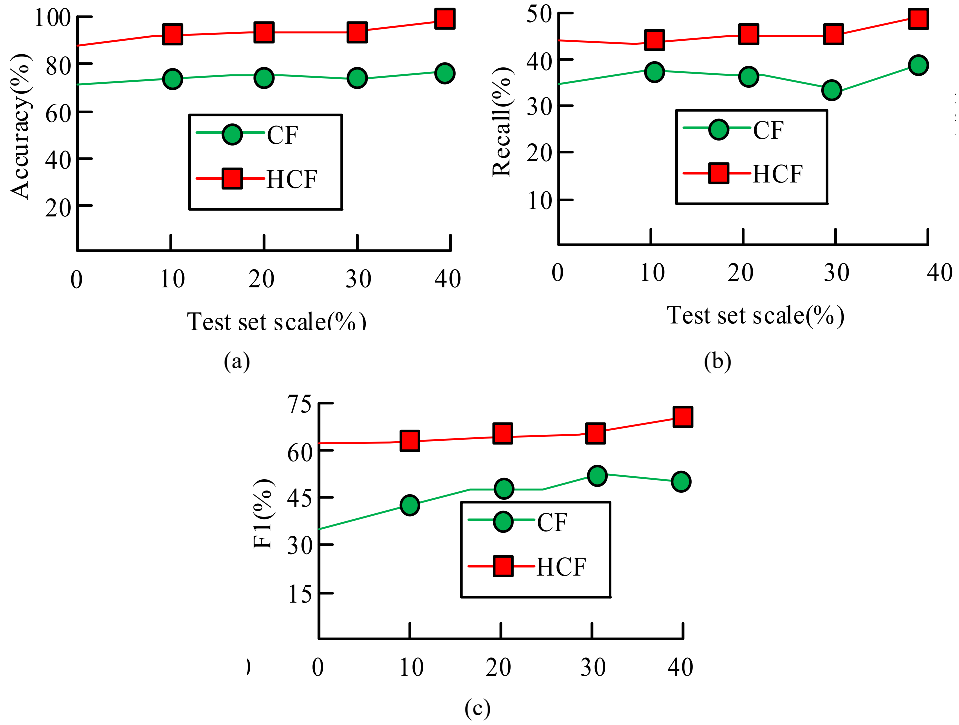
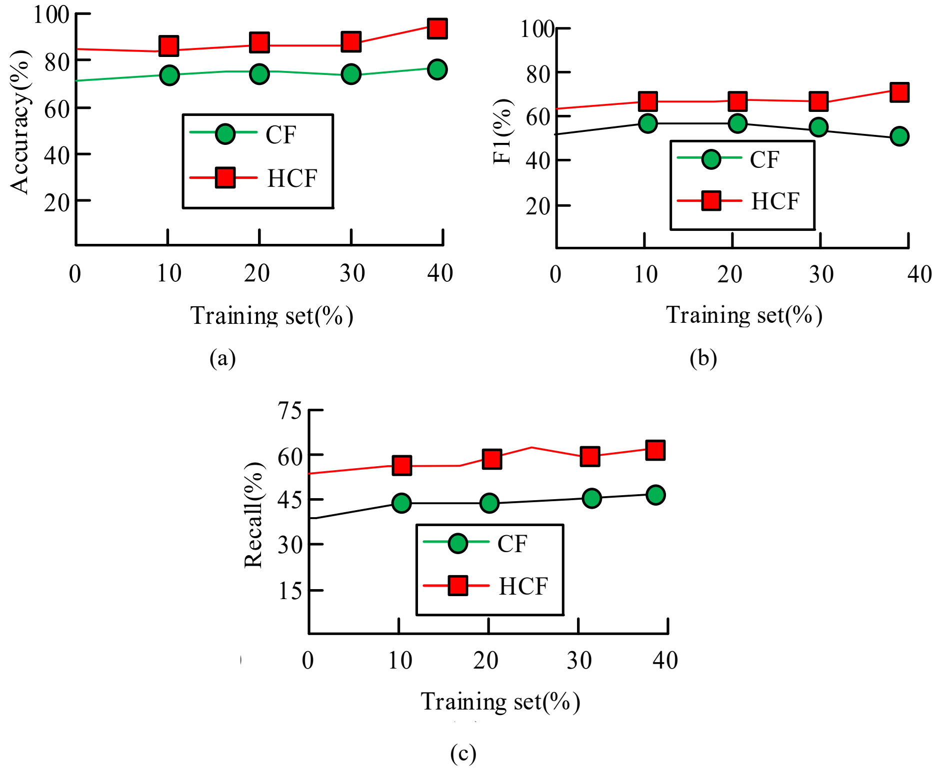



**Figure 8** Training set results (a) accuracy (b) recall (c) F1 (see online version for colours)




Therefore, the academic performance of students in both groups improved to a certain extent, and the learning efficiency of group A1 was significantly lower than that of group A2. t-test results between groups A1 and A2 are shown in Table 3.

From Table 3, we can see that the mean score of the pre-test of group A1 was 71.81 and the mean score of group A2 was 71.71, which were relatively close to each other. The significance value of the difference between group A1 was 0.977 (> 0.05), which means that there was no significant difference between the mean score of group A1 and group A2. However, the difference between the post-test means of the two groups was greater, with the mean post-test score of the A1 group being 75.71, much lower than that of the A2 group, with a significant difference of 0.003 (< 0.05), indicating that there was a significant difference between the mean post-test scores of the A2 and A1 groups.

## 5 Conclusions

The collaborative filtering algorithm was improved under cognitive diagnosis to optimise the recommendation service for English teaching resources. The experimental results show that with the gradual increase of the parameter $\beta$, the accuracy rate of the test set all gradually decreases and the recall rate gradually increases, which proves the effectiveness of the exercise recommendation method; in the test set experiments, the accuracy rate, F1 value and recall rate of the hybrid collaborative filtering algorithm under cognitive diagnosis are stable above 90%, 40%~50% and 40%~60% respectively, with the accuracy rate as high as 98%, while the collaborative filtering In the training set experiments, the F1 value of the hybrid collaborative filtering algorithm was stable at 60%~70%, the accuracy was stable at 80%~95% and the recall was stable at 50%~65%, all of which were better than the collaborative filtering algorithm, indicating that the algorithm used in the study could improve the recommendation accuracy and the recommendation effect; in practical application, the results of the within-group The t-test results showed that the mean values of the pre-test and post-test of the A1 group were 71.81 and 75.71 respectively, while the post-test of the A2 group was 11.31 higher than that of the pre-test, which showed that the academic performance of students in both groups had improved, and the learning efficiency of the A2 group was significantly higher than that of the A1 group. The mean score of 83.020 was much higher than that of group A1, and there was a significant difference between the mean post-test values of group A2 and group A1, proving that the algorithm has a high accuracy rate, can achieve personalised recommendation services for English teaching resources, and can improve learning outcomes. However, the study did not evaluate the results in combination with indicators such as novelty and satisfaction of the recommendation results, so the recommendation result indicators need to be enriched for evaluation.

## Acknowledgements

The research is supported by: the national social science foundation '13th FIVE-YEAR PLAN' educational project of the year 2017,' Constructional and Applicational Research of Intelligent 3D Network Learning Space ' (project number: BCA170078).

## References

Azmi, A.K. (2020) 'A hybrid knowledge-based and collaborative filtering recommender system model for recommending interventions to improve elderly wellbeing', *International Journal of Advanced Trends in Computer Science and Engineering*, Vol. 9, No. 4, pp.4683–4689.

Baskaran, S. and Murugan, K. (2019) 'A study on collaborative filtering for hybrid recommender system for service discovery', *International Journal of Scientific Research in Computer Science Engineering and Information Technology*, Vol. 2019, No. 3, pp.235–240.

Dang, M.P., Dao, T.P., Chau, N.L. et al. (2019) 'Effective hybrid algorithm of taguchi method, FEM, RSM, and teaching learning-based optimisation for multiobjective optimization design of a compliant rotary positioning stage for nanoindentation tester', *Mathematical Problems in Engineering*, Vol. 2019, pp.1–16.

Dong, S. (2020) 'Intelligent English teaching prediction system based on SVM and heterogeneous multimodal target recognition', *Journal of Intelligent and Fuzzy Systems*, Vol. 38, No. 153, pp.1–10.

Hattori, S., Misawa, R., Ishikawa, H. et al. (2017) 'Hybrid recommender system based on collaborative filtering employing personal values-based user model', *Journal of Japan Society for Fuzzy Theory and Intelligent Informatics*, Vol. 29, No. 4, pp.628–636.

Heeryong, N., Seulbi, C. et al. (2017) 'Social network-based hybrid collaborative filtering using genetic algorithms', *Journal of Intelligence and Information Systems*, Vol. 23, No. 2, pp.19–38.

Li, H. and Han, D. (2020) 'A novel time-aware hybrid recommendation scheme combining user feedback and collaborative filtering', *Mobile Information Systems*, Vol. 2020, No. 1, pp.1–16.

Li, X. and Li, D. (2019) 'An improved collaborative filtering recommendation algorithm and recommendation strategy', *Mobile Information Systems*, Vol. 2019, No. 13, pp.1–11.

Liu, Q., Wu, R. et al. (2018) 'Fuzzy cognitive diagnosis for modelling examinee performance', *ACM Transactions on Intelligent Systems and Technology*, Vol. 9, No. 4, p.296.

Liu, Y., Wang, S., Khan, M.S. et al. (2018) 'A novel deep hybrid recommender system based on auto-encoder with neural collaborative filtering', *Big Data Mining and Analytics*, Vol. 1, No. 3, pp.211–221.

Milli, M. and Bulut, H. (2017) 'The effect of neighborhood selection on collaborative filtering and a novel hybrid algorithm', *Intelligent Automation and Soft Computing*, Vol. 23, No. 2, pp.261–269.

Ortega, F., Rojo, D., Valdiviezo-Diaz, P. et al. (2018) 'Hybrid collaborative filtering based on users rating behavior', *IEEE Access*, Vol. 6, No. 99, pp.69582–69591.

Rao, P. and Mounika, A. (2018) 'Recommended system of collaborative filtering based in social networks', *Neural, Parallel and Scientific Computations*, Vol. 26, Nos. 3–4, pp.247–256.

Rojo, D., Ortega, F., Valdiviezo-Diaz, P. et al. (2018) 'Hybrid collaborative filtering based on users rating behavior', *IEEE Access*, Vol. 6, No. 99, pp.69582–69591.

Sharma, N., Kolekar, M.H., Jha, K. et al. (2019) 'EEG and cognitive biomarkers based mild cognitive impairment diagnosis', *Innovation and Research in Biomedical En*, Vol. 40, No. 2, pp.113–121.

Tm, A., Amb, A., Re, B. et al. (2019) 'Efficient clustering in collaborative filtering recommender system: hybrid method based on genetic algorithm and gravitational emulation local search algorithm', *Genomics*, Vol. 111, No. 6, pp.1902–1912.

Wang, X., Dai, Z., Li, H. et al. (2021) 'Research on hybrid collaborative filtering recommendation algorithm based on the time effect and sentiment analysis', *Complexity*, Vol. 2021, No. 2, pp.1–11.

Xiong, R., Wang, J., Zhang, N. et al. (2018) 'Deep hybrid collaborative filtering for web service recommendation', *Expert Systems with Applications*, Vol. 110, No. 11, pp.191–205.

Yang, S. (2021) 'Innovative design and application of mind mapping in English writing class –taking English teaching of senior grade in primary school as an example', *Education Study*, Vol. 3, No. 3, pp.405–409.