



Research article

Online learning resources recommendation model based on improved NSGA-II algorithm

Hui Li¹, Rongrong Gong¹, Pengfei Hou¹, Libao Xing¹, Dongbao Jia^{1,*} and Haining Li^{2,*}

¹ School of Computer Engineering, Jiangsu Ocean University, Lianyungang, China

² Department of Neurology, General Hospital of Ningxia Medical University, Yinchuan, China

* **Correspondence:** Email: dbjia@jou.edu.cn, lhnwww@126.com.

Abstract: Due to the characteristics of online learning resource recommendation such as large scale, uneven quality and diversity of preferences, how to accurately obtain various personalized learning resource lists has become an urgent problem to be solved in the field of online learning resource recommendation. This paper proposes an online learning resource recommendation model based on the improved NSGA-II algorithm, which integrates the Tabu search algorithm to improve the local search ability of NSGA-II algorithm. It takes background fitness, cognitive fitness and diversity as the objective functions for optimization. The dynamic updating of crowding degree is used to avoid the risk that the individuals with low crowding degree in the same area are deleted at the same time. Meanwhile, an adaptive genetic algorithm is applied to assign the optimal crossover rate and the mutation rate according to individual adaptability level, which ensures the convergence of genetic algorithm and the diversity of population. The experimental results show that the proposed model is superior to the traditional recommendation algorithm in terms of accuracy index, mean fitness, recall rate, F1 mean, HV, GD and IGD, etc., thus verifying the feasibility and effectiveness of the algorithm.

Keywords: Tabu search; NSGA-II algorithm; learning resources recommendation

1. Introduction

With the spread of COVID-19, the emergence of online learning [1] platform has effectively solved the dilemma that universities and educational institutions cannot carry out offline teaching and ensured the safety and reliability of learning scenarios, which has become an important way for people

to acquire knowledge. How to use the Internet and various artificial intelligence technologies to select the appropriate resource list from massive learning resources and accurately recommend the learning content to provide personalized learning experience for people has become one of the research hotspots in the field of online learning.

Personalized recommendation system can not only satisfy the preferences of learners, but also effectively improve the satisfaction of online learning. The recommendation system [2] generates the appropriate resource recommendation list mainly according to learning time [3], knowledge [4], learning preference [5], demand [6] and social situation [7]. The methods of learning resource recommendation can be basically divided into content-based resource recommendation [8], collaborative filtering-based resource recommendation [9], knowledge-based resource recommendation methods [10] and session-based resource recommendation [11]. The content-based learning resource recommendation makes similarity recommendation by comparing the characteristics between learners and resources, which is intuitive but ignores the correlation between resources. The collaborative filtering-based learning resource recommendation makes recommendation through the similarity matrix between learners or resources. This method has the defects of cold start and high time complexity, and the calculation method of similarity will affect the results. The knowledge-based learning resource recommendation considers the relevant knowledge of learners and resources, and matches the knowledge demands of learners with resource knowledge. This method conforms to the law of psychology and can save a lot of learning time, but its recommendation process requires large amount of manual intervention, which has subjective influence. The session-based learning resource recommendation relies on the interaction information between learners and resources, which can effectively improve the recommendation accuracy of resources. However, it requires the internal structure processing and the modeling of relationship between sessions on session data, and there are few researches on this aspect.

Learners have multiple attributes, and there are associations between resources, while users' demands for learning resources are multi-dimensional. For example, different users have different requirements on cognitive ability, media preference, content preference and other aspects. Therefore, the problem of learning resource recommendation with multidimensional constraints can be regarded as an NP-hard problem, and the main solutions include ant colony algorithm [12], genetic algorithm [13] and particle swarm optimization algorithm [14]. In addition, the ϵ -constraint method [15] and the minimum-maximum method [16] solve the multi-objective optimization problem by constraining other objective functions and optimizing one objective function. These research methods all need to set complex weights. In order to avoid the above problems, NSGA-II and its improved algorithm [17–25], as a solution to the multi-objective optimization problem, have successfully solved the multi-objective optimization problem in several research fields. However, there are few related researches in the field of online learning resource recommendation.

From the view of balancing the various demands of learners, this paper proposed a learning resource recommendation model based on the improved NSGA-II algorithm. The learning resources were filtered with background fitness, cognitive fitness and diversity as objective functions. The Tabu search algorithm was applied to improve local search ability. At the same time, the dynamic updating of crowding degree and the adaptive changing of crossover rate and mutation rate were adopted to ensure individual diversity. The main contributions of this paper are as follows:

- 1) In order to solve the problem that low-density individuals are easy to be deleted and homogeneous populations, dynamic updating crowded distance and adaptive crossover mutation

operators are proposed.

2) According to the multi-dimensional needs of learners, a three-objective optimization model of background fitness, cognitive fitness and diversity is established.

3) While retaining the advantages of individual diversity and uniformity of NSGA-II algorithm, Tabu search strategy is added to improve the search ability of solution set and increase the diversity.

The rest of this paper is structured as follows: Section 2 describes the algorithm and modeling process of this paper; Section 3 details the solution process of model. Section 4 discusses the experiments and results. Section 5 presents conclusion and future research direction.

2. Problem description and model framework

This section first makes conventions on the symbols used in the full text, and formalizes the selection of personalized learning resources, including the necessary symbol definition. After that, the factors that affect the results of personalized resource recommendation are briefly analyzed to clarify the rationality of the three factors considered in this paper, i.e., background fitness, difficulty fitness and content preference.

2.1. Problem description

The online resource recommendation technology in online teaching judges the preferences of students according to their pre-test characteristics, so as to provide students with resources that meet their demands [26]. The characteristics of students in online teaching mainly include age, gender, educational background, cognitive ability and other information, as well as the acceptable learning time of learners. The characteristics of online resources mainly include the average age of learning groups [27], female proportion, educational threshold, difficulty level, the feature of resource content and time to complete learning resources.

This paper proposes an online learning resource recommendation model based on improved NSGA-II algorithm, which does not convert the multi-objective problem into a single objective problem by the weighted method, but uses the multi-objective method to improve the online learning resources recommendation. The improved NSGA-II algorithm adopts the ternary combination mode, and takes background fitness, cognitive fitness and diversity as the selection targets.

Definition 1. Characteristics of learner: The personalized characteristics of learners include background information (BI), cognitive ability (CA) and time limit (T). The learner set is expressed as $U = \{u_1, u_2, \dots, u_{|U|}\}$, where the $|U|$ represents the size of learner set. For a specific learner u_i , its background information (BI_i) is defined as $BI_i = \{AI, GI, EI\}$, which is composed of age information (AI), gender information (GI) and education Information (EI). According to the relevant researches on the influencing factors of cognitive ability in recent years, it is concluded that learners' cognitive ability in the field of education has relation to their age [28], gender [29] and educational background [30].

Definition 2. Characteristics of learning resource: The characteristics of learning resources include preference information (PI), difficulty information (DI), content information (CI), and time (t). The learning resource set is expressed as $L = \{LO_1, LO_2, \dots, LO_{|LO|}\}$, of which $|LO|$ denotes the size of learning resource set. The preference information (PI) of specific learning resource LO_i is defined as

$PI_i = \{MA, WP, UP\}$, which is composed of mean age (MA) of learner group, women proportion (WP) and undergraduates percentage (UP).

Definition 3. Content information of learning resource: content information (CI) of learning resources is expressed as $CI = \{PD, AD, ND\}$, which determines the selection quality of learning resource recommendation list. It is codetermined by popularity degree (PD), authority degree (AD) and novelty degree (ND). The background fitness (BF) between learners and learning resources is decided by learners' background information (BI) and learning resources' preference information (PI). The difficulty fitness (DF) of learner to learning resources is determined by the cognitive ability (CA) of learners and the difficulty information (DI) of learning resources (DI).

The background fitness, cognitive fitness and diversity screening of learner affect the final resource recommendation list from three dimensions. Based on the above three characteristics, a multi-dimensional online learning resource recommendation model combined with time constraints is constructed. According to the background fitness and the difficulty fitness of learners, the diversity of resources is regarded as the auxiliary screening objective to provide suitable online learning resources list to learners.

2.2. Construction of multi-objective recommendation model

Aimed at the above problems, an online learning resource recommendation model based on the improved NSGA-II algorithm is proposed, as shown in Figure 1.

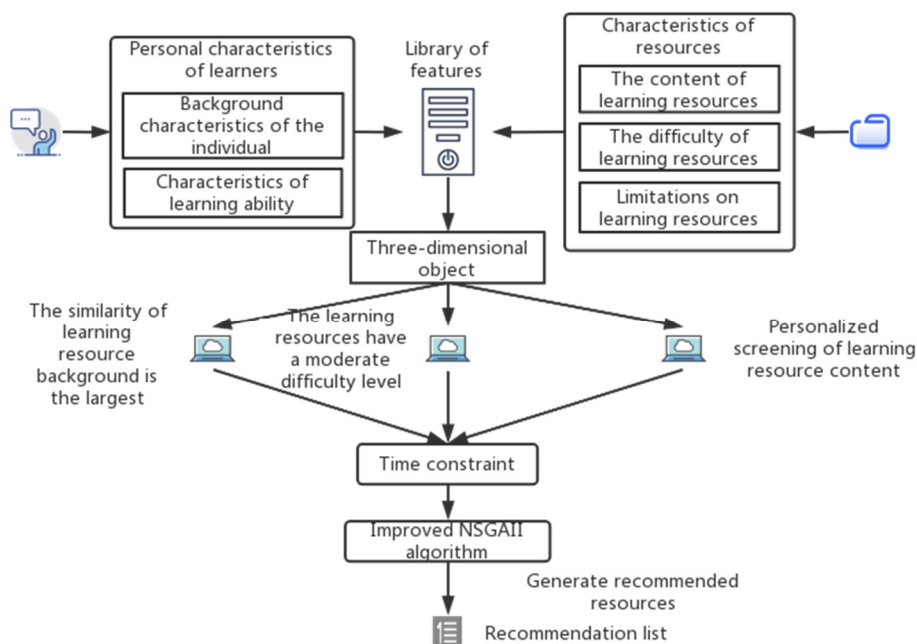


Figure 1. Learning resource recommendation.

In the above recommendation model, the three-dimensional objectives of resource screening include background fitness, cognitive fitness and diversity. The screening conditions of diversity

contain popularity degree, authority degree and novelty degree. The three-dimensional objectives are constrained by learning time limit. The three-dimensional objectives and time constraints jointly determine the final result, the higher the fitness of result, the better the experience of learners.

2.3. Mathematical modeling of multi-objective programming model

The background characteristics of learners include age, gender and education, which affect the choice of learning resources to a large extent. The background information is expressed as $BI = [BI_1, BI_2, \dots, BI_{|U|}]$, where BI_i is a three-dimensional vector. The background characteristics of the i th learner are represented by $BI_i = \{AI_i, GI_i, DI_i\}$. The preference information of learning resources are described by $PI = [PI_1, PI_2, \dots, PI_{|LO|}]$, of which PI_j is a three-dimensional vector, $PI_j = [MA_j, WP_j, UP_j]$ shows the preference information of the j th learning resource. The background fitness of the i th learners to the j th learning resource is formulated as:

$$BF_{ij} = sim(BI_i, PI_j) = \omega_1 \frac{|AI_i - MA_j|}{\min(AI_i, MA_j)} + \omega_2 |GI_i - WP_j| + (1 - \omega_1 - \omega_2) |DI_i - UP_j| \quad (1)$$

where, AI_i indicates the age information of learners, values for actual age. GI_i figures the gender information of learner, with value of 0 or 1, 0 for female and 1 for male. DI_i denotes the education information of learners. The value can be 0 or 1, indicating that the education of learner is below the bachelor degree and above the bachelor degree, respectively. MA_j represents the mean age of learner that choose the resource. WP_j shows the proportion of women that study the resource, and UP_j represents the undergraduate proportion of the population studying the resource.

The cognitive ability of learner is expressed as $CA = [CA_1, CA_2, \dots, CA_{|U|}]$, where CA_i shows the learning ability of the i th learner, value for $CA_i \in (0,1)$. The difficulty of learning resource is related to the learning stage and the percentage of people who have got achievement. The difficulty DI_j of the j th learning resource can be calculated by

$$DI_j = \frac{\text{phase}_j \times \text{com}_j}{\text{Max}(\text{com})} \quad (2)$$

where, phase_j represents the course duration of learning resource, $\text{phase}_j \in [0,4]$ and phase_j is an integer. com_j denotes the number of people who have completed the learning task and successfully took the exam; $\text{Max}(\text{com})$ indicates the maximum number of participants for all learning resources. The difficulty fitness DF_{ij} of learners is the difference between the cognitive ability of the i th learner CA_i and the difficulty of the j th learning resource D_j :

$$DF_{ij} = CA_i - D_j \quad (3)$$

If the difficulty of learning resources exceeds the cognitive ability of learners too much, learners will not be able to learn smoothly. If the difficulty of learning resources is too simple, there will be time waste of learners. Therefore, the difficulty of learning resources should be as close as possible to the cognitive limit of learners, the calculation result $|DF_j|$ the closer to zero, the better.

The content preference CP of learning resource is determined by its popularity degree, authority degree and novelty degree. The popularity degree of learning resources is related to the percentage of participants, the viewing frequency and score. The percentage of learning participants, the viewing frequency and the comprehensive score of resources can be expressed as:

$$f_1(LO_j) = \frac{S_j}{\text{Max}(S)} \quad (4)$$

$$f_2(LO_j) = \frac{V_j}{\text{Max}(V)} \quad (5)$$

$$f_3(LO_j) = \frac{1}{n} \sum_{i=1}^n (P_{ij} + \overline{P_{LO_j}} - \overline{P_{u_i}}) \quad (6)$$

where, S_j indicates the participant number of learning resource LO_j that exceeds 50%. $\text{Max}(S)$ indicates the maximum visits time of learning resource. V_j represents the viewing times of learning resource LO_j . $\text{Max}(V)$ denotes the maximum viewing times of resource. P_{ij} shows the score of i th learner for the j th learning resource. $\overline{P_{LO_j}}$ is the average score of learning resources. $\overline{P_{u_i}}$ is the average score of all learners.

The popularity degree (PD) of learning resource is calculated by weighted $f_1(LO_j)$, $f_2(LO_j)$ and $f_3(LO_j)$ in the above formula, with formula as follows:

$$PD(LO_j) = \omega_3 f_1(LO_j) + \omega_4 f_2(LO_j) + (1 - \omega_3 - \omega_4) f_3(LO_j) \quad (7)$$

The authority of learning resources will affect the trust of learner in the resources. The authoritative resources can improve the learning motivation of learner. The influences of issuing institution and lecturer affect the authority of resources. H-index is a method to evaluate academic achievements. The H-index of a scientific researcher means that he has H papers cited at least H times. The influence of institution can be related to the H-index of its staff, so the influence of agency can be expressed as:

$$f_4(LO_j) = \frac{H_P - \text{Min}(H_P)}{\text{Max}(H_P) - \text{Min}(H_P)} \quad (8)$$

where, H_p represents the H-index of institution. $\text{Min}(H_p)$ and $\text{Max}(H_p)$ are the minimum and maximum H-index of the staff in institution. The influence of resource lecturers is expressed as:

$$f_5(LO_j) = \frac{H_a - \text{Min}(H_a)}{\text{Max}(H_a) - \text{Min}(H_a)} \quad (9)$$

where, H_a is the H-index of resource lecturers. $\text{Min}(H_a)$ and $\text{Max}(H_a)$ represent the minimum and maximum H-index of lecturers. Their H-index is closely related to the citation rate of their published scientific papers, which can be found and calculated from research platforms like CNKI. The authority degree AD of learning resources is expressed as:

$$AD(LO_j) = \omega_5 f_4(LO_j) + (1 - \omega_5) f_5(LO_j) \quad (10)$$

Novel learning resources can improve the positivity of learners. The novelty degree of learning resources $N(LO_j)$ is related to the release time of resource Time_{LO_j} , and the novelty degree of resources is expressed as:

$$ND(LO_j) = \frac{\text{Time}_{LO_j} - E(\text{Time})}{T - E(\text{Time})} \quad (11)$$

where, $E(\text{Time})$ represents the earliest release time of the knowledge point contained in the learning resource, and T represents the current year. By integrating the PD, AD and ND of learning resources, the content selection of learning resources can be expressed as:

$$Cl_{ij} = \omega_6 PD(LO_j) + \omega_7 AD(LO_j) + (1 - \omega_6 - \omega_7) ND(LO_j) \quad (12)$$

3. Model solution

In this paper, a new hybrid algorithm is proposed, which combines the Tabu search algorithm with excellent local search efficiency with the NSGA-II algorithm. The children obtained from the Tabu search algorithm are inserted into the population selection of NSGA-II algorithm to enhance the diversity of population. The basic idea of NSGA-II algorithm is to sort the population by non-dominated sorting and calculate the crowding distance to maintain the diversity of population. The improved NSGA-II algorithm takes the resources as the parent population P_0 with the size of N , and carries out non-dominated sorting on them. Each resource individual obtains the fitness, then doing tournament selection and genetic operator operation on each resource individual to generate a child population G_t with the size of N .

In the elitist strategy, the parent population P_t of generation t , the children population G_t generated by crossover mutation and the new individual T_t generated by Tabu search are combined into the new population R_t . The non-dominated Pareto solution set with decreasing rank is obtained by fast non-dominated sorting of R_t , expressed as F_1, F_2, \dots . The first N solutions are selected into the population P_{t+1} until there are $F_1 + F_2 + \dots + F_i \geq N$, where F_i is called the critical layer. Calculating the crowding distance of the critical layer, the individuals with high crowding degree is preferentially selected until the maximum number of iterations. The flow chart is shown in Figure 2.

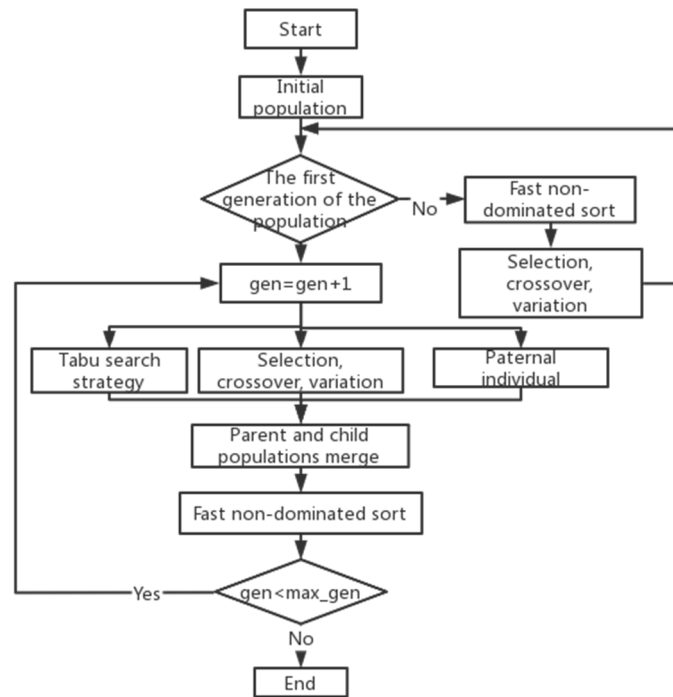


Figure 2. Flowchart of the improved NSGA-II algorithm.

3.1. Calculation of crowding distance

The sorting mechanism of NSGA-II algorithm is based on the calculation of crowding distance of individual resources. Crowding distance represents the density of resource solutions near a specific resource individual. In order to maintain diversity, individuals with large crowding distance at the same layer are preferentially selected. Although this strategy can retain individuals with low crowding degree, there is a risk that individuals with low crowding degree in the same area will be deleted at the same time.

In order to solve the above problems, the dynamical updating of crowding distance is adopted. The calculation process is as follows:

Step 1: Calculate the Euclidean distance between adjacent individuals. Let the crowding distance between the k th individual and the $k+1$ th individual be:

$$I[k, k + 1]_{\text{distance}} = \sqrt{\sum_1^m \left(\frac{f_i(k) - f_{i_{\min}}}{f_{i_{\max}} - f_{i_{\min}}} - \frac{f_i(k+1) - f_{i_{\min}}}{f_{i_{\max}} - f_{i_{\min}}} \right)^2} \quad (13)$$

where, m is the target number, $f_i(k)$ is the value of the k th individual on the i th target. $f_{i_{\min}}$ is the minimum value of the i th target, and $f_{i_{\max}}$ is the maximum value of the i th target.

Step 2: Find the minimum crowding distance $I[k, k + 1]$ and compare $I[k - 1, k]$ with $I[k +$

1, $k + 2$]. If $I[K - 1, k]$ is smaller, select individual $k + 1$; otherwise, select individual k .

Step 3: Update the crowding distance.

Step 4: Determine whether the number of individuals is greater than or equal to N , otherwise return to Step 2.

3.2. Crossover mutation operator

When the genetic algorithm approaches the termination generation, there will be duplicate individuals with good fitness in the mating pool, and the population is close to homogeneity. Although the mutation operation can improve this situation, traditional mutation operations would set a fixed value and the probability of mutation is very small. Even with the mutation operation, premature population convergence will occur.

The performance of genetic algorithm is closely related to the selection of crossover rate H_c and mutation rate H_m . Therefore, an adaptive genetic algorithm (AGA) is proposed. The optimal crossover rate H_c and mutation rate H_m are allocated adaptively according to the adaptive level of individuals. This algorithm ensures the convergence of genetic algorithm and the diversity of population:

$$H_c = \begin{cases} H_{c_{\min}} + \frac{(H_{c_{\max}} - H_{c_{\min}})(\text{Gen}_{\max} - \text{Gen}')}{2(\text{Gen}_{\max} - \text{Gen}_{\text{avg}})}, & \text{Gen}' > \text{Gen}_{\text{avg}} \\ \frac{H_{c_{\min}} + H_{c_{\max}}}{2} + \frac{(H_{c_{\max}} - H_{c_{\min}})(\text{Gen}_{\text{avg}} - \text{Gen})}{2(\text{Gen}_{\text{avg}} - \text{Gen}_{\min})}, & \text{Gen}' \leq \text{Gen}_{\text{avg}} \end{cases} \quad (14)$$

$$H_m = \begin{cases} H_{m_{\min}} + \frac{(H_{m_{\max}} - H_{m_{\min}})(\text{Gen}_{\max} - \text{Gen}')}{2(\text{Gen}_{\max} - \text{Gen}_{\text{avg}})}, & \text{Gen}' > \text{Gen}_{\text{avg}} \\ \frac{H_{m_{\min}} + H_{m_{\max}}}{2} + \frac{(H_{m_{\max}} - H_{m_{\min}})(\text{Gen}_{\text{avg}} - \text{Gen})}{2(\text{Gen}_{\text{avg}} - \text{Gen}_{\min})}, & \text{Gen}' \leq \text{Gen}_{\text{avg}} \end{cases} \quad (15)$$

where, Gen_{avg} represents the mean remaining generations of individuals. Gen_{\max} and Gen_{\min} represent the minimum and maximum remaining generations of the population, respectively; Gen' denotes the larger residual generation of the two individuals participating in the crossover; Gen expresses the remaining generations of individuals involved in the mutation; $H_{c_{\min}}$ and $H_{c_{\max}}$ refer to the minimum and maximum crossover rates set by the algorithm, respectively. $H_{m_{\min}}$ and $H_{m_{\max}}$ are the minimum and maximum mutation rates set by the algorithm, respectively.

3.3. Tabu search strategy

The candidate solutions generated by Tabu search algorithm during the search process are not random, which can accept sub-optimal solutions, so it has strong search ability. Therefore, the advantages of individual diversity and uniformity in NSGA-II algorithm are retained while maintaining the genetic operator operation of NSGA-II algorithm. The individual variables in this paper are generated from the neighborhood of current solution, rather than randomly generated. When the optimal solution meets the aspiration criterion, we have $P_m = 1$, and the current individual is added

to the children population as a mutant; otherwise, $P_m = 0$.

The solution set after merging is set as $R_t = P_t \cup G_t \cup T_t$, where P_t is the parent individual, G_t is the individual obtained by genetic operator, and T_t is the individual obtained by Tabu search operator. Secondly, the first N individuals are selected as the next generation parent individuals R_{t+1} by non-dominated sorting and crowding distance. Through this operation, the advantage of NSGA-II algorithm that avoids losing excellent individuals in the parent generation can be retained and the local search ability can be improved, so as to reduce the number of duplicate individuals in the offspring and enhance the diversity of solution set.

The Tabu search strategy is as follows:

1) Initial solution

Tabu search algorithm is highly dependent on the initial solution, so the initial solution is randomly selected from the individuals whose parent individual non-dominated order equal to 1.

2) Neighborhood solution

During the initial iteration, the search scope should be enlarged to increase the probability of finding the optimal solution. As the iteration continues, the obtained solution gradually approaches the optimal solution, so the search scope should be narrowed to avoid repeated search. The objective functions (1), (3) and (12) are used as fitness functions to calculate the fitness values of each candidate solution.

3) Candidate solutions

For a specific resource LO_j , it takes a lot of computation to calculate all the neighborhood solutions, so some strategies are required for choosing the candidate solutions. In this paper, a number of solutions $N(X)$ are selected from a given set of candidate solutions by using non-dominated sorting and crowding operators.

4) Aspiration criterion

The solution with the best fitness is selected from the candidate solutions and compared with the current optimal solution. If it is better than the current optimal solution, the best candidate solution is applied to update the current optimal solution regardless of whether it is taboo or not. It is also used as the current solution for the next iteration. Then, the corresponding operation is inserted to the taboos table. If it is not better than the current optimal solution, the optimal solution that is not in the Tabu state is selected from all candidate solutions as the new current solution, and then the corresponding operation is added to the Tabu table.

4. Experimental performance and analysis

In order to evaluate the efficiency of the proposed course personalization algorithm, three evolutionary algorithms (NSGAI, IBEA, PESA2) are compared with the proposed algorithm, so as to verify the effectiveness of the personalized learning resource recommendation proposed in this paper.

4.1. Dataset

The study was based on 290 online courses from Harvard and MIT on the edX platform. The dataset contains 250,000 certifications, 4.5 million participants, and 28 million hours of participant data. The feasibility and accuracy of the proposed algorithm are verified on this dataset. The dataset

of this online public course contains data like issuing institution, course number and release date, which fully meet the research requirements of this paper. Table 1 shows the description of parameters in the edx dataset, and Table 2 presents the information of four datasets generated by course subject according to the edx dataset, denoted as S1, S2, S3 and S4, respectively.

Table 1. Description for the parameters generated by edx dataset.

parameter	instructions
Institution	Issuing agency
Course number	Course ID
Launch Date	Creation time
Course title	Course title
Instructors	Keynote Speaker
Course Subject	Course subject
Year	Duration of Course
Audited	Students who learn more than 50% of their progress
Participants	Number of visitors
Played Video	Number of viewers
%Posted in Forum	Percentage of forum discussion
Median age	Average age of users
%female	Percentage of women
Total course hours	Total course hours
%bachelor's degree or higher	The proportion of users with bachelor degree or above
M	Scale of learning resources
N	Learner scale

Table 2. Course dataset.

NO.	Course Subject	M	N
S1	Computer Science	30	171,502
S2	Government, Health, and Social Science	75	240,794
S3	Humanities, History, Design, Religion, and Education	94	163,729
S4	Science, Technology, Engineering, and Mathematics	91	163,235

4.2. Evaluation index

This section introduces the performance evaluation index of the algorithm, which mainly evaluates learners' satisfaction with the resource list. We determined the weight value through

comparative analysis of several experiments, and then compared the accuracy, mean fitness, recall rate, F1 mean, HV, GD and IGD of each algorithm in different datasets.

1) Accuracy refers to the proportion of resources in a given resources recommendation list that can correctly predict the demand of learners. The computation formula of accuracy is as follows:

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN} \quad (16)$$

2) Mean fitness refers to the fitness between learners and learning resources. The mean fitness with different scales reflects the satisfaction of learners with the recommendation list. The computation formula is as follows:

$$\text{Fitness} = \frac{\sum_{i=1}^m \sum_{j=1}^n (BF_{ij} + DF_{ij} + CI_{ij})}{nm} \quad (17)$$

3) Recall rate refers to the proportion of samples that accurately predict resources required by learners. The computation formula is as follows:

$$\text{Recall} = \frac{TP}{TP+FN} \quad (18)$$

4) F1 value is used to better represent the accuracy rate and recall rate of learning resource recommendation. The computation formula is as follows:

$$F1 = \frac{2 \times \text{Precise} \times \text{Recall}}{\text{Precise} + \text{Recall}} \quad (19)$$

5) GD measures the average minimum distance between each point in solution set P and reference set P. The smaller the GD value is, the better the convergence effect will be. The computation formula is as follows:

$$GD = \frac{\sqrt{\sum_{y \in P} \min_{x \in P^*} \text{dis}(x,y)^2}}{|P|} \quad (20)$$

6) IGD measures the average distance between each reference point and the nearest solution. The smaller the IGD value is, the better the comprehensive performance of the algorithm is. The computation formula is as follows:

$$IGD = \frac{\sqrt{\sum_{y \in P^*} \min_{x \in P} \text{dis}(x,y)^2}}{|P^*|} \quad (21)$$

4.3. Parameter verification experiment

The weight in Eq (1) is the weight factor to calculate the background fitness between learners and resources, which is used to control the proportion of age, gender and educational background when calculating the background similarity. In order to analyze the influence of weight ω_1 and ω_2 on the algorithm in this paper, we calculate the accuracy values under different values of weight ω_1 and ω_2 in the datasets with different sizes. By comparing the accuracy values under three different weight

values in Figure 3, it can be seen that when $\omega_1 = 0.38$ and $\omega_2 = 0.1$, the fitness between learners and resources has nothing to do with gender, and the importance of age and education for learning resources is much higher than gender.

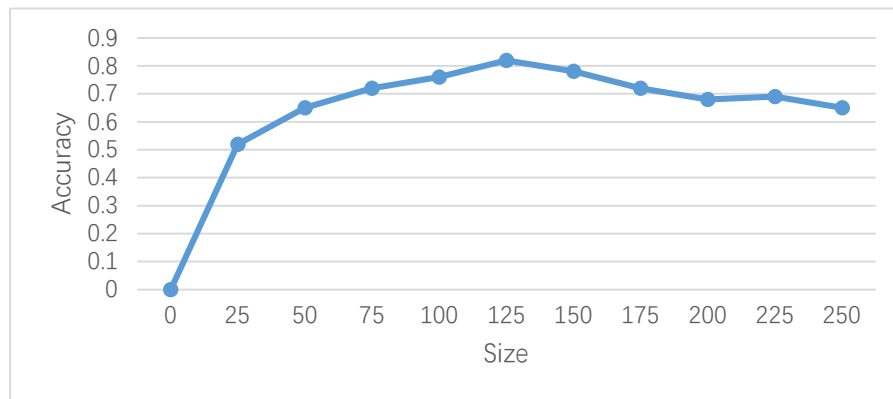


Figure 3. Influence of ω_1 and ω_2 values on the background fitness.

The weight in Eq (7) is the weight factor to calculate the popularity degree of learning resources, which is used to control the proportion of the learning participant percentage, viewing frequency and score when calculating the popularity degree. In order to analyze the influence of weight ω_3 and ω_4 on the algorithm in this paper, we calculate the accuracy values of resources under different values of weight ω_3 and ω_4 in datasets with different size. By comparing the accuracy values under three different weight values in Figure 4, it can be seen that when $\omega_3 = 0.47$ and $\omega_4 = 0.3$, the average popularity degree of learning resources is the highest.

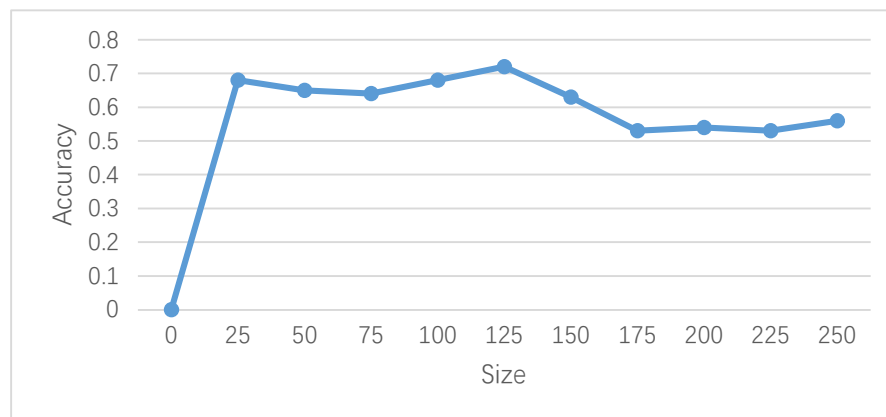


Figure 4. Influence of ω_3 and ω_4 values on the popularity degree.

The weight in Eq (10) is the weight factor to calculate the authority degree of learning resources, which is used to control the proportion of the influence of institutions and the influence of lecturers when calculating the authority degree of resources. In order to analyze the influence of weight ω_5 on the algorithm in this paper, we calculate the accuracy values of resources under different values of weight ω_5 in datasets with different sizes. By comparing the accuracy values under two different

weight values in Figure 5, it can be seen that when $\omega_5 = 0.65$, the average authority degree of learning resources is the highest.

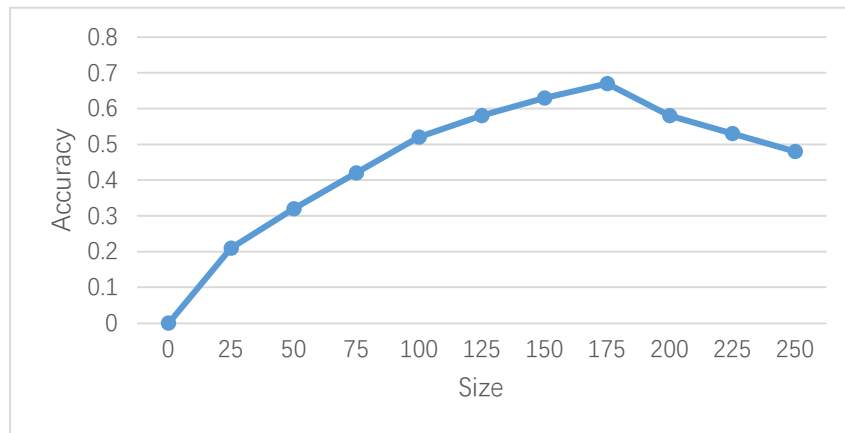


Figure 5. Influence of ω_5 values on the authority degree.

The weight in Eq (12) is the weight factor to calculate the content quality of learning resources, which is used to control the proportion of popularity degree, authority degree and novelty degree when calculating the content quality of learning resources. In order to analyze the influence of weight ω_6 and ω_7 on the algorithm in this paper, we calculated the accuracy value of resources under different values of weight ω_6 and ω_7 in datasets with different sizes. By comparing the accuracy values under three different weight values in Figure 6, it can be seen that when $\omega_6 = 0.22$ and $\omega_7 = 0.52$, the average accuracy value of learning resources is the highest.

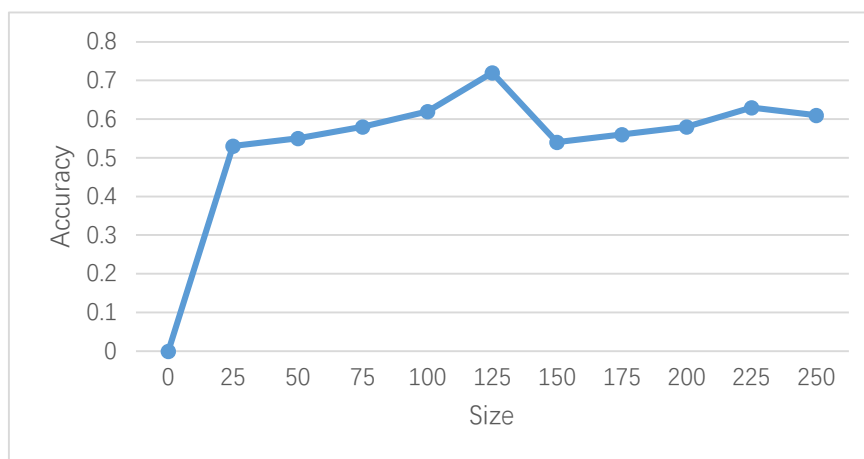


Figure 6. Influence of ω_6 and ω_7 values on the content quality.

4.4. Performance verification

The parameter setting of the multi-objective algorithm is an important step, which has huge impact on the results. Therefore, we configured probabilities for the basic MOEA operations, including mutation and crossover, population size, and generation. Table 3 summarizes the parameter settings of

four algorithms used in the experiment, and Figure 7 shows the Pareto frontier of a certain user.

Each point in Figure 7 represents the recommendation lists for target users. The algorithm is based on the tradeoff between learner fitness and resource diversity, which is contradictory. In order to verify the effectiveness of the algorithm, the proposed algorithm is compared with the other three algorithms. Table 4 shows the performance results of each algorithm in each dataset.

In this paper, 30 independent operation (iterative) experiments were performed to verify the efficiency of four algorithms in the generation of personalized learning resource recommendation list, so as to obtain the average accuracy of the model. For each iteration, we evaluated 10 different data partitions that were cross-validated.

Table 3. Parameter setting.

	Improved NSGA-II	NSGA-II	IBEA	PESA2
Target quantity	2	2	2	2
Number of population	150	150	150	150
Algebra	1000	1000	1000	1000
Crossover rate	0.9	0.9	0.9	0.9
Mutation rate	0.1	0.1	0.1	0.1

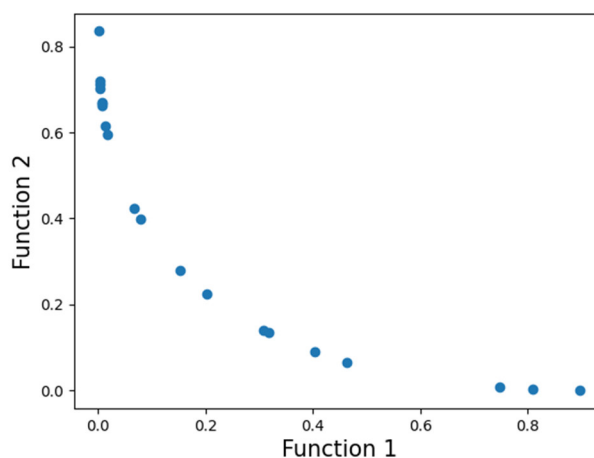


Figure 7. Pareto frontier of a specific user.

Table 4. Performance comparison experiments of algorithms in S1 dataset.

	Accuracy	Recall	F1
Improved NSGA-II	71.3	80.4	75.58
NSGA-II	68.6	78.2	73.09
IBEA	65.3	68.3	66.77
PESA2	66.5	76.8	71.28

Table 5. Performance comparison experiments of algorithms in S2 dataset.

	Accuracy	Recall	F1
Improved NSGA-II	82.4	83.6	83
NSGA-II	78.5	81.5	79.97
IBEA	71.5	75.3	73.35
PESA2	68.8	75.2	71.86

Table 6. Performance comparison experiments of algorithms in S3 dataset.

	Accuracy	Recall	F1
Improved NSGA-II	71.5	82.4	76.56
NSGA-II	68.2	73.6	70.8
IBEA	65.4	75.1	69.92
PESA2	60.1	71.4	62.26

Table 7. Performance comparison experiments of algorithms in S4 dataset.

	Accuracy	Recall	F1
Improved NSGA-II	72.5	65.8	68.99
NSGA-II	69.1	62.7	65.74
IBEA	75.1	62.5	68.22
PESA2	70.5	63.1	66.6

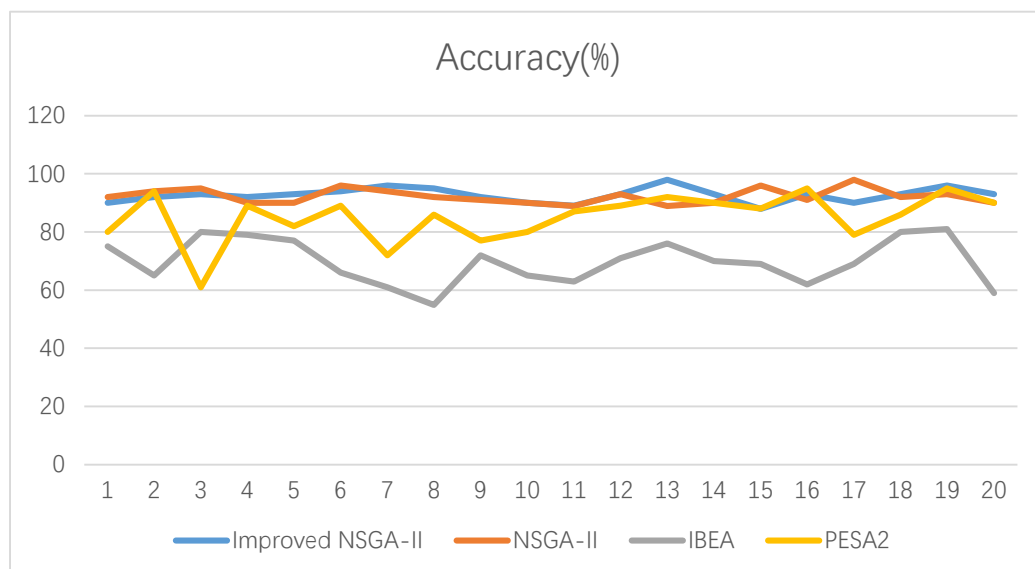
**Figure 8.** Accuracy comparison of experimental results.

Figure 8 demonstrates the accuracy results of the algorithm for generating personalization rules. The improved NSGA-II provides optimal accuracy in 85 to 98%, with an average accuracy of 92%, slightly higher than the average accuracy of NSGA-II algorithm. However, IBEA and PESA2 are less accurate.

The evaluation of multi-objective evolutionary algorithm needs to consider some evaluation

indexes, so various performance indicators are introduced. This paper focuses on only three performance indicators, GD and IGD, to evaluate the efficiency of the algorithm. The IGD algorithm that provides minimum GD value is considered as the best algorithm.

Table 8. Algorithm comparison.

	Improved NSGA-II	NSGA-II	IBEA	PESA2
IGD	0.010	0.012	0.115	0.141
GD	0.065	0.086	0.108	0.26

Table 8 describes the results achieved when applying performance indicators. It can be seen that the improved NSGA-II algorithm provides the minimum value of 0.010 and 0.065 for IGD and GD respectively. and its superior convergence and diversity are verified by calculating the distance between the individuals in the real individuals of non-dominated front. The indicator GD is used to determine the number of solutions placed at the intersection of the obtained Pareto frontier and the true Pareto frontier, which is achieved by calculating the distance between each recommended resource. To sum up, the improved NSGA-II can generate a personalized resources recommendation list better than INSGA-II, IBEA and PESA2 due to its good performance in indicators.

5. Conclusions

Online learning is an emerging education method that aims to meet the preferences of each learner by engaging with them and providing personalized resources. Ultimately, the goal of personalized learning is to provide appropriate learning resources according to the demands of learners. Therefore, this paper utilizes the online learning resource recommendation model based on the improved NSGA-II algorithm to transform the personalized screening of massive learning resources into an optimization problem, and defines two competing objectives. The quality of the resource list gets maximum improvement by maximizing the difficulty and the learning threshold. Besides, the credibility also gets enhanced by maximizing the popularity degree, authority degree, and novelty degree of each learning resource. The improved NSGA-II algorithm takes the dataset, personalized characteristics of learner and the resource characteristic of learner as input. In the process of model construction, a large number of weights are set for the feature processing of the learning resource model in the early stage, which is not intelligent enough. In future work, we plan to add correlation researches between resources to build a model that can customize learning paths according to learners' preferences.

Acknowledgments

The research is supported by the National Natural Science Foundation of China (No. 12105120, No. 72174079 and No. 72101045), Natural Science Foundation of the Jiangsu Higher Education Institutions of China (No. 21KJB520033), Jiangsu Qinglan Project Teaching Group, Lianyungang "521 projects", The Research Project of Graduate Education and Teaching Reform of Jiangsu Ocean University (YJG202201).

Conflict of interest

The authors declare there is no conflict of interest.

References

1. H. Chatti, S. Hadoussa, Factors affecting the adoption of E-learning technology by students during the COVID-19 quarantine period: the application of the UTAUT model, *Eng. Technol. Appl. Sci. Res.*, **11** (2021), 6993–7000. <https://doi.org/10.48084/etasr.3985>
2. H. Ezaldeen, R. Misra, S. K. Bisoy, R. Alatrash, R. Priyadarshini, A hybrid E-learning recommendation integrating adaptive profiling and sentiment analysis, *J. Web Semant.*, **72** (2022), 100700. <https://doi.org/10.1016/j.websem.2021.100700>
3. H. Yu, Q. Dai, Non-stationary financial time series prediction based on self-adaptive incremental ensemble learning, *J. Data Acquis. Process.*, **36** (2021), 1030–1040. <https://doi.org/10.16337/j.1004-9037.2021.05.018>
4. N. Ma, L. Du, Y. Zhang, Z. Cui, J. Guo, Research on influence of group knowledge map construction on teachers' online learning and interaction, *e-Educ. Res.*, **42** (2021), 55–62. <https://doi.org/10.13811/j.cnki.eer.2021.02.008>
5. F. Liu, F. Tian, X. Li, L. Lin, A collaborative filtering recommendation method for online learning resources incorporating the learner model, *CAAI Trans. Intell. Syst.*, **15** (2021), 1117–1125. <https://doi.org/10.11992/tis.202009005>
6. H. Li, L. Yang, P. Zhang, Method of online learning resource recommendation based on multi-objective optimization strategy, *Pattern Recognit. Artif. Intell.*, **32** (2019), 306–316. <https://doi.org/10.16451/j.cnki.issn1003-6059.201904003>
7. H. Li, Q. Chen, Z. Zhong, R. Gong, G. Han, E-word of mouth sentiment analysis for user behavior studies, *Inf. Process. Manage.*, **59** (2022), 102784. <https://doi.org/10.1016/j.ipm.2021.102784>
8. C. L. Tang, J. X. Liao, H. C. Wang, C. Y. Sung, W. C. Lin, ConceptGuide: Supporting online video learning with concept map-based recommendation of learning path, in *Proceedings of the 30th The Web Conference*, New York, (2021), 2757–2768. <https://doi.org/10.1145/3442381.3449808>
9. H. Li, X. Wang, Y. Chen, G. Han, P. Hou, X. Liu, et al., Research of second class learning system structure based on collaborative filtering recommendation algorithm, *J. Jiangsu Ocean Univ. (Nat. Sci. Ed.)*, **2021** (2021), 87–93. <https://doi.org/10.3969/j.issn.2096-8248.2021.04.014>
10. D. Q. Shi, T. Wang, H. Xing, H. Xu, A learning path recommendation model based on a multidimensional knowledge graph framework for e-learning, *Knowledge-Based Syst.*, **195** (2020), 105618. <https://doi.org/10.1016/j.knosys.2020.105618>
11. M. Zhang, S. X. Liu, Y. F. Wang, STR-SA: Session-based thread recommendation for online course forum with selfattention, in *2020 IEEE Global Engineering Education Conference (EDUCON)*, (2020), 374–381. <https://doi.org/10.1109/EDUCON45650.2020.9125245>
12. W. Kong, S. Han, Z. Zhang, Construction of adaptive learning path supported by artificial intelligence, *Mod. Distance Ed. Res.*, **32** (2020), 94–103. Available from: <https://www.cnki.com.cn/Article/CJFDTOTAL-XDYC202003011.htm>.
13. M. Dong, Multi-objective learning resource recommendation algorithm based on knowledge graph, *Comput. Syst. Appl.*, **30** (2021), 139–145. Available from: <http://www.c-s-a.org.cn/1003-3254/7884.html>.

14. H. Mai, Y. Zhong, Application of FRP + GA in adaptive learning system, *J. Liaoning Tech. Univ. (Nat. Sci. Ed.)*, **39** (2020), 459–464. Available from: <https://fxky.cbpt.cnki.net/WKD/WebPublication/paperDigest.aspx?paperID=843f9272-630e-47c4-927e-d702a95d78b5>.
15. F. Gembicki, Y. Haimes, Approach to performance and sensitivity multiobjective optimization: The goal attainment method, *IEEE Trans. Autom. Control*, **20** (1975), 769–771. <https://doi.org/10.1109/TAC.1975.1101105>
16. X. Ma, Y. Li, L. Yan, Comparison review of traditional multi-objective optimization methods and multi-objective genetic algorithm, *Electr. Drive Autom.*, **32** (2010), 48–50. Available from: <https://www.cnki.com.cn/Article/CJFDTOTAL-DQCD201003013.htm>.
17. X. Xie, Y. Chen, X. Li, L. Mo, Collaborative recommendation algorithm of online learning based on trust-combined similarity model with variable weight, *J. Chin. Comput. Syst.*, **39** (2018), 525–528. Available from: <http://xwxt.sict.ac.cn/CN/Y2018/V39/I3/525>.
18. J. Fan, T. Lei, N. Dong, R. Wang, Multi-objective UAV path planning based on an improved NSGA-II algorithm, *Fire Control Command Control*, **47** (2022), 43–48. Available from: <https://info.cqvip.com/Qikan/Article/Detail?id=7106793688>.
19. J. D. Schaffer, Multiple objective optimization with vector evaluated genetic algorithms, in *Proceedings of the First International Conference on Genetic Algorithms and their Applications*, 1985.
20. N. Srinivas, K. Deb, Multiobjective optimization using nondominated sorting in genetic algorithms, *Evol. Comput.*, **2** (1994), 221–248. <https://doi.org/10.1162/evco.1994.2.3.221>
21. K. Deb, S. Agrawal, A. Pratap, T. Meyarivan, A fast elitist non-dominated sorting genetic algorithm for multi-objective optimization: NSGA-II, in *Parallel Problem Solving from Nature PPSN VI*, (2000), 849–858. https://doi.org/10.1007/3-540-45356-3_83
22. K. Deb, H. Jain, An evolutionary many-objective optimization algorithm using reference-point-based nondominated sorting approach, Part I: Solving problems with box constraints, *IEEE Trans. Evol. Comput.*, **18** (2014), 577–601. <https://doi.org/10.1109/TEVC.2013.2281535>
23. V. L. Vachhani, V. K. Dabhi, H. B. Prajapati, Improving NSGA-II for solving multi objective function optimization problems, in *2016 International Conference on Computer Communication and Informatics (ICCCI)*, (2016), 1–6. <https://doi.org/10.1109/ICCCI.2016.7479921>
24. A. H. M. Pimenta, H. A. Camargo, NSGA-DO: Non-dominated sorting genetic algorithm distance oriented, in *2015 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, (2015), 1–8. <https://doi.org/10.1109/FUZZ-IEEE.2015.7338080>
25. J. A. Vrugt, B. A. Robinson, Improved evolutionary optimization from genetically adaptive multimethod search, *Proc. Natl. Acad. Sci.*, **104** (2007), 708–711. <https://doi.org/10.1073/pnas.0610471104>
26. H. Li, Z. Zhong, J. Shi, H. Li, Y. Zhang, Multi-objective optimization-based recommendation for massive online learning resources, *IEEE Sens. J.*, **21** (2021), 25574–25281. <https://doi.org/10.1109/JSEN.2021.3072429>
27. H. Li, X. P. Ma, S. Zhang, J. Shi, C. Li, Z. Zhong, Research of overlap community detection algorithm based on time-weighted, *Acta Autom. Sin.*, **47** (2021), 933–942. <https://doi.org/10.16383/j.aas.c180559>

28. D. Seblova, R. Berggren, M. Lövdén, Education and age-related decline in cognitive performance: Systematic review and meta-analysis of longitudinal cohort studies, *Ageing Res. Rev.*, **58** (2020), 101005. <https://doi.org/10.1016/j.arr.2019.101005>
29. M. Bloomberg, A. Dugravot, J. Dumurgier, M. Kivimaki, A. Fayosse, A. Steptoe, et al., Sex differences and the role of education in cognitive ageing: analysis of two UK-based prospective cohort studies, *Lancet Public Health*, **6** (2021), e106–e115. [https://doi.org/10.1016/S2468-2667\(20\)30258-9](https://doi.org/10.1016/S2468-2667(20)30258-9)
30. Q. Zhang, P. Li, Family background, parental expectations and child’s cognitive abilities: Empirical evidence from Chinese education panel survey, *J. Wuhan Univ. Technol. (Social Sci. Ed.)*, **30** (2017), 97.



AIMS Press

©2023 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)