Research Article

Research on English Online Education Platform Based on Genetic Algorithm and Blockchain Technology

Xuezhong Wu

English College, Zhejiang Yuexiu University, Shaoxing, Zhejiang 312000, China

Correspondence should be addressed to Xuezhong Wu; 20131109@zyufl.edu.cn

Received 17 September 2020; Revised 12 October 2020; Accepted 31 October 2020; Published 24 November 2020

Academic Editor: Hongju Cheng

Copyright © 2020 Xuezhong Wu. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Nowadays, with the rapid development of network technology, the online examination mode has gradually replaced the traditional paper examination mode. This paper introduces computer technology into English teaching and studies and designs an English online test system for English subjects. The system introduces the design module and introduces the genetic algorithm for analysis. By using the artificial intelligence of the genetic algorithm to analyze the examination process, the online examination system completes a series of tasks from questions to examination results, making the examination work intelligent, standardized, and highly efficient. At the same time, it also gives candidates greater fairness and flexibility, making the entire examination process more efficient and convenient.

1. Introduction

With the rapid development of network technology, all walks of life have entered the era of network intelligence; the traditional manual operation and management model is gradually replaced by various IT technologies, and people are enjoying the convenience and ease of this change [1]. The current education industry is also following the trend of network technology to break the traditional teaching methods of teaching, examination and question answering only in the classroom, and the way of manual reading in the classroom. And the mode of distance education has been introduced not only to students but also to social personnel to make more people who want to learn knowledge through the network can be at any time and an arbitrary location of the network to choose the subject they are interested in [2]. Distance education is not only concerned with the content of teaching and the interrogative of the learners, but also more attention is paid to the feedback of the learners on the knowledge they have learned in order to provide guidance to the learners. So the development of the online examination system is particularly important [3]. With the rapid development of computers and networks, the traditional patterns of all walks of life will gradually be replaced by the network technology. Network technology is changing people’s learning, life, work, and way of thinking [4]. Nowadays, network technology is developing rapidly, and the traditional examination mode is gradually replaced by the online examination model [5]. The staff will be liberated from the heavy examination work. They use the online examination system to complete a series of tasks from the problem to the results of the test to make the examination work intellectualized, standardized, and efficient [6]. At the same time, it also gives the examinee greater fairness and flexibility. Moreover, the resources are saved, the environment is protected, and the paperless examination is realized.

2. Related Work

The online examination is becoming more and more popular in foreign countries and has become the mainstream of the form of examination. At the same time, the technology based on the system is constantly updating, and the development of the system is becoming more and more perfect. The United
States is the first country to apply computer technology to education testing, and it has stepped into the stage of mature application [7]. At present, the research and development of online examination system in China are not very long. But because of the mature software and hardware and the rapid development in recent years, the online examination system has been put into use earlier and has a computer test [8]. As early as in the 1993 “Shanghai City computer application ability assessment (primary)” examination has begun to try this new test method. At that time, it was an application under the DOS system because of the limited function of the condition. As the operating system is upgraded from DOS to Windows, the application has been successfully upgraded to the Windows platform [9]. Although the domestic Internet test is developing so fast, it is still the weak link of distance education or online teaching system. At present, there are only objective questions, such as selection, judgment, and filling, which can be fully implemented and other subjective questions, such as short answer and writing which still need people to score. While the subjective score is flexible, it consumes more energy and time than the objective score. So it makes the reader not completely free from the rewording [10]. Now, there is also a study on the score of the subject but the algorithm is still in the primary stage, and the subjective score cannot be completed independently, only a preliminary score and finally a manual examination [11].

3. Method

3.1. Genetic Algorithm. The English online test system is a unified test for all candidates. It requires that every test takers get the same indexes on each test paper [12]. That is, the difference is controlled in a smaller range. The people of setting questions should first set several constraints on the system, such as examination time, each question score, the type, the difficulty coefficient of each question, the knowledge point of each question, and the teaching requirement of each question [13]. According to the direction of the constraint, the genetic algorithm is used to search for the optimization continuously until a test paper near the set condition is produced [14]. Each computer can get a test paper with higher reliability according to the algorithm. Since the papers are all in accordance with the unified constraints, the test paper at the terminal will be the same, for example, the difficulty closed to but not to repeat the test [15, 16]. It also brings unanimous fairness in the examination. Therefore, in many fields of examination, the genetic test paper algorithm is more commonly used. In the traditional genetic algorithm, it is not for the use of the group so it is necessary to improve the genetic algorithm in the application of the paper [17]. First, the mathematical model of this examination paper is constructed. Nine questions will be extracted in the system (2 listening, 2 vocabulary and grammar, 1 reading comprehension, 1 finished fill, 2 tenses, and 1 writing questions). And each question has 4 attributes (topic, number, value, difficulty), so they can construct a $9 \times 4$ order target matrix $A$ to represent the test paper structure [18].

\[
A = \begin{bmatrix}
a_{11} & a_{12} & a_{13} & a_{14} \\
a_{21} & a_{22} & a_{23} & a_{24} \\
a_{31} & a_{32} & a_{33} & a_{34} \\
a_{41} & a_{42} & a_{43} & a_{44} \\
a_{51} & a_{52} & a_{53} & a_{54} \\
a_{61} & a_{62} & a_{63} & a_{64} \\
a_{71} & a_{72} & a_{73} & a_{74} \\
a_{81} & a_{82} & a_{83} & a_{84} \\
a_{91} & a_{92} & a_{93} & a_{94}
\end{bmatrix},
\]

in which $a_{11} \sim a_{94}$ are the title of the question; $a_{12} \sim a_{92}$ are the number of the title; $a_{13} \sim a_{93}$ are the score of the test; and $a_{14} \sim a_{94}$ are the difficulty of the test. The 4 attributes of each problem represent the constraints of 4 aspects [19]. It points out the direction for people to select excellent examination paper. The constraints that the matrix should satisfy to the maximum limit are as follows: the total score of the test paper is bound to $\sum_{i=1}^{9} a_{i3} = 100$ and number of T type scores $\sum_{i=1}^{9} a_{i3} t_i$. The system of hearing problems should meet that every question is 5 points, a total of which is 10 points. Vocabulary and grammar should meet every question is 5 points, a total of which is 10 points [20]. Reading comprehension should meet the total score of 20. The finished fill should meet the total score of 20. The tenses should be 5 points per channel, with a total of 10 points. The writing question should meet the total score of 30 [21, 22]. The sequence number of the test question is not repeated: that is, in the matrix $A$, there cannot be a case of the same two or a few elements in the matrix. Otherwise, it shows that the test is repeated. The difficulty of test paper constraint is calculated as $\sum_{i=1}^{9} a_{i3} a_{i4}/\sum_{i=1}^{9} a_{i3}$. The difficulty of setting up the test paper is 3.5 (the highest degree of difficulty is 5) [23]. Although the final examination papers are often difficult, to meet every constraint we set, there are three hard conditions to be achieved for the system in which the first is that the total score of the exam is 100, the second is that the score of each type is the prescribed value, and the third is that there cannot be the same topic. In the case, all of the three conditions can be reached. If a test paper with a 3.5 degree of difficulty can be found, the test paper will be the optimal solution of the genetic algorithm. If you cannot find a test paper that is just 3.5 of the difficulty coefficient, we have to set the difficulty factor with $3.0 \leq \sum_{i=1}^{9} a_{i3} a_{i4}/\sum_{i=1}^{9} a_{i3} \leq 4.0$ so that the suboptimal solution can be found [24]. The number of each item and the number of each type of question can be entered by an artificial method in the setting of a test paper, and the total score is 100 points. Therefore, each test paper meets the total score constraints and the total score constraints for each class of questions at the beginning of random volume extraction [25]. Then, judge whether there is the same question or not, which can be avoided in the cross algorithm so the only thing that needs to be controlled is the difficulty of the test paper. In other words, at the initial
stage, each test paper is difficult to find a test paper that is difficult to meet the requirements of the setting by the genetic algorithm. In this way, the target function is only set for the difficulty of the test paper. The objective function can be defined as [26]

\[ f = \sum_{j=1}^{4} f_j w_j, \quad (2) \]

in which \( f_j \) is the absolute value of the actual value of the property of the \( j \) volume of the current test paper and the difference between the attribute constraint values of the \( j \) paper volume [27]; \( w_j \) is the weight of the difference factor. Now, only the difficulty coefficient is considered; then, the objective function can be simplified to:

\[ f = f_4 w_4, \quad (3) \]

in which \( f_4 = |\sum_{j=1}^{4} a_{1j} a_{2j} \sum_{j=1}^{4} a_{1j} - \text{ND}| \) is the absolute error between the average difficulty of generating the test paper and the difficulty constraint of the set test. ND is an artificially set test difficulty constraint. \( w_4 \) is the weight of the difference factor of the difficulty coefficient. Generally, the fitness function is designed to be inversely proportional to the objective function. In order to prevent the denominator from being zero, the fitness function can be designed as:

\[ F_j = \frac{1}{1 + \sum_{j=1}^{4} f_j w_j}. \quad (4) \]

In the same way, only the difficulty coefficient can be simplified to

\[ F_4 = (1 + f_4 w_4). \quad (5) \]

It can be seen that the better the \( F \) value is, the better the quality of the 1 test paper. When \( F = 0 \), the best solution is found that fits all the constraints.

3.2. Roulette Selection Algorithm. Crossover is to make test papers produce more quality new individuals. In general, the system needs to be paired randomly to complete the two pairs. The test paper code is divided into several coding segments according to the type of the questions. Each group is divided into eight groups according to the type of the coding. A group can be regarded as a gene point. Here, the system adopts the way of single point crossover. The general operation is first to form two pairs and produce a random number RC of \([0, 1]\) for each pair of test papers and select the probability value \( P_c \). If RC < \( P_C \), a crossover point is generated randomly in the coding segment of a certain type of topic. Then, the part of the point is exchanged to produce a new pair of papers. It can be seen that the crossover does not occur in each pair of test papers and is determined by the crossprobability \( P_c \). Therefore, the value of \( P_c \) is very important. If the value is too large, it will increase the probability of destroying the genetic pattern; if it is too small, the efficiency of producing new individuals will be too slow. The traditional method is to control the \( P_c \) in the range of 0.60~0.8, and this design uses the crossprobability to adjust the value with the change of individual fitness. The formula is as follows:

\[ P_c = \begin{cases} K_c \times \frac{f_{\text{max}} - f'}{f_{\text{max}}} \times f' \geq f_{\text{avg}}, \\ K_c \times f' < f_{\text{avg}}, \end{cases} \quad (6) \]

In which \( f' \) is the one with smaller fitness values in two crosstest papers; \( f_{\text{max}} \) is the maximum fitness value in a test paper individual; \( f_{\text{avg}} \) is the average fitness value of all the test papers; and \( K_c \) is the crossprobability coefficient, \( K_c \leq 1 \). When the larger fitness value in a test paper is less than the average fitness value, the crossprobability is increased; on the contrary, the crossprobability is reduced. There may be repeated questions in the offspring generated after crossing. In this case, we need to replace the 8-bit encoding in the paragraph (one question) to the 8-bit code that has not appeared before and get the new generation again.

3.3. Variant Operation of Tissue Paper. The mutation operation is introduced to solve the high-quality individual reduction produced by the late crossover operation of the algorithm. It is a small probability to change one question (8-bit code) of a test paper to another (another 8-bit code). This system adopts the single point variation in the segment, so the variation does not affect the change of the model. The general operation is generate a random number of \([0, 1]\), such as \( r_m \) for each test paper that enters the mutation operation and select the value of the probability \( P_m \). If \( r_m \leq P_m \), a variation point is generated randomly in a certain type of coding section, and the 8-bit encoding on this variation is replaced by another 8-bit encoding of the same type to produce a new test paper. As a result, the mutation operation does not happen to every test paper but depends on the mutation probability \( P_m \). Therefore, the value of \( P_m \) is very important; if the value is too large, then the genetic algorithm becomes a random search algorithm; if it is too small, it is not easy to produce a new test paper individual. The traditional way is to control the value of \( P_m \) in the range of 0.01 to 0.02, and the design adopts the way of adjusting the mutation probability with the change of individual fitness. The formula is as follows:

\[ P_m = \begin{cases} K_m \times \frac{f_{\text{max}} - f'}{f_{\text{max}}} \times f' \geq f_{\text{avg}}, \\ K_m \times f' < f_{\text{avg}}, \end{cases} \quad (7) \]

in which \( f \) is the fitness value of an individual in a variant test paper; \( f_{\text{max}} \) is the maximum fitness value in a test paper individual; \( f_{\text{avg}} \) is the average fitness value of all the test papers; and \( K_m \) is probability coefficient of variation, \( 0.01 \leq K_m \leq 0.05 \). When the larger fitness value in a test paper is less than the average fitness value, the mutation probability is increased; on the other hand, the mutation probability is
reduced. The offspring of the mutant may have repeated questions. In this case, the repeated 8-bit coding (one problem) in the segment will be replaced by an 8-bit encoding (another problem) that has not appeared, and the new generation will be recovered. Based on the analysis of the genetic algorithm, the following flow chart is obtained, as shown in Figure 1.

4. Result Analysis and Discussion

For the time complexity analysis of the improved genetic algorithm, the number of text in the process of searching the initial cluster center for the extracted sample text is less, the number of iterations is very small, and the speed is very fast. Compared with the original algorithm, the time consumed is not much increased. The improved algorithm performance stability simulation diagram before and after the signal is shown in Figure 2.

The steady-state maladjustment is proportional to the forgetting factor \(\lambda\). In order to observe the stable performance of the algorithm, using the invariant parameters of the algorithm, the signal to noise ratio (SNR) is 10 dB. Then, the steady-state forgetting factor \(\lambda\) of the two improved algorithms is compared. This reflects the state of the algorithm’s steady-state imbalance. The real line \(e_1\) represents an improved algorithm, and the dashed line \(e_2\) represents an
unmodified algorithm. Because in the range of $0 < \lambda < 1$, the smaller the forgetting factor $\lambda$ is, the stronger the tracking ability of the system is, and the more stable the system is, so the advantage of the algorithm in a steady-state misalignment is obvious. The improved cluster diagram account is shown in Table 1.

From Table 1, we can see how large the clusters are and it provides the basis for the next approximation experiment. Each result is recorded, and the experiment is repeated on the data experiment platform. Finally, the satisfactory results are obtained. The accuracy and stability of the improved genetic algorithm are greatly improved. Using the common $K$-means algorithm, the $F$ value of the clustering results is between 0.10 and 0.15 but with the improved algorithm, its value is stable between 0.05 and 0.15. It is also possible to further analyze the results of clustering combined with other related information; thus, there is a more profound understanding of the sample wood points and characteristic variables.

In a word, clustering analysis applied in practice is a process that needs multiple participation, and it cannot be divorced from the user’s participation. Genetic algorithm is only an important step in the whole cluster process, and a satisfactory result is obtained by clustering. The genetic algorithm is obviously better than the random algorithm. In the front of the article, a very detailed introduction to the initialization technique of the test is made. This technique reduces the search scope of the algorithm and improves the efficiency of the algorithm. It also has a great effect on the homogenization of the population. By comparing the search time of the genetic algorithm using the test problem initialization technology, the difference between the fitness and efficiency is obtained. Then, we make a comparison of the aggregation time and evaluate the performance of the system according to the cluster and the different expectations of the number of different individuals.

If the number of individuals taking part in the cluster is 10, 20, 30, 40, 50, and 60, each individual in the cluster chooses only one free time and each period of free time is set to 3 hours. In the process of teaching, we record the time of concentration of focus and the number of three types of agent negotiation. The concentration time is defined from pressing the "prepare" button to the focus. Although the data are test data, the results are more close to the actual data for the results of the test. We randomly extract a number of individuals from a group of individuals to test, and the time to get the focus is determined by the average number of tests. The test results are shown in Figure 3. The test results show that there is a monotonous increase in the number of individuals involved in the concentration of focus. Research on the indexes of English online education platform in time series is shown in Figure 4.

The relationship between the number of running times of the statistical algorithm and the length of the free time period...
and the time to determine the focus was determined. We separately check the number of handover tasks between agent (MA) and identity agent (IA) and the number of transactions between identity agent (IA) and the other party’s intention to cluster agent (HPA). The result is shown in Figure 5. Thus, it can be seen that the compression of the length of an individual free time period and the number of coordination between various types of transactions are also increased accordingly. It will also cause the increase of the number of times and the time of the communication between each event. Data storage comparison of English online education platforms is shown in Figure 6.

The above focuses on how to apply genetic algorithms to the English online teaching system, and the correctness of the theoretical algorithm proposed in this paper is verified. And through the experiment, we find the deficiency of the algorithm and improve it. The final set of test paper algorithm has the following characteristics: the algorithm flow of the test paper is determined; the test question bank classification initialization technology is used, and the questions are selected from the general library according to the constraints and are redeposited into different test database tables according to the type of questions; using the test number as the chromosome coding, it saves the tedious coding and decoding; the error value of the knowledge point is added to the fitness, which will make the location of the knowledge point more accurate; the crossover and mutation operation is upgraded to an adaptive operation, so that the algorithm can automatically match the appropriate operator. In addition, in order to achieve the best results in the overall scope of the search, the small habitat law should be selected when choosing the strategy.

5. Conclusion

Undoubtedly, teaching information has injected a new vitality into the reform of the traditional educational model. And the online examination algorithm is also a symbol of the application of education information. This paper is based on ASP’s online English test algorithm design and exploration. Online examination algorithm has intelligent examination management, intelligent test paper algorithm, intelligent marking function, and so on to emancipate the examiners and teachers from the traditional heavy examination work. In addition, it was aimed at improving work efficiency, enhancing the fairness and flexibility of the examination, and making the examination work algorithmic, standardized, and paperless. This algorithm is based on the features of the English test type and the examination syllabus and set-up test, content, and difficulty condition, establishing a reasonable structure of the database. And through the improvement of the genetic test paper algorithm, the efficiency and quality of the test paper are improved. The main module test algorithm is designed, and the entire process to the results of the examination from the volume can be completed. Each module of the online examination algorithm is tested, and the program is constantly revised to improve the function of the algorithm, and the user experience is more humanized.

Data Availability

All the data is available online.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

References


