6

Research Article

Jianlan Lyu*

Writing assistant scoring system for English second language learners based on machine learning

https://doi.org/10.1515/jisys-2022-0009 received June 24, 2021; accepted November 21, 2021

Abstract: To reduce the workload of paper evaluation and improve the fairness and accuracy of the evaluation process, a writing assistant scoring system for English as a Foreign Language (EFL) learners is designed based on the principle of machine learning. According to the characteristics of the data processing process and the advantages and disadvantages of the Browser/Server (B/S) structure, the equipment structure design of the project online evaluation teaching auxiliary system is further optimized. The panda method is used to read the data, the clean method is used to realize the data preprocessing, the model test is carried out, the cross validation method is selected, the data is divided in advance, and the process of programming the problem scoring system is further optimized, the automatic scoring technology is constructed by English teaching recognition module, feature extraction module and scoring module, the table structure of programming problems is designed, the auxiliary evaluation program of English writing is designed, and the design of writing auxiliary scoring system is completed. The analysis of the experimental results shows that the accuracy of the system is close to 90%, and the total average difference is 0.56. The system can normally take out a variety of test papers. Considering the subjectivity of manual scoring and the impact of key code setting on scoring, the carefully set key code can effectively improve the scoring accuracy of the system. The scoring strategy of the automatic scoring system is effective and the scoring effect is good, and it can be used in practical application.

Keywords: auxiliary scoring, learner writing, machine learning, second language, scoring system

1 Introduction

To facilitate the test of students' comprehensive language application ability, the writing assistant scoring system for English second language learners has been widely used because this system can realize the paperless examination and automatic evaluation of English second language courses, especially for programming questions. Automatic evaluation reduces the workload of paper reviews and improves the fairness and accuracy of the evaluation process. It can also provide a complete self-test system for the second language teaching website to provide better support for students to learn independently. The application of the online scoring system plays an active role in the teaching of English writing and greatly improves students' enthusiasm for self-conscious learning. However, the current writing assistant scoring system for English second language learners has problems such as slow operation and response speed and low scoring accuracy.

^{*} Corresponding author: Jianlan Lyu, College of Foreign Language, Hengyang Normal University, Hengyang 421002, China, e-mail: lyujianlan@hynu.edu.cn

Scholars in related fields have done some research on the writing assistant scoring system for English second language learners. Lalouani and Younis propose a new method of assigning reputation scores based on the observable value of Muser [1]. MuSeR is a novel approach to assign reputation scores for observables, even when no prior information is available, and flag suspicious sessions by conducting inter-observable analysis of user requests. To determine the request score, Muser maps the classifier probabilities to adaptive subjective logic and then uses polynomial fusion to leverage evidence from different observations. Given the request score, Muser further promotes a novel session reputation scoring model that uses three-valued subjective logic to handle trust propagation and aggregation on user requests. The model can effectively score automatically, but the model does not optimize the model hardware, and the scoring results are inaccurate. Zhang et al. proposed a scoring system based on a combination of traditional screening guidelines and random forest algorithm, developed a hybrid scoring system, screened each reservoir/fluid characteristic case, and presented the graphical screening results, and determined the explanation of review (EOR) by applying the random forest algorithm [2]. As the weighting factor, the EOR type and incremental oil recovery are selected as the objective function, and then a scoring system is established through the fuzzification of reservoir/fluid property scores and comprehensive screening scores. The system does not consider the fairness of the evaluation process, so that the accuracy of the scoring results is low.

Machine learning algorithms study how computers simulate or implement human learning behaviors to acquire new knowledge or skills, and reorganize the existing knowledge structure to continuously improve its own performance. Various forms of integrated learning systems that combine various learning methods are emerging. In particular, the coupling of learning symbol learning can better solve the problem of acquisition and refinement of knowledge and skills in continuous signal processing, which has attracted attention. In view of the problems in the above scoring results, considering the fairness of the model hardware and the evaluation process, this study designs a writing aid scoring system for English second language learners based on machine learning, researches and designs the automatic test paper generation, test taker examination, and automatic scoring system in the examination system, and focuses on the automatic scoring method of programming questions. Based on the comparison result scoring method, a method combining repair compilation score and key code comparison is designed to make the scoring result more fair and accurate. For test-taking programs with few errors, the lexical and grammatical analysis methods are applied in the compilation principle to find out the errors and modify them to make them run. Candidates are scored according to the result information and error conditions, which solve the problem of a large number of points lost due to small errors in programming [3]. In the system design process, we attach great importance to the needs of users and the practicability of the software [4]. After a large number of tests and a certain range of trials, it is proved that the system is stable, accurate, and unified and has good practical value and application prospects.

2 Design of writing assistant scoring system for English second language learners

2.1 Hardware structure of English second language learner's writing assistant scoring system

The system is B/S architecture, and the client is only one browser. Considering the ease of deployment, the server side adopts the windows 2003 server +iis +asp framework, and makes the static page with adobe dream weaver: dynamic hyper text markup language (DHTML), JavaScript, and VBScript to make dynamic web pages [5]. The backstage database adopts structured query language (SQL) server and active server pages (ASP) technology to realize the dynamic connection between website and database. The main method of composition of test paper mainly adopts the method of random question extraction to ensure

different students' test papers. To ensure the same scope and difficulty of the same test paper, the system adopts the scheme that the management provides as the unified test paper parameters. When the examinee enters the examination system, the system generates a different test paper from the question bank according to the algorithm of the test paper [6]. Considering the compatibility of traditional habits and the need of some special situations, the system also maintains the way that the paper is composed of specific writing designated by the proposition teacher. Considering that the system will face different users, the writing question base is designed to be highly flexible and extensible [7]. In addition to the relatively fixed type of questions, all the writing of the question bank can be supplemented, deleted, and modified by the users who actually use it. Hence, writing is to set attributes according to chapter, section, knowledge point, and difficulty coefficient. Chapter, section, and knowledge point can be changed, deleted, and added by the course provided by the user. The difficulty coefficient of the system is also set by the user when entering writing [8]. The system uses Adobe Dreamweaver to make static web pages. JavaScript, and VBScript to make dynamic web pages. The backstage database adopts SQL server, and ASP technology realizes the dynamic connection between website and database. The application of B/S structure is established by using local area network (LAN), and the database application is relatively easy to grasp and low cost through Internet mode [9]. This is a continuous development process, which can enable different people to access and operate public databases in different places and in different ways. It can effectively protect the data platform and management access rights, and the server database is also very secure. Especially after the emergence of cross platform language such as Java, the B/S architecture management software is more convenient, fast, and efficient. On this basis, the B/S structure of the scoring assistant system is optimized, as shown in Figure 1.

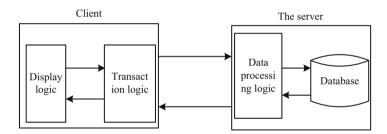


Figure 1: B/S structure of scoring assistant system.

B/S structure is used in the design of the online evaluation system. When users view the page through the browser, the browser needs to display the relevant data but does not carry out complex data processing. It is a long process for the server to process the data submitted by users, which is not suitable for instant display [10]. Considering the characteristics of the above data processing process and the advantages and disadvantages of B/S structure, the interaction between the foreground and the user, that is, through the browser interaction, uses B/S structure, background data processing, that is, automatic evaluation, write back the evaluation results, and other operations. Based on this, the design of the equipment structure of the online evaluation and teaching aid system program is optimized, as shown in Figure 2.

This system is a B/S mode examination system, which can make full use of the Internet resources, not only for the course examination service but also for oral teaching service. To achieve the above goals, the system adopts open registration and closed examination [11]. Students can log in to the system at any time to complete the registration of personal information, use the resources provided by the system to carry out self practice test in oral teaching, and use the simulation test paper provided before the test to be familiar with the test process and answer method. The corresponding process and system structure are as shown in Figure 3.

The system database is divided into two parts: the question bank part and the management part. The writing related part mainly includes two kinds of objective question tables, which are multiple-choice questions and judgment questions, two kinds of semiobjective question tables, which are program filling

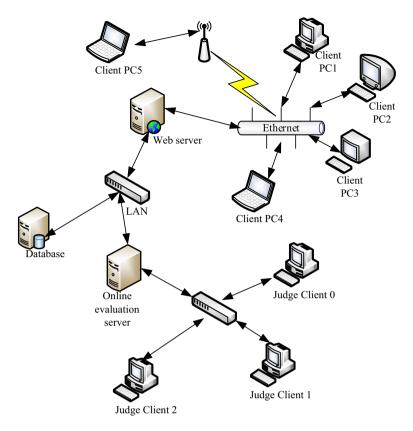


Figure 2: Equipment structure of online evaluation and teaching system of program.

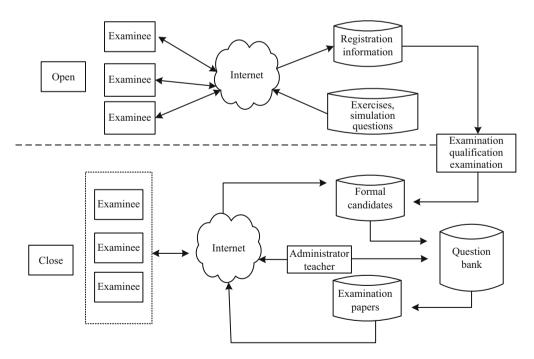


Figure 3: Examination scoring organization and management system structure.

questions and program reading questions, and subjective question tables, which are specially designed to realize the change of subjective questions, as well as other kinds of tables to save relevant information of

examination papers. The management related part mainly includes information tables of various users, user rights management, and other related tables and examination management. The system hardware structure and function structure are as Figure 4.

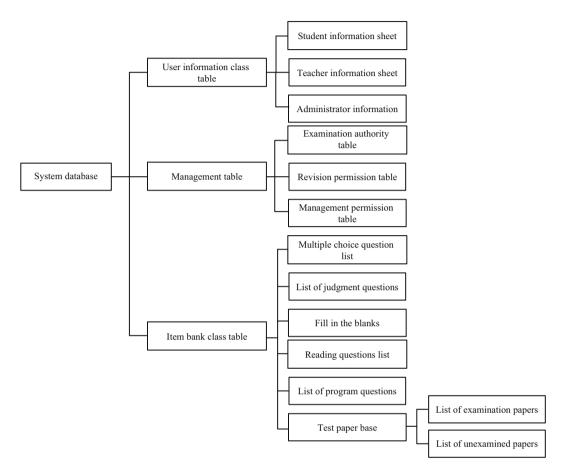


Figure 4: System hardware structure and function structure.

According to the needs of users, the system has two main applications: the first is the course examination system as "English second language programming" and the second is the self-test and practice system as a teaching assistant website. Based on the application of these two aspects, the server of this system adopts the following two configurations: Host hosting is placed in the network management center computer room of the school [12]. The network management center is used to ensure the stable operation of the system and effectively prevent the external network attacks. Meanwhile, the system backup rules and server management rules shall be formulated, and the system shall be backed up regularly and managed by specially assigned personnel. For important examinations, the system is installed in the LAN of the computer room used for the examination by using the characteristics of easy deployment of the system, and the external network is physically isolated. Clean up server system services such as dynamic host configuration protocol (DHCP) Client and task scheduler [13]. The background database of the system adopts Microsoft SQL Server 2000 and uses the security mechanism of SQL Server 2000. The main measures are as follows: when installing SQL Server 2000, mixed mode is used, and the password of system analyst (SA) is required. SA password ensures a certain length and complexity and is modified regularly. The specified database administrator periodically checks whether there are accounts that do not meet the password requirements. Avoid using account in database applications. Create a new super user account with the same permissions as SA for database application and database management. Strictly restrict the administrator's rights, only

give users the permission to meet the application requirements. In the database log record setting, for database login event "failure and success," select "security" in the instance attribute, select all audit levels to record login events of all accounts. In SQL Server 2000, the Tabula Data Stream protocol is used to exchange network data, and the data transmitted in the network including password and content are encrypted to prevent network interception [14]. Pay attention to the release of the latest security patch of SQL Server 2000, download, and install it in time.

2.2 Algorithm design of writing assistant scoring system for English second language learners

Machine learning is a method that enables computers to automatically summarize laws from some data, obtain a prediction model, and then use the model to predict unknown data. It is a way to realize artificial intelligence. It is an interdisciplinary subject, which integrates statistics, probability theory, approximation theory, convex analysis, computational complexity theory, and so on. The outlier detection algorithm in machine learning is designed and adopted in the platform. The outlier detection algorithm learns the location of outliers through the data group structure and data group structure of the dynamic data generation layer, constructs the wireless network dynamic data integrity detection algorithm, and constructs the wireless network dynamic data integrity detection index to complete the wireless network dynamic data integrity detection. The design of writing assistant scoring system is realized by detecting the target, and its calculation formula is as follows.

$$\eta_{i}(T) = \frac{\sum_{i=1}^{n} \int_{0}^{T} g_{i}(t) \nu_{i}(t) dt}{\sum_{i=1}^{n} \int_{0}^{T} P_{i}(t) dt}.$$
(1)

Among them, $g_i(t)$ is the dynamic data set group function, $v_i(t)$ is the label information function in the data block, $P_i(t)$ is the characteristic information content of the dynamic data set group function, and T is the change time of the dynamic data information. Using the outlier detection algorithm to calculate dynamic data, discrete scheduling information flow can strengthen the detection of vocabulary and grammar.

Use pandas method to read data (raw data), use clean to achieve data preprocessing, for model testing, first select the cross-validation method, divide the data in advance, use the processed data to build a training model, test the parameters, and calculate the classification to save the number, original result, predicted result, and predicted probability in comma-separated values (CSV), and store the error classification. Based on this, the machine learning model training process is completed.

To facilitate the use of the system, highlight the integrity, and practicability of the system application function, the system software structure and function structure is optimized, and the system has integrated the examination function and management function. The function of system software module is as Figure 5.

The functions of each module of the system are as follows:

- (1) Paper composition module: According to the writing parameter table given by the proposition teacher, the writing questions can be extracted from the question bank, which is difficult to form and have the same scope of examination, but the writing contents are different. It is also necessary to extract the same content of writing, even allow teachers to directly assign writing papers from the question bank for students to practice and simulate the test.
- (2) Examination module: This module mainly designs various test interfaces to guide students to complete the test of various writing. The module also provides the functions of examination monitoring and test timing.

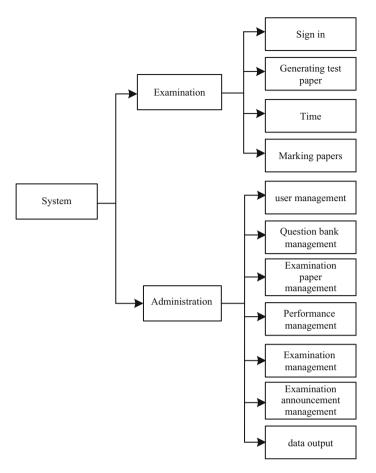


Figure 5: System software module function.

- (3) Automatic scoring system: This module designs different algorithms for different types of questions, evaluates all kinds of questions reasonably, and gives the scoring results according to the set scoring standards.
- (4) System management module: This module includes the management of question bank, announcement management, student management, teacher management, and system management.

The evaluation of software programming questions is the key of the system, and the key point whether the system can realize the completely paperless examination and whether the system functions are advanced and practical [15]. After unremitting efforts, in-depth exploration, and careful design, the system has made some breakthroughs compared with the current examination system. Under the B/S mode, the program design questions of English L2 can be automatically evaluated, and the results of the evaluation are reasonable and accurate, which is basically consistent with the results of manual review [16]. According to the user needs and the in-depth analysis of the user needs, the system of the program design test method is the same as the national level examination, that is, the candidates to complete part of the program code (usually a key function). For this form, the code of nonexaminee filling in part of the program needs to be read from the database, and the candidates' answers can be combined into a complete procedure file to be executed correctly. Based on this, the process of the scoring system of programming questions is further optimized, as shown in Figure 6.

The whole process of program design question evaluation includes generating the examinee's source program file and using tiny C compiler (TCC) to compile and link the program, program operation and control, program operation result comparison, repair compilation score, and key code comparison score [17,18]. Test paper parameters refer to the basic characteristics of the test paper to be filled in by the test

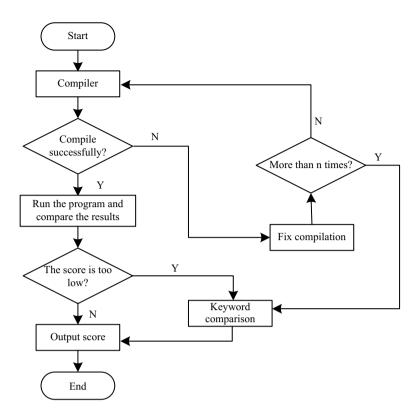


Figure 6: Process of program design question scoring system.

paper composition personnel in the process of test paper composition, which roughly includes the following aspects: test paper name, course name, test paper type, test paper type, test paper code, difficulty coefficient distribution, test paper composition time, number of writing items of each question type, and knowledge points of each question type [19]. It must be noted that the different ways of generating test paper means that the test paper parameters to be filled in by the personnel generating test paper are different. The specific parameters include paper name, course name, paper code, test time, paper type, evaluation score of each writing type in the paper, as well as the number of evaluation writing, difficulty coefficient, and evaluation knowledge points of each type of questions. Based on this, this study explains the writing difficulty coefficient as shown in Table 1.

Table 1: Description of writing difficulty coefficient

Easier	Easily	Secondary	More difficult	Difficult
0	0.1	0.2	0.3	0.4

Through the structure of the auxiliary scoring system, the comprehensive evaluation index data of the database capacity and bearing state of the auxiliary scoring system are obtained. According to the obtained comprehensive evaluation index data, the safety matrix is established, which is expressed as:

$$\mathbf{G} = \begin{bmatrix} G_{11} & G_{21} & \dots & G_{n1} \\ G_{12} & G_{22} & \dots & G_{n2} \\ \dots & \dots & \dots & \dots \\ G_{1m} & G_{2m} & \dots & G_{nm} \end{bmatrix}.$$
 (2)

Combined with the above matrix, the strain capacity of the auxiliary system is calculated as follows.

$$\bar{G}_{i} = 2g_{ij} \frac{\prod k_{ij}}{\sum_{i=1}^{a} (\Pi k_{ij})^{\frac{1}{n}}}.$$
(3)

In formula (3), *i* and *j* are the maximum and minimum parameters in the effective storage range of the database, k_{ij} is the height of the system, a is the judgment scale within the system, and g_{ij} is the relative importance of structural elements. By calculating the judgment scale of system structural safety, the standard value of the strain capacity of the database structure is calculated. Then the deformation matrix of scoring parameters is as follows:

$$\mathbf{V} = \begin{bmatrix} v_{11} & v_{12} & \dots & v_{1n} \\ v_{21} & v_{22} & \dots & v_{2n} \\ \dots & & & & \\ v_{n1} & v_{n2} & \dots & v_{nn} \end{bmatrix}.$$
(4)

The value of the matrix ranges from 1 to 9. These values represent different meanings. The evaluation and calculation of the quality indicators of the database can be summarized as the calculation of the eigenvectors and results of the matrix. The component Q in the characteristic equation represents the weight of each quality index of the database. The specific steps of calculating the matrix characteristic equation are as follows. Calculate the security elements of each row of matrix Q as follows:

$$Q = \prod_{i=1}^{j} \overline{G}_{i} V_{ij}, \quad (j = 1, 2, ..., n).$$
 (5)

If *m* is the reference value of calculation matrix, the calculation method of system load is as follows:

$$\overline{U} = m^2 \sqrt{Q_{ij}}, \quad (i = 1, 2, ..., n).$$
 (6)

After normalizing the values calculated in the above steps, the database capacity data can be obtained. The calculation method is as follows:

$$W_i = \frac{1}{n} \sum_{j=1}^{i} \overline{U} Q \frac{\delta_{\text{max}}}{m-1} \tag{7}$$

In formula (7), δ_{max} in the equation represents the characteristic equation root of database quality detection, *n* and *m* represent the maximum and minimum value of the characteristic vector in the range of change respectively. According to the formula, the quality evaluation model B is calculated:

$$B = \frac{1}{2} \sqrt{\overline{(G_i + \overline{U_j})}^{m-n} \delta_{\max}}.$$
 (8)

To judge whether the matrix satisfies the average random consistency, the assignment test is carried out on the matrix. If an undefined identifier is found in the program, it is likely that the variable name, function name, or keyword is wrong due to the wrong input of candidates. In this case, the correction strategy is to look up the keyword table and the defined identifier table. If a word with high similarity is found (usually the examinee will only make a mistake, omit or add one letter), it will be replaced and then compiled. The process is as shown in Figure 7.

The parameters of parameter test paper generation are the same as those of specific test paper generation, but the processing mode of the system is different. The parameters of the manual test paper generation include test paper name, course name, test paper code, test time, test paper type (test/test), assessment score of each writing type in the test paper, and the number of assessment writing [20]. According to the requirements of different users for the system and the analysis results of the system requirements, the test paper generation methods are divided into automatic test paper generation and manual test paper generation. Automatic test paper generation is further divided into parameter test paper

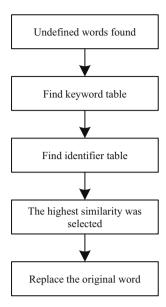


Figure 7: Undefined identifier correction process.

generation and specific test paper generation. The writing aided scoring model for English second language learners is as Figure 8.

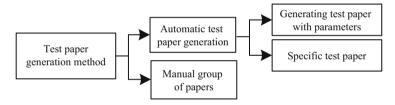


Figure 8: Writing aided scoring model for second-language learners of English.

The original intention of dividing parameter test paper and concrete test paper is to avoid the phenomenon of plagiarism caused by the same content of test paper in the process of online examination. Parameter test paper generation means that the system collects the main parameters of the test paper after setting the main parameters of the test paper (such as the name of the test paper, the knowledge points tested, and the distribution of the difficulty coefficient of the test paper) and then directly saves them in the database. At the beginning of the examination, the system takes out the parameters of the test paper and forms a different test paper for each candidate through the test paper generation algorithm.

2.3 Realization of English second language learner's writing assistant score

The system provides two different test modes for users, namely self test and online test. Self test is used to improve the practice in ordinary times, whereas the online examination is a collective examination organized by the teaching teacher. For online examination, the system is further divided into two types: the same test and the different examination. The system mainly adopts the different examination. So-called different papers refer to the test papers used by examinees are different from each other. The scheme is to set the unified test paper parameters (i.e., parameter test papers) for a certain test by the proposition teacher. When the examinees enter the exam, the system generates a test paper for each candidate

according to the algorithm of the group paper. These papers are different from each other, which can avoid the abuse of copying each other by the adjacent examinees. The same test paper, that is, the test paper used by examinees is the same; this way is designed according to the needs of users. In addition, the system also keeps the manual volume formation method according to the user requirements. The classification of system test mode is as Figure 9.

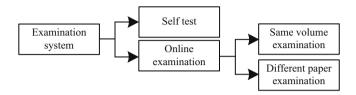


Figure 9: Classification of writing category scoring model.

The automatic scoring technology of English writing is based on the technology of English teaching recognition and English teaching analysis, and it integrates the knowledge of natural language understanding, data mining, and other fields. The computer automatically scores the candidates' English answers. According to whether the answer is known and unique, the types of English test are often divided into closed type and open type. Pets low level English test is mainly short question and answer, which belongs to the open type. The block diagram of the automatic scoring technology for open-ended questions is shown in the figure. The automatic scoring technology mainly includes three modules: English teaching recognition module, feature extraction module, and scoring module. The automatic scoring technique for the writing test is as shown in Figure 10.

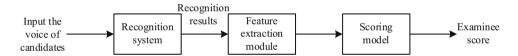


Figure 10: Block diagram of automatic scoring technology for writing test.

The online scoring system helps teachers to make periodic evaluation. In addition to the personalized evaluation of students, it also provides some reference data for teachers' teaching. Through the class data, teachers can understand the students' writing level, adjust the teaching plan, and make the teaching better combined with the practical application. The author uses the marking network to train the freshmen in writing. This semester, there are four times. The full score of writing is 15, and the number of participants is 125. At the end of the semester, the analysis of students' writing is as follows. The analysis of students' final writing is as shown in Table 2.

Table 2: Students' final writing

Assessment seal	Assessment section	Assessment of knowledge points	Number of examination questions		
C program overview	Steps of running C language	The method of running C program on computer	3		
AA	ВВ	С	DD		

English teaching recognition module is mainly to convert the input examinee's English teaching into text for the use of feature extraction module. For open-ended questions, in the case that the correct answer is not unique and there is no correct answer as a priori, we need to use the English teaching recognition module to convert the input examinee's English teaching into words, and extract the characteristics of the acoustic layer at the same time. The accuracy of the output text of the English teaching recognition module will directly affect the effectiveness of the language level features extracted by the feature extraction module. Feature extraction module is very important in the whole English test automatic scoring technology because the effectiveness of the features used will be the key factor to determine the performance of the whole system. The difference of feature extraction results is also one of the main reasons and decisive factors for the different performances of different research institutions. The open-ended question type is more to test the candidates' language application ability, and the characteristics of the language level are indispensable and account for a considerable proportion. The scoring module takes the features extracted by the feature extraction module as the input data, and finally outputs the examinee's English score. Generally, the nonlinear prediction method is used to score the candidates' English answers. Among them, the training data of scoring model is generally the comprehensive score of odd number of reliable manual scoring. These training scores need to be rescored by experienced raters. The purpose of this system is to avoid the mistakes in the design of the whole test paper and to ensure the correctness of the whole test paper. Therefore, the table structure of programming questions is specially designed. It not only contains the test data, test output results, test data file, output result file, and other data in the automatic revision of general program design questions but also adds some key codes and corresponding weights. Other field information is the same as the above tables. The system also divides the programming questions into two tables to save, one of which saves the key codes and the corresponding weights. Similarly, the difficulty coefficient and the knowledge points of the chapters are mainly convenient for the operation of the test paper. The structure optimization of English writing auxiliary scoring program is as shown in Figure 11.

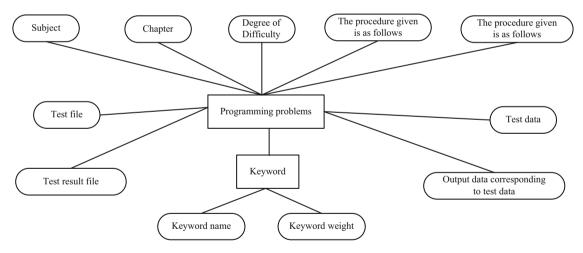


Figure 11: Structure optimization of English writing auxiliary scoring program.

In addition, this study also designs the auxiliary evaluation procedure of English writing, as shown in Table 3.

In Table 3, when the integer value exceeds the int data range, Bigint can be used. For implementation, the int data type is still the main integer data type in Microsoft SQL Server 2005. Varchar is a more flexible data type than Char. It is also used to represent data, but Varchar can hold variable-length strings. Where M represents the maximum length of the string allowed to be stored in the data type, as long as the string length is less than the maximum value, it can be stored in the data type. Real records are the difficulty of the problem. Int is used to record the chapter to which the problem belongs, the part to which the problem

Table 3: Problem assessment program for English Writing

Data type	Explain					
Bigint	The table can uniquely identify a question					
Varchar	Record the title of the program question					
Varchar	Record the procedure given					
Varchar	Record the procedure given					
Varchar	Record the front of the program required for automatic marking					
Varchar	Record the back of the program required for automatic marking					
Varchar	Record test data					
Varchar	Record the results of corresponding test data					
Varchar	Test file and install test data					
Varchar	Test result file and install the corresponding test data of student program					
Real	Record the difficulty coefficient of the problem					
Int	Record the chapter to which the question belongs					
Int	Record the section to which the question belongs					
Int	Record the knowledge points to which the question belongs					
Smaldatetir	Record the time when the question was uploaded					
Bit	Delete tag					
Char	Picture, spare					

belongs, and the knowledge point to which the problem belongs. Smaldatetir is used to record the upload time of the problem, and Bit is used to delete the tag.

The examination paper is mainly used to save the students' examination papers, which is convenient for teachers to query the examination situation and students' scores. In the examination paper, not only the basic information of the test paper but also the examination situation of the students is saved. In the corresponding table, the answers given by the students and the corresponding scores are also saved. Because the students' examinations are all in the form of specific papers that have been formed, the parameter table is no longer needed in the tested papers, but only the specific writing table is needed in the tested papers. The tested papers are generated from the untested papers. One untested paper can correspond to multiple tested papers. Refer to the data flow chart of test paper generation for the relationship among template test paper, nontest paper, and tested paper. The process flow of test paper generation evaluation data is as shown in Figure 12.

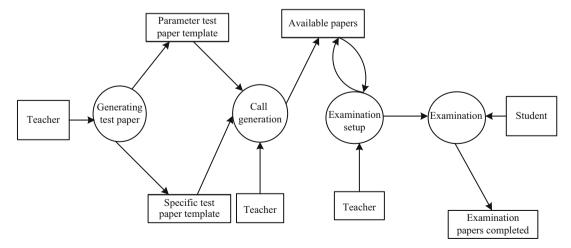


Figure 12: Test paper generation evaluation data processing flow.

In this system, if the examinee is not set as the test object, there is no new record on the online examination page, some of which have been tested before, that is, the record in the locked state of the test. If the examinee has been set as the test object, but the time to enter the system is not within the set range, the corresponding test paper questions will be displayed in the online examination page, but locked. Only when the time set for the test is waiting, the examinee can enter the examination interface for the test. When a candidate enters the exam, if other students log in with the account number of the examinee, they can enter the online examination interface, but the corresponding test papers are in the examination state and cannot be repeated into the test. Different permissions to different users are assigned to ensure the public and private ownership of data. Illegal operation of users is prohibited to prevent the system from being damaged intentionally or unintentionally. The users of this system are divided into three levels: administrators, teachers, and students. The administrator has the highest system authority, namely the system administrator. The teacher is a low-level administrator, and the system administrator can assign different operation rights to different teachers. For example, the system strictly limits the user account number and password length in the registration page, makes strict judgment on the data type of age, and makes strict judgment on the mailbox format, and uses Chinese verification code. The system program to correct the user's error operation and prohibit the execution to maintain the normal operation of the system.

3 Experimental results

To improve the reliability of the system, the system is tested. Statistics show that 40% of the total workload of the software development is used for system testing and maintenance. At present, formal methods and program correctness proof techniques have not yet become practical methods. In this case, software testing is still an effective method to ensure software reliability for a long time. The system is a web application software of B/S mode, and there may be a large number of visitors online at the same time. Therefore, it is necessary to carry out stress test to check whether the network pressure relief strategy designed by the system can cope with the situation of large number of visitors. The system is an online examination system. To ensure the effectiveness, fairness, and justice of the examination, it is necessary to test the test paper generating function of the system. Taking the number of users of English second language learners' Writing Assistant scoring system as the experimental variable, this study verifies the effectiveness of English second language learners' Writing Assistant scoring system based on machine learning. Automatic marking system is the key to realize paperless examination. Therefore, it is necessary to test the scoring function of the system, especially the automatic scoring function of programming questions. In the campus network and LAN, under the premise of ensuring the performance of the server, the system has undergone all the stress tests, and the system has a good running speed and response speed. For those less than 50 people, the server is configured with integrated development environment (IDE) hard disk with central processing unit (CPU) of Celeron d2.0 g, memory of 1 G and average reading speed of 40 m, which can meet the demand. For the test of 80 people, there is a significant lag in the selection of questions (but it is still within the acceptance range of candidates). After replacing the high performance server (CPU 3.2 g, memory 4 G, SCSI hard disk), the system speed is normal under all stress tests. The test results are as shown in Table 4 (taking the system test for different papers and 50 people taking part in the test as an example).

Table 4: System pressure test results

System hardware configuration response effect	Celeron D 2.0 g, memory 1 G, average read speed of hard disk 40 M	CPU 3.2 G, memory 4 G, SCSI hard disk			
No lag	43	50			
It lags slightly, but it can	7	0			
Serious lag, unacceptable to users	0	0			

To analyze the spatial complexity of the method in this study and make the number of states of the three methods increase evenly with time, the change of the number of state transitions of the three methods is compared. The change curve of the number of state transitions with time is shown in Figure 13.

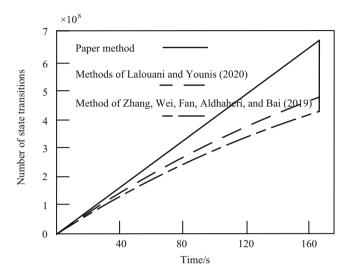


Figure 13: Change curve of state transition number with time.

With the growth of this method over time, the number of state transitions increases steadily, whereas with the increase or decrease of methods of Lalouani and Younis [1] and Zhang et al. [2], the growth of state transition number is lower than that of this method. It can be analyzed that the methods will produce a large number of ineffective state transitions. It is proved that the proposed method can effectively reduce the space complexity.

Table 5: Results of automatic scoring test for program design questions

Candidate number											
	No. 1									No. 2	
Automatic scoring results	10		5		10		10		0	4	4 104 10
Manual scoring results	10		4		10		10		1	3	2 105 10
Absolute difference	0		1		0		0		1	1	2 0 1 0
Candidate number	No. 3									No. 4	
Automatic scoring results	6		10		10		0		5	10	100 8 4
Manual scoring results	8		10		10		1		4	10	101 7 5
Absolute difference	2		0		0		1		1	0	0 1 1 1
Candidate number	No. 5									No. 6	
Automatic scoring results	4		2		10		6		0	10	104 1010
Manual scoring results	6		3		10		5		1	10	105 1010
Absolute difference	2		1		0		1		1	0	0 1 0 0
Candidate number	No. 7									No. 8	
Automatic scoring results	3		10		0		10		2	5	6 2 8 10
Manual scoring results	3		10		1		10		3	5	7 3 7 10
Absolute difference	0		0		1		0		1	0	1 1 1 0
Candidate number	No. 9									No. 10	
Automatic scoring results	10	4		10		10		2		10	4 101010
Manual scoring results	10	5		10		10		1		10	2 101010
Absolute difference	0	1		0		0		1		0	2000

According to the data in Table 4, for the correct examinee program, can correctly score, scoring results are 100% correct. For the examinee program with errors, the average difference between automatic marking and manual marking is 1.18 points (the score of each procedure is 10 points), the accuracy is close to 90%, and the total average difference is 0.56. In all simulated examinations, the system can take out all kinds of test papers normally, and the function of different test papers and the same test paper is normal. After a large number of tests, the automatic scoring system of the system can score the judgment questions, multiple-choice questions, program reading questions and program blank filling questions perfectly and correctly. In the above program design question automatic marking test, all the examinee programs can achieve automatic scoring. The detailed test results are as shown in Table 5.

According to the experimental data in Table 5, when the candidate number is 10, the absolute error of the score of the proposed system is only 2. It can be seen that, considering the subjectivity of manual scoring and the influence of key code setting on the scoring, the carefully set key code can effectively improve the scoring accuracy of the system. Therefore, it can be considered that the scoring strategy of the automatic scoring system is effective and the scoring effect is good, which can be used in practical application.

4 Analysis and discussion

This study designs a writing aid scoring system for English second language learners based on machine learning, optimizes the B/S structure of the scoring assistant system, and further optimizes the equipment structure design of the program online evaluation assistant system, and adopts the B/S model examination system, the realization of the system adopts open registration and closed review, and the system database is divided into two parts: the question bank part and the management part. The server adopts the following two configurations: the hosting is placed in the computer room of the school network management center, the back-end database of the system adopts Microsoft SQL Server 2000, the security mechanism of SQL Server 2000 is used for database application and database management, the latest security patch of SQL Server 2000 is downloaded and installed, the paper writing module and exam are designed Module, automatic scoring system module and system management module. The above hardware design can ensure the stability of the system, ensure the safety of the system, prevent the system from loopholes, and strengthen fairness for students. Through the test of the system's scoring function, in the campus network and local area network, under the premise of ensuring the performance of the server, the system has passed all the stress tests and has a good running speed and response speed. In the software design part, through the analysis of user needs, the system of the programming test method is the same as that of the national exam, that is, the examinee completes part of the program code. The whole process of program design question evaluation includes generating candidate source program files, using TCC to compile the link program, program operation and control, program operation result comparison, repair compilation score, key code comparison score, and calculate the weight of each quality indicator in the database. Look up the keyword table and the defined identifier table to avoid plagiarism in the online examination process, divide the parameter test paper and the specific test paper, and use the test paper generation algorithm to form a different test paper for each examinee. It can be seen through experiments that the simulated test scores are evaluated by the system in this article. The average difference between automatic scoring and manual scoring is 1.18 points (each program scores 10 points), the accuracy rate is close to 90%, and the total average difference is 0.56.

The scoring system designed in this study is basically similar to manual scoring. The system scoring is basically reasonable, and the scoring results meet the basic requirements. Moreover, the scoring system designed in this study is more fair, and the scoring efficiency is not comparable to that of humans. It has a high application value and can be replaced. Manual examination of papers reduces the burden on teachers to mark papers. The scoring strategy of the automatic scoring system is effective, the scoring effect is good, and it can be used in practical applications.

5 Conclusion

As an important auxiliary tool for writing teaching, the online writing evaluation system has a very significant advantage. On the premise of ensuring fairness, it replaces the teacher's evaluation of papers. It reduces the burden on teachers for marking papers and increases the efficiency of marking papers, which can be widely used. However, due to technical limitations, there are still several problems. There was more feedback on vocabulary and grammar, but less feedback on thought content, logic, and text structure. Most of the feedback is suggestive and not specific enough. If the teacher only relies on the system without intervention, it is difficult to improve the weak links between the students. Students constantly modify and submit their essays based on the teacher's comments and systematic sentence-by-sentence comments, but only the last essay they submitted is saved on the Internet. Students are unable to review the progress of their studies, which affects the results of their studies.

Acknowledgment: This study was supported by the 2019 Hunan Provincial Department of Education's General Higher Education Reform Research Project (project number: Xiangjiaotong [2019] 291 No. 662) and the School-Enterprise Cooperation Horizontal Project (project number: 20200070303).

Conflict of interest: Authors state no conflict of interest.

References

- Lalouani W, Younis M. Multi-observable reputation scoring system for flagging suspicious user sessions. Comput Netw. 2020;182(2):107474.
- [2] Zhang N, Wei M, Fan JY, Aldhaheri M, Bai B. Development of a hybrid scoring system for EOR screening by combining conventional screening guidelines and random forest algorithm. Fuel. 2019;256:115915.
- Hajihosseini M, Andalibi M, Gheisarnejad M, Farsizadeh H, Khooban MH. DC/DC power converter control-based deep machine learning techniques: real-time implementation. IEEE Trans Power Electron. 2020;35(10):9971-7.
- Miller TH, Gallidabino MD, MacRae JI, Owen SF, Bury NR, Barron LP. Prediction of bioconcentration factors in fish and invertebrates using machine learning. Sci Total Env. 2019;648(10):80-9.
- Wellmann T, Lausch A, Scheuer S, Haase D. Earth observation based indication for avian species distribution models using the spectral trait concept and machine learning in an urban setting. Ecol Indic. 2020;111(5):106029-30.
- [6] Muraro C, Polato M, Bortoli M, Aiolli F, Orian L. Radical scavenging activity of natural antioxidants and drugs: development of a combined machine learning and quantum chemistry protocol. J Chem Phys. 2020;153(11):114117.
- Mishra M, Bhatia AS, Maity D. Predicting the compressive strength of unreinforced brick masonry using machine learning techniques validated on a case study of a museum through nondestructive testing. J Civ Struct Health Monit. 2020;10(3):389-403.
- Liu JX, Xu BL, Zheng C, Gong YH, Garibaldi J, Soria D, et al. An end-to-end deep learning histochemical scoring system for breast cancer TMA. IEEE Trans Med Imaging. 2019;38(2):617-28.
- Schwartz O, Talmy T, Olsenc CH, Dudkiewicz I. The landing error scoring system real-time test as a predictive tool for knee injuries: a historical cohort study. Clin Biomech. 2020;73:115-21.
- [10] Cfl A, Vm B. Design and development of MLERWS: a user-centered mobile application for English reading and writing skills. Proc Comput Sci. 2019;161(7):1002-10.
- [11] Lian Q, Zhao T, Jiao T, Huyan Y, Gu H, Gao L. Direct-writing process and in vivo evaluation of prevascularized composite constructs for muscle tissue engineering application. J Bionic Eng. 2020;17(3):457-68.
- [12] Fogel Y, Josman N, Rosenblum S. Functional abilities as reflected through temporal handwriting measures among adolescents with neuro-developmental disabilities. Pattern Recognit Lett. 2019;121:13-8.
- [13] Kumar R, Joanni E, Savu R, Pereira MS, Singh RK, Constantino CJL, et al. Fabrication and electrochemical evaluation of micro-supercapacitors prepared by direct laser writing on free-standing graphite oxide paper. Energy. 2019;179:676-84.
- [14] Nguyen CT, Nguyen HT, Mita K, Nakagawa M. Robust and real-time stroke order evaluation using incremental stroke context for learners to write Kanji characters correctly. Pattern Recognit Lett. 2019;121:140-9.
- [15] Ibrahim NK, Hammed H, Zaidan AA, Zaidan BB, Alaa M. Multi-criteria evaluation and benchmarking for young learners' English language mobile applications in terms of LSRW skills. IEEE Access. 2019;7:146620-51.

- [16] García Nieto PJ, García-Gonzalo E, Alonso Fernández JR, Muñiz CD. Water eutrophication assessment relied on various machine learning techniques: a case study in the Englishmen Lake (Northern Spain). Ecol Model. 2019;404:91–102.
- [17] Ellina G, Papaschinopoulos G, Papadopoulos BK. Research of fuzzy implications via fuzzy linear regression in data analysis for a fuzzy model. J Comput Methods Sci Eng. 2020;20(3):879–88.
- [18] Orlenko EV, Barilov AK, Evstafev AV, Orlenko FE. Peculiarities of a helium interaction with hydrogen and free electrons from the point of view of the exchange perturbation theory solutions. J Comput Methods Sci Eng. 2020;20(4):1183–1209.
- [19] Hao YW, Lee KS, Chen ST, Sim SC. An evaluative study of a mobile application for middle school students struggling with English vocabulary learning. Comput Hum Behav. 2019;95:208–16.
- [20] Ying L, Jia Y, Li W. Research on state evaluation and risk assessment for relay protection system based on machine learning algorithm. IET Gener Transm Distrib. 2020;14(18):6552-8.