

Research Article

Design of English Intelligent Simulated Paper Marking System

Lina Yang¹ and Wei Liu² 

¹*School of Foreign Languages, Xinyang College, Xinyang, Henan 464000, China*

²*Intercultural Language Research Center, Xinyang Normal University, Xinyang, Henan 464000, China*

Correspondence should be addressed to Wei Liu; xylw@xynu.edu.cn

Received 11 February 2021; Revised 7 April 2021; Accepted 13 April 2021; Published 21 April 2021

Academic Editor: Wei Wang

Copyright © 2021 Lina Yang and Wei Liu. 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.

In this paper, we analyze the intelligent subdesign of the simulated marking system through an in-depth study of it. This paper proposes a correlation analysis-based quantification of N-element sense values and a rationality enhancement-based scoring fitting algorithm for English essays. This paper also extracts word features, sentence features, and chapter structure features in essays to fit English composition scores. Since not all students can complete the essays according to the topic requirements, a triage scoring model is used to separate the normal essays from the low-scoring essays. Statistically, it was found that the essay scores also showed a certain normal distribution. The standard support vector regression algorithm is prone to data skewing problems, so this paper addresses this problem by using a rationality enhancement method that gives a corresponding penalty factor according to the distribution of the dataset. The results show that the English essay scoring fitting algorithm proposed in this paper can well improve the prediction accuracy of some data and solve the problem of skewed data where the scores show a normal distribution. This paper designs and implements an online mock examination system that incorporates an intelligent scoring system for essays, enabling it to meet the needs of teachers and students for online examinations and intelligent scoring.

1. Introduction

In recent years, with the continuous evolution of IT technologies, especially the wave of smart education represented by artificial intelligence, cloud computing, big data, and the Internet of Things, these technologies have triggered a new round of educational revolution, which is also a complete overturning of the traditional educational evaluation field, showing a magnificent blueprint of educational assessment and evaluation for education experts and scholars, and the majority of front-line teachers [1]. Online examinations based on computer networks are characterized by convenient organization, independent of time and space constraints, and rapid access to examination results, which have brought about a significant increase in marking efficiency and an intuitive decrease in assessment costs [2]. However, among the many online examination systems, it is common that there is little or no automatic machine grading involving subjective question types [3]. After research, the algorithms used for automatic

machine marking of objective questions in online examination systems are relatively simple and easy to implement and only need to compare candidates' answers with standard answers to determine the correctness and give the examination score in real-time; it is obvious that the application of automatic marking techniques for objective question types represented by multiple-choice, judgment, matching, and fill-in-the-blank questions has greatly improved the markers' performance, making the probability of misclassification or omission close to zero [4]. In the field of automatic machine marking of subjective questions, especially in the field of Chinese subjective marking [5], the intelligent marking technology has become more than a word-by-word comparison and verification process. Given the different levels of understanding of each candidate and the diversity of Chinese language expressions, even if students can answer accurately, it is difficult to fully harmonize with the narrative and logic of standard answers [6], which undoubtedly poses a significant problem for the "less intelligent" computer assessment.

The development of the mock exam platform eliminates the need for frequent printing of test papers. Teachers only need to select the topics to be assessed in the system's question bank, assemble the papers when an exam is required, and release the papers. The mock test system saves teachers' time in printing test papers and saves paper to protect the environment [7]. There is no need to fill in answer cards for online exams, and the accuracy rate of marking objective questions is 100%. This prevents students from losing marks due to nonstandard answer cards and allows teachers to be more accurate when counting error rates. The intelligent marking system can be used to mark essay questions, which reduces the work pressure for teachers and allows them to summarize students' mistakes promptly so that they can be more focused when explaining their essays; for students, they can get an objective and fair score, and they can also correct the mistakes in their essays according to the feedback results of the system.

Our goal is to make the marking more reasonable, fair, and open and to make our English scores more accurate. The English essay test is an extremely important type of question in the comprehensive examination of students' English application skills and helps to provide a comprehensive understanding of students' proficiency in English. Automatic machine-based marking will help to establish a consistent marking standard for the marking process, which will no longer be implemented by subjective "humans," but by objective machines, ensuring consistency throughout the marking process. This will ensure that students' English composition papers are marked objectively and fairly with high quality, which will provide a rigorous basis for teaching improvement and talent selection.

2. Related Studies

The Project Essay Grade (PEG) essay scoring system has been developed by a team at Duke University in the United States [8]. At this time, natural language processing technology was still in its infancy and was based on working life experience with natural language, so the PEG system was a rule-based approach to grading essays that focused on form [9]. The system uses word length to determine students' mastery of vocabulary, sentence length to predict students' mastery of sentence structure, and so on. It extracts these simple and easily quantifiable text characteristics and then performs multiple linear regression analysis on these variables to produce a linear regression equation. When the essay is to be graded, each feature variable is entered to determine the score of the essay. This extraction of only simple features without involving the content features of the essay was easy for the test takers to find loopholes, which was the reason why the PEG system was not widely available at that time.

By extracting multidimensional features and setting the corresponding weights according to the usefulness, Wu stitched the output of the LSTM unit layer into the feature vector of the current article after iterating through a multilayer deep recurrent neural network and obtained

the score using statistical analysis such as ridge regression [10]; Li extracted grammatical errors using a language model and detected errors in the use of coronal and prepositional words [11], Malak used a multivariate rule-based grammar check and fused N-element grammar model to detect grammar [12]; Lasen used CNN to extract deep features and LSTM to process temporal information combined with traditional automatic marking methods to score essays, and so on [13]. After the teacher logs in and sets the essay topic, the essay number is given to the student, who then responds online and submits the essay to the system, which automatically grades the essay score [14]. The system automatically scores the essay by calculating the distance between the student's essay and the corpus in the system, and the closer the student's essay is to the corpus, the higher the score [15]. The feedback is given on a sentence-by-sentence basis, and any errors or hints in each sentence are listed. In terms of the evaluation of the essays, the feedback is mainly on the spelling and usage errors of words, while the shallow evaluation is done mainly on the sentence and chapter structure. The natural flow of English and Chinese is completely different; Chinese is about a word right; basically there is no weak reading, let alone hyphenation, while English is completely different; in regular English, a large number of syllables will be weak and connected together, resulting in a phrase or a sentence that sounds like a long word and is a completely unfamiliar word, dictation, no matter how to rewind are I cannot understand it, but when I look at the text, I find that it is entirely an unusual phrase made up of a few simple words.

Through this kind of evaluation, students can have a clear understanding of the degree of mastery of what they have learned at this stage based on the content of the examination so that they can adjust their learning style and focus according to the assessment results, and teachers can also adjust their teaching content and teaching style based on the distribution of students' error-prone questions and grading results and provide in-depth explanations of the content that most students cannot master. The teacher can also adjust the teaching content and teaching style according to the distribution of students' error-prone questions and grading results, explain in-depth the content that most students cannot master, correct their mistakes in learning, remind students to focus on reviewing the related content, and propose the goals to be achieved in the next stage. Through these functions, it can be found that the application of the cloud marking system is not a tool for teaching evaluation, but a supporter of teachers' teaching, helping them to make objective changes in teaching content and teaching methods and making the classroom more relevant and efficient by reacting to the assessment of the exams. It can also be used to discover the learning styles of different students by analyzing their performance and discovering their nearest developmental zone, to improve their enjoyment of learning through targeted teaching by selecting methods that provide indispensable help for both individualized and group teaching.

3. Intelligent Marking System Design

3.1. Natural Language Processing Techniques. A natural language is a language that evolves naturally with culture and is distinct from an artificial language, while programming languages such as C and VB are artificial languages. Natural language processing is the processing of the language we use to communicate in our daily lives. It is a cross-cutting discipline, with linguistics as the subject of study, and uses computer science to process it with the help of statistical and other disciplinary methods [16]. Although rule-based natural language processing methods have certain drawbacks, they are quite effective in word separation. In terms of counting the number of words, Chinese can directly count the length of the string, while English requires a word separation step. In well-written texts, there are spaces between words in English to indicate the independence of words. The same method is used to count the number of sentences in English, with each sentence ending followed by its punctuation mark. In general, English sentences usually end with a dot, question mark, exclamation point, and so on. The number of sentences can be counted according to the type of punctuation in the sentence.

Shallow text features such as the number of words and sentences can be extracted by rule-based methods, and the annotation of lexical properties, although it can be handled by retrieving the database of already marked words, is not accurate due to some words having multiple lexical properties and other reasons and statistical methods are used [17]; for the article content of subjective nature, using rule extraction is even less suitable. The scattered nature of article content topics makes it impossible to set rules comprehensively, and the setting of rules requires first reading a large amount of text to gain experience in summarizing the subject matter. This method is laborious and difficult to determine whether it fits the topic. As a result, a statistical-based approach was developed.

The early annotation of lexical properties was also realized by the method of rules, and linguists realized the creation of a rule base for lexical properties by summarizing the rules. One of them, the TAGGED system, was developed in the United States, which compiled 86 lexical classification rules. The accuracy rate is only 77% when the words in the Brown corpus are automatically annotated. In contrast, the Viterbi and CLAWS algorithms based on statistical methods can achieve an accuracy of more than 90%. The lexical annotation tool used in this paper is Stanford's POS Tagger, which is an open-source lexical annotation tool that can be used not only in the eclipse environment but also in python.

TF-IDF is a way to evaluate the importance of a word to the whole text or corpus, where the number of occurrences of an important word increases as the length of the text grows, and the frequency of its occurrence in the corpus tends to decrease. TF stands for term frequency, which means the frequency of occurrence of a word in the text. The formula for calculating TF is shown in (1), where $\text{count}(w)$ denotes the number of occurrences of word w in the document and $|D_i|$ denotes the number of all words in the document:

$$TF_{w,D_i} = \frac{\text{count}(w)^2}{|D_i|_2^3}. \quad (1)$$

IDF stands for inverse document frequency, which is a measure of the general importance of a word [23]. The IDF calculation formula is shown in equation (2), where N is the total number of all documents, and $I(w, D_i)$ indicates whether a keyword exists in document D_i , and its value is 1 if it exists; otherwise, it is 0:

$$IDF_w = \log \frac{M}{\sum_{i=1}^N I D(w, D_i)^2}. \quad (2)$$

Since the denominator in (2) cannot be 0 and the numerator cannot be 0 and the presence of some hot words will make the IDF value 0, the smoothing method is used, as shown in the following:

$$IDF_w = \log \frac{M + 1}{\sum_{i=1}^M I D(w, D_i)^2 + 1}. \quad (3)$$

In summary, the TF-IDF of a word is the product of the calculated TF and IDF values. A comprehensive analysis of this algorithm shows that the purpose of this algorithm is to filter out the common words and keep the important words in the article. It is not difficult to find a random article to read, and it is easy to find that the word "the" appears very frequently in Chinese, and in English, the word "the" is also a high-frequency word. But these words or words do not have actual meaning, so the data that has been divided into words can not only simply use the TF-IDF algorithm directly but also need to preprocess the text. The main task of text preprocessing is to remove the stop words, which mainly include some meaningless words such as auxiliary words, coronals, and conjunctions. Removing these meaningless stop words can greatly speed up the algorithm and also improve the efficiency of text keywords.

If all the important words in the document are used as keywords of the topic, then these keywords will be so many that they will not be particularly prominent for the real keywords in the document and will not reach the goal of detecting whether the topic is off-topic. Therefore, further feature selection is needed to extract the real keywords of the article. Feature selection can effectively reduce the dimensionality of the text space vector, simplify the text model, and improve the efficiency and accuracy of keyword extraction. The chi-square method is a statistical test to calculate the degree of correlation between words and article categories. Its basic idea is to determine the correctness based on the deviation between the theoretical and actual values, and its correctness is calculated as shown in the following:

$$\chi^2(t_i, c_j) = \frac{N \times (A \times D - C \times B)^2}{(A + C) \times (B + D) \times (A + B) + (C + D)}. \quad (4)$$

It is a good choice to filter out keywords by TF-IDF of cardinality statistics using the vector space model. It can map the keywords into spatial vectors for operation, by calculating the similarity of two vectors over space, thus indicating that the semantics of both are similar. Cosine

similarity is calculated using the cosine theorem, which can measure the similarity between two-word vectors [18]. The smaller the angle between the two-word vectors, the more similar the two-word vectors are. Similarly, two vectors are overlapping; then the angle between them is 0 and the cosine value is 1. If two vectors are orthogonal, the cosine value is 0. According to the magnitude of the cosine value, it can determine whether two keywords are related or not. The cosine similarity is calculated as shown in the following:

$$\sin(\theta) = \frac{a \cdot b}{\|a\|^2 \times \|b\|^2}. \quad (5)$$

English grammar checking usually uses statistical methods. If rules are used to check, not only is it time-consuming and laborious, but also some of the rules themselves can be conflicting. Suppose the number of English words present is N . If, according to the principle of the Bayesian algorithm, each word is independent of each other, the probability of a word appearing in any sentence should be $1/N$. But according to statistics, this is not the case. A complete English sentence generally begins with a subject, which rules out the possibility that many words are at the beginning. Second, in determining the first word, the words used later also cannot be judged based on the high probability of the word occurring in the corpus. The word “the” in English is the highest probability and cannot be followed by any word. The current state of the model depends only on the first few states, and Figure 1 shows the framework of natural language processing techniques.

For the extraction of sentence features, it is not possible to accurately fit the scores only from the form as an indicator of intelligent marking, but also from the grammatical aspects of the sentences. The tense shift of verbs and the singular and plural of nouns cannot be judged from the words alone but from the tense and number of words in the sentences where the words are located. The correctness of tense and singular-plural can be determined by locating the relevant words based on the lexical features extracted from the words.

Compound sentences include parallel sentences and main and subordinate sentences, and subordinate sentences are a plus in English composition. The complexity of various subordinate clause structures is more complicated than simple sentences, which means that it is more difficult for students to write a correct subordinate clause than a simple one. The difficulty is the only way to show the student’s writing ability. According to the grammar section of the English Advanced Placement syllabus, students need to master five types of main and subordinate complex sentences. If students use multiple subject-subject complex sentences several times in their essays, it will add a lot of colour to the essay.

3.2. Marking System Design. To study the intelligent marking algorithm, this paper obtained the dataset of senior high school mock exams. By collecting essay data and counting the scores, we found that the essay scores showed a certain normal distribution. The essays in the low and perfect score bands occupy a small portion, and the remaining score

bands occupy different proportions [19]. Because of the variability of the data in each score band, an intelligent marking model was designed to classify normal essays and low-scoring essays by using a triage scoring method. To ensure that the experimental model has a good fit, the essay scores selected for the training dataset involved all score bands. Since the school numbers of these high school students were randomly sorted according to the computer before enrollment, this paper selects the training and test sets in the dataset based on the score sorting, while the same score sorting strategy is based on the size of the school number. Firstly, the scores of the acquired sample sets are sorted in order from smallest to largest, and 70% of the sample sets in each score band are taken consecutively as the training set for this experiment, and the remaining samples are used as the test set. This ensures the comprehensiveness and randomness of the sampling.

The English simulation platform development and intelligent marking system involve three major parts: paper formation, mock exams, and paper marking [20]. The previous part of this paper mainly introduces and studies the intelligent marking algorithm part and does not introduce the design and implementation of the examination and marking part. To improve the whole system function of examination and marking, this chapter mainly introduces the design and implementation of the simulation platform and intelligent marking of the college entrance examination.

The main function of this system is to provide a way of independent learning for high school students, especially those who have been out of school but want to take the college entrance examination, and have some difficulties in getting timely marking from teachers; even for school students, there is a lot of pressure due to the tutoring of teachers who are facing the students’ college entrance examination. Students may not have anyone to mark the test questions they have done, and teachers may not have time to mark a large number of test papers. This system is a study system developed to address this actual social need for a simulated college entrance examination.

Students are the direct demanders of the system. The significant difference between using this system and traditional self-study is, first, time control and, second, timely feedback from the “teacher.” The so-called time control is that in the traditional learning process, learners often focus on whether they can solve the problem or not but often ignore the ability to complete the problem within a certain period. This system especially simulates the whole paper of the college entrance examination and the answer of each question type. And the so-called “timely feedback from teachers,” the system is to make up for the manual marking want to do but difficult to do “timely feedback:” on. You can say that the system can give you the results of the test “instantly” when the students submit their papers. Compared to manual marking, teachers usually take at least a day to give feedback.

In this system, students can take online exams after successfully logging into the system and selecting the exam function. The exam page should be set up with a countdown function so that if time runs out, the system automatically

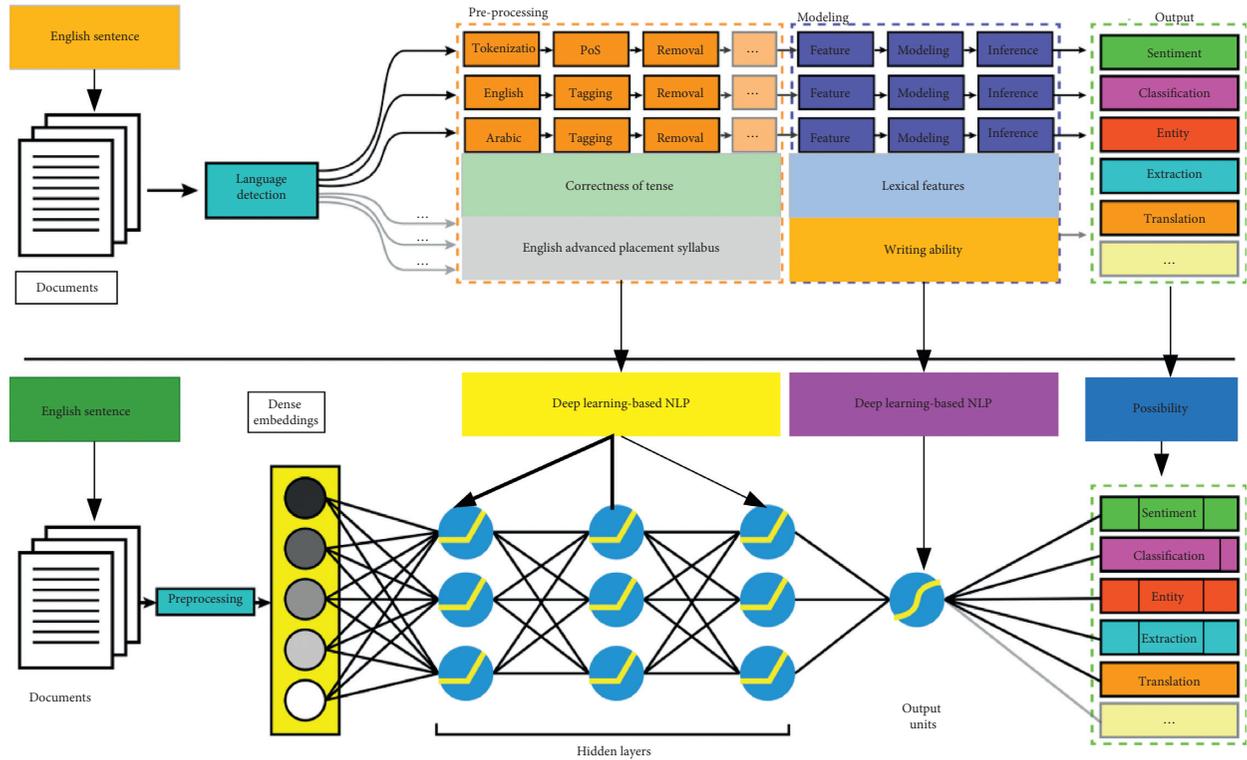


FIGURE 1: Natural language processing technology framework.

saves the answer sheet and submits the exam paper. At the end of the exam, you will be able to view the results in time and understand the reasons for your mistakes based on the analysis of the test questions. This allows students to quickly understand the correct answers to test questions based on the test questions while they are still fresh in their minds. In the online mock test system, students' test questions and answers can be recorded. When students need to find the test papers, they can quickly find the test papers directly through the search function, and such test papers are not easily lost. It not only saves students' time in searching for paper papers but also can be saved permanently without taking up limited desktop space. For students' mistakes, they can be added to the error book for future review. After a period of online exams, students' exam records will generate a graph showing the current trend of students' exam performance. To efficiently transmit multiple high-precision navigation signals in the L1 band, the GPS global navigation satellite system is designed to transmit multiple signals on the same center frequency, and the signal spectrums overlap severely, so internal interference becomes the primary consideration of the system. Through detailed analysis of the attenuation values of the carrier-to-noise ratio of the M, LIC, C/A, and $P(Y)$ codes on the GPS L1 band, it was found that the CDMA interference was the main source of internal interference, except for the C/A code, which was subject to more serious interference, and the other three signals, which were subject to less interference.

The system considers the difficulties of teachers in the marking stage and students in the correction stage after answering the papers and solves the problems in these two

aspects well. First, the auxiliary question system can analyze whether the text is suitable for the current test taker based on the reading comprehension test questions given by the teacher. It is mainly based on the proportion of word levels in the text. For example, if a high school student has a low level 6 vocabulary and the article contains more level 6 vocabulary, the test taker will have some difficulty in answering the questions. Second, students' responses are scored using intelligent marking technology after they answer the paper. The objective questions are marked using the string matching method, while the essay questions are marked using the intelligent marking technology proposed in this paper. The administrator, as the system administrator, is responsible for the management of teachers, students, and classes [21–24]. The administrator in this system is equivalent to the position of the head of the teaching and research team and reviews the group papers completed by the teachers and approves them before submitting the papers to the system. Exam management contains several important functions, setting up exams, and opening and closing exam rooms. Setting up exams allows you to set up the status, model, name of the exam room, and so on. After submitting the settings, the test papers are released to the system.

The most important function of the teacher role is the management of the question bank and the assembly of the papers. English test questions are slightly different from those of other subjects. While our common test system is one type of question with several subtopics, the English test is one type of question with several big questions and several subtopics under the big questions, the most typical of which

is reading comprehension. When you look at the English test in recent years, the reading comprehension is divided into four short essays, each with a different number of subtopics. For this kind of question with a variable number of subtopics, it is not possible to simply use a fixed number of subtopics under each big question. Therefore, this system adopts a dynamic adjustment method by first entering the stem part of the test questions, and the remaining subtopics form a correlation with the major questions. The database table design for the major and minor questions of the test is shown in Table 1.

The English simulation platform and intelligent marking system of the college entrance examination involve three major roles of teachers, administrators, and students, and its main flow chart is shown in Figure 2.

To improve the security of the system and prevent unscrupulous people from using machine methods to crack user passwords, the system requires a verification code to be entered at login. Captcha is a good way to prevent machines from combining user names to try passwords multiple times and thus achieve the purpose of password cracking. The machine cannot recognize or recognize the verification code well enough to crack the password in a short period. At the same time, the system will temporarily stop accepting logins for each user with more than 5 incorrect login passwords. This feature is implemented by adding the number of login errors and the time of the 5th login error to the user database. If the user fails to login 5 times in a row, the login failure time will be registered and a pop-up message will appear that the user can try to login again after 5 minutes. Five minutes later, if the user logs in successfully, the login failure time will be changed to 0 and the 5th login failure time will be cleared.

Based on the above, this system uses separate entries of large and small questions and sets the difficulty of the questions when entering the test questions. When teachers enter the test questions, the first screen they enter is the large question entry screen. The questions are entered as either the stem information of the question or the text content of the reading comprehension. To make it easier to search by conditions, you need to select the range of questions when entering the questions.

The drop-down list contains all the question types in the GCE English exam, and you can set the category to which the questions belong by selecting the question type. To control the difficulty of the test questions when grouping the papers, you need to select the test difficulty of the current test questions. The entry of large questions also has requirements for entry, and all necessary fields need to be filled in before the current entry of large questions can be submitted. If there are required fields that are not filled, the system will use the red font to indicate the content items that do not meet the requirements.

4. Analysis of Results

4.1. Intelligent Algorithm Results Analysis. In this paper, we test the scoring runtime of the traditional VSM and the hybrid intelligent scoring model of this paper with different

TABLE 1: Data table of test papers.

Field name	Type	Value
RId	Int	1
Num	Int	3
QPIId	Int	4
StemId	Int	5

numbers of student response texts according to the description of the test criteria, as shown in Figure 3. Also, the scoring runtimes of the traditional VSM and this paper's hybrid intelligent scoring model are tested for different numbers of student response text feature vectors, as shown in Figure 3. The execution time of both the hybrid model and the traditional VSM model increases as the number of student response texts and the number of feature word vectors in each response text increases. However, due to the introduction of LDA for text dimensionality reduction in this model, the running time of this model increases linearly and rapidly when the number of test texts and feature word vectors increases, while the running time of the traditional VSM has almost no significant effect. Runtime is inferior to that of the traditional vector space model (VSM).

The accuracy test in this paper refers to the agreement between the scores evaluated by the hybrid intelligent scoring model and the scores evaluated by experts. The root mean square error (RMSE) can accurately and effectively reflect the small errors between the scores of the intelligent scoring system and the scores of the experts; that is, the smaller the RMSE, the closer the scores of the intelligent scoring system and the scores of the experts. Therefore, the root mean square error value is chosen as the basis for text accuracy testing and analysis in this paper. The intelligent scoring model in this paper is very close to the scores given by the experts at $\lambda = 0.1$, which indicates that the hybrid intelligent scoring model in this paper has high accuracy. To further test the accuracy of the hybrid intelligent scoring model, we tested the root mean square error (RMSE) for a different number of student responses and traditional VSM scores when $\lambda = 0.1$, as shown in Figure 4. When the value of λ is appropriate, the RMS error of this model is smaller than that of the traditional model, indicating that the accuracy of this model is better than that of the traditional VSM.

At the same time, because of the linear combination of the intelligent scoring models in this paper, the root means a square error of the two models when the λ value is changed in the model can also reflect the accuracy of the traditional VSM and the model in this paper, as shown in Figure 5. This further indicates that the accuracy of the intelligent scoring model is better than that of the traditional VSM when the λ value is chosen appropriately.

The performance of building the hybrid intelligent scoring model and the traditional VSM model in terms of time efficiency is compared according to the predefined test cases, such as changing the number of student response texts and the number of student response text feature vectors, and the performance of building the hybrid intelligent scoring model and the traditional VSM model in terms of space consumption is compared under the same test cases. At the

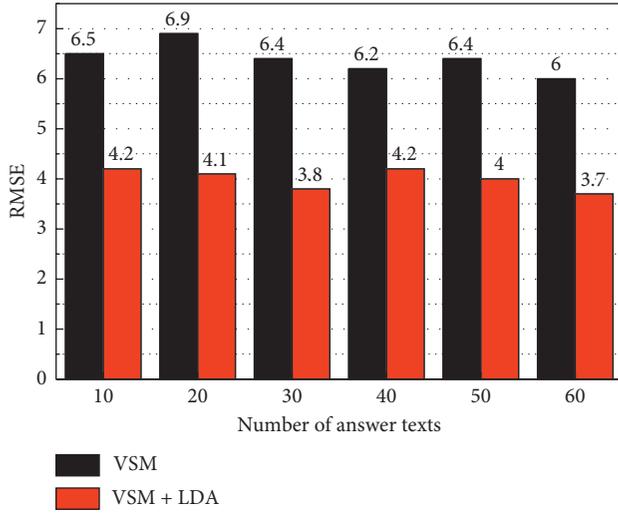
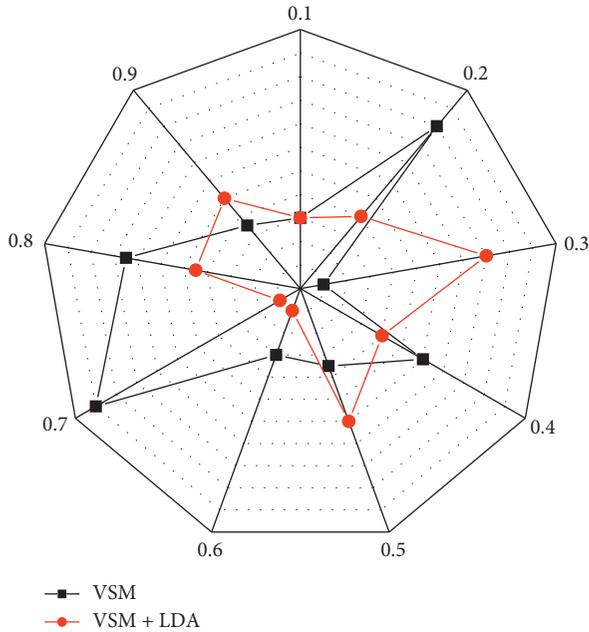


FIGURE 4: Root mean square error of scores.

FIGURE 5: Effect of varying λ values on the root mean square error of the scores.

be covered by testing according to what the code expresses. Black-box testing is mainly to test the implementation of various functions to ensure that the software can achieve the designed purpose.

The experimental data of the four core algorithms were obtained by running them 30 times independently on the MATLAB bench, and the maximum number of iterations was set to 100 for each run. Figure 7 shows the mean and variance of the four algorithms obtained after running them 30 times independently, and the optimal values are bolded.

The smaller the value, the better the learning path matches the actual needs of the learners and the higher the learning quality; on the contrary, it does not match the needs of the learners. By comparing these three sets of

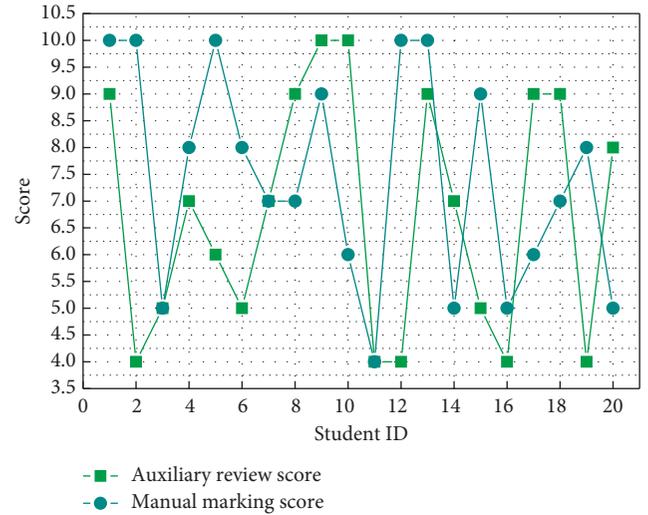


FIGURE 6: Comparison of scores of sample application problems.

experimental data in Figure 7, we find that with the increase of the number of knowledge points, the MABPSO algorithm has the best convergence accuracy, but the variance is the best only in Experiment 1, which is not much different from other algorithms, indicating that the stability of the MABPSO algorithm is acceptable. The data of Experiment 2, Experiment 4, Experiment 5, and Experiment 6 differ only in the number of learners. By observing the data of these three groups of experiments in Figure 7, we find that the MABPSO algorithm has the best convergence accuracy as the number of learners increases, and the same variance is not the best, but not much different from other algorithms, which indicates that the stability of the MABPSO algorithm is acceptable.

The convergence plots of the algorithms can clearly show the optimization search process of each algorithm, and the comparison of the optimization search process can observe the optimization search performance of each core algorithm. Figure 8 shows the convergence graph of the experimental repulsive degree function.

By comparing the mean convergence curves of the total optimization function fitness under different numbers of knowledge points, we can see that the MABPSO core algorithm has the best convergence speed and convergence stability, which indicates that the personalized learning paths optimized by the MABPSO core algorithm meet the needs of learners more with the increase of the number of knowledge points and show better optimization speed and matching degree. Firstly, the idea of the personalized learning path optimization method is introduced as a whole; secondly, the personalized learning path is regarded as a combinatorial optimization problem with minimization solution, and the solution process of the personalized learning path optimization method is elaborated; finally, the operational performance of the personalized learning path optimization method is analyzed in terms of the optimization-seeking accuracy and the optimization-seeking process, by comparing with the basic BPSO algorithm, LPSO algorithm, RPSO algorithm, and RPSO algorithm.

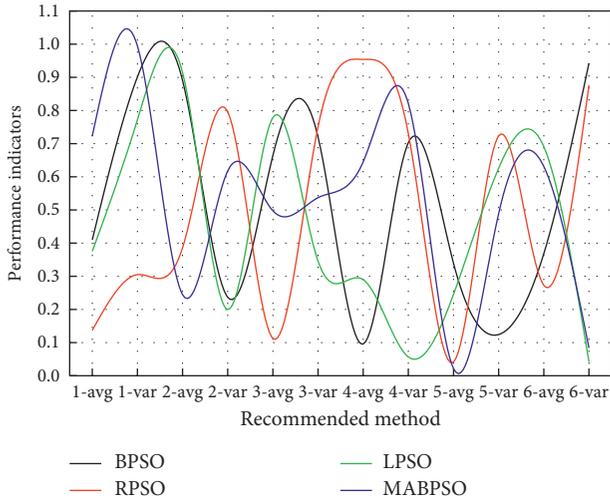


FIGURE 7: Performance indicators.

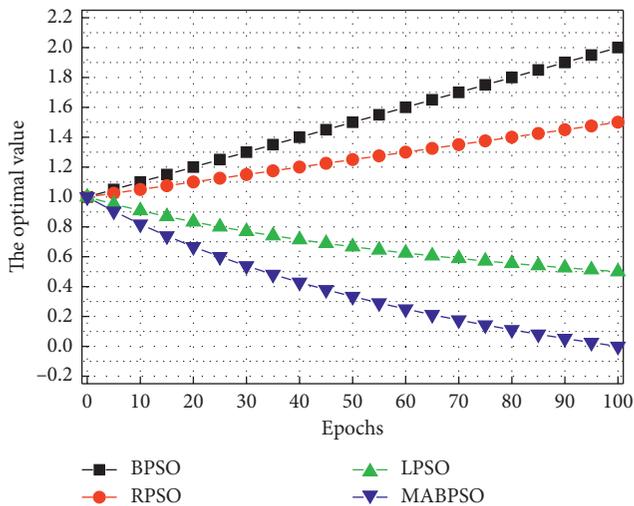


FIGURE 8: Convergence curve.

5. Conclusion

For the essay scoring algorithm, the support vector regression algorithm was used for normal essays that meet the requirements of the questions. Due to the high score of the essay, using the classification algorithm requires classifying each scored segment, which may result in high time complexity affecting the efficiency of scoring execution. Therefore, a support vector regression algorithm was used to fit a regression curve to the scores of normal essays, while the scores queried by students were presented using rounding. A study of English writing and reading of domestic and international literature revealed that there is no such item as the sense of voice in the features of essay scoring. This paper reflects students' English learning level from the side by examining English language sense. It also draws on domestic and international extracts of composition features to assess students' compositions comprehensively. This system adopts the correlation of major and minor questions and realizes

the entry of English test questions by cyclically entering minor questions associated with the corresponding major questions. Through the organic combination of the two, students can both take the test online according to the teacher's needs and to achieve the marking of the whole set of test papers through the computer. It solves the problem that the mainstream test systems in the market cannot mark subjective questions, adds the extraction of English language features, and improves the model of the scoring algorithm to achieve the purpose of improving accuracy.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] A. Winkler-Schwartz, V. Bissonnette, N. Mirchi et al., "Artificial intelligence in medical education: best practices using machine learning to assess surgical expertise in virtual reality simulation," *Journal of Surgical Education*, vol. 76, no. 6, pp. 1681–1690, 2019.
- [2] B. A. Talkhestani, T. Jung, B. Lindemann et al., "An architecture of an intelligent digital twin in a cyber-physical production system," *At-Automatisierungstechnik*, vol. 67, no. 9, pp. 762–782, 2019.
- [3] O. Almousa, J. Prates, N. Yeslam et al., "Virtual reality simulation technology for cardiopulmonary resuscitation training: an innovative hybrid system with haptic feedback," *Simulation & Gaming*, vol. 50, no. 1, pp. 6–22, 2019.
- [4] L. Mora, R. Bolici, and M. Deakin, "The first two decades of smart-city research: a bibliometric analysis," *Journal of Urban Technology*, vol. 24, no. 1, pp. 3–27, 2017.
- [5] F. Corradini, F. De Angelis, A. Polini, C. Castagnari, J. De Berardinis, and G. Forcina, "Tangramob: an agent-based simulation framework for validating urban smart mobility solutions," *Journal of Intelligent Systems*, vol. 29, no. 1, pp. 1188–1201, 2019.
- [6] L. Mora, M. Deakin, A. Reid, and M. Angelidou, "How to overcome the dichotomous nature of smart city research: proposed methodology and results of a pilot study," *Journal of Urban Technology*, vol. 26, no. 2, pp. 89–128, 2019.
- [7] S. Mehmi, H. K. Verma, and A. L. Sangal, "Simulation modeling of cloud computing for smart grid using Cloud-Sim," *Journal of Electrical Systems and Information Technology*, vol. 4, no. 1, pp. 159–172, 2017.
- [8] A. Towers, J. Field, C. Stokes, S. Maddock, and N. Martin, "A scoping review of the use and application of virtual reality in pre-clinical dental education," *British Dental Journal*, vol. 226, no. 5, pp. 358–366, 2019.
- [9] Y. Ma, Z. Wang, H. Yang, and L. Yang, "Artificial intelligence applications in the development of autonomous vehicles: a survey," *IEEE/CAA Journal of Automatica Sinica*, vol. 7, no. 2, pp. 315–329, 2020.
- [10] W. Wu, A. Berestova, A. Lobuteva, and N. Stroiteleva, "An intelligent computer system for assessing student performance," *International Journal of Emerging Technologies in Learning (ijET)*, vol. 16, no. 02, pp. 31–45, 2021.

- [11] Q. Li, F. Qiao, X. Wang, and L. Yu, "Drivers' smart advisory system improves driving performance at STOP sign intersections," *Journal of Traffic and Transportation Engineering (English Edition)*, vol. 4, no. 3, pp. 262–271, 2017.
- [12] J. S. Malak, H. Zeraati, F. S. Nayeri, R. Safdari, and A. D. Shahraki, "Neonatal intensive care decision support systems using artificial intelligence techniques: a systematic review," *Artificial Intelligence Review*, vol. 52, no. 4, pp. 2685–2704, 2019.
- [13] M. Lasen, S. Evans, K. Tsey, C. Campbell, and I. Kinchin, "Quality of WIL assessment design in higher education: a systematic literature review," *Higher Education Research & Development*, vol. 37, no. 4, pp. 788–804, 2018.
- [14] P. R. Smart, "Mandevillian intelligence," *Synthese*, vol. 195, no. 9, pp. 4169–4200, 2018.
- [15] M. Banerjee, J. Lee, and K.-K. R. Choo, "A blockchain future for internet of things security: a position paper," *Digital Communications and Networks*, vol. 4, no. 3, pp. 149–160, 2018.
- [16] W. J. González, "From intelligence to rationality of minds and machines in contemporary society: the sciences of design and the role of information," *Minds and Machines*, vol. 27, no. 3, pp. 397–424, 2017.
- [17] J. Böke, U. Knaack, and M. Hemmerling, "State-of-the-art of intelligent building envelopes in the context of intelligent technical systems," *Intelligent Buildings International*, vol. 11, no. 1, pp. 27–45, 2019.
- [18] I. Lochhead and N. Hedley, "Mixed reality emergency management: bringing virtual evacuation simulations into real-world built environments," *International Journal of Digital Earth*, vol. 12, no. 2, pp. 190–208, 2019.
- [19] R. A. Rojas and E. Rauch, "From a literature review to a conceptual framework of enablers for smart manufacturing control," *The International Journal of Advanced Manufacturing Technology*, vol. 104, no. 1–4, pp. 517–533, 2019.
- [20] G. Jing, T. Tosun, M. Yim, and H. Kress-Gazit, "Accomplishing high-level tasks with modular robots," *Autonomous Robots*, vol. 42, no. 7, pp. 1337–1354, 2018.
- [21] K. Macauley, T. J. Brudvig, M. Kadakia, and M. Bonneville, "Systematic review of assessments that evaluate clinical decision making, clinical reasoning, and critical thinking changes after simulation participation," *Journal of Physical Therapy Education*, vol. 31, no. 4, pp. 64–75, 2017.
- [22] W. Wei, J. Wu, and C. Zhu, "Special issue on deep learning for natural language processing," *Computing*, vol. 102, no. 3, pp. 601–603, 2020.
- [23] J. Li, G. Feng, W. Wei et al., "Psotrack: a rfid-based system for random moving objects tracking in unconstrained indoor environment," *IEEE Internet of Things Journal*, vol. 5, no. 6, pp. 4632–4641, 2018.
- [24] H. Hu, B. Tang, X. Gong, W. Wei, and H. Wang, "Intelligent fault diagnosis of the high-speed train with big data based on deep neural networks," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 4, pp. 2106–2116, 2017.