Network Algorithms for Intelligent Evaluation of Composition in Middle School English Cloud Classrooms

Yaohua Huang, School of Foreign Languages, Suqian University, Suqian, China*
Chengbo Zhang, Suqian University, Suqian, China

ABSTRACT

This study combines link grammar (LG) detector with N-grammar model to analyze and evaluate grammar in compositions. And then the composition level is judged through information entropy. Finally, the composition score is calculated based on the overall composition level and grammar weight. The experimental results show that the combined weight of recall and accuracy of the proposed method in this study is 89.9%, which is 26.6% higher than LG and 9.7% higher than Grammarly. In the performance test of scoring the entire essay, the proportion of error between the proposed method and manual evaluation is 87.29%, with a lower overall mean square error of only 3.08 and a shorter average running time of only 22.69 seconds. The method proposed in this study has a high accuracy and strong applicability in the evaluation of English compositions for middle school students, providing a new approach for teaching English writing for middle school students.

KEYWORDS

Cloud Classroom, Intelligent evaluation system, LG, N-grammar model

INTRODUCTION

In recent years, due to the increasingly heavy task of English learning for middle school students, there has been an urgent need to rapidly improve students' English writing ability (Kong & Lee, 2021). When scoring English compositions, due to significant differences among students, teachers need to spend a lot of time reviewing and explaining and cannot balance the discrepancies in student abilities. Teaching students according to their aptitude results in students being unable to receive timely feedback and engage in large-scale writing exercises (Albiansyah & Minkhatunnakhriyah, 2021). At present, the development of automatic scoring (AC) systems for objective questions in foreign countries is relatively mature, but the research on scoring systems for subjective questions such as compositions is not yet as developed (Zhao, 2021). The Link Grammar (LG) analyzer is a key component of an AC system. It can quickly evaluate English sentences and is robust. However, it only targets complete sentences, and the number of dictionaries is limited, making it difficult to

DOI: 10.4018/IJWLTT.337967
*Corresponding Author

accurately recognize abbreviated and complex sentences. It is easy to misjudge correct sentences, which affects the accuracy of the entire composition score (Chen & Liang, 2022). The N-grammar model can not only filter out misjudged sentences using binary standards but also improve the output results. To design a more accurate composition scoring model, this research introduces an N-grammar model, combines LG with the N-grammar model, and analyzes and evaluates grammar in composition; it then uses information entropy (IE) to judge composition level, synthesizes the grammar weight of each sentence to score grammar, and finally calculates composition score by combining composition level with overall grammar weight. The model aims to shorten the time for teachers to evaluate compositions, improve their work efficiency, and increase opportunities for students to practice English writing. Detailed analysis of grammar in research can also help students to improve their grammar and improve their English writing ability. The innovation of this research lies in the design of a human-machine integrated English composition intelligent scoring (CIS) system, which improves the accuracy of composition scoring. The research content is divided into four parts: The first part is a brief introduction to the relevant research on cloud classrooms and scoring systems; the second part first introduces two types of grammar detectors in detail and then constructs a CIS model based on these grammar detectors; the third part is the application of the CIS model, during which performance testing and comparative analysis experiments are conducted, and the results of this application; and the fourth part is a summary and outlook of the research content.

RELATED WORKS

Recently, an increasing number of courses have transformed into cloud classroom teaching models. Jiao and Liu (2021) applied cloud classrooms to ideological and political learning. The study delved into the actual meaning and implementation of cloud classrooms and conducted a detailed analysis of specific cases. The results showed that the cloud classroom was fully flexible in managing the classroom, and the user experience was good (Jiao & Liu, 2021). Liang et al. (2021) designed a survey questionnaire for intelligent classrooms to investigate various stages of the classroom. The intelligent teaching model was improved after analyzing and processing the survey results. This study utilized cloud technology to redesign the division of functional modules and the overall system architecture and put the model into practical development and use. The results indicated that the method had high accuracy (Liang et al., 2021). Xiong and Li (2021) found that most scholars in their research on English cloud classrooms overlooked temporary information and did not take into account students' changing states in the classroom. This study investigated the opinions of teachers from various countries on classroom management and managed cloud classrooms based on moderating and mediating variables. Finally, the relationship between cheating and teachers' perception of the classroom was analyzed (Xiong & Li, 2021). Jing et al. (2020) found that traditional teaching methods were insufficient to assess students' learning status and made timely adjustments to teaching methods. Therefore, research was conducted on improving the Kmeans algorithm and constructing an online learning platform model to monitor students' learning status. The test results showed that the algorithm had certain applicability in flipped classrooms (Jing et al., 2020). Li and Juan (2021) proposed a new video denoising method in their research on English cloud classrooms. This method utilized a camera motion compensated flow model to generate denoised videos. The results showed that the proposed algorithm had stronger robustness and achieved good denoising results (Li & Juan, 2021). Wang et al. (2022) collected and created a dataset on students' reactions to textbooks in English cloud classrooms. The results showed that textbooks were of great help to students' self-awareness and English learning, which could enhance their reading energy while also improving their English proficiency (Wang et al., 2022).

The development of cloud classrooms has become relatively mature, and as a key part of English cloud classrooms, CIS is still in the initial stage of development. Liu et al. (2021) redesigned a model to address insufficient intelligence in the English CIS system. Firstly, this study utilized intelligent image recognition technology to improve machine learning algorithms and proposed a pseudo-character region filtering algorithm, which was improved on the basis of convolutional neural networks. Then it selected an essay to input into the model and made performance verification analysis on the proposed model. The results showed that the model performed well and had good applicability in English CIS systems (Liu et al., 2021). Zhang (2021) has constructed a new English CIS model aiming to accurately grade and fully reflect students' writing. Firstly, this study referred to existing AC systems both domestically and internationally. It conducted a detailed analysis and discussion on the feature selection methods of these systems, taking into account the impact of linguistics on scoring systems. Then, multiple regression methods were used to evaluate the scoring results. Finally, performance testing was conducted on the model proposed in this study. The results showed that this method was superior to traditional methods (Zhang, 2021). In the research of English grammar detectors, Gakis et al. (2021) investigated the use of detectors in teaching and their specific contributions to teaching for middle school students. The survey results showed that the errors in the text were mainly template and syntax errors, so the detector could be further adjusted and studied (Gakis et al., 2021). Gaillat et al. (2022) designed a new CIS system based on students' English proficiency, aiming to identify the characteristics of standards to which English writers commonly refer. The research method was based on the concept of microsystems, linking learners' language system characteristics and conducting grammar analysis. The results showed that different microsystems were able to help group compositions classify, with a classification accuracy of 82% (Gaillat et al., 2022).

In summary, although a large number of scholars have conducted research on cloud classrooms and designed numerous improved models, there is still relatively little research on middle school students' English CIS model, which has strong potential application value for middle school students' English learning.

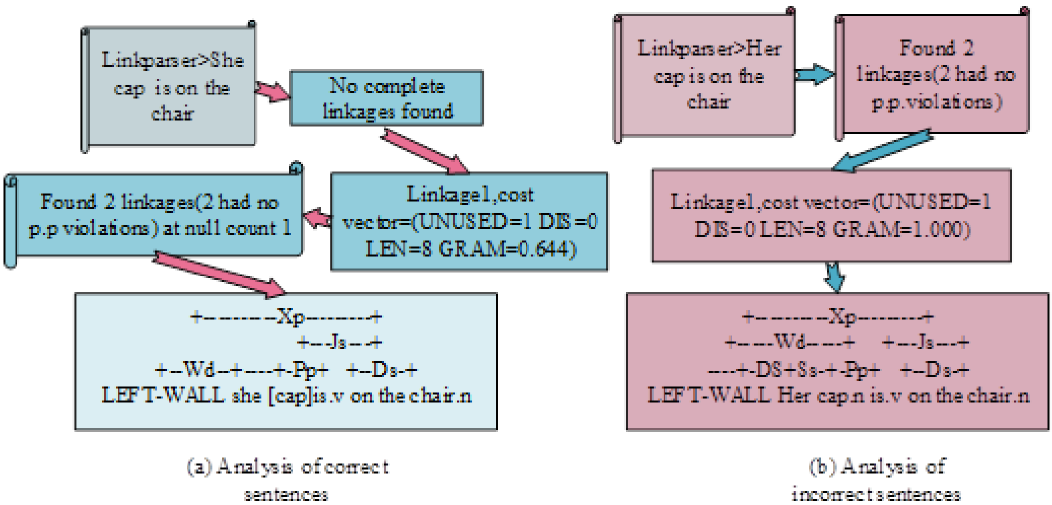
IMPROVEMENT OF CIS MODEL

A Grammar Detector Based on N-Grammar Model

Grammar detectors are an important part of CIS, since they can detect syntax errors in English compositions. A qualified grammar detector should also accurately identify the location and category of syntax errors and give corresponding suggestions for modification (Liang et al., 2021). This study redesigns a grammar detection system based on LG and N-grammar models. Chain grammar is a unique syntax structure that compares the value of an object, value, or variable with the value of other objects or variables and combines them into a new object or variable. In chain syntax, comparison and combination are achieved through linked lists between pointers, which are ordered nodes of pointers. N-grammar is a grammatical structure based on metalinguistics which is expressed in the form of natural language. In N-grammar, each word has a unique meta grammar symbol, which indicates that the word starts with "meta" and can be extended with "_". N-grammar simplifies complex grammatical structures into more intuitive and understandable forms, thereby achieving better language communication. The detection system studied not only has the characteristics of the traditional grammar detector mentioned above but also has a high response speed because of its unique statistical method, which shortens the running time. The following will provide a detailed introduction to the operation of LG and N-grammar models (Mansfield & Barth, 2021). LG is written in C language and has strong robustness. It mainly focuses on whether the sentence satisfies connectivity, planarity, and order. If these three principles are met, it judges the sentence as grammatically correct, otherwise it is incorrect. By using this method, LG can quickly determine the correct or incorrect grammar in a sentence. For example, the analysis results of the sentences, "[s]he cap is on the chair" and "[h]er cap is on the chair," are shown in Figure 1.

Figure 1 (a) shows the output form of LG's sentence, "[s]he cap is on the chair." Because there are syntax errors in the sentence, the complete link set cannot be found during the analysis. The figure shows the chaining situation after skipping the unlinked word, "cap." The number of unlinked words

Figure 1. LG's analysis of sentences



is represented by UNUSED, and the grammatical weight of sentences is represented by GRAM. The UNUSED value of this sentence is 1, and the GRAM value is 0.644. Figure 1 (b) shows the analysis results of, "[h]er cap is on the chair." Because the sentence is completely correct, the GRAM value is 1 and the UNUSED value is 0.

LG still has some shortcomings, such as low accuracy in judging complex sentences and inaccurate analysis of abbreviated sentences. Introducing the N-grammar model can improve the judgment of LG, improve the output results, and thus improve the accuracy of sentence diagnosis. The running mode of the N-grammar model is to predict the probability of the next word appearing based on historical words (Näf, 2021). $W = W_1W_2W_3...W_n$ stands for sentence, and $P(W_n)$ stands for the probability of the word W_n appearing in $W = W_1W_2W_3...W_{n-1}$. According to the conditional probability, the expression is shown in (1).

$$P(W) = P(W_1)P(W_2 \mid W_1)P(W_3 \mid W_1W_2)...P(W_n \mid W_1W_2...W_{n-1}) \quad (1)$$

As shown in (1), if it is necessary to calculate the probability $P(W_n)$ of word occurrence, all conditional probability needs to be calculated. The calculation is complex and further simplification of the model is needed (Sasaki et al., 2021). It is not necessary to review all words in history if assumptions are introduced using the Markov method (Tavakol et al., 2022).

$$P(W) = \prod_{i=1}^{n} P(W_i \mid W_{i-N+1}...W_{i-1}) \quad (2)$$

As shown in (2), $N-1$ in the model represents words that have already appeared in history. N indicates the word that needs to be predicted.

$$P(W_i \mid W_{i-1}) \approx count(W_i \mid W_{i-1}) / countW_{i-1} \quad (3)$$

As shown in (3), the frequency of words appearing in the corpus is represented by $count(\ldots)$. When using the N-grammar model in practice, it is conventional to use the binary model of $N = 2$ or the ternary model of $N = 3$. The accuracy of the N-grammar model increases as N increases, and the computational complexity also increases exponentially. The feedback flowchart of the grammar detector combining LG and N-grammar model is shown in Figure 2.

In Figure 2, after students submit their English compositions, the computer first starts initialization work and then preprocessing, the main purpose of which is to avoid the impact of special characters on subsequent grammar analysis. Then, it uses LG for grammar analysis to determine its chain formation. If the chain is not formed, it will be transferred to Stanford Parser analysis. Stanford Parser is used to extract binary structures of non-linked words and match them in the database. If the frequency of F is greater than T, it means that there is no syntax error; otherwise, it means that there is an error. Finally, it outputs the error category and binary structure.

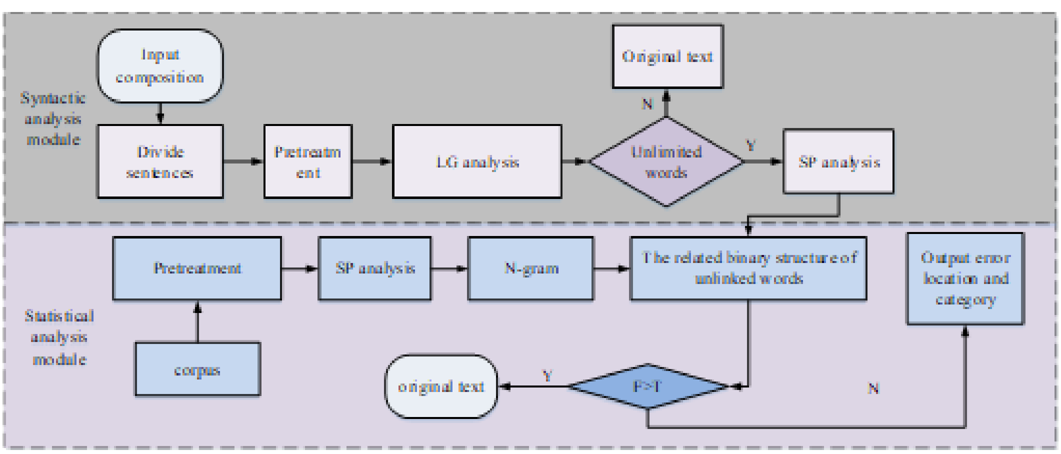
After completing the grammar testing of the essay, it is necessary to classify and quantify the topic of the essay, providing an analytical basis for subsequent composition scoring (Mastick et al., 2022). Traditional methods cannot perform quantitative analysis. This study first preprocesses the composition and collects word sets related to the theme words. The IE of composition is calculated by evaluating the correlation between the composition word set and the topic set. Finally, based on the total number of words in the composition, the goal of classifying and quantifying the topic of the article is achieved (Lai et al., 2022). The calculation of IE is shown in (4).

$$H(X) = \sum_{x \in X} q(x) I(x) = -\sum_{x \in X} q(x) \log_b q(x) \tag{4}$$

As shown in (4), $I(X)$ denotes the random variable in IE. It follows that $\{X, q(x)\}$ refers to the mathematical expectation of random variables.

$$X = \begin{cases} 1, P = q \\ 0, P = 1 - q \end{cases} \tag{5}$$

Figure 2. Feedback flow chart of LG-N-grammar detector



In (5), the random variable X obeys the binomial distribution. P represents the probability of occurrence of the event. q means the probability of non-occurrence of the event.

$$H(X) = -q\log q - (1-q)\log(1-q) \tag{6}$$

In (6), when q approaches 1, the value of $H(X)$ decreases at an accelerated rate; When $q = \frac{1}{2}$, $H(X)$ reaches its maximum value.

$$H_{rate} = \frac{1}{n}H(X_1 ... X_n) = -\frac{1}{n}\sum_{i=1}^{n} q(x_i)\log q(x_i) \tag{7}$$

In (7), n indicates the length of the information. H_{rate} denotes character entropy.

$$H = -\sum_{j}^{n}\sum_{i}^{h} p_{ij} Inp_{ij} \tag{8}$$

In (8), H refers to the IE of the composition. It sets the total number of words as h, and $w_i \left(0 \le i \le h\right)$ is the i th word; $k_j \left(1 \le j \le h\right)$ denotes the j th theme word. The correlation between the i th word in the composition and the j th word in the topic set is expressed in p_{ij}. To avoid the increase of IE caused by words unrelated to the theme of the composition, the model takes the average value of IE for more detailed scoring. The average IE, H_{ave}, is shown in (9).

$$H_{ave} = H / W_{total} \tag{9}$$

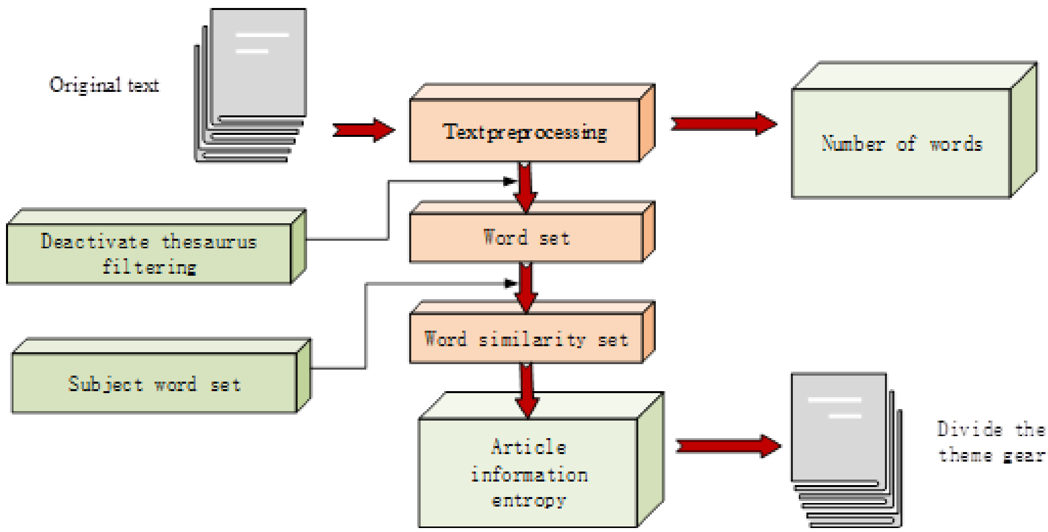
As shown in (9), the number of words in the article is represented by W_{total}.

Figure 3 is the flowchart of English CIS model. As shown in Figure 3, after submitting the text, the characters and abbreviations need to be processed first, and then the word set is obtained through word and sentence segmentation of the text. Then the word set is compared with the topic word set to calculate the degree of correlation. Finally, the IE of composition is calculated according to the correlation degree of the word set to define the level of composition.

Grammar Scoring and Automatic CIS Algorithms

An LG grammar detector is a computer-aided language model that can detect the correctness of statements based on their grammar rules. According to syntax rules, an LG detector can determine whether a statement conforms to a specific form, such as whether there should be a punctuation mark. When using LG detectors, errors in statements can be detected more quickly and corresponding solutions can be adopted. Based on the LG detector, this study determined the weight distribution and size of grammar based on the grammar score of each sentence. The overall idea is that for each complete sentence in the composition, non-linked words divide the sentence into separate parts. Firstly, the study evaluates the completion of sentences by using the number of non-linked words. Then, for the complete part, the grammar weight of the sentence is determined by the number of violations of post-processing rules and the consumption of grammar operators. Finally, the grammar weight of

Figure 3. Composition scoring process



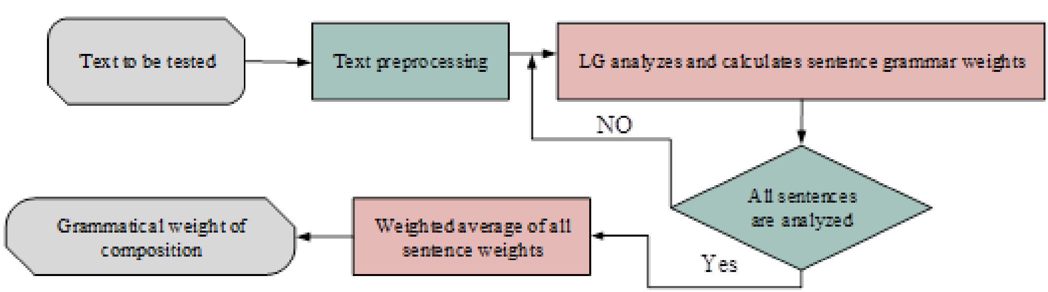
the composition comes from the weighted average of the grammar weight of each sentence. On this basis, the calculation of grammar score weights for compositions is shown in Figure 4.

Figure 4 shows that the grammar scoring of each sentence determines the size of the grammar weight of the composition. The specific algorithm for the grammar score of the sentence is shown below. Normal chained words are represented by W, and the smaller value between the two is represented by the $\min$ function. It assumes that W_0 represents a non-chained word in a sentence, and P_M means its position in the sentence.

$$B = \min(P_M - P_L, P_R - P_M) \tag{10}$$

In (10), B stands for the contribution value of W_0 to sentence fragmentation; P_L means the position of unused words on the left side of P_M. P_R refers to the position of unused words to the right of P_M.

Figure 4. Grammar scoring flow chart



$$BrokenDegree = \begin{cases} 1 / N, & m = 1 \\ (\sum_{i=1}^{m} B_{W_i}) / N, & m > 1 \end{cases} \tag{11}$$

In (11), N indicates the length of the sentence. $m(0 \leq m \leq N)$ is the number of non-linked words; the fragmented contribution value of the i non-linked word is represented by B_{w_i}. When using LG for analysis, multiple analysis results will be obtained. Each result corresponds to a grammar operator cost value. The cost value reflects the degree to which the results conform to English grammar conventions. The higher the value, the less in line with grammar conventions. Therefore, LG selects the result with the lowest cost value as the optimal output in the final output. The calculation is shown in (12).

$$DisjunctDegree = m / LEN \tag{12}$$

In (12), LEN means the chain length. m is the cost value. $DisjunctDegree$ denotes the degree of consumption of grammar operators. When using LG to analyze fixed grammar, it is not possible to determine the legality of sentence chaining. Here, a post-processing rule is introduced, which divides sentences into different domains and judges the legality of their chains based on its rules.

$$ViolationDegree = m / N \tag{13}$$

In (13), m indicates the number of violations of the rule; $ViolationDegree$ stands for the degree of violation of the rule. From the above three formulas, the grammar weight comprehensively considers the completeness of sentences, the consumption of grammar operators, and the degree of violation of post-processing rules. The comprehensive weight of the sentence is shown in (14).

$$GRAM = (1 - BrokenDegree) * (1 - ViolationDegree) * (1 - DisjunctDegree) \tag{14}$$

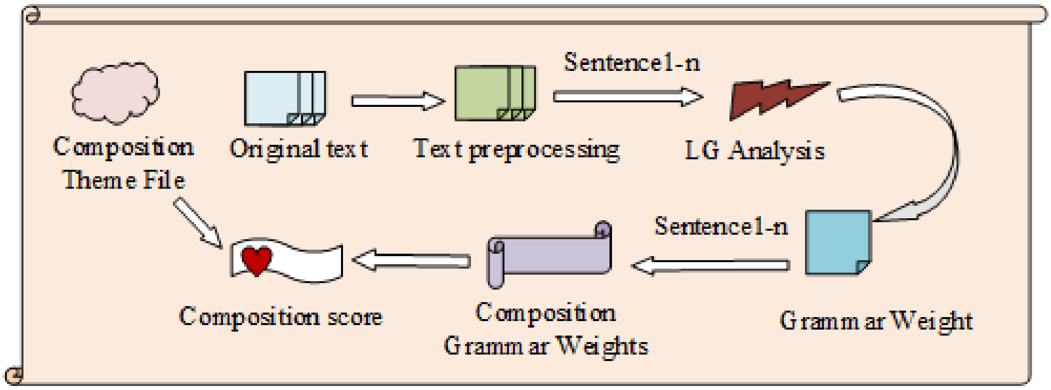
In (14), $GRAM$ is the grammatical weight of the sentence. After calculating the weight of each sentence in the composition, the grammatical weight of the entire composition is calculated using a weighted average method.

$$\bar{G} = \frac{1}{N} \sum_{i=1}^{N} G_i \tag{15}$$

In (15), N is the total number of sentences in the composition. $G_i(0 \leq i \leq N)$ shows the weight of each sentence, and $\bar{G}$ is the average grammatical weight of all sentences. Finally, the entire composition is graded based on the theme scoring and grammar weights, as shown in Figure 5.

In Figure 5, composition score mainly involves four steps. The first step is to preprocess and segregate the input text and mark the sentence after the segmentation as $T = (T_1 T_2 ... T_n)$. The second step is to use the LG grammar detector for analysis, outputting the cost of link grammar operators and the number of post-processing violations and non-linked words. After comprehensive calculation, the grammar weight of each sentence is output. Step three is to calculate the grammar weight of the entire composition based on the grammar weight of each sentence. Step four is to calculate the

Figure 5. Composition scoring flow chart



composition score based on the level and grammar weight of the composition (Zhu & Schlick, 2021). The calculation is shown in (16).

$$Score = ((6 - Level) / 5) * GRAM * F\_mark \tag{16}$$

In (16), F_mark denotes the full score of the composition. $Score$ indicates the actual score of the composition; the composition is divided into five levels, and the lower the level, the more relevant it is to the topic. Assuming that the composition belongs to the first tier, with a weight of 0.5 and a maximum score of 30, the final score of the composition is 15 points. This study mainly focuses on the specific scoring of compositions, and the actual essay scoring also includes sections such as manual supplementation and modification (Liu et al., 2021). Figure 6 shows the framework of a CIS system.

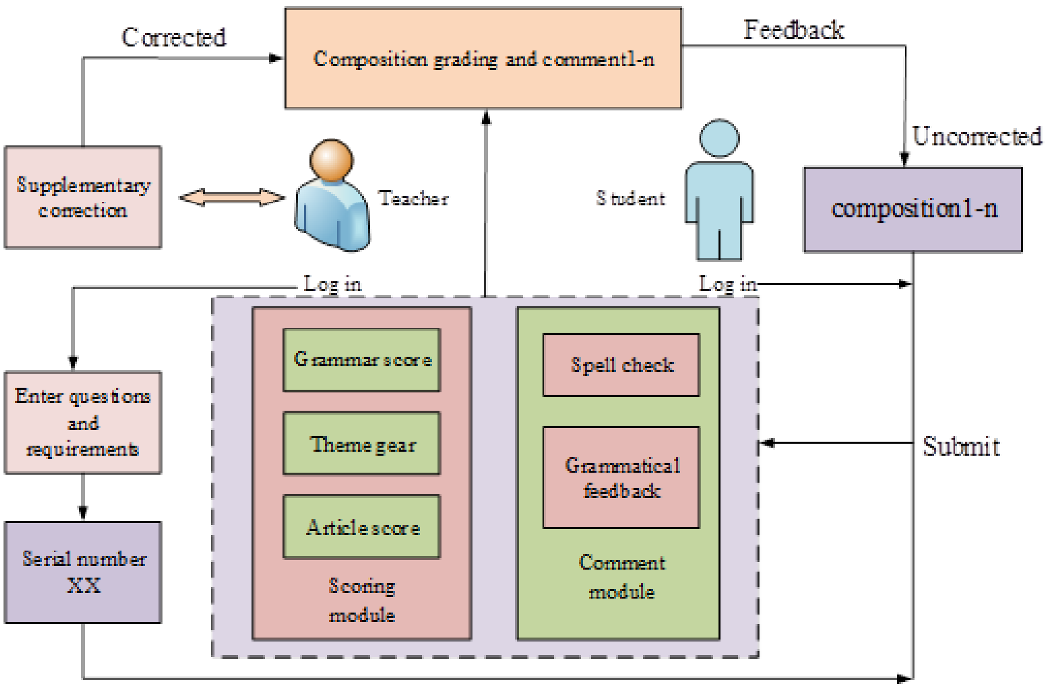
In Figure 6, the actual composition scoring mainly consists of six sections. The first step entails teachers logging in to the system and providing the questions and requirements for their compositions. This sets up a correction mode, and students open the system to write and submit according to the requirements. In the second part, the computer combines various algorithms to grade and comment on the composition (Yuan, 2021). The scoring part includes the composition level, grammar score, and the score of the whole composition. The comment module includes syntax error correction and spelling check and finally generates a rating and comment chart report. The third step is to supplement and modify manually based on machine modifications. The fourth step is to automatically generate a review record of the composition and directly give the results feedback to students and teachers. Finally, it generates a historical review record and a chart of all composition scores and content changes for students.

PERFORMANCE TEST OF CIS MODEL

Performance Test of Grammar Detectors and Scoring Algorithms

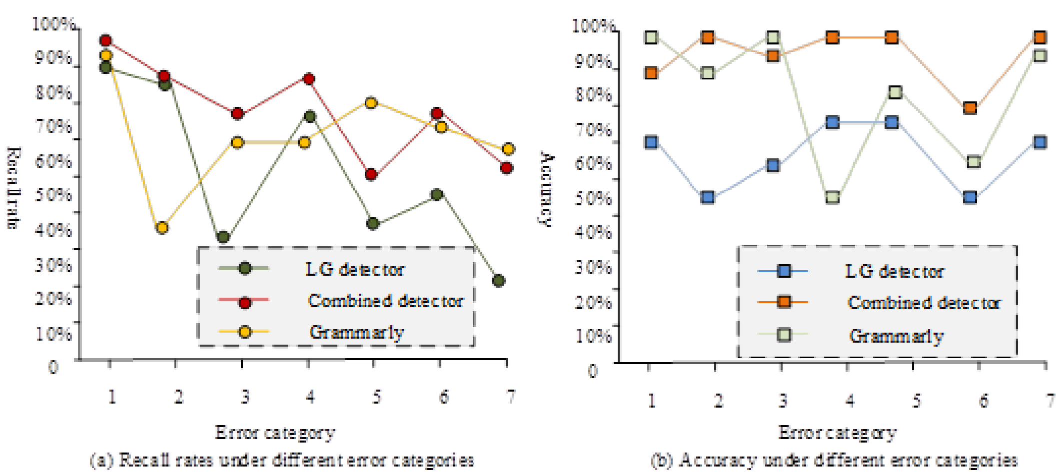
To verify the performance of the grammar detector combined with LG and N-grammar model, a testing experiment was conducted on the Windows platform. To avoid experimental errors caused by different devices, the same computer was used throughout the experiment. The computer was configured with dual cores and four threads, with a DDR3 processor and 20GB of RAM. The test samples selected for this study were from the Chinese Learner Corpus. There were detailed manual annotation results in the corpus. It randomly selected 10 sets of 800 sentences from the corpus

Figure 6. Frame diagram of CIS system



for testing, each consisting of correct and incorrect grammar sentences. There were 423 incorrect grammatical sentences in the sample, including 465 syntax errors. 10 sets of samples were input into the traditional LG detector, which combined with N-grammar model and the Grammarly grammar detection software for analysis and scoring. The scoring indicators recall and accuracy are introduced. The test results are shown in Figure 7.

Figure 7. Accuracy and recall under different error types



In Figure 7, the vertical axis of Figures 7 (a) and (b) represent the recall rate and the accuracy rate, respectively. The types of syntax error represented by abscissa 1 to 7 in Figure 7 are collocations, sentences, conjunctions, adjectives, adverbs, verb phrases, and noun phrases. In Figure 7 (a), the recall rate of the proposed combined type grammar detector in this study was significantly higher than that of traditional LG detectors. In the case of syntax error category 5 and 7, the recall rate was lower than that of Grammarly, and other categories were higher than that of Grammarly. According to Figure 7 (b), the accuracy of the combined class grammar detector was significantly higher than that of traditional LG. When the syntax error category was 1 or 3, the accuracy rate was lower than that of Grammarly, and other categories were higher than that of Grammarly. To sum up, compared with the traditional LG detector, the syntax detector combined with the N-grammar model had higher recall and accuracy, better performance, and better recognition of syntax error. Compared to Grammarly, most of the recall and accuracy rates were better, while a small portion of the results were not significant. To further validate the superiority of this combined model, the experiment introduced a comprehensive indicator F1, taking into account both recall and accuracy. The results are shown in Figure 8.

In Figure 8, the vertical axis represented the calculation results of the comprehensive indicators, and the higher the value, the better. Horizontal coordinates 1 to 7 still represented collocations, sentences, conjunctions, adjectives, adverbs, verb phrases, and noun phrases. The average F1 value of combined detector, LG, and Grammarly was 89.9%, 63.3%, and 80.2%, respectively. The F1 value of the combined detector was 26.6% higher than LG and 9.7% higher than Grammarly. Overall, the addition of the N-grammar model significantly improved the performance of LG compared to the three parsers. Compared with the common grammar analysis software on the market, Grammarly, the error correction performance of the combined detector was also better. After analyzing the performance of grammar detection, the research continued to test the performance of horizontal partitioning. Using 150 manually-graded compositions as the test criteria, it categorized the compositions according to the number of words and input them into the algorithm used in this study and Microsoft Love Writing's scoring system and compared the results of the two levels. Figure 9 is a comparison of the results of correctly dividing the level of composition.

As shown in Figure 9, in different word count intervals, the correct division level of the algorithm proposed in this study was relatively close to the total amount of compositions, with a high accuracy

Figure 8. Comprehensive indicators under different error types

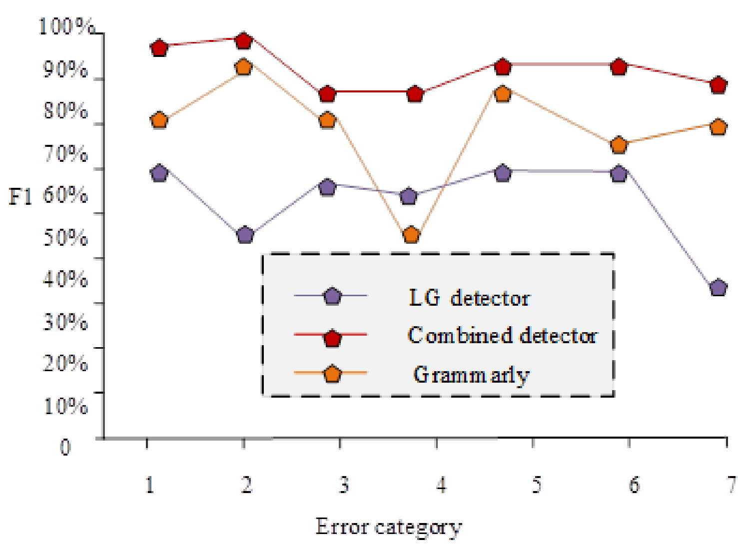
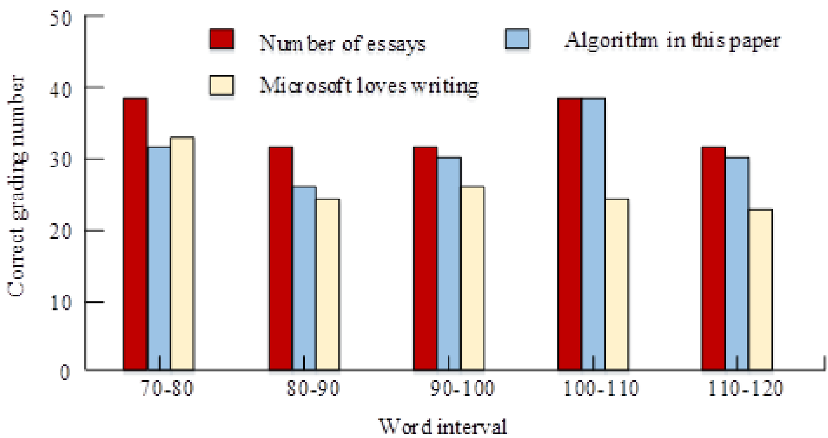


Figure 9. Comparison of test results between this algorithm and Microsoft Love writing system



rate at an average of 94.92%. When the number of words was between 70 and 80, the algorithm proposed in this study correctly divided the number of writing proficiency compositions compared to Microsoft Love Writing. This indicated that its accuracy was lower than that of Microsoft Love Writing. When the number of words was between 80 and 120, the proposed algorithm correctly divided the number of compositions at the writing level, which was significantly higher than Microsoft Love Writing, indicating a higher accuracy rate. When the number of words was between 100 and 110, the proposed algorithm correctly divided the number of composition levels that overlap with the total amount, with an accuracy of 100%. Overall, the accuracy of the proposed algorithm was higher than that of Microsoft Love Writing, so this algorithm can practically be used in the application of CIS.

Performance Testing of Scoring Model Application

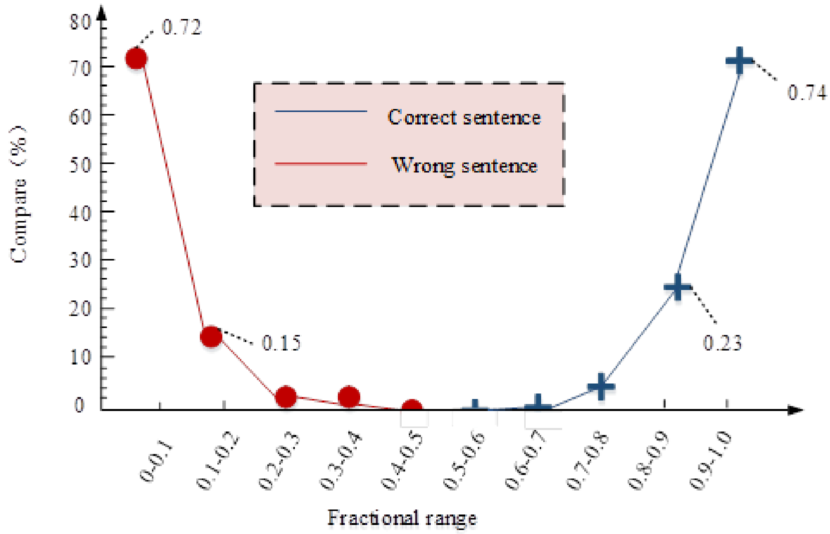
To test the accuracy of the grammar scoring proposed in this study, 1276 correct sentences and 1539 incorrect sentences were selected from the middle school student English material library and analyzed using an LG analyzer. The grammar scoring algorithm proposed in this article was used for scoring. The results are shown in Table 1.

As shown in Table 1, after analyzing different samples, a score has been obtained, with a maximum score of 1 and a minimum score of 0. The score distribution of incorrect sentences was between 0 and 0.5 points, and the closer the score of incorrect sentences was to 0, the better, indicating that the error of incorrect sentences was within 0.5 points. The distribution of correct sentence scores was between 0.7 and 1 points, and the closer the correct sentence score was to 1, the better, indicating that the error of incorrect sentences was within 0.3 points. To analyze the obtained data more intuitively, the table data was plotted into a graph. Figure 10 shows the distribution of scores for correct and incorrect sentences.

Table 1. Error analysis of grammar scoring

Fractional interval	[0,0.1)	[0.1,0.2)	[0.2,0.3)	[0.3,0.4)	[0.4,0.5)	[0.5,0.6)	[0.6,0.7)	[0.7,0.8)	[0.8,0.9)	[0.9,1.0)
Correct sentence distribution ratio	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.23	0.74
Error sentence distribution ratio	0.72	0.15	0.11	0.01	0.01	0.00	0.00	0.00	0.00	0.00

Figure 10. Proportion of score distribution between wrong and correct sentences

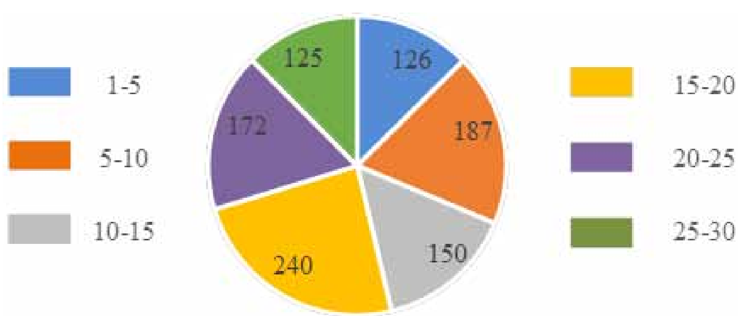


In Figure 10, the horizontal axis represents the score range of a sentence, with a maximum score of 1. The vertical axis represents the proportion of sentence numbers. For correct sentences, the more higher scoring intervals the score occupied, the better. For incorrect sentences, the higher the proportion of low scores, the better. As shown in the figure, analyzing and scoring correct sentences accounted for 96% of the high scoring range of 0.8 to 1, with a high accuracy rate. Analyzing and scoring incorrect sentences accounted for 87% ranging from 0 to 0.2 in the low partition, with average accuracy. In summary, this grammar algorithm was more sensitive and accurate in judging correct sentences. This indicated that there were shortcomings in the grammar scoring of incorrect sentences in this study, and the accuracy needed to be improved.

For the English composition scoring system proposed in the method section, corresponding performance tests were conducted. The test sample came from a composition in a high school English exam, with a maximum score of 30 points. It randomly selected 1000 essay samples from students for testing. The distribution of scores in the samples is shown in Figure 11.

In Figure 11, there were 126 compositions rated from 1 to 5, 187 compositions rated from 5 to 10, 150 compositions rated from 10 to 15, 240 compositions rated from 15 to 20, 172 compositions rated from 20 to 25, and 125 compositions rated from 25 to 30. After organizing the samples, the

Figure 11. Distribution of samples



test inputted them into the research algorithm for AC. Finally, the obtained results were compared and analyzed, and the scoring errors for different scoring stages were calculated as shown in Table 2.

As shown in Table 2, there was a significant difference in scoring errors among different score ranges. For compositions with an error of less than one point, the proportion of compositions in the 20-25 score range was relatively high, reaching 56.78%. This indicated that there was not much difference between manual correction and program correction in this score segment, and the accuracy was high. For compositions with an error of more than five points, the proportion of compositions in the 5-10 score range was relatively high, reaching 37.12, indicating that there was a large error in this score stage, and there was a significant difference between program correction and manual correction. This might be because, for correct grammatical sentences, the research model had a higher recognition rate, resulting in higher accuracy. The research model did not have a large amount of data on incorrect sentences and could not accurately identify incorrect sentences, resulting in relatively low accuracy. The mean square error value was the smallest in the 20-25 score stage. In terms of runtime, the average analysis time for an article was around 22 seconds, which met the real-time requirements. Although the accuracy of judging incorrect sentences was not high enough, overall, the proportion of program scoring and manual scoring errors within five points was relatively high, reaching 87.29%, and the overall mean square error was 3.08 which reaches the accuracy required for essay scoring, indicating that the research method had high applicability.

CONCLUSION

Aimed at addressing the weakness of English grammar among middle school students in China, this research was based on the combination model of LG and N-grammar model to detect and analyze grammar. Then IE was used to grade the composition and finally calculate the score of the whole composition. Comparing the proposed grammar detection performance with traditional LG detectors and Grammarly software, the performance test experimental results showed that the combined model's recall and accuracy comprehensive weight of the proposed method was 89.9%, which was 26.6% higher than LG and 9.7% higher than Grammarly. The results also showed that the accuracy rate of the CIS algorithm based on IE proposed in this study was significantly higher than that of Microsoft's Love Writing, reaching 94.92%. In grammar scoring, the experimental results showed that the accuracy of analyzing and scoring correct sentences was higher than that of incorrect sentences, and the accuracy of grammar scoring for incorrect sentences needed to be improved. Finally, the proposed method was applied to the English composition exam of a selected high school. The results showed

Table 2. Composition scoring test results

Fractional segment	Proportion(%)			Mean square deviation of error	Average analysis time(ms)
	Within 1 score	Within a score of 1 to 5	More than 5 points		
0-5	33.22	53.27	13.15	2.73	2012
5-10	8.23	36.78	54.99	4.01	1998
10-15	6.54	56.34	37.12	4.23	2544
15-20	56.78	38.22	5.00	1.65	2571
20-25	66.24	30.08	3.68	1.53	1922
25-30	4.24	51.32	44.44	4.56	2548
Total	38.35	48.94	12.71	3.08	2269

that the error value between the proposed method and manual scoring was relatively small, with a proportion of 87.29% within five points. The overall mean square error value was low (only 3.08), and the average running time was short (only 22.69 seconds). The CIS model constructed by Zhang (2021), referring to the AC system, discussed the feature selection method in detail. This model also analyzed the impact of the test rate on the composition score from the perspective of linguistics and finally evaluated and analyzed the score results using a multiple regression method. Compared with this model, this research model took into account the shortcomings of machine learning and adopted a man-machine combination method for composition scoring, which had higher accuracy and could also teach students according to their aptitude (Zhang, 2021). This indicated that this study has strong applicability in the scoring of middle school students' English compositions. However, there are still shortcomings in the research, as the module design of the scoring system is not rich enough. Further research is needed on the overall scoring of compositions, which can reflect the grammar habits of Chinese middle school students. In future research, sentence coherence and vocabulary classification will be added to the design of the scoring system to make it more comprehensive. In addition, it will continue to combine more algorithms to build more efficient and accurate composition scoring models.

FUNDING

The research is supported by Philosophy and Social Science Research Project of Jiangsu Universities: "On the Construction Mode of Blended Teaching in Moral Education of Foreign Language Courses from the Perspective of Multimodality" (No.2022SJYB2404).

REFERENCES

Albiansyah, A., & Minkhatunnakhriyah, M. (2021). The needs analysis of English in Bina Informatika Vocational High School of Bintaro as a basis to design English teaching materials. *Edunesia: Jurnal Ilmiah Pendidikan*, *2*(1), 142–152. doi:10.51276/edu.v2i1.88

Chen, G., & Liang, M. (2022). Verb form error detection in written English of Chinese EFL learners: A study based on Link Grammar and Pattern Grammar. *International Journal of Corpus Linguistics*, *27*(2), 139–165. doi:10.1075/ijcl.19107.che

Gaillat, T., Simpkin, A., Ballier, N., Stearns, B., Sousa, A., Bouyé, M., & Zarrouk, M. (2022). Predicting CEFR levels in learners of English: The use of microsystem criterial features in a machine learning approach. *ReCALL*, *34*(5), 130–146. doi:10.1017/S095834402100029X

Gakis, P., Kokkinos, T., & Tsalidis, C. (2021). Towards the development of a grammar checker and its utilization in the teaching of Modern Greek as mother tongue. *Literary & linguistic computing*. *Journal of the Alliance of Digital Humanities Organizations*, *36*(1), 87–100.

Jiao, Y., & Liu, Y. (2021). The teaching optimization algorithm mode of integrating mobile cloud teaching into ideological and political courses under the internet thinking mode. *Scientific Programming*, *2021*(13), 1–8. doi:10.1155/2021/6492009

Jing, L., Bo, Z., Tian, Q., Xu, W., & Shi, J. (2020). Network education platform in flipped classroom based on improved cloud computing and support vector machine. *Journal of Intelligent & Fuzzy Systems*, *39*(99), 1–11. doi:10.3233/JIFS-179952

Kong, J. E., & Lee, D. (2021). A study on middle school students' perceptions on remote English learning during COVID-19. *Journal of English Teaching through Movies and Media, 22*(2), 41-57.

Lai, Y., Ma, Z., Xu, S., & Easa, S. (2022). Information entropy evaluation model of bus-line reliability considering the combination of bus stops and bus travel time. *Canadian Journal of Civil Engineering*, *49*(1), 64–72. doi:10.1139/cjce-2020-0666

Li, S., & Juan, W. (2021). Intelligent English classroom video clarity improvement based on motion compensation and grid flow. *Journal of Intelligent & Fuzzy Systems*, *40*(4), 6985–6995. doi:10.3233/JIFS-189528

Liang, S., Chen, C., & Zou, G. (2021). Research on English grammar recognition system based on combination of genetic algorithm and KNN algorithm. *Journal of Intelligent & Fuzzy Systems: Applications in Engineering and Technology*, *41*(5), 5770–5770.

Liang, X., Haiping, L., Liu, J., & Liu, L. (2021). Reform of English interactive teaching mode based on cloud computing artificial intelligence–a practice analysis. *Journal of Intelligent & Fuzzy Systems*, *40*(2), 3617–3629. doi:10.3233/JIFS-189397

Liu, J., Lin, L., & Liang, X. (2021). Intelligent system of English composition scoring model based on improved machine learning algorithm. *Journal of Intelligent & Fuzzy Systems*, *40*(2), 2397–2407. doi:10.3233/JIFS-189235

Mansfield, J., & Barth, D. (2021). Clause chaining and the utterance phrase: Syntax–prosody mapping in Matukar Panau. *Open Linguistics*, *7*(1), 423–447. doi:10.1515/opli-2021-0023

Mastick, N. C., Welicky, R. L., & Leslie, K. L. (2022). Parasite communities in English Sole (Parophrys vetulus) have changed in composition but not richness in the Salish Sea, Washington, USA since 1930. *Parasitology*, *149*(6), 786–798. doi:10.1017/S0031182022000233 PMID:35238289

Näf, A. (2021). Zentrum und Peripherie in der deutschen Syntax: Erprobung des theoretischen Modells. *Linguistik online, 106*(1), 115-147.

Sasaki, K., Nair, A., Firl, D. J., McVey, J. C., Moro, A., Diago Uso, T., Fujiki, M., Aucejo, F. N., Quintini, C., Kwon, C.-H. D., Eghtesad, B., Miller, C. M., & Hashimoto, K. (2021). Conditional probability of graft survival in liver transplantation using donation after circulatory death grafts a retrospective study. *Transplant International*, *34*(8), 1433–1443. doi:10.1111/tri.13846 PMID:33599045

Tavakol, A. V., Agababaoglu, A., Yildirim, S., & Onat, A. (2022). A real-world application of Markov chain Monte Carlo method for Bayesian trajectory control of a robotic manipulator. *ISA Transactions*, *125*(125), 580–590. doi:10.1016/j.isatra.2021.06.010 PMID:34148651

Wang, C. H., Lee, Y. J. J., Wu, W. S., & Armstrong, J. (2022). Educational bibliotherapy for developing undergraduates' bibliotherapeutic energy in an Advanced English Reading classroom. *Journal of Poetry Therapy*, *35*(2), 98–113. doi:10.1080/08893675.2021.2004373

Xiong, L., & Li, M. (2021). Behavioral modeling based on cloud computing and target user recommendation for English cloud classroom. *Microprocessors and Microsystems*, *80*(2), 1–5. doi:10.1016/j.micpro.2020.103587

Yuan, Z. (2021). Interactive intelligent teaching and automatic composition scoring system based on linear regression machine learning algorithm. *Journal of Intelligent & Fuzzy Systems*, *40*(2), 2069–2081. doi:10.3233/JIFS-189208

Zhang, Y. (2021). Interactive intelligent teaching and automatic composition scoring system based on linear regression machine learning algorithm. *Journal of Intelligent & Fuzzy Systems*, *40*(2), 2069–2081. doi:10.3233/JIFS-189208

Zhao, Y. (2021). Research and design of automatic scoring algorithm for English composition based on machine learning. *Scientific Programming*, *2021*(1), 1–10. doi:10.1155/2021/3429463

Zhu, Q., & Schlick, T. (2021). A Fiedler vector scoring approach for novel RNA motif selection. *The Journal of Physical Chemistry B*, *125*(4), 1144–1155. doi:10.1021/acs.jpcb.0c10685 PMID:33471540