Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
 Volume 25, Issues 1, pp. 451–463, DOI 10.12694/scpe.v25i1.2333

RESEARCH ON THE APPLICATION OF INTELLIGENT GRADING METHOD BASED ON IMPROVED ML ALGORITHM IN SUSTAINABLE ENGLISH EDUCATION

LEI HUANG*AND LI MA†

Abstract. The current intelligent grading methods in English education have issues of feedback lag and incompatibility with manual grading, and manual grading is easily influenced by subjective consciousness. Based on this, this research selects Adaboost algorithm in ML algorithm to realize intelligent grading. The experiment improves the Adaboost algorithm according to the actual needs, constructs the Adaboost/CT algorithm, and verifies its effectiveness. The experimental results show that in the English intelligent scoring module, the adjacency accuracy of the Adaboost/CT algorithm for high- and low-quality English is 95.33%; 95.45% in the middle score; 94% in the breakdown. The comparison between Adaboost/CT model and DFA method shows that the accuracy and proximity accuracy of Adaboost/CT are 79.66% and 94% respectively, which are much higher than 55% and 92% of DFA. In addition, compared with Adaboost, the accuracy of Adaboost/CT is also significantly better than Adaboost. In practical application, the use of Adaboost/CT in the evaluation of English compositions can not only get more accurate scores, but also find out the shortcomings of each student, so as to improve them pertinently. Meanwhile, the accuracy, recall, and F1 values of the Adaboost/CT algorithm are 96%, 95%, and 95%, respectively, which are higher than those of the comparison algorithm. Overall, the improved Adaboost/CT algorithm has shown high effectiveness and practicality, and has high applicability and effectiveness in practical intelligent grading.

Key words: Machine learning; Intelligent marking; Sustainable English education; Adaboost/CT

1. Introduction. The intelligent development of computers has promoted the gradual prevalence of "machine marking". Machine grading can not only overcome the low efficiency of manual grading, but also effectively avoid the negative impact of teachers' subjective factors in the grading process [1]. In the context of sustainable English education, manual grading will not only add unnecessary burden to the grading teachers, but also bring some adverse effects. Therefore, intelligent grading is gradually widely used [2]. Zhou built an automatic composition scoring system under the mixed mode of English writing on the basis of machine learning (ML) for the related problems of manual grading [3]. Zhang et al. started with the automatic scoring of semi-open short answer questions, combined with general information and specific information in relevant fields, built an automatic scoring model for semi-open short answer questions [4]. Fang designed a scoring system for oral English intelligence based on dynamic time warping algorithm to solve the problem of low accuracy of oral English intelligence scoring [5]. In this context, this study selected Adaptive Boosting (Adaptive Boost) algorithm by analyzing natural language processing and ML, and introduced the concept of Centralized Trend (CT) to improve it, and constructed the Adaptive Boost/CT algorithm. The Adaboost/CT algorithm is mainly used to solve the problems of low efficiency and insufficient fairness in English manual grading, aiming to help teachers find effective ways to improve students' learning performance, and also provide reference for the development of sustainable English education.

2. Related Work. At present, the grading of objective questions in English has gradually matured, but there is no clear unified answer for the intelligent grading method of subjective questions [6]. At the same time, due to the low efficiency of manual grading, students' writing ability has not been improved, and the intelligent reading of English compositions has become a hot topic [7]. Aimed at the problems of high time consuming and low reliability of manual grading, Ramesh et al. made a comprehensive study on the automatic grading system and proposed a ML technology for intelligent grading. This technology provides help to judge student performance and improve student performance [8]. Liu proposed a hybrid scoring model for intelligent scoring

*General Education School, Chongqing Youth Vocational & Technical College, Chongqing, 400712, China (Corresponding Author: huangsweet@163.com)

†School of Artificial Intelligence, Chongqing Youth Vocational & Technical College, Chongqing, 400712, China

in English writing courses by combining real-time feedback from machines with manual evaluation. This model not only improves the accuracy of grading, but also improves student performance [9]. In order to evaluate learners' English level, Gaillat et al. built an intelligent grading system for English compositions based on the supervised learning method. The proposed system not only improves the accuracy of assessment, but also provides help to truly master learners' English level [10]. Liu et al. introduced convolutional neural network as an assistant in English automatic grading, aiming at the problem that the subjective consciousness of the rater will have a negative impact on the final grading. Finally, the experiment proposed a new automatic English composition scoring system, which provided help for the improvement of foreign students' English level [11]. In order to improve the spoken English ability of English learners, Wu et al. proposed a new oral English scoring method based on computer neural network. This method improves students' oral English ability and also promotes the standardization of students' oral English pronunciation [12].

In addition, Phophohuangpairoj et al. proposed a new method of English article grading and feedback by using automatic text analysis to solve the problem of poor writing performance of students when learning English. The proposed method can help teachers work effectively and also improve English learners' writing level [1]. Taskiran et al. conducted an in-depth study on students in the open education department of a university in Türkiye in order to verify the impact of intelligent marking on students' writing ability. The experiment verified the positive impact of intelligent grading, and also provided help for improving students' performance [14]. In order to solve the problem that the current intelligent scoring cannot achieve subjective scoring, Miao et al. proposed an intelligent scoring model for English subjective questions based on deep neural network and linear regression. This model can improve the accuracy of subjective scoring and also provide direction for the training system of English skills [15].

From the research of scholars at home and abroad, it can be seen that the current English intelligent scoring method is not very mature, and the research on natural language feedback is not deep enough. In the current intelligent scoring system built by intelligent scoring method, the connection between scoring module and feedback module is not smooth enough. Therefore, this research innovatively proposes the Adaboost/CT algorithm, which makes the English intelligent scoring model have a deeper connection between the two parts of scoring and feedback. In addition, another innovation of the study is the construction of the comment model, which can achieve the real meaning of promoting learning by evaluation.

3. Research on English intelligent grading method based on ML algorithm.

3.1. Research on natural language processing and intelligent grading technology. In order to improve the efficiency and fairness in the assessment of sustainable English education, Adaboost algorithm is selected in the ML algorithm. In the experiment, the concept of CT is introduced into the actual English intelligent grading to improve Adaboost, and the Adaboost/CT algorithm is obtained. As an important branch of artificial intelligence, natural language processing occupies a large proportion in computer science. Therefore, in order to understand machine learning, we need to fully understand natural language processing. Natural language processing is a discipline integrating linguistics, computer and mathematics. This discipline not only studies natural language, but also provides help for the development of computer systems capable of efficient natural language communication [16]. In the statistical language model of natural language processing, the relevant sequence model of words or sentences in the English corpus is essentially a probability model. The probability calculation expression of the word string in the complete sentence is shown in equation 3.1.

$$P(T) = P(r_1, r_2, r_3, , r_n) \tag{3.1}$$

In equation 3.1, $P(T)$ represents the probability of occurrence of English word strings in sentence T; r indicates the English words that make up the sentence; indicates the number of English words. In addition, you can also choose to use the chain rule inside the conditional probability to decompose the probability of the word string. Therefore, the probability calculation expression obtained by expanding equation 3.2 is shown in equation 3.2.

$$P(r_1, r_2, r_3, \ldots, r_n) = P(r_1) \cdot P(r_2 \,|\, r_1) \cdot P(r_3 \,|\, r_1, r_2) \cdot \ldots \cdot P(r_n \,|\, r_1, r_2, r_3, \ldots, r_{n-1}) \tag{3.2}$$

In equation 3.2, $P(r_1)$ represents the probability of the occurrence of the first word r_1 in the sentence; $P(r_1|r_2)$ indicates the probability of the occurrence of the second word on the basis that the first word is known. Based on equation 3.1 and equation 3.2, the simplified calculation expression is shown in equation 3.3.

$$P(T) = P(r_1) \cdot P(r_2 \,|\, r_1) \cdot P(r_3 \,|\, r_2) \cdot \ldots \cdot P(r_n \,|\, r_{n-1}) \tag{3.3}$$

Equation 3.3 is essentially the relevant calculation expression of the binary grammar model. Suppose a is used to represent the beginning of a sentence. If you want to calculate the possibility of "Yesterday was a bad day", just multiply the probability of binary grammar of two adjacent words. Therefore, the model probability calculation expression of the sentence is shown in equation 3.4.

$$\begin{aligned}(Yesterday\,\text{was a bad day}) = {} & P(Yesterday \,|\, \langle a \rangle) \cdot P(was \,|\, Yesterday) \\ & \cdot P(a \,|\, was) \cdot P(bad \,|\, a) \cdot P(day \,|\, bad)\end{aligned} \tag{3.4}$$

On the basis of equation 3.4 the research first determines the corpus of English training, and according to the corpus, the number of occurrences of a binary grammar can be obtained. Normalize it on this basis. The calculation expression of conditional probability obtained is shown in equation 3.5.

$$P(r_n \,|\, r_{n-1}) = \frac{C(r_{n-1} r_n)}{\sqrt{C(r_{n-1} r)}} \tag{3.5}$$

In equation 3.5, C represents counting. According to equation 3.5, the model of multiple grammar can be deduced. For general multivariate grammar, the calculation expression of its parameter estimation is shown in equation 3.6.

$$P(r_n \,|\, r_{n-N+1}^{n-1}) = \frac{C(r_{n-N+1}^{n-1} r_n)}{C(r_{n-N+1}^{n-1})} \tag{3.6}$$

In equation 3.6, N represents the total number of words. Correspondingly, when the number of English words is more than two, the sentence probability calculation expression is shown in equation 3.7.

$$P(T) = \prod_{k=1}^{k+1} P(r_k \,|\, r_{k-n+1}^{k}) \tag{3.7}$$

In equation 3.7, k represents the serial number of the word, and the maximum is n. In addition, in the statistical analysis model of natural language processing, correlation analysis plays an important role in practical problems. According to the background of sustainable English education, the Pearson correlation coefficient is mainly summarized here. This coefficient mainly represents the linear correlation coefficient between variables, and its calculation expression is shown in equation 3.8.

$$c = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}} \tag{3.8}$$

In equation 3.8, c represents correlation; x_i and y_i represent the values of the two under the sample; $\bar{x}$ And $\bar{x}$ represent the sample mean of both. In addition, in the manual grading of English, the composition grading standard is one of the main bases. The reviewers evaluate the quality of an English composition according to the general requirements of the grading criteria, and then determine an appropriate score through the grading criteria. The rules for grading CET-6 composition in the conventional sense are shown in Figure 3.1.

As can be seen from Figure 3.1, the grading rules for English compositions in the conventional sense mainly divide the grades of compositions into six levels. Each level is divided according to whether the composition is relevant to the topic, whether the thought is clearly expressed, whether the language is wrong, and whether the semantics are coherent. Therefore, it is necessary to convert the scoring elements in the scoring rules into characteristic indicators for intelligent English grading under sustainable education. The research proposes a three-level indicator system based on actual needs, and its contents are shown in Figure 3.2.

It can be seen from Figure 3.2 that the research has established a new evaluation index system on the basis of comprehensive consideration of some manual scoring standards, composition scoring rules at home and

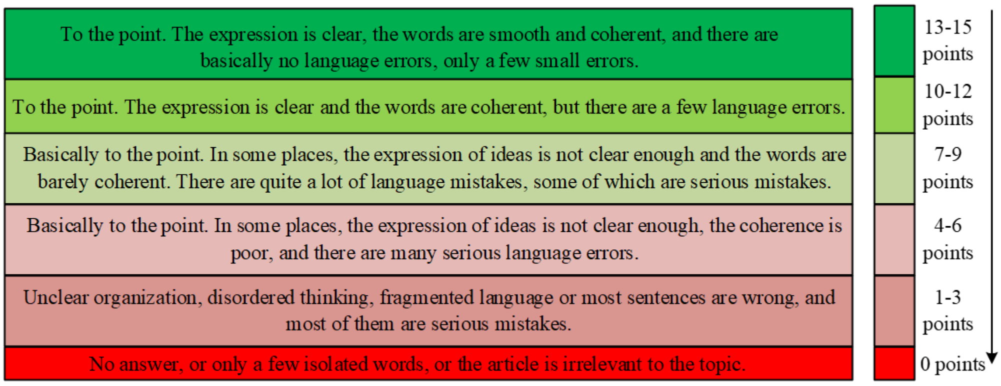


Fig. 3.1: Detailed Rules for Grading CET-6 English Composition

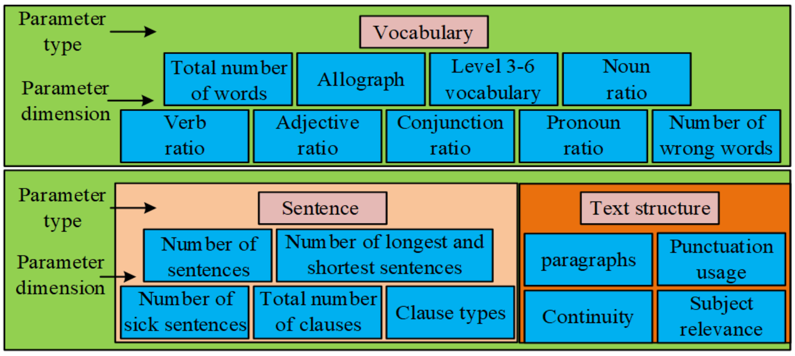


Fig. 3.2: Three-layer index system considering manual evaluation criteria

abroad and more mature scoring standards abroad. Among them, the characteristic index is calculated and classified from hard to soft, from large to small, and from shallow judgment to specific score. In addition, feature indicators also include lexical, syntactic and textual feature indicators. These indicators are based on the level of words, the total number of sentences, and the coherence of the text as the basic elements of the characteristic indicators.

3.2. Adaboost algorithm based on ML and its improved algorithm analysis. ML is interdisciplinary research, including probability theory, statistics, approximation theory, convex analysis, algorithm complexity, etc. It is specialized in studying how computers imitate or implement human learning behaviors to acquire new knowledge and skills. ML can reorganize the existing knowledge to continuously improve its own capabilities [17]. As an iterative algorithm in ML, the core idea of Adaboost algorithm is to train different classifiers (weak classifiers) in the same training set and combine them to form a stronger classifier (strong classifier) [18]. The main advantage of the Adaboost algorithm is that there are no requirements for the actual design of weak classifiers in the provided framework, so various methods can be used to construct weak classifiers without the need for prior knowledge of relevant experience. At the same time, its performance requirements for weak classifiers are not high, and the algorithm application is relatively simple. It does not need to be used for feature filtering, nor does it need to worry about overfitting. Therefore, it is studied as the basic algorithm for intelligent grading. However, when the traditional Adaboost algorithm is applied to English intelligent grading, the weak classifier is prone to repeat and error. Therefore, in order to remedy this

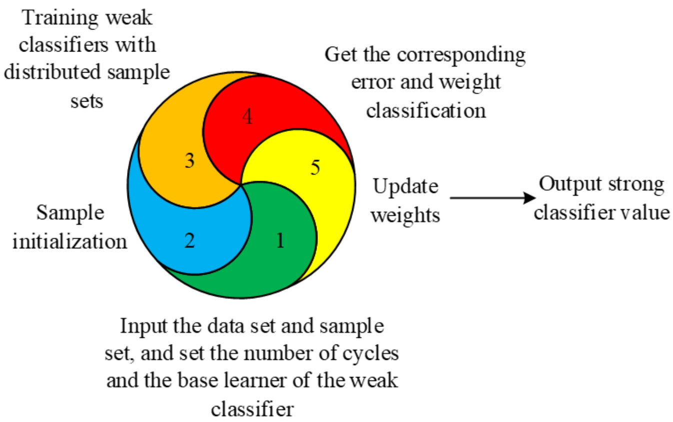


Fig. 3.3: Flow diagram of Adaboost/CT algorithm

defect, the experiment introduced the concept of CT and proposed an improved Adaboost algorithm, namely Adaboost/CT algorithm.

CT is a statistical concept, which represents a group of data close to a central value. At the same time, it also reflects the center of a data set [19]. The representative values of CT are numerical average and position average. In addition, the main factors reflecting CT include mean, central and median. It is worth noting that extreme data will break the balance of a group of data, so the average value does not represent the average level. The flow of the improved Adaboost/CT algorithm is shown in Figure 3.3.

From Figure 3.3, the Adaboost/CT algorithm process first inputs the corresponding dataset and sample set, and sets the number of cycles and the base learner of the weak classifier; The second step is to initialize the sample weights and use them to train weak classifiers using distributed sample sets, while using distributed sample sets to train base learners in the dataset. Then calculate the corresponding error and weight classification, and update the weights, that is, update the distribution; Finally, output the strong classifier value. Using Adaboost/CT algorithm can better avoid the over-fitting of traditional Adaboost, and can also effectively delete extreme data, thus solving the trap of stacking errors of weak classifiers. In the process, the equation expression of the dataset is shown in equation 3.9.

$$R = \{B_1, B_2, \ldots, B_K\} \tag{3.9}$$

In equation 3.9, R represents the input data set; B represents the elements in the data set, and the expression of their subscript related equations is shown in equation 3.10.

$$v = \{1, 2, \ldots, K\} \tag{3.10}$$

In equation 3.10, v represents the number of cycles, and its maximum value is K. In addition, the equation expression of the sample set is shown in equation 3.11.

$$B_v = \{(x_1, y_1), (x_2, y_2), \ldots, (x_m, y_m)\} \tag{3.11}$$

In equation 3.11, x and y represent the characteristic value and result value of the data set samples respectively, and the maximum number of samples of the two is m . At the same time, in the Adaboost/CT algorithm flow, the calculation expression of error is shown in equation 3.12.

$$\lambda_v = \Pr_x \sim B_{(v,y)} I \tag{3.12}$$

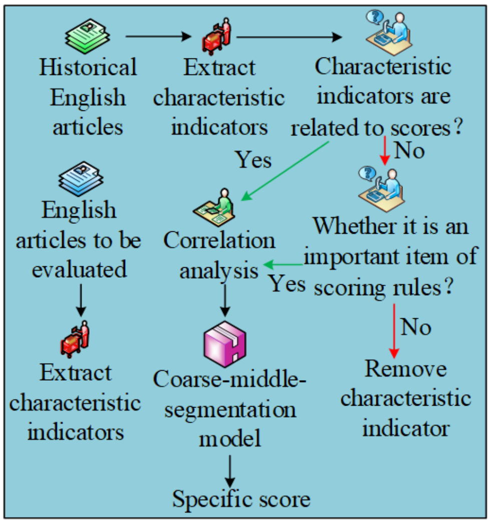


Fig. 3.4: Overall flow diagram of scoring model

In equation 3.12, λ represents error; I indicates the indicator function, which is 1 when it is correct and 0 when it is wrong. The calculation expression of weight classification is shown in equation 3.13.

$$\omega_v = \frac{1}{2} \ln \left(\frac{1-\lambda_v}{\lambda_v} \right) \tag{3.13}$$

In equation 3.13, ω represents the weight classification. Based on equation 3.13, the calculation expression of weight update is shown in equation 3.14.

$$B_{v+1}(i) = \frac{B_v(i) \exp\left(-\omega_v y_i h_v(x_i)\right)}{Z_v} \tag{3.14}$$

In equation 3.14, i also represents the maximum number of samples; h_v represents a weak classifier; Z_v represents a normalized constant. Finally, the output strong classifier numerical calculation expression is shown in equation 3.15.

$$U(x) = \text{sign}\left(\sum_{v=1}^{K} \omega_v h_v(x) \right) \tag{3.15}$$

In equation 3.15, $U(x)$ represents the strong classifier value. In the actual intelligent grading of English, the scoring model mainly adopts correlation analysis technology. This technology mainly removes irrelevant characteristic indicators and determines whether it is the key feature of the scoring standard [20]. The technology finally retains the characteristic indicators with high prediction ability, and establishes a scoring model based on the corresponding characteristic indicators. Finally, the model is integrated into the Adaboost/CT algorithm. The Adaboost/CT algorithm classifies the scores obtained by using the method of rough score, middle score and subdivision to get the final score. The overall process of the scoring model is shown in Figure 3.4.

It can be seen from Figure 3.4 that the scoring model will be based on the English article samples that have been reviewed and reviewed by experts. At the same time, the sample will be statistically analyzed and the characteristic index of the composition sample will be extracted. The experiment uses the correlation analysis method to retain the characteristic index with high correlation ($correlation\ coefficient \geq 0.6$). Then, according to the actual situation of sustainable English education, the model adds a feature index with high prediction ability to the mature automatic English article scoring system. According to the actual situation, the model

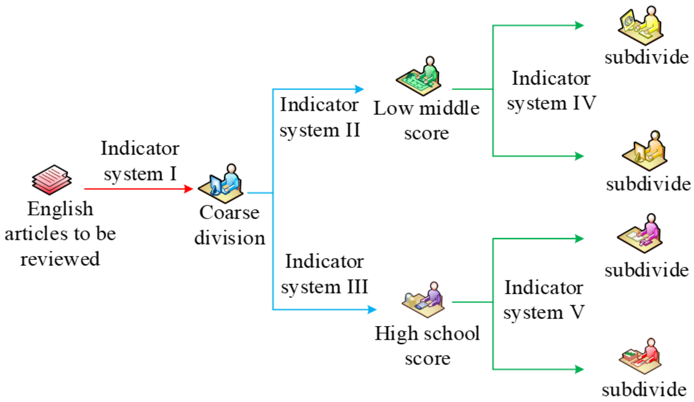


Fig. 3.5: Schematic diagram of course-middle-segmentation model

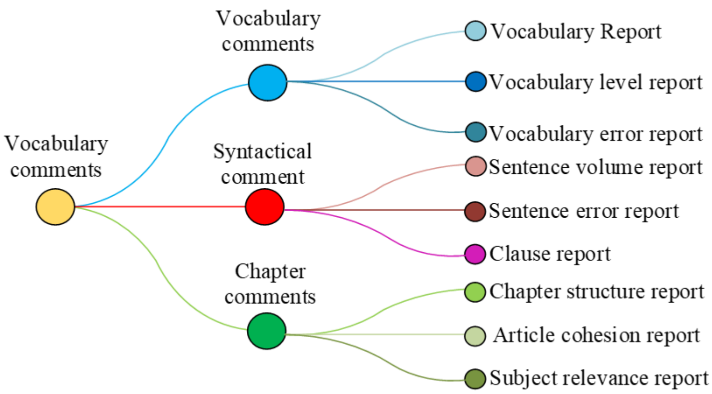
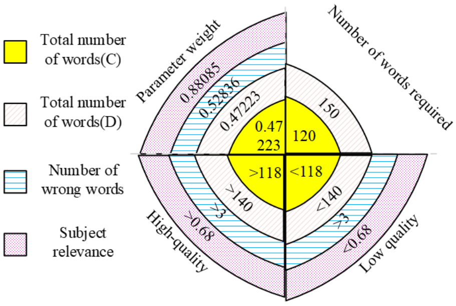


Fig. 3.6: Sketch map of comment generation model

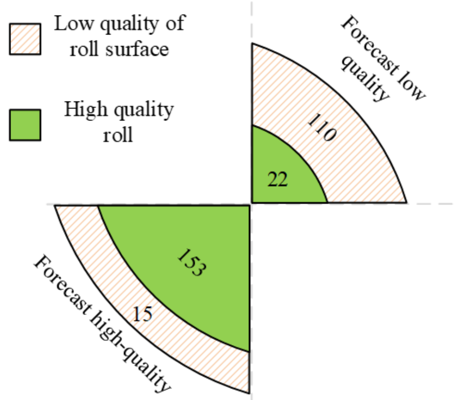
can include the required characteristic indexes into the corresponding index set and, finally, remove the remaining characteristic indexes. In this process, the final score will be obtained through the rough score-medium score-segmentation model. The coarse-middle-segment model built by the research is shown in Figure 3.5.

As can be seen from Figure 3.5, first input the English articles to be evaluated into the rough score-medium score-segment model. This process follows the scoring index system 1 composed of hard indicators such as the total number of rough scoring modules and the number of wrong words. Then, the article gives a low and middle score in index system 2, which is composed of the number of whole sentences and the number of grammatical errors. Then, the article carries out senior high school scores in the index system 3, which is composed of vocabulary cohesion and complex sentence patterns. Finally, the paper subdivides the index system 4 which is composed of sub-dispersion and punctuation frequency; The article is subdivided according to the index system 5 of lexical cohesion, complex sentence patterns and lexical relevance. In addition, real-time feedback in English articles can effectively help students find problems in time and improve their writing ability. Therefore, on the basis of intelligent English grading, the research also adds real-time natural language feedback to students. Among them, the generation of English article comments includes three aspects of comments: vocabulary, syntax and text. Based on this, the comment generation model built in the study is shown in Figure 3.6.

As can be seen from Figure 3.6, in the English article comment generation model, the vocabulary comment



(a) Content of Rough Scoring Index System



(b) High and low quality matching matrix content

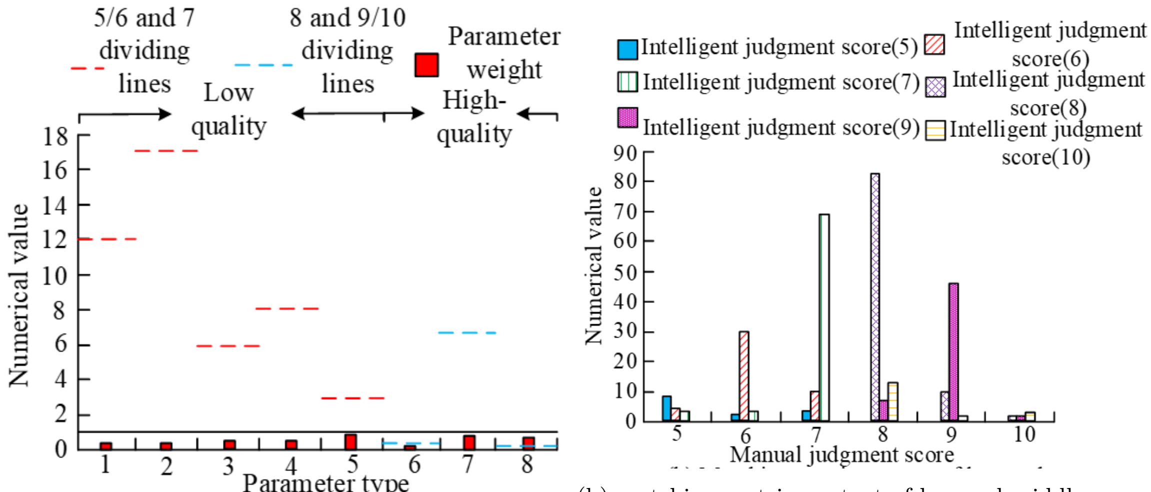
Fig. 4.1: Rough classification index system and high- and low-quality prediction classification content

includes the report of vocabulary size, vocabulary grade and vocabulary errors. Syntactic comments include the report of sentence size, syntactic errors and clauses. The text comments include reports on the partial structure, the cohesion of the article and the relevance of the topic. It is worth noting that the prerequisite for entering the comment generation model is to meet the relevant requirements, that is, whether the number of words in English articles meets the minimum requirements of the scoring model.

4. The practical application of Adaboost/CT algorithm in English intelligent grading. To verify the effectiveness of the Adaboost/CT algorithm in English intelligent scoring, the study validated the performance results of the algorithm in the scoring module and the commenting module, respectively. In the scoring module, the study used 312 English compositions from two grades of students in a certain university as test samples, and a total of 300 English compositions were included in the ”coarse medium subdivision” model constructed in the study. 300 essays cover low, medium, and high levels, and are relatively comprehensive. At the same time, 300 English compositions that are higher than the mandatory indicators are considered high-quality compositions, while those that are lower than the mandatory indicators are considered low-quality compositions, denoted by A and B respectively. In the coarse grading - medium grading - subdivision model, the coarse grading index system and the content of the high- and low-quality matching matrix are shown in Figure 4.1.

In Figure 4.1a, C represents 120 words; D means 150 words. It can be seen from Figure 7 that the coarse score indicator system focuses on topic relevance, and the parameter weight reaches 0.88085. At the same time, the high- and low-quality limits of 120 words and 150 words are 118 and 140; The high- and low-quality limits of the number of wrong words and topic relevance are 3 and 0.68. As can be seen from Figure 4.1b, 110 samples of low quality and 15 samples of high quality are predicted under the low quality of the roll surface; There are 22 samples with low quality and 153 samples with high quality. Overall, 132 English compositions were divided into Group A and 168 into Group B. The accuracy rate of high- and low-quality English composition classification was 87.66%, and the adjacency accuracy was 95.33%. Therefore, Adaboost/CT algorithm shows good performance. When the coarse score - middle score - subdivision model is divided into stages, the low middle score and high middle score indicator system and the low middle score and high middle score matching matrix are shown in Figure 4.2.

In Figure 4.2, the horizontal axis 1-8 represents the parameter types of low, middle and high scores and high scores. Among them, 1-5 represents the integer sentence required by 120 words, the integer sentence required by 150 words, the number of words required by 120 words at level 3 to 4, the number of words required by 150 words at level 3 to 4, and the number of syntax errors. 6 8 means Latent Semantic Analysis (LSA), the



(a) Content of low and middle score and high score index system

(b) matching matrix content of low and middle scores and high scores

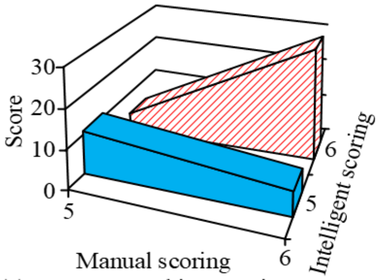
Fig. 4.2: The index system of low middle score and high high score and the content of the matching matrix of low middle score and high score

Table 4.1: Index system content of Group P and Group Q

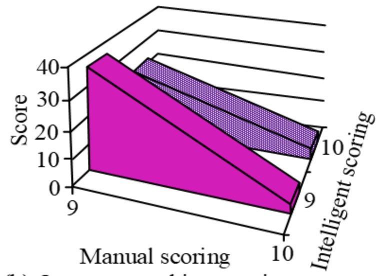
Group P indicator system			
	Sentence dispersion	Punctuation frequency	Part of speech
Parameter weight	0.549	0.805	0.908
5 points	>4.710	>0.670	Many pronouns
6 points	<4.710	<0.670	Many adjectives
Group Q indicator system			
Parameter weight	0.446	0.675	0.876
9 points	>0.763	<4.000	<0.550
10 points	<0.763	>4.000	>0.550

number of words at level 5 6 and grammatical cohesion of all sentences. In addition, there are high scores above the dotted line in the figure. It can be seen from Figure 8 that in the low and middle stages, 107 articles were classified using the Adaboost/CT algorithm, and the accuracy and adjacency accuracy were 81.06% and 95.45% respectively. Among them, 3 compositions with 5 points for manual review were divided into 7 points by algorithm; 10 manually-rated compositions with 6 scores were divided into 7 scores by algorithm; Three manually-reviewed compositions with 7 points were divided into 5 points by the algorithm; Two compositions with a manual score of 5 points were divided into 6 points by the algorithm. In the high score, the Adaboost/CT algorithm is used to classify 132 articles, and the accuracy rate and proximity accuracy rate are 78.57% and 91.07% respectively. In general, Adaboost/CT algorithm has high classification accuracy. In the subdivision stage, the study set the 5-6 score as group P, and the 9-10 score as group Q. The contents of the indicator system of Group P and Group Q are shown in Table 4.1.

It can be seen from Table 4.1 that the focus of the index system of Group P is on part of speech, and its parameter weight reaches 0.908. In the final prediction score, the pronouns under 5 are mostly, and the adjectives under 6 are mostly. The index system of group Q focuses on the verb related characteristics, and the

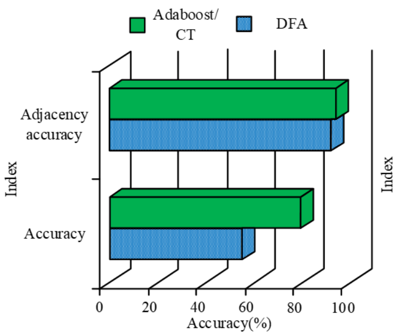


(a) P group matching matrix content

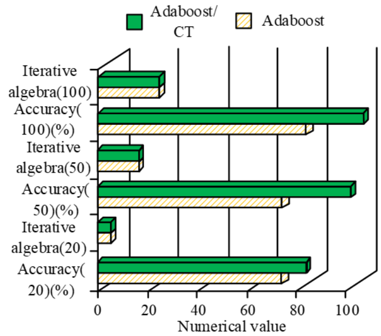


(b) Q group matching matrix content

Fig. 4.3: Matching matrix content of group P and group Q



(a) Accuracy comparison results of Adaboost/CT algorithm and DFT algorithm



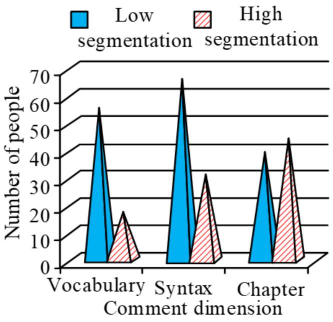
(b) Comparison of classification results between Adaboost and Adaboost/CT

Fig. 4.4: Comparison of algorithm accuracy and classification results under different parameters

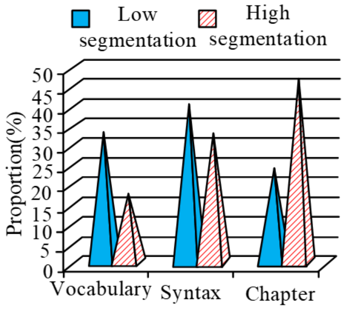
parameter weight reaches 0.87612. Under this indicator system, the contents of two sets of matching matrices obtained by using Adaboost/CT algorithm are shown in Figure 4.3.

According to Figure 4.3, 40 English compositions in Group P were correctly classified according to Adaboost/CT algorithm, with an accuracy rate of up to 80%. There are also 40 English compositions in Group Q that are correctly classified according to Adaboost/CT algorithm, with an accuracy rate of 61.54%. In addition, the predicted score of Group P is 5 or 6, while that of Group Q is 9. Therefore, there is no problem of adjacency accuracy, and there is no problem of adjacency accuracy of 100%. In general, the results of the evaluation of 300 English compositions by Adaboost/CT algorithm show that 239 of them are the same as those of manual evaluation. The prediction accuracy of Adaboost/CT algorithm reached 79.66%, and the adjacency accuracy reached 94%, showing high performance. In order to further verify the results, the Discriminant Function Analysis (DFA) algorithm was introduced and compared with the Adaboost/CT algorithm in accuracy [21]. The classification results of Adaboost and Adaboost/CT are shown in Figure 4.4.

It can be seen from Figure 4.4 that the accuracy and adjacency accuracy of Adaboost/CT algorithm are 79.66% and 94% respectively, which are higher than 55% and 92% of DFA. In addition, under different sample parameters, the classification accuracy of Adaboost/CT algorithm is higher than that of Adaboost algorithm under the same iteration. When the sample parameter is 100 and the number of iterations is 15, the accuracy rate of Adaboost/CT algorithm is up to 94%, far higher than 71.99% of Adaboost algorithm. In



(a) Number of people to be improved under different dimensions



(b) Proportion of people to be improved

Fig. 4.5: Comparison of algorithm accuracy and classification results under different parameters

Table 4.2: Comparison results of different indicators for different algorithms

Algorithm	Performance Indicators		
	Accuracy	Recall	F1
Adaboost/CT	96%	95%	95%
TA	85%	78%	81%
TF-IDF	93%	93%	81%

general, Adaboost/CT algorithm effectively solves the interference of singular value, and its performance is far higher than that of the comparison algorithm. In the evaluation module, we study the use of Adaboost/CT algorithm to evaluate 300 pre-processed English test papers here. On the basis of ensuring the reliability, the experiment invites experts to re-evaluate the scores with objections, so as to obtain reliable scores. Among them, the differences between the processed high and low score segments at different comment levels are shown in Figure 4.5.

It can be seen from Figure 4.5 that after the test paper is reviewed again using the Adaboost/CT algorithm, 163 people need to be improved in the low section, while 5 people need to be improved in the high section. In general, high-level students have a relatively stable grasp of vocabulary level, and only 10.2% of students still need to improve their vocabulary level. Among the students with low syntactic level, 53.6% of them need to make syntactic adjustments, which requires them to conduct deeper research at the grammatical level. In discourse, both junior and senior students must work hard on discourse in order to improve their writing ability. The experimental results show that the use of Adaboost/CT algorithm for English intelligent grading can not only obtain more accurate scores, but also give feedback. The result can help students improve their English level more pertinently. In addition, in the subsequent natural language module, the review results can be used to achieve positive language feedback for articles with elements up to standard; Articles that fail to meet the standards can also be targeted according to the review results. To further validate the performance of the Adaboost/CT algorithm, accuracy, recall, and F1 values were introduced to evaluate its classification performance. At the same time, the Triplet Algorithm (TA) and Term Frequency Inverse Document Frequency (TF-IDF) algorithms were introduced for comparison, and the results are shown in Table 4.2.

From Table 4.2, the classification accuracy of Adaboost/CT is 85%, which is higher than the TA algorithm's 85% and TF-IDF93%; And the recall rate is 95%, which is higher than 78% of TA algorithm and 93% of TF-IDF; The F1 value is 95%, which is much higher than 81% of the TA algorithm and 81% of the TF-IDF algorithm. Overall, the Adaboost/CT algorithm has high accuracy in essay classification. Based on this, the study applied it to actual English grading. A certain grade of students from a certain university covering low, medium, and

Table 4.3: Practical Application Results of Adaboost/CT Algorithm

Algorithm	Number of Tests	Correct Quantity	Accuracy	Average Accuracy	
				1	2
TA	1	400.00	190.00	48.71%	57.26%
	2	390.00	279.00	71.46%	
	3	385.00	166.00	43.00%	
	4	390.00	257.00	65.88%	
TF-IDF	1	400.00	246.00	61.53%	69.43%
	2	390.00	312.00	80.09%	
	3	385.00	231.00	59.88%	
	4	390.00	297.00	76.22%	
Adaboost/CT	1	400.00	308.00	76.91%	76.82%
	2	390.00	301.00	77.22%	
	3	385.00	275.00	71.49%	
	4	390.00	318.00	81.64%	

high levels of English test papers were selected and divided into four groups. The accuracy of the scoring results of the three algorithms on the four groups compared to manual grading was verified, and the results are shown in Table 4.3.

From Table 4.3, the accuracy of Adaboost/CT has decreased in practical applications, but it basically meets expectations. Its accuracy in the four groups is 76.91%, 77.22%, 71.49%, and 81.64%, respectively, most of which are higher than the comparison algorithm, and its average accuracy is also higher than the comparison algorithm.

5. Conclusion. In order to ensure the fairness of English grading and promote the development of sustainable English education, the research has improved and constructed the Adaboost/CT algorithm based on the actual English intelligent grading. In order to verify the proposed model, relevant experiments were carried out. The experimental results show that in the English intelligent grading module, the adjacency accuracy of Adaboost/CT in classifying high and low quality English compositions at the segmentation stage is 95.33%; The accuracy rate and adjacency accuracy rate of high and low quality English compositions in the middle stage are 81.06% and 95.45% respectively; In the subdivision stage, the Adaboost/CT algorithm evaluated 300 English compositions, 239 of which were the same as the manual evaluation, with the prediction accuracy of 79.66% and the adjacency accuracy of 94%. Compared with other DFA algorithms, the results show that the accuracy and adjacency accuracy of Adaboost/CT algorithm are 79.66% and 94% respectively, which are higher than 55% and 92% of DFA. Compared with Adaboost, the accuracy of Adaboost/CT algorithm also has advantages. In the English composition evaluation module, the use of Adaboost/CT algorithm can not only obtain more accurate scores, but also give feedback to each student, so as to improve it pertinently. Meanwhile, its average accuracy in practical applications is 76.82%, which is higher than the comparison algorithm. In general, the intelligent grading method using Adaboost/CT algorithm has shown high effectiveness and practicability in sustainable English education. It is worth noting that when using the Adaboost/CT algorithm in research, there is no syntactic or semantic knowledge involved in English discourse coherence issues. It is necessary to add these knowledge in the future to enhance discourse level coherence issues and further enhance the persuasiveness of intelligent grading.

Funding. The research is supported by: Science and Technology Research Program of Chongqing Municipal Education Commission, Design and Implementation of Cross-border E-commerce English Intelligent Learning Platform Based on Knowledge Map, (No., KJQN202204107); The research is supported by the 13th Five-Year Plan for Education Science of Chongqing Municipality in 2019, The International Development of Chongqing Higher Vocational Colleges under the Belt and Road Initiative: Current Status, Function and Path, (No., 2019-GX-047); The research is supported by Scientific Research and Innovation Team of Chongqing Youth Vocational & Technical College: A Research and Innovation Team for the Construction of the Ideological and

Political System of the English Curriculum, (CQY2021CXTDB01).

REFERENCES

[1] Phoophuangpairoj, R. & Pipattarasakul, P. Preliminary Indicators of EFL Essay Writing for Teachers' Feedback Using Automatic Text Analysis. *International Journal Of Educational Methodology.* **8**, 55-68 (2022)
[2] Dong, Y., Yu, X., Alharbi, A. & AI-and, A. and application of English multimode online reading using multi-criteria decision support system. *Soft Computing.* **26**, 10927-10937 (2022)
[3] Zhou, L. Construction of English Writing Hybrid Teaching Model Based on Machine Learning Automatic Composition Scoring System. *Procedia Computer Science.* **208** pp. 384-390 (2022)
[4] Zhang, L., Huang, Y., Yang, X., Yu, S. & Zhuang, F. An automatic short-answer grading model for semi-open-ended questions. *Interactive Learning Environments.* **30**, 177-190 (2022)
[5] Fang, Y. Design of Oral English Intelligent Evaluation System Based on DTW Algorithm. *Mobile Networks And Applications.* **27**, 1378-1385 (2022)
[6] Canoy, D. & Loquias, A. Identifying Reading Miscues and Reading Performance in the Oral Reading Verification Test in English: Basis for an Intensive Reading Program. *International Journal Of English Language Studies.* **4**, 38-46 (2022)
[7] Štajner, S., Sheang, K. & Saggion, H. Sentence simplification capabilities of transfer-based models. *Proceedings Of The AAAI Conference On Artificial Intelligence.* **36**, 12172-12180 (2022)
[8] Ramesh, D. & Sanampudi, S. An automated essay scoring system: a systematic literature review. *Artificial Intelligence Review.* **55**, 2495-2527 (2022)
[9] Liu, S. Research on the Blended Evaluation Mode in College English Writing Course. *Journal Of Language Teaching And Research.* **13**, 763-771 (2022)
[10] Gaillat, T., Simpkin, A., Ballier, N., Stearns, B., Sousa, A., Bouye, M. & Zarrouk, M. Predicting CEFR levels in learners of English: the use of microsystem criterial features in a machine learning approach. *ReCALL.* **34**, 130-146 (2022)
[11] Liu, L. & Li, Y. An automatic composition scoring system for oversea Chinese students//International Conference on Artificial Intelligence and Intelligent Information Processing (AIIIP 2022). *SPIE.* **12456** pp. 479-484 (2022)
[12] Wu, X. & Zhang, Y. English Speech Scoring System Based on Computer Neural Network. *International Journal Of Education And Humanities.* **5**, 213-216 (2022)
[13] Phoophuangpairoj, R. & Pipattarasakul, P. Preliminary Indicators of EFL Essay Writing for Teachers' Feedback Using Automatic Text Analysis. *International Journal Of Educational Methodology.* **8**, 55-68 (2022)
[14] Taskiran, A. & And, G. and teacher feedback: Writing achievement in learning English as a foreign language at a distance. *Turkish Online Journal Of Distance Education.* **23**, 120-139 (2022)
[15] Miao, L. & Zhou, Q. Research on application of deep neural network model in college English skill training system. *International Journal Of Wireless And Mobile Computing.* **22**, 274-280 (2022)
[16] Regin, R., Rajest, S. & Shynu, T. An Automated Conversation System Using Natural Language Processing (NLP) Chatbot in Python. *Central Asian Journal Of Medical And Natural Science.* **3**, 314-336 (2022)
[17] Sheth, S., Giancardo, L., Colasurdo, M., Srinivasan, V., Niktabe, A. & And, K. and acute stroke imaging. *Journal Of Neurointerventional Surgery.* **15**, 195-199 (2023)
[18] Irfan, M., Ayub, N., Althobiani, F., Ali, Z., Idrees, M., Ullah, S. & Gas, P. Energy theft identification using AdaBoost Ensembler in the Smart Grids. *CMC-Computers, Materials & Continua.* **1**, 2141-2158 (2022)
[19] Walker, S., Solberg, S., Schneider, P. & Guerreiro, C. The AirGAM 2022r1 air quality trend and prediction model. *Geoscientific Model Development.* **16**, 573-595 (2023)
[20] Asif, M., Sheeraz, M. & Sacco, S. Evaluating the Impact of Technological Tools on the Academic Performance of English Language Learners at Tertiary Level: A Pilot Investigation. *Pegem Journal Of Education And Instruction.* **12**, 272-282 (2022)
[21] Vassallo, S., Davies, C. & Biehler-Gomez, L. Sex estimation using scapular measurements: discriminant function analysis in a modern Italian population. *Australian Journal Of Forensic Sciences.* **54**, 785-798 (2022)

Edited by: Mudasir Mohd
Special issue on: Scalable Computing in Online and Blended Learning Environments: Challenges and Solutions
Received: Jun 7, 2023
Accepted: Oct 9, 2023