# Design of Online Learning Efficiency Evaluation Algorithm for College English Based on Data Mining

Hui Li(✉)

Ordos Institute of Technology, Ordos 017000, China
lq96311@126.com

**Abstract.** The current learning efficiency evaluation algorithm has low accuracy and speed due to the singleness of the indicators and the neglect of the management of the indicators. To this end, this study designs an evaluation algorithm for college English online learning efficiency based on data mining. After analyzing the factors that affect the online English learning efficiency of college students, the evaluation indicators are abstracted and an evaluation system is established. Then use the analytic hierarchy process to determine the weight of the indicators in the evaluation system, and build a data warehouse according to the indicators. Finally, ES-ANN integrated sampling neural network is used to mine and analyze the data in the data warehouse, and the evaluation results of the students' learning efficiency are obtained. The experimental results show that the evaluation rate of the algorithm is fast and the evaluation accuracy is higher than 93%, which proves that the method greatly improves the evaluation performance.

**Keywords:** Data mining · College english · Online learning · Efficiency evaluation · Evaluation algorithm

## 1 Introduction

Online teaching is a teaching mode formed by relying on modern information and communication technology and mobile Internet technology. It can realize a ubiquitous teaching situation with the help of rich teaching resources, so that students can break through the limitations of time and space, and become better, more active, and more effective. Participate more fully in the whole process of online teaching activities.

With the outbreak and development of the new crown pneumonia epidemic in the country and even around the world, online teaching and online learning have moved from behind the scenes to the front desk, and have become the mainstream for a while, and have also become the teaching and learning methods used by teachers and students [1]. However, online teaching has also caused new teaching problems. Compared with traditional offline English teaching, students have higher requirements for self-control and concentration when learning, and online learning cannot provide students with a good


W. Fu and L. Yun (Eds.): ADHIP 2022, LNICST 468, pp. 537–548, 2023.
https://doi.org/10.1007/978-3-031-28787-9_40

learning atmosphere. As a result, students' learning efficiency is low, which seriously affects the quality of English online teaching.

The so-called learning efficiency is the ratio of the time and energy consumed by learning to the quantity and quality of learning obtained. Efficient learning can enable college students to acquire more and better knowledge more easily and happily during online English learning, and achieve a good learning effect of cultivating ability and promoting the all-round development of English ability. Therefore, evaluating the learning efficiency of college students who are studying English online can effectively measure the degree to which teaching objectives are achieved, control the teaching process, and identify the effect of online teaching [2].

The randomness of students' online learning time and learning methods makes management teachers lack strong basis and standards when evaluating students' learning status. At present, the learning efficiency evaluation method adopted by many domestic scholars mainly evaluates the learning efficiency of learners effectively by collecting the data of online learning input and output. For example, the online learning efficiency is analyzed by establishing a data envelopment analysis model (DEA). This method mainly analyzes the learning efficiency of the learners by collecting the input and output data of the learners during the online learning process. However, this analysis method considers a single influencing factor, and the analysis results have a large deviation [3]. However, these evaluation methods are not comprehensive enough to consider the influencing factors of online learning efficiency. It is difficult to directly measure the learning efficiency of college English online learners through the above methods only by collecting the data of learners' online learning input and output evaluate.

With the development of the epidemic, online teaching and online learning have become educational hotspots. More and more scholars pay attention to online teaching and online learning and conduct a lot of research on this theme and opportunity.

The scale and time of college students' English learning continues to expand, and the factors that affect students' learning efficiency continue to increase. The conventional evaluation index selection method can no longer meet the requirements of learning efficiency and accuracy. Data mining processes and analyzes a large amount of business data automatically, through a series of reasoning and classification, which helps to improve the evaluation accuracy [4].

Based on the above analysis, considering the large scale of online teaching and the advantages of data mining technology in evaluation work, this paper will design a data mining-based online learning efficiency evaluation algorithm for college English. It is of great significance to improve the quality of students' English learning, online teaching and other progressive research, and to help teachers of online teaching to adjust their teaching methods in a timely manner.

## 2 Algorithm Design

### 2.1 Analysis of Factors Affecting Students' Online English Learning Efficiency

There are many factors that affect students' learning efficiency. This article will analyze the factors that affect college students' online English learning efficiency from multiple dimensions.

Students' interest in English learning, students' attention span and duration of online learning, students' self-control ability in online learning, students' cognitive style, and students' homework completion; teacher factors include teachers' information technology literacy, teachers' class style, teacher's pre-class preparation, teacher's severity, and teacher's inspection and feedback on the content learned; curriculum factors include the duration of online English class, the type of English class, the design style of English class, and the difficulty level of class content; Environmental factors include the degree of family environment interference; equipment factors include class hardware equipment, network quality, teaching platform and web page interference [5].

From the perspective of students, their learning ability and English foundation will affect their English learning efficiency. Students' interest and confidence in English learning, motivation, effort, class participation and family background will also affect their English learning results. In the process of online learning, college students pay more attention to the external needs brought by the fluency of the online teaching process, which motivates them to improve their learning motivation. And the dedication in foreign language learning activities and the positive attitude reflected in the process of this activity. English learning motivation is divided into five categories: effort level, intrinsic interest, extrinsic needs, learning situation and learning value. Online English training makes it difficult for students to concentrate for a long time, and it is difficult to lock in the important information of learning materials, resulting in poor online learning effect. Students' self-confidence in English learning goals, competence in English courses and adaptability to English courses will also affect students' learning efficiency [6].

Due to remote teaching, teachers and students are separated from the screen. Teachers have certain difficulties in the organization and management of online teaching, and the teaching effect is even more difficult to detect. However, the arrangement of teachers' teaching methods and the design of teaching content will affect students' enthusiasm for learning and thus affect students' learning efficiency. The better the comprehensive performance of the learners' emotional participation, cognitive participation and behavioral participation during the learning of knowledge using the online platform, the higher the recognition of the course. The cognitive structure of high-efficiency English learning students has the characteristics of integrity, integrity and connectivity. With the increase of students' age, the influence of learning motivation on learning efficiency is more obvious. From the perspective of the characteristics of students' motivation development, whether it is the level of motivation, the content of motivation, or the intensity of motivation, etc., will change with time and conditions. Fixed place, proper lighting conditions, color, temperature and humidity of the study place also have some influence on study efficiency.

Among the many factors that affect students' online learning efficiency, English learning motivation and self-efficacy, as the two most basic internal influencing factors in students' online learning process, have a significant impact on students' online learning participation.

### 2.2 Establish an Online Learning Efficiency Evaluation System for College English

According to the above multi-dimensional analysis of the influence of college students' online English learning efficiency, the influencing factors are selected as learning efficiency evaluation indicators, and the corresponding college students' English online learning efficiency evaluation index system is established, and the index weights are determined. Table 1 below is the evaluation system of college English online learning efficiency established in this paper [7].

**Table 1.** Evaluation system of college english online learning efficiency

| Target layer | Criterion layer | Indicator layer | Meaning |
|---|---|---|---|
| Efficiency evaluation of college English online learning | Student's own factors | Interest and Confidence in English Learning | From the perspective of students themselves, evaluate students' learning efficiency from multiple perspectives such as interest, method, purpose and initiative |
| | | Motivation to learn | |
| | | Study method | |
| | | Study effort | |
| | | Learning content awareness | |
| | | Class participation | |
| | Teaching Impact | Course content suitability | Judging the efficiency of online learning from the perspective of courses according to the contents, progress, teaching methods and other contents |
| | | Teacher teaching method | |
| | | Course content design | |
| | | Reasonable teaching arrangement | |
| | | Attractiveness of teaching content | |
| | Other factors | Learning environment | Judge online learning efficiency mainly from the perspective of environment |
| | | Learning hardware | |
| | | Interaction with other course schedules | |
| | | Parents concern | |

After establishing the college English online learning efficiency evaluation system shown in Table 1, the fuzzy comprehensive evaluation method is used to establish the initial model of the indicators. According to the quantitative evaluation index system and scoring results, mathematical statistics are carried out, and comprehensive evaluation results are given. The method adopted is to quantify the evaluation grades, and obtain the weighted algebraic sum of grade scores with membership as the weight as the comprehensive evaluation value.

After the layering is established, the pairwise comparison matrix should be established from the second layer of the layered model to the last layer by the method of pairwise comparison. In order to compare the accuracy of judgment under different scale conditions, a comparison method with the scale set to 1 to 9 is used.

When calculating the maximum eigenroot and eigenvector of the matrix, if the requirements are not very high and high accuracy is not required, the method of calculating the method root can be used instead. The specific process is as follows [8]:

Step 1: Multiply all elements of each row of the judgment matrix to get the product $M_i$:

$$M_i = \prod p_{ij} \tag{1}$$

where, $p_{ij}$ represents an element of each row.

Step 2: Calculate the $n$ root of $M_i$ to get $\overline{M}_i$.

The third step: normalize the vector $\left[\overline{M}_1, \overline{M}_2, \ldots, \overline{M}_i\right]$:

$$M_i' = \frac{\overline{M}_i}{\sum \overline{M}_i} \tag{2}$$

After obtaining the normalized comparison matrix, it is necessary to carry out the consistency test, obtain the eigenvectors of the paired matrices through calculation, and then carry out the index consistency test, and then normalize the qualified eigenvectors. Finally, the weight vector of the matrix is obtained. After the weight vector is calculated, the consistency index $CI$ can be obtained by the following formula, as shown below:

$$CI = \frac{t_{\max} - n}{n - 1} \tag{3}$$

where, $n$ represents the dimension of the comparison matrix; $t_{\max}$ represents the maximum value of the comparison matrix.

According to the average random consistency index RI table, to calculate the consistency ratio. The weight vector is obtained according to the above content, and then the feature vectors of each layer are aggregated to form a combined weight vector, and then the consistency detection is performed on the combined weight vector. When the calculated result satisfies the consistency index value, it indicates that the combined weight vector is qualified and can be referred to.

Afterwards, the overall ranking of the hierarchy can be used to obtain the priority of each factor relative to the decision-making target, and the ranking result can be used as the basis for decision-making. Therefore, whether the result of the total ranking of the hierarchy is reasonable or not will directly affect the effect of decision-making. To carry out the total ranking of the hierarchy, it is necessary to perform top-to-bottom layer-by-layer operations on the constructed analytic hierarchy, to obtain the importance value of the lowest-level factor relative to the sub-total target level, and to sort each factor according to the obtained results. The weights of the scheme layer relative to the total target layer are obtained through mathematical operations, and these values are sorted, so that the pros and cons of specific schemes can be analyzed, and the scheme can be selected. After determining the above-mentioned evaluation indicators of influencing factors of learning efficiency, in order to facilitate data mining processing, a data mining warehouse of influencing factors of learning efficiency is constructed.

### 2.3 Building a Data Mining Warehouse for Influencing Factors of Learning Efficiency

Data preparation is the foundation of data mining, and only sufficient data preparation work can ensure the effect of data mining. Determining the goals and objects of data mining is the primary task of data mining. Only when the goals and objects of data mining are established, can data mining work be carried out accurately. Therefore, this paper uses Python language and PyCharm compilation platform as data mining tools to mine and analyze the behavior data of students during English online learning, and establish a data warehouse of influencing factors that affect students' learning efficiency.

As the most popular language now, Python's biggest advantage is that it is easy to use, has an intuitive syntax, and also has many data mining-related class libraries. The Pandas library is a powerful time series processing library that provides the tools needed to efficiently manipulate large datasets; Scikit-learn is an algorithm library for implementing machine learning in Python, providing commonly used machine learning algorithms. Two Python libraries, the Pandas library and the Scikit-learn library, provide the foundation for data mining with Python. After determining the evaluation indicators of students' online English learning efficiency, establish a data warehouse dimension table. The online English learning efficiency evaluation data warehouse dimension table established in this paper is shown in Table 2 below (only part) [9].

According to the teaching practice experience, statistical analysis and design from two different perspectives, vertical and horizontal, can comprehensively grasp the students' online learning situation. Longitudinal refers to taking individual students as the object, and aims to describe the overall situation and details of students' online learning. Horizontal refers to taking the course as the object and aims to describe the learning situation of the students of the course. The horizontal angle is helpful for teachers to understand the overall situation of students' learning in each course, and use this as a reference to adjust the teaching content of the course. According to the objective laws of study, students usually study hard and conscientiously, and correspondingly, they will get better final grades. Therefore, the historical data of students' online learning behavior can be mined to find out the relationship between students' learning behavior and their corresponding final grades, and use this as a reference for evaluating students' online learning efficiency in the future. In the process of college students' online learning of English, the data collected by the online teaching platform is stored in the data warehouse, and the integrated sampling neural network is constructed to conduct in-depth mining according to the established evaluation index system to realize the evaluation of students' online English learning efficiency.

**Table 2.** Data warehouse dimension table

| Field name | Field Type | Field Description |
|---|---|---|
| studentno | varchar | student ID |
| username | varchar | Name |
| studytype | int | student type |
| bd_id | bigint_not null | Student Learning Behavior Data Fact ID |
| courseid | int | course code |
| svalue | numeric | Starting time |
| evalue | numeric | departure time |
| stayt | int | dwell time |
| lmo | int | motivation to learn |
| evi | int | learning environment |
| tme | int | teaching method |
| tc | int | Teaching content |
| t date | date | function |
| tgrade | int | function |

### 2.4 Realize the Evaluation of Online English Learning Efficiency

In this paper, the data mining method is applied, mainly using the analytic hierarchy process and sampling neural network. According to the common steps of data mining, first extract the necessary data sources in the system, then clean the data, and then perform data mining. All the data originally obtained are the data of all students in the school obtained from the school database, and some database tables have different coding formats, which need to be uniformly transcoded for reading, and then filter the required data entries and data items according to the student numbers of the students participating in the experiment., and then check whether there are omissions in each data, some data can be filled, and other missing data can be obtained from other related tables.

Teaching work is a dynamic link, and various factors in each link can affect the quality of teaching, which in turn affects the learning efficiency of students. According to the corresponding data in the dimension table of the data warehouse, extract the most direct relationship between the behavioral characteristics of the students during online English learning and the students' learning efficiency.

This paper introduces ES-ANN ensemble sampling neural network model to mine and analyze the student learning efficiency evaluation index data in the data reference, and obtain the evaluation results of learning efficiency.

After obtaining all the data features required for model training, the specific data is substituted into the model for calculation. For the supervised neural network model, that is, the integrated sampling neural network, the ten-fold cross-validation method is used to divide the data into ten parts. Each time, nine of them are taken as the training set, the other is the test set, and the average is taken as the final result after ten times of training and testing, so that the entire data set can be fully utilized. D represents all datasets, DX train represents the training set, DT test represents the test set, DZ represents the set of positive samples, and DF represents the set of negative samples [10].

The composition of the ten-fold cross-validation dataset is shown in Table 3.

**Table 3.** The composition of the ten-fold cross-validation dataset

| Order | Data set | Make up subsets |
|---|---|---|
| 1 | DT | DT1, DT2, …, DTk |
| 2 | DN | DN1, DN2, …, DNk |
| 3 | D | DT, DX |
| 4 | DT | DT1, DT2, …, DTk-1; DN1, DN2, …, DNk-1 |
| 5 | DX | DNk, DTk |

Because the problem of unbalanced samples needs to be solved, five weak classifiers are used for ensemble sampling, so the negative samples in the remaining training set are randomly divided into five parts, and then the positive samples are copied into the set of each negative sample to form training set of five weak classifiers.

The neural network is built using the keras library, the tensorflow library as the lower layer support, three hidden layers with 64, 8, and 2 nodes respectively. The hidden layer uses relu as the activation function, and the output uses softmax as the classification. The loss function uses category cross entropy, with $o$ and $\hat{o}$ as the actual output and label results, respectively. The calculation formula is as follows:

$$C(w,t) = -\left[\left(o \ln \hat{o}\right) + (1-o)\ln\left(1-\hat{o}\right)\right] \tag{4}$$

where, $w$ represents the network connection weight; $t$ represents the loss parameter.

The model uses the adam optimization algorithm, which is a method that computes an adaptive learning rate for each of its parameters. Its algorithm idea is equivalent to the combination of RMSprop + Momentum algorithm. In addition to saving the exponentially decaying average of the square $\delta_t$ of the past gradients like the Adadelta algorithm and the RMSprop algorithm, the Adam algorithm also stores the exponentially decaying average of the past gradients $\delta_{mt}$ like the momentum algorithm:

$$\begin{aligned} \delta_t &= \alpha\delta_{t-1} + (1-\alpha)\varepsilon^2 \\ \delta_{mt} &= \alpha\delta_{mt-1} + (1-\alpha)\varepsilon \end{aligned} \tag{5}$$

where, $\alpha$ represents the gradient squared bias parameter; $\varepsilon$ represents the gradient descent bias parameter.

If mt and vt are initialized to 0 vectors, their biases will be biased towards 0, and bias correction is made to prevent this from happening.

The gradient update rules are as follows:

$$\delta'_{t+1} = \delta_{t+1} - \frac{\alpha}{\sqrt{\varepsilon + \vartheta}}\delta_{mt} \tag{6}$$

where, $\vartheta$ represents a hyperparameter. According to the above processing process, the relevant index data of the online English learning of the college students are obtained, and the evaluation result of the online English learning efficiency of the college students can be obtained through the analysis and processing of the algorithm.

All the above is the theoretical research process of the efficiency evaluation algorithm of College English online learning based on data mining. The specific evaluation process is shown in Fig. 1.

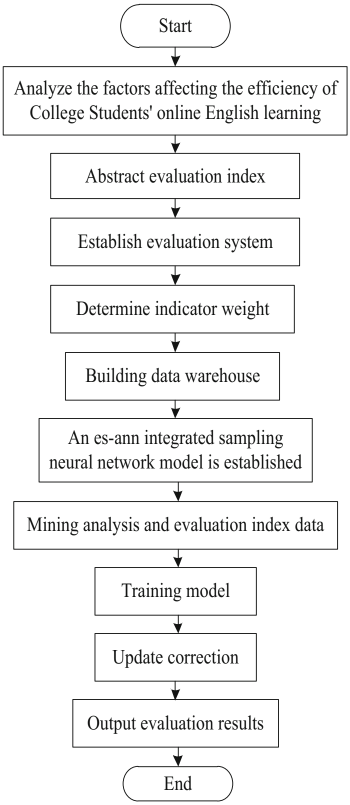


**Fig. 1.** Evaluation flow chart

## 3 Experimental Study

Before the evaluation algorithm is applied to the actual college English online teaching platform, it needs to pass the algorithm performance test. Therefore, this section will use a comparative experimental form to test the performance of the learning efficiency evaluation algorithm.

### 3.1 Experimental Content

Among all the freshman to senior year students who were taught online English in a university, the students were divided into two parts according to English majors and non-English majors, and each part was randomly divided into an experimental group and a control group according to the grades of the students. In the experimental group, the evaluation algorithm designed in this paper was used to evaluate the students' learning efficiency, and in the comparison group, DEA-based evaluation algorithm was used to evaluate the students' learning efficiency. The student's academic performance is used as the evaluation standard of the evaluation algorithm, and the evaluation accuracy of the algorithm on the student's learning efficiency is compared. At the same time, the evaluation time cost of the algorithm is counted to judge the evaluation rate when the algorithm is applied.

### 3.2 Experimental Results

Table 4 is a comparison of the evaluation results of two evaluation algorithms for English majors.

**Table 4.** Comparison of evaluation results of english majors

| Student number | Test group | | Comparison group | |
|---|---|---|---|---|
| | Evaluation accuracy/% | Time costs/s | Evaluation accuracy/% | Time costs/s |
| 100 | 98.6 | 1.3 | 96.7 | 2.4 |
| 150 | 98.7 | 1.5 | 97.2 | 2.6 |
| 200 | 98.4 | 1.9 | 97.3 | 34 |
| 250 | 97.7 | 2.5 | 96.5 | 3.8 |
| 300 | 97.2 | 2.7 | 93.4 | 4.1 |
| 400 | 96.9 | 2.6 | 94.1 | 4.6 |
| 500 | 96.5 | 2.8 | 92.0 | 4.9 |
| 1000 | 96.3 | 3.9 | 90.6 | 6.7 |

Table 5 compares the evaluation results of the two evaluation algorithms for non-English majors.

Comparative analysis of Table 4 and Table 5 shows that the evaluation accuracy of the algorithm in this paper for the learning efficiency of students in different majors in the experimental group has reached more than 93%. There is a big difference in accuracy. Combined with the students' English test scores, the evaluation results of the algorithm in this paper are more reliable. From the time cost of the algorithm, the evaluation rate of the algorithm in this paper is increased by at least 49.6%, and the performance is better.

Summarizing the above experimental analysis content, it can be seen that the evaluation accuracy and evaluation efficiency of the online college English learning efficiency

**Table 5.** Comparison of evaluation results for non-English majors

| Student number | Test group | | Comparison group | |
|---|---|---|---|---|
| | Evaluation accuracy/% | Time costs/s | Evaluation accuracy/% | Time costs/s |
| 100 | 97.5 | 1.4 | 90.6 | 2.7 |
| 150 | 97.2 | 1.6 | 92.4 | 2.9 |
| 200 | 96.8 | 1.9 | 87.5 | 3.4 |
| 250 | 96.4 | 2.3 | 88.6 | 3.8 |
| 300 | 96.1 | 2.5 | 89.3 | 4.2 |
| 400 | 95.6 | 2.6 | 84.1 | 5.7 |
| 500 | 94.7 | 3.1 | 76.5 | 6.3 |
| 1000 | 93.5 | 3.7 | 80.2 | 8.9 |

evaluation algorithm designed in this paper based on data mining are relatively higher, which can more effectively help teachers optimize online teaching methods and content and improve English. The quality of online teaching. The reason for this result is that after establishing the evaluation system, this method uses AHP to give weight to the evaluation indicators, and builds a data warehouse according to the indicators, which improves the pertinence of the evaluation process. Finally, the es-ann integrated sampling neural network is used to mine and analyze the data in the data warehouse, and the learning efficiency is accurately analyzed.

## 4 Conclusion

Traditional classroom teaching has always been dominant in English teaching, and online teaching has always existed as an auxiliary tool as a new teaching method. Under the influence of the new crown pneumonia epidemic, many problems have arisen in students' online learning in the online teaching conducted by various primary and secondary schools and colleges and universities. From the overall situation of online teaching, students participate in online teaching and online learning. The degree is obviously not high, which makes the efficiency of students' online learning very low.

The efficiency evaluation algorithm of college English online learning based on data mining proposed in this paper provides methods and means for objectively, fairly and reasonably evaluating students' online learning, This algorithm utilizes es-ann integrated sampling neural network to mine and analyze relevant evaluation data, with fast evaluation speed and high evaluation accuracy, which can help teachers deeply grasp students' online English learning situation and guide students to reasonably arrange online English learning.

However, due to the limitation of research time and other conditions, the algorithm proposed in this paper still has some shortcomings. In future research, the algorithm will be further optimized from the perspective of improving the diversity of assessment.

## References

1. Jiang, M.: Learning assessment: the path choice to improve college students' learning quality. Heilongjiang Res. High. Educ. **37**(8), 45–48 (2019)
2. Yao, Y., Xu, J., Zhu, X.: Online learning evaluation based on processing technologies in 2D and 3D images. Comput. Technol. Dev. **31**(12), 128–134 (2021)
3. Lijuan, D.: Research on university education quality grading evaluation based on data mining. Mod. Electron. Tech. **43**(15), 101–104 (2020)
4. Liu, J., Huang, Y., Y, L.: The mining and analysis of the evaluation data of classroom teaching. J. Educ. Sci. Human Normal Univ. **18**(2), 118–124 (2019)
5. Lu, Z., Liu, Z., Zheng, Q.: Evaluation of the efficiency of online learning learners based on data envelopment analysis. J. Open Learn. **24**(02), 30–38 (2019)
6. Peng, L.: How does evaluation promote learning?——Empirical analysis on the effectiveness of vocational education learning evaluation. Vocat. Tech. Educ. **42**(22), 37–44 (2021)
7. Gou, R., Ye, X., Wang, B., et al.: Design of learning quality of intelligent evaluation system based on deep learning. Microcomput. Appl. **37**(09), 23–26 (2021)
8. Huang, B., Xie, Y., Tang, Y., et al.: Data storage information serialization completeness and efficiency evaluation simulation. Comput. Simul. **37**(4), 159–163 (2020)
9. Huang, T., Zhao, Y., Geng, J., et al.: Evaluation mechanism and method for data-driven precision learning. Mod. Distance Educ. Res. **33**(1), 3–12 (2021)
10. Hui, C.H.E.N.: Application research of multi-objective decision algorithm in evaluation of data mining. J. Guiyang Coll. (Nat. Sci.) **15**(3), 5–9 (2020)