J. COMBIN. MATH. COMBIN. COMPUT. 127b (2025) 2705--2720

Journal of Combinatorial Mathematics and Combinatorial Computing
www.combinatorialpress.com/jcmcc

Combinatorial Press

A Study on the Metricized Assessment of Foreign Language Talent Cultivation Goal Achievement in Applied Colleges and Universities under the Concept of OBE Based on Decision Tree Algorithm

Shuai Wu[1,✉]

[1] *College of Foreign Studies, Guangdong University of Science and Technology, Dongguan, Guangdong, 523083, China*

ABSTRACT

The evaluation of English course goal attainment is an important basis for colleges and universities to judge whether the goal of cultivating foreign language talents has been achieved. This paper proposes a method for quantitative assessment of course goal attainment according to the OBE concept. Calculating the importance of attributes about classification, the decision tree algorithm based on rough set is proposed, combined with association rules for deep mining of educational data. Collect quantitative educational data and questionnaire data of a university, modeling relying on SPSS Modeler 14.2, and outputting decision tree of influencing factors. Using the evaluation of course goal achievement to analyze the achievement of A4 course goals, and exploring the association rules of influencing factors based on the decision tree. The traditional decision tree algorithm is introduced as a control group to evaluate the performance of the rough set-based decision tree algorithm. The results show that the achievement degree of each sub-objective of A4 course is higher than 0.70, and students who have the achievement degree of A4 course objective greater than 0.7, the nature of their major is foreign language and they have passed the Grade 4 test have a higher possibility of achieving the final foreign language talent cultivation goal of the university. The precision of the assessment method based on rough set decision tree is maintained at about 88%, and the accuracy rate is basically maintained at about 90%.

Keywords: OBE concept, decision tree, association rule, attainment evaluation, foreign language talents

✉ Corresponding author.
E-mail address: christy15243629893@163.com (S. Wu).
Received 05 August 2024; Revised 25 October 2024; Accepted 15 February 2025; Published Online 16 April 2025.
DOI: 10.61091/jcmcc127b-152

1. Introduction

With the in-depth development of reform and opening up, China's contacts and exchanges with other countries in the world are getting closer and closer, and the trend of globalization makes international exchanges become more and more frequent. Foreign language teaching has gone through the ups and downs in the changes of the times, and has shown its vigorous momentum and fruitful results under the contemporary theme of peace and development [1-4]. However, although most colleges and universities have made a lot of reforms and practices in the cultivation of applied foreign language talents, they still face some common problems, such as the mismatch between the real needs and the cultivation of talents, the cultivation concept and the cultivation effectiveness, the cultivation objectives and the cultivation methods [5-8]. As the employment field of foreign language majors gradually develops in the direction of diversification, the cultivation of high-quality applied foreign language talents adapted to the needs of regional economic construction and social development has become an important direction for the reform of talent cultivation of foreign language majors, and the concept of OBE has gradually been emphasized [9-12].

OBE, or "Output-Based Education", is a teaching concept that focuses on students' learning outcomes as the goal of instructional design and implementation. It focuses on what kind of outcomes students ultimately obtain and can transform them into a kind of competence [13-16]. Compared with the traditional teaching mode, OBE education concept emphasizes student-centeredness, which helps to promote the improvement of students' practical level and innovative spirit, promote the process of quality education with students as the main body and teachers as the guides, and promote the growth of teaching and learning at the same time [17-20]. In the guidance of this teaching concept, we no longer over-emphasize the role of the teacher's "teaching", but focus on highlighting the "output capacity" of students, so as to accurately grasp the individual characteristics of students. Under this teaching concept, the students' professional ability and comprehensive quality have been improved in a subtle way [21-24].

Based on the OBE concept, this paper uses quantitative and qualitative evaluations to assess the degree of achievement of course objectives. Data mining technology is utilized to explore the influencing factors of goal attainment, and the process of association rule mining is described. The decision tree algorithm is briefly sorted out, and the rough set-based decision tree algorithm is designed. Starting from the two directions of course goal attainment evaluation and data mining, the goal attainment of foreign language talent cultivation in colleges and universities is studied. The rough set-based decision tree algorithm is used to analyze the importance of the factors affecting the achievement degree of course objectives. Based on quantitative data and qualitative data, the degree of achievement of each teaching objective and the degree of achievement of the total objectives of the course were calculated. With the help of Apriori correlation algorithm model, the correlation between each influencing factor and the degree of achievement of foreign language talent cultivation objectives in colleges and universities is analyzed. Set up controlled experiments to verify the effectiveness of the decision number algorithm based on rough set.

2. The basis and manner of evaluating the degree of achievement of course objectives under the concept of OBE

The OBE concept means that the goal of instructional design and implementation is the final result achieved by students through the educational process. According to the OBE concept, there are two ways of evaluating the achievement of course objectives: one is direct evaluation, which accounts for 80% of the total and is mainly based on quantitative evaluation; the other is indirect evaluation, which accounts for 20% of the total, and is mainly evaluated by qualitative evaluation. The method of direct evaluation is to analyze the course examination results, including process assessment (midterm quiz, homework results), outcome evaluation (final examination) and comprehensive evaluation, calculated based on the teaching design, in which the process assessment accounts for 40% and the outcome assessment accounts for 60% of the comprehensive evaluation. The data of indirect evaluation comes from the feedback results of questionnaires and interviews, which belongs to the basis of qualitative evaluation of the achievement of course objectives, and the main way to take is to carry out targeted questionnaire surveys on the students of the courses studied.

3. The application of data mining technology in the assessment of the achievement of teaching objectives in colleges and universities

Although the evaluation of course goal attainment can realize the quantitative assessment of talent cultivation goal attainment, it cannot find out the factors affecting the cultivation goal attainment. In order to explore more valuable information, data mining techniques are used to further process and analyze the degree of achievement of talent training objectives.

3.1. Association Rule Mining

Association rules are one of the important data mining paradigms, which find out the little-known relationships that exist between different data and bring a lot of unexpected rewards to people. In the field of education, this paper focuses on the process of educational data mining, which can be divided into four phases: the preprocessing phase, the phase of finding out all frequent itemsets, the rule discovery phase and the post-processing phase.

1) Pre-processing phase. Educational data is very rich, but the data involved often contain many incomplete, noisy, and even inconsistent data, so the data to be used must be cleaned, integrated, transformed, discretized, and normalized and other related processing before data mining, in order to make the results of data mining is the useful information and knowledge we want.

2) Finding all frequent itemsets. This stage is mainly to find out all the frequent itemsets from the educational and teaching data set. Frequent items means that a certain item set must appear frequently in all the items, and the frequency of its appearance must satisfy certain requirements, then we call the frequency of the appearance of a set of item sets as the degree of support. Now find a contain X and Y two classes of students of teaching materials as an example, we use the formula $sup(X,Y)=P(X\cap Y)$ calculation, can quickly come up with contains $\{X,Y\}$ item set of the degree of support. According to the support degree formula, if the support degree is not less than the minimum support degree set, then the item set $\{X,Y\}$ can be called a frequent item set.

3) Rule discovery phase. The rule discovery phase is the statistical filtering of the discovered frequent item sets to generate association rules. The main idea of this phase is to use the formula to calculate the support and confidence must not be less than the minimum support and minimum confidence given in advance. For example, the confidence level of rule XY generated by frequent item set $\{X,Y\}$ can be obtained by formula $conf(X\rightarrow Y)=sup(XnY)/sup(X)=P(Y/X)$, if the confidence level is greater than or equal to the minimum confidence level, then it is called XY as the association rule.

4) Post-processing stage, The post-processing phase includes rule evaluation and visual presentation of the rules. The result of the post-processing stage is to output the association rules that the user cares about in various forms, including data tables, various statistical graphs and so on. The mining process of association rules is shown in Figure 1.

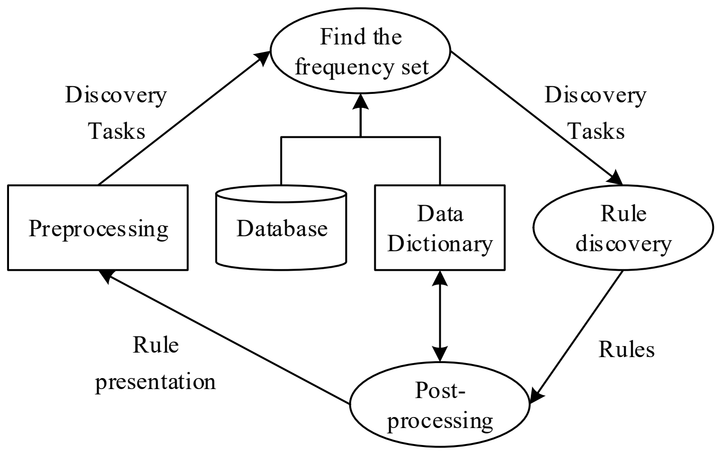

Fig. 1. Association rule mining process

3.2. Decision trees

3.2.1. Overview of decision trees. Association rule mining and decision tree classification are both important research directions in data mining, decision tree is a data mining method based on real cases, using decision tree model can show the key features of classification decisions, and then the meaning and laws behind the features can be further explained by association rules. Decision tree is a method in machine learning with a tree-like structure that approximates a discrete-value structure, in which each internal node represents a judgment on an attribute, each branch represents the output of the result of each judgment, and finally, each leaf node represents the result of a classification. The basic decision tree structure is shown in Fig. 2, where rectangles are used to represent the internal nodes and the root node of the decision tree, and ellipses are represented as the leaf nodes of the decision tree, and the structure as a whole is presented as a tree structure, where A, B represent different judgment attributes, $a1$, $a2$, $b1$, $b2$ represent decision conditions, and $d1$, $d2$, $d3$ represent the results of classification. As shown in Fig. 2, the root node has only outgoing edges and no incoming edges; the internal nodes have one incoming edge and at least two outgoing edges; and the leaf nodes have only one incoming edge and no outgoing edges.

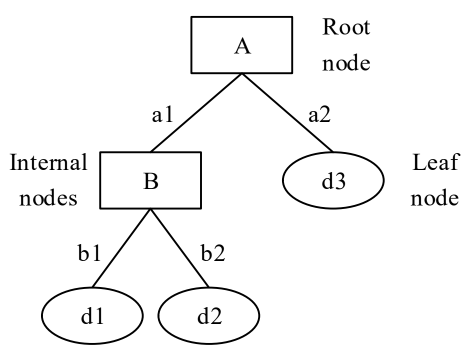

Fig. 2. Decision tree concept

Decision tree will be a set of messy, irregular data in accordance with certain attributes and conditions for the construction of the decision tree model, in the whole process, the use of the top-down recursive way, starting from the root node, along the branch to reach the leaf nodes, which, each path is a classification rule, the combination of all the rules, constitutes a classifier, can be used for predicting problem-solving scenarios.

3.2.2. Decision Tree Correlation Algorithm. C4.5 algorithm is a classification decision tree algorithm among data mining algorithms, and its base algorithm is ID3 algorithm.ID3 algorithm is a classification prediction algorithm, which is based on information theory, and uses information entropy and information gain degree as the measure, so as to realize the inductive classification of data.C4.5 algorithm and CART algorithm are algorithms which are improved and optimized on the basis of ID3 algorithm, and it mainly uses the information gain rate to classify the data. After that, C5.0 algorithm is proposed on the basis of C4.5 algorithm, C5.0 is the commercial version of C4.5, which has improved in its operation rate, while CART algorithm mainly uses the Gini coefficient. The algorithms are not the same, and the corresponding data measurements are also not the same, and the relevant data standard attributes are explained as follows.

1) Information Entropy. Information entropy refers to the measure of uncertainty of random variable $Y = \{c1, c2, c3, \ldots\ldots ck\}$, the greater the information entropy, the greater its uncertainty, in the classification category, that is to say, the more information contained, the greater the entropy, calculated as shown in the following equation (1), X represents a certain classification, $p(x)$ represents the probability of choosing the classification, $H(X)$ represents the sum of all categories, if there is only one category, $p(x)=1$, the entropy that is, the entropy is 0, the smaller entropy means the higher the certainty.

$$(X) = -\sum_{i=1}^{n} p\left(x_i\right) \log p\left(x_i\right) \tag{1}$$

2) Information Gain. Information gain indicates the difference in information entropy before and after the data set is divided according to the feature T attribute. When dividing the dataset, the method of obtaining the maximum information gain should be selected, and the larger the information gain, the stronger the classification ability. The specific calculation is shown in equation (2) below.

$$Information\ Gain\ (T) = Entropy(S) - Entropy(S/T) \tag{2}$$

3) Information gain rate. C4.5 adopts the information gain rate as a branching criterion for selecting branching attributes for the decision tree, which indicates the ratio of useful information generated by the branch, and the larger the value, the more useful information the branch contains, and its expression is shown in the following equation (3), Gain is the information gain, and $H\left(S,A\right)$ is the information entropy.

$$ainRatio(S,A) = \frac{Gain(S,A)}{H(S,A)} \tag{3}$$

4) Gini coefficient. The Gini coefficient is used to express the uncertainty of the data, the larger the Gini coefficient, the larger the uncertainty of the sample set. Specifically in the classification problem, set there is K category, the probability of the K category is pk , its Gini coefficient expression is shown in the following formula (4).

$$Gini(p) = \sum_{k=1}^{k} pk(1-pk) = 1 - \sum_{k=1}^{k} p^2 k \tag{4}$$

If it is a two-class classification problem, the calculation will be relatively simple, if the probability of belonging to the first sample output is p, the Gini coefficient is expressed as the following equation (5):

$$Gini(p) = 2p(1-p) \tag{5}$$

3.2.3. Rough set based decision tree algorithm. Definition (6) Given an information system $\langle U, C\cup D, V, f\rangle$, $R\subseteq(C\cup D)$, $\forall X\subseteq U$ and a division $\pi(U)=\{X_1,X_2,\ldots,X_n\}$ of the domain U independent of the equivalence relation R, knowledge R about the set X is defined in terms of importance:

$$sig_R(X)=\frac{|U-bn_R(X)|}{|U|} \tag{6}$$

Knowledge R is defined with respect to the importance of division $\pi(U)$:

$$sig_R(\pi(U))=\frac{\sum_{i=1}^{n}\left|U-bn_R\left(X_i\right)\right|}{n|U|} \tag{7}$$

For a decision table, a division $\pi(U)$ of U can be obtained from D. If knowledge R is a single attribute $a_{(}(i=1,2,\ldots,m)$, then $\mathrm{sig}_x(\pi(U))$ describes the importance of knowledge R with respect to classification.

From the importance of attributes about classification, then the rough set based decision tree induction algorithm is described as follows:

Input: decision tables $\langle U,C\cup D,V,f\rangle$, $C=\{a_1,a_2,\ldots,a_m\}$, decision attributes with values $V_D=\{d_1,d_2,\ldots,d_n\}$;

Output: Decision Tree

Step 1: For a given decision table $\langle U,C\cup D,V,f\rangle$, compute its classification with respect to the decision attributes

$\pi(U)=\{X_1,X_2,\ldots,X_n\}$. where $X_i=[x]_{di}, i=1,2,\ldots,n$.

Step 2: $for(i=1;i<=m;i++)$

{

$$for(j=1;j<=n;j++) \tag{8}$$

{

$$\left\{\underline{ai}\left(X_j\right)=\left\{x\,|\,(\forall x\in U)\wedge\left([x]_{ai}\subseteq Xj\right)\right\}\right. \tag{9}$$

$$\overline{a}_i\left(X_j\right)=\left\{x\,|\,(\forall x\in U)\wedge\left([x]_{ai}\cap Xj\neq\varnothing\right)\right\} \tag{10}$$

$$bn_{ai}\left(X_j\right)=\overline{a}_i\left(X_j\right)-\underline{ai}\left(X_j\right) \tag{11}$$

$$sig_{ai}\left(X_j\right)=\frac{\left|U-bn_{ai}\left(X_j\right)\right|}{|U|} \tag{12}$$

}

$$sig_{ai}(\pi(U))=\frac{\sum_{j=1}^{n}sig_{ai}(X)}{n} \tag{13}$$

}

Step 3: Calculation $a_j=\arg\max_{1\le i\le m}\{siga(\pi(U))\}$

Step 4: Calculate $U/a_j=\{U_1,U_2,\ldots,U_k\}$

Step 5: $for\left(i=1;i<=k;i++\right)$

{

If the samples in U_i all belong to the same class, end;

Otherwise, repeat steps two through five for U_i;

}

This algorithm mainly constructs a decision tree table by calculating the importance of the attributes relative to the division, and the following is an example of a training sample set of playing tennis to verify the effectiveness of this algorithm.

Suppose a date attribute is added to the training sample set, and assume that the date attribute values are all different, so that if the decision tree is constructed with the ID3 algorithm, since the algorithm tends to take the attribute with more values as the extended attribute, for such a data set, it is obvious that the date attribute has the greatest information gain, and it will be chosen as the extended attribute. The algorithm is terminated by using it to split the sample set and obtaining 14 sample sets, while each sample set contains only one sample. The final decision tree obtained is a one layer decision tree with one root node, 14 branches and leaf nodes. Obviously such a decision tree does not have any generalization ability and the rules it represents are meaningless to us. Therefore the ID3 algorithm fails for such a sample set. Next we construct a decision tree using the rough set based decision tree algorithm proposed above. The construction steps are as follows:

Argument domain U on the division $\pi(U)=\{Y_1,Y_2\}$ of decision attributes D.

Among them: $Y_1=\{D_1,D_2,D_6,D_8,D_{14}\}\,Y_2=\{D_3,D_4,D_s,D_7,D_9,D_{10},D_{11},D_{12},D_{13}\}$

First calculate the 5 conditional attributes weather, temperature, humidity, and wind speed and test the importance of these 5 attributes with respect to division $\pi(U)=\{Y_1,Y_2\}$.

Calculate the importance of attribute weather with respect to the division $\pi(U)=\{Y_1,Y_2\}$ by first calculating the importance of attribute weather with respect to the Y_1 Equivalence classes for attribute weather:

$$U/\text{Weather}=\{\{D_1,D_2,D_8,D_9,D_{11}\},\{D_3,D_7,D_{12},D_{13}\},\{D_4,D_5,D_6,D_{10},D_{14}\}\}$$

The lower approximation of the attribute weather with respect to Y_1 is calculated as: $\varnothing$

The upper approximation of the property weather with respect to Y_1 is calculated as: $\{D_1,D_2,D_4,D_s,D_6,D_8,D_9,D_{10},D_{11},D_{14}\}$

Thus the property weather with respect to Y_1 is bounded by: $bn_{\text{Weather}}(Y_1)=\{D_1,D_2,D_4,D_5,D_6,D_8,D_9,D_{10},D_{11},D_{14}\}$

So the importance of attribute weather with respect to Y_1 is:

$$sig_{\text{Weather}}(Y_1)=\frac{|U-bn_{\text{Weather}}(Y_1)|}{|U|}=2/7 \tag{14}$$

Similarly calculate the importance of attribute weather relative to Y_2 as:

$$sig_{\text{Weather}}(Y_2)=\frac{|U-bn_{\text{Weather}}(Y_2)|}{|U|}=2/7 \tag{15}$$

So the importance of attribute weather with respect to division $\pi(U)=\{Y_1,Y_2\}$ is:

$$sig_{\text{Weather}}(\pi(U))=\frac{sig_{\text{Weather}}(Y_1)+sig_{\text{Weather}}(Y_2)}{2}=2/7 \tag{16}$$

Similarly, one can calculate the importance of temperature, humidity, wind speed, and test relative to the division respectively:

$$sig_{\text{Temperature}}(\pi(U))=0, sig_{\text{Humidity}}(\pi(U))=0$$
$$sig_{\text{Wind speed}}(\pi(U))=0, sig_{\text{Test}}(\pi(U))=0 \tag{17}$$

Obviously, the attribute weather has the greatest importance with respect to division $\pi(U)=\{Y_1,Y_2\}$. Therefore, the attribute weather is chosen as the extended attribute, and the domain is divided into three subsets with weather as the root node, respectively:

$$U_1=\{D_1,D_2,D_8,D_9,D_{11}\}U_2=\{D_3,D_7,D_{12},D_{13}\}$$
$$U_3=\{D_4,D_s,D_6,D_{10},D_{14}\} \tag{18}$$

Notice that sample $U_2=\{D_3,D_7,D_{12},D_{13}\}$ belongs to the same class, so end.

Calculate the importance of attributes temperature, humidity, wind speed, and test relative to division $\pi(U_1)=\{Y_1,Y_2\}$ for subset $U_1=\{D_1,D_2,D_8,D_9,D_{11}\}$, where $Y_1=\{D_1,D_2,D_8\},Y_2=\{D_9,D_{11}\}$, respectively.

Continue to compute the importance of attribute temperature with respect to division $\pi(U_1)=\{Y_1,Y_2\}$, first compute the importance of attribute temperature with respect to Y_1. The equivalence class of attribute temperature is:

$$U_1/\text{Temperature}=\{\{D_1,D_2\},\{D_8,D_{11}\},\{D_9\}\}$$

The lower approximation of the property temperature with respect to Y_1 is: $\{D_1,D_2\}$

The upper approximation of the property temperature with respect to Y_1 is: $\{D_1,D_2,D_8,D_{11}\}$

The boundary of the property temperature with respect to Y_1 is: $\{D_8,D_{11}\}$

So the importance of calculating the attribute temperature with respect to Y_1 is:

$$sig_{\text{Passing}}(Y_1)\frac{|U_1-bn_{\text{Passing}}(Y_1)|}{|U_1|}=3/5 \tag{19}$$

Similarly one can calculate the importance of attribute temperature with respect to Y_2 as

$$sigas(Y_2)\frac{|U_1-b_{\text{Passing}}(Y_2)|}{|U_1|}=3/5 \tag{20}$$

So the importance of attribute temperature with respect to division $\pi(U_1)=\{Y_1,Y_2\}$ is:

$$\mathrm{sig}_{\text{Passing}}(\pi(U_1))=\frac{\mathrm{sig}_{\text{Humidity}}(Y_1)+\mathrm{sig}_{\text{Wind Speed}}(Y_2)}{2}=3/5 \tag{21}$$

Similarly, the importance of the other three attributes humidity, wind speed, and test relative to division $\pi(U_1)=\{Y_1,Y_2\}$ can be obtained as:

$$sig_{\text{Humidity}}(\pi(U_1))=1, sig_{\text{Wind Speed}}(\pi(U_1))=0, sig_{\text{Passing}}(\pi(U_i))=0 \tag{22}$$

Obviously, the attribute humidity has the greatest importance relative to the division $\pi(U_1)=\{Y_1,Y_2\}$, and it is chosen as the extended attribute to divide the domain $U_1=\{D_1,D_2,D_2,D_2,D_{11}\}$ into 2 subsets with humidity as the root node, respectively:

$$U_{11}-\{D_1,D_2,D_2\}U_{12}=\{D_0,D_{11}\} \tag{23}$$

The algorithm ends since the objects in both subsets U_{11} and U_{12} belong to the same class. The process is repeated for $U_3 = \{D_4, D_s, D_6, D_{10}, D_{16}\}$ and the final decision tree generated is shown in Fig. 3.

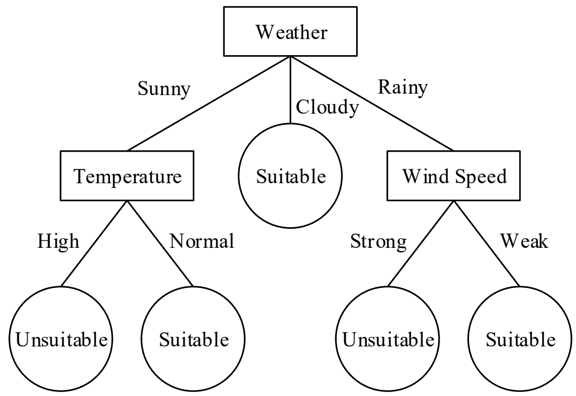

Fig. 3. Playing tennis decision tree

From the final generated decision tree, it can be seen that the generated decision tree is exactly the same as the one generated with the classical ID3 algorithm, but it can overcome the ID3 bias of selecting attributes with more values as the root node of the decision tree, i.e., it can overcome some of the shortcomings of the ID3 algorithm.

4. Analysis of the Achievement Degree of Foreign Language Talent Cultivation Objectives in Colleges and Universities Based on Data Mining Technology

This paper takes a 985 college in city A as the research object, and collects the data of a total of 4,267 students in the class of 2020, which includes the basic information of each student such as school number, name, gender, ID number, college, major, grade, and nature of specialty. According to the concept of OBE, this paper adopts both quantitative evaluation and qualitative evaluation for the analysis of the degree of achievement of the objectives of the English course in this school, and utilizes the rough set-based decision tree algorithm combined with the association rule Apriori algorithm for data mining.

The teaching objectives of foreign language courses in this university are numbered as O1~O4. Considering that students' foreign language proficiency is affected by the nature of their majors and the gender factor, this paper analyzes the degree of achievement of foreign language talent cultivation objectives of the university in terms of the following four indicators: (1) the nature of the students' majors (C1); (2) the gender of the students (C2); (3) the students' passing status of the four-level test (C3); and (4) college English course study (C4). Among them, the nature of their majors is divided into two categories: foreign language and non-foreign language, and there are four college English courses for students, which are represented by A1, A2, A3, and A4, respectively. In order to assess the degree of achievement of the objectives of college English courses, this paper uses E1, E2, E3, E4 and E5 to denote the midterm examination results, homework results, final examination results, overall assessment results and students' self-assessment results, respectively.

4.1. Data collection and processing

4.1.1. Access to quantitative data. For the competencies that can be directly quantified using the test scores taken by the students, the achievement data showing the competencies were collected through the collection of students' grades in each session. In this paper, the grade 4 results and the midterm exams, assignments, final exams, and overall grades of the university English course were collected respectively, and the specific data collected are shown in Table 1.

Table 1. Quantifiable partial data of students' achievements

Serial number	Cet-4 score	A1				...
		Mid-term exam score	Homework score	Final exam score	Overall score	
1	628	94	100	91	93.4	...
2	602	90	92	88	89.2	...
3	578	90.5	93	92	91.9	...
4	546	91	90	85.5	87.5	
5	619	89	97	98	96	...
6	507	95	91.5	82	86.5	...
7	0	81	98	76.5	81.7	...
8	487	88	100	92	92.8	
9	644	98	95.5	95	95.7	...
10	565	92	78	89	87.4	...
...	...	...	...	...	...	...

4.1.2. Access to non-quantitative data. The realization of the course teaching objectives is evaluated by distributing questionnaires to the students, who should first understand the course objectives and graduation requirements, conduct self-evaluation, and then make choices according to the questionnaire set up question options, and the evaluation standard of each course objective is divided into five grades in ascending order according to the realization of the course objectives, and the scores are as follows: complete completion of 5 points, better completion of 4 points, basic completion of 3 points, partial incomplete completion of 2 points, and completely incomplete completion of 1 point. The specific data collected are shown in Table 2.

Table 2. Statistics of students' goal achievement evaluation questionnaire

Serial number	O1	O2	O3	O4
1	4	5	3	2
2	5	3	4	2
3	4	4	5	3
4	4	5	4	4
5	3	4	2	5
6	3	4	3	5
7	3	4	4	3
8	5	3	4	3
9	3	4	3	5
10	5	4	3	5
...	...	...	...	...

4.2. Assessment of Talent Cultivation Goal Achievement Based on Decision Tree Algorithm

4.2.1. Decision Tree Analysis of Influencing Factors. This section analyzes the importance of the impact of each evaluation index on the degree of achievement of the training objectives of foreign language talents, as well as the results of the impact of important evaluation indexes. A rough set-based decision tree algorithm was used to classify each evaluation indicator and modeled using SPSS Modeler 14.2. The parameters of the model are set as follows: the output type is decision tree and interactive validation 10 is selected, global pruning is not used and the severity of pruning is set to 22, and the minimum number of records in each sub-branch is 2. The system outputs the factors that have a greater impact on the achievement of the cultivation objectives of foreign language talents as A4, C1, and C3, and the decision tree is shown in Fig. 4, and the modeling accuracy and precision is verified in 4.3 again.

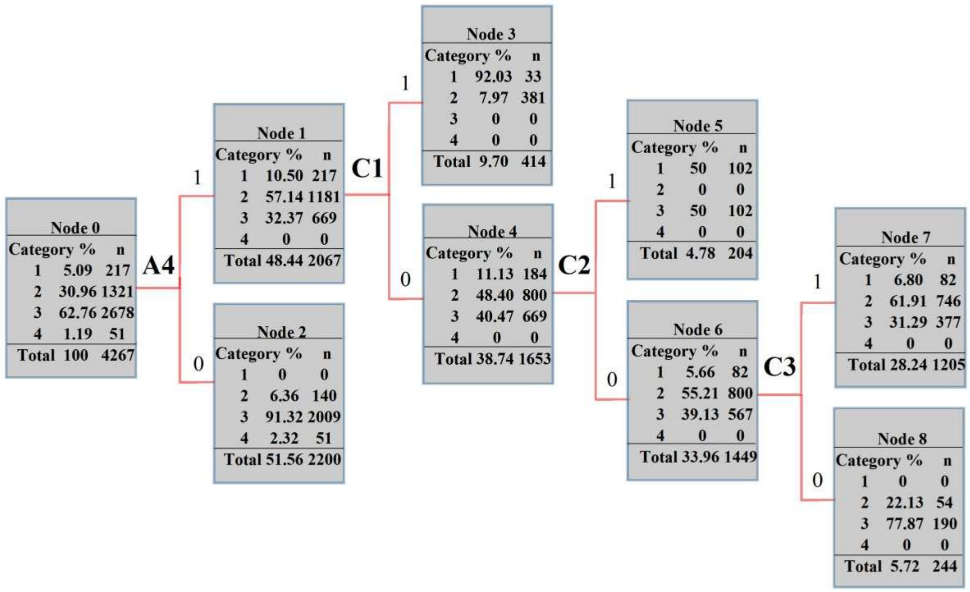

Fig. 4. Decision tree of each influencing factor

4.2.2. Analysis of the degree of achievement of course objectives. A4 course student's final examination results in addition to the entry of the total score value, but also detailed entry of the score of each sub-question, final examination results E5 statistical results shown in Table 3. Sample classification statistics on the scores of each link, according to the weight of the assessment link developed in the syllabus, the teaching objectives of the course and the corresponding relationship with the teaching content to calculate the full marks of the assessment link of the teaching objectives and the actual average score of the students, and then calculate the degree of achievement of the teaching objectives and the degree of achievement of the overall objectives of the course.

Table 3. Statistical analysis results of final exam results

Question type	Test center	Corresponding course objectives	Perfect score	Actual average score	Scoring average
Writing	Essay writing	O1	15	12.6	0.84
Listening comprehension	Long conversation listening	O2	8	7.3	0.91
	Long passage listening	O2	7	5.8	0.83
	Lecture/report listening	O2	20	15.7	0.79
Reading comprehension	Vocabulary understanding	O3	5	3.3	0.66
	Long reading	O3	10	9.1	0.91
	Careful reading	O3	20	14.5	0.73
Translation	Paragraph translation	O4	15	10.2	0.68

The results of the overall assessment of the grades of the A4 course are shown in Table 4. It can be calculated that the evaluation values of the achievement of course objectives O1, O2, O3 and O4 are 0.89, 0.80, 0.84 and 0.83 respectively, and the achievement of each sub-objective is higher than 0.70, and the course objectives are all effectively achieved.

Table 4. Overall evaluation result

Assessment item	Score	O1	O2	O3	O4
E1	Perfect score	25	25	25	25
	Actual score	21.3	22.9	23.2	20.6
E2	Perfect score	25	25	25	25
	Actual score	21.9	20.1	23.3	22.8
E3	Perfect score	15	35	35	15
	Actual score	12.6	28.8	26.9	10.2
E4	Perfect score	19	31	31	19
	Actual score	16.2	25.9	25.4	14.8
E5	Perfect score	25	25	25	25
	Actual score	21.8	21.9	22.7	21.2

4.2.3. Association Rule Mining. Apriori algorithm is the classic association rule data mining algorithm, the core is based on the idea of two-stage frequency set recursion, which searches the dataset layer by layer in order to iteratively identify all the frequent itemsets, and constructs association rules accordingly. After analyzing using Apriori algorithm, each association rule presents $A \Rightarrow B$ form. Support and confidence are two important constraint parameters, support indicates the probability of A and B occurring at the same time, describes the frequency of the association rule, and is a measure of the importance of the association rule, with the expression $Support(A \to B) = \mathrm{P}(A \cup B)$; confidence indicates the strength of the association rule, and is a measure of the reliability of the association rule, with the expression $Confi-dence(A \to B) = \mathrm{P}(A \mid B)$. The students' performance data and the corresponding questionnaire data are analyzed using the Apriori algorithm, where The lower bound of support is set to 0.5, the lower bound of confidence is set to 0.9, and 12 association rules are obtained; then the association rules are sorted according to the size of support, and the results of the first 7 association rules are shown in Table 5, which reveals that if the students' A4 course goal achievement degree is greater than 0.7, their major is foreign language and they have passed the Grade 4 test, then the possibility of achieving the goal of English course is greater, i.e., the possibility of achieving the goal of cultivation of foreign language talents is greater in colleges and universities. Talent cultivation goals are more likely to be reached.

Table 5. Association rules

Association rule	Support(%)	Confidence(%)
A4>0.7,C1=1⇒Goal achievement	100	100
A4>0.7,C2=1⇒Goal achievement	83	100
C2=1,C3=1⇒Goal achievement	83	100
A4>0.7,C2=1,C3=1⇒Goal achievement	83	100
C1=1,C2=1⇒Goal achievement	72	100
C1=1,C3=1⇒Goal achievement	72	100
C1=1⇒Goal achievement	61	100

4.3. Validation of effectiveness

In order to verify the effectiveness of the quantitative evaluation method proposed in this paper, a test work is carried out on Fastone to compare the accuracy and precision of this paper's evaluation method based on rough set decision tree with the traditional decision tree method.

4.3.1. Accuracy. The accuracy can reflect the overall performance, the higher the accuracy, the better the evaluation effect, and vice versa. The comparison of the accuracy of the rough set decision tree-based evaluation method and the traditional method is shown in Table 6. From the test, it can be seen that with the increase of time, the accuracy of the rough set decision tree-based evaluation method is basically stable, remaining at about 88%, while the accuracy of the traditional method decreases rapidly after stabilization, with a maximum of only about 65%. It can be seen that the use of rough set-based decision tree algorithm can significantly improve the accuracy, which has a positive significance in improving the efficiency of the degree of achievement of training objectives.

Table 6. Accuracy comparison(%)

Time/s	Decision tree-based evaluation method	Traditional evaluation method
3	87.98	65.44
6	87.92	65.01
9	87.87	64.41
12	87.76	63.82
15	87.64	63.19
18	87.58	62.43

4.3.2. Accuracy. The comparison of the accuracy of the rough set decision tree-based assessment method and the traditional method in this paper is shown in Figure 5. From the test, it can be seen that when the number of samples increases from 100 to 600, the accuracy of the rough set decision tree-based assessment method basically stays around 90%, with more stable ups and downs; with the increase in the number of samples, the accuracy of the traditional method continues to decline, indicating that it is greatly affected by the number of samples, and the more the number of samples is, the lower the accuracy is.

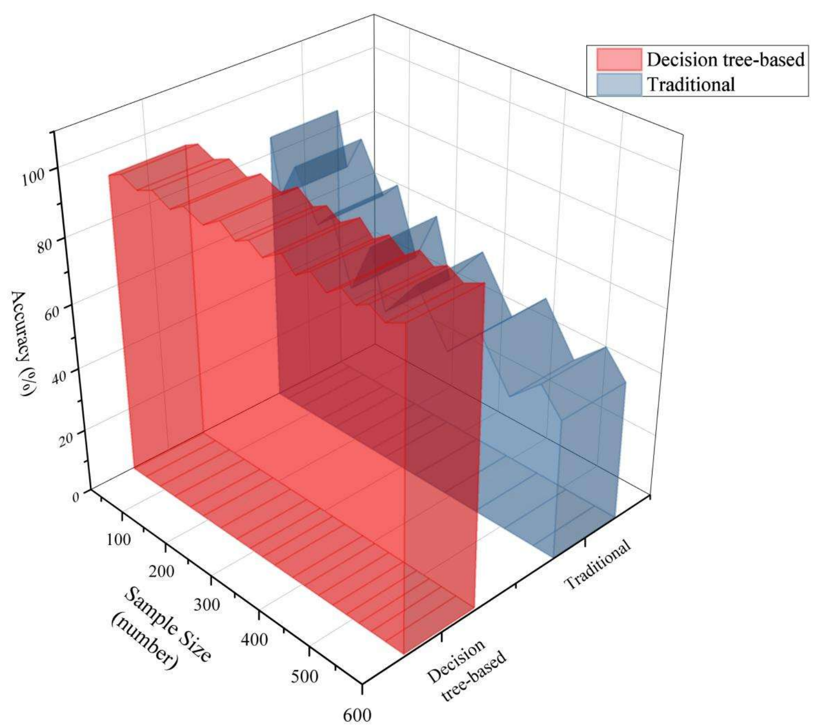

Fig. 5. Accuracy comparison

5. Conclusion

This paper proposes to use rough set-based decision tree algorithm to quantitatively assess the achievement of foreign language talent cultivation objectives in colleges and universities, and comparatively test its effectiveness.

The system outputs the factors that have a greater impact on the achievement of foreign language talent training goals are A4, C1 and C3, and the assessment values of the achievement of the course objectives O1, O2, O3 and O4 of A4 course are 0.89, 0.80, 0.84, 0.83, respectively, and the achievement degree of each sub-objective is higher than 0.70.The results of the correlation rule mining show that the students' achievement degree of the objectives of the A4 course is greater than 0.7 if the student's The nature of their major is foreign language and they have passed the Grade 4 test, then the possibility of reaching the final college foreign language talent cultivation goal is higher.

The accuracy of the assessment method based on rough set decision tree remains around 88%, while the accuracy of the traditional method decreases rapidly after stabilization, with a maximum of only 65.44%. The accuracy of the rough set decision tree based assessment method basically stays around 90%, while the accuracy of the traditional method continues to decline with the increase in the number of samples.

Funding

This work was supported by 2024 Guangdong Provincial Education and Science Planning Project: A Research on the Construction of an Evaluation System for the Achievement of Foreign Language Talent Training Goals in Application-oriented University under the OBE Concept (Grant No.: 2024GXJK578) and 2022 Quality Project of the Department of Education of Guangdong Province, “Integrated English Teaching and Research Department” (Project Approval Number: No. [2023]4: 244).

References

[1] Lu, Y., Zhang, Y., Cao, X., Wang, C., Wang, Y., Zhang, M., ... & Zhang, Z. (2019). Forty years of reform and opening up: China’s progress toward a sustainable path. Science advances, 5(8), eaau9413.

[2] Liu, X. (2020). Structural changes and economic growth in China over the past 40 years of reform and opening-up. China Political Economy, 3(1), 19-38.

[3] Lei, L., & Qin, J. (2022). Research in foreign language teaching and learning in China (2012–2021). Language Teaching, 55(4), 506-532.

[4] Qian, L., & Garner, M. (2019). A literature survey of conceptions of the role of culture in foreign language education in China (1980–2014). Intercultural Education, 30(2), 159-179.

[5] Kong, Y. (2018). A Study of the Training Mode of Practical Talents of Foreign Languages in Colleges and Universities. Journal of Language Teaching & Research, 9(5).

[6] Shujie, W. (2017). How to Cultivate Application-Oriented Foreign Language Talents to Meet the Needs of the Times in Local Universities and Colleges. In Asia International Symposium on Language, Literature and Translation (Vol. 2010, No. 2020, p. 542).

[7] Wang, H. (2023, September). Nurturing of Application-oriented Foreign Language Talents in Local Tertiary Education in China: An Outcome-based Perspective. In SSEME workshop on Social Sciences and Education (SSEME-SSE 2023) (pp. 4-13). Atlantis Press.

[8] Yin, W. (2020). Research on the Transformation and Development of College English Teachers under the Mode of Applied Talents Training. Journal of Contemporary Educational Research, 4(6).

[9] Wang, F. (2020). How to use the internet thinking to cultivate applied English talents in colleges and universities. TWP Series in Education, Sport Sciences and Physiology, 1, 91-93.

[10] Liu, W. (2023). Exploration of Training Approaches for Applied English Talents in the Context of New Liberal Arts. International Journal of Mathematics and Systems Science, 6(3).

[11] Liu, F. (2020). Research on the Cultivation of International Talents in Applied Universities under the Concept of OBE. Frontiers in Educational Research, 3(15).

[12] Zhou, J. (2020). The Trend of the Reform and Development of the Training of Professional Teachers under the Concept of OBE. Lifelong Education, 9(3), 109-112.

[13] Jin, F. (2021). Exploring the blended teaching mode under the guidance of OBE theory—taking the course of English newspaper reading I in Zhejiang Yuexiu university as an example. Open Journal of Modern Linguistics, 11(4), 511-519.

[14] Zhang, L., Xuan, Y., & Zhang, H. (2020). Construction and application of SPOC-based flipped classroom teaching mode in Installation Engineering Cost curriculum based on OBE concept. Computer applications in engineering education, 28(6), 1503-1519.

[15] Lu, X., Xie, Y., & Cui, D. (2021). Research on the Integration of Curriculum Ideology and Politics into Electronic Practical Teaching Based on OBE. International Journal of Frontiers in Sociology, 3(8.0).

[16] He, L., Wang, C., Han, Y., Sun, L., Wang, H., & Gao, X. (2023, October). Research on the cultivation mode of innovative talents in mechanical and electronic engineering based on OBE concept. In 2023 7th International Seminar on Education, Management and Social Sciences (ISEMSS 2023) (pp. 1045-1052). Atlantis Press.

[17] Li, N., Jiang, P., Li, C., & Wang, W. (2022). College teaching innovation from the perspective of sustainable development: the construction and twelve-year practice of the 2p3e4r system. Sustainability, 14(12), 7130.

[18] Liao, Y., Gao, J., Zhao, M., & Qiu, T. (2024, November). Teaching Reform of New Engineering Artificial Intelligence Course Oriented to OBE Concepts. In Proceeding of the 2024 International Conference on Artificial Intelligence and Future Education (pp. 62-67).

[19] Lin, H. (2023). Action research on the application of OBE based hybrid teaching mode of Basic Nursing. Pacific International Journal, 6(1), 12-19.

[20] Irvani, A., Rochintaniawati, D., Riandi, R., Sinaga, P., & Henukh, A. (2024). Analysis of Quantum Physics Lectures from the Perspective of the MBKM and OBE Based Higher Education Curriculum. Jurnal Pendidikan Fisika Dan Teknologi, 10(1), 44-54.

[21] De Guzman, M. F. D., Edaño, D. C., & Umayan, Z. D. (2017). Understanding the Essence of the Outcomes-Based Education (OBE) and Knowledge of its Implementation in a Technological University in the Philippines. Asia Pacific Journal of Multidisciplinary Research, 5(4), 64-71.

[22] Ni, C. (2021). Research on the training path of professional core literacy matching ability of private universities under the background of new liberal arts. Open Journal of Social Sciences, 9(10), 163-172.

[23] Damit, M. A. A., Omar, M. K., & Puad, M. H. M. (2021). Issues and challenges of outcome-based education (OBE) implementation among Malaysian vocational college teachers. International Journal of Academic Research in Business and Social Sciences, 11(3), 197-211.

[24] Zamir, M. Z., Abid, M. I., Fazal, M. R., Qazi, M. A. A. R., & Kamran, M. (2022). Switching to outcome-based education (OBE) system, a paradigm shift in engineering education. IEEE Transactions on Education, 65(4), 695-702.