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

Integration of College English Teaching Resources Based on Divide and Conquer Algorithm in the Era of Digital Transformation

Shali Zhou[1,] ✉

[1] *School of General Education, Hunan University of Information Technology, Changsha, Hunan, 410000, China*

ABSTRACT

The construction of university English teaching resources is an inevitable requirement to adapt to the development of the times and educational reform. Based on the concept of knowledge and classification, this paper puts forward the theory of Rough set, and applies the idea of partition to the data simplification based on Rough set. Based on the applicability of the partition strategy, the partition idea is added in the process of attribute simplification to achieve the purpose of reducing the complexity of the data simplification algorithm about Rough set. After deriving the decision table, the attribute approximation algorithm based on the attribute order and the partition method is given, i.e., the efficient knowledge approximation method based on the partition method for Rough set. Analyze the performance of Rough set efficient knowledge reduction method based on partitioning method in multiple datasets. To build a knowledge acquisition system platform for university English teaching resources using the efficient knowledge reduction method based on the Rough set of the partition method. In the Heart dataset, the classification accuracies of DIDS method, IV-FS-FRS method, and this paper's method are 0.5936, 0.5536, and 0.6689, respectively, and this paper's method outperforms the classification accuracies of DIDS method, IV-FS-FRS method 0.0753, and 0.1153, respectively. The knowledge acquisition system platform of university English teaching resources constructed by using this algorithm has operational advantages in instance analysis.

Keywords: rough set, partition algorithm, knowledge approximation, English teaching resources

1. Introduction

Since the large-scale reform of university English teaching, the level of educational technology has been constantly improved, teaching methods and means have been constantly improved, and the

✉ Corresponding author.
E-mail address: zhoushali0959@163.com (S. Zhou).
Received 15 January 2024; Revised 25 March 2024; Accepted 30 November 2024; Published Online 16 April 2025.
DOI: 10.61091/jcmcc127b-387

industry has gradually recognized the university English teaching mode based on network technology. However, the problems of too many people in too large university English classes, students' varying English proficiency and lack of enthusiasm for learning, insufficient number of teachers and low level of educational technology still exist [1-3]. In recent years, due to the rapid development of mobile Internet technology and the extensive use of mobile devices, a series of new teaching methods and modes have been promoted, such as catechism, micro-courses, flipped classroom, etc., and university English teaching is not willing to lag behind in joining the wave of reform [4-7]. After the continuous exploration and practice in recent years, the college English teaching mode based on multimedia and network environment has been basically formed [8]. Universities and colleges have not only established multimedia classrooms, network independent learning rooms and teaching resource libraries, but also built and improved informatization and networked digital campuses, which have built a brand-new platform for college English teaching and provided a strong guarantee for improving the quality of college English teaching [9-12]. Not only that, university English teachers also give full play to their subjective initiative and creativity, make full use of the advantages brought by computers and networks, constantly enrich and improve their own knowledge structure and technical level, and are committed to changing their roles and practicing new teaching modes [13-15]. Based on this, the university English teaching resources can not be said to be rich and diverse, but due to a variety of factors, these teaching resources have not been effectively utilized, and the reform effect is not as expected [16-17]. The lack of coordinated planning, the scarcity of high-quality ecological resources and the serious repetitive construction make the integration and sharing of resources an urgent task for the current teaching reform and development [18-19].

This paper tries to decipher the foundation of the partition algorithm, including the basic idea of the partition algorithm, applicable conditions, and solution steps. Based on the necessity of the construction and integration of university English teaching resources, from the perspective of knowledge and classification, the relevant features of Rough set are proposed, and the partition method is used to optimize the data simplification algorithm based on Rough set, forming a new knowledge simplification algorithm, i.e., efficient knowledge simplification method based on the partition method for Rough set, which is applied to the knowledge acquisition of university English teaching resources. Multiple datasets are selected from the UCI dataset, and multiple comparative experiments are conducted to analyze the number of simplification results, classification accuracy, efficiency of experimental results, and running time of the efficient knowledge simplification method based on Rough set by partitioning method proposed in this paper.

2. Simplification of resources for teaching English at the university level using the partition methodology

2.1. Foundations related to partitioning algorithms

2.1.1. Basic Ideas of Partitioning Algorithms. The basic idea of a partition algorithm is to decompose a problem of size n into k smaller subproblems. These subproblems are independent of each other and are the same as the original problem, and these subproblems are solved recursively. Then the solutions of the subproblems are combined to obtain the solution of the original problem.

Divide and conquer is a very important algorithm, and the literal interpretation is "divide and rule", which contains two steps: "divide" and "govern", how to divide and how to "govern" after dividing is the key to research.

For the original problem that is large and difficult to solve, it is recursively decomposed into two or more subproblems of the same type until they can be directly calculated.

Finally, an efficient method is used to merge the solution of the sub-problem into the solution of the original problem. This technique is the clever application of "dividing" and "curing". But not all problems can be solved with a divide and conquer algorithm, and some conditions must be met.

2.1.2. Conditions of application of the partitioning strategy. Divide and conquer is a method of problem solving, and the basic idea is: "If a big problem is complex, you can break it down and break it down individually." Partition literally contains the meaning of "division" and "governance", so how to divide and how to "govern" after division has become the key to solving the problem. Usually not all problems can be divided and conquered, only those that can be broken down into sub-problems that are similar to the original problem and close to the meaning. In addition, only those problems that still conform to the meaning of the original issue after the merger can be subject to the divide-and-conquer strategy. In general, problems that can be solved using divide and conquer algorithms have the following characteristics:

1) This problem can be easily solved by reducing it in size to a certain point. This condition is satisfied for general problems because the computational complexity of a general problem increases with the size of the problem.

2) This problem can be decomposed into a number of similar problems of smaller size, i.e., the problem can be decomposed into a number of subproblems to be solved. This is the premise of the application of the partition strategy, generally most of the problems can be satisfied, this feature fully reflects the application of recursive ideas in the partition strategy.

3) A number of decomposed subproblems can be merged into the solution of the original problem. This point is the key, whether to utilize the partition method depends entirely on whether the problem has the third feature, if the first and second features, but not the third feature, you can consider using the greedy method or dynamic programming method.

4) Several subproblems resulting from the decomposition of this problem must be independent of each other, i.e., the decomposed subproblems do not have common subproblems with each other. This involves the efficiency of the use of the partition strategy, if the sub-problems are not independent of each other, there are common parts, the partition method will have to do a lot of repetitive work, the common sub-problems are solved repeatedly. At this point, although you can use the partition algorithm, but its efficiency is low, this time instead of using the dynamic programming algorithm is better.

2.1.3. Solution Steps for the Partitioning Algorithm. The decomposition and merging of sub-problems must not be too complicated; if the decomposition and merging take more time than the violent solution, the loss will outweigh the gain.

1) Decomposition: first decompose the original problem into a number of smaller, independent of each other, the same form as the original problem or similar subproblems.

2) Solve: according to the decomposition rules of the previous step, continue to follow the previous step of the method recursively split the problem until the size of the problem can be simply calculated.

3) Merge: This step is the last and key step of the algorithm, in order to make the algorithm efficient, it is necessary to use a suitable algorithm to solve the original problem using the results of the subproblems.

2.2. Construction and Integration of University English Teaching Resources

2.2.1. Construction of resources for teaching English at university level. The construction of university English teaching resources is an inevitable requirement to adapt to the development of the times and educational reform. In the era of information explosion, the ways and means for students to acquire knowledge have undergone profound changes, and the traditional single teaching mode can no longer meet the diversified needs of students. Teaching resource construction aims to integrate high-quality teaching resources, innovate teaching methods and means, and provide students with rich, diverse and personalized learning experiences [20-22].

First of all, the construction of teaching resources helps to enhance students' learning interest and enthusiasm. High-quality teaching resources can attract students' attention, stimulate students'

learning interest, and make students more active and positive in the learning process.

Secondly, the construction of teaching resources helps to cultivate students' comprehensive use of English. Through rich curriculum resources, students can get in touch with more language materials and cultural background, improve their abilities in listening, speaking, reading and writing, so as to cultivate students' comprehensive use of English.

Finally, the construction of teaching resources helps to promote teaching reform and innovation. Through the introduction of new teaching concepts, methods and means, teaching resources can promote the transformation of college English teaching from the traditional teaching mode to the modernized and informationized teaching mode.

2.2.2. Integration of university English teaching resources. The integrated development of university English teaching resources is to take the improvement of students' English application ability as the leading role, to screen the teaching-valuable resource data from a large number of text, image, audio, video, animation, microclasses, catechisms and other network resources, and to integrate the teaching resource modules suitable for the needs of college students and with the characteristics of the university, so as to improve the comprehensive application ability of the English language of the students, and to realize the knowledge, ability and quality objectives of the English teaching in the university. To realize the goal of knowledge, ability and quality of college English teaching.

According to the current construction of university English teaching resources, it can be divided into three major parts: teaching platform, learning center and material library. The teaching platform generally contains: user management, interactive Q&A, homework review, question bank management, online test, online search and search engine, etc. The learning center mainly contains: courseware indexing, online search and search engine. Learning center mainly contains: courseware index, case library, test bank, literature and so on. The material library mainly provides various teaching resources for teaching, including text, images, animation, audio, video, games and so on. Teaching platform is the foundation of network teaching, learning center is the extension of knowledge, material library for teaching to provide a wealth of material, the organic combination of the three constitute a complete network teaching resources system.

2.3. Knowledge and Classification

Knowledge is the sum of cognition and experience acquired by human beings in the practice of transforming the objective world, which has the characteristics of abstraction, universality and regularity, and is also a guideline for people's behavior. It is a collection of propositions, rules, etc., and can be generally categorized into descriptive knowledge, control knowledge, and process knowledge. In Rough set theory, knowledge is considered to be directly associated with different categorizations of the real or abstract world. Any objective thing in reality is able to be described by knowledge. Therefore, knowledge can be considered as the ability to categorize things as well as the ability to categorize knowledge can be described by the form of expression of a collection of knowledge systems.

Definition 1: Given a set A, the set consisting of all subsets of A is said to be the power set of the set A, denoted as $P(A)$ or 2^A, i.e.:

$$P(A) = 2^A = \{X \mid \forall X \subseteq A\} \tag{1}$$

Definition 2: Given a set X whose powerset is 2^X, the set consisting of any of the elements in the powerset is said to be a set, i.e., each subset $\varnothing$ of the powerset is a cluster of subsets of X.

Definition 3: Let U be a non-empty finite set of objects called an argument domain. Any subset $X \subseteq U$ of the domain U is called a concept or category of the domain U.

For normalization purposes, $\varnothing$ is also regarded as a null concept. Any subset of clusters in an argument domain U is called abstract knowledge about U, or knowledge for short. In practice, this denotes a categorization of individuals in U, and a division $Par(U)$ is defined as follows: given a

non-empty subset of clusters $\pi = \{A_1, A_2, \ldots, A_m\}$ of the thesis domains U and U. If satisfied: ① $A_i \varnothing$.② $A = \bigcup_{i=1}^{m} A_i$. ③ $A_i \cap A_j = \varnothing$, $i \neq j$, $1 \leq i$, $j \leq m$, is said to be a division of the subset cluster π into sets U.

Definition 4: A relation R on a set A satisfies the following conditions if:

1) Self-reversibility, i.e., $\forall x \in A$, xRx.

2) Symmetry, i.e., $\forall x, y \in A$, if xRy, there must be yRx.

3) Transmissibility, i.e., $\forall x, y, z \in A$, if $xRy \wedge yRz$, there must be xRz, then R is said to be an equivalence relation on the set A.

Definition 5: Given an ontology U and a cluster of equivalence relations S on U, a binary group $K = (U, S)$ is said to be a knowledge base on the ontology U.

It can be seen from the above related definitions that the equivalence relation on the domain actually represents division and knowledge, and the knowledge base represents the various kinds of knowledge on the domain derived from the equivalence relation, i.e., the division or categorization schema, and at the same time, it also represents the ability to categorize the domain.

Definition 6: For a knowledge base $K = \{U, P\}$, where U is the thesis domain and P is a cluster of equivalence relations on U, if $Q \subseteq P$ and $Q \neq \varnothing$, then $\cap Q$ (the intersection of all equivalence relations of Q) is also an equivalence relation, denoted $IND(Q)$.

Definition 7: Let there be a knowledge base $K_1 = (U, P)$ and a knowledge base $K_2 = (U, Q)$ if there exists a relationship between the intersection of $IND(P) = IND(Q)$, then the knowledge bases K_1 and K_2 are said to be equivalent, denoted as $K_1 \cong K_2$.

When two knowledge bases are equivalent, it means that objects of the domain can be described with different sets of attributes expressing exactly the same knowledge about the thesis.

2.4. Efficient Knowledge Approach for Rough Set Approximation Based on Partitioning Method

2.4.1. Rough set correlation features. Definition 8: Given a knowledge base $K = (U, S)$, where U is the thesis domain, and S is a cluster of equivalence relations on thesis domain U, then $\forall X \subseteq U$ and an equivalence relation on thesis domain U $R \in IND(K)$, define the lower approximation $\underline{R}(X)$ and the upper approximation $\overline{R}(X)$ of the subset X with respect to the knowledge R, respectively:

$$\begin{aligned} \underline{R}(X) &= \{x \mid (\forall x \in U) \wedge ([x]_R \subseteq X)\} \\ &= \cup \{Y \mid \forall Y \in U / R) \wedge (Y \subseteq X)\} \end{aligned} \tag{2}$$

$$\begin{aligned} \overline{R}(X) &= \{x \mid (\forall x \in U) \wedge ([x]_R \cap X \neq \varnothing)\} \\ &= \cup \{Y \mid (Y \in U / R) \wedge (Y \cap X \neq \varnothing)\} \end{aligned} \tag{3}$$

Based on the above definition, the following concepts are defined for a more concise and clear representation:

The set $bn(X) = \overline{R}(X) - \underline{R}(X)$ is called the R boundary domain of X. $pos_R(X) = R(X)$ is called the R positive domain of X. $neg_R(X) = U - \overline{R}(X)$ is called the R negative domain of X.

The above definitions lead to the following conclusions:

1) The lower approximation $\underline{R}(X)$ is the set of elements in the domain U which, according to the knowledge R, definitely belong to X.

2) The upper approximation $\overline{R}(X)$ is the set of elements in the domain U which, according to the knowledge R, definitely belongs or may belong to X.

3) The R boundary domain $bn_R(X)$ is the set of elements in the domain U which by the knowledge of R cannot be judged to belong either definitely to X or definitely not to X.

4) The negative domain $neg_R(X)$ of R is the set consisting of those elements in the thesis domain U which, according to the knowledge R, are judged to belong definitely not to X.

Definition 9: Let there be given an ontology U and an equivalence relation R on the ontology U, and a division $\pi(U)=\{X_1,X_2,\ldots,X_n\}\in\Pi(U)$ of the ontology U, and that this division is independent of knowledge R. where the subset $X_i\,(i=1,2,\ldots,n)$ is the equivalence class of the division $\pi(U)$, defining the lower and upper approximations of $\pi(U)$ as respectively:

$$\begin{aligned}\underline{R}(\pi(U))&=\underline{R}(X_1)\cup\underline{R}(X_2)\cup\ldots\cup\underline{R}(X_n)=\bigcup_{i=1}^{n}\underline{R}(X_i)\\ \overline{R}(\pi(U))&=\overline{R}(X_1)\cup\overline{R}(X_2)\cup\ldots\cup\overline{R}(X_n)=\bigcup_{i=1}^{n}\overline{R}(X_i)\end{aligned}\tag{4}$$

The $R-$ approximate classification accuracy and approximate classification quality of the division $\pi(U)$ are defined on top of the above formulas respectively as follows:

$$\alpha_R(\pi(U))=\frac{\sum_{i=1}^{n}\left|\underline{R}(X_i)\right|}{\sum_{i=1}^{n}\left|\overline{R}(X_i)\right|}=\frac{card(\underline{R}(\pi(U)))}{card(\overline{R}(\pi(U)))}=\frac{\left|\underline{R}(\pi(U))\right|}{\left|\overline{R}(\pi(U))\right|}\tag{5}$$

$$\gamma_R(\pi(U))=\frac{\sum_{i=1}^{n}\left|\underline{R}(X_i)\right|}{|U|}=\frac{card(\underline{R}(\pi(U)))}{card(U)}=\frac{\left|\underline{R}(\pi(U))\right|}{|U|}\tag{6}$$

Approximate classification accuracy expresses the percentage of possible decisions that are correct when applying knowledge R to classify an object. Approximate classification quality expresses, on the other hand, the percentage of objects that can be classified exactly into $\pi(U)$ categories by applying knowledge R.

The knowledge representation system can be viewed as a relational data table whose rows correspond to the objects to be studied and whose columns correspond to the attributes of the objects, and whose information is expressed by specifying the values of each attribute of the objects. The formal definition of a knowledge representation system is given below:

Definition 10: $S=(U,A,V,f)$ is a system of knowledge expressions where:

U : a non-empty finite set of objects, called an argument domain.

A: a non-empty finite set of attributes, also denoted as A_i.

V : the value domain of the entire property, $V=\bigcup_{a\in A}V_a$, V_a denotes the value domain of the property $a\in A$.

f : denotes a mapping of $U\times A\to V$, called the information function.

Note: The information function $f=\left\{f_{o_j}\middle|(j=1,2,\ldots,n)\right\}$ denotes that each attribute of each object is assigned an information value, and that the information of an object is represented by the value of each attribute of the specified object. The information function expresses the connection between the set of objects U and the set of attributes A, and is the informational basis for knowledge representation, approximation, and discovery. For example, $f_{a_j}(x_i)=v(j=1,2,\ldots,n)$ expresses the value of the object x_i under the attribute a_j v, where $f_{a_j}:U\to V_{a_j}\,(j=1,2,\ldots,n)$.

2.4.2. Rough set based data approximation algorithm:

1) Data simplification. The huge volume and high dimensionality of actual large database data makes the process of performing data approximation more important in database knowledge discovery. It plays a very crucial role in the performance of the data mining system.

Data approximation includes deletion of columns (attributes), i.e. attribute selection. Deleting rows and merging similar records. Reducing the number of values in a column, including continuous attribute discretization. In the classification problem of KDD, the attributes of the database table can be categorized into conditional attributes and decision attributes. Attribute selection refers to

selecting a better and representative subset of the full set of conditional attributes that consistently describes the data in the table. In practice, the number of attributes of a data record may be very large, while the category of a tuple may depend on only a few attributes. At this point, if the original data records with all the attributes are fed into the classifier, unnecessarily complex rules will be generated. Attribute selection, on the other hand, can reduce the time taken to obtain the rules, making the generated rules more concise and having a higher classification accuracy.

2) Attribute Simplification. Attribute simplification is an important part of data simplification. Through attribute simplification, we can get a simplified set of the original data, which is much less than the original data and can produce more concise knowledge.

There are three main methods for attribute selection:

(1) Exhaustive search, testing starting from one attribute, then combining two attributes until finding the attribute set that breaks through the conflict analysis.

(2) Heuristic search, which tests attribute sets sorted by their likelihood of being a marquee set, with higher-order attributes prioritized.

(3) Randomized search.

Approximation is an important concept of Rough set for data analysis, however the computation of minimum approximation is NP-hard, this use of heuristic information to simplify the computation is necessary.

Attribute approximation algorithms based on Rough sets have a definition that shows that an information table usually has more than one approximation and $core(P) = \cap core(P)$, where $red(P)$ is a subset of all the approximations of P, and $core(P)$ is the kernel, and it can be seen that the kernel is useful in two ways: firstly it can be used as a computational basis for all simplifications because the kernel is necessarily included in all simplifications and the computation can be performed directly. Secondly it is the set of characteristic parts of knowledge that cannot be eliminated in a knowledge reduction.

Define the algorithm for solving kernels:

Input: decision table $T = \langle U, C \cup D \rangle$, where C and D are the set of conditional and decision attributes, respectively.

Output: kernel $core(C, D)$ of the decision table.

Steps:

Step1: Let $R = \varnothing, B = C$ and $B = C$.

Step2: Take any $r \in B$ and calculate $POS_C(D)$, $POS_{C-(r)}(D)$.

Step3: If $POS_C(D) \neq POS_{C-(r)}(D)$, then $R = R \cup \{r\}$ and $B = B - \{r\}$. Otherwise $R = R$, $B = B - \{r\}$.

Step4: If $B \neq \varnothing$, then turn to Step2. otherwise make $core(C, D) = R$ which is the desired. Comment, the time complexity of this algorithm is $0(card(C))$.

2.4.3. Attribute Approximation Algorithm Based on Attribute Sequence and Partitioning Methods. In attribute approximation based on the invariance of the positive region of the decision table, computing the positive region of the decision table is one of the most important computations. For the computation of positive region can use the partition method (the method of fast sorting belongs to the partition method), the average time complexity result of fast sorting for multidimensional tables is $O(|U| \times (|C| + \log|U|))$, with a space complexity of $O(|U|)$. Considering the space complexity, this paper adopts the method of fast sorting to compute the positive region of the decision table, and the attribute approximation algorithm under the attribute order is given below.

Definition 11: It is a sufficient condition that the discriminant matrix element B_{xy}^s in M is not empty:

$$\begin{aligned}&\left(x \in Pos_C(D) \wedge y \in Pos_C(D) \wedge (d(x) \neq d(y))\right.\\&\vee\left(x \in Pos_C(D) \wedge y \notin Pos_C(D)\right)\\&\left.\vee\left(x \notin Pos_C(D) \wedge y \in Pos_C(D)\right)\right.\end{aligned} \tag{7}$$

Definition 12: Let $[k] \in M / L(S)$, then $[k] \neq \phi$ is sufficient for $\exists x, y \in U$ to satisfy the following two conditions:

The discriminant matrix element B_{xy}^s in M is not empty.

$$\forall_{k_1(1 \leq k_1 < k)}\left(c_{k_1}(x) = c_{k_1}(y)\right) \wedge \left(c_k(x) \neq c_k(y)\right) \tag{8}$$

Definition 13: Given a decision table:

$$S = \langle U, A = C \cup D, V, f \rangle \tag{9}$$

Given the attribute order SO: $c_1 \prec c_2 \prec \cdots \prec c_{|C|}$, the attribute $c_k\left(c_k \in C \wedge 2 \leq k \leq |C|\right)$ is a non-empty labeled attribute if, as a sufficient condition, after sequentially removing from the Skowron discrimination matrix the elements of the discrimination matrix containing the attributes $c_1, c_2, \ldots, c_{k-1}$, $\exists B_{xy}^s \in M$ satisfies $c_k \in B_{xy}^s$.

The time complexity and space complexity of the algorithm to compute the positive region is reduced by using fast sort, and the complexity of the algorithm can be reduced if the idea of partitioning method is also added to the attribute approximation process. After obtaining the set of non-empty labeled attributes of the decision table under the attribute order, the attribute approximation algorithm based on attribute order and partition method is given next.

The steps of attribute approximation algorithm based on attribute order and partition method are as follows:

Input: decision table:

$$S = \langle U, A = C \cup D, V, f \rangle \tag{10}$$

Output: attribute approximation Red for decision table S.

Step1: Set:

$$C = \{c_1, c_2, \ldots, c_{|C|}\} \tag{11}$$

Given an attribute order SO: $c_1 \prec c_2 \prec \ldots \prec c_{|C|}$, $r = 1$, $U_1^1 = U$, $\mathrm{Re}\, d = \phi$.

Step2: Compute the positive region $POS_c(D)$ of the decision table using fast sort.

Step3: Call Algorithm 5.2 to compute the set of non-empty labeled attributes R_1 of the decision table.

Step4: Let $c_{N'}$ be the labeled attribute with the largest label in R_t. If $c_{N'} \in Red$, Then turn to Step5. otherwise $Red = Red \cup \{c_{N'}\}$ and $c_{N'}$ is placed in the last position of Red. $R_1 = R_1 - \{c_{N'}\}$ and $C' = \phi$. To merge the attribute sets Red and R_1, first add the attributes in Red to C' in descending order of label attribute subscripts. Then the attributes in R_1 are added to C'' in ascending order of label attribute subscripts. $C = \phi$, $C = C'$. Call Calculate the set of non-empty labeled attributes R_1 of the decision table under the new attribute order, go to Step4.

Step5: Jump to Red.

Complexity analysis of attribute approximation algorithms based on attribute ordering and partitioning methods:

In the algorithm, Step2 has an average time complexity of $O\left(|U| \times \left(|C| + \log|U|\right)\right)$ and a space complexity of $O\left(|U|\right)$. Step3 has an average time complexity of $O\left(|U| \times \left(|C| + \log|U|\right)\right)$ and a space complexity of $O\left(|U| + |C|\right)$. The average time complexity of Step4 is $O(U| \times |C| \times (|C| + \log|U|))$ and the space complexity is $O\left(|U| + |C|\right)$. Therefore, the average time complexity of the algorithm is:

$$O\left(|U|\times|C|\times\left(|C|+\log|U|\right)\right) \tag{12}$$

The space complexity is:

$$O\left(|U|+|C|\right) \tag{13}$$

3. Algorithmic feasibility and realization of a knowledge-reducing platform

3.1. Experiments with Knowledge Approximation Algorithms

All the programs in this paper are written in Python and the hardware environment for the experiments is Inter(R) Core(TM) i7-4790 CPU @ 3.60 GHz, computer running memory 8G, and operating system Windows 11. 10 datasets from the UCI dataset are selected for the experiments, as described in Table 1.

Table 1. Experimental data set

Serial number	Data set name	Sample size	Attribute number	Classification number	Data type
1	Heart	300	12	2	Classification, real number
2	Wdbc	580	30	3	Real type
3	Balance	640	6	2	Classification type
4	Australian	700	17	3	Classification, real number
5	Germen	1000	20	2	Boolean, classification, actual type
6	Car	1536	10	3	Classification type
7	Segment	2296	20	8	Classification, real number
8	Gender	4500	22	4	Boolean, classification, actual type
9	Mushroom	5809	25	3	Classification type
10	Callt2	11000	6	2	Classification, real number

Ten datasets on UCI were selected to test the efficient knowledge approximation method based on partitioning method for Rough sets and validate its effectiveness with a change in the sample set.

In the datasets, Wdbc is a real-type dataset, Balance and Mushroom are sub-type datasets, and Heart, Australian, German, Segment, Gender, Car and Callt2 are all mixed-type datasets. Each dataset was divided into two parts, the original dataset, and the incremental dataset.

The original dataset consists of 50% of the number of samples, and the remaining 50% of the number of samples, make up the incremental dataset.

Comparison and analysis with static attribute approximation algorithm (SFDA), fuzzy rough set-based incremental approximation algorithm (IV-FS-FRS), and Gaussian kernel approximation-based incremental approximation algorithm (DIDS) were conducted in terms of the number of approximation results, classification accuracy, and runtime length, respectively. For the classification accuracy of the approximation results, SVM and KNN classifiers will be used for evaluation in this paper. 1) Comparison of the number of attributes before and after simplification. The comparison of the number of attributes before and after simplification in four different algorithms is shown in Figure 1. From the figure, it can be analyzed that all the different approximation algorithms can ensure that all the datasets can eliminate the irrelevant attributes from the original attributes. The efficient knowledge approximation method for Rough set based on partition method proposed in this paper has the highest number of approximations for all datasets, and for the Heart dataset 20 original

number of attributes, the value of attributes after processing by this paper's method is 10.

It outperforms Algorithm IV-FS-FRS and DIDS in most of the cases.In Wdbc dataset, Algorithm IV-FS-FRS outperforms DIDS mainly because Wdbc dataset is a real-valued dataset and IV-FS-FRS is good at dealing with real-valued dataset, so the algorithm performs better.

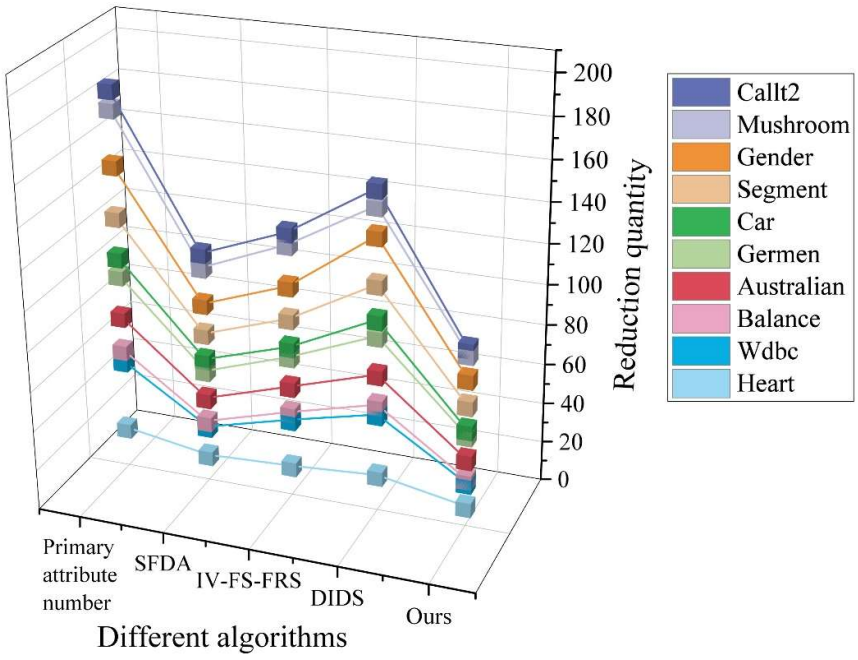

Fig. 1. Comparison of the number of different algorithms

2) Comparison of classification accuracy under the classifier before and after simplification. In order to test whether the attribute simplification algorithm proposed in this paper is accurate and effective, the next step is to compare the classification accuracy of the original data with the simplified data. The classifier is selected to be a combination of SVM and KNN, and ten-fold cross-validation is used, and the experimental results are based on the average classification accuracy of ten times.

The comparison of classification accuracy of different algorithms based on SVM is shown in Figure 2. As can be seen from the trend in the figure, the classification accuracy of different algorithms is different for different datasets with different original data. The classification accuracies of this paper's method based on SVM classifier after ten datasets are 0.8123, 0.956, 0.6689, 0.8652, 0.89, 0.9124, 0.9605, 0.8825, 0.5211, 0.6631, respectively.

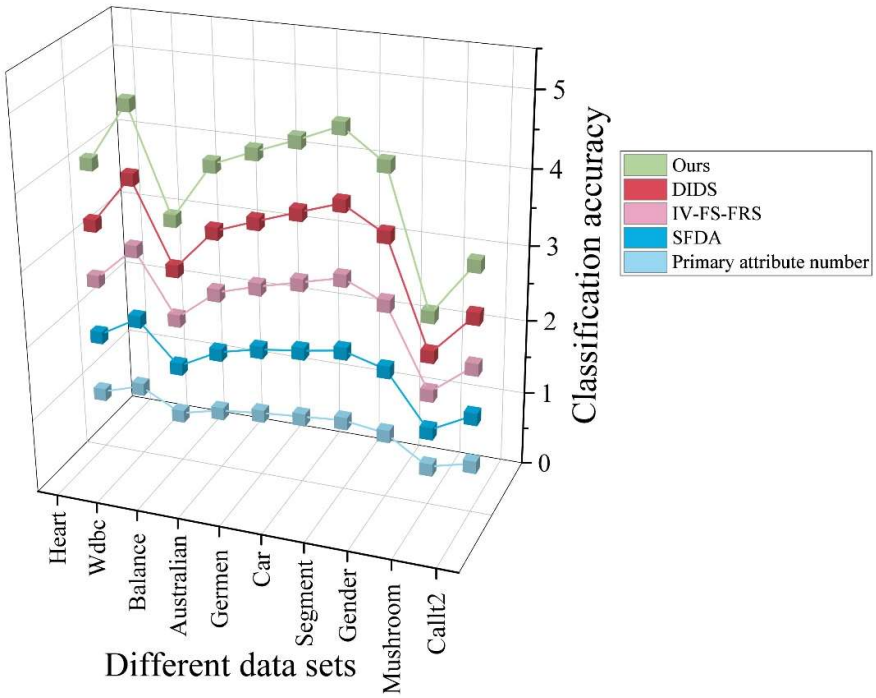

Fig. 2. Classification accuracy of different algorithms based on SVM

The comparison of classification accuracy of different algorithms based on KNN is shown in Fig. 3.

In the Heart dataset, the classification accuracy of the DIDS method is 0.5936, which is lower than that of the method of this paper (0.6689).The method of this paper has a higher classification accuracy of 0.1153 than that of the IV-FS-FRS method (0.5536).

In the Balance dataset, the classification accuracies of algorithm SFDA, algorithm IV-FS-FRS, algorithm DIDS, and this paper's method are 0.5111, 0.5123, 0.5181, and 0.5236, respectively, and the gap between the classification accuracies of the algorithms is small.

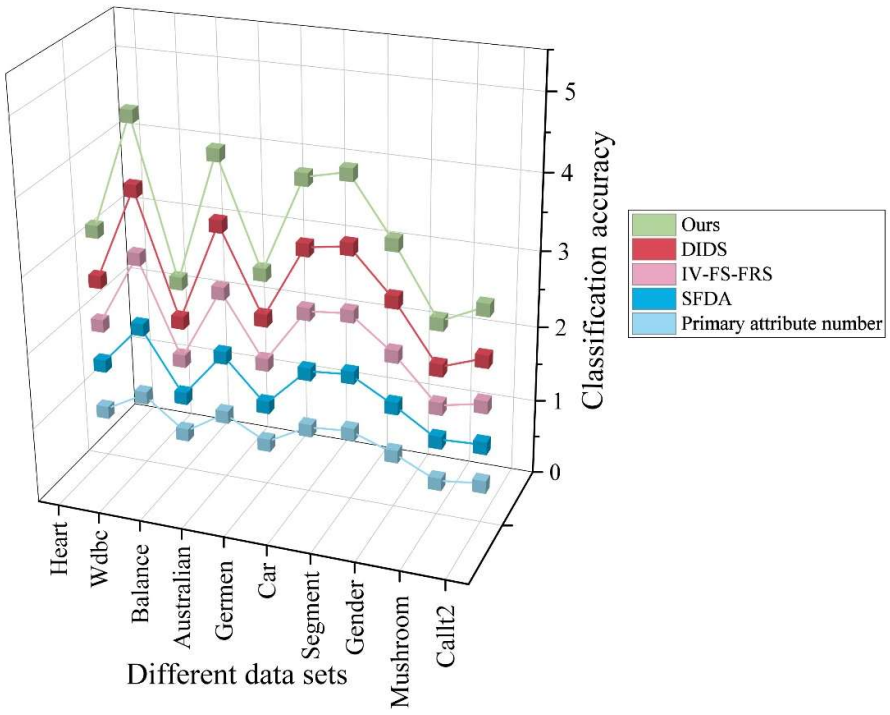

Fig. 3. The classification accuracy of different algorithms based on KNN

3) Accuracy analysis of experimental results. In this experiment, 10 datasets of different sizes are selected from the open source UCI dataset for comparison experiments, and the dataset is doubled to get a larger dataset and to check the effectiveness of the parallel algorithm. The experimental datasets are shown in Table 2.

Table 2. Experimental data set

Data Set	|U|	|C|
Heart	300	2
Wdbc	580	3
Balance	640	2
Australian	700	3
Germen	1000	2
Zoo	120	20
Cancer	713	10
Car	1520	8
Chess	3524	32
Nursery	13560	10

Due to the high complexity of the discriminant matrix knowledge approximation algorithm, this experiment first selects a smaller 10 UCI dataset to compare the number of kernels and the number of knowledge approximations computed by the serial algorithm and the parallel algorithm. |COR| and |R| denote the number of attribute kernels and the number of knowledge parsimony computed serially, respectively. |PCOR| and |PR| denote the number of attribute kernels and knowledge parsimony computed in parallel, respectively. The kernel versus approximation example is shown in Fig. 4.

From the figure, it can be seen that for the same dataset, the computational results of the parallel algorithm are unchanged and are the same as those of the serial algorithm. This shows that the efficient knowledge approximation method for Rough sets based on partition method proposed in this paper has no effect on the computational results.

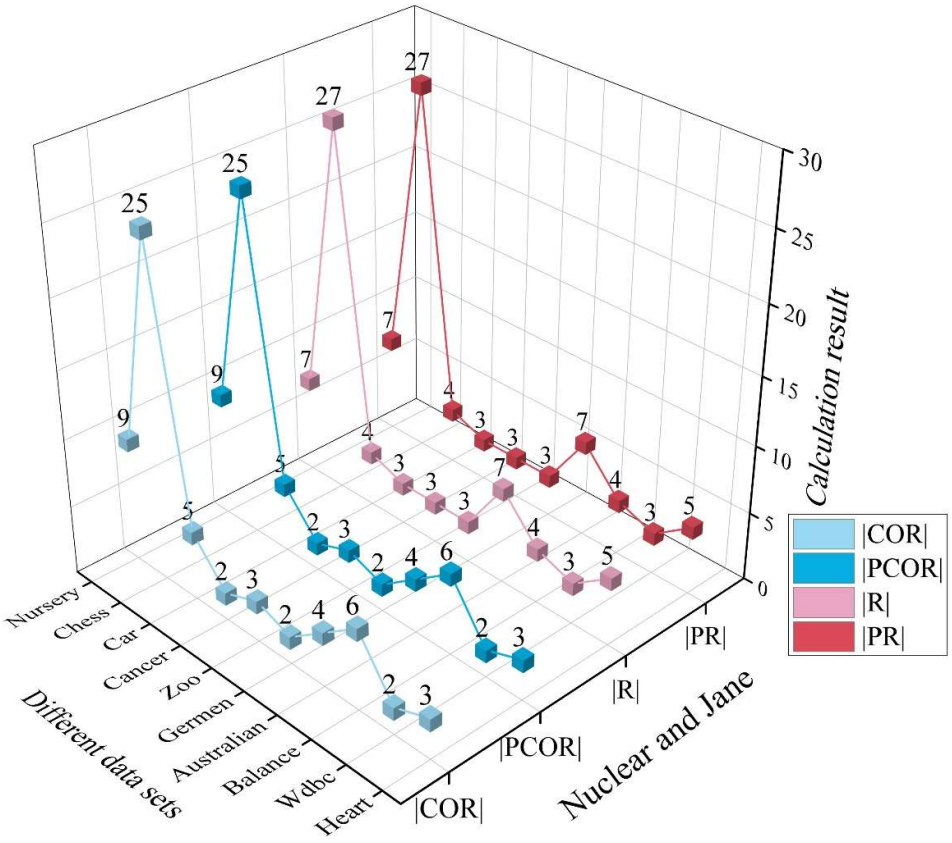

Fig. 4. The correlation between the kernel and the reduction

To do accuracy comparison on the dataset, R1 denotes the blind deletion knowledge approximation algorithm, R2 denotes the attribute importance based knowledge approximation algorithm, R3 denotes the discriminant matrix knowledge approximation algorithm, and PR3 denotes the efficient knowledge approximation method based on the partition method for the Rough set.

The classification accuracies are shown in Fig. 5, which shows that the experimental accuracies of R3 and PR3 for the datasets Zoo, Cancer, Car, Chess, and Nursery are the same, which are 0.7828, 0.8596, 0.7936, 0.9235, and 0.8364, respectively.The classification accuracies of parallel algorithms are the same as the serial algorithms, which are unchanged, and the classification accuracies of the partition-based method of Rough set efficient knowledge approximation method can ensure the accuracy remains unchanged.

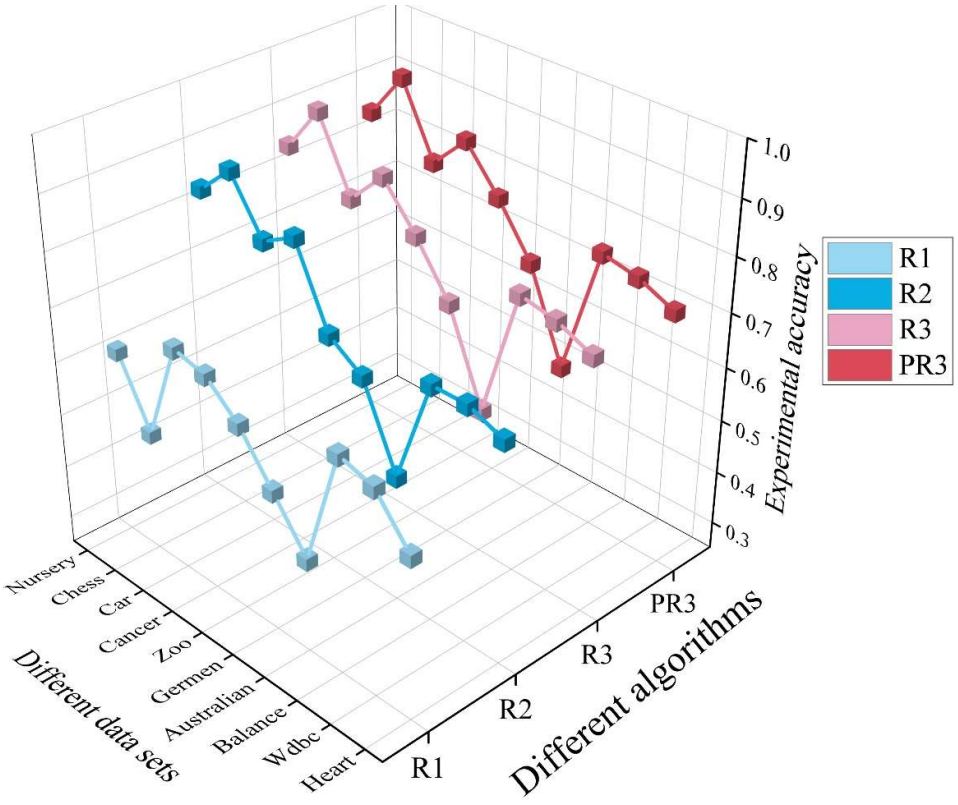

Fig. 5. Classification accuracy

4) Efficiency analysis of experimental results. Running time is an intuitive evaluation of the parallel effect, and this part of the experiment doubles the UCI dataset to get a larger dataset.

The comparison of experimental running time is shown in Table 3.Red denotes the serial running time and PRed denotes the running time of the efficient knowledge approximation method based on the partition method for Rough set.The parallel algorithm uses 5 computing nodes for different datasets, and the running time is as statistic in the table. The experiments show that when using the Rough set efficient knowledge approximation method based on partitioning method to do knowledge approximation, it can improve the efficiency and even solve the problems that cannot be solved by a single machine, such as memory overflow and downtime, while keeping the approximation results unchanged.

Table 3. Experiment run time comparison

Data Set	|U|	|C|	Red(s)	PRed(s)
Nursery	13560	10	35.048	20.117
Nursery*2	27120	10	68.227	25.359
Nursery*4	54240	10	141.364	56.004
Nursery*6	81360	10	212.496	72.658
Nursery*8	108480	10	278.384	90.154
Nursery*12	162720	10	420.255	121.663
Nursery*20	271200	10	701.390	210.207

Twenty percent of the data from the datasets Car, Segment, Gender, Mushroom, and Callt2 were selected as the base dataset, and the remaining 80% of the dataset was divided into four equal-length datasets, Part I, Part II, Part III, and Part IV. The first part is treated as the first incremental dataset, and the combination of the first and second parts is treated as the second incremental dataset. Stacked in this way, the whole combination of the four parts is treated as the fourth incremental dataset. The

number of conditional attributes is denoted by |C|, r denotes the number of attributes remaining after the approximation, R denotes the remaining attributes after the approximation, and t denotes the algorithm execution time. A comparison of the approximation results obtained from the algorithm run is shown in Table 4.

R1 denotes the blind deletion knowledge approximation algorithm, R2 denotes the attribute importance based knowledge approximation algorithm, R3 denotes the discriminant matrix knowledge approximation algorithm, and PR3 denotes the efficient knowledge approximation method based on the partition method for Rough sets. It can be seen that this paper's method is better in the quality of approximation, the number of attributes remaining after approximation r is 4, 7, 5, 8, 2 in order, and the number of attributes remaining after approximation is the least.

The overall analysis shows that the method of this paper is better than algorithms R1 and R2 in terms of the quality of simplification, and the main reason for the difference between the quality of simplification of this paper and that of algorithm R1 is the difference in the method of gradually selecting the remaining attributes to be added to the set of simplification using the kernel as an expansion.

Table 4. Comparison of the approximate results of different algorithms

Data set	n	|C|	Ours		R1		R2	
			r	R	r	R	r	R
Car	1536	10	4	1,2,7,9	6	1,2,3,5,9,10	8	1~4,6,8,9,10
Segment	2296	20	7	1,3,4,8,9,11,15	8	1~5,8,12,17	10	1~3,6~12,18
Gender	4500	22	5	1,8,12,15,21	12	1,2~7,9~11,12,16,22	15	1~5,6,9,12,15~22
Mushroom	5809	25	8	1,2,7,8,9,10,12,18	10	1,3,8,9,12,15~18,21,22	11	1~3,5,9~14,16,23
Callt2	11000	6	2	1.3	4	1,3,4,5	5	1,2,3,4,6

The time comparison of the algorithm running for adding different datasets is shown in Table 5.

It can be seen that the running time of this method is significantly better than that of Algorithm R1 and Algorithm R2.The main reason that the running time of this method is better than that of Algorithm R1 is that the method in this paper gets the picking attribute by deleting the corresponding rows in the CBM and determining whether the CBM is empty. While Algorithm R2 obtains the picking attribute and then determines whether the positive domains are equal by calculating the positive domains, which takes a relatively long time to calculate the positive domains.

This paper's method compared to the algorithm R1 has one more branch to deal with group data, from the running time can be seen in this paper's method in response to the increase of a group of data do not have to add one by one to deal with the increase of data, so it has a very good effect.

For the data set CAR three algorithms running time is comparable, while for the data set MUSHROOM, etc. this paper's method and Algorithm R2 has a huge difference, the main reason for this is that Algorithm R2 in the selection of attributes of the quality of the poorer, so that the number of attributes selected more, which leads to the calculation of the number of positive domains is more times, which leads to the whole algorithm of the running time is longer.

Algorithm R1 needs to use calculate the positive domain in reverse elimination, and Algorithm R2 needs to calculate the positive domain in judgment after picking the attributes, thus leading to the running time of the whole algorithm is significantly larger in the Gender dataset than in the mushroom dataset. In contrast, the method in this paper has a processing group data branch, so only one direction elimination is performed when processing an increased group of data, i.e., the number of times of calculating the positive domain is less, so the number of decision attribute values does not affect the running time of the method in this paper. The running time of this method is longer in mushroom than in Gender, mainly because the number of instances and attributes in mushroom is larger than that in Gender.

Table 5. The time comparison of the algorithm runtime for increasing different data sets

Data set	Algorithm	First batch	Second batch	Third batch	Fourth batch
		t/s	t/s	t/s	t/s
Car	Ours	2.36	7.11	12.36	15.63
	R1	21.63	26.38	121.32	133.45
	R2	36.25	92.13	152.08	210.36
Segment	Ours	0.98	3.21	5.63	14.57
	R1	62.53	223.58	645.88	1502.36
	R2	526.36	1800.63	425.36	21536.2
Gender	Ours	3.00	12.26	20.00	32.04
	R1	425.26	1053.26	2354.87	3365.01
	R2	5526.3	15362.2	26369.21	49901.00
Mushroom	Ours	5.36	15.70	24.90	35.05
	R1	72.63	160.35	272.00	366.00
	R2	389.26	452.89	893.23	11243.05
Callt2	Ours	7.55	20.63	62.24	121.36
	R1	78.93	130.42	270.82	352.14
	R2	210.36	332.6	668.55	942.09

3.2. Design of the Resource Acquisition System for College English Teaching

Based on the research of the above algorithms, this paper designs a platform for knowledge acquisition system of university English teaching resources. The platform not only integrates the algorithms proposed in this paper, but also implements some typical algorithms proposed in the paper, which provides a convenient way for the subsequent experimental analysis and lays the foundation for the subsequent work.

3.2.1. System design objectives. This paper mainly studies the problem of efficient knowledge reduction and rule acquisition of Rough set based on the partition method, and proposes the attribute reduction algorithm and rule acquisition algorithm. This paper designs a knowledge acquisition system platform based on the partition method for university English teaching resources, which integrates the attribute simplification and rule acquisition algorithms proposed in this paper, and the design objectives of the system are as follows:

1) The system interface is simple, beautiful and easy to operate.

2) Users can independently select the data sets to be involved in the operation, and need to realize the browsing and reading of the data files in the local folder, and can display the data set information in the window.

3) Users can change the precision parameter β and the coverage parameter α as needed.

4) The user can change the algorithm used to perform attribute simplification and rule acquisition on the selected dataset as needed, and save the attribute simplification and rule extraction results to a local file and display them in the window.

5) Users can manually input an example, change the precision and coverage parameters, and obtain the attribute simplification and rule extraction results.

3.2.2. System development environment. Hardware development environment:

Lenovo PC, CPU: 2.5GHz, RAM: 4GB

Software development environment:

Windows 10, MATLAB2019a

3.2.3. Example analysis. In order to facilitate the understanding, a knowledge acquisition example is used to illustrate the use of the knowledge acquisition system platform for university English teaching resources based on the efficient knowledge reduction method of the Rough set of the partition method.
1) Select a data set: Click “University English Teaching Resources Data Files” to enter the local folder directory interface. Select the desired dataset “Iris”.
2) Parameter setting: Set the variable precision parameter β to 0.7 and the coverage parameter to 0.8.
3) Algorithm selection: Select the efficient knowledge reduction method for Rough set based on partition method proposed in this paper.
4) Display results: Click “Run”, the results of the algorithm will be displayed in the display box.
5) Export rules: Click “Export Rules” and select the directory in the folder where the rules are saved. The exported rule set is an Excel file, which is displayed as the result of the knowledge related to university English teaching resources.

4. Conclusion

This paper draws on the idea of partition algorithm to establish a Roguh set efficient knowledge approximation method based on partition method for integrating university English teaching resources, analyzes the experimental accuracy and running time of the proposed algorithm, and builds a platform for knowledge acquisition system of university English teaching resources. Multiple datasets are selected from the UCI dataset, and the performance of the Roguh set efficient knowledge approximation method based on the partition method is compared with different algorithms. Compared with the number of simplification results, classification accuracy, experimental efficiency and running time of multiple algorithms, the Roguh set efficient knowledge simplification method based on the partition method can realize more attribute simplification, and achieve the highest classification accuracy in both SVM and KNN based classifiers. And the classification accuracy is unchanged. The Roguh set efficient knowledge reduction method based on the partition method to build a knowledge acquisition system platform for university English teaching resources can quickly realize the integration of university English teaching resources.

Funding

Research on the Construction and Practice of a "Four-Dimensional Synergy" Teaching New Ecosystem for College English Empowered by Digital Intelligence (XXYJGE2439).

References

[1] Wright, M. C., Bergom, I., & Bartholomew, T. (2019). Decreased class size, increased active learning? Intended and enacted teaching strategies in smaller classes. Active learning in higher education, 20(1), 51-62.

[2] Duan, R., & Qiu, P. (2022). Survey on the Current Situation of College English Teaching in China’s Universities and the Direction of College English Teaching Reform and Development. Mobile Information Systems, 2022(1), 5073924.

[3] Zhiyong, S., Muthukrishnan, P., & Sidhu, G. K. (2020). College English Language Teaching Reform and Key Factors Determining EFL Teachers' Professional Development. European Journal of Educational Research, 9(4), 1393-1404.

[4] Kuang, Q. (2024). A Study of the Production-oriented Approach-based Application of Micro Lecture in a College Oral English Course in the Blended Teaching and Learning Environment. The Educational Review, USA, 8(1).

[5] Bing, W. (2017). The College English Teaching Reform Based on MOOC. English Language Teaching, 10(2), 19-22.

[6] Zhang, F. (2017). Quality-improving strategies of college English teaching based on microlesson and flipped classroom. English Language Teaching, 10(5), 243-249.

[7] Du, Y. (2020). Study on Cultivating College Students' English Autonomous Learning Ability under the Flipped Classroom Model. English Language Teaching, 13(6), 13-19.

[8] Zou, D. (2017). Research on college English teaching model based on multimedia and network. DEStech Transactions on Social Science Education and Human Science, (icsste), 19-21.

[9] Zhang, L. (2019). Development of an information-based online foreign language teaching platform with ASP. NET. International Journal of Emerging Technologies in Learning (Online), 14(13), 117.

[10] Zhang, Y. (2021, September). Research on Innovative Approaches to College English Teaching Based on Information Technology. In 2021 4th International Conference on Information Systems and Computer Aided Education (pp. 1500-1503).

[11] Wang, J., & Li, W. (2021). The construction of a digital resource library of English for higher education based on a cloud platform. Scientific programming, 2021(1), 4591780.

[12] Shen, G. R. (2018). Chinese college English teachers' ability to develop students' informationized learning in the era of big data: status and suggestions. EURASIA Journal of Mathematics, Science and Technology Education, 14(6), 2719-2728.

[13] Li, Y., Zhou, C., Wu, D., & Chen, M. (2023). Evaluation of teachers' information literacy based on information of behavioral data in online learning and teaching platforms: an empirical study of China. Library Hi Tech, 41(4), 1039-1062.

[14] Kassymova, G. M., Tulepova, S. B., & Bekturova, M. B. (2023). Perceptions of digital competence in learning and teaching English in the context of online education. Contemporary Educational Technology, 15(1), ep396.

[15] Rahimi, A. R. (2024). Beyond digital competence and language teaching skills: The bi-level factors associated with EFL teachers' 21st-century digital competence to cultivate 21st-century digital skills. Education and Information Technologies, 29(8), 9061-9089.

[16] Jiang, D., Pei, Y., Yang, G., & Wang, X. (2022). Research and analysis on the integration of artificial intelligence in college English teaching. Mathematical Problems in Engineering, 2022(1), 3997573.

[17] Yang, Y. (2022). Integration and utilization strategy of university english teaching information resources based on fuzzy clustering algorithm. Mobile Information Systems, 2022(1), 1321660.

[18] Wang, Z. (2022). Integration and sharing of college English teaching resources using cloud computing platform. Mobile Information Systems, 2022(1), 8202229.

[19] Li, R. (2024). A data mining-based approach to integrating multimedia English teaching resources. International Journal of Computational Systems Engineering, 8(1-2), 1-9.

[20] Xin Han & Kun Jiao. (2024). Study on multimedia network aided English teaching resource integration system based on cloud storage. International Journal of Continuing Engineering Education and Life-Long Learning,34(1),39-52.

[21] Han Qiao. (2024). Personalized Recommendations of Online English Teaching Resources for Higher Vocational Education. Applied Mathematics and Nonlinear Sciences,9(1).

[22] Kelu Wang & Dexu Bi. (2024). Integrating MOOC online and offline English teaching resources based on convolutional neural network. International Journal of Business Intelligence and Data Mining,25(3-

4),271-291.