Journal of Information Technology Review

DLINE JOURNALS

Print ISSN: 0976-3511
Online ISSN: 0976-2922
JITR 2025: 16 (2)
https://doi.org/10.6025/jitr/2025/16/2/55-62

The Learning Platform System of Marxist Theory and Education based on Association Rule Learning Algorithm

Xiaojuan Chen
School of Marxism, Chongqing Vocational College of Transportation
Chongqing, 402247, China
cxj3408@163.com

ABSTRACT

The integration of traditional advantages, such as promoting Marxist theory and educational courses, with information technology is necessary for academic development. Achieving true "teaching based on individual differences" requires the scientific application of educational big data. This paper utilizes association rule mining technology to address the demand for student task point learning management in the current personalized learning platform system. It analyzes the current status and research trends of personalized learning and association rules. Then, it discusses the principles, advantages, and disadvantages of the classic Apriori algorithm and proposes an improved algorithm. Finally, the feasibility and actual effects of the algorithm are verified through simulations on the learning platform system, providing certain assistance for future research on Marxist theory and educational courses.

Keywords: Association Rule Learning Algorithm, Data Mining, Higher Education, Marxist Theory, Learning Platform System

Received: 7 November 2024, Revised 19 January 2025, Accepted 30 January 2025

1. Introduction

With the widespread adoption of big data, we are entering a critical period of transformation. Therefore, educators in universities must actively embrace this transformation and combine it with existing traditional advantages to better meet the needs of current social development. Based on this, a more creative approach to big data thinking that is more conducive to social development should be developed. This can help us gain a better understanding of the essence of education and carry out related preaching and training more effectively. With the advancement of science and technology, data mining has become a vital tool that enables us to uncover valuable information hidden within massive amounts of data. It uses advanced mathematical algorithms and computer science techniques to support our decision-making. However, due to the vast amount of information, traditional data mining techniques still have shortcomings. The Apriori algorithm is considered a vital tool that helps us identify relationships between different events, enabling better anticipation of future developments. The core idea of this method is to establish a framework of support, confidence, and interest and use it to identify different rules, thereby better analyzing and solving problems. The Apriori algorithm takes support as the threshold and adopts a multi-level iterative method to effectively detect and analyze the information content in educational data sources. It decomposes the information content into multiple subsets through pruning and other cleaning operations for quick discovery and analysis. With the continuous advancement of science and technology, new calculations have been developed to significantly reduce the search range and shorten the search process, thereby substantially improving search efficiency. The development of association rules has become a hot topic in data analysis today and has deeply integrated into human daily life. By utilizing association rule mining technology, we can extract valuable insights from a large dataset and gain a deeper understanding of future trends, such as students' learning backgrounds and achievements upon graduation. This information can be applied to college graduation employment planning and management to achieve better results.

2. Related Work

Agrawal and other researchers initially regarded association rule mining as a significant technology aimed at gaining in-depth insights into consumer needs by exploring the relationships between different types of products. This enables companies to better manage and monitor purchasing efficiency, resulting in more efficient sales [4]. In recent years, an increasing number of algorithms have been developed to enhance the efficiency, flexibility, real-time performance, and universality of rule mining algorithms. These include hierarchical mining, incremental updates, distributed and parallel mining, multi-level, multi-value, concept lattice mining, and more [5]. According to the U.S. National Educational Technology Plan of 2017, "Reimagining the Role of Technology in Education," the proposal suggested that mobile data collection tools and network collaborative platforms could bring personalized learning approaches to students, thereby providing them with more opportunities. Furthermore, the book highlighted that personalized learning will be one of the core areas of effective leadership in future development, which can help students achieve better learning outcomes by collecting and integrating formative and summative assessment statistics during individualized digital teaching processes [6]. With the introduction of the "New Generation Artificial Intelligence Development Plan," China is vigorously developing an education system based on intelligent learning. Through teaching interventions and decisions, it aims to promote innovation and development in teaching methods, advancing the sustainable and healthy growth of higher education [7]. With the advancement of artificial intelligence technology, collecting and utilizing data analysis from different sources have become necessary steps for personalized learning, and maximizing the value of these data analyses is crucial for achieving this goal. Currently, research on personalized learning has evolved from traditional dimensions, such as personal information collection, preferences, and interests, to more comprehensive levels [8], and continues to explore and improve. By using artificial intelligence systems in schools and combining them with the latest research in the education field, such as big data analysis and deep neural networks, it is possible to track students' behavior better *g* and accurately predict their academic performance. As times evolve, people are paying more attention to precise guidance, no longer limiting themselves to imparting content. Therefore, refined education has become a new development trend, encompassing not only management but also a broader range of social factors. It emphasizes practice, social development, progress, stability, and overall societal improvement. By combining the characteristics of association rule algorithms, we can better apply them to educational courses at universities and more effectively utilize knowledge in mathematics and computer science [9]. Furthermore, we can enhance academic outcomes through data mining, such as by presenting information more effectively and conducting more accurate behavioral predictions. By combining traditional education with the latest big data technology, we can make more accurate predictions, comprehensively understand the needs of different periods, achieve overall educational effectiveness, and accurately grasp the relationship between them, ultimately conducting practical operations more effectively [10]. Before conducting educational data analysis, several necessary preparations must be made, including multiple mathematical pre-processing steps, selecting appropriate analysis algorithms, visualizing the results of data analysis, and making effective predictions and decisions.

3. Association Rule Mining Technology

3.1 Generation of Association Rule Learning Algorithms

In a frequent itemset, if itemset *I* contain multiple items, such as a tennis racket, tennis ball, sports shoes, and shuttlecock, the rule can be used to describe the interactions between these items. We can use a more complex rule to describe such interactions; for example, in an itemset, if the XY of an item is 3, then the D of this rule is 6, and its XY will result in a higher confidence. If α=0.5, β=0.6, we can assert that the selection of a tennis racket and the selection of tennis service have a close mutual influence.

$$\text{Confidence}(X \Rightarrow Y) = P(X \mid Y) = \frac{\text{support}(X \cup Y)}{\text{support}(X)} \times 100\% \tag{1}$$

By applying the above methods, we can formulate a series of corresponding association rules:

Step 1: Select appropriate frequencies from the availability levels of D.

Step 2: Use the similarity between these frequencies and availability levels to construct corresponding relationships.

3.2 Apriori Algorithm

The Apriori algorithm is considered the first association rule mining algorithm and is regarded as a very effective method. This algorithm can extract valuable content from a large number of association rules and effectively mine the associations between these rules. To achieve this goal, the Apriori algorithm follows two main steps: firstly, extracting valuable content from the candidate set, and secondly, conducting an in-depth evaluation of this content to determine the effectiveness of the association rules. The Apriori algorithm is typically seen as a one-dimensional, multi-level classification algorithm that uses multi-level iterations to achieve effective recognition of multiple sub-objectives. As shown in Figure 1, it is named Apriori. The core idea of this algorithm lies in discovering the similarity between various sub-objectives and combining them to achieve effective recognition of multiple sub-objectives.

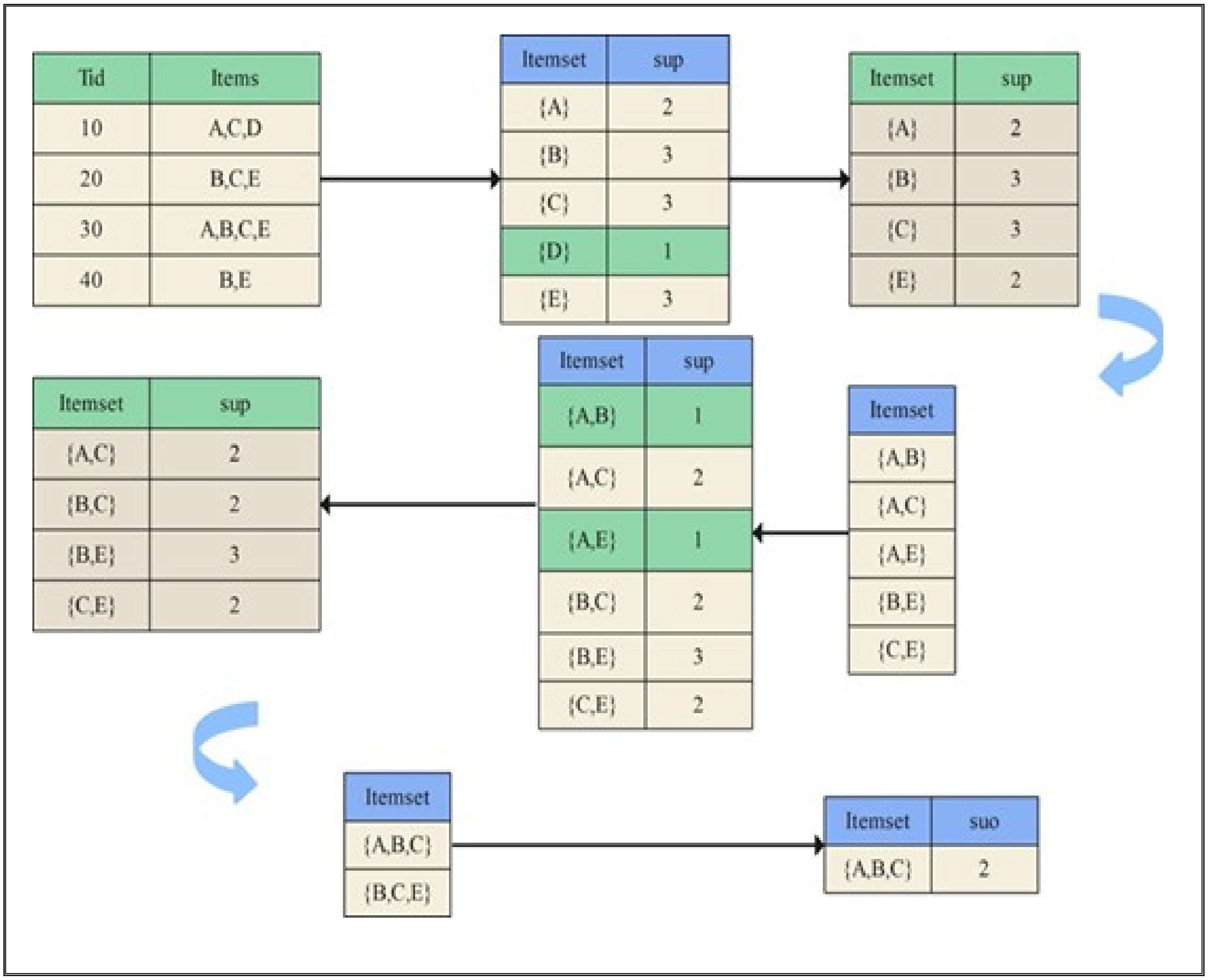

Figure 1. Apriori Algorithm Example

In some cases, we can search for frequent itemsets through database scanning. First, we can find *L*; then, we use it to find two Ls, and finally, we continue the search until we find *k* frequent itemsets. This way, we can determine that the support of each *Lx* exceeds the minimum threshold. The Apriori algorithm has been widely applied due to its simple and understandable operations. Its association rules are based on multiple repetitions of item combinations, ensuring that their support meets the expected standards and has generality and credibility. Although this approach has many advantages, such as significantly reducing the number of database system scans, reducing the formation of intermediate itemsets, improving the reliability of individual items, extending the algorithm's scalability, and enhancing its operability, it still faces challenges such as heavy I/O load, inability to meet large-scale data processing requirements, and lack of high reliability.

3.3 Improvement Methods

Several optimization techniques have been developed for the Apriori algorithm. For example, dynamic counting of itemsets involves checking, evaluating, and filtering the candidate sets of each itemset to determine which itemsets meet the requirements and which ones do not, thus achieving fast and accurate evaluation of Apriori. The specific approach is as follows:

(1) Statistically determine the minimum support itemsets among those containing a single element to construct a one-dimensional frequent itemset *Lt*.

(2) Repeat this process until no more complex frequency itemsets are generated. For example, in step *k*, generate *k*-dimensional candidate itemsets based on *k-1* frequent itemsets generated in step *k-1*. By performing *p* partial classifications, we can obtain a classifier containing *p* rules. When the test set and the training set data are identical, the misidentification rate can be calculated as follows: [Note: The content after ” is missing and needs to be provided for a complete translation.

$$\text{error}(r_i) = 1 - \text{Conf}(r_i) \quad (2)$$

The Apriori algorithm can help us determine which k-dimensional frequent itemsets have been accurately collected, thus avoiding duplicate searches and improving search efficiency. It allows us to filter out those k-dimensional frequent itemsets that do not meet the requirements, thereby enhancing search efficiency. The number of instances covered by rule *ri* is *ri.Sup* × |*Di*| / *ri.Conf*, and the number of misidentified instances is (1 - *ri.Conf*) × *ri.Sup* × |*Di*| / *ri.Conf*. Therefore, the overall classifier misidentification rate is: [Note: The content after “ Therefore, the overall misclassification rate of the classifier is missing and needs to be provided for a complete translation.]

$$\text{error} = \sum_{i=1}^{p}\left[\left(1/\text{conf}(r_i) - 1\right) \times \sup(r_i) \times |D_i|\right] \quad (3)$$

(4) When scanning each transaction TID in database *D*, if at least one candidate itemset *CK* is found, keep that record; If not, exchange the record with the undeleted record at the end of the database and mark it as deleted. After this process, database *D* will become more complete and have more available information.

4. Design of Experiments and Result Analysis

4.1 Design of Experiments and Data Collection

This article uses a student learning behavior record database of educational courses for 19 level computer science students in vocational and technical colleges to extract possible correlations between chapters, and then obtains functional correlations between knowledge points and exam scores. Finally, association rule mining algorithms are used to find the relationship between students’ knowledge points and exam scores. There are several main processes. (1) Collect students’ learning records and convert them into a searchable database of learning outcomes. (2) Extract data mining objects from the learning results database, compile code, and convert the relationship list into a business processing database. (3) For business databases, frequent item sets are generated based on a given minimum support level, and then combined with a given minimum confidence level to generate association rules. Data mining is based on the objects provided in the code table and the simplified relationship table in the learning log database, which includes the number of students, school hours, teaching situation, exam scores, and learning method data. Convert the relationship table into a corresponding transaction library and write the code. For example, selecting a portion of students’ “Database Principles” as a transactional database, and using a coding table to convert corresponding items to chapters with good grades. After generating transactional database D, | D |=9, K={K01, K02, K03, K04, K05, K06, K07, K08, K09}. Assuming a given minimum support of 0.25, use the Apriori algorithm to obtain all frequency term sets in *D*. After generating a frequency itemset, for any *K* types of frequent itemsets, search for all possible true subsets and calculate the corresponding rule reliability. When it is greater than the given minimum value, the rule is output until it is finally connected to the *k-1* itemized set.

4.2 Result Analysis

As shown in Figure 2, when the minimum confidence threshold is set to 0.75, the rule is K01 Ë K08 '! K03. Specifically, when studying educational courses, if students achieve good grades in the "Introduction" and "Database Programming" chapters, their grades in the "Relational Database Standard Language SQL" chapter will also improve. Therefore, in personalized learning systems, it can be considered to place "Database Programming" before "Relational Database Standard Language SQL", Alternatively, when students choose to learn the "Standard Language for Relational Database SQL", the personalized learning system may suggest that they first review and consolidate the content of the "Introduction" and "Database Programming" chapters. When the unoin value gradually increases to 100%, the confidence threshold will first increase and then decrease. For example, when students with "academic" Learning styles learn the chapter of "relational data theory", they use more theoretical text materials for learning, and most students have excellent academic performance. On the contrary, students with "operational" Learning styles, when learning the chapter on "SQL, the standard language of relational databases", tend to use video animations more for practical learning, and most of these students also exhibit excellent academic performance. Therefore, when students with an "academic" learning style choose learning content, the personalized learning system will prioritize recommending theoretical, text-based learning materials for them. When students with "operational" Learning styles choose learning content, the system prioritizes presenting video animation-based learning materials to them.

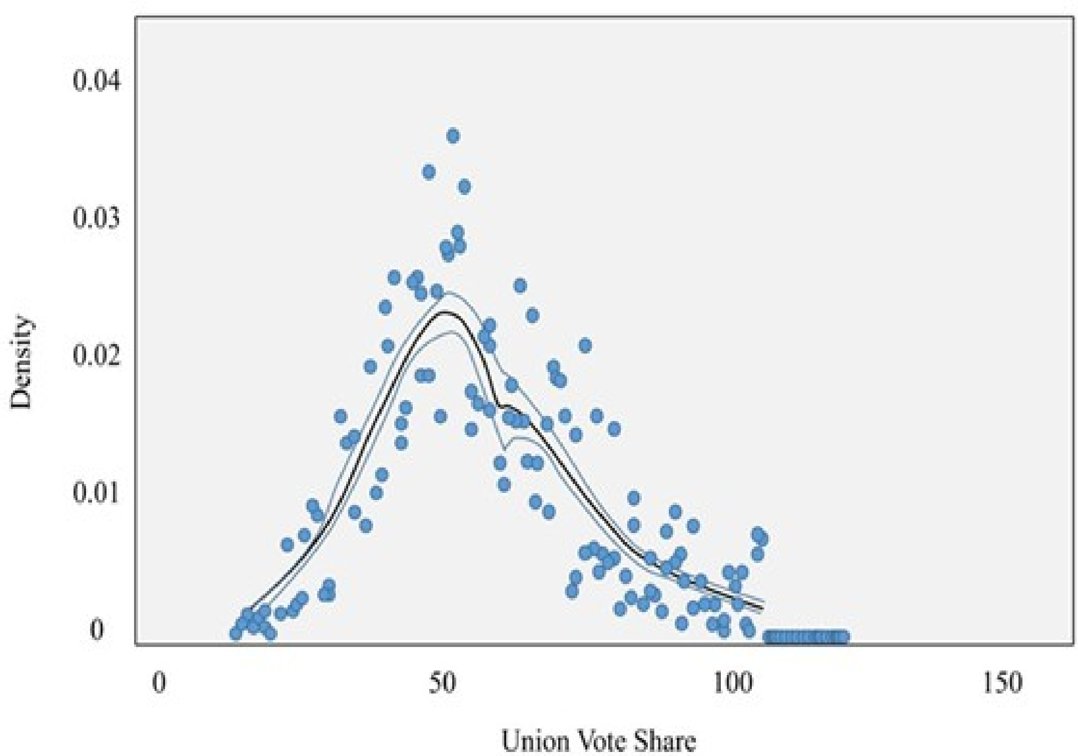

Figure 2. Confidence of Association Rules Generated from Transaction Database

Through the Apriori algorithm, we can not only explore the associations between course contents in depth but also discover how students with different learning styles utilize the most effective learning strategies to improve their learning effectiveness. The experiments have demonstrated that optimizing the Apriori algorithm can significantly enhance computational efficiency, and the resulting rules can effectively assist teachers in providing learning supervision and guidance. By studying the vertical layout dataset based on bit mapping, this paper aims to investigate the learning efficiency of the CAAR (Classification Model for Association Rules) and its application in distributed computing. To obtain more accurate results, this paper used 33.2 seconds rather than the original 33.2 seconds. By leveraging the method of bit vectors, we can partition and efficiently transmit the data to other locations. To construct a more effective classification system, a new algorithm can be employed that automatically adjusts algorithm parameters according to different requirements, thereby achieving a more efficient completion of classification tasks. Through the utilization of Master/Slave technology, we can synchronize the processing of multiple datasets in a single Agent, thereby constructing an efficient multi-agent ensemble. To achieve this, we simulated the working states of different Agents to explore the additional time they require in the process of extracting and processing similar information, including communication and information aggregation. In this experiment, we set *minsup* and *minconf* as evaluation criteria, with *minsup* set to 0.01 and minconf set to 0.50.

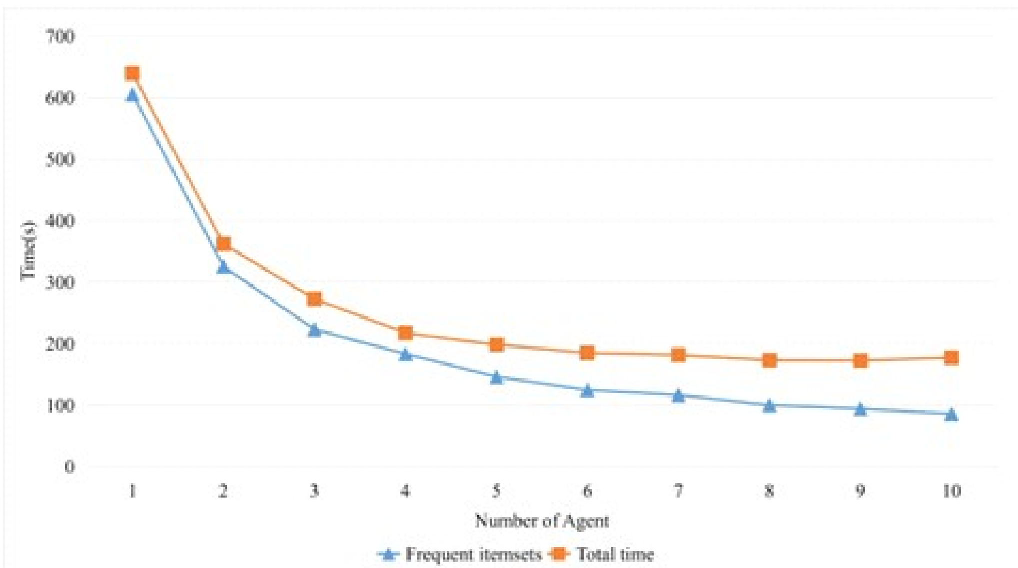

Figure 3. Number of Agents and Mining Time

Based on the experimental results in Figure 3, we found that sufficient agent numbers significantly improve the processing efficiency of classification association rules. For example, if the Agents reach 6, the processing efficiency will decrease to 188 seconds. However, suppose the number of agents exceeds 6, even if they are processed more frequently. In that case, it will result in additional processing costs, which will not lead to any improvement in processing efficiency. With the continuous expansion of agents, even in the case of 6, the data mining efficiency of this dataset cannot be significantly improved.

5. Conclusions

In recent years, due to the rapid popularization of artificial intelligence technology, many new data have come from various types of models and algorithms. Among them, the Apriori algorithm is one of the most effective, as it can provide accurate and reliable models, helping us quickly identify and predict future trends. Therefore, the Apriori algorithm has been proven to be a suitable model for educational courses. It can help us quickly identify and predict future trends, and its reliability and maneuverability are also extreme. By introducing the Apriori algorithm, it performs well in supervised image content analysis and large-scale data mining. Its characteristics include fast computation, excellent analysis accuracy, clear rules, compact rule specifications, and convenient model operation, thus enriching future research fields.

References

[1] Mao, C., Wang, Q., Feng, K., others. (2021). Research on the mechanism of multi-resource mapping matching and transmission of road system based on meme theory. *Journal of Physics: Conference Series, 1838*(1), 01–20. https://doi.org/10.1088/1742-6596/1838/1/012020

[2] Zhang, H., Padua, S. A., Li, Y. (2021). Research on the design of preschool education management information system based on computer technology. *Journal of Physics: Conference Series, 1915*(2), 02–03. https://doi.org/10.1088/1742-6596/1915/2/022003

[3] Xu, J. (2022). Research of online courses construction based on the theory of interaction. *Journal of Higher Education Research, 3*(1), 26–28. https://doi.org/10.32629/jher.v3i1.624

[4] Wang, Jianhua., Zhang, Yuanrui., Zhang, Jianbin. (2021). Research on ideological and political education of e-commerce course for marketing major. *Education Study, 3*(1), 50–54.

[5] Nakabayashi, K., Morimoto, Y. (2023). Design and implementation of a distributed multi-platform learning environment based on an extensible learning support system architecture. *Transactions of Japanese Society for Information and Systems in Education, 40*(1), 28–41. https://doi.org/10.14926/jsise.40.1_28

[6] Li, H., Dong, X. (2022). Construction of online ideological and political education platform based on artificial intelligence technology. *International Conference on E-Learning, E-Education, and Online Training*. Cham: Springer Nature Switzerland, 129–144. https://doi.org/10.1007/978-3-030-96307-3_12

[7] Zhang, Y. (2021). Research on ideological and political education of college students from the perspective of system theory. *Frontier of Higher Education, 2*(2), 1–17.

[8] An, N., Wang, J., Wang, H. (2018). Research and practice on innovative methods of ideological and political education for college students based on the internet of things+ technologies. *Educational Sciences: Theory & Practice, 18*(5). https://doi.org/10.12738/estp.2018.5.023

[9] Wang, Y. (2020). Analysis on the construction of ideological and political education system for college students based on mobile artificial intelligence terminal. *Soft Computing, 24*(11), 8365–8375. https://doi.org/10.1007/s00500-019-04190-6

[10] Wu, X. (2022). Research on the reform of ideological and political teaching evaluation method of college English course based on "online and offline" teaching. *Journal of Higher Education Research, 3*(1), 87–90. https://doi.org/10.32629/jher.v3i1.624