Eng Int Syst (2025) 2: 131–140

Engineering
Intelligent Systems

Association Rule Mining for English Digital Archive System Based on Improved Apriori Algorithm

Xiaoling Fang[1*]

[1]*School of Foreign Languages, Chongqing College of Humanities, Science and Technology, Chongqing, 40000, China*

With the widespread use and implementation of digital technology, the English digital archive system has accumulated a significant amount of data encompassing various dimensions such as learning behavior, teaching processes, and student feedback. Extracting valuable information and knowledge from this vast data has now become crucial for educational research and management. To this end, it is suggested that an enhanced convolutional neural network be combined with an Apriori algorithm to design and optimize a digital archive management system and association rule mining for English language. The improved Apriori algorithm takes into account the data's peculiarities and mining demands, thereby yielding comprehensible and high-quality results. The study's outcomes revealed that when the system reached a maximum iteration of 114 times, the proposed method attained the highest fitness value of 98.25 on the training set, in comparison with other fitness values. Similarly, when the proposed method was tested on the validation set, it achieved a fitness value of 98.74 after reaching a maximum iteration of 61 times, compared to fitness values obtained by other methods. The change in the overfitting curves shows that the model's performance in managing the data was stabilized after 30 iterations. With the optimized training model, the accuracy of data manipulation progressively increased to a consistent 90% and eventually converged to 99.99%. In practical applications, the test scores obtained by the proposed system for data transmission, preservation, retrieval, and management, all surpassed 84 points –a significant improvement. Notably, the system's data management score exceeded 92 points. The research findings demonstrate the clear advantages of the new methodology over the traditional approach in regards to accuracy and operational efficiency. It also indicates the proposed method's capacity to effectively manage the vast amount of data in the English digital archives of colleges and universities, thus yielding robust data support for research and management purposes in English education.

Keywords: Improved Apriori algorithm; English language; digital archiving system; association rules; educational management

1. INTRODUCTION

With the rapid advancement of information technology, digital storage and information management have become essential tasks for modern businesses and research institutions. In the realm of education, the proliferation and expansion of English language education (ELE) have led to the gradual integration of the English Language Digital Archiving System (ELDAS) into educational management [1–2]. ELDAS systems offer significant data support for educational activities and research. They provide vast data resources for researchers, creating a strong foundation for ELE research. The information in ELDAS is extensive and complex, covering aspects such as students' learning history, grades, feedback, and textbook usage. Mining valuable information from this data to enhance educational management and teaching is a crucial concern in current ELE research [3–4]. Association rules (AR) mining, as a significant subset of data mining, offers an efficient approach to achieving this aim. The Apriori algorithm, one of the most traditional AR mining algorithms, performs data mining

*Email: 13752888802@163.com

by gradually identifying frequent item sets and creating ARs based on these sets [5]. However, the conventional Apriori algorithm is inefficient and computationally onerous when handling large datasets, particularly those with numerous items and transactions. This is because the entire dataset needs to be scanned multiple times in order to generate a substantial set of candidate frequent items, resulting in a significant computational overhead when big data is involved. In this study, an optimized method is proposed based on the improved Apriori algorithm, specifically for AR mining in ELDAS. This tool is expected to provide new insights to ELE administrators and researchers, helping them extract valuable information from digital archives and make data-driven decisions to improve education.

2. RELATED WORKS

In the age of continuously evolving Internet technology, it is inevitable for various management systems to adopt information technology. Furthermore, with the advent and increase of big data, an artificial intelligence boom is occurring once again. Deep neural networks (DNN), one of the quintessential machine learning networks, are extensively employed in construction, sales, and management control. Chen et al. proposed a DNN-based method for detecting wooden structures in response to the problem that wooden structures are susceptible to termite activity resulting in internal holes. By means of DNN, the method analyzed the acoustic waves generated by knocking on the wooden structure to determine whether it contained holes. Test results showed that the method has high accuracy and versatility in identifying internal damage to wooden structures [6]. Wu et al. proposed a three-layer DNN-based fingerprint biological key generation framework to address the problem of unstable fingerprint key generation. The framework can eliminate the inconsistency of fingerprint samples by using DNN for feature extraction and layer-by-layer convolutional projection of fingerprint samples. Testing confirmed that the framework could achieve an accuracy of 98% for fingerprint recognition, with a false recognition rate of only 1.5% [7]. Preethi proposed a DNN-based e-commerce recommendation system facilitating the way that shopping sites recommend goods to potential customers. First, the system generated a glossary of terms related to the goods, and then the unimportant statements in the glossary were eliminated by Dirichlet allocation. Test results revealed that the goods recommended by this system were more in line with customers' preferences [8]. Walters et al. proposed a DNN-based ice rink linear segmentation method that needs to take into account real-time sports analysis, which segmented the rink linearly by BenderNet and RingerNet. The method was tested for real-time localization of hockey rinks with high accuracy while maintaining high efficiency [9]. Shafiee and his team proposed a DNN-based assisted driving system to address the problem that drivers are prone to traffic accidents due to fatigue and other factors. This system can largely avoid traffic accidents and injuries caused by driver's operational errors, and the system can also provide assisted driving for disabled people [10].

Data mining is widely used in various industries for its ability to process large amounts of data and transform them into useful information and knowledge. Pan and Yang proposed a marketing method based on data mining algorithms in order to develop personalized marketing strategies. The method obtains the intervals with the highest marketing success rate by calculating the set of frequent closed items and positive and negative sample support. The success rate of the marketing strategy proposed by this method was tested and found to be 8% higher than the traditional method, and it demonstrated better performance with a smaller number of intervals [11]. Wu proposed an information screening method, based on data mining and XBRL technology, for the processing of business information and to provide support for company decision making. Test results indicated that the method has a high accuracy rate for processing business information and can provide strong support for the development of corporate development strategies [12]. Latif et al. proposed a data mining method based on linear regression and particle swarm optimization for the purpose of estimating software workload. The test results showed that the RMSE value of this method for software workload estimation was 1552, and the estimation accuracy was improved by 3.12% [13]. Matondang et al. proposed a data-mining forecasting model for the fishing industry in order to forecast the quantity of growth in the fishing industry and assess the factors affecting it. The model used the C45 algorithm to categorize the fishing industry data as a way of determining the factors the growth quantity of the fishing industry. After testing, the model was found to be highly accurate in predicting the amount of growth of the fishing industry and determining the factors affecting it [14]. Fan et al. proposed a data mining method based on frequency analysis, AR and entropy clustering so as to mine the laws of traditional Chinese medicine treatment of psoriasis and pharmacological exploration. The method revealed the potential configuration laws of core drugs by screening out 349 psoriasis treatment prescriptions in which the frequency of use was greater than 30 times. After testing, the model effectively mined the signaling pathways of psoriasis treatment prescriptions and provided a powerful reference for psoriasis treatment [15].

In summary, DNNs and data mining, which utilize commonly-used network structures and data processing methods, are extensively employed in various fields, impacting the growth and development of diverse industries. Currently, digital file management systems are the primary method used to manage courses, grades, and other student-related information in colleges and universities. These systems have been significantly impacted by advancements in information technology, and artificial intelligence has the potential to offer robust decision-making support for English teaching in higher education. By processing student and administrative data more accurately, AI can facilitate more informed decision-making in this field. However, the system frequently produces considerable quantities of data, the processing of which cannot be accomplished efficiently by traditional algorithms. To address this issue, the study suggests a data mining approach that combines CNN and Apriori algorithms to process effectively and efficiently the data from English file management systems in colleges and universities. The goal

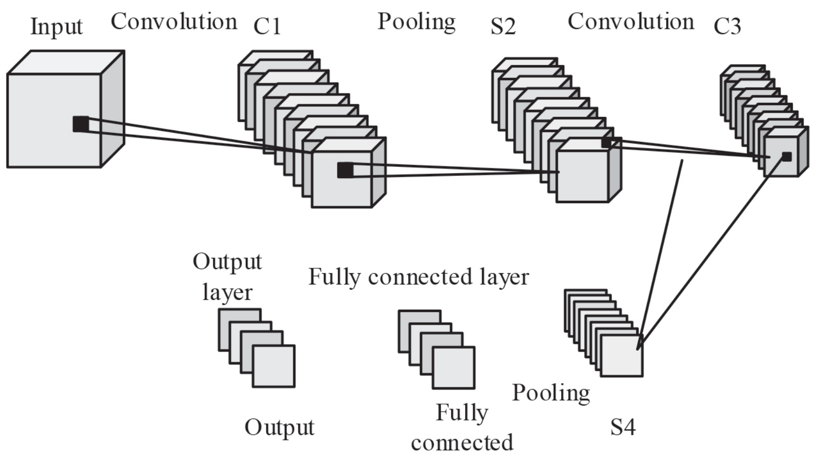

Figure 1 Schematic diagram of CNN framework.

is to accurately locate and categorize relevant information in a logical manner.

3. DESIGNING AN ENGLISH DIGITAL ARCHIVE MANAGEMENT METHOD BASED ON DATA MINING AND ASSOCIATION RULES

3.1 File Management Design by Incorporating DNN and Data Mining Algorithms

The management of English digital archives in colleges and universities usually produces a huge amount of data, and the efficient processing of these data is highly problematic. In order to effectively solve the problem, in this paper, a combination of deep learning and a data mining Apriori algorithm is proposed for the building of an English digital archive management model [16–17]. DNN, an important technology in the field of machine learning, is essentially an unsupervised learning approach. Common types of networks include convolutional neural networks, recurrent neural networks, deep belief networks, deep self-encoders, and generative adversarial networks, etc. CNNs are widely used in many fields due to their superior feature representation capabilities. In this study, CNN is adopted as the basic framework of the model constructed for the experiment. The framework of this model is depicted in Figure 1.

As Figure 1 clearly shows, the CNN model consists of a convolutional layer, a pooling layer and a fully connected layer. Of these, the convolutional layer is responsible for capturing and extracting local features in the data [18]. In this study, a one-dimensional convolutional approach was chosen for the experiments. This convolutional method can handle vector operations involving filters with numerical sequences. The linear sequence of data after this transformation can be computed by using equation (1).

$$c_j = f^T s_j - m + 1 \quad j + b \tag{1}$$

In equation (1), c_j denotes the new sequence and f denotes the filter. s_j denotes the original sequence, m denotes the filter length, and j denotes the index. b denotes the bias term and also acts as a threshold in the model. Considering the complexity of the input to the convolutional layer, the input data is usually not only a simple sequence of values, but may also be a sequence of multidimensional vectors. In order to handle this situation, the experiment needs to convert the dimensionality of each vector into matrix form. Therefore, it is then necessary to use a filter set of b filters and a bias set consisting of b biases [19–20]. The new matrix transformation can be referred to equation (2).

$$conv(S, F, B) \quad R^{d\times|s|} \rightarrow R^{d\times(|s|+m-1)} \tag{2}$$

In equation (2), S denotes a sequence of multiple values, F denotes the filter set. B denotes the bias group, and $R^{d\times|s|}$ denotes the set of real numbers. Because the CNN model needs to better express the nonlinear relationship, the activation function is often added to the output of each layer. The detailed calculation of the activation function is obtained with equation (3).

$$f(x) = \max(0, x) \tag{3}$$

The pooling layer plays a very crucial role in the neural network, which is able to aggregate the input data, thus enhancing the tolerance of the model to small changes in the data. The specific operation of pooling is calcultated with equation (4).

$$Pooling(a){:}R^{|a|} \rightarrow R^{[|a|/k]} \tag{4}$$

In equation (4), a denotes the vector and k denotes the number of values to be aggregated, which is actually a hyperparameter. The new matrix P can be obtained when applying the pooling operation to the matrix, as shown in equation (5).

$$Pooling(A) \quad R^{d\times|a|} \rightarrow R^{d\times[^{|a|}/k]} \tag{5}$$

In equation (5), A denotes the vector matrix. When the data is convolved and pooled, a primary representation or called first-order expression of the original data can be obtained. By repeating the above operation several times, a more complex, higher-order data expression can be further obtained. This higher order data expression is computed with equation (6).

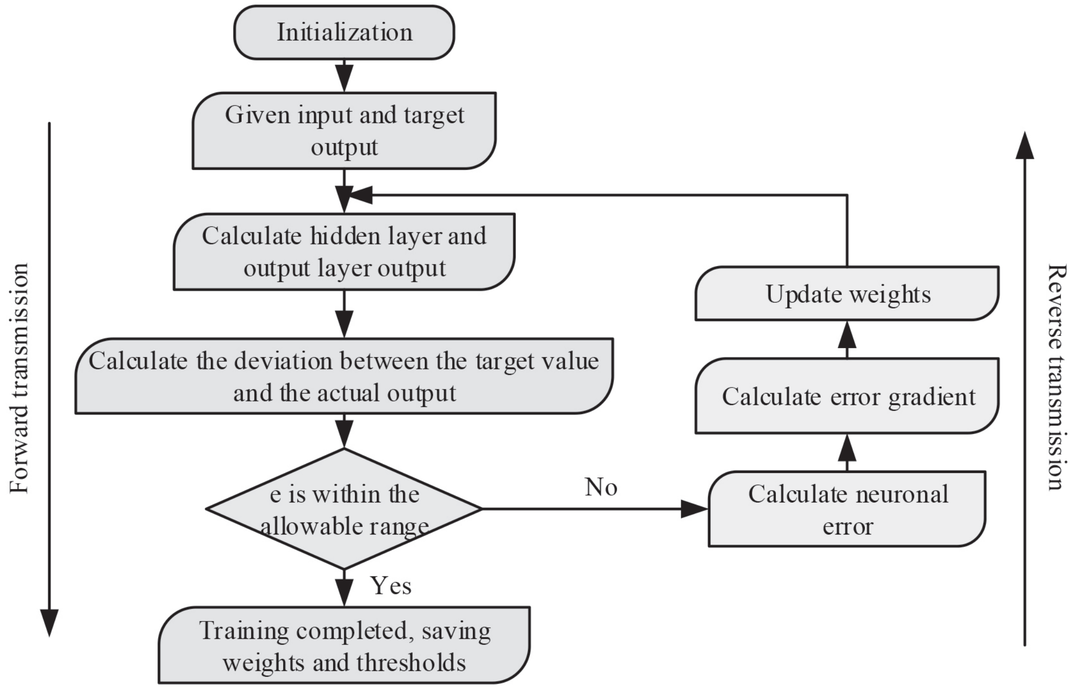

Figure 2 CNN Training Process.

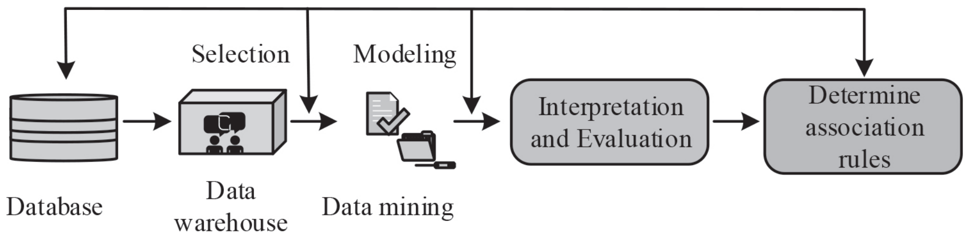

Figure 3 Data Mining Process.

$$P_j^i = Pooling\left(\alpha\left(\sum_{k-1}^{K_i=1} conv(P_k^{i-1}, F_{j,k}^i, B_{j,k}^i)\right)\right) \quad (6)$$

In equation (6), P_j^i denotes the first order expression of the j th value. α denotes the activation function up and K_j denotes the number of values. The core task of the fully connected layer is to integrate and compress the information from the uppermost layers in the network into a single comprehensive vector. This process is computed with equation (7).

$$\hat{x} = \alpha(p^T H) \quad (7)$$

In equation (7), $\hat{x}$ denotes a vector composed of features of the original data. Matrix H denotes the parameters to be optimized during the training process, and p denotes the vector composed of higher-order expressions. The CNN training process is shown in Figure 2.

The data accumulated in the English digital archive management system needs to be transformed into practical strategies and information by appropriate means [21]. For this reason, several data mining techniques were developed because of their ability to deeply analyze and extract patterns from the data. Some of the more common algorithms among the many data mining methods include clustering, classification, AR, genetic, and Bayesian. The data mining steps and methods are shown in Figure 3.

In Figure 3, data mining first requires collecting raw data, performing data cleaning and preprocessing to obtain a database. Secondly, store the preprocessed data in a data warehouse for quick querying and analysis. Then, use data mining algorithms to discover patterns and relationships in the data, and establish models based on the results of data mining. Subsequently, explain the results of data mining and evaluate the performance of the model. Finally, determine the association rules. AR is well suited for analyzing transactional or transactional databases with the aim of discovering possible patterns hidden in the data [22]. In this research, an AR mining algorithm (Apriori algorithm) was used as it is widely recognized as being effective. The Apriori algorithm works by identifying frequently occurring itemsets by means of a level-by-level search to find out potential correlations in the data. The flow of the Apriori algorithm is shown in Figure 4.

As can be seen in Figure 4, before applying the Apriori algorithm, specific support and confidence thresholds need to be established. Here, the support is the frequency of occurrence of a particular itemset in a transaction, while the confidence is the ratio of the frequency of occurrence of the itemset in a set ofARs in terms of the leading itemset. After the thresholds have been established, the algorithm further

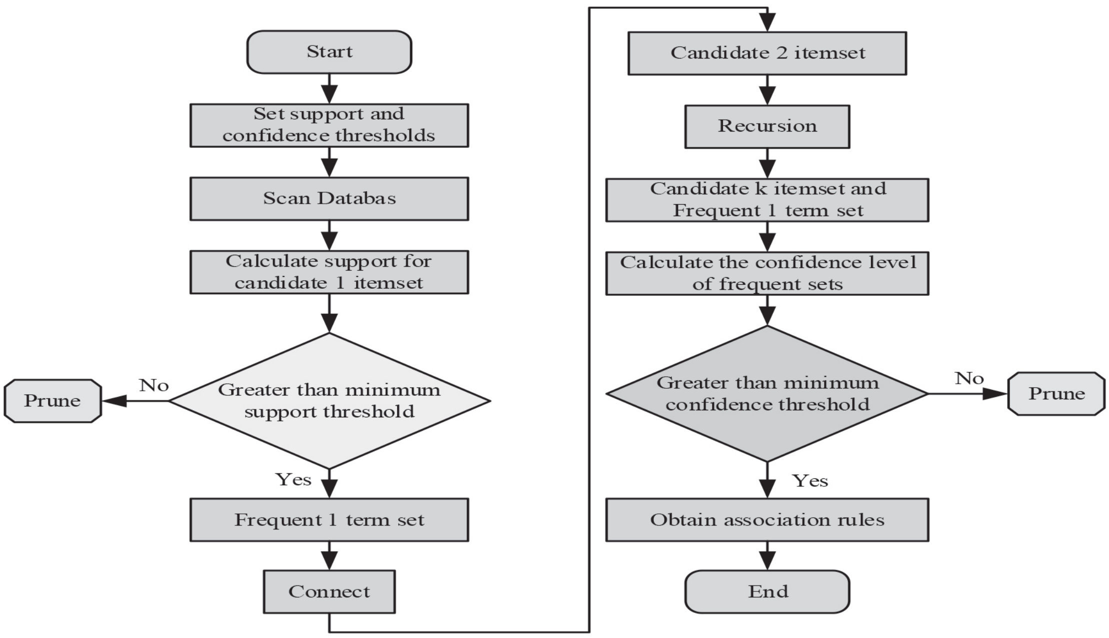

Figure 4 Apriori Algorithm Process.

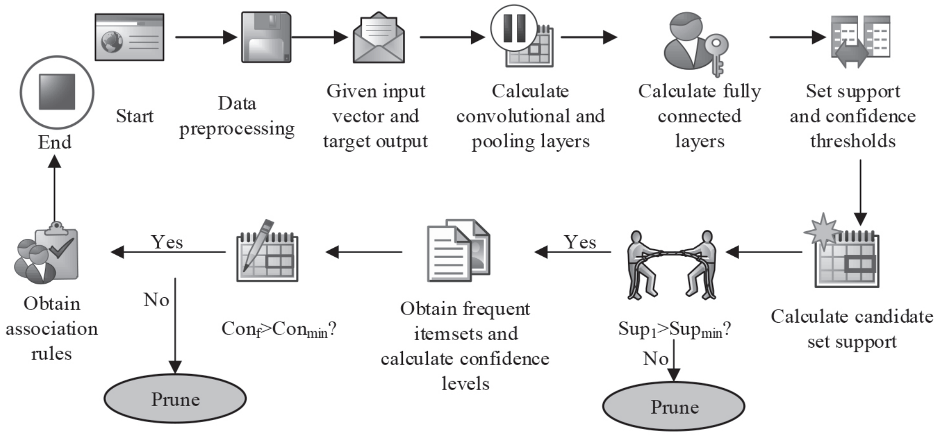

Figure 5 CNN-APriori Model Process.

calculates the support degree for the candidate set and the confidence degree for the frequent set. Combining the set minimum support and confidence thresholds, the algorithm generates a series of ARs. Among these generated rules, the experiment selects those that fulfill the practical needs. If the results do not meet expectations, consider adjusting the thresholds and then re-execute the above process.

3.2 Mining Model Construction for English Digital Archives Management in Colleges and Universities Counting CNN-Apriori

Due to the large number of students in colleges and universities, the corresponding English digital archive management system usually contains a huge amount of data, and all the data are mostly very complicated, so the experimental processing of archive management data is actually a relatively tedious task. If the traditional CNN and AR algorithms are utilized, it is basically impossible to complete the effective mining and accurate classification of massive data [23]. In view of this, the experiment proposes to combine the above two algorithms to jointly construct a data mining model to improve the performance of the system. The corresponding obtained process is schematized in Figure 5.

Based on the process in Figure 5, it can be seen that the raw data is first normalized, which ensures that the data is on a uniform scale and reduces the adverse effects of anomalous data. After normalization, the experiment is able to provide the model with input vectors and expected outputs through which the hidden layers and the outputs of each cell can be calculated. Next, the error between the expected output and the actual output is measured. If the error is within acceptable limits, the process continues. Otherwise, the parameters of the hidden layer need to be readjusted using the back propagation algorithm. In addition, feature extraction is an important part of this process and the extracted data will be used as the output of the subsequent steps. These output data can be used as input to the Apriori algorithm. Once the specific support and

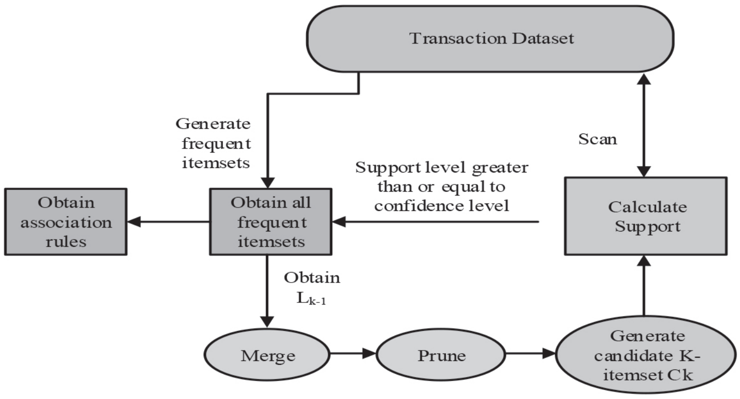

Figure 6 Framework for Running Association Rules.

confidence thresholds are given, one can start scanning the entire database and calculate the support for each candidate set. The existence of frequent itemsets can be confirmed only if these support degrees exceed the thresholds set. Calculate the confidence level of each frequent itemset, and only when their confidence level exceeds a specific threshold can the existence of a valid AR be considered. Once an AR that meets the requirements is obtained, it can be considered as an output result. Where the support and confidence are calculated in equation (8).

$$\begin{cases} Sup(X) = \frac{count(X \subseteq Y)}{|TS|} * 100\% \\ Con(Y = Z) = \frac{Sup(Y=Z)}{Sup(Y)} * 100\% \end{cases} \tag{8}$$

In equation (8), $Sup(X)$ denotes the degree of support and *count* $(X \subseteq Y)$ denotes the number of supports. $|TS|$ denotes the thing set mode, $Con(Y \implies Z)$ denotes the confidence level, and $Sup(Y \implies Z)$ denotes the support in AR. It is assumed that T_i is the thing, TS is the thing set, $m \times n$ is the thing set order, and AR is the $Y \implies Z(S\%, C\%)$. The equation for support in AR is shown in equation (9).

$$Sup(Y \implies Z) = S\% = \sum_{i=1}^{m} \{(Y \implies Z) \subseteq T_i\} \tag{9}$$

After setting the appropriate confidence and support thresholds, the experiment can then further calculate the support of the candidate set and generate the frequent item sets accordingly. After completing the computation of all frequent item sets, the generation of AR is continued. The computation of the associated AR is shown in equation (10).

$$\frac{Supcount(L)}{Supcount(s)} > \min_{-} conf \tag{10}$$

In equation (10), L denotes the set of frequent terms. S denotes the non-empty subset of L. And $\min_{-} conf$ denotes the minimum confidence threshold. The normalization equation is shown in equation (11).

$$X = \frac{x - \min}{\max - \min} \tag{11}$$

In equation (11), X represents a single data point after normalization, min represents the minimum value of this data point, max represents the maximum value of the data point, and x simply represents a single sample data. In order to more accurately reveal the interrelationships among different variables, the study used the correlation coefficient as a metric, the equation for which is (12).

$$r = \frac{\sum_{i=1}^{n} (X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{n} (X_i - \overline{X})^2} \sqrt{\sum_{i=1}^{n} (Y_i - \overline{Y})^2}} \tag{12}$$

In equation (12), X and Y denote two different variables, and r A denotes the correlation coefficient, the larger its absolute value, the stronger the correlation between the variables. The model supervises the training through a loss function, which is calculated using equation (13).

$$L_s = -\sum_{k=1}^{m} \log \frac{\exp(W_{y_k}^T x_k + b_{y_k})}{\sum_{i=1}^{n} \exp(W_i^T x_k + b_i)} \tag{13}$$

In equation (13), x_k denotes extracted features and W_{yk} denotes learning weights. b_i denotes the bias value, m denotes the training batch size, and n denotes the number of categories. The loss function is the key to assessing the effectiveness of the model, as it reveals the deviation of the model output from the real data. In order to provide a comprehensive assessment of the model's prediction effect, the experiment introduces core evaluation metrics such as accuracy, precision and F-measure. The purpose of the study is to create an integrated CNN-Apriori model that combines the deep learning properties of CNN and the rule mining capabilities of the Apriori algorithm. This integrated approach aims to address the respective limitations of the CNN model and the Apriori algorithm, especially to reduce the intermediate sets generated by Apriori in the data mining process. Ultimately, this integrated approach provides a more efficient and accurate analysis of big data related to English digital archive management in universities and colleges. The operational framework of the English digital archive management system under AR obtained is shown in Figure 6.

Table 1 Basic Environmental Parameters of the Experiment.

Parameters variables	Parameter selection
The overall implementation platform of the system	Simulink
System PC side memory	36G
Data storage	Sheepdog
Data regression analysis platform	SPSS Excel
Simulation experimental environment	MRDS
Operating system	Microsoft Windows 10
CPU model	Intel Xeon E5-2665x2
CUDA	3584
Operating system	Ubuntu16.04LTS
GPU	GTX1080Ti llGx4

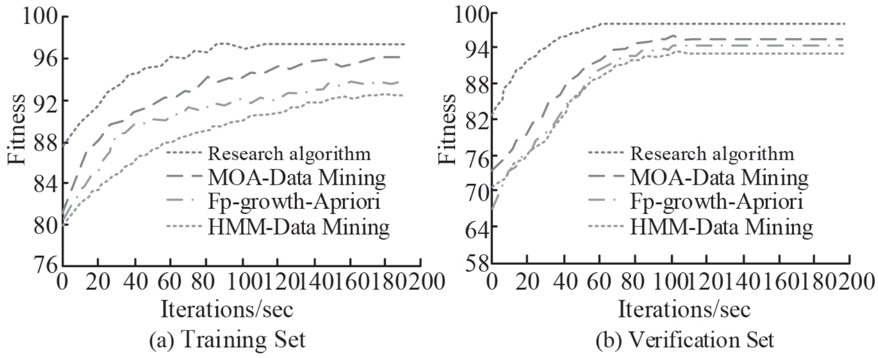

Figure 7 Comparison of Convergence Changes.

4. PERFORMANCE TESTING AND APPLICATION EFFECTIVENESS OF THE ENGLISH LANGUAGE DIGITAL ARCHIVING SYSTEM

Before the experiment was formally carried out, tests were conducted to determine the feasibility and validity of the ELDAS constructed by the Institute, , in order to ensure the fairness and rationality of all subsequent experiments. Firstly, the basic parameters of the relevant software and hardware equipment were established, as shown in Table 1.

Multi-objective optimization and data mining(MOA-Data Mining), frequent itemsets and improved Apriori algorithm (Fp-growth-Apriori) in AR, Fusion of Data Mining and Hidden Markov Models (HMM-Data Mining) (used for the evaluation of PE teaching quality) and various research methods were selected for performance comparison [24–26]. The experiments were conducted with the exclusion of certain parameters, and all remaining parameters remained the same. The number of system iterations for all four algorithms was set to 200, and 2112 data items from the English file management system of a university were selected; after removing erroneous data through data preprocessing, 1800 data items remained. The study divided 35% of the data into a validation set and 65% of the data was used as the training set. The convergence obtained by running the four models on the two data sets was compared, which is shown in Figure 7.

Figure 7(a) shows the change curve of the adaptation values of different models on the training set. It was found that when the system keeps running iteratively up to 114 times, the proposed method achieves the maximum adaptation, with a value of 98.25. At this time, the convergence adaptations of the other three models are still constantly changing, iand the system does not have a relatively smooth adaptation value. Figure 7(b) shows the convergence variation curves of the different algorithms on the validation set. When the management system runs up to 58 times, the research method has the maximum adaptation value, with the value of 61 times, and the adaptation value of the proposed method is 98.74. The three methods compared with the proposed method all have smooth adaptation values after 100 iterations, and all of the adaptation values are smaller than that of the proposed method. These results indicate that the proposed method always has the highest fitness value, has a more significant and faster convergence rate, and the system reaches stable operation more quickly under the same experimental conditions. The overfitting curves of the selected English file management system training data were then analyzed, as shown in Figure 8.

In Figure 8, the horizontal coordinate indicates the number of iterations of the system (100); the vertical coordinates vary. The curve change in Figure 8(a) indicates that when the number of iterations is 30, the ability of the proposed model to manage the data is gradually stabilized, and it can map all the selected training samples to the corresponding location labels. In Figure 8(b), with the increase of the number of iterations, the validation accuracy of the research model starts to decrease and the overall accuracy is less than 65%; the accuracy of the training set also starts to stabilize and remains at a higher value. In Figure 8(c) and Figure 8(d), with the iteration of the system, the optimized accuracy of the training model starts to increase linearly, and the accuracy of the final

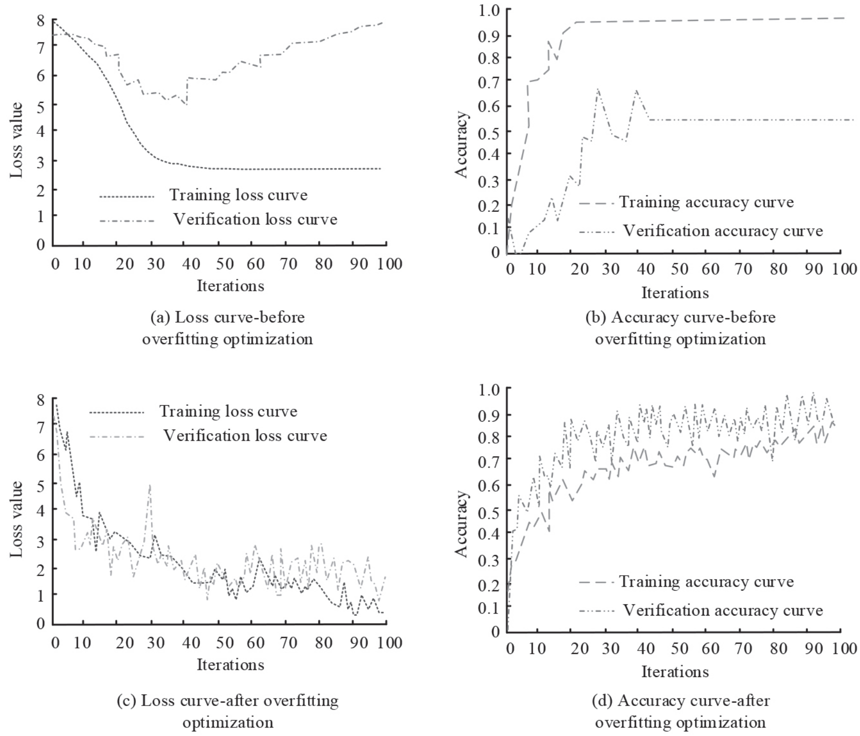

Figure 8 Overfitting diagram of optimized training data for English file management system.

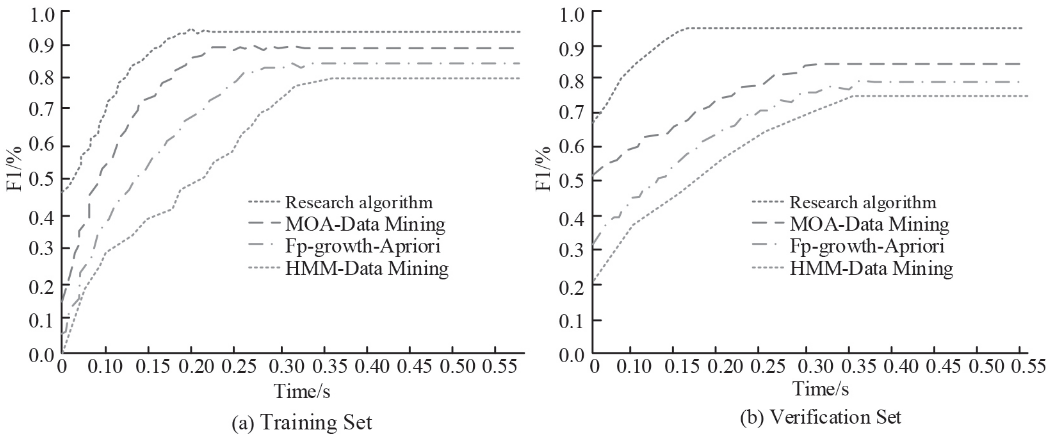

Figure 9 Comparison of the Fl Values of Different Models.

model can be maintained stably at 90% and constantly tends to reach 99.99%. Combining the above results, it can be seen that the proposed model can accurately manage the data and solve problems such as overfitting caused by the small data set during the training process. Then the Fl values obtained by running the four models on different data sets were analyzed. The results are shown in Figure 9.

Figure 9(a) shows the variation of Fl values of the models on the training set. It was found that the Fl values of all the models, in terms of the running time of the system, begin to increase. When the digital archive system runs up to 0. 22s, the Fl value of the research method reaches the maximum value of 0.947; after that, it remains in that state and continues to be stable. However, correspondingly, the Fl values of the other methods are smaller than those obtained by the proposed method. Figure 9(b) shows the variation curves of Fl values of different algorithms applied to the validation set. The Fl value of the proposed method is always at the maximum and has the best Fl value at 0. 17s. Comparison shows that there is a very significant difference between the research method and the other three algorithms, which also indicates that the proposed method has a smaller error in the operation of the English digital archive management system, and its overall performance is superior. Then the research constructed ELDAS was applied to the ELDAS of students in a domestic university to evaluate the application of the system in regard to data transmission, preservation, retrieval and management. The statistical results for these aspects are shown in Figure 10.

Figure 10 (a) and (b) enable a comparison of the practical effect of applying the ELDAS on the training set and validation set, respectively. Under the operation of the system constructed by the Institute, the test effect scores of the system in regard to data transmission, preservation, retrieval, and

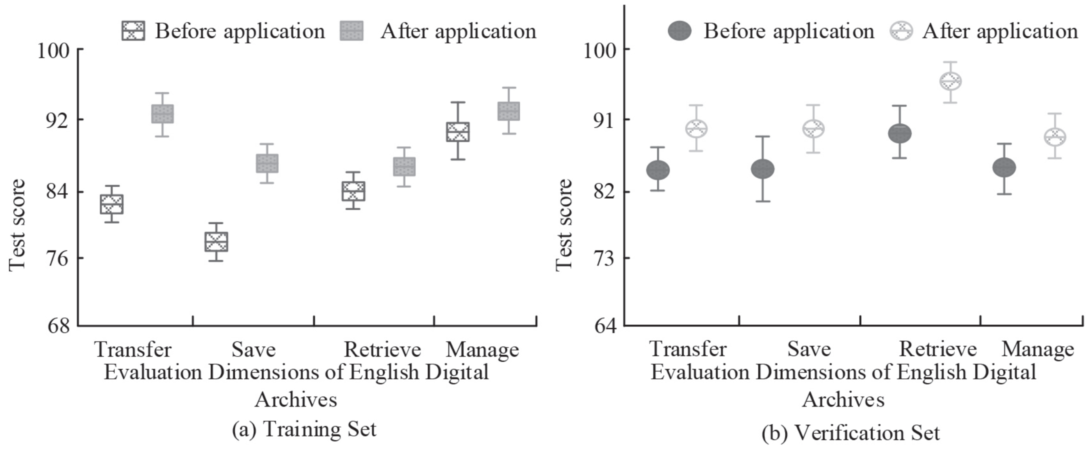

Figure 10 Application Effects of the System on Different Datasets.

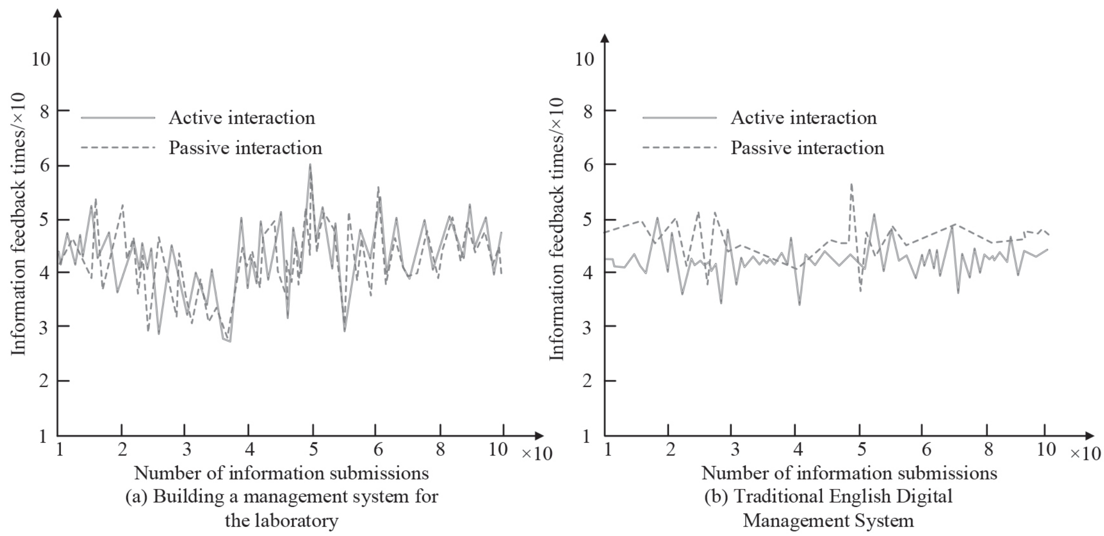

Figure 11 Comparison of Interactive Results between Two Management Systems.

management show various degrees of improvement, all of which are greater than 84, with the score for the system's management of data being greater than 92. Meanwhile, by applying the system to the validation set, the scores of all four aspects were obtained, and showed an increase significantly greater than 91. Figure 11 depicts the application of the institute-built system to English digital management, demonstrating that it can effectively improve the performance of data transmission, preservation, retrieval, and management. Finally, the interactive results obtained for the two different English digital management systems are compared, as shown in Figure 11.

As can be seen in Figure 11(a), there is a small difference between the number of information submissions and feedback during the implementation of the system constructed by the experiment; however, the overall difference is not very large. In addition, the active information interaction time is slightly more than the passive interaction. This indicates that the digital management system for English in higher education operates more efficiently and is more effective in managing English data. However, in Figure ll(b), the efficiency of the institute-constructed system fluctuates less and the number of passive interactions is greater, suggesting that the interactivity of the system needs to be further strengthened.

5. CONCLUSION

To provide valuable insights for ELE research and university management, and enhance system performance, the researchers propose that CNN be combined with an improved Apriori algorithm and applied to the system. CNN will boost the Apriori algorithm, and a new data structure will be introduced to achieve system control. In this study, the research method was applied 114 times to the training set and 61 times to the validation set. The maximum fitness values achieved were 98.25 and 98.74, respectively. The overfitting curve shows that the performance of the model gradually stabilized after 30 iterations. It can accurately map all selected training sample locations to their corresponding labels. During the experiment, the proposed method achieved a peak Fl score of 0.947 on the training set when the system was operating for 0.22 seconds. Furthermore, the efficiency of the system in managing data transfer, preservation, retrieval, and administration exhibited a range of enhancements, all of which exceeded 84, with the system achieving a score greater than 92 in the management of data. The system constructed by the Institute operated with high stability, showing minimal fluctuation and more passive interactions. The research findings confirmed the Institute's method as superior concerning operational efficiency, accuracy, and speed of convergence. This method is an effective means of mining AR in ELDAS. Further research could enhance the effectiveness of the Apriori algorithm on larger and more complex datasets. Furthermore, the combination of other data mining techniques, such as clustering and classification, could also be considered for more in-depth analysis and mining.

Funding

No funding was provided for this study.

Competing Interests

The authors have no relevant financial or non-financial interests to disclose.

Data availability statements

The data will be made available on request.

REFERENCES

1. Qianna S. Evaluation model of classroom teaching quality based on improved RVM algorithm and knowledge recommendation. Journal of Intelligent& Fuzzy Systems, 2021, 40(2): 2457–2467.
2. Jain A, Yadav D, Arora A. Particle swarm optimization for Punjabi Text summarization. International Journal of Operations Research and Information Systems (IJORIS), 2021, 12(3): 1–17.
3. Cui J. Application of deep learning and target visual detection in English vocabulary online teaching. Journal of Intelligent & Fuzzy Systems, 2020, 39(4): 5535–5545.
4. Luo X. efficient English text classification using selected machine learning techniques. Alexandria Engineering Journal, 2021, 60(3): 3401–3409.
5. Gao P Li J, Liu S. An introduction to key technology in artificial intelligence and big data driven e-learning and e-education. Mobile Networks and Applications, 2021, 26(5): 2123–2126.
6. Chen L, Xiong H, Sang X, Yuan C, Li X, & Kong Q. An innovative deep neural network based approach for internal cavity detection of timber columns using percussion sound:. Structural Health Monitoring, 2022, 21(3): 1251–1265.
7. Wu Z, Lv Z, Ding W, Zhang J, & Kang J. Fingerprint bio key generation based on a deep neural network. International Journal of Intelligent Systems, 2022, 37(7):4329–4358.
8. Preethi S. ERS: latent dirichlet allocation based e-commerce recommendation system using deep neural network. Psychology (Savannah Ga 2021, 58(2):953–959.
9. Walters P Clausi D, Wong A, & Fani M. BenderNet and RingerNet: highly efficient line segmentation deep neural network architectures for ice rink localization. Journal of Computational Vision and Imaging Systems, 2021, 6(1):1–4.
10. Shafiee MJ, Jeddi A, Nazemi A, Fieguth P, & Wong A. Deep neural network perception models and robust autonomous driving systems: practical solutions for mitigation and improvement. IEEE Signal Processing Magazine, 2021, 38(1):22–30.
11. Pan Q, Yang G Application of mining algorithm in personalized Internet marketing strategy in massive data environment. Journal of Intelligent Systems, 2022, 31 (1):237–244.
12. Wu P Data mining and XBRL integration in management accounting information based on artificial intelligence. Journal of Intelligent and Fuzzy Systems, 2020, 40(1): 1–12.
13. Latif A, Fitriana LA, Firdaus M R. Comparative analysis of software effort estimation using data mining technique and feature selection. JITK (Jurnal Ilmu Pengetahuan dan Teknologi Komputer), 2021, 6(2):167–174.
14. Matondang MR, Lubis MR, Tambunan H S. Analisis data mining dengan metode C.45 pada Klasifikasi Kenaikan Rata-Rata Volume Perikanan Tangkap. Brahmana Jurnal Penerapan Kecerdasan Buatan, 2021, 2(2):74–81.
15. Fan YH, Yang RQ, He JY, Gou X P Zhuang ZN, & Guo X J. Explore the rule and mechanism of psoriasis herbal treatment based on data mining and network pharmacology. Asian Toxicology Research, 2021, 3(2):P.34–43.
16. Hemeida A M, Alkhalaf S, Mady A, Mahmoud EA, Hussein ME, & Eldin A M B. Implementation of nature-inspired optimization algorithms in some data mining tasks. Ain Shams Engineering Journal, 2020, 11(2): 309–318.
17. Swathy M, Saruladha K. A comparative study of classification and prediction of Cardio-Vascular Diseases (CVD) using Machine Learning and Deep Learning techniques. ICT Express, 2022, 8(1): 109–116.
18. Khamparia A, Pandey B. Association of learning styles with different e-learning problems: a systematic review and classification. Education and Information Technologies, 2020, 25(2): 1303–1331.
19. Fang Y Luo B, Zhao T, He D, Jiang B, & Liu Q. ST-SIGMA: Spatio-temporal semantics and interaction graph aggregation for multi-agent perception and trajectory forecasting. CAAI Transactions on Intelligence Technology, 2022, 7(4):744–757.
20. Agarwal A, Sharma S, Kumar V, et al. Effect of E-learning on public health and environment during COVID-191ockdown. Big Data Mining and Analytics, 2021, 4(2): 104–115.
21. Chen X, Tang Y Data center energy management based on cloud computing and artificial intelligence. Engineering Intelligent Systems, 2024, 32(3): 257–266.
22. Chen X, Tang Y Data center energy management based on cloud computing and artificial intelligence. Engineering Intelligent Systems, 2024, 32(3): 257–266.
23. Dixon-Román E, Nichols T P, Nyame-Mensah A. The racializing forces of/in AI educational technologies. Learning, Media and Technology, 2020, 45(3): 236–250.
24. Patel D H, Kotadiya H, Vasant A R. Privacy-preserving association rule mining considering multi-objective through an evolutionary algorithm. International Journal of Computer Theory and Engineering, 2022, 14(1): 1–8.
25. Wang C, Zheng X. Application of improved time series Apriori algorithm by frequent itemsets in association rule data mining based on temporal constraint. Evolutionary Intelligence, 2020, 13(1): 39–49.
26. Zeng Y Evaluation of physical education teaching quality in colleges based on the hybrid technology of data mining and hidden mMarkov model. International Journal of Emerging Technologies in Learning (iJET) , 2020, 15(1): 4–15.