Research on deep mining of MOOC multimodal resources based on improved Eclat algorithm

Yu Cao, Shu-Wen Chen and Yu-Xi Wang*

School of Mathematics and Information Technology,
Jiangsu Second Normal University,
Nanjing 210013, China
Email: 5486217@qq.com
Email: 7542893@qq.com
Email: yuxiw@mls.sinanet.com
*Corresponding author

Abstract: In order to overcome the problems of low recall and precision in traditional MOOC multimodal resource mining methods, this paper proposes a new MOOC multimodal resource deep mining method based on improved Eclat algorithm. Based on cloud computing technology, according to MOOC resource pool structure, MOOC multi-modal knowledge map is constructed, and hash chain is used to analyse the attribute connection rules between knowledge maps. Based on the attribute connection rules, the improved Eclat algorithm is used to transform the captured modal information of resources, so as to design the MOOC multi-modal resource deep mining process and get the results of resource deep mining. The experimental results show that the recall and precision of this method are above 97%, the mining effect is better, and the mining time is always less than 0.7 s, the mining efficiency is higher, and the actual application effect is better.

Keywords: modal mining; association rules; MOOC resources; Eclat algorithm.

Reference to this paper should be made as follows: Cao, Y., Chen, S-W. and Wang, Y-X. (2023) 'Research on deep mining of MOOC multimodal resources based on improved Eclat algorithm', *Int. J. Continuing Engineering Education and Life-Long Learning*, Vol. 33, No. 1, pp.99–113.

Biographical notes: Yu Cao received his Bachelor's from Nanjing Normal University and Master's in Software Engineering from Nanjing University in 2008. She is currently a Second Lecturer in Jiangsu Normal University. He is mainly engaged in the research of open course teaching theory.

Shu-Wen Chen received his BS in Information Engineering from the Nanjing University of Posts and Telecommunications, Nanjing, China in 2005, and MS and PhD from the State Key Laboratory of Millimetre Waves, Southeast University, Nanjing, in 2009 and 2016. He has been a Lecturer with the School of Mathematics and Information Technology, Jiangsu Second Normal University, Nanjing since 2016.

Yu-Xi Wang is an Associate Professor of School of Mathematics and Information Technology of Jiangsu Second Normal University. He is mainly engaged in the application of computer network technology, computer education and information teaching research.

1 Introduction

Data mining technology is a popular technology of network data processing in recent ten years. Big data mining is based on cloud computing technology and provides high-performance computing, large-scale storage services and various applications through virtual computer server cluster, storage cluster and medium and high-level cluster (SDN) supported by software defined network technology. Among them, the modal data mining of MOOC resources has been paid more and more attention. In the cloud environment, MOOC resources (such as knowledge map) generate a lot of modal information through specific classification. The analysis of historical data shows that there are potential association rules in a variety of modal information. For example, in the distributed MOOC system, using cloud technology to mine its internal data, we need to complete the modal data mining on the simulator. When using virtual machine to analyse modal data, there will be repeated analysis of the same kind of modal data, resulting in data redundancy and high business load of virtual machine. At this time, it is necessary to use association rule algorithm to effectively mine the data in MOOC resource pool. The current academic research in this field has achieved initial results.

In Cao et al. (2016), a method of deep mining MOOC multimodal resources based on negative sequence pattern is proposed. When analysing the modal data, the hierarchical data is determined according to the correction coefficient of the threshold iteration algorithm, and the modal prediction model of the main engine is established. This method also improves the efficiency of the algorithm, but the model is only for the parameters of virtual host. On the basis of genetic network algorithm (GNP), a method of MOOC multi-modal resource deep mining based on sky pattern is proposed in Zhiyong et al. (2017). The method is from pattern aggregation representation to dynamic constraint satisfaction problem mining method. By using association rules and data independence feature analysis method, XML files and XML files are integrated by data independence. The MOOC data is analysed comprehensively, and then the resource status of MOOC is analysed through the comparative modal window. In Yao et al. (2017), a method of MOOC multimodal resource depth mining considering density weighted distance threshold is proposed. The knowledge map is integrated into a virtual cloud by using neural network algorithm, and the location is based on the contribution of various indicators.

The above methods can effectively complete data mining, but when they are applied to MOOC multimodal resource mining, they do not consider the running state of the data, and cannot accurately mine the hidden rules between knowledge maps, which has the problem of low recall and precision. In order to solve the problems of traditional methods, this paper proposes a deep mining method of MOOC multimodal resources based on improved Eclat algorithm, which has the characteristics of high recall and precision and low mining time

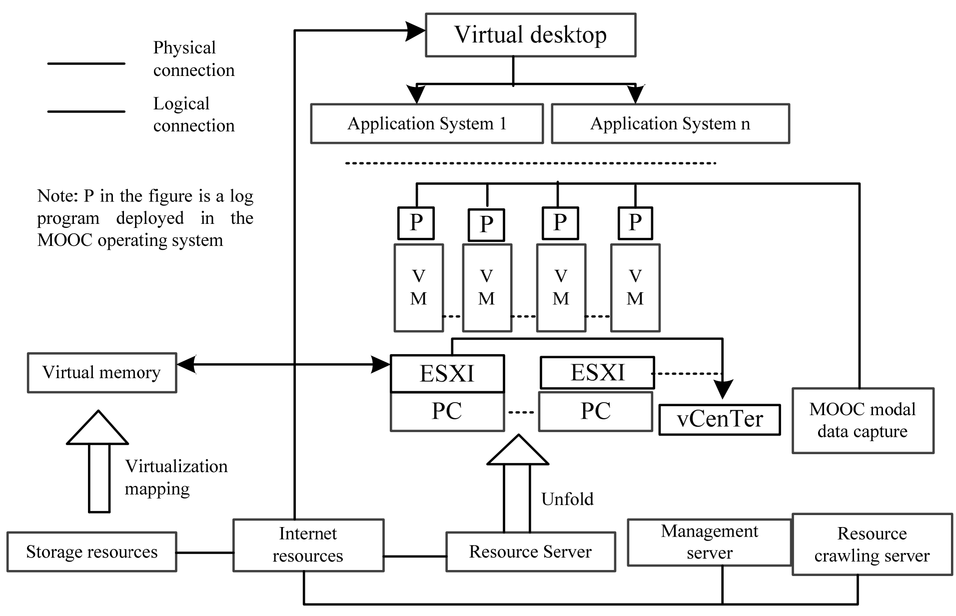
2 Design of deep mining method for MOOC multimodal resources

In order to solve the problems of traditional MOOC multi-modal resource deep mining methods, and make the new method have the characteristics of high recall and precision, as well as low mining time, this paper proposes a MOOC multi-modal resource deep mining method based on improved Eclat algorithm. According to the MOOC resource pool structure, the MOOC multi-modal knowledge map is constructed, and the hash chain is used to analyse the knowledge map Attribute connection rules between spectra. The improved Eclat algorithm is used to transform the captured resource modal information to get the final mining result.

2.1 Generation of MOOC multimodal knowledge map

MOOC course is a multi-modal system composed of video, voice, image and text, which includes micro video (including video subtitles and handouts), test assignments, question discussion records and other modes of knowledge content expression. Multimodal knowledge mapping has relevance, structure and navigation, which can provide multimodal curriculum resources with knowledge entity as the core and non-linear organisation for MOOC curriculum reconstruction. MOOC curriculum reconstruction is a process of reconstructing the organisational structure of MOOC curriculum resources through the cross-platform retrieval and integration of multi-modal curriculum resources based on the knowledge entities and their relationships. Therefore, the generation of multimodal knowledge map is the key to MOOC curriculum reconstruction (Campbell et al., 2020).

Figure 1 MOOC resource pool structure



Supported by cloud computing technology, data mining takes MOOC resources as data objects, and uses computing integration, storage integration and network integration as data mining forms of MOOC resource pool, and allocates them according to different business and configuration requirements.

VMwarevSphere virtual software provides a good basic function for resource utilisation of MOOC and resource mining of MOOC. On this basis, the combined medium, high and low-end MOOC resources can be flexibly divided into modes and applied to storage, computing and network.

Implicit and complex intelligent data analysis resources must capture data and create source code set through different methods. It is necessary to install probes in computing servers, storage servers and network switches to complete data capture activities, such as deploying SDK host operating system virtualisation software to obtain the log information of MOOC resource running status, and using SDN technology to capture MOOC network resource database and storage virtualisation integrated management tool to complete the capture of MOOC resource operation mode (Pan et al., 2019; Lee and Kong 2016). Analyse API performance analysis system application, needle data source, research and development, realise virtual integration (integration of integrated computing, storage and network integration), connect information resources with virtual host, network equipment and other equipment and store them.

The steps of multi-modal knowledge map generation for MOOC curriculum reconstruction are as follows:

1 A directed graph $L = \{L_i\}$, $t = 1,2,3, \dots, w$ is constructed to store the multi-modal knowledge map. Node L_t contains a three-level knowledge entity $[L_{t1}, L_{t2}, L_{t3}]$ and its hierarchical relationship attribute $[a_t, b_t, c_t]$ and word vector $[v_{t1}, v_{t2}, v_{t3}]$.

2 A course C_i is extracted from C, which is composed of n MOOC courses C_i, and then a knowledge entity K_j $[k_{j1}, k_{j2}, k_{j3}$ is extracted from C_i. Its hierarchical relationship attribute is $[a_j, b_j, c_j]$, and the word vector of each level of knowledge entity is divided into $[v_{j1}, v_{j2}, v_{j3}]$.

3 If there is a node with $v_{t1} = v_{j1}$ in the child node of the starting node, perform step 4; otherwise, insert K_j as the child node of the starting node into L, and keep the hierarchical relationship attributes of the inserted secondary knowledge entity in a non-decreasing order (Yang et al., 2018).

4 If there is a secondary node satisfying $v_{t2} = v_{j2}$, then step 5 is executed; otherwise, K_j is inserted into L with L_{t1} as the child node, and the third level relationship attributes of the inserted L_{t1} child knowledge entity are kept in non-decreasing order.

5 If $v_{t3} = v_{j3}$, the curriculum resources corresponding to K_j will be integrated into the knowledge entity attributes of L_t leaf nodes; otherwise, the knowledge entity K_j will be inserted into L, and the third level relationship attributes of the child knowledge entities after L_{t2} will remain non decreasing (Xiao et al., 2020).

6 Iteratively perform steps 2, 3, 4 and 5 until all knowledge entities in the course are added to the multimodal knowledge map.

To sum up, this paper mainly takes cloud computing technology as the support, takes MOOC resources as data objects, constructs MOOC multi-modal knowledge map according to MOOC resource pool structure, and uses hash chain to analyse the attribute connection rules between knowledge maps. The next step is to use the improved Eclat algorithm to design data mining process and realise data mining based on attribute connection rules.

2.2 Deep mining of MOOC multimodal resources based on improved Eclat algorithm

On the basis of the above, the basic principle of Eclat algorithm is analysed, and the hash chain is used to analyse the attribute connection rules between knowledge maps. The improved Eclat algorithm with vertical format is constructed to reduce the operation process and improve the data redundancy. The improved Eclat algorithm obtains the parameter list, judges the storage mode structure of MOOC multimodal resource data, and realises the mining of MOOC multimodal resource data dig.

2.2.1 Improvement of Eclat algorithm

The improved Eclat algorithm data set consists of Item Item and TIDSet, a set of transaction ID. The algorithm uses intersection operation to get the candidate item set, and calculates the support degree and the minimum support degree to get the frequent item set. The Eclat algorithm in this paper is divided into three modules, namely, data transformation, model combination and modal mining (Gen et al., 2020).

According to the characteristics of traditional Eclat algorithm, combined with the feature of relational databases is proposed a new method for multidimensional association mining, increase IO overhead algorithms don't have to scan the database many times, also need not frequent pattern tree structure, at the same time, will not be lost strong association rules mined commonly used algorithm, effectively improve the efficiency of mining algorithm. For each dimension, Eclat algorithm is used to find out all 1-frequent item sets L1, and then join operation is carried out to produce k-frequent item sets Lk. The algorithm generates the candidate set Ck of k through the connection between LK-1 dimensions. For each k candidate I $\in$ Ck, check whether its support is greater than the minimum support, and put the one that meets the requirements into Lk.

The improved Eclat algorithm directly performs breadth first search and cross count on the data set represented vertically to calculate the support degree of the candidate item set (Ahmed and Ulah 2018a; Cheng et al., 2020). Meanwhile, the items in the frequency set and frequency set are sorted in lexicographical order. Its structure is shown in Figure 2.

As can be seen from Figure 2, the initial data stores model data of each device in the time series. Through the conversion module, the master data object is in vertical format, and the original modal data in horizontal format is converted into the data in vertical format. The model combination module is mainly combined with the model knowledge analysis storage structure to reduce the number of redundant connections. This structure preserves the raw data and helps generate intelligent data analysis results (Ahmed and Ulah 2018b; Yun et al., 2017). At the same time, the two elements of infrequent combination are deleted in advance, so as to reduce the computational flow of infrequent combination of associative set analysis. Model mining modules are mainly divided into high-high mode mining and low-high mode mining. The algorithm adopts continuous iteration method to obtain all the association rules satisfying the conditions.

The improved Eclat algorithm is different from Apriori algorithm and other data formats. The vertical data format reduces the requirement of complex computation and improves the retrieval speed of rules between knowledge graphs. Figure 3 shows the transformation process (Acosta et al., 2016).

Figure 2 Improved Eclat algorithm structure

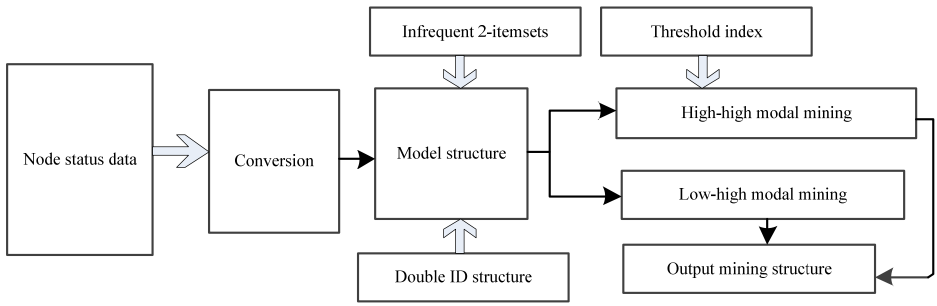


Figure 3 Data format conversion

Time	Node	State type
	IP1	S1,s2,...sm
	IP2	S1,s2,...sm
	...	...
	IPn	S1,s2,...sm
2	...	...
...	...	...
t	...	...

Data conversion

IP/Time	1	2	...	t
IP1	S1	S1	...	S1
	S2	S2	...	S2
	...	...	...	...
	Sm	Sm	...	Sm
IP2	...	...	...	...
...	...	...	...	...
IPn	...	...	...	...

The improved Eclat algorithm uses infrequent two-item sets to cut resources, and the retention of infrequent two-item sets is helpful to reduce the operation of comparing resource sets and improve the efficiency of the algorithm. This article uses a double-hash linked list to show an infrequent two-item set of stored procedures, whose structure is shown in Figure 4.

In the first knowledge graph IP, *key*1 stores the first infrequent two-item set of the knowledge graph, and *model*1 in the hash linked list is the second knowledge graph IP storage result of the infrequent two-item set. In structure *model*2, a storage structure parameter identifier is mapped to each knowledge map; Determine the subbendex structure storage parameters, the IP parameter of the first knowledge graph is firstsubid, and the IP parameter of the second knowledge graph is secsubid (Li et al., 2020).

Therefore, the connection between 3 item sets and more than 3 item sets requires the mining of the relationship between their knowledge maps and their recall rate. For example, in groups *A* and *B*, the value of the last parameter is <*IP*1, s1> and <*IP*2, *s*2>, respectively. Then its connection judgment condition is:

1 Hash linked list location $key1 = IP1$ in the first layer of the connection.

2 After obtaining model1, look for the related symbol key= $IP2$

3 After obtaining the parameter list, determine whether $secs1firstsubID$ and $s2$ are equal (Damevski et al., 2017).

Figure 7 shows the data storage modal structure defined in this article. The identifier of the knowledge map is stored in structure *ID*, namely main*ID*, while the module number is *subID*. The *ID* groups are defined in the node structure and are mainly used to store the storage connection in the algorithm iterative knowledge map. The resource storage modal data is located on the model position *t* (Sandhu et al., 2018; Karan and Samadder, 2018). The cloud structure also defines the identifier category of the graph. The flag sets of knowledge maps between different modes basically exclude the evaluation parameters of the same knowledge map and reduce the redundancy of connection information.

Figure 4 Infrequent 2-item set stored procedures

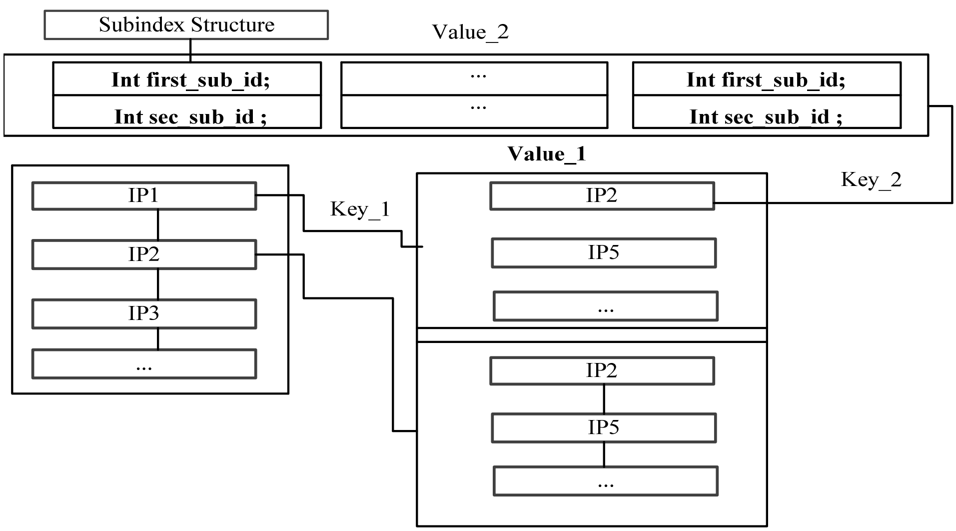
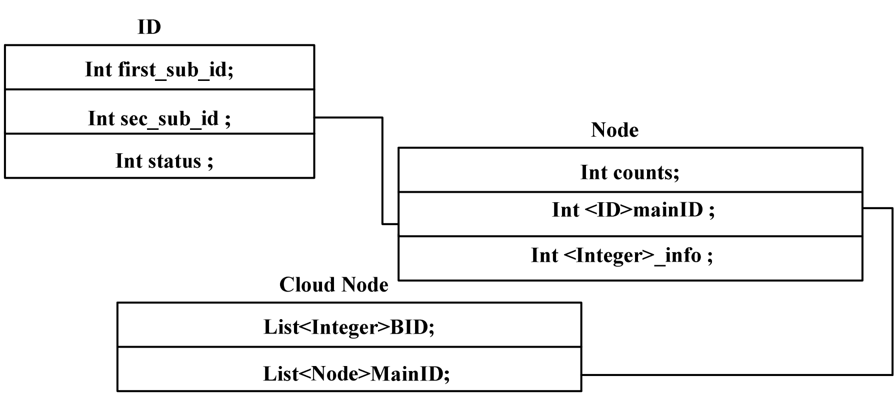


Figure 5 Resource storage modal structure



The improved algorithm does not need to scan the database for many times to increase IO overhead, nor need to construct frequent pattern tree, and at the same time will not lose the strong association rules mined by common algorithms. It can judge that the result of this operation is not frequent items before the intersection operation, so it can directly skip this intersection operation and improve the efficiency of the algorithm.

Through the above process, the Eclat algorithm is improved, in the next step, we need to use the improved Eclat algorithm for MOOC multimodal resource deep mining.

2.2.2 Implementation process of MOOC multi-mode resource deep mining

Because different physical devices or virtual hosts run under different conditions, data modal parameters do not match, so the first need to determine the evaluation criteria for a variety of modal resources. The proportion of low-level, ordinary and high-level resources in historical multi-resource data is calculated by modal separation method, and all modal information is converted into a data model that can be processed (Chou et al., 2017; Rashid et al., 2016).

For cloud computing conditions $P = \{IP1, IP2, IP3, \ldots, IPn\}$, P represents the set of all knowledge maps in MOOC resources under cloud computing conditions. The knowledge map in P includes virtual host, network device, storage server and so on. This paper uses its IP address as its unique symbol. Data set $D = \{D1, D2, D3, \ldots, Dt\}$, t represents timestamp, and represents data set to capture t timestamp modal resource information sets, $t \in [1, T]$.

$IPi = \{s1, s2, s3, \ldots, sm\}$, where KG represents the parameter modes existing in the knowledge graph, such as CPU mode, memory mode, virtual network card port mode, etc., where $G \in [1, m]$.

The original MOOC data of MOOC resources are numerical and the data parameters are different. For example, the CPU's data parameter is the utilisation, and the network mapping port's model parameter is the traffic rate. In the case of cloud computing, numerical data needs to be converted to modal bit data s, where $S = \{mode1, mode2, mode3\}$. At the same time, even among different knowledge graphs, the modal target information of numerical data also has different meanings according to different generation conditions.

Suppose that IP1 and IP2 both represent a virtual host and they have the same set of modes $\{s1, s2, s3, s4\}$. Although the modal classes are the same, they operate differently in the actual generation environment, and the meaning of continuous modal information is different. Therefore, the steps of digital data conversion in data processing are as follows:

The numerical distribution of $maxvj$ and $minvj$ of the knowledge map modal information in the time period of $t, j \in [1, m]$, j and t are obtained, and is used to mark the maximum and minimum values of the modes. According to the distribution of resources, the numerical data is converted into a set of $V = \{mode1, mode2, mode3\}$ modes.

In the cloud computing environment, there are association rules between the characteristic constants of knowledge graph parameters and those of other knowledge graph parameters. The following will compare whether the two modal data can be divided into $mode3$ at time t under the same conditions. If the modes are the same, the recall rate will increase. The mining process of the higher-order modal data is as follows:

a Obtain resources and establish a new resource structure, with Cu representing the storage set and u representing the number of iterations.

b Capture the knowledge map set Au = {Cu–1, IPi}, Bu = {Cu–1, IPj}, where i < *j*.

c Judge whether the prefixes of Au and Bu are the same, that is, whether the data values of C1 and C × u1 are equal, then enter step d); If not, continue the link loop, returning to the initial step a).

d Connect the internal parameters of Au and Bu knowledge graphs successively to determine whether they are infrequent in the 2-item set. If it does not implement connections and supports statistics; If not, follow the e steps.

e If the modal parameters Au and Bu are equal to mode3 at time t, the precision increases. When the statistics are complete, check that the Au and Bu parameter sets are connected frequently. It is normally stored to Cu memory cells, otherwise the connection will continue.

f When u = u + 1, return step a).

The modal input position of obtaining a frequent set of one item at the time of t is captured. In this paper, when obtaining a frequent set of one item, the modal positions that mark the knowledge map parameters appearing most frequently at t times represent the parameter characteristics of the connection. Assuming that the modal bit mode1 of parameter s1 marked in IP1 appears most frequently, then in the next parameter connection, a parameter of another knowledge graph is connected with the parameter s1 in IP1, only the modal bit mode1 of parameter s1 needs to be referred.

To sum up, through the analysis of cloud computing environment mining MOOC stored data, the characteristics of network data, using hash chain to analyse the knowledge map between the properties of the connection rule, build a vertical format Eclat algorithm reduce data redundancy operation process improvement and transformation by fetching modal information resources, and according to the data mining process implementation MOOC resources in a variety of modal mining. The next step is to test the effectiveness of the method, so experimental test is needed.

3 Experimental research

3.1 Experimental scheme

In order to test the practical application effect of MOOC multi-modal resource deep mining method based on improved Eclat algorithm, experimental test is needed:

All the experiments in this paper were completed under the experimental conditions listed in Table 1 and 2. Table 1 shows the simulation of the experimental environment. Table 2 Experimental conditions for data conversion. It contains the experimental equipment information of MOOC resources under the cloud computing environment. The structure of MOOC resources. Figure 6 is the extension diagram of some MOOC resource structures, restoring the process of reading data from the background database of VMwarev Sphere control Centre.

Figure 6 MOOC resource topology (see online version for colours)

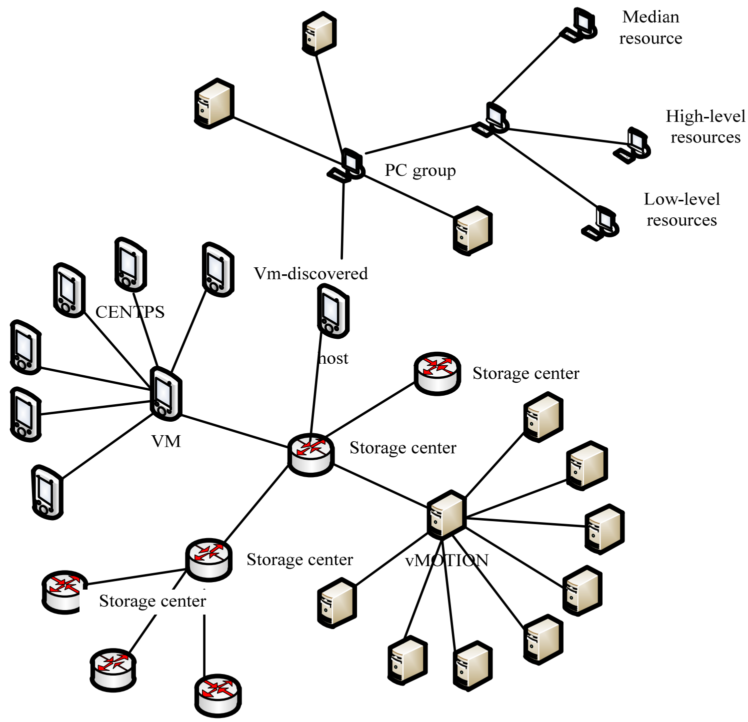


Table 1 Simulation table of experimental environment

Instrument configuration	*Experimental environment*
Cacti + centre on + Nagios	Visual monitoring
VMwarevSphere7.0	Virtual software
PostgreSQL	Original database
CPU4 × Intel® Memory 512 GB, SSD 5.6TB	Server

Table 2 Experimental conditions of data processing

Instrument configuration	*Experimental environment*
AMDA88640K3.60 GHZ	CPU
Windows7	PC system
Java	Programming language
8 GB	Memory
Eclipse	Programming software

The data used in the experiment is the actual running data set captured from the MOOC resource pool under cloud computing conditions, and 50000 original data sets with time stamp are selected. These original data sets come from the MOOC operation data of 6 months in March, 2019 and August, 2019, including 9 computing server modes, 8 storage server modes, and 15 modes for switching different switch port modes. Taking the receiving parameters of virtual host knowledge map as an example: CPU usage mhz, disk usage, mem usage, net usage, some experimental environment data are selected, as shown in Table 3.

Table 3 Experimental environment data

Netusage	*Diskusage*	*CPUusage*	*Entity*	*CPUusagemhz*	*Memusage*	*Time*
0	20	528	192.168.1.121	695	5,727	2019 9 13 10:00
152	26	2,632	192.168.131	3,478	5,744	2019 9 13 10:00
92	93	7,380	192.168.142	1,764	6,538	2019 9 13 10:00

3.2 Experimental methods and performance indicators

The methods of Cao et al. (2016), Zhiyong et al. (2017) and Yao et al. (2017) and this paper are selected for experimental test. The setting process of experimental performance index is as follows:

Association rule is a logic with A-mode and mode connection, A is called leader in association rule, *B* is called successor in association rule. $A \to B$ represents the assumption that *A* can grab, then *B* can also grab under certain conditions, among which *A* and *B* are the collection of MOOC resources, and recall ratio and error rate are generally used as evaluation criteria to measure the effect of data grabbing.

1 Recall ratio sup port ($A \to B$) represents the percentage of resources with $A \cup B$ resource sets in the MOOC resource pool in total database resources, that is:

$$sup\ port(A \cup B) = sup\ port(A \to B) = P(A \cup B) \tag{1}$$

where $A \cup B$ represents a resource covering A and B at the same time.

2 Precision ratio confidence($A \to B$) represents the proportion of the number of resources in the resource database covering $A \cup B$ in the number of resource sets covering *A*. The formula is as follows:

$$confidence(A \to B) = \frac{\text{sup}(A \cup \text{B})}{\text{sup}(A)} \tag{2}$$

3 Mining time: mining time refers to the time required to mine a certain number of MOOC resources.

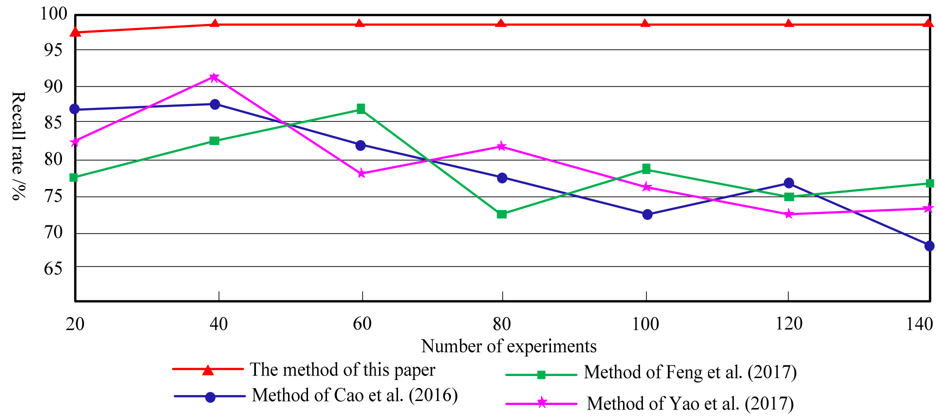
3.3 Performance test

3.3.1 Recall ratio

The recall rates of the four methods are compared, and the results are shown in Figure 7.

According to the analysis of Figure 7, the recall rate of Cao et al. (2016) varies from 73% to 87%, that of Zhiyong et al. (2017) varies from 72% to 87%, and that of Yao et al. (2017) varies from 78% to 92%. Compared with the literature method, the recall rate of this method is always above 97%, which indicates that the recall rate of this method is higher and the practical application effect is better.

Figure 7 Recall ratio comparison results (see online version for colours)

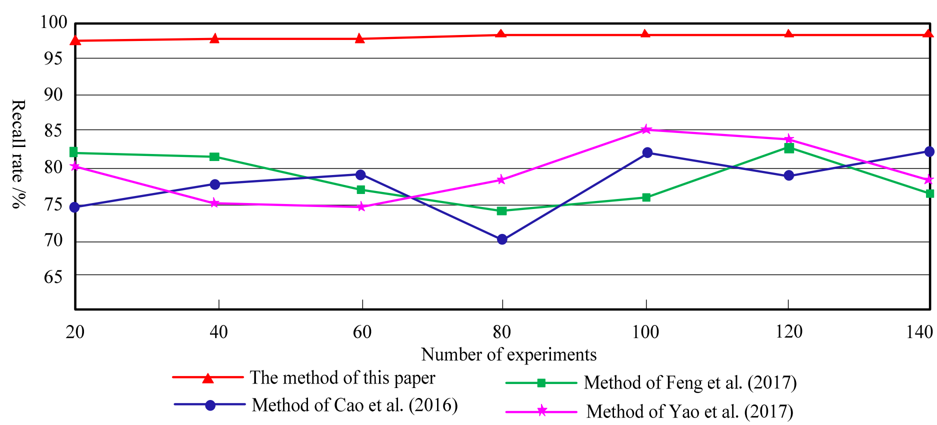


3.3.2 *Precision ratio*

On the basis of the above experiments, the precision of the four methods is analysed, and the results are shown in Figure 8.

As can be seen from Figure 8, the precision of Cao et al. (2016) varies from 70% to 82%, that of Zhiyong et al. (2017) varies from 74% to 82%, and that of Yao et al. (2017) varies from 75% to 85%. Compared with the literature method, the precision of this method is always above 97%, which indicates that the precision of this method is higher and the practical application effect is better.

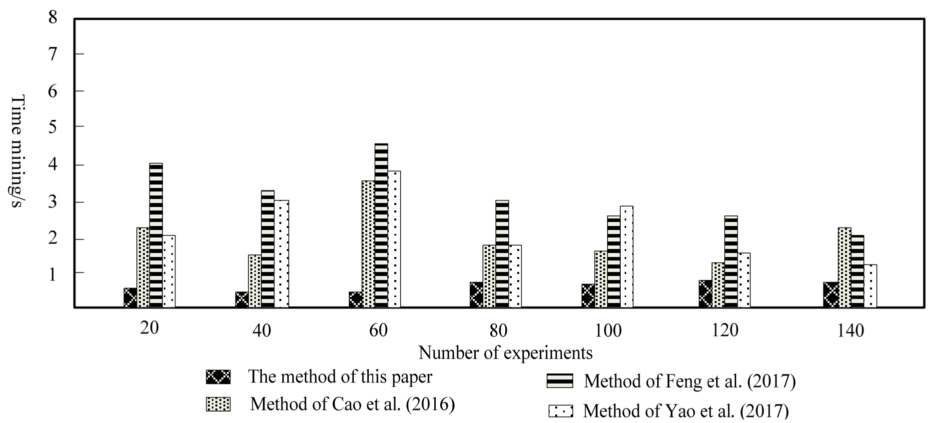
Figure 8 Precision ratio comparison results (see online version for colours)



3.3.3 Excavation time

Finally, in order to fully verify the application performance of the method, the mining time comparison test is carried out, and the results are shown in Figure 9.

Figure 9 Comparison results of mining time



Analysis of the above figure shows that the mining time of Cao et al. (2016) method changes in the range of 1.6 s–3.6 s, the mining time of Zhiyong et al. (2017) method changes in the range of 2.3 s–4.7 s, and the mining time of Yao et al. (2017) method changes in the range of 1.8 s–3.8 s. Compared with the literature method, the mining time of this method is always less than 0.7 s, the mining time is shorter, and the mining efficiency is higher.

4 Conclusions

Because the traditional method does not consider the running state of data, it cannot accurately mine the hidden rules between knowledge maps, and has the problem of low recall and precision. In this paper, an improved Eclat algorithm is designed in the cloud computing environment, which mines the characteristics of MOOC resources' stored data and network data, and uses hash chain to analyse the attribute connection between knowledge maps. According to the rules, the Eclat algorithm with vertical format is constructed to reduce the operation process, improve the data redundancy problem, convert the captured resource modal information, obtain the parameter list, and then judge the infrequent two item sets to judge the data storage mode structure, and complete the multi-modal mining of MOOC resources through the modal symbols in the structure. The experimental results show that the recall and precision of the method are above 97%, the mining effect is better, and the mining time is always less than 0.7s, the mining efficiency is higher, which can be further promoted in practice.

Acknowledgements

This project is supported by the online open course of Jiangsu Provincial Department of Education and Research on the reform of university teaching mode based on MOOC under the background of internet plus.

Reference

Acosta-Mendoza, N., Gago-Alonso, A., Ariel Carrasco-Ochoa J, Francisco Martinez-Trini D, Jose, Eladio Medina-Pagola, J. (2016) 'A new algorithm for approximate pattern mining in multi-graph collections', *Knowledge-Based Systems*, Vol. 109, No. 1, pp.198–207.

Ahmed, M. and Ullah, A.S.B. (2018a) 'Barkat, Ullah. Infrequent pattern mining in smart healthcare environment using data summarization', *Journal of Supercomputing*, Vol. 78, No. 2, pp.73–86.

Ahmed, M. and Ullah, A.S.B. (2018b) 'Infrequent pattern mining in smart healthcare environment using data summarization', *The Journal of Supercomputing*, Vol. 74, No. 10, pp.1–19.

Campbell. E.A., Bass, E.J. and Masino, A.J. (2020) 'Temporal condition pattern mining in large, sparse electronic health record data: a case study in characterizing pediatric asthma', *Journal of the American Medical Informatics Association*, Vol. 13, No. 2, pp.152–163.

Cao, L., Dong, X. and Zheng, Z. (2016) 'E-NSP: efficient negative sequential pattern mining', *Artificial Intelligence*, Vol. 12, No. 5, pp.1–12.

Cheng, Z., Trépanier, M. and Sun, L. (2020) 'Probabilistic model for destination inference and travel pattern mining from smart card data', *Transportation*, Vol. 68, No. 1, pp.490–506.

Chou, C.A., Bonates, T.O. and Lee, C. (2017) 'Wanpracha art chaovalitwongse. multi-pattern generation framework for logical analysis of data', *Annals of Operations Research*, Vol. 249, Nos. 1–2, pp.329–349.

Damevski, K., Shepherd, D.C., Schneider, J. and Pollock, L. (2017) 'Mining sequences of developer interactions in visual studio for usage smells', *IEEE Transactions on Software Engineering*, Vol. 43, No. 4, pp.1–10.

Gan, W., Chun-Wei Lin, J., Fournier-Viger, P. and Chao, H-C. (2020) 'HUOPM: high-utility occupancy pattern mining', *IEEE Transactions on Cybernetics*, Vol. 50, No. 3, pp.1195–1208.

Karan, S.K. and Samadder, S.R (2018) 'A comparison of different land-use classification techniques for accurate monitoring of degraded coal-mining areas', *Environmental Earth Sciences*, Vol. 77, No. 20, pp.713.1–713.15.

Lee, Y. and Kong, M. (2016) 'Analysis of tree-based uncertain frequent pattern mining techniques without pattern losses', *Journal of Supercomputing*, Vol. 86, No. 9, pp.35–42.

Li, Q., Zhao, L., Lee Y-C., Avesta, S. and Lin, J. (2020) 'CPM: a general feature dependency pattern mining framework for contrast multivariate time series', *Pattern Recognition*, Vol. 21, No. 1, pp.177–186.

Pan, L., Mu, S. and Wang, Y. (2019) 'User click fraud detection method based on Top-Rank- k frequent pattern mining', *International Journal of Modern Physics B*, 2019, 33(1, p.1950150.

Rashid, M.M., Amar, M., Gondal, I. and Kamruzzaman, J. (2016) 'A data mining approach for machine fault diagnosis based on associated frequency patterns', *Applied Intelligence*, Vol. 45, No. 3, pp.1–14.

Sandhu, R., Kaur, N., Sood, S.K. and Buyya, R. (2018) 'TDRM: tensor-based data representation and mining for healthcare data in cloud computing environments', *The Journal of Supercomputing*, Vol. 74, No. 2, pp.592–614.

Xiao, Z., Fu, X., Zhang, L. and Goh, R.S.M. (2020) Traffic pattern mining and forecasting technologies in maritime traffic service networks: a comprehensive survey', *IEEE Transactions on Intelligent Transportation Systems*, Vol. 21, No. 5, pp.1796–1825.

Yang, T., Ren, M. and Zhou, K. (2018) 'Identifying household electricity consumption patterns: a case study of Kunshan, China', *Renewable & Sustainable Energy Reviews*, Vol. 91, No. 1, pp.861–868.

Yao, X., Chen, L., Peng, L. and Chi, T. (2017) 'A co-location pattern-mining algorithm with a density-weighted distance threes holding consideration', *Information Ences*, Vol. 39, No. 6, pp.144–161.

Yun, U., Lee, G. and Yoon, E. (2017) 'Efficient high utility pattern mining for establishing manufacturing plans with sliding window control', *IEEE Transactions on Industrial Electronics*, vol. 29, No. 9, pp.1–12.

Zhiyong, F., Xi, L., Qixun, Z. and Wei, L. (2017) 'Proactive radio resource optimization with margin prediction: a data mining approach', *IEEE Transactions on Vehicular Technology*, Vol. 51, No. 10, pp.1–10.