

## **Analysis and application of the computer network information data mining based on clustering algorithm**

Zequan Shi, Ying Wang

Zequan\_Shi@yeah.net

Department of Information Engineering, Chongqing Institute of Engineering, 402260, China

**Keywords:** data mining; computer network information; clustering algorithm

**Abstract.** With the traditional algorithm to process computer network information data analysis, assuming the existence of diversity in multi-dimensional space, which will greatly reduce the efficiency of mining. For this reason, a data mining method based on clustering algorithm is proposed, to study computer network information data mining problems, the method can polymerize all computer network information data from up to down, and then classify the common data. The experimental results show that, with the proposed algorithm to process computer network information data mining, can access to the relationship of the computer network data from the low level concept to higher level concepts, which demonstrate the effectiveness of the proposed algorithm.

### **Introduction**

With the popularization of computer network information, for data mining, the most important task is to discover the same kind of information in the huge data system, and search the target data, with clustering algorithm can complete the computer network information data mining process described above[1,2]. Clustering algorithm is in accordance with some common similarity of the objective, they will be summed up as a class, and then data mining analysis is processed for the information network [3,4]. Study method and issues of cluster analysis is very extensive, such as: if clustering method can be utilized to analyze complex shapes of high difficulty, data type and high clustering technology and so on, and for the information network data mining analysis, MAXNET (maximum likelihood classification SOFM network) clustering is selected for computer network information data mining and analysis [5,6].

### **Analysis of information network data clustering hierarchy**

Method of computer data information mining is through the hierarchy analysis method to decompose data, and then create a hierarchy for a specific data object, according to the decomposition level to form the decomposed data structure. Computer network information data level can be divided into two types: from up to down and from down to up. The hierarchy analysis method from up to down, regard every computer network information data as a single individual, then fuse these individuals sequentially, until the merger meet clustering constraint conditions. Clustering method from down to up, on the contrary, all the computer network information data are regarded as a whole, decomposition is applied in sequence, so as to divide it into smaller overall.

The detailed analysis of computer network data clustering method is as follows:

(1) computer network information data clustering method from down to up, is to regard each computer network information data as an individual, and then integrate the individual above into a bigger group until all data is integrated into a whole data group, or stop when meet some specific requirements. During the process of the information network data mining and analysis, most of the clustering algorithms belong to this method, but the specific requirements are different.

(2) computer network information data clustering method from up to down, just opposite to it, it starts with a whole, and then decompose into smaller whole, until all objects are able to become a cluster separately, or terminate the decomposition when meet some requirements.

### Analysis principle of computer network information data based on clustering mining

In the process of network information data mining, MAXNET network need to be established, this network is mainly applicable to select a node of the highest data output, and then increase the output.

Computer network information data samples are saved in  $u_{ij}$  right through storage, for an arbitrary pattern of  $x$ , all of the characteristics  $x_i$  are input to each node, and these features are weighted and summed through the computer network, in the network, inputting a node from higher level can learn that, computer network information value of higher level is exactly equal to the value we input and the same value of corresponding connection right.

In the sub network, let  $t_{ik}$  be the connection right from  $j$ -th node to  $k$ -th node, take

$$t_{ik} = \begin{cases} 1 & j = k \\ -\varepsilon & j \neq k \end{cases} \quad (1)$$

$$\varepsilon < \frac{1}{M}, k, j = 1, 2, \dots, M \quad (2)$$

Among them,  $M$  represents the value may have classification in the class, that is the number of nodes in the network.

According to the output data of each node, sub network matches with the connection weights and be treated repeatedly, the output value of node  $j$  at time  $t+1$  depends on the output values of each node at  $t$  time, i.e.

$$y_j(t+1) = f_1 \left( y_j(t) - \varepsilon \sum_{k \neq j} y_k(t) \right) j, k = 1, 2, \dots, M \quad (3)$$

Wherein  $f_1(a) = a \quad a > 0$

$f_1(a) = 0 \quad a < 0$

Regardless of any  $y_j(t)$  will receive lateral data inhibitory of the its output, if the variables  $a$  of function  $f_1(a)$  is positive, then the output  $y_j(t+1)$  is positive; since all variables  $a$  will become negative node, and its output will be 0. This treatment method, will make output data of the node with the largest output values rise, and possess very small difference from the original output data, and the computer network information data will make their difference bigger, after a stage, network output nodes outside of a node, the output value become 0 soon, but that the only node without value 0, is node with the largest input value at the beginning, as it will stop the change along to the change stopping of network.

According to the above principle, a tissue clustering algorithm can be obtained, the procedure is as follows:

the  $x = |x_i|$  in the computer network information data is input to node, wherein,  $i, x_i = 0$  or  $1$ .

With from down to up of the hierarchy analysis method to process computer network information data analysis, the equation is as follow

$$y = \sum b_{ji} x_i \quad (4)$$

(3) Searching the node with maximum value  $y_i$  ;

(4) With from up to down of the hierarchy analysis method to verify if  $x$  is really belongs to the  $j$  class, that is, to calculate the weighted sum  $\sum_i t_{ji} x_i$ , if it meet the following conditions:

$$\frac{\sum_i t_{ji} x_i}{\|x\|} \rangle \rho \quad (5)$$

Stop operation

(5) The value of  $b_{ji}$  and  $t_{ji}$  is updated;

### Experiment results and analysis

In order to verify the effectiveness of computer network information data mining method based on clustering algorithm proposed in this paper, there is the need for an experiment.

The traditional algorithm and the proposed algorithm are adopted separately for computer network information data mining, mining precision rate can be expressed as follow:

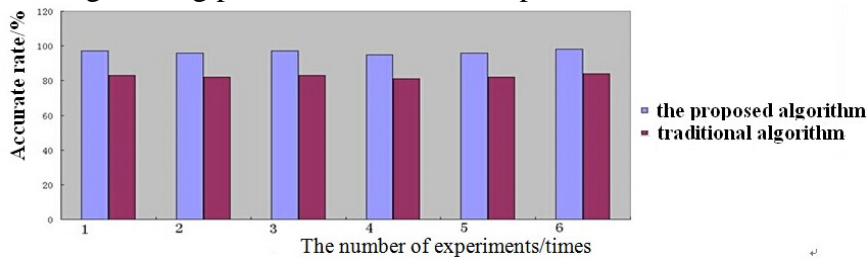


Figure 1 accuracy comparison graph of different algorithm

According to the above graph to arrange the table below:

Table 1 data table of accuracy comparison of different algorithms

The number of experiments (times)	Accuracy of the proposed algorithm (%)	Accuracy of the traditional algorithm (%)
1	97	83
2	96	82
3	97	83
4	95	81
5	96	82
6	98	84

Based on the above experimental contents, it can be learned that with the proposed algorithm to process computer network information data mining, can greatly improve the mining accuracy, fully demonstrate the advantages of the proposed algorithm.

### Conclusion

This paper presents a method of computer network information data mining based on clustering algorithm, which is used for computer network information data mining problems, the method can polymerize all computer network information data from up to down, and then classify the common data. Through experiments, this algorithm can discover the computer network data effectively, and compared with the traditional data mining algorithm, it is more accurate, effective, and has strong advantages.

### References

- [1] Zeng Yu, Xiao Renbin. Entropy optimization on scale-free networks' resilience [J].Journal of system engineering, 2013.143-150.
- [2] Li Jianchun, Wu Xueli, Han Bing, Li Jianyong. Scale-free network with robustness of attack. Journal of Henan University (Natural Science).2013.324-327.

- [3] Liu Qun, Li Sujian. Study on CNKI word similarity computation [J].Computational Linguistics and Language Processing.2002.2-76
- [4] Zhou Yafu, Ma Li, Dong Luobing. Design and implement of a system extracting keywords using SWN Theory [J]. Journal of Xi'an University of post and telecommunications. 82-85.
- [5] Liu Jingming, Han Lichuan, Hou Liwen. Cluster Analysis Based on Particle Swarm Optimization Algorithm [J]. Systems engineering-theory & practice.2005.54-58.
- [6] Li Zongfu, Deng Qiongbo, Li Huan. Research on kohonen SOFM neural network and its evolution [J]. COMPUTER ENGINEERING AND DESIGN.2004.1729-1730.