Study on the method of fast mining for intelligence data in a library database

Jie Yu

Library of Handan College, Handan Hebei 056005, China

Keywords: improved fuzzy genetic algorithm; data mining; intelligence data;

Abstract: The fast mining method for intelligence data in library database occupies an important position in the field of data processing. With traditional algorithm to dig intelligence, when encounter the interference of characteristic similarity, fuzzy rules are utilized to establish mining association rules, however, during the establishment of fuzzy rules, once characteristics are too close, a lot of constraints need to be added to establish rules, which leads to complicated rules and strong limitations, the calculation process is complicated, thus, the information data mining speed is reduced. For this, library database intelligence data mining method based on improved fuzzy genetic algorithm is proposed. Information characteristics model of intelligence data of library database is constructed, the update smoothing is conducted for data training samples under iterative genetic state, based on the rule of smaller square difference function value to update the central point of clusters, to obtain the power spectral density function of intelligence data as feature, with the improved fuzzy genetic algorithm to select intelligence data features, and compute fuzzy clustering center of intelligence data flow information focused on the multi space, associate the training set and the category to obtain the attribute set classification and information gain of intelligence data, so as to improve the mining performance for intelligence data in a library database. Simulation results show that intelligence data mining method based on the improved fuzzy genetic algorithm has high efficiency and precision.

1 Introduction

With the continuous development of database technology, usage scope of database is gradually expanded. How to obtain and search intelligence data from the library database, is the problem need to be solved [1,2,3]. In view of this problem, mining method for intelligence data in a library database is proposed [4,5,6]. At present, the mainstream of intelligence data mining methods in library database includes intelligence data mining method in library database based on genetic algorithm, intelligence data mining method in library database based on ant colony algorithm and intelligence data mining method in library database based on multi-level data analysis algorithm [7,8,9]. Among them, the most commonly used method is the intelligence data mining method in library database based on genetic algorithm [10]. As the intelligence data mining method in library database has a very broad space for developing in the data processing field, therefore, attracts a lot of attention from the experts, and has become a hot issue for expert to research, have great potential of development.

2 The principle of the method for mining intelligence data in library database

2.1 The relevant principles of improved fuzzy genetic algorithm

The improved fuzzy genetic algorithm is utilized for intelligence data mining in database, an algorithm is described as:

Fitness function and genetic operation of a population is designed, , including the crossover, mutation and convergence condition etc.. Assuming the intelligence data training set constituted by n samples in library database is :

$$L = \{(w_m, x_i) | i = 1, 2, \dots, n, m = 1, 2, \dots, c\} \quad (1)$$

(1) Assuming that the corresponding samples classes ω_m of c category is known, the variable

need to be solved is set as Q , X_i is solution of variable Q in the space, there is a function corresponding to each X_i , f_i is the fitness function of X_i , the calculated optimal solution is the value making maximum(or minimum) X_m and function value f_m .

The chromosome in improved fuzzy genetic algorithm is decomposed into three sub modules, first, the principle of PMX is utilized, with evenly random algorithm distribution to generate two cross points $sw(u)$, the region included by the cross points is setting as matching region, as:

$$\varepsilon_i(i, j) = \frac{\alpha_i(i) a_{ij} b_j(o_{t+1}) \beta_{t+1}(j)}{\sum_{i=1}^N \sum_{j=1}^N \alpha_i(i) a_{ij} b_j(o_{t+1}) \beta_{t+1}(j)} \quad (2)$$

For the fuzzy genetic iterative process, genetic execution operator data $S(t)$ is added as spread control variable of genetic variation:

$$\begin{aligned} r_i'(t) &= S(t) * p_{pi}(-t) \\ &= S(t) * p(-t) * h_i'(t) * h_i(-t) + n_{li}(t) \end{aligned} \quad (3)$$

:

$$\begin{aligned} n_{li}(t) &= S(t) * h_i'(t) * n_{pi}(-t) \\ &\quad + n_{si}(t) * p(-t) * h_i(-t) \\ &\quad + n_{si}(t) * n_{pi}(-t) \end{aligned} \quad (4)$$

In order to get dispersion coefficient of the genetic variation, it is defined as:

$$\begin{cases} u(t) = \frac{1}{\sqrt{T}} \text{rect}\left(\frac{t}{T}\right) \exp\{-j[2\pi K \ln(1 - \frac{t}{t_0})]\} \\ \text{rect}(t) = 1, \quad |t| \leq 1/2 \\ K = T f_{\max} f_{\min} / B \\ t_0 = f_0 T / B \end{cases} \quad (5)$$

The improved fuzzy genetic algorithm is utilized to calculate the fuzzy clustering center in multi space, j is the activator of i , then:

$$b_{ij}(p_j(t)) = a_{ij} \frac{(p_j(t) / \beta_j)^{H_j}}{1 + (p_j(t) / \beta_j)^{H_j}} \quad (6)$$

Associating the training set with the categories they belong to acquire attribute set classification of intelligence data, so as to obtain gain expression:

$$\begin{aligned} \text{Info}(B) &= - \sum_{i=1}^m p_i \times \log_2 p_i \\ \text{Info}_A(B) &= \sum_{j=1}^v \frac{|B_j|}{|B|} \times \text{Info}(B_j) \\ \text{Gain}(A) &= \text{Info}(B) - \text{Info}_A(B) \end{aligned} \quad (7)$$

It can be seen from the above formula that improved fuzzy genetic algorithm used for feature mining of intelligence data, can change random variation into transposition operation, two alleles are exchanged, data chromosome can skip repair process after crossing, the selection gene and semi connected gene of the copy have adopted the uniform crossover algorithm, which is beneficial to improve the performance of the intelligence data mining.

2.2 The establishment of the intelligence data mining model

Based on improved fuzzy genetic algorithm to construct library database intelligence data mining model, according to the above method to acquire a lot of library data as input, the input intelligence data is $w_1(t) \cdots w_N(t)$, through a genetic fitness function to calculate fitness of each feature value in groups, the size of population M , the maximum iteration N , the crossover probability P_c and mutation probability P_m , are selected, based on the generalized likelihood ratio test (GLRT) method, the maximum likelihood estimation $\hat{s}_{0ML}$, $\hat{\tau}_{0ML}$ of S_0 , τ_0 are used to detect, i.e.:

$$\begin{aligned}
l_1(r) &= \int r(t) \sqrt{\hat{s}_{0ML}} f^*(\hat{s}_{0ML}(t - \hat{\tau}_{0ML})) dt \\
&= \max_{s, \tau} \left| \int r(t) \sqrt{s} f^*(s(t - \tau)) dt \right| \\
&= \max_{a, b} \left| W_f r(a, b) \right| \begin{matrix} > \\ < \end{matrix} \lambda_i
\end{aligned} \tag{8}$$

Setting the extraction ratio parameter of each sample as BL , and class number of all sample are acquired. If the membership degree of $j(j \in N)$ class which $i(i \in p)$ samples corresponding to is largest, the tap coefficients $\tilde{g}_1(t) \cdots \tilde{g}_n(t)$ of library database intelligence data mining is obtained. The difference between maximum membership degree F_{ij} of samples and $E_j(j=1, 2, \dots, N)$ is calculated, according to the value of difference to do ascending order, pervious $BL \times K_j(j=1, 2, \dots, N)$ samples are extracted as training samples of the categories. As the intelligence data feature have an large impact on LSSVM performance, less influence on other features, so normalization is needed to eliminate the negative impacts, in particular as:

$$x'_i = \frac{x_i - \bar{x}}{x_{std}}, i = 1, 2, \dots, n \tag{9}$$

In the formula, n represents the number of training samples, x_{std} represents standard deviation.

Finally, using the improved fuzzy genetic algorithm to identify intelligence data in library database, the training process of improved fuzzy genetic algorithm for intelligence data in library database as shown in figure 1.

Table 1 Comparison of performance of different genetic algorithm

Algorithm	Evenly distributed iteration from stability to overall optimal point	Maximum iteration from stability to overall optimal point	Minimum iteration from stability to overall optimal point	Number from stability to partial best point of all
Standard genetic algorithm	278	397	240	5
Fuzzy big mutation operation	199	264	138	1
Adaptive genetic algorithm	165	198	144	0
Improved fuzzy genetic algorithm	134	139	122	0

It can be seen from table 1, the improved fuzzy genetic algorithm in independent 30 times operation, not only can get the optimal solution, and have faster convergence speed than other algorithms, and strong stability.

The XOR problem is a typical linear inseparable problem, also called the decision problem of parity. Considering a relatively simple three bit XOR problem, that is, the input three bit binary has odd numbers of "1", "1" is output; otherwise, the output is "0". The traditional genetic algorithm and improved fuzzy genetic algorithm are respectively adopted for simulation, then the intelligence data mining error comparison results as shown in Figure 2.

Figure 2 Comparison of mining error curve of different algorithms

It can be known from the above experiments that the improved fuzzy genetic algorithm have strong convergence and high accuracy in the field of intelligence data mining.

4 Conclusion

Considering the problems of the traditional algorithm, like large limitation and low accuracy, library database intelligence data mining method based on improved fuzzy genetic algorithm is proposed in the paper. Information characteristics model of intelligence data of library database is constructed, the update smoothing is conducted for data training samples under iterative genetic state, based on the rule of smaller square difference function value to update the central point of clusters, to obtain the power spectral density function of intelligence data as feature, with the improved fuzzy genetic algorithm to select intelligence data features, and compute fuzzy clustering center of intelligence data flow information focused on the multi space, associate the training set and the category to obtain the attribute set classification and information gain of intelligence data, so as to improve the mining performance for intelligence data in a library database. Simulation results show that intelligence data mining method based on the improved fuzzy genetic algorithm has high efficiency and precision.

Acknowledgement

In 2014 in hebei province development of social science research subject research subject of the people's livelihood "based on the path of university library building public culture service system research" (201401933).

References

- [1] Chen Rongxin. Data mining and application of R software [J]. Journal of Industrial and Commercial University of Chongqing: Natural Science Edition, 2011.6:602-607.
- [2] Zheng Hong, Lu Yang, Xu Zhaonong. The implementation of intrusion detection system classifier based on BP neural network [J]. Journal of HeFei University of Technology, 2003, 26 (2): 150-155.
- [3] Liu Mingjing. Data mining [J]. Southern China finance computer, 2001.4:31-33.
- [4] Yao Qing, Qu Zhiqiang, Qiao Jing. Analysis of the data mining [J]. Science and technology information, 2010.24:I0211-I0212.
- [5] Li Hai Yan, Peng Shimi. By using genetic neural network to research the diagenetic reservoir facies of low permeability [J]. Oil and gas geology, 2006, 27 (1):111-117
- [6] Wang Anhui, Yu Shuying, Zhang Yingkui, et al. the application of nerve network in diagenetic reservoir facies of the low permeability [J]. Oil and gas geology, 2004, 25 (3): 338 – 343
- [7] Jin Jin. Data mining technology [J]. China science and technology review, 2010.20:311.
- [8] Shi Tiefeng. Application of support vector machine in e-mail classification [J]. Computer simulation, 2011, 28 (8): 156 - 158
- [9]Lu Yunyan. Data mining technology [J]. Journal of Chongqing Education College, 2006.3:44-47.
- [10]Zhang Jianhong. Data mining of electronic commerce website [J]. Journal of Huzhou Vocational and Technical College, 20051:86-88.