Research Journal of Applied Sciences, Engineering and Technology 4(18): 3533-3538, 2012
ISSN: 2040-7467

Submitted: April 16, 2012 Accepted: May 18, 2012 Published: September 15, 2012

Creating Time-Varying Fuzzy Control Rules Based on Data Mining

Zhongkui Sun, Zhang Hao, Chen Lei and Wang Lili
College of Qinggong, Hebei United University, Tangshan 063000, China

Abstract: In this study, we propose the method creating fuzzy control rules based on experiential knowledge. With the advent of data mining techniques, a new way is provided to create fuzzy control rules. With the introduction of incremental data mining techniques, fuzzy control rules can be updated dynamically in time on the condition that massive new data is producing.

Keywords: Data mining, fuzzy control rules, incremental data mining

INTRODUCTION

With the development of the theory of control, there have been lots of researches about combination of advanced intelligent control strategies of neural network control, fuzzy control and expert control, which have already been applied to practical production control process. During the control process, the relationships of all kinds of data attributes tend to have fuzzy attributes. Fuzzy language variables expressed by fuzzy sets more accord with the objective things and can deal with some uncertain and incomplete information. As an important branch of modern intelligent control theory, the practicability caused by the fuzzy control theory is highly control rules (namely control rules table), so set up the fuzzy control rules and then the method of fuzzy logic reasoning based on rules has become one of i hotspots of intelligent control. These rules are used to imitate experts' association reasoning during control process. The fuzzy control rules are generated based on experiential knowledge, rather than on system structure or behavioral procedure. Therefore, the fuzzy control system which is based on rules exits problems as follows (Lai and Chen, 2005; Zhigang and Bo, 2007):

In the first, the bottleneck problem is knowledge acquisition. Processing power of this kind of system mainly depends on its knowledge quantity and quality. At present, the main way to obtain these knowledge is to summarize the experts' experience and then to change the knowledge into the rules which the computer can identify. In fact, this procedure is an inductive process. It is a very complicated and interactive process with randomness and personality. This kind of knowledge obtains indirectly. It is not only time-consuming and laboursome, but also inefficiency.

In the second, the generation knowledge base is static. The system is lack of ability of automatic study and automatic perfect. It is not easy to get and expand new knowledge from new instances during the operation process. As the nonlinear time-varying controlled object, information of new batch data more reflects the current system state than historical data. So the new data in the dynamic clustering process should have the corresponding weights to reveal its special function and realize the nonlinear and time-varying system dynamic renewal.

On the other hand, whereas the burgeoning data mining technology, more and more fields have been adopting this technology to extract previously unknown knowledge from the mass data, which achieves good effect and provides great help for reight decisions. The fuzzy control rules depend on all kinds of language variables and clustering theory in the data mining can realize automatic generation language variables based on data features, which can make full use of the accumulation of historical data and the current data combining with the production process requirement, find potentially valuable information and rules of data (Li and Sun, 2006). Finally, we can get the fuzzy control rules and propose new solutions for complex industry process control system with a new perspective:

- The way breaks the bottleneck problem of knowledge acquisition, which is no longer limited to knowledge of experts in the field, also reduces the subjectivity and random of the control rules.
- The knowledge base can be generated dynamically. With the application of incremental data mining technology, the knowledge base realize the ability of self-learning and self-improvement, even the system can get and expand new knowledge base from new instances during the operation process, also realize dynamic renewal of rules (Zhigang and Bo, 2007).

Corresponding Author: Zhongkui Sun, College of Qinggong, Hebei United University, Tangshan 063000, China

This study proposes the method creating fuzzy control rules based on experiential knowledge. A new way is provided to create fuzzy control rules. With the introduction of incremental data mining techniques, fuzzy control rules can be updated dynamically in time on the condition that massive new data is producing.

CREATING STATIC FUZZY CONTROL RULES BASED ON K-MEANS CLUSTERING ALGORITHM

Clustering is an unsupervised classification and its input is a set of unclassified records. Clustering can reasonably divide record collection in accordance with the similarity of each sample and determine which category each record should be in, which makes the property of internal data more nearer and the data similar distance among each class larger. Because the automatic classification method has "no supervision" property, so that the clustering method gets a great application in huge amounts of information processing, the genetic data analysis, the field of pattern recognition. K-means is a commonly used clustering method based on division, which achieves good results about many applications.

K-means clustering algorithm belongs to one of clustering analysis methods, which is a basic and most widely used partitioning method and can find classes and the center of the classes in data without class label. The basic algorithmic idea: given a database which contains N numbers data objects and the number of generating class "k". Then k numbers data objects are randomly selected to be regarded as initial clustering centers and then calculate the distance between the remaining samples and each cluster center. A sample is classified to the clustering whose center point distances the sample recently, then recalculate a new clustering center with average value method. If the center values of adjacent twice clustering are no change or a change less than one given value, which indicates the end of adjusting samples and the converge of clustering average error criterion function E Liu and Wen (2005). The criterion function E is determined by the following formula):

$$\mathrm{E} = \sum_{j=1}^{k} \sum_{i \in C_j} \left| i_l - w_j \right|^2 \tag{1}$$

where,

i_l : Clustering object

w_j : The average of objects in Clustering C_j

The average of each clustering can be expressed as:

$$w_j = \sum_{i_l \in C_j} i_l / |C_j| \tag{2}$$

where,

$|C_j|$: The number of objects in clustering C_j

The pressure of gas-collecting pipe is affected by multiple factors in coking and gas plant factory. Some great influencing factors to gas-collecting pipe pressure are gained through analyzing various factors. In this study, data of gas-collecting pipe pressure and various influencing factors is adopted from two coke oven in certain coking and Gas Plant factory. Through building a mechanism model, the gas-collecting pipe pressure system can be simplified as a two input four output variables fuzzy control system. Take the fuzzy control rules about the NO.1 gas-collecting pipe pressure P1 and NO.1 butterfly valve K1 as an example, we will introduces the whole process of creating fuzzy control rules based on data mining in this study (Sun *et al.*, 2010; Ian and Eibe, 2000).

Data preparation: The original collected data often has the following problems:

- The collected data comes from multiple actual systems and has heterogeneous problem.
- Because the various properties of variable samples usually use different measurement units, the observations may have great difference. Such the variables which have bigger absolute value may hide the smaller one and the function of latter one can not reflect.
- As there are the noise data and the irrelevant data hidden in the mass of the data. So the original collected data must have a data pretreatment processing for data mining.

After the collected NO.1 gas-collecting pipe pressure data and NO.1 butterfly valve opening data was pretreated, we can get data format as in shown in Table 1.

Table 1: Data format of being pretreated gas-collecting pipe pressure P1 and valve degree K1

P1	K1
-1.6098	0.0562
1.5829	-1.6060
⋮	⋮
0.8600	0.9351
-0.3402	-0.2713

Table 2: Data team of creating fuzzy controller

e_1 (t)	e_c (t)	u_1 (t)
-1.6098	1.1	0.0562
1.5829	2.7696	-1.6060
⋮	⋮	⋮
0.8600	1.6303	0.9351
-0.3402	-1.0441	-0.2713

After pretreated, we can only get gas-collecting pipe pressure value and butterfly valve opening value, that is to have the input variables e1 (t) and the output variables u1 (t), because fuzzy controller adopts the single variable two-dimensional pattern in this study, therefore also needs derivative value ec1 (t) of the gas-collecting pipe pressure. Here ec1 (t) is got by backward differentiation formula:

$$ec1\ (t) = e1\ (t) - e1\ (t-1)$$

So needed data has prepared ready for establishing fuzzy controller. As shown in Table 2.

Setting language values and its membership based on model: The data records (e1 (t), ec1 (t), u1 (t)) are clustered separately by K-means algorithm. In this study, the values ranges of language variables e1 (t), ec1 (t) and u1 (t) are in {PB, PM, PS, ZE, NS, NM, NB}, so we can set that the number of clustering model class is seven. The center value and range of each clustering can be determined after clustering. As shown in Table 3.

The center value of each variable model is one-to-one correspondence the values ranges of language variables {PB, PM, PS, ZE, NS, NM, NB} according to from big to small order and determines the corresponding membership functions. The membership functions of language variable are various. Here the triangular membership function is adopted. In every language variables of the membership functions, the center value of membership functions is set to “1”, the maximum and minimum value of the membership is set to “0”. The ranges of variation of each variable are set as the domain of discourse of the membership functions. Because their ranges of variation are not strictly symmetrical about "0" point, such as [-3.5, 3.9], [-4.7, 3.9], [-5.3, 5.1], so their domains of discourse are not strictly symmetrical about "0" point. So the domain of discourse of e1 is [-4, 4], the domain of discourse of ec1 is [-5, 4], the domain of discourse of u1 is [-6, 6].

There is not clear limit among each language variable fuzzy sets (linguistic value), that is to say, the membership functions must be overlapping in the membership functions of fuzzy set. Because we generate each language variables with clustering analysis method, according to the characteristic of clustering: the objects in a same class have the most similarity, but the objects in different classes have the less. So we can know that there is only a small overlap among the membership functions of each language variable, which are got by clustering method.

Clustering is unsupervised classification, the initial clustering center value is randomly generated from all the data and, in general, the each measured variable is normal in normal distribution, so the membership function of each fuzzy variable is the uniform asymmetric in the whole domain and it is dense nearby “0” and sparse around to extend.

Creating fuzzy control rules: For each language variables performs 7 models with clustering analysis,

Table 3: Center value and the range of every pattern

(a) $e_1(t)$							
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Center value	-1.238	-0.535	0.006	3.105	0.660	1.654	-2.310
Max value	-0.894	-0.269	0.329	3.849	1.127	2.340	-1.797
Min value	-1.755	-0.886	-0.264	2.387	0.334	1.157	-3.490
(b) $ec_1(t)$							
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Center value	2.577	-0.435	1.214	0.536	0.065	-1.317	-2.688
Max value	3.836	-0.187	1.875	0.861	0.300	-0.88	-2.012
Min value	1.907	-0.856	0.876	0.304	-0.184	-1.974	-4.623
(c) $u_1(t)$							
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Center value	3.3602	1.3488	0.5651	0.051	-0.496	-1.566	-4.093
Max value	5.0758	2.2835	0.9515	0.307	-0.222	-1.047	-3.263
Min value	2.3927	0.9843	0.3101	-0.217	-0.997	-2.731	-5.231

Table 4: Transformed data team format

e_1	ec_1	u_1
-2	2	0
¤2	3	-2
⋮	⋮	⋮
-2	-2	-1
1	2	1

Table 5: Classed and trimmed data team format

	e_1	ec_1	u_1
1st class	-3	-3	0
2nd class	-3	-2	0
3rd class	-3	-1	0
4th class	-3	0	-1
⋮	⋮	⋮	⋮
47th class	3	1	-1
48th class	3	2	0
49th class	3	3	1

and the center values of each languages variable are sorted in accordance with from big to small order and use the sequence of {+ 3 + 2, + 1, 0, -1, -2, 1} is used to replace, respectively every model from big to small order. The data format in Table 2 is converted as shows in Table 4.

As the data is shown in Table 4, two columns data e1, ec1 as the main attribute is classified and the data of e1 and ec1 which are equivalent simultaneously are grouped a class. For the dataset in every class, we can calculate the memberships of u1 and compare them, then find out the maximum membership degree u1 (i). To delete this kind of data group and only to keep the one that belongs to the maximum membership degree. The format of the sorted data is transformed as shown in Table 5.

According to the data in Table 5, the sequence of {+3, +2, +1, 0, -1, -2, -1} is one-to-one correspondence with the language variables sequence {PB, PM, PS, ZE, NS, NM, NB}, we can get a typical fuzzy control rules table, such as shown in Table 6.

DYNAIC BATCH INCREMENTAL FUZZY CLUSTERING ALGORITHM APPLICATION

In process of coking, mass now data is constantly increasing, the fuzzy control rules should be adjusted for adapting the new industry circumstances. Some new rules are obtained by clustering analysis to new collecting data. If the entire newly increased data is re-clustered, not only a waste of time, but also make the most memory-based algorithms cannot be achieved. So the re-clustering method is generally not used. Then the incremental clustering can be used, it can not only improve efficiency and can take advantage of the existing clustering results. The new clustering results can be incremental changed and improved in the base of the existing clustering results.

Table 6: P_1K_1 fuzzy control rules

e1/u1/ec1	PB	PM	PZ	ZE	NS	NM	NB
PB	PB	PB	PM	PS	ZE	ZE	ZE
PM	PB	PM	PS	PS	PS	ZE	ZE
PS	PM	PS	PS	ZE	ZE	NS	NS
ZE	PS	PS	ZE	NS	NS	NM	NB
NS	PS	ZE	NS	NM	NB	NB	NB
NM	PS	ZE	NS	NM	NB	NB	NB
NB	ZE	ZE	NM	NM	NB	NB	NB

In general, the information in a database is collected and updated in a regular interval, batch way, so the batch mode is used to update of the fuzzy control rules. In this study, a based-on parathion batch incremental clustering algorithm is used on the basis on the method of dynamic K-means algorithm. This algorithm also solves a trouble of requiring all data are loaded into memory in partitioning algorithm, limiting its drawbacks of large-scale data application.

The basic idea of the algorithm is that, at first, a batch new increasing data is clustered and then the clustering results are integrated into the original clustering results based on a restraint of distance and ultimately achieve the goal of incremental clustering.

The main process of the algorithm can be described as:

Step 1: The original data is clustered using the K-means algorithm, then calculate each class cluster center c_i (i = 1, 2, ..., k) and the cluster centers are looked as the initial center point; set recently distance threshold d_{min} and the minimum number of threshold ε.

Step 2: A batch of new gained data is clustered and generate Q piece of new classes. The specific steps are as follows: the new increasing data objects are used one by one to calculate the distance between the data and all kinds of centers using the formula (3):

$$d_{ij} = \left(\sum_{k=1}^{p} \left| x_i - x_{jk} \right|^2\right)^{1/2} \tag{3}$$

Finding the smallest distance $d_{ij\min}$, if $d_{ij\min}$ is less than the set threshold $d_{\min}$, the data will

be add into the nearest class and modify the center of the class; otherwise, the data is looked as a new class and as the beginning central point of the new class. And then the next data is used to do it until all objects in this batch of data are classified.

Step 3: After generating the new clustering results, calculate successively the distance between the new data center point and the original type of the center, which is set as d_{ij} ($i \le Q$, $j \le k$). the minimum distance between the center of new set and the center of each original class is set as $\min d_i$, $\min d_i = \min\{d_{i1}, d_{i2}, \cdots, d_{ik}\}$ Comparing the distance between the center of all new data set and the center of original clustering results, the center of all new data set may produce the following four cases:

- One new clustering is fell into a original clustering. If the smallest distance of one new clustering $d_{ij\min} < d_{\min}$, then the new clustering is fell into the original clustering and re-calculate the center of the class.
- One new clustering is looked as noise. If the smallest distance of one new clustering $d_{ij\min} > d_{\min}$, but the number of objects in such clustering is less than the certain threshold ε, then such as is the noise, it may be caused by the sudden change in certain conditions.
- Creating a new cluster. If the smallest distance of a certain clustering $d_{ij\min} > d_{\min}$, but the number of objects in such clustering is less than the certain threshold ε, then such as a new class and later be used to generate a new decision-making.
- Merging adjacent clustering. After disposing the very clustering of the new increasing data and re-calculating the center of the original clusters, which have been changed, we will calculate the distance between the various clustering. If the minimum distance is less than $d_{\min}$, then two corresponding classes are combined and then recalculate the center of the class, until the distance between the centers of all the classes are larger than $d_{\min}$.

Using the method provided in section II, new fuzzy control rules can be obtained by the clustering results, which are gained by dynamic batch incremental fuzzy clustering algorithm, thus completing the update of the fuzzy controller with the newly generated data.

EXPERIMENTAL ANALYSIS

In this study, a total of 10 million about collecting pipe pressure data of coke oven 1 and the 1st butterfly valve open data have been adopted, using K-means clustering algorithm for one-time clustering methods and bulk incremental fuzzy clustering method respectively, pretreatment of raw data clustering comparison. When using K-means clustering algorithm, the initial class number is set to 7. Bulk incremental clustering using K-means clustering algorithm for clustering the 70000 data, initialize the class number of N0 = 7, the results obtained in the initial cluster is of the history of clustering results volume increment data set and then the remaining data, bulk amount of data for 5000, initialize the class = 30, ε = 0.01. The clustering results are shown in Table 7.

From Table 7, the two clustering methods can be evidently shown as: two methods have good equivalence class vector central value one-time clustering results contained in the bulk incremental clustering method results corresponding to all existence; Bulk incremental clustering method has obvious dynamic. Collector pressure system is a nonlinear time-varying systems, the control model changes over time, so the bulk incremental clustering method to generate some of the historical data does not exist in the class center values, so that the clustering results to better reflect the current actual situation.

The simulation process is in MATLAB 6.5 platform. Because coke oven collector pressure system is a non-linear multi-variable real-varying systems, we have adopted the unit step response of the second order system transfer function

$$\frac{1}{5s^2 + 8s + 1}$$

Table 7: Bulk incremental fuzzy clustering and a one-time clustering method

Performance indicators	Curve 1	Curve 2	Curve 3
The maximum overs hoot σ_P	19.1%	15.1%	13.3%
Peak time t_P (S)	18	8.8	11.7
Rise time t_r (S)	10.1	7.3	9.6
Adjustment time t_s (S)	47.9	32.4	27.6

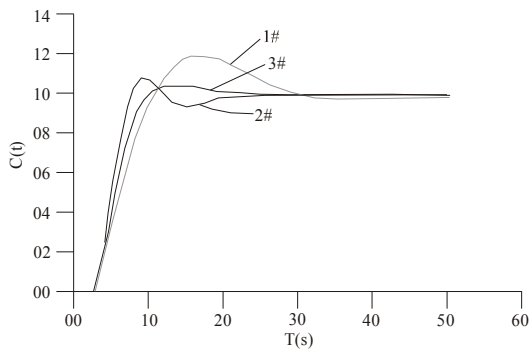

Fig. 1: Simulation figure

Table 8: Comparison of collect effect

One-time clustering results	Bulk incremental clustering results
17.684	15.4330
	50.9170
73.047	74.0182
140.381	145.1240
183.634	186.1020
274.128	270.6840
301.836	298.1030
	314.8430
351.524	348.7130
	365.5810

to simulate the collector pressure system, simulation diagram shown in Fig. 1.

The second-order system transfer function is used as the controlled object, from the maximum overshoot, peak time, rise time and steady-state error of the four performance indicators to evaluate the control effect. It is shown in Fig. 1:

- Curve 1 represents using the empirical inductive method to generate the fuzzy control rules to control the curve of the controlled object.
- Curve 2 represents using K-means clustering algorithm and one-time clustering method to generate the fuzzy control rules controlled object after the control curve.
- Curve 3 represents the bulk incremental fuzzy clustering method to update the curve after the fuzzy control rules to control the controlled object.

Compare the three curves in Fig. 1, giving them control effect data analysis, as shown in Table 7.

Comprehensive control effect curves (Fig. 1) and controls the effect of data analysis. Table 8 shows that, the use of data mining fuzzy control rules generated coke oven collector pressure system with its rationality is entirely feasible in practical applications. Fuzzy clustering method to introduce bulk incremental, real-time update of the fuzzy control rules, the newly generated fuzzy control rules to better reflect the current system state.

CONCLUSION

This study proposes the method creating fuzzy control rules based on experiential knowledge. A new way is provided to create fuzzy control rules. With the introduction of incremental data mining techniques, fuzzy control rules can be updated dynamically in time on the condition that massive new data is producing.

REFERENCES

Ian, H. and F. Eibe, 2000. Weka machine learning algorithms in java [M]. Morgan Kaufmann Publishers, American, pp: 265-230.

Lai, B. and X. Chen, 2005. A method for Multi-atrribute Classifying Based on Decision Tree [J]. Comp. Eng., (05): 111-113.

Li, Z. and Z. Sun, 2006. Data mining application in controlling the pressure of gas-collecting pipe [C]. Proceedings of the 3nd International Conference on Impulsive Dynamic Systems and Applications, 3(5): 201-203.

Liu, Z. and Q. Wen, 2005. K-means clustering algorithm analysis and achievement. Comp. Knowl. Technol. [J], 6: 20-24.

Sun, Z., L. Zhongkui, C. Lei and W. Jinhua, 2010. Creating Fuzzy Control Rules Based on Data Mining. Sixth International Conference on Natural Computation (ICNC), China, Yantai, Vol. 3, pp: 1635-1639.

Zhigang, L. and Z. Bo, 2007. An incremental partition-based clustering algorithm in a batch mode [C]. The 4th International Conference on Impulsive Dynamical Systems and Apllications, China, Nanning, pp: 1116-1119.