Proceedings of the American Control Conference
Anchorage, AK May 8-10, 2002

Soft sensing modeling using neurofuzzy system based on Rough Set Theory

Luo Jianxu Shao Huihe
Institute of Automation, Shanghai Jiao Tong University, 200030

Abstract

This paper contributed to develop soft sensor models combining the theory and methodology of data mining technology. Rough set theory, which can extract reductive rules out of data, is an efficient tool of data mining. In this paper, rough set theory is used to obtain the reductive rules, which are used as the fuzzy rules of the fuzzy system. Then the fuzzy system is represented via an equivalent artificial neural network (ANN). Because the initial parameter of the ANN is reasonable, the convergence of the ANN training is fast, and since the rules are reducts, the structure size of the ANN becomes small. The neurofuzzy approach based on RST is used to building a soft sensor model for estimating the freezing point of the light diesel fuel in Fluid Catalytic Cracking unit (FCCU).

Key words: soft sensor, neurofuzzy, rough set theory, rule extraction

1.Introduction

A major problem in product quality control is the lack of online quality sensors. Although in some cases analytical instruments are available, they possess substantial measurement delays, which make timely control impossible. Soft sensing technology is an economical and efficient solution for this problem [1]. Soft sensor is a modeling approach to estimating hard-to-measure process variables (primary variables) from easy-to-measure, online process sensors (secondary variables). There are three main kinds of approaches to building models used as soft sensors in processes: mechanistic modeling, statistical regression modeling and artificial intelligence modeling. Due to the uncertainty, complexity and nonlinearity of industrial processes, mechanical models are often unavailable, thus, data-driven empirical models are useful alternatives. Data-driven modeling approach makes use of certain technologies (statistical regression, artificial intelligence, etc) to get the relationship between the primary variables and the secondary variables from data samples, and predict the value of the primary variable consequently.

Since distributed control system (DCS) is widely used in chemical processes, the amounts of data historians about the process increase, while our ability to analyze that data has not kept up proportionately. Knowledge Discovery from Databases (KDD), which has been developed to extract knowledge from data, is a rapidly developing discipline in computer science. Data mining is a part of KDD research. KDD methods could be applied for the analysis of process systems, and the results could be used for process design, process improvement, process monitoring, operator training, and so on [2].

Since rough set theory (RST) was exposed by Pawlak as a method of set approximation in1982 [3], it has continued to flourish as a tool of data mining [4]. Due to its advantages, which include the elimination of the need for additional information about data and the ability to extract rules directly from data itself, this theory has been used in more and more domains. The integral part of RST is the construction of rule using reducts, which are particular subsets of the attributes providing classifications with the same quality of approximation as the full set of attributes [5].

Neurofuzzy approach has been widely used in nonlinear modeling [10], but few people make use of it to build soft sensor models. One of the main reasons is that for high-dimension problem, fuzzy rules set may be very large, which makes the structure of the network very huge. In this paper, Neurofuzzy system is used to develop soft sensor model. To avoid the above problem, the fuzzy rules are generated by RST, hence, the number of rules decreases, and each rule dose not need all input attributes.

The rest of the article is structured as follows: section 2 represents the steps of rule extraction using RST, including obtaining decision table from data samples, discretizing continuous-valued attributes, and generating minimal reducts; section 3 presents the structure and learning algorithm of the neurofuzzy system based on the minimal reducts; section 4 uses the modeling approach to a soft sensor modeling for estimating the freezing point of the light diesel fuel in Fluid Catalytic Cracking unit (FCCU).

2. Rule extraction

The neurofuzzy system has the ability of self -learning, and it can auto generated rules from data historians other than from expert's prior knowledge. The aim of the rule extraction is to get rules in the following form from input-output data historians:

$R^{(K)}$: If x_1 is A_1, x_2 is A_2, ..., x_m is A_m, then Y is Y_K.

Where $R^{(K)}$ represents the *kth* rule, x_1, x_2, x_m are input variables, Y is the output variable and A_1 A_2 A_m are input fuzzy sets and Y_K is corresponding output fuzzy sets.

When the input dimension increases, the number of the fuzzy rules will increase exponentially, which makes the structure of the neurofuzzy system large and the learning speed slow. Hence, rough set theory is used here to extract reductive rules from sample data set. The number of rules decreases, and each rule does not need all condition attributes values, so the structure becomes small and the number of parameters needed to be trained decreases, and the speed of the network training improves.

The steps of rule extraction are as follows:

2.1 Get initial information system using the sample data set

Consider MISO system. Suppose there are m pieces of sample data, in the form $\{x_1, x_2, x_n ....y\}$. The sample data set is used to generate the initial information system (decision table): $S = < U, C, D, V, f >$, where U is finite universe of objects, (here is the m pieces of data), $C \bigcup D = R$ is the set of finite attributes , $C = \{x_i\}$ is the set of condition attributes,(here is the input variable x_i), $D = \{y\}$ is the set of decision attributes,(here is the output variable y), $V = \bigcup_{r \in R} V_r$ is the set of all possible attribute values , V_r is the range of attribute r, $r \in R$, and $f : U \times R \to V$ is the information function. The initial information system can be represented with table 1.

Tablel 1 Initial information system

U	x_1	x_2	x_3	x_i	x_n	y
1	x_{11}	x_{12}	x_{13}	x_{1i}	x_{1n}	y_1
2	x_{21}	x_{22}	x_{23}	x_{2i}	x_{2n}	y_2
3	x_{31}	x_{32}	x_{33}	x_{3i}	x_{3n}	y_3
k	x_{k1}	x_{k2}	x_{k3}	x_{ki}	x_{kn}	y_k
m	x_{m1}	x_{m2}	x_{m3}	x_{mi}	x_{mn}	y_m

2.2 Discretization of continuous attribute values

There exist a great number of successful discretization algorithms and applications [6], such as equal-width-intervals, equal-frequency-intervals [8], Minimal Entropy Method [9] and fuzzy method etc. In this paper, we apply fuzzy method to condition attributes, and equal-width-interval method to the decision attribute.

a. Fuzzy discretization of condition attribute value

For each condition attribute x_i , ($x_i \in [x_i^{\min}, x_i^{\max}]$), $[x_i^{\min}, x_i^{\max}]$ is divided into m_i parts. Each partition point corresponds a fuzzy set A_{ij} $(j = 1,2,....m_i)$, and the membership function value of the point is 1. The membership function of A_{ij} is Gaussian function.

Suppose x_{ij} is the value of the attribute x_i,

$$\text{Let } \mu_{A_{is}}(x_{ij}) = \underset{k \in \{1,...,m_i\}}{MAX}\{\mu_{A_{ik}}(x_{ij})\} \qquad (2.1)$$

Where A_{ik} is the fuzzy subset of x_i, and it is also corresponding to a discrete value k. Here, A_{is} has the maximal membership function value, then select s as the discrete value of x_{ij}.

b. Equal-width-interval discretization of decision attribute value

For decision attribute, $y \in [y_{\min}, y_{\max}]$, divide $[y_{\min}, y_{\max}]$ into n_i parts , and each partition point is Y_j , $j \in (1,2,3...,n_i)$ .Given a decision value y_i, seek the partition point Yr , who satisfies the following equation:

$$|y_i - Y_r| = \underset{j \in (1,2,\dots,n_i)}{MIN} \{|y_i - Y_j|\}, \ r \in (1,2,\dots n_i) \quad (2.2)$$

Then, r is the discrete value of y_i.

Substitute the continuous attribute value with discrete attribute value in table 1. Each row in the table can be considered as a piece of rule, and each rule has a confidence μ_k, μ_k is defined as follows:

$$_{k} \quad \underset{i\ 1,\dots n}{} \ {}^{(k)}_{A_{ki}}(x_{ki}) \qquad 2.3)$$

Here, k is the kth rule. $k \quad 1,2,\dots.m$.

*2.3 **Get reductive rules from discrete decision table***

a. Dealing with inconsistent rules

If there are two rules with the same condition attribute values, but different decision attribute value, then these two rules are inconsistent. Keep the one with bigger confidence, and delete the other.

b. Dealing with superfluous attributes

For attribute set c , $c \quad \{x_1, x_2, \dots. x_n\}$, if $\cap(c - \{x_i\}) \quad \cap c$, then x_i is superfluous in c. How to judge weather x_i is superfluous or not? If no inconsistent rules will appear after x_i is removed, then x_i is superfluous. Remove all redundant attributes from attribute set c.

c. Dealing with redundant attribute values and obtaining reductive rules set

There are several methods to generate reductive rules set. Maximal cover algorithm (MCA) [7] is used in this paper.

First, some definitions are introduced.

Definition 2.1: Given a decision rule d_x, if $[x]_{C\ \{r\}} \not\subset [x]_D$, then r is indispensable for d_x, and r is called *core* attribute of d_x ; if $[x]_{C\ \{r\}} \subseteq [x]_D$, then r is omissible for d_x, and r is not the core attribute of d_x.

$$core(d_x) \quad \{r \mid \forall r \in C, [x]_C \not\subset [x]_D\} \quad (2.4)$$

Definition 2.2: Let $Y \quad \{Y_1, Y_2, \quad , Y_n\}$ be decision class in U, defined according to decision attribute D. There is a *decision rule class (DRC)* for each decision class defined as follows:

$$DRC(Y_j) = \{d_x : Des([x]_C) \to Des([x]_D) |, x \in U, [x]_C \quad Y_j\}$$
$$\forall Y_j \in Y \qquad (2.5)$$

Definition 2.3: Let $C \quad \{C_1, C_2, \quad , C_m\}$, then the probability of the attribute $C_i \in C$ appearing in the $DRC(Y_j)$ is $P(C_i, Y_j)$:

$$P(C_i, Y_j) \quad \frac{card(core(C_i) \cap DRC(Y_j))}{\underset{i\ 1}{\overset{m}{.}}\ card(core(C_i) \cap DRC(Y_j))} \quad (2.6)$$

When we compute the reduct of a rule, start from the core attributes of the rule, then add attribute with bigger significance gradually according to certain criterion, thus make the rule cover more positive examples, and exclude all negative examples. In this way, the reduction rule set can be obtained. This method is called maximal cover algorithm. The criterion applied here is the maximal probability criterion: Compute the probability of each attribute appearing as the core attribute in the $DRC(Y_j)$, viz. $P(C_i, Y_j)$, the bigger the probability is, the bigger the attribute's significance is.

3. The neurofuzzy system based on the reductive rules

Substitute the discrete condition attribute value with the corresponding fuzzy subset and the decision value with the corresponding partition point, we get the initial fuzzy system.

$R^{(K)}$: If x_i is A_i, x_j is A_j, then Y is Y_K.

Where x_i, x_j are the input variables (condition attributes) used in the kth rule. $i, j = 1,2,..n$.

Then the output of the fuzzy system is[10]:

$$\hat{y} \quad \frac{\underset{j\ 1,\dots,R}{} \ {}_j * Y_j}{\underset{j\ 1,\dots,R}{} \ {}_j} \qquad 3.1)$$

Where ${}_j$ is the confidence of the j th rule ,

$$_j \quad \underset{i\ (1,2,\dots,n)}{} \ {}^{(j)}_{A_{ij}}(x_{ij}) \qquad (3.2)$$

Then we use the equivalent neural network to realize the fuzzy system.

3.1 Neural network structure

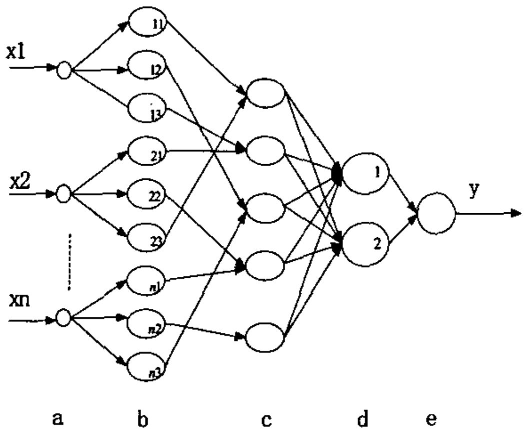

Fig.1 network structure

a is the input layer, b is the fuzzyfier layer, in witch each neuron perform the membership function of x_{ij}, c is the infer layer, in witch each neuron is a multiplier calculating the confidence of each rule d and e are defuzzifier layer, which are used to calculate the equation 3.1. Neurons in layer d are adders, the weight connecting $\Sigma 1$ and layer c is each rule's conclusion part, y_j, and the weight connecting $\Sigma 2$ and layer c is 1.

3.2 Learning algorithm

The neurofuzzy system uses BP learning algorithm. Choose Gaussian function as membership function. Then the output of neurofuzzy system has the form of equation 3.3.

$$\hat{y} = \frac{\sum_{r=1}^{R} y_r \mu_r}{\sum_{r=1}^{R} \mu_r} = \frac{\sum_{r=1}^{R} y_r \{\prod_{i\in\{1,2,..n\}} \exp[-(\frac{x_i - \bar{x}_i}{\sigma_i})^2]\}}{\sum_{r=1}^{R} \{\prod_{i\in\{1,2,..n\}} \exp[-(\frac{x_i - \bar{x}_i}{\sigma_i})^2]\}} \tag{3.3}$$

The general BP learning formula is as follows:

$$\omega(k+1) = \omega(k) - \beta \frac{\partial e_k}{\partial \omega}\Big|_{\omega = \omega_k} \tag{3.4}$$

Where $e_k = (y_k - \hat{y}_k)^2/2$, β is learning rate, ω is the parameter to be modified, including the rule's conclusion part y_r, the center of the Gauss function $\bar{x}_i$ and the width of the membership function σ_i. Learning algorithm of each parameter is as follows:

$$y_r(k+1) = y_r(k) + \beta(y_k - \hat{y}_k)\frac{\mu_r}{\sum_{r=1}^{R} \mu_r} \tag{3.5}$$

$$\bar{x}_i^r(k+1) = \bar{x}_i^r(k) - \beta \frac{y_k - \hat{y}_k}{\sum_{r=1}^{R} \mu_r}(y_r - \hat{y}_k)\mu_r \frac{2[x_i^k - \bar{x}_i^r(k)]}{[\sigma_i^r(k)]^2} \tag{3.6}$$

$$\sigma_i^r(k+1) = \sigma_i^r(k) - \beta \frac{y_k - \hat{y}_k}{\sum_{r=1}^{R} \mu_r}(y_r - \hat{y}_k)\mu_r \frac{2[x_i^k - \bar{x}_i^r(k)]^2}{[\sigma_i^r(k)]^3} \tag{3.7}$$

4. Application in soft sensing modeling

Fluid catalytic cracking unit (FCCU), which converts heavy gas oils into a range of lighter hydrocarbon products, is an essential equipment of the refinery. The ability of a typical FCCU to produce lighter and more valuable products, such as gasoline, light diesel fuel, etc, from low market value feedstock makes the FCCU play an important role in the overall economic performance of the refinery. A typical FCCU in Shijiazhuang refinery includes reactor-regenerator, fractionator and gas processing facilities. Vapor products from the reactor are sent to the main fractionator where various boiling products are withdrawn such as gasoline, diesel, etc. In practical processes, the control of the boiling point is the control of the dry point of the rough gasoline and the freezing point of the light diesel. But these two variables are hard to measure on line, and are usually analyzed offline with big time delay, which makes timely control of product quality impossible. We applied the above-described modeling approach to predict the freezing point of the light diesel.

According to the analysis of technological mechanisms, five variables are selected as secondary variables for estimating the freeze point of the light diesel fuel (Fp). They are light diesel fuel extracting temperature (T110_44), gas temperature at 19th column tray (T211_3), reflux flow rate (G215), reflux extract temperature (TI_72), and the cycle return temperature at the first middle sidetrack in main fractionator (TI_20). The source of data acquisition is the process data, which are recorded and collected from the DCS system and the corresponding daily laboratory analysis. We collected 200 samples of each variable, and used 150 samples to train and 50 samples to test.

First, the initial information table is obtained from the sample data set.

Table 2 Initial information system.

S	G_215	TI_72	TI_20	T110_4	T211_3	Fp
1	156.0	282.0	151.0	244.0	223.0	1
2	145.0	266.5	145.5	234.0	213.5	5
3	151.0	269.0	128.5	233.0	209.0	10
4	118.0	286.0	164.0	257.0	218.0	3
5	147.0	275.0	165.0	245.0	226.0	-1
6	151.5	259.5	145.5	228.0	209.0	-9
7	156.0	267.0	154.0	240.0	227.0	1
8	150.5	265.5	129.5	229.5	207.0	-10
9	150.5	268.0	135.0	233.5	210.0	-9
10	151.0	265.0	124.5	228.0	205.5	-11
150	150.0	262.0	136.5	230.5	207.5	-10

The following discretized table is got when the partition number of each input variable is 3, and that of the output is 5.

Table 3 Discretized information system

S	G_215	TI_72	TI_20	T110_4	T211_3	Fp
1	2	2	3	3	2	4
2	2	2	3	1	3	3
3	2	2	1	1	1	2
4	1	2	2	2	3	5
5	2	2	2	2	4	4
6	2	3	3	1	1	2
7	2	2	3	2	2	4
8	2	3	1	1	1	2
9	2	2	1	1	1	2
10	2	3	1	1	1	2
150	2	3	3	1	1	2

Then the Maximal Cover Algorithm is used to get the reductive rules set, and a rules set containing 19 rules is obtained.

Table 4 Reductive rules set

Rule	G_215	TI_72	TI_20	T110_4	T211_3	Fp
1	2		1	1		1
2		2	2	2	2	2
3	3		2	1	2	2
4	3				1	2
5			2		1	2
6				2	1	2
7	3	3	2		2	3
8			1		2	3
9		2	3		2	3
10	2			1	2	3
11		1				3
12			3		3	4
13		2			3	4
14	2				3	4
15	3	3	3		2	4
16			3		1	4
17	1	3			2	5
18	2	3			2	5
19	3	3	2		3	5

A neurofuzzy system is constructed according to these rules. The generalization ability and the learning ability of the model can be evaluated by the following two criterions:

$$lmse=\frac{1}{N_s}\sum_{i=1}^{N_s}\left(y_i-\hat{y}_i\right)^2 \quad (x_i,y_i) \in S_{samp} \quad (4.1)$$

$$rmse=\sqrt{\frac{1}{N_t}\sum_{i=1}^{N_t}\left(y_i-\hat{y}_i\right)^2} \quad (x_i,y_i) \in S_{stest} \quad (4.2)$$

Where $S_{samp}=\{(x_i,y_i), i=1,2,.....N_s\}$ is the learning sample set, and $S_{test}=\{(x_i,y_i), i=1,2,.....N_t\}$ is the testing sample set.

The discretization of the continuous attribute value is essential for the whole system. If the discretizaion interval is big, then the number of attributes contained in the attribute set is few, and important attributes may be left out; if the interval is small, then inconsistent decision table may be obtained, and there are more attributes contained in attributes set, so concise decision table is difficult to get. The following table represents the generation results and the learning ability of soft sensor models under different discretization conditions.

Table 5: model comparison under different discretizing conditions

Discretizing degree	Extracted rules	Iteration times	lmse	rmse
In:3 Out:3	14	194	0.01	1.6521
In:3 Out:4	14	1020	0.01	1.9346
In:3 Out:5	19	178	0.01	1.5514
In:4 Out:3	23	233	0.01	1.7854
In:4 Out:4	22	226	0.01	1.7301
In:5 Out:3	22	1205	0.01	1.9002
In:5 Out:4	22	430	0.01	1.8191
In:5 Out:5	20	401	0.01	1.7560
In:5 Out:6	21	1150	0.01	2.0847

So, when the partition number of each input variable is 3, and that of output variable is 5, the best results are obtained. 19 reductive rules are extracted from sample data sets by RST according to the above-described rule extraction approach. A neurofuzzy system is built according to these rules. After 178 times training, $lmse$=0.01 $rmse$=1.5514 and the generalization results of the neurofuzzy system is illustrated as the following figure.

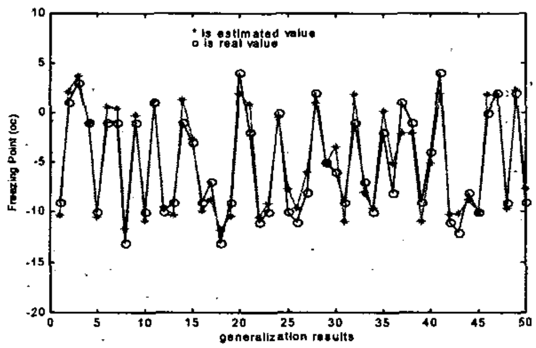

Fig.2 Generalization results

We consider using Genetic Algorithm to get the optimized discretization in the future research.

5. Conclusion

A modeling approach, neurofuzzy system based on Rough Set Theory, is explored in this paper. The soft sensor model developed by this approach has a satisfied result in learning ability and generalization ability. So, introducing the concept and methodology of KDD and data mining into soft sensing modeling is a promising direction.

Literature Cited

[1] Macvoy T. J., Contemplative stance for chemical process. Automatica,1992,28(2):441 442

[2] Wang, X. Z., Data Mining and Knowledge Discovery for Process Monitoring and Control. Springer, London,1999.

[3] Pawlak Z., Rough Sets. International Journal of Information and Computer Science, 1982, 11(5): 341 356

[4] S.K.Pal, A.Skowron, Rough Fuzzy Hybridization: A New Trend in Decision Making, Springer, Singapore, 1999

[5] Pawlak Z., Rough Sets, Theoretical Aspects of Reasoning about Data. Kluwer Academic Publishers, Dordrecht, 1991

[6] R. Susmaga. Analyzing Discretizations of Continuous Attributes Given a Monotonic Discrimination Function. Intelligent Data Analysis, 1997, 1:157 179

[7] Wang Y.Y., Research of Knowledge Discovery Methods Based on Rough Set Theory, Ph.D Thesis, Shanghai Jiao Tong University, 2001

[8] Shan N., Hamilton H.J., Ziarko W., Cercone N., Discretization of Continuous Valued Attributes in Classification Systems, in: Proceedings of the Fourth International Workshop on Rough Sets, Fuzzy Sets and Machine Discovery (RfSD'96). (Tokyo, Japan), 1996, 74 81.

[9] Fayyad U.M., Irani K.B., On The Handling of Continuous-valued Attributes in Decision Tree Generation, Machine Learning, 1992, 8: 87 .102.

[10] Jang, J.R.S, Nuro-fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence. Prentice Hall, Englewood Clis, NJ, 1997