KAPPA NUMBER PREDICION BY HYBRID MODEL FOR BATCH PULP COOKING PROCESS

Yan Li, Jian Zhang, Xuefeng Zhu, Daoping Huang

College of Automation Science & Engineering,
South China University of Technology, Guangzhou, China

Abstract: In batch pulp cooking process the wood chips are converted into pulp by lignin dissolution in cooking acid. The percentage of non-dissolved lignin is often expressed by so called Kappa number. To obtain desired quality of the pulp, Kappa number of the pulp should be decreased to the desired value at the end of batch cycle. Since reliable on-line commercial sensors of Kappa number are still unavailable, developing the soft sensor for measuring Kappa number in batch pulp cooking process is of practical significance. In this paper, a kinetic hybrid model is developed to predict the Kappa number for the batch cooking process. The effectiveness of the proposed hybrid model can be illustrated by the predicted errors for a actual cooking process.

Keywords: Soft sensing, Hybrid mode, Radial basis function network, Pulp cooking process.

1. INTRODUCTION

Pulp and paper industry bears the stamp of exhaustive energy and raw material consumption. To achieve better yield at lower production costs, many researchers have been working on the measurement and control of Kappa number, an important quality index of pulp cooking. Although lots of research workers have been conducted in the field of Kappa sensor technology in recent years, on-line reliable Kappa number measurements in batch digesters is still very difficult. Therefore developing Kappa number model and the model-based control strategy of batch pulp cooking process is a challenge task for the pulp industry.

By analyzing the physical-chemical mechanism of the cooking process, our research team has developed a simplified model for predicting the Kappa number, in which the initial charge conditions are expressed and correlated with an initial effective alkali concentration (sampled at the time of H factor equal to a certain number). The model can achieve very good predictive result in laboratory condition but while the model is used in practical cooking process the performances of the model are not very satisfied. Because of lack of sufficient off-line data to provide comprehensive knowledge of the complicated industrial process, the model is only useful over a narrow range of operating conditions. For this reason, in this paper it is to try developing a hybrid Kappa number model, which will provide the predictive Kappa number for the last phase of the batch cook process by means of learning from the history data.

2. BACKGROUND

2.1 Soft Sensing Technology

Soft sensing technology is a measurement method which employing easily measured variables (auxiliary variables) and their relationship (soft sensing model) to acquire some variables (primary variables), which hard to measure directly. We can say that, soft sensing technology is a method of information utilizing and rule discovery, data classification and variable prediction. Data classification extents the data space by estimating the unknown data class experientially. Simultaneously, variable prediction extents the data temporal space by forecasting the variable development. During the process of soft sensing modeling, different kind of theory and method should be utilized comprehensively to dig out useful information.

Artificial neural network technology has been introduced in the control field because there are many systems whose rigid mathematical models are

hard to acquire, such as highly non-linear chemical processes including those found in the pulp industry. In this case, ANN may be an effective tool to cope with these problems, especially for systems whose characteristics and uncertainties are difficult to identify using mathematical models. To model such systems, ANN can provide some promising solutions (Thompson and Kramer, 1994).

In this paper, an empirical predictive hybrid model is provided which based on neural network. It also takes advantage of Rough Set Theory and fuzzy theory to construct the model. First, certain rules and uncertain rules are acquired by analyzing the history data using Rough Set Theory. Then, Radical Based Neural Network is employed to realize the fuzzy model. The main steps are:

(1) Rules Extraction by attribute reduction strategy based on rough set theory.
(2) Using the rules as the node centres of the hidden layer to train the RBF.

2.2 Rule Extraction

Rough Set Theory

Rough Set Theory was introduced by Z. Pawlark, a polish mathematician, in 1982. It is a relatively new soft computing-tool to deal with vagueness and uncertainty (Pawlark, 1996). It has received much attention of the researchers around the world. Rough Set Theory has applied to many area successfully including pattern recognition, machine learning, decision support, process control and predictive modeling.

Rule Extraction

The concept of un-differentiate relationship is the basement of the RS. The other important concepts include upper approach, lower approach, boundary area and rough abstract function. The main steps are listed below using Rough Set methods to discover knowledge and decision rules by analyzing and simplifying the great amount of measure data (Wang, *et al.*, 1998).

(1) Disperse the continuous data interval into discrete intervals, using the code of the sub-area as the value of the continuous data.
(2) Acquire the discrete decision table and begin attributes reduction, using reduction strategy based on differentia matrix, which defines the times of the attribute appeared in differentia matrix as the attribute significance.
(3) Based on the simplified result, look for the upper approximate set and lower approximate. Then sum up the logic rules.

Review these rules and compare them with expert experience to acquire the final results. These rules will supervise the training of the RBFNN.

2.3 Radical Basis Function Network

In a sense, radical basis function has common ground with fuzzy system, or we can realize a certain kind of fuzzy system by means of Radical Basis Function Network. The number of the nodes in input layer is the number of reduction attribute vectors. The number of the nodes in hidden layer is decided by number of the rules. The transform function of hidden layer is Gauss function.

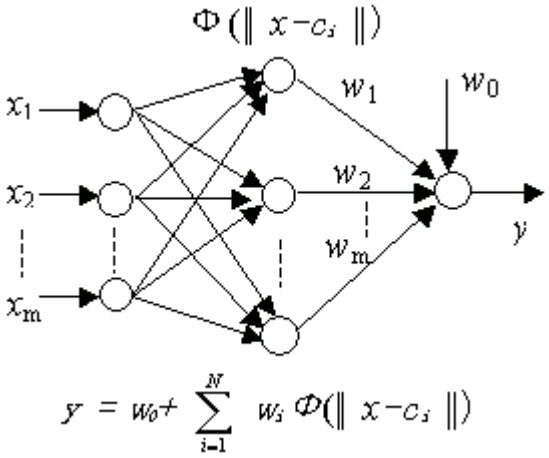

Fig.1. Structure of RBFNN

$$\Phi_i(x) = e^{\frac{\|x-c_i\|^2}{\sigma_i^2}} \quad (1)$$

Contract to number i rule of a simple system:
IF x_1 IS A_{1i} and x_2 IS A_{2i} ... and x_n IS A_{ni} , THEN y_1 IS w_{i1} and y_2 IS w_{i2} ...and y_m IS w_{im}

The centres of Gauss Function are decided by the precondition vectors of the fuzzy rules. The weights of the output layer are corresponding with the posterior parameter of the fuzzy rule. That is RBFNN is of some comparability with fuzzy system. Comparing with the classical BP neural network, RBFNN is more apprehensible (Krzyzak and Linder, 1998).

3. HYBRID MODEL

Choosing initial effective alkali (sampled at time of H=200), sulfide degree, H factor and wood chip eligible rate as the input variables and Kappa number as the output the hybrid model can be developed.

3.1 find the centre value of hidden layer node by Rough Set Theory

Using 160 groups data from a factory cooking process as an example to illustrate the construction of the hybrid model. The first 120 groups data are chosen as learning data, last 40 groups of data are used to verify the effect.

construct binary decision table

To construct a decision table, the consecutive variables should be converted to be discrete firstly, so Equal Interval Division method and Equal Probabilities method have been applied, but results are not so ideal. These methods are difficult to determine the discrete grade. Too rough grade leads to appear large amount of inconsistent data. Consequently, more inconsistent part of the

constructed decision table will be produced. On the other hand, if the grade is too precise, rules can't be abstracted effectively from the decision table. To solve the problem, code the decision table, which is dispersed by equal interval division method, and the table is converted to an approximately binary attribute table, then conduct attribute reduction using differential matrix method (Wang, Y.Y., 2001). As the result, some neighbourhood intervals would join together. The steps in details are listed as following:
(1) Evaluate frequency distribution table of condition variables using statistic analysis software, as presented in table 1.

Table 1 Variable frequency distribution table

fre-quency	dividing point			
	effective alkali	sulfide degree	wood chip eligible	H factor
-10%	25.43	25.90	69.80	1826.20
-20%	26.97	26.80	73.00	1890.80
-30%	27.90	27.43	74.33	1954.60
-40%	28.68	27.64	75.68	2016.80
-50%	29.37	28.20	77.80	2048.00
-60%	29.92	28.60	79.38	2148.40
-70%	30.54	29.00	80.41	2234.20
-80%	31.16	29.50	81.36	2299.20
-90%	32.55	29.90	84.00	2480.60

(2) Utilize the frequency distribution table and maximum limit of variables Condition variables are divided into 10 intervals, then each interval is coded, for example, the coding result of effective alkali variable is shown in table 2.

Table 2 code of effective alkali interval

Inter-val	[22.00 25.44)	[25.44 26.97)	[26.97 27.90)	[27.90 28.68)	[28.68 29.38)
Code	0	1	2	3	4
Inter-val	[29.38 29.92)	[29.92 30.54)	[30.54 31.16)	[31.16 32.55)	[32.55 43.00)
Code	5	6	7	8	9

(3) Construct a binary decision table, for a decision system, any condition attribute among the system can be represented by 9 binary attributes $Z_q^0,\ldots,Z_q^8$, whose value domain is $\{0,1\}$.(as shown in table 3)

For example, if the value of a sample effective alkali is 28.72, the interval code will be 4, then the attribute is $\{Z_e^0, Z_e^1,\ldots, Z_e^8\} = \{0,0,0,0,1,1,1,1\}$ after conversion.
Decision attributes can be divided into 3 intervals by equal frequency. (as presented in table 4)

Table 3 binary table of attribute q

Value	Z_q^0	Z_q^1	...	Z_q^7	Z_q^8
0	0	0	...	0	0
1	0	0	...	0	1
...	...	...	...	...	...
8	0	1	...	1	1
9	1	1	...	1	1

Table 4 Discretization of Decision Attributes

interval	[22.00, 33.27]	[33.27, 36.90]	[36.90, 49.00]
code	0	1	2

Consequently, there are 36 condition attributes and 1 decision attribute in the constructed binary decision table (as seen in table 5), while the original table only has 4 condition attributes and 1 decision attribute., the value domain of condition attributes is $\{0,1\}$, while the value domain of decision attributes is $\{0,1,2\}$

Table 5 binary attribute decision table

Case no.	$Z_e^0 Z_e^1 Z_e^2$ $Z_e^3 Z_e^4 Z_e^5$ $Z_e^6 Z_e^7 Z_e^8$	...	$Z_H^0 Z_H^1 Z_H^2$ $Z_H^3 Z_H^4 Z_H^5$ $Z_H^6 Z_H^7 Z_H^8$	K_a
1	0 0 0 1 1 1 1 1 1	...	0 0 0 1 1 1 1 1 1	0
2	0 0 0 0 0 0 1 1 1	...	0 0 0 0 1 1 1 1 1	2
...	...	...	...	...
119	0 0 1 1 1 1 1 1 1	...	0 0 0 0 0 1 1 1 1	0
120	0 0 0 0 0 0 0 1 1	...	0 0 0 1 1 1 1 1 1	1

(4) Reduce the attributes of the constructed binary decision table using the decision attributes reduction technique presented in paper (Wang, *et al*., 1998), reduplicate samples reduction is made in vertical and reduction based on differential matrix method is made on horizontal. As a result, two reduplicate samples are reduced and the reduced condition attributes are:

effective alkali $\{Z_e^1, Z_e^3\}$

sulfide degree $\{Z_s^1, Z_s^3, Z_s^5, Z_s^6, Z_s^8\}$

wood-chip eligible rate $\{Z_m^3, Z_m^4\}$

H factor $\{Z_H^2, Z_H^4, Z_H^5, Z_H^7, Z_H^8\}$

How to merge conditions variables is shown in table 6.

In the next section, we will train the network by using centre value of rule precondition interval as centre value of hidden layer nodes of RBF neural network, so the logic rules induction is not needed here. We can recode each interval based on the condition variables interval table above, the coding sequence can be 0,...,n-1, where n is number of intervals, then the reduction decision table is created.

Table 6 Discretzation of Decision Attributes

Variable	Inter-val Num-ber	Intervals
Effective alkali	8	[22.00,25.44], [25.44,26.97], [26.97,27.90], [27.90,28.68], [28.68,29.38], [29.38,30.54], [30.54,32.55], [32.55,43.00],
Sulfide degree	5	[23.00,26.80], [26.80,28.20], [28.20,29.00], [29.00,29.90], [29.90,33.00]
Wood chip eligible rate	8	[58.00,69.80], [69.80,73.00], [73.00,74.33], [74.33,75.68], [75.68,80.41], [80.41,81.36], [81.36,84.00], [84.00,89.00],
H factor	5	[1300.0,1954.6], [1954.6,2148.4] [2148.4,2299.2], [2299.2,2480.6], [2480.6,5400.0]
Kappa number	3	[22.00,33.27], [33.27,36.90], [36.90,49.00]

3.2 Determination of RBF neural network hidden layer nodes and learning samples

Although equal interval coding method is applied to discrete continues interval, the data still may be lost or distorted, which will produce inconsistent rule, which is that two samples have the same attribute value while the decision value is difference. There are 3 pairs of inconsistent rules in previous process. The appearance of inconsistent rules is a common problem when constructing model with data from industry fields, especially, when constructing model with data from some complex process. Such situation does often appear: a pair of input data is nearby in distance while the corresponding output is reverse by distance meaning, so the extension capability of the model is not so ideal. There are many reasons, including improper selection of raw pulp sample spot, miss-manipulate by operators and some baffling reason etc. Consequently, the error between evaluation value of model and measurement value (off-line) is great when the soft-sensor model is running.

To solve above problem, we should select proper discretazation method, separate interval reasonable and consult technical mechanism. The necessary rules and possible rules should be compared with expert experiences and then made proper adjustment. Applying distributed RBF network structure, we can select the number of decision classification as number of subnet, and then divide the original problem with m decision attributes into m sub problems. To set up soft-sensor model for paper pulp Kappa value, decision values are divided into 3 interval and 3 sub networks are constructed. The last result is weighted average of decision values, so can keep balance between inconsistent rules.

When training RBF sub network, the number of hidden layer's nodes is determined by number of samples in decision table, while the centre value of transform function is chosen based on centre of interval of condition attributes, for example, a RBF sub network with 39 hidden layer nodes, whose decision value is 0, the centre value corresponding to node i is normalization of centre values for interval of sample i. Variances of the Gauss function are all set as 1.

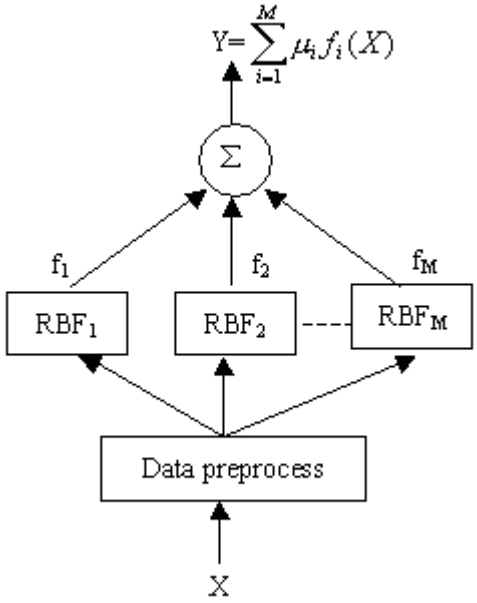

Fig.2. Distributed RBF network structure

(1) Normalization of data

Normalization of data is necessary for that measurement scale is difference between variables. Firstly we can easily get the maximum value domain by using statistical analysis because the technical variables are limited strictly in certain range. The transform formula is as following:

Provided that variety range is $[x_{\min}, x_{\max}]$, after normalizing, $x' = (x - x_{\min})/(x_{\max} - x_{\min})$.

To satisfy the requirement of precision, we can adjust the weight of output. For subnet work 1, where there are 38 hidden layer nodes and 41 training samples. The following table 7 illustrates how to train the three RBF subnet works.

Table 7 Training of the RBF sub-network

RBF Sub-network	Nodes of Hidden layer	Training samples	Training error
1	39	41	0.080
2	32	32	0.032
3	47	47	0.087

(2) Distributed model structure

we should keep to the following rules when select proper μ_i,

$$\alpha_i = \frac{\sum_{k=1}^{M_i} \frac{1}{d_k}}{M_i}, \mu_i = \frac{a_i}{\sum_{l=1}^{M} a_l} \qquad (2)$$

where M is number of sub-network (in the paper, M=3); M_i is number of training samples for sub-network i, d_k is square of Euclidean distance between testing samples and learning samples.

3.3 Verify the model

In previous work, we have developed a regressive equation for Kappa number of paper pulp based on kinetics mechanism of delignification (Luo and Liu, 2000).

$$K_a = A - B\ln[(H - H_b)(CEA_b)^n] \qquad (3)$$

where K_a is Kappa value, H is H factor

H_b, EA_b are H coefficient and effective alkali on mass delignification period separately.

C is function of sulfide degree, $C = \ln(S)$

A, B, R are coefficients to be determined.

Table 8 performance compare

Absolute Error	Mini-mum	Maxi-mum	mean	Standard deviation
RBFNN	0.08	9.20	2.332	2.1213
Mechanism-regression model	0.10	12.70	3.592	2.8041

Analyzing 40 groups of error tolerances absolute values, comparing them with predicted results of mechanism-regressive equation, it can be seen that the result of RBF neural network is better than former (table 8).

4. CONCLUSIONS

In this paper, a kinetic hybrid model is developed to predict the Kappa number on the lat phase of the batch cook process. The hybrid model is consisted of two modules: the lower essential module and the upper module. The essential part of the model is Radical Basis Function Neural Networks, which is composed by three sub networks. Considering some non-linear factors, such as the conflicts existing in the sample data, undetectable initial conditions, disturbances in cooking and so on, the upper part divides the whole secondary variables space into different fuzzy subspaces by applying expert knowledge and RS (Rough Set) data mining. In each space, the sub RBF network is trained to get better prediction. The final result is given by synthesize the outputs of the three sub network. Effectiveness of the proposed hybrid model is illustrated by the error between the predictive values and the data obtained from the lab. analysis in actual cooking process. It also indicates that the empirical model is effective for certain non-linear and complicated processes.

REFERENCES

Krzyzak, A. and Linder, T. (1998). Radical Basis Function Networks and Complexity Regulation in Function Learning. *IEEE Transactions on Neural Networks.* **Vol. 9(2)**, 247-256.

Luo, Q. and Liu, H.B.(2000). Modelling of the batch kraft pulping. Journal of South China University of Technology(Natural Science Edition). 28(1), 25.

Pawlak, Z. (1996). Rough Sets, Rough Relations and Rough Functions. *Fundamenta Informaticae.* 27: 103-108

Thompson, L.M., and Kramer, A.M. (1994). Modeling Chemical Process Using Prior Knowledge and Neural Networks. *AIChE Journal*, **Vol. 40(8)**, 1328-1339.

Wang, J. (1998). Data Enriching Based on Rough Set Theory. *Chinese J. Compuers.* **Vol. 21(5),** 393-400

Wang, Y.Y., and Shao H.H. (2001). Binary Decision System-based Rough Set Approach for Knowledge Acguisition. *Chinese J. Control and Decision.* **Vol. 16(3),** 374-377

Acknowledgement

The authors wish to acknowledge the support of this project by National 863 Project (2001AA413110) and National Science Foundation of China (60274033).