Sensors & Transducers, Vol. 21, Special Issue, May 2013, pp. 211-217

www.sensorsportal.com

Sensors & Transducers

http://www.sensorsportal.com

Research on Soft Sensing Modeling Method Based on the Algorithm of Adaptive Affinity Propagation Clustering and Bayesian Theory

[1,2] LIU Ranran, [1*] LI Zhengming, [1] PAN Tianhong

[1] School of Electrical and Information Engineering, Jiangsu University, Zhenjiang 212013, China
[2] Changzhou Liu Guojun Vocational Technology College, Changzhou 213004, China
[1] Tel.: 0511-88790618, fax: 0511-88791245
[1] E-mail: lzming@ujs.edu.cn

Received: 9 April 2013 /Accepted: 14 May 2013 /Published: 30 May 2013

Abstract: The industrial data has the characteristic of clustering and migrating with the operating point. The accuracy and generalization ability of the single-model prediction are poor because of the large amount of information lost in single-model modeling. In order to overcome these problems, a modeling method of multi-model soft sensor was proposed based on adaptive affinity propagation clustering (ADAP) and Bayesian filtering. ADAP algorithm was utilized in this method to realize the clustering and tracking of multiple operating points. The sub-models of various types of samples were established utilizing Bayesian filtering method, and the joint output and estimation were carried out based on the model of the subclass of current working point. The soft sensor models of CO and CO_2 in PX oxidation side reaction were utilized in the method. The simulation results show that the estimation and generalization ability of soft sensing model is significantly improved by the method. *Copyright © 2013 IFSA.*

Keywords: Adaptive affinity propagation clustering algorithm, Bayesian filtering, Soft sensor, Multi-model.

1. Introduction

In practical production process, the manipulated variables of each device are controlled around the given values. These given values of the operating point often need to be changed to meet the production requirement, which leads to a problem that the production data cluster and migrate with the working point. Single model has problems like long studying time, low accuracy, poor process characteristic match and poor generalization ability. Therefore, the multi-model modeling method was proposed [1-3]. The basic idea of this modeling method is to cluster the training samples by clustering algorithm and establish the soft sensor sub-model for different types of samples. When the process operating point changes in a wide range, the model can identify this change, and the prediction and output could be conducted in accordance with the model of the corresponding operating point.

The common clustering algorithms are K-means clustering and fuzzy clustering, etc. The initial cluster centers and the number of clusters are determined based on prior knowledge. However, prior knowledge is generally unknown for the operation data of actual production. Affinity propagation clustering algorithm, a new clustering algorithm [4] proposed in recent years, could determine the centers and the number of clusters according to the characteristics of the data. This method has been widely applied so far

[5-7]. A multi-model modeling algorithm of support vector machine based on affinity propagation clustering algorithm (ADAP) was proposed for the modeling of arachidonic acid fermentation process in literature [5]. A multi-model soft-sensing modeling method based on online clustering and association vector machine was proposed in literature [8] which has achieved good results in the final-boiling-point prediction system of light naphtha in hydrocracking unit fractionator. A multi-model algorithm of affinity propagation clustering with supervision was proposed in literature [6] to circularly adjust the clustering based on the output errors, and this method was better than traditional multi-modeling methods in term of modeling effect.

Affinity propagation clustering algorithm, however, is difficult in determining the values of key parameters for the optimal clustering results, and the concussion of the cluster number generated in the iterative process does not automatically disappear to make algorithm converged [9-13]. In allusion to these problems, a soft-sensing modeling method based on Bayesian filtering and adaptive affinity propagation clustering was proposed in the study. The ADAP algorithm was utilized for the classification of training samples in this method; Bayesian estimation method was utilized to create the sub-models of the small classified sample and carry out global output in accordance with the model of the belonging subclass of current operating point. The effectiveness of this method was proved by the CO and CO2 soft-sensing model in PX oxidation side reaction.

2. Algorithm of Adaptive Affinity Propagation Clustering

2.1. Affinity Propagation Clustering Algorithm

Affinity propagation clustering is a new fast and effective clustering method proposed in literature [4]. First, all sample points are regarded as potential cluster centers, and each sample point competes for the cluster center in the iteration loop. So there is no need to determine the number of clusters in advance.

A given sample set $\{x_i, y_i\}, i=1,2,...,N, x_i \in R^m$ is the input of m-dimension sample; $y_i \in R^m$ is the output of the sample. Negative Euclidean distance is utilized to measure the similarity of x_i and x_j, which is $S(i,j) = -\|x_i - x_j\|$., Samples are standardized before solving the similarity to eliminate the effect of dimension. $R(i,k)$ is defined as the attraction degree that x_k is suit for the cluster center of x_i; $A(i,k)$ is defined as the membership degree that x_k is chosen as the cluster center of x_i. AP algorithm continuously collects the evidences ($R(i,k)$ and $A(i,k)$) from data samples. The iteration formula of $R(i,k)$ is shown in formula 1:

$$\begin{aligned} R^{old}(i,k) &= R(i,k), \\ R^{new}(i,k) &= S(i,k) - \max\left[A(i,j) + S(i,j)\right], \\ R(i,k) &= (1-\lambda)R^{new}(i,k) + \lambda R^{old}(i,k) \end{aligned} \tag{1}$$

The iteration formula of $A(i,k)$ is shown in formula 2:

$$\begin{aligned} A^{old}(i,k) &= A(i,k), \\ A^{new}(k,k) &= \sum \max\left[0 + R(j,k)\right], \\ A^{new}(i,k) &= \min(0, R(k,k) - \sum \max\left[0 + R(j,k)\right] \\ A(i,k) &= (1-\lambda)A^{new}(i,k) + \lambda A^{old}(i,k) \end{aligned} \tag{2}$$

If $R(i,k) + A(i,k) = \max\{R(i,j) + A(i,j)\}, j = 1,2,...,n$, data point x_k is the cluster center of x_i. The update speed of the iteration could be changed by the adjustment of damping factor λ, λ=0~1. The element on the diagonal of similarity matrix is a bias parameter P which meets.

$$R(k,k) = p(k) - \max_{j \neq k}\left[A(k,j) + S(k,j)\right]$$

The bigger $p(k)$ is, the bigger $R(k,k)$ and $A(k,j)$ are. So cluster head x_k has the biggest chance of becoming the cluster center. The more and bigger $p(k)$ means that the more cluster heads could finally become the cluster center. Thus the change of P could affect the cluster number of AP. When prior knowledge is unknown, the value of $p(k)$ is the median of S, which is

$$p(k) = p_m = \frac{\sum_{i,j=1}^{n} S(i,j)}{n^2}, k = 1,...,n \tag{3}$$

An indicator Silhouette, introduced to achieve the effect of evaluation cluster, indicates the inner-cluster tightness and inter-cluster separability of cluster structure. Assume that the data set is divided into k cluster subset $C_i (i = 1,2,...,k)$, the indicator Silhouette of a sample t is

$$Sil(t) = \frac{\min\{d(t,C_j)\} - a(t)}{\max\{a(t), \min\{d(t,C_j)\}\}}, \tag{4}$$

where in $d(t,C_j)$ is the average Euclidean distance from t to another cluster C_j; $a(t)$ is the average Euclidean distance from t to the cluster C_i. The average value of the indicator Silhouette of all the samples in a sample set is $S_{il-av} = mean\left[sum(Sil(t))\right]$. S_{il-av} can reflect the quality of all the data and clusters. A bigger S_{il-av} means better cluster effect, and the corresponding cluster number of the maximum is the optimal cluster number. If S_{il-av} is bigger than 0.5, each cluster could be apparently separated; if S_{il-av} is smaller than 0.2, then, a relevant cluster structure is needed.

2.2. Clustering Algorithm of Adaptive Affinity Propagation

AP algorithm can determine the clustering results based on the characteristics of the data with a fast arithmetic rate. But the algorithm has the following problems: the increase or decrease of the bias parameter p can increase or decrease the cluster number of AP algorithm, so it is difficult to determine the value of P to generate the optimal clustering results of algorithm; algorithm cannot automatically eliminate the concussion and converge. In order to solve the above problems, the clustering algorithm of adaptive affine propagation was proposed in literature [9].

The iteration steps of the clustering algorithm of adaptive affine propagation are:

Step 1: initialize S and P and start the algorithm with a relative big P. The initial value of P in this paper is $p=0.5\times p_m$ (the element in S is negative, so p_m is negative);

Step 2: execute AP algorithm once, then K cluster heads and the types of each sample are generated, wherein $\lambda=0.5$;

Step 3: examine whether the K cluster heads are converged (the convergence condition is to meet the preset continuous constant frequency $\omega=0.5$); if they are converged, evaluate the clustering results with effectiveness indicator S_{il-av} and reduce P with a step of

$$p_{step} = \frac{0.01 p_m}{0.1\sqrt{K+50}},$$

if they are not converged, increase λ with a step of $\lambda_{step} = 0.05$; if they are not converged when $\lambda \geq \lambda_{max}$ (λ_{max} is 0.85), the concussion of P is obstinate, and a relatively big λ could not restrain the concussion. This P needs to be abandoned. Reduce P with a step of

$$p_{step} = \frac{0.01 p_m}{0.1\sqrt{K+50}}.$$

Step 4: determine whether the algorithm meets the termination condition, which is to meet that the maximum number of loop iteration or the number of clusters reaches 2. Terminate the iteration if the condition is met; otherwise, go to step 2. ADAP algorithm searches for the number of clusters space to find the optimal clustering results by adaptive scanning bias parameter space. Adaptively adjust the damping factor to eliminate the concussion in the iterative process and reduce the value of P to get rid of the concussion when the damping factor adjustment method fails.

3. Establishment of Soft Sensing Multi-model

In order to overcome the problems of single-model modeling, the structure of soft sensing multi-model based on Bayesian theory is shown in Fig. 1. Firstly, cluster the training samples and find the optimal clustering results utilizing ADAP algorithm. Secondly, establish the sub-models of each sample utilizing Bayesian theory. Lastly, predict the output based on the corresponding sub-model of current operating point in the process. Its structure is shown in Fig. 1.

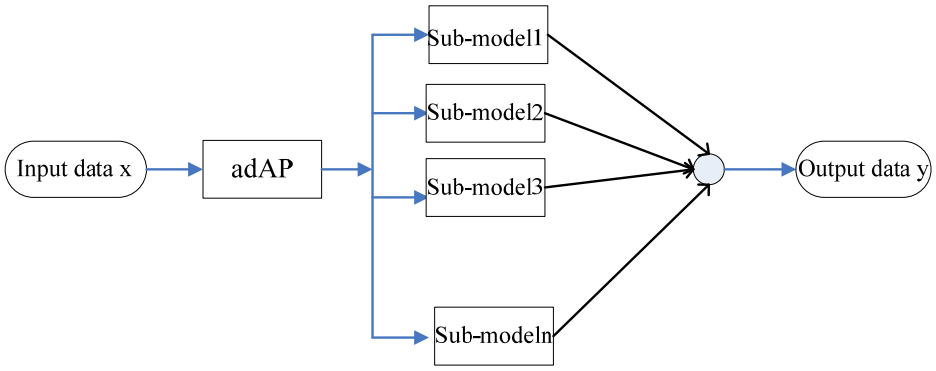

Fig. 1. Structure of multiple models soft-sensing method.

3.1. Bayesian Estimation

In this research, the content of CO and CO2 was thought as a state estimation problem. The state-space model was utilized to describe this problem. Assume that the position state of the target is $(x_t \mid t = 1,..., N)$, and the collection for each moment observation is $(x_t \mid t = 1,..., N)$. Then the state of the target can be described by the motion equation and observation equation in formula 5:

$$\begin{cases} x_t = F(x_{t-1}) + \omega_t \\ z_t = H(x_t) + \upsilon_t \end{cases} \tag{5}$$

where F is the impact parametric equation describing the model input;

H is the observation model which describes the relationship between the observed quantity and model output;

ω_t is the noise, which is utilized to describe the uncertainty of movement;

υ_t is the observed noise, which is utilized to describe the uncertainty caused by outside interference and the noise of the detecting element.

Bayesian estimation is a state estimation method utilizing state priori distribution and the observation of likelihood function to determine the posterior probability distribution. For a first-order Markov process, assume that the observation of each moment is mutually independent. If the posterior distribution at time t-1 is $p(x_{t-1} \mid z_{t-1})$, the prior distribution at time t can be expressed as shown in formula 6:

$$p(x_t \mid z_{t-1}) = \int p(x_{t-1} \mid z_{t-1}) p(x_t \mid z_{t-1}) dx_{t-1} \tag{6}$$

where $p(x_t \mid z_{t-1})$ is the transition probability density function, which is determined by F and the probability distribution of the noise ω_t ($p(\omega_t)$) in motion equation 5. The definition is:

$$p(x_t \mid x_{t-1}) = \int p(\omega_t) \delta(x_t - F(x_{t-1})) d\omega_t \tag{7}$$

where in $\delta()$ is the Dirac function. After the prior distribution $p(x_t \mid z_{t-1})$ is obtained, the state posterior distribution $p(x_t \mid z_t)$ could be expressed as:

$$p(x_t \mid z_t) = \frac{p(z_t \mid x_t) p(x_t \mid z_{t-1})}{\int p(z_t \mid x_t) p(x_t \mid z_{t-1}) dx_t} \tag{8}$$

where $p(z_t \mid x_t)$ is the observation likelihood function, which is determined by H and the probability distribution of the noise υ_t ($p(\upsilon_t)$). The definition is:

$$p(z_t \mid x_t) = \int p(\upsilon_t) \delta(z_t - H(x_t)) d\upsilon_t \tag{9}$$

The above formula 6 and 7 constitute the prediction process of Bayesian estimation; formula 8 and 9 constitute the update process of Bayesian estimation. The two processes are determined by the equation and observation equation of state space model parameters in formula 5, respectively. The tracking of the posterior distribution of state x_t could be achieved through the iterative and recursive solution of above prediction and update process.

3.2. Simulation Study

The source of our data was the PTA production process data in a chemical plant. The model of para-xylene (PX) oxidation side reaction through soft sensor model provided the basis for production optimization of operating parameters and the transformation of the production process. The main factors that will affect the combustion side reaction are: the reaction temperature (x_1, °C), solvent ratio (x_2, Kg. HAc/Kg. PX), the concentration of cobalt catalyst (x_3, wt%), the concentration of manganese catalyst (x_4, wt%), the concentration of bromine accelerator (x_5, wt%) and the residence time (x_6, S). Regard these main factors after data preprocessing as the input data; the total content of CO and CO_2 (ΦCOX) of the reactor exhaust gas as the output data y. Generally 250 sets of data were obtained. 170 sets of data are the training data, and the remaining 80 sets of data are the current running data in the process, which were utilized to test the generalization capability of the model.

Fig. 2 is the cluster number of ADAP clustering algorithm and the corresponding indicator of effectiveness S_{il-av}. Fig. 2 shows the ADAP algorithm was started with a large value of P to obtain several clustering results. The optimal clustering results could be selected according to the indicator S_{il-av}. The indicator S_{il-av} was the largest when the training samples were divided into two clusters in terms of the data utilized in this study. Therefore, the training samples were divided into two clusters.

Further simulation analysis on the training results and the linear regression analysis of the output and target output of network simulation are shown in Fig. 3. The figure shows that the correlation coeffi-

cient is 0.99, which means that the performance of the network is satisfying.

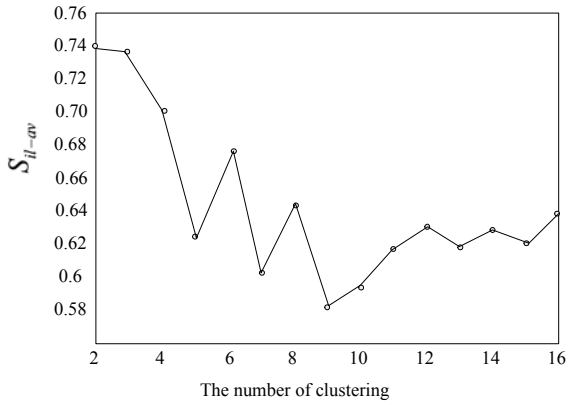

Fig. 2. The effective index of clustering algorithm.

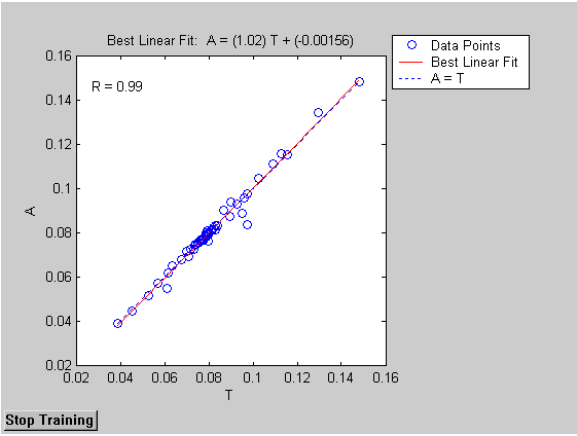

Fig. 3. The analysis result of network output.

3.3. The Combustion Loss Model of Acetic Acid and Xylene

Establish the combustion loss model of acetic acid and PX based on the determined generation model of CO_x. The combustion loss model of acetic acid:

$$m_{HAc}^{consume} = \frac{m_{gas} \times x_{CO_x} \times \frac{x_{HAc}}{100} \times \frac{60}{1000}}{2 \times \frac{m_{HAc}^{CO_x}}{100} \times m_{CTA}}$$

The combustion loss model of PX is:

$$m_{PX}^{consume} = \frac{m_{gas} \times x_{CO_x} \times \frac{x_{PX}}{100} \times \frac{106}{1000}}{8 \times \frac{m_{PX}^{CO_x}}{100} \times m_{CTA}}$$

x_{CO_x} : the total content of generated CO and CO_2 of the exhaust gas;

x_{HAc} : the percentage of CO_x generated by acetic acid (%);

$m_{HAc}^{CO_x}$: the percentage of CO_x accounting for the combustion product of acetic acid side reaction (%);

m_{gas} : the flow of exhaust gas from the reactor (ft3/hr);

m_{CTA} : CTA product yield (tons / Hr);

The molecular weight of acetic acid: 60 (g / mol);

$m_{HAc}^{consume}$: the amount of acetic acid loss (Kg / ton. CTA);

x_{PX} : the percentage of CO_x generated by PX (%);

$m_{PX}^{CO_x}$: the percentage of CO_x accounting for the combustion product of PX side reaction (%);

The molecular weight of PX: 106 (g / mol);

$m_{PX}^{consume}$: the amount of PX loss (Kg / ton CTA);

From the above text, the value of x_{HAc}, $m_{HAc}^{CO_x}$, x_{PX} and $m_{PX}^{CO_x}$ are known as 60 %, 75 %, 40 % and 60 %, respectively; m_{gas} and m_{CTA} were obtained through the actual production data.

4. Analysis of the Effect on Process Operating Parameters on Combustion Loss

Establish a model to reflect the effect of operating parameters on the HAC and PX combustion loss in PX oxidation reaction process. Then, analyze the effect of various operating parameters on HAC and PX combustion loss in the reaction process. Thereafter, analyze the effect of residence time, reaction temperature, solvent ratio, the concentration of catalyst like cobalt, manganese, bromine and other operating parameters on the HAC and PX combustion loss through the model, respectively. The results are shown in Figs. 4-8.

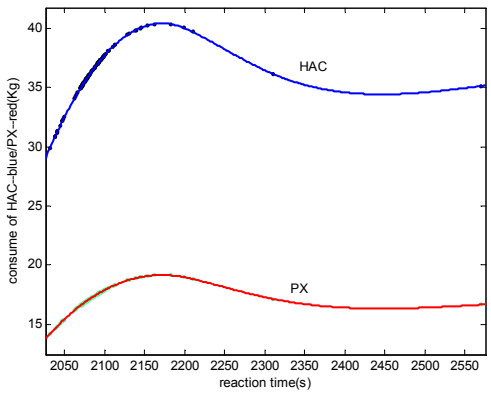

Fig. 4. Effect of reaction time on oxidative side-reaction.

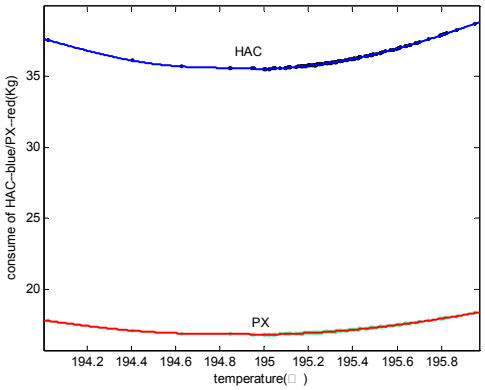

Fig. 5. Effect of reaction temperature on oxidative side-reaction.

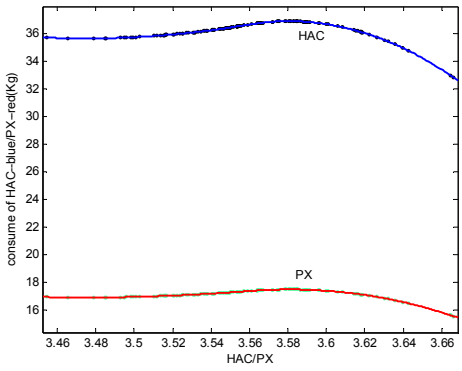

Fig. 6. Effect of HAC/PX on oxidative side-reaction.

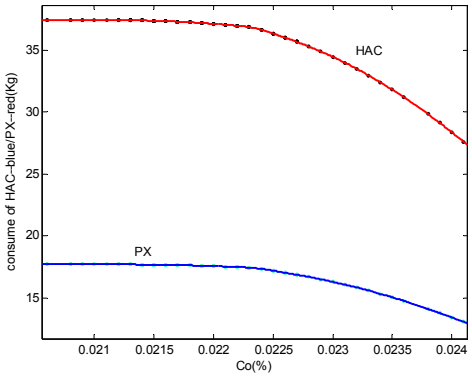

Fig. 7. Effect of Co concentration on oxidative side-reaction.

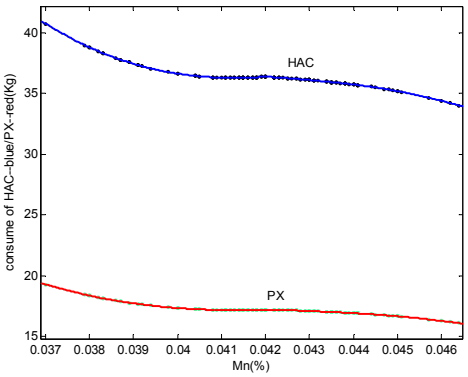

Fig. 8. Effect of Mn concentration on oxidative side-reaction.

Fig. 4-9: reaction residence time has a great effect on the combustion loss; the effects of reaction temperature and the concentration of Co, Mn, Br are small, but the effect of reaction temperature is slightly bigger. The effects of each operating parameter on the combustion loss can be quantitatively analyzed from these curves to provide guidance for the adjustment and optimization of operating parameters in practical production process.

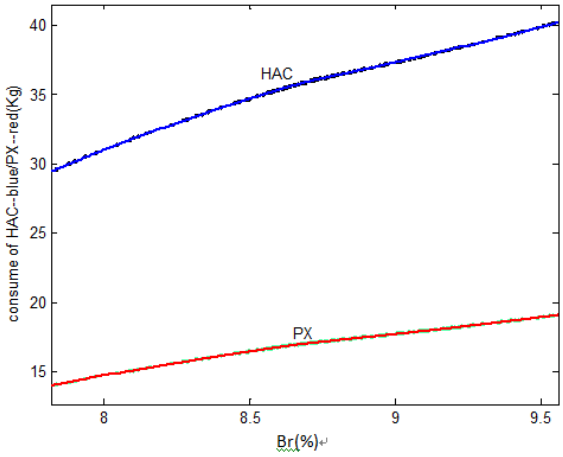

Fig. 9. Effect of Br concentration on oxidative side-reaction.

5. Conclusions

Utilize the algorithm of adaptive affine propagation clustering to divide the field data into two categories. The combustion loss model of HAC and PX was established based on Bayesian filtering. And this neural network model regarded the main adjustable process parameters (residence time, the concentration of Co, Mn, Br, reaction temperature and solvent ratio) as the independent variables and the total content of CO and CO2 as the dependent variable. The simulation results show that the network is well performed. The effects of process operating parameters on HAC and PX combustion loss are analyzed based on the model. Wherein, the effects of residence time, reaction temperature on the combustion loss are bigger, while the effects of other factors are relatively small.

Acknowledgement

The authors would like to thank the financial support provided by Research and Innovation Project for College Graduates of Jiangsu Province of China under Grant CXLX12_0648, National Natural Science Foundation under Grant 61273142 and Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD).

References

[1]. Li Wei, Yang Yupu, Wang Na, LSSVM multi-model regression modeling based on the fuzzy kernel clustering, *Control and Decision*, Vol. 23, Issue 5, 2008, pp. 560-562.

[2]. Suo Xingyi, Shi Hongbo, The soft sensor modeling of sewage treatment process based on multi-model fuzzy kernel clustering algorithm, *East China University of Technology: Natural Science*, Issue 5, 2010, pp. 732-736.

[3]. Petr K., Bogdan G., Sibylle S., Data-driven soft sensors in the process industry, *Computers and Chemical Engineering*, Vol. 33, Issue 4, 2009, pp. 795-814.

[4]. Frey B. J., Dueck D., Clustering by passing messages between data points, *Science*, Vol. 315, Issue 5814, 2007, pp. 972-976.

[5]. Li Lijuan, Song Kun, Zhao Yingkai, ARA fermentation process modeling based on affinity propagation clustering, *Journal of Chemical Industry and Engineering,* Vol. 62, Issue 8, 2011, pp. 2116-2121.

[6]. Deng Weiwei, Yang Hui, A supervision multi-modeling method of affinity propagation clustering, *Shanghai Jiao Tong University*, Vol. 45, Issue 8, 2011, pp. 1172-1175.

[7]. Li X., Su H., Chu J., Multiple Model Soft Sensor Based on Affinity Propagation, Gaussian Process and Bayesian Committee Machine, *Chinese Journal of Chemical Engineering*, Vol. 17, Issue 1, 2009, pp. 95-99.

[8]. Liu Xiuliang, Su Hongye, Chu Jian, Multiple model soft sensor based on online clustering and relevance vector machine, *Control and Instruments in Chemical Industry,* Vol. 35, Issue 3, 2008, pp. 34-37.

[9]. Wang KaiJun, Zhang Junying, Li Dan, Adaptive Affinity Propagation Clustering, *AAS*, Vol. 33, Issue 8, 2007, pp. 1242-1246.

[10]. Shang F., Jiao L. C., Shi J., et al., Fast affinity clustering: a multilevel approach, *Pattern Recognition*, Vol. 45, Issue 1, 2012, pp. 474-486.

[11]. Su Z., Wang P., Shen J., et al., Multi-model strategy based evidential soft sensor model for predicting evaluation of variables with uncertainty, *Applied Soft Computing*, Vol. 11, Issue 2, 2011, pp. 2595-2610.

[12]. Tan S., An effective refinement strategy for KNN text clusterifier, *Export Systems with Applications,* Vol. 30, Issue 2, 2006, pp. 290-298.

[13]. Ge Z., Song Z., A comparative study of just-in-time-learning based methods for online soft sensor modeling, *Chemometrics and Intelligent Laboratory Systems*, Vol. 104, Issue 2, 2010, pp. 306-317.
