2011 American Control Conference
on O'Farrell Street, San Francisco, CA, USA
June 29 - July 01, 2011

Soft Sensing of Sodium Aluminate Solution Component Concentrations via On-line Clustering and Fuzzy Modeling

Wei Wang, Tianyou Chai, *Fellow*, *IEEE*, Lijie Zhao and S. Joe Qin

Abstract—**The component concentrations measurement of sodium aluminate solution are critical to the process of alumina production, they affect the product quality. However, they can not be measured online at present, thus the control and optimal operation is hardly to be achieved. This paper presents an on-line fuzzy modeling method to predict the component concentrations. It includes an on-line clustering approach which can be applied in a general class of fuzzy TKS models. Stable learning algorithms for the premise and the consequence parts of fuzzy rules are also given. A measuring device is developed to achieve the proposed method and industry experiments are conducted in the alumina production process, the predicted results show the effectiveness of the proposed method.**

I. Introduction

Sodium aluminate solution component concentrations are important quality indices for the alumina production. At present, the main chemical compositions of sodium aluminate solution i.e. caustic hydroxide, alumina and sodium carbonate are measured by manually sampling and chemical analysis in a laboratory. Due to hard conditions, they are sampled every two hours, and the results can not guide alumina production on time. Therefore, online measurement of sodium aluminate solution component concentrations becomes especially important.

In the open literature, many methods have been tried, but most of them are based on acid-base titrations and the measurements are off-line [1], [2]. Several on-line devices have been developed [3]. However, they are not adopted into practice owing to the differences between bauxites at home and abroad, and sodium aluminate solution in alumina process has the characteristics of high concentration, high viscosity and easy precipitation which lead to pipeline blocking and so on. With the development of computer technology, a large number of soft sensing models based on intelligent method for concentration measurement are proposed, such as look-up table algorithm [4], mechanism analysis [5], [6], reconstruction algorithm based on the natural cubic spline and neural networks [7] and so on. They are not adopted because of the different solution characters.

Motivated by solving the on-line problem of component concentrations measurement in alumina process, this paper presents a novel soft sensing method for them. In this method, the component concentrations of sodium aluminate solution can be estimated by the solution temperature and conductivity. Based on the analysis of many chemical experiments, we have proved that, to proportional sodium aluminate solution, there is an approximate linear relationship between its temperature and conductivity. Then, an on-line fuzzy model which is composed of on-line clustering and some TSK fuzzy models are proposed to predict the component concentrations of sodium aluminate solution. A stable learning algorithm is also applied for parameters identification. Industrial experiment results show that the proposed method has the merits such as accurate prediction and good tracking performance.

II. Description of Component Concentrations Measurement in Alumina Production

Alumina production is a complex continuous process which aims at converting bauxite to alumina. At present, alumina production mainly depends on alkaline process which uses NaOH or Na_2CO_3 to dissolve bauxite ore. The reaction of alumina in the bauxite ore and the alkali liquor gives sodium aluminate solution. Bayer alumina production is one of the most useful methods in alkaline alumina process, and its flow chart is shown in Fig. 1.

Original ore pulp batching is the first significant procedure in Bayer method and plays an important role in alumina production. The quality of the original ore pulp processing affects the indexes of high-pressure digestion directly and the other subsequent procedures. But the component concentrations of sodium aluminate solution which are manual timed sampling and titration in chemical laboratory can not guide production on time and affect the quality of original ore pulp.

Manuscript received September 20, 2010. This work is supported by the 111 project of Ministry of Education of China (B08015); Natural Science Foundation of China (61020106003); Natural Science Foundation of China (60904079, 61004009); China Postdoctoral Science Foundation (20100471464).

W. Wang and T. Y. Chai are with the State Key Laboratory of Integrated Automation for Process Industry (Northeastern University), Shenyang, 110819, China. (e-mail: wangwei_wei@yahoo.cn; tychai@ mail.neu.edu. cn)

L .J .Zhao is with the Key Laboratory of Integrated Automation of Process Industry, Ministry of Education and Information Engineering School, Shenyang Institute of Chemical Technology. (e-mail: zlj_lunlun@163.com)

S. J. Qin is with the Mork Family Department of Chemical Engineering and Materials Science, the Ming Hsieh Department of Electrical Engineering, and the Daniel J. Epstein Department of Industrial and Systems Engineering, University of Southern California, Los Angeles, CA 90089-1211 USA. (e-mail: sqin@usc.edu)

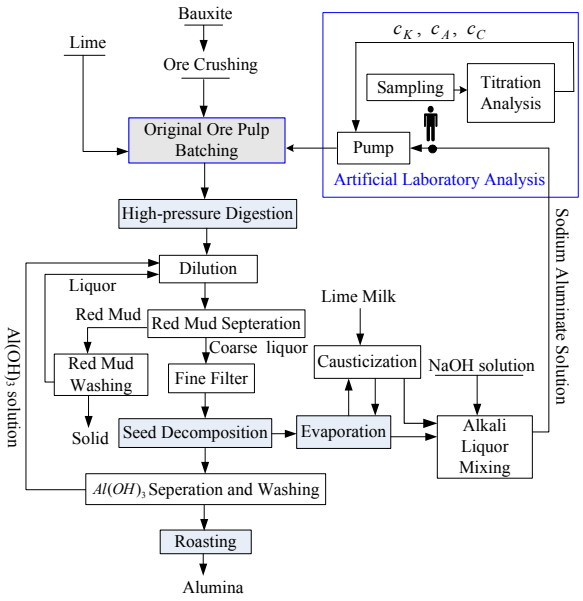

Fig. 1. The flowchart of Bayer alumina production

As the solution source of this procedure is disperse and uncertain, it is difficult to measure the component concentrations using the general process variables like pressure, flow and so on. Therefore, using a new on-line approach to measure the component concentrations of sodium aluminate solution is an important work in alumina production. It is also crucial for reducing labor intensity and increasing composite automatic level of alumina process.

III. Soft Sensing for Component Concentrations in Sodium Aluminate Solution

A. Soft Sensing Strategy

According to the physicochemical property of sodium aluminate solution, the solution conductivity, temperature and its chemical compositions (caustic hydroxide, alumina and sodium carbonate) have the following relationship [8].

$$d = f(T, c_K, c_A, c_C) \tag{1}$$

where $f(.)$ is a nonlinear function, d is the conductivity of sodium aluminate solution, T is the temperature, c_K is the concentration of caustic hydroxide, c_A is the concentration of alumina, c_C is the concentration of sodium carbonate.

Due to the high complexity, there is no mechanism model for the component concentrations of sodium aluminate solution. Fuzzy systems have been widely used because of simple topological structure and universal approximation ability, so an on-line fuzzy modelling method is adopted as shown in Fig.2.

B. Data Preprocessing

In order to overcome the influences of bubbles and impurities in the process of measuring temperature and conductivity, we use median average filter algorithm to deal with the sampling data:

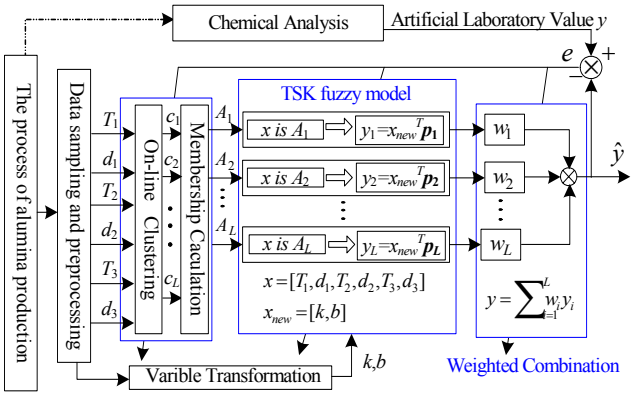

Fig. 2. Strategy of component concentrations measurement

$$T = \frac{\sum_{i=1}^{n} \vec{T}_i - T_{max} - T_{min}}{n-2}, \quad d = \frac{\sum_{i=1}^{n} \vec{d}_i - d_{max} - d_{min}}{n-2} \tag{2}$$

where T and d is the heating (T_1 and d_1), cooling (T_2 and d_2) or mixing (T_3 and d_3) temperature and conductivity, $\vec{T}_i$ and $\vec{d}_i$ (i=1,...,n) is the sampled data from sensor, n is the number of one sampling group, T_{max}, T_{min}, d_{max} and d_{min} is the maximum and minimum temperature and conductivity in one sampling interval, respectively.

C. On-line Clustering

There are many off-line clustering methods, such as fuzzy C-means [9]. Mountain clustering [10], and subtractive clustering [11]. Off-line clustering methods require that data be ready before the modelling which is not easy to obtain in real cases. Therefore, on-line clustering method [12] is used here for temperatures and conductivities classification.

The basic idea of on-line clustering is, if the distance from a point to the center of a group is less than a required length, then the point belongs to this group. When new data are obtained, the center and the group should also change. The Euclidean distance at time k is defined by d_k as follows:

$$d_k = \left(\sum_{j=1}^{6} \left[\frac{x_j(k) - c_{ij}}{x_{j,\max} - x_{j,\min}} \right]^2 \right)^{1/2} \tag{3}$$

where x=[T_1 d_1 T_2 d_2 T_3 d_3] is the data vector, c_i is the center of the ith group G_i, j is the jth variable in the vector. For group G_i, the center c_i is updated as

$$c_{ij} = \frac{1}{l_2^i - l_1^i + 1} \sum_{l=l_1^i}^{l_2^i} x_j(l) \qquad i=1,2,\ldots,L;\ j=1,2,\ldots,6 \tag{4}$$

where l_1^i and l_2^i are the start and end time indices of group G_i respectively, and the time length is $m^i = l_2^i - l_1^i + 1$. The on-line clustering algorithm is as follows:

Step 1: The first group G_1 is composed of only the first data [$T_1(1)$ $d_1(1)$ $T_2(1)$ $d_2(1)$ $T_3(1)$ $d_3(1)$]. So, it is the center of group G_1, i.e., $c_{1j} = x_j(1)$, and $l_1^i = l_2^i = 1, j=1$.

Step 2: When new data are obtained, we use (3) to calcu-

late the distance d_k. If no new data come, goto Step 5.

Step 3: If $d_k \le \theta$ then $x_j(k)$ is kept in group G_i, goto Step 2.

Step 4: If $d_k > \theta$ then $x_j(k)$ belongs to a new group G_{i+1}, and the center of group G_{i+1} is $x_j(k)$, $l_1^i = l_2^i = k$, goto Step 2.

Step 5: Check the distances between all centers c_j, if $d_k \le \theta$, then the two groups are combined into one group, the center of the new group may be any of the two centers.

where θ is the threshold of distance for creating new rules, which is the lowest possible value of similarity. How to choose the user-defined threshold is a trade-off problem. If θ is too small, then there are too many groups and some of them maybe singletons. Conversely, if θ is too large, many objects that are not similar may be partitioned in the same cluster. If the clustering number is more than one, i.e., $L > 1$, then it must be guaranteed that $\theta < d_{\max}$.

D. Variable Transformation

Many experiments have been conducted to measure the temperatures and conductivities of different solution. At last we proved that, to proportional sodium aluminate solution, the temperature and conductivity have such relationship which is similar to a straight line, and this conclusion is also shown in [13]:

$$d = k(c_K, c_A, c_C)T + b(c_K, c_A, c_C) \tag{5}$$

where k is the slope, b is the intercept of the straight line, and they are nonlinear function of c_K, c_A, c_C. Then we transform T_1, d_1, T_2, d_2, T_3, d_3 to k and b, and use them to establish the model of c_K, c_A, c_C. The formula of transformation is as follows:

$$\begin{cases} d_1 = kT_1 + b \\ d_2 = kT_2 + b \\ d_3 = kT_3 + b \end{cases}$$

$$\Rightarrow \begin{bmatrix} k \\ b \end{bmatrix} = \left(\begin{bmatrix} 1 & 1 & 1 \\ T_1 & T_2 & T_3 \end{bmatrix} \begin{bmatrix} 1 & T_1 \\ 1 & T_2 \\ 1 & T_3 \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 & 1 & 1 \\ T_1 & T_2 & T_3 \end{bmatrix} \begin{bmatrix} d_1 \\ d_2 \\ d_3 \end{bmatrix} \tag{6}$$

where d_1, d_2, d_3 are conductivities under the condition of T_1, T_2, T_3 respectively.

E. TSK Fuzzy Modeling Based on Stable Learning Rate

The fuzzy inference system proposed by Takagi, Sugeno and Kang is well known as the TSK model. It provides a powerful tool for modeling complex nonlinear system [14]. Typically, a TSK model consists of IF-THEN rules of the form

$$R_i : \text{if } x_1 \text{ is } A_{i1} \text{ and } \ldots \text{and } x_r \text{ is } A_{ir},$$
$$\text{then } y_i = p_{i0} + p_{i1} x_1 + \ldots + p_{ir} x_r$$
$$\text{for } i = 1,2,\ldots,L$$

where L is the number of rules, x_j for $j = 1,2,\ldots,r$ are input variables, and in this paper $x = [T_1\ d_1\ T_2\ d_2\ T_3\ d_3]$. y_i are local output variables and p_{ij} are real-valued parameters, A_{ij} are fuzzy sets that are characterized by the membership function $A_{ij}(x_j)$

$$A_{ir}(x_r) = \exp\left(-\frac{(x_r - c_{ir})^2}{2\sigma_i^2}\right) \qquad i = 1,\cdots L \tag{7}$$

After variable transformation, $x = [k\ b]$, and the overall output of the model can be written as

$$\hat{Y}(k) = P(k)\Phi[X(k)] \tag{8}$$

where $\Phi(\cdot)$ is a Gaussian function, $\hat{Y}(k) = [\hat{y}_1 \cdots \hat{y}_m]^T$. The qth (q=1,2,…,L) output of the fuzzy model can be calculated by

$$\hat{y}_q = \frac{\sum_{i=1}^{L}\left(\sum_{k=0}^{n} p_{qk}^i x_k\right)\prod_{j=1}^{n}\exp\left(-\frac{(x_j - c_{ji})^2}{\sigma_{ji}^2}\right)}{\left[\sum_{i=1}^{L}\prod_{j=1}^{n}\exp\left(-\frac{(x_j - c_{ji})^2}{\sigma_{ji}^2}\right)\right]} \tag{9}$$

where $x_0 = 1$.

Let $z_i = \prod_{j=1}^{n}\exp\left(-\frac{(x_j - c_{ji})^2}{\sigma_{ji}^2}\right)$, $a_q = \sum_{i=1}^{L}\left(\sum_{k=0}^{n} p_{qk}^i x_k\right) z_i$, $b = \sum_{i=1}^{L} z_i$, then there is

$$\hat{y}_q = \frac{a_q}{b} \tag{10}$$

According to the approximation theorem of fuzzy inference system and the Stone-Weierstrass theorem [15], the identification process can be written as

$$y_q = \frac{\sum_{i=1}^{L}\left(\sum_{k=0}^{n} p_{qk}^{i*} x_k\right)\prod_{j=1}^{n}\exp\left(-\frac{(x_j - c_{ji}^*)^2}{\sigma_{ji}^{*2}}\right)}{\left[\sum_{i=1}^{L}\prod_{j=1}^{n}\exp\left(-\frac{(x_j - c_{ji}^*)^2}{\sigma_{ji}^{*2}}\right)\right]} - \mu_q \tag{11}$$

where p_{qk}^{i*}, c_{ji}^*, σ_{ji}^{*2} are unknown parameters which can minimize the unmodeled dynamic μ_q.

In the case of three independent variables, the Taylor series expansion is as follows:

$$\begin{aligned} y_q + \mu_q = \hat{y}_q &+ \sum_{i=1}^{L}\left(\sum_{k=0}^{n}(p_{qk}^{i*} - p_{qk}^i)x_k\right)\frac{z_i}{b} \\ &+ \sum_{i=1}^{L}\sum_{j=1}^{n}\frac{\partial}{\partial c_{ji}}\left(\frac{a_q}{b}\right)\left(c_{ji}^* - c_{ji}\right) \\ &+ \sum_{i=1}^{L}\sum_{j=1}^{n}\frac{\partial}{\partial \sigma_{ji}}\left(\frac{a_q}{b}\right)\left(\sigma_{ji}^* - \sigma_{ji}\right) + R_{1q} \end{aligned} \tag{12}$$

where R_{1q} is the remainder of the Taylor formula. Using the chain rule, we get

$$\begin{aligned} \frac{\partial}{\partial c_{ji}}\left(\frac{a_q}{b}\right) &= \frac{\partial}{\partial z_i}\left(\frac{a_q}{b}\right)\frac{\partial z_i}{\partial c_{ji}} \\ &= \left(\frac{1}{b}\frac{\partial a_q}{\partial z_i} + \frac{\partial}{\partial z_i}\left(\frac{1}{b}\right)a_q\right)\left(2z_i\frac{x_j - c_{ji}}{\sigma_{ji}^2}\right) \end{aligned} \tag{13}$$

$$= \left(\frac{\sum_{k=0}^{n} p_{qk}^{i} x_k}{b} - \frac{a_q}{b^2} \right) \left(2 z_i \frac{x_j - c_{ji}}{\sigma_{ji}^2} \right)$$

$$= 2 z_i \frac{\sum_{k=0}^{n} p_{qk}^{i} x_k - \hat{y}_q}{b} \frac{x_j - c_{ji}}{\sigma_{ji}^2}$$

and

$$\frac{\partial}{\partial \sigma_{ji}} \left(\frac{a_q}{b} \right) = \frac{\partial}{\partial z_i} \left(\frac{a_q}{b} \right) \frac{\partial z_i}{\partial \sigma_{ji}}$$

$$= 2 z_i \frac{\sum_{k=0}^{n} p_{qk}^{i} x_k - \hat{y}_q}{b} \frac{(x_j - c_{ji})^2}{\sigma_{ji}^3} \tag{14}$$

In matrix form

$$y_q + \mu_q = \hat{y}_q - Z(k) \widetilde{P}_q - D_{Zq} \overline{C}_k E - D_{Zq} \overline{B}_k E + R_{1q} \tag{15}$$

where

$$Z(k) = \left[\frac{z_1}{b} \cdots \frac{z_L}{b} \right]^T, \; P_q = \left[\sum_{k=0}^{n} p_{qk}^{1} x_k \cdots \sum_{k=0}^{n} p_{qk}^{L} x_k \right],$$

$$\widetilde{P}_q = P_q - P_q^*,$$

$$D_{Zq} = \begin{bmatrix} 2 z_1 \dfrac{\sum_{k=0}^{n} p_{qk}^{1} x_k - \hat{y}_q}{b}, \cdots, 2 z_L \dfrac{\sum_{k=0}^{n} p_{qk}^{L} x_k - \hat{y}_q}{b} \end{bmatrix},$$

$$E = [1, \cdots, 1]^T,$$

$$\overline{C}_k = \begin{bmatrix} \dfrac{x_1 - c_{11}}{\sigma_{11}^2} (c_{11} - c_{11}^*) & \dfrac{x_n - c_{n1}}{\sigma_{n1}^2} (c_{n1} - c_{n1}^*) \\ & \ddots \\ \dfrac{x_1 - c_{1L}}{\sigma_{1L}^2} (c_{1L} - c_{1L}^*) & \dfrac{x_n - c_{nL}}{\sigma_{nL}^2} (c_{nL} - c_{nL}^*) \end{bmatrix},$$

$$\overline{B}_k = \begin{bmatrix} \dfrac{(x_1 - c_{11})^2}{\sigma_{11}^3} (\sigma_{11} - \sigma_{11}^*) & \dfrac{(x_n - c_{n1})^2}{\sigma_{n1}^3} (\sigma_{n1} - \sigma_{n1}^*) \\ & \ddots \\ \dfrac{(x_1 - c_{1L})^2}{\sigma_{1L}^3} (\sigma_{1L} - \sigma_{1L}^*) & \dfrac{(x_n - c_{nL})^2}{\sigma_{nL}^3} (\sigma_{nL} - \sigma_{nL}^*) \end{bmatrix},$$

so

$$e_q = \hat{y}_q - y_q = Z(k) \widetilde{P}_q + D_{Zq} \overline{C}_k E + D_{Zq} \overline{B}_k E + \mu_q - R_{1q},$$

In vector form

$$e(k) = \widetilde{P}_k Z(k) + D_Z(k) \overline{C}_k E + D_Z(k) \overline{B}_k E + \varsigma(k) \tag{16}$$

where

$$e(k) = [e_1 \cdots e_m]^T,$$

$$\widetilde{P}_k = \begin{bmatrix} \sum_{k=1}^{n} p_{1k}^{1} x_k - \sum_{k=1}^{n} p_{1k}^{1*} x_k & \sum_{k=1}^{n} p_{mk}^{1} x_k - \sum_{k=1}^{n} p_{mk}^{1*} x_k) \\ \ddots & \\ \sum_{k=1}^{n} p_{1k}^{L} x_k - \sum_{k=1}^{n} p_{1k}^{L*} x_k & \sum_{k=1}^{n} p_{mk}^{L} x_k - \sum_{k=1}^{n} p_{mk}^{L*} x_k) \end{bmatrix},$$

$$D_z(k) = \begin{bmatrix} 2 z_1 \dfrac{\sum_{k=1}^{n} p_{1k}^{1} x_k - \hat{y}_1}{b} & 2 z_L \dfrac{\sum_{k=1}^{n} p_{1k}^{L} x_k - \hat{y}_1}{b} \\ \ddots & \\ 2 z_1 \dfrac{\sum_{k=1}^{n} p_{mk}^{1} x_k - \hat{y}_m}{b} & 2 z_L \dfrac{\sum_{k=1}^{n} p_{mk}^{L} x_k - \hat{y}_m}{b} \end{bmatrix},$$

$\varsigma(k) = \mu - R_1$, $\mu = [\mu_1 \cdots \mu_m]^T$, $R_1 = [R_{11} \cdots R_{1m}]^T$.

By the bound of the Gaussian function Φ and the plant is BIBO stable. The following algorithm makes identification error $e(k)$ bounded [16]

$$p_{qk}^{i}(k+1) = p_{qk}^{i}(k) - \eta_k (\hat{y}_q - y_q) \frac{z_i}{b} x_k \tag{17}$$

$$c_{ji}(k+1) = c_{ji}(k) - 2 \eta_k z_i \frac{\sum_{k=0}^{n} p_{qk}^{i} x_k - \hat{y}_q}{b} \frac{x_j - c_{ji}}{\sigma_{ji}^2} (\hat{y}_q - y_q) \tag{18}$$

$$\sigma_{ji}(k+1) = \sigma_{ji}(k) - 2 \eta_k z_i \frac{\sum_{k=0}^{n} p_{qk}^{i} x_k - \hat{y}_q}{b} \frac{(x_j - c_{ji})^2}{\sigma_{ji}^3} (\hat{y}_q - y_q) \tag{19}$$

Define $D_1 = \frac{z_i}{b} x_k$, $D_2 = \frac{x_j - c_{ji}}{\sigma_{ji}^2}$, $D_3 = \frac{(x_j - c_{ji})^2}{\sigma_{ji}^3}$,

$\Phi_k = \|D_1\|^2 + \|D_z\|^2 \|D_2\|^2 + \|D_z\|^2 \|D_3\|^2$, and let $\eta_k = \frac{\eta}{1 + \Phi_k}$, $0 < \eta \le 1$. We selected a positive-defined matrix L_k

$$L_k = \left\| \widetilde{P}_k \right\|^2 + \left\| \widetilde{C}_k \right\|^2 + \left\| \widetilde{B}_k \right\|^2 \tag{20}$$

According to the updating algorithm and definition,

$$\begin{aligned} \Delta L_k &\le -\frac{\eta}{1 + \Phi_k} \|e(k)\|^2 \left[1 - \frac{\eta}{1 + \Phi_k} \Phi_k \right] + \eta_k \|\varsigma(k)\|^2 \\ &\le -\frac{\eta}{1 + \Phi_k} \|e(k)\|^2 \left[1 - \frac{1}{1 + \Phi_k} \Phi_k \right] + \eta_k \|\varsigma(k)\|^2 \\ &\le -\pi \|e(k)\|^2 + \eta \|\varsigma(k)\|^2 \end{aligned} \tag{21}$$

Because $n[\min[(\widetilde{p}_{qk}^{i})]^2 + \min(\widetilde{c}_{ji}^2) + \min(\widetilde{b}_{ji}^2)] \le L_k \le n[\max[(\widetilde{p}_{qk}^{i})]^2 + \max(\widetilde{c}_{ji}^2) + \max(\widetilde{b}_{ji}^2)]$, where $n[\min[(\widetilde{p}_{qk}^{i})]^2 + \min(\widetilde{c}_{ji}^2) + \min(\widetilde{b}_{ji}^2)]$ and $n[\max[(\widetilde{p}_{qk}^{i})]^2 + \max(\widetilde{c}_{ji}^2) +$

$\max(\tilde{b}_{ji}^2)]$ are κ_∞ -functions, and $\pi\|e(k)\|^2$ is an κ_∞ -function, $\eta\|\varsigma(k)\|^2$ is a κ function. From (21), we know L_k is the function of $e(k)$ and $\delta(k)$, so L_k admits a smooth ISS-Lyapunov function. From Theorem 1, the dynamic of the identification error is input-to-state stable. Because the INPUT $\delta(k)$ is bounded and the dynamic is ISS, the STATE $e(k)$ is bounded [17]. Inequation (21) can be rewritten as

$$\Delta L_k \le -\pi\|e(k)\|^2 + \eta\|\varsigma(k)\|^2 \le -\pi\|e(k)\|^2 + \eta\bar{\varsigma} \qquad (22)$$

where $\bar{\varsigma} = \max_k(\|\varsigma(k)\|^2)$. Summating (22) from 1 to *T,* and by using $L_T > 0$ and L_1 is a constant, we obtain

$$L_T - L_1 \le -\pi\sum_{k=1}^{T}\|e(k)\|^2 + T\eta\bar{\varsigma}$$

$$\pi\sum_{k=1}^{T}\|e(k)\|^2 \le L_1 - L_T + T\eta\bar{\varsigma} \le L_1 + T\eta\bar{\varsigma}$$

The average of the identification error satisfies

$$J = \limsup_{T\to\infty}\frac{1}{T}\sum_{k=1}^{T}\|e(k)\|^2 \le \frac{\eta}{\pi}\bar{\varsigma} \qquad (23)$$

where $\pi = \dfrac{\eta}{[1+\max(\Phi_k)]^2} > 0$, $\bar{\varsigma} = \max_k(\|\varsigma(k)\|^2)$.

IV. Industrial Application

A. Data Sampling

A measuring device is designed and developed to measure temperature and conductivity, so that the soft sensing method proposed in Section 3 can be experimentally evaluated. It is shown in Fig.3.

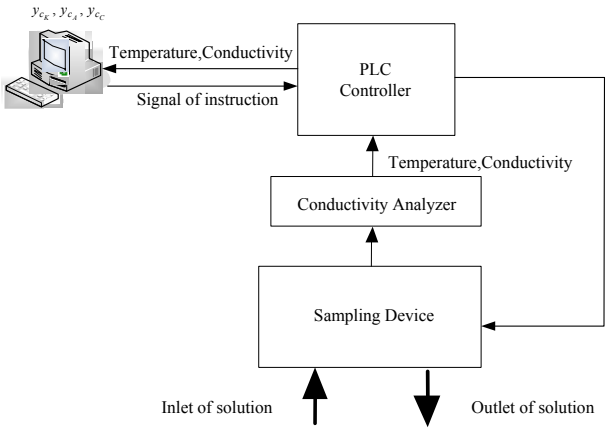

Fig.3 The measurement device for component concentrations

First, sodium aluminate solution enters the sampling device. During the process of heating, cooling and mixing, the conductivity analyzers measure the temperatures and conductivities. The obtained signal is transmitted to the PLC controller, then to the computer which is used to implement the proposed fuzzy modelling method and calculate the component concentrations of sodium aluminate solution.

Applying this device to the process of alumina production and using median average filter to deal with the sampling data, we collect 400 records of sampled data from the industrial process and partial of them are listed in Table I. Among them 300 records are used for modeling, and 100 records for predicting.

TABLE I
PARTIAL OF ORIGINAL DATA

Variables	1	2	3	…	399	400
T_1(°C)	90.97	90.66	89.44	…	92.94	96.13
d_1(ms/cm)	653.44	647.81	634.69	…	660.7	701.72
T_2(°C)	74.48	75.32	74.13	…	74.25	72.52
d_2(ms/cm)	514.77	513.75	501.25	…	486.33	463.2
T_3(°C)	78.92	78.68	78.33	…	78.95	80.53
d_3(ms/cm)	531.72	532.66	528.28	…	521.41	552.19
c_K (g/l)	203	203	206	…	218	217
c_A (g/l)	93.74	93.74	94.40	…	104.27	103.94
c_C (g/l)	30.2	30.2	30.8	…	29.8	29.0

B. Algorithm Realization

Using the on-line clustering algorithm to classify the normalized data and 25 classes are got. Partial of the clustering results are shown in Table II. The threshold of distance θ is 0.2 which is chosen based on the best predicted precision. Then temperatures T_1, T_2, T_3 and conductivities d_1, d_2, d_3 are transformed to slope k and intercept b to establish the TSK fuzzy models. The initial values of the parameters are random, and then they are updated by the rules as equations (17), (18) and (19).

TABLE II
PARTIAL OF CLUSTERING CENTERS

Clustering centers	1	2	3	…	24	25
T_1(°C)	81.33	87.02	81.36	…	90.65	86.26
d_1(ms/cm)	569.67	591.12	520.40	…	582.97	517.18
T_2(°C)	65.26	67.17	66.23	…	67.46	71.51
d_2(ms/cm)	433.16	417.06	390.93	…	434.61	425.08
T_3(°C)	70.11	65.999	71.81	…	75.37	74.29
d_3(ms/cm)	493.89	420.2	466.69	…	448.28	428.83
c_K (g/l)	186.55	200.59	204.53	…	191.55	198.55
c_A (g/l)	71.49	82.81	84.79	…	89.35	95.25
c_C (g/l)	26.1	25.52	26.28	…	23.3	20.8

Most multiple modelling methods are using FCM and TSK fuzzy model, so we compare the measured values to the predicted values by FCM and TSK model without stable learning algorithm, and the proposed on-line fuzzy modeling method. As shown in Fig.4, from the curves we can see the accuracy of the proposed method is relatively high, and good

effect is achieved by using on-line clustering and stable learning.

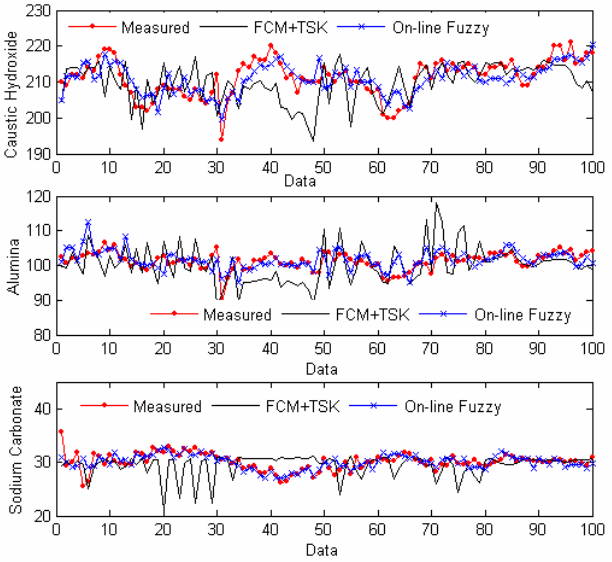

Fig.4 The predicted results of different methods

C. Results Analysis

Compared the proposed method with FCM and TSK model algorithm, the prediction accuracy is shown in Table III, among them there is *MAX*, *MSE* criteria defined as follows:

$$MAXE = \max_{i=1}^{n}(\|\hat{y}_i - y_i\|) \tag{24}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{25}$$

where $\hat{y}$ and y are value of prediction and measurement respectively.

TABLE III
ACCURACY COMPARISON OF DIFFERENT METHODS

Method		FCM+TSK	On-line fuzzy
c_K	*MAXE*	12.56	8.37
	RMSE	5.69	3.29
c_A	*MAXE*	10.51	6.67
	RMSE	5.09	2.73
c_C	*MAXE*	11.88	5.09
	RMSE	2.92	1.05

From the results we can see the accuracy is improved by using nonlinear and on-line fuzzy modeling method. The above analysis indicates that our method is accurate enough to satisfy the industrial requirements in alumina production, and it is also a guideline for process engineers to arrive at the control and optimum operational strategy.

V. CONCLUSION

A novel fuzzy modelling method is proposed for measuring component concentrations of sodium aluminate solution. It presents a new solution for the on-line measurement problem in alumina production process. Considering the same temporal interval between the input and output spaces, an on-line clustering algorithm is applied. Together with some fuzzy TSK models, it is used to capture the nonlinear between the inputs and outputs. The industrial experiment results indicate that the prediction accuracy of the proposed method is relatively high and it is also a basis of control and optimization operation in the alumina production.

REFERENCES

[1] Z. Kowalski and W. Kubiak. "Potentiometric titration of hydroxide, aluminate and carbonate in sodium aluminate solutions," *Analytica Chimica Acta*, vol.140, no.1, pp.115-121, 1982.

[2] Aimin Tan and Cailin Xiao. "Direct determination of caustic hydroxide by a mincro-titration method with dual-wavelength photometric end-point detection," *Analytica Chimica Acta*, vol.341, pp.297-301, 1997.

[3] Farkas Ferenc, Klug Otto and Nemeth Bela. "Measurement methods and instruments developed on basis of physico-chemical properties of sodium-aluminate solutions used in alumina production," *Light Metals*: *Proceedings of Sessions*, *AIME Annual Meeting*, pp.157-170, 1982.

[4] Alessandra Flammini, Daniele Marioli and Andrea Taroni."Application of an Optimal Look-Up Table to Sensor Data Processing," *IEEE Transactions on Instrumention and Measurement*, vol.48, no.4, pp.813-816, 1999.

[5] Bernd Henning, Peter-Christoph Daur and Peter Hauptmann. "In-line concentration measurement in complex liquids using ultrasonic sensors," *Ultrasonics*, vol.38, pp.709-803, 2000.

[6] Guo Wei and Katsunon Shida. "A New Multifunctional Sensor for Measuring the Concentration and Temperature of Dielectric Solution," *SICE 2002*. *Proceedings of the 41st SICE Annual Conference*, vol.1, pp.575- 580, 2002.

[7] Guo Wei and Katsunori Shida, "Estimation of Concentrations of Ternary Solution With NaCl and Sucrose Based on Multifunctional Sensing Technique," *IEEE Transactions on Instrumention and Measurement*, vol.55, no.2, pp.675-681, 2006.

[8] G.R.Browne and C.W.P.Finn. "Determination of Aluminum Content of Bayer Liquors by Electrical Conductivity Measurement," *Metallurgical Transaction B*, vol.8, pp349, 1977.

[9] S. Mitra, Y. Hayashi, Neuro-fuzzy rule generation: survey in soft computing framework, *IEEE Transactions on Neural Networks*, vol.11, no.3, pp.748-769, 2000.

[10] Ronald R Yager and Dimitar P Filev. "Approximate Clustering via the Mountain Method," *IEEE Trans on Systems, Man and Cybernetics*, vol.8, pp.1274-1284, 1994.

[11] Chiu S L. "Fuzzy model identification based on cluster estimation," *Journal of Intelligent and Fuzzy Systems*, vol.2, no.3, pp.267-278, 1994.

[12] Wen Yu, Xiaoou Li. "On-line fuzzy modeling via clustering and support vector machines," *Information Sciences*, vol.178, pp.4264-4279, 2008.

[13] G.R.Browne and C.W.P.Finn. "The effects of aluminum content, temperature and impurities on the electrical conductivity of synthetic bayer liquors," Metallurgical Transaction B, vol.12B, pp.487-492, 1981.

[14] Chang Kyoo Yoo, Peter A, Vanrolleghem and In-Beum Lee. "Nonlinear modeling and adaptive monitoring with fuzzy and multivariate statistical methods in biological wastewater treatment plants," *Journal of Biotechnology*, vol.105, pp.135–163, 2003.

[15] G. Cybenko. "Approximation by superposition of sigmoidal activation function," *Math. Control Sig. Syst*, vol.2, no. 4, pp.303–314, 1989.

[16] Wen Yu, Xiaoou Li, "Fuzzy Identification Using Fuzzy Neural Networks with Stable Learning Algorithms," *IEEE Transactions on Fuzzy Systems*, vol.12, no.3, pp.411-420, 2004.

[17] Wen Yu, Xiaoou Li, "Discrete-time neuro identification without robust modification," *IEE Proceedings-Control Theory and Applications*, vol.150, no.3, pp.311-316, 2003.