Proceedings of the 17th World Congress
The International Federation of Automatic Control
Seoul, Korea, July 6-11, 2008

Cascade Process Modeling with Mechanism-Based Hierarchical Neural Networks

Qiumei Cong * **Tianyou Chai** *,** **Wen Yu** ***

* *Key Laboratory of Process Industry Automation, Northeastern University, Ministry of Education, Shenyang, 110006, China.*
** *Research Center of Automation, Northeastern University, Shenyang, 110006, China.*
*** *Departamento de Control Automatico, CINVESTAV-IPN,A.P. 14-740, Av.IPN 2508, Mexico D.F., 07360, Mexico.*

Abstract: Cascade process, such as wastewater treatment plant, includes many nonlinear subsystems and many variables. When the number of sub-systems is big, the input-output relation in the first block and the last block cannot represent the whole process. In this paper we use two techniques to overcome the above problem. First we propose a new neural model: hierarchical neural networks to identify the cascade process. Then we use serial structural mechanism model based on the physical equations to connect with neural model. A stable learning algorithm and theoretical analysis are given. Finally, this method is used to model a wastewater treatment plant. Real operational data of wastewater treatment plant is applied to illustrate the modeling approach.

1. INTRODUCTION

Neural networks can approximate any nonlinear function to any prescribed accuracy provided with sufficient hidden neurons. It can also be applied to control problem, Speer et al. [1998] presented a neuro-control method for mobile robots. The stability of neural identification is very important in applications. It is well known that normal identification algorithms (gradient descent, least square etc.) are stable under ideal conditions. They might become unstable in the presence of unmodeled dynamics. Lyapunov approach can be used directly to obtain robust training algorithms of continuous-time and discrete-time neural networks. By using passivity theory, Yu et al. [2001], Jin et al. [1999] and Polycarpou et al. [1992] successfully proved that gradient descent algorithms of continuous-time recurrent neural networks were stable and robust to any bounded uncertainties.

Cascade process, such as wastewater treatment plant, includes many nonlinear sub-systems. The input-output relation between the first block and the last block cannot represent the whole process when the number of subsystems are big. Hierarchical models can be used to deal with this problem. Hierarchical fuzzy systems consisting of a number of low-dimensional fuzzy systems have been presented by Raju et al. [1991] and Wang [1997] in order to avoid rule-explosion problem. To the best of our knowledge, the training method of hierarchical neural system still is gradient descent. Normal approaches of analyzing learning algorithms' stability are to investigate identification errors. The key for the training of hierarchical neural model is how to get explicit expression of each internal error.

The common used design for neural modeling is a black-box approach which does not include mechanism knowledge. When we have some prior knowledge, grey-box identification which is the combination of mechanism modeling and intelligent identification may show better results, see Lee et al. [2005]. The mechanism model usually represents the physical properties which can be described by nonlinear functions or nonlinear dynamic equations. The residual uncertainties between the mechanism model and the plant can be modeled by black-box approaches such as neural networks. The above technique is called parallel compensation. On the other hand, the mechanism model and compensator can be in serial form. In this paper, we will discuss this new modeling approach and propose a novel mechanism-based hierarchical neural model. The stability of hierarchical neural model is also proven.

The efficiency of wastewater treatment plant is the most important factor relating to environmental protection. Activated sludge technology is widely used in this field, which the removals of contaminations are implemented by biological reactions of microorganisms in activated sludge. A good model of activated sludge system is beneficial for supervision operation, prediction of effluent quality and realizing model-based optimal control. The model of activated sludge process, mainly 'hard modeling' or parametric modeling is too complex to be described in exact mathematical expression because of unknown behaviors of microorganisms, the complexity of biological reactions, strong nonlinearity and time-varying parameters,

⋆ This work was supported by the State Key Program of National Natural Science of China (No. 60534010), the National Fundamental Research Program of China (No. 2002CB312201), the Funds for Creative Research Groups of China (No. 60521003), the 111 project(B08015), and the National High-tech Program(2006AA040307).

see Henze et al. [1987] and Henze et al. [1999], here nonlinearity is introduced into activated sludge model by Monod equation. But the reliability of these models depends on an increasing number of kinetic and stoichiometric constants in the experimental assessment of different kinds of wastewater. Model order-reduction method based on times scale separation was presented by Weijers [2000]. Lee et al. [2002] introduced high-dimensional nonlinear partial differential equations for wastewater treatment process. Multi-Model interpolation procedure was proposed by Smets et al. [2003]. Multiple linear approximate model with switching algorithm was obtained by Anderson et al. [2000]. Multivariable modeling with subspace-based algorithm had been used and a linear time-invariant dynamic model was given by Sotomayor et al. [2003]. Recently, some intelligent methods are investigated. Zhu et al. [1998] used time-delay neural network to model wastewater treatment plant. The combination of activated sludge model and neural network had improved the modeling accuracy by Lee et al. [2005]. A statistical learning method was employed to develop a learning machine based on the physical mechanisms of biological wastewater treatment systems by Guergachi et al. [2006]. But there are two problems: 1) Wastewater treatment plant is a big cascade process, one model cannot represent a plant with many sub-systems. 2) Grey-box identification is better when we have prior knowledge. In this paper, mechanism-based hierarchical neural networks is applied to model wastewater treatment plant. Real application results show that the new modeling approach is effective for cascade process.

2. CASCADE PROCESS MODELING VIA HIERARCHICAL NEURAL NETWORKS

For each subprocess of cascade process, such as wastewater treatment plant, it can be described using the following general nonlinear dynamic equation

$$\dot{x}_t = f\left(x_t, u_t, \gamma\left(x_t\right)\right) \tag{1}$$

where $x \in R^{n_c}$ is inner state, $u \in R^{m_c}$ is input, f is a vector function, $\gamma\left(x_t\right)$ is unknown term. The mechanism model is $\dot{x}_t = f\left(\cdot\right)$, we will use neural networks to model $\gamma\left(x_t\right)$. We concern discrete-time model of (1) in order to calculate model (1) with computer

$$x(k+1) = \Psi\left[X_1\left(k\right)\right], \quad y\left(k\right) = h\left[X_2\left(k\right)\right] \tag{2}$$

where $\Psi\left(\cdot\right)$ is an unknown nonlinear function representing $\gamma\left(x_t\right)$, $h\left[X_2\left(k\right)\right]$ is known model containing the plant dynamics, and $y \in R^r$ is output, $y\left(k\right) = \left[y_1 \cdots y_r\right]^T$,

$$\begin{array}{c} X_1\left(k\right) = \left[x\left(k\right), x\left(k-1\right), \cdots x\left(k-n_1\right),\right. \\ \left. u\left(k\right), u\left(k-1\right), \cdots u\left(k-m_1\right)\right]^T \\ X_2\left(k\right) = \left[x\left(k\right), x\left(k-1\right), \cdots x\left(k-n_2\right),\right. \\ \left. u\left(k\right), u\left(k-1\right), \cdots u\left(k-m_2\right)\right]^T \end{array} \tag{3}$$

where $n_c \times n_1 + m_c \times m_1 = n_d$. It is a NARMAX model. If h^{-1} exists (we will not calculate it)

$$\begin{array}{c} X_1\left(k\right) = \left[h^{-1}\left[y\left(k\right)\right], h^{-1}\left[y\left(k-1\right)\right],\right. \\ \left. \cdots h^{-1}\left[y\left(k-n_1\right)\right], u\left(k\right), u\left(k-1\right), \cdots u\left(k-m_1\right)\right]^T \\ X_2\left(k\right) = \left[h^{-1}\left[y\left(k\right)\right], h^{-1}\left[y\left(k-1\right)\right],\right. \\ \left. \cdots h^{-1}\left[y\left(k-n_2\right)\right], u\left(k\right), u\left(k-1\right), \cdots u\left(k-m_2\right)\right]^T \end{array} \tag{4}$$

The basic idea of the mechanism-based neural model is that we use neural networks to identification the unknown parts of the plant, then the outputs x of neural networks are sent to the mechanism model as parts of inputs X_2. Here single-output multilayer perceptrons is adopted:

$$\hat{x}\left(k\right) = W\left(k\right)\phi\left[V\left(k\right)\hat{X}\left(k\right)\right] \tag{5}$$

where the neural output $\hat{x}\left(k\right)$ is scalar, input vector $\hat{X}\left(k\right) \in R^n$ is defined as

$$\begin{array}{c} \hat{X}\left(k\right) = \left[y_1\left(k\right), y_1\left(k-1\right), \cdots y_r\left(k-n_1\right),\right. \\ \left. u\left(k\right), u\left(k-1\right), \cdots u\left(k-m_1\right)\right]^T \end{array}$$

the weights in output layer are $W\left(k\right) \in R^{1\times m}$, the weights in hidden layer are $V\left(k\right) \in R^{m\times n}$, ϕ is m-dimension vector function. The typical presentation of the element $\phi_i(\cdot)$ is sigmoid function. The mechanism model is

$$\hat{y}\left(k\right) = h\left[\hat{X}_2\left(k\right)\right] \tag{6}$$

First we discuss how to model one block of cascade process with one mechanism-based neural network. The objective is to update the weights of neural network so that the error between the output of neural model (6) and the output of the plant (2) is minimized. The performance index is defined as

$$J = \frac{1}{2}\left\|e\left(k\right)\right\|^2 = \frac{1}{2r}\sum_{i=1}^{r}(\hat{y}_i(k) - y(k))^2, e\left(k\right) = \hat{y}\left(k\right) - y\left(k\right)$$

We use gradient descent learning law $\Delta w_i\left(k\right) = -\eta\frac{\partial J}{\partial w_i}$, $\Delta v_{i,j}\left(k\right) = -\eta\frac{\partial J}{\partial v_{i,j}}$, where $W\left(k\right) = \left[w_i\left(k\right)\right]$, $V\left(k\right) = \left[v_{i,j}\left(k\right)\right]$. Now we use the chain rule

$$\begin{aligned} \frac{\partial J}{\partial w_i} = \frac{\partial J}{\partial \hat{y}}\frac{\partial \hat{y}}{\partial \hat{x}}\frac{\partial \hat{x}}{\partial w_i} &= \sum_{t=1}^{r}\left(e_t(k)\frac{\partial \hat{y}_t}{\partial \hat{x}}\right)\frac{\partial \hat{x}}{\partial w_i} \\ &= e^T(k)h'\phi_i \end{aligned} \tag{7}$$

where ϕ_i is the output at the ith hidden node. The same method can be applied to $V\left(k\right)$. The chain rule of the weights at hidden layer can be given as:

$$\frac{\partial J}{\partial v_{i,j}} = \frac{\partial J}{\partial \hat{y}}\frac{\partial \hat{y}}{\partial \hat{x}}\frac{\partial \hat{x}}{\partial \phi_i}\frac{\partial \phi_i}{\partial v_{i,j}} = \sum_{t=1}^{r}(e_t(k)h'_t)w_i\phi'_i x_j \tag{8}$$

So for one block, learning law is

$$\begin{array}{l} W\left(k+1\right) = W\left(k\right) - \eta e^T\left(k\right)h'\phi^T \\ V\left(k+1\right) = V\left(k\right) - \eta e^T\left(k\right)h'\phi' W^T(k)\hat{X}^T \end{array} \tag{9}$$

where η is learning rate.

Secondly, we discuss modeling of cascade process with hierarchical neural networks. Cascade nonlinear process can be described as

$$\begin{array}{ll} \dot{x}_t^{(1)} = f\left(x_t^{(1)}, u_t^{(1)}\right), & y_t^{(1)} = h\left(x_t^{(1)}\right) \\ \dot{x}_t^{(2)} = f\left(x_t^{(2)}, y_t^{(1)}\right), & y_t^{(2)} = h\left(x_t^{(2)}\right) \\ \vdots & \\ \dot{x}_t^{(p)} = f\left(x_t^{(p)}, y_t^{(p-1)}\right), & y_t^{(p)} = h\left(x_t^{(p)}\right) \end{array}$$

Corresponding discrete-time format is as follows

$$\begin{array}{ll} x^{(1)}(k+1) = \Psi\left[X_1^{(1)}\left(k\right)\right], & y^{(1)}\left(k\right) = h\left[X_2^{(1)}\left(k\right)\right] \\ \vdots & \\ x^{(p)}(k+1) = \Psi\left[X_1^{(p)}\left(k\right)\right], & y^{(p)}\left(k\right) = h\left[X_2^{(p)}\left(k\right)\right] \end{array} \tag{10}$$

The modeling scheme is shown as Fig. 1. We can train each block independently if we know the modeling errors be-

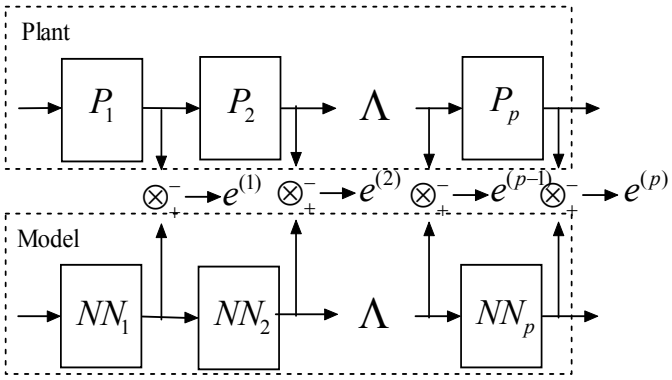

Fig. 1. Cascade process modeling with hierarchical neural networks

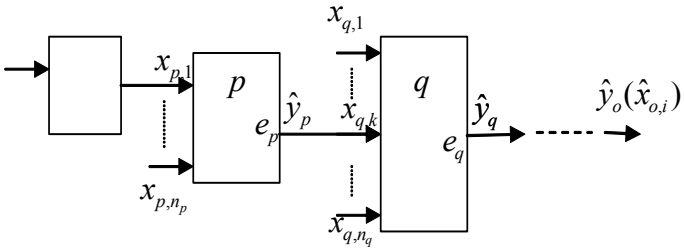

Fig. 2. General hierarchical neural networks

tween neural networks and corresponding process blocks, $e^{(1)}, e^{(2)} \cdots e^{(p)}$.

Now we discuss how to propagate virtual modeling errors. The general hierarchical neural networks in Fig. 2 are Considered. When the errors are back propagated from block q to block p, how can we get $e^{(p)}$ from the hierarchical model if $e^{(q)}$ is known.

We can obtain the relationship between $e^{(q)}$ and $e^{(p)}$ by the chain rule:

$$\begin{array}{c} e^{(q)}(k) = [e_i^{(q)}(k)] \\ e_i^{(q)}(k) = \widehat{y}_i^{(q)}(k) - y_i^{(q)}(k),\ i = 1, \cdots, r \end{array} \tag{11}$$

where $\widehat{y}_i^{(q)}(k)$ is output of the qth neural network, $y_i^{(q)}(k)$ is virtual output of the qth block of the plant.

We use the error of the qth block to calculate $\frac{\partial J}{\partial w_i} = \frac{\partial J}{\partial \hat{y}} \left(\frac{\partial \hat{y}}{\partial \hat{x}} \frac{\partial \hat{x}}{\partial w_i} \right)$ for the pth block, so

$$\frac{\partial J}{\partial w_i} = \frac{\partial J}{\partial \hat{y}^{(q)}} \frac{\partial \hat{y}^{(q)}}{\partial \hat{y}^{(p)}} \left(\frac{\partial \hat{y}^{(p)}}{\partial \hat{x}} \frac{\partial \hat{x}}{\partial w_i} \right)$$

We can see the last two terms is the same as single block case, $\frac{\partial J}{\partial \hat{y}^{(q)}}$ is the virtual error of the qth block, $\frac{\partial J}{\partial \hat{y}^{(q)}} = e^{(q)}$, the term $\frac{\partial \hat{y}^{(q)}}{\partial \hat{y}^{(p)}}$ realize the error backpropagation

$$\frac{\partial \hat{y}^{(q)}}{\partial \hat{y}^{(p)}} = \frac{\partial \hat{y}^{(q)}}{\partial \hat{x}^{(q)}} \frac{\partial \hat{x}^{(q)}}{\partial \phi^{(q)}} \frac{\partial \phi^{(q)}}{\partial \hat{y}^{(p)}} = h^{(q)\prime} \Phi^{(q)\prime} W^{(q)T} V^{(q)}$$

Finally, we obtain

$$e^{(p)} = e^{(q)T} h^{(q)\prime} \Phi^{(q)\prime} W^{(q)T} V^{(q)} \tag{12}$$

The proposed algorithm can be extended to general hierarchical structure, see Fig. 2. We explain the training procedure.

(1) (Feedforward propagation) According to the structure of hierarchical neural networks, we calculate the output of each mechanism-based neural network as (5) and (6). The output of multilayer neural networks should be the inputs of the next level.
(2) Calculate the modeling error for each block. We start from the last block, the identification error is

$$e(k) = \hat{y}(k) - y(k) \tag{13}$$

where $\hat{y}$ is the output of the whole hierarchical neural networks, y is the final output of the cascade plant. Then we propagate the error back from the structure of the hierarchical neural networks as in Fig. 2. We can calculate the error of the block p (defined as $e^{(p)}$) from its former block q (defined as $e^{(q)}$) by (12).
(3) Train the weights matrices for each block independently. Backpropagation-like algorithm of the pth block is

$$\begin{array}{l} W^{(p)}(k+1) = W^{(p)}(k) - \eta e^{(p)T} h^{(p)\prime} \phi^{(p)T} \\ V^{(p)}(k+1) = V^{(p)}(k) - \eta e^{(p)T} h^{(p)\prime} \phi^{(p)\prime} W^{(p)T} \hat{X}^T \end{array} \tag{14}$$

3. STABLE LEARNING

Gradient descent algorithm as in (14) is a general algorithm that includes least-square and backpropagation as special cases. The fixed learning rate yields poor performance. In contrast, time-varying learning rate shows faster convergence, see Ljung et al. [1983]. While it results in slow convergence to bad solutions when time is small and cannot guarantee stability. In this paper we propose a new time-varying learning rate, which takes advantages of Yu et al. [2003] and Moody et al. [1989] to assure stable and fast learning.

We first consider one mechanism-based neural network as (5) and (6). According to the Stone-Weierstrass theorem, see Cybenko [1989], one sub-block of the cascade process (10) can be written as

$$x(k) = W^* \phi[V^* \hat{X}(k)] - \mu(k) \quad y(k) = h[X_2(k)]$$

where W^* and V^* are the unknown weights matrices which may minimize the modeling error $\mu(k)$ of this block. We use Taylor series to obtain error dynamic. In the case of two independent variables, smooth function f has the following Taylor series expansion near the point $\left[x_1^0, x_2^0\right]$,

$$\begin{array}{c} f(x_1, x_2) = \sum_{k=0}^{l-1} \frac{1}{k!} [(x_1 - x_1^0) \frac{\partial}{\partial x_1^{'}} \\ + (x_2 - x_2^0) \frac{\partial}{\partial x_2^{'}}]^k f(x_1^{'}, x_2^{'})_{\substack{x_1^{'} = x_1^0 \\ x_2^{'} = x_2^0}} + \varepsilon \end{array} \tag{15}$$

where ε is the remainder of the Taylor formula. For neural network $\hat{x}(k) = W(k) \phi \left[V(k) \hat{X}(k) \right]$, if we let x_1 and x_2 correspond to $W(k)$ and $V(k)$, x_1^0, x_2^0 correspond to W^* and V^*, we have

$$\begin{array}{c} \hat{x}(k) = W^* \phi[V^* \hat{X}(k)] + [W(k) - W^*] \phi \\ + [V(k) - V^*] \phi' W^T(k) \hat{X}^T(k) + \varepsilon(k) \\ = x(k) + [W(k) - W^*] \phi \\ + [V(k) - V^*] \phi' W^T(k) \hat{X}^T(k) + \varepsilon(k) + \mu(k) \end{array}$$

So error of single-output multilayer neural network

$$e_n(k) = \tilde{W}(k)\,\phi\left[V(k)\,\hat{X}(k)\right] + \tilde{V}(k)\,\phi' W^T(k)\,\hat{X}^T(k) + \zeta(k) \tag{16}$$

where $\tilde{W}(k) = W(k) - W^*$, $\tilde{V}(k) = V(k) - V^*$, $\zeta(k) = \varepsilon(k) + \mu(k)$,

$$\begin{aligned} e_n(k) &= \hat{x}(k) - x(k) \\ e(k) &= \hat{y}(k) - y(k) = h\left[\hat{X}_2(k)\right] - h\left[X_2(k)\right] \end{aligned} \tag{17}$$

In this paper we are only interested in open-loop identification. We may assume that the plant (10) is bounded-input and bounded-output stable, *i.e.*, $y(k)$, $x(k)$ and $u(k)$ in (10) are bounded. Because of the boundness of the sigmoid function ϕ, we may assume that $\varepsilon(k)$ and $\mu(k)$ in (16) are bounded. The following theorem gives a stable backpropagation-like algorithm for the training of discrete-time multilayer neural network.

Theorem 1. If we use the multilayer neural network (5) to identify nonlinear plant (2), the following backpropagation-like algorithm can make identification error $e(k)$ bounded

$$\begin{aligned} W(k+1) &= W(k) - \eta_k e_n(k)\,\phi\left[V(k)\hat{X}(k)\right] \\ V(k+1) &= V(k) - \eta_k e_n(k)\,\phi' W^T(k)\,\hat{X}^T(k) \end{aligned} \tag{18}$$

where $\eta_k = \dfrac{\eta_0}{1 + \left\|\phi' W^T(k)\hat{X}^T(k)\right\|^2 + \|\phi\|^2}$, $0 < \eta_0 \le 1$.

The average of the identification error satisfies

$$\limsup_{T\to\infty} \frac{1}{T}\sum_{k=1}^{T} \|e_n(k)\|^2 \le \frac{\bar{\zeta}}{1-\eta_0} \tag{19}$$

where $\overline{\zeta} = \max_k \left[\zeta^2(k)\right]$

4. APPLICATION TO WASTEWATER TREATMENT

The wastewater treatment plant used as a case study throughout this paper consists of two biodegradation tanks in series in a back-to-back scheme with multi recycle streams, see Fig. **??**. Denitrification and nitrification occur in anoxic and aerated tanks in order to remove nitrogenous and carbonaceous contaminations respectively. Dissolved oxygen in aerated tank is under tight control by a PI controller towards set point $S_O^* = 2mg/L$ through controlling the air flowrate Q_{air} into aerated tank. The secondary settler is the device for separation of liquid and solid, which has the ten-layer structure with double exponential settling rate of solid components within each layer, see Takacs et al. [1991]. Mixed liquid is recycled to the inlet of anoxic tank at the rate of Q_r to guarantee smooth denitrification. Activated sludge is recycled to the inlet of anoxic tank at the rate of Q_R to retain the concentration of biomass sufficient for biological reactions. The surplus activated sludge is discharged from the bottom of the secondary settler at the rate of Q_w to maintain appropriate organic load (F/M). Besides, the influent is fed into anoxic tank at the rate of Q_{in}, and the effluent is discharged to recipient river at the rate of Q_e from the top of the secondary settler.

Nitrification reactions aim to oxidate ammonia into nitrite or nitrate, and consume biodegradable COD (Chemical Oxygen Demand) by the heterotrophic organisms; In this tank, two major reaction processes occurred are

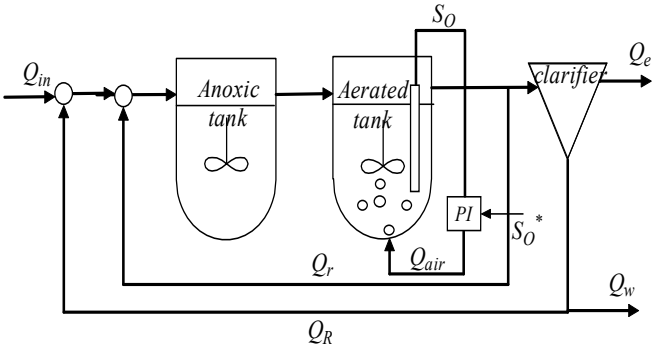

Fig. 3. Plant configuration

$$\begin{gathered} NH_4^+ + 1.5O_2 \rightarrow NO_2^- + H_2O + 2H^+ \\ NO_2^- + 0.5O_2 \rightarrow NO_3^- \\ COD + O_2 \rightarrow CO_2 + H_2O + AS \end{gathered} \tag{20}$$

where COD denotes carbonous contamination, AS denotes activated sludge. Nitrite or nitrate that is recycled from the aerobic tanks are deoxidized into nitrogen air by the autotrophic organisms in denitrification phase which results in release of alkalinity and hence increase of pH, the reaction is

$$2NO_3^- + 2H^+ \rightarrow N_2 + H_2O + 2.5O_2 \tag{21}$$

These reactions are implemented by microorganisms existing in activated sludge under appropriate circumstances. Water quality indices such as COD, BOD_5 (Biological Oxygen Demand), $NH_4 - N$ (ammonia), nitrate and SS (Suspended Solid) are decomposed into those components of ASM1 listed in Table 1 according to rough proportion. The general mass-balance equation for a component Z in the typical activated sludge model is given as follows:

$$\frac{dZ(k)}{dt} = r(k) + \frac{1}{V}(F_{in}Z_{in}(k) - F_{out}Z(k)) \tag{22}$$

where F_{in} and F_{out} are influent and effluent flows respectively, Z is component concentration listed in Table 1, Z_{in} and Z are corresponding influent and effluent concentrations, V is the volume of reactor, $r(k)$ is reaction rate of component. The description of reaction rates in (22) as well as kinetic and stoichiometric parameters are given by Henze et al. [1987].

In this section, we will use real data of North wastewater treatment plant (NWTP) in Shenyang, China and the mechanism-based hierarchical neural networks proposed in this paper to model wastewater treatment plant. The plant including 6 water lines in parallel mainly deals with municipal wastewater, where No.1~3 lines use traditional activated sludge technics and No.4~6 lines use anoxic-aerobic (A/O) technics with nitrogen removal. The case study focuses on A/O process of No.4 line which consists of an anoxic reactor with the volume of $7772.5m^3$, an aerobic reactor with the volume of $10326m^3$, and a secondary settler with the height of $4.2m$ and the diameter of $57m$. The NWTP set-up is essentially similar to Fig. **??**. The main operational parameters from 1999 to 2004 are listed in Table 2.

Table.2. The main operational parameters from 1999 to 2004

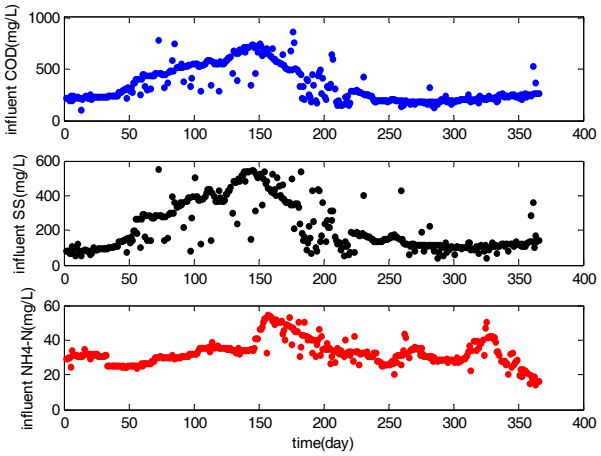

Fig. 4. Influent qualities in 2003

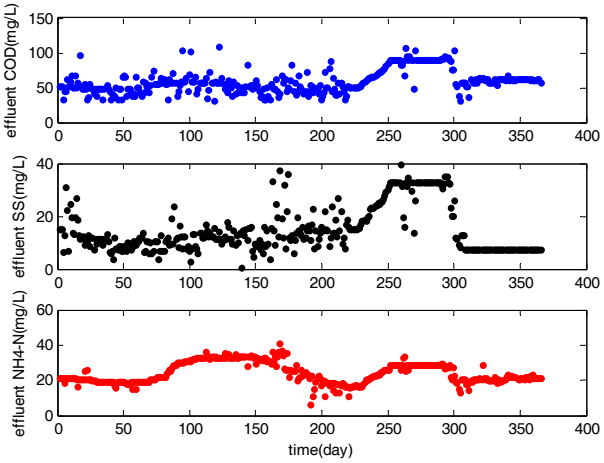

Fig. 5. Effluent qualities in 2003

Description	COD	BOD_5	SS	$NH_4{-}N$
influent quality(mg/L)	292	98	123	45
effluent quality(mg/L)	61	17	17	22
removal ratio$(\%)$	80	83	86	51

Unit	Q_{in}	Q_R	Q_r	Q_w
m^3/h	2278~3700	1389~3700	2778~7400	760

Steady simulation should be implemented before dynamic one to obtain the initial values of dynamic simulation. The average influent qualities are taken as inputs of steady simulation. The resulting steady values of anoxic and aerobic reactors are shown in Table 3:

Table.3.Steady values of anoxic and aerobic reactors

	S_S	X_{BH}	X_S	X_I	S_{NH}	S_I
anoxic	1.2518	3249	74.332	642.4	7.9157	38.374
aerobic	0.6867	3244.8	47.392	643.36	0.1896	38.374
S_{ND}	X_{ND}	S_O	X_{BA}	S_{NO}	X_P	$Salk$
0.7868	5.7073	0.0001	220.86	3.9377	822.19	4.9261
0.6109	3.7642	1.4988	222.39	12.819	825.79	3.7399

There are many missing data and outliers in measured all-year data in 2003. Principal component analysis based on EM (Expectation-Maximum) algorithm with iterative robust least square method was employed to reduce the influence of noise, outliers and missing data inherent in measured values, see Zhao et al. [2005]. Pretreated real data are shown as Fig. 4 and Fig. 5. Heavy repair of equipment led to mass missing data of effluent qualities from August to October. Real data still have large offset after being coordinated.

Real data of influent COD, SS display ascending trend firstly and then descending one. It is the increase of influent loads brought by the snow. NH_4-N is not effected by this phenomenon, while shows a fluctuation at the first ten days of May. Effluent COD, SS are relatively smooth after A/O biological treatment. It indicates the plant possesses redundancy ability to fluctuation of water qualities. The removal efficiency of NH_4-N is low relative to other contaminations, which should be improved by optimal control using existing equipment. Dynamic simulation uses the resulting steady values of steady simulation as initial values with hydraulic residence time of $10.8h$ and sludge age of $15d$.

100 input/output data pairs from the records in 2003 are used as the training data, and 30 input/output pairs as the testing data. The input data of NWTP are $\Omega = [S_I\ S_S\ X_I\ X_S\ X_P\ X_{BH}\ X_{BA}\ S_{NO}\ S_{NH}\ S_O\ S_{ND}\ X_{ND}\ S_{alk}\ Q_{in}\ Q_w]$ converted from influent indices $[COD_i, NH_{4,i}]$ of No.4 line and control inputs $U = [Q_{in}, Q_w]$, the output is effluent indices $[COD_e]$. The relationship between components and water quality COD is:

$$COD = S_S + S_I + X_S + X_{BH} + X_{BA} + X_P + X_I$$

Modeling NWTP in Fig. **??** via hierarchical neural networks is to use neural networks to identify the uncertain reaction rates of cascade reactor models. The inputs of anoxic neural blocks are input data Ω of NWTP, the outputs are reaction rates $[r_2\ r_7\ r_8]$ under anoxic condition and the whole outputs of mechanism-based anoxic block is effluent components of anoxic reactor. Here we use three single-output multilayer neural networks as anoxic neural blocks, then the outputs of neural blocks are used as the inputs of corresponding mechanism model. The inputs of aerobic neural blocks are from the outputs of mechanism-based anoxic block, which is in accordance with the structure of hierarchical neural networks and its outputs are reaction rates $[r_2\ r_6\ r_9\ r_{10}]$ under aerobic condition. Four single-output multilayer neural networks are used as aerobic neural blocks. Similarly, the outputs of mechanism-based aerobic block are effluent components of corresponding model, then taken as the inputs of secondary settler. The final output of cascade process is effluent COD. 30 nodes in hidden layer of each neural network are selected, and the learning algorithm is error back-propagation learning law similar to (14). Activation vector functions ϕ are selected as $\phi_i(\cdot) = \tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$, so $\phi_i'(\cdot) = \sec h(x) = \frac{2}{e^x + e^{-x}}$. The modeling results of effluent COD are shown in Fig.6.

The modeling error is introduced by the following reasons: 1) Wastewater treatment process suffers from external disturbances such as temperature, influent qualities, influent flow, operational status as well as internal factors like microorganism activities etc.; 2) Each reactor is supposed to be completely mixed regardless of the influences of stream status like back-mixing on the microorganism distributions

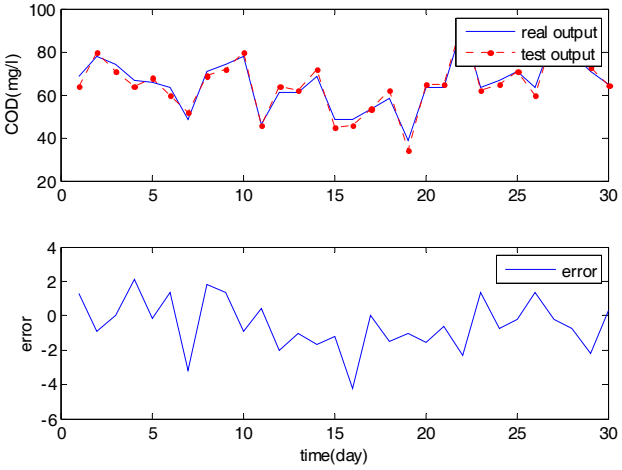

Fig. 6. The modeling results via hierarchical neural networks

and reaction rates; 3) The number of imaginary reactor series is less, which restricts the model precision; 4) It is supposed that there are no biological reactions in secondary settler; 5) Secondary settler adopts solid flux model of one-dimension gravity, regardless of the influences of stream status like diffusion on the solid settling velocity; 6) The number of separated settler layers is small, while 30~50 layers are appropriate for considerable accuracy; 7) Real data are corrupted by noise as well as missing data and outliers. 8) The offset is introduced when real data are coordinated using statistical methods. These influencing factors to modeling error can be minimized further via other intelligent methods as error compensator like neural network, fuzzy rule and expert system.

5. CONCLUSIONS

The contributions of this paper are that a new neural model for cascade process is proposed based on a mechanism model which has hierarchical structure; the mechanism model is connected with neural network in serial form; the stability of the identification algorithm is proven. Real data of wastewater treatment plant is applied to illustrate the modeling approach.

REFERENCES

J.S. Anderson, Hyunook Kim, T.J. McAvoy, and O.J. Hao. Control of an alternating aerobic–anoxic activated sludge system - Part 1: development of a linearization-based modeling approach. *Control Engineering Practice*, 8:271-278, 2000.

G.Cybenko. Approximation by superposition of sigmoidal activation function. *Math.Control, Sig Syst*, 2: 303-314, 1989.

A.A. Guergachi, and G.G. Patry. Constructing a model hierarchy with background knowledge for structural risk minimization: Application to biological treatment of wastewater. *IEEE Trans. Systems, Man and Cybernetics - Part A: Systems and Humans*, 36:1-11, 2006.

M. Henze, C. Grady, W. Gujer, G.v.R. Marais, and T. Matsuo. *Activated Sludge Model No.1.* IAWQ Scientific and Technical Report No. 1. IAWQ, London, 1987.

M. Henze, W. Gujer, T. Mino, T. Matsuo, M.C. Wentzel, G.v.R Marais, and M.C.M.van Loosdrecht. Activated Sludge Model No.2d, ASM2D. *Wat. Sci. Tech.*, 39:165–182, 1999.

L. Jin and M.M. Gupta. Stable dynamic backpropagation learning in recurrent neural networks. *IEEE Trans. Neural Networks*, 10:1321-1334, 1999.

D.S. Lee, P.A. Vanrolleghem, and J.M. Park. Parallel hybrid modeling methods for a full-scale cokes wastewater treatment plant. *Journal of Biotechnology*, 115:317–328, 2005.

T.T. Lee, F.Y. Wang, and R.B. Newell. Robust model-order reduction of complex biological processes. *Journal of Process Control*, 12:807–821, 2002.

L. Ljung and T. Soderstrom. *Theory and Proactice of Recursive System Identification*, The MIT Press, 1983.

J. Moody and C.J. Darken. Fast learning in networks of locally-tuned processing units. *Neural Computation,* 1:281-294, 1989.

M.M. Polycarpou and P.A. Ioannou. Learning and convergence analysis of neural-type structured networks. *IEEE Trans. Neural Networks*, 3:39-50, 1992.

G.V.S. Raju, J. Zhou, and R.A. Kisner. Hierarchical fuzzy control. *Int. J. of Control*, 54:1201-1216, 1991.

I.Y. Smets, J.V. Haegebaert, R. Carrette, and J.F.V. Impe. Linearization of the Activated Sludge Model ASM1 for fast and reliable predictions. *Wat. Res.*, 37:1831-1851, 2003.

O.A.Z. Sotomayor, S.W. Park, and C. Garcia. Multivariable identification of an activated sludge process with subspace-based algorithms. *Control Engineering Practice*, 11:961-969, 2003.

R.K. Speer, and W.E. Moore. Hierarchical artificial neural network architecture. *IEEE World Congress on Computational Intelligence.* volume 3, Pages 2146-2150, May 1998.

I. Takacs, G.G. Patry, and D. Nolasco. A dynamic model of the clarification thickening process. *Wat. Res.*, 25:1263-1271, 1991.

L.X. Wang. *A course in fuzzy systems and control.* Prentince Hall Inc., 1997.

S.R Weijers. Modelling, identification and control of activated sludge plants for nitrogen removal. PhD thesis, Technology University of Eindhoven, The Netherlands, 2000.

W. Yu, and X. Li. Discrete-time neuro identification without robust modification. *IEE Proceedings - Control Theory and Applications*, 150:311-316, 2003.

W. Yu, A.S. Poznyak, and X. Li. Multilayer dynamic neural networks for nonlinear system on-line identification. *Int. J. of Control*, 74:1858-1864, 2001.

W. Yu and X. Li. Some new results on system identification with dynamic neural networks. *IEEE Trans. Neural Networks*, 12:412-417, 2001.

L.J. Zhao, and T.Y. Chai. Wastewater BOD forecasting model for optimal operation using robust time-delay neural network. *Lect. Notes Comput. Sci.*, 3498:1028-1033, 2005.

J.B. Zhu, J. Zurcher, M. Rao, and M.Q-H. Meng. An online wastewater quality predication system based on a time-delay neural network. *Engineering Applications of Artificial Intelligence*, 11:47-758, 1998.