17

Reinforcement Learning-Based Supervisory Control Strategy for a Rotary Kiln Process

Xiaojie Zhou, Heng Yue and Tianyou Chai
Key Laboratory of Integrated Automation of Process Industry, Northeastern University
P. R. China

1. Introduction

Rotary kiln is a kind of large scale sintering device widely used in metallurgical, cement, refractory materials, chemical and environment protection industries. Its complicated working mechanism includes physical change and chemical reaction of material, procedure of combustion, thermal transmission among gaseous fluid, solid material fluid and the liner. The automation problem of such processes remains unsolved because of the following inherent complexities. A rotary kiln is a typical distributed parameter system with correlative temperature distribution of gaseous phase and solid phase along its axis direction. Limited by device rotation and technical design, sensors and actuators can be installed only at the kiln head and kiln tail, and lumped parameter control strategies are employed to deal with distributed parameter problems. Thus the rotary kiln process is a multivariable nonlinear system with strong coupling, large lag and uncertain disturbances. Moreover, the key controlled variable of burning zone temperature is measured with serious disturbances. Most of rotary kilns are still under manual control with human operator observing the burning status. As a result, the product quality is hard to be kept consistent and energy consumption remains high, kiln liner is easy to wear out, the kiln running rate and yield is low.

Although several advanced control strategies including fuzzy control (Holmblad & Østergaard, 1995) , intelligent control (Jarvensivu et al., 2001a; Jarvensivu et al., 2001b) and predictive control (Zanovello & Budman, 1999) have been introduced into process control of rotary kiln, all these researches focused on stabilizing some key controlled variables but are valid only for cases that boundary conditions do not change frequently. As a matter of fact, the boundary conditions of a rotary kiln often change. For example, the material load, water content and components of the raw material slurry vary frequently and severely. Moreover, the offline analysis data of components of raw material slurry reach the operator with large time delay. Thus conventional control strategy cannot reach automatic control and keep the product quality consistent. To deal with the complexity of operation conditions, the authors have proposed an intelligent control system based on human-machine interaction for an alumina rotary kiln in (Zhou et al., 2004; Zhou et al., 2006), in which human intervention function was design so that, if the operation condition changed largely, the human operator observing burning status can intervene the control actions when the system is in the automatic control mode to enhance the adaptability of the control system.

Source: Reinforcement Learning: Theory and Applications, Book edited by Cornelius Weber, Mark Elshaw and Norbert Michael Mayer
ISBN 978-3-902613-14-1, pp.424, January 2008, I-Tech Education and Publishing, Vienna, Austria

This chapter develops a supervisory control approach for burning zone temperature based on Q-learning, in which the signals of human intervention are viewed as the reinforcement learning signals. Section 2 makes brief descriptions of process and supervisory control system architecture. Section 3 discusses the detailed methodology of Q-learning-based supervisory control approach. The implementation and industrial applications are shown in Section 4. Finally, Section 5 draws the conclusion.

2. Process description and supervisory control system architecture

The alumina rotary kiln process is described as follows. Raw material slurry is sprayed into the rotary kiln from upper end (the kiln tail). At the lower end (the kiln head), the coal powders from the coal injector and the primary air from the air blower are mixed into bi-phase fuel flow, which is sprayed into the kiln head hood and combusts with the secondary air, which comes from the cooler. The heated gas was brought to the kiln tail by the induced draft fan, while the material moves to the kiln head via the rotation of the kiln and its self weight, in counter direction with the gas. After the material passes through the drying zone, pre-heating zone, decomposing zone, burning zone and cooling zone in sequence, soluble sodium aluminate is generated in the clinker, which is the product of the kiln process. This process aims to reach high digesting rate of alumina in the following digestion procedure.

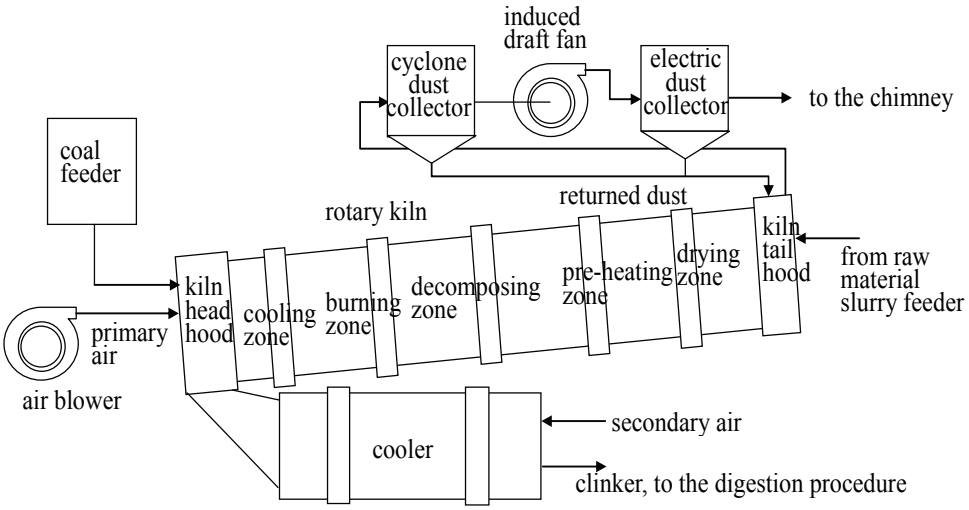

Fig. 1. Schematic diagram of the alumina rotary kiln

The control problem of quality index of kiln production is how to keep the liter weight of clinker being qualified under fluctuated boundary conditions and operating conditions. The liter weight of clinker is hard to measure online and cannot be controlled directly. This paper employs the following strategy to deal with this problem. Some online measurable technologic parameters with closed relations to the final quality index are chosen and

controlled into certain ranges governed by technical requirement so that the quality index control is realized indirectly.

In the sintering process, the normal range of sintering temperature T_{sinter} of raw material depends upon components of raw material slurry. Variations of components of raw material slurry require corresponding variations of sintering temperature. Inconsistency of real sintering temperature range with requirement of raw material will results in over burning or under burning, and clinker quality is not satisfactory. Thus we conclude that components of raw material slurry and sintering temperature are the main aspects influencing clinker quality. Besides, other factors include particle size of raw material and residing time under T_{sinter}. The relationship between desired T_{sinter} and components of raw material slurry can be viewed as a unknown nonlinear function

$$T_{\sin ter} = f([A/S],[N/R],[C/S],[F/A]) \tag{1}$$

where [*A/S*] is the alumina silica ratio of raw material slurry, [*N/R*] is the alkali ratio, [*C/S*] is the calcium silica ratio, [*F/A*] is the iron alumina ratio. Among them, the alumina silica ratio of raw material slurry has the strongest influence on T_{sinter}, the latter must be enhanced along with the enhancement of the former.

From above analysis, one may conclude that there are two key issues about the control problem of quality index of kiln production. One is how to keep the kiln temperature distribution satisfing technical requirement under fluctuated boundary conditions and operating conditions, i.e. how to keep burning zone temperature, kiln tail temperature and residual oxygen content in combustion gas in their technical required ranges. The other is how to adjust the setpoint range of burning zone temperature so that the liter weight of clinker may be kept qualified under fluctuated boundary conditions and operating conditions.

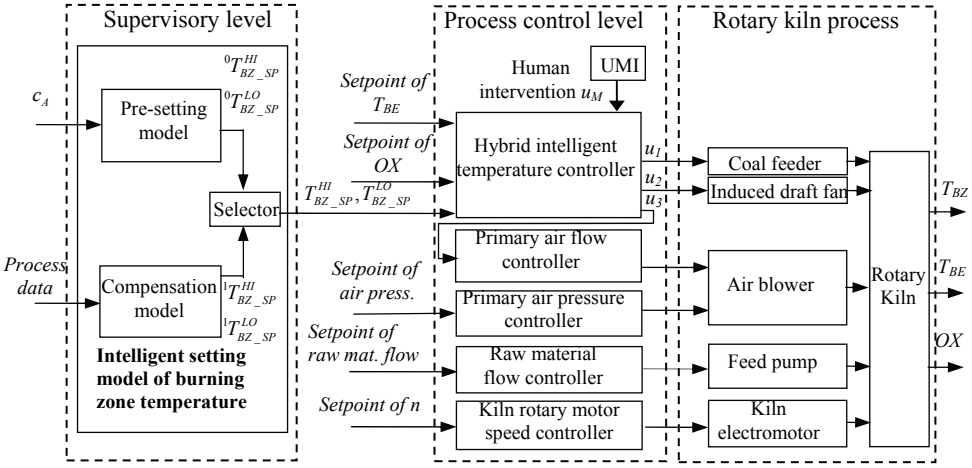

Fig. 2. General structure of the supervisory control system for rotary kiln process

This paper has constructed a supervisory control system consisting of a supervisory level and a process control level, whose general structure is shown in Fig. 2. The final target of this supervisory control system is to keep the production quality index, i.e. the clinker unit weight, being acceptable even if the boundary conditions changed. The related process control strategies in process control level include, 1) a hybrid intelligent temperature controller was designed, which coordinated the coal feeding u_1, damper position of the induced draft fan u_2, and primary air flow u_3 to make the burning zone temperature T_{BZ}, the kiln tail temperature T_{BE}, and the residual oxygen content in combustion gas OX satisfy technical requirements; T_{BZ} is indirectly measured by an infrared pyrometer located at kiln head hood, and T_{BE} is obtained through a thermocouple; 2) individual PI controllers were assigned to basic loops of primary air flow, primary air pressure and flow rate of raw material slurry; and 3) a human-machine interaction(HMI) mechanism was designed so that certain human interventions to coal feeding control from experienced operator can be introduced in the mode of automatic control when the operating conditions changed significantly. The aforementioned process control strategies were depicted in our previous study (Zhou et al., 2004).

The main part of the supervisory level is an intelligent setting model of T_{BZ}, which adjusts the setpoint range of T_{BZ} according to the variations of components of raw material slurry. The setpoints of T_{BE}, OX, primary air pressure, flow rate of raw material slurry and the kiln rotary speed n are given by the operators according to production scheduling and production experiences.

The intelligent setting model of burning zone temperature consists of a pre-setting model of burning zone temperature, a compensation model and a setting selector mechanism. The pre-setting model is to give the upper and lower limits of setpoint range of burning zone temperature, denoted by ${}^0T_{BZ\_SP}^{HI}$ and ${}^0T_{BZ\_SP}^{LO}$, calculating from the offline analysis data of components of raw material slurry. The fuzzy clustering analysis combined with case-based inference learning is employed to build up the pre-setting model of burning zone temperature. The core of the pre-setting model is a case base containing different upper and lower limits of setpoint range of burning zone temperature corresponding to different components of raw material slurry. Such case base is established through fuzzy clustering based data mining from vast process data samples under various components of raw material slurry. Details are not described in this paper.

As a matter of fact, the main problem we are facing is that the components of raw material slurry often change due to unstable raw material mixing process and the offline analysis data reach to the operator with large time delay so that the operator or the pre-setting model cannot directly adjust the setpoint of T_{BZ} duly. As a result, a single intelligent temperature controller and a single pre-setting model of T_{BZ} cannot maintain satisfactory performance. In such a case, a human operator usually rectifies the output of the temperature controller, i.e. the coal feeding, based on the experience of observing burning status through the HMI embedded in the control system. Such interventions can adapt the variation of operating conditions to a certain degree to sustain the quality of the product.

To deal with such a problem, a compensation model and a setting selector are appended. When the offline analysis data of components of raw material slurry are known and input into the system, i.e. the lth sampled time, the setting selector mechanism triggers the pre-setting model to calculated the proper setpoint range of T_{BZ}. When the components of raw material slurry are unknown, the compensation model is triggered to calculated the proper

upper and lower limits of setpoint range of the burning zone temperature, denoted by ${}^1T_{BZ\_SP}^{HI}$ and ${}^1T_{BZ\_SP}^{LO}$ respectively. In the following section, a Q-learning strategy is employed to construct compensation model to learn the self-adjusting knowledge about the setpoint of T_{BZ} through online self-learning from the human intervention signals.

3. Setpoint adjustment approach based on Q-learning

3.1 Bases of Q-learning

Reinforcement learning is learning with a critic instead of a teacher. The only feedback provided by the critic is a scalar signal *r* called reinforcement, which can be regarded as a reward or a punishment. Reinforcement learning performs an online search to find an optimal decision policy in multi-stage decision problems.

Q-learning (Watkins & Dayan, 1992) is a reinforcement learning method where the learner builds incrementally a Q-function which attempts to estimate the discounted future rewards for taking actions from given states. The output of the Q-function for state *x* and action a is denoted by $Q(x,a)$ .When action a has been chosen and applied, the environment moves to a new state, x', and a reinforcement signal, *r*, is received. $Q(x,a)$ is updated by

$$Q_k(x,a) \leftarrow Q_{k-1}(x,a) + \alpha_k \{r + \gamma \max_{a' \in A(x')} Q_{k-1}(x',a') - Q_{k-1}(x,a)\} \tag{2}$$

where

$$\alpha_k = \frac{1}{1 + visits_k(x,a)} \tag{3}$$

where $A(x')$ is the set of possible actions in state x', γ is discount factor , α_k is the learning rate, and $visits_k(x,a)$ is the total number of times this state-action pair (x,a) has been visited up to and including the *k*th iteration.

3.2 Principle of setpoint adjustment approach based on Q-learning

In this section, we may design an online self-learning system based on reinforcement learning to gradually establish the optimal policy of setpoint adjustment of T_{BZ}. Although it cannot reach to the operator in time, the changes of components of raw material slurry may be indirectly reflected through certain measurements of the rotary kiln process. The measurements can be used to construct the environment state set of the learning system. Moreover, information of human interventions can be regarded as evaluations about whether the setpoint of T_{BZ} is proper or not, for human interventions often occur when the performance is unsatisfactory. Thus this kind of information can be defined as reward signal from environment.

For the learning system, the environment includes the rotary kiln process, the temperature controller and the operator. The environment provides current states and reinforcement payoffs to the learning system. The learning system produces the compensated upper and lower limits of setpoint range of T_{BZ} to temperature controller in the environment. The learning system consists of a state perceptron, a critic, a learner and an action selector, as shown in Fig. 3. The state perceptron firstly samples and processes selected measurements

to construct the original state vector, and then converts the original continuous state vector into a discrete feature vector **x** based on a defined feature extraction function. The action selector employs a ε-greedy action selection strategy to produce an amendment of setpoint of T_{BZ} , i.e. ΔT_{BZ_SP} and the critic serves to calculate an internal practicable reward *r* relying on some heuristic rules. The learner updates value function of the state-action pair based on tabular Q-learning. The final outputs of the learning system are the compensated upper and lower limits of setpoint range of T_{BZ}, which are calculated respectively by

$$ {}^{1}T^{HI}_{BZ\_SP}(k) = \Delta T_{BZ\_SP}(k) + {}^{1}T^{HI}_{BZ\_SP}(k-1) \tag{4} $$

$$ {}^{1}T^{LO}_{BZ\_SP}(k) = \Delta T_{BZ\_SP}(k) + {}^{1}T^{LO}_{BZ\_SP}(k-1) \tag{5} $$

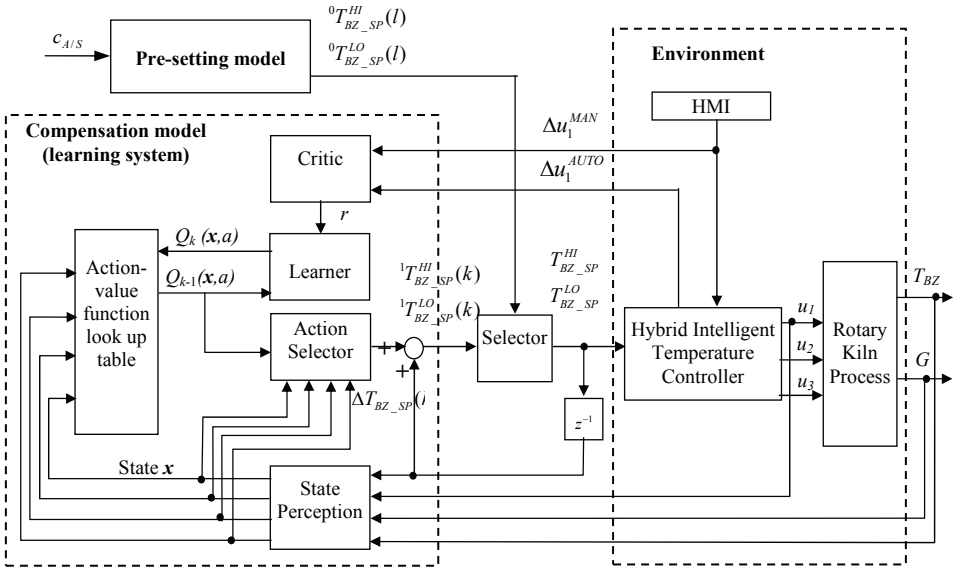

Fig. 3. Schematic diagram of setpoint adjustment approach for T_{BZ} based on Q-learning

In a Markov decision process (MDP), only the sequential nature of the decision process is relevant, not the amount of time that passes between decision stages. A generalization of this is the semi-Markov decision process (SMDP) in which the amount of time between one decision and the next is a random variable. For the learning process, we define τ_s as state perception time span for the perceptron to get the state of the environment and τ_r as reward calculation time span, also named as action execution time span, for the critic to calculate internal reward. The shortest time span from one decision to the next is $\tau = \tau_s + \tau_r$.

The design of the learning system concerns the following key issues:

1. Construction of the environment perception state set;
2. Determination of the action set;
3. Determination of the immediate reward function;
4. Determination of the learning algorithm.

3.3 Construction of the state set

When components of raw material slurry fluctuate and related offline analysis data are unavailable, we hope that the learning system can estimate the changes of the components of raw material slurry through the percepted information about the environment state. From this idea, some related variables are selected from online measurable variables of the kiln process based on human experience, with which the state vector $\mathbf{s}$ is defined to buildup the original state space $\mathbf{S}$ of the learning system, where $\mathbf{s}=[s_1,s_2,s_3,s_4,s_5]$, $\mathbf{s}\in\mathbf{S}$. s_1 is defined as the averaged burning zone temperature $\overline{T}_{BZ}$, s_2 is the averaged flow rate of raw material slurry $\overline{G}$, s_3 is the averaged coal feeding $\overline{u}_1$, s_4 and s_5 are the averaged upper and lower limit of the setpoint range of T_{BZ}, named as $\overline{T}^{HI}_{BZ\_SP}$ and $\overline{T}^{LO}_{BZ\_SP}$ respectively, all during τ_s. They are calculated from

$$\overline{T}_{BZ}=\sum_{j=1}^{J}T_{BZ}(j)/J \tag{6}$$

$$\overline{G}=\sum_{j=1}^{J}G(j)/J \tag{7}$$

$$\overline{u}_1=\sum_{j=1}^{J}u_1(j)/J \tag{8}$$

$$\overline{T}^{HI}_{BZ\_SP}=\sum_{j=1}^{J}T^{HI}_{BZ\_SP}(j)/J \tag{9}$$

$$\overline{T}^{LO}_{BZ\_SP}=\sum_{j=1}^{J}T^{LO}_{BZ\_SP}(j)/J \tag{10}$$

where $T_{BZ}(j)$, $G(j)$, $u_1(j)$, $T^{HI}_{BZ\_SP}(j)$, $T^{LO}_{BZ\_SP}(j)$ denote the jth sampling values of T_{BZ}, flow rate of raw material slurry, coal feeding, upper and lower limits of the setpoint range of T_{BZ} during τ_s respectively. J is the total number of sampling values during τ_s.

Since the state space $\mathbf{S}$ defined above is continuous, it is impossible to compute and store value functions for every possible state or state-action pair due to the curse of dimensionality. The issue is often addressed by generating a compact parametric representation, such as an artificial neural network, that approximates the value function and can guide future actions. we practically choose to use a feature extraction method (Tsitsiklis & Van Roy, 1996) to map the original continuous state space into a finite feature space, then we can employ tabular Q-learning to solve the problem.

By identifying one partition per possible feature vector, the feature extraction mapping $F(\mathbf{s}) = [f_1(s_1, s_4, s_5), f_2(s_1), f_3(s_2), f_4(s_3)]$ defines a partitioning of the original state space. The burning zone temperature biasing (from the setpoint range) level feature f_1, the temperature level feature f_2, flow rate of raw material slurry level feature f_3, the coal feeding level feature f_4 are defined respectively by

$$f_1(s_1, s_4, s_5) = \begin{cases} -2, & (\bar{T}_{BZ} - \bar{T}^{LO}_{BZ\_SP}) < -L2 \\ -1, & -L2 \le (\bar{T}_{BZ} - \bar{T}^{LO}_{BZ\_SP}) < -L1 \\ 0, & (\bar{T}_{BZ} - \bar{T}^{HI}_{BZ\_SP}) \le L1 \text{ and } (\bar{T}_{BZ} - \bar{T}^{LO}_{BZ\_SP}) \ge -L1 \\ 1, & L1 < (\bar{T}_{BZ} - \bar{T}^{HI}_{BZ\_SP}) \le L2 \\ 2, & (\bar{T}_{BZ} - \bar{T}^{HI}_{BZ\_SP}) > L2 \end{cases} \tag{11}$$

$$f_2(s_1) = \begin{cases} 0, & \bar{T}_{BZ} < 1250 \\ 1, & 1250 \le \bar{T}_{BZ} < 1280 \\ 2, & \bar{T}_{BZ} \ge 1280 \end{cases} \tag{12}$$

$$f_3(s_2) = \begin{cases} 0, & 70 \le \bar{G} < 75 \\ 1, & 75 \le \bar{G} < 80 \\ 2, & \bar{G} \ge 80 \\ 3, & else \end{cases} \tag{13}$$

$$f_4(s_3) = \begin{cases} 0, & 800 \le \bar{u}_1 < 1000 \\ 1, & 1000 \le \bar{u}_1 < 1200 \\ 2, & 1200 \le \bar{u}_1 < 1400 \\ 3, & else \end{cases} \tag{14}$$

where $L1$ and $L2$ are the thresholds scaling the burning zone temperature bias from setpoint range level.

Each feature function maps the state space $\mathbf{S}$ to a finite set $P_m, m = 1,2,3,4$. Then we associate the feature vector $\mathbf{x} = [x_1, x_2, x_3, x_4] = F(\mathbf{s})$ to each state $\mathbf{s} \in \mathbf{S}$. The resulting set of all possible feature vectors, also defined as feature space $\mathbf{X}$, is the Cartesian product of the sets P_m.

Because the compensation model for the setpoint of burning zone temperature needs only to be applicable for the normal kiln operating conditions, the design of state set needs certain filtration in the feature space $\mathbf{X}$. The appearence of $x_3 = 3$ or $x_4 = 3$ might means the abnormal operating conditions such as low load of flow rate of raw material slurry during kiln starting phase or abnormal coal components. The state set excludes such valued feature vectors.

3.4 Action set

The learning system aims to deduce the proper or best actions of setpoint adjustment of T_{BZ} from specified environment state. The problem to be handled is how to choose ΔT_{BZ_SP} according to the changes of environment state. Thus the action set can be defined as $A = \{a^1, a^2, a^3, a^4, a^5\} = \{-30, -15, 0, 15, 30\}$.

3.5 Immediate reward signal

During τ_r after the action selection based on current state judgment, the learning system determines the immediate reward signal $r = R(\Delta u_1^{MAN}, \Delta u_1^{AUTO})$, which represents the satisfactory degree of the environment about the action execution under current state, using the human intervention regulation of coal feeding Δu_1^{MAN} and temperature controller regulation Δu_1^{AUTO}.The reward signal *r* is determined in table 1.

r	$\lvert\Delta Coal_{AUTO}\rvert \le L3$	$\Delta Coal_{AUTO} > L3$	$\Delta Coal_{AUTO} < -L3$
$\lvert\Delta Coal_{MAN}\rvert \le L3$	0.4	0.4	0.4
$\Delta Coal_{MAN} > L3$	-0.2	0.2	-0.4
$\Delta Coal_{MAN} < L3$	-0.2	-0.4	0.2

Table 1. Definition of immediate reward function *R*

where *L*3 is the threshold constant, $\Delta Coal_{MAN}$ denotes the total regulation of coal feeding from human intervention during τ_r, which is calculated by

$$\Delta Coal_{MAN} = \sum_{\tau_r} \Delta u_1^{MAN} \tag{15}$$

and $\Delta Coal_{AUTO}$ denotes the total regulation from temperature controller during τ_r, which is calculated by

$$\Delta Coal_{AUTO} = \sum_{\tau_r} \Delta u_1^{AUTO} \tag{16}$$

The immediate reward function R in Table 1 is from the following heuristic rules:
During τ_r, if $|\Delta Coal_{MAN}| \le L3$, which means the operator is satisfied with the regulation action of the control system and little human intervention occurs, then a positive reward r=0.4 is returned. If $\Delta Coal_{MAN}$ and $\Delta Coal_{AUTO}$ have same regulation directions , which means the direction of regulation action of the control system fits with the operator expectation with short amplitude, then a positive reward r=0.2 is returned. If $\Delta Coal_{MAN} > L3$ or $\Delta Coal_{MAN} < -L3$, and $|\Delta Coal_{AUTO}| \le L3$, which means little regulation action of the control system occurs while large human intervention occurs, then r=-0.2. If $\Delta Coal_{MAN}$ and $\Delta Coal_{AUTO}$ have contrary regulation directions, which means the operator is not satisfied with the regulation action of the control system, then a negative reward r=-0.4 is returned.

3.6 Algorithm summary

The whole learning algorithm of the learning system under learning mode is summarized as follows:

Step 1: If it is in initialization, then the Q value table of state-action pairs is initialized according to expert experience, otherwise goto step 2 directly;
Step 2: During τ_s, the state perceptron obtains and saves measured burning zone temperature, flow rate of raw material slurry, coal feeding, upper and lower limits of the setpoint range of the burning zone temperature, and calculates related averaged values by using (6)-(10), then transfer them into related level features to construct feature vector $\mathbf{x}$ by using (11)-(14) .
Step 3: Search in the Q table to make state matching, if unsuccessful then goto step 2 to make state judgement again, if successful then go ahead;
Step 4: The action selector chooses an amendment of setpoint of T_{BZ} as its output according to ε-greedy action selection strategy (Sutton & Barto, 1998), where ε=0.1;
Step 5: During τ_r, the critic determines the reward signal r of this state-action pair according to Table 1.
Step 6: When the current τ_r finishes, entering the next τ_s, the state perceptron judges the next state $\mathbf{x}'$, state matching is made in the Q table, if unsuccessful then goto step 2 to start the next learning round, if successful then using the reward signal r, the learner calculates and updates the Q value of the last state-action pair by using (2)-(3), where $\gamma = 0.9$.
Step 7: Judge if the learning should be finished. When all evaluation values of state-action pairs in the Q table do not change obviously, it means that the Q-function have converged, and the compensation model is well trained.

The problem of Q table initialization: there is no explicit tutor signal in reinforcement learning, the learning procedure is carried out through constant interaction with

environment to get the reward signals. Usually, less information from environment will results low learning efficiency of reinforcement learning. In this paper different initial evaluation values are given for different actions under same state based on expert experience so that the convergence of the algorithm has been speedup, and online learning efficiency has been enhanced.

3.7 Technical issues

The main task of the learning system is to estimate the variations of the kiln operating conditions continuously, and to adjust the setpoint range of burning zone temperature accordingly. Such adjustments should be made when the burning zone temperature is fairly controlled smooth by the temperature controller. Such a judgment signal is given out from the hybrid intelligent temperature controller. If the temperature control is in the abnormal conditions, the learning procedure must be postponed. In this case the setpoint range of the burning zone temperature is kept constant.

Moreover, setpoint adjustments should be made when the learning system make accurate judgment about the kiln operating conditions. Because of complexity and fluctuation of kiln operating conditions, accurate judgment for current state usually needs long time, and the time span between two setpoint adjustments cannot be too short, otherwise the calculated immediate reward cannot reflect the real influence of the above adjustment upon the behaviour and performance of the control system. Thus special attention should be paid to selection of τ_s and τ_r. This makes solid foundation, on which obtained environmental states and reinforcement payoffs are effective.

After long term running, large characteristic changes of components of raw material slurry, coal and kiln device may appear. The previous optimal designed compensation model for the setpoint of burning zone temperature might become invalid under new operating conditions. This needs new optimal design to keep good performance of control system for long term. In this case, the reinforcement learning system should be switched into the learning mode, and above models can be established through new learning to improve the performance, so that the control system has strong adaptability for long term running. This is an important issue drawing the attentions of the enterprise.

4. Industrial application

Shanxi Alumina Plant is the largest alumina plant in Asia with megaton production capacity. It has 6 oversize rotary kilns of φ4.5×110m. Its production employs the series parallel technology of Bayer and Sintering Processes. Such a production technology makes components of the raw material of rotary kilns often vary in large range. It is more difficult to keep a stable kiln operation than ordinary rotary kiln.

A supervisory control system has been developed in the #4 rotary kiln of Shanxi Alumina Plant based on the proposed structure and the setpoint adjustment approach of burning zone temperature. It is implemented in the I/A Series 51 DCS of Foxboro. The Q-learning-based strategy has been realized in the configuration environment of Fox Draw and ICC of I/A Series 51 DCS. Related parameters are chosen as τ_s = 30min, τ_r =120min.

Fig. 4. The setpoint of burning zone temperature is properly adjusted after learning

Fig. 4 shows the condition that, after a period of learning, a set of relatively stable strategies of setpoint adjustment has been established so that the setpoint range of T_{BZ} can be automatically adjusted to satisfy the requirement of sintering temperature, according to the level of raw material slurry flow, the level of coal feeding, the level of T_{BZ} and the level of temperature biasing. It can be seen that the setpoint adjustment happened only when T_{BZ} is controlled smoothly. The judgment signal, denoted by "control parameter" in Fig. 4, takes value of 0 when the burning zone temperature is fairly controlled smooth, and vice versa.

The adjustment actions of the above reinforcement learning system result in satisfactory performance of the kiln temperature controller, with reasonable and acceptable regulation amplitude of coal feeding and regulation rhythm, so that the adaptability for variations of operating conditions has been significantly enhanced and the production quality index, liter weight of clinker, can be kept to reach the technical requirement even if the boundary conditions and operation conditions change. Meanwhile, human interventions become weaker and weaker since the model application has improved the system performance.

In the period of test run, the running rate of supervisory control system has been up to 90%. Negative influences on the heating and operating conditions from human factors have been avoided, rationalization and stability of clinker production has been kept, and operational life span of kiln liner has been prolonged remarkably. The qualification rate of clinker unit weight has been enhanced from 78.67% to 84.77%; production capacity in unit time per kiln has been increased from 52.95t/h to 55t/h with 3.9% increment. The kiln running rate has been elevated up to 1.5%. Through the calculation based on average 10℃ reduction of kiln tail temperature and average 2% decrease of the residual oxygen content in combustion gas, it can be concluded that 1.5% energy consumption has been saved.

5. Conclusion

In this chapter, we focus on the discussion about an implementation strategy of how to employ reinforcement learning in control of a typical complex industrial process to enhance control performance and adaptability for the variations of operating conditions of the automatic control system.

Operation of large rotary kilns is difficult and relies on experienced human operators observing the burning status, because of their inherent complexities. Thus the problem of human-machine coordination is addressed when we design the rotary kiln control system, and the human intervention and adjustment can be introduced. Except for emergent operation conditions that need urgent human operation for system safety, the fact is observed that human interventions to the automatic control system usually imply human's discontent to the performance of the control system when the variation of boundary conditions occurs. From this idea, an online reinforcement learning-based supervisory control system is designed, in which the human interventions might be defined as the environmental reward signals. The optimal mapping between rotary kiln operating conditions and adjustment of important controller setpoint parameters can be established gradually. Successful application of this strategy in an alumina rotary kiln has shown that the adaptability and performance of the control system have been improved effectively.

Further research will focus on trying to improve the setting model of the burning zone temperature by introducing the offline analysis data of clinker liter weight to reject the other uncertain disturbances in the quality control of kiln production.

6. References

Holmblad, L. & Østergaard, J. (1995). The FLS application of Fuzzy logic, *Fuzzy Sets and Systems*, Vol. 70, No. 2-3, (March 1995) 135-146, ISSN: 0165-0114

Jarvensivu, M.; Saari, K. & Jamsa-Jounela, S. (2001a). Intelligent control system of an industrial lime kiln process, *Control Engineering Practice*, Vol. 9, No. 6, (June 2001) 589-606, ISSN: 0967-0661

Jarvensivu, M.; Juuso, E. & Ahava, O. (2001b). Intelligent control of a rotary kiln fired with producer gas generated from biomass, *Engineering Applications of Artificial Intelligence*, Vol. 14, No. 5, (October 2001) 629-653, ISSN: 0952-1976

Sutton, R. & Barto, A. (1998). *Reinforcement Learning: An Introduction*, MIT Press, ISBN: 0262193981, Cambridge, MA

Tsitsiklis, J. & Van Roy, B. (1996). Feature-based methods for large scale dynamic programming, *Machine Learning*, Vol. 22, No. 1-3, (Jan./Feb./March 1996) 59-94, ISSN:0885-6125

Watkins, J. & Dayan, P. (1992). Q-Learning, *Machine Learning*, Vol. 8, No. 3-4, (May 1992) 279-292, ISSN:0885-6125

Zanovello, R. & Budman, H. (1999). Model predictive control with soft constraints with application to lime kiln control, *Computers and Chemical Engineering*, Vol. 23, No. 6, (June 1999) 791-806, ISSN: 0098-1354

Zhou, X.; Xu, D.; Zhang, L. & Chai, T. (2004). Integrated automation system of a rotary kiln process for Alumina production, *Journal of Jilin University* (*Engineering and*

Technology Edition), Vol. 34, No. sup, (August 2004)350-353. ISSN:1671-5497(in Chinese)

Zhou, X.; Yue, H.; Chai, T. & Fang, B. (2006). Supervisory control for rotary kiln temperature based on reinforcement learning , *Proceedings of* 2006 *International Conference on Intelligent Computing*, pp. 428-437, ISBN: 3 540 37255 5, Kunming, China, August, 2006, Springer-Verlag, Berlin, Germany

Reinforcement Learning

Edited by Cornelius Weber, Mark Elshaw and Norbert Michael Mayer

ISBN 978-3-902613-14-1
Hard cover, 424 pages
Publisher I-Tech Education and Publishing
Published online 01, January, 2008
Published in print edition January, 2008

Brains rule the world, and brain-like computation is increasingly used in computers and electronic devices. Brain-like computation is about processing and interpreting data or directly putting forward and performing actions. Learning is a very important aspect. This book is on reinforcement learning which involves performing actions to achieve a goal. The first 11 chapters of this book describe and extend the scope of reinforcement learning. The remaining 11 chapters show that there is already wide usage in numerous fields. Reinforcement learning can tackle control tasks that are too complex for traditional, hand-designed, non-learning controllers. As learning computers can deal with technical complexities, the tasks of human operators remain to specify goals on increasingly higher levels. This book shows that reinforcement learning is a very dynamic area in terms of theory and applications and it shall stimulate and encourage new research in this field.

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821