Q-Learning-Based Controller for Fed-Batch Yeast Fermentation

H. S. E. Chuo, M. K. Tan, H. J. Tham and K. T. K. Teo

Abstract Industrial fed-batch yeast fermentation process is a typical nonlinear dynamic process that requires good control technique and monitoring to optimize the yeast production. This chapter explores the applicability of Q-learning in determining the feed flow rate in a fed-batch yeast fermentation process to achieve multiobjectives optimization. However, to develop such control system, the complex nature of the yeast metabolism that will affect the system stability has to be considered. Q-learning is well known for its interactive properties with the process environment and is suitable for the learning of system dynamic. Therefore, the utilization and performance of Q-learning to seek for the optimal gain for the controller is studied in this chapter. Meanwhile, the performance of Q-learning under the process disturbance is also tested.

Introduction

The fed-batch application in industrial bioprocess is mainly for the substrate feeding control. A good control on substrate inflow to the bioreactor system is able to avoid serious overflow metabolism and increase the cell productivity

H. S. E. Chuo (✉) · M. K. Tan · H. J. Tham · K. T. K. Teo
School of Engineering and Information Technology, University Malaysia Sabah, Jalan UMS, 88400 Kota Kinabalu, Sabah, Malaysia
e-mail: hsechuo@gmail.com

M. K. Tan
e-mail: mktan@ieee.org

H. J. Tham
e-mail: hjtham@ums.edu.my

K. T. K. Teo
e-mail: kenteo@ums.edu.my

R. Pogaku et al. (eds.), *Developments in Sustainable Chemical and Bioprocess Technology*, DOI: 10.1007/978-1-4614-6208-8_28,

(Sonnleitner and Henes 2007). The overflow metabolite, i.e., alcohol in yeast fermentation, if highly concentrated, would be toxic and therefore suppress the cell growth. For a fed-batch fermentation process in which product concentration precision is highly required, a normal three-mode controller is bound by the sensors problem to maintain the control action (Smets et al. 2002).

In recent years, the research focus on artificial intelligence in fermentation optimization has brought about breakthroughs in classic control methods. For example, differential evolution technique (Yüzgeç 2010; Kapadi and Gudi 2004) has been applied for the stochastic search of product optimization. This approach requires the knowledge of effective optimal range in order to minimize the computation time and avoid trap in local optimization. Neural-network-based model training for model predictive control has faced challenges in training and mapping the large number of data until small-error model of satisfactory can be obtained (Ławryńczuk 2011).

The Q-learning algorithm, well known for its explorative and interactive properties with the process environment, is a worth-trying alternative to handle the optimization of the dynamic fermentation. In this chapter, the applicability and the development of Q-learning to adapt the controller gain in a fed-batch yeast fermentation is of the major interest for the multiobjectives optimization. The robustness of Q-learning-based controller under the influence of disturbance will also be discussed in this chapter, in comparison with nominal exponential feeding (Chuo et al. 2011; Teo et al. 2010) and scheduled-gain proportional control under the same process basis.

Dynamic Metabolism of Baker's Yeast Fermentation

The dynamic in the yeast fermentation process is represented by the summarized and re-arranged material balances using the parameters in Yüzgeç (2010):

$$\begin{aligned} \mathrm{d}C_s = {} & \frac{F}{V}(S_0 - C_S) \\ & - C_X\left(1.08547Q_S - 0.08547Q_{S,\mathrm{OX}} + Q_M + 1.228457Q_{E,\mathrm{OX}} + Q_M\right) \end{aligned} \tag{1}$$

$$\mathrm{d}C_X = -\frac{F}{V}C_X + C_X\left(0.05Q_S + 0.535Q_{S,\mathrm{OX}} + 0.7187Q_{E,\mathrm{OX}}\right) \tag{2}$$

$$\mathrm{d}C_E = -\frac{F}{V}C_E + C_X\left[0.4859\left(Q_S - Q_{S,\mathrm{OX}}\right) + Q_{E,\mathrm{OX}}\right] \tag{3}$$

$$\mathrm{d}C_O = -\frac{F}{V}C_O + C_X\left(0.3857Q_{S,\mathrm{OX}} + 0.8896Q_{E,\mathrm{OX}}\right) + k_La\left(C_O^* - C_O\right) \tag{4}$$

$$\mathrm{d}V = F \tag{5}$$

where dC_S, dC_X, dC_E, dC_O, and dV denote the change of concentration of substrate, yeast, ethanol, oxygen, and bioreactor volume, respectively. Assume Monod growth kinetics, the substrate consumption rate, $Q_{S,}$ can go through either oxidative, $Q_{S,\mathrm{OX}}$ or fermentative, $Q_{S,\mathrm{RED}}$ consumption. The rate of oxidation of substrate in the cell, $Q_{S,\mathrm{OX},}$ is determined by the smallest rates at which glucose and oxygen are taken up by the cells. On the other hand, the rate of oxidation of ethanol, $Q_{E,\mathrm{OX},}$ in cell is determined by the smallest of the rates at which ethanol and limited oxygen are taken up by the cells. Specific growth rate, μ, and respiratory quotient, RQ, expressed in terms of Q_S, $Q_{S,\mathrm{OX},}$ and $Q_{E,\mathrm{OX}}$ are shown as below.

$$Q_S = 2.943 \frac{C_S}{0.612 + C_S}\left(1 - e^{-\frac{t}{2}}\right) \tag{6}$$

$$Q_{O,\mathrm{lim}} = 0.255 \frac{C_O}{9.6 \times 10^{-5} + C_O} \frac{3.5}{3.5 + C_e} \tag{7}$$

$$Q_{E,\mathrm{UP}} = 0.238 \frac{C_e}{0.1 + C_e} \frac{3.5}{3.5 + C_S} \tag{8}$$

$$Q_{S,\mathrm{OX}} = \min\begin{pmatrix} C_S \\ 0.359 \\ Q_{O,\mathrm{lim}}/0.3857 \end{pmatrix} \tag{9}$$

$$Q_{E,\mathrm{OX}} = \min\begin{pmatrix} Q_{e,\mathrm{up}} \\ 1.1236 Q_{O,\mathrm{lim}} - 0.4334 Q_{S,\mathrm{OX}} \end{pmatrix} \tag{10}$$

$$\mu = 0.05 Q_S + 0.535 Q_{S,\mathrm{OX}} + 0.7187 Q_{e,\mathrm{OX}} \tag{11}$$

$$RQ = \frac{0.1124 Q_{S,\mathrm{OX}} + 0.462 Q_S + 0.645 Q_{e,\mathrm{OX}}}{0.3857 Q_{S,\mathrm{OX}} + 0.8896 Q_{e,\mathrm{OX}}} \tag{12}$$

Methodology

In a Q-learning algorithm, a learning agent actively interacts with the dynamic environment. The agent decides the best action that causes transition of state to the environment and the latter responds in a new state to the agent. A desirable action is taken based on the reward function which specifies the overall objective of the learning. A reward function assigns rewards or penalties depending on the incorporated preference indices that tell the agent the best way to achieve the goal (Syafiie et al. 2008). The difficulty of designing the reward function by putting it in the shoe of different dynamic situation lies in the determination of these preference indices and the effective range.

Multistep Action Q-Learning

Multistep action (MSA) Q-learning (Schoknecht and Riedmiller 2003) considers the inborn delay nature of the fermentation system to react to the substrate feeding. In MSA, the algorithm will look into the effect (resulting rewards) after executing a sequence of m number of actions.

In the proposed algorithm, constant action is executed throughout m number of steps before the agent decides on the next action for the next round of multistep. The execution of MSA, a^m, causes transitions to the environment, resulting in the state, s^m. The maximum Q-value comes from the execution of the best action, $(a^m)'$, determined by the reward function, r^m, resulting in the maximum state, $(S^m)'$. The Q-value is updated using the Q-learning function:

$$Q_{t+1}(s^m, a^m) \leftarrow (1-\alpha)Q_t(s^m, a^m) + \alpha\left[r^m + \gamma^m \max_{a_n \in A} Q\big((s^m)', (a^m)'\big)\right] \tag{13}$$

where $Q_t\,(s^m,a^m)$ is the Q-value for state-action at time t, α is the learning rate, and γ is the discount factor. The learning rate determines the importance of past experience, and the discount factor weighs the importance of near term rewards (Chuo et al. 2011).

Q-Learning-Based Controller

In this chapter, Q-learning (QL)-based controller is used to react to the gain, K_P, that multiplies the feedback error, E_t, to the substrate feed flow rate, F, as shown in Fig. 1.

The various states at time t and the calculated error, E_t, are related to the process gain, K_P, by the agent using reward function, r^m, and the Q-learning function. The argument maximum Q-value determines the optimized K_P used to tune the magnitude of the error to determine the next feed flow rate, F, for the m steps time.

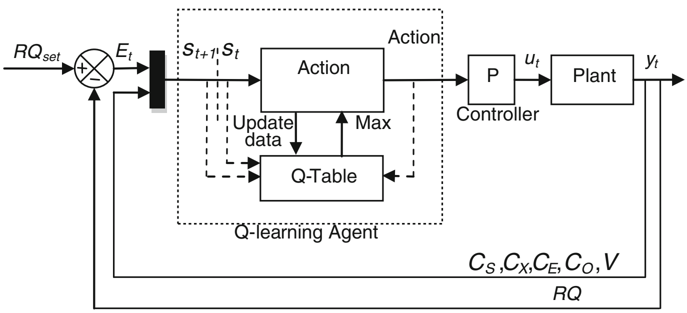

Fig. 1 Q-learning-based control process flow diagram

$$u_t = F_0 K_p E_t \tag{14}$$

$$E_t = RQ_{\text{set}} - RQ \tag{15}$$

The fermentation process was run for 10 h with a total number of 500 actions, each for m steps time. The process step size is 0.001 h and m is 0.02 h or 72 s. Within m steps, the amount of accumulated yeast (q_x) and ethanol (q_e) was stored. The accumulated values instead of the final concentrations at m steps were taken because the dynamicity of the system within a period should not be represented by the final concentration alone. The difference of average specific growth rate at m steps with the critical specific growth rate at 0.21 h^{-1} (q_μ) was also calculated for reward calculation using the reward function:

$$r^m = \beta_1 q_x - \beta_2 q_e - \beta_3 q_\mu \tag{16}$$

where β_1, β_2, and β_3 indicates the weight of the preference index qx, qe, and $q\mu$, respectively. In this case, $\beta_1 = 1$, $\beta_2 = 100$, and $\beta_3 = 10$. More yeast production will result in higher r^m, and more ethanol production will result in lower r^m. Therefore, only actions with higher rewards will be chosen, which has represented the multiobjectives optimization. The K_P with the highest Q-value will be chosen as the gain for the QL-based proportional controller at each m step.

Results and Discussion

For the nominal case, the initial concentration of glucose (C_s), yeast (C_x), ethanol (C_e), and oxygen (C_o) are 3.5, 9.0, 0, and 0.008 g/l, respectively. The substrate feeding stream has a concentration of 325 g/l, initial volume of 50,000 l, and initial feed flow rate 100 l/h. The following results showed the concentration profiles of glucose, yeast (to be maximized), and ethanol (to be minimized) for (1) nominal exponential feeding, $F = 100e^{0.3t}$ (Fig. 2a), (2) scheduled-proportional control, with linear K_P increment of 3 for each m steps, starting at initial $K_P = 0$ (Fig. 2b), and (3) QL-based control (Fig. 2c and d).

Under optimum scheduled-proportional control, the final yeast and ethanol concentrations are 32.91 g/l and negligible, respectively, more robust compare to nominal exponential feeding of 28.21 and 0.0025 g/l, respectively. Under the same initial conditions and using the range of gain, K_P, suggested by trial-and-error tuning scheduled-proportional controller, i.e., $0 \leq K_P \leq 1{,}500$, Q-learning-based controller can adjust and seek for the optimal gain. Maximum final yeast concentration and minimum final ethanol concentration, i.e., 40.42 g/l and negligible, respectively, are obtained for Q-learning-based control. On the other hand, disturbance in substrate feeding concentration can happen in fermentation and overfeeding of substrate tends to trigger the overflow metabolism. In this case, it is introduced as substrate concentration change of 525 g/l at $1 < t \leq 1.5$ h. Scheduled-proportional controller is unable to control the increasing ethanol

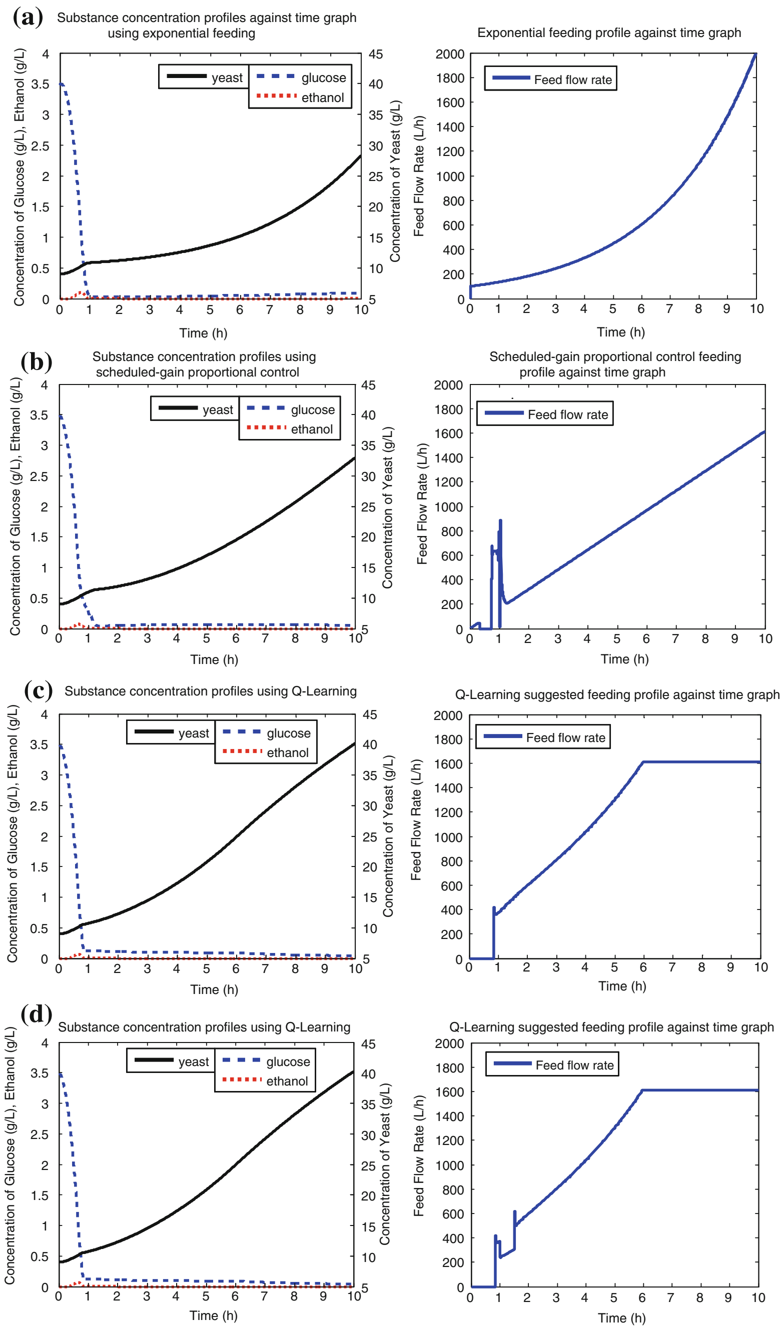

Fig. 2 Concentration profiles of glucose, yeast, and ethanol (*left*) and the substrate feeding profile (*right*) for **a** nominal exponential feeding; **b** scheduled-proportional control; **c** Q-learning-based control and **d** Q-learning-based control under disturbance of $S_0 = 525$ g/l at $1 < t \leq 1.5$

concentration in the case of disturbance. The final concentrations of yeast and ethanol are 33.13 and 19.02 g/l, respectively, for scheduled-proportional controller; 40.49 g/l, negligible, respectively, for Q-learning-based control, as shown in Fig. 2d. Q-learning is able to adapt the process gain when necessary and achieve its goal of maximizing yeast and minimizing ethanol even under the disturbance, as shown in the feeding profile in Fig. 2d.

Conclusions

In this study, Q-learning has shown result of satisfactory in tuning the controller gain to achieve multiobjectives optimization, compare to nominal exponential feeding and scheduled-proportional controller. Meanwhile, Q-learning is able to seek for optimal process gain that can reduce the disturbance effect in the substrate feeding stream rather than increasing-gain scheduled-proportional controller.

Acknowledgments Gratitude is expressed to the Universiti Malaysia Sabah and the Ministry of Higher Education, Malaysia, for their financial support on this research.

References

Chuo, H. S. E., Tan, M. K., Tham, H. J., & Teo, K. T. K. (2011). Optimization of fed-batch baker's yeast fermentation process using learning algorithm. In *Proceedings of 18th Regional Symposium on Chemical Engineering*, 27–28 Oct, Ho Chi Minh City, Vietnam.

Kapadi, M. D., & Gudi, R. D. (2004). Optimal control of fed-batch fermentation involving multiple feeds using differential evolution. *Process Biochemistry, 39*, 1709–1721.

Ławryńczuk, M. (2011). Online set-point optimization cooperating with predictive control of a yeast fermentation process: A neural network approach. *Process Engineering Applications of Artificial Intelligence, 24*, 968–982.

Schoknecht, R., & Riedmiller, M. (2003). Learning to control at multiple time scales. In *Proceedings of ICANN 2003*, 26–29 June, Istanbul, Turkey.

Smets, I. Y., Bastin, G. P., & van Impe, J. F. (2002). Feedback stabilization of fed-batch bioreactors: non-monotonic growth kinetics. *Biotechnology Progress, 18*, 1116–1125.

Sonnleitner, B., & Henes, B. (2007). Controlled fed-batch by tracking the maximal culture capacity. *Journal of Biotechnology, 132*, 118–126.

Syafiie, S., Tadeo, F., & Martinez, E. (2008). Model-free learning control of chemical processes. In C. Weber, M. Elshaw & N. M. Mayer (Eds.), *Reinforcement learning: theory and applications.* Vienna: I-Tech Education and Publishing.

Teo, K. T. K., Tham, H. J., & Tan, M. K. (2010). Optimization of substrate feed flow rate for fed-batch yeast fermentation process. In *Second International Conference on Computational Intelligence, Modelling and Simulation*, 28–30 Sept, Bali, Indonesia.

Yüzgeç, U. (2010). Performance comparison of differential evolution techniques on optimization of feeding profile for an industrial scale baker's yeast fermentation process. *ISA Transactions, 49*, 167–176.