Comparative Study on the Application of Q-learning in Fed-batch Bioproces

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

School of Engineering and Information Technology, Universiti Malaysia Sabah, Jalan UMS, 88400 Kota Kinabalu, Sabah, Malaysia. E-mail: hjtham@ums.edu.my

ABSTRACT: This paper aims to determine the optimal substrate feeding that is able to give high yeast production and low alcoholic content. 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 (QL) is implemented to optimize the yeast production. The main purpose of QL is to reduce the dependency of bioprocess control on human expertise and the amount of data needed for effective control. QL has the learning ability that builds up experience of the process by interacting with the process environment whereby the optimal route seeks through the reward and penalty calculation. In this work, the optimal trajectory suggested by QL was also compared to the exponential feeding rate and their concentration profiles. The performance of the QL controller in rejecting disturbance is also shown.

Key words: fed-batch, Q-learning, baker’s yeast, fermentation

1. INTRODUCTION

Baker’s yeast is among the most studied strain in fermentation process and can be used as an overall benchmark for other strain of yeast fermentation process control [1]. There are few reasons that yeast fermentation is carried out in fed-batch operation: (i) to cope with the metabolic behavior of yeast by intermittent substrate feeding, (ii) to control the quality of the production, and (iii) to maximize the production in the most economic way.

To understand the control issues that were raised throughout the decades in yeast fermentation control, some significant metabolic behavior of yeast will have to be recognized beforehand. In general, yeast relies on how the substrate and the amount of substrate fed into the system to determine which metabolic pathway to be taken [2,3].

With the understanding on the metabolic behavior of yeast, control strategies had been developed to maintain the yeast metabolism at its critical state [2,3,4]. Several ways were proposed to maximize yeast production through a fed-batch fermentation process. Traditional industrial fed-batch production is done through open loop control using predetermined substrate feeding [5]. Often, this strongly relies on the experience of the human operator to deal with the process uncertainties and optimization is seldom achieved [6]. Other control strategies include attempting on maintaining the critical state of yeast metabolism by using critical specific growth rate ($\mu$), critical respiratory quotient (RQ) and ethanol concentration as reference point [2,5]. The RQ and $\mu$ values are generally time-varying [4,5].

To catch up with the time-varying process environment, various intelligent control strategies have also been considered. Knowledge-based fuzzy control has been used to settle the system ambiguity problem [7]. Optimization through this control requires thorough understanding of the process to set the rules and membership functions to obtain desired process output. Alternatively, neural network can work in pair with fuzzy control or genetic algorithm to map the process variables relationship via sufficient data training [8,9,10]. Using the black-box model concept, recent year control schemes have been focused on developing intelligent modeling and control strategies such as dynamic programming, evolutionary algorithm, adaptive control and etc to tackle the fermentation process dynamics. Disturbance rejection methods and error estimation have also been discussed to optimize the process [11,12].

In this paper, Q-learning (QL) is introduced to study the fed-batch yeast fermentation system
dynamics in simulation. This work is interested in reporting the ability of QL in learning its surroundings environment and making decision for the optimization process. The aim of QL development is with least process knowledge, least pre-settings, and can work under control personnel lacking and data insufficiency. This work also compares the performances of the proposed QL and exponential feeding in yeast fermentation process. A QL-optimized predetermined feeding strategy was implemented to test its performance with comparison to exponential feeding strategy.

2. YEAST FERMENTATION PROCESS

In this simulation, the Karakuzu et al. [6] model was used as the plant to represent the yeast fermentation process. This model was chosen because it merges the theoretical background studies related to bottleneck theory [13] and Monod kinetics with real time industrial process dynamics.

The process model consisted of two parts, i.e. the reaction kinetics within yeast cells and the overall system dynamics. The model and parameters can be referred in Appendix A.

The assumptions for this fermentation model are as followed:
- There is sufficient nutrients and dissolved oxygen supply.
- The gaseous phase concentration is negligible.
- The system is in quasi steady state.
- The system is well-mixed.
- The substrate, oxygen and ethanol metabolism follows the Monod kinetics.

The simulation was run for 10 h with sampling time of 0.001 h. In this case, feeding rate ranges from 0 L/h to 4500 L/h. The volume of the reactor was 100 m^3 with the initial broth of 50 m^3. The initial substrate, yeast and oxygen concentration in broth were 50 m^3 of rate range, sampling time of 0.001 h.

3. Q-LEARNING

Q-learning (QL) is a step-to-step exploring and learning algorithm which determines the ultimate goal of the process based on the route that returns the maximum reward, as shown in Eq. (1).

\[ Q(s,a) = (1-\alpha)Q_{t-1}(s,a) + \alpha[R_t + \gamma \max_a Q_{t+1}(s,a)] \] (1)

There are altogether three main components in Q-learning process, which are past experience, current reward and future state, with referring to equation (1). The learning agent needs to accumulate its experience through exploring and interacting with the environment until it finds the best pathway to take to achieve its goal, which is the purpose of the fermentation process. This is indicated by \( \alpha \), the learning rate of the system. The larger the value of \( \alpha \), the learning of the process depends more on the past experiences. While more attempts are made by the learner, its experience will increase with respect to time and the time taken to learn and make decision will be shortened throughout the process. Along the learning process, rewards is accumulated, indicated by the term \( R \) and the final state is determined by the route that gives the most rewards. The future state is represented by the term \( \gamma \), i.e. the discount factor which indicates the importance of process accuracy and where it is heading towards.

In other words, the Q-learning algorithm attempts to learn a state-action pair value \( Q(s,a) \) by starting in state \( s \), taking an action \( a \), and following the optimal policy thereafter using value iteration [14]. The flow of iterations of QL were as followed [15]: (i) Let the current state be \( s \), (ii) Select an action \( a \) to perform, (iii) Let the reward received for performing \( a \) be \( R \), and the resulting state be \( Q(s,a) \), (iv) Update \( Q(s,a) \) to reflect the observation \( < s, a, R, Q(s,a) > \). (v) Repeat from step (i) again.

The role of QL was to study the dynamics of the fermentation plant in section 2 to determine the optimal feeding profile. The input and output into the plant would be sent to Q-table to update the Q-value in each iteration. From there the maximum Q-value that suggested the highest reward was chosen. Q-table consists of several important variables studies, i.e. the feed rate, volume, response of yeast, ethanol glucose and oxygen, specific growth rate (\( \mu \)) and respiratory quotient (RQ). With responding to various types of states, for example, increasing ethanol concentration or fluctuating \( \mu \), the various feed rate trials are the actions taken to test which feed flow rate is giving the best reward. The maximum reward is the action taken that produces the maximum amount of yeast.

Appendix A.
4. RESULTS AND DISCUSSION

In case 1, simulation was run under exponential feeding, $F = 500e^{0.05t}$, which is the nominal feed rate in open loop response, as shown in Fig. 1. 500 L/h was the initial feed rate chosen as the value should be big enough to support the exponential growth. The exponential parameter 0.05 decides the magnitude of exponential curve to a reasonable scale to avoid overfeeding or starvation.

The results are determined. After 1 h, the feeding rate was the initial feed rate decided to update the optimized feeding profile in every 0.5 h. In the initial process where environmental adaptation and low reproduction took place, low substrate feed was determined. After 1 h, the feeding rate increased to produce more yeast.

The response of yeast, ethanol and glucose concentration for case 1 is shown in Fig. 2. The glucose concentration drops after 1 h after being consumed for the production of yeast. The overall ethanol production was little in amount when appropriate initial concentration of yeast and glucose were chosen. The initial glucose to yeast concentration ratio is approximately 2.5 to 3. The final yeast concentration is 27.5 g/L.

Case 2 is the application of QL to determine the best feeding rate profile without disturbance. The results are shown in Fig. 3 and Fig. 4. The final yeast production reaches 46 g/L, which is approximately 67% increment compared to case 1. The QL was determined to update the optimized feeding profile in every 0.5 h. In the initial process where environmental adaptation and low reproduction took place, low substrate feed was determined. After 1 h, the feeding rate increased to produce more yeast.

The suggested feed rate lies in the range of 0 to 2500 L/h. Less ethanol was produced in case 2 compare to case 1. This is because through the learning and iterations updates, QL was able to deal with ethanol concentration rises and minimize the ethanol production. Furthermore, compare to the open loop exponential feeding, the closed loop properties of QL has more corrective actions.

Case 3 utilized the predetermined substrate feed suggested by QL to verify that the feeding strategy was able to perform and obtain the same results as learned by QL in an open loop. It was carried out in open loop feeding to compare with the performance of exponential feeding. The results of the yeast, glucose and ethanol concentration profiles for case 3 were very similar to the profiles optimized using QL as shown in Fig. 4. Therefore, the predetermined substrate feeding was able to perform well in maximizing yeast production and minimizing ethanol concentration in the bioreactor even in an open loop.

Under the influence of glucose disturbance in the substrate feeding stream, the increasing substrate concentration was noticed at 1 h in case 4. There was no change of the feeding rate profile as it is assumed that the glucose disturbance only changed the concentration of substrate in the inlet substrate stream without changing the feeding rate, as shown in Fig. 5. The additional substrate concentration...
exceeding the critical substrate concentration due to the overflow metabolism [2,3] will be converted into ethanol, therefore the concentration of ethanol increment was triggered by the overflow of substrate as shown in Fig. 6. High ethanol concentration will degrade the quality of yeast culture, which is undesirable.

The testing of the performance of predetermined substrate feeding under the influence of disturbance was shown in Fig. 7 and 8. In Fig. 7, the predetermined profile is very similar to that of the QL in case 2. In this case, the initial substrate feeding was 50 L/h compare to nil in case 2 because the optimal substrate feeding profile determination of QL only started at 0.5 h. In the first 0.5 h, there is no substrate feed rate for QL because of the initiation of the learning process.

From Fig. 8, it can be seen that the predetermined substrate feeding can eliminate part of the glucose disturbance and maintain the maximum yeast production, compare to the exponential feeding in Fig. 6. In this paper, the QL-optimized substrate feeding profile for the disturbance case was not used to cope for case 5 due to the variations in the type and magnitude of disturbances in real time process. The predetermined substrate profile is able to cope with normal operations with tolerance to low amount of disturbance.

**5. CONCLUSION**

In this work, QL is implemented to suggest the optimal substrate feeding for fed-batch yeast fermentation process through learning. The optimized predetermined substrate feeding can perform better than nominal exponential feeding under and without the influence of disturbance. In the future work, research will be focusing on optimizing the QL algorithm development in handling with various uncertainties and to improve the robustness of the QL algorithm.

**NOMENCLATURE**

\( a \) : action
\( F \) : feed flow rate \([\text{L/h}]\)
\( Q_i \) : present Q-value
\( Q_{i-1} \) : past Q-value
\( Q_{i+1} \) : future Q-value
\( R_i \) : reward
\( s \) : state
\( \alpha \) : learning rate
\( \gamma \) : discount factor

**APPENDICES**
The Kazakuzu et al. [6] model as the process plant.
\[ Q_s = Q_{s,\text{max}} \frac{c_s}{K_s + c} (1 - e^{-\frac{t}{\tau_s}}) \]

\[ Q_{s,\lim} = \frac{\mu_{\text{up}}}{\tau_{s,\text{up}}} \]

\[ Q_{o,\lim} = Q_{o,\text{max}} \frac{c_o}{K_o + c_o} \]

\[ Q_{s,\text{ox}} = \min \left( \frac{Q_s}{Q_{s,\text{lim}}/Y_{o/s}} \right) \]

\[ Q_{s,\text{red}} = Q_s - Q_{s,\text{ox}} \]

\[ Q_{e,\text{up}} = \frac{c_e}{K_e + c_e} \]

\[ Q_{e,\text{ox}} = \min \left( \frac{Q_{e,\text{up}}}{Q_{e,\text{lim}} - Q_{s,\text{ox}} Y_{o/s}} \right) \]

\[ \mu = Q_{s,\text{ox}} Y_{o/s} + Q_{s,\text{red}} Y_{x/s} + Q_{e,\text{ox}} Y_{x/e} \]

\[ RQ = Q_c/Q_o \]

\[ Q_c = Q_{s,\text{ox}} Y_{c/s} + Q_{s,\text{red}} Y_{c/s} + Q_{e,\text{ox}} Y_{c/e} \]

\[ \frac{dc_s}{dt} = \frac{F}{V} (S_0 - C_s) - \left( \frac{\mu}{\tau_{s,\text{up}}} + \frac{Q_{e,\text{pr}}}{Y_{v/e}} + Q_{\text{m}} \right) C_s \]

\[ \frac{dc_o}{dt} = (Q_{e,\text{pr}} - Q_{e,\text{ox}}) C_x - \frac{F}{V} C_s \]

\[ \frac{dc_x}{dt} = \mu C_x - \frac{F}{V} C_x \]

\[ \frac{dc_o}{dt} = -Q_{r,\text{ox}} C_x + k_{L,A} (C_A - C_o) - \frac{F}{V} C_o \]

REFERENCES


