**Table 21**. Fussy rules [89] for the fuzzy control decoupling in Section 4.1.1

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| E | Ec | | | | | | |
| NB | NM | NS | ZO | PS | PM | PB |
| NB | PB/NB | PB/NB | PM/NM | PM/NM | PS/NS | ZO/ZO | ZO/ZO |
| NM | PB/NB | PB/NB | PM/NM | PS/NS | PS/NS | ZO/ZO | NS/ZO |
| NS | PM/PB | PM/NM | PM/NS | PS/NS | ZO/ZO | NS/PS | NS/PS |
| ZO | PM/NM | PM/NM | PS/NS | ZO/ZO | NS/PS | NM/PM | NM/PM |
| PS | PM/NM | PS/NS | ZO/ZO | NS/PS | NS/PS | NM/PM | NM/PB |
| PM | PS/ZO | ZO/ZO | NS/PS | NM/PS | NM/PM | NM/PB | NB/PB |
| PB | ZO/ZO | ZO/ZO | NM/PS | NM/PM | NM/PM | NB/PB | NB/PB |

# Appendix 2

**Table 22.** Key parameters of the DDPG decoupling model

|  |  |
| --- | --- |
| Relevant parameters in DDPG | Value or range |
| Agents | RL Agent 1(For the first loop)  RL Agent 2 (For the second loop) |
| States of each RL Agent | There are three state variables in each RL Agent:  1) Real output value of each loop,  2) Deviation between real output value and set point value of each loop,  3) Changing rate of deviation of each loop |
| Actors of each RL Agent | There are two actors in each RL Agent:  1) Proportion coefficient of PI controller of each loop  2) Integral coefficient of PI controller of each loop |
| Searching range of actor space | Considering the RL Agent1:  1)  2)  Considering the RL Agent2:  1)  2) |
| Critic network | Each RL Agent in this system has one critic network and one target critic network.  Two networks have the same structure, and the specific structure of each network is as follows:  Step 1: Three states and two actions are designed as the network inputs, and Q values correspond to the network outputs.  Step 2: When the 3 states are input to the network, the output result is used as input to the full connection layer of 50 neurons, and then the output result is input into the full connection layer of 25 neurons, followed by the ReLU activation function.  Step 3: When the 2 actors are input to the network, it passes through the full connection layer with 25 neurons.  After that, the results of step 2 and step 3 are input into the added layer of two neurons together, followed by the ReLU activation function, and finally, the Q values are output through the full connection layer.  Thereafter, we set the network optimizer as Adam optimizer, set the learning rate to 0.001, and set the L2 regularization to 0.0002. |
| Target critic network |
| Actor network | Each RL Agent in this system has one actor network and one target actor network.  Two networks have the same structure, and the specific structure of each network is as follows:  Three states are designed as the network inputs, and two parameters （Kp, Ki）of the PI controller correspond to the network outputs.  The network mainly includes a three-layer network: 3 full connection layers with 50 neurons, 25 neurons, and 2 neurons, respectively. The activation function of each fully connected layer is ReLU, ReLU, and tanh, respectively.  Thereafter, we set Adam optimizer as network optimizer, set the learning rate to 0.0001, and set the L2 regularization to 0.00001. |
| Target actor network |
| Other key parameters during the Reinforcement learning training process | 1) Each agent sampling time is 1 s,  2) The target update smoothing factor is 0.001  3) The discount factor is 1  4) The Mini batch\_size is 64  5) The experience area is 1 × 106,  6) The maximum simulation time in each episode is 200 s  7) The maximum training episode is 80 times  8) The training stops when the rewards of the two agents both reach 1500 |

# Appendix 3

**Table 23.** Parameters of reinforcement learning system (DDPG) of the batch beer fermentation

|  |  |
| --- | --- |
| Relevant parameters in DDPG | Value or range |
| Agents | One RL Agent |
| States of each RL Agent | There are four state variables in the RL Agent: concentrations of etac, diacetyl, sugar, and ethanol |
| Actors of RL Agent | In this experiment, the actor is a vector , which is a temperature value sequence of seven sampling times (0 s, 44s, 90 s, 120 s, 150 s, 180 s, 200 s) in a whole control process. |
| Searching range of actors | For RL Agent, the searching range of each element of ***Temp*** is:  ,,,, ,, |
| Critic network | The RL Agent in this system has one critic network and one target critic network.  Two networks have the same structure, and the specific structure of each network is as follows:  Step 1: Four states and seven elements of the actor vector ***Temp*** are designed as the network inputs, and Q values correspond to the network outputs.  Step 2: When the 3 states are input to the network, the output result is input into the full connection layer of 50 neurons, and then the output result is input into the full connection layer of 25 neurons, followed by the ReLU activation function.  Step 3: When the 7 elements of the actor vector ***Temp*** are input to the network, it passes through the full connection layer with 25 neurons.  Subsequently, the results of step 2 and step 3 are input into the added layer of two neurons together, followed the ReLU activation function, and finally the Q values are output through the full connection layer.  Thereafter, we set network optimizer as Adam optimizer, set the learning rate to 0.001, and set the L2 regularization to 0.0002. |
| Target critic network |
| Actor network | The RL Agent in this system has one actor network and one target actor network.  Two networks have the same structure, and the specific structure of each network is as follows:  Four states are designed as the network inputs, and the seven elements of the actor vector ***Temp*** correspond to the network outputs.  The network mainly includes a three-layer network: 3 full connection layers with 50 neurons, 25 neurons, and 2 neurons, respectively. The activation function of each fully connected layer is ReLU, ReLU, and tanh, respectively.  Thereafter, we set Adam optimizer as network optimizer, set the learning rate to 0.0001, and set the L2 regularization to 0.00001. |
| Target actor network |
| Other key parameters during the reinforcement learning training process | 1) Each agent sampling time is 1 s  2) The target update smoothing factor is 0.001  3) The discount factor is 1  4) The MiniBatchSize is 64  5) The experience area is 1 × 106  6) The maximum simulation time in each episode is 200 s  7) The maximum training episode is 80 times  8) The training stops when the rewards of the two agents both reach 1500. |