

Reinforcement Learning

Parameter	Description	Parameter	Description
i	Type of the Pipe	e_t	Efficiency Percentage
j	Transfer Location	T_i	Standard Production Time
t	Time (day)	$F_{i,t}$	Forecast Order Quantity
$m_{i,t}$	Stock Level	$Q_{i,j,t}$	Total Order Quantity
n	Order Quantity Limit	$a_{i,j,t}$	Transferred Order Quantity
c	Stock Capacity	p_t	Penalty

Table 1: RL Parameters.

State Vector:

$$s_t = (F_{1,t}, \dots, F_{I,t}, Q_{1,1,t}, \dots, Q_{I,J,t}, m_{1,t}, \dots, m_{I,t})$$

Action Vector:

$$a_t = (a_{1,1,t}, \dots, a_{I,J,t})$$

Reward Function:

$$\begin{aligned} p_t - T_i &\leftarrow p_t \quad \text{if } c > m_{i,t} + a_{i,1,t} \\ p_t + T_i &\leftarrow p_t \quad \text{if } Q_{i,t,j} > n \end{aligned}$$

State Transitions:

$$\begin{aligned} m_{i,t+1} &= m_{i,t} + a_{i,1,t} + \min(Q_{i,j,t+1} - a_{i,j,t}, 0) \\ Q_{i,j,t+1} &= Q_{i,j,t} - a_{i,j,t} \end{aligned}$$

Transfer Locations (j):

$$\begin{aligned} \text{Stock} : j &= 1 \\ \text{Production} : j &= 2 \end{aligned}$$