On-Policy Learning for the Swing Process Control of a Cutter Suction Dredger

Changyun Wei, Xiujing Chen and Fusheng Ni
College of Mechanical and Electrical Engineering, Hohai University, China
E-mail: c.wei@hhu.edu.cn

Abstract. It is hard to describe the swing process of a cutter suction dredger with accurate mathematical models because the dynamic characteristics of the swing process are complex, and the relationship between the parameters is not clear. Currently, the swing process control depends entirely on the operators, and it is sometimes difficult to obtain a high production efficiency and construction accuracy. In this paper, an approach that combines SARSA-Lambda with a linear neural network is proposed to solve the intelligent control of the swing process. The dynamic model of the swing process is built using the generalization capability of the linear neural network that solves the output problem of the continuous state space. SARSA-Lambda is used to realize the adaptive control of the swing process by means of learning. The simulation results show that the proposed approach can quickly and effectively learn and achieve goals in uncertain environmental conditions and achieve better control consequents.

1. Introduction
Dredging equipment is widely used in the national economic infrastructure such as the construction and maintenance of port and waterways, flood control and dredging as well as environmental renovation [1-3]. Among them, the cutter suction dredger is a kind of dredging equipment that is widely used in dredging projects (see Figure 1). It mainly cuts the underwater soil by means of a rotary cutter, and then uses a centrifugal pump to carry out long-distance transportation in the slurry mixture. It can be used in river navigation, rivers, lakes and offshore operations. In the dredging operation of cutter suction dredgers, various factors must be considered to adjust the lateral movement speed in real time to ensure safe and efficient operation of dredgers [4].

![Figure 1. A cutter suction dredger and its swing process.](image_url)
At present, the control of the swing speed is still in the manual operation stage. Due to the dredging area environment, in order to make the dredger production relatively stable, operators need to continuously adjust the swing speed by operating the traverse handle according to their long-term dredging experience [5]. However, the dredging operation environment is not always constant. Different environments and soil quality have a great impact on dredging. Therefore, there are certain limitations that the operator uses the previous construction experience to handle the current dredging construction. In addition, the production efficiency and completion of dredging operations are affected by the operator’s physical fatigue and personal work ability [6,7]. In order to reduce the dependence of the dredging construction process on personnel, it is necessary to develop a swing process control system that is highly efficient, accurate, and highly automated.

Reinforcement learning (RL) is an important machine learning method, which was first proposed by Minsky, emphasizing the environment-based action to maximize the expected benefit [8]. In RL algorithms, SARSA-Lambda is an on-policy learning algorithm that can perceive the environment and choose the optimal action through self-learning [9]. The learning agent has strong robustness and adaptability and this is suitable for dealing with multi-dimensional nonlinear systems. Because of the need to connect to the environment in SARSA-Lambda, this paper uses the generalization ability of neural network in the simulation experiment. The linear neural network is applied to SARSA-Lambda and realizes the continuous state space value within a certain range. In this paper, SARSA-Lambda is applied to control the swing system of a cutter suction dredger, which is more efficient and more automatic than conventional control.

2. SARSA-Lambda Learning for Swing Process Control

SARSA-Lambda is a value-based on-policy method for learning the optimal policy [10]. It mainly obtains corresponding rewards in the environment through the action of agent output, and then adjusts the agent behaviour policy according to the reward [11]. SARSA-Lambda establishes the Q table and eligibility traces E table, and selects actions based on the strategy where all Q values corresponds to the current state of the Q table. After that, the Q table will be updated according to the E table and the E table will also be updated at the same time.

2.1. SARSA-Lambda learning model

The learning agent in SARSA-Lambda chooses actions according to Q value in Q table and action space under an unknown environment. The basic process can be described as follows: at the current time t, the agent receives the environmental feedback state \( s_t \), reward \( r_t \) and action space \( \tilde{A} \), and selects the action \( a_t \) in the Q table through the strategy \( \pi \). Afterwards, the action \( a_t \) is feed back to the environment, and then the state is transferred to the next moment, in addition to determining its action space and rewards. Finally, according to the E table, the Q value is updated, and the E table is updated at the same time. Based on the given strategy \( \pi \), \( Q(s,a) \) is a function of the value of the state \( s \) and the action \( a \), then:

\[
Q^\pi(s,a) = E^\pi \left( \sum_{t=0}^{\infty} \gamma^t r_t \mid s_0 = s, a_0 = a \right)
\]  

where \( s_0 \) is initial state, \( a_0 \) is the initial action. According to the Bellman equation [11], \( Q(s,a) \) satisfy:

\[
Q^\pi(s,a) = \sum_{s_{i+1}} \left[ p(s_i, a_i, s_{i+1}) r(s_i, a_i, s_{i+1}) \right] + \gamma \sum_{s_{i+1}} \left[ p(s_i, a_i, s_{i+1}) Q^\pi(s_{i+1}, a_{i+1}) \right]
\]  

where \( p(s_i, a_i, s_{i+1}) \) is the probability of transferring it to the next state \( s_{i+1} \) in the agent’s current state \( s_i \) and action \( a_i \); \( r(s_i, a_i, s_{i+1}) \) is reward which is obtained when transferring it to the next state.

\[
Q^\pi(s,a) = \sum_{s_{i+1}} \left[ p(s_i, a_i, s_{i+1}) r(s_i, a_i, s_{i+1}) \right] + \gamma \sum_{s_{i+1}} \left[ p(s_i, a_i, s_{i+1}) Q^\pi(s_{i+1}, a_{i+1}) \right]
\]
in the agent’s current state $s_t$ and action $a_t$; $\gamma$ is a discount factor which weights immediate rewards relative to future rewards, and its value is between $[0, 1]$.

After each action, the Q value needs to be updated. The Q table is updated according to the following rule:

$$Q_{t+1}(s, a) = Q_t(s, a) + \alpha \delta_t(s, a)$$

(3)

where $\alpha$ is the learning rate. Then the $\delta_t$ and $e_t(s, a)$ can be updated as follows:

$$\delta_t = r_{t+1} + \gamma Q_t(s_{t+1}, a_{t+1}) - Q_t(s_t, a_t)$$

(4)

$$e_t(s, a) = \left\{ \begin{array}{ll} \gamma e_{t-1}(s, a) + 1, & s = s_t, a = a_t \\ \gamma e_{t-1}(s, a), & \text{otherwise} \end{array} \right.$$  

(5)

Where $e_t(s, a)$ is the current eligibility traces, $\lambda$ is the trace decay parameter. Decrease the $E(s, a)$ value with time, the farther away from the reward, the smaller the “indispensability”. The E table is updated for all $s, a$ on the basis of the following rule:

$$E = E \ast \gamma \ast \lambda$$

(6)

Because SARSA-Lambda emphasizes action based on environment, the environment is essential. Thus, the environment setup of the swing process is introduced in the next section.

2.2. Swing Process Analysis

To analyse the swing process of a cutter suction dredger, the state and action parameters need to select so as to build the environment model. In dredging operations, the performance of the dredger is directly reflected on the output to some extent, and also directly determines the project benefits [5-7].

The production of a cutter suction dredger is calculated by:

$$W = QC_t \pi r^2 v t$$

(7)

where $W$ is the production; $Q$ is the flow in the pipe; $C_w$ is the mud concentration; $r$ is the radius of the mud pipe; $v$ is the flow rate in the pipe; $t$ is the time. The flow rate of the pipeline changes little during the dredging process, so the mud concentration directly reflects the production. The formation of mud within the pipeline goes through three main processes:

Cutter cutting soil: The output of the dredger is determined by the quality of the soil that is cut by the cutter. The volume of the soil $V$ is calculated as follows:

$$A_c = kb h_c$$

$$V = A_c V_s$$

(8)

In the above equation, $A_c$ is the cutting cross-sectional area perpendicular to the swing speed; $k$ is the scale factor; $h_c$ is the cutting depth; $b_c$ is the cutting width; $V_s$ is the swing speed. After cutting, small particles of the soil will be sucked up through the suction pipe to form a slurry. The particle size of silt particles is related to the thickness $d_d$:

$$d_d = \frac{60V_s}{Zn}$$

(9)

In the form, $Z$ is the number of the cutter teeth, and the $n$ is the cutter speed. In the pipeline, the percentage of soil volume is $f$, the volume of sediment $V_m$ entering the pipeline per unit time can be:

$$V_m = fV_m$$

(10)

In the formula, $V_m$ is the volume of sediments. However, there are missing losses $\eta$ in this process, which can be characterized by the slippage rate $(0 \leq \eta \leq 1)$. The final dredging output is:
\[ W = (1 - \eta)A \cdot V \]  \hspace{1cm} (11)

Combining Equation (7) and (11), we can get:
\[ C_w = f(\lambda, \eta, A, \nu)V \]  \hspace{1cm} (12)

Where \( \lambda \) is the deposition loss rate during the slurry pipeline transportation. The motor-driven cutter performance directly reflects its cutting ability. By analysing the data collected at the construction site, the cutter motor current is selected. In the dredging operation, the suction vacuum is installed on the upper part of the cutter and is very sensitive to the change of mud concentration in the pipeline. Once the cutter sucks the suction, the suction vacuum will change. Therefore, the suction vacuum number reflects the production rate of the dredger in real time. In actual operation, the operator adjusts the swing speed to ensure that the slurry concentration is stable within a certain range.

In summary, the action in reinforcement learning is the swing speed \( V_s \), and the environmental status parameters include the cutter motor current \( I_d \), suction vacuum \( p_v \), and mud concentration \( C_v \).

2.3. Learning Model for Swing Process Control

The SARSA-Lambda learning model for swing process control is depicted in Figure 2.

![Figure 2. SARSA-Lambda learning model for swing process control.](image)

We use \( S = [I_d, p_v, C_v] \) to define the state space of the swing process, and the reward function \( r \) is expressed by the following formula:
\[
    r = \begin{cases} 
        -10 & I_d \in (0,500) \cup (950,960), b \in (0,45) \\
        C_v & I_d \in (500,950), b \in (0,45) \\
        10 \cdot C_v & I_d \in (500,950), b \in (-20,-70)
    \end{cases}
\]  \hspace{1cm} (13)

Before performing intelligent control, it is necessary to establish and improve the Q and E tables in the agent. In the SARSA(\( \lambda \)) process, the action \( a_s \) is selected according to the state \( s_t \) and the action space \( A_t \) under the strategy \( \pi \). Afterwards, the Q table and E table are updated using the reward value generated by the action \( a_t \).

Due to too many states in the process of swing process, it is impossible to set the Q table and E table directly, so the Q table and the E table need to be explored by agents themselves. The setting of the strategy \( \pi \) directly affects the exploration ability of the agent, and its value is between \([0,1]\), which is expressed as the random selection action in the current state. And the probability of \( 1-\pi \) chooses the behaviour according to the optimal Q value of the Q table. If \( \pi \) is too small, the agent will not be able to fully explore the environment; otherwise, it will cause the agent to be unstable in system
control. Therefore, this paper sets that as the number of explorations continues to increase, strategy \( \pi \) gradually decreases in the Q table exploration process. The strategy \( \pi \) can be calculated as:

\[
\pi = \pi_0 - \frac{e}{\epsilon_{\max}} \pi_0 
\]

In the formula, \( \pi_0 \) is the initial strategy; \( e \) is the current exploration round number; \( \epsilon_{\max} \) is the maximum number of exploration rounds. When the Q table is established and perfected, set the initial strategy \( \pi_0 = 0.5 \), the maximum number of exploration rounds \( \epsilon_{\max} = 500 \), the agent constantly explores the state space, obtain and update the Q value. After the Q table and E table is complete, a swing process intelligent control simulation test can be performed.

3. Simulation and Results

In the simulation experiment, we set the SARSA(\( \lambda \)) parameters as follows: \( \alpha = 0.01 \), \( \gamma = 0.9 \), \( \pi = 0.1 \). The initial state for the swing process simulation is \( s_0 = [939.08, -54.09, 30.91] \), and the swing speed is \( V_s = 0 \). The simulation test results are compared with the actual manual operation shown in Figure 3. We can see that the swing speed under the proposed model control fluctuates greatly because there is a 10% probability of randomly selecting actions. However, the maximum change range of the swing speed is \([-2.2]\).

![Figure 3](image_url)

**Figure 3.** Comparison between manual operation with intelligent control.

In the manual operation, the swing speed fluctuates early, and the swing speed is greater at lower concentrations. However, under the control of the SARSA(\( \lambda \)), the swing speed fluctuates less, and the swing speed is lower than the manual operation. In addition, the cutter motor current is smaller than manual operation and fluctuates without overload. At the same time, suction vacuum fluctuations
are small, so it is not easy to have suction or over-absorption. In the same way, the mud concentration also reaches relatively fast and stabilizes at high concentrations. In summary, we can claim that the overall SARSA(\lambda) operation is more stable and efficient, and the total production volume can be higher than manual operation.

4. Conclusions
The operation of a cutter suction dredger is affected by various uncertainties such as soil quality, environment, and ship status. It is difficult to describe the complex and dynamic characteristics of the swing process with an ideal model. Therefore, it is often not possible to maintain optimal control during actual construction. In this paper, the proposed approach is adaptive to the swing process control model of a cutter suction dredger based on SARSA(\lambda). When the swing process environment model is changed, the swing speed can be adjusted independently to achieve the best efficiency.

The comparison between the simulation results of the SARSA(\lambda) model with the actual operational results shows that the control effect of this model is obviously better than the manual operation, and the optimization speed is fast and relatively stable. This model connects the SARSA(\lambda) algorithm theory with the actual control object, and thus it has certain significance and application value.

Acknowledgments
This research is supported by the National Natural Science Foundation of China (Grant No. 61703138), and Natural Science Foundation of Jiangsu Province (Grant No. BK20170307).

References
[1] Li W. Analysis of China Dredging Market. Port and Economy, 2006(1):38-39.
[2] NI F. Review of dredging equipment development at home and abroad. Journal of Hohai University Changzhou Branch, 2004(1):1-9.
[3] Wang Y, Xie H and Wu Y. Present State and Development of Domestic and International Dredging Markets. Journal of SSSRI, 2004(2):109-113.
[4] Zhu W, Ni F, Wei C, and Wang S. Modelling and analysis of dredging control system based on state space method. Journal of Computer Applications, 2016, 36(S2): 120-124.
[5] Wei C, Ni F, Dai W. The predictive control for the swing process of a cutter suction dredger. Machine Design and Manufacturing Engineering, 2018(1):70-74.
[6] Jiang S, Ni F, Wei C, and Li H. Research on the Multi-loop Control for Traverse System of Cutter Suction Dredger. Process Automation Instrumentation, 2016(12):62-66.
[7] Zhang D, Jiang S, Li Z. Research of an Improved Smith Production Control Method for Cutter Suction Dredger. Machinery and Electronics, 2015(7):57-60.
[8] Ge J, Zhou L. Real-time traffic signal control model based on SARSA. Computer and Engineering and Applications, 2015, 51(24):244-248.
[9] Xue L, Yin C, and Li L. Fused Sarsa-lambda learning algorithm based on multi-agent. Computer Engineering and Application, 2008, 44(4):182-184.
[10] Lin L, Wang H, and Wang Y. Sarsa Reinforcement Learning Algorithm based on Neural Networks. Computer Technology and Development, 2006(1):30-32.
[11] Eamat P, Mohsen P, Maziar P. Multi-criteria expertness based cooperative method for SARSA and eligibility trace algorithms. Business Media New York, 2015(43):487-49