Optimization of Welding Route by Automatic Machine Using Reinforcement Learning Method

Yasuhisa Okumoto[1]

Summary

A ship hull block is generally composed of skin plates, longitudinals and transverse webs as a grillage structure, and longitudinals and transverse webs are joined by fillet weld on a skin plate panel. Then much labour time is necessary for this welding work because many welding lines exist. Now automatic welding machine of simple type using truck system is applied widely as well as semi-automatic CO_2 weld or gravity weld. Since the automatic welding machine needs help of workers when initial setting, turning, and shifting, the efficient routing has to be investigated for the improvement of the productivity. However, it is very difficult to find an optimal weld sequence when weld lines increases and then the combination number of welding sequence increases. Such type of research is called combinatorial problem. This paper has examined how to decrease the work time using the reinforcement learning method which imitated the behavior pattern of animals.

Preface

The hull block is such structure that the primary supporting members and the secondary supporting members are welded on a plate panel as a grillage shape. Though the former includes transverse webs, girders and stringers, and the latter includes stiffeners, longitudinals and beams, the former is called transverse webs, and the latter are called longitudinals respectively in this paper. Since many man-hours have been taken in these welding, some automatic machines have introduced recently: such as semi-automatic CO_2 welding, gravity welding, and simple welder of car truck type. Though the welding machine of car truck type is popularized now, the worker has to carry it in the welding location (egg box), and also rotate it for another welding line. Therefore, when a lot of longitudinals and transverse webs exist, the weld lines increase and much costs are needed. Hence it is more important to select the most efficient welding route. However, the welding sequence has been decided irrationally and experimentally by the skilled workers now.

---

[1] Professor, Kinki University, Faculty of Engineering,
1 Takayaumenobe, Higashihiroshima, Hiroshima 739-2116 Japan

If the welding lines increase, the number of combinations of the execution order among much welding lines become big and it is the most difficult to find the optimum route so as to minimize the working hour. Such problem using computer simulation is called combinatorial optimization. Though meta- heuristic techniques are widely applied, the reinforcement learning method, which simulates the learning of animal action, is proposed for the optimization recently. This paper describes the root optimization of a welding robot using the reinforced learning method to reduce total welding hours, in the case that plural longitudinals and transverse webs were welded to the skin plate panel.

Outline of reinforcement learning method [1]

Reinforcement learning has been used for the meta-heuristic optimization technique recently. It seems to be a kind of dynamic programming and learning without teacher, and used in the search of the robot action under unknown environment. The reinforcement learning method is the technique that simulates a learning process of an animal. A reward like food is given when the animal succeeds or achieves a certain problem, while a punishment is administered when it fails. Through such experiments, the animal gradually learns the suitable action in the given environment. Reinforcement learning imitates such a process of optimization using a computer calculation. Though this method has been used mainly in the field of route optimization or maze problems etc., such application has not yet been seen in the field of production. This report sets out to confirm the effectiveness of this method.

Reinforcement learning is handled using a reward ***r***, which is provided from environment ***E***, when an agent, i.e. worker or robot, carries out an action **a** for a state **s** of the environment, and acts so as to receive the maximum reward from an initial state to a goal state as shown in Fig.1. The learning is so-called "learning without teacher's data", and it is not needed to prepare the right knowledge data about the environment beforehand. It is an optimization technique that can cope with moderate dynamic change in the environment, because it is a kind of learning by repeating and evaluating the action.

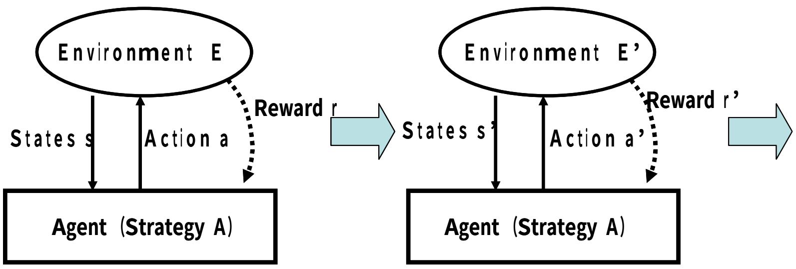


Fig.1 Mechanism of reinforcement learning method

The Q-learning method, one of the algorithms of reinforcement learning, was applied here. It is an algorithm which estimates the value function Q (s,a), obtained by the agent's repeating of the action by trial and error, for the environment. Q (s,a) expresses the expectation of a gain when the agent takes the most suitable action after having chosen an action in a state **s**. When Q (s,a) was given, the action **a** that becomes greatest is the most suitable action. The Q value is updated as in expression (1) when **r** is a reward after having chosen an action **a** in the state **s**, and **s '** is a state after transition.

$$Q(s,a) \leftarrow Q(s,a) + \alpha\{r(s,a) + \gamma \max_{a' \in A(s')} Q(s',a') - Q(s,a)\} \tag{1}$$

$Q(s,a)$: Value of action **a** at present condition **s**

$Q(s',a')$: Value of action **a'** at **s'** after transition

α : Learning rate($0<\alpha<1$)

γ: Discount rate($0<\gamma<1$)

The learning rate α indicates which is emphasized: a present Q-value or the value got by action. When it approaches 1, the result newly gained is emphasized, and the change in the Q-value increases, and when it approaches 0, the present Q-value is emphasized, and the change in the Q-value decreases. The discount rate γ shows which one is emphasized: the present reward or the future (estimated) reward, and when it is near to 1, the future reward is emphasized and when near to 0, the present reward is emphasized. Generally 0.1 is used for α and 0.9 - 0.99 is used for γ. The selection of a route is shown in Fig.2. It is necessary to choose one action executed among the many possible actions that exist. Though a number of selection methods have been proposed, the ε-greedy method is used here. This method stochastically adopts a lesser reward to prevent a fall in the local optimum solution. The ε-greedy method chooses the result for which the Q-value is a maximum at probability (1-ε) as shown in formula (2), and it is randomly chosen at probability ε (0<ε<1). When **a** is chosen in a condition **s**,

$$p(s,a) = \begin{cases} 1-\varepsilon & \text{When } Q(s,a) \text{ is maximum} \\ \varepsilon & \text{Else} \end{cases} \tag{2}$$

The procedure for Q learning is as follows:

1) Initializing of Q (s,a)
2) Initializing of state **s**
3) Selection of action **a**
4) Execution of **a** and observing of **s'**
5) Receiving reward **r**
6) Renewal of Q (s,a) using formula (1)

When the next state **s'** is the goal, the execution is terminated, otherwise the calculation repeats from above 3) as S←S'

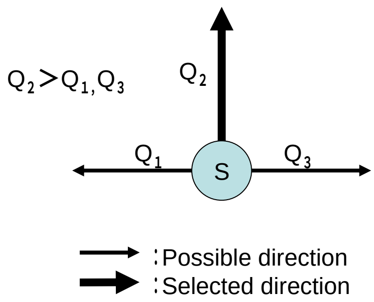


Fig.2 Selection of route

Optimizing Simulation

Fig.3 shows a typical ship hull block. As an example, the division between two transverse webs (one transverse space) was selected, and the fillet welding of the longitudinals and transverse webs by simple automatic welding machines was simulated. The welding lines in a division are modeled as shown in Fig.4. A portable welding machine is shown in Fig.5. The machine starts by a start button and stops by the touch sensing.

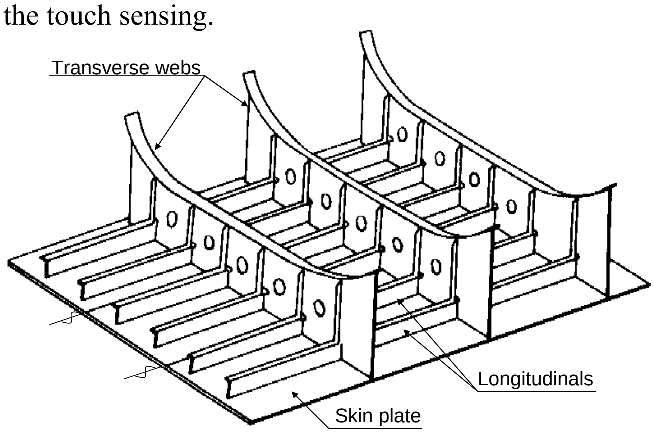


Fig.3 Typical hull block

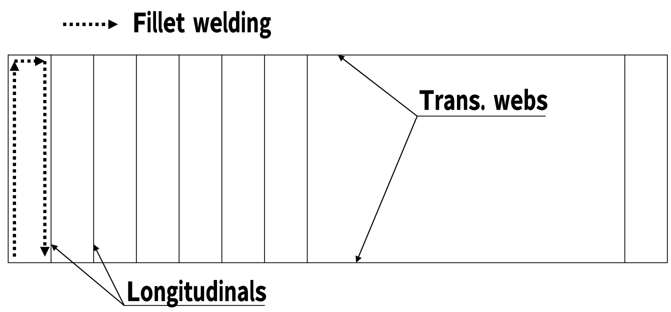


Fig.4 Fillet welding line

In the reinforcement learning method, the welding machine corresponds to the agent, and the optimum order in which the agent traces all weld lines is more easily simulated. As a criterion of the optimum welding route, scores of work which shows difficulty can be assigned in this simulation as follows:

Fig.5 Welding machine

1) Lifting and turning the machine: 0.3
2) Lifting the machine and transfer to the next space on a longitudinal: 0.6
3) Carrying the machine laterally without welding: 1.0
4) Carrying the machine longitudinally without welding: 1.5

The transverse movement corresponds to a shift of one longitudinal space, the longitudinal direction corresponds to a shift of a transverse space, and the latter is 1.5 times the former, since the latter distance is usually longer than the former. These relative numerical values were decided by direct communication with welding staffs in the field.

The above numerical value was summed for the total route, and many cases were compared to each other, and the route where the sum of the scores becomes a minimum was obtained:

$$W = \sum w_i \rightarrow \mathbf{min}. \tag{3}$$

In the work simulation calculation, the constant values were taken to be; learning rate α=0.1, discount rate γ= 0.99, and ε= 0.01, and the program was made in C language.

Though the calculation was carried out for many cases, the optimum route is shown in Fig.6 as one example. In this case the robot starts from an arbitrary position in the middle. If the initiation point is located on the left side in the subdivision, it is better to go the right part as shown in Fig.6(a), while if it is on the right side, it is better to go to left part as shown in Fig.6(b). The total work is equal for both cases (a) (b). If the initiation point is on the right side of the subdivision and the welding movement is on the right side as shown in Fig.6(c), one pass over the longitudinal is added in comparison with case (b). In the case of other starting points, or other cases of multiple machines, better solutions could be calculated. Fig.7 shows the results of the latter.

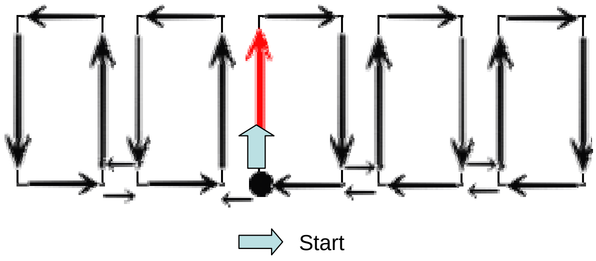
Conclusions

In this study, the root optimization of fillet welding was examined for a simple automatic machine used in the assembly of a ship hull block using the reinforcement learning method, since such fillet welding of longitudinals and transverse webs on a skin plate needs many work man-hours.

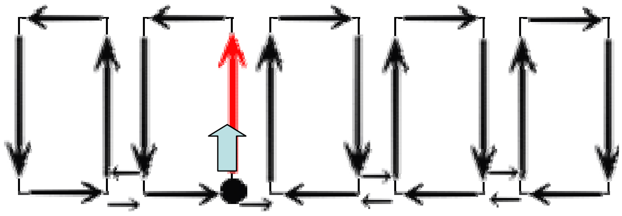
As a result, the computer simulation by the reinforcement learning method was showed to be an effective method to reduce the work time.

References

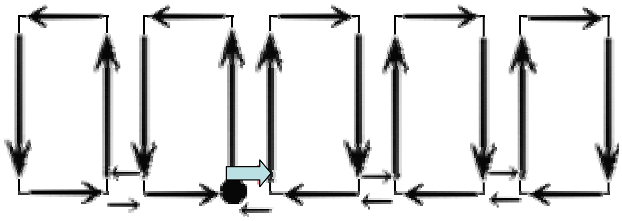
1 Araya S.: Artificial Intelligence (2nd Edition), (2004), pp.113-117, (in Japanese).



(a) Case 1

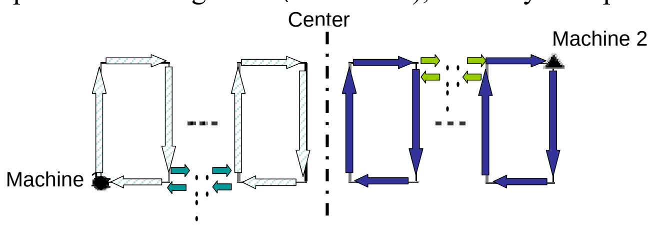


(b) Case 2

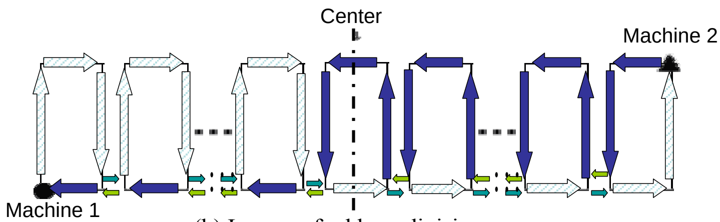


(c) Case 3

Fig.6 Optimum welding route (1 machine), arbitrary start point



(a) In case of even no. division



(b) In case of odd no. division

Fig.7 Optimum welding route (2 machines)