Reinforcement Learning Based Search for Ships' Courses Controlled by Safety

Masahiro Nakayama†, Takeshi Kamio†, Kunihiko Mitsubori††,
Takahiro Tanaka†††, and Hisato Fujisaka†

† Hiroshima City University, 3-4-1, Ozuka-higashi, Asaminami-ku, Hirhoshima, 731-3194, Japan
†† Takushoku University, 815-1, Tatemachi, Hachioji-shi, Tokyo, 193-0985, Japan
††† Japan Coast Guard Academy, 5-1, Wakaba-cho, Kure-shi, Hiroshima, 737-8512, Japan
Email: kamio@hiroshima-cu.ac.jp

Abstract– Although the ship transportation is important for low cost mass transit, the optimality of ships' courses and the interaction between maneuvering actions have not been sufficiently discussed yet. In order to brisk up these discussions, we have developed multi-agent reinforcement learning system (MARLS) to find ships' courses [1]-[4]. Although our basic MARLS [3] can keep navigation rules [5], it may get inefficient courses including larger avoidance of collisions between ships.

In this paper, we clarify the causes of this problem and propose a new MARLS controlled by the safety to overcome it. From numerical experiments, we have confirmed that our proposed MARLS can get more efficient courses than our basic MARLS.

1. Introduction

Deciding efficient and safe courses of ships before actual navigation is important. Multi-ship course problem has been treated in maneuvering simulation and automatic operation, where the course has been given as a guideline which the ship should trace and the procedures to avoid collisions between ships have been discussed. But, the optimality of the course and the interaction between maneuvering actions have not been sufficiently discussed yet. We regard multi-agent reinforcement learning system (MARLS) as a useful tool to brisk up these discussions, since ships have the special conditions in the maneuvering [1]-[4]. The conditions are as follows: 1) the dynamics is nonlinear, 2) there is no way to brake and go backward effectively, 3) the attitude is unstable at a low speed, and 4) the control tower with the strong authority does not exist. Our basic MARLS [3] can avoid collisions between ships based on navigation rules (NRs), which are international regulations for collision avoidance [5]. Our basic MARLS tends to search efficient courses after it has obtained courses which may include larger avoidance than necessary. Therefore, if learning is continued for a long time, efficient courses can be obtained. However, it becomes impractical according as the difficulty of a given problem increases. Although we have modified our basic MARLS to improve the course efficiency, it cannot consider the safety explicitly [4]. Therefore, our modified MARLS may get danger courses.

In this paper, to get efficient courses in limited learning time, we propose the way to suppress larger avoidance using safety. In numerical experiments, we compare our new proposed MARLS with our basic MARLS. As a result, it has been confirmed that our proposed MARLS can get more efficient courses than our basic MARLS.

2. Basic MARLS to Find Ships' Courses [3]

2.1. Multi-Ship Course Problem

Fig.1(a) shows the model of ship maneuvering motion. To simplify the discussion, there is no external force (e.g., tidal current). But, using our previous work [1], we can consider the tidal current effects. O_S is the center in turning the ship's head and shows the ship's position (i.e., $O_S=(x, y)$). ϕ is the heading angle. L_S is the ship's length. $\mathbf{v}_0$ is the velocity and its size is V_0. The dynamics is given by KT model [6] as follows:

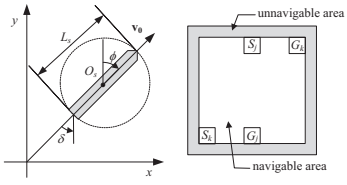
$$T\ddot{\phi}+\dot{\phi}=K\delta,\ \ \dot{x}=V_0\sin\phi,\ \ \dot{y}=V_0\cos\phi, \qquad (1)$$

where δ is the rudder angle. T and K are the maneuvering performance parameters and they are given by $K=K_0/(L_S/V_0)$ and $T=T_0(L_S/V_0)$. Each ship has individual values of K_0 and T_0. When many ships are in a limited sea area, actual navigators tend to avoid collisions by only changing the direction before changing the speed. From this fact, we fix V_0 at the standard value.

Fig.1(b) shows the model of sea area. It defines the start (S) and goal (G) for each ship in the navigable area (white). Also, it defines the unnavigable area (gray) which represents obstacles. Therefore, we judge that MARLS has obtained a solution of multi-ship course problem if the following conditions are satisfied: 1) all the ships arrive at their goals without entering the unnavigable area, 2) there is no collision between ships.

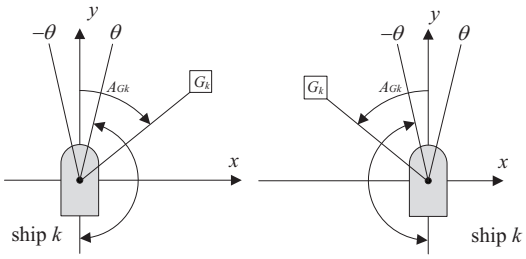
2.2. Basic Structure of MARLS

We show the basis of our MARLS which uses Q-learning. There are some assumptions to solve multi-ship course problem by MARLS. A navigator is regarded as an agent and the number of agents is N. The perceptual input of agent k consists of the own ship's information $\boldsymbol{I}_k$ and other ship's information $\boldsymbol{D}_k$. The action is defined by the rudder angle δ_k. If the ship k is in the goal G_k, unnavigable area, and the others, the agent k receives $r_A=1$, $r_F=-1$, and zero as the reward, respectively. Also, when the ship k collides with other ships, the agent k receives r_F. The way to judge the collision is described in Sect. 2.3.2.



(a) Model of ship maneuvering motion. (b) Model of sea area.

Fig.1 Models of ship maneuvering motion and sea area.



(a) $\theta < A_{Gk} \leq \pi$. (b) $-\pi < A_{Gk} < -\theta$..

Fig.2 Limited action selection based on GO.

2.3. Prior Knowledge

Our basic MARLS [3] uses goal orientation (GO) and navigation rules (NRs) as prior knowledge. They help to improve the learning efficiency.

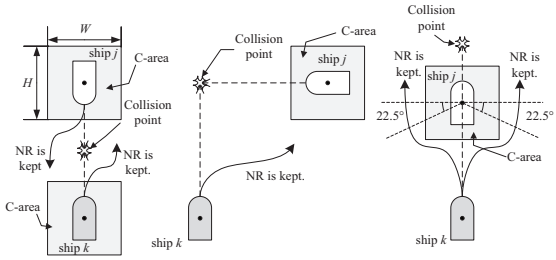
2.3.1. Goal Orientation (GO)

GO is based on the idea that a ship ought to move to the goal if there is no danger of collisions. GO is implemented by limiting the action selection when the ship's heading angle differs widely from the goal direction and there is no danger of collisions. We show limited action selection based on GO (LAS_{GO}). Fig.2 shows the criteria. These are applied to the ship which has no need to avoid other ships. If $\theta < A_{Gk} \leq \pi$ as shown in Fig.2(a), the action selection is limited so that $\delta_k \geq 0$ (i.e., turn to the right). If $-\pi < A_{Gk} < -\theta$ as shown in Fig.2(b), the action selection is limited so that $\delta_k \leq 0$ (i.e., turn to the left).

2.3.2. Navigation Rules (NRs)

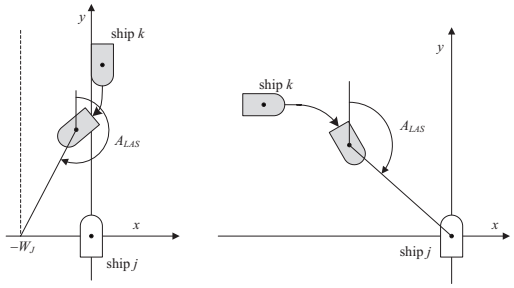
Fig.3 illustrates the collision situations, NRs, and C-area. Fig.3(a) shows Head-on-situation and each ship must change the course to the right to avoid the collision. Fig.3(b) shows Crossing situation and the ship which has the other ship on the right side must change the course to the right. Fig.3(c) shows Overtaking and the overtaking ship must change the course to the right or the left. When the ship k must avoid the collision with the other ship j according to NRs, C-area is placed around the ship j. If the ship k enters C-area around the ship j, then only the ship k receives a penalty (i.e., negative reward r_F).

Our MARLS limits the action selection in the execution of Q-learning to keep NRs strongly. We explain limited action selection based on NRs (LAS_{NR}). If observing Fig.3 carefully, you can see that the avoiding ships in Head-on-



(a) Head-on-situation. (b) Crossing situation. (c) Overtaking.

Fig.3 Collision situation, NRs, and C-area.



(a) Head-on-situation. (b) Crossing situation.

Fig.4 Limited action selection based on NRs.

situation and Crossing situation must change the course to the right. That is to say, the action selection should be limited so that $\delta_k \geq 0$. But, to avoid turning to the right unnecessarily, LAS_{NR} is not available if $\phi_k \geq A_{LAS}$ as shown in Fig.4.

2.4. Process Flow of Basic MARLS

Here, we review the process flow of our basic MARLS [3]. Our MARLS is based on Q-learning and uses GO and NRs as prior knowledge. Since they are implemented by limiting action selection, our MARLS can easily get the courses which satisfy NRs. Moreover, since the limited action selection (LAS) based on NRs and GO prevents each agent from learning extra states, the leaning efficiency will also be improved. The following processes are iterated until the end condition is satisfied:

1) At the beginning of each episode, the judgment status for collision situation (J_{kj}) is set to free.
2) After starting each episode, the agent k always detects other ships in the view circle of the radius R_k.
3) If the ship j is in the view circle and the status J_{kj} is free, the agent k judges the collision situation by NRs.
4) The status J_{kj} is made free according to the relationship between ships k and j.
5) Q-learning is executed applying LAS designated by the status J_{kj}.

3. Proposed Method

3.1. Drawback of Basic MARLS and its Causes

Even if a learning trial is successful in our basic MARLS, the courses to avoid collisions may be longer than necessary. Here we consider the reason why such inefficient courses are obtained.

It is assumed that the own ship k has detected the other ship j in the view circle of radius R_k, the collision situation is Head-on-situation or Crossing situation, and the ship k must avoid the collision with the ship j according to navigation rules (NRs). In this case, since $\mathrm{LAS_{NR}}$ is applied, the action selection of the ship k is limited so that $\delta_k \geq 0$, which means go straight or turn right. Although $\mathrm{LAS_{NR}}$ can make the ship k keep NRs strongly, the ship k begins to turn right at an early timing (in other words, at the position that is quite far from the ship j) and the course to avoid the ship j becomes longer than necessary. This phenomenon is often observed at the early stage of learning. Since larger avoidance makes it easy to reach the goal, it is natural that $\mathrm{LAS_{NR}}$ should induce the phenomenon. Therefore, our basic MARLS tends to search efficient courses after it has obtained courses which may include larger avoidance than necessary but connect the start with the goal without collisions. However, the end condition of learning aims at obtaining feasible courses in limited time. As a result, learning may be stopped before efficient courses are found.

These above facts are considered the reason why our basic MARLS may obtain inefficient courses including larger avoidance than necessary.

3.2. $\mathrm{LAS_{NR}}$ Controlled by Degree of Safety

The simplest way to solve the drawback of our basic MARLS is to continue learning until searching courses is sufficiently executed. However, it becomes impractical according as the difficulty of a given problem increases. Therefore, we propose to control the execution timing of $\mathrm{LAS_{NR}}$ by the degree of safety in order to obtain the efficient courses in limited time.

To implement the above proposition, we construct the evaluation model of safety as follows. For example, as shown in Fig.5(a), it is assumed that the own ship k must avoid the other ship j in Crossing situation.

If the ship j is out the view circle of the ship k, the ship k cannot detect the ship j. In this case, the ship k judges that there is no danger of collision. Therefore, it is valid that the degree of safety S_{kj} is set to 100%:

$$S_{kj} = 100 \quad if \quad D_{kj} \geq R_k , \tag{2}$$

where D_{kj} is the distance between ships k and j. Also, the ship k has the personal area PA_k detailed in Fig.5(b). If the ship j is in the area PA_k, the actual navigator of ship k feels very strongly danger of collision [7]. Therefore, if the position of ship j (i.e., O_{Sj}) is included in the area PA_k, it is valid that S_{kj} is set to zero:

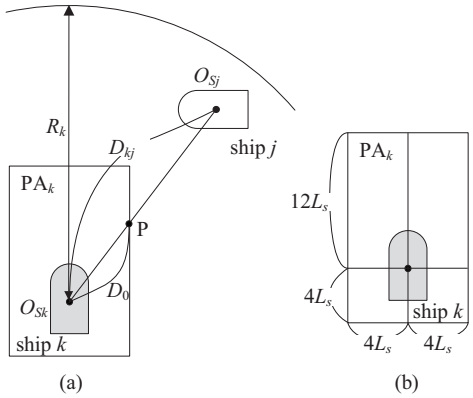


Fig.5 Evaluation model of degree of safety.

$$S_{kj} = 0 \quad if \quad O_{Sj} \in PA_k . \tag{3}$$

If these two situations are not true, we consider that S_{kj} should be given according to the relationship between PA_k and O_{Sj}. So, we propose the following S_{kj}:

$$S_{kj} = 100 \times \frac{D_{kj} - D_0}{R_k - D_0} \quad if \quad O_{Sj} \notin PA_k \text{ and } D_{kj} < R_k , \tag{4}$$

where D_0 is the length of the segment $O_{Sk}P$ and P is the intersection of the segment $O_{Sk}O_{Sj}$ and the boundary of area PA_k. Eq.(4) means that S_{kj} decreases linearly according as the ship j approaches the personal area of the ship k.

When the ship k must avoid the ship j according to NRs, our new proposed MARLS controls the execution of LAS_{kj} by the degree of safety S_{kj} as follows:

$$\mathrm{LAS}_{kj} = \begin{cases} \mathrm{LAS_{GO}} & if \quad S_{kj} \geq S_{ref} \\ \mathrm{LAS_{NR}} & if \quad S_{kj} < S_{ref} \end{cases} , \tag{5}$$

where S_{ref} is a parameter to decide the execution timing of $\mathrm{LAS_{NR}}$. Although S_{ref} is given heuristically in this paper, Eq.(5) prevents the ship k avoiding the ship j at an early timing. Therefore, it is expected that our proposed MARLS can get more efficient courses than our basic MARLS.

4. Numerical Experiments

Experiments have been carried out to investigate the performance of our proposed MARLS. Fig.6 shows the test problem which includes 6 ships in $42L_S \times 42L_S$ sea area. To simplify the discussion, all the ships have common parameters except for their start and goal positions. The parameters of ships are L_S=107(m), V_0=6.17(m/s), K_0=1.310, T_0=1.085, $\delta \in \{0.0, 10.0, -10.0, 20.0, -20.0\}$ (deg.). The policy of Q-learning is ε-greedy policy. The parameters of C-area are $H=2L_S$ and $W=2L_S$. The radius of view circle R_k is $40L_S$. The parameters of LAS are $W_J=L_S$, θ=1.0(deg.), S_{ref}=70. Moreover, there are other parameters which are same as ones in Ref.[3]. The maximum number of episodes in each learning trial is 300000.

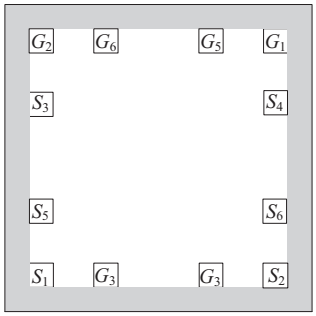
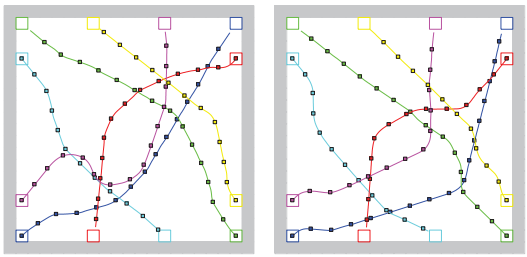


Fig.6 Test problem (6 ships in $42L_S \times 42L_S$ sea area).



(a) Basic MARLS. (b) Proposed MARLS.

Fig.7 Examples of courses obtained by our basic MARLS and our proposed MARLS.

Table 1 Comparison results in terms of N_{SLT}, L_{AVE}, L_{MIN}, and L_{MAX}.

	N_{SLT}	L_{AVE} [m]	L_{MIN} [m]	L_{MAX} [m]
Basic MARLS	30	30156	29344	32207
Proposed MARLS	30	29231	28227	31244

Table 2 The classification of learning trials based on the range of length of courses.

		Range of length of courses [m]				
		case 1 [28000, 29000)	case 2 [29000, 30000)	case 3 [30000, 31000)	case 4 [31000, 32000)	case 5 [32000, 30000)
The number of successful learning trials	Basic MARLS	0	18	7	4	1
	Proposed MARLS	18	4	6	2	0

The end condition is as follows: a learning trial is successful if the task achievement ratio is over 80% for 20000 successive episodes. Task achievement means that all ships arrive at their goals without collisions in an episode. Also, if a learning trial is successful, we have estimated the course of ship whose initial heading angle is the goal direction. The number of learning trials is 30.

Fig.7(a) shows an example of inefficient courses which are often obtained by our basic MARLS. On the other hand, Fig.7(b) shows a typical example of courses obtained by our proposed MARLS. Table 1 shows comparison results between our basic MARLS and our proposed MARLS in terms of N_{SLT}, L_{AVE}, L_{MIN}, and L_{MAX}. N_{SLT} is the number of successful learning trials. L_{AVE} is the average length of courses. L_{MIN} is the minimum length of courses. L_{MAX} is the maximum length of courses. Table 2 shows the classification of learning trials based on the range of length of courses. From these results, we can find following. Fig.7 shows our proposed MARLS has suppressed larger avoidance of 5th ship obtained by our basic MARLS. Also, Table 1 shows our proposed MARLS can get shorter courses than our basic MARLS. Therefore, we can judge that our proposed MARLS is superior to our basic MARLS in terms of the course efficiency. On the other hand, Table 2 shows our proposed MARLS sometimes gets inefficient courses which correspond to cases 3-5. The following are considered as the causes. Our proposed MARLS has the possibility that ships are closer than necessary. In this case, ships need larger avoidance of collisions. Therefore, to overcome this problem, we will consider setting S_{ref} according to the collision situation.

5. Conclusions

Our basic MARLS [3] may obtain inefficient courses including larger avoidance than necessary. To overcome this problem, we have proposed the way to suppress larger avoidance using safety. From numerical experiments, we have confirmed that our proposed MARLS can get more efficient courses than our basic MARLS. However, our proposed MARLS have not always obtained efficient courses. In the future, we will consider the way to set S_{ref} according to the collision situation to always obtain efficient courses.

Acknowledgments

This work was supported by JSPS KAKENHI Grant Number 24500179.

References

[1] K. Mitsubori, T. Kamio, and T. Tanaka, “Finding the shortest course of a ship based on reinforcement learning algorithm,” Journal of Japan Institute of Navigation, 110, pp.9-18, 2004.

[2] K. Mitsubori et al., “Finding the course and collision avoidance based on reinforcement learning algorithm,” NAVIGATION, 170, pp.26-31, 2009 (in Japanese).

[3] T. Kamio et al., “Effects of prior knowledge on multi-agent reinforcement leaning system to find courses of ships,” Australian Journal of Intelligent Information Processing Systems, vol.12, no.2, pp.18-23, 2010.

[4] T. Tanigawa et al., “Modified multi-agent reinforcement learning system to find ships’ courses,” Proc. of NOLTA, pp.487-490, 2013.

[5] International Maritime Organization, “International regulations for preventing collisions at sea,” 1972.

[6] T. I. Fossen, “Guidance and control of ocean vehicle,” John Wiley & Sons Ltd., pp.172-174, 1994.

[7] M. Fuchi et al., “Differences in allowable distance between ships due to ship’s specific type,” Japan Ergonomics Society, vol.46, pp.144-145, 2010 (in Japanese).