2022 International Symposium on Nonlinear Theory and Its Applications,
NOLTA2022, Virtual, December 12-15, 2022

Identification of Avoidance Starting Points by Reinforcement Learning-Based Multi-Ship Course Search Method with Target Courses as Actions

Takeshi Kamio†, Hiroki Kimura†, Takahiro Tanaka††, Kunihiko Mitsubori†††, and Hisato Fujisaka†

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

Abstract– Since navigation rules (NRs) only roughly define how to avoid collisions between two ships, the actual navigators must make decisions about the direction and timing for avoidance based on their experience. Therefore, the decisions of the unskilled navigators tend to be ambiguous. Against this background, we have discussed course efficiency and safety using a multi-agent reinforcement learning system (MARLS) to search ships' courses. However, we have not discussed avoidance timing. In this paper, we propose a method to identify avoidance starting points using our MARLS. Through numerical experiments, we have confirmed that our proposed MARLS can find efficient courses and converge the avoidance starting points corresponding to each avoidance to a small area.

1. Introduction

Recently, many researches on unmanned autonomous ships have been conducted and their purpose is to build a versatile automatic collision avoidance system using deep reinforcement learning [1], [2].

On the other hand, we have developed multi-agent reinforcement learning system (MARLS) to search ships' courses [3], [4] and have considered how to use it as marine traffic assessment tools [5], [6]. This is because it remains important to pre-select safe and efficient courses for complex collision situations where even actual navigators are at a loss to make a decision, and to get useful knowledge from the process of the course selection.

The reason for the navigator's confusion is due to the lack of clarity in the collision avoidance method prescribed by navigation rules (NRs) [7]. For example, NRs order that the ship which has the other ship on the right side must change the course to the right in the collision situation called crossing situation. However, NRs do not request the direction and timing for the avoidance. Therefore, actual navigators must make appropriate decisions based on their experience. Against this background, researches on collision avoidance systems have long been conducted. Recent researches on unmanned ships are also essentially the construction of collision avoidance systems. However, these researches do not aim to get useful knowledge through the identification of avoidance starting points.

In this paper, we propose a method to identify avoidance starting points using our MARLS. Through numerical experiments, we have confirmed that our proposed MARLS can find efficient courses and converge the avoidance starting points corresponding to each avoidance to a small area. Also, we mention the possibility of getting useful knowledge from the courses and avoidance starting points obtained in the test problem.

2. MARLS to Search Ships' Courses

2.1. Basic MARLS for Multi-Ship Course Problems

Fig.1 is a model of ship maneuvering motion. **O** is the center in turning the ship's head and shows the ship's position (i.e., **O**=(x, y)). ϕ is the heading angle. L_S is the ship's length. **v** is the velocity and its size is V. The motion equation is given by TK model as follows:

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

where δ is the rudder angle. T and K are the maneuvering performance parameters which are given by K=K_0/(L_S/V) and T=T_0(L_S/V). Each ship has individual K_0 and T_0. Also, since actual navigators tend to avoid collisions by only changing the direction before changing the speed in congested sea area, we fix V at the standard value.

Fig.2 is a model of sea area. Fig.2(a) is a common sea area which all ships share and it defines the start (S) and the goal (G) for each ship in the navigable area (white). Also, it defines the unnavigable area (gray) which represents obstacles. Fig.2(b) is an individual sea area which each ship occupies and it is based on the common sea area. It consists of grids whose side length is fixed at L_G (=2L_S in this paper). Each grid is numbered for Q-learning (QL). There are 4 kinds of grids: start one (S), goal one (G), navigable one (white), and unnavigable one (gray). Each ship is permitted to move every grid except for unnavigable ones. Therefore, we judge that MARLS has obtained a solution if all the ships arrive at their goal grids without entering any

ORCID iDs First Author: 0000-0002-7661-0898, Second Author: 0000-0002-3956-1951, Third Author: 0000-0002-6956-4838, Fourth Author: 0000-0002-6962-0596, Fifth Author: 0000-0003-2619-5224 unnavigable grids in their individual sea area and there is no collision between ships in the common sea area.

Next, we explain the basis of our basic MARLS [3], [4] which uses QL (hereafter B-MARLS). There are some assumptions to solve multi-ship course problems by MARLS. A navigator is regarded as an agent. The perceptual input of agent k consists of its own ship's information $\mathbf{I}_k=(x_k, y_k, \phi_k, \dot{\phi}_k)$ and other ships' information $\mathbf{D}_k$. If there are other ships which the ship k needs to avoid according to navigation rules (NRs), $\mathbf{D}_k$ is generated based on the directions where they exist. The state is defined by $\mathbf{I}_k$ and $\mathbf{D}_k$. The action is defined by the rudder angle δ_k. If the ship k is in the goal grid G_k, unnavigable ones, 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 another ship, the agent k receives r_F. The judgment of the collision is executed as follows. When the ship k must avoid the collision with the ship j according to NRs, the collision area (C-area) is placed around the ship j. If the ship k enters the C-area around the ship j, then only the ship k receives a penalty (i.e., r_F). When all the agents reach their terminal states (i.e., receive r_A or r_F), the present episode is finished and the next one is started. Therefore, the agent k optimizes Q-value by iterating episodes until the end condition is satisfied. The end condition of a learning trial is based on the task achievement rate detailed in Sect.4. Also, the task achievement means that all the ships arrive at their goals in an episode. Moreover, B-MARLS uses the limited action selections (LASs) based on NRs and the goal orientation (GO) to keep NRs, improve the learning efficiency, and suppress the influence of the concurrent learning problem. These LASs are detailed in Refs. [3], [4].

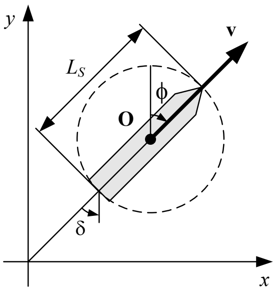
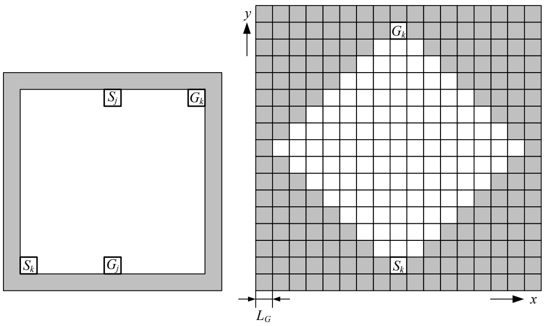


Fig.1 A model of ship maneuvering motion.



(a) Common sea area (b) Individual sea area

Fig.2 A model of sea area.

2.2. TC-MARLS

To further improve the learning efficiency of B-MARLS, we have proposed MARLS with target courses instead of rudder angles as actions (hereafter TC-MARLS) [6]. TC-MARLS is based on B-MARLS except for the definition of actions. The target courses are defined based on actual courses in congested sea area and correspond to maintaining the course, avoiding collision, suppressing the amount of avoidance, and recovering the course. Also, they are basically designed to satisfy NRs or GO.

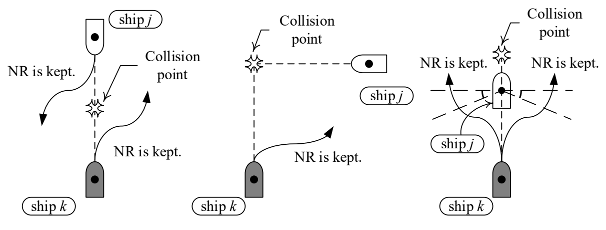
TC-MARLS is constructed to keep NRs for three typical collision situations: head-on-situation, crossing situation, and overtaking shown in Fig.3. As an example, this section describes the definition of the target courses for the crossing situation.

Fig.3(b) shows that if ships k and j continue to move straight ahead, the crossing situation will eventually occur. According to NRs, the ship k must avoid the collision with the ship j by changing the course to the right at arbitrary timing. To satisfy this request, the target courses A to D are prepared for the ship k as shown in Fig.4(a). Since the ship cannot change the course in a moment, the starting point of each target course is d [m] away from the present position of the ship. The course A means maintaining the present course toward the goal (G_k) and the courses B to D correspond to avoiding collision. The courses B to D are set based on the course A.

However, if the target course for collision avoidance is continuously taken, the ship k may generate large avoidance unnecessarily. To overcome this problem, the target courses A to D are prepared for the ship k as shown in Fig.4(b). The course A means maintaining the present course and the courses B to D correspond to suppressing the amount of avoidance. Although the courses B and C are set based on the course A, the course D is set parallel to the straight line connecting the start (S_k) and G_k.

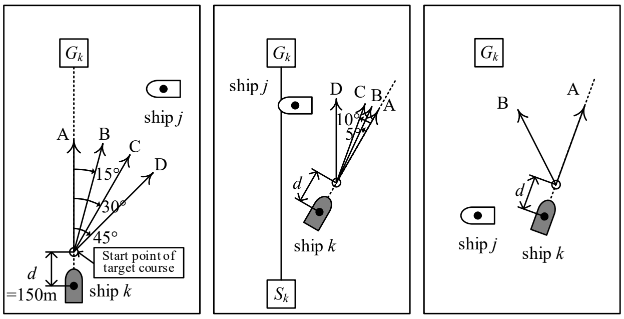
Moreover, when the ship k achieves collision avoidance with the ship j, the ship k is allowed to recover the course. Therefore, the target courses A and B are prepared for the ship k as shown in Fig.4(c). The course A means maintaining the present course and the course B corresponds to recovering the course toward G_k.

As mentioned above, TC-MARLS switches the set of target courses according to the situation.



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

Fig.3 Typical collision situations and NRs.



(a) Avoiding collision (b) Suppressing the amount of avoidance (c) Recovering the course

Fig.4 Definition of target courses for crossing situation.

3. Detecting Method of Avoidance Starting Points

As mentioned in Sect.2.2, the actions of TC-MARLS include target courses corresponding to collision avoidance. Therefore, the moment when the agent selects a target course for collision avoidance corresponds to the avoidance starting point. However, since we want to obtain some useful knowledge by converging the avoidance starting points to a small area, it is necessary to set conditions that the avoidance starting points must satisfy.

Therefore, we propose a method for detecting the avoidance starting point under the condition that the current course is maintained as long as possible while ensuring safety. The degree of the safety can be controlled by the shape of the collision area (C-area) mentioned in Sect.2.1. Hereafter, the detected point is considered to be the avoidance start limit point.

The basic idea for detecting the avoidance start limit point is as follows. We assume that each ship tracks a goal-oriented course at the start of the episode and at the time when it chooses to recover the course. In this situation, even if NRs requires the ship k to avoid the ship j, TC-MARLS forces the ship k to continue tracking the goal-oriented course and induces a collision. However, if the collisions between ships k and j are repeated, the range within which the ship k must track the goal-oriented course is gradually reduced.

Based on the above basic ideas, we implement the following method for detecting the avoidance start limit point. First, we list the assumptions.

– Each ship tracks a goal-oriented course at the start of the episode and at the time when it chooses to recover the course.
– The target course for collision avoidance can only be selected for switching from a goal-oriented course.
– $C_{kj}\in[0, 1]$ is given as the permission criterion for the ship k to start avoiding the ship j according to NRs; the larger C_{kj}, the earlier the ship k can avoid the ship j. At the start of learning, C_{kj} is set to zero.

Next, the procedure for detecting the avoidance start limit point is shown below.

(1) The relative distance between the ships k and j at the moment when it is judged that the ship k, which is tracking the goal-oriented course, should avoid the ship j according to NRs is stored as D_{kj}. On the other hand, when it is judged that the ship k does not need to avoid the ship j, $D_{kj} = \infty$.

(2) For the ship j satisfying $D_{kj}\neq\infty$, the relative distance $d_{kj}(t)$ between the ships k and j at time t is calculated. If there exists the ship j that satisfies Eq.(2), then the ship k can select either a goal-oriented course or a target course for collision avoidance. Otherwise, the ship k maintains the goal-oriented course.

$$d_{kj}(t)/D_{kj} \leq C_{kj}\,. \qquad (2)$$

(3) Each time the ship k collides with the ship j, C_{kj} is increased by Eq.(3), where Δ is a small positive number and $C_{kj} = 1$ if C_{kj} exceeds one.

$$C_{kj} \leftarrow C_{kj} + \Delta\,. \qquad (3)$$

(4) The steps (1) to (3) are repeated while executing the current learning trial. The avoidance start limit points are detected by checking when the target course for collision avoidance is selected in the obtained courses.

4. Numerical Experiments

Fig.5 is a test problem including six identical ships in $42L_S\times42L_S$ sea area. For simplification of discussion, each agent has common parameters except for the start and the goal positions. The important parameters in this paper are as follows. The action of B-MARLS is defined by the rudder angle $\delta\in\{0, 5, -5, 10, -10\}$ [deg.]. The action of TC-MARLS is defined by the target course, as shown in Fig.4. The rudder angle is determined by tracking control [6] to track the selected target course. The range of the rudder angle is given by [−10, 10] [deg.]. As mentioned in Sect.3, C_{kj} is the criterion for allowing the ship k to avoid the ship j. The parameter to adjust C_{kj} is $\Delta=2\times10^{-4}$. The other parameters are set according to Refs. [3], [6]. The maximum number of episodes in each learning trial is 100000. The end condition is as follows: a learning trial is successful if the task achievement rate SR is over 80% for 20000 successive episodes. SR is calculated using recent 5000 episodes. The number of learning trials is 30. After a learning trial is successful, MARLS calculates a set of courses without learning and randomness. We call it an obtained course.

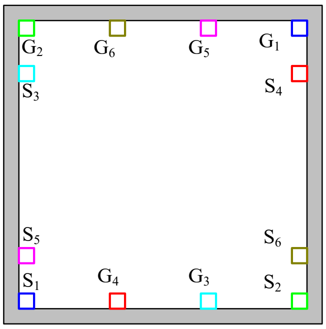
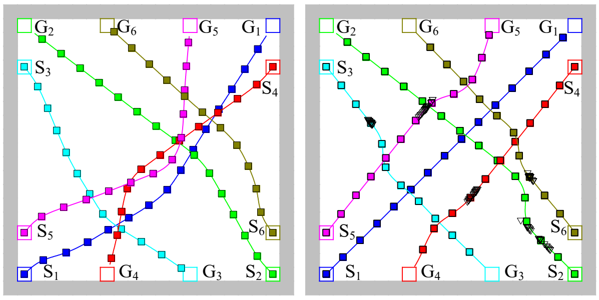


Fig.5 Test problem.

Table 1 Learning and course efficiencies.

	N_{SLT}	N_{EPS}	N_S	N_{GET}	L_{ave}(m)	L_{min}(m)	L_{max}(m)
BM	30	36748	12871	28	29138	28718	29780
TM_{ori}	29	25535	2204	29	27653	27508	28026
TM_{asp}	30	25740	909	30	26813	26619	26965



(a) TM_{ori} (b) TM_{asp}

Fig.6 Examples of obtained courses in TM_{ori} and TM_{asp}.

Table 1 shows the learning and course efficiencies in B-MARLS (BM), the original TC-MARLS (TM_{ori}), and TC-MARLS for detecting avoidance starting point (TM_{asp}). N_{SLT} is the number of successful learning trials. N_{EPS} is the average number of episodes executed in successful trials. N_S is the average number of states used actually. N_{GET} is the number of obtained courses without collisions. L_{ave}, L_{min}, and L_{max} are the average, minimum, and maximum lengths of obtained courses, respectively.

From Table 1, we can find following. As mentioned in Sect.2.2, we have proposed TM_{ori} to improve the learning efficiency of BM. Comparing N_{EPS} of BM and TM_{ori}, we can reconfirm the above purpose is achieved. Although N_S of TM_{asp} is much smaller than that of TM_{ori}, both of them have almost the same N_{EPS}. This means that TM_{asp} pay a lot of calculation cost for detecting the avoidance starting points. Improving the learning efficiency of TM_{asp} is one of our future works. From L_{ave}, L_{min}, and L_{max}, it can be seen that the course efficiency improves in the order of BM, TM_{ori}, and TM_{asp}. The reason why TM_{ori} is better than BM is that TM_{ori} can suppress large avoidance and meanders observed in BM, and this fact has already been confirmed in Ref.[6]. On the other hand, the reason why TM_{asp} has better course efficiency than TM_{ori} is revealed by comparing their obtained courses. Fig.6 shows examples of the obtained courses in TM_{ori} and TM_{asp}. Square marks show the position of each ship every 60 seconds. Triangular marks shows the avoidance starting points detected by TM_{asp} during the 30 learning trials. Fig.6(a) shows that TM_{ori} gets the obtained course which keeps NRs. Fig.6(b) shows that TM_{asp} gets the obtained course which ignores NRs if the safety is ensured. Therefore, it can be confirmed that TM_{asp} has better course efficiency than TM_{ori} as a result of suppressing the amount of avoidance by ignoring NRs.

Next, we consider the avoidance starting points detected by TM_{asp}. From Fig.6(b), we can find following. There are no triangular marks on the course of ship 1. This means that ship 1 does not need to avoid any ships. In the case of ships 3 and 6, the convergence area of the avoidance starting points corresponding to each avoidance is quite small. However, the convergence areas of ships 2, 4, and 5 are not as small as those of ships 3 and 6. Observing ships 2, 4, and 5 during the learning process, we can confirm that they repeatedly make trial-and-error attempts to avoid collisions. In other words, it is considered that the convergence area has expanded due to the complexity of the collision situation. These facts suggest the possibility that the complexity of the collision situation may be quantified from the shape and size of the convergence area, which is expected to provide useful knowledge for actual navigators.

5. Conclusions

We have proposed a method to identify avoidance starting points using TC-MARLS. We have confirmed that proposed MARLS can find efficient courses and converge the avoidance starting points corresponding to each avoidance to a small area. Also, we have found the possibility that the shape and size of the convergence area provide useful knowledge for actual navigators. In the future, we will investigate the relationship between the degree of the safety and the avoidance starting points in detail. Moreover, we will quantify the complexity of the collision situation from the convergence area.

Acknowledgments

This work was supported by JSPS KAKENHI Grant Number JP19K12100.

References

[1] H. Shen, et al., “Automatic collision avoidance of multiple ships based on deep Q-learning,” Applied Ocean Research vol.86, pp.268-288, 2019.

[2] C. Wang, et al., “Research on intelligent collision avoidance decision-making of unmanned ship in unknown environments,” Evolving Systems, vol.10, no.4, pp.649-658, 2019.

[3] T. Kamio, et al., “Effects of prior knowledge on multi-agent reinforcement learning system to find courses of ships,” AJIIPS, vol.12, no.2, pp.18-23, 2010.

[4] T. Kamio, et al., “Finding course of ships by multi-agent reinforcement learning with a priori knowledge,” IEICE Technical Report, NLP, 109(458), pp.21-26, 2010 (in Japanese).

[5] T. Tomihara, et al., “Modification of near-miss courses by reinforcement learning to search ships’ courses,” Proc. of NOLTA, pp.633-636, 2019.

[6] H. Kimura, et al., “A reinforcement learning based approach to search ships’ courses using tracking control,” Proc. of NOLTA, pp.199-202, 2020.

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