2017 International Symposium on Nonlinear Theory and Its Applications,
NOLTA2017, Cancun, Mexico, December 4-7, 2017

Influence of Reference Courses on Reinforcement Learning to Search Ships' Courses

Takeshi KAMIO†, Takahiro TANAKA††, Kunihiko MITSUBORI†††, and Hisato FUJISAKA†

† Hiroshima City University, 3-4-1, Ozuka-higashi, Asaminami-ku, Hirhoshima, 731-3194, Japan
†† Japan Coast Guard 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– Deciding efficient and safe courses of ships before actual navigation is very important. We have developed multi-agent reinforcement learning system (MARLS) to search ships' courses as a useful tool to discuss the appropriateness of courses and the interaction between maneuvering actions. In this paper, we design a novel MARLS to search ships' courses based on the reference courses. Also, we evaluate the influence of the reference courses on MARLS through the numerical simulations.

1. Introduction

Although the ship transportation is important for low cost mass transit, the appropriateness 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]-[7]. Especially, the MARLS in Ref. [4] is the basic model and it has been used in our subsequent studies [5]-[7].

The most important feature of our basic MARLS [4] is that navigation rules (NRs) and goal orientation (GO) are implemented by limiting the action selection in the reinforcement learning. NRs and GO are practical rules in the ship maneuvering. NRs are international regulations [8] for collision avoidance. GO is a rule for course recovery based on the idea that a ship ought to move to the goal if there is no danger of collisions. Therefore, we have called the limited action selections based on NRs and GO $\mathrm{LAS_{NR}}$ and $\mathrm{LAS_{GO}}$ respectively.

Although our MARLS is modeled in the distributed learning environment, Q-learning (QL) is used as the reinforcement learning. In general, it is well-known that the concurrent learning problem gives serious damages such learning systems. However, we have confirmed that NRs and GO can effectively suppress the influence of concurrent learning problem in our MARLS [4]-[7].

In our previous studies on MARLS, we have developed several methods to improve the leaning efficiency and the quality of courses (e.g., length of courses and degree of safety of courses). However, considering the application of our previous MARLS to the real multi-ship course problem, we have to judge that our MARLS is an awkward tool.

In this paper, we propose a novel MARLS to search ships' courses based on the reference courses. Because we expect that our new MARLS grows into a valuable tool to improve the actual courses by using them as the initial reference ones appropriately. The main design idea is to introduce the limited action selection to trace the reference courses ($\mathrm{LAS_{RC}}$) into our basic MARLS. However, $\mathrm{LAS_{NR}}$ and $\mathrm{LAS_{GO}}$ have priority over $\mathrm{LAS_{RC}}$. Also, updating the reference courses is executed.

As the preparatory stage for our expectation, we evaluate the influence of the reference courses obtained by our basic MARLS on our proposed MARLS through the numerical simulations. From the results of simulations, we have found that our proposed MARLS can improve the reference courses with larger avoidance than necessary and can also decrease learning time drastically.

2. Basic MARLS [4]

2.1. Multi-Ship Course Problem

Fig.1 is the 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 dynamics is given by KT model [9] as follows:

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

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.

Fig.2 is the model of sea area. Fig.2(a) is a common sea area which all ships share and 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. 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. Each grid is numbered for 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 unnavigable ones. Therefore, we judge that MARLS has obtained a solution if all the ships arrive at their goal grids without entering the unnavigable grid in their individual sea area and there is no collision between ships in the common sea area.

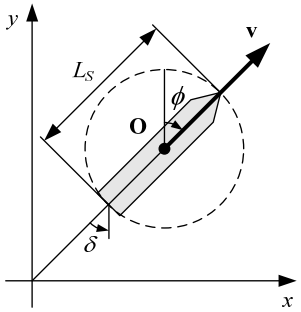
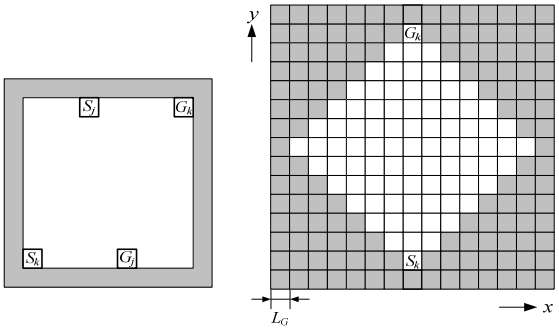


Fig.1 Model of ship maneuvering motion.



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

Fig.2 Model of sea area.

2.2. Basic Structure of MARLS

We show the basis of our basic MARLS which uses QL. There are some assumptions to solve multi-ship course problem by MARLS. A navigator is regarded as an agent. The perceptual input of agent k consists of the own ship's information $\mathbf{I}_k=(x_k,\ y_k,\ \phi_k,\ \dot{\phi}_k\ ,\ V_k)$ and other ship's 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. For example, Fig.3 shows that there are 2 ships j_2 and j_3 in the view circle of the ship k. Moreover, the circle is divided N_{Dk}=4 regions (i.e., $D_k[0]$~$D_k[3]$). Since the ship k needs to avoid only the ship j_2 in the region $D_k[0]$ according to NRs, the agent k generates $\mathbf{D}_k$=[1, 0, 0, 0]. Therefore, the number of stats (N_{Sk}) is given by

$$N_{Sk} = N(x_k)N(y_k)N(\phi_k)N(\dot{\phi}_k)N(V_k)\times 2^{N_{Dk}}, \quad (2)$$

where $N(I)$ is the number of division of each element of $\mathbf{I}_k$.

The action is defined by the rudder angle δ_k and the increment of speed Λ_k. Therefore, the number of actions (N_{Ak}) is given as follows,

$$N_{Ak} = N(\delta_k)N(\Lambda_k). \quad (3)$$

$N(\delta_k)$ is the number of δ_k and $N(\Lambda_k)$ is the number of Λ_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 other ships, the agent k receives r_F. Therefore, the agent k optimizes Q-table with the size of $N_{Sk}\times N_{Ak}$ until the end condition is satisfied. The end condition of a learning trial is based on the task achievement ratio detailed in Sect.4. Also, the task achievement means that all the ships arrive at their goals in an episode.

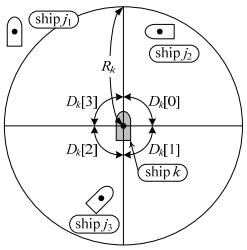
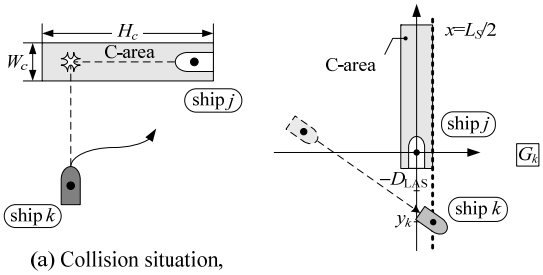
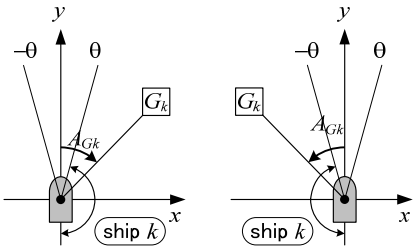


Fig.3 Generation of $\mathbf{D}_k$.



(a) Collision situation, NR, C-area. (b) LAS based on NR.

Fig.4 Crossing situation and its LAS based on NR.



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

Fig.5 LAS based on GO.

2.3. Limited Action Selection Based on NRs and GO

2.3.1. LAS Based on Navigation Rules (*NRs*)

Fig.4(a) illustrates an expample of collision situation with the collision area (C-area) and NR. It shows Crossing situation and the ship which has the other ship on the right side must change the course to the right. 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 the C-area around the ship j, then only the ship k receives a penalty (i.e., negative reward r_F).

Our basic MARLS limits the action selection in the execution of QL to keep NRs strongly. We explain the limited action selection based on NRs (LAS_{NR}). If observing Fig.4(a) carefully, we can see that the avoiding ships 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, the agent k judges if $\mathrm{LAS_{NR}}$ is available as follows. As shown in Fig.4(b), the agent k assumes that the ship k goes straight to the line $x=L_S/2$ and predicts the position $(L_S/2, y_k)$ when the ship k arrives at the line. If y_k is smaller than $-D_{\mathrm{LAS}}$, $\mathrm{LAS_{NR}}$ is inactivated temporarily. However, it does not mean that the agent is completely released from $\mathrm{LAS_{NR}}$. Also, D_{LAS} is defined as follows,

$$D_{\mathrm{LAS}} = \frac{\ln(\mu d_{kj} / R_k + 1)}{\ln(\mu + 1)}(D_{\max} - D_{\min}) + D_{\min}, \quad (4)$$

where, d_{kj} is the distance between ships k and j, R_k is the radius of view circle of the ship k, and μ is a positive constant value. $D_{\min}$ and $D_{\max}$ are minimum and maximum values of $\mathrm{D_{LAS}}$, respectively.

2.3.2. LAS Based on 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 (A_{Gk}) and there is no danger of collisions. We explain the limited action selection based on GO ($\mathrm{LAS_{GO}}$). Fig.5 shows the criteria. These are applied to the ship which has no need to avoid other ships. If $\theta < A_{Gk} \le \pi$ as shown in Fig.5(a), the action selection is limited so that $\delta_k \ge 0$ (i.e., turn to the right). If $-\pi < A_{Gk} < -\theta$ as shown in Fig.5(b), the action selection is limited so that $\delta_k \le 0$ (i.e., turn to the left).

3. Introduction of Reference Courses to MARLS

In the following sections, we explain the main design idea to introduce the reference courses to basic MARLS. The idea consists of the limited action selection to trace the reference courses ($\mathrm{LAS_{RC}}$), the synthesis of three LASs, and updating reference courses.

3.1. LAS Based on Reference Courses

It is assumed that $\mathbf{P}_k(t)$ is the position in time t (= 0, h, $2h, \cdots$) on the reference course of the ship k, where h is the time step.

To trace the reference course, the agent k must select not only the rudder angle δ_k but also the increment of speed Λ_k appropriately. Also, if the ship's position $\mathbf{O}_k(t)$ is far from $\mathbf{P}_k(t)$, the agent k must stop tracing the reference course. To satisfy these requests, we design $\mathrm{LAS_{RC}}$ as follows.

First, we explain the limitation of δ_k. This limitation is same as $\mathrm{LAS_{GO}}$ by considering $\mathbf{P}_k(t + N_{\mathrm{TL}}h)$ as the goal, where N_{TL} is a positive integer. Therefore, if the goal direction ($A_{\mathbf{P}k}$) satisfies $\theta_{\mathrm{RC}} < A_{\mathbf{P}k} \le \pi$, the selection is limited so that $\delta_k \ge 0$. Similarly, if $-\pi < A_{\mathbf{P}k} < -\theta_{\mathrm{RC}}$, the selection is limited so that $\delta_k \le 0$.

Next, we explain the limitation of Λ_k. The speed of the ship k in time t is given by $V_k(t)$ and the speed of the reference point (i.e., $\mathbf{P}_k(t)$) is given by

$$V_{\mathbf{P}k}(t) \equiv \frac{\mathbf{P}(t+h) - \mathbf{P}(t)}{h}. \quad (5)$$

When $\mathbf{O}_k(t)$ is close to $\mathbf{P}_k(t)$, or the ship k can trace the reference point, it should be satisfied that $V_k(t) \approx V_{\mathbf{P}k}(t)$. To satisfy it, the selection of Λ_k is limited as follows:

$$\begin{cases} \Lambda_k \ge 0, & \text{if } V_k(t) < V_{\mathbf{P}k}(t) - \Delta \\ \Lambda_k \le 0, & \text{if } V_k(t) \ge V_{\mathbf{P}k}(t) + \Delta, \\ \Lambda_k \text{ is unlimited}, & \text{otherwise} \end{cases} \quad (6)$$

where Δ is a positive constant value.

Moreover, the condition to apply $\mathrm{LAS_{RC}}$ is as follows. At the start of each episode (i.e., t=0), $\mathrm{LAS_{RC}}$ is inevitably activated because $\mathbf{O}_k(0)$ is identical with $\mathbf{P}_k(0)$. On the other hand, when the distance between $\mathbf{O}_k(t)$ and $\mathbf{P}_k(t)$ is larger than the threshold value D_{C}, $\mathrm{LAS_{RC}}$ for the agent k is inactivated until the end of present episode.

3.2. Synthesis of $\mathrm{LAS_{NR}}$, $\mathrm{LAS_{GO}}$, and $\mathrm{LAS_{RC}}$

Although $\mathrm{LAS_{NR}}$ and $\mathrm{LAS_{GO}}$ do not compete each other, $\mathrm{LAS_{RC}}$ might compete with them. Considering the efficiency and safety of ships' courses, $\mathrm{LAS_{NR}}$ and $\mathrm{LAS_{GO}}$ must have priority over $\mathrm{LAS_{RC}}$. Therefore, $\mathrm{LAS_{RC}}$ is executed when one of the following conditions is satisfied.

- $\mathrm{LAS_{NR}}$ is activated. Also, $\mathrm{LAS_{RC}}$ and $\mathrm{LAS_{NR}}$ demand the same limitation of δ_k.
- $\mathrm{LAS_{NR}}$ is inactivated temporarily as mentioned in 2.3.1. Also, the limitation of δ_k by $\mathrm{LAS_{RC}}$ is $\delta_k \le 0$.
- $\mathrm{LAS_{GO}}$ is activated. Also, $\mathrm{LAS_{RC}}$ and $\mathrm{LAS_{GO}}$ demand the same limitation of δ_k.
- $\mathrm{LAS_{GO}}$ is activated. Also, both $\mathrm{LAS_{RC}}$ and $\mathrm{LAS_{GO}}$ do not limit δ_k.
- Both $\mathrm{LAS_{NR}}$ and $\mathrm{LAS_{GO}}$ are inactivated.

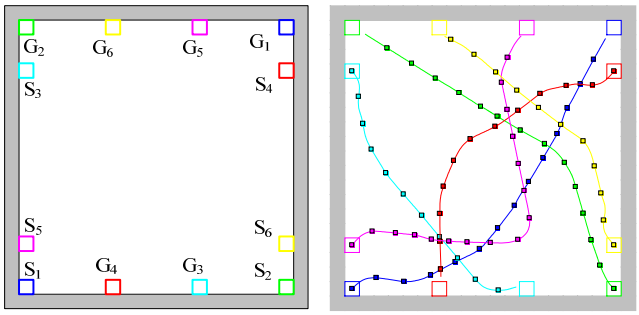
3.3. Updating Reference Courses

It can be expected that the appropriate use of reference courses improves the leaning efficiency and the quality of obtained courses. Moreover, updating the reference courses may generate better performances. Our proposed MARLS updates them when not only the length of courses but also the number of steps of QL are improved.

4. Numerical Experiments

Experiments have been carried out to investigate the influence of reference courses on MARLS by comparing our proposed MARLS and basic MARLS. Fig.6(a) is the test problem including 6 same ships in $42L_S \times 42L_S$ sea area. Fig6(b) is the reference courses obtained by our basic MARLS[4] with the temporal cancellation of $\mathrm{LAS_{NR}}$ using Eq.(4). The total length of the courses is 30041(m). There are squared marks on each course. The marks are drawn from the start to the coal every 60 seconds. Also, Fig6(b) shows that the course of 5-th ship is ineffective.

The main parameters are as follows. The parameters of ships are L_S=107(m), $V(0)$=12(knots), K_0=1.310, T_0=1.085,



(a) Test problem (b) Reference courses

Fig.6 Test problem and reference courses.

$\delta \in \{0, 10, -10, 20, -20\}$(deg.), $\Lambda \in \{0, 0.1, -0.1\}$(knot). The initial heading angle is given by adding the random value within [−2.5, 2.5](deg.) to the goal direction with probability ρ=0.8. QL uses ε-greedy policy with ε=5.0×10^{-3}. The state variables are divided as follows: x and y are divided by L_G(=2L_S), $\phi \in [0, 2\pi]$ is divided into 12 equal parts, $\dot{\phi}$ is divided into 2 equal parts based on its sign, and N_D=4. The speed $V \in [0.5V_S, 1.3V_S]$ is divided into 4 equal parts, where V_S=12(knots) is the standard speed. The parameters of LASs are as follows: $\mathrm{LAS_{NR}}$ uses R=40L_S, μ=10, $D_{\min}$=4L_S, and $D_{\max}$=20L_S in Eq.(4); $\mathrm{LAS_{GO}}$ uses θ=1.0(deg.); $\mathrm{LAS_{RC}}$ uses θ_R=1.0(deg.), Δ=0.1, and D_C=30V_S. The numerical analysis has been done by fourth order Runge-Kutta. The time step is h=1.0(sec.). QL is executed every 4h(sec.). Also, there are other parameters which are same as ones in Ref.[4]. The maximum number of episodes in each learning trial is 150000. The end condition is as follows: a learning trial is successful if the task achievement ratio is over 80% for 20000 successive episodes. The ratio is calculated using recent 5000 episodes. The number of learning trials is 30. If a learning trial is successful, MARLS calculates a set of courses without any randomness. We call it the obtained course.

Table 1 shows comparison results in terms of learning efficiency. 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 in successful trials. N_{GET} is the number of obtained courses without collisions. Table 2 shows comparison results in terms of course efficiency in 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.

From these results, we can find following. Table 1 shows that our proposed MARLS possesses much better learning efficiency than our basic MARLS. However, our proposed MARLS cannot always get the courses without collisions (i.e., $N_{\mathrm{SLT}} > N_{\mathrm{GET}}$). Tables 2 shows that our proposed MARLS can get shorter courses than our basic MARLS. Also, our proposed MARLS can improve the reference courses with larger avoidance than necessary.

Table1 Comparison results in terms of leaning efficiency.

	N_{SLT}	N_{EPS}	N_S	N_{GET}
Proposed	30	31349	13788	29
Basic	28	99486	49892	28

Table2 Comparison results in terms of course efficiency.

	L_{ave}(m)	$L_{\min}$(m)	$L_{\max}$(m)
Proposed	28756	28012	29666
Basic	29559	28667	31549

5. Conclusions

We have proposed a novel MARLS to search ships' courses based on the reference courses. From numerical experiments, we have found that our proposed MARLS can improve the reference courses with larger avoidance than necessary and can also make learning time smaller than our basic MARLS. In the future, we will consider how the reference courses should be traced accurately.

Acknowledgments

This work was supported by JSPS KAKENHI Grant Number JP16K00309.

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, T. Kamio, and T. Tanaka, "Finding the course and collision avoidance based on reinforcement learning algorithm," NAVIGATION, 170, pp.26-31, Sep. 2009.

[3] T. Kamio, S. Sugeo, K. Mitsubori, T. Tanaka, C. J. Ahn, H. Fujisaka, and K. Haeiwa, "A reinforcement learning approach to course decision of ships under navigation rules," Proc. of NOLTA, pp.141-144, Oct. 2009.

[4] T. Kamio, K. Mitsubori, T. Tanaka, H. Fujisaka, K. Haeiwa, "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.

[5] T. Tanigawa, T. Kamio, K. Mitsubori, T. Tanaka, H. Fujisaka, and K. Haeiwa, "Modified Multi-Agent Reinforcement Learning System to Find Ships' Courses," Proc. of NOLTA, pp.487-490, Sep. 2013.

[6] M. Nakayama, T. Kamio, K. Mitsubori, T. Tanaka and H. Fujisaka, "Reinforcement Learning Based Search for Ships' Courses Controlled by Safety," Proc. of NOLTA, pp.28-31, Sept. 2014.

[7] M. Nakayama, T. Kamio, K. Mitsubori, T. Tanaka and H. Fujisaka, "Multi-Agent Reinforcement Learning System to Find Efficient Courses for Ships," Proc. of IEEE IWCIA, pp.89-94, Nov. 2014.

[8] International Maritime Organization, "International regulations for preventing collisions at sea," 1972.