IEICE Proceeding Series

Modified Multi-Agent Reinforcement Learning System to Find Ships' Courses

Tomohiro Tanigawa, Takeshi Kamio, Kunihiko Mitsubori, Takahiro Tanaka, Hisato Fujisaka, Kazuhisa Haeiwa

Vol. 2 pp. 487-490
Publication Date: 2014/03/18
Online ISSN: 2188-5079

Modified Multi-Agent Reinforcement Learning System to Find Ships' Courses

Tomohiro Tanigawa†, Takeshi Kamio†, Kunihiko Mitsubori††, Takahiro Tanaka†††,
Hisato Fujisaka†, and Kazuhisa Haeiwa†

† 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. However, our previous MARLS cannot search the courses, considering both efficiency and safety.

In this paper, we modify our MARLS to overcome this problem. From numerical experiments, we have confirmed that our modified MARLS can get more efficient courses than our previous MARLS, although both of them have similar performance with regard to safety.

1. Introduction

Deciding efficient and safe courses of ships before actual navigation is important. The importance deeply relates to the value of ship transportation and the special conditions in ship maneuvering. 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. 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.

From these backgrounds, we have developed several MARLSs to find ships' courses [1]-[4]. Especially, our previous MARLS proposed in Ref. [4] uses the navigation rule (NR) and goal orientation (GO) as prior knowledge. NR is the knowledge to avoid 3 typical collisions between 2 ships, which is given by international regulations [5]. GO is the common rule that a ship should move to the goal or return to the course line if there is no danger of collisions. Since they are implemented by limiting action selection, our MARLS can easily get the courses which satisfy NR. However, our MARLS cannot search the courses, considering both efficiency and safety.

References [6] and [7] are the studies on ships' behaviors using multi-agent system (MAS). However, since the purpose is to represent ships' behaviors as a swarm, it is difficult to consider both safety and efficiency for multi-ship course problem by the system. Unfortunately, we have not found researches similar to ours except them.

In this paper, we modify our MARLS to search the courses, considering both efficiency and safety. In numerical experiments, we compare our modified MARLS with our previous MARLS. As the result, it has been confirmed that our modified MARLS can get more efficient courses than our previous MARLS, although both of them have similar performance with regard to the safety.

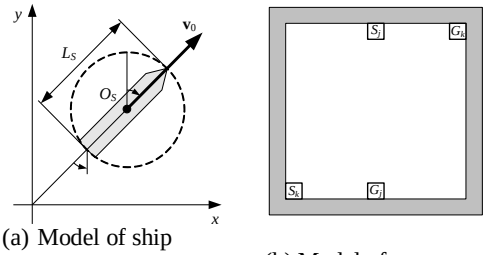
2. Previous MARLS to Find Ships' Courses

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, we can consider the tidal current effects [1]. 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 [8] as follows:

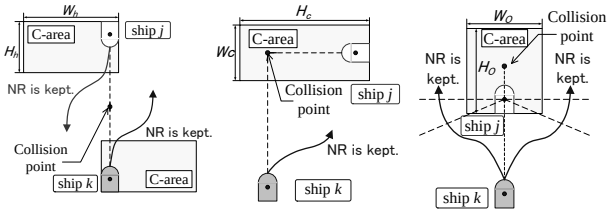
$$T\ddot{\phi}+\dot{\phi}=K\delta,\ \ \dot{x}=V_0\sin\phi,\ \ \dot{y}=V_0\cos\phi \qquad (1)$$

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.



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

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



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

Fig.2 Collision situation, NR, and C-area.

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 and 3) the obtained courses keep NR mentioned in Sect. 2.2.

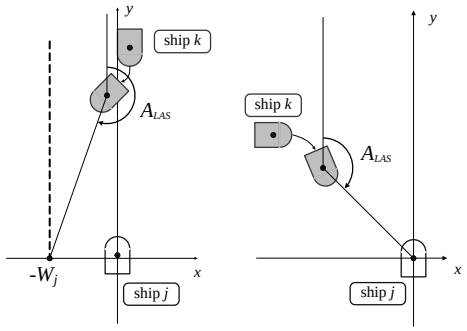
2.2. Collision Area Based on Navigation Rule

Fig.2 illustrates the collision situation, NR, and collision area (C-area). Fig.2(a) shows Head-on-situation and each ship must change the course to the right to avoid the collision. Fig.2(b) shows Crossing-situation and the ship which has the other ship on the right side must change the course to the right. Fig.2(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 NR, C-area is placed around the ship *j*. The shape of C-area depends on the corresponding collision situation. If the ship *k* enters C-area around the ship *j*, then only the ship *k* receives a penalty (i.e., negative reward). However, we have confirmed that C-area is not enough to keep NR [4].

2.3. Process Flow of Previous MARLS

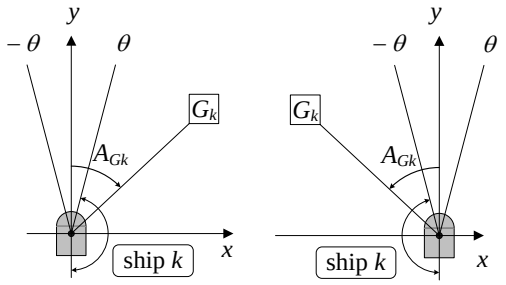
Here, we review our previous MARLS proposed in Ref. [4]. Our MARLS is based on Q-learning and uses NR and GO as prior knowledge. Since they are implemented by limiting action selection, our MARLS can easily get the courses which satisfy NR. Moreover, since the limited action selection (LAS) based on NR 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 with 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 NR.
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}.



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

Fig.3 Limited action selection based on NR.



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

Fig.4 Limited action selection based on GO.

2.4. Limited Action Selection

Our previous MARLS limits the action selection in the execution of Q-learning to keep NR forcibly and improve the possibility that ships arrive at their goals. First, we explain LAS based on NR. If observing Fig.2 carefully, one can see that the avoiding ships in Head-on-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, this LAS is not available if $\phi_k \geq A_{LAS}$ as shown in Fig.3. Next, we show LAS based on GO. Fig.4 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.4(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.4(b), the action selection is limited so that $\delta_k \leq 0$ (i.e., turn to the left).

3. Modified MARLS to Find Ships' Courses

3.1. Drawbacks of Previous MARLS

Our previous MARLS cannot search the courses, considering both efficiency and safety. Through a lot of numerical experiments, we have confirmed that our MARLS often gets inefficient courses and rarely gets the courses including near miss. Therefore, in our MARLS,

the drawback of efficiency is more serious than that of safety.

Here, we show the cause that our MARLS gets inefficient courses. As mentioned above, our MARLS limits the action selection in the execution of Q-learning to keep NR forcibly. In other words, our MARLS puts avoiding collisions between ships before shortening the courses. Even if lengthened courses are obtained temporally, the ideal Q-learning can redress the inefficiency. However, our MARLS uses the successful learning condition which helps to obtain the courses in limited time. As a result, inefficient courses are often obtained.

3.2. MARLS with Modified LAS based on NR

To solve the drawback of our previous MARLS, we have to improve the exploration of Q-learning so that the efficient courses are obtained in limited time.

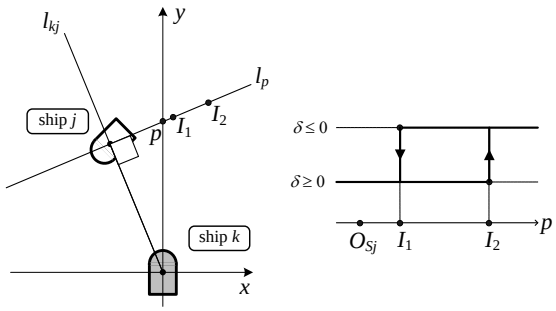
We modify LAS based on NR to achieve the above demand. As shown in Fig.5(a), our modified LAS has two criteria I_1 and I_2 for the avoiding ship. If the ship k must avoid the ship j, they are on the line (l_p) which passes through O_{Sj} (i.e., the position of the ship j) and is perpendicular to the line (l_{kj}) which connects O_{Sj} and O_{Sk} (i.e., the position of ship k). In the case of Head-on-situation, I_1 and I_2 are put on the left side of the ship j. In the case of Crossing-situation, they are put on the rear side of the ship j. The point p is the crossing one between l_p and the line which passes through O_{Sk} and is parallel to the heading angle direction of the ship k. Fig.5(b) shows the method to switch the direction of LAS by the relationship between the point p and two criteria (i.e., I_1 and I_2). If $p<I_1$, $\delta_k \geq 0$ is set to LAS. This helps the ship k to go away from the ship j. If $p>I_2$, $\delta_k \leq 0$ is set to LAS. This helps the ship k to approach the ship j. If $I_1 \leq p \leq I_2$, $\delta_k \geq 0$ or $\delta_k \leq 0$ is set to LAS according to the hysteresis characteristics. I_1 is the criterion to avoid the collision with the other ship according to NR and I_2 is the criterion to shorten the inefficient course.

However, if our modified LAS is applied to the situation shown in Fig.6, the efficiency of the course becomes worse. In this case, even if the ship k moves to the goal, it can necessarily avoid the collision with the ship j. Therefore, if $A_{jGk}>A_{jI1}$, our new MARLS uses LAS based on GO shown in Fig.4 instead of the modified LAS shown in Fig.5.

Therefore, our modified MARLS search the courses, considering both efficiency and safety.

4. Numerical Experiments

Experiments have been carried out to investigate the performance of our modified MARLS. Fig.7 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 initial heading angle is equal to the goal direction plus random value within [−2.5, 2.5] (deg.). The parameters of Q-learning are α=0.9, γ=0.99, r_A=1.0, r_F=−1.0, ε=10^{-3}. The state variables are divided as follows: x_k and y_k are divided by $2L_S$, $\phi \in [0, 2\pi]$ is divided into 12 equal parts, $\dot{\phi}$ is divided into 2 equal parts based on its sign. The parameters of C-area are H_h=$2L_S$, W_h=$5L_S$, H_c=$8L_S$, W_c=L_S, H_o=$5L_S$, W_o=$2L_S$. The parameters to perceive other ships are R=$40L_S$, D_h=D_o=$10L_S$, N_D=4. The parameters of LAS are as follow: W_J=L_S, θ=1.0(deg.), I_1=$4L_S$, I_2=$8L_S$. Most of these parameters are defined in Ref.[4]. The numerical analysis has been done by fourth order Runge-Kutta. The time step is Δt=1.0(sec.). The maximum number of episodes in each learning trial is 300000. The learning in a trial is successful if the task achievement ratio is over 80% for 20000 successive episodes. Also, if the learning is successful, we have estimated the course of ship whose initial heading angle is the goal direction. The number of trials is 30.



(a) Definition of criteria to control LAS. (b) Method to switch the direction of LAS.

Fig.5 Modified LAS.

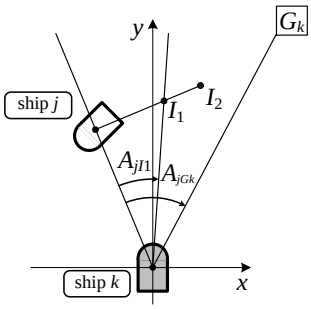


Fig.6 Cancel condition of modified LAS.

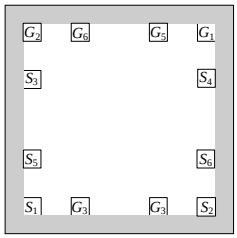
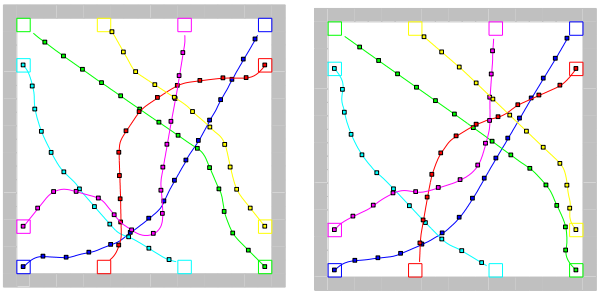


Fig.7 Test problem (6 ships in $42L_S \times 42L_S$ sea area).



(a) Previous MARLS (b) Modified MARLS

Fig.8 Examples of courses obtained by our previous MARLS and our modified MARLS.

Table 1 Comparison results between our previous MARLS and our modified MARLS.

	N_{SLT}	N_{EPS_AVE}	N_{EPS_MIN}	N_S	D_{AVE} [m]	D_{MIN} [m]
Previous MARLS	29	44070	33236	22744	29280	28560
Modified MARLS	30	30003	25167	6588	28507	28227

Fig.8(a) shows a example of inefficient courses which are sometimes obtained by our previous MARLS. On the other hand, Fig.8(b) shows a typical example of courses obtained by our modified MARLS. Also, our modified MARLS has never obtained inefficient courses which are similar to Fig.8(a). Table 1 shows comparison results between our previous MARLS and our modified MARLS in terms of N_{SLT}, N_{EPS_AVE}, N_{EPS_MIN}, N_S, D_{AVE}, and D_{MIN}. N_{SLT} is the number of successful learning trials. N_{EPS_AVE} is the average number of episodes executed in successful trials. N_{EPS_MIN} is the minimum number of episodes executed in successful trials. N_S is the average number of states used in successful learning trials. D_{AVE} is the average distance of courses which keep NR. D_{MIN} is the minimum distance of courses which keep NR. From these results, we have confirmed that our modified MARLS is superior to our previous MARLS in terms of not only the course efficiency but also the learning efficiency. In addition, we comment on the safety. According to Ref.[9], if the distance between ships is smaller than $4L_S$, actual navigators feel this situation to be dangerous. Therefore, we have investigated the distances between ships in all successful trials. As a result, we have confirmed that the frequencies of such dangerous situations are almost same in both our previous MARLS and our modified MARLS. In other words, both of them have similar performance with regard to the safety. However, they have not always satisfied the safety which Ref.[9] demands.

5. Conclusions

Our previous MARLS [4] cannot search the courses, considering both efficiency and safety. To overcome this problem, we have modified the limited action selection based on NR. From numerical experiments, we have confirmed that our modified MARLS can get more efficient courses than our previous MARLS, although both of them have similar performance with regard to the safety. However, they have not always satisfied the safety which Ref.[9] demands. In the future, we have to improve the safety of courses obtained by our MARLS in order to quantify both efficiency and safety.

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, T. Kamio, and T. Tanaka, “Finding the course and collision avoidance based on reinforcement learning algorithm,” NAVIGATION, 170, pp.26-31, 2009 (in Japanese).

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

[6] M. Inaishi, H. Kondo, A. Kawaguchi “Multiple Ship Clusters Behavior Simulation” Journal of Japan Institute of Navigation, 110, pp.1-7, 2004.

[7] M. Inaishi, H. Kondo, M. Kondo, A. Kawaguchi “Marine Traffic Simulation Using Ship Agent Clusters : Northbound Traffic in Tokyo Bay” Journal of Japan Institute of Navigation, 115, pp.11-16, 2006.

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

[9] M. Fuchi, S. Usui, S. Fujimoto, K. Hirono and T. Mochida, “Differences in allowable distance between ships due to ship’s specific type,” Japan Ergonomics Society, vol.46, pp.144-145, 2010 (in Japanese).