Cooperative Multi-Knowledge Learning Control System for Obstacle Consideration

Syafiq Fauzi Kamarulzaman[1] and Seiji Yasunobu[2]

[12]Dept. of Intelligent Interaction Technologies, University of Tsukuba, Japan
[1]Faculty of Computer Systems and Software Engineering, Universiti Malaysia Pahang, Malaysia
syafiq@fz.iit.tsukuba.ac.jp
yasunobu@iit.tsukuba.ac.jp

Abstract. A safe and reliable control operation can be difficult due to limitations in operator's skills. A self-developing control system could help assist or even replaces the operators in providing the required control operations. However, the self-developing control system is lack of flexibility in determining the necessary control option in multiple conditions where a human operator usually prevails by experiences in optimizing priority. Here, a cooperative multi-knowledge learning control system is proposed in providing flexibility for determining priority in control options, within multiple conditions by considering the required self-developing control knowledge in fulfilling these conditions. The results show that the system was able to provide consideration in prioritizing the use of the required control knowledge of the condition assigned.

Keywords: Multi-knowledge, Learning Control, Reinforcement Learning

1 Introduction

Human operators are prone to be inefficient during any control operation due to the lack of skills in unfamiliar operation's environment and parameters. Learning Control System provides a self-developing Control Knowledge that changes according to interaction with unfamiliar environment, thus reduces the possible risk related to skills inefficiency [2]. The Control Knowledge will be continuously updated through the control operations that later provides instructions for controlling the related machine to perform at a high efficient manner. The Learning Control System reduces the dependency on human command since the Control Knowledge provides most of the required control instruction, learned from successes in previous operations [1].

However applying Learning Control operation on various conditions of state parameters is difficult and slow since the Control Knowledge is needed to be constructed for each condition. Here, issues concerning application of multiple knowledge in learning are brought to establish a way of learning an optimum action from multiple sources of individual Control Knowledge [7][9]. This study

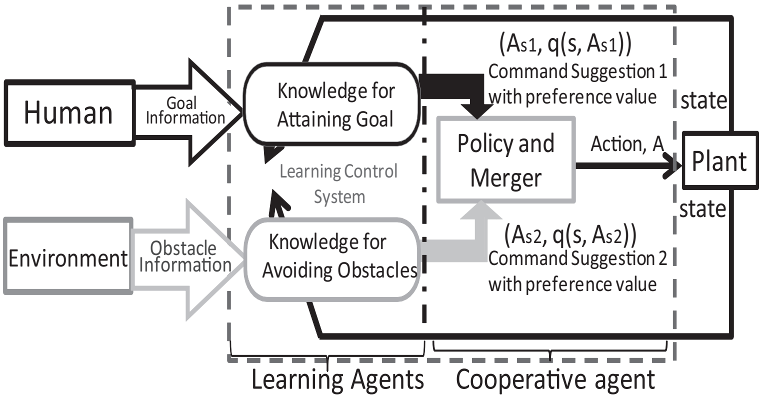


Fig. 1: The structure of cooperative multi-knowledge learning system

is driven by the need to combine multiple Control Knowledge into a single control output while putting the necessity of individual Control Knowledge into consideration. In order to combine multiple knowledge protocol, a mechanism that involves cooperation of multiple Learning Control System for producing an optimum action is presented.

In this research, the cooperation mechanism is developed by evaluating the output preference value from two sets of Control Knowledge with specific input parameters as shown in Fig. 1. Focusing on two different types of states; termed as Goal Distance and Obstacle Distance, separated learning process was conducted to obtain a Cooperative Knowledge that operates to satisfy both state condition. The first set consists of a Control Knowledge that controls the operation upon achieving a goal while the second set consists of a Control Knowledge that controls the operation upon avoiding detected obstacles. Each Control Knowledge provides the preference value of each output depending on current state. Later, a policy and merger agent evaluates the preference value from both knowledge and produce a control output.

Applying both Control Knowledge on the system provides an efficient, safe and successful operation. The success is achieved, due to the application of Cooperative Knowledge for learning and establishing a preference between two Control Knowledge specializing in two different state parameters at the same time. Learning and application of multiple Control Knowledge in a control system are simplified using the proposed method. Simulations were conducted to evaluate the proposed control technique. The results prove that the system was able to cooperate the learning process between the two Control Knowledge that results in a successful control operation.

2 Cooperative Multi-Knowledge Learning Control

Multi-knowledge Learning Control is defined by having a control system that learns multiple Control Knowledge. The knowledge will then be used to perform a certain control operation. These Control Knowledge are specified, each for different task with different concerned parameters. Usually, Control Knowledge is applied by control operators consecutively and manually depending on the requirement and conditions. However, the Cooperative Multi-knowledge Learning Control proposed in this research applies cooperation between multiple Control Knowledge that is developed using preference values stored in each Control Knowledge. The preference values represents the importance of each Control Knowledge in a given situation. These preference values are obtained from Learning Control, where the Control Knowledge is constructed in a form of value function through trial and error.

As an example, the Cooperative Multi-Knowledge Learning Control will be able to configure around an obstacle and achieved the desired goal by using two types of Control Knowledge; Control Knowledge for goal attainment and Control Knowledge for obstacle avoidance. Using "Achieving Goal" as main command, the system uses "Avoiding Obstacle" for creating a cooperative command, heading to the goal while avoiding closing obstacles. Here, Learning Control is applied twice in the protocol. Firstly, Learning Control is applied to construct two specific Control Knowledge. Then, the knowledge gained from the first iteration is combined and relearn for cooperative control operation.

2.1 Learning Control

Learning Control is a method of obtaining Control Knowledge by repeatedly construct and correct the Control Knowledge depending on the outcome of a control operation in several trials [1]. In this research, this method applies reinforcement learning where Control Knowledge is constructed in a form of value functions Q. The value will then be updated depending on the reward r that is received after performing a control operation trial [3].

The Control Knowledge is divided into state S, which defines the current situation of the control object and action A, which defines the next move of the control object. State S and Action A is divided into a set of numbers as State $S = s_1, s_2, , s_n$ and Action $A = a_1, a_2, , a_n$. During the update phase, the preference value q of the combination between state s and action a is defined through the reward obtained after performing action a. In the case successful operation, the preference value q increases, should the action contribute to a fail operation, the value decreases. A set of preference value q can be defined as value function Q, where all preference value for the combination between state S and action A is recorded. The value function Q is defined as Control Knowledge.

In this research, this Control Knowledge is updated based on Q-learning algorithm,

$$Q(S, A) = (1 - \alpha)Q(S, A) + \alpha[r + \gamma Q_{max}], \tag{1}$$

$$Q_{max} = \max_{A} Q(S, A) \tag{2}$$

where α is the learning rate and γ is update value discount rate [3]. The algorithm is applied in updating all the Control Knowledge constructed.

Two Control Knowledge were constructed to confirm the effectiveness of the system through a simulation. The first Control Knowledge is for Goal Attainment Control, while another is for Obstacle Avoidance Control. For each Control Knowledge, various parameters for state S were used, without imposing any changes to action A.

Goal Attainment Control Goal Attainment Control consists Learning Control method for operating the control object towards the goal [5]. Here, the Goal Attainment Control applies goal distance $\Delta G = \{\Delta X_{goal}, \Delta Y_{goal}\}$ as state S while control output u and rotation θ as action A_{GA}. Therefore, the value function Q for Goal Attainment Control can be defined as $Q(\Delta G, A_{GA})$.

The update equation for Goal Attainment Control alone is,

$$Q_{Goal}(\Delta G, A_{GA}) = (1-\alpha)Q_{Goal}(\Delta G, A_{GA}) + \alpha[r + \gamma Q_{max}], \tag{3}$$

$$Q_{max} = \max_{A_{GA}} Q_{Goal}(\Delta G, A_{GA}) \tag{4}$$

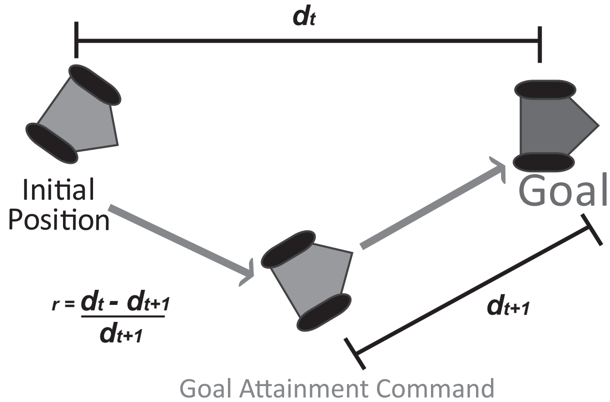


Fig. 2: Method for applying reward r to the Control Knowledge for goal attainment.

Obstacle Avoidance Control Obstacle Avoidance Control consists Learning Control method for operating the control object away from obstacles. Here, the Obstacle Avoidance Control applies obstacle distance $\Delta O = \{\Delta X_{obs}, \Delta Y_{obs}\}$ as state S while control output u and rotation θ as action A_{OA}. Therefore, the value function Q for Obstacle Avoidance Control is defined as $Q(\Delta O, A_{OA})$.

The update equation for Obstacle Avoidance Control alone is,

$$Q_{obs}(\Delta O, A_{OA}) = (1-\alpha)Q_{obs}(\Delta O, A_{OA}) + \alpha[r + \gamma Q_{max}], \tag{5}$$

$$Q_{max} = \max_{A_{OA}} Q_{obs}(\Delta O, A_{OA}) \tag{6}$$

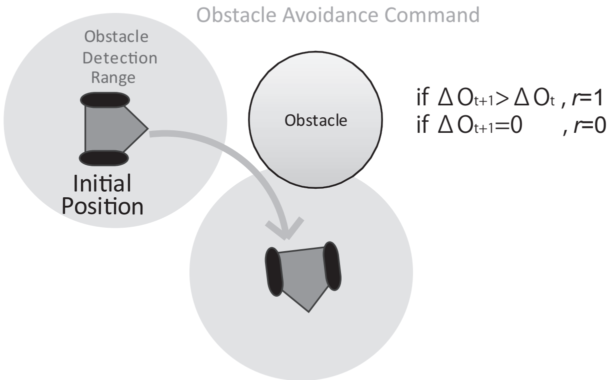


Fig. 3: Method for applying reward r to the Control Knowledge for obstacle avoidance.

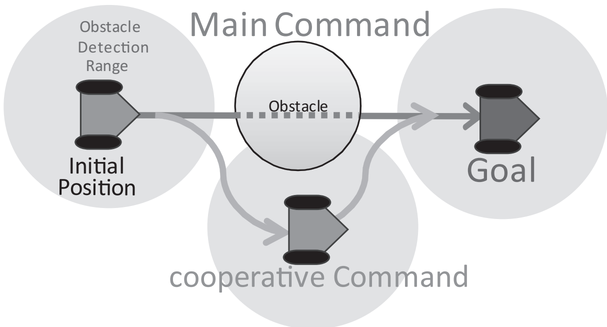


Fig. 4: The control method of an object using cooperative multi-knowledge learning control around obstacles

2.2 Cooperative Multi-Knowledge Learning

Applying Learning Control for multiple Control Knowledge requires the system to analyze the value functions of both knowledge prior to the execution of the control output. The preference value of outputs supplied from each control method at a moment of state are needed to determine which Control Knowledge is more preferred at the current state. Here, update method for both Control Knowledge is modified so that the preference value is limited between 0(bad) and 1(Good) [2]. Therefore, the updated value discount rate γ in the update equation for each knowledge is applied as,

$$\gamma_{Goal} = 1 - Q_{Goal}(\Delta G, A_{GA}) \tag{7}$$

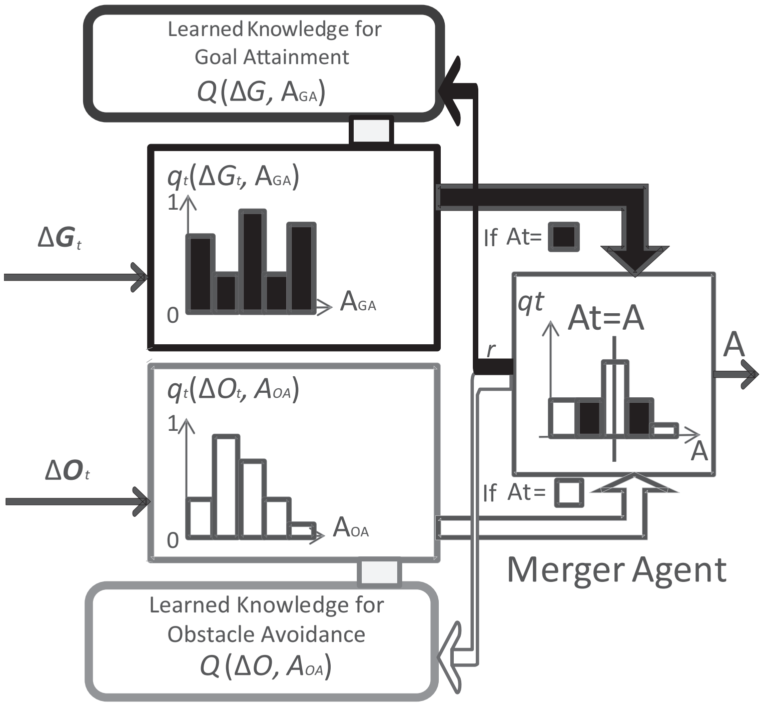


Fig. 5: The application and update process for all Control Knowledge

and

$$\gamma_{obs} = 1 - Q_{obs}(\Delta O, A_{OA}), \tag{8}$$

so that the preference value will be restricted between 0 and 1.

Fig. 5 described the method of cooperating both Control Knowledge in constructing an output. The preference values of a set of output A from both Control Knowledge are used to construct a new set of preference value for output A with the identity of the source knowledge attached; termed as Merger Output. Merger Output is constructed by selecting the minimum preference value among the two Control Knowledge for each element in the set of output A. The control output is determined from Merger Output by a *greedy* policy where the action a with the maximum preference value among the options in the set is chosen as the output at current state. The result after the output been operated will determined the rewards. Rewards will be given to the Control Knowledge based on the identity of the source knowledge in Merger output that supplies the preference value of the executed output.

The system structure involving cooperative multi-knowledge learning control is later applied in a series of simulation as validation of its effectiveness in producing an efficient, safe and reliable operation.

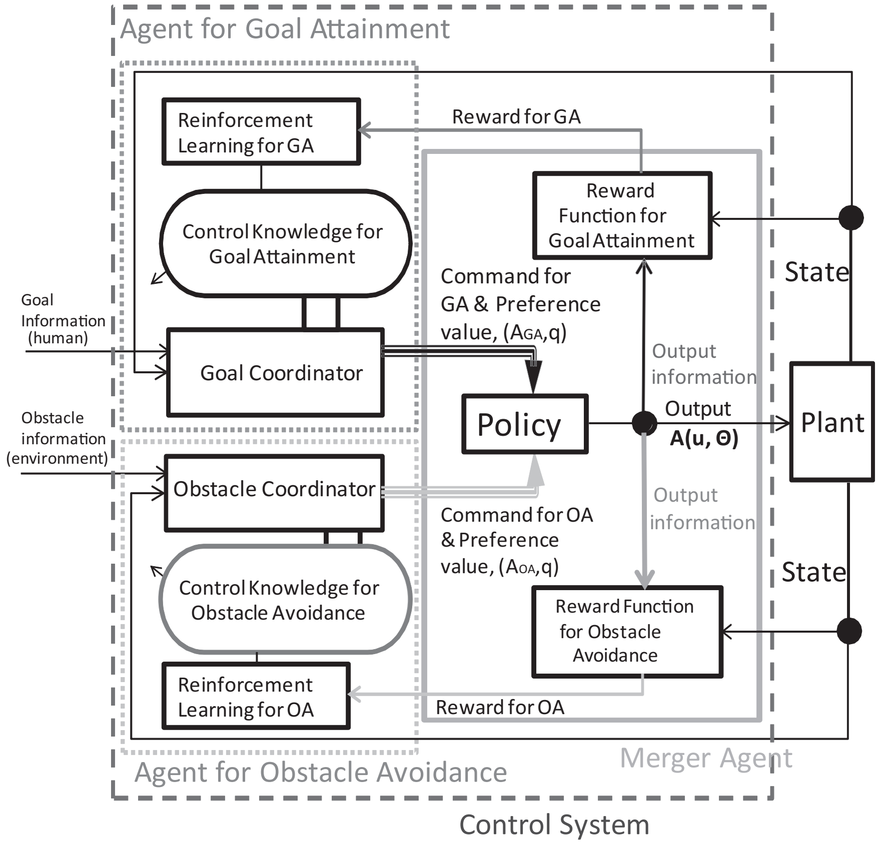


Fig. 6: The system structure for simulation experiment

3 Simulation Experiment

Simulation regarding the proposed method was constructed based on the system structure shown in Fig. 6. The simulation was conducted based on a small robot with the parameters shown in table 1 and Fig. 7b. The simulation was operated in Matlab Simulink and the result was based on operation with and without obstacles in a pre-constructed map.

The map was constructed as Fig. 7a, where 4 different goals were prepared before the simulation was done. The simulated operation results were divided into two sections for easy comparison between an operation with and without obstacles.

The simulation starts by constructing the Control Knowledge of goal attainment and obstacle avoidance separately using the Learning Control. Learning Control was done in 350 episodes in separate simulation. The simulation was continued with cooperative simulation where the constructed knowledge was integrated in the simulation. Here, the Cooperative Learning Control was done for 100 trials and the results were taken after.

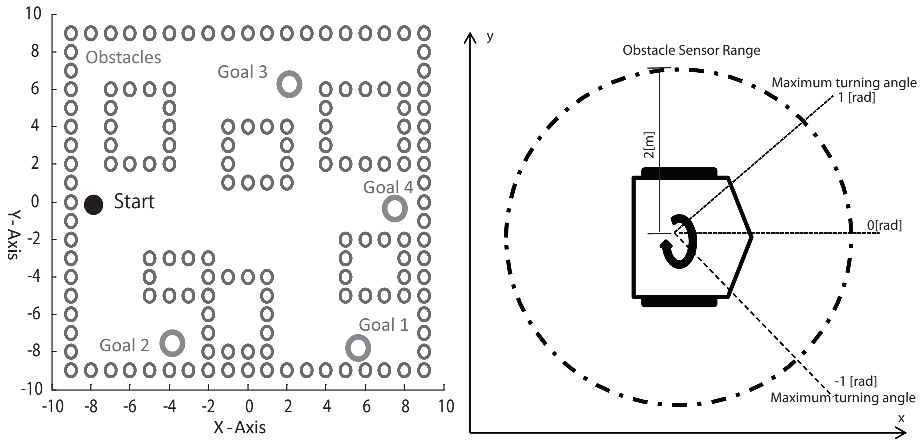
Table 1: Specifications of the simulated control object

Parameters	Value
Weight	0.42 [kg]
Size:	
Length	0.53 [m]
Width	0.52 [m]
Height	0.1 [m]
Turning Radius	$-1 < \theta < 1$ [rad]
Torque Force	$-10 < V < 10$ [volt]

Table 2: Q-learning parameters

	Parameters	Range	Intervals
State (Goal)	Goal Distance, ΔG[m]	$-10 < \Delta G(x, y) < 10$	2
State (Obstacle)	Obstacle Distance, ΔO[m]	$-2 < \Delta O(x, y) < 2$	0.5
Action	Target Angle, θ[rad]	$-1 < \theta < 1$	0.5
	Torque Force, V[volt]	$-10 < V < 10$	2

Learning rate, α	0.5	Discount rate, γ	0.3

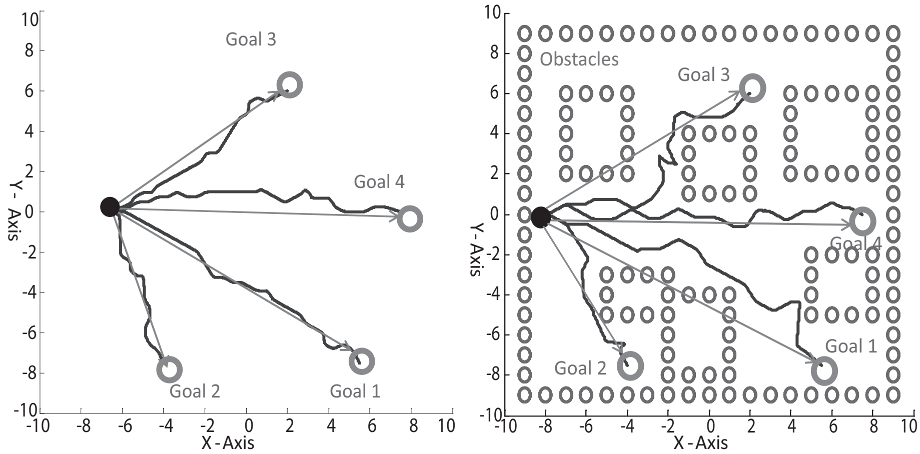


(a) Field map for simulation (b) Robot specifications

Fig. 7: Simulation setup for field and robot

4 Experiment Results

Here, the results for control operation without obstacles and control operation with obstacles are presented. Control operation without obstacles was done to



(a) without obstacle

(b) with obstacle

Fig. 8: Results of Control Operation Simulation

confirm the effectiveness of the Learning Control process in constructing the most effective Control Knowledge for the system. Control operation with obstacles was done to confirm the effectiveness of the whole system in utilizing the cooperative multi-knowledge learning control by using two Control Knowledge constructed prior to the simulation operation.

4.1 Control Operation Without Obstacles

Fig. 8a shows the results of control operation without any obstacles. Here, The system was needed to achieve the goal through Learning Control. Four goals were assigned prior to the simulation. The simulation shows that the system was able to learn and operate the control object to achieve the designated goals.

4.2 Control Operation with Obstacles

Fig. 8b shows the results from the simulation of control operation with obstacles. In this simulation, obstacles were assigned prior to the simulation with four goals assigned to be achieved. The robot was set to reach the assigned goals in the domain. The results show that the system was able to operate the control object towards the designated goals while avoiding existing obstacles.

5 Conclusion

This study presents a cooperative multi-knowledge learning control to overcome operations with obstacles. The proposed method applies learning control for multiple Control Knowledge at the same time and cooperatively shares the Control

Knowledge by referring to the preference value sustained by each Control Knowledge. Control Knowledge is continuously updated through the Learning Control process. Rewards are given to the knowledge that provides the preference value of the operated control action.

The proposed method was applied in a control system for simulation of a small robot in a virtually constructed field map. The control system simulation was applied in a field, both with and without the obstacles. Results show that the system is able to utilize both Control Knowledge in performing a control operation with obstacle consideration. Therefore, Obstacle Consideration was achieved by calibrating two Control Knowledge in creating a safer cooperative command during the control operation.

Acknowledgement This research was supported by JSPS KAKENHI Grant Number 24500272

References

1. Schaal, S., Atkeson, C.G.: Learning Control in Robotics; Trajectory-Based Optimal Control Techniques. IEEE Robotics and Automation Magazine, Volume 7 Issue 2 pp.20-29 (2010)
2. Matsubara, T., Yasunobu S.: An Intelligent Control Based on Fuzzy Target and Its Application to Car Like Vehicle. SICE Annual Conference (2004)
3. Xu, X., Zuo, L., Huang, Z.: Reinforcement learning algorithms with function approximation: Recent advances and applications. Information Sciences, Journal of. 261 pp. 1-31 (2014)
4. Nakamura, Y., Ohnishi, S., Ohkura, K. ,Ueda, K.: Instance-Based Reinforcement Learning for Robot Path Finding in Continuous Space. IEEE SMC, pp.1229-1234 (1997)
5. Kamarulzaman, S.F., Shibuya, T., Yasunobu, S.: A Learning-based Control System by Knowledge Acquisition within Constrained Environment. IFSA World Congress, FC-104 pp.1-6 (2011)
6. Chang, D., Meng, J. E.: Real-Time Dynamic Fuzzy Q-Learning and Control of Mobile Robots. 5th Asian Control Conference, pp.1568 - 1576 Vol.3 (2004)
7. Busoniu, L., Babuska, R., Schutter B.D.: A Comprehensive Survey of Multiagent Reinforcement Learning. IEEE Trans. Systems, Man and Cybernetics, pp.156 - 172 Vol.38 No.2 (2008)
8. Gullapalli, V.: Direct Associative Reinforcement Learning Methods for Dynamic Systems Control. Neurocomputing 9, pp.271 - 292 (1995)
9. Sun. R, Sessions, C.: Multi-agent reinforcement learning with bidding for automatic segmentation of action sequences. 4th International Conference on MultiAgent Systems, pp.445 - 446 (2000)
10. Yu, J.: An Adaptive Gain Parameters Algorithm for Path Planning Based on Reinforcement Learning. 4th International Conference on Machine Learning and Cybernetics, pp.3557 - 3562 (2005)