Pertanika J. Sci. & Technol. 25 (S): 305 - 314 (2017)

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Compound Learning Control for Formation Management of Multiple Autonomous Agents

Syafiq Fauzi Kamarulzaman[1]* and Hayyan Al Sibai[2]

[1]Faculty of Computer System & Software Engineering, Universiti Malaysia Pahang, Lebuhraya Tun Razak 26300, Kuantan, Pahang Malaysia
[2]Faculty of Engineering Technology, Universiti Malaysia Pahang, Lebuhraya Tun Razak 26300, Kuantan, Pahang Malaysia

ABSTRACT

Having cooperation between multiple autonomous devices against one task is difficult due to each device having their own decision management based on self-deterministic protocol. Within the self-deterministic protocol, a formation management task should be considered along another task in order to provide cooperation and consideration between the operating autonomous devices. In this research, a compound learning control system for formation management of multiple control agents is proposed by managing coordination between multiple autonomous agents along with other tasks simultaneously in an operation. A series of simulation based on an autonomous robot was conducted to evaluate the effectiveness of learning through compound knowledge for providing consideration among achieving goals or coordination configuration against partner robot. The proposed system was able to provide consideration in coordination among operating partners in a task of achieving goal.

Keywords: Learning control, multi-agent, formation management, reinforcement learning, intelligent control

INTRODUCTION

Manual control is when a human performs tasks such as monitoring the state of the system, generating performance options, selecting the options in decision making and physical implementation on a device (Sutton et al.,1998). A device that is able to handle tasks done manually can be categorized as an autonomous device, capable of monitoring, deciding a control action and operating its own actuators for an assigned task. Having cooperation between multiple autonomous devices against one task is difficult due to each

ARTICLE INFO

Article history:
Received: 24 August 2016
Accepted: 02 December 2016

E-mail addresses:
syafiq29@ump.edu.my (Syafiq Fauzi Kamarulzaman),
mhdhayyan@gmail.com (Hayyan Al Sibai)
*Corresponding Author

ISSN: 0128-7680 © 2017 Universiti Putra Malaysia Press.

device having their own decision management based on self-deterministic protocol (Busoniu et al., 2008).

Various research has been conducted on formation management where the operation of multiple control agents is monitored by one intelligent system (Egerstedt et al., 2001) (Rui, 2010). A multi- agent control that is constrained by formation provides control of multiple devices that constrained the movement of the agents based on certain formation (Egerstedt et al, 2001). More application can be seen in (Rui, 2010) where multiple control agents in a form of unmanned aerial vehicle, UAV was operated according to formation assigned by an intelligent system. Within these research, each control agent was operated based on a primary decision that was provided by the main protocol. Thus, lack of the self-deterministic property of an autonomous device.

In this research, a compound learning control system for the management of multiple control agents is proposed for managing coordination between multiple autonomous agents. Earlier research was focused on multi-tasking by compound learning function for goal attainment as obstacle avoidance (Syafiq et al., 2013, 2014). From the previous research, it is understood that the compound learning function could consider multiple control knowledge (state-action rule) with each monitoring different state for producing the optimum action that can satisfy the rules within each control knowledge. Here, the research objective is for the compound learning function to consider a new task, which is the distance from operation partners, where the rules for partner's coordination has to be satisfied along with rules from other control knowledge concerning other tasks; In this case, a control knowledge for goal attainment. Therefore, the compound learning control system that is proposed in this research would consider goal attainment, and partner's coordination as dynamic constraints for having an effective autonomous control.

METHOD & DESIGN

The tasks performed by the Compound Learning Control System is operated by Learning Agents. These Learning Agents provide information on the control knowledge for each task to a compound function, with which the collection of this information is analysed and rearranged into a compound state-action rule defined by Compound Knowledge. Construction of control knowledge through Reinforcement Learning applies Q-learning, where value function Q is developed through states S and actions A conditions. Actions are selected among the options through preference value stored in the value function Q (S, A) as q and later updated after a control attempt through rewards r. In this research, Learning Control is used to apply control action to the main controller during a control operation, where here, the control object is a two-wheeled differential robot. Therefore, control action A is in the form of target rotation angle θ and target transition $\mathbf{\Delta r}$, where $A = \{\theta, \mathbf{\Delta r}\}$. Here, two learning control functions were applied on the control device where the control decision is unified by compound function. These two Learning Control functions require the functions to develop at almost similar phase by which minimum and maximum preference value q in the value function $Q\,(S, A)$ had to be constrained between 0 and 1. Therefore, the applied algorithm in the whole system is as,

$$Q(S,A) = (1-\alpha)Q(S,A) + \alpha\left[r + (1-Q(S,A))\max_{A} Q(S,A)\right]. \tag{1}$$

Learning Agent for Goal Attainment

The Learning Agent for goal attainment updates its value function according rewards r, obtained according to the distance between the control device and the goal location as shown in Figure1. Rewards are determined by comparing the distance of the control device to the goal from its last position with the distance to the goal from its current position. Based on Q-learning, the value function of the Learning Agent is updated by the distance towards goal $\mathbf{\Delta G}$ according to the conducted action $\boldsymbol{A}_{\mathrm{G}}$ that propels the control device closer to the goals, as

$$Q_1(\mathbf{\Delta G},\boldsymbol{A}_{\mathbf{G}}) = (1-\alpha)Q_1(\mathbf{\Delta G},\boldsymbol{A}_{\mathbf{G}}) + \alpha\left[r + (1-Q_1(\mathbf{\Delta G},\boldsymbol{A}_{\mathbf{G}}))\max_{A_G} Q_1(\mathbf{\Delta G},\boldsymbol{A}_{\mathbf{G}})\right]. \tag{2}$$

Learning Agent for Partner Consideration

The Learning Agent for partner consideration updates its value function according rewards r, obtained according to the distance between the control device and the operation partner as shown in Figure 2. Rewards are determined by comparing the distance of the control device to partner from its last position with the distance to partner from its current position. Based on Q-learning, the value function of the Learning Agent is updated by the distance towards partner $\mathbf{\Delta P}$ according to the conducted action $\boldsymbol{A}_{\mathbf{p}}$ that propels the control device to maintain the distant to the partner, as

$$Q_2(\mathbf{\Delta P},\boldsymbol{A}_{\mathbf{P}}) = (1-\alpha)Q_2(\mathbf{\Delta P},\boldsymbol{A}_{\mathbf{P}}) + \alpha\left[r + (1-Q_2(\mathbf{\Delta P},\boldsymbol{A}_{\mathbf{P}}))\max_{A_G} Q_2(\mathbf{\Delta P},\boldsymbol{A}_{\mathbf{P}})\right]. \tag{3}$$

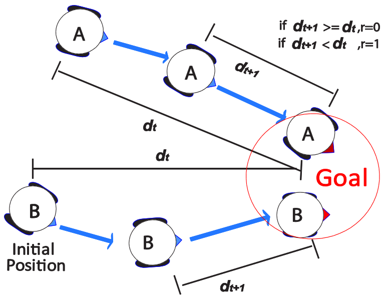

Figure 1. Rewarding in goal attainment function

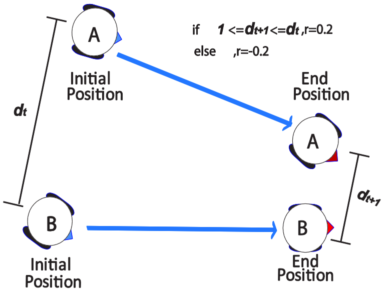

Figure 2. Rewarding in partner consideration function

Compound Learning Control System with Formation Management

The proposed system utilizes a Compound Function in a Learning Control System. Learning agents for Goal Attainment Q_G, and Partner Consideration Q_P which are the control knowledge for specific tasks, contains value function that is converted into a new value function denoted as Compound Control Knowledge, **CQ**. The learning agents are converged and a new value functions, listing actions A according to the minimum preference value q when comparing the list of actions in those learning agents based on state $\boldsymbol{S}_t$, creating the Compound Knowledge as,

$$Q_{All} = \min_{n=1,2} Q_n(\boldsymbol{S}_t, \boldsymbol{A}) \tag{4}$$

where the Learning Agent accountable for the list of actions is denoted as *N*,

$$N_t(\boldsymbol{S}_t, \boldsymbol{A}) = n, \tag{5}$$

defining the Compound Control Knowledge **CQ** as,

$$f(x) = \sum_{i=0}^{N-1} \alpha_i y_i \exp(-\| x - x_i \|^2 / 2\sigma^2) + b \tag{6}$$

Therefore, the overall system of Compound Learning Control for Formation Management of multiple autonomous agents is designed as shown in Figure 3.

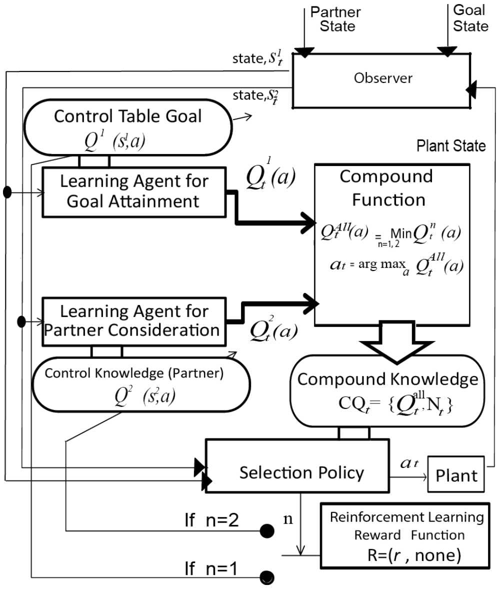
Partner State
Goal State
state, s_t^1
state, s_t^2
Observer
Control Table Goal
$Q^1(s^1,a)$
Plant State
Learning Agent for Goal Attainment
$Q_t^1(a)$
Compound Function
$Q_t^{All}(a) = \text{Min}_{n=1,2} Q_t^n(a)$
$a_t = \arg\max_a Q_t^{All}(a)$
Learning Agent for Partner Consideration
$Q_t^2(a)$
Control Knowledge (Partner)
$Q^2(s^2,a)$
Compound Knowledge
$CQ_t = \{Q_t^{all}, N_t\}$
Selection Policy
a_t
Plant
If n=2
n
Reinforcement Learning Reward Function
R=(r , none)
If n=1

Figure 3. Diagram of Compound Function System for Partner Consideration

SIMULATION SETTINGS

The Compound Learning Control System was applied in simulation of multiple two-wheeled differential robots. The properties of the robots are determined by the specification of the device in Figure 4 and Figure 5. Details of the device specification are described in Table 1. Both Learning Control functions applied in the Compound Learning Control System are based on the parameters given in Table 2. The controllers involved in managing the target transition and turning was created based on Proportional-Derivative (PD) controller. Intervals defined in Table 2 described that the states and targets were analysed in discrete form, where, for example, there are five options of turning angle and three options of transition.

Table 1
Robot Physical Specifications

Parameter	Values
Weight	0.515 [kg]
Size	Diameter: 0.19 [m] Height: 0.85 [m]

Table 2
Robot Physical Specifications

	Parameters	Range	Intervals
State, ***S***	(*Q_1*) Goal	-	2
	Distance **ΔG**(x_i,y)[m]	10 < **ΔG** < 10	
	(*Q_2*) Partner	1	0.5
	Distance **ΔG**(x_i,y)[m]	< **ΔG** < 10	
Action, ***A***	Target Angle, θ [rad]	-1< 0<1	0.5
	Target Position	0.1 < r <0.5	0.2
	Distance, r [m]		
Discount Rate, $\gamma = 0.3$		Learning Rate. $\alpha = 0.5$	

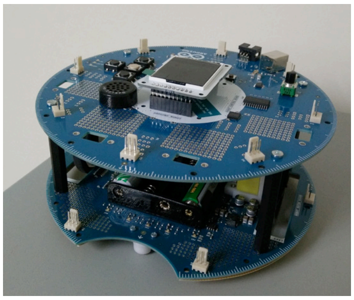

Figure 4. The robot which the simulation is based on (Arduino Mobile Robot Kit)

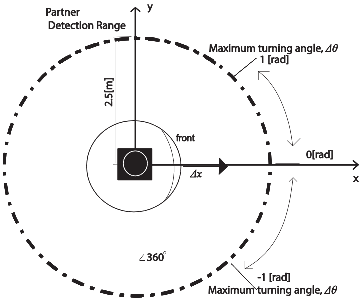

Figure 5. Specification of robot for simulation

The simulation was conducted in two phases; the first phase was the development phase, where value functions of both Learning Control Functions were developed through a series of basic simulation of goal training and partner coordination training. Two devices were simulated at once, where one of the devices is controlled through PD control entirely, and another one is embedded with the proposed system. Simulation concerning the development of both control function was conducted in 50 episodes for three assigned target without obstacles. However, during this phase, only the value functions for position control is updated during simulations with the goal, and only the value function for partner consideration is updated during the second simulation. The second phase was the operation phase, where value functions that have been developed in development phase were applied, united in the Compound Learning Control System, and was conducted in 50 episodes of iteration for two targets.

SIMULATION RESULTS

The simulation provides results based on the movement of the simulated device after completing the assigned 50 episodes of operation iteration. Learning agent developed in the first phase is compared with the movement results in the second phase.

During the first phase of simulation, from the development of goal attainment control knowledge through learning control. The control operation successfully achieved the goals assigned as shown in the Figure 6. Through this result, the value function of the Learning Control function for goal attainment has been successfully developed towards creating an expert control knowledge for controlling the simulated device. Thus, shows that the Compound Learning Control System should be able to successfully operate goal attainment control in case of an absence of an operation partner. Figure 7 provides information on the control knowledge of goal attainment during simulation. The proposed system was required to reach all three targets consecutively after each 50 iterations. Win =1 describes success on reaching the goal, while win=0 describes failure in reaching the target. Total Value TA describes the improvement occurred in the control knowledge where rewards r increases the overall value of the goal attainment knowledge as iteration increases, which means, with more stable Total Value TA, more consistent control it will make to achieve the goal. Here, the crossing point

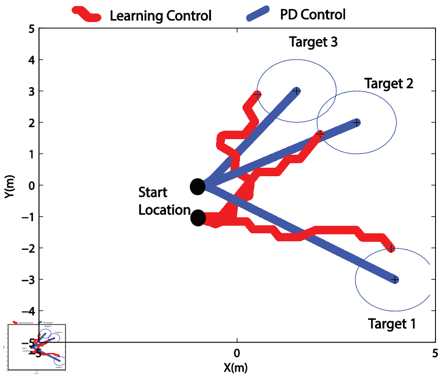

Figure 6. Control Result for target attainment

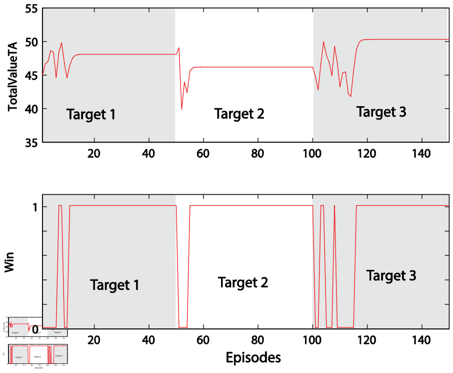

Figure 7. Improvement for Target Attainment Knowledge Learning

did not highlight the collision, since the time reach at that point is different according to each robot. The system was set to halt the robots, in the event of collision. .

During the first phase of the simulation, continues with the development of partner coordination control knowledge through learning control, the control operation successfully achieved the goals assigned as shown in the Figure 8. Through this result, the value function of the Learning Control function for partner coordination has been successfully developed towards creating an expert control knowledge for controlling the simulated device to follow the partner that was operated by PD control. Thus, shows that the Compound Learning Control System should be able to successfully operate partner coordination control in case of no goal information exists. The movement path is curved due to the requirement provided in the Learning Control where the control device has to maintain at least distance of 1[m] around the partner (save distance to avoid collision) to maintain the maximum reward, using path that helps avoid collisions, and gives movement room to the partner with PD control. Figure 9 provides understanding on how the control knowledge of partner coordination is developed during the simulation. The proposed system was required to reach all three targets consecutively after each 50 iterations of episodes. Total Value FM describes the improvement occurred in the control knowledge where rewards r increases the overall value of the partner coordination knowledge as iteration increases.

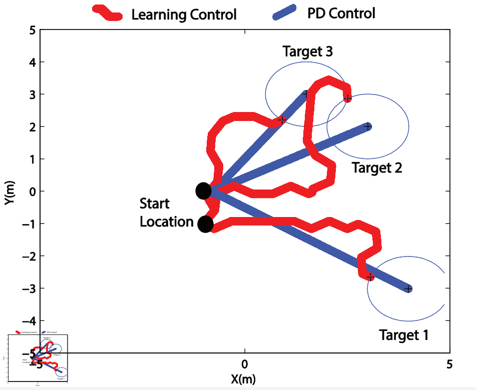

Figure 8. Improvement for Target Attainment Knowledge Learning

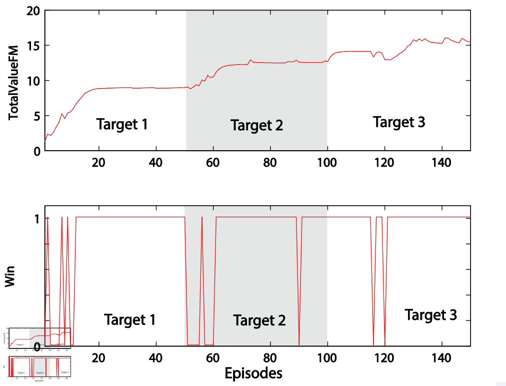

Figure 9. Control Knowledge Improvement for Partner Consideration Learning

During the second phase of the simulation, the control operation successfully achieved the goals assigned as shown in Figure 10. Through this result, the value function of the Learning Control function for partner coordination and goal attainment has been successfully applied and developed. towards creating priority for controlling the simulated device to coordinate with the partner that was operated by PD control while reaching the target area at an optimum movement. Thus, shows that the Compound Learning Control System were able to successfully operate partner coordination control along with goal attainment. The movement path is curved due to the requirement provided in the Learning Control where the control device has to maintain at least distance of 1[m] around the partner to maintain the maximum reward and conserve safety but still provide efforts to reach the target at the optimum movement.

Figure 11 shows how the control knowledge of partner coordination and goal attainment is developed during the second phase of simulation. The proposed system was required to reach all two targets consecutively for 50 iterations. Total Value TA and Total Value FM describes the improvement achieved in the control knowledge where rewards r increases the overall value of the partner coordination knowledge as iteration increases. Figure 12 provides understanding on how the control knowledge of partner coordination influence the goal attainment in the simulation. The movement path when applying compound function is slightly curved and delayed to preserve the distance between the robot partner but still completes the goal attainment task by optimum movement. Therefore, based on this figure, it is understood that the compound learning control method is applicable for formation management of multiple robots when completing a required task. Further study must include environmental constraints and also increase the number of partner robots in the operation.

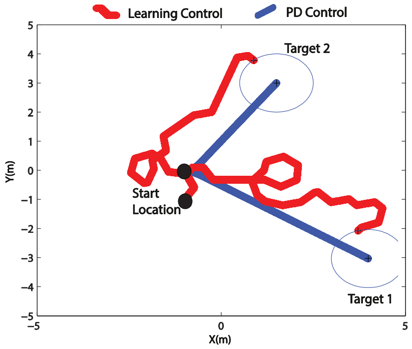

Figure 10. Result for Target Attainment and Partner Consideration using Compound Knowledge with partner using PD control)

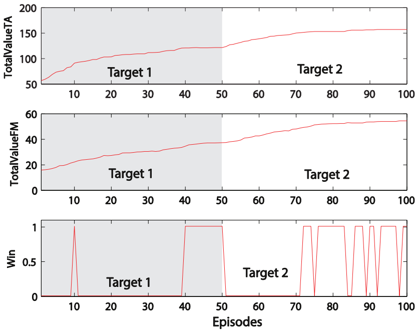

Figure 11. Control Knowledge Improvement using Compound Learning

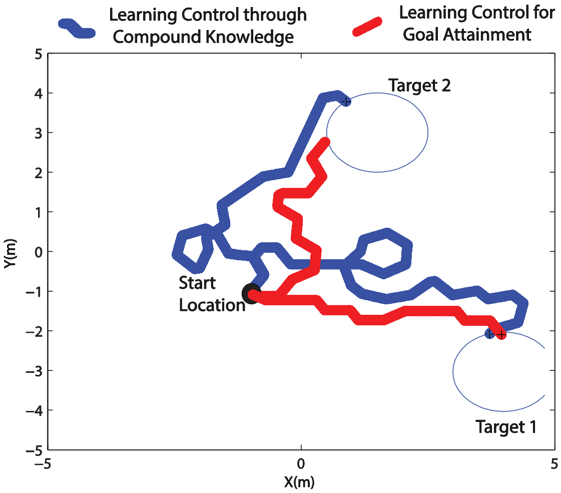

Figure 12. Comparing results using Compound Knowledge and normal Target Attainment Learning Control

CONCLUSION

In this research, a compound learning control system for the management of multiple control agents by managing coordination between multiple autonomous agents along with other tasks simultaneously was proposed. The Compound Learning Control System makes use of multiple Learning Control functions to provide expert control knowledge of position transition and partner coordination for autonomous two-wheeled differential robots, united by compound function for selecting the best control target options for operating the device. Simulation based on an autonomous robot was conducted to evaluate the effectiveness of learning through compound knowledge for achieving goals or coordination configuration against partner robot. The results of applying compound knowledge was undertaken by analysing the differences in coordination among partner robots. Results show that the Compound Learning Control System was able to provide successful controls towards the goal position with consideration of operation partners location. Therefore, the Compound Learning Control System for position and formation control was achieved. Further study will be conducted in the future to evaluate the reliability in more than two devices within constrained environment.

REFERENCES

Busoniu, L., Babuska, R., &Schutter, B. D. (2008). A Comprehensive Survey of Multi-Agent Reinforcement Learning. *IEEE Transaction on Systems, Man, and Cybernetics, 38*(2), 156-172.

Egerstedt, M., & Hu, X. (2001). Formation Constrained Multi-Agent Control. *IEEE Transactions on Robotics and Automation, 17*(6), 947-951.

Endsley, M. R., & Kaber, D. B. (1999). Level of Automation Effects on Performance, Situation Awareness and Workload in Dynamic Control Task. *Ergonomics, 42*(3), 462-492.

Rui, P. (2010). Multi-UAV Formation Manoeuvring Control Based on Q-Learning Fuzzy Controller. *Advanced Computer Control (ICACC), 4*, 252- 257.

Sutton, R. S., & Barto A. G. (1998). *Reinforcement Learning: An Introduction*. MIT Press.

Syafiq, F. K., & Yasunobu, S. (2014). Cooperative Multi- Knowledge Learning Control System with Obstacle Consideration. In *Proceedings of IPMU International Conference* (pp. 505-515).

Syafiq, F. K., & Yasunobu, S. (2015). Compound Learning Control for Autonomous Position and Obstacle Control of Aerial Hovering Vehicles. In *Proceedings of Asian Control Conference* (pp. 797-802).