Anadolu Üniversitesi Bilim ve Teknoloji Dergisi A- Uygulamalı Bilimler ve Mühendislik
**Anadolu University Journal of Science and Technology A- Applied Sciences and Engineering**

ANADOLU UNIVERSITY

**2017 - Volume: 18 Number: 4**
**Page: 864-875**
**DOI: 10.18038/aubtda.340101**
**Received: 15 March 2017 Revised: 08 August 2017 Accepted: 11 September 2017**

# EXPERIENCED TASK-BASED MULTI ROBOT TASK ALLOCATION

**Hatice Hilal EZERCAN KAYIR** [1]

[1] Electrical and Electronics Engineering Department, Engineering Faculty, Pamukkale University, Denizli, Turkey

## ABSTRACT

In multi robot system applications, it is possible that the robots transform their past experiences into useful information which will be used for next task allocation processes by using learning-based task allocation mechanisms. The major disadvantages of multi-robot Q-learning algorithm are huge learning space and computational cost due to generalized state and joint action spaces of robots. In this study, experienced task-based multi robot task allocation approach is proposed. According to this approach, robots believe to be experienced about the tasks most frequently done. Robots prefer to do these tasks rather than the inexperienced ones. Then, robots refuse to execute inexperienced tasks over time. This means that the system has reduced learning space. The proposed approach plays a crucial role to achieve required system performance and provides effective solutions to learning space dimensions. The effectiveness of the proposed algorithm is demonstrated by simulations on multi-robot task allocation problem.

**Keywords:** Multi robot task allocation, Multi-Agent Q-Learning, Adaptive system architecture

## 1. INTRODUCTION

Multi robot systems (MRS) are widely used especially in complicated applications, because they provide concurrent processing ability, faster task execution feature and robust system architecture [1]. Despite these advantages, the major drawback of MRS is precise and accurate coordination requirement [2]. The working environments of MRS are dynamic and partially observable in nature due to robots' independent decision-making and acting mechanism. For this reason, estimation of all possible situations and problems to be encountered during execution becomes impossible. To achieve desired system performance needs that the robots adapt themselves and adjust their decisions to dynamic working conditions. MRS with robots having learning ability provides robust system structure against uncertainties in environments and unpredictable problems [2].

In most real-life MRS applications, a prior knowledge about the time sequence of tasks is not accessible. Generally, tasks appear immediately and in random sequence during system execution. In such a case, these tasks can be allocated to the robots which are not busy with another task at that moment. In such a case, an effective coordination providing desired system performance cannot be realized. If robots having learning ability transform past task allocation experiences to a useful knowledge, it becomes possible to overcome these problems [3].

Q-learning is reinforcement learning method that is widely preferred in MRS due to its dynamic and adaptive structure [4]. Theoretically, Q-learning method is defined for single-agent systems satisfying Markov Decision Process properties [5]. In fact, most MRS environments are not MDP in nature. Centralized Q-learning method is an effective approach that is a direct application of single-agent Q-learning to multi-agent systems. The major disadvantage of centralized Q-learning is huge learning space dimension and computational cost due to generalized state and joint action spaces [3]. Especially for MRS with large number of robots, this causes difficulties in application.

*Corresponding Author: hezercan@pau.edu.tr

In this study, experienced task-based multi robot task allocation, ExpT-MRTA, approach is proposed. According to this approach, robots would rather execute more experienced tasks and drop less experienced tasks over an enough time period. So, the structure of MRS becomes less complicated. The dimension of learning space gets smaller and computational cost is reduced.

The structure of the paper is organized as follows: In Section 2, a brief explanation of market-based multi robot task allocation approach is given. In Section 3, Q-learning theory is presented. The problem studied in this paper is stated in Section 4. In Section 5, the proposed approach, Experienced Task-Based Multi Robot Task Allocation, is explained in detail. Application environment and experimental results are given in Section 6 and Section 7 respectively. Lastly, conclusion part is in Section 8.

## 2. MARKET-BASED MULTI ROBOT TASK ALLOCATION

Multi robot task allocation (MRTA) problem is defined as the process of deciding which task should be performed by which robot in which order [6]. In most cases, it is not possible to execute all tasks because of insufficiency in system resources, namely robots and their abilities [7]. Effective use of system resources plays a crucial role to get desired system performance. So, task allocation should be realized by maximizing overall system gain or minimizing system cost.

Market-based approaches provide efficient solutions for coordination problems in MRS by combining the advantages of centralized and distributed coordination approaches [8]. In these approaches, robots have independent decision-making mechanisms but overall coordination is realized by participation of all robots [9]. Auction protocols are widely used market-based task allocation methods [10]. An auction process starts with the announcement of tasks to be done by auctioneer. The robots having the ability of executing the announced tasks determine bid values and send to the auctioneer. In mobile robot applications, these bid values are calculated by using travelled distance, needed time [11] or required energy [12]. The auctioneer determines the winner robot in a manner that maximizing gain or minimizing cost and informs the robots. The assignment of tasks to appropriate robots is defined as Optimal Assignment Problem (OAP) in operations research studies [2]. Hungarian Algorithm is an efficient way to solve OAP [13].

## 3. Q-LEARNING

Reinforcement learning methods are the machine learning approaches that do not require any input-output data sets or a supervisor [14]. In reinforcement learning methods, the agents are directly connected to the environment by their perception and action units. Action of the agent causes the state transition of the environment. Agent is informed by a feedback signal called reward which indicates the effect of the action on the environment. Learning process is performed only through trial-and-error by using this reward value. Because of no supervisor requirement, simple structure of the algorithms and the possibility of using it in partially observable and dynamic environments, reinforcement learning approaches are preferred especially in multi robot system applications [5].

Theoretically, reinforcement learning approaches are defined on the Markov Decision Process (MDP) environments. An MDP is a tuple of $< S, A, P, \rho >$, where $S$ is finite and discrete set of environment states, $A$ is finite and discrete set of agent actions, $P: S \times A \times S \to \Pi(S): [0,1]$ is state transition function for each state-action pair and $\rho: S \times A \times S \to \mathbb{R}$ is reward function of the agent [15].

For any discrete step $k$, the environment state changes from $s(k) \in S$ to $s(k+1) \in S$ by the action $a(k) \in A$ of the agent. The reward value of $r(k) = \rho(s(k), a(k), s(k+1))$, which the agent receives as the result of $a(k)$, represents the instantaneous effect of action on the environment [14]. The agent

in an MDP aims to maximize the expected value of the overall reward for each step. For $k$th step, the expected value of overall gain is defined as in equation (1) [16].

$$Q^h(s,a) = E\left\{\sum_{i=0}^{\infty} \gamma^i\, r(k+i) \middle| s(k)=s, a(k)=a, h\right\} \tag{1}$$

$\gamma \in [0,1)$ is the discount factor. $Q$-function is expressed as the optimal action-value function and given in equation (2).

$$Q^*(s,a) = \max_h Q^h(s,a) \tag{2}$$

A learning agent should determine the optimal $Q$-value, $Q^*$ firstly and then should find required action by using action policy providing $Q^*$ [15].

Q-learning algorithm is a widely-used value function-based model- free reinforcement learning algorithm and proposed by Watkins [4]. According to Q-learning algorithm, the optimal $Q$-values for each state-action pair is calculated by the following recursive equation [17]:

$$Q(s(k),a(k)) = Q(s(k),a(k)) + \alpha\left[r(k) + \gamma \max_{a' \in A} Q_k(s(k+1),a') - Q(s(k),a(k))\right] \tag{3}$$

It is shown that learned $Q$-values converges the optimal $Q$-values with the probability '*1*', if each state-action pair is repeated infinitely many and learning rate $\alpha$ is decreased in each step $k$ for MDP environments [18].

Stochastic Game (SG) is the extended form of MDP to multi-agent case. An SG is defined as the tulle of $< S, A, P, \rho_j >$, where $S$ is the set of finite and discrete environment states, $A = A_1 \times A_2 \times \ldots \times A_m$ is the generalized action set for all agents, $m$ is the number of agents, $P: S \times A \times S \to \Pi(S): [0,1]$ is the state transition function for each state-action pair and $\rho_j: S \times A \times S \to \mathbb{R}, j = 1 \ldots m$ is reward function for each agent [17]. For an SG, the state transitions are realized by joint actions of all agents.
One solution approach in an SG is to get the Nash equilibrium [15]. The Nash equilibrium is defined as the joint action policy such that each agent's action policy provides maximum total reward value against others' action policy [5]. In the Nash equilibrium, it is not possible to increase the total reward by changing one agent's action policy while all other agents' action policies remain same. Hu and Wellman developed Nash-Q-learning algorithm which is based on reaching the Nash equilibrium [18]. It is shown that the optimal solutions are acquired under some certain conditions [19]. For each agent $j$, the $Q$-values are updated by equation (4).

$$\begin{aligned} V_N &= Nash_j(s, Q^1, \ldots, Q^j, \ldots, Q^m) \\ Q^j(s,a_1,\ldots,a_m) &= Q^j(s,a_1,\ldots,a_m) + \alpha\left[\rho_j + \gamma V_N - Q^j(s,a_1,\ldots,a_m)\right] \end{aligned} \tag{4}$$

It is shown that a fully cooperative SG is assumed as an MDP [20]. However, there exist more than one Nash equilibrium in an SG. It can be difficult to find joint actions of robots which result in Nash equilibrium due to agents' independent decision making ability.

## 4. PROBLEM STATEMENT

When a prior knowledge about tasks and their time sequence is accessible, it is possible to optimize system performance by preplanning the order of tasks performed by each robot. However, in most MRS applications, tasks appear in unpredictable time steps and order. System performance is negatively influenced by instantaneous allocation of tasks to the robots that is not busy with another

task at that moment. When there exist a hierarchical order among the tasks in terms of priority, emergency or sensitivity, the tasks which must be completed primarily and unconditionally cannot be executed if the robots are busy with the low-ordered tasks. For example, consider a two-robot system with ability to do all tasks. A high-priority task announced when all robots are occupied by low-priority tasks cannot be processed. This means that aimed system performance is not achieved and overall gain is reduced [3].

In auction-based MRTA approaches, the robots bid for the announced tasks if they have the ability to do such a task and they are not busy at that time. These approaches have no mechanism for reasoning about future task sequence. If robots have expectations about future task sequence, some robots will not be willing to do low-ordered tasks and will wait for a high-ordered tasks. Having such information about future task sequence is possible if robots have learning ability which transforms past experiences to a useful advice. For this purpose, Q-learning based multi robot task allocation approach is studied in recent studies [3, 21]. In this approach, robots decide whether they bid for announced low-ordered tasks or wait for future high-ordered tasks by using past experiences. It is shown that successful results are obtained [3]. However, the application of Q-learning algorithm, which is defined theoretically on single-agent systems, to multi-agent systems causes some problems such as huge learning space dimension and greater computational load. Learning space dimension increases exponentially depending on the number of agents because of generalized state and joint action spaces of agents.

ExpT-MRTA approach, proposed in this study, aims to obtain adaptive system architecture by means that the robots prefer to do more performed tasks and refuse to execute less performed tasks. According to this approach, robots have more experience about the tasks perform more than expected times. Robots are eager to execute the experienced tasks because these tasks are done with lower cost. On the contrary, robots do not want to execute inexperienced tasks and they refuse them over enough time period. Thus, robots start to perform experienced tasks only. This means reduced learning space dimension, reduced computational load and lower system cost as desired.

## 5. ExpT-MRTA: EXPERIENCED TASK-BASED MRTA APPROACH

A heterogeneous MRS with $m$ robots ($R_j, j = 1, \dots, m$) carries out $n$ different types of tasks ($T_i, i = 1, \dots, n$). Task experience parameter (TEP), ${}^{j}v_{\mathrm{i}}$, gives instantaneous experience information of robot $R_j$ about task $T_i$ and is defined as in equation (5).

$$ {}^{j}v_i = \frac{2}{1 + e^{-({}^{j}\eta_i - {}^{j}\mu_i)}} \tag{5} $$

${}^{j}\eta_i$ is the number of total $T_i$ tasks allocated to robot $R_j$ and ${}^{j}\mu_i$ is the expected value of $T_i$ tasks to be assigned to robot $R_j$. TEP means that a robot is experienced about a task when it executes that task more than the expected. TEP value changes between zero and two and it is one at the beginning. The experience status of robot $R_j$ about task $T_i$ is determined by task experience measure (TEM), ${}^{j}u_i$. TEM is calculated as the arithmetic mean of TEP values as shown in equation (6).

$$ {}^{j}u_i = \frac{\sum_{1}^{{}^{j}\eta_i} {}^{j}v_i}{{}^{j}\eta_i} \tag{6} $$

By using TEP and TEM values, robots derive three different behavior that are regular bidding, eager / hesitant to execute and refuse task. System architecture is reconstructed according to these behaviors as explained below.

### 4.1. Behavior-0: Regular Bidding

At the beginning, all robots are in behavior-0 state which means there is no knowledge about tasks. Robots in behavior-0 bid for announced tasks by using task cost as bid value if they are able to do announced task and are not busy at that time. In behavior-0, TEP value is in near neighborhood of one. A robot transits behavior-1, when it has any knowledge about experience status for a task. This behavior change occurs if the condition in equation (7) holds:

$$\left| {}^{j}v_{i} - 1 \right| < \delta_v \tag{7}$$

$\delta_v$ is a threshold value for TEP.

### 4.2. Behavior-1: Eager / Hesitant to Execute Tasks

The proposed approach thinks that more experienced tasks can be done by lower cost e.g. faster execution, less time or less energy. This assumption leads that the robots are eager to execute more experienced tasks and hesitant to execute otherwise. Eager / hesitant to execute tasks behavior affects bid values in auction process. In the case bid values are determined as cost of performing tasks, these bid values are re-calculated according to equation (8).

$$T_i \; bid \; value = \frac{T_i \; cost}{{}^{j}u_i} \tag{8}$$

Consequently, bid value decreases in eager to execute tasks case because of ${}^{j}u_i > 1$, whereas it increases in hesitant to execute tasks case where ${}^{j}u_i < 1$.

Eager / hesitant to execute tasks behavior plays an important role in the determination of reward values of state-action pairs in learning process. If a robot is eager to execute a task, the reward value related to do that task is increased proportional to TEM value and the reward value related not to do that task is decreased in the same manner. On the contrary, the reward value related to do that task is decreased by TEM value and the reward value related not to do that task is increased in hesitant to execute task behavior. In according to this, new reward value is calculated as below:

$$r_i' = \begin{cases} {}^{j}u_i \cdot r_i, & action\ to\ do\ announced\ task \\ \dfrac{r_i}{{}^{j}u_i}, & action\ to\ wait\ next\ task \end{cases} \tag{9}$$

$r_i$ represents the reward value of a state-action pair.

### 4.3. Behavior-2: Refuse Task

When TEM value calculated over enough time period is less than a threshold value as expressed in equation (10), robots think that they are inexperienced and refuse to perform that task. $\delta_u$ is TEM threshold value.

$${}^{j}u_i < \delta_u \tag{10}$$

In multi robot Q-learning approach, it is required that the learning process should be reorganized if any robot refuses to execute any task. Generalized state and joint action spaces of overall system are reconstructed by excluding the local state and action spaces of refused tasks. This procedure has advantages of simplified system structure and reduced learning space dimension, whereas it seems as a drawback.

The algorithm of proposed approach is given as below:

**ExpT-MRTA Algorithm**

***for*** *each announced task*
  ***if*** $R_j$ *bids for* $T_i$ *task*
    *Behavior-0: Regular Bidding*
      $T_i$ *bid value*=$T_i$ cost
    *Behavior-1: Eager/Hesitant to Execute Task* $T_i$
      $T_i \ bid\ value = T_i\ cost\ /\ {}^{j}u_i$
  ***end if***
  ***for*** *each task bidden*
    ***if*** $T_i$ *allocated to robot* $R_j$
      *Update* ${}^{j}\eta_i$
    ***end if***
    *Calculate* ${}^{j}v_i$
    *Update* ${}^{j}u_i$
  ***end for***
  ***if*** ${}^{j}u_i < \delta_u$
    *Behavior-2: Refuse Task* $T_i$
    *Re-arrange state and action spaces*
  ***end if***
  ***for*** *each task bidden*
    *Q-Learning Process*
    *Behavior-1: Eager/Hesitant to Execute Tasks*
      *Adapt reward values*
  ***end for***
***end for***

## 6. APPLICATION ENVIRONMENT

To show the effectiveness of the proposed approach, an experimental environment on which the applications are realized is prepared. The multi-robot system used in applications consists of ten robots, ($R_1$, $R_2$,..., $R_{10}$). The system is said to be fully heterogeneous because the robots have different skills. There exist eight different type of tasks ($T_1$, $T_2$,..., $T_8$). The robots and the tasks which the robots are capable of executing are given in Table 1 by '+' sign.

Tasks are generated randomly and with equal probability. The number of tasks announced at any time could be between three and seven. Each task has two different priority degrees, low-priority and high-priority. Low-priority tasks and high-priority tasks are 65% and 35% of all tasks, respectively.

**Table 1.** Robots and tasks performed by robots.

| Robots | Tasks | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | $T_1$ | $T_2$ | $T_3$ | $T_4$ | $T_5$ | $T_6$ | $T_7$ | $T_8$ |
| $R_1$ | + | | + | + | | + | | |
| $R_2$ | + | + | | | + | | + | + |
| $R_3$ | | | | + | | + | + | |
| $R_4$ | | | | + | | | | + |
| $R_5$ | | + | | | + | | + | + |
| $R_6$ | + | + | | | | | | |
| $R_7$ | | | + | | | + | + | |
| $R_8$ | | + | | | + | | | + |
| $R_9$ | | + | | | | + | + | |
| $R_{10}$ | | | + | + | + | | | + |

The applications are realized by using 45 experimental sets, each having nearly 90 tasks. First 30 sets are used for learning process and last 30 sets are used for test purpose. It is assumed that the tasks assigned to robots are completed. None of the allocated tasks is left unfinished.

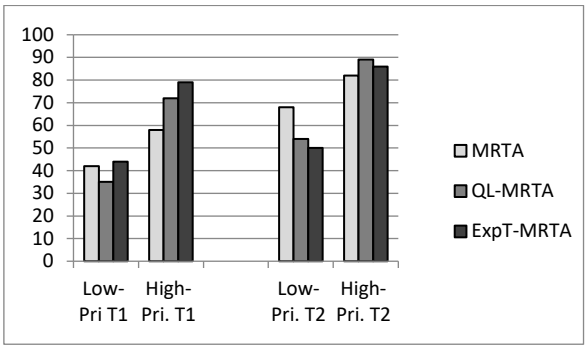
## 7. EXPERIMENTAL RESULTS

The purpose of the proposed approach is to increase system performance by using learning-based MRTA. The essential goal of the study is to increase the number of executed high-priority tasks in addition to the number of completed low-priority tasks remains as high as possible. To show the effectiveness of the proposed approach, the applications are realized in the experimental environment whose details are given in the previous section and the results are compared in terms of the task completion ratio. The task completion ratio is defined as the ratio of the number of tasks assigned to any robot to the total number of tasks announced. Applications are executed for three different approaches as traditional market-based task allocation (MRTA), Q-learning-based task allocation (QL-MRTA) and experienced task-based task allocation (ExpT-MRTA). The results of low-priority and high-priority tasks for each task type are given separately in Figure 1.

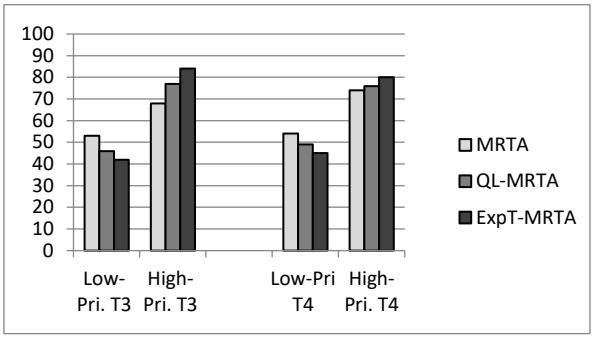
It is clear that the MRS in this study is much more complicated system with large number of robots and their task execution abilities. The application of Q-learning algorithm in MRS causes huge learning space dimension and high computational cost due to generalized state and joint action spaces. According to the proposed approach, ExpT-MRTA, robots want to execute the experienced tasks because of their lower cost. On the contrary, robots do not want to execute tasks that are not experienced enough and they refuse them enough time later. So, robots' individual state and action space dimensions become smaller. Consequently, learning space dimension and computational cost decrease.

The graphs in Figure 1 indicate that the completion ratios of high-priority tasks of all task types are higher than low-priority tasks because of MRTA algorithms' nature.

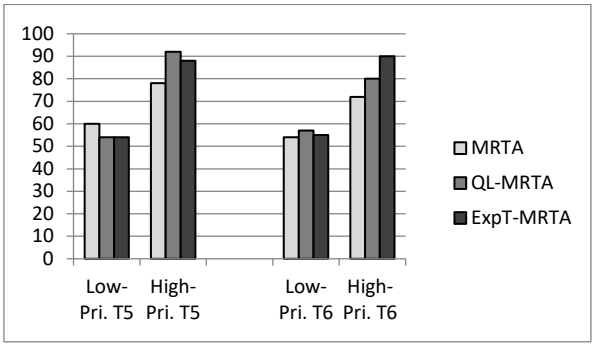
The results obtained by QL-MRTA approach indicates that the robots are successfully use their past task-allocation experiences by means of learning ability. The completion ratios of high-priority tasks



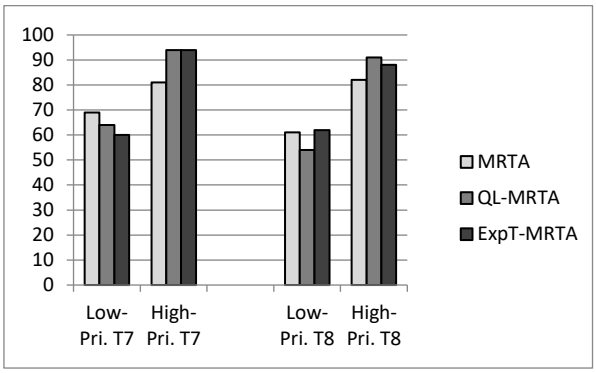

(a)




(b)




(c)




(d)

**Figure 1.** Task completion ratios of all task types: (a) for $T_1$ and $T_2$; (b) for $T_3$ and $T_4$; (c) for $T_5$ and $T_6$*;* (d) for $T_7$ and $T_8$.

get higher. So, increased system performance is achieved. However, a small amount decrease occurs in the completion ratio of low-priority tasks. In fact, the number of total tasks to be completed is limited due to restricted system resources.

In the proposed approach, ExpT-MRTA, a great change in the task completion ratios of both high-priority and low-priority tasks does not occur although a lot of robots refuse to execute one or more tasks. The robots and their experienced tasks are shown in Table 2. According to this table, it is seen that all robots except robots $R_6$ and $R_7$ give up executing at least one task.

**Table 2.** Robots and experienced tasks.

| Robots | Tasks | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | $T_1$ | $T_2$ | $T_3$ | $T_4$ | $T_5$ | $T_6$ | $T_7$ | $T_8$ |
| $R_1$ | + | | + | | | | | |
| $R_2$ | + | | | | + | | + | |
| $R_3$ | | | | + | | + | | |
| $R_4$ | | | | | | | | + |
| $R_5$ | | + | | | | | | + |
| $R_6$ | + | + | | | | | | |
| $R_7$ | | | + | | | + | + | |
| $R_8$ | | | | | + | | | + |
| $R_9$ | | + | | | | | + | |
| $R_{10}$ | | | + | + | | | | |

As an example, robot $R_5$ has the ability to do four different tasks, namely $T_2$, $T_5$, $T_7$ and $T_8$ at the beginning. After a while, robot $R_5$ thinks that become inexperienced about task $T_7$ because task $T_7$ is rarely assigned. TEM value of robot $R_5$ for task $T_7$ gets lower than the threshold. After that, robot $R_5$ refuses to execute task $T_7$. While the dimension of individual learning space of robot $R_5$ is 512 initially, it reduces to 162 after the dropping of $T_7$. Later, task $T_5$ is refused by robot $R_5$ in a similar manner. At the end of the ExpT-MRTA algorithm, robot $R_5$ performs only two tasks, $T_2$ and $T_8$, instead of four tasks. So, the dimension of individual learning space of robot $R_5$ decreases to 32. The reduction of learning space dimension of robot $R_5$ from 512 to 32 while the task completion ratios of mentioned tasks remain nearly unchanged indicates the success of the proposed algorithms. TEM values of $T_2$, $T_5$, $T_7$ and $T_8$ tasks for robot $R_5$ is shown in Figure 2(a). Behavior transition of robot $R_5$ and change in learning space dimension of robot $R_5$ are drawn in Figure 2(b) and Figure 2(c) respectively.

In multi robot Q-learning approach, the dimensions of generalized state and joint action space increases exponentially depending on the number of robots. ExpT-MRTA approach provides a considerable reduction in the learning space dimension as a result of the decrease in the number of tasks performed by each robot. For the system evaluated in this study, the dimension of learning space for QL-MRTA approach is found as 3498. The dimension of learning space reduced to 550 when ExpT-MRTA approach is used. As a result, by taking into account the significant difference between these values and the task completion ratios it is evident that the proposed approach yields successful results.

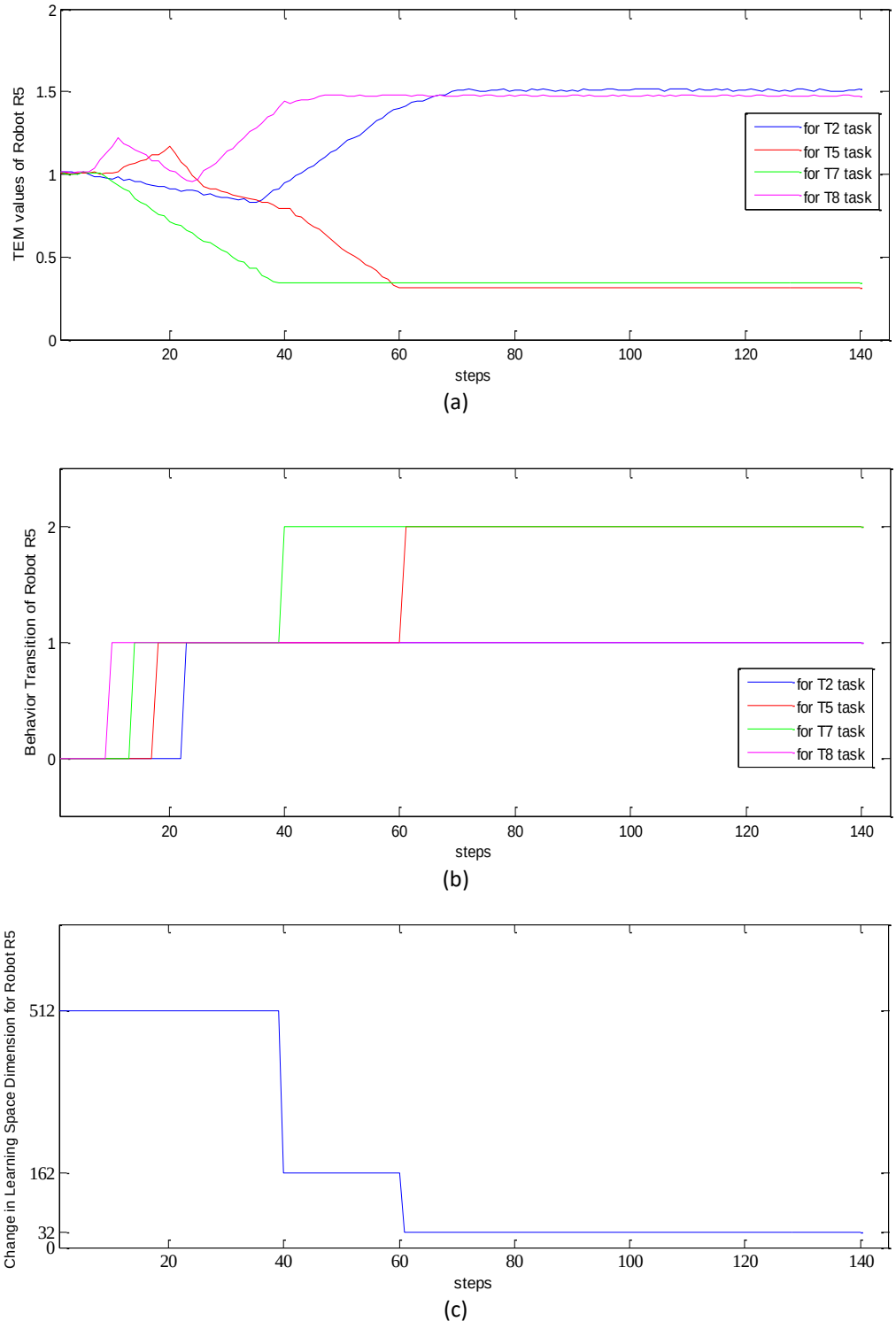



**Figure 2.** Result of ExpT-MRTA for robot $R_5$: (a) TEM values of $T_2$, $T_5$, $T_7$ and $T_8$ tasks for robot $R_5$; (b) Behavior transition of robot $R_5$; (c) Change in learning space dimension of robot $R_5$.

## 7. CONCLUSION

In this study, experienced task-based multi robot task allocation, ExpT-MRTA, is proposed. In this approach, robots believe that they are experienced about more executed tasks and inexperienced for less performed tasks. For this, robots derive Behavior-1:eager / hesitant to execute task and Behavior 2: refuse task. Behavior-1 influences the decision of bidding and calculated bid value in auction process. In addition, reward value of state-action pairs is directly related to this behavior. Behavior-1 feeds positively the eager or hesitant performance of the robots. Behavior-2 affects the structure of learning process. The reconstruction of generalized state and joint action spaces is required because of Behavior-2. As a result of Behavior-2, learning space dimension is decreased to a reasonable level and also the desired system performance is achieved. Experimental results indicate that ExpT-MRTA proposes a successful solution for huge learning space dimension and high computational load problems of multi robot Q-learning applications.

## REFERENCES

[1] Arkin R C. Behavior-Based Robotics. Cambridge, MIT Press, 1998.

[2] Mataric M J. Reinforcement learning in multi-robot domain. Autonomous Robots 1997; 4(1): 73-83.

[3] Ezercan Kayır H H, Parlaktuna O. Strategy planned Q-lerning approach for multi-robot task allocation. Procs of ICINCO 2014; 2: 410-416.

[4] Sutton R S, Barto A G. Reinforcement Learning: An Introduction. Cambridge, MIT Press, 1998.

[5] Yang E, Gu D. Multiagent Reinforcement Learning for Multi-Robot Systems: A Survey. Technical Reports of the Dept. of Computer Science, Univ. of Essex, 2004.

[6] Gerkey B P, Mataric M J. A formal analysis and taxonomy of task allocation in multi-robot systems. International Journal of Robotics Research 2004; 23(9): 939-954.

[7] Jones E G, Dias M B, Stentz A. Learning-enhanced market-based task allocation for oversubscribed domains. Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems; 2007; San Diego, CA, USA: pp. 2308-2313.

[8] Dias M B, Zlot R M, Kaltra N, Stentz A. Market-based multirobot coordination: a survey and analysis. Proceedings of the IEEE 2006; 94(7): 1257-1270.

[9] Zlot R, Stentz A. Market-based multirobot coordination for complex tasks. Int. Journal of Robotics Research, Special Issue on the 4th Int. Conf. on Field and Service Robotics 2006; 25(1): 73-101.

[10] Gerkey B P, Mataric M J. Sold!: auction methods for multi robot coordination. IEEE Transactions on Robotics and Automation 2002; 18(5): 758-768.

[11] Mosteo A R, Montano L. Comparative experiments on optimization criteria and algorithms for auction based multi-robot task allocation. Proceedings of the IEEE International Conference on Robotics and Automation 2007; pp. 3345-3350.

[12] Kaleci B, Parlaktuna O, Ozkan M, Kırlık G. Market-based task allocation by using assignment problem. IEEE Int. Conf. on Systems, Man, and Cybernetics; 2008; pp. 135-14.

[13] Hatime H, Pendse R. A comparative study of task allocation strategies in multi-robot systems. IEEE Sensors Journal 2013; 13(1): 253-262.

[14] Russel S, Norvig P. Artificial Intelligence a Modern Approach. New Jersey, Prentice Hall, 2003.

[15] Buşoniu L, Babuška L, Schutter B. A comprehensive survey of multiagent reinforcement learning. IEEE Trans. on Systems, Man, and Cybernetics 2008; 38(2): 156-172.

[16] Kaelbling L P, Littman M L, Moore A W. Reinforcement learning: a survey", Journal of Artificial Intelligence Research 1996; 4: 237-285.

[17] Watkins C J C H. Learning from delayed rewards. PhD Thesis, University of Cambridge, UK, 1989.

[18] Watkins C J, Dayan P. Q-learning, Machine Learning 1992; 8.

[19] Hu J, Wellman M P. Nash Q-learning for general sum games. Journal of Machine Learning Research 2003; 4: 1039-1069.

[20] Boutlier C. Planning learning and coordination in multiagent decision processes. Proceedings of the 6th Conference on Theoretical Aspects of Rationality and Knowledge, TARK '96; 1996; pp. 195-210.

[21] Ezercan Kayır H H. ContQL-MRTA: Continuous-learned Q-learning based task allocation approach in multi robot systems. TOK 2015; 10-12 Eylül 2015; Denizli, Turkey. (article in Turkish with an abstract in English).