Machine Learning Research
2017; 2(4): 133-147
http://www.sciencepublishinggroup.com/j/mlr
doi: 10.11648/j.mlr.20170204.14

Performance Evaluation of Cooperative RL Algorithms for Dynamic Decision Making in Retail Shop Application

Deepak Annasaheb Vidhate[1, *], Parag Arun Kulkarni[2]

[1]Department of Computer Engineering, College of Engineering, Pune, India
[2]iKnowlation Research Labs Pvt. Ltd., Pune, India

Email address:

dvidhate@yahoo.com (D. A. Vidhate), parag.india@gmail.com (P. A. Kulkarni)
[*]Corresponding author

To cite this article:

Deepak Annasaheb Vidhate, Parag Arun Kulkarni. Performance Evaluation of Cooperative RL Algorithms for Dynamic Decision Making in Retail Shop Application. *Machine Learning Research*. Vol. 2, No. 4, 2017, pp. 133-147. doi: 10.11648/j.mlr.20170204.14

Received: September 27, 2017; **Accepted**: October 20, 2017; **Published**: December 12, 2017

Abstract: A novel approach by Expertise based Multi-agent Cooperative Reinforcement Learning Algorithms (EMCRLA) for dynamic decision-making in the retail application is proposed in this paper. Performance evaluation between Cooperative Reinforcement Learning Algorithms and Expertise based Multi-agent Cooperative Reinforcement Learning Algorithms (EMCRLA) is demonstrated. Different cooperation schemes for multi-agent cooperative reinforcement learning i.e. EQ learning, EGroup scheme, EDynamic scheme and EGoal driven scheme are proposed here. Implementation outcome includes a demonstration of recommended cooperation schemes that are competent enough to speed up the collection of agents that achieve excellent action policies. This approach is developed for three retailer stores in the retail marketplace. Retailers are able to help with each other and can obtain profit from cooperation knowledge through learning their own strategies that exactly stand for their aims and benefit. The vendors are the knowledgeable agents in the hypothesis to employ cooperative learning to train helpfully in the circumstances. Assuming significant hypothesis on the vendor's stock policy, restock period, arrival process of the consumers, the approach is modeled as Markov decision process model that makes it possible to design learning algorithms. Dynamic consumer performance is noticeably learned using the proposed algorithms. The paper illustrates results of Cooperative Reinforcement Learning Algorithms of three shop agents for the period of one-year sale duration and then demonstrated the results using proposed approach for three shop agents for the period of one-year sale duration. The results obtained by the proposed expertise based cooperation approach show that such methods can put into a quick convergence of agents in the dynamic environment.

Keywords: Cooperation Schemes, Multi-Agent Learning, Reinforcement Learning

1. Introduction

The retail store sells the household items and gains profit by that. Retailers are interested in their selling, their profit. By accepting certain steps, the portion that can reason break or decrease the revenue can be prohibited. The aim of predicting the sales business is to collect data from various shops and analyze it by machine learning algorithms. The proficient significance of the practical information by ordinary ways is not practically achievable because the information is extremely vast [1]. Retail shops example is considered here. Walmart is an example for huge shops, big bazaars etc. Most of the time retailers will not be doing well in getting the consumer's requests because they will be unable in the estimation of marketplace perspective. In some particular occurrences, the speed of sale or shopping is more. Sometimes it might reason insufficiency of the items. The relationship between the consumers and the shops is evaluated and the modifications that require gaining extra yield are prepared. The history of buy of each item in each shop and department is maintained. By examining these, the sales are predicted that facilitate the understanding of yield and loss happened throughout the year [1-2]. Let us consider example Christmas in some branch for the period of the specific session. In Christmas celebration, the sales are more in shops like clothing, footwear, jewelry etc. Throughout summertime the purchase of cotton clothing is more; in winter the purchase for sweaters is more. The purchase of

items alters as indicated by the season. By examining this past record of purchases, the sales can be forecasted for the future [2]. That discovers the result to predict the highest revenue in the industry of retail shop market. The retailers monitor the behavior of consumers and attract them by offering several beautiful schemes. In order, they will be back to the shop and pay for more time and money. The major target of retail shop market preparation is to acquire the highest revenue by significant the knowledge and where to provide gainfully and in which shops [2-3].

There are many challenges in the retail shop forecasting. Some of them are retailers be unsuccessful in the estimating the possibility of the market. Retailers disregard the seasonal changes. The human resources are insufficient and the workers do not exist as and when required. The retailers experience the complexity in storage management system. The retailers sometimes pay no attention to the competition or cooperation in the market. Retailers build the strategies that encourage the success and the extremely target plan. The strategies should be such that they facilitate to achieve the highest revenue [3].

Generally, the income of the sale of a specific product is kept which is the result of forecasting the maximum potential of the quantity of sales in given period of time and under uncertain environment. Market sale determined by the customer's behavior, the cooperation, facilities support etc. These take effect on the sales of future of a particular shop. Shop and inventory scheduling is significant and is organized policy method in individual shop level [4]. Goods to buy and sale, store management, and space management are the major work in the planning of a shop. By monitoring the past history of the shop it helps to put up a scheme of sales of the shop and build any changes in the idea so that it can be highest cost-effective. The fundamental information presented by the existing shop is extremely useful in the forecasting of sales [5].

The paper is arranged as: Section 2 provides the proposed approach toward dynamic decision-making in retail shop application by Expertise Based Multi-agent Cooperative Reinforcement Learning Algorithms (EMCRLA). Section 3 illustrates expertise based multi-agent cooperative learning schemes. Section 4 presented Mathematical Model of Cooperative Learning for the system of retail shops. Section 5 demonstrates Implementation Results of Cooperative Reinforcement Learning Algorithms and in Section 6 implementation Results of proposed approach Expertise based Multi-agent Cooperative Reinforcement Learning Algorithms (EMCRLA) are given. Result Analysis of Cooperative Reinforcement Learning Algorithms and proposed approach Expertise based Multi-agent Cooperative Reinforcement Learning Algorithms (EMCRLA) are given in Section 7.

2. Expertise Based Multi-Agent Cooperative Reinforcement Learning Algorithms (EMCRLA)

Three shop agents cannot obtain the maximum profit without cooperation. Cooperative Reinforcement Learning Algorithms for these shop agents certainly increases the saleofitems due to cooperation between them that gives a significantrise in profit. Convergence of reinforcement learning becomes important as a numberof states increases. Adding expertness factor into cooperative reinforcementlearning would surely enhance its performance in terms of profit and also can put into a quick convergence of agents in the dynamic environment. Hence the proposed approach Expertise based Multi-agent Cooperative Reinforcement Learning Algorithms (EMCRLA) is given.

The communication in multi-agent reinforcement can build a sophisticated collection of accomplishments achieved from the agents' proceedings. The part of accomplishments set is allocated to the agents via an *Incomplete Action Plan* (Q_i) [5-6].

Normally similar incomplete policies maintain incomplete information about the state. These strategies can be incorporated to improve the sum of the partial rewards received using satisfactory association method. The action plans are generated via the way of multi-agent cooperative reinforcement training through gathering such rewards with constructing these agents to go nearer to the excellent policy Q*. Once the plans Q_1,……, Q_x is incorporated, it is possible to build up a new strategy that is *Complete Action Plan* (*CAP* = *{CAP$_1$,..., CAP$_x$}*), in which C*AP$_i$* denotes the best reinforcement received by agent iall over the training algorithm [7].

The *Splan* algorithm 1 gives out the agents' training particulars. Strategies are considered by the Q-learning method for every algorithm. The best reinforcements are distributed to *CAP* with the aim to create a gathering of such best-collected rewards by every agent. Such rewards are one more time given by the way of the extra agents [8]. Coordination is implemented through the changing of incomplete rewards because *CAP* is forecasted by the best reinforcements. A *val*task is applied in the direction of the discovery of excellent strategy among the previous states and last state for a specified plan that calculates *CAP* with the best reinforcements. The *val* task is found out as adding of phases the agent demand to reach at final-state and the sum of the acquired amount in the plan amongst every initial state and final state [8-10].

Algorithm 1 Multi-agent cooperative RL Algorithm

Algorithm *Splan* (I, technique)

Consider state s, action a, agent iϵ I,α learning rate, γ discount factor and Q table Q(s, a)

Begin

(1). Initialization Q_i (s, a) and CAP_i (s, a)
(2). Communication by the way of the agents *i*ϵ I*;*
(3). episode ← episode +1;
(4). Revise policy which determines the reward value;
(5). Q (s, a)← Q (s, a) +α (r + γQ (s', a') – Q(s, a))
(6). Fco-op (epi, scheme, s, a, i);
(7). Q_i←CAP that is Q_i of agent *i* ϵ I is customized by the way of CAP_i.

End

The F*co-op* function decides a coordination scheme. *epi,*

*scheme, s, a, I*are the parameters, in which *epi i*s a current episode, coordination *scheme* is {EG*roup*, ED*ynamic, EGoal-driven*}, *s* and *a* are state and action chosen accordingly;

2.1. Expertise Rewards

More skilled agents discover additional reinforcements and penalty of the set. As an effect, if the set achieves reinforcement then expertise agents will obtain additional rewards as compared to another agent. On the opposite, further agents receive more punishments as measured to expert agents when the group gets punishment. Expert agents normally execute better than other agents [9-10]. They find extra chance to conduct correct action as measured apart from less expert agents. Agents acquire rewards (rewards and penalty) as follows:

$$r_i = Rx \frac{e_i}{\sum_{j=1}^{N} e_j} \quad (1)$$

where r is a reward, R is a cumulative reward, N is a number of agents, e_i is the expertise of agent i and $e_j$are theexpertise of agent j.

2.2. Expertise Criteria

Expertise criteria consider both reinforcements and penalty as a symbol of being knowledgeable. It indicates that negative and positive results, calculated based upon the cost of reinforcement and penalty indication, are together important for the agent. It is an addition to the complete cost of the reinforcement signal [11-12].

$$e_i = \sum_{t=1}^{now} |r_i(t) \quad (2)$$

3. Expertise Based Multi-Agent Cooperative Schemes

Various expertise based multi-agent cooperative schemes for cooperative reinforcement learning are given below [12].

1. *EGroup scheme* – reinforcements are issued in a sequence of steps.
2. *EDynamic scheme* – reinforcements are issued in each action.
3. *EGoal-driven scheme* – issuing the addition of reinforcements when the agent reaches the goal-state (S_{goal}) [13-14]

Algorithm 2 Cooperation schemes Consider state s, action a, agent i, reward r, number of agents N, α learning rate, γ discount factor, expertise of agent i is e_i and expertise of agent j aree_j, and Q table Q(s, a).

Begin
Fco-op (epi, scheme, s, a, i)
switch between schemes
In case of EGroup scheme
if episode mod N = 0 then
get_Ereward (e_i, e_j, N, R);
get_Policy (Q_i, Q*, CAP_i);
In case of EDynamic scheme
get_Ereward (e_i, e_j, N, R);
r ←$\sum_{j=1}^{x} Qj(s,a)$;
Q_i (s, a)← r;
get_Policy (Q_i, Q*, CAP_i);
In case of EGoal-driven scheme
if S = S_{goal}then
get_Ereward (e_i, e_j, N, R);
r ←$\sum_{j=1}^{x} Qj(s,a)$;
Q_i (s, a)← r;
get_Policy (Q_i, Q*,CAP_i);
End
Algorithm 3 get_Policy(Q_i, Q*,CAP_i)
Begin
Function get_Policy(Q_i, Q*,CAP_i)
While each agent i ϵ I do
while each state s ϵ S do
if value(Q_i, s) ≤ value(Q*,s) then
CAP_i(s,a) ← Q_i(s,a);
done
End
Algorithm 4 get_Ereward (e_i, e_j, N, R)
Begin
Function get_Ereward(e_i, e_j, N, R)
while agent i ϵ I do
while state s ϵ S do
get_Expertise(e_i);
$r_i = Rx \frac{e_i}{\sum_{j=1}^{N} e_j}$
done
done
return r_i
End
Algorithm 5 get_Expertise (r_i)
Begin
Function get_Expertise(r_i)
while agent i ϵ I do
whilestate s ϵ S do
$e_i = \sum_{t=1}^{now} |r_i(t)$
done
done
return e_i
End

The *EGroup scheme* appears to be extremely strong meeting extremely quick to the best action plan Q*. Reinforcements obtained by the agents are produced in series of pre-identified stages. They gather reasonable reward values that cause a good convergence. In the *EGroup scheme,* the global policy converges to the best action strategy as there is an intermission of series necessary to gather good reinforcements [14-16]. The global action policy of the *EDynamic scheme* is able to gather excellent reward values in small learning series. It is observed that after some series, the performance of global strategy reduces. This takes place because the states neighboring to the goal state begin to gather much-advanced reward values giving to a local maximum. It punishes the agent as it will no longer stay in the other states. In the *EDynamic scheme* as the

reinforcement learning algorithm renews learning values, actions with higher gathered reinforcements are chosen by the top possibility than actions with small gathered reinforcements. Such a policy is recognized as *greedy* search [15-16]. In the *EGoal-driven scheme,* the agent distributes its learning in a changeable number of sequences and the cooperation acquired when the agent arrives at the goal-state. The global action strategy of the *EGoal-driven* scheme is capable to gather excellent reward values, agreed that there is a sum of iteration series to gather values of acceptable rewards. The execution of the cooperative learning algorithms is generally small in the early series of the learning process with the *EGoal-driven* scheme [16-17].

4. Model of Cooperative Learning

Wedding period situation is considered for the development. Beginning from choosing a site, invitation cards, decoration, booking the caterers, purchase of clothing, gifts, jewelry and additional items for bride and groom, so many actions are concerned. Such periodical conditions are able to be practically executed like follow: Consumer purchasing in clothing shop surely go for the purchase of jewelry, footwear, and further related items. Retailers of various items can come jointly and in cooperation fulfill consumer demands and can acquire the profit by an enhancing in the item sale [17]. Below are mathematical notations for above model.

(1) Customer arrival at Poisson flow λ.
(2) Assume the number of shops / seller agents l = 1,2,3
(3) The consumer's need, D at any seller shop for a said item can be represented as $D=\sum_{i=1}^{N} d$. Here d_i denotes the need of the i^{th} consumer, that have a discrete uniform distribution U(a,b).
(4) N is a Poisson distributed random variable with parameter λ. N stands for the number of consumers that arrive at the time period. The cost of λ relies on i the retailer.
(5) It is assumed that f1=1/2, f2=1/3, f3=1/5, that means 50% of customers are visiting shop 1, 30% of customers visiting shop 2 and 20% of customers visiting shop 3 [17-18].
(6) Refill times at three shops are spread evenly with mean 1/μ [18].
(7) Let the shop 1 price estimate p and refill time estimate w (equal to projected refill time 1/μ) represented by U1(p,w)
(8) Learning parameter =0.2 and Discount rate =0.99.
(9) Seller has limited stock capability I_{max} and pursues a permanent interchange strategy [19-20].
(10) States (s): Assume maximum stock level at each shop = I_{max} = i_1,i_2,i_3= 20. State for agent 1 become (x_1,i_1) e.g. (5,0) that means 5 customer requests with 0 stock in shop 1. Similarly state for agent 2 become (x_2,i_2) and state for agent 3 become (x_3,i_3). State of the system become Input as (x_i,i_i) [21-22]
(11) Actions (a): Assume set of possible actions i.e. action set for agent 1 is (that means Price of products in shop 1) A1 = Price p = {8 to 14} = {8.0*;*9.0*;* 10.0*;* 10.5*;* 11.0*;* 11.5*;* 12.0*;* 12.5*;* 13.0*;* 13.5}. Set of possible actions i.e. action set for agent 2 is A2 = Price p = {5 to 9} = {5.0*;*6.0*;* 7.0*;* 7.5*;* 8.0*;* 8.5*;* 9.0}. Set of possible actions i.e. action set for agent 3 is A3 = Price p = {10 to 13} = {10.0*;* 10.5*;* 11.0*;* 11.5*;* 12.0*;* 12.5*;* 13.0} [22-23].
(12) The output is the possible action taken i.e. price in this case. It is now the state-action pair system can be easily modeled using Q learning i.e. Q(s, a) [24-25]. Here we need to define the reward calculation.
(13) Reward (r): Reward is calculated in our system as given below: Assume current state $i=(x_i,i_i)$ and next state $j=(x_j,i_j)$. Four transactions are possible for reward calculation. (f1=1/2, f2= 1/3, f3=1/5) [25]

current state i → next state j
Case 1: $[x_i,i_i]\rightarrow[x_i,i_{i-1}]$
That means: one product is sold
Case 2: $[x_i,i_i]\rightarrow[x_{i+1},i_{i-1}]$
That means: one customer request served & one product is sold
Case 3: $[x_i,i_i]\rightarrow[x_i,i_{i-3}]$
That means: Three products are sold "Buy One Get Two"
Case 4: $[x_i,0]\rightarrow[x_{i+1},0]$
That means: new customer request arrived, no stock available.

Depending on above state transitions from current state to next state, reward is calculated as

Reward is $r_p(i, p, j)$ = p if $x`_1 = x_1 + 1$..... Case 4
= p if $i`_1 = i_1 - 1$..... Case 1
= 2p if $i`_1 = i_1 - 3$..... Case 2 & 3
= 0 otherwise

5. Implementation Results of Cooperative Reinforcement Learning Algorithms

The experiments were carried out into environments with dimensions between 120 to 350 states.

5.1. Shop Agent 1

The result of shop agent 1 for the period of one-year sale duration is given below. All the states with zero (0) profit entries in each method are excluded from the resultant table. It shows the profit obtained without cooperation methods (Q-learning) and with cooperative schemes (i.e. group, dynamic, goal driven schemes). By following Q learning method (without cooperation) shop agent 1 cannot obtain the maximum profit. Amount of profit received without cooperation is less as compared to the amount of profit received with cooperation method

The graph in Figure 1 for Shop agent 1 shows that profit margin vs a number of states is given by four methods. Profit obtained by cooperative schemes i.e. group, dynamic and goal driven schemes is much more than that of without

cooperation method for agent 1. Agent 1 obtains more profit by applying cooperation methods i.e. group and dynamic methods throughout the year.

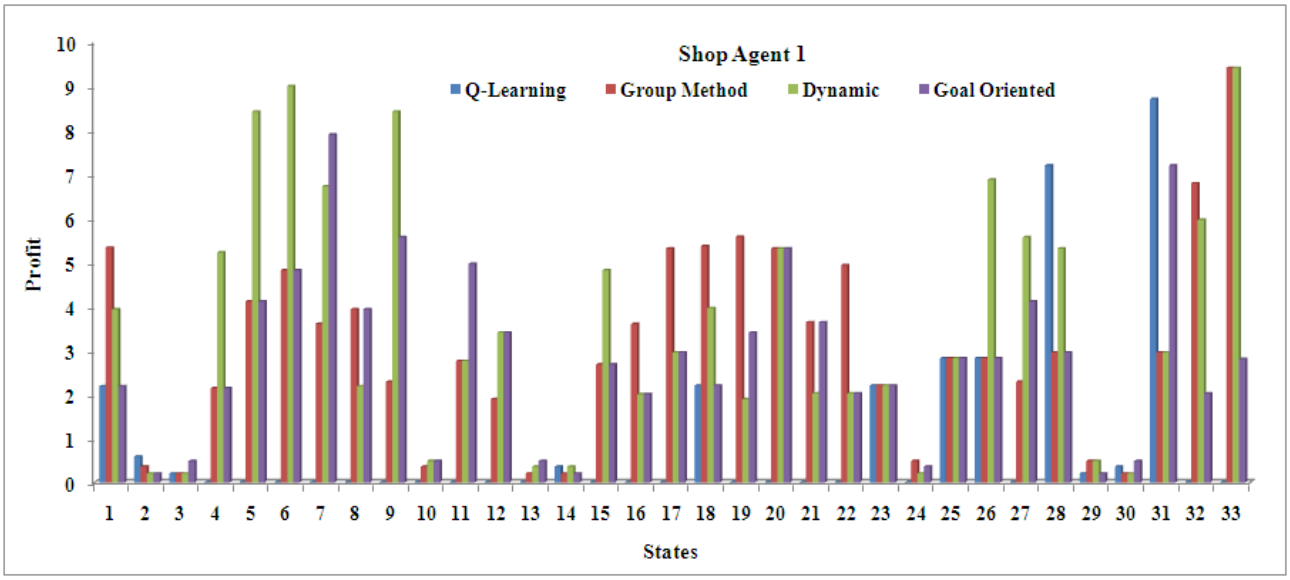

Figure 1. *Graph of Shop agent 1 using with and without cooperation methods.*

5.2. Shop Agent 2

The result of shop agent 2 for the period of one-year sale duration is given below. All the states with zero (0) profit entries in each method are excluded from the resultant table. It shows the profit obtained without cooperation methods (Q-learning) and with cooperative schemes (i.e. group, dynamic, goal driven schemes). By following Q learning method (without cooperation), shop agent 2 cannot obtain the maximum profit. Amount of profit received without cooperation is less as compared to the amount of profit received with cooperation method.

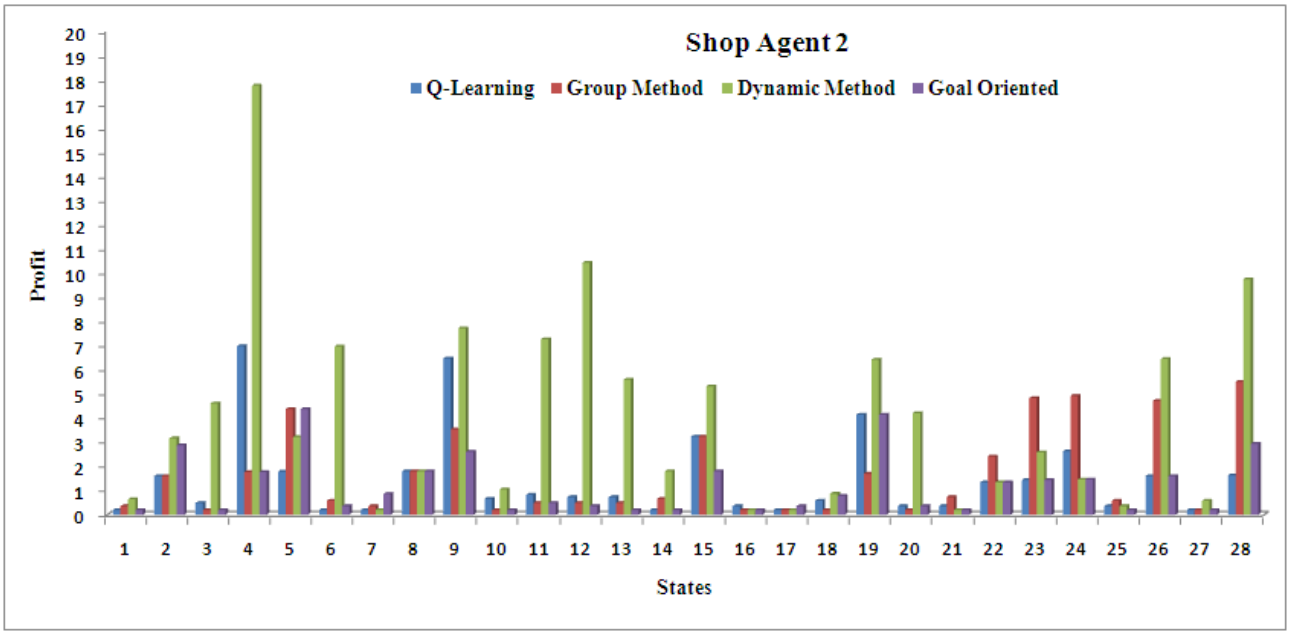

Figure 2. *Graph of Shop agent 2 using with and without cooperation methods.*

The graph in Figure 2 for Shop agent 2 shows that profit margin vs a number of states is given by four methods. Profit obtained by the cooperative schemes i.e. group, dynamic and goal driven schemes is much more than that of without cooperation method i.e. simple Q learning for agent 2. Agent 2 obtains more profit by applying cooperation methods i.e. dynamic methods throughout the year.

5.3. Shop Agent 3

The result of shop agent 3 for the period of one-year sale duration is given below. All the states with zero (0) profit

entries in each method are excluded from the resultant table. It shows the profit obtained without cooperation methods (Q-learning) and with cooperative schemes (i.e. group, dynamic, goal driven schemes). By following Q learning method (without cooperation), shop agent 3 cannot obtain the maximum profit. Amount of profit received without cooperation is less as compared to the amount of profit received with cooperation method.

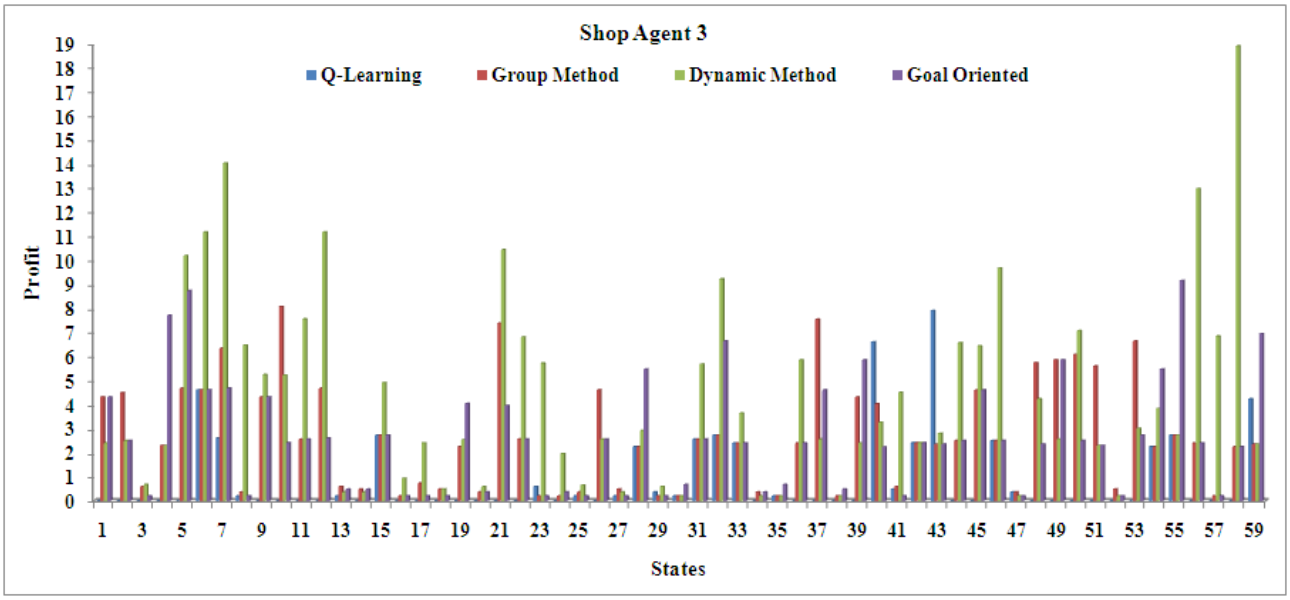

Figure 3. *Graph of Shop agent 3 using with and without cooperation methods.*

The graph in Figure 3 for Shop agent 3 shows that profit margin vs a number of states is given by four methods. Profit obtained by the cooperative schemes i.e. group, dynamic and goal driven schemes is much more than that of without cooperation method i.e. simple Q learning for agent 3. Agent 3 obtains more profit by applying cooperative schemes i.e. dynamic and goal driven schemes throughout the year.

6. Implementation Results of Expertise based Multi-agent Cooperative Reinforcement Learning Algorithms (EMCRLA)

The experiments were carried out into environments with dimensions between 120 to 350 states.

6.1. Shop Agent 1

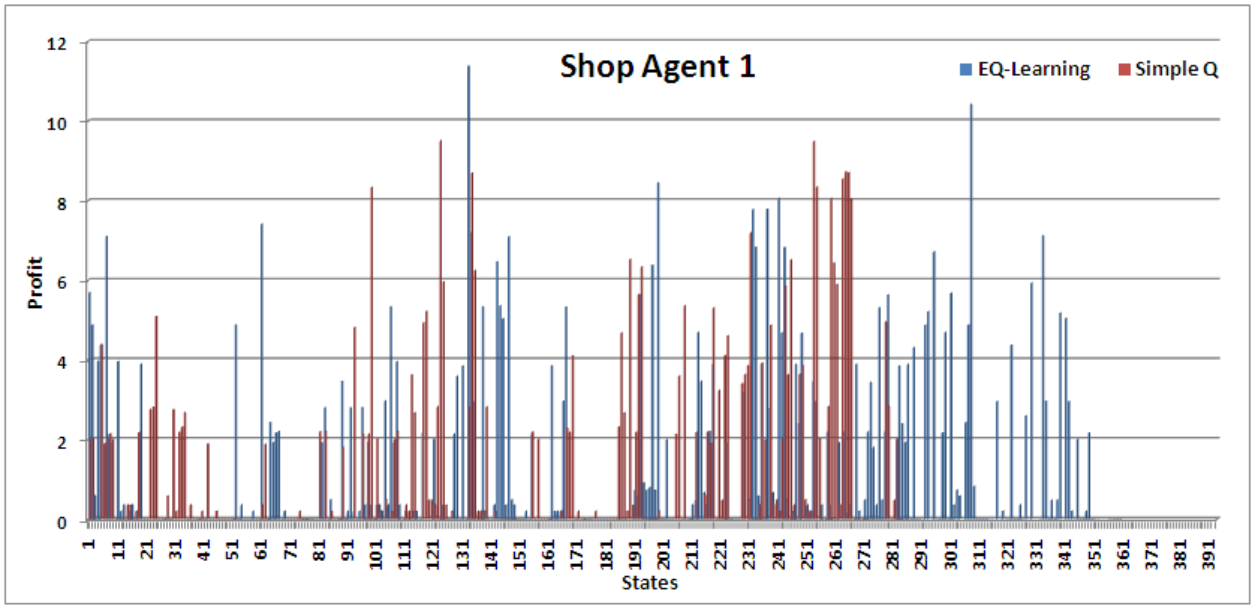

Figure 4. *Graph of Shop agent 1 using simple Q-Learning and EQ-Learning methods.*

The result of shop agent 1 for the period of one-year sale duration using proposed cooperative expertness methods is given below. The graph in Figure 4 for Shop agent 1 describes the comparison between simple Q learning and proposed expertness based Q learning (EQ-Learning) algorithms. It shows that expertness based Q learning algorithm gives better results in terms of profit vs states as compared to simple Q learning algorithm.

The graph in Figure 5 for Shop agent 1 describes the comparison between simple group learning and proposed expertness based group learning (EGroup) method. It shows that expertness based group learning algorithm gives better results in terms of profit vs states as compared to simple group method.

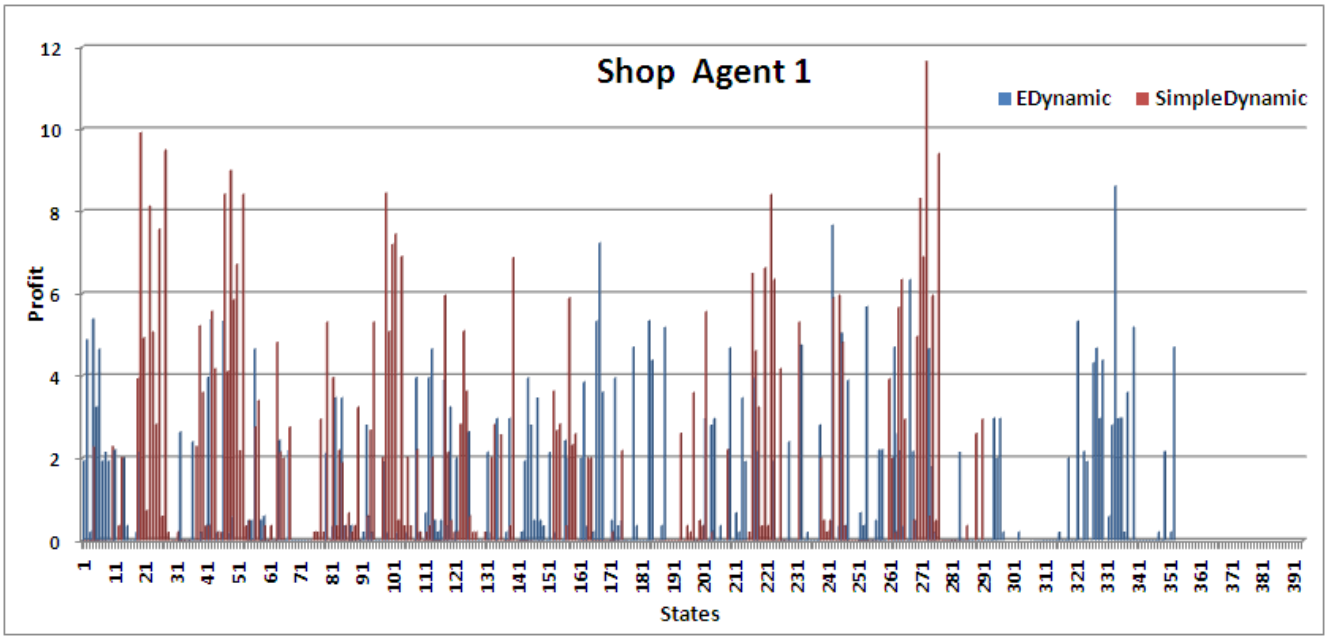

Figure 5. *Graph of Shop agent 1 using simple Group learning and EGroup Learning methods.*

The graph in Figure 6 for Shop agent 1 describes the comparison between simple dynamic learning method and proposed expertness based dynamic learning (EDynamic) method. It shows that expertness based dynamic learning algorithm gives better results in terms of profit vs states as compared to the simple dynamic method.

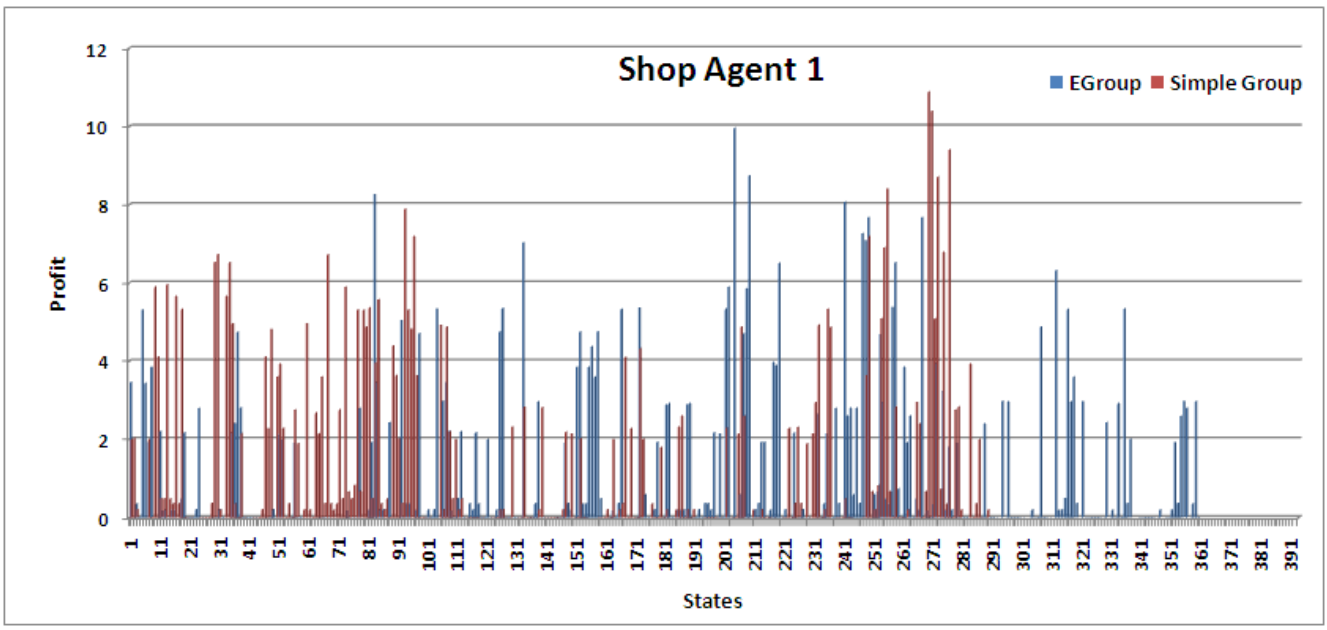

Figure 6. *Graph of Shop agent 1 using simple Dynamic learning and EDynamic Learning methods.*

The graph in Figure 7 of Shop agent 1 describes the comparison between simple goal-driven learning method and proposed expertness based goal-driven learning (EGoal) method. It shows that expertness based goal-driven learning algorithm gives

better results in terms of profit vs states as compared to the goal-driven method.

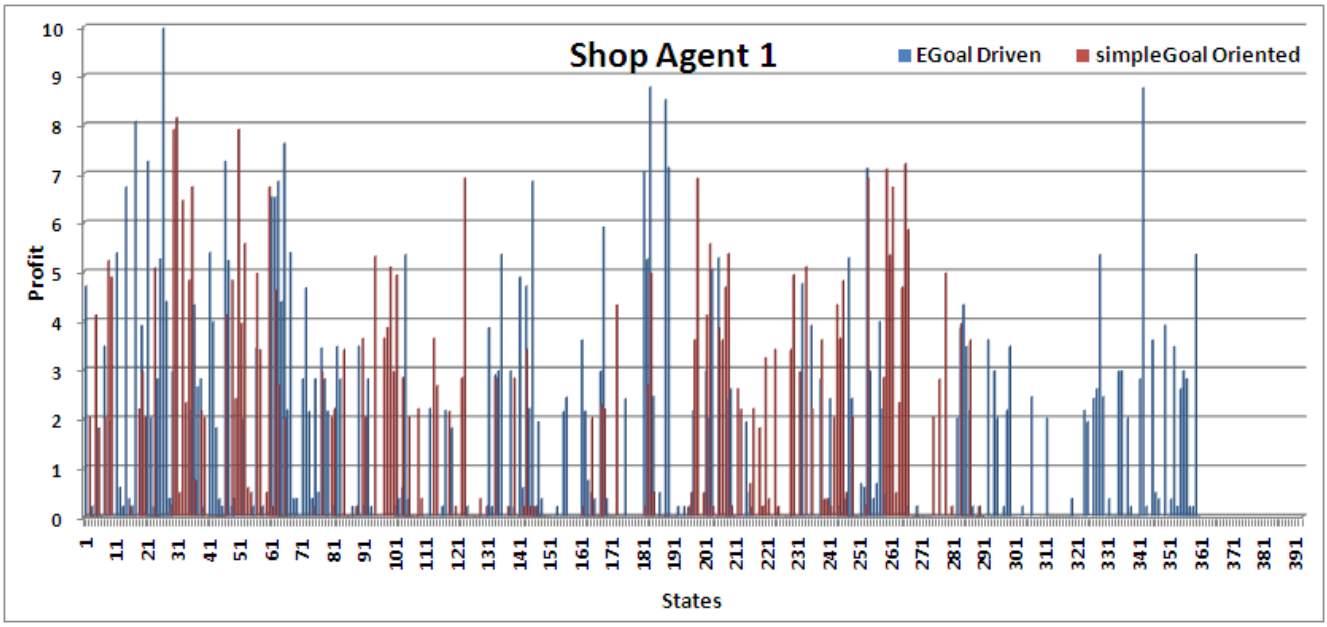

Figure 7. *Graph of Shop agent 1 using simple Goal Driven and EGoal Driven Learning methods.*

6.2. Shop Agent 2

The result of shop agent 2 for the period of one-year sell duration using proposed cooperative expertness methods is given below. The graph in Figure 8 for Shop agent 2 describes the comparison between simple Q learning and proposed expertness based Q learning (EQ-Learning) algorithms. It shows that expertness based Q learning algorithm gives better results in terms of profit vs states as compared to simple Q learning algorithm.

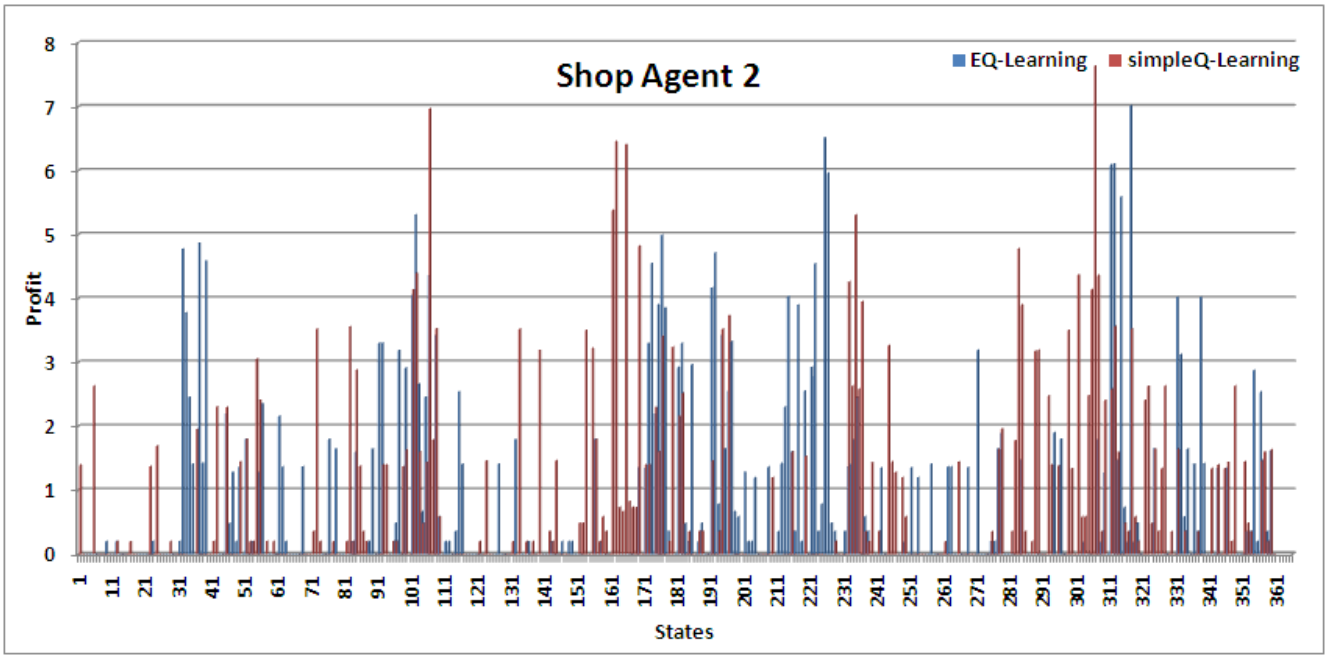

Figure 8. *Graph of Shop agent 2 using simple Q-Learning and EQ-Learning methods.*

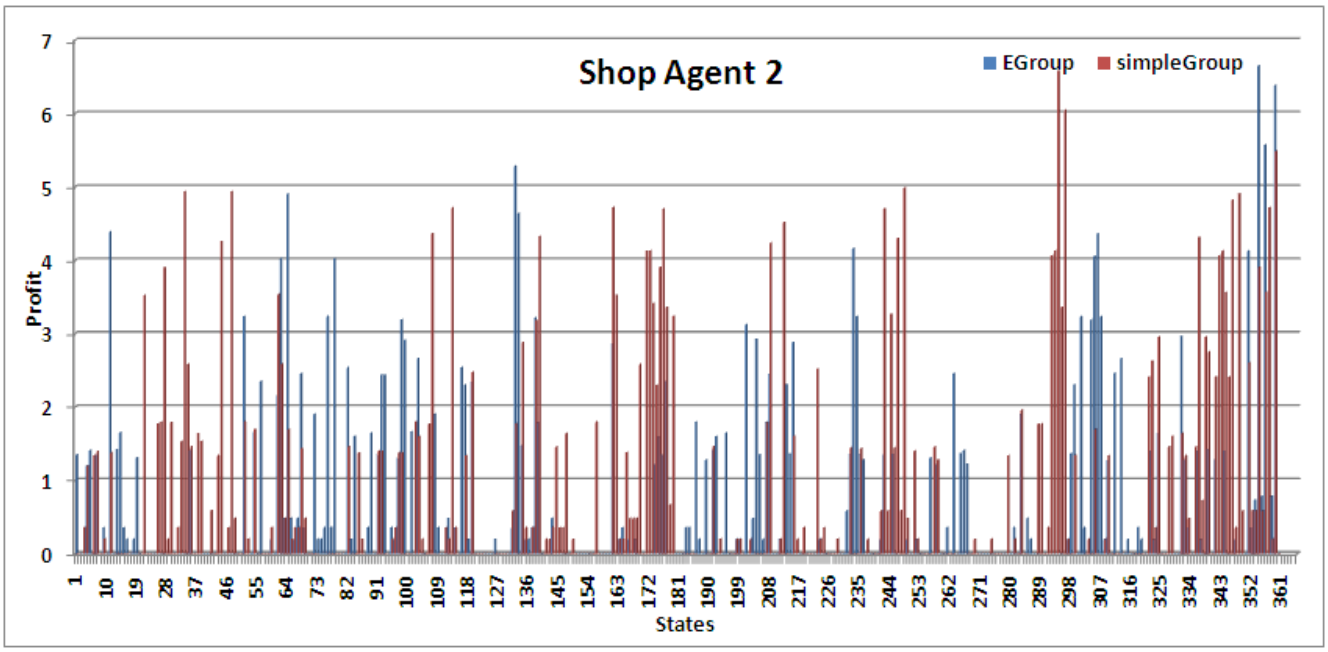

Figure 9. *Graph of Shop agent 2 using simple Group learning and EGroup learning methods.*

The graph in Figure 9 for Shop agent 2 describes the comparison between simple group learning and proposed an expertness based group learning (EGroup) method. It shows that expertness based group learning algorithm gives better results in terms of profit vs states as compared to simple group method.

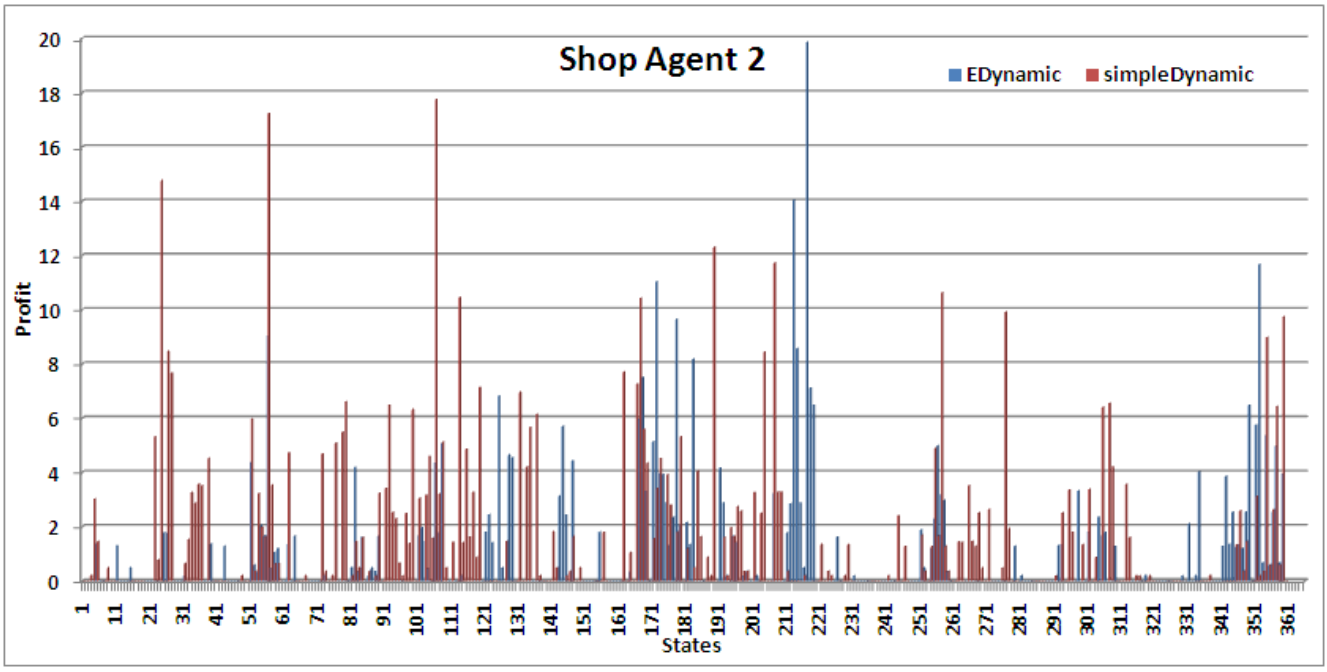

Figure 10. *Graph of Shop agent 2 using simple Dynamic learning and EDynamic learning methods.*

The graph in Figure 10 for Shop agent 2 describes the comparison between simple dynamic learning method and proposed expertness based dynamic learning (EDynamic) method. It shows that expertness based dynamic learning algorithm gives better results in terms of profit vs states as compared to the simple dynamic method.

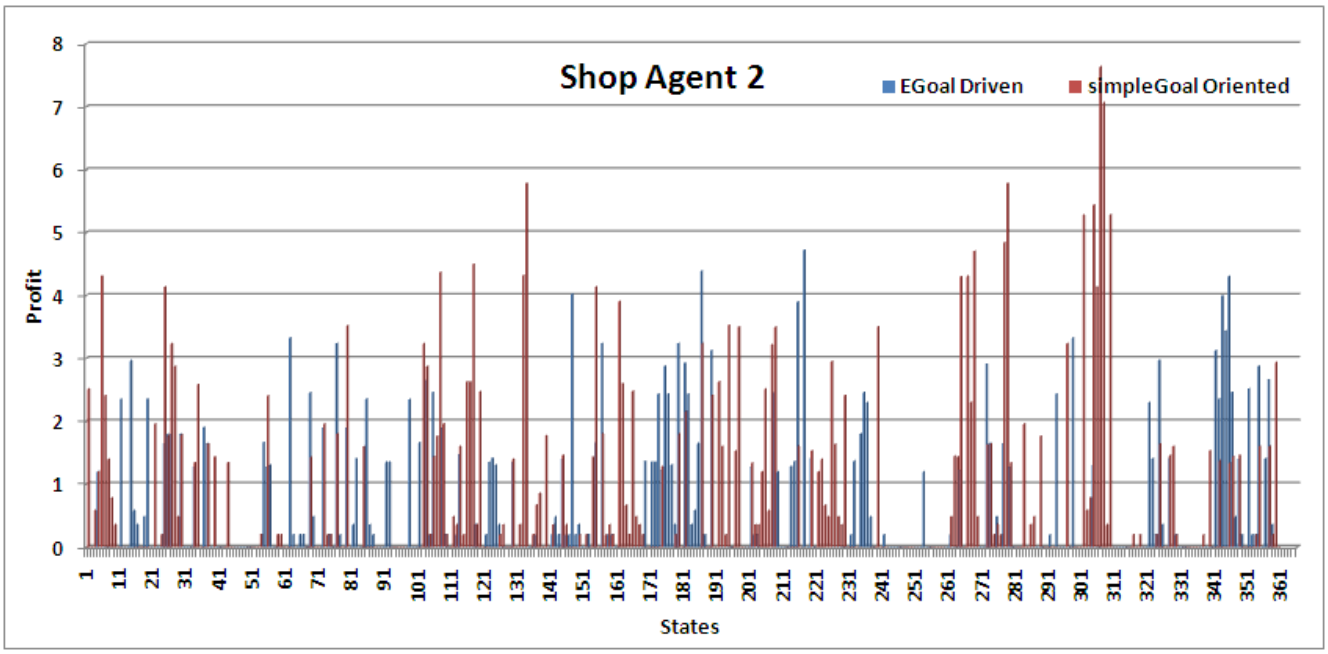

Figure 11. *Graph of Shop agent 2 using simple Goal Driven and EGoal Driven learning methods.*

The graph in Figure 11 of Shop agent 2 describes the comparison between simple goal-driven learning method and proposed expertness based goal-driven learning (EGoal) method. It shows that expertness based goal-driven learning algorithm gives better results in terms of profit vs states as compared to the goal-driven method.

6.3. Shop Agent 3

The result of shop agent 3 for the period of one-year sell duration using proposed cooperative expertness methods is given below. The graph in Figure 12 for Shop agent 3describes the comparison between simple Q learning and proposed expertness based Q learning (EQ-Learning) algorithms. It shows that expertness based Q learning algorithm gives better results in terms of profit vs states as compared to simple Q learning algorithm.

The graph in Figure 12 for Shop agent 3 describes the comparison between simple group learning and proposed an expertness based group learning (EGroup) method. It shows that expertness based group learning algorithm gives better results in terms of profit vs states as compared to simple group method.

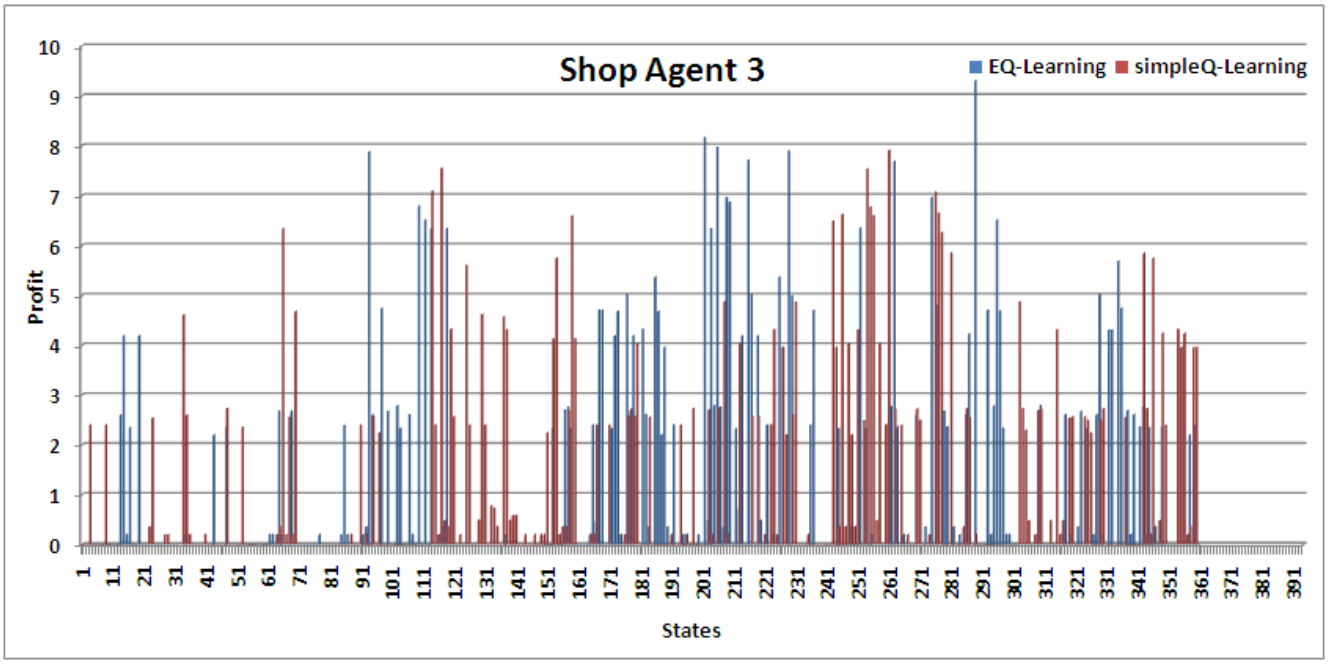

Figure 12. *Graph of Shop agent 3 using simple Q-Learning and EQ-Learning methods.*

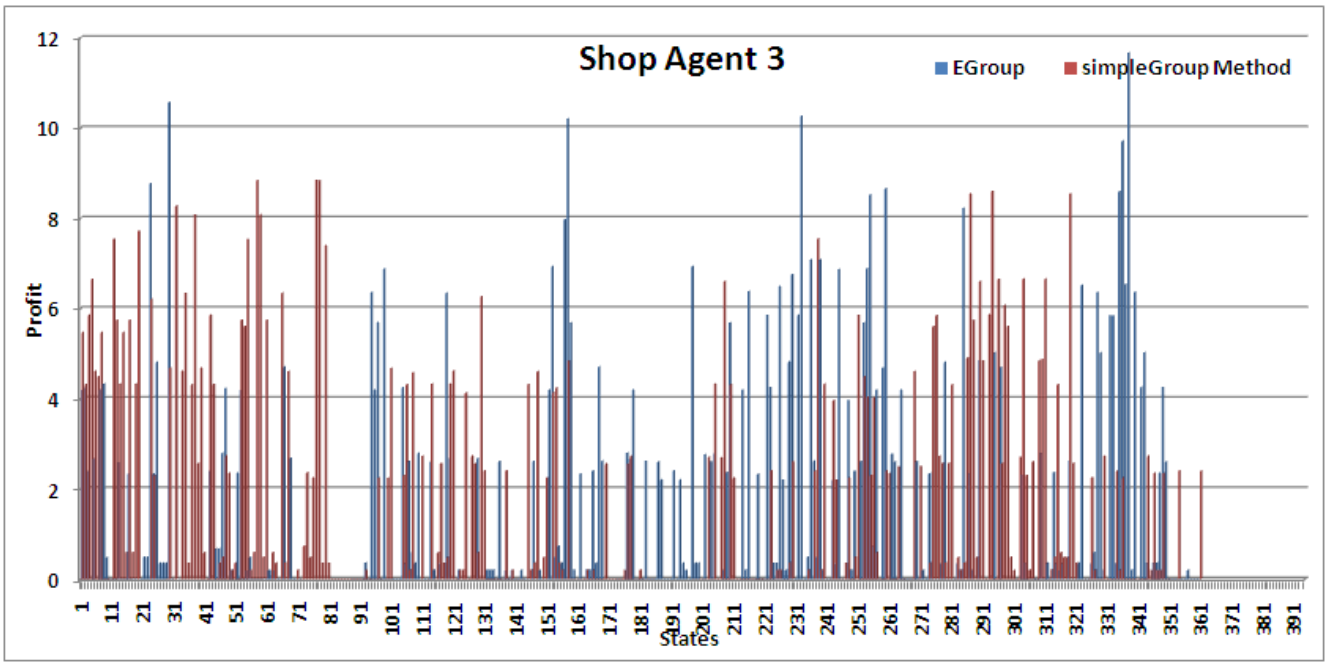

Figure 13. *Graph of Shop agent 3 using simple Group learning and EGroup learning methods.*

The graph in Figure 14 for Shop agent 3 describes the comparison between simple dynamic learning method and proposed expertness based dynamic learning (EDynamic) method. It shows that expertness based dynamic learning algorithm gives better results in terms of profit vs states as compared to the simple dynamic method.

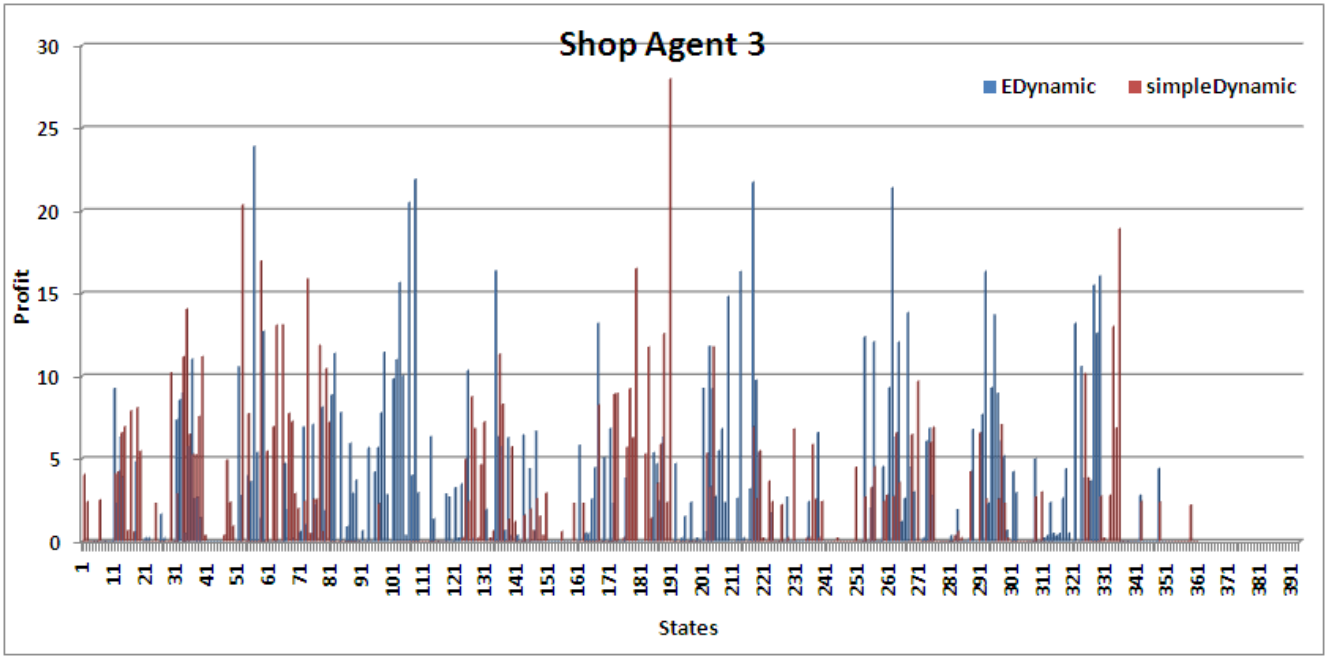

Figure 14. *Graph of Shop agent 3 using simple Dynamic learning and EDynamic Learning methods.*

The graph in Figure 15 of Shop agent 3 describes the comparison between simple goal-driven learning method and proposed expertness based goal-driven learning (EGoal) method. It shows that expertness based goal-driven learning algorithm gives better results in terms of profit vs states as compared to the goal-driven method.

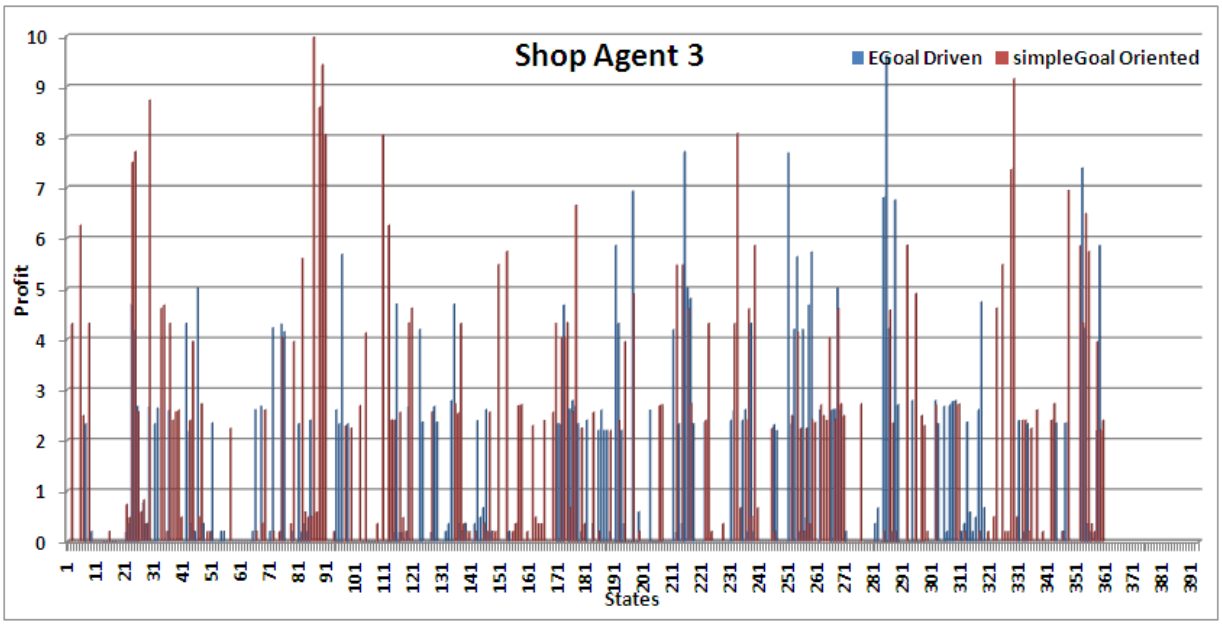

Figure 15. *Graph of Shop agent 3 using simple Goal Driven and EGoal Driven Learning methods.*

7. Result Analysis of Cooperative Reinforcement Learning Algorithms

During one year period, for agent 1, dynamic method, group method, goal driven method and Q-learning method gives good profit as per the decreasing order. During one year period, for agent 2, dynamic method, group method, goal driven method and Q-learning method gives good profit as per the decreasing order. During one year period, for agent 3, dynamic method, goal driven method, group method and Q-learning method give good profit as per the decreasing order.

Table 1. *Yearly, Half Yearly & Quarterly Profit obtained by with and without cooperation methods for Shop Agent 1.*

Period	Profit without CL (Simple QL)	Profit with Cooperation by three CL Methods		
		Group	Dynamic	Goal Driven
One Year	6.58	7.64	17.74	7.64
Half Year 1	4.94	6.97	17.74	5.78
Half Year 2	7.64	6.58	12.3	7.64
Quarter 1	3.56	4.94	17.22	4.32
Quarter 2	5.78	4.72	17.74	6.97
Quarter 3	4.72	4.99	12.31	5.31
Quarter 4	7.64	6.58	9.91	7.64

Table 2. *Yearly, Half Yearly & Quarterly Profit obtained by with and without cooperation methods for Shop Agent 2.*

Period	Profit without CL (Simple QL)	Profit with Cooperation by three CL Methods		
		Group	Dynamic	Goal Driven
One Year	8.13	10.9	11.65	9.51
Half Year 1	8.13	7.89	9.91	9.51
Half Year 2	7.19	10.9	11.65	9.51
Quarter 1	5.09	6.72	9.91	8.13
Quarter 2	6.91	7.89	8.45	9.51
Quarter 3	6.89	4.88	6.87	8.71
Quarter 4	9.51	10.9	11.65	7.19

Table 3. *Yearly, Half Yearly & Quarterly Profit obtained by with and without cooperation methods for Shop Agent 3.*

Period	Profit without CL (Simple QL)	Profit with Cooperation by three CL Methods		
		Group	Dynamic	Goal Driven
One Year	6.58	7.64	17.74	7.64
Half Year 1	4.94	6.97	17.74	5.78
Half Year 2	7.64	6.58	12.3	7.64
Quarter 1	3.56	4.94	17.22	4.32
Quarter 2	5.78	4.72	17.74	6.97
Quarter 3	4.72	4.99	12.31	5.31
Quarter 4	7.64	6.58	9.91	7.64

During half year period, for agent 1, dynamic method and group method give good profit as compared to goal driven and Q-learning method. During half year period, for agent 2, dynamic method and goal driven method give good profit as compared to group and Q-learning method. During half year period, for agent 3, dynamic method and goal driven method give good profit as compared to group and Q-learning method. During the quarterly period, all the agents 1, agent 2 and agent 3 get good profit from the dynamic method.

7.1. Result Analysis of Expertise Based Multiagent Cooperative Reinforcement Learning Algorithms (EMCRLA)

During one year period, for agent 1, expertness based dynamic method, expertness based group method, and Q-learning method gives good profit as per the decreasing order. New method proposed expert agent gives satisfactory results as listed in Table 4, Table 5 and Table 6 for Shop Agent 1, Shop Agent 2 and Shop Agent 3 respectively.

For Shop Agent 1, it is clear from Table 4, that for one year duration profit obtained without cooperation (EQL) method is moderate compared to profit with cooperation by expert methods i.e EGroup, EDynamic, EGoal Driven. The profit range (lowest & highest) for four expertness based cooperative methods are given as: for EGroup is 10.21 to 11.67, for EDynamic is 16.32 to 23.91, for EGoal Driven is 3.33 to 7.11. The profit range obtained by without cooperation method EQL is 4.19 to 9.29.

Table 4. *Yearly & Quarterly Profit obtained by with and without cooperation expertness methods for Shop Agent 1.*

Period	Profit without CL (EQL)	Profit with Cooperation by two CL Expert Methods		
		EGroup	EDynamic	EGoal Driven
One Year	9.29	11.67	23.91	7.11
Quarter 1	4.19	10.57	23.91	3.33
Quarter 2	7.88	10.21	21.89	4.02
Quarter 3	8.17	10.27	21.73	4.72
Quarter 4	9.28	11.67	16.32	4.31

For Shop Agent 2, it is understood from Table 5, that for one year duration profit obtained without cooperation (EQL) method is reasonable as compared to profit with cooperation by expert methods i.e EGroup, EDynamic, EGoal Driven. The profit range (lowest & highest) for four expertness based cooperative methods are given as: for EGroup is 6.32 to 9.96, for EDynamic is 5.38 to 11.38, for EGoal Driven is 5.01 to 9.63. The profit range obtained by without cooperation method EQL is 8.45 to 11.38.

Table 5. *Yearly & Quarterly Profit obtained by with and without cooperation expertness methods Shop Agent 2.*

Period	Profit without CL (EQL)	Profit with Cooperation by two CL Expert Methods		
		EGroup	EDynamic	EGoal Driven
One Year	8.61	9.96	11.38	9.63
Quarter 1	7.41	8.26	5.38	5.01
Quarter 2	11.38	7.03	7.23	5.69
Quarter 3	8.45	9.96	7.67	7.71
Quarter 4	10.42	6.32	8.61	9.63

For Shop Agent 3, it is understood from Table 6, that for one year duration profit obtained without cooperation (EQL) method is reasonable as compared to profit with cooperation by expert methods i.e EGroup, EDynamic, EGoal Driven. The profit range (lowest & highest) for four expertness based cooperative methods are given as: for EGroup is 4.17 to 6.65, for EDynamic is 9.06 to 19.86, for EGoal Driven is 7.03 to 9.96. The profit range obtained by without cooperation method EQL is 4.18 to 7.01.

Table 6. *Yearly & Quarterly Profit obtained by with and without cooperation expertness methods Shop Agent 3.*

Period	Profit without CL (EQL)	Profit with Cooperation by two CL Expert Methods		
		EGroup	EDynamic	EGoal Driven
One Year	4.18	6.65	19.86	9.96
Quarter 1	7.01	4.91	9.06	9.96
Quarter 2	5.31	5.29	11.02	7.03
Quarter 3	6.52	4.17	19.86	8.76
Quarter 4	7.01	6.65	11.68	8.75

8. Conclusion

It claims that reinforcement learning methods with cooperation outperform those without cooperation and expertise based cooperative learning algorithms are surely enhance the performance of cooperative learning algorithm. The paper illustrates results of Cooperative Reinforcement Learning Algorithms of three shop agents for the period of one-year sale duration. Profit obtained without cooperation methods (Q-learning) and with cooperative schemes (i.e. EGroup, EDynamic, EGoal driven schemes) is calculated. By

following without cooperation methods shop agents cannot obtain the maximum profit. Amount of profit received without cooperation methods is less as compared to the amount of profit received with cooperation methods. Graphical results show profit margin vs a number of states for four methods. The paper also demonstrated the results using proposed approach i.e. Expertise based Multi-agent Cooperative Reinforcement Learning Algorithms (EMCRLA) for three shop agents for the period of one-year sale duration. Expertness based Q learning (EQ-Learning) method presents improved results in comparison with simple Q learning method in profit vs states. The expertness based group learning (EGroup) method presents improved results in comparison with simple group method in profit vs states. The expertness based dynamic learning (EDynamic) method presents improved results in comparison with a simple dynamic method in terms of profit vs states. Expertness based goal-driven learning (EGoal) method presents improved results in comparison with a goal-driven method in profit vs states. Comparison between expertness based cooperative methods and without expertness based cooperative method for the period of one year is calculated. In more than 70% months the proposed methods i.e. cooperation with expertness gives better results than without expertness. The results obtained by the proposed expertise based cooperation methods show that such methods can put into a quick convergence of agents in the dynamic environment. It also shows that cooperative methods give a good presentation in dense, incompletely and composite circumstances.

References

[1] Deepak A. Vidhate and Parag Kulkarni, "Expertise Based Cooperative Reinforcement Learning Methods (ECRLM)", *International Conference on Information & Communication Technology for Intelligent System, Springer book series Smart Innovation, Systems and Technologies*(SIST, volume 84), Cham, pp 350-360, 2017.

[2] Abhijit Gosavi, "Simulation-based Optimization: Parametric Optimization Techniques and Reinforcement Learning" *Kluwer Academic Publishers*, 2003.

[3] Andrew Y. Ng, "Sharding and Policy Search in Reinforcement Learning", *Ph.D. dissertation. The* University of California, Berkeley, 2003.

[4] Deepak A Vidhate and Parag Kulkarni, "Enhanced Cooperative Multi-agent Learning Algorithms (ECMLA) using Reinforcement Learning" *International Conference on Computing, Analytics and Security Trends* (CAST), IEEE Xplorer, pp 556 - 561, 2017.

[5] Antanas Verikas, Arunas Lipnickas, Kerstin Malmqvist, Marija Bacauskiene, and Adas Gelzinis, *"*Soft Combination of Neural Classifiers: A Comparative Study*"*, *Pattern Recognition Letters* No. 20, 1999, pp429-444.

[6] Deepak A. Vidhate and Parag Kulkarni "Innovative Approach Towards Cooperation Models for Multi-agent Reinforcement Learning (CMMARL) "*International Conference on Smart Trends for Information Technology and Computer Communications* Springer, Singapore, 2016 pp. 468-478.

[7] Babak Nadjar Araabi, Sahar Mastoureshgh, and Majid Nili Ahmadabadi "A Study on Expertise of Agents and Its Effects on Cooperative Q-Learning" *IEEE Transactions on Evolutionary Computation, vol:14, pp:23-57*, 2011.

[8] C. J. C. H. Watkins and P. Dayan, "Q-learning", Machine Learning, 8 (3): 1992.

[9] Deepak A. Vidhate and Parag Kulkarni "New Approach for Advanced Cooperative Learning Algorithms using RL methods (ACLA)" *VisionNet'16 Proceedings of the Third International Symposium on Computer Vision and the Internet, ACM DL* pp 12-20, 2016.

[10] Deepak A Vidhate and Parag Kulkarni, Parag "Enhancement in Decision Making with Improved Performance by Multi-agent Learning Algorithms" *IOSR Journal of Computer Engineering, Vol. 1,No. 18,*pp 18-25,2016.

[11] Ju Jiang and Mohamed S. Kamel *"*Aggregation of Reinforcement Learning Algorithms", *International Joint Conference on Neural Networks,* Vancouver, Canada July 16-21, 2006.

[12] Lun-Hui Xu, Xin-Hai Xia and Qiang Luo "The Study of Reinforcement Learning for Traffic Self-Adaptive Control under Multi-agent Markov Game Environment"*, Mathematical Problems in Engineering,* Hindawi Publishing Corporation, Volume 2013.

[13] Deepak A. Vidhate and Parag Kulkarni, "Implementation of Multi-agent Learning Algorithms for Improved Decision Making", *International Journal of Computer Trends and Technology (IJCTT),* Volume 35 Number 2- May 2016.

[14] Lun-Hui Xu, Xin-Hai Xia and Qiang Luo "The Study of Reinforcement Learning for Traffic Self-Adaptive Control under Multi-agent Markov Game Environment" *Hindawi Publishing Corporation, Mathematical Problems in Engineering, Volume* 2013.

[15] Deepak A Vidhate and Parag Kulkarni, "Performance enhancement of cooperative learning algorithms by improved decision-making for context-based application", *International Conference on Automatic Control and Dynamic Optimization Techniques (ICACDOT) IEEE Xplorer*, pp 246-252, 2016.

[16] Deepak A. Vidhate and Parag Kulkarni, "Design of Multi-agent System Architecture based on Association Mining for Cooperative Reinforcement Learning", *Spvryan's International Journal of Engineering Sciences & Technology (SEST),* Volume 1, Issue 1, 2014.

[17] M. Kamel and N. Wanas, "Data Dependence in Combining Classifiers", *Multiple Classifiers Systems, Fourth International Workshop,* Surrey, UK, June 11-13, pp1-14, 2003.

[18] V. L. Raju Chinthalapati, Narahari Yadati, and Ravikumar Karumanchi, "Learning Dynamic Prices in Multi-Seller Electronic Retail Markets With Price Sensitive Customers, Stochastic Demands, and Inventory Replenishments", *IEEE Transactions On Systems, Man, And Cybernetics—Part C: Applications And Reviews*, Vol. 36, No. 1, January 2008.

[19] Deepak A. Vidhate and Parag Kulkarni, "Multilevel Relationship Algorithm for Association Rule Mining used for Cooperative Learning*", International Journal of Computer Applications* (0975 – 8887), volume 86, number 4, pp 20--27, 2014.

[20] Y. S. Huang and C. Y. Suen, "A method of combining multiple experts for the recognition of unconstrained handwritten numerals." *IEEE Trans. on Pattern Analysis and Machine Intelligence* 17(1), 1995, pp90-94.

[21] Deepak A. Vidhate, Parag Kulkarni, "To improve association rule mining using new technique: Multilevel relationship algorithm towards cooperative learning", *International Conference on Circuits, Systems, Communication and Information Technology Applications (CSCITA)*, IEEE pp 241—246, 2014.

[22] Young-Cheol Choi, Student Member, Hyo-Sung Ahn "A Survey on Multi-Agent Reinforcement Learning: Coordination Problems", *IEEE/ASME International Conference on Mechatronics and Embedded Systems and Applications,* pp. 81–86, 2010.

[23] Deepak A Vidhate, Parag Kulkarni, "A Novel Approach to Association Rule Mining using Multilevel Relationship Algorithm for Cooperative Learning" *Proceedings of 4th International Conference on Advanced Computing & Communication Technologies (ACCT-2014),* pp 230-236, 2014.

[24] Zahra Abbasi, Mohammad Ali Abbasi "Reinforcement Distribution in a Team of Cooperative Q-learning Agent", *Proceedings of the 9th ACIS International Conference on Artificial Intelligence, Distributed Computing, IEEE* 2012.

[25] Deepak A Vidhate, Parag Kulkarni, "Cooperative machine learning with information fusion for dynamic decision making in diagnostic applications", *International Conference on Advances in Mobile Network, Communication and its Applications (MNCAPPS),IEEE*, pp 70-74, 2012.