International Journal of Engineering & Technology, 7 (1.5) (2018) 269-273

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Implementation of modified Q learning technique in EMCAP control architecture

D. Ganesha[1], Vijayakumar Maragal Venkatamuni[2]

[1]Bharathiar University, Department of ISE, PVP Polytechnic,
Dr.AIT campus Outer Ring Road, Malathahalli, Nagarabhavi, Bangalore – 560056, Karnataka, India;
[2]Department of Computer Science, Research Progress Review Committee[RPRC],
Dr. Ambedkar Institute of Technology, Visvesvaraya Technological University,
Bengaluru – 560056, Karnataka, India;
**Corresponding author E-mail: ganesh207d@gmail.com*

Abstract

This research introduces a self learning modified (Q-Learning) techniques in a EMCAP (Enhanced Mind Cognitive Architecture of pupils). Q-learning is a modelless reinforcement learning (RL) methodology technique. In Specific, *Q*-learning can be applied to establish an optimal action-selection strategy for any respective Markov decision process. In this research introduces the modified Q-learning in a EMCAP (Enhanced Mind Cognitive Architecture of pupils). EMCAP architecture [1] enables and presents various agent control strategies for static and dynamic environment. Experiment are conducted to evaluate the performace for each agent individually. For result comparison among different agent, the same statistics were collected. This work considered varied kind of agents in different level of architecture for experiment analysis. The Fungus world testbed has been considered for experiment which is has been implemented using SwI-Prolog 5.4.6. The fixed obstructs tend to be more versatile, to make a location that is specific to Fungus world testbed environment. The various parameters are introduced in an environment to test a agent's performance.his modified q learning algorithm can be more suitable in EMCAP architecture. The experiments are conducted the modified Q-Learning system gets more rewards compare to existing Q-learning.

Keywords: *Self learning, Cognitive Control, Q Learning.*

1. Introduction

The machine learning has classified into supervised and un supervised and Reinforcement learning, the reinforcement learning is expressed in artificial intelligence cognitive control in the form of the computational representations to manage systems being robotic [2]. This is a model learnt by benefits or charges through particular stimulating requests for handling complex structures, so that can optimize numerical performance that reveals a goal known as durable. Current results also have exposed just how critical information supplies the intellectual controller using the information required in the atmosphere to trigger the system by using q-learning [3]. The reinforcement learning algorithm had been used to cope with Process industrial [4], [5] goal guideline in identity evolutionary modern organization and modern execution organization in the industrial procedures. In [6] the authors projected an interoperable well-informed dynamic development system predicated on multi-gent, where in an adaptive scheduling device in line with the national account grade weighted Q-learning for directing the device agent towards choose scheduling is suitable in a robust environment , so that can be deal with the doubt of the modern environment in modern systems. This paper is devoted to the algorithm called q-learning. The explanation is the foremost, choice could be the easiness of their technique, its nature is model less also effective expositions of Q learning methods currently described within the works [7]. Q-learning finds an action and selection optimized for virtually by any Markov decision process [8]. It works by learning a function called action-value finally produces the predictable energy of experiencing a supplied action in a supplied state plus after it follows the perfect policy [8] know as optimal. Whenever the sort of function called action-value learned the policies which are suitable built by just choosing the action with the best principles in every state.

The paper work is arranged in four units. Following an introduction, unit II labels the Artificial Cognitive Design Emcap. The RL technique is presented in unit III, particularly, the anticipated customizations connected to the modified Q learning procedure. IV describes an experimental simulation linked to the proposed algorithm performed in prolog computer software. Assumptions and effort are further drawn into location at latter.

2. Artificial cognitive architecture EMCAP

Reinforcement devices that stand learning had a need towards rival genuine functionalities of intellectual schemes. The task described in this paper is concentrated on scheming and employing a reinforcement that can be a strategy is raised is suitable functionality of artificial architecture called intellectual.

The appearance of the self-learning system might bring advantages in regards to the other control is biologically-inspired. Making use of the capability is self enhance in designs which may be used is learning of self competences such as self association, self adaptation, and the self optimization. The architecture comprises of cognitive and amounts that are done. The points that can easily be main the EMCAP architecture is found in [1].

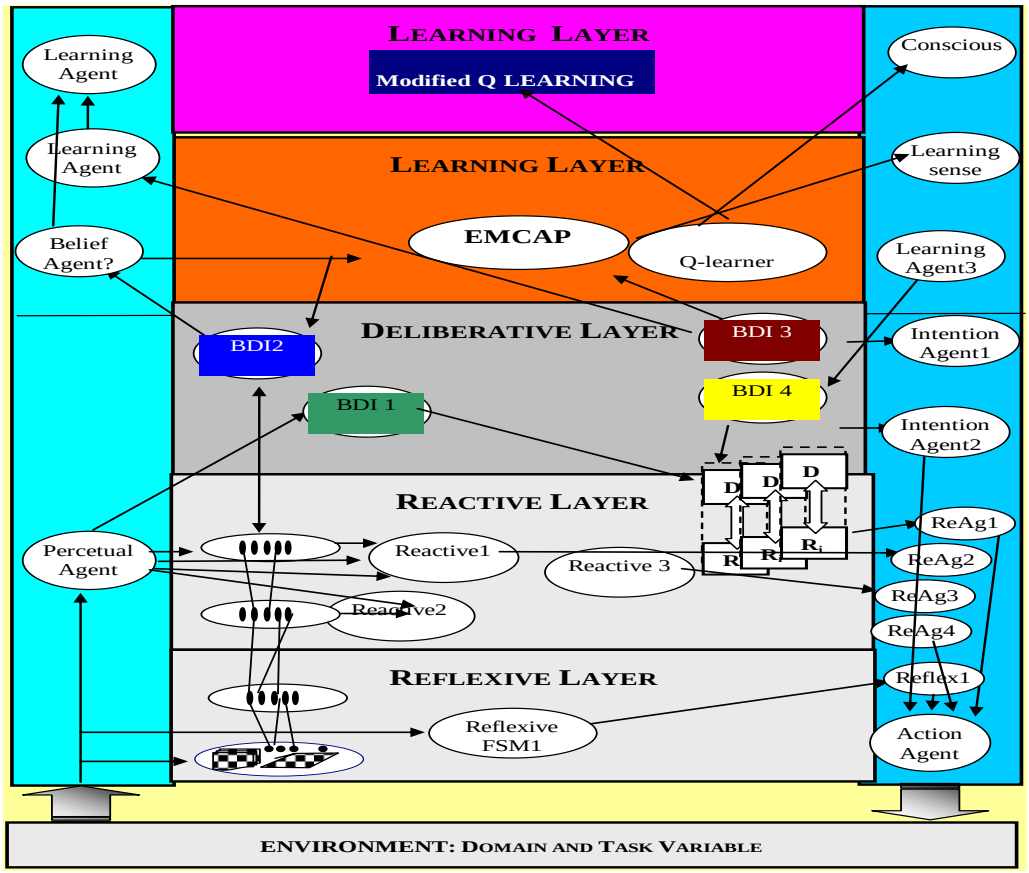


Fig. 1: Emcap modified q learning architecture

3. Modified Q learning algorithm

The fundamental Q learning technique plus and the alterations familiarized to enhance the outcome also to enable the implementation and near optimal operations are labeled in this part. The Markov choice issue contains a real agent, a real amount of S states and a couple of actions for each state A. By doing an action, a €A, the representative can go from a single state to a different. Performing an action in circumstances is certainly the representative by having a reward. The aim of the representative is always to optimize its cumulative reward. It will, by learning the action is most beneficial for every state. Consequently a function is held by the algorithm which determines the caliber of a condition action grouping

Q: S x A→R

Before initializing of learning, Q can get back any parameter is fixed opted for by the design regarding the issue. Before every instance the representative chosses an action; it gets its benefits and comes into their state is brand new. The essential associated with algorithm is just a value is straightforward enhanced. It considers the parameter is old applies a modification based on the knowledge is brand new:

$$Q_t \square 1 \square s_t, a_t \square \square Q_t \square s_t, a_t \square \square \square_t (s_t, a_t)(R_t \square 1 \square \square \max Q_t \square s_t \square 1, a_t \square \square Q_t (s_t, a_t)) \, a \square A \quad (1)$$

where, St is the constant state with instance t, it may be the action drawn in time t, R (t+1) is the bonus received after doing an action at a, could be the rate of learning and γ is the reward element which craft from the importance of earlier versus late on benefits. It could be well-known that Q-learning does perhaps not stipulate a strategy to choose the action to do in every state. However, there could be numerous strategies to choose an action, i.e., the renowned policies which are greedy algorithm 1.

Typically, Q-learning is performed within an occasional way where an incident concludes whenever state S (t+1) is a declare last. But, q-learning can discover in non incident additionally tasks. To be able to utilize Q-learning within the system environment, is intellectual factors have already remained occupied. Primarily, these factors are with respected to this is for the ideas of state and action together with bonus operation algorithm 2. First off most, each consistent state happens to be mapped by having a group of values for the systems, hence every state is recognized clearly having a pair of constraints values:

$$S_t \leftrightarrow (k_1^{(t)}, K_2^{(t)}, ...k_N^{(t)}) \quad (2)$$

Therefore, those things to alter from a state to some other are the ones that modify one or more parameter regarding the set. Hence the stable scope of the factors needs to be discredited as currently reported in [9] so that we can have the ability to utilize Q-learning with latest concept of state. Therefore, each parameter Ki has its restrictions, which can be very own by the model to that parameter belongs. Then, if found M feasible values of each and every parameter, the number of values with this constraint value is

$$k_{i_1} = k_i^{\min}, k_{i_2} = k_{i_1} + \frac{k_i^{\min} - k_i^{\max}}{M-1}, ...k_{i_M} = k_i^{\max} \quad (3)$$

As an outcome of the above capacity, the space of states with a measurement of MN is limited, so it can be utilized as a part of Q-learning.

It is assumed for a given state $s_t \leftrightarrow \left(k_1^t, k_2^t, \ldots\ldots k_N^t\right)$, its accessible activities will be those that change st to $s_{t+1} = \leftrightarrow \left(k_1^{(t+1)}, k_2^{(t+1)}, \ldots\ldots k_N^{(t+1)}\right)$, where

$$K_i^{(t+1)} \in [\max(K_i^{\min}, K_i^{(t)} + step), \min(K_i^{\max}, K_i^{(t)} + step)] \quad (4)$$

The ki (t+1) parameter constantly would be involving the period Kmin-Kmax or at a sub-interval privileged of the, prohibited an action roots uncertainty within the commercial procedure in this way. Into the intellectual system, will find various instance period scales using the bandwidth/data rate matched. The training process operates at a diminished regularity compared to the control guidelines for the reinforcement learning procedure needs to run for the enough size to ensure that proper learning how to occur. The hypothesis considered was that when the control apparatus features a period is sampling controller of pH control, the educational needs to be done at the very least ten folds slow compared to the controller, in other words.,

$$p_{learning} = \delta . p_{control}, \quad (5)$$

where δ∈N,δ≥10 Finally, the bonus operation is understood to be R=90(1-ϕ(t))

where, the outcome key for the action is obtained, -ϕ(t), has got the

$$\phi(t) = \sum_{i=1}^{\delta} \left(\frac{ref_i^{(t)} - y_i^{(t)}}{ref_i^{(t)}} \right)^2 \quad (6)$$

phrase is after

Where, $ref_i^{(t)}$ is the reference parameter in time t+1. $p_{control}$ And $y_i^{(t)}$ is the production of the procedure in instance period t+1. $p_{control}$ With the value set $K_1^{(t)}, K_2^{(t)}, \ldots\ldots, K_N^{(t)}$. As seen, ϕt is the

Mean Square Error (MSE) computed in [t,t+1. $p_{control}$].

The bonus operation is plumped for establishing the greatest q-parameter that can be done 100. Because of the presumption concerning the kind of procedure, i.e., very fast problematic is powerful that the bonus needs to be restructured on a period, αand (are initialized to 0.1. Taking into consideration every one of these changes, the big event to upgrade the Q values

$$Q_{(t+1)}(s_{t+1}) = Q_{(t+1)}(s_{t+1}) + \alpha_t (R_{t+1} + \gamma \max_{\alpha \in A} Q_t(s_{t+2}) - Q_{(t)}(s_{t+1})) \quad (7)$$

Greedy approach is more accurate for first to pick an action, because it precisely functions in lots of positions and genuine position. The greedy strategy is displayed in algorithm 1. A lot of the action to the modified Q-learning technique are offered. The alterations being main in the algorithm q-learning linked to state and action, along with all the current depiction connected with reward function.

Algorithm 1: Greedy policy Algorithm

R=random ()
If i $<\in$
Take random action between the points
Else
Take the action that produces the state with most q values
End

The key change introduced into the algorithm is q-learning related to this is associated with the principles of state and action, in addition to aided by the depiction associated with reward function.

Algorithm 2 (modified q learning technique for EMCAP)

Initialize $Q(s_t)$ arbitrary (or with a static parameter attained with some technique)

Initialize s_0 to an random or static state.

Iterate, until s_t is a terminal state.
For each step do
Choose a, using the $\in$–greedy algorithm;
Perform action α_t and change to s_{t+1};
Wait $\delta . p_{control}$ and receive R;
Update Q-values with equation (7);
$s_t \leftarrow s_{t+1}$;
End

4. Experimental setup

The proposed technique is evaluated in a term is fungus utilizing SWI-Prolog software. The fungus world environment happens to be intended to have both power and fixed (Figure 1). The fixed obstructs are far more versatile, to produce a location is specific of environment. You will find the various values in the environment for the agent's engine is biochemical evaluation. Set values within the environment are manufactured through the checkbox composed of 1. Little fungus 2. Standard fungus 3. bad fungus 4. Ore 5. Crystal 6. Golden ore 7. Medication. The agents are manufactured into the environment by making use of Prolog (Figure 2 and Figure 3). Every one of the parameters is changed in accordance with demands which can be experimentally as they are defined in a setup module. The agent cannot distinguish among standard fungus, tiny fungus, and bad fungus until it gathers or consumes them. The experiments had been carried out for a similar amount of agents, the sort of equivalent amount of ore (including standard and golden ore) and fungi (including standard, tiny, and bad), together with the exact same things (including hurdles). Enough time scale and maximum rounds had been held constant with the addition of the kind is the exact same of in each test.

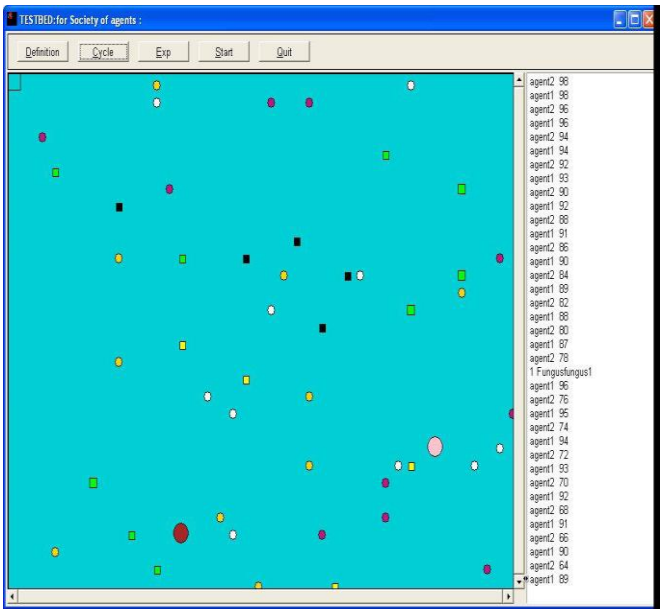


Fig. 2: Fungus world Testbed

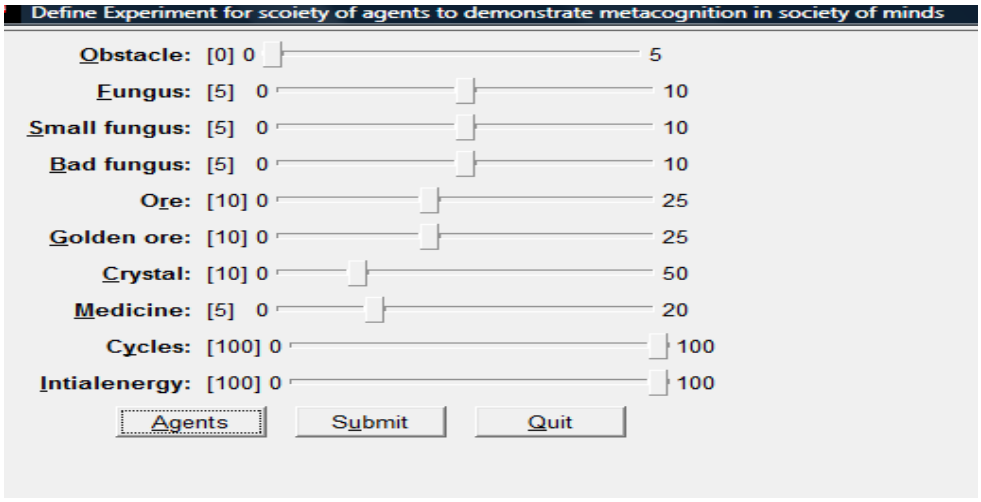
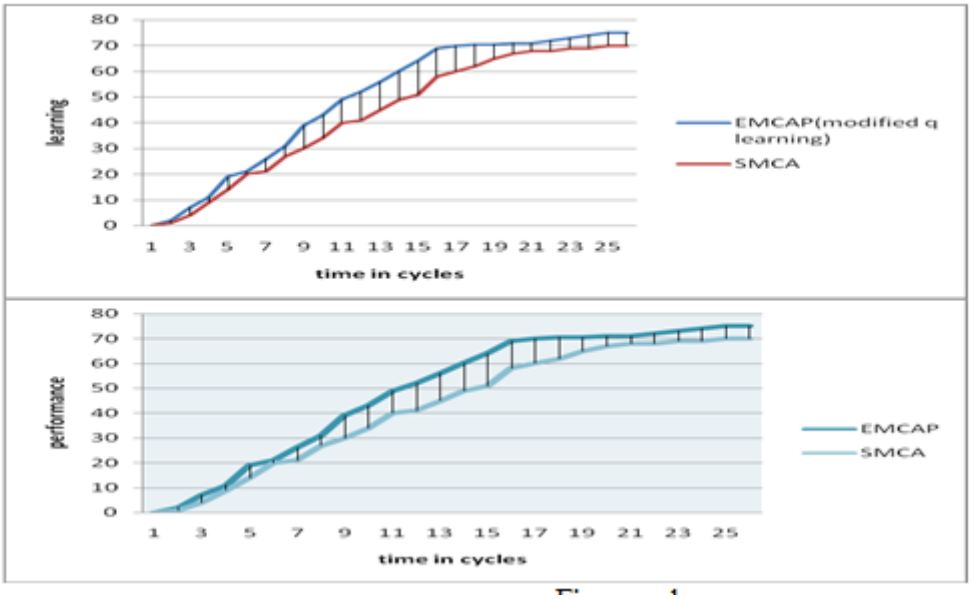


Fig. 3: Types of food

Standard Fungus: Fungus is really a energy efficent for the agents. Every fungus is standard a real estate agent 10 power devices. Initially, each representative has predetermined power devices. The representative uses a hard and fast amount of power devices for every single period. Fungus: The little fungus provides a realtor 5 power devices. In the event that representative uses a fungus is little 5 power devices (standard) are put into the vitality storage space. Bad Fungus: The bad fungus has 0 power devices. In the event that representative uses bad fungus, it gets power is worthless. Moreover, bad fungus advances the breakdown rate price, and changes the breakdown influence. Ore: The gathering of ore could be the objective is ultimately of representative. Each representative team attempts to gather just as much ore as you possibly can into the environment. Golden Ore Assortment of golden ore escalates the outcome of a representative. One little bit of golden ore is corresponding to five ore is standard Crystal: Assortment of crystal advances the performance of representative with a component is dull compared to ore. Medication: The medication disturbs the breakdown of the agent in the system network. The gathering of medication declines the breakdown rate.

Table 1: Table for learning

Time	EMCAP	SMCA
0	0	0
1	2	1
2	7	4
3	11	9
4	19	14
5	21	20
6	26	21
7	31	27
8	39	30
9	43	34
10	49	40



Graph:[1.1,1.2]

The 5th layer in EMCAP, Learning modifies decision making at one degree about actions at another degree for tasks defined at yet an additional degree this layer is in place managed through connections towards the met control degree. The modified learning determines considering exactly what and exactly how to map situations to action for making the most of a reward. Modified Q-Learning system discovers or attempts to locate a reward for optimum in the use of fungus for the action. The policy describes the stimulus- response rules for representative behavior. Policy is really a core element of reinforcement learning.If the agent's action is low reward, then policy will likely to be changed to another, also is actively seeking the reward is high. Benefits determine the instant and desirability is intrinsically of states. The very best actions for the representative are discovered by mistake and test. The representative is relocated to do the duty, utilizing the modified q algorithm is learning performance is calculated with respect to the reward; the reward is determined with regards to the levels of energy regarding the representation. The performance associated with the modified q learning algorithum about EMCAP architecture is calculated with SMCA that has shown in the graph.

5. Conclusions

This work demonstrated how to devolop and adopt a modified q_learning mechanism in cognitive system in the wide area of AI. This work gives frame rules of modified Q-learning and investigates EMCAP architecture .

Experiments are conducted with agents in EMCAP using Fungus world tested [1]. Agents from different levels of Architecture were employed for this experiment. In this experiment to show the effect of modified Q-learning on an agent, we considered comparisons between Q-learning and modified Q-learning The agents are experimenting with the same percentage (100%) life expectancy and resources (refer graph 1.1,1.2) in learning with respect to time in cycles .The proposed algorithm in a EMCAP has been simulated by using a prolog.The modified Q-learning technique had been tested in a Fungus world environment. The, modified q learning algorithm gets more rewards and gives more performance as shown in the graphs.

References

[1] D. Ganesha , Vijayakumar Maragal Venkatamuni “Design and development of hybrid Architecture model named Enhanced Mind Cognitive Architecture of pupils for implementing the learning concepts in Society of Agents” * Indian Journal of Science and Technology, Vol 10

[2] A. Sánchez Boza, R. H. Guerra, and A. Gajate, "Artificial cognitive control system based on the shared circuits model of sociocognitive capacities. A first approach," Engineering Applications of Artificial Intelligence, vol. 24, pp. 209-219, 2011.

[3] Q. Hu, Sh. Jie, and D. Yu, "Application of Fuzzy Self-learning Sliding Mode Variable Structure Control in Linear AC Servo System," IEEE Power Electronics and Motion Control Conference (IPEMC 2006), vol.3, pp.1-5, 14-16 August 2006.

[4] Tamayo-Torres, L. Gutierrez-Gutierrez, and A. Ruiz-Moreno, "The relationship between exploration and exploitation strategies, manufacturing flexibility and organizational learning: An empirical comparison between Non-ISO and ISO certified firms," European Journal of Operational Research, vol. 232, pp. 72-86, 2014.

[5] L. M. Hercog, "Better manufacturing process organization using multi-agent self-organization and co-evolutionary classifier systems: The multibar problem," Applied Soft Computing, vol. 13, pp. 1407-1418, 2013.

[6] Z.-P. Su, J.-G. Jiang, C.-Y. Liang, and G.-F. Zhang, "Path selection in disaster response management based on Q-learning," Int. J. Autom. Comput., vol. 8, pp. 100-106, 2011.

[7] K. Lakshmanan and S. Bhatnagar, "A novel Q-learning algorithm with function approximation for constrained Markov decision processes," Communication, Control, and Computing (Allerton), pp.400-405, 1-5 October 2012.

[8] A. Hariri and O. P. Malik, "22 - Self-Learning Knowledge Systems and Fuzzy Systems and Their Applications," in Knowledge-Based Systems, C. T. Leondes, Ed., ed San Diego: Academic Press, 2000, pp. 675-707.

[9] Ganesha D*,Dr. Vijayakumar Maragal Venkatamuni”Application Of Reinforcement Learning Methodologies In Society Of Mind Cogntive Architecture” International Journal of Advances in Engineering & Scientific Research (IJAESR) ISSN: 2349 –3607 (Online) , ISSN: 2349 –4824

[10] T. Padmapriya and V. Saminadan, “Distributed Load Balancing for Multiuser Multi-class Traffic in MIMO LTE-Advanced Networks”, Research Journal of Applied Sciences, Engineering and Technology (RJASET) - Maxwell Scientific Organization , ISSN: 2040-7459; e-ISSN: 2040-7467, vol.12, no.8, pp:813-822, April 2016.

[11] S.V.Manikanthan and D.Sugandhi “ Interference Alignment Techniques For Mimo Multicell Based On Relay Interference Broadcast Channel ” International Journal of Emerging Technology in Computer Science & Electronics (IJETCSE) ISSN: 0976-1353 Volume- 7 ,Issue 1 –MARCH 2014.

[12] Rajesh, M., and J. M. Gnanasekar. "GCCover Heterogeneous Wireless Ad hoc Networks."Journal of Chemical and Pharmaceutical Sciences (2015): 195-200.