International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-8 Issue-6, March 2020

A Hybrid Algorithm in Reinforcement Learning for Crowd Simulation

K.Pavithra, G.Radhamani

Abstract - Exploiting the efficiency and stability of Dynamic Crowd, the paper proposes a hybrid crowd simulation algorithm that runs using multi agents and it mainly focuses on identifying the crowd to simulate. An efficient measurement for both static and dynamic crowd simulation is applied in tracking and transportation applications. The proposed Hybrid Agent Reinforcement Learning (HARL) algorithm combines the Q-Learning off-policy value function and SARSA algorithm on-policy value function, which is used for dynamic crowd evacuation scenario. The HARL algorithm performs multiple value functions and combines the policy value function derived from the multi agent to improve the performance. In addition, the efficiency of the HARL algorithm is able to demonstrate in varied crowd sizes. Two kinds of applications are used in Reinforcement Learning such as tracking applications and transportation monitoring applications for pretending the crowd sizes.

Keywords: Artificial Intelligence, Reinforcement Learning, Reinforcement Learning Agent, Crowd Simulation

I. INTRODUCTION

The autonomous, interacting entities' collection which are allocated in a common environment and these entities uses sensors to recognize this environment and upon the sensors item with actuators are used which collectively is known as multi-agent system. There exists various domains where applications of multi-agent system are being used which includes data mining, collaborative decision support systems, resource management, distributed control, robotic teams, etc. These systems stand up as one of the common way that is used for looking the scheme, or as another perspective for schemes which are formerly known as centralized. For example, generally the control authority is distributed between the robots in the robotic teams [1].

Due to the increased complications in the real-world problems, there exists various situations in which a single deep RL agent cannot deal with such situations effectively. Therefore, in these situations, the MAS (Multi-Agent System) applications are indispensable [2]. In such systems, agents cooperate or compete with each other so that best results can be obtained. These

Manuscript received on March 15, 2020.
Revised Manuscript received on March 24, 2020.
Manuscript published on March 30, 2020.

* Correspondence Author

K.Pavithra*, Research Scholar, School of IT and Science, Dr.G.R.D College of Science, Coimbatore 641014. kpavithramsc@gmail.com

G.Radhamani, Professor & Director, School of IT and Science, Dr.G.R.D College of Science, Coimbatore 641014. radhamani@grd.edu.in

systems include autonomous military systems like spacecraft, surveillance, and unmanned aerial vehicles; traffic control systems; cooperative robots in the production factories, and multi-player online games. Including several presentations of Reinforcement learning in the literature, around various studies that uses RL in MAS, later forth multi-agent RL (MARL).
Therefore, a system with multiple agents' various semi-autonomous or autonomous modules is referred as multi-agent system. Furthermore, multi-agent learning cannot be considered as single agent learning's simple addition. The multi-agent learning process is straightly being contingent on the numerous agents' interaction that will growth the problematic system's complexity. Established the complete algorithm's performance can be enhanced by linking the multi-agent expertise's rewards along with the reinforcement learning.

Furthermore, a RL (Reinforcement Learning) agent uses a scalar reward signal as response for its performance and absorbs through interrelating with the atmosphere. Also, for the multi-agent culture, it becomes much more attractive due to its settings related to the summary as well as its simplicity. Although, one of the greatest challenge in MARL (Multi-Agent RL) consists of explicit consideration of other learning agents (non-stationary) by each of the learning agents as well as coordinating agent's behavior with others so that it results in the coherent joint behavior.

II. LITERATURE SURVEY

The survey is carried out into two kinds of applications namely Tracking application and Transportation monitoring system using crowd sources GPS data. Hereby discuss each application below.
(i) Tracking application- In the recent decade, much attention has been gained by the object tracking [3-6]. A principled technique is intended by RL (Reinforcement Learning) to the issues related to temporal decision making. The object tracking mainly emphasizes on localizing an object that is in uninterrupted video frame and in the first frame it is provided with initial annotation. A standard RL framework consists of an agent who absorbs a policy purpose from the environment which helps in aping the state in exploiting every distinctive time step which focuses on the maximizing the collected plunders which environment has returned. Usually, the effective application of RL has been seen in game playing, path planning, list management, etc.

In the object tracking of training active visual explanations (that are frame sequences) are preceded by a tracker through some efforts as well as for an output control signals are created by the corresponding camera (that is turn left, move forward, etc.). Camera controller tasks as well as tracking is discretely tackled by the conventional procedures and there exists some difficulty in jointly tune the resulted system. In the real-world, substantial human pains are required by these methods for classy trial-and-error system tuning as well as image labeling.

Demonstrate the successful examples of such transfer, via experiments as well as with the help of the proposed active tracker trained in simulation real-world robot's deployment is achieved [7].

For the multi-object tracking a CRL (Collaborative Reinforcement Learning) method is intended to propose in this paper. Furthermore, for showing this approach's effectiveness, the challenging MOT15 as well as MOT16 benchmark's experimental results are presented [9].

Chu et al. [10] used a single object tracker and concentrates on studying the robust appearance model for every target. Furthermore, a prediction-decision link is used for reporting noisy and occlusion detection difficulty as well as for online multi-object tracking decisions are made by using above approach.

Also, for localizing the objects collectively in some iterations a collaborative deep reinforcement learning method is proposed by Kong et al. [11]. Spontaneously tracing elongated structures, such as axons and blood vessels, is an interesting problem in the ground of biomedical imaging but single with many downstream presentations.

Finally, we demonstrate that our model's uncertainty measure a feature lacking in engineered trackers corresponds with how well it tracks the structure [6].

Pinto et al. (2017) [7] introduced the asymmetric actor-critic algorithm, where the critic is given extra information that is available only during training, can improve the learning of the true value function. The methodology used in tracking application with corresponding performance is shown in Table1 as follows:

Table1: Tracking Application

Author (Year)	Title	Method used	Advantage
Pinto et al. (2017)	"Asymmetric actor critic for image-based robot learning"	asymmetric actor critic algorithm	It allows to learn the underlying value function more easily.
Kong et al. (2017)	"Collaborative deep reinforcement learning for joint object search"	collaborative deep reinforcement learning method	fine-tuning on real data improves performance
Chu et al. (2017)	"Online multi-object tracking using cnn-based single object tracker"	robust appearance model	for making decisions for online multi-object tracking prediction-decision network is used

III. PROPOSED SYSTEM

Q-Learning is an Off-Policy algorithm, SARSA algorithm is an On-Policy algorithm for Temporal Difference learning. One of the main dissimilarity among Q-Learning and SARSA algorithm is that for Q-value updation, the next state maximum reward is not utilized as essentially. The SARSA as well as Q-learning algorithm hybrid is that, which allows the global as well as local search operations' dynamic selection which depends on the reward mechanism is within the HARL (Hybrid Agent Reinforcement Learning) methodology's framework.

A. Q-Learning Algorithm

It is an off policy reinforcement learning algorithm which examines for discovering the most appropriate actions that must be used for specific current state. Furthermore, this is believed as q-learning function's off-policy which absorbs the value function according to the action derived from another strategy.

Q-Learning algorithm process

It is an off policy reinforcement learning algorithm which examines for discovering the most appropriate actions that must be used for specific current state. It's measured off-policy then the function of q-learning studies from the actions which are not related to the current policy, for example, taking chance actions, as well as hence there is no need of a policy. To be more specific, q-learning try to find the policy which increases the total reward i.e. as shown in Fig.1 below

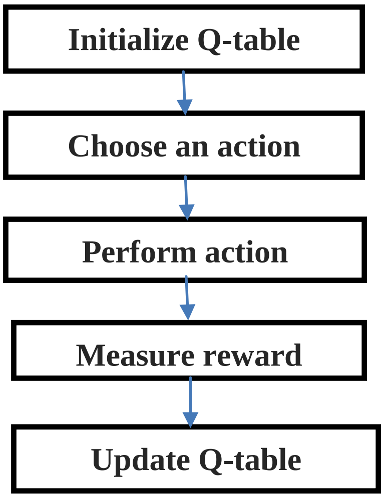


Fig.1 Q-Learning Process

B. State-Action-Reward-State-Action (SARSA) Algorithm

A SARSA agent relates with the environment and updates the policy established on actions taken; hence this is referred as an on-policy reinforcement learning algorithm. Q values denotes the receiving the possible rewards in the following time stamp so that particular action a must be taken in state s, plus the reduced future reward acknowledged from the next state-action opinion.

Sarsa Algorithm Process

The major alteration in Q-Learning is stated as for appraising the Q-values, resulting state's maximum reward is not essentially used. Whereas, it uses a new action as well as consequently reward, Fig.2 which is elected spending the similar policy which determines the actual action.

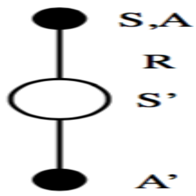
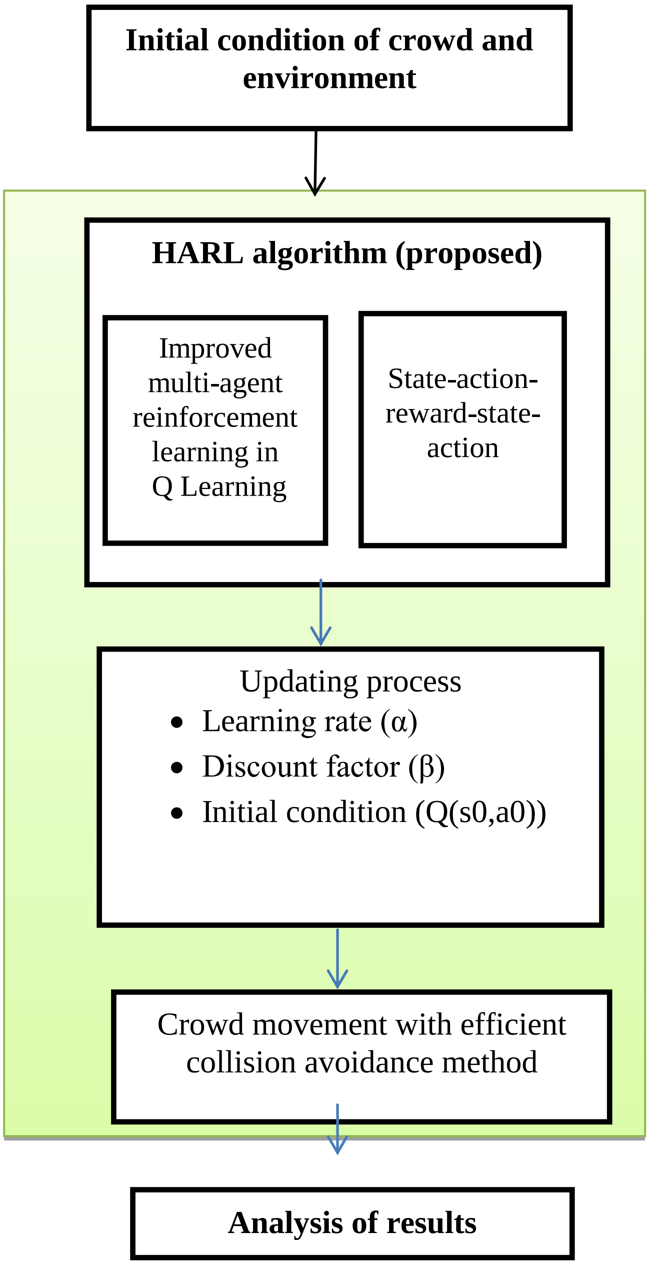


Fig.2 S-A-R-S-A process

C. Hybrid Algorithm

Reinforcement Learning (RL) offers a promising new approach to improve the performance by using hybrid algorithms. Those algorithms targeted for improving the performance of the learning process.



The methodology as shown in Fig.3, explains the updating process of simulating the crowd in an effective way by hybrid algorithms. Q Learning off policy and S-A-R-S-A on policy function value for crowd progress with no collisions.

This Hybrid framework multiples the agents to advise each other while learning in a shared environment. If in any state an agent is uncertain about what to do, it be able to ask for guidance to new agents and may accept answers from agents that have more confidence in their actuation for that state.

It is designed in the way to minimize the inconsistency of the Q-values of consecutive state-action pair. It improves the learning speed by updating Q-values and S-A-R-S-A values from past states at every time step. Tracking application and Transportation monitoring application for simulating the crowd to their objective, new focus, and their element approaches.

In addition this also updates the process to retrieve the values by taking learning rate, discount factor and initial condition for simulate the crowd. It is applied to optimize the route selection of the crowd while capturing dynamism in crowd. Crowd movements with collision are used for avoidance of noise in crowd.

Our formulation executes long-range collision avoidance for isolated agent groups to powerfully compute trajectories that are smoother than those obtain with state-of-the-art techniques and at quicker rates. Evaluation to real-world crowds demonstrate that crowds pretend with our algorithm display improved speed sensitivity to density related to human crowds.

IV. CROWD APPLICATIONS

Crowdsourcing permits large-scale and flexible supplication of human input for data collecting and analysis, which presents a new pattern of data mining process. Traditional data mining approaches often wants the experts in logical domains to annotate the data.

One of the key abilities agents should have is course plotting, which is a success goal without rear-ending with other means or problems.

Furthermore, CASCADE algorithm is presented by the Landay, Weld, Edge, Little, and Chilton (2013) [6], for creating a general reliable classification by HITs distribution to several individuals.

Also, CASCADE based improved workflow, namely, DELUGE is presented by the Bragg and Weld (2013).

Viktor Putrenko (2017) defined the relation among the particular region population's quantitative characteristics and their several social activities.

Also, the distribution relationships basic geographical peculiarities as well as their usage possibility in regional studies were marked [7].

An "Adaptive Voting Noise Correction (AVNC)" algorithm was proposed by Zhang et al. (2015), which takes benefits from the information gained through the inference stage to supervise noise identification.

It uses for evaluating the environment's safety as well as performance in each life cycle phase i.e. from design phase to operation phase.

Chaoqun Li et al (2019) proposed a novel class noise correction method that takes benefits from multiple noisy labels sets information. Fig.4 and Fig.5 shows the evacuation of crowd with its time. Thus the numbers of individual's increases, the advantages of our proposal model gradually emerge.

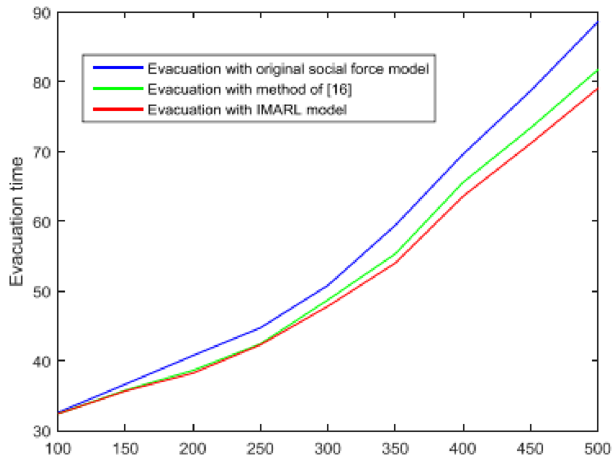


Fig.4 Evacuation with IMARL

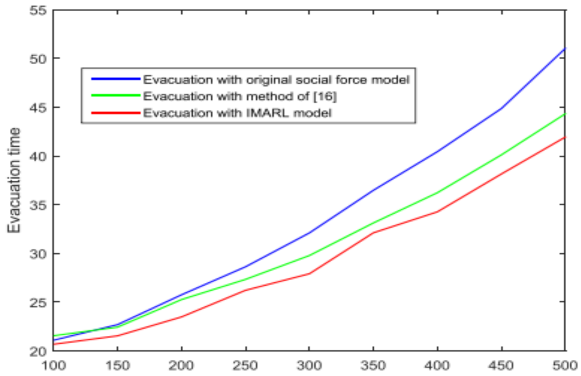


Fig.5 Evacuation with Social Force model

For making the freight truck a "smart freight" truck, Shailesh Chandra et al., 2019, developed a simulation framework combining it with the probabilistic modeling. The HARL algorithm is utilized for guiding the evacuation effect and quickly evacuate the crowd. Thus, for enhancing the truck freight as well as passenger operation, this crowd sourced data-based simulation framework might be collaborated with the traffic simulation software packages. Fig.6 shows the evacuating with HARL process is applied in crowd evacuation, simulation and quick migration with increasing the exit way when the crowd is more.

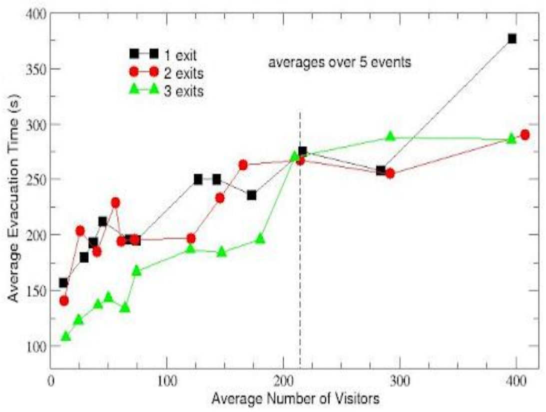


Fig.6 Evacuation with HARL process

In the recent years, agents' crowd is simulated through several methods that are being planned. Unfortunately, most of them are not computational scalable as the production of simulate agents is more. For immersive actuality platforms, gaming, and virtual production these values are predominantly vital.

Dynamics Crowd is an inclusive crowd simulation software function. It is being utilized for large simulation's conception and implementation among the complex infrastructures like airports as well as hospitals. With the help of flow simulations, the evacuation planning can be improved. Also, it has been proved to be a powerful tool for analyzing, predicting, and measuring the people's flow via. Infrastructure and helping in its management "people flow." Furthermore, evacuation plans examination as well as exploration enables the evacuation simulation modeling.

One of the most advanced crowd simulation and analysis tools available anywhere. Capable of simulating hundreds of thousands of citizens within a substance of hours, enabling rapid design exploration. Table 2: illustrates the survey of crowd applications which helps to increasing the simulation in crowd by using specific methods with advantages.

Table 2: Crowdsourcing applications

Author Name (Year)	Title	Method Used	Advantages
Viktor Putrenko (2017)	Data Mining of relationship in Crowdsourcing Projects and Social Activities of Citizens	crowdsourcing project OpenStreetMap (OSM)	Social activity of the population and creating volunteered geographic information was performed.
Shailesh Chandra et al. (2019)	Crowdsourcing-based traffic simulation for smart freight mobility	Discrete-time Markov chain (DTMC)-based simulation model	Able to detour to avoid a downstream congestion on the path.
Chaoqun Li et al (2019)	Noise correction to improve data and model quality for crowdsourcing	Novel between-class margin-based noise correction (BMNC)	BMNC can significantly reduce the class noise level in integrated labels and thereby enhance the performance of target classifiers.
Zhang et al.(2015)	Learning from crowdsourced labeled data	Adaptive voting noise correction algorithm	The labeling qualities of labelers, is used to guard noise identification and filtering.

V. DISCUSSION

To solve the optimal tracking problem for continuous-time non-linear systems with completely unknown dynamics. Conventional approaches to tackle the camera controller tasks distinctly and resultant system is problematic to adjust jointly.

Reinforcement learning resolves the difficult problem of correlating immediate actions with the delayed returns they produce, also difficult to understand which action leads to which outcome over many time steps. Furthermore, multi-agent learning cannot be considered as single agent learning's addition.

On the dissimilar, the procedure of multiagent learning is openly depends on the interaction of numerous agents, which will growth the complexity of the problem.

Transportation difficulties develop a challenge when the scheme with user's behavior is too hard to model in decision making, planning, and managing.

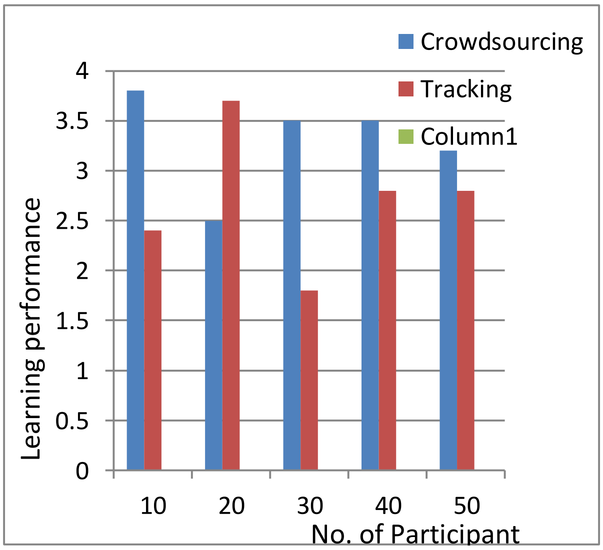


Fig.7 Overall performances based on learning performance

As Fig.7 will explain the performance of learning based on the number of participant in the application. This is important to overcome the concern that is related with the continuous increasing need with the restricted road supply as well as it also involves accurate prediction's better utilization.

Agent-based Approach is characterized by autonomous, interacting individuals. In this approach, crowd's every agent is assigned with an intelligence degree; and depending on their decision rule set they will be reacting to a particular situation. The surroundings of agent helps in providing the information that is necessary for deciding an action. To determine the technology usage for improving economy and its vital assets' efficiency, more reliable travel time for their customers, making a speedy improvement in congestion relieving.

VI. CONCLUSION

Two schemes of combining QLearning and SARSA Algorithm based on corrective instructions were presented: Recent advances in communication technologies have enabled researchers to collect travel data based on technologies. Technically and economically feasible to review driving behavior in natural surroundings on a large scale, through unobtrusive data collection and without experimental control. A brief review of reinforcement learning algorithms, followed by advanced learning methods. The above mentioned schemes of two algorithms are hybrid and it's tracking application and transportation monitoring system using crowd sources data. Based on these two kinds of application, the problems are faced and methods can be listed. To stimulate dynamic crowd, the HARL algorithm quickly evacuate the crowd.

REFERENCES

1. ThanhThi Nguyen, Ngoc Duy Nguyen, and SaeidNahavandi, "Deep Reinforcement Learning for Multi-Agent Systems: A Review of Challenges, Solutions and Applications", 2018.
2. Bertinetto, Luca, Valmadre, Jack, Golodetz, Stuart, Miksik, Ondrej, and Torr, Philip HS. Staple: Complementary learners for realtime tracking. In The IEEE Conference on Computer Vision and Pattern Recognition, pp. 1401–1409, 2016.
3. Danelljan, Martin, Bhat, Goutam, Shahbaz Khan, Fahad, and Felsberg, Michael. Eco: Efficient convolution operators for tracking. In The IEEE Conference on Computer Vision and Pattern Recognition, July 2017.
4. Cui, Zhen, Xiao, Shengtao, Feng, Jiashi, and Yan, Shuicheng. Recurrently target-attending tracking. In The IEEE Conference on Computer Vision and Pattern Recognition, pp. 1449–1458, 2016.
5. Zhu, Gao, Porikli, Fatih, and Li, Hongdong. Beyond local search: Tracking objects In The IEEE Conference on Computer Vision and Pattern Recognition, pp. 943–951, 2016.
6. Luo W, Sun P, Zhong F, Liu W, Zhang T, Wang Y. End-to-end active object tracking and its real-world deployment via reinforcement learning. IEEE transactions on pattern analysis and machine intelligence. Feb 14, 2019.
7. LiangliangRen, Jiwen Lu , Zifeng Wang , Qi Tian, Jie Zhou," Collaborative Deep Reinforcement Learning for Multi-Object Tracking", Springer, 2018.
8. Chu, Q., Ouyang, W., Li, H., Wang, X., Liu, B., Yu, N.: Online multi-object tracking using cnn-based single object tracker with spatial-temporal attention mechanism. In: ICCV. pp. 4836–4845, 2017.
9. Kong, X., Xin, B., Wang, Y., Hua, G.,"Collaborativedeep reinforcement learning for joint object search." In: CVPR. pp. 1695–1704, 2017.
10. Tianhong Dai, Magda Dubois, et al., "Deep Reinforcement Learning for Sub pixel Neural Tracking", Proceedings of Machine Learning Research 102:130–150, 2019.
11. L. Bus¸oniu, R. Babuska, and B. De Schutter, "A comprehensive survey of multi-agent ˇ reinforcement learning," IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, vol. 38, no. 2, pp. 156–172, Mar. 2018.

AUTHOR PROFILE

K.Pavithra has received her M.Sc in computer Science, Dr. G R Damodaran College of Science, affiliated to Bharathiar University, India . Formerly, she worked as an Assistant Professor for last 7 years from 2012. She pursuing Ph.D in the discipline of Computer Science from Dr. G R Damodaran College of Science, affiliated to Bharathiar University, India and undergone her research activities in data mining and passionate to work in the field of reinforcement relearning. Further the author has widely published in various National and peer reviewed International Conferences.

Dr.G.Radhamani is presently working as Professor and Director, School of Information Technology and Science, Dr. G R Damodaran College of Science, affiliated to Bharathiar University, India. Formerly, she worked as Head, Department of IT, Ministry of Manpower, Sultanate of Oman. She also served as a Research Associate in IIT (India) and as a faculty in Department of Information Technology, Multimedia University, Malaysia. She received her M.Sc and M.Phil degrees from P.S.G College of Technology, India, Ph.D degree from the Multimedia University, Malaysia. Also invited to be Keynote Speaker and Chair for International conferences in India and abroad.