

IMPROVED PARTICLE SWAM OPTIMIZATION FOR CROWD SIMULATION USING HYBRID AGENT REINFORCEMENT LEARNING ALGORITHM

K. Pavithra, Research Scholar, School of IT and Science, Dr.G.R.D College of Science, Coimbatore, Tamilnadu, India.

Mail Id: kpavithramsc@gmail.com

G. Radhamani, Professor & Director, School of IT and Science, Dr.G.R.D College of Science, Coimbatore, Tamilnadu, India. Mail Id: radhamani@grd.edu.in

Senthil Kumar Thangavel, Department of Computer Science and Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, Tamilnadu, India, Mail Id: t_Senthilkumar@cb.amrita.edu

Abstract

In an emergency route planning technique, simulating the dynamic crowd has route capacity constraints and global target of evacuating all crowd evacuees. To stimulate the crowd, the new arena is developed to know the real-time situation to face the crowd evacuation on exit point. The crowd evacuation is done with the process of Hybrid Agent Reinforcement Learning (HARL) algorithm consisting of Improved Multi-Agent Reinforcement Learning (IMARL) and State-Action-Reward-State-Action (SARSA). In the proposed work, the appropriate route selection mechanism focused on finding optimum evacuation route(s) is done in the first phase. Dynamic crowd can also be evacuated to find its way with the support of the HARL process in the second phase. The proposed HARL method can also be implemented with multi-objective improved particle swarm optimization (IPSO) technique for crowd simulation. The experimented results demonstrate the effectiveness of stability in the HARL process, which provides an improved performance for crowd simulation.

Keywords: *Crowd evacuation, Agent, SARSA, particle swam optimization.*

I. Introduction

Crowds are formed by several or thousands of people that move around the environment with respect to their individual goals in space, avoiding obstacles or blocking and remaining close to friends or family (Yersin et al., 2008). In addition, crowds may regulate their movement in groups or individual. This is dependent on three aspects: (1) goals and needs; (2) social and physical attributes (i.e, level of interaction, age, or social differentiation); and (3) psychological and situational aspects (i.e, stress levels at a respective time or place). The crowd would be out of control effortlessly when density exceeds a certain threshold, whereas public security would be threatened seriously. The evacuation path is scheduled to condense the evacuation time in thickly populated areas, especially complex environments with obstacles and multiple path is one of the significant issues of crowd simulations caused by emergency tragedies [1]. In such a condition, most pedestrians will track to the nearer exit or follow the crowd to reach exit [2], the above will cause overcrowding delay, trampling mortalities and other safety incidents. Additional experimental data in the evacuation process, it is essential for a large number of people to conduct real world experiments on crowd evacuation [3]. Computer simulation technology not only guaranteed safety and convenience in reality, but also provide the meaningful information for crowd evacuation [4]. Therefore, it is indispensable to apply a computer simulation method for the study of crowd evacuation in large public places in emergencies. Considering the problem of the large scale planning of an evacuation

path, Multi-Agent Reinforcement Learning (MARL)[5] agents interact with the environment. To obtain the reward and punishment information is transmitted from the environment to maximize the reward through trial and error, (i.e.) closely related to the path knowledge accumulation of pedestrians in reality. Thus, combining Reinforcement Learning (RL) with interactive machine learning strategies might be a synergetic relationship, in which the learning processes are speed up as RL benefits from the human knowledge of the task. The main challenge in this type of learning is to correct associate actions with their effects, even though some of the effects are typically delayed. A multi agent system (MAS) comprises independent, autonomous and interactive agents. An agent is determined by desire (goal) and directed by own opinions to achieve the goal [2]. For instance, when an emergency evacuation is compulsory for stimulate the crowd from the dangerous part and move to the safe area. Considering an optimal path to the current stage of the safe area, is the necessary action taken by the crowd. When an emergency happens, all the status of pedestrians will be changed as an evacuation pattern. They will discard their current movement patterns and begin to follow the static crowd, towards the exit. The scenario of crowd evacuation is shown in Fig. 1. In an emergency situation the static crowd evacuation tends to move towards exit by following the shortest distance rule, depends on the pedestrian's familiarity of the scenario.

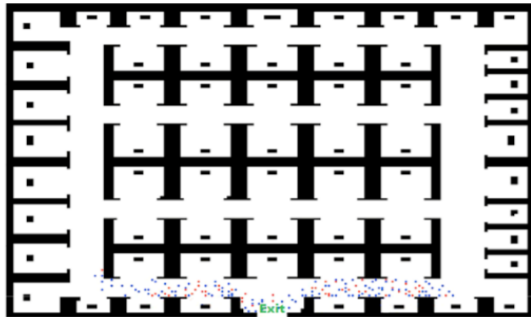


Fig.1 Static Crowd Evacuation

In a multi-agent system, the basic behaviour is difficult to understand the good when the environment is too complex. Many tasks required for agents to learn new behaviours for the improved performance of a multi-agent system [8]. For Instance, in simulation, the emergency evacuation is clearly notified inside the shopping centre to facilitate fast evacuation in dynamic crowd. Thus, the majority of them have arrived to the main evacuation exits. The flow of dynamic crowd evacuated in any directions to get exit as shown in Fig.2.

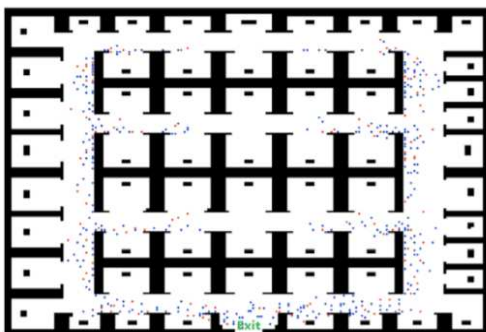


Fig.2 Dynamic Crowd Evacuation

RL is a type of algorithm in the field of machine learning that aims to allow agents to learn how to maximize future rewards in the environment. Specifically, an agent changes his state through actions and obtains a reward signal in the form of a scala [9]. RL can be regularized by the model of Markov decision process framework $\langle S, A, T, R \rangle$, included state S , action A , the transformation function T , the reward equation R , etc. For different spatial problems, the Markov decision making process is divided as continuous domain and the discrete domain. However, Busoniuetal.[16] clearly pointed out that Improved Multi-Agent Reinforcement Learning (IMARL) still facing some challenges, (i.e.) the problem in curse of dimensionality and the instability of the environment becoming more serious. Also, the multi-objective optimization problem of signal timing, some heuristic or metaheuristic methods were adopted to deal with it. In existing method, a nonlinear programming model to solve the signal timing problem for crowd evacuation and presented a heuristic algorithm to solve the traditional model. In order to solve pedestrian phase patterns, a HARL was proposed to obtain solutions for signal setting to optimized signal timings and to solve traffic

signal optimization problems. The Improved Particle Swam Optimization (IPSO) technique adopted to improve the exploitation power of HARL. The effectiveness of the HARL method is designed for an evacuation plan. The researchers have mainly focused on improving the algorithm performance to optimized the screening of Pareto solutions. The three contributions are made in our proposed work as mentioned: (1) the signal timing optimization model with the goals of HARL (2) in order to solve a multi-objective signal timing problem, by applying an improved Particle Swarm Optimization technique (IPSO) in HARL algorithm and (3) a dynamic relaxation strategy, which improves convergence and ensures the uniformity of solution set as distribution. The existing algorithm is compared with the traffic signal control strategies, and the results show that the algorithm has control effect.

The rest of the article is organised accordingly: Section II provides an overview of related works to analyse the evacuation process and each has unique methods for detecting a human in the crowd. Section III presents the challenges of Emergency Route Planning (ERP) in crowd evacuation for signal timing optimization. Section IV describe about the HARL algorithm based on QLearning and SARSA. Section V suggests an IPSO in HARL algorithm for path planning in crowd evacuees. Section VI describes the Improved particle swam optimization and Crowd Evacuation using HARL. Section VII expresses the results analysis for simulating the crowd. Finally, Section VIII provides a summary of our work and providing some suggestions for future study.

II. RELATED WORK

Xueying Huang (2019), a hybrid framework is used for detecting human faces in a crowd. It gives more accuracy and improves the speed by hybriding in it. Bindita Chaudhuri, Noranart Vesdapunt, Baoyuan Wang (2019), recommends building blocks to detect human faces in any pose through mobile devices. Initially they design a novel multitask learning framework that learns a disentangled representation of 3DMM parameters for a single face.

ShuaiShao, ZijianZhao Boxun, LiTete Xiao, GangYu Xiangyu, ZhangJian, SunMegvii Inc. (2018) focuses on crowdhuman methodology for detecting large crowds with powerful training datasets. To realize the behavioural choices, they calculate the position of each individual at the next step to obtain a conflict free moving path in a global scenario.

Other approaches of human detection Wang, X.; Hosseinyalamdary, S. (2019) and traditional object detection methods by Zhong-QiuZhao, PengZheng, Shou-TaoXu, Xindong Wu (2019) can be well applied for detecting human body appearance and its shape. Chengliang Chai, Ju Fan, Guoliang Li, Jiannan Wang, YudianZheng (2018),

CrowdPowered tool is used to combine various mining tasks for improving the performance and providing more accuracy.

Steffen Huber, HajoWiemer, Dorothea Schneider, Steffen Ihlenfeldt (2018), CRISP-DM methodology for data mining applications with scenarios. Adrian-CostinTundrea (2019) multi-agent-based algorithm for detecting large crowd. Rongyong Zhao, Qiong Liu, Cuiling Li, Daheng Dong, Qianshan Hu, Yunlong Ma, Qin Zhang (2019) self-organizing mapping network (SOM) using specific simulation tools for increasing accuracy. They proposed a novel knowledge representation model for the evolution mechanism of large crowd stability states.

Qingqing Wang, Hong Liu, Kaizhou Gao, And Le Zhang (2019) using Improved SFM and K-medoid algorithms for improving convergence speed and simulation. These algorithms are described for crowd simulation in complex contexts. Carlos Celemin, Javier Ruiz-del-Solar and Jens Kobe (2018) using PSRL and ML algorithms for specific techniques increasing the speed and getting human feedback. Combining RL with interactive machine learning strategies might be a synergetic relationship, in which the learning processes are sped up as RL benefits from the human knowledge of the task.

III. Challenges of Emergency Route Planning (ERP) Problem

The cognition of the ERP approach involves three distinct and interrelated components: (1) crowd modeling, (2) route assignment, and (3) emergency's scenario modeling. Our study is tuned towards fulfilling the first two aspects: crowd evacuating and path planning.

To evacuate crowds effectively is a challenging issue because emergency events may propagate in uncertain ways due to the effect of the perceived environment, space capacity constraining the speed of crowd movement, and modifications in crowd performance due to psychological aspects. In the context of planning, emergency evacuation also known as emergency route planning (ERP), focuses on three important factors as the routing of the residents, scheduling the resident regression and regulating the resident's flow rates. These influences are naturally consolidating on two interrelated ranges are the crowd dynamic and their supposed environment. In the context of crowd evacuation modeling and path planning, dynamic concepts are important characteristics that aim to fulfil either level of service, design elements of pedestrian facilities, or planning guidelines. In a more detailed term, a dynamic concept is an occurrence of synthetic human behaviour when compared to actual human behavior (Fridman et al., 2011). In our study, the dynamic perception refers to the cohesiveness of collections in a crowd, particularly the grouping phenomenon (group formation) within the crowd and the level of interaction within the groups (intra-group compliance). Also, a static or dynamic crowd may

regulate their movement in assemblies or separately. The crowd dynamic considered in this particular study, is tuned towards the first two aspects. Group cohesion, which is defined as the tendency for a group to be in unity while working towards a goal or to fulfill the demands of its members (Carron and Brawley, 2000), is one of the crowd dynamics that is measured for this training.

With additional information pertaining to initial locations of all evacuees and their final destinations, the ERP problem produces a set of origin and destination routes for evacuees. Consider a simple building ERP problem. Each room, corridor, staircase, and exit of the building is represented as a node, and each pathway from one node to another node is represented as an edge [9]. The main objective of the ERP problem is to minimize the computational cost of producing the evacuation plan while minimizing evacuation time. With respect to the previously mentioned ERP factors, determining the best route while regulating crowd flow within the acceptable performance of an evacuation plan, poses as another computational challenge. Therefore, adopting an appropriate ERP method is vital for consecutive risk moderation prior to the occurrence of a disastrous situation. An emergency plan specifies procedures for handling sudden or unexpected situations. The impartial is to be prepared to:

- Reduce damage to buildings, stock, and equipment.
- Protect the environment and the community.

However, the studies on crowd dynamics are rarely emphasized in the ERP community, which induces significant impacts on the evacuation efficiency and crowd survivability (Wang et al., 2008). In order for an effective emergency evacuation plan to be produced, an optimal routing of the evacuation route, which considers crowd dynamic should be emphasized. It also provides an optimum solution despite their stochastic nature (Yusoff et al., 2018). As such, this study helps to adopt a hybridization model (Hybrid Agent of Reinforcement Learning) with improved PSO as the dominating solution for solving the ERP problem. Challenges arise during evacuations due to the spread of people over space and time and the multiple paths that can be taken to reach them.

IV. Methods and Methodology

QLearning

Q-learning is the reinforcement algorithm in which the agent takes the most optimal action under different circumstances. The off-policy is a method of reinforcement learning in which the next state is calculated using the current state along with a greedy step. This policy predicts the next coordinate/state instead of calculating it, as seen in the on-policy method. The coordinate with the greatest reward is chosen as the next state.

Even though the next state is chosen based on the reward system, the program processes it as though the greedy algorithm wasn't used at all. On using this algorithm, even though we get the most optimal path possible, it might be risky in some situations, in which we end up with a high negative reward.

QLearning Algorithm

This algorithm uses the off-policy model to obtain its Q-values, and hence, reward is calculated based on the greedy action. Q-learning is also represented with the attributes $\{S, A, \gamma, \alpha\}$, where, S is the set of states, A is a set of actions, $\gamma(0 \leq \alpha \leq 1)$ is the discount and $\alpha(0 < \alpha \leq 1)$ is the step size. These are used to calculate the Q-values(Q [S, A]).

$$Q_{\text{new}}(st, at) \leftarrow Q_{\text{old}}(st, at) + \alpha[rt + \gamma \cdot \max_a Q_{\text{estimated future value}}(st+1, a) - Q_{\text{old}}(st, at)]$$

Where,
st represents the initial state,
st+1 represents the final state,
rt represents the reward

SARSA

SARSA(State-Action-Reward-State-Action) is called so, because it experiences to updating Q-values. It uses the on-policy method. The on-policy is a method of reinforcement learning in which the next state is calculated using the current state along with the data received by the exploration steps. The SARSA algorithm uses this policy. Although, this policy may be a bit more expensive, when compared to the off-policy method, it is safer and has lower risks.

SARSA Algorithm

SARSA is represented with the attributes $\{S, A, \gamma, \alpha\}$, where, S is the set of states, A is a set of actions, $\gamma(0 \leq \alpha \leq 1)$ is the discount and $\alpha(0 < \alpha \leq 1)$ is the step size. These are used to calculate the Q-values (Q [S, A]).

$$Q_{\text{new}}(st, at) \leftarrow Q_{\text{old}}(st, at) + \alpha[rt + \gamma \cdot Q_{\text{exploration value}}(st+1, at+1) - Q_{\text{old}}(st, at)]$$

Where,
st represents the initial state,
st+1 represents the final state,
rt represents the reward

HARL

In RL, Q-learning algorithms are that try to perform the most profitable action given the current state. However, these powerful set of algorithms are not fully exploited at their full potential. Hybrid Agent reinforcement learning (HARL) explains the updating process of simulating the crowd in an effective way by two different algorithms. Q Learning off policy and S-

A-R-S-A on policy function value for crowd progress with no collisions. This Hybrid algorithm multiplies 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 is able to ask for guidance to new agents and may accept answers from agents that have more confidence in their actuation for that state. To attain this, the action selection strategy of the standard RL algorithm is modified to take into account an experiential running in parallel with the RL process. Hybrid Agent Reinforcement Learning (HARL) a new method, which is successive to combine with the Sarsa algorithm and the Q-learning. For the simulation of Crowd, it can find that Sarsa algorithm converges faster than Q-learning, but Sarsa algorithm cannot find that the optimum path planning. 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. Thus, using the HARL algorithm enhances the learning speed, also it improves the final performance.

HARL Algorithm

Step1: Initialize arbitrarily all Q(s,a) values

Repeat (for each episode)

Step2: Choose a random state st or initialize state st

Repeat (for each step in the episode)

Step3: Select an action ar in A(s) arbitrarily

Step4: Select an action ap in A(s) according to the policy

Generate random value $\xi A(0,1)$

IF $\xi < \exp((Q(s,ar) - Q(s,ap))/T)$

THEN $a \leftarrow ar$ **ELSE** $a \leftarrow ap$

Step5: Execute the action a, receive an immediate reward $rt+1$, then observe the new state $st+1$

Step6: Update Q(st,at) according to step2

Step7: $st \leftarrow st+1$

Until st is one of the goal states

Step8: Recalculate the temperature parameter using the temperature-dropping criterion

Step9: All the episodes have to be searched until the desired number is found.

This HARL is a new class of algorithm that combines the advantages of advanced Qlearning with SARSA. The new action policies are done without delaying the learning process by hybridation; it reduces the probability of conflicts among actions from different policies. Also, it illustrates that the Q-learning based RL algorithm performs the well-known of Q-learning and the SARSA algorithm. Randomizing batches break correlations between samples and behaviour distribution is averaged over many of its previous states. The main difference between Q-learning and SARSA is that Q-learning is an off-policy algorithm whereas SARSA is an on-policy methods. The off-policy algorithms would not base the learning solely on the values of the policy, but would rather use an

optimistic estimation of the policy (in this case the maxa' selection condition), where as an on-policy algorithm bases its learning solely on the values of the policy at a given time.

V. Crowd Evacuation

RL based on single agents has remained widely used in most former cases of path planning, such as Q-learning and Sarsa. The rewards are usually set to find exits and avoid obstacles. However, IMARL has been developed more in the field of planning an evacuation path. In a small case the IMARL and SARSA algorithm is applied to crowd path optimization.

The Proposed hybrid algorithm, where the multiple agents can advise each other while learning in a shared environment. If in any state an agent is uncertain about what to do, it can ask for advice to other agents and may receive answers from agents that have more confidence in their actuation for that state.

Path planning is one of the critical issues of evacuation simulations, which purposes to find the optimal path from the starting position to the target position. At present, the more commonly used path planning methods in the crowd simulation field include swarm intelligence algorithms, the methods based on RL, etc. The improved swarm intelligence technique in HARL algorithm is general in crowd path planning but these algorithms often face early results in crowd simulation.

The method of hybridization is used to simplify the calculation to some extent, which leads to lower precision, reduces the maximum possible reward in path planning, and leads to a convergence to the local optimum when the input changes are small. Lee [10] proposed a method based on HARL for predicting crowd congestion to achieve crowd path planning in the event of a disaster.

In this study, the state-action-reward-state-action (SARSA(λ)) and Q-learning algorithms are innovatively used in different stages. In the selection, expansion, and simulation section of the HARL search, the SARSA(λ) algorithm is used to update the quality of each action. In the back propagation stage, the Q-learning algorithm is used to update the quality of each action. (e quality of each action is denoted by its Q-value. Probability distribution function based on a two-dimensional (2D) normal distribution matrix is used as feedback to SARSA(λ) and Q-learning for the layout

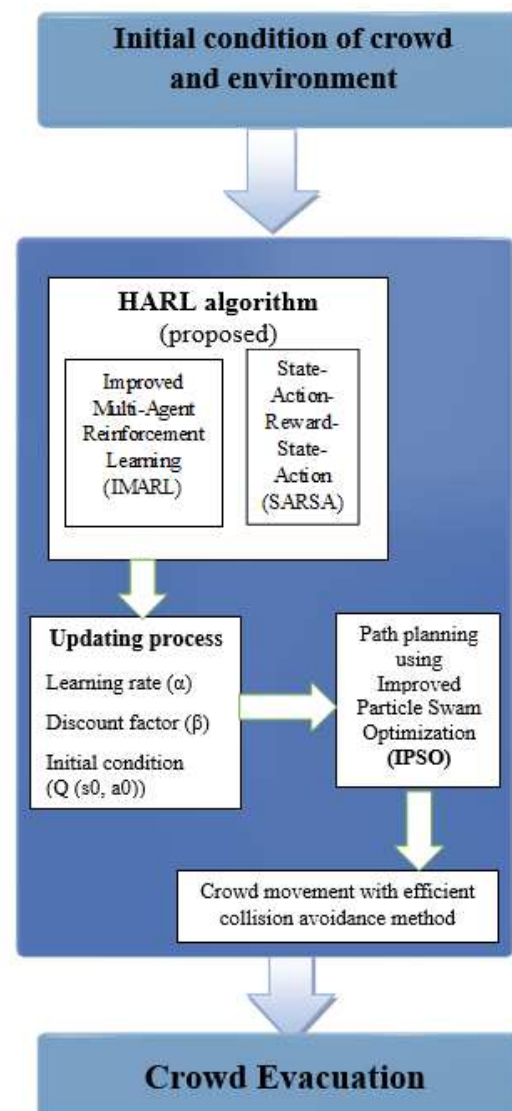


Fig.3 HARL Framework with improved PSO

The HARL framework in Fig.3 is proposed by considering the congestion detection area of the reward function based IMARL, which is beneficial to realize reasonable congestion avoidance and the selection of an evacuation exit. It improves the learning speed by updating Q-Values from past states at every time step. At present, IPSO technique has been widely used in optimization problems in various fields due to its good optimization ability.

Hybrid Agent Reinforcement Learning (HARL) algorithm is a proposed approach that can prevent low value nodes generated by the deep learning resources. Using SARSA(λ) and Q-learning algorithms in different stages of HARL search provides guarantee for learning steps faster and improves the algorithm's learning efficiency; besides, final result can be feedback to all steps to improve the accuracy of the return estimation value. It is designed in the way to minimize the inconsistency of the Q-values of consecutive state-action pairs. The IPSO algorithm is based on particle swarm intelligence [12], the movements of the particles are determined by their own best-

known position in the search space as well as the entire swarm's best-known position. In addition, this also updates the process to retrieve the values by taking the learning rate; discount factor and initial condition to simulate the crowd. It is applied to optimize the route selection of the crowd while capturing dynamism in the crowd. Crowd movements with collision are used for avoidance of noise in crowd [17]. A new HARL algorithm is a combination of IMARL algorithm with SARSA algorithm is proposed, also adds the mean field function to maximize the returns of their agents and uses the mean field idea to reduce the complexity of training additional samples, it can enable all agents to work better and maximize performance of a collaborative planning task. An improved particle swarm optimization is used for path planning method and also applied in the field of crowd evacuation which is proposed to reduce evacuation time, under the framework of congestion detection based HARL. The temporary target on the path is determined by the grid as each step unit, and the human actions in the simulation scene are simply separated into five actions. Then, according to the extracted direction, the current state and adjacent feature information are driven to predict the degree of congestion. The path direction with a small crowd density is selected for evacuation. The above methods provide a new idea for the research on crowd path planning based on RL to some extent. However, because of the curse of dimensionality, the continuous behaviour and observation space of the crowd are discretized. During the evacuation process, the reactions of different people are unpredictable. Since this model is simulated in a large enclosed public venue with multiple exits, it also describes the need to and a balance between applying the policy that is currently estimated to be best and exploring unknown regions of the state-space. One way to ensure this balance is to employ QLearning that makes the agent choose a random action with probability and the action maximizing $Q(s, a)$ otherwise. When an emergency occurs, pedestrians are all rational and always want to move closer to the exits.

VI. Discussions

Q-Learning improves by adding SARSA algorithm to the existing Q-learning system. The idea of HARL with a method designed for the hybrid control of a both off policy and on policy values. The Hybrid Agent Reinforcement Learning algorithm is the concept of hybridization to make agents/people to reach exit. Crowd simulation by applying IPSO in HARL which gives better performance.

a) Improved Particle Swarm Optimization

Particle swarm optimization (PSO) was first invented by Dr.Eberhart and Dr. Kennedy [11]. It's a crowd-based heuristic algorithm used to model social comportment, like birds that cluster to suitable sites to

find specific targets in multi-dimensional environments. In order to search with evolutionary algorithms and update the particles from one iteration to the next, PSO uses personal collections (called particles) [6-10]. Particle is searched according to two factors to find the optimal solution: its earliest position (pbest) and the best position of all the other members. (gbest) [11-13]. Shi et al. called pbest the cognitive part and gbest the social part. Evacuation preparation is important in order to represent both geo and social actions. It is not only physical (i.e. age and ability), but also psychologically significant, vary the flow of large numbers of feet. This movement of the crowd is especially difficult to avoid panic and achieve highly dense ordinary urgent evacuation move.

Emergency evacuation in disaster circumstances like cliffs or flash flooding made this much more difficult to grab huge numbers of people. In addition, the physical and socio-psychological situations of the pedestrians, the great mass gathering event requires public health planning and basic human needs, with fresh food, portable water, and accommodation.

Finally, flow segmenting structure for open spaces must be established to group moving crowds and plan spatio sequential crowd flows. Imagine a scenario where a bird flies around in a space toward its goal. Adjust its position according to the own experience and its surroundings. By learning both social and individual behaviors, the bird tends to move towards the most promising position. While improved PSO, the best person (Global best) is the leader who can rotate in the swarm during the evolution process.

b) Crowd Evocations using HARL with IPSO

This process describes the population distribution characteristics vary as the evolution goes on. For example, in the evacuation process the overcrowd disperse is not possible at a time. The path planning is easily recognized for human to evacuate the crowd and pursue their collective goals. However, on implementation of HARL along with IPSO technique the crowd effectively reach to exit. The crowd evacuation happened in an emergency situation as shown in Fig.4.

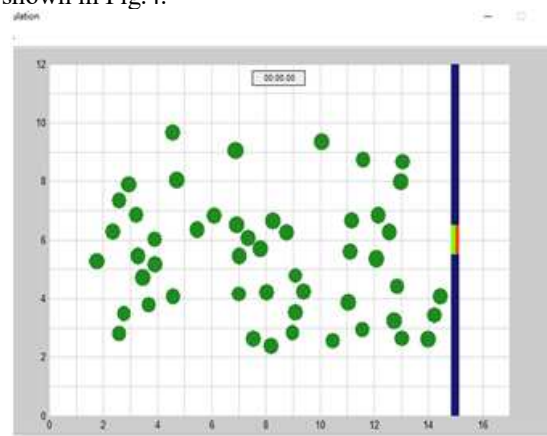


Fig.4 Crowd Occurrences in Building

In Fig.5, the crowd dispersing in exit. Each crowd is considered as a particle in the swarm, and the collective goal is to approach exit.

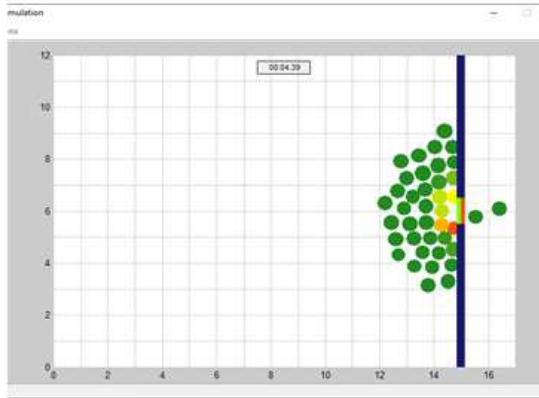


Fig.5 Crowd moving to an Exit

However, crowd simulation is based on the two assumptions: (i) In large enclosed spaces, most crowd are not familiar with an environment, (i.e.) A vague impression on the location of the exit. As a consequence, pedestrians in each swarm tend to follow the crowd or other people. (ii) The crowd as seen in either selected each swarm or designated as professional workers who are more qualified and familiar with this area. Herein, instead of following others, they likely to head straight to the desirable exits in proximity at a relatively high speed.

To address this problem, a proposed HARL method is used for the crowd evacuations and path planning. The behaviors of surrounding pedestrians will automatically seek the best routes for simulation. In this model, HARL algorithm is combined with Improved PSO in a discrete manner.

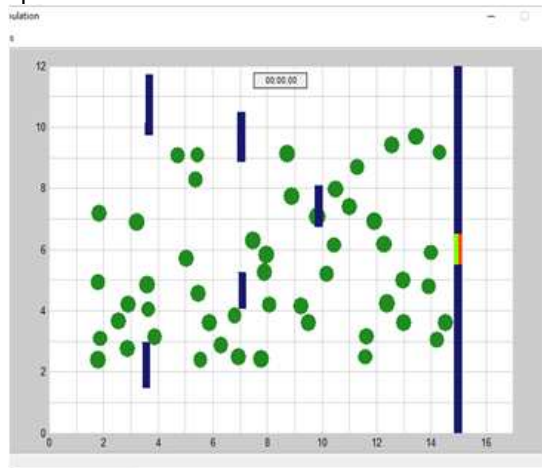


Fig.6 Crowd Occurrences in obstacle state

In Fig.6, the crowd behavior has changed due to occurrence of obstacles and move towards the exit.

The reason of changing in module occurs to adapt and get into exit by overcoming the obstacles. Assume changing module is a simple binary logic formula that decides, if crowd changes their chosen exit. This depends on directional attributes and a prefixed parameter that adjust into exit-choice changing. The changes of module in exit will happen while implementing IPSO in HARL algorithm to

reach an exit. Results showed a substantial difference in enhancing the accuracy of the simulation output. The proposed method has concluded that an intermediate degree of decision changing is the most beneficial strategy.

VII. Experimental Analysis

In Fig.7, the performance of the proposed path planning is effectively evacuating the crowd. The agent has diverted in obstacle state and retarget to a new scenario of our simulation method as observed.

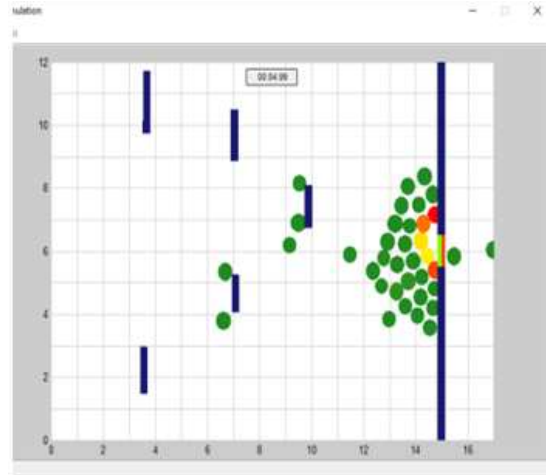


Fig.7 Crowd Evacuations in obstacle state

Initially, some typical swarm intelligence algorithms are introduced and their performances are analysed. This idea is to use crowd expression programming to find a heuristic rule. This rule is used to indicate people in the same sub region to move towards the same exit.

With the improvement of crowd simulation in obstacle state, walking speed of agents are calculated. In Fig.8, the crowd time has found while hassled in Obstacle state. HARL method is focused on the agent environment interactions and collision avoidance for simulating the crowd.

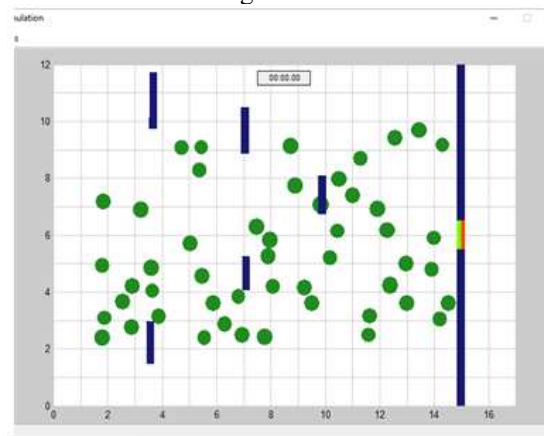


Fig.8 Crowd hassled in Obstacle state

The crowd evacuation planning problem is converted to find the optimal heuristic rule that minimizes the total evacuation time. Crowd simulation is used to compute routes by iteratively partitioning graph edges at critical division points.

However, the adaptive evacuation plan is capable of responding to the changes in environmental conditions. The agents reached an exit at a similar speed to evacuate crowd through HARL algorithm as shown in Fig.9.



Fig.9 Obstacle crowd reached exit

The HARL based crowd evacuation planning developed and adapts to changing conditions, but the formula is complex and challenging to interpret. It assumes that a crowd distribution is known in advance also considering safety or dynamic changes during evacuation. This model is based on dynamic grouping principle, driven by improved particle swarm optimization in HARL framework. The proposed methodology is not only limited to generate navigation rules. It can also be used to produce other behavioural rules of crowd models. This model focuses on simulating the gathering and flowing characteristics, regarding human crowd movement. Our proposed method was implemented in MATrix LABORatory. MATLAB is a high performance language for technical computing. It integrates computation, visualization, and programming environment. It's easy to use graphics command that makes the visualization of results immediately.

HARL algorithm is tested on publicly available datasets involving crowded scenarios, and on synthetic data produced by a crowd simulation algorithm, which allows the creation of controlled environments. With a global perspective, this model can be used for rough analysis and its advantage for fast simulation. Based on Q-learning, SARSA and HARL algorithm performances the number of agent's / people, time of execution, and path description was explained in each algorithm as shown in Table-1 as follows,

Algorithm	Number. of Agents/ People	Execution Time (sec)	Path Description
QLearning	90	11.89	Optimal Path with Executional Risks
SARSA	92	17.22	Safe but not Optimal
HARL with IPSO	91	14.94	Safest Optimal Path possible

Table-1: Analysis of crowd evacuation

The analysis for the crowd evacuation was done by creating a fixed environment. The experimental results showed that the improved evacuation model can simulate crowd evacuation on the complicated scene. Each agent then changes its velocity maximize this time to collision. By continuously updating the velocities in this manner, collisions are avoided. The key idea is to generate a set of input-output pairs from the real trajectory data and then utilize HARL methodology to search for optimal explicit formulas that describe the relationships between inputs and outputs. Increasing time pressure causes a severe reduction in the capacity of exits and a phase transition from efficient free flow to congested flow. The main cause of capacity drop for path planning is mentioned in time as sec. Moreover, when density becomes too large, the dynamics of flow are governed by physical interactions generating hazardous situations. To this end, the Hybrid Reinforcement Agent Learning (HARL) has proven to be a powerful concept in understanding and controlling path planning for crowd flow dynamics.

The simulated scenario consists of agents evacuating from a crowd environment with a single exit. The value of walking speed has calculated in practical terms to a collision of an agents with an obstacle. Crowd evacuees are reached exit by monitoring the walking speed of a people or agent as per Table-2,

S. No.	Number of Agents / People	Walking Speed (in sec's)
1	0-50	0.5
2	50-100	1
3	100-150	1.5
4	150-200	2
5	200-250	2.5
6	250-300	3

Table-2: Crowd evacuees by walking speed

The HARL with IPSO technique is applied in order to identify an impact of crowd dynamism in the ERP problem. The result indicates that dynamism has direct influence on the evacuation plan performance. It fully combines the advantages of improved particle swarm optimization (IPSO) in HARL to simulate the crowd more realistically with the advantages of particle swarm optimization algorithm for global optimization.

VIII. Conclusion

In our proposed work, the path planning using improved particle swarm optimization (IPSO) technique in Hybrid Agent Reinforcement Learning can be applied for simulating the crowd. This technique act well and provide guaranteed collision free crowd movement. The movement of the human crowd is influenced by several factors during the evacuation process. The HARL algorithm will be apt for the number of people in a low to medium density crowd with optimization. If detection of accuracy is improved will leads to better count and estimate population density. It has been proven to be very efficient as regards the evacuation time and decision changes. The dynamic crowd evacuation is to provide insights to support decisions when planning an evacuation procedure. Finally, the usage of models can be allowed for a more automated process for assessing the conformity of the standard process. The experimental evaluation declare that HARL performs better than existing algorithms. Also, Improved PSO technique in HARL provided more detection accuracy.

To summarise, our effort improves the performance in reinforcement learning for crowd simulation. On the above technique, the work carries in future will strengthen the study of learning and try to combine the perception of deep learning with the decision making for crowd evacuation.

References

- [1] H. Shi, S. Yang, K.-S. Hwang, J. Chen, M. Hu and H. Zhang, "A sample aggregation approach to experiences replay of dyna-Q learning," *IEEE Access*, vol. 6, pp. 3717337184, 2018.
- [2] Y. Liu, A. Xu, and Z. Chen, "Map-based deep imitation learning for obstacle avoidance," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 8644–8649, 2018.
- [3] P. Long, T. Fanl, X. Liao, W. Liu, H. Zhang, and J. Pan, "Towards optimally decentralized multi-robot collision avoidance via deep reinforcement learning," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, pp. 6252–6259, 2018.
- [4] Y. Gao, J. Wang, W. Wu, A. Sangaiah, and S.-J. Lim, "A hybrid method for mobile agent moving trajectory scheduling using ACO and PSO in WSNs," *IEEE Sensors*, vol. 19, no. 3, p. 575, Jan-2019.
- [5] Y. Mu, B. Li, D. An, and Y. Wei, "Three-dimensional route planning based on the beetle swarm optimization algorithm," *IEEE Access*, vol. 7, pp. 117804_117813, 2019.
- [6] Q.Wang, H. Liu, K. Gao, and L. Zhang, "Improved multi-agent reinforcement learning for path planning-based crowd simulation," *IEEE Access*, vol. 7, pp. 7384173855, 2019.
- [7] O. Vinyals, I. Babuschkin, W. M. Czarnecki, M. Mathieu, A. Dudzik, J. Chung, D. H. Choi, R. Powell, T. Ewalds, P. Georgiev et al., "Grandmaster level in star craft ii using multi-agent reinforcement learning," *Nature*, vol. 575, no. 7782, pp. 350–354, 2019.
- [8] E. S. Low, P. Ong, and K. C. Cheah, "Solving the optimal path planning of a mobile robot using improved Q-learning," *Robot. Auton. Syst.*, vol. 115, pp. 143161, May 2019.
- [9] B. M. Booth, T. Feng, A. Jangalwa, and S. S. Narayanan, "Toward robust interpretable human movement pattern analysis in a workplace setting," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, pp. 7630–7634, 2019.
- [10] Xue, H.X.; Bai, Y.P.; Hu, H.P.; Xu, T.; Liang, H.J," A Novel Hybrid Model Based on TVIW-PSO-GSA Algorithm and Support Vector Machine for Classification Problems". *IEEE Access*, vol.7, pp.27789–27801, 2019.
- [11] Liu, H.M.; Yan, X.S.; Wu, Q.H, "An Improved Pigeon-Inspired Optimisation Algorithm and Its Application in Parameter Inversion", *Symmetry* 2019.
- [12] X. Li, Z. Lv, S. Wang, Z. Wei, and L. Wu, "A reinforcement learning model based on temporal difference algorithm," *IEEE Access*, vol. 7, pp. 121922–121930, 2019.
- [13] Salleh, Belkourchia, Y, Azrar.L, "Optimization of the shape parameter of RBF based on the PSO algorithm to solve nonlinear stochastic differential equation," 5th International Conference on

Optimization and Applications (ICOA), Kenitra, Morocco, April 2019.

[14] B. Liu, N. Xu, H. Su, L. Wu, and J. Bai, "On the observability of leader-based multiagent systems with fixed topology," *Complexity*, vol. 2019, pp. 1–10, 2019.

[15] H. Su, J. Zhang, and X. Chen, "A stochastic sampling mechanism for time-varying formation of multiagent systems with multiple leaders and communication delays," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 12, pp. 3699–3707, 2019.

[16] X. Li, S. Wang, Z. Lv, Y. Li, and L. Wu, "Strategies research based on chess shape for Tibetan JIU computer game," *International Computer Games Association Journal (ICGA)*, vol. 40, no. 3, pp. 318–328, 2019.

[17] K.Pavithra, G.Radhamani, "A Hybrid Algorithm in Reinforcement Learning for Crowd Simulation," *International Journal of Recent Technology and Engineering (IJRTE)*, ISSN: 2277-3878, Volume-8 Issue-6, pp.5251-5256, March 2020.

[18] Xu Xu, Youwei Jia, Yan Xu, Zhao Xu, Songjian Chai, Chun Sing Lai, "A Multi-Agent Reinforcement Learning-Based Data-Driven Method for Home Energy Management", *IEEE Transactions on Smart Grid* vol. 11, Issue. 4, July 2020.

[19] Libin Liu, Urbashi Mitra," On Sampled Reinforcement Learning in Wireless Networks: Exploitation of Policy Structures", *IEEE Transactions on Communications* (Volume: 68, Issue: 5,) May 2020.

[20] Mingliang Xu , Chaochao Li , Pei Lv , Wei Chen , Senior Member, IEEE, Zhigang Deng , Bing Zhou , and Dinesh Manocha, Fellow," Emotion-Based Crowd Simulation Model Based on Physical Strength Consumption for Emergency Scenarios", *IEEE Transactions on Intelligent Transportation Systems*, 2020.