**CSE 564 - AI Physical Agents**

**Team Number 2 :**

Manju Bisht

1210416246

**Team Number 3 :**

Abhijith Krishnan Radhakrishna Kurup

1211577910

Abhishek Nagaraj

1211417906

Anant Srivastava

1209427008

Ankit Patel

1205972621

Ismael Mercado

1205147771

Prateek Shrivastava

1211160064

Shibani Singh

1209404128

N D B Vachan

1211214768

**INTRODUCTION**

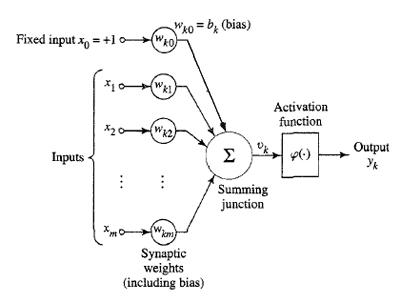
Up to now, we have had major developments for Artificial Intelligence (AI) agents and tools. AI agents take responsibility, take initiatives do not normally interact with others on behalf of a client. An AI tool responds directly to its user, takes responsibility or initiative. Tools do not normally interact with others on behalf of a client. While the field has advanced in fits and starts since, numerous basic artificial intelligence tools exist today, including virtual assistants Siri (Apple) and Cortana (Microsoft), Google Translate, and Netflix’s viewing recommendation engine. Other applications are emerging too, such as self-driving cars, autonomous trading platforms, and emotion-reading humanoid robots.

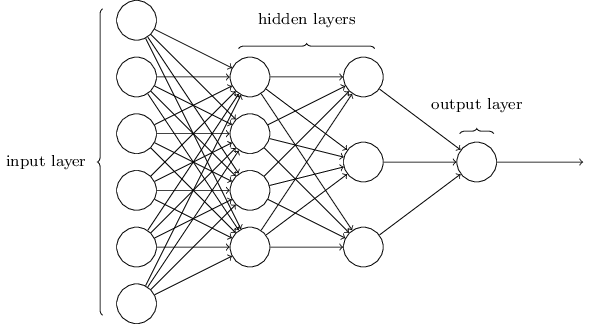
In 1400 BC, Clepsydra, a clock that measured time using the flow of water was created. It's considered one of the first "robotic" devices in history. In 1737, French inventor Jacques de Vaucanson build a clockwork duck capable of flapping its wing quacking, eating and digesting food. In 1961, Unimate, the world’s first industrial robot, goes to work on a General Motors assembly line. In 2000, Honda ASIMO was created. It is a physical agent that is able to walk freely (can change stride speed), able to balance on one foot, able to climb stairs, able to manipulate objects, space & cost-efficient. Later developments include:

1.) Robot Assisted Surgeries  
2.) Military Robots

3.) Swarm Robotics (a new approach to the coordination of multi robot systems which consist of myriad simple physical robots to display collective behavior)  
4.) Nanorobotics: close to the scale of a nanometre (10-9 metres)  
5.) Waalbot (adapting dry-adhesion that lizards use)   
6.) Water-Runner (robot that runs on water, just like the bacillus)  
7.) Slugbot (robot to hunt slugs, ferment them in a separate off-board digester unit. The gas would then be passed through fuel cells to generate electricity that was stored in batteries and could be downloaded to a 'hungry' robot.)   
8.) Ecobot II (Microbial fuel cells, to digest flies. Gets enough energy every 12 minutes to take a step forwards by 2 centimeters).

Robots with consciousness can be developed using Neural Networks. Neural networks consist of neurons. A single neuron is shown in the figure (Figure 1) below, with multiple inputs and one output:

  
**Figure 1: A neuron with inputs and outputs**

  
  
**Figure 2: An Artificial Neural Network**

An artificial neural network (Figure 2) can be defined by three parameters:

1.) Interconnection patterns – Connection patterns between different layer neurons in the network  
2.) Learning Process: Also known as the optimization function/loss function which is used to update the weights of the interconnections  
3.) Activation function: This is applied by every neuron in the network and it converts neuron’s weighted inputs to its output activations

We further discuss Reinforcement Learning, which is unsupervised learning (Learning that is done without any training samples that are manually labelled)

**COLLECTIVE ROBOT REINFORCEMENT LEARNING WITH DISTRIBUTED ASYNCHRONOUS GUIDED SEARCH POLICY**

Global and local workers represent a hierarchical learning pattern. The paper that explained the process used door knob recognition. The paper showed that individual robots could acquire the intricacies of their own situation while contribution to a more generalized understanding of what door knobs are. Frequency of training the robots was also explored, demonstrating the individual agents acquire data very rapidly and the global model is updated much more slowly. Of course, there are some drawbacks to the global-local model, while a robot may spend its time understanding a unique case, it may ultimately not have much to contribute to the global model. Therefore, one must balance the breadth of training data with the level of exploitation of that data. Progress of any one robot may be very high, but ultimately the global model grows at a logarithmic pattern.

The way this system was implemented in the paper was through a communication model consisting of replay models and global parameter optimization. When a robot successfully classifies the door handle, it builds a replay model that it will use for all subsequence door handles. When this fails, the replay model is now shifted more in favor of the new door knob. Of course, this volatile replay model occasionally communicates up a level and parameters concerning the identification of door knobs are tweaked based on the trials.

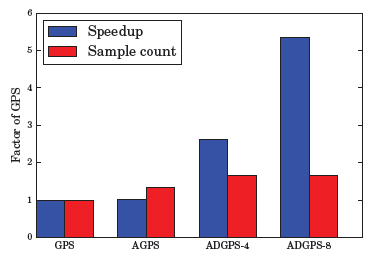
Experimental Evaluations on Asynchronous distributed learning: The experimental evaluations carried out basically answers two questions:

1.       Does distributed asynchronous learning accelerate training of neural network policies?

2.       Does training across groups of robots improve the generalization of the resulting policies?

Two types of evaluations were carried out in order to answer these questions:

i.                    Simulated Evaluation: Number of robots are varied to evaluate training times and study the effects of asynchronous learning. Various GPS techniques were used to carry out these experiments. Speed up and sample counts were recorded and analyzed which resulted in accelerated learning while using Asynchronous Distributed GPS. Hence this approach can be used to reduce the neural network training times.



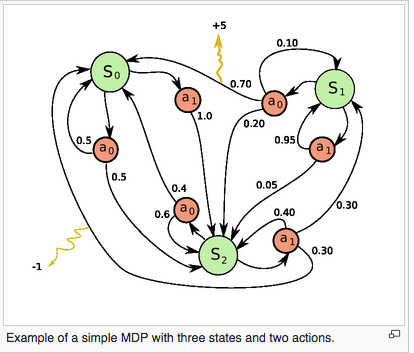
ii                   Real World evaluation: Aims that distributed asynchronous system can effectively learn neural network policies using visual inputs. Vision based door opening policy is used in which cameras and actuators with different calibrations are made to learn opening of doors with different types of knob. This improves the generalization capability of policies in the neural network.



Robots continuously collects new experiences which is used to train neural network policies. Distributed asynchronous policy learning with multiple robots can be thus used as a single generalized skill and can accelerate the learning process.

**DEEP REINFORCEMENT LEARNING FOR ROBOTIC MANIPULATION**

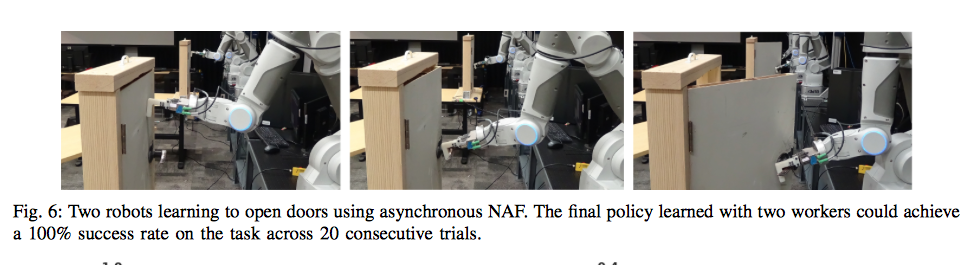
Coming on to the next research, this paper talks in general about the challenges faced by diverse researchers and scientists when talking about Reinforcement learning in regards to human computer interaction when you have multiple physical agents. Q learning is a model free reinforcement learning technique. Specifically Q-Learning can be used to find an optimal action selection policy for any given (finite) Markov decision process (MDP).



This figure is an example of an MDP and in a Deep Q-function there exist an interaction between states, actions and rewards. The figure alongside illustrates the concept that state S0 can act in two ways a0 and a1 if S0 takes action a1 there is a 100% chance of going to S2 if S0 takes the action a0 then there is a 0.5 probability of going back to state S0 and a 0.5 probability of going to S2 again. Using this general idea to form a neural network able to converge a deep Q-Function may be a solution to a supervised learning problem, meaning that to train your network or set of physical agents you need to give solution to the problem before training so that the cost function converges accordingly.

The novelty of this paper is the application of this concept of deep Q-Learning to a set of physical agents in an unsupervised structure, which implies that the agents learn without any supervision. Another highlighted issue was the training time and the real-time challenges faced by researchers before. To tackle this problem the authors designed an asynchronous parallel network that learns collectively from many worker threads and has a core learner thread. Each worker thread is connected to a physical agent and maintain a relay buffer to manage outgoing information. Each worker thread updates the learner thread independently and the learner thread makes necessary changes to the workers as and when the weights and biases changes.

The figure below is a robotic arm training to open a door by pushing the lever down and then pulling the door outwards.



The robotic arm has seven degrees of freedom and they are capable many complex operations, there is a safety concern atop this as the learning algorithm has a huge domain to explore and we should restrict some movements of the robotic arm. The maximum commanded velocity per joint, position limit and bounding sphere for end effector position are regulated.

The system learns how to open a door from scratch in 2.5 hours and without any human supervision. This is the prime of modern computer science in Artificial Intelligence.

**DEEP VISUAL FORESIGHT FOR PLANNING ROBOT MOTION**

Further, we speak about the AI agents in particular. When we talk about AI agents, then the most challenging and inevitable question that arises is - how can we make an AI agent learn on its own? Certainly, we want the AI agents to think and reason like a human. We want them to learn from their own experiences. Around the world, there are many robotics research labs that are working towards this goal. These labs are developing algorithms to make AI agents learn from their experiences with or without human supervision.

In the presentation so far we have seen two such algorithms: reinforcement learning and guided policy search. They both are the best examples of unsupervised learning in AI agents but with some drawbacks. Reinforcement learning can be explained as, to do a task ‘n’ number of times till you reach your goal. An AI agent is given a task and a reward state (goal) is defined, and AI agent will perform the task till it reaches the reward state. The drawback for this algorithm is that it might not provide an optimal solution for the task in hand. The other algorithm for AI agents, Guided Policy Search is a little advance algorithm that makes sure that the optimal solution is found for the task in hand. This also has a drawback that could be dangerous in some cases. To reach the optimal solution, an AI agent could pose as a safety hazard.

Our next algorithm is based on a predictive model. In this the AI agent first perceives the task and predicts the result of its actions, and then act on it. It becomes important to mention here that the AI agent is capable of changing its actions based on the prediction, that is, if optimal path does not result in a safe path. Let’s discuss an experiment to understand the working of this algorithm.

To learn the benefits or advantages about the Model Predictive Control Learning an experiment with the following two questions were conducted.

(1) Is it possible to use action-conditioned video prediction models to manipulate novel objects that were not previously seen during training?

(2) Can video prediction models trained entirely on raw image pixels make meaningful and nontrivial inferences about the behavior of physical objects?

The two questions are answered by conducting qualitative and quantitative experiments.

The question (1) was answered by evaluating our method on new objects not seen during training, and the question (2) were answered through comparison to baseline methods that either move the arm to user-specified positions, or use optical flow to perform continuous re-planning. The question (2) were answered through qualitative experiments aimed to construct physically nuanced pushing scenarios that require reasoning about rotations and centers of mass.

The experiment was conducted with a 7-DoF robot arm to perform the pushing tasks in our experiments, with an RGB camera positioned over the shoulder. The pushing task consists of episodes of length T = 15 where the goal is to move M pixels from their current locations f(x(i)s;y(i)s )g to corresponding goal locations f(x(i)g ;y(i)g )g. Unless otherwise specified, all objects in the experiments were not seen previously in the training set. Images with a resolution of 64\*64 pixels and a planning horizon of H = 3, corresponding to about 800 ms. This allows to re-plan in real time, with new controls computed about every 200 ms. Video prediction model was forced to learn a calibration-agnostic predictive strategy. An optical flow solver supported the system to track the position of the pixel for re-planning. The optical flow is used only during the re-planning phase to provide the initial P(st) distribution. The flow solver is also used to quantitatively evaluate the distance between the final position of the pixel and the goal position of the pixel.

The quantitative comparison were given to three baselines:

1) Select actions randomly from a uniform distribution.

2) Servo the end-effector to the goal pixel position (xg;yg), using a known camera calibration.

3) Servo the end-effector along the vector from the current pixel position dt = (xt ;yt ) to the goal (xg;yg), with continuous re-planning based on the current pixel position estimated using optical flow.

The performance of the experimental method compared to the last two baselines suggests that the experimental method is making meaningful inferences about the motion of objects in response to the arm. Although these results leave significant room for improvement, they suggest that predictive models can be used with minimal prior knowledge to perform robotic manipulation tasks. As video prediction models continue to improve, is expected that the performance of the experimental method will improve with them.

The goal of these experiments is to determine whether or not the model can perform more complex manipulations that require reasoning about object rotations and centers of mass. The benefits of planning actions with a predictive model of all pixels is that the model can be used to plan actions that affect multiple pixels, causing them to move in different directions in a coordinated fashion. The experimented had some limitations which included the difficulty with handling self-occlusions and incorrect model predictions also cause failures in performance.

The method used in the experiment can be used in the future for a lot of advancements in the domain. Since the model is trained entirely through a self-supervised procedure, the method is well suited for continuous self-improvement through constant data collection. The LSTM network is well-suited to learn from experience to classify, process and predict time series. As the accuracy of deep video prediction models improves, we expect the capabilities of this approach to also improve.

**CONCLUSION**

To summarize, we covered various topics such as Use of Neural Networks for training the robots and how they teach other. We saw the concept of a Global worker and the local worker and how local workers optimize their work by learning from the experiences of others. We saw the idea of artificial agents uses the technique of exploration and exploitation to learn and the importance of achieving a balance between these two methods.

We saw the importance of Reinforced Learning and how the increase in the number of workers improved the performance of each one of them. We saw how Predictive Modeling helped robots in moving the objects from one location to another or in picking up a specific object from a pile of distinct objects.

The key feature which is preventing the widespread usage of artificial agents in our day to day life is common sense. Having a large database containing the knowledge of how everything works, the ability to learn and teach other robots, and the ability to understand the context are the three main factors are three main factors which will help in getting the artificial agents close to thinking with common sense. While we have seen the ability of robots to learn and teach other, the other two factors are tough to achieve. Some strides have been made in helping the robots understand the context in which they are working but these are far from the desired levels. And it is anyone’s guess if it is possible to have a database with information about how everything in the world works.

**FUTURE WORKS**

We all have seen the latest research work that is on-going in the field of Artificial intelligence (AI). We are in the era of robots training each other, self-driven cars like Google cars, personal assistants etc. Neural network is the prime topic of AI research and many industrial giants like Google, Microsoft and Facebook are using it to build advanced AI agents. But what is the future of Artificial Intelligence? Where are we really headed?

Father of deep learning, Jurgen Schmidhuber says, ‘By the year 2050, self-replicating robot factories will be present on the asteroid belt’. Self-replicating robots are the robots which are able to build other robots from the scratch. These can be considered as a form of reproduction that humans are capable of. When these are placed on the asteroid belt, these can be used in mining the asteroid belt for many metals and metal ore that are very valuable on the earth. As mentioned previously, the robots are learning more and the ability to learn from the database. By this, the robots can take their own decisions without any human intervention. This trait will be very useful in space exploration as not even 0.01% of the total universe is explored and these AI agents can take their own decisions which exploring the universe.

Google’s plan of building the next generation personal assistant is going on to build an agent which is able to recognize human emotions and reciprocate to it. For that, the AI agent needs to understand the facial expressions. Recently, an AI agent in MIT has learnt how to recognize faces as human do. This is a big step in building human-like robots. This will enable the robot to understand the context and understand how everything works.

In conclusion, the AI agents which are being built in the future can do three main things, self-replicate, understand emotions and reciprocate to it and think independently. Give it a thought, these are the exact unique things humans have the ability to do. Going by this trend, Artificial intelligence agents can be just like humans or may be, even better than us.

**References**

1.<https://arxiv.org/pdf/1610.00633v1.pdf>

2. <https://arxiv.org/pdf/1603.02199.pdf>

3.<http://futurism.com/dawn-of-synthetic-reason-deepmind-can-learn-from-its-own-memory/>

4.<http://futurism.com/ai-will-colonize-the-galaxy-by-the-2050s-according-to-the-father-of-deep-learning/>

5.<https://arxiv.org/pdf/1610.00673v1.pdf>

6.<https://arxiv.org/pdf/1610.00696v1.pdf>