Overview of Thesis/ Project

Teaching a robot to do a particular task. Example picking up garbage or cans in a given room. Such a task can be done using traditional AI algorithms like reinforcement learning.

RL is basically defining a Markov Decision Process which is some states and reward to be given when a particular action is taken in a state. The RL algorithm learns an optimal policy ie a state action pair for every state. But in practical conditions when the state space is very high using reinforcement learning is not feasible because the numbe of iterations taken by the robot to try every state action combination is almost infinite. Making the robot learn time critical problems like, avoiding obstacles or driving on a highway is not practical using RL.

Thesis:

Learning from demonstrations is a growing area of machine learning/AI in which programming of autonomous agents is done by demonstrating the desired behaviour or task. In deomnstration based approach a teacher typically a human shows the agent how to perform the task. The agent/robot records the state/action pair to learn a policy that reproduces an observed behaviour. Learning from demonstations is inspired by the way humans and animals teach each other, and aiming to prvide an intuitive way to transfer knowledge to autonomous systems. compare to exploration based methods like traditional RL , LFD reduces the learning time and also eliminates the frequent difficult task of defining detailed reward functions.

Example:

The demonstration part would be to tell the robot/agent whether the object it is looking at is a can or not.

Now the problem is reduced to building an image classifier to classify whether the given image has a can or not. Further what I have done is I have combined reinforcement learning and an image classification task and taking demonstration only when the classifier is uncertain. What do I mean by the classifier is uncertain?

Example:

A classifier takes in some information(features/attributes) and provides some output (label).

What is the hypothesis for a patient given his/her symptoms.

We have collected information from the past like what is the probability of having symptoms like

headache, cough, fever and the hypothesis is flu,meningitis , lupus.

$$P(h \mid E_1, E_2 \ldots E_n) = \frac{P(E_1, E_2, \ldots E_n \mid h) \times P(h)}{P(E_1, E_2 \ldots E_n)}$$

$$\arg\max_h P(h \mid E_1, E_2 \ldots E_n) = \arg\max_h P(E_1 \mid h) \ldots P(E_n \mid h) \times P(h)$$

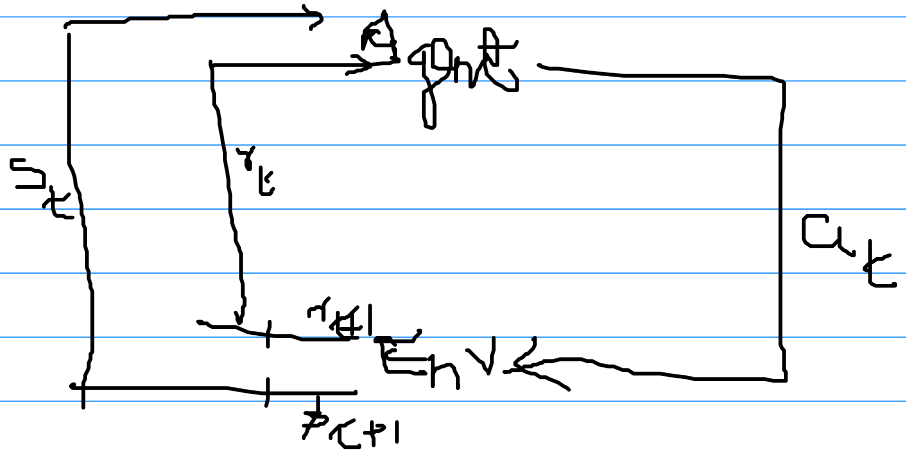
sc

$$P(h_i|E_1 \ldots E_n) = \alpha P(h_i) \times \prod_j P(E_j|h_i)$$

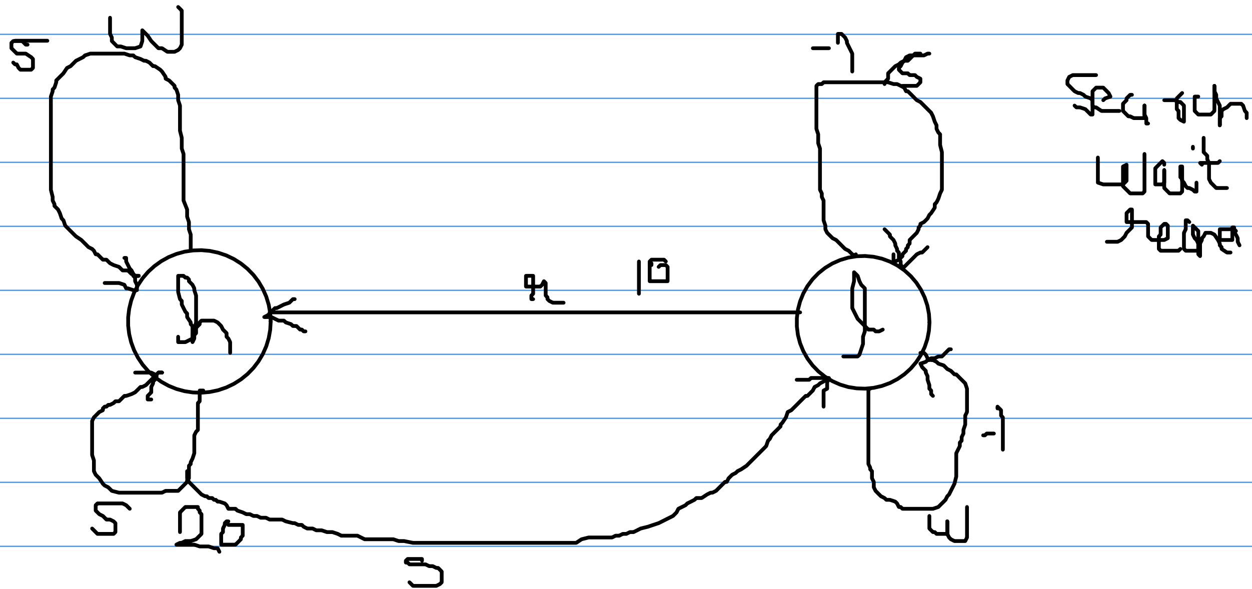
$$\alpha = \frac{1}{\sum_i P(h_i) \prod_j P(E_j|h_i)}$$

Reinforcement Learning

Reinforcement learning is learning from interaction with the environment to achieve a specific goal.



The goal of the agent is formalized in terms of special rewards passed from the enviromment to the agent.



Agents Goal is to maximize rewards it receives in the long run.

R = r1 + r2 + r3 +...

Discounted return.

R = r1 + $\gamma R_2 + \gamma^2 R_3 + \ldots$

$0 \leq \gamma \leq 1$

If discount factor close to 0 action with high immediate reward is selected.
If discout factor close to 1 is selected actions with high future rewards is selected.

$$Q(s,a) = Q(s,a) + \alpha\left[r_{t+1} + \gamma Q(s,a') - Q(s,a)\right]$$

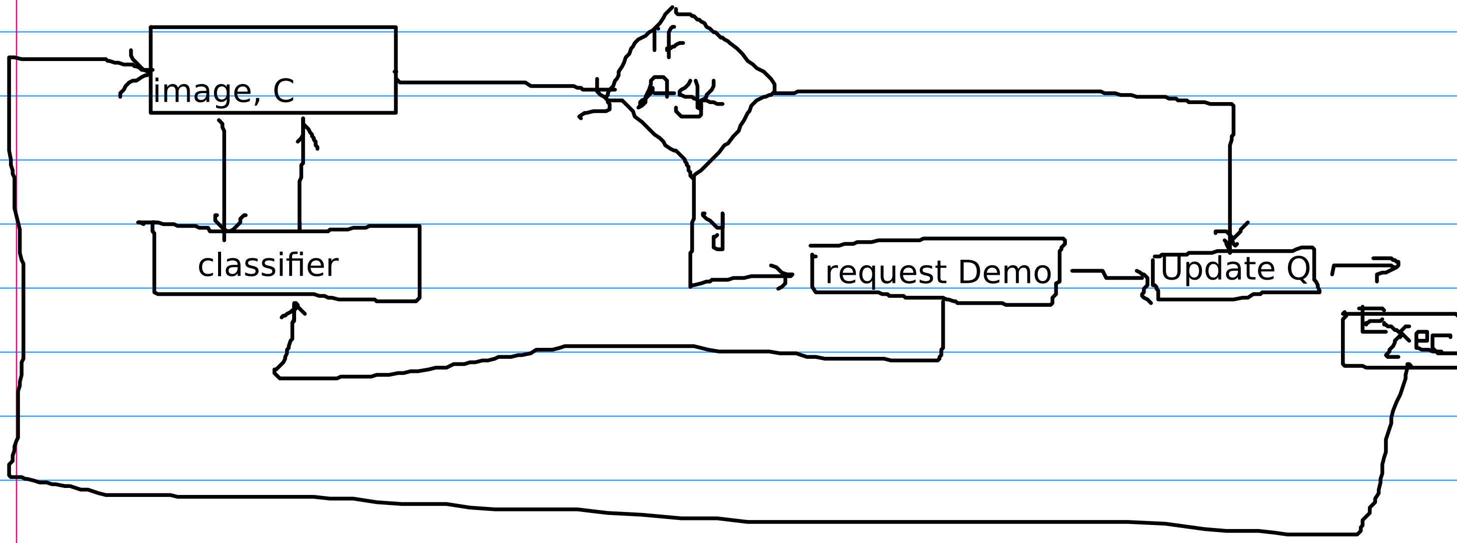
$$s \leftarrow s', \quad a \leftarrow a'$$

Q(s,a) - estimate of how good it is to take an action in state S.

Thesis:
I have taken a two dimensional grid. And each cell in this grid has a handwritten digit image. Hand written digit images from 0-9 are scatteredrandomly in the grid and the task of the robot is to pick a specific digit. Initially the LR classifier is trained using a few 100 images, but the classifer is not very matured. Even the policy of the reinforcement learning is not yet optimal.

My state in QLearning is made up of binary image and a discretized certainty value returned by the classifer.

		L	R	P	A	(Q values)
Cell 0	1					
	2					
	3					
	⋮					
	10					
cell 1	1					
	2					
	⋮					
	10					



there will be a lot of queries to the human to classify a given image. As the iterations increase the reinforcement learner learns from the reward structure a policy to make human queries when the classifier certainity for a particular instance is low and if the classifier probability or certainity is high the robot goes ahead with the classifed value and takes the action of picking up if there is an image with the specific digit.

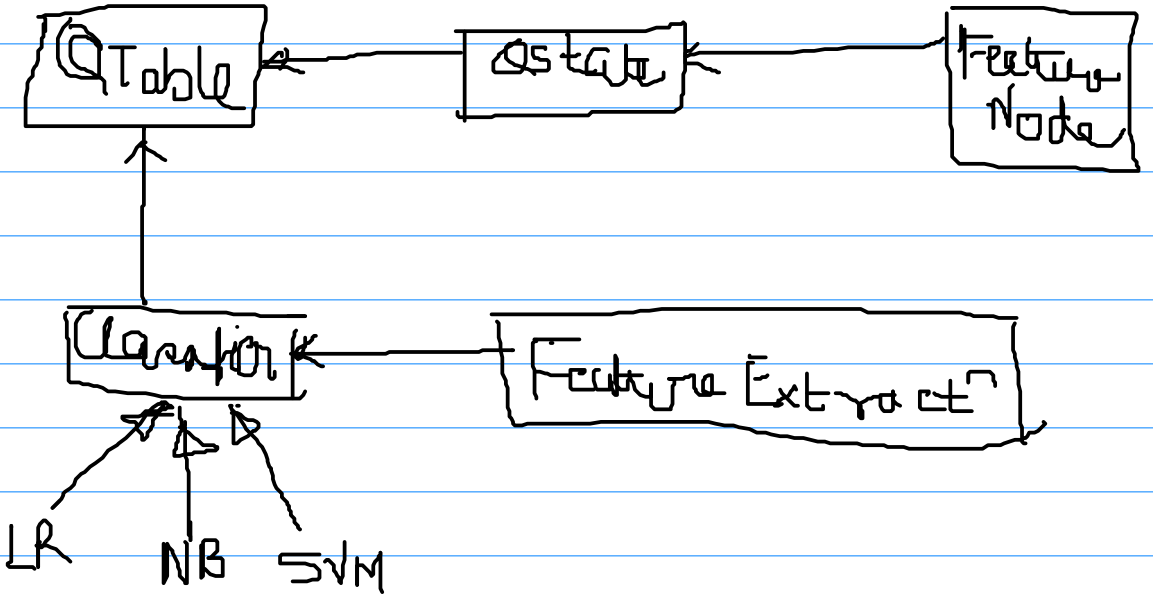
Goals Achieved: Less number of iterations of the RL. Also the rewards learnt are better compared to the basic RL. When the task is critical, for example what if the robot is trying to avoid obstacles on ots path.
The classification needs to be real time. A matured robot does minimum to none human queries.

.

Problems faced:

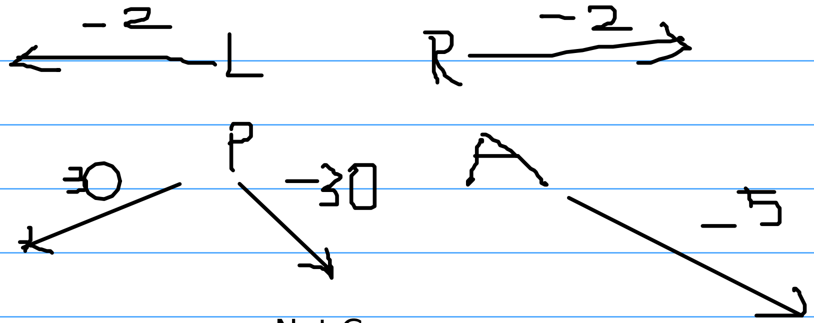
What if big data is involved. The response for an image classification task would not be in real time. I have used the hadoop framework to divide the machine learning task to multiple cores. The current training data is divided into number of cores/cpus available. The Logistic regression is plit into map and reduce jobs which helps in paralleg computation of classification.

Classes:



Module:
Q learning
Logistic Regression
Image Processing

Reward Structure - Intuition difficult



Can

Not Can

Because the reward for pickup is high if there is a can, the RL learns to pickup when certainty provided by classifier is high. The RL learns not to pickup and just go left or rig when certainty provided by the classifier is low due to the very large negative reward.

Depending on how matured the classifier is the RL makes optimal ask queries when the classifer is uncertain ie certainty level between 4-8

Outline:
1) Teachine a robot - example - why not feasible
2) Learning from demonstrations

3) Problem reduced to image classification

4)What is a classifier with simple patient example
5)What is RL with simple example
6) Actual combination of Classification and LR - with current domain
7) Class diagram
8) Goals Achieved
9)Problems faced

10) Intuition of reward structure if time