ARTIFICIAL INTELLIGENCE & MACHINE LEARNING

UNIT – 4

REINFORCEMENT LEARNING

Dr. G. Ravi & Dr. S. Peerbasha
PG & Research Department of Computer Science
Jamal Mohamed College (Autonomous)
Trichy-620 020.

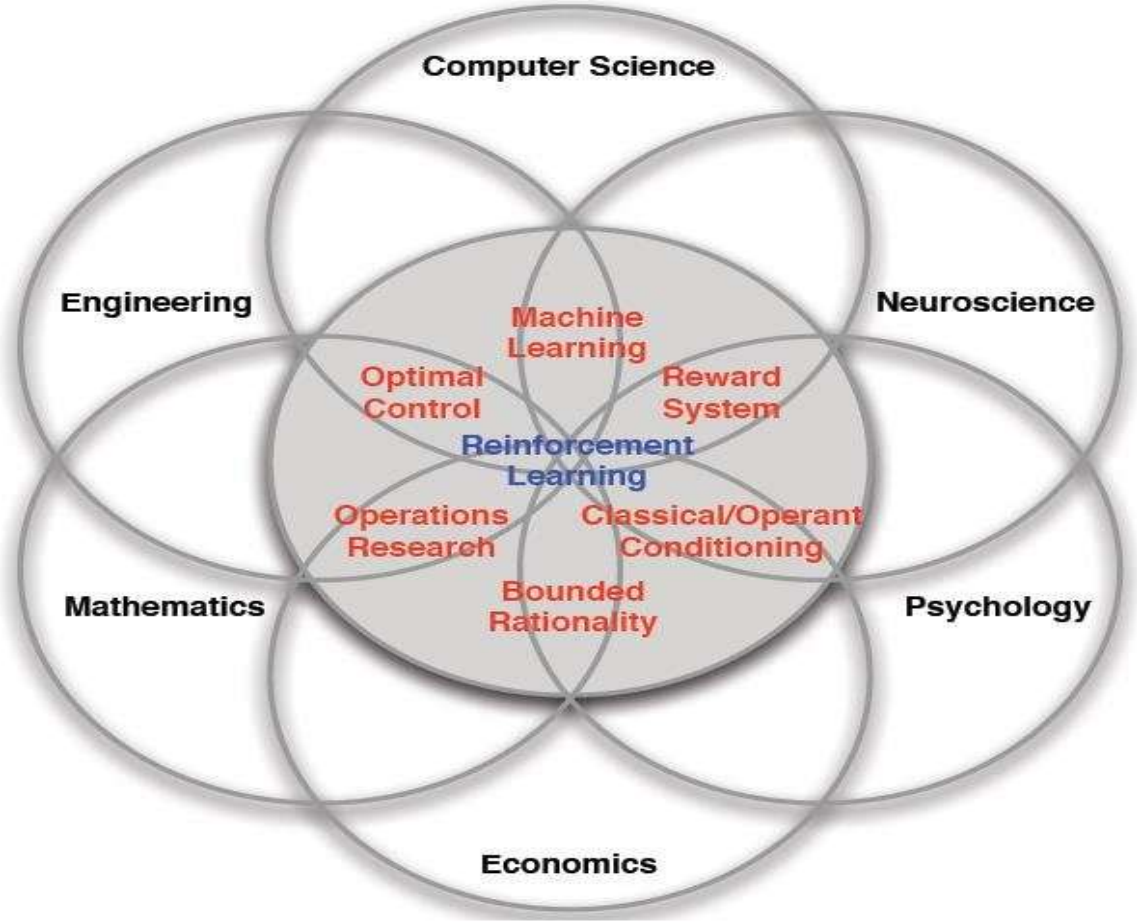
UNIT - 4

REINFORCEMENT LEARNING

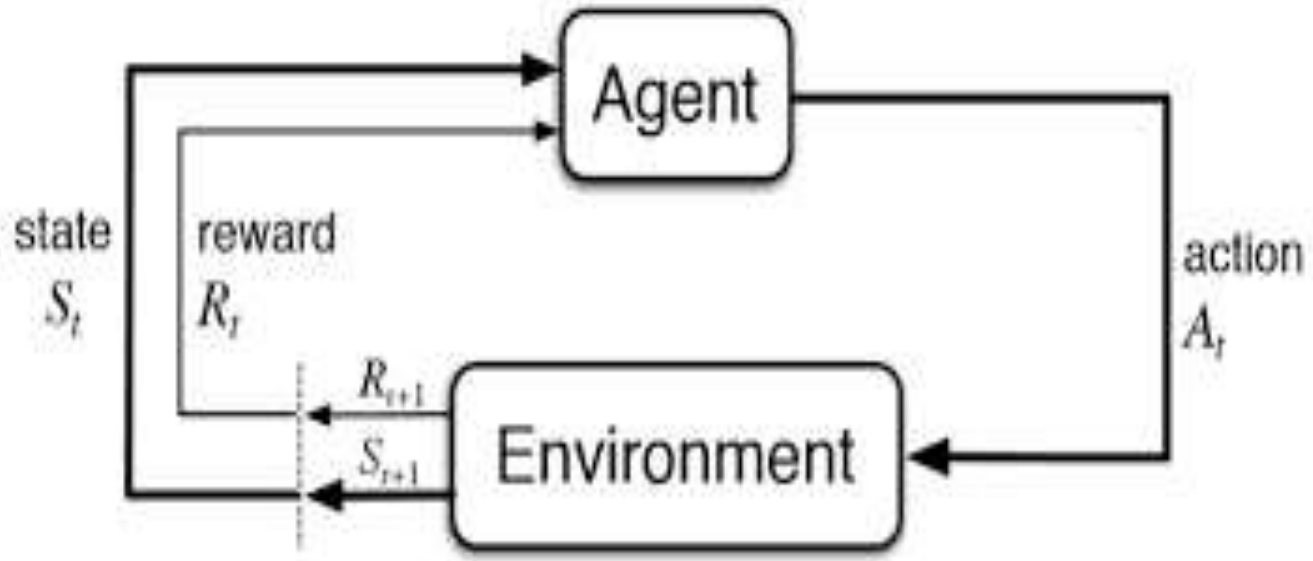
REINFORCEMENT LEARNING (Learning by environment)

- It is a type of machine learning where learning occurs by interacting with the environment.
- Goal-directed learning process
- Reinforcement Learning has an agent which learns from Cause and effect.
- This agent selects the actions from its past experiences and through choices, which can be treated as a TRIAL-AND-ERROR method.

MANY FACES OF REINFORCEMENT LEARNING MODEL



REINFORCEMENT LEARNING MODEL



REINFORCEMENT LEARNING

It is mainly aimed at solving the so-called MARKOV DECISION PROBLEM (MDP).

We can also say that the problems that are being solved by reinforcement learning will either be a Markov's Decision Problems

MARKOV DECISION PROBLEM

Markov Decision Problem (MDP) also has a set of framework elements. They are:

1. State of the system

2. Actions

3. Transition probabilities

4. Transition rewards

STATE

The state describes the present condition of the system with the help of a set of attributes or parameters.

It can either be a time dependent or time independent system.

If the system of time dependent, it is called a dynamic system. (i.e) the state changes as the time moves.

Example:-

Queue in a Supermarket. Here, the state of such a system can be defined by the number of people in the queue.

Thus, it is sure that the state changes with time; when a customer is being served and moved out or when a new person comes to the queue, the state changes.

ACTIONS

Actions may be defined as the changes that need to be done when the system attains a particular condition.

For the above-mentioned example, we can say that whenever the number of customers exceeds a particular number, the remaining people need to be diverted to a new counter that is opened.

So here, we can have two actions, namely opening a new counter and not opening a new counter.

TRANSITION PROBABILITY

Consider the situation where one state changes to another under the influence of an action.

Let a be the action at a state i.

By this action a, it changes to another state j.

We can denote this as p(i,a,j) which denotes the probability of going from state i to j under the actions a. This is termed as Transition probability.

3 states and 2 actions in a Markov decision problem will result in 9 transition probabilities per action.

IMMEDIATE REWARDS

The system always gets an immediate reward in transition from one state to another.

It is denoted by r(i,a,j).

The system changes its state frequently; the action that needs to be selected in every state is defined by the term policy.

There are cases where no action is needed to be done for some states.

But, the states which make some actions to be chosen are called decision-making states.

PERFORMANCE METRIC

Every policy will be associated with a performance metric which judges the performance of the policy.

Our aim is to have the policy which gives the best performance matrix.

One such performance matrix is average reward of a policy.

TIME OF TRANSITION

Time of transition is the time taken by the Markov decision problem for transition from one state to another.

Normally, it is assumed to be one.

It does not have any meaning like 1 hour, 1 minute or 1 second. It is just a fixed quantity by analyst.

CURSES

We have generally two concepts: curse of modelling and curse of dimensionality.

In cases where we have a large number of governing variables, it is quite difficult to derive the exact values. It is called the curse of modelling.

LIMITATION OF REINFORCEMENT LEARNING

When the problems become complex, it is difficult to store the values of each state in memory. To solve this, we can introduce techniques for value estimation like neural networks and decision trees.

CURSES

We have generally two concepts: curse of modelling and curse of dimensionality.

In cases where we have a large number of governing variables, it is quite difficult to derive the exact values. It is called the curse of modelling.

APPLICATIONS OF REINFORCEMENT LEARNING

➔Robotics: Navigation, Robo-soccer, Walking, Juggling

➔Control: Factory processes, admission control in telecommunication, resource control in multimedia networks

➔Games: Backgammon, chess, othello

➔Operations research: Warehousing, transportation, scheduling

➔Others: Adaptive treatment design, biological modelling

➔Combinational optimisation: VLSI Placement and routing, elevator dispatching

THANK YOU