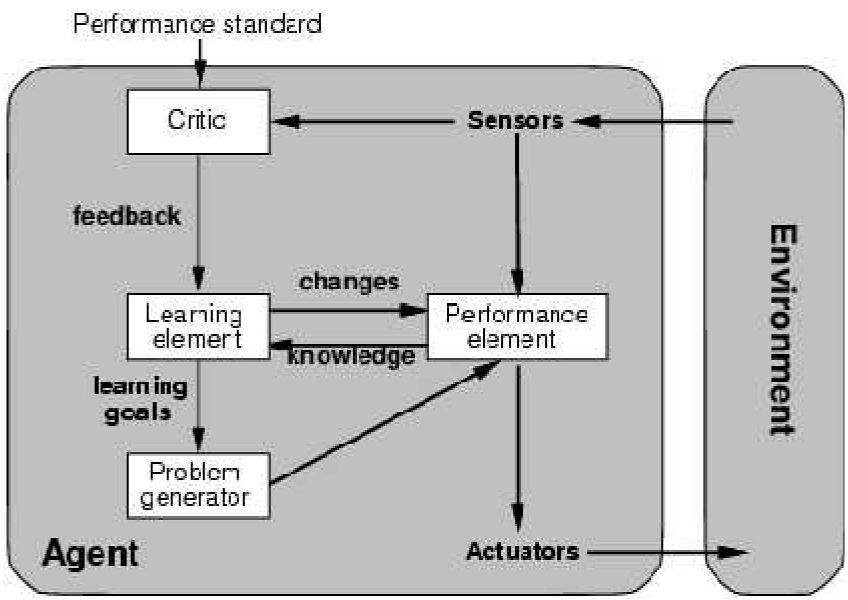
GKMCET

Lecture Plan

Subject code & Subject Name: CS2351 & AI **Unit Number: V**

UNIT V-LEARNING

LEARNING AGENT



All agents can improve their performance through **learning.**

A learning agent can be divided into four conceptual components, as shown in Figure 1.15 The most important distinction is between the **learning element**, which is responsible for making improvements, and the **performance element**, which is responsible for selecting external actions. The performance element is what we have previously considered to be the entire agent: it takes in percepts and decides on actions. The learning element uses feedback from the **critic** on how the agent is doing and determines how the performance element should be modified to do better in the future.

The last component of the learning agent is the **problem generator**. It is responsible for suggesting actions that will lead to new and **informative experiences**

1 Forms of learning

- ***Supervised***
- ***unsupervised***

- ***Supervised***

Need supervisor to carry out the learning process

- ***UnSupervised***

Does not Need supervisor to carry out the learning process

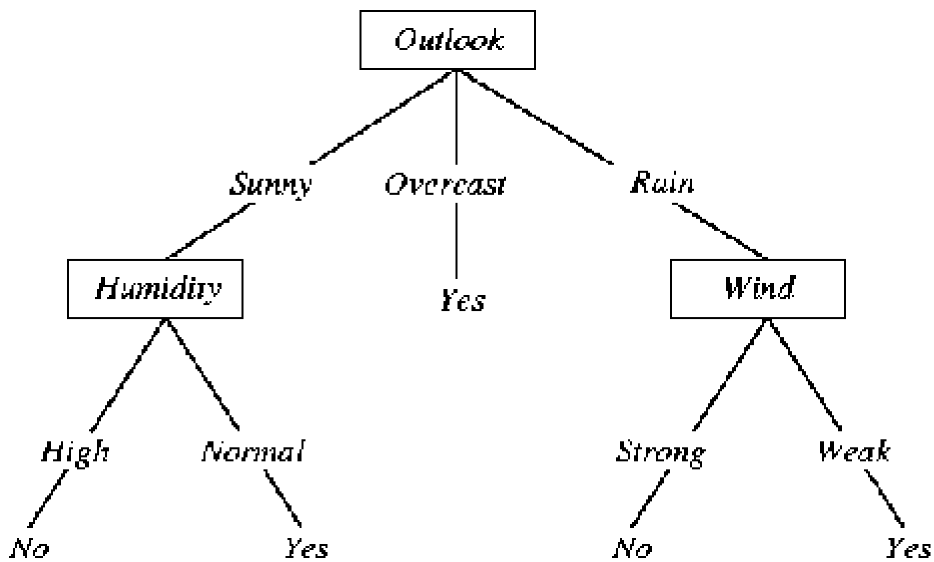
http://csetube.co.nr/

GKMCET

Lecture Plan

Subject code & Subject Name: CS2351 & AI **Unit Number: V**

Decision tree learning



Decision Trees

_ Decision tree representation

_Each internal node tests an attribute

_Each branch corresponds to attribute value

_Each leaf node assigns a classification

GKMCET

Lecture Plan

Subject code & Subject Name: CS2351 & AI **Unit Number: V**

Day	Outlook	Temperature	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

5 Neural networks

- *Artificial neuron vs biological neuron*
- *neural net architecture*

GKMCET

Lecture Plan

Subject code & Subject Name: CS2351 & AI **Unit Number: V**

Reinforcement learning

- ***Active Reinforcement learning***
- ***Passive Reinforcement learning***

Reinforcement learning

•Frequency of rewards:

–E.g., chess: reinforcement received at end of game

–E.g., table tennis: each point scored can be viewed as reward

http://csetube.co.nr/

learninggoalsknowledgeEnvironmentSensorsActuatorsCriticAgentLearning elementPerformance elementProblem generatorPerformance standardchangesfeedback•reward part of the input percept•agent must be hardwired to recognize that as reward and not as another sensory input•E.g., animal psychologists have studied reinforcement on animals

Passive reinforcement learning

•Direct utility estimation

•Adaptive dynamic programming

•Temporal difference learning

–Active reinforcement learning

•Exploration

•Learning an Action-Value Function

Active Reinforcement learning

http://csetube.co.nr/

CSE Tube

GKMCET

Lecture Plan

Subject code & Subject Name: CS2351 & AI **Unit Number: V**

The agent‘s policy is fixed

–in state s, it always executes the action π(s)

•Goal: how good is the policy?

•The passive learning agent has

–no knowledge about the transition model T(s,a,s‘)

–no knowledge about the reward function R(s)

•It executes sets of trialsin the environment using its policy π.

–it starts in state (1,1) and experiences a sequence of state transitions until it reaches one of the terminal states (4,2) or (4,3).

•E.g., (1,1)-0.04 (1,2)-0.04 (1,3)-0.04 (2,3)-0.04 (3,3)-0.04 (3,2)-0.04 (3,3)-0.04 (4,3)+1

•Use the information about rewards to learn the expected utility Uπ(s):

Utility is the expected sum of (discounted)rewards obtained if policy πis followed

http://csetube.co.nr/

Adaptive dynamic programming (1)

•Idea: Learn how states are connected

•Adaptive dynamic programming (ADP) agent

–learns the transition modelT(s, π(s), s’)of the environment

–solves the Markov decision process using a dynamic programmingmethod

•Learning transition model is easy fully observable environment

–supervised learning taskwith input = state-action pair, output = resulting state

–transition model can be represented as table of probabilities

•how often do action items occur estimate transition probability T(s,a,s‘) from the frequency with which s‘is reached when executing a in s.

GKMCET

Lecture Plan

Subject code & Subject Name: CS2351 & AI **Unit Number: V**

•E.g., from state (1,3) *Right*is executed three times. The resulting state is two times (2,3) → T((1,3) , *Right*, (2,3)) is estimated to be 2/3.

Temporal difference learning (1)

•Idea: Use the best out of Direct Utility Estimation and AdaptiveDynamic Programming

•Use the observed transitions to adjust the values of the observed states so that they agree with the constraint equations

•Example:

–<1>: (1,1)-0.04→ (1,2)-0.04→ (1,3)-0.04→ (1,2)-0.04→ (1,3)-0.04→ (2,3)-0.04→ (3,3)-0.04→ (4,3)+1

–<2>: (1,1)-0.04→ (1,2)-0.04→ (1,3)-0.04→ (2,3)-0.04→ (3,3)-0.04→ (3,2)-0.04→ (3,3)-0.04→ (4,3)+1

–After first trial, $U^{\pi}(1,3) = 0.84$ and $U^{\pi}(2,3) = 0.92$

–If the transition from (1,3) to (2,3) occured all the time, the utilities should obey:

$U^{\pi}(1,3) = -0.04 + U^{\pi}(2,3)$

so $U^{\pi}(1,3)$ would be 0.88

→ current estimate of 0.84 too low, should be increased

•Generally:

$U^{\pi}(s) \leftarrow U^{\pi}(s) + \alpha(R(s) + \gamma U^{\pi}(s') - U^{\pi}(s))$

with αas learning parameter

•Update rules uses difference in utilities between successive states → temporal-difference(TD) equation

http://csetube.co.nr/

GKMCET

Lecture Plan

Subject code & Subject Name: CS2351 & AI **Unit Number: V**

http://csetube.co.nr/

***Q*-Learning**

•*Q*-learning: alternative TD method

–learns action-value representationinstead of utilities

–*Q(a,s)*denotes value of doing action *a*in state *s*

–A TD agent that learns a *Q*-function does not need a model for either learning or action selection model-free

•*Q*-learning agent learns optimal policy, but much slower than ADP agent

–TD does not enforce consistency among values via the model

- ***STATISTICAL learning methods***
- ***Instance based learning***
- ***EXPLANATION BASED LEARNING***
- ***LEARNING WITH HIDDEN VARIABLES***
- ***LEARNING USING RELEVANCE INFORMATION***
- ***INDUCTIVE LEARNING***

GKMCET

Lecture Plan

Subject code & Subject Name: CS2351 & AI **Unit Number: V**

http://csetube.co.nr/

http://csetube.co.nr/