Neurological and Computational aspects of Reinforcement Learning

Jessy John
Kiran Kumar Joseph
Sreejith P. K.

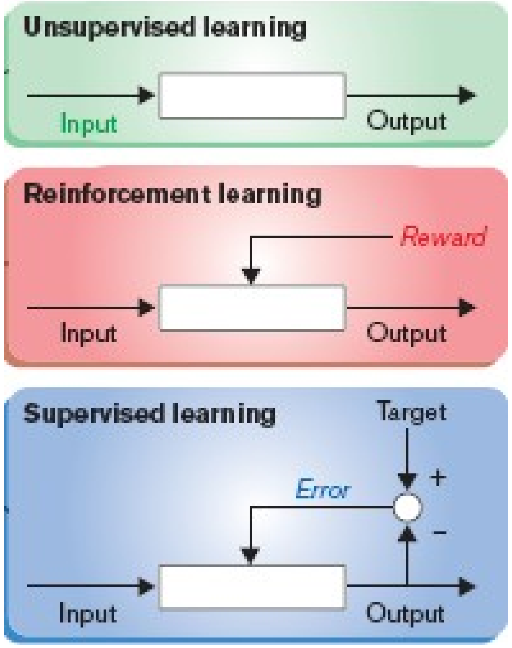
Organization of the seminar

- ➢ Introduction
- ➢ Neurological aspects of reinforcement learning
- ➢ Computational aspects of reinforcement learning
- ➢ Conclusion

Introduction

- Learning
  - ✓adaptive changes in system
  - ✓ enable the system to do the same or similar tasks more effectively the next time

- Types
  - ➢Unsupervised
  - ➢Supervised
  - ➢Reinforcement



Types of learning (Source: Doya, 2000)

Reinforcement learning

- Defined by Sutton and Barto

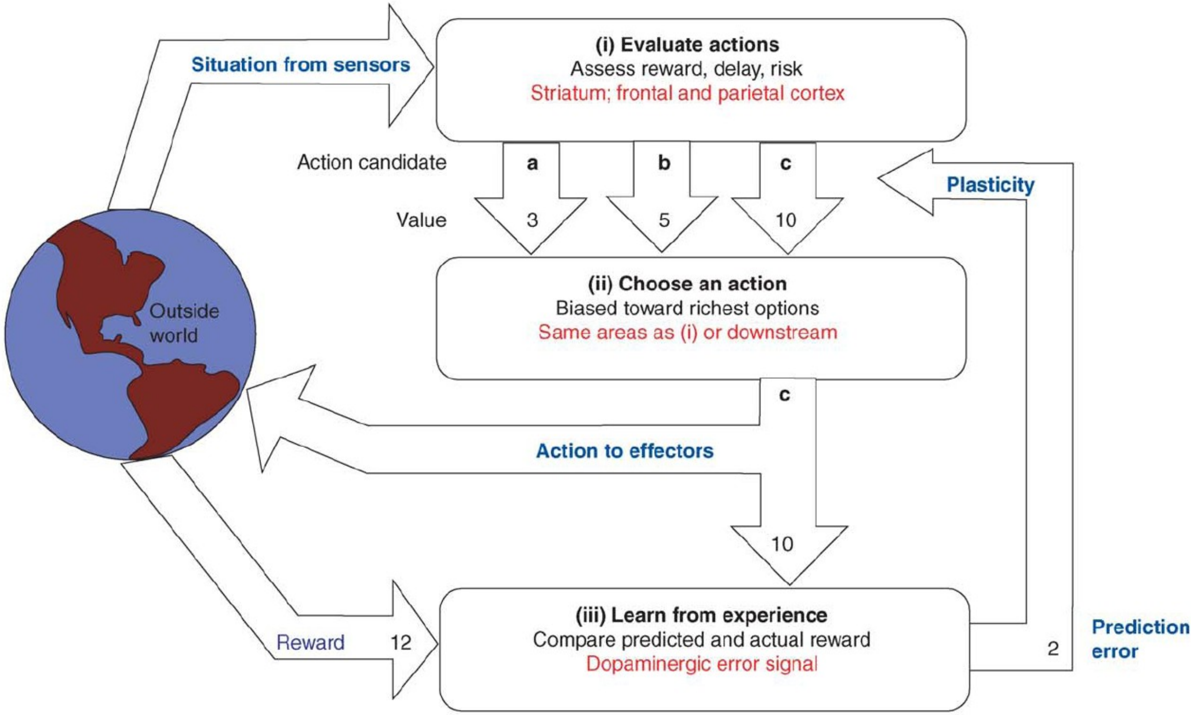
“branch of AI that deals with how an agent e.g. robot learns by trial and error to make decisions in order better to obtain rewards and avoid punishments”

Daw et. al., 2006

Reinforcement learning contd….

- Trial and error process using predictions on the stimulus

- Predict reward values of action candidates
- Select action with maximum reward value
- After action, learn from experience to update predictions so as to reduce error between predicted and actual outcomes next time


Situation from sensors
(i) Evaluate actions
Assess reward, delay, risk
Striatum; frontal and parietal cortex
Action candidate
a
b
c
Value
3
5
10
Plasticity
Outside world
(ii) Choose an action
Biased toward richest options
Same areas as (i) or downstream
c
Action to effectors
10
(iii) Learn from experience
Compare predicted and actual reward
Dopaminergic error signal
Reward
12
2
Prediction error

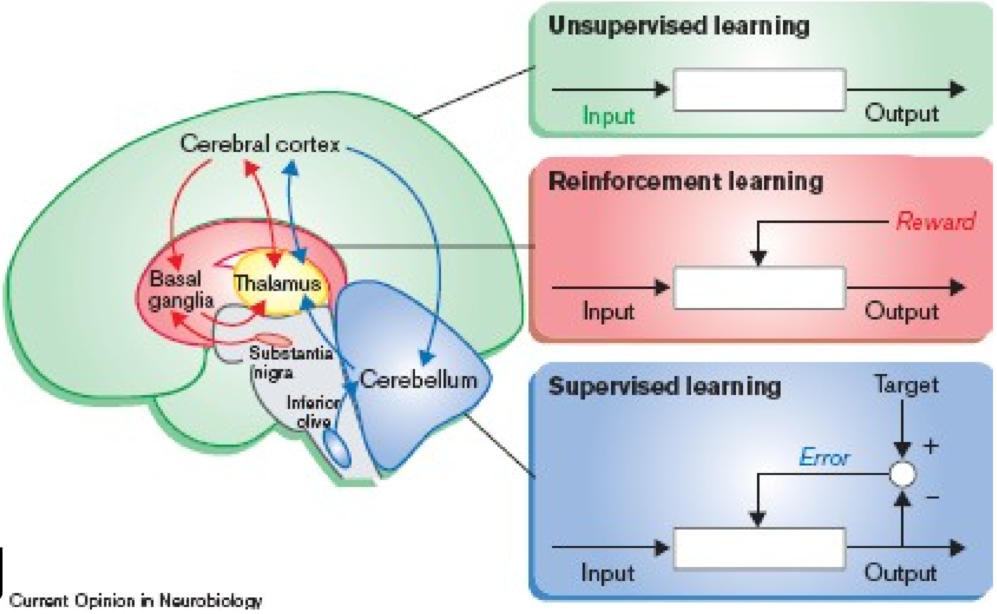
Schematic showing the mechanism of reinforcement learning

(Source: Daw et. al. 2006)

Neurological aspects of reinforcement learning

Learning

- Learning
  - Cerebral cortex
  - Cerebellum
  - Basal ganglia



- Reinforcement/
  Reward-based learning

Types of learning and the associated brain areas

(Source: Doya, 2000)

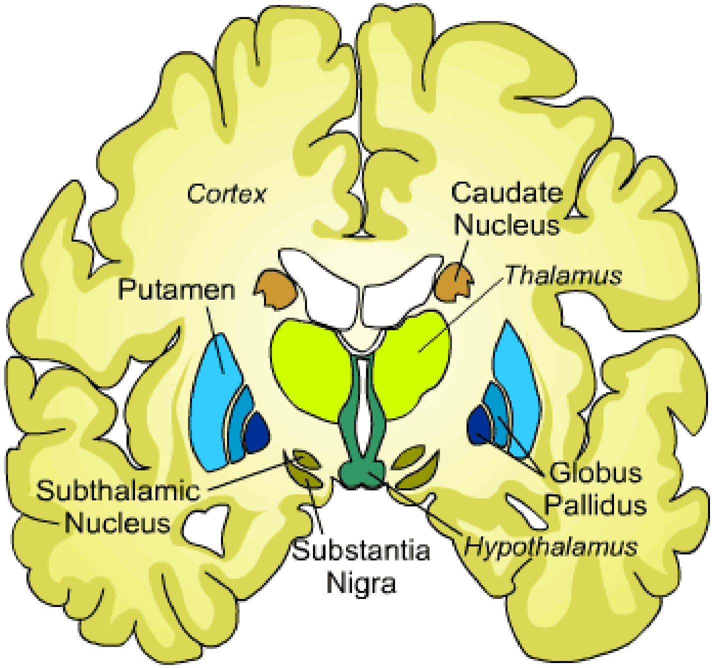
- Methodologies
  - Prediction learning using classical or Pavlovian conditioning
  - Action learning using instrumental or operand conditioning

Structures of the reward pathway

- Areas involved in reward-based learning and behavior
    - Basal ganglia
    - Midbrain dopamine system
    - Cortex
- Additional areas of reward processing
    - Prefrontal cortex
    - Amygdala
    - Ventral tegmental area
    - Nucleus accumbens
    - Hippocampus

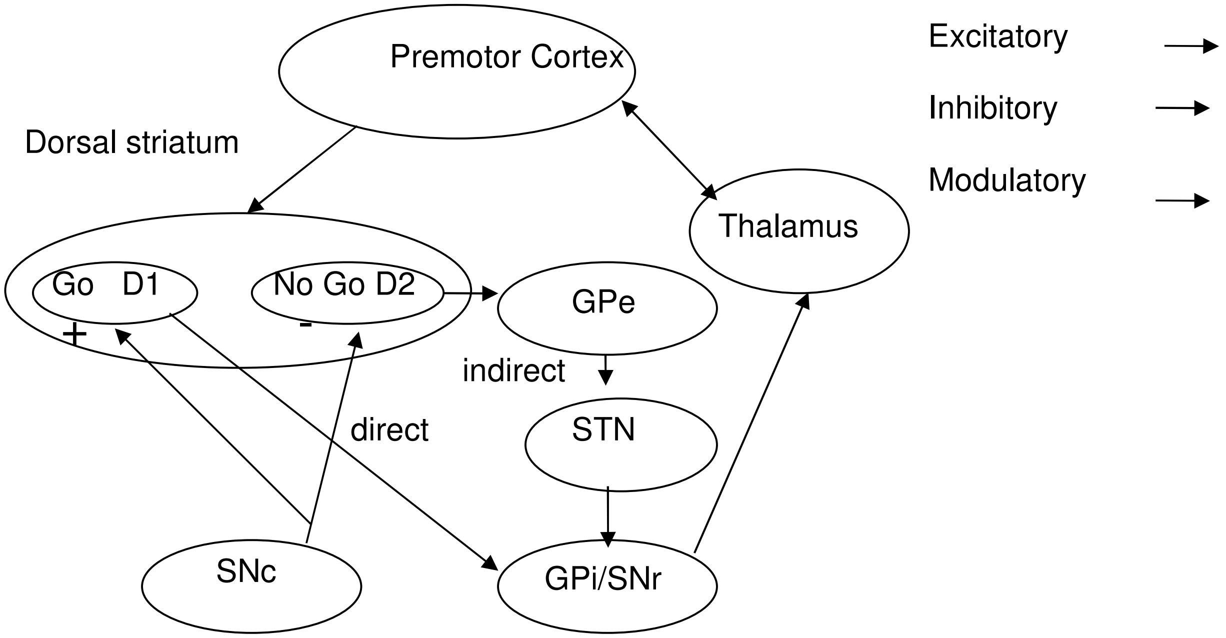
Basal Ganglia

- Group of subcortical structures
  - ➢Striatum
    - ▪ Major input structure - from cortex, thalamus
    - ▪ Dorsal and ventral striatum
    - ▪ MSP neurons
      - o Having D1 receptors
      - o Having D2 receptors
    - ▪ Output to cortex via thalamus through two pathways
  - ➢Globus pallidus – GPi and GPe
  - ➢Substantia Nigra – SNc and SNr
  - ➢Subthalamic nucleus



Basal ganglia and constituent structures (Source: http://www.stanford.edu/group/hopes/basics/braintut/f_ab18bslgang.gif)

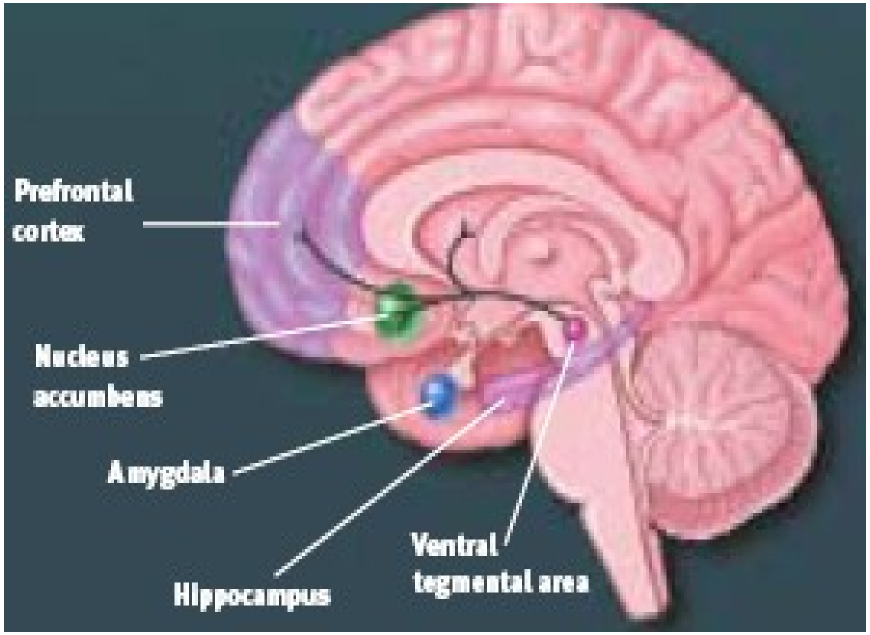
Output pathways of basal ganglia



Direct and indirect pathways of basal ganglia for deciding the behavioral strategy

Other brain areas

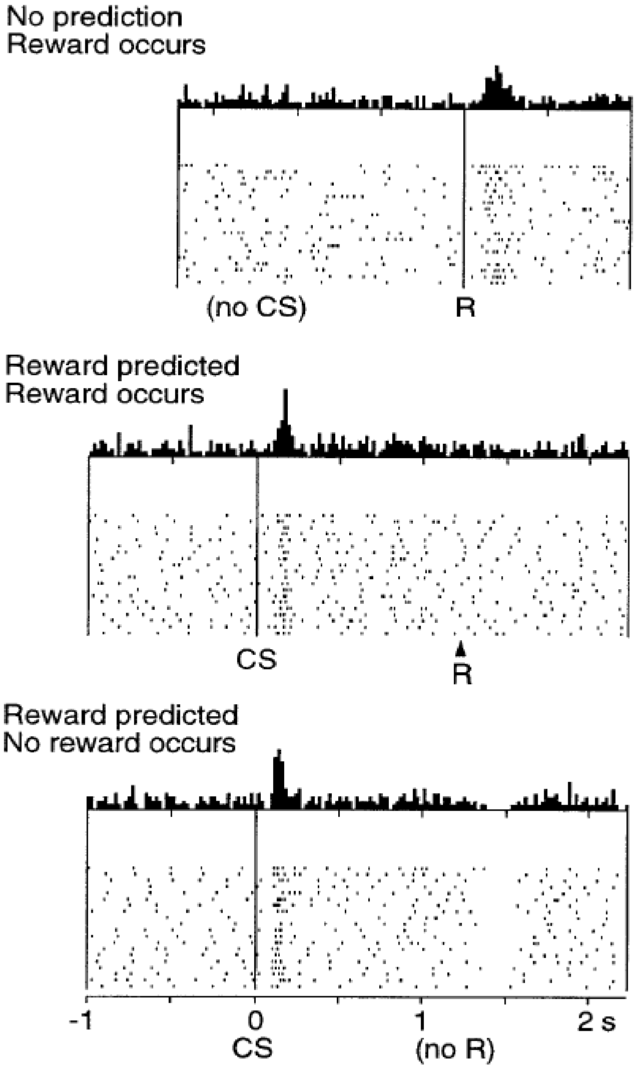
- Prefrontal cortex
  - ➢ Working memory to maintain recent gain-loss information
- The amygdala
  - ➢ Processing both negative and positive emotions
  - ➢ Evaluates the biological beneficial value of the stimulus
- Ventral tegmental area
  - ➢ Part of DA system
- Nucleus accumbens
  - ➢ Receives inputs from multiple cortical structures to calculate appetitive or aversive value of a stimulus
- Subiculum of hippocampal formation
  - ➢ Tracks the spatial location and context where the reward occurs



Structures of reward pathway (Source: Brain facts: A primer on the brain and nervous system)

Role of dopamine neurons

- Two types
  - From VTA to NAc
  - From SNc to striatum
- Phasic response of DA neurons to reward or related stimuli
  - Process the reward/ stimulus value to decide the behavioral strategy
  - Facilitates synaptic plasticity and learning
- Undergo systematic changes during learning
  - Initially respond to rewards
  - After learning respond to CS and not to reward if present
  - If reward is absent depression in response



Phasic response of dopamine neurons to rewards (Source: Schultz, 1998)

Role of dopamine neurons contd….

  - Response remain unchanged for different types of rewards
  - Respond to rewards that are earlier or later than predicted

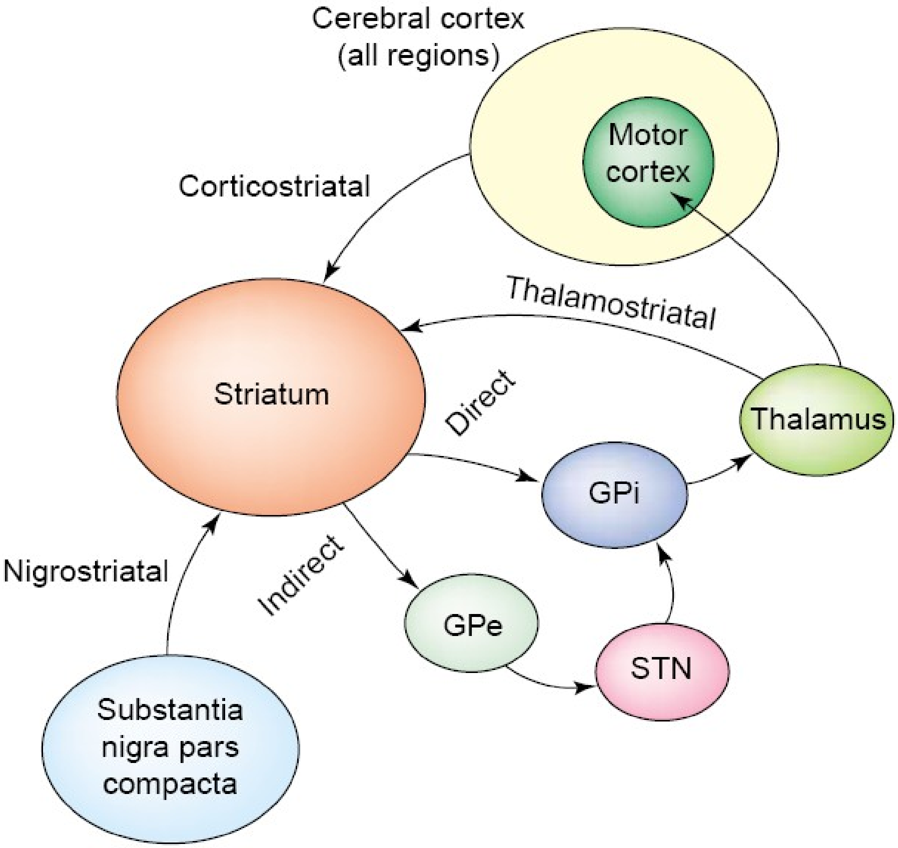
- Parameters affecting dopamine neuron phasic activation
  - Event unpredictability
  - Timing of rewards

- Initial response to CS before reward
  - initiates action to obtain reward

- Response after reward (= Reward Occurred – Reward Predicted)
  - reports an error in predicted and actual reward → learning signal to modify synaptic plasticity

Striatum and Cortex in learning

- Integration of reward information into behavior through direct and indirect pathways
- Learning related plastic changes in the corticostriatal synapses



Connections between striatum and cortex through direct and indirect pathways
(Source: Wickens et. al. 2003)

Conclusion

- DA identifies the reward present
- Basal Ganglia initiate actions to obtain it
- Cortex implements the behavior to obtain reward

- After obtaining reward DA signals error in predictions that facilitate learning by modifying plasticity at corticostriatal synapses

Computational aspects of reinforcement learning

General Approaches

- Modeled as a Markov process
  - $P(X_{n+1} = x / X_n=x_n, X_{n-1}=x_{n-1}, X_{n-1} = x_{n-2},\dots) = P(X_{n+1} = x \mid X_n=x_n)$
- Different approaches
  - Monte Carlo Method
  - Dynamic programming
  - Temporal difference

Model definition

- Environment
- Agent
- Value function (Policy)
- Reward
- Feedback

Target is to learn the policy

Temporal difference learning (TD)

- Markov process combined with the notion of reward
- Chosen state s_1 from s_0

=> current policy favors s_1 (path towards more reward)

=> next time s_0 should be chosen with ease

=> Update the value function for s_0

TD (λ) Algorithms

$V(t)_{new} = V(t)_{old} + \alpha (d_0 + \gamma \lambda d_1 + \gamma^2 \lambda^2 d_2 + \ldots)$

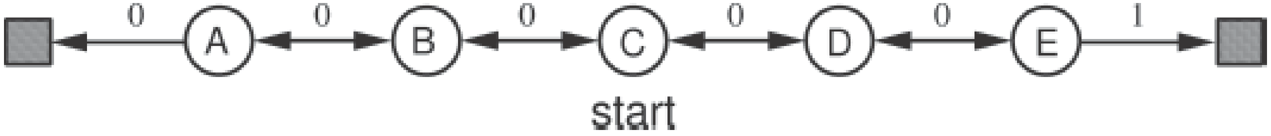
- $d_k = R(t+k+1) + \gamma V(t+k+1) - V(t + k)$

- α - Learning rate
- γ - Discount on the future reward
- λ - Discount on the future temporal differences

TD(0) Algorithm

$$V(t)_{\text{new}} = V(t)_{\text{old}} + \alpha\,(R(t+1) + \gamma\, V(t+1) - V(t))$$

- Next state only is considered.
- Example of random walk



Reinforcement learning and Neural Networks

- Problems with very large state spaces.
  - ➢ memory requirement
  - ➢ Time requirement
  - ➢ Necessary to visit all state spaces to learn how to play game
- Uses approximation function
- Using neural nets as an approximation function in reinforcement learning

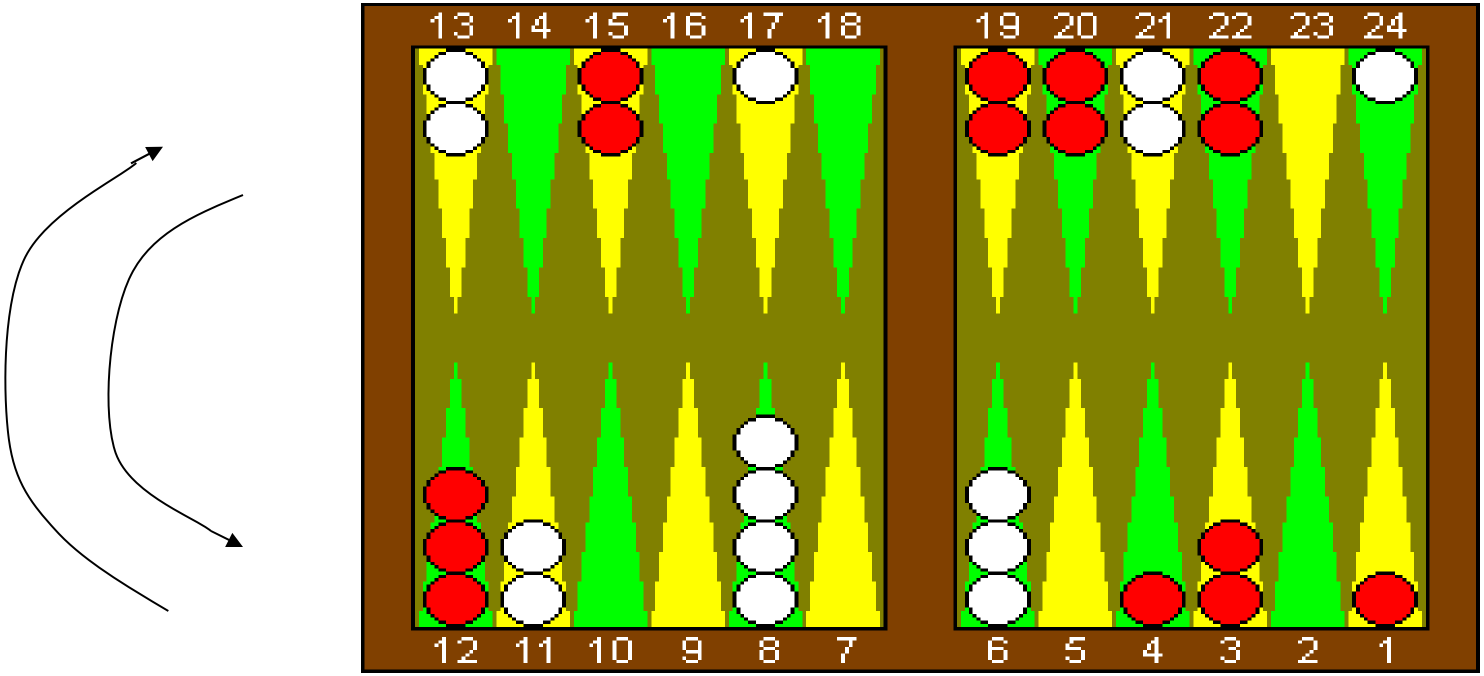
Neural nets as approximation function

- Learning
  - supervised learning
  - unsupervised, using reinforcement learning paradigm
- RL in Neural networks
  - instantaneous reward
  - delayed reward
    - combines temporal difference with BP

$$\mathbf{w} \leftarrow \mathbf{w} + \alpha[r + U_w(j) - U_w(i)]\, \Delta\mathbf{w}\, U_w(i)$$

An application using Reinforcement learning and Neural networks : TD-Gammon

- TD-Gammon : a backgammon game developed using neural net and reinforcement learning
- Backgammon

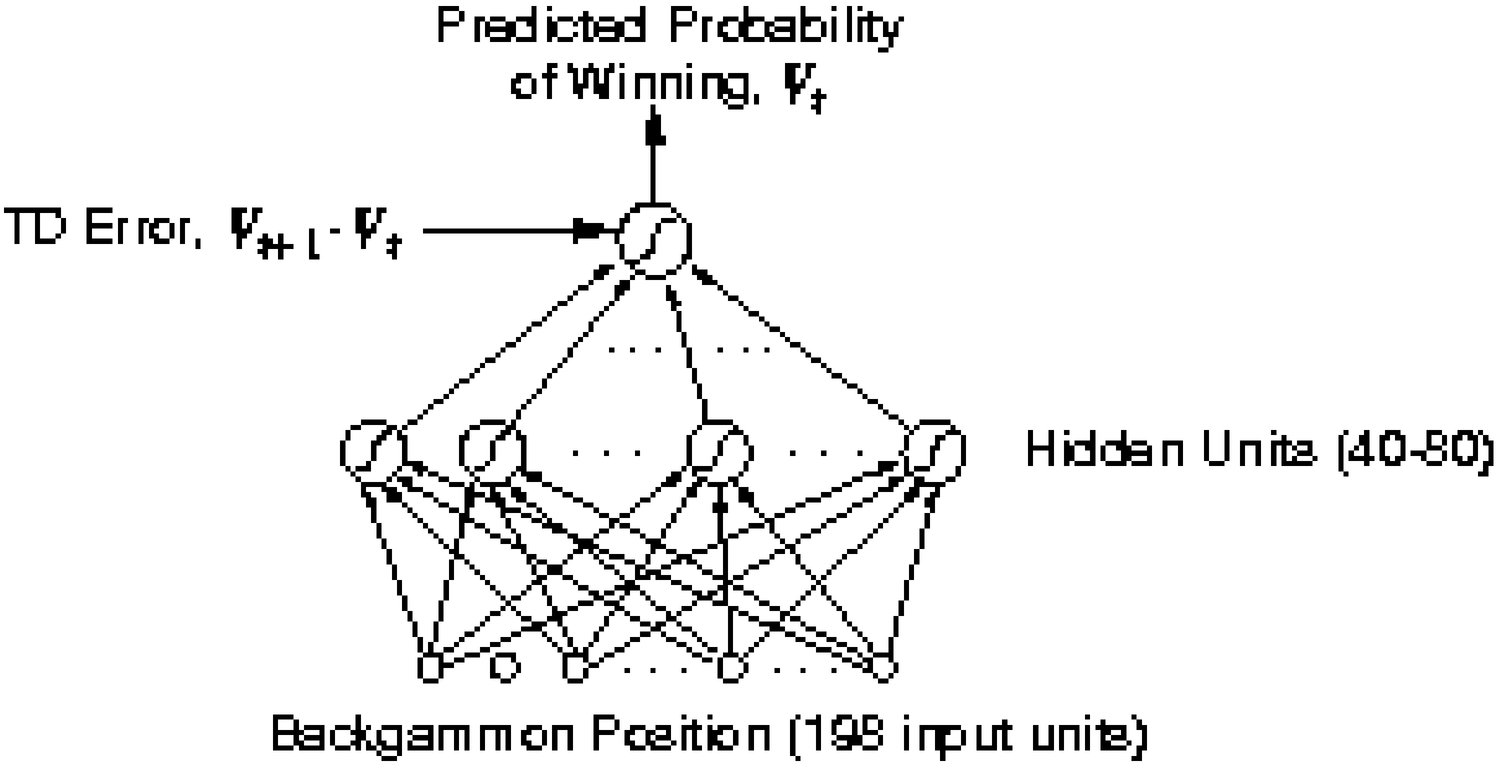


TD-Gammon

- Number of possible backgammon positions is very large.
- Learning algorithm
  - ➢ combination of TD(λ) algorithm and non linear function approximation using neural networks.
  - ➢ Output of neural net as value of each backgammon position
  - ➢ Networks weight change is governed by TD(λ) algorithm

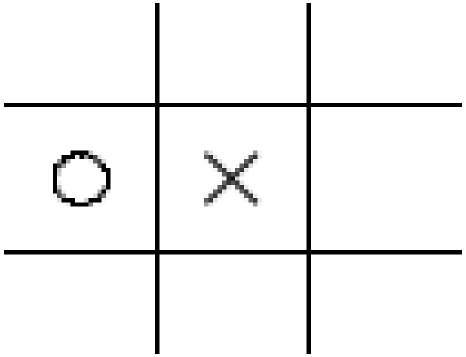
$$w_{t+1} - w_t = \alpha(Y_{t+1} - Y_t) \sum_{k=1}^{t} \lambda^{t-k} \nabla_w Y_k$$

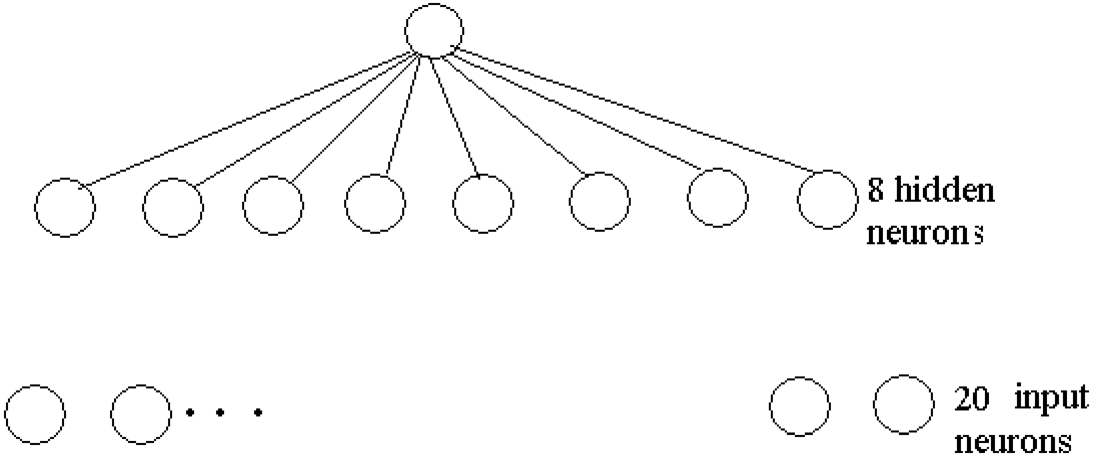
NN used in TD-Gammon


Predicted Probability
of Winning, V_t
TD Error, $V_{t+1} - V_t$
Hidden Units (40-80)
Backgammon Position (198 input units)

Project proposal

- Implementing the game Tic-tac-toe using self learning neural networks.





THANK YOU

References

- Reinforcement Learning: An Introduction (1998)

  -- Richard S. Sutton, Andrew G. Barto
- A Tutorial on Reinforcement LearningTechniques

  -- Carlos Henrique, Costa Ribeiro
- Learning to Predict by the Methods of Temporal Differences (1988) -- Richard S. Sutton
- Complementary roles of basal ganglia and cerebellum in learning and motor control (2000)

  -- Kenji Doya
- The computational neurobiology of learning and reward (2006) -- Nathaniel D Daw1 and Kenji Doya
- What are the computations of the cerebellum, the basal ganglia and the cerebral cortex? (1999)

  -- K. Doya

References contd….

- Predictive Reward Signal of Dopamine Neurons (1998)

  -- Wolfram Schultz

- Neural mechanisms of reward-related motor learning (2003)

  -- J.R. Wickens, J.N.J. Reynolds and B.I. Hyland

- Temporal difference learning and TD gammon (1995)

  -- Gerald Tesaro

- Temporal difference learning in game playing (2001)

  -- Tom Hauk