San Jose State University
Fall 2019
CMPE 255 Sec 02
Group 6 - Significant Paper Report

Significant paper: Simple random search provides a competitive approach to reinforcement learning

Key Learning

Augmented Random Search : ARS is a shallow reinforcement learning algorithm. It is shallow because unlike many other reinforcement learning algorithms, it only has a single perceptron instead of an entire neural network.

Key Points:

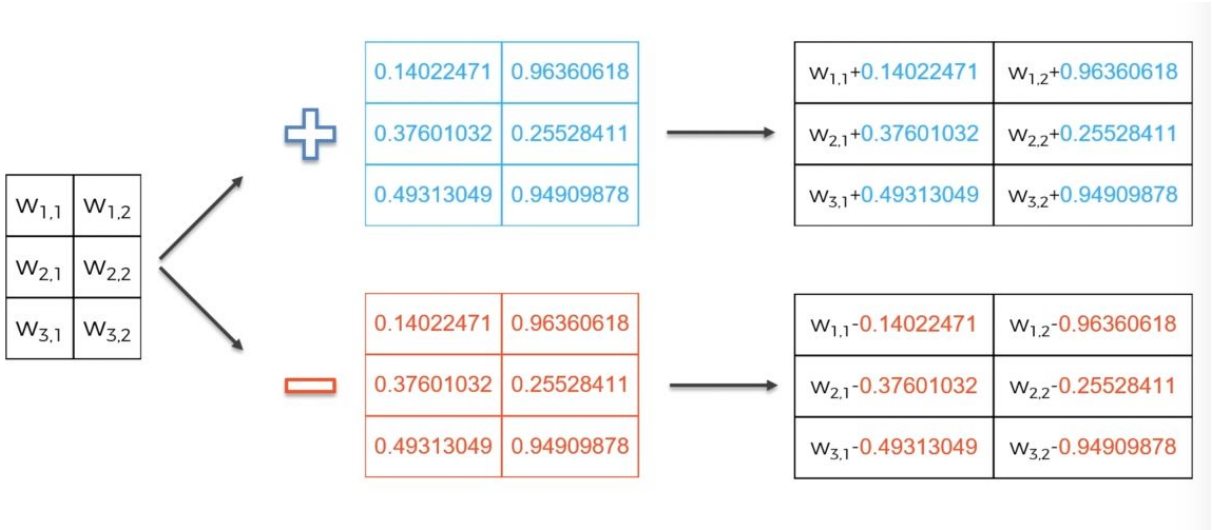
1. Perceptron : The perceptron is a mathematical computing model which consists of 4 parts
    1. Input values
    2. Weights
    3. Sum
    4. Activation function

All the inputs are multiplied with given weights and added. This gives us a weighted sum. This sum is then used to correct the activation function.

In case of ARS, there might be multiple weights and hence we represent it in the form of a matrix. So the output sum is the result of scalar matrix multiplication.

2. Maximizing rewards : The environment provides a reward for each episode. An episode starts when the humanoid in our environment starts to walk and ends when it falls or it might be according to certain constraints like timeout or reaching a certain goal. In order for our AI to learn how to walk, certain feedback is provided to the model which is called reward. This reward is used by the AI to adjust the weight.

3. The algorithm starts by calculating perturbations from the original weight matrix.



Two sets of 16 matrices are calculated by adding and subtracting very small random numbers from the original matrix. Every matrix gets its own cycle of episodes.

$w_{1,1}+d_{1,1}$	$w_{1,2}+d_{1,2}$
$w_{2,1}+d_{2,1}$	$w_{2,2}+d_{2,2}$
$w_{3,1}+d_{3,1}$	$w_{3,2}+d_{3,2}$

$w_{1,1}+e_{1,1}$	$w_{1,2}+e_{1,2}$
$w_{2,1}+e_{2,1}$	$w_{2,2}+e_{2,2}$
$w_{3,1}+e_{3,1}$	$w_{3,2}+e_{3,2}$

$w_{1,1}+f_{1,1}$	$w_{1,2}+f_{1,2}$
$w_{2,1}+f_{2,1}$	$w_{2,2}+f_{2,2}$
$w_{3,1}+f_{3,1}$	$w_{3,2}+f_{3,2}$

$w_{1,1}+g_{1,1}$	$w_{1,2}+g_{1,2}$
$w_{2,1}+g_{2,1}$	$w_{2,2}+g_{2,2}$
$w_{3,1}+g_{3,1}$	$w_{3,2}+g_{3,2}$

$w_{1,1}-d_{1,1}$	$w_{1,2}-d_{1,2}$
$w_{2,1}-d_{2,1}$	$w_{2,2}-d_{2,2}$
$w_{3,1}-d_{3,1}$	$w_{3,2}-d_{3,2}$

$w_{1,1}-e_{1,1}$	$w_{1,2}-e_{1,2}$
$w_{2,1}-e_{2,1}$	$w_{2,2}-e_{2,2}$
$w_{3,1}-e_{3,1}$	$w_{3,2}-e_{3,2}$

$w_{1,1}-f_{1,1}$	$w_{1,2}-f_{1,2}$
$w_{2,1}-f_{2,1}$	$w_{2,2}-f_{2,2}$
$w_{3,1}-f_{3,1}$	$w_{3,2}-f_{3,2}$

$w_{1,1}-g_{1,1}$	$w_{1,2}-g_{1,2}$
$w_{2,1}-g_{2,1}$	$w_{2,2}-g_{2,2}$
$w_{3,1}-g_{3,1}$	$w_{3,2}-g_{3,2}$

The matrices are represented as Rd-pos, Rd-neg etc…Here R represents rewards. Some perturbations help to improve the current results which some make it worse. We try to randomly search for new weights by this methods. Now, we need to calculate the reward. This is done by calculating the finite differences as follows

$$\begin{bmatrix} w_{1,1} & w_{1,2} \\ w_{2,1} & w_{2,2} \\ w_{3,1} & w_{3,2} \end{bmatrix} = \begin{bmatrix} w_{1,1} & w_{1,2} \\ w_{2,1} & w_{2,2} \\ w_{3,1} & w_{3,2} \end{bmatrix} + \Big((R_{d\text{-}pos} - R_{d\text{-}neg}) * \begin{bmatrix} d_{1,1} & d_{1,2} \\ d_{2,1} & d_{2,2} \\ d_{3,1} & d_{3,2} \end{bmatrix} + (R_{e\text{-}pos} - R_{e\text{-}neg}) * \begin{bmatrix} e_{1,1} & e_{1,2} \\ e_{2,1} & e_{2,2} \\ e_{3,1} & e_{3,2} \end{bmatrix}$$

$$+ (R_{f\text{-}pos} - R_{f\text{-}neg}) * \begin{bmatrix} f_{1,1} & f_{1,2} \\ f_{2,1} & f_{2,2} \\ f_{3,1} & f_{3,2} \end{bmatrix} + (R_{g\text{-}pos} - R_{g\text{-}neg}) * \begin{bmatrix} g_{1,1} & g_{1,2} \\ g_{2,1} & g_{2,2} \\ g_{3,1} & g_{3,2} \end{bmatrix} \Big)$$

Mathematically, it is represented as

$$M_{j+1} = M_j + \frac{\alpha}{b\sigma_R} \sum_{k=1}^{b} \left[r(\pi_{j,(k),+}) - r(\pi_{j,(k),-}) \right] \delta_{(k)}$$

This evaluation is done on the completion of each episode.