Report For Solving MDP using Linear Programming:

For A matrix , i iterated through all the states and for each state i checked all the actions which are possible at that state. For each action from a particular state we will insert a new coloumn in the A matrix. For this coloumn we will add We will add probability of leaving current state from current state cell and then for all states we will subtract the probability of reaching it after this current action.

Now We made a x vector which will basically store the expected no of times , the ij will take a particular action from a particular state. So summation(ri.xi) is total expected reward. We have to maximize this to find best policy. So we will maximize matmul(r.x) .

Analyzing the results of policy that we got as an output from our computation :
So from starting state ij will decide to choose up action to reach north square. Probaby the reason is that he don’t want to get hit by the mm’s attack and therefore moved upwards.
Now it is possible that mm attacks in the starting state only or if it doesn’t then ij will reach north and have all his arrows with him.
If mm attacked in starting state then ij will be in (C,2,0,0,100) state , where ij will like to choose go right , probably because in east his shooting accuracy increases and now mm is in dormant state so he might want to plan a counter attack on mm using this opportunity.
When mm does not attack then ij will reach north safely and will then decide to take action “Stay ” as he don’t wish to get attacked by mm in future. Similarly we can analyze all the actions of ij in a state.

Yes , there can be multiple policies because a policy depends on many factors. If we change the initial probability distribution i.e.

alpha matrix then we will get different policy , because ij would like to choose different routes depending on from where he starts.
Also if we change rewards then policies will be changed , because now it is possible that a state where you were earning very high rewards start giving you loss instead.
Also if we change transition probabilities then we can get different policies. For example if you reduce the shooting accuracy chances to 0.1 from 0.9 in east , then ij might now wish to move to east from centre as he have more accuracy in the centre as compared to east.
[Note: changes in transition probabilities => changes in A matrix]