European Journal of Molecular & Clinical Medicine
ISSN 2515-8260 Volume 7, Issue 11, 2020

A Novel Approach To Doctor's Decision Making System Using Q Learning

K.C.Sreedhar[a*] , M.Swathi[b],

[a]Asst Professor, Dept of CSE, Sreenidhi Institute of Science and Technology, Hyderabad, India.
simplykakarla@gmail.com
[b]Asst Professor, Dept of CSE, VMTW, Hyderabad, India

Abstract

The e-Health care system enables us to store patient's personal health record online. Now a days, doctor's decisions on health of patients is gaining importance in treating serious diseases. The overall health of human body can be subjected to many clinical parameters like random blood sugar level, white blood cell count etc. In addition to clinical parameters, the state of set of symptoms of all diseases contributes to overall well-being of a human being. Due to this the health of a human body can be decided by a set of parameters which include clinical parameters that decide the health of various organs in our body and symptoms associated with various diseases. Each of the clinical parameter can be associated with a reward based on its value being fallen in a particular bin. Also symptoms can be associated with a reward based on its intensity. The doctor will take many actions against a patient such as giving appropriate medication in course of tablets, operating surgeries, giving salination etc. So this system consists of set of clinical parameters and symptoms together as states in a model of machine learning. The set of actions taken by the doctor constitute actions of an agent where doctor is treated as an agent in this model. So a set of clinical parameters and symptoms were taken and a specified number of actions is taken to assess the performance of model in basic reinforcement learning learning and epsilon-greedy approach of machine learning. Results show that Q learning outperforms reinforcement learning and epsilon-greedy approach and these results enable the doctor for better decision making.

Keywords: *E- Health Care Data, Reinforcement learning, Q learning, Symptoms, Diseases.*

1 Introduction

Electronic Health Records (EHRs) have become natural and the data collected has been stored on a particular platform. The tracking of a patient and his history of health profile can be managed in online health care system. The data can be accumulated as patient is facing different stages in his process of treatments by various health practitioners(doctors).In India, it happens frequently that a patient who is from economically weaker sections at first consults a small hospital where fee of doctor's low and eventually he migrates to better hospitals where his expenses are more compared to earlier one to get a better treatment. In the first hospital he visited, the doctor takes some actions on his health state which is a collection of set of clinical parameters and symptoms which are diagnosed in diagnostic centers. Based on this state information the doctor takes some actions such as giving medication in view of capsules, salination etc. In the case of medication being given to the patient, there can be a number of options available to the doctor so that these parameters will get changed as the patient takes medicines. As the days passing on, he may again go to diagnostic test may be after a couple of weeks to get a new state information. Even if the patient goes to another hospital for a better treatment the new diagnostic tests may be conducted by the doctor and new state parameters are likely to be gathered. In this process, Q tables of machine learning models will get updated in the course of patients' journey through various hospitals.

The structure of the paper is: A survey on applications of Q learning in health care has been given in Section 2. Section 3 explores the concept of reinforcement learning; Section 4 explores of the methodology of Q learning approach and Epsilon-Greedy approach. Section 5 explores the proposed

approach; Section 6 shows the experimental results of the work proposed and conclusion is given in Section 7.

2 Related works

A review on application of techniques of reinforcement learning has been given by Chao Yu .,et al. [2]. Zhao., et al.[3] have applied Q learning for optimal chemotherapy drug dosage for cancer treatment to assess trade-off between toxicity and efficacy after simple reward structure. Hassani et al. [4] applied Q learning to the same problem by taking naïve discrete actions and states. Humphrey [5] applied three machine learning models in subgroup scenarios and high dimensional models. In this way, lot of work happened to propose solutions to particular problems. But there is no general framework to study various actions of doctors and various state parameters that represent clinical parameters and symptoms which convey overall well health of human body.

3 Reinforcement learning

Reinforcement learning is an area of machine learning where agents take actions in the environment which eventually lead to update of state information or state transition as explained by Barto.A.[1].There is a reward associated with for each state transition which occurs as a result of an action by the agent. Reinforcement learning differs from supervised learning in such a way that it does not require labelled output. Reinforcement learning, due to its generality, is studied in many other disciplines, such as control theory, game theory, operations research, game theory, information theory, simulation-based optimization, swarm intelligence, genetic algorithms and statistics.

The reinforcement learning model is described as follows.

1. The environment E.

2. The set of states S.

3. The set of actions of an agent,A.

4. An immediate reward $R(S_t,A_t,S_{t+1})$ associated with an action A_t,that when applied on state S_t will result in state S_{t+1}.

5. Alpha ,the parameter that tells the learning rate of the agent which is multiplied by temporal difference as given in Tesauro et.al.[6]

6. A meachanism to calculate the benefit of applying a particular action A_t in state S_t using a table called Q table.

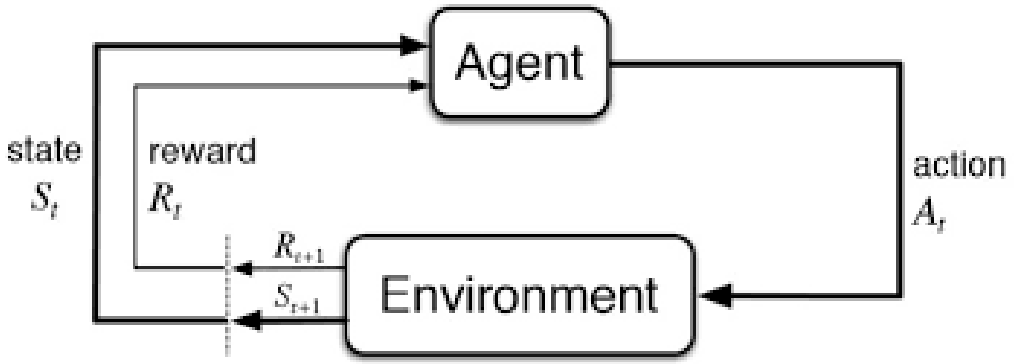


Figure 1:The workflow of Reinforcement Learning

In the basic reinforcement learning algorithm, the 5th step which is a mechanism to evaluate the ultimate learning of model based on past experience is given by following equation(1).And sample Q table is given in

$$Q^{new}(S_t,A_t)=Q^{old}(S_t,A_t)+Alpha*[R(S_t,A_t,S_{t+1})- Q^{old}(S_t,A_t)] \quad \text{------(1)}$$

Action	↑	↓	→	←
Start	0	0	1	0
Idle	0	0	0	0
Correct Path	0	0	0	0
Wrong Path	0	0	0	0
End	0	0	0	0

Figure 2:Sample Q Table

4 Q learning

Q learning is a model-free and off-policy reinforcement learning algorithm. In the process of Q learning a model is not required and so it is called as model-free as given in Strehl et.al [8].It is called off-policy because it does not require a policy which says which action to taken in which state. In other words, at the end of Q learning process we learn the policy. The letter 'Q' in Q learning stands for quality. Q learning tries to optimize the overall gain of future rewards for an action.

An agent interacts with environment in two ways. They are 1)Explore 2)Exploit. First thing is to use our Q-Table and to see all possible actions for a given state. Being present at a given state, the agent selects the action where Q-value in the table is maximum. This process is known as exploiting as we have used our past experience or the knowledge what we already have. The second way is to take a random action instead of selecting the action where Q-value is at maximum. This is known as exploring because we explore unvisited states and any further better gain from those states will lead to best long-term rewards. This exploration/exploitation trade off can be managed with the help of a parameter called epsilon. Let us randomly generate a probability value p between 0 and 1.If this probability value is less than or equal to epsilon, then we choose a random action, otherwise we choose the action with maximum Q-value. This approach is known as Epsilon-Greedy approach. Instead if we select always the action with maximum Q-value ,this approach is known as Q-learning. We use equation (2) in Q learning for updating Q-table. Gamma parameter tells the discount factor that indicates how higher we have to weight future rewards.

$$Q^{new}(S_t,A_t)= Q^{old}(S_t,A_t)+Alpha*[R(S_t,A_t,S_{t+1})+Gamma*maxQ^{old}{}_{A(t+1)}(S_{t+1},A_{t+1})- Q^{old}(S_t,A_t)] \quad \text{----------(2)}$$

5 Proposed Approach

The clinical laboratory parameters/symptoms that are being calculated for complete blood picture (CBP) are considered for example.

Table-1:Set of Parameter/Symptoms with their normal range values

Parameter/Symptom	Result	Normal Range
Haemoglobin	14	13.0-17.0 gm%
Haemotocrit(PCV)	32	40.0-50.0 Vol%
RBC Count	3.5	4.5-5.5 Millions/cumm
WBC Count	6460	4000-11000

		cells/cumm
Platelet count	2.0	1.5-4.5 Lakhs/Cumm
Parameter/Symptom	Result	Normal Range
Chest Pain	0.2	0.0 -1.0
Abdominal Pain	0.1	0.0 -1.0
Sudden Weight Loss or Gain	0.2	0.0-1.0
Fatigque	0.1	0.0 -1.0
Shortness of Breath	0.1	0.0-1.0

We have taken the value of a symptom based on its severity to represented in the scale of 0.0 to 1.0.For each clinical parameter we assess its value and we assign a reward based on its fallen range in any one of following categories. We assume following three categories.

1) Below Normal range

2) Normal Range

3) Above Normal Range

We associate a reward for each category of each clinical parameter. Similarly for symptoms we assume three levels as follows.

1) Mild Symptomatic(0.0 to 0.3)

2) Medium Symptomatic(0.4 to 0.6)

3) High Symptomatic(0.7 to 1.0)

For the experimentation, we have taken 5 parameters out of which three parameters are chosen as clinical parameters

and 2 parameters are chosen as symptoms. The rewards are as follows for each of the parameter/symptom.

Variable-1/Parameter-1 : -3 +5 +6

Variable-2/Parameter-2: -4 +6 -7

Variable-3/Parameter-3: -2 +4 -3

Variable-4/Symptom-1: +8 -4 -8

Variable-5/Symptom-2: +7 -2 -10

The values for parameter-1 indicate that if patient's clinically calculated value is below normal range then we get a reward of -3.If the value is in normal range then we get a reward of +5 and if it is above normal range we get a reward of +6.The data of all variables of a particular patient constitute as part of state information of patient. Based on above data, there are totally 3*3*3*3*3 (243) states are possible. One important point here is that we have to select divergent set of clinical parameters and symptoms related to different main organs in human body. Then our combination of different states will yield a meaningful outcome for Q learning. Out of all 243 states, the ideal or final or goal state is one which has state information as the values for each parameter/symptom lying in normal range i.e. (5,6,4,8,7).Typically there are 242 states in which atleast one parameter is not in normal range. So there we require 242 unique actions in the environment. But in real world, generally most of the actions overlap in their state transitions we took 80 actions in the experiment as an example. We have calculated the Reward matrix as state to state transition matrix as follows. The reward of action on

state S_i resulting in state S_j is calculated as distance between two state vectors is sum of differences of respective positions of state variables.

Example:S_i =(-3,-4,-2,8,7)

S_j=(-3,6,4,-4,-2)

Reward(S_i,Action,S_j)=Distance(S_j,S_i)=((-3+3)+(6+4)+(4+2)+(-4-8)+(-2-7))= -5

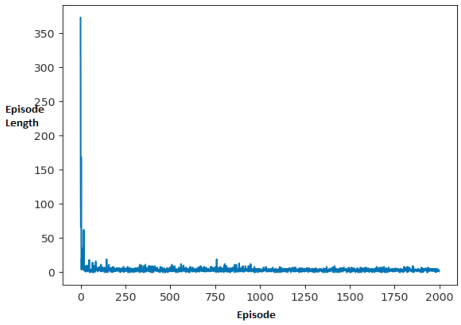
The set of actions of doctor given for example are as follows.

1. Medication
2. Surgeries
3. Salination
4. Rate of continuous Insulin supply
5. Bolus dose

Under medication the doctor may suggest many combinations of tablets/injections which form as part of action space.

6 Experimental Results and Analysis

We have implemented the reinforcement learning, Q learning and Epsilon-Greedy approach in python 3.8 on Intel Core i3-7020U CPU with 4GB RAM. The graphs are as follows. We have taken around 2000 episodes. RL stands for reinforcement learning in the graph.We have taken Alpha as 0.6, Gamma as 0.75 and epsilon as 0.1.



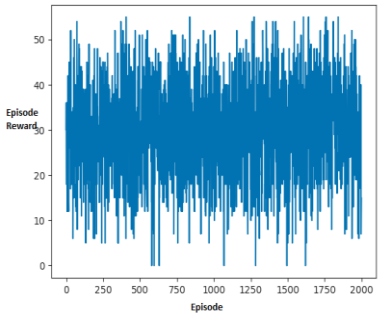
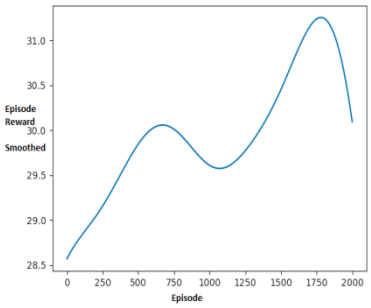


Figure 3: Episode Length in Basic RL
Figure 4:Episode Reward in basic RL



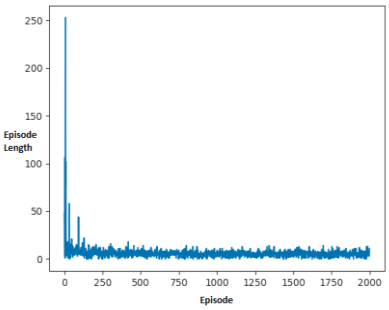


Figure 5: Episode Reward Smoothed in basic RL

Figure 6: Episode Length in Q learning

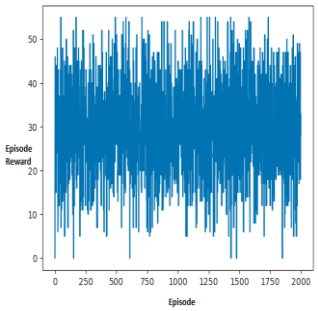


Figure 7:Episode Reward in Q Learning

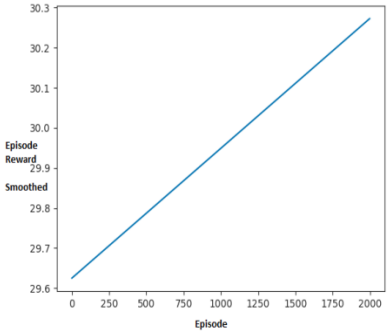


Figure 8:Episode Reward Smoothed in Q learning

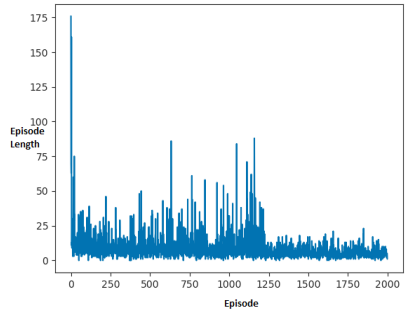


Figure 9: Episode Length in Epsilon-Greedy Approach

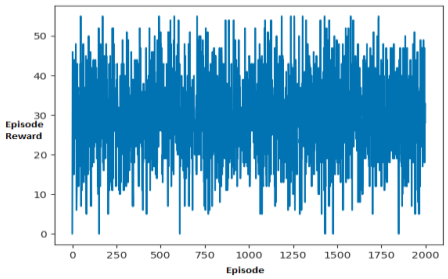


Figure 10:Episode Reward in Epsilon-Greedy Approach

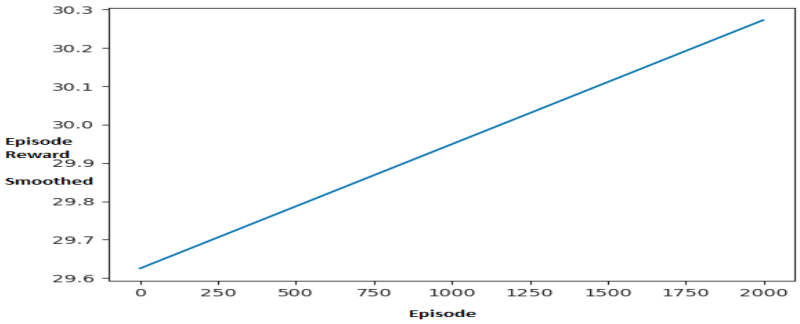


Figure 11: Episode Reward Smoothed in Epsilon-Greedy Approach

7 Conclusion

The paper focused on doctor's actions and patient's health profile as state information together to form as a model for machine learning. The model or Q-table gained as the final outcome tells the action to take at particular state. Obviously the doctor takes the action at which maximum Q-value is obtained. If we represent state information either clinical parameters or symptoms as the representative of health of various important organs in our body this model definitely a great benefit to doctors. The actions are not only limited to medication but also may be extended to physical exercises and spiritual practices which are other forms of healing for various health issues. Further,

this paper can be extended to implement deep reinforcement learning to be suitable to large number of parameters in which case there is an explosion of number of states as stated in Van Hasselt et.al.[7].

References

1. Barto, A. (24 February 1997). "Reinforcement learning". In Omidvar, Omid; Elliott, David L. (eds.). Neural Systems for Control. Elsevier. ISBN 978-0-08-053739-9.
2. C Yu, J Liu, S Nemati –(2020),Reinforcement learning in health care :A Survey, arXiv preprint arXiv:1908.08796, 2019 - arxiv.org.
3. Y. Zhao, M. R. Kosorok, and D. Zeng, "Reinforcement learning design for cancer clinical trials," Statistics in Medicine, vol. 28, no. 26, pp. 3294–3315, 2009.
4. A. Hassani et al., "Reinforcement learning based control of tumor growth with chemotherapy," in 2010,International Conference on System Science and Engineering (ICSSE). IEEE, 2010, pp. 185–189.
5. K. Humphrey, "Using reinforcement learning to personalize dosing strategies in a simulated cancer trial with high dimensional data," 2017.
6. Tesauro, Gerald (March 1995). "Temporal Difference Learning and TD-Gammon". Communications of the ACM. 38 (3): 58–68. doi:10.1145/203330.203343. Retrieved 2010-02-08.
7. Van Hasselt, Hado; Guez, Arthur; Silver, David (2015). "Deep reinforcement learning with double Q- learning" (PDF). AAAI Conference on Artificial Intelligence: 2094–2100.
8. Strehl, Alexander L.; Li, Lihong; Wiewiora, Eric; Langford, John; Littman, Michael L. (2006). "Pac model-free reinforcement learning".