Optimal Resource Allocation Policy for Breast Cancer Intervention using Reinforcement Learning

Vijeta Deshpande
CMPSCI 687: Project Report

***Abstract*—Cancer state transition models are Markov process models defined for a specific type of cancer, which provides probabilistic information on population's transition rate into pre-clinical (asymptomatic) states of cancer and clinical (diagnosed) states of cancer. Transition parameters of such models are essentially specific cancer's characteristics for population under consideration (might be a country or set of countries). Late diagnosis of cancer is being a serious issue for Low & Middle Income Countries (*LMIC*). Budget and limited infrastructure availability in LIMC puts constraints on executable intervention strategies. State transition models of cancer provide a sophisticated mathematical platform to compare different intervention strategies and select an optimal one. In this project, problem of finding specific age groups to undergo an expensive screening strategies is solved using Reinforcement Learning.**

I. Introduction

Decision of screening for a specific person largely depends on the biomedical dynamics of that person. Longitudinal data; i.e patient's medical history data; can be accessed to make a decision on screening. State transition models developed for High Income Countries (*HIC*) are based on longitudinal data and hence can be used as tool for analyzing decision of screening. Lavieri et. al. [1], investigated the problem of optimal allocation in limited screening resource availability using longitudinal data and reinforcement learning methods. But in case of LMIC, unavailability of longitudinal data poses a problem in modeling the state transition model and in formulating optimal resource allocation problem. Therefore, two step Markov Process approach is used to develop the state transition model of cancer. Section II, will briefly discuss the two step Markov process approach of developing state transition model for breast cancer. Transition parameters of the state transition model provide mean of performing sensitivity analysis, e.g. change in diagnostic rate vs. change in incidence of cancer (or simply, diagnosed cases of cancer). To tackle the late diagnosis of cancer, intervention t=strategies in LMIC are aim at increasing the total number of diagnosed cases and increasing the early diagnosis. Hence, keeping a goal of certain increment in total diagnosed cases with limited budget, age groups producing maximum result can be found out. Mammography screening of breast cancer is an expensive intervention strategy and falls into problem structure same as mentioned above. In this project, problem of finding age groups which produce the maximum positive change in diagnosed cases, under certain intervention with limited budget, is solved using Q-Learning. Section III, will present the problem formulation and problem implementation. Then in Section V results are presented followed by discussion on results and conclusion in Section VI.

II. Two-Step Markov Model

Two-Step Markov model adapts approach of representing the underlying Markov process in two different ways. Advantage of representing the same process in two different ways is that it not only enables calculation of the different transition parameters of the Markov process but also ensures a unique combination of the parameters; as there could be many possible combinations of transition parameters which can result in same cross-sectional data. State transition diagram of first step is represented in Figure 1, where Pre-Clinical state of cancer is that state where cancer is asymptomatic; i.e. not yet diagnosed; and Clinical state is that where the patient is diagnosed of cancer. O_a, I_a, m_a; called onset rate, cancer incidence and cancer mortality respectively; are the transition rates of the Markov process, for age $'a'$. Aim of this step is to find onset rate. Onset rate defines the exponential distribution of time taken to make transition from healthy state to Pre-Clinical state of cancer.

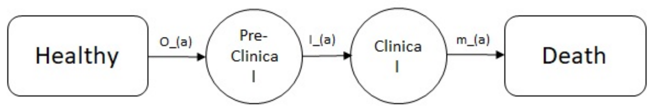

Fig. 1. State transition diagram for step-1 in Two-Step Markov Process

The second step in Two-Step Markov model uses fully expanded state space, considering the sub stages of cancer, refer Figure 2. In figure, PC-i stands for i^{th} Pre-Clinical state and C-i stands for i^{th} Clinical state of cancer. Also,

$l_{i,a}$ = Dwell time for PC-i state at age $'a'$,
$d_{i,a}$ = Diagnostic rate of cancer for PC-i state at age $'a'$,
$m_{i,a}$ = Mortality rate of cancer for C-i state at age $'a'$

Dwell time for state PC-i is time spent in the state PC-i before making transition to state PC-(i+1), for PC-4 it can be considered as life expectancy of the person in PC-4. Similarly, diagnostic rates define the rate of exponential distribution of time spent in Pre-Clinical state before getting diagnosed. And mortality rates are rate of exponentially distributed life expectancy.

Therefore, the Markov process for step 2 can expressed mathematically as follows,

$$\{X_t; t \geq 0, Z, Q, \chi\} \tag{1}$$

where, Q is rate matrix, χ is steady state distribution and Z is state space,

$$\text{Z} = \{\ H_a, (PC-i)_a, (C-i)_a\ \}$$

subscript $'a'$ stands for age and $'i'$ stands for cancer stage.

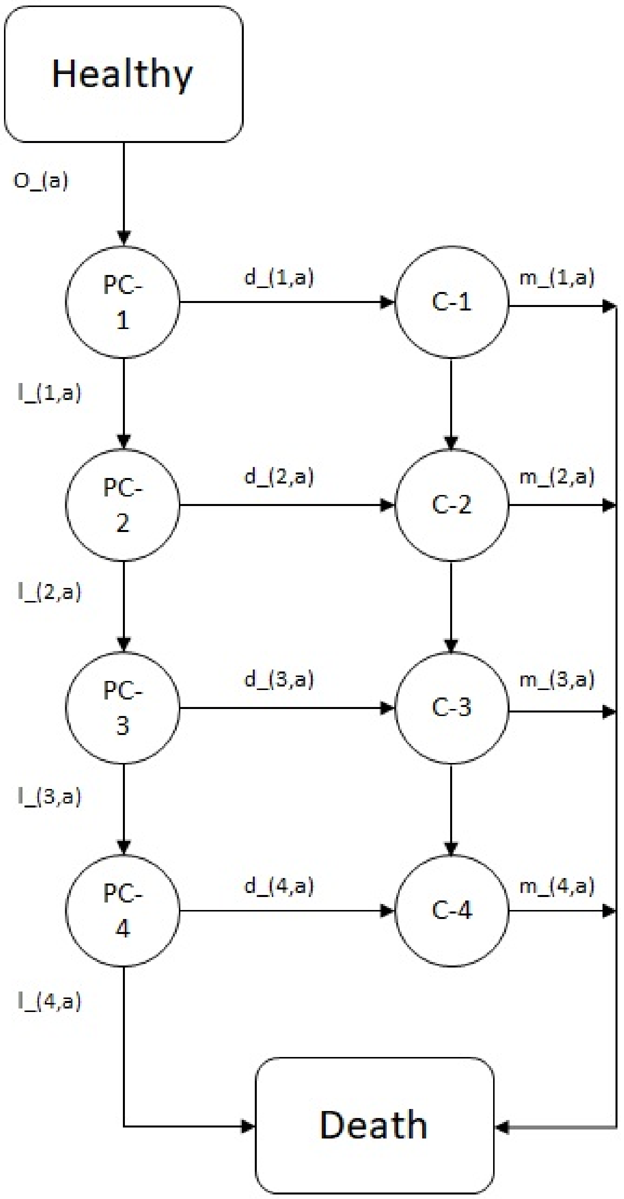

Fig. 2. State transition diagram for step-2 in Two-Step Markov Process

III. Problem Formulation

A. State set and Action set

The problem investigated in this project is finding an optimal resource allocation policy for breast cancer intervention. Resource allocation should be such that it will maximize the change in diagnosed cases of breast cancer due to an intervention. Intervention considered here is mammography for breast cancer in Peru. Due to expensive nature of this intervention strategy, not all age groups can be screened with mammography and hence the age groups which results in highest change in diagnosed cases are found out with Q-Learning.

State set of the MDP is defined as follows,

$$S = \{s_1, s_2\} \tag{2}$$

where,
s_1 = percentage of budget remaining, $\in [0, 1]$, and
s_2 = current age for which budget is being analyzed, $\in [0.01, 1]$, after each time step in an episode, s_2 will increase by 0.01

As s_1 in state set can be continuous, Fourier Basis is used for approximating the state of the agent.

Action set of the MDP is defined as follows,

$$A = \{0.10, 0.20, ..., 1.00\} \tag{3}$$

where, all the values in the action set are percentage increment in the current number of diagnosed cases in $age = s_2$.

For example, when agent is in state $[0.10, 0.60]$ and takes $action = 0.10$, it means that when agent reached 60^{th} time step in an episode he was left with 10% of starting budget and he is aiming to increase the incidence of the 60^{th} age by 10%. Transition from this state to state at next time step is entirely defined by environment and environment is defined by the Markov state transition model for breast cancer. Transition parameters of breast cancer state transition model is calculated by the two step Markov process approach for Peru's cross-sectional data (incidence, mortality and stage distribution), refer Section IV for discussion on data.

B. Environment

Optimal transition parameters were found for the Peru's data and with those parameters a steady state population was generated according to following formula,

$$\rho_{t+1} = \rho_t + Q \cdot \Delta t \cdot \rho_t \tag{4}$$

where, Q is rate matrix for the Markov process, Δt is time step and ρ matrix contains population for every state in Markov process defined by Equation 1. After steady state was reached, ρ matrix was directly taken as input data for environment calculations. In addition to steady state population, incidence of breast cancer in Peru (defined for each age) was also considered, as the *action* is basically percentage increment in incidence itself. Zelle et. al. [6], presented cost-effective analysis study for different possible intervention strategies that can be conducted in Peru. One scenario from this study was selected, where

$$\text{Cost/patient-year} = 15{,}611 \text{ USD}$$

if triennial mammography was conducted for population from 45 years of age to 69 years of age. The additional population undergoing screening as a result of an action taken by agent, gets multiplied by this cost to calculate expenditure of the action taken. Therefore,

$$s_{1,t+1} = s_{1,t} - \frac{\rho_{healthy:pre-clinical4} \cdot (15611)}{budget} \tag{5}$$

and budget will be updated for next time step,

$$budget_{t+1} = s_{1,t+1} \cdot budget_t \tag{6}$$

Starting budget was taken to be 70 million USD.

C. Reward

With the current formulation, rewards are based only on the number of cancer cases diagnosed due to the intervention,

$$reward = (10 \cdot \Delta\rho_{clinical-1:clinical-4})^{1.2} \tag{7}$$

where, $\Delta\rho_{clinical-1:clinical4}$ are the new diagnosed cases of cancer due to action taken by agent. Power was chosen to be 1.2 to have the increasing rate of reward.

IV. Data

The data input for the whole model is cross-sectional cancer data data for Peru. Cross-sectional data includes *Incidence, Mortality and Stage Distribution* of cancer cases diagnosed in Peru. Incidence of cancer is number of new cases diagnosed per 100,000 population per year, Mortality is number of deaths due to cancer per 100,000 population per year and Stage Distribution of cancer is percentage of cases diagnosed in a specific sub-cancer state out of total incidence of the population under consideration. GLOBOCAN 2012 is World health organization project aimed towards estimating incidence, mortality and prevalence from major types of cancer, at national level, for 184 countries of the world. The estimates are presented for year 2012. Incidence and Mortality data required for the model is collected from GLOBOCAN 2012's website [2]. Stage distribution data is very sensitive to the country's current cancer control strategies.

Carlos S.Vallejos et. al. [3], presented a study which took Peruvian cancer patient database and discussed different clinical and statistical results about different types of tumor and cancer stage distribution for breast cancer. Zelle S. G. et. al. [4] presented the cost-effective analysis of different interventions options that Peru could adopt for cancer control. In doing so, authors have presented a base stage distribution and calculated the stage distribution for different intervention programs, if conducted. Also, Instituto Nacional de Enfermedades Neoplsicas (INEN)'s project Plan Esperanza is aimed towards cancer prevention and control in Peru. Plan Esperanza's report [5], presents stage distribution values from public medical system and private medical system. Taking all those values into careful consideration, stage distribution data was tuned.

V. Results

A. Hyper-parameters

Following hyper-parameters were found to be the optimal, refer Figure 3,
Alfa = 0.001,
Epsilon = 0.1,
Gamma = 0.9778,
Fourier Order = 3,
Episodes = 50,000,
Trials = 20

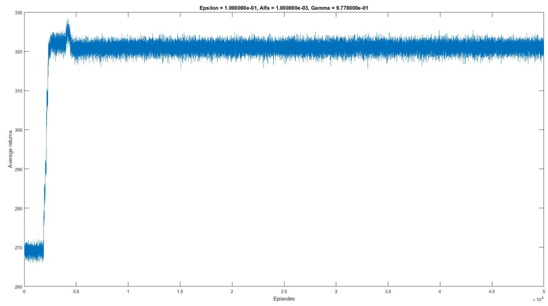

Fig. 3. Learning Curve

B. State-Action Values

With the help of trained weights and the approximated function for state space, state-action values were plotted for $Age = 45, 47, 49, 51, 53$. With the budget under consideration, Zelle et. al. [6] presented a policy to screen the population in age group from 45 to 69 triennially. Hence, to cross-verify the results from reinforcement learning algorithm with results given in [6], $Age = 45, 47, 49, 51, 53$ were selected to plot the state-action values.

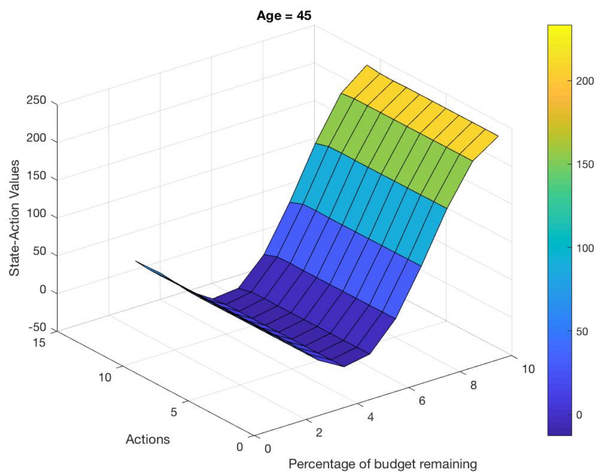

Fig. 4. State Action values for $Age = 45$

From the Figure 4, it can be seen that the state-action values for the states with high amount of percentage budget remaining are better than the other states at 45ˆth age. As

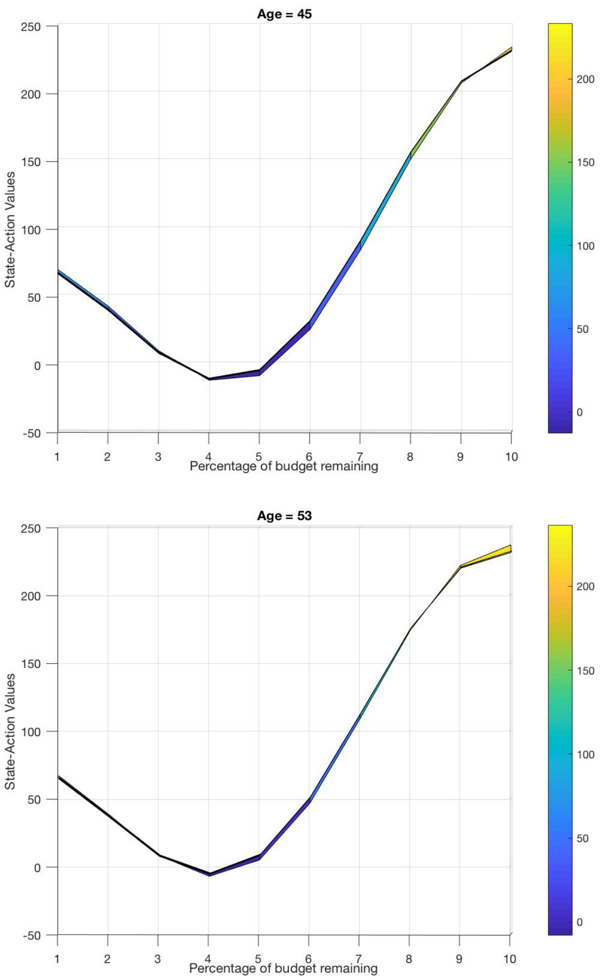

Fig. 5. State Action values for $Age = 45$

agent is maximizing the discounted reward, agent will always ensure that high amount of budget is remaining when he is in 45ˆth age/time-step of episode. Which also ensure minimal expenditure on all the ages before the 45ˆth.

C. Minimum State-Action Values

Figure 5 represents the comparison between optimal action values for age 45 and 53. One important observation from this comparison was that the minimum values of all the action values were different for both the ages. Action values for state (40%,53) are slightly better than the action values at (40%,45). This relative difference represents the agent's preference for investment in age group from 45 to 52, which is positive.

VI. Conclusion

From the results mentioned above two important observations are made. First, agent ensures minimal expenditure of budget till the age 45. Second, the difference between the minimum action values for age 45 and 53, represents a positive preference towards resource allocation in age group from 45 to 52. These two results are matching with the results presented in [6]. Therefore, reinforcement learning methods can potentially be used in designing the resource allocation policies for cancer intervention.

References

[1] Lee, E., Lavieri, M. S., Volk, M. L., & Xu, Y. (2015). Applying reinforcement learning techniques to detect hepatocellular carcinoma under limited screening capacity. Health care management science, 18(3), 363-375.

[2] http://globocan.iarc.fr/Pages/age-specific˙table˙sel.aspx

[3] Vallejos, C. S., Gmez, H. L., Cruz, W. R., Pinto, J. A., Dyer, R. R., Velarde, R., ... & Vigil, C. E. (2010). Breast cancer classification according to immunohistochemistry markers: subtypes and association with clinicopathologic variables in a peruvian hospital database. Clinical breast cancer, 10(4), 294-300.

[4] Zelle, S. G., Vidaurre, T., Abugattas, J. E., Manrique, J. E., Sarria, G., Jeronimo, J., ... & Baltussen, R. (2013). Cost-effectiveness analysis of breast cancer control interventions in Peru. PLoS One, 8(12), e82575.

[5] Plan Esperanza 2016

[6] Zelle, S. G., Vidaurre, T., Abugattas, J. E., Manrique, J. E., Sarria, G., Jeronimo, J., ... & Baltussen, R. (2013). Cost-effectiveness analysis of breast cancer control interventions in Peru. PLoS One, 8(12), e82575.