Replicating and Revising Current Literature on Reinforcement Learning For Strategic Chemotherapy Dosages

Kristina Beck, Daniel Henry, Jimmy Le

CS221 - Fall 2019

Problem Statement

- Treatment scheduling and drug dosages for cancer chemotherapy vary tremendously according to the stage of tumor, patient wellness, white blood cell levels, external illnesses, age, and more.
- Current literature builds virtual environments for Reinforcement Learning agents using ODEs to simulate the effects of chemotherapy on patients.
- Researchers train these agents in hope of a future where RL improves the decisions made during cancer treatment.

Objective

- We explored, replicated, and revised these models to generate our own environments and train our own agents to deliver optimal chemotherapy dosages.

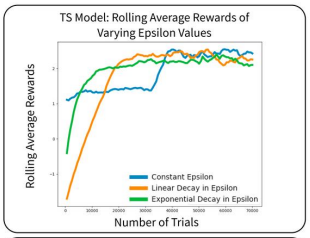
Approach

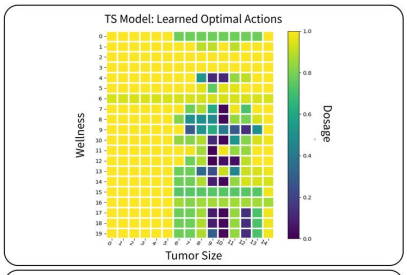
- We chose to replicate and revise two papers to create two models: the Tumor Size Model [3] (simple) and the Cells Count Model [1] (complex).
- We implement Watkins' Q-Learning algorithm, where we update feature weights with the following equation:

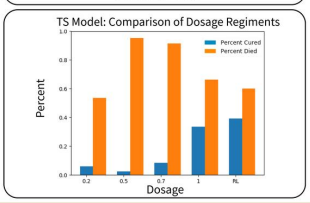
$$\mathbf{w} \leftarrow \mathbf{w} - \eta\big(\hat{Q}_{opt}(s,a;\mathbf{w}) - (r + \gamma \hat{V}_{opt}(s'))\big)\phi(s,a)$$

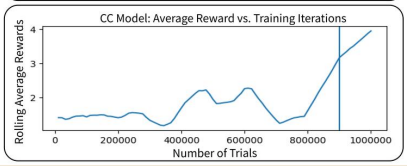
- For feature selection, we discretized our continuous state space to provide indicators for action selection.
- We custom tune hyperparameters with guidance from [1] and [3].

Results









Model 1: Tumor Size (TS)

- ***States***: $S_t = (W_t, M_t, t)$, where W is the patient's wellness, M is tumor size, and t is months
- ***Actions***: dosages *d* where $0 \leq d \leq 1$
- ***State Succession / ODEs***: Hyperparameters influence changes of wellness and tumor size with each dosags

$$\dot{W}_t = \alpha M_t + \beta(d - \lambda)$$
$$\dot{M}_t = \chi W_t - \gamma(d - \lambda)$$

$V = 1$ ends simulation (death) where $V \sim \text{Bernoulli}(e^{(-W+M)} + \psi)$

- ***Reward:***

$$R(s,a) = \begin{cases} 5 & \text{M}_t \leq 0 \\ -5 & \text{V} = 1 \\ R_W(s,a) + R_M(s,a) & t < 6 \\ 2M_0 & t = 6 \end{cases}$$

$$R_W(s,a) = \begin{cases} -\dot{W}_t & \dot{W}_t < -.5 \\ \dot{W}_t & \dot{W}_t > .5 \\ 0 & otherwise \end{cases} \qquad R_M(s,a) = \begin{cases} -\dot{M}_t & \dot{M}_t < -.5 \\ \dot{M}_t & \dot{M}_t > .5 \\ 0 & otherwise \end{cases}$$

Model 2: Cell Count (CC)

- ***States***: $S_t = (N_t, T_t, I_t, C_t)$ where N is normal cells count, T is tumor cells count, I is immune cells count, C is drug concentration, and *t* is months
- ***Actions***: dosages *d* where $0 \leq d \leq 1$
- ***State Succession / ODEs***: Hyperparameters influence cell kill rate, cell carrying capacity, intra-cell competition, and cell death/growth/influx rates

$$\dot{N}_t = r_2 N_t(1 - b_2 N_t) - c_4 N_t T_t - a_3 N_t C_t$$
$$\dot{T}_t = r_1 T_t(1 - b_1 T_t) - c_2 I_t T_t - c_3 T_t N_t - a_2 T_t C_t$$
$$\dot{I}_t = s + \frac{\rho I_t T_t}{\alpha + T_t} - c_1 I_t T_t - d_1 I_t - a_1 I_t C_t$$
$$\dot{C}_t = d - d_2 C_t$$

- ***Reward:***

$$R(s,a) = \begin{cases} 3 & \text{T}_t \leq 0 \\ -3 & \text{V} = 1 \\ \frac{T_t - \dot{T}_t}{T_t} & \dot{T}_t < T_t \\ 0 & otherwise \end{cases}$$

Results Analysis

- We implemented a number of learning rates in the TS model and found them all to converge around the same average rewards of between 2-3 at different rates.
- The TS agent gave aggressive dosages when it was possible to cure the patient instantly, but became cautious when the tumor size was large and the patient was unwell.
- The CC agent discovered a fault in the environment setup and attained high rewards without curing the patient.
- Discretization factor had negligible impact, implying a need for a better simulation.

Future Work

- Our agent should outperform the constant dosages by a greater factor than it currently does.
- Simulation realism leaves much to be desired. Further tuning and ODEs are needed.

Conclusions + Social Impact

- This project highlights the breakthroughs at the intersection of cancer/chemotherapy research and reinforcement learning.
- Developing the sophistication of the mathematical models behind chemotherapy and effects on patients is crucial. Future research with simulation modeling in mind will be required.
- Our results show that it may be possible to eventually determine treatment schedules across various forms of cancer using reinforcement learning.

References: **[1]** Padmanabhan, Regina, Nader Meskin, and Wassim M. Haddad. 2017. \Reinforcement Learning-Based Control of Drug Dosing for Cancer Chemotherapy Treatment." Mathematical Biosciences 293 (November):11{20. doi:10.1016/j.mbs.2017.08.004. **[2]** Yauney, Gregory, and Pratik Shah. "Reinforcement learning with action-derived rewards for chemotherapyand clinical trial dosing regimen selection." In Machine Learning for Healthcare Conference,pp. 161-226. 2018. **[3]** Zhao, Yufan, Michael R. Kosorok, and Donglin Zeng. "Reinforcement learning design for cancerclinical trials." Statistics in medicine 28, no. 26 (2009): 3294-3315.