Leveraging AI to help people become more productive

Problem Statement: Write down the MDP model that Lieder, Chen, Krueger, and Griffiths (2019) used to compute optimal incentives for to-do list gamification assuming that there are five to-dos that take 5,10,15,20,25 minutes to complete respectively and that the user values their time at about $8 per hour. Assuming that the discount factor is less than 1 what do you think is the optimal policy? And what do the optimal incentives look like qualitatively?

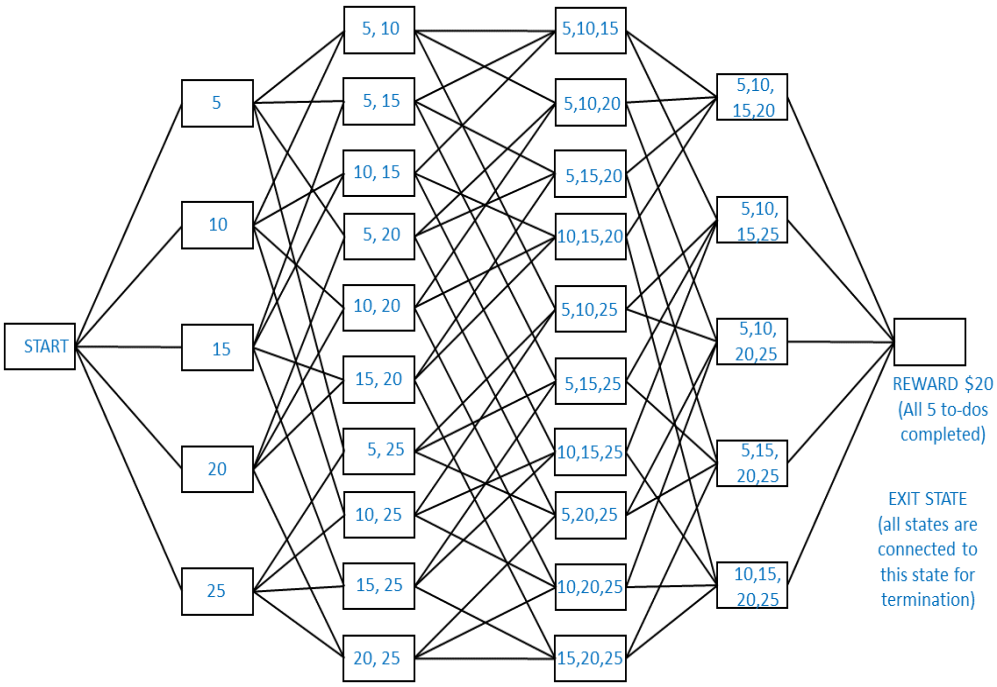
Approach:

To Do List Gamification

As specified in the paper, a sequential decision problem can be modeled as a Markov Decision Process (MDP):

$$M = (S, A, T, \gamma, r, P0)$$

Hence, the MDP model that Lieder, Chen, Krueger, and Griffiths (2019) used to compute optimal incentives for to-do list gamification assuming that there are five to-dos that take 5,10,15,20,25 minutes would consist of 32 states. The figure below is the state diagram of the to-do list gamification:



States

In any MDP, S stands for total number of states and **this model consists of 32 states** – In the figure above, the values written in the state depict that the to-dos that required that amount of value of time to be completed have been done and the to-dos requiring times apart from the values written in the state are yet to be completed.

Actions

'A' represents set of actions and in this model for to-do list gamification, **the actions at every state comprises doing one of the five to-do tasks or doing nothing i.e. if the agent is in a certain state where he has done a certain set of tasks and is trying to repeat a previously done task he will stay in the same state. Also, there is an exit state where the user can go to from any state for termination when the user loses motivation and leaves a task in between or stops doing tasks. Hence there is an option of TERMINATION at every state.**

Transition Probability between states

$T(s, a, s_0)$ is the probability that the agent will transition from state s to state s_0 if it takes action a, and **in experiment 3 in the paper, it is stated that participants were free to complete as few or as many of those assignments as they wanted, so we assume equal probability for all actions at every state**. The value of gamma is **given as less than one in the question**.

Rewards

The rewards i.e. **number of points assigned to each task was computed by optimal gamification,** which takes into account the **time required to complete the to-do** and the **value of the goals that each task contributes and how much it contributes to it.** This helps people procrastinate less and looking at the optimal reward structure, users tend to do the more difficult, tedious and long tasks first. According to the optimal reward function derived in the paper:

$$r'(s, a, s') = r(s, a, s') + \gamma \cdot V_M^\star(s') - V_M^\star(s).$$

The reward function after inclusion of pseudo rewards, has the original reward and difference in state value functions of next state and current state. Hence, even though the negative reward or cost of the 25-minute task may be the highest, the difference in state value of the next state and current state in this case is also the highest. This is because the state values depend upon the percentage of task left to achieve the end goal i.e. in this case the $20 reward. Hence, the 25-minute task completion contributes most towards reaching the final goal and hence the state value function of the state after completion of 25-minute task will be higher as compared to completion of any other task after the START state resulting in highest pseudo rewards for the 25-minute task and thus encouraging the user to complete the 25-minute task first.

General explanation of the model

So, at the START state, the agent has five actions (to-do tasks) and according to the action (to-do task) the agent performs, it reaches the next state where it's left with four actions (to-do tasks). Again, according to the action, the agent performs at this state it reaches another state where it is left with three actions and so on. Finally, after having performed all the five to-do tasks it reaches the TERMINAL state. Also, in any state, if the agent does nothing it stays in the same state or else it can take the EXIT state and terminate. The problem states that there are five to-dos that take 5,10,15,20,25 minutes to complete and that the user values their time at about $8 per hour, hence the cost for completing these to-dos is **$0.67, $1.34, $2, $2.67, $3.34 respectively**. Yet the reward for the actions in optimal gamification would be in decreasing order of times, because so that people can get motivated to do the difficult tasks and procrastinate less.

Optimal Policy and optimal incentives:

A rational decision-maker should follow the optimal policy π_M^* which maximizes the expected sum of discounted rewards, that is

$$\pi_M^\star = \arg\max_\pi \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t \cdot r\left(S_t, \pi(S_t), S_{t+1}\right)\right].$$

So, in the experimental conditions, as rewards were computed by optimal gamification, the 25 minutes to-do will have the highest optimal incentive whereas the 5-minute to-do will have the lowest optimal incentive as explained above in the rewards section. Also, as the equation above states that the optimal policy is choosing the action which has maximum discounted rewards, so the optimal policy would be completing the to-dos in descending order of times. Also, since the discount is less than one, it captures the possibility that the episode described by the MDP can end early so that future rewards might become unavailable. So, higher incentive actions should be chosen and finished first and hence the optimal policy at every step is choosing actions with highest immediate reward i.e. doing the to-dos in descending order of times (i.e. complete the to-dos in the order 25,20,15,10,5 minutes).

Also, the optimal incentives if they are represented as virtual dollars would look like that, they would be positive and high for the 25 minute to-do and the value of the rewards would be decreasing in descending order of time of the remaining to-dos. Similarly, at every step qualitatively the rewards values would be proportional to the time required to complete the to-do, hence higher reward for longer tasks than short and simple ones.