Decision Making System for Daily Activity Scheduling based on Markov Decision Process Approach

RUNZE YAN, Engineering Systems and Environment

This paper presents a modeling framework for the choice of activity type in daily activity scheduling based on human rhythms. We define rhythm factors which can affect the performance of decision makers. In this frame, we consider the interplay between rhythms and activities, and the forward-looking behavior is also included. For example, the decision maker realize the impact of the current choice on the future utility and take into account the future utility that he can obtain. The focus of this paper is thus capturing these dynamics in activity scheduling and replicating the decision maker's daily activity plans. The activity scheduling behavior is formulated as a Markov Decision Process. Solution algorithms are developed for the model and numerical examples are presented in a simulation environment.

Additional Key Words and Phrases: activity scheduling, Markov decision process, human rhythm

ACM Reference Format:

Runze Yan. 2018. Decision Making System for Daily Activity Scheduling based on Markov Decision Process Approach. 1, 1 (May 2018), 6 pages. https://doi.org/10.1145/1122445.1122456

1 INTRODUCTION

Human rhythm refers to the periodic changes in physical, psychological, or behavioral patterns. At present, research on human rhythm is relatively mature, and existing research has confirmed that human rhythm is closely related to all aspects of human body functions. For example, patients with depression usually show irregular changes in circadian rhythm. Adjusting the circadian rhythm is an effective auxiliary method for treating depression. Antidepressant drugs have a significant effect on circadian rhythm and sleep [3]. Besides, sleep and circadian rhythm disruption is also observed in people with bipolar disorder and schizophrenia [1, 6]. In addition to being used to detect emotional barriers, human rhythm is highly related to other diseases such as cancer, diabetes, and aging. Christos, et al [4] discussed the relationship between circadian rhythm between the causes of cancer and human rhythms. The clock gene that controls the day and night alternation also controls the synthesis of proteins from other genes. If the circadian genes are destroyed, the cell proliferation in the body will be disordered, which will lead to cancer in the long run. Doryab, et al [2], studied the rhythm of patients with pancreatic cancer before surgery, in hospital and after surgery and found that patients with cancer recurrence usually do not have a regular circadian rhythm. Although there is a growing interest in researching human rhythms, a limited number of researches jointly focus on circadian rhythms and human behaviors. In this paper, we build a framework to utilize human rhythms to automatically guide the daily life of people. Our framework is based on Markov Decision Process (MDP) [5], and under this framework, people could achieve reasonable allocation of time, and maximal daily benefits. Then, we demonstrate our framework in a simulation environment, and we implement three policies to solve the MDP problem.

Author's address: Runze Yan, ry4jr@virginia.edu, Engineering Systems and Environment, P.O. Box 400747, Charlottesville, VA, 22904-4259.

Manuscript submitted to ACM

2 METHOD

In this paper, we define three inner types of rhythms (cognitive, energy, and emotion) from the physical, psychological, or behavioral aspects. All these three internal rhythms can interact with each other to control the overall external performance of the human bodies. Fig 1 ignores the interaction between the rhythms, and uses the cosine curves to show how the rhythms changes with time.

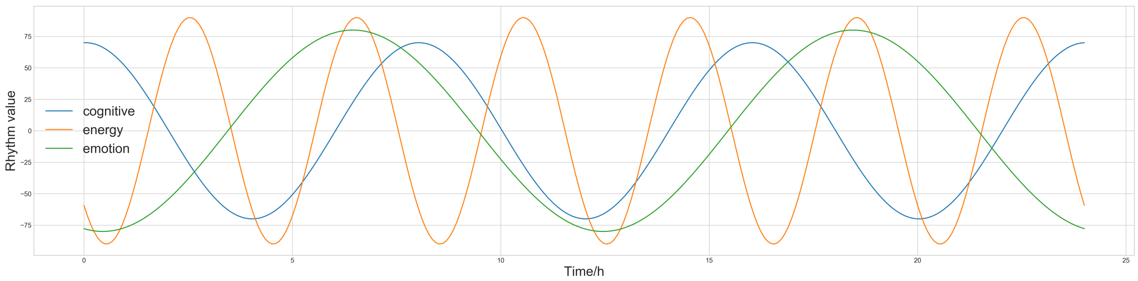

Fig. 1. Waveform of three rhythms including cognitive, energy, and emotion

If we assume that each hour forms a stage of human bodies, the levels of the cognitive, energy, and emotion hold the state of humans. And it is reasonable to implement different interventions at each state with a different combination of interval rhythms. For example, when human bodies have enough energy and are in good emotion, it's a good choice to do homework or do something that requires full attention, on the contrary, rest would be better when energy is low, or feeling sad.

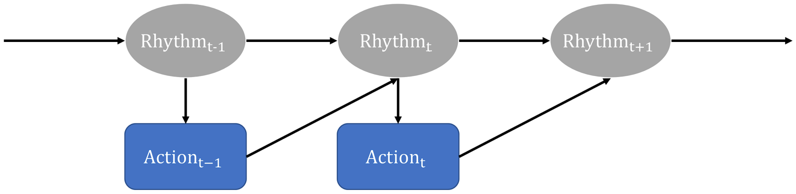

Fig. 2. The effects of action taken at each stage on rhythms

However, in fact, the rhythms would not change forever. In Fig 2, the rhythm at time $t-1$ would determine which action will be taken, and the action which is chosen at time $t-1$ would make an impact on the rhythm at time t. In the model formulation part, we will consider this scenario, and make the decision system more realistic.

3 MODEL FORMULATION

The activity-based models commonly have hierarchical structures. These complicated structures can be divided into two levels. The upper level is to assign the activity patterns onto the status of human behaviors, involving energy level, mood level, and cognitive level. The lower level is concerned with activity participation over a period of time and produces a set of activity patterns. Each activity pattern is represented by a set of activities, for example, {Sleep, Class, Exercise, Study, Rest}. The typical activity patterns can be extracted from the above action set. The choices of activities in the lower level is determined by the status of behavior in the upper level. The choices in the lower level constitute a set of activity plans for each activity pattern. The activity-scheduling problem focuses on the generation and choice of activity plans.

The decision makers gain utility by engaging in activities. The amount of utility obtained is determined by the activity duration, the attributes of the activity, and status of behaviors. The total utility obtained from an activity plan is the sum of the utility. We assume that decision makers choose the daily activity plan that provides the maximum total utility. The model introduced in this section is dynamic in the sense that the decisions made by the decision makers at a particular period would affect the states as well as the decisions of the decision makers in future periods. The reminder of this section first briefly reviews Markov Decision Process, then presents the formulation of the activity-scheduling problem.

3.1 An overview of markov decision process

Suppose there are N activity patterns, for each activity pattern $n \in \{1, \dots, N\}$, let $C_n = \{1, \dots, J_n\}$ be the set of activities included in this pattern. The superscript n is dropped for simplicity if this does not cause ambiguity.

In the activity-scheduling problem, time is discretized and the whole day (24 hours) is divided into T equal time intervals. In most MDP problems the time period is used to index the decision process, while in the activity-scheduling problem the order of the decision process is specified by a stage number k. We incorporate the time period t as part of the definition of state variable. At each decision stage, the decision maker chooses a particular type of activity. The activity sequencing is simultaneously determined by these choices. There are therefore J decision stages, each of which corresponds to an activity in $\{1, \dots, J\}$.

At each decision stage k, a decision d_k is selected from a choice set D. The set D initially concludes five types of activities: sleep, class, exercise, study, rest. But the available activities depend on the state of the decision makers and varies over decision stages. Once the decision is made, the decision maker receives an immediate utility, and the decision maker's participation of activity in subsequent state is determined by the decision. The sequence of decisions constitutes the daily activity plan of the decision maker.

In summary, a finite-horizon Markov Decision Process consists of the following objects:

- A decision stage index $k \in \{1, ..., J\}$;
- A set of state variable $s_k \in S$;
- A set of feasible decisions $d_k \in D(s_k)$;
- A immediate utility function $R_k(s_k, d_k)$;
- A discounted ratio for future utility β;

The decision maker's objective is to find a series of decisions $d = (d_1, ..., d_J$ to maximize the expected overall utility:

$$E[\sum_{k=1}^{J} \beta^k \cdot R(s_k, d) | s_0 = s] \tag{1}$$

where s is the initial state which is fixed and known.

3.2 Choice set

How can the problem of daily activity planning be encoded in a way that it becomes a finite state problem? For this, we assume that the day is segmented into a number of time slices, $k = 1...J$, each time slice corresponds to a stage in MDP. This paper will investigate time slices of 20 minutes. There are a few assumptions to be made for regulating the generation of daily activity plan. Firstly, we assume that the decision maker will engage in classed at known and fixed time. Secondly, we assume that if the decision maker fall asleep at night, the decision maker will not wake until next morning. In other words, the activity plan of the decision maker is a cycle beginning from awaking and terminating at falling asleep. Finally, we ruled out some impossible transitions of activities. For example, it is impossible to sleep after class. Usually, people always take classes during the day and sleep at night.

3.3 State variables

The state of a decision maker, s_k, includes a set of variables that provide all the information for making decision at stage k. The state variables in a MDP problem satisfy the Markov property. That is, given current state and decision: (s_k, d_k), the subsequent state s_{k+1} is conditionally independent of all previous states and decisions. Thus, at any period, the sequences of previous decisions and states are irrelevant, and all their influences on subsequent decisions have been incorporated into the current state variables. In probability theory, Markov property is expressed as:

$$P(x_k|x_{k-1}, d_{k-1}, K, x_0, d_0) = P(x_k|x_{k-1}, d_{k-1}) \tag{2}$$

Note that satisfying Markov property does not require the current choice of activity to be irrelevant of the history of activity participation. This is because the set of completed or uncompleted activities can enter into the state variable.

3.4 Immediate reward

In stage k, once the decision maker takes an activity, he will receive an immediate reward. The reward is not fixed for each type of reward. On the contrary, the immediate reward is determined by the statues of behavior in stage k. We use energy, mood, and cognitive to represent the behavior. To simplify this problem, we level the energy, mood, and cognitive. The energy has three levels: high, low, and neural; the mood has three levels: good, bad, and neural; the cognitive has five levels: alert, not alert, focus, not focus, and neutral. For the combination of different levels of energy, mood, and cognitive, the immediate reward is also different. In this way, we design a reward table which provide the immediate reward corresponds to each combination of levels of rhythms.

3.5 Expected maximum utility

If the decision maker only care about current immediate utility or the current decision does not affect the future utility received by the decision maker, the decision maker would take the decision with the greatest immediate utility. However, these assumptions are too restricting. First, the decision maker is assumed to schedule his daily activity plan to maximum the total utility derived from all the daily activities. Second, the current decision influences the future choice set: a decision maker cannot engage in the activity that has been pursed before. Furthermore, current decision partly determines the subsequent state of the decision maker and thus, affects the utility that can be obtained in future periods.

Following this line of argument, we assume that the decision maker would take into account the future utility when he makes his decision at current decision stage. The decision maker cannot foresee the future states, and he does not know with certainty his future utility. But the decision maker knows that, as well as right now, in future decision stages he can adjust his activity plan to obtain the highest overall level of utility. Hence, although he does not know the exact value of future utility, he can make decisions based on the expectation of maximum future utility. This consideration reflects the forward-looking behavior of the decision maker.

According to the above arguments, we assume that the decision maker made a decision to maximize the weighted sum of the immediate utility and the expected future utility at any decision stage. Specifically, for given state variable s_k at decision stage k, the decision maker makes a decision d_k to maximize the overall utility:

$$E[\sum_{n=k}^{J} \beta^{n-k} \cdot (r(x_k, d) + \varepsilon_k(d))|s_k, d_k] \tag{3}$$

Based on Bellman's principle of optimality, the maximum utility can be obtained by solving the recursive equation:

$$V(s_k) = E[\max_{d \in D(x)} \{r(x_k, d) + \varepsilon_k(d) + \beta V(s_{k+1}|s_k, d)\}] \tag{4}$$

where $\beta \in (0, 1)$ is the discount factor for future utility and is constant over time in this study. Different values of β reveal a variety of behaviors. If $\beta = 0$, the decision maker is only concerned with immediate utility. If $\beta = 1$, the decision maker values the future utility the same as immediate utility. As long as $\beta > 0$, then the current decision depends on the future utility, which implies forward-looking behavior.

4 EXPERIMENT

In the experiment part, we test three types of polices which guides decision at each stage. The first type is random policy, a decision d_k is randomly selected from the decision set $D(s_k)$ at stage k within limits. The second type is greedy policy. The decision which cause the highest reward at stage k is selected. This kind of policy could only lead to the local optimum but the global optimum. In order to get the global optimal solution, the third type of policy is solved by the Monte Carlo method. In the simulation scenario, we generate a student's activity plan for one day using the three policies. The reward received at each stage is shown dynamically for each of the three polices. And the initial energy will be set to 100 at first, for different action, high, low, and neural level will be set differently. This is because we consider that the energy required to take different actions will also be different, and Table 1 show the energy levels. The cognitive is similar to energy with a 100 initial value, but we add a random factor to set the cognitive to the level of alert.

	high level	neural level	low level
sleep	>= 20	5 ~20	<5
class	>= 70	50 ~70	<50
exercise	>= 80	60 ~80	<60
study	>= 60	40 ~60	<40
rest	>= 40	10 ~40	<10

Table 1. Energy levels correspond to different activities

As for the emotion, we set them random, shown in Table 2, and they will be in a good state under a high probability. At the current stage, we do not consider the interaction between rhythms.

	good level	neural level	bad level
Probability	20%	75%	5%

Table 2. Probability corresponds to different levels

As shown in Fig 3, we can conclude that the average of reward for random policy is the lowest, and is export many negative rewards, which is what we don't want to happen. In compaision with the greedy policy, the Monte-Carlo method provides the global optimal solution, and in most stages, rewards will be positive under the Monte-Carlo policy.

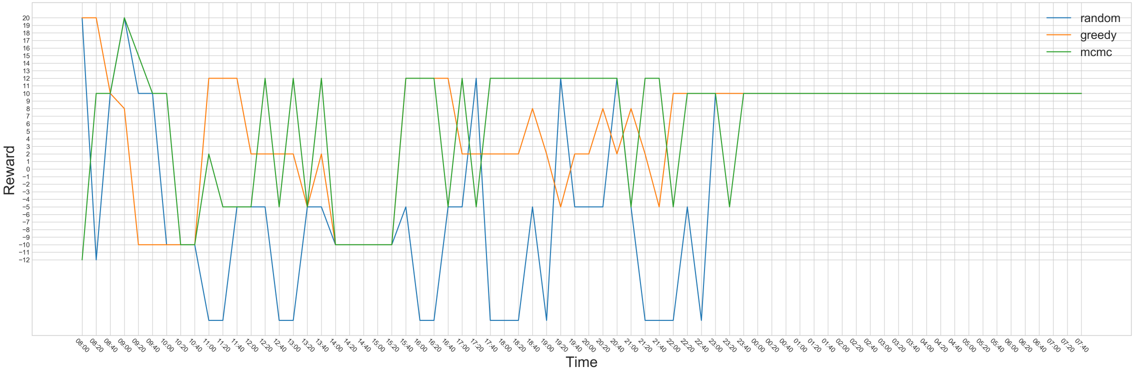

Fig. 3. Evaluation of three policies including random random policy, greedy policy, and global optimal policy exported by Monte-Carlo Method

5 CONCLUSION

In this paper, a dynamic model based on human rhythms has been proposed to plan daily activity. The decision maker allocates time to each activity to maximize the cumulative reward in the whole day. The proposed model contains three rhythm features including cognitive, energy, and emotion. The interaction between the current choice of activity and the history of activity participation is explicitly considered. The temporal utility function is employed in the models, and we add limits to the reward function. Finally, the activity-scheduling problem is formulated as a Markov Decision Process. Hence ,the results are consistent with the random utility maximization framework. At last, the optimal policy could be obtained through the Monte-Carlo method. Further study will incorporate the interaction between rhythms. The best thing to do in the future is that we can get better models of rhythms through surveys.

REFERENCES

[1] Frank Bellivier, Pierre Geoffroy, Bruno Etain, and Jan Scott. 2015. Sleep- and circadian rhythm–associated pathways as therapeutic targets in bipolar disorder. *Expert Opinion on Therapeutic Targets* 19 (03 2015), 1–17. https://doi.org/10.1517/14728222.2015.1018822

[2] Afsaneh Doryab, Anind Dey, Grace Kao, and Carissa Low. 2019. Modeling Biobehavioral Rhythms with Passive Sensing in the Wild: A Case Study to Predict Readmission Risk after Pancreatic Surgery. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3 (03 2019), 1–21. https://doi.org/10.1145/3314395

[3] Anne Germain and David Kupfer. 2008. Circadian rhythm disturbances in depression. *Human psychopharmacology* 23 (10 2008), 571–85. https://doi.org/10.1002/hup.964

[4] Christos Savvidis and Michael Koutsilieris. 2012. Circadian Rhythm Disruption in Cancer Biology. *Molecular medicine (Cambridge, Mass.)* 18 (07 2012). https://doi.org/10.2119/molmed.2012.00077

[5] Marco Wiering and Martijn Otterlo. 2012. *Reinforcement Learning: State-Of-The-Art*. Vol. 12. https://doi.org/10.1007/978-3-642-27645-3

[6] Katharina Wulff, Derk-Jan Dijk, Benita Middleton, Russell Foster, and Eileen Joyce. 2011. Sleep and circadian rhythm disruption in schizophrenia. *The British journal of psychiatry : the journal of mental science* 200 (12 2011), 308–16. https://doi.org/10.1192/bjp.bp.111.096321