Duty Cycle Optimization for Blood Pressure Sensors in Wireless Body Area Networks Based on Reinforcement Learning

First A. Author, *Fellow, IEEE*, Second B. Author, and Third C. Author, Jr., Member, IEEE

*Abstract*—Blood Pressure (BP) is a rhythmic biological parameter which reflects the health state of individual's cardiovascular system. In this paper, the issue of duty cycle optimization for BP sensor in a Wireless Body Area Network (WBAN) scenario is studied. The sticking point is how to achieve a balance between monitoring reliability and energy conservation according to the BP status. Motivated by this, we use the peak and the trend of BP values over a period of time to characterize the BP status of a human, and establishes a two-dimensional BP evolution model based on the circadian and the seasonal variation of BP. Then we formulate the duty cycle optimization problem as a Markov Decision Process (MDP) and a Q-learning based algorithm is proposed as the solution. To test the validity of the algorithm, BP values of 180 days are generated based on the basic medical statistics. Simulation results show that the proposed algorithm significantly reduces energy consumption and effectively restrains delay in sensing abnormality.

*Index Terms*—Wireless Body Area Network, blood pressure, duty cycle, reinforcement learning

# INTRODUCTION

Modern society is today threatened by current aging trends and chronic diseases. These require escalating medical management and improved infrastructure, which can also lead to a surge in health care costs. WBAN is an effective technology that utilizes wireless sensor nodes to enable real-time health monitoring of patients. These sensor nodes can be used to remotely monitor a variety of biological parameters (such as blood oxygen saturation, BP and heart activity) [1]. Thus, WBAN is commonly used to implement a home health monitoring system [2]. Taking BP monitoring as an example, Fig. 1 is a schematic diagram of the WBAN.

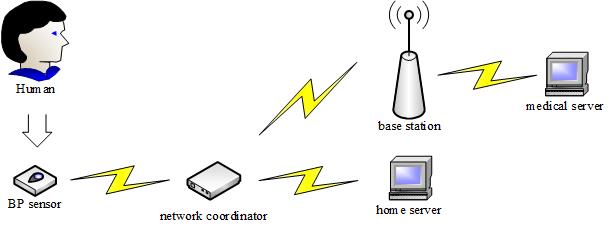
BP is an important biological parameter that affects many chronic diseases. As more people experience elevated BP, the importance of continuously monitoring BP increases. WBAN can be used to achieve continuous BP monitoring. For example, article [3] uses the changes in Pulse Transit Time (PTT) to derive changes in BP, which is a new method for detecting continuous noninvasive BP monitoring. BP values can be sent over the wireless network to the home server and medical server. Patients and medical staff can analyze and process the collected data .This network does not interfere with the patient's daily life and can achieve 24-hour continuous BP monitoring.

Fig. 1 WBAN

Like the general wireless sensor network (WSN), the battery life of the sensor nodes in the WBAN is limited, so it is also necessary to consider the service life of the sensor nodes [4]. Sensor nodes consume energy in the processes of data collection, storage, processing and transmission . Thus, the sensor nodes can reduce energy consumption by switching between active and dormant states. In article [4] the duty cycle is defined as the ratio between the time during which a sensor node is in active state and the total time of active and dormant states. It considered the upper-bound of the WSN lifetime based on random duty cycle. This paper focuses on optimizing the duty cycle of the sensor to save energy while restraining delay in sensing abnormality.

Article [5] proposes a Viterbi-based context-aware motion-sensing mechanism that adaptively finds an optimized perceptual schedule to determine when to trigger the sensor for data collection, while weighing the perceived energy and delay to detect state changes. This method effectively balances energy consumption and delay. Article [6] proposes a method for predicting the onset of acute hypotension using a hidden Markov model, which can predict future BP status based on observational data. Inspired by these two articles, we use the regularity and predictability of BP to find a method to optimize the duty cycle of the sensor to balance energy consumption and delay in sensing abnormality. Our goal is to reduce energy consumption as much as possible while ensuring the effectiveness of monitoring.

The major contributions of this paper can be summarized as follows:

* We establish a two-dimensional BP model based on circadian and seasonal variation of BP, using the peak and trend to characterize the state of BP over a period of time.
* We design an algorithm based on Q-Learning to optimize the duty cycle of the sensor for energy saving and low delay. It is verified by simulation that the algorithm is effective.

# Observation Windows Model

Biological parameters are mainly affected by nerves and body fluids (hormones), and have the characteristics of circadian and seasonal variation. We divide a cycle into multiple observation windows, and calculate several statistics of BP values in the windows. Due to the rhythm of BP, the corresponding statistics also have regularity.

## Circadian and Seasonal Variation

continuous BP measurement plays an important role in monitoring certain diseases. Article [7] studies the rhythm in BP of patients and gives some images of circadian pattern of systolic BP of normotensive and hypertensive patients. For most people, BP rises significantly after getting up in the morning and falls at night.

Article [8] studied the seasonal variation in BP over a longer span of time and concluded that systolic BP and diastolic BP are higher in winter than in summer. This can be seen as a process of gradual change between day and day, essentially related to climatic factors such as outdoor temperature.

## Observation Windows

As mentioned above, changes in biological parameters are periodic and have certain regularity in each cycle. For example, the cycle of BP values is 24 hours, higher in the morning and lower in the evening. To describe the state of BP more accurately, it is necessary to divide a cycle into smaller intervals. As a result the magnitude and complexity of the numerical changes in each interval are reduced, which is beneficial for us to observe the BP condition of the patient.

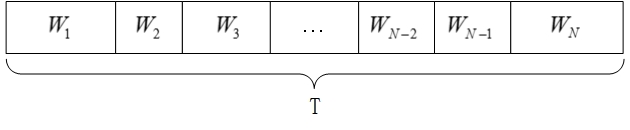
As shown in Fig. 2, we divide a cycle T into several windows (*i*=1,2, ,N). The length of each window is determined according to requirements. Lengths of different windows may be equal or unequal. 

Fig. 2 window lengths may be equal or unequal

In each window we define multiple time slots （m=1,2,3,…）of equal length and control the sensor to work or sleep in each time slot. Each window has its own unique sensing schedule to decide when to collect data. Our goal is to find a balance between energy consumption and effectiveness of monitoring. Fig. 3 is a schematic diagram of time slots.

Fig. 3 time slots

## Two-dimensional Windows Model

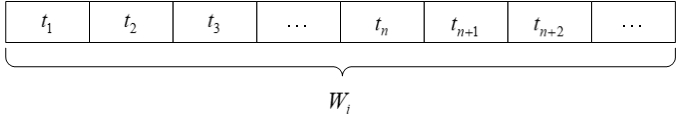
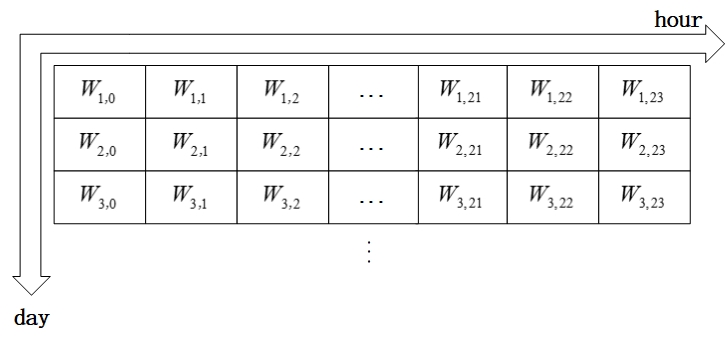
Taking into account the circadian and seasonal variation of BP, we consider the change of BP values within one day as one dimension, and the change between day and day as another dimension. The lengths of different windows are set equal, and then we can obtain the two-dimensional model of observation windows. Fig. 4 is a schematic diagram of the two-dimensional windows model. 

Fig. 4 two-dimensional windows model

There are a series of BP data in each window, and we will analyze the data to extract several features in next section.

# Window Feature Vector

The BP data in the window are usually too many and not intuitive. In order to better describe the condition of BP, we make statistical analysis of the data, which involves calculating the peak and the trend of the data. We divide these two statistics into multiple intervals, corresponding to multiple states of peak *v* and states of trend *z*, and finally get the window feature vector (*v*, *z*).

## States of Peak

The peak of BP values within each window characterizes the maximum value reached during that window. For the patient, the time when BP value reaches its peak is the most dangerous moment in the window. When the is within the interval exceeding the normal value, both the medical staff and the patient should pay sufficient attention to it. Table 1 is the latest WHO standard, which divides systolic BP into multiple intervals.

Table 1 WHO standard of systolic BP

|  |  |
| --- | --- |
| Classes | systolic BP(mmHg) |
| Ideal BP | <120 |
| Normal BP | <130 |
| Normal high BP | 130-140 |
| Hypertension | >140 |

As shown in Table 1,the patient is hypertension when the systolic BP exceeds 140 mmHg. 130 mmHg to 140mmHg is normal high BP, and 120 mmHg to 130mmHg is normal BP. Systolic BP of less than 120 mmHg is ideal. Based on Table 1, we use a similar partitioning method to extract the feature from the peak and obtain the states of peak *v*. Considering that the hypotensive standard is generally less than 90 mmHg, we divide the peak of BP value into multiple intervals, as shown in (3.1).

 (3.1)

The value of *v* can characterize the condition of peak in each window. Obviously, when *v* equals 4, it indicates that the patient's BP exceeds the normal value in current window，which needs to be vigilant. When *v* is equal to other values, it is also beneficial for us to judge the physical condition.

## Trend State

There are many statistical tests that can be used to assess the importance of time series trends. One of the commonly used nonparametric trend tests is the Mann-Kendall trend test. It is widely used to distinguish whether a natural process is in natural fluctuations or there is a definite trend. This paper uses this test to determine whether the BP values in each window shows a significant trend.

The rank correlation test for a set of observations is formulated as follows. Eq. (3.2) is the calculation method of *S*:

 (3.2)

where

 (3.3)

When n is large enough, the statistic *S* tends to be normally distributed and the variance of *S* is calculated as in Eq. (3.4):

 (3.4)

The Mann-Kendall trend test is calculated as in Eq. (3.5):

 (3.5)

satisfies the standard normal distribution. When the absolute value of is greater than or equal to 1.28, 1.64, 2.32, it means that the confidence is 90%, 95%, and 99% significance test respectively.

With a confidence of 99%, we define three intervals for as three states, as in Eq. (3.6):

 (3.6)

Obviously, the BP values decrease significantly when, the trend is flat when , and the BP values rises rapidly when . We use the value of z to judge the trend of BP values in each window. Combined with the *v* mentioned above, we obtain the window feature vector (*v*, *z*), which can effectively describe the condition of BP in each window. In this paper, we try to optimize the duty cycle in each window by controlling the sensing schedule according to the vectors.

# Coupled Markov Chain Theory

Due to the circadian and seasonal variation of BP, the sequence of vectors can be established as a Markov chain. Because the BP values change in two dimensions, the one-dimensional Markov chain cannot fully describe the process. Thus, we need to establish a two-dimensional coupled Markov chain and list the transition probability. Article [9] gives the concept of a coupled Markov chain.

The Markov chain is a probabilistic model that exhibits special dependencies. The future depends on the current and has a transition probability, but is not associated with the past. In the model of this paper, the Markov property is expressed as: the future window feature vector *W* depends on the current *W.* For a single-dimensional Markov chain, such a property can be expressed as in Eq. (4.1):



(4.1)

where is the sequence of window feature vectors and is the state space.

Obviously, represents the probability of transitioning from state to state , where n is the number of states in the model. Thus we can use to represent the one-dimensional state transition probability, and there are two constraints expressed as shown in Eq. (4.2):

 (4.2)

The coupled Markov chain describes the joint behavior of Markov chains in both dimensions. The current state depends on the previous state in both dimensions and has a joint transition probability. Similar to the one-dimensional Markov chain, we describe the Markov chain in the horizontal dimension as shown in Eq. (4.3):

 (4.3)

where is the window feature vector of the *j*th window of the *i*th day.

The Markov chain in the vertical dimension is described as shown in Eq. (4.4):

 (4.4)

Coupled Markov chains are obtained by coupling horizontal and vertical chains, but forcing the chains in both dimensions to reach an equal state. Thus we get the joint transition probability:

 (4.5)

where *C* is a normalizing constant whose value is equal to that shown in Eq. (4.6):

 (4.6)

Thus we get the joint transition probability by combining Eq. (4.5) and Eq. (4.6):

 (4.7)

The joint state transition probability indicates the effect of the previous day and the previous hour on the current state. This means that the current state can be predicted to some extent. In the following sections we will use the RL method to optimize the duty cycle of sensor based on the window feature vector and joint transition probability.

# Optimization Goal

The goal of this paper is to optimize the duty cycle of the sensor based on the states of the windows and the transition probability. In general, the duty cycle of a sensor is the ratio between the time during which a sensor node is in active state and the total time in a cycle. A high duty cycle means that the sensor sleeps less. As mentioned before, the patient's BP values are generally low at night. If the duty cycle is maintained high, the energy of the sensor is wasted actually. Thus, we use the RL method to adaptively optimize the duty cycle to reduce energy consumption while ensuring the effectiveness of monitoring.

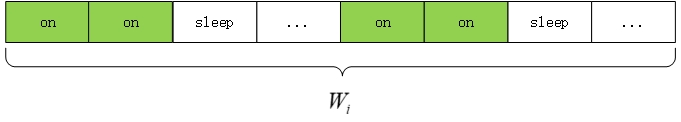
In order to judge the effect of a certain duty cycle, we assume the sensing schedule of sensor as shown in Fig. 5. In each window, it is assumed that the number of time slots in which the sensor is active in one cycle is , and the number of time slots in which the sensor sleep is . The process starts from the first time slot of the window, and after slots, the sensor turns to sleep. After slots, it starts sampling again. The process loops until the end of the window.

Fig. 5 sensing schedule (,)

Since and are small relative to the length of the window, we calculate the duty cycle of the window as shown in Eq. (5.1):

 (5.1)

In this paper, we use two calculations to judge whether the duty cycle in the window is appropriate: energy consumption and delay *.* The final goal is to find the optimal duty cycle based on these two calculations.

Article [4] introduced a formula for calculating the energy consumption (J/s) of a sensor for WSN as shown in Eq. (5.2):

 (5.2)

where is the average sensing rate, is the energy consumption to sense a bit of data, is the transceiver relay data rate, is energy consumed per bit by the transmitter electronics, is the energy consumed per bit in the transmit op-amp, is the transmission distance and is the sleep state energy consumption per second.

In general, the energy consumed by the sensor during sleep is relatively low, so the energy consumption of the sensor is positively related to the duty cycle . Higher the duty cycle is, more the energy consumption will be, and vice versa. For most sensor nodes in WSN, the energy of transmitting per bit of data is much greater than the energy consumed by sensing per bit of data. The energy consumption of a sensor node in WSN per second across a distance *d* with path loss exponent *n* is as shown in Eq. (5.3).

 (5.3)

In a specific scenario and application, the transmission power has a maximum and minimum limit as shown in Eq. (5.4).

 (5.4)

According to our specified sensing schedule, we can define the delay as the interval between the time when sensor first senses the abnormal BP and the time when the BP actually reaches abnormal in one window. Obviously, indicates the monitoring effectiveness of the current duty cycle. If is high, the current duty cycle may cause the patient's condition to be undetected and dangerous. So we want to control the within the allowed range as shown in Eq. (5.5).

 (5.5)

Therefore, our goal is to find the optimal duty cycle to achieve the best balance of energy consumption and delay. To quantify the plausibility of the duty cycle of each window, we define a reward function based on energy consumption and delay as shown in Eq. (5.6).

 (5.6)

The function  is a decreasing function with respect to and . So our goal is to maximize the sum of the reward functions of all windows by decreasing and as shown in Eq. (5.7).

 (5.7)

# Algorithm based on reinforcement learning

Reinforcement Learning is a common machine learning method whereby the agent tries to maximize the total reward when interacting with dynamic and complex environments. For specific problems, the goal is to get the optimal policy , which is the sequence of best actions, by assessing the impact of actions on the states. We design an algorithm based on Q-Learning in RL and use it to solve our optimization problem.

## Q-Learning

A Markov decision process (MDP) can be characterized by the following variables: state space (a finite set), action space (a set of finite actions), state transition probability matrix, reward function, and discount factor. State refers to the state in which the agent is currently located, usually consisting of a series of discrete values that make up its state space. Action is an action taken by agent that affects the state and is usually a finite set. The state transition probability describes the probability of a transition between states. The state transition probability is one of the components describing the Markov process. The reward function is used to evaluate current state. A positive value indicates an acceptable state and a negative value indicates a state that should be avoided.

Using these elements of MDP, we can create a Q-table. The Q-table maps the state space to the action space one by one, and the values in the table represent the maximum future reward expectations for taking the action in that state. The initial value is set to 0. Based on this table, the agent can choose the appropriate decision for each state.

In order to generate the Q-table, we use the Q-Learning method. In each episode, we explore the action by using the *ε*-greedy strategy. The *ε*-greedy strategy chooses exploitation with the probability of *ε* and chooses exploration with a probability of .If it is the exploration phase, we randomly select an action corresponding to the current state. If it is the exploitation phase, we select the action with the highest . After the action is selected, the next state and the corresponding reward function are obtained, and the of the current state in the Q-table is updated by Eq. (6.1)

 (6.1)

where is the reward for taking specific action in current state, is the expected maximum future reward given the new state and all possible actions at that new state,α is learning rate and is discount rate.

## Optimization Algorithm

Based on the Q-Learning, we design an optimization algorithm to solve our problem. We initialize the Q-table and every coefficient, and set the *ε-*greedy algorithm to be related to the number of episode. When the number of episode increases, the probability of exploitation will be higher. The reward function is calculated and the Q-table will be updated according to Eq. (6.1) until the training is completed.

In our algorithm, the state vector *s* is set as (*v*, *z*, ,). For each window feature vector (*v*, *z*), the final goal of our training is to have a optimal duty cycle expressed as a specific combination of and. Action *a* is set to multiple values, corresponding to changing and .

|  |  |
| --- | --- |
| **Algorithm** Q-Learning basedoptimization of duty cycle | |
| 1: | Initial Q-table and *s*; |
| 2: | Initial α,γ, and N; |
| 3: | Initial ; |
| 4: | repeat |
| 5: | Initial ,; |
| 6: | Initial and the sum of reward *R*=0; |
| 7: | Initial and ; |
| 8: | ; |
| 9: | repeat |
| 10: | Randomly generate ; |
| 11 | Calculate *v* and *z* of ; |
| 12: | Calculate ; |
| 13: | If : |
| 14: | Choose a randomly; |
| 15: | If: |
| 16: | ; |
| 17: | Move to next window; |
| 18: | Change and according to *a*; |
| 19: | Obtain *s*’ ,and *r*; |
| 20: | ; |
| 21: | If >: |
| 22: | Back to step 6 |
| 23: | Update Q-table by Eq. (6.1); |
| 24: | until( is the last observation window) |
| 24: | ; |
| 25: | Until() |

At the beginning of the algorithm, the parameters are initialized, and the values in the Q-table are all set to 0. The process will be started from the first observation window .In each window, we calculate *v* and *z* to obtain the state vector. The sum of energy consumption is calculated based on the duty cycle of the current window. Then we find the corresponding state vector *s* in the Q-table and choose an action to change the duty cycle of next window according to the *ε*-greedy strategy. We set *ε* to the expression associated with episode as shown in Eq. (6.2).

 (6.2)

where is a constant used to control the convergence speed of the algorithm.

In each episode, a random number is generated and compared with *ε* to determine the selection strategy of the action. According to the *ε*-greedy strategy, if in the exploration phase, one action is randomly selected, and in the exploitation phase, the action with the largest  is selected. After that, the and of the next window are changed according to the action. Then we observe the next window and calculate the reward function *r*, which can be used to update the.If the delay exceeds the maximum , we return to the first window and restart the process. This process will loops until the number of episode reaches the maximum value N we set at the beginning.

# Simulation results

## Parameter setting

To verify the feasibility of the algorithm, we tested it using the generated systolic BP data. The generated systolic BP data is derived from the regularity described in articles [5] and [6]. The systolic BP has circadian and seasonal variation. During one day, systolic BP has a peak in the morning and evening, and a trough occurs in the night. On the other hand, the systolic BP in summer is lower than that of winter. Therefore, based on the regularity, we generate 180-day data to simulate systolic BP values of hypertensive patients from summer to winter. There is one data per second.

As described in the algorithm, we need to define the action space and the reward function *r*. The action *a* is a change on and . When the action is selected, the and of the sensor will be changed, and the duty cycle of the sensor changes at the same time. Since and may be low, the change on them should also be low. Thus, the action space we designed is shown in Eq. (6.3).

 (6.3)

The reward function designed needs to conclude energy consumption and delay. Our goal is to minimize energy consumption when the delay is not greater than the maximum. According to Eq. (5.2), the difference of the energy consumption after the duty cycle of the sensor is changed can be calculated by Eq. (6.4).

 (6.4)

where and are the duty cycles of adjacent windows.

We divide the reward function into two parts. One is for controlling delay and the other is for energy consumption. The reward function of delay can be designed as shown in Eq. (6.5).

 (6.5)

Within each episode, we calculate of each window, and there is a penalty if the duty cycle leads to a delay that exceeds the limit.

The reward function of energy consumption can be designed as shown in Eq. (6.6).

 (6.6)

When the energy consumption is reduced, the reward is positive, otherwise it is negative. By this way we can control the duty cycle to change in the direction of reducing the energy consumption.

Thus, our reward function *r* can be written as shown in Eq. (6.7).

 (6.7)

Through continuous training, a duty cycle that ensures less delay and less energy consumption can be obtained.

Learning rate α is set to 0.1 and discount rate is set to 0.9.

## Results

Based on our algorithm and parameters, we use the generated 180-day BP data to test. In order to make our method more effective, we test 30 days in summer and 30 days in winter. The variation of the sum of reward , total energy consumption and maximum delay are used to analyze the effectiveness of the algorithm. We set the values of to 5, 15 and 25 to observe the convergence speed and the final convergence value of the algorithm.

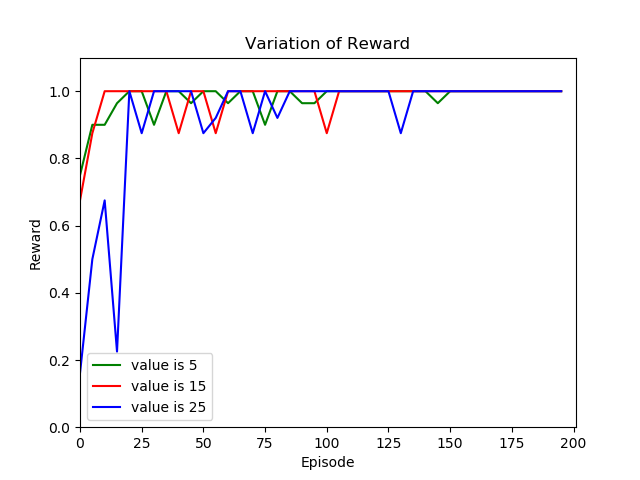
Fig. 6 shows the relationship between the sum of reward and training episodes with different constants . In order to make the image easier to interpret, we used the maximum generated during the training as the denominator and normalize all the .It can be seen from Fig. 6 that has an influence on the convergence speed of the algorithm, and the higher the value, the slower the convergence speed. On the other hand, when is higher, it is more likely to get better results.

Fig. 6 sum of reward

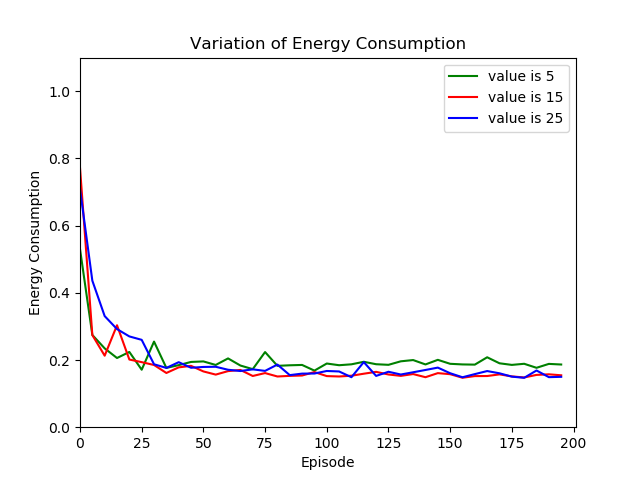
As shown in Fig. 6, starts with a lower value , gradually increases and eventually stabilizes. When the convergence speed is low, the algorithm has a greater probability of exploring to find a better solution. However, if is too large, the training time will be multiplied when the windows is too many.

Fig. 7 energy consumption

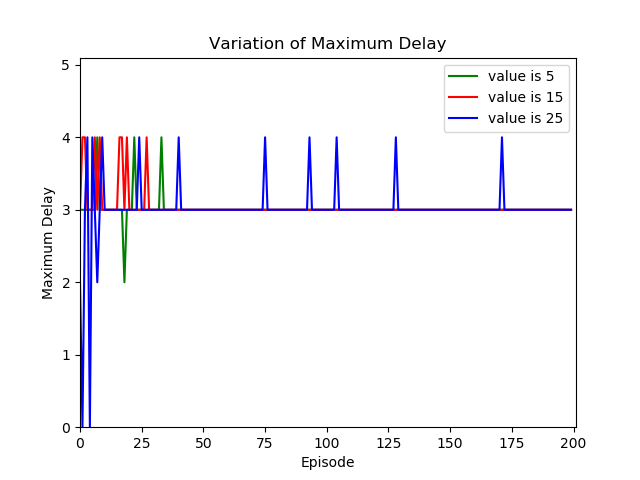
Fig. 7 shows the relationship between total energy consumption and training episodes with different constants . Obviously, gradually declines and eventually reaches a low value. This result shows that our algorithm can effectively reduce the energy consumption of the sensor. The final energy consumption can be reduced to around 20%.

Fig. 8 maximum delay

The maximum delay in each episode is as shown in Fig. 8. As mentioned before, we required that the delay not exceed the specified maximum when optimizing the duty cycle. In our tests we set a maximum of 4 s. The final result shows that the maximum delay does not exceed 4 s, which meets our requirements.

Then, we use our algorithm for 30 days in summer and 30 days in winter to compare the differences in energy consumption. Fig. 9 shows the differences in the energy consumption values. The is set to 15.In winter, because the overall value of BP is higher than summer, the energy consumption should also be higher to ensure the effectiveness of monitoring. As shown in Fig. 9, our algorithm finally achieves higher about 10% than that in summer. This result is in line with our expectations.

# Conclusion

With the application of WBAN, many biological parameters will be monitored more intelligently. In practical applications, optimizing the duty cycle of the sensor can increase the service life of the data acquisition end, which is of great significance. In this paper we build a two-dimensional Markov model to better understand the variation of BP. And an optimization algorithm based on RL is proposed. Tests have shown that our method can effectively reduce energy consumption and ensure that the delay in sensing abnormality is within a reasonable range.

References

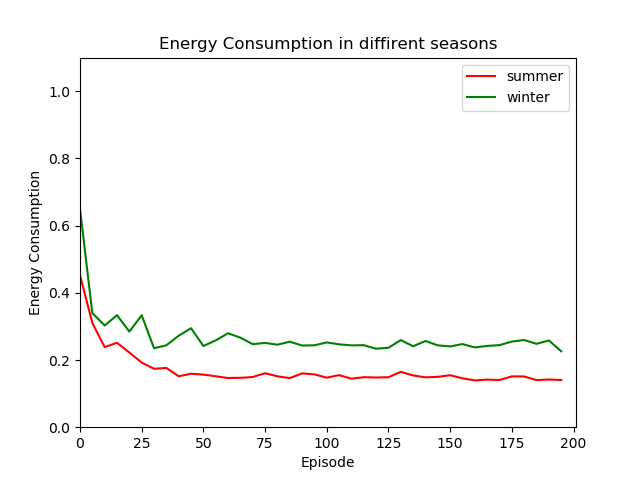
1. Agyei-Ntim, Frank , and K. Newman. “Lifetime Estimation of Wireless Body Area Sensor Network for Patient Health Monitoring,” *Conference proceedings:. Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference*, pp.1659-1662,2009.
2. Otto, C. A., E. Jovanov., and A. Milenkovic. “A WBAN-based System for Health Monitoring at Home,”  *IEEE/EMBS International Summer School on Medical Devices & Biosensors*, 2006.

Fig. 9 energy consumption in different seasons

1. Sheng, H. , M. Schwarz , and J. Boercsoek.“A wireless wearable body sensor network for continuous noninvasive blood pressure monitoring using multiple parameters,”*International Conference on Circuits*,2011.
2. Rout, Rashmi Ranjan, and S. K. Ghosh , “Enhancement of Lifetime musing Duty Cycle and Network Coding in Wireless Sensor Networks,”*IEEE Transactions on Wireless Communications*, pp. 656–667,Dec.2013.
3. Taleb S., Hajj H., Dawy Z., “VCAMS: Viterbi-based context aware mobile sensing to trade-off energy and delay,” *IEEE Transactions on Mobile Computing*, pp.99:1-1, 2017.
4. Singh, Abhishek, et al. “Hidden Markov Models for modeling blood pressure data to predict acute hypotension,” *IEEE International Conference on Acoustics IEEE*, 2010.
5. Hermida, R. C., Ayala, D. E., & Portaluppi, F.,“Circadian variation of blood pressure: the basis for the chronotherapy of hypertension,” *Adv Drug Deliv Rev*, vol. 59, no. 9, pp. 904–922, 2007.
6. Woodhouse, P. R., Khaw, K. T., & Plummer, M., “Seasonal variation of blood pressure and its relationship to ambient temperature in an elderly population,” *Journal of Hypertension*, vol. 11, no. 11, pp. 1267–1274, 1993.
7. Chessa, A. G., “A markov chain model for subsurface characterization: theory and applications,” *Mathematical Geology*, vol. 38, no. 4, pp. 503–505, 2006.