

REM Estimation Based on Combination of Multi-Timescale Estimations and Automatic Adjustment of Personal Bio-vibration Data of Mattress Sensor

Iko Nakari¹, Naoya Matsuda¹, Keiki Takadama²

The University of Electro-Communications
1-5-1 Chofugaoka, Chofu, Tokyo, Japan 182-8585
{iko0528, matsuda.naoya}@cas.lab.uec.ac.jp¹, keiki@inf.uec.ac.jp²

Abstract

This paper proposes the novel REM estimation method based on the combination of REM estimations with multi-timescale logarithmic spectrums calculated from overnight bio-vibration data acquired from mattress sensor. Concretely, this paper learns each Random Forests for multiple scale spectrums, and counts the number of REM estimation in the length of the window, and estimates REM if the counted number exceeds the threshold. The threshold is automatically determined based on the REM estimation ratio to the total sleep length for each person to consider individual differences. Through the human subject experiments, the following implications have been revealed: (1) the combination of RFs learned with each scale spectrum improves the Precision and Recall of REM estimation, and Accuracy, Precision, Recall and Specificity are 80.2%, 51.4%, 47.0% and 48.5%, respectively; and (2) the automatic adjustment of the threshold can be flexibly adapted to data with large individual differences without the need to retrain the model.

Introduction

According to the survey conducted by the Ministry of Health, Labour and Welfare, it is estimated that about one in three Japanese adults feel sleepy during the day at least three times a week. In addition to that, Japan has the shortest sleep time among the OECD member countries (Organization for Economic Cooperation and Development 2019), which suggests that many people in Japan are sleep-deprived. The accumulation of sleep deprivation (especially 4-6 hours of sleep) leads to a state of sleep debt. In the state of sleep debt, the ability to think and make decisions is equivalent to staying up all night (Van Dongen et al. 2003), and it is a factor in the increased risk of industrial and traffic accidents. It also decreases immune function and increases the risk of developing lifestyle-related diseases such as depression and dementia (Mullington et al. 2009; Holingue et al. 2018). For individuals to stay healthy and for the government to reduce health care costs, these sleep problems should be solved as soon as possible.

To solve these sleep problems, it is important to increase the amount of sleep time, but many people suffer from the problem of poor sleep quality even if they sleep for a long time. For the facts, it is necessary to understand sleep quality. The standard method for measuring sleep quality (sleep stage) is to evaluate biological data acquired by

Polysomnography (PSG) test based on the Rechtschaffen & Kales (R&K) method (Rechtschaffen and Kales 1968). However, the PSG test is a highly restrictive method and requires a person to attach multiple electrodes to his/her head and body, which burdens physical and mental on his/her and prevents obtaining data of sleep as usual. To address the problems, the demand for sleep stage estimation methods by simple sensors (such as mattress sensors) has increased as an alternative to the PSG test. For example, Watanabe developed a mattress sensor and focused on the relation between heart rate variability and sleep stage (Watanabe and Watanabe 2004). The accuracies of the method are reported as follows: 42.8% in three stages (NREM/REM/WAKE) estimation; 82.6% in NREM estimation; 70.5% in WAKE estimation; and 38.3% in REM estimation. As the results show, the accuracy of the method is not high, especially the accuracy of REM sleep estimation. This is because that REM sleep estimation is mainly based on rapid eye movements in the R&K method, and mattress sensors cannot measure eye movements. Even though REM sleep has other characteristics (i.e., unstable heart and respiration rate) acquired from the mattress sensor, it is hard to estimate REM sleep because of the following points. (1) The characteristics that appear in REM sleep appear intermittently rather than all the time during REM sleep. (2) The heart rate gets unstable by body movements. (3) The heart rate is easily affected by individual differences and daily physical condition.

To tackle the problems, it is necessary to estimate REM sleep from a new perspective, including physiological characteristics. However, since we do not know what to focus on, machine learning (ML) is a good way to estimate REM sleep from a new perspective. In this study, Random Forests (Breiman 2001) is employed for the ML model because it is easier to analyze what the model learned from the data than deep learning (Goodfellow et al. 2016), which is widely employed because of its high prediction accuracy. It is essential for analyzing models easier because it leads to the interpretability of the model in the future. However, it cannot deal with the problem (1) mentioned before by applying the ML to sleep stage estimation because each epoch (30 seconds) is estimated without considering before/after the corresponding epoch. Due to this, it is difficult to estimate REM sleep when the characteristics do not appear at the epoch. In addition to that, the ML is not good at learning

In T. Kido, K. Takadama (Eds.), Proceedings of the AAAI 2022 Spring Symposium "How Fair is Fair? Achieving Wellbeing AI", Stanford University, Palo Alto, California, USA, March 21-23, 2022. Copyright © 2022 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

data with individual differences.

To overcome these problems, this paper aims to improve the accuracy of REM sleep estimation and proposes the novel REM sleep estimation method with a mattress sensor that can consider before/after the corresponding epoch and be automatically adjusted the REM sleep estimation threshold for each person. Concretely, this paper employs TANITA sleep scan SL511 (Japan) as the mattress sensor for acquiring bio-vibration data, and prepares several RFs for learning multi-timescale logarithmic spectrums. It is combined that the REM sleep estimations by each RF for the final output of estimation, and the estimation sensitivity is automatically adjusted based on the REM sleep estimation ratio out of all epochs in an overnight sleep.

This paper is organized as follows. The next section describes the sleep mechanism especially REM sleep. Section 3 describes the related works of non-contact sleep stage estimation and RF which is the main ML method in our proposed method and Section 4 proposes our multiple scales REM estimation method. The experiment is conducted in Section 5 and the result are analyzed in Sections 6. Finally, our conclusion is given in Section 7.

Sleep Mechanism

Sleep Stage

The sleep stage is an indicator of the depth of sleep defined by the R&K method. The depth of sleep is classified into six stages in each epoch (30 seconds), i.e., WAKE, REM, Non-REM1 (N1), N2, N3, and N4 (N4 is often included in N3). The proportion of each sleep stage in healthy young adults per night is as follows: WAKE is 1-5%; REM is 15-25%; N1 is 5-20%; N2 is 45-75%; and N3 is 10-22% (depending on age and physical condition on that day). In order to determine the sleep stage, the R&K method needs biological data such as electroencephalography (EEG), electrooculogram (EOG), and electromyogram (EMG) acquired by the PSG test. Figure 1 shows the example of the overnight sleep stages, where the vertical axis indicates the sleep stage and the horizontal axis indicates the time. As shown in Figure 1, the structure of the sleep stage in a healthy person repeats deep sleep (N3 sleep) and shallow sleep (above N3 sleep) alternately, and the regular sleep repeats this cycle (about 90 to 120 minutes) three to five times a night. Each cycle is connected by about 20 to 30 minutes of REM sleep, and this cycle is called the ultradian rhythm.

Characteristics of REM Sleep

The physiological characteristics of REM sleep are as follows:

- rapid eye movement;
- decreased skeletal muscle activity;
- increased or unstable heart rate and respiratory rate;
- changes in autonomic function.

In particular, REM sleep is determined by focusing on “rapid eye movement” and “decreased skeletal muscle activity” in the R&K method. This is because that the two characteristics are clearly expressed in the biological data acquired by

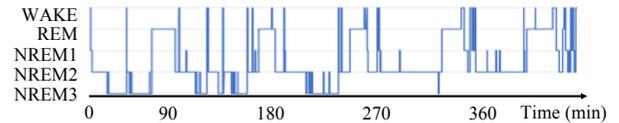


Figure 1: Example of the overnight sleep stage.

the PSG test and are easy to evaluate. Note that these characteristics occur intermittently, rather than continuously.

Related Works

Sleep Stage Estimation by mattress sensor

Watanabe et al. tried to extract the relation between the change in the heart rate and sleep stages through the frequency band containing the multiple biological rhythms of a human to build a foundation of sleep stage estimation from heart rate variability (HRV) (Watanabe and Watanabe 2001). They focused on two biological rhythms that the ultradian rhythm and the circadian rhythm, which is an approximate 25 hours cycle. From their study, the relations between the frequency of HRV and sleep stage have been revealed, and they built a sleep stage estimation method based on the heart rate data acquired from the air mattress sensor (Watanabe and Watanabe 2004).

Random Forests

This study employs Random Forests (RF) (Breiman 2001). The RF model repeats random sampling from training data, randomly construct decision trees with different conditional branches, and classify them by majority rule of those results. In this research, Gini impurity is the splitting condition, it becomes low when all the samples contained in each node of the decision tree are the same. RF processing is as follows:

1. Generate bootstrapped sample (S_j) from training data set (S).
2. One-third of the original data is called Out-Of-Bug (OOB), and it is used for constructing decision tree. Each node processing is as follows:
 - (a) Extract m_{try} features randomly with not allowing duplicate value.
 - (b) Choose the feature that minimizes Gini impurity, and divide nodes.
3. Repeat 1. to 2. N_{tree} times.

Where N_{tree} is the number of decision trees to be constructed. In the classification problem, it is recommended to use the square root of the total number of features for the variable m_{try} , which used to divide the nodes of decision trees.

Proposed Method: Multi-Timescale REM Estimation

The proposed method, Multi-Timescale REM estimation starts from learning several RFs with each scale bio-vibration data, then combines REM predictions by each RF,

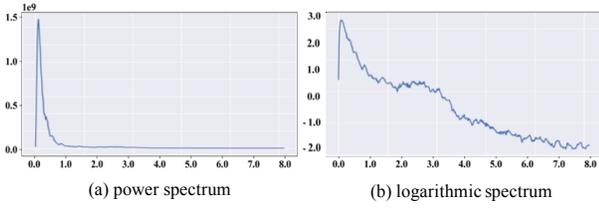


Figure 2: Examples of (a) power spectrum and (b) logarithmic spectrum ($L = 64$).

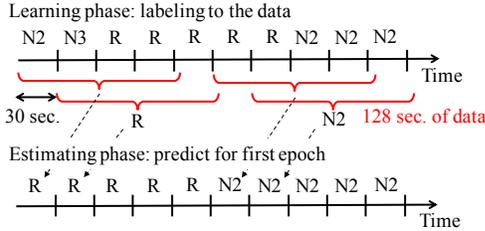


Figure 3: Example of strides and how to label sleep stage to the spectrum.

and estimates REM sleep based on the REM estimation threshold adjusted for each person.

Input data

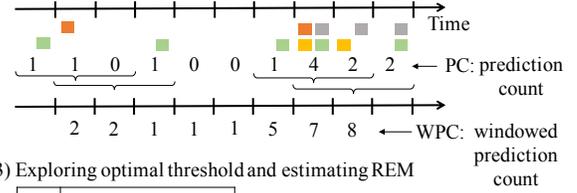
To extract characteristics of bio-vibration data by the ML, the frequency analysis is applied for decomposing each vibrations (i.e., heartbeats, respiration and body movement) to frequencies. This process is conducted as follows.

1. Applying the Fast Fourier Transform (FFT) (Cooley and Tukey 1965) to the bio-vibration data in a L -second window to convert the data to a power spectrum (note that the sampling frequency of the mattress sensor is 16Hz, and data size is $L \times 16$). In this study, window size (L) is set as next for capturing several scales of REM. $L = \{32, 64, 128, 256\}$. According to the sampling theorem (Shannon 1949), the frequency that can be analyzed by FFT is up to 8Hz, so that the data size of power spectrum is $L \times 8$, and the frequency resolution is $1/L$ Hz. Figure 2(a) shows the example of power spectrum ($L = 64$) calculated from bio-vibration data, where the vertical axis indicates the density of power spectrum and the horizontal axis indicates the frequency. In particular, the frequency band between 0.1Hz and 0.3Hz is related to the respiration, and the frequency band between 0.6Hz and 1.5Hz is related to the heartbeats. Regarding the BM, the larger/smaller BM, the higher/lower density of the power spectrum. However, as shown in Figure 2(a), it is difficult to understand the shape of the power spectrum above 1Hz because of the high density of frequencies below 1Hz.
2. In order to make it easier to understand above 1Hz and for RF to learn, power spectrum is converted into a logarithmic spectrum (\log_{10}). Figure 2(b) shows the example of the logarithmic spectrum converted from the power spectrum of Figure 2(a), where the vertical axis indicates

- (1) Learning each scale spectrum by each RF



- (2) Combining the REM predictions



- (3) Exploring optimal threshold and estimating REM

| TH | REM estimation ratio |
|----|----------------------|
| 1 | 60% |
| 2 | 38% |

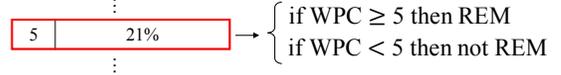


Figure 4: Overview of Multi-Timescale REM Estimation.

the density (logarithmic value) of the spectrum and the horizontal axis indicates the frequency. Furthermore, the density of each frequency in the logarithmic spectrum is normalized to 0, 1 based on the value of the density of the overall frequency.

3. This logarithmic spectrum is calculated per 30-second (stride size is 30-second) and labeled with the correct sleep stage (REM/Not-REM) determine by R&K method for RF to learn. Figure 3 shows the example of strides (window size is 128 seconds) and how to label sleep stage to the spectrum. When labeling sleep stage to the spectrum, bio-vibration data often have multiple sleep stages, so that, in this study, the sleep stage which is labeled to the spectrum is determined by a majority vote of the proportions occupied by those sleep stages. Note that, when using RF for REM prediction (not learning phase), the logarithmic spectrum is not labeled with the correct sleep stage, and the output of the prediction is for first epoch. The number of the input data that can be calculated from one subject (in case of seven hours of sleep) is about 840.

REM estimation based on multiple scales spectrum

Figure 4 shows the overview of the proposed Multi-Timescale REM Estimation. The flow of the method is as follows: (1) preparing RFs for a number of scales (each window size spectrum), and learning each scale spectrum by each RF; (2) combining the number of REM predictions by each RF in each window (note that, this window is different from window size of spectrum); (3) exploring optimal threshold for REM estimation from overnight data, and REM sleep is detected when the number of REM predictions in a window counted in (2) exceeds the threshold.

Combining REM predictions by each RF: Our method outputs four REM predictions for each epoch from RF of 32 sec., 64 sec., 128 sec. and 256 sec., and the REM Prediction Count (PC) is counted for each epoch as shown in top of Figure 4(2). Since REM sleep do not occur singly (one epoch)

Algorithm 1: Exploring optimal threshold

```

1:  $r\_list[]$ : REM estimation ratio for each threshold
2:  $WPC\_list[]$ : WPC for each epoch in overnight sleep
3: N_EPOCH: Number of epoch in overnight sleep
4:  $current\_TH \leftarrow 1$ : Start to explore from 1
5: IS_CONTINUE  $\leftarrow true$ 
6: while IS_CONTINUE do
7:   REM_COUNT  $\leftarrow$  count epochs detected as REM
   from  $WPC\_list$ 
8:    $r\_list.append(REM\_COUNT / N\_EPOCH)$ 
9:   if REM_COUNT is equal to 0 then
10:    IS_CONTINUE =  $false$ 
11:   end if
12:    $current\_TH++$ 
13: end while
14: optimal_TH: Optimal threshold found by the exploring
15: for  $i = 0$  to  $r\_list.length-1$  do
16:   if  $r\_list[i] > 0.25$  then
17:     break
18:   end if
19:   optimal_TH =  $i$ 
20: end for

```

but in clusters (successive epochs), in order to consider the state before/after the epoch wanted to estimate, the method prepare a window to count PC for N_e epochs before/after the epoch, which called Windowed Prediction Count (WPC) as shown in bottom of Figure 4(2).

Automatic adjustment of REM estimation threshold: The finally output of the proposed REM estimation for an epoch is determined by the value of WPC. If the WPC exceeded a certain threshold, then the epoch is detected as REM sleep, as shown in the Figure 4(3). It has a proportional relationship between the size of the threshold and the REM estimation ratio, which means the ratio of estimation out of all epochs in an overnight sleep (without considering correct and incorrect answers). According to the physiological characteristics of sleep, the proportion of REM sleep per overnight sleep is about 20%, so that the proposed method explores the threshold which the REM estimation ratio is about 20% for each person to avoid excessive or negative estimation.

The algorithm of exploring optimal threshold is described in Algorithm 1 The algorithm counts the number of epoch detected as REM sleep by a threshold in overnight sleep, and calculate the REM estimation ratio, while the REM sleep count is equal to 0 (see line 6 to 13). Then, the optimal threshold is extracted as the previous threshold where the REM estimation ratio exceeds 25% (see line 15 to 20).

Experiments

To investigate the effectiveness of the proposed Multi-Timescale REM Estimation, this paper conducted the human subject experiment of the nine of healthy subjects. The performance of the REM estimation is compared with RFs learned with each window size of logarithmic spectrum. The information of subjects is summarized in Table 1.

Table 1: Information of healthy subjects.

| ID (Age) | WAKE | REM | N1 | N2 | N34 | Total |
|----------|------|-----|----|-----|-----|-------|
| A (20's) | 46 | 176 | 41 | 421 | 164 | 848 |
| B (20's) | 53 | 113 | 48 | 368 | 2 | 584 |
| C (30's) | 103 | 184 | 35 | 382 | 0 | 704 |
| D (40's) | 75 | 65 | 46 | 419 | 2 | 607 |
| E (40's) | 44 | 85 | 53 | 390 | 23 | 595 |
| F (40's) | 34 | 121 | 40 | 225 | 0 | 420 |
| G (40's) | 53 | 110 | 80 | 407 | 1 | 651 |
| H (50's) | 35 | 249 | 56 | 520 | 0 | 860 |
| I (60's) | 98 | 159 | 27 | 436 | 0 | 720 |

The column ‘‘ID (Age)’’ indicates the ID of the subject and age of that. The columns from ‘‘WAKE’’ to ‘‘N34’’ indicate the number of epochs in each sleep stage (WAKE, REM, NREM1, NREM2 and NREM34), and the column ‘‘Total’’ indicates the total number of epochs in one night. The average number of epochs (30 seconds) of sleep is 664 ± 130 . As evaluation criteria, this study employs five evaluation indicators, Accuracy, Precision, Recall, F-measure and Specificity of REM estimation. In addition, this study evaluates the REM estimation ratio to see if REM estimations are being made at an appropriate frequency.

setup

The electrodes were attached to the body and head of each subject to acquire EEG, EOG and EMG, and mattress sensor was placed under the mattress in the bed to acquire bio-vibration data in one night. After sleep, the correct sleep stages for each subject were determined according to the R&K method based on the data measured by PSG (helped by medical specialist), and the bio-vibration data measured by mattress sensor is converted to logarithmic spectrums of several scales (i.e., window sizes are $L = \{32, 64, 128, 256, 512\}$) of which are labeled with the correct sleep stage in each epoch.

The logarithmic spectrum of each scale is learned with different RFs from each other. The training data is generated from the eight subjects, and the validation data is the other subjects. The ratio of REM sleep and not-REM sleep of training data is 1:3 because REM sleep accounts for 20% of one night sleep and to prevent excessive REM estimation. The data which have large BM are excluded because it affect the shape of the spectrum and difficult to learn with RF. The parameters of RF are set as follows: (i) the maximum depth of decision tree is 10; (ii) the number of decision tree is 50; (iii) the number of the features employed to construct the decision tree is 16, 23, 32, 46 and 64 for window size 32, 64, 128, 256 and 512 respectively. The window size N_e for counting WPC is set as 3.

Results

Table 2 shows the results of REM estimations. The column ‘‘Type of RF’’ indicates the RF learned with each window size of spectrum and rows ‘‘proposed 1’’ and ‘‘proposed 2’’ are the combination of multiple RF results, and the other

Table 2: Results of REM estimation by RFs learned with each window size and the proposed method. Values are expressed as mean \pm standard deviation.

| Type of RF | Accuracy | Precision | Recall | F-measure | Specificity | REM estimation ratio |
|-------------------------------------|----------------------------------|-----------------------------------|-----------------------------------|-----------------------------------|----------------------------------|----------------------|
| 32 sec. window | 74.1 \pm 8.5 | 33.0 \pm 21.3 | 9.8 \pm 13.6 | 11.3 \pm 10.3 | 91.5 \pm 12.2 | 8.8 \pm 12.3 |
| 64 sec. window | 76.0 \pm 7.1 | 38.5 \pm 21.3 | 12.9 \pm 18.9 | 14.2 \pm 12.3 | 92.9 \pm 11.6 | 8.4 \pm 13.3 |
| 128 sec. window | 76.6 \pm 8.3 | 47.3 \pm 14.8 | 20.6 \pm 16.5 | 24.8 \pm 9.8 | 91.5 \pm 12.3 | 11.0 \pm 13.3 |
| 256 sec. window | 78.0 \pm 7.8 | 52.8 \pm 20.8 | 24.5 \pm 17.4 | 29.9 \pm 12.6 | 92.0 \pm 11.2 | 11.6 \pm 12.4 |
| 512 sec. window | 76.3 \pm 5.9 | 25.8 \pm 16.8 | 4.6 \pm 3.8 | 7.2 \pm 5.3 | 95.6 \pm 4.4 | 4.5 \pm 4.1 |
| Proposed 1 (same TH (= 2)) | 79.0 \pm 8.1 | 52.8 \pm 20.7 | 55.4 \pm 21.3 | 51.1 \pm 16.4 | 84.9 \pm 9.8 | 23.3 \pm 11.1 |
| Proposed 2 (consider individual) | 80.2 \pm 5.5 | 51.4 \pm 15.0 | 47.0 \pm 15.5 | 48.5 \pm 13.7 | 89.0 \pm 3.0 | 18.5 \pm 3.4 |

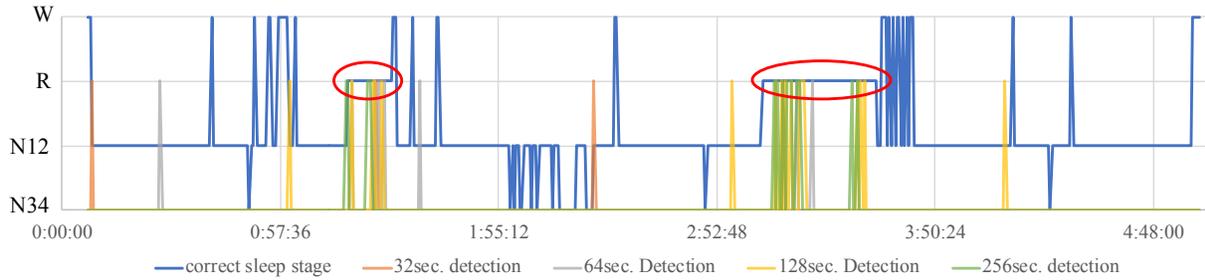


Figure 5: Estimation result of subject “E”.

columns are the evaluation indicators. Each value of indicators are expressed as mean value of nine subjects \pm standard deviation value of those, and the value is a percentage. As shown in Table 2, the Accuracy, Precision, Recall and F-measure are getting high rate as window size of spectrum getting wide, but when window size of spectrum is too wide (i.e., 512 sec.), these evaluation indicators get worsen than any other results. The reason why the value of Specificity in 512 seconds is most largest than any other results is that the REM estimation ratio of the RF leaned with 512 seconds is small (i.e., the REM estimation is passive).

Focusing on the results of two proposed methods, the values of Accuracy, Recall and F-measure outperform any other results (i.e., single RF leaned with each window size of spectrum). The values of Precision and Specificity are not the best among the result but these values are close to the best. In addition, the REM estimation ratio is larger than any other results, and the ratio is close to the average ratio of REM sleep per one night (about 20%). The standard deviation of each evaluation indicators are smaller for the proposed method 2, which considers individual differences, than for the proposed method 1, which does not considers individual differences, and this fact suggests the results are stable for each subject.

Discussion

How the combination of multiple RF results contributes?

Figure 5 shows the estimation result of subject E, where the vertical axis indicates sleep stage (WAKE, REM, NREM12,

NREM34), the horizontal axis indicates the time, the blue line indicates the correct sleep stage, and orange line, gray line, yellow line and green line are REM estimation result of 32, 64, 128 and 256 seconds respectively. As shown in Figure 5, the areas of the actual REM sleep marked by red cycles tend to have a concentration of REM estimation by four types of RF. On the other hand, the actual not REM sleep areas tend to have few REM estimations. Based on the proposed method, the number of REM estimations by each RF in a window interval of 3 epochs (1.5 minutes) before/after is shown in Figure 6. In Figure 6, the left and right vertical axes indicate the number of windowed REM prediction count (WPC) and sleep stage, respectively, the horizontal axis indicates the time, and the orange and blue lines indicate the WPC and correct sleep stage, respectively. As shown in Figure 6, the WPC tends to be high in the areas of the actual REM sleep, and it tends to be low in that of actual not REM sleep. The proposed method (the combination of multiple RF results) exploits this tendency and estimates REM sleep when the WPC exceeds a certain threshold. The threshold for REM estimation in the proposed method was determined by the sensitivity analysis of REM estimation threshold as shown in the Figure 8, where the vertical axis indicates the percentage, horizontal axis indicates the threshold, and blue, orange, gray, yellow, purple and green lines indicate Accuracy, Precision, Recall, F-measure, Specificity and REM estimation ratio, respectively. From the figure, the smaller the threshold value, the higher the REM estimation ratio, and the better the estimation of the actual REM sleep (as shown in the Figure 8, Recall), while the larger the threshold value,

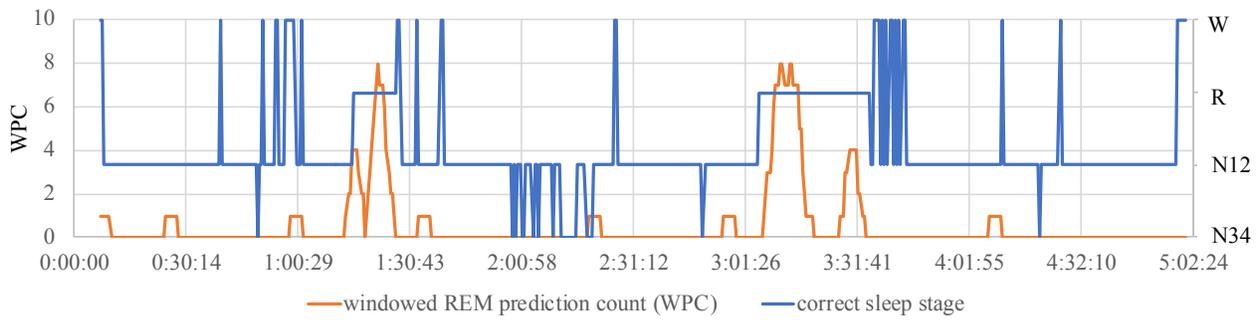


Figure 6: The number of windowed REM prediction count (WPC): subject “E”.

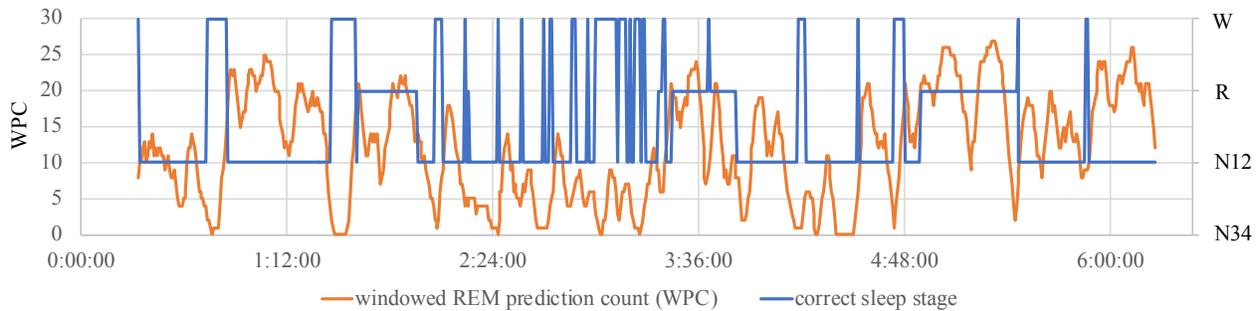


Figure 7: The number of windowed REM prediction count (WPC): subject “I”.

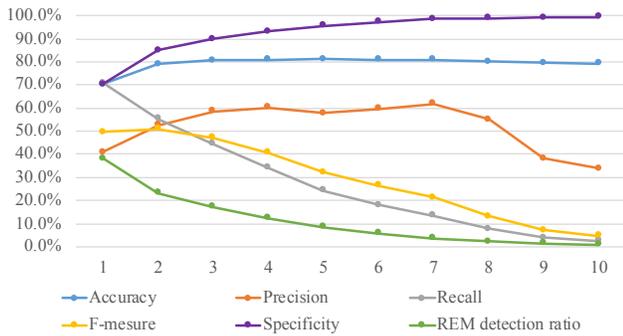


Figure 8: Sensitivity analysis of REM estimation threshold in the proposed method.

the lower the REM estimation ratio and the better the estimation of actual not-REM sleep (as shown in the Figure 8, Specificity). In this study, the threshold value ($= 2$) was chosen based on the Precision and Recall are almost equal and F-measure is the largest. However, this threshold is susceptible to individual differences (e.g., age and physical condition on the day) and must be carefully determined for each subject.

Consideration of individual differences in the proposed method

As mentioned above, this section discusses the importance of the threshold setting in the proposed method with subject

“I” who have to set a significantly different threshold compared to other subjects. Figure 7 shows the number of WPC of subject I, where the left and right vertical axes indicate the number of REM estimations and sleep stage, respectively, the horizontal axis indicates the time, and the orange and blue lines indicate the WPC and correct sleep stage, respectively. Compared to the results of subject “E” in Figure 6, the WPC of subject “I” is excessive, and it is not desirable to set a threshold of the same value. To deal with the individual differences the proposed method focused on the physiological characteristics about sleep that REM sleep accounts for about 20% of the total sleep in one night.

Table 3 shows the thresholds and results that were automatically adjusted for each subject so that the REM estimation ratio to the overall sleep time is about 20%. In the Table 3, the column “ID (Age)” indicates the subject ID and age, the columns “TH” and “REM estimation ratio” indicate the threshold (the value is an integer) for REM estimation and REM estimation ratio (the value is percentage) based on the threshold. The other columns indicate each evaluation indicators (the value is a percentage). Each threshold is set to a value close to 20% without the REM estimation ratio exceeding 25%. As shown in Table 3, the thresholds for all subjects, except subject “I”, are set between 1 and 4. If the threshold value of 4 is given to the subject “I” like the other subjects, the REM estimation ratio will be 86.1% and the Accuracy, Precision, Recall, F-measure and Specificity will be 35.9%, 25.7%, 99.4%, 40.9% and 17.7%, respectively. This situation should be avoided in real applica-

Table 3: The results of proposed method considering the individual differences (the row “proposed 2” in Table 2) in detail.

| ID (Age) | TH | REM estimation ratio | Accuracy | Precision | Recall | F-measure | Specificity |
|----------|----|----------------------|----------|-----------|--------|-----------|-------------|
| A (20’s) | 3 | 17.2 | 87.0 | 73.1 | 60.2 | 66.0 | 94.1 |
| B (20’s) | 4 | 19.4 | 72.4 | 29.5 | 29.2 | 29.3 | 83.0 |
| C (30’s) | 3 | 19.4 | 77.8 | 60.7 | 44.6 | 51.4 | 89.7 |
| D (40’s) | 2 | 9.3 | 84.8 | 23.2 | 21.3 | 22.2 | 92.0 |
| E (40’s) | 1 | 19.6 | 87.4 | 54.8 | 74.1 | 63.0 | 89.7 |
| F (40’s) | 4 | 20.1 | 77.7 | 63.9 | 46.1 | 53.5 | 89.9 |
| G (40’s) | 4 | 20.0 | 84.0 | 52.7 | 61.8 | 56.9 | 88.6 |
| H (50’s) | 1 | 21.1 | 72.7 | 54.4 | 39.4 | 45.7 | 86.4 |
| I (60’s) | 20 | 20.2 | 78.0 | 50.7 | 45.9 | 48.2 | 87.2 |

tions, and the proposed method makes the REM estimation with fewer wrong estimation by the condition that the keep the REM estimation ratio around 20%.

In general, to improve the accuracy of the estimation for such data, it is necessary to retrain the model by collecting similar data or devising new input features, which are difficult tasks and take a long time to do. By contrast, the proposed method does not need to do these things to improve the accuracy, and the only thing needed to do is set the REM estimation threshold. In addition, the threshold can be automatically determined based on the REM estimation ratio so that the proposed method makes it easy to adapt to individual differences. Therefore, the differences in the automatically determined thresholds, as shown in column TH of Table 3, represent individual differences. Since the heart rate is increased or unstable during REM sleep, it can be inferred, for example, that if the value of the threshold is high, heart rate of overnight may be higher or more unstable than an average person.

Conclusion

This paper proposed the novel REM estimation method that combination of multiple RF learned with different timescale of spectrums and investigates its effectiveness through comparison of the REM estimation by single RF learned with each scale spectrums. Concretely, the proposed method learns several RFs with each scale spectrums, then counts the number of REM estimation in the length of the window and estimates REM sleep if the counted number exceeds the threshold. Furthermore, the threshold is automatically determined for each person based on the REM estimation ratio to the overall sleep time for considering individual differences. The results of the human subject experiments, the Accuracy, Precision, Recall and Specificity of the REM estimation are $80.2(\pm 5.5)\%$, $51.4(\pm 15.0)\%$, $47.0(\pm 15.5)\%$ and $48.5(\pm 13.7)\%$, respectively. Through experiments, the following implications have been revealed: (1) the combination of RFs learned with multiple window sizes spectrum separately improves the Precision of REM estimation and Recall of that, rather than RF learned with only a particular window size spectrum; (2) the automatic adjustment of the threshold based on the REM estimation ratio to the total sleep length can be flexibly adapted to data with large individual differences without the need to retrain the model.

The future task is following: (1) to investigate the validity of the combination of multiple RFs; (2) to validate whether it is effective for other sleep stages.

References

- Breiman, L. 2001. Random forests. *Machine learning*, 45(1): 5–32.
- Cooley, J. W.; and Tukey, J. W. 1965. An algorithm for the machine calculation of complex Fourier series. *Mathematics of computation*, 19(90): 297–301.
- Goodfellow, I.; Bengio, Y.; Courville, A.; and Bengio, Y. 2016. *Deep learning*, volume 1. MIT press Cambridge.
- Holingue, C.; Wennberg, A.; Berger, S.; Polotsky, V. Y.; and Spira, A. P. 2018. Disturbed sleep and diabetes: A potential nexus of dementia risk. *Metabolism*, 84: 85–93.
- Mullington, J. M.; Haack, M.; Toth, M.; Serrador, J. M.; and Meier-Ewert, H. K. 2009. Cardiovascular, inflammatory, and metabolic consequences of sleep deprivation. *Progress in cardiovascular diseases*, 51(4): 294–302.
- Organization for Economic Cooperation and Development. 2019. GENDER EQUALITY, Gender Data Portal. <https://www.oecd.org/gender/data/>.
- Rechtschaffen, A.; and Kales, A. 1968. *A Manual of Standardized Terminology, Techniques and Scoring System for Sleep Stages of Human Subjects*. Washington DC.
- Shannon, C. E. 1949. Communication in the presence of noise. *Proceedings of the IRE*, 37(1): 10–21.
- Van Dongen, H.; Maislin, G.; Mullington, J. M.; and Dinges, D. F. 2003. The cumulative cost of additional wakefulness: dose-response effects on neurobehavioral functions and sleep physiology from chronic sleep restriction and total sleep deprivation. *Sleep*, 26(2): 117–126.
- Watanabe, T.; and Watanabe, K. 2001. Estimation of the sleep stages by the non-restrictive air mattress sensor relation between the change in the heart rate and sleep stages. *Transactions of the Society of Instrument and Control Engineers*, 37(9): 821–828.
- Watanabe, T.; and Watanabe, K. 2004. Noncontact method for sleep stage estimation. *IEEE Transactions on biomedical engineering*, 51(10): 1735–1748.