Multimed Tools Appl
DOI 10.1007/s11042-015-3053-z

A Bayesian approach for sleep and wake classification based on dynamic time warping method

Chunxiao Fu[1] · Pengle Zhang[1] · Jiang Jiang[1] · Kewei Yang[1] · Zhihan Lv[2]

Received: 17 September 2015 / Revised: 12 October 2015 / Accepted: 28 October 2015

Abstract Sleep plays a significant role in human' smental and physical health. Recently, the associations between lack of sleep and weight gain, development of cancer and many other health problems have been recognized. Then monitoring the sleep and wake state all night is becoming a hotspot issue. Traditionally it classified by a PSG recording which is very costly and uncomfortable. Nowadays, with the advance of internet of things, many convenient wearable devices are being used for medical use, like measuring the heart rate (HR), blood pressure and other signals. With the sleep quality monitor problem, the key question is how to discriminate the sleep and weak stage from these signals. This paper proposed a Bayesian approach based on dynamic time warping (DTW) method for sleep and wake classification. It used HR and surplus pulse O2 (SPO2) signals to analyze the sleep states and the occurrence of some sleep-related problems. DTW is an algorithm that searches an optimal alignment between time series with scaling and shifting and Bayesian methods have been successfully used for object classification in many study. In this paper, a three-step process is used for sleep and wake classification. In the first step, the DTW is used to extract features of the original HR and SPO2 signals. Then a probabilistic model is introduced for using the Bayesian classification for

✉ Chunxiao Fu
chunxiaosky@hotmail.com

✉ Zhihan Lv
lvzhihan@gmail.com

Pengle Zhang
paulplayergg@gmail.com

Jiang Jiang
jiangjiangnudt@163.com

Kewei Yang
kayyang27@nudt.edu.cn

[1] College of Information System and Management, National University of Defense Technology, Changsha, Hunan, People's Republic of China

[2] Shenzhen Institutes of Advanced Technology, Chinese Academy of Science, Shenzhen, China

Published online: 09 November 2015

uncertain data. And in the classification step, the DTW features are used as the training dataset in the Bayesian approach for sleep and wake classification. Finally, a case study form a real-word applications, collected from the website of the Sleep Heart Health Study, is presented to shown the feasibility and advantages of the DTW-based Bayesian approach.

Keyword Multimedia · Sleep and wake classification · Dynamic time warping · Bayesian classification · Wearable devices · Internet of things

1 Introduction

Sleep plays a significant role in human's mental and physical health. Recently, the associations between lack of sleep and weight gain, development of cancer and many other health problems have been recognized. A lot of people live with sleep-related problem that have a serious influence on one's health condition. Then monitoring the sleep and wake states throughout nights is becoming a hotspot issue [20]. Traditionally it classified by a PSG recording which is very costly and uncomfortable. Nowadays many convenient wearable devices are being used for medical use, like measuring the heart rate (HR), blood pressure and other signals. With the sleep quality monitor problem, the key question is how to discriminate the sleep and weak stage from these signals.

Sleep shows a complex, highly organized pattern of diverse physiological variables. Overnight Polysomnography (PSG) recordings with manually annotated hypnograms are regarded as the "gold standard" for analyzing the sleep states and occurrence of some sleep-related problems [29]. A PSG system is usually placed in a sleep laboratory and includes at least eleven channels including SPO2, electroencephalogram (EEG) and electrocardiogram (ECG) [4]. Signals are recorded for every 30-seconds epoch. PSG is not only very costly and not widely available but also obstructive and uncomfortable for the patients and thus not proper to monitor for long term.

Actigraphy measures activity from accelerometer data. It also satisfies many requirements of unobtrusive sleep monitoring [5]. It is typically worn on wrist and some on fingers. Recently such wearable devices become more popularized for the low cost as the pre-test before PSG. However the signals collected are relatively in low accuracy and not sound as PSG signals, which makes the method widely debated. An essential job related with this issue is to improve the classification accuracy with less requirements of signals to better discriminate between sleep and wake states at night.

In this paper, we try to discriminate the sleep and wake states with less requirements. Firstly the Dynamic Time Warping (DTW) method is used for extracting the "features" from the original signals which are collected from the devices (Heart rate and SPO2 signals from Sleep and Heart Health Study). After that multilayer perceptron Bayesian network classifiers are used to process the input series/signals that the features extracted by DTW.

Although Bayesian methods have been studied for years, it is only recently that their practical applications have become truly widespread because of fast computers. Experiments [3] show that the Bayesian approach is efficient to improve the performance of DTW based sleep and wake classification.

The rest of the paper is organized as follows. Chapter 2 briefly introduces Dynamic time warping and probabilistic classification methods especially Bayesian classification, also the basic situation of how they are used in signal processing. Chapter 3 presents how DTW and Bayesian approach are improved and used in this paper. Further experiments and results are shown in Chapter 4. Chapter 5 conclusion is made about proposed methods and our contribution.

2 Background

2.1 Dynamic time warping

Dynamic time warping is recently proposed for the use of similarity measure on time series data (also referred to as sequences hereafter). In this section we reflect the DTW similarity measure.

Consider two sequences: $X = < x_1, \ldots x_n >$, $Y = < y_1, \ldots, y_m >$, There series can be arranged and form a warping path in a m*n "warping matrix", in which each element is given by a distance function, presenting the squared distance between x_i and y_j:

$$\delta(i,j) = \sqrt{x_i - y_j{}^2} \tag{1}$$

The warping path presents alignments of the elements of the two series so that the total cumulative distance between them is minimized. The warping path W is one path of all the possible warping paths, and is denoted as:

$$W = w_1, \ldots, w_K$$

Where $w_k - \delta i, j_k$ is the kth alignment of the warping path W and max(m,n)≤K≤m+n–1. The Dynamic warping distance base on path W is:

$$DTW(X,Y) = \min\left[\frac{1}{K}\sqrt{\sum_{k=1}^{K} w_k}\right] \tag{2}$$

Figure 1 illustrates an example of dynamic warping process between series X and Y. And the process to find the optical warping path (alignment) is recursively computing the dynamic warping distance [16, 20].

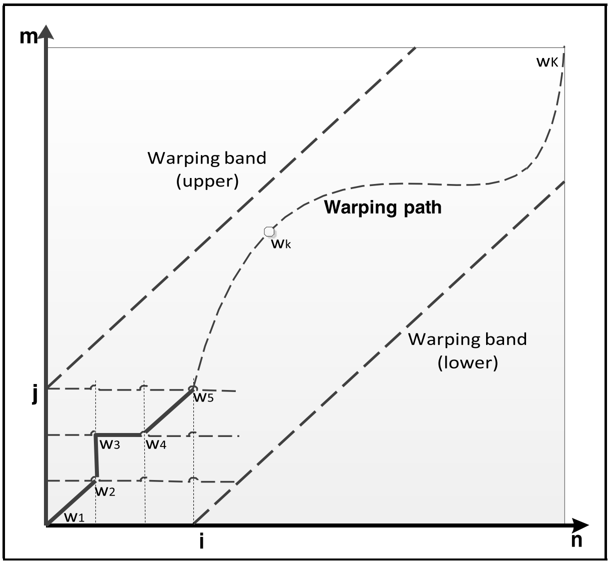


Fig. 1 The warping path map of series X and Y

2.2 Naive Bayes classifier

Probabilistic models are always used for describing pattern recognition problems. The Bayesian classifier estimates the class-conditional probability by assuming that the attributes are conditionally independent, given the class label Y_k. Consider there are M classes in total, $Y_1, Y_2, \ldots, Y_m$, thus $P(X/Y_k) = \prod_{i=1}^{n} P(X_i/Y_k)$, where X contains n attributes considered in the model [15, 30].

In the Bayesian approach, under the conditional independent assumption, the posterior probability for each class Y_k is computed as follow:

$$PY_k/X = P(Y_k)\prod_{i=1}^{n}(X_i/Y_k)/P(X) \tag{3}$$

The naïve Bayes classifier combines the probabilistic model with a decision rule. One common rule is to pick the hypothesis that is most probable, which is known as the maximum a posteriori or MAP decision rule. The corresponding classifier, a Bayes classifier, is the function that assigns a class label $y\hat{} = Y_k$ for one k as follows:

$$\hat{Y} = \text{argmax} PY_k \prod_{i=1}^{n} X_i/Y_k, k \in \{1, 2, \ldots, m\} \tag{4}$$

Most realistic problems focus on the uncertainty in attributes and certainty in class types. Thus in this paper a Bayesian approach is proposed or uncertain data base on the naïve Bayes classifier theorem. And the probabilistic model is describes in section 3.2 and 3.3.

3 Methods

In this chapter, sleep and wake classification started by describing the DTW-based classification method. Then focusing on the over-lapping problem in the DTW-based method, we setup our Bayesian approach with the assumption of the distribution prior information exists.

3.1 DTW-based classification analysis

DTW-based method solves the sleep and wake classification issue as a time-series classification problem. It is known that the breath rhythm and heart rate are usually more stable and more regular during sleep than when awake. And the DTW-based method assumes that a sleep epoch is more similar to another sleep epoch and less similar to wake epochs from the perspective of "series shape", regardless of being in the time or in the frequency domain. When used with respiratory effort, or heart rate, or some other signal, DTW-based method is expected to find a good match between the waveforms of the signal in two separate sleep periods, which is in contrast that it should not find any good match of the signal waveform between a sleep and a wake period, or even between two distinct wake periods [19]. And the method have been shown to be effective for using respiratory effort signals in some certain scenarios/data sets, as shown in Figs. 2 and 3 [11, 18, 20].

However, account for the possibly different time scales in measurement, different rates of changes, and noise or gaps, the signal sequence is potentially non-equal-length sequences aligned in time, typically in a non-parametric fashion. In our

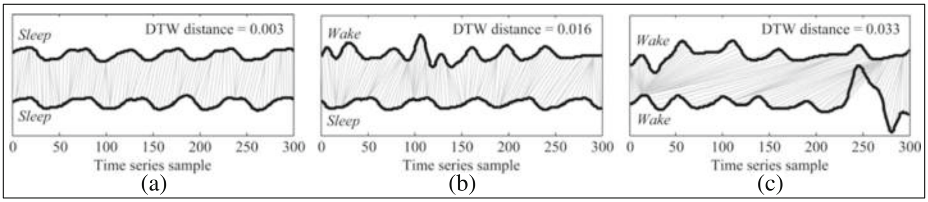


Fig. 2 DTW-based method of the respiratory time series. [20]

research it is found that using the DTW-based feature independently to distinguish the sleep and wake pattern would lead a wide overlapping, after checking out with a data set of 20 people's heart rate (HR) and oxygen saturation signals (SPO2) in Fig. 4.

Focusing on the overlapping problem in DTW-based method, we setup our Bayesian Approach to make the sleep and wake classification more accurate.

3.2 Model preliminaries and overview

In this paper, it is considered that the sleep and wake classification problem in Bayesian regression and classification method as time-series data. Consider a data set of input series $\{a_n\}^N{}_{n-1}$, noted as vector x , along with corresponding output targets $\{b_n\}^M{}_{n=1}$, noted as vector y . And assume that the input-targets pairs x,y are samples from a model by (5) with additive noise ε.

$$y = f(x) + \varepsilon \tag{5}$$

Least squares regression is a classic method to estimate the function f, typically for candidate functions f , given by (6)

$$f(x,w) - \sum_{i=0}^{M} w_i \bullet \phi_i(x) - w^T \phi(x) \tag{6}$$

where $\phi(x) = \phi_0(x), \phi_1(x), \ldots, \phi_M(x)^T$, $\phi_1(x)$ to $\phi_M(x)$ are M nonlinear fixed basis functions, which can be considered as characteristic features, w are the weight parameters, in which $w = (w_0, w_1, \ldots, w_n)$. Least squares regression is used here to determine a specific values for the parameter vector w, with minimizing this error equation, $\varepsilon(w) = \frac{1}{2}\sum_{t=1}^{T} \|f(x_t, w) - y_t\|^2$. Then the estimate

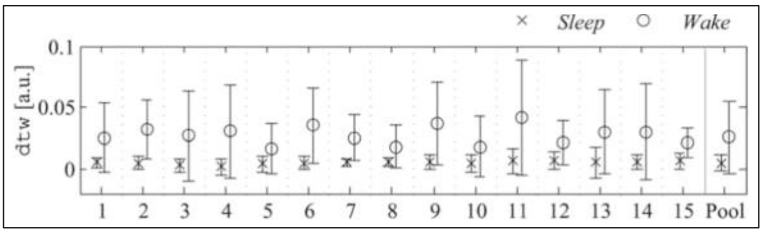


Fig. 3 Box plot of the DTW-based feature values. [20]

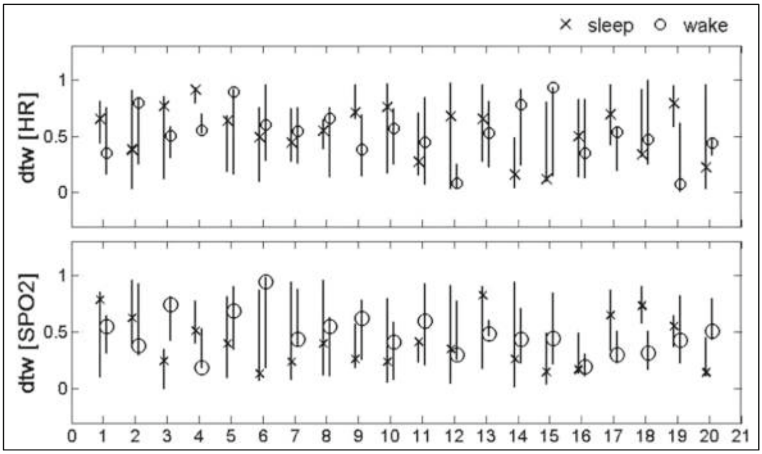


Fig. 4 Box plot of the DTW-based feature values of the 20 samples. These samples are provided by the Sleep Heart Health Study (SHHS) [17] shared on National Sleep Research Resource (NSRR) web [28]

w^* is used to make predictions for new values of input series x . As for the classification problems, the corresponding error function can be given by the cross-entropy [1, 2].

Bayesian regression method is another classic method to seek a point estimation of the unknown parameter vector w. And the difference is that the Bayesian approach focuses on the uncertainty in w through a probability distribution. Mostly we define a probability distribution $p(w)$ as the prior information, and then the distribution can be modified/improved by the observation of the pairs x,y data set in a maximum likelihood method. According to Bayesian theorem, the posterior distribution is the production of the prior distribution and the likelihood function. Then the predictive inferences can be made with the posterior distribution. Especially for a two-class classification problem, a logistic activation function can be used to interpret the output vector y as probabilities in each class, and make the classification with Bayesian decision rule [33].

A well-known problem with error function minimization is that complex and flexible models can 'over-fit', leading to poor generalization [3]. For a function in the form of (6), when the data set is small, nearly equals the number of parameters, the least square method can achieve a perfect fit, but having very bad generalization to new data point; when the data set is large, far more bigger than the number of parameters, the least square method can lead a poor fit, which can hardly be qualified for new data point generalization. However, we often have to handle with data set of limited size or big size. We wish to build a flexible models with adjustable parameters to fit the data set and to generalize new data point. And we shall see the Bayesian model with adjustable hyper-parameters can be applied to both small and big data set without encounter problems of over-fit.

As mentioned in the previous section, the DTW-based feature is applied in our Bayesian approach as the feature value. Though it may perform inefficiency in some certain scenarios, the feature combined with Bayesian network can reduce the "overlapped" problem. Thus a Bayesian network classifier is used to process the input series/signals. For sleep and wake classification problem, we establish a Bayesian network with one hidden layer and two output unit as shown in Fig. 5.

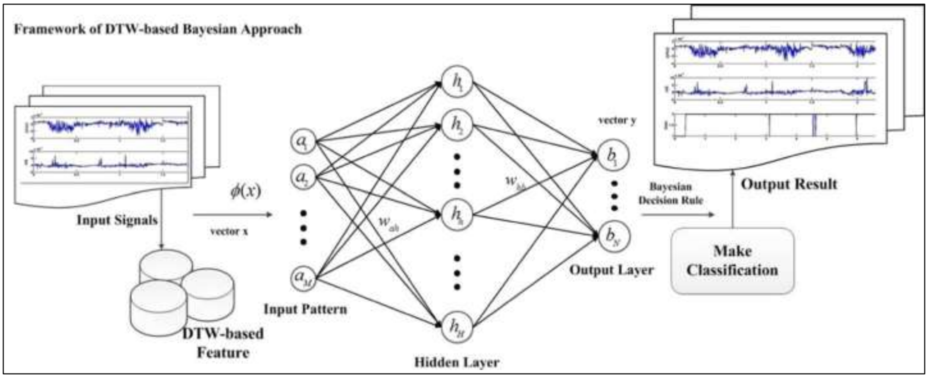


Fig. 5 Framework of the Bayesian network with DTW-based feature

In this Bayesian network, the general form of function f is given by (7),

$$y{=}f(x,w,h) = W_{HB}^{T}\left(H + W_{AH}^{T}\phi^{'}(x)\right) \tag{7}$$

Where $\phi'(x)-\phi_1^{'}(x),\phi_2^{'}(x),\ldots,\phi_M^{'}(x)^T$, $\phi_1^{'}(x)$ to $\phi_M^{'}(x)$ calculating the DTW-based feature from input original signal vector x. W_{AH} is the weight matrix between the input pattern and the hidden layer, H is the constant vector of the hidden layer, $H=(h_1,h_2,\ldots,h_H)^T$. W_{HB} is the weight matrix between the hidden layer and the output layer. That is, $W_{AH} = \begin{bmatrix} w_{a_1h_1} & w_{a_2h_1} & \cdots & w_{a_Mh_1} \\ w_{a_1h_2} & w_{a_2h_2} & & w_{a_Mh_2} \\ \vdots & & \ddots & \vdots \\ w_{a_1h_H} & w_{a_2h_H} & \cdots & w_{a_Mh_H} \end{bmatrix}$, $W_{HB} = \begin{bmatrix} w_{h_1b_1} & w_{h_1b_2} & \cdots & w_{h_1b_N} \\ w_{h_1b_1} & w_{h_2b_2} & \cdots & w_{h_2b_N} \\ \vdots & \vdots & \ddots & \vdots \\ w_{h_Hb_1} & w_{h_Hb_2} & \cdots & w_{h_Hb_N} \end{bmatrix}$. And the output y is a vector as the probability of sleep or wake.

Let $\phi_{(x)} = \begin{pmatrix} 1 \\ \phi_{(x)} \end{pmatrix}$, w as the general weight vector/matrix, $w'{=}W_{HB}^{T}H\ W_{HB}^{T}W_{AH}{}^{T}$ then Eq. (7) can be written as (8).

$$y{-}f(x,w,h){-}W_{HB}^{T}H + W_{HB}^{T}W_{AH}\phi(x){-}W_{HB}^{T}H\ W_{HB}^{T}W_{AH}\begin{pmatrix} 1 \\ \phi_{(x)} \end{pmatrix}{-}w^{'T}\phi(x) \tag{8}$$

Thus, commonly denoted as $y{=}f(x,w)$ to make it unify and simple to describe.

3.3 The Bayesian classification approach

From a probabilistic viewpoint, supposing the noise value ε in (5) is normally distributed by a zero-mean Gaussian process with variance σ^2, so that

$$p\left(\varepsilon/\sigma^2\right) = N\left(\varepsilon/0,\sigma^2\right) = \left(\frac{1}{2\pi\sigma^2}\right)^{1/2}\exp\left\{\frac{\varepsilon^2}{2\sigma^2}\right\} \tag{9}$$

From Eqs. (5) to (9), we have $p(y/x){=}N(y/f(x,w),\sigma^2)$, that is:

$$p\left(y/w,\sigma^2\right) = \left(\frac{1}{2\pi\sigma^2}\right)^{1/2}\exp\left\{-\frac{1}{2\sigma^2}\left\|y_t{-}w^T\phi(x)\right\|^2\right\} \tag{10}$$

Due to the assumption of the independence of the input series/vector x , so the likelihood function with the complete data set can be written as (11).

$$L(w) - \left(\frac{1}{2\pi\sigma^2}\right)^{T/2} \exp\left\{-\frac{1}{2\sigma^2}\sum_{t=1}^{T} \left\|y_t - \left|w^T\phi(x_t)\right|\right\|^2\right\} \tag{11}$$

Before observing y, we will process some prior information of w. We characterize w with a probability distribution $p(w)$. Specifically we define a Gaussian prior distribution over w in the form

$$p(w/\alpha) = \left(\frac{\alpha}{2\pi}\right)^{M/2} \exp\left\{-\frac{\alpha}{2}\|w\|^2\right\} \tag{12}$$

where α, as the hyper parameter, is the inverse variance of the Gaussian.

Using Bayesian theorem, the posterior distribution for w is the product of the prior distribution and the likelihood function.

$$p\left(w|\alpha,\sigma^2\right) \propto p(w|\alpha)L(w) \tag{13}$$

That is,

$$p\left(w|\alpha,\sigma^2\right) - K\frac{\alpha^M}{\sigma^T} \exp\left\{-\frac{\alpha}{2}\|w\|^2 - \frac{1}{2\sigma^2}\sum_{t=1}^{T} \left\|y_t - w^T\phi(x_t)\right\|^2\right\} \tag{14}$$

In a Bayesian approach, we make inference by integrating with respect to the posterior distribution of w. And we suppose to use the posterior distribution to find a point estimate for w, one classic statistic technique used for estimating w is *maximum likelihood*. For convenience, we can take the negative log of (14). Since the negative log function is a monotonically decreasing function, to maximum the equation (14) is equivalent to minimize (15)

$$\frac{\alpha}{2}\|w\|^2 + \frac{1}{2\sigma^2}\sum_{t=1}^{T} \left\|y_t - w^T\phi(x_t)\right\|^2 \tag{15}$$

Therefore, we can make the predictive distribution for y^*, given a new value of x,

$$p(y^*|x) = \int p\left(y^*|w,\alpha,\sigma^2\right) p\left(w,\alpha,\sigma^2|x\right) dw\, d\alpha\, d\sigma^2 \tag{16}$$

However, practically it is not possible to perform the computations in full analytically, and some approximation has to be done. The specific approximating method is introduced in [3], thus we can get the computable predictive distribution, for a new x, using: [24, 25]

$$p(y^*|x) \approx \int p\left(y^*|w,\sigma^2\right) p\left(w|\alpha,\sigma^2\right) dw \approx \int f(x,w) p\left(w|\alpha,\sigma^2\right) dw \tag{17}$$

However, the hyper parameter controls the scale and shape of the prior distribution of w, and the value of can be specified by modeling the distribution of w associated with each input x in order to minimize the probability of misclassification. A typical method for this procedure is known as the automatic relevance determination (ARD) [26, 27]. And the feature selection is also implemented in the training process [27, 34, 36].

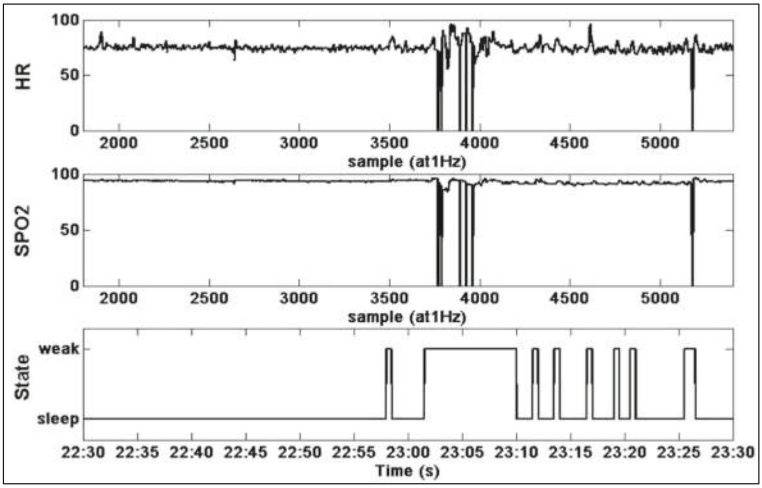


Fig. 6 Example of the PSG records with sleep state in 30 min with sample 1#

Once the weight of network has been trained, the Bayesian decision rule can be applied to perform pattern classification. It states as (18), where g represents the criteria state vector, sleep vector or wake vector. Find a vector for g that minimizing the DTW distance to vector y. We can decide the input pattern x for sleep if g=*sleep vector*, else decide x for wake.

$$\operatorname{argmin} p\left(\|y\text{-}g\|^2 \middle| w, \sigma^2\right), g \subset \{\textit{sleep vector},\ \textit{wake vector}\} \tag{18}$$

4 Experiment and evaluation

In this section, the method proposed in this paper is tested with the Sleep Heart Health Study (SHHS), a comprehensive set of experiments, implemented by the National Heart Lung &

Table 1 Demographic and clinical statistic of all the 20 samples. BMI: body mass index, in the unit of $kg\ m^{-2}$; Hour: recording hours. Data information is manually collected from the data reliability demonstration file, and some values are missing

All samples, n=20

Age	Male	BMI	Hour		Age	Male	BMI	Hour		Age	Male	BMI	Hour	
1#	56	M	26.07	8	2#	64	M	25.10	8	3#	62	F	27.15	8
4#	63	F	–	6	5#	56	M	29.89	–	6#	68	F	34.9	8
7#	70	M	24.31	6	8#	72	F	–	6	9#	56	M	–	8
10#	–	F	33.96	8	11#	55	M	33.53	8	12#	-	M	–	–
13#	59	F	28.85	8	14#	66	F	–	8	15#	52	M	33.64	6
16#	62	F	27.91	8	17#	60	M	–	6	18#	62	M	33.51	8
19#	53	M	24.23	8	20#	53	M	24.84	8					

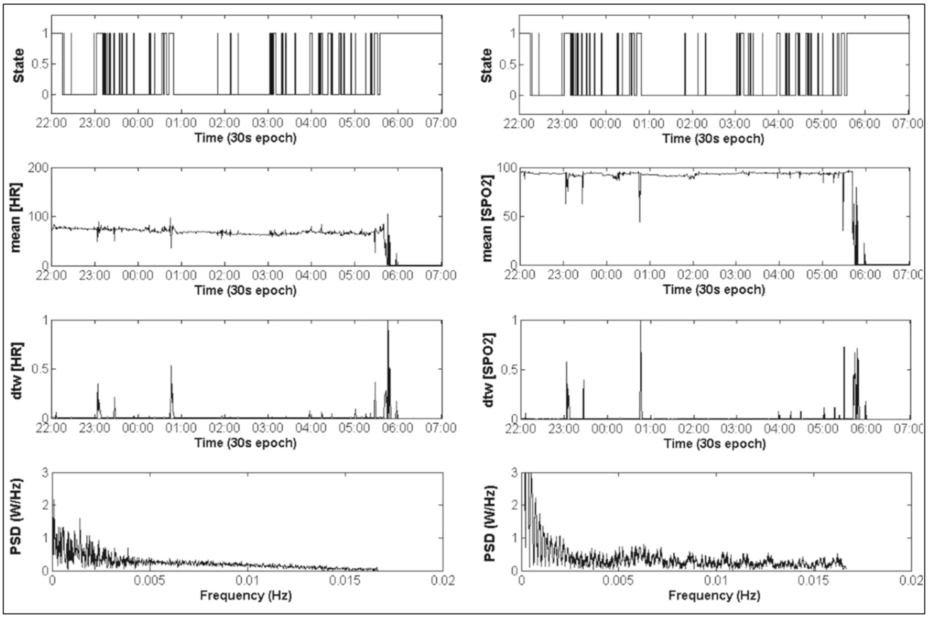


Fig. 7 Example of the features of the disposed signal for each 30s epoch, with sample 1# recorded at the frequency of 1 Hz, from 10:00 pm to 7:00 Am of the next day, by a conventional PSG device provided by SHHS [28]

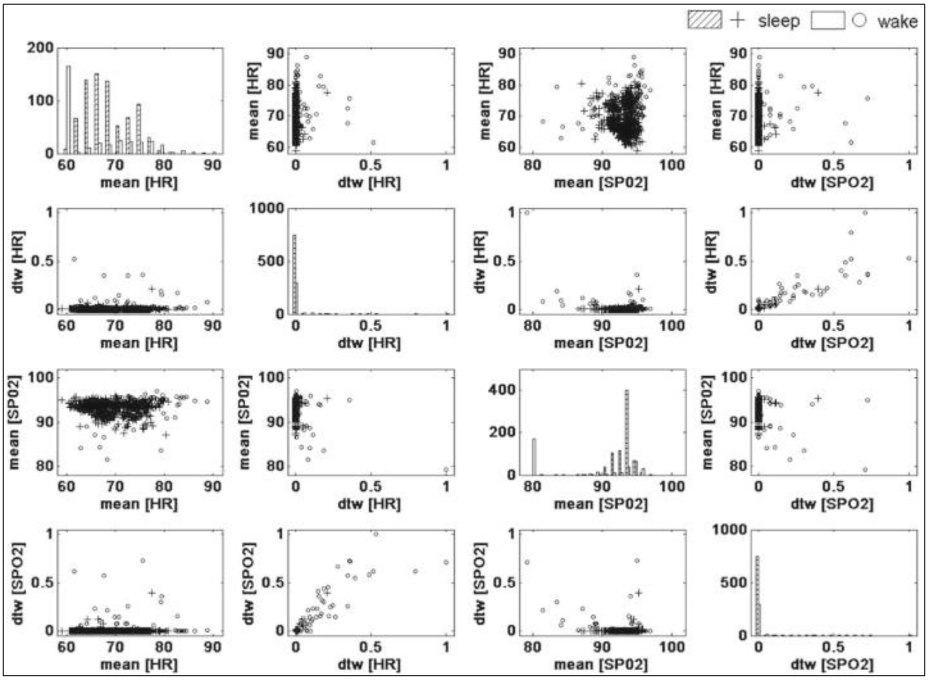
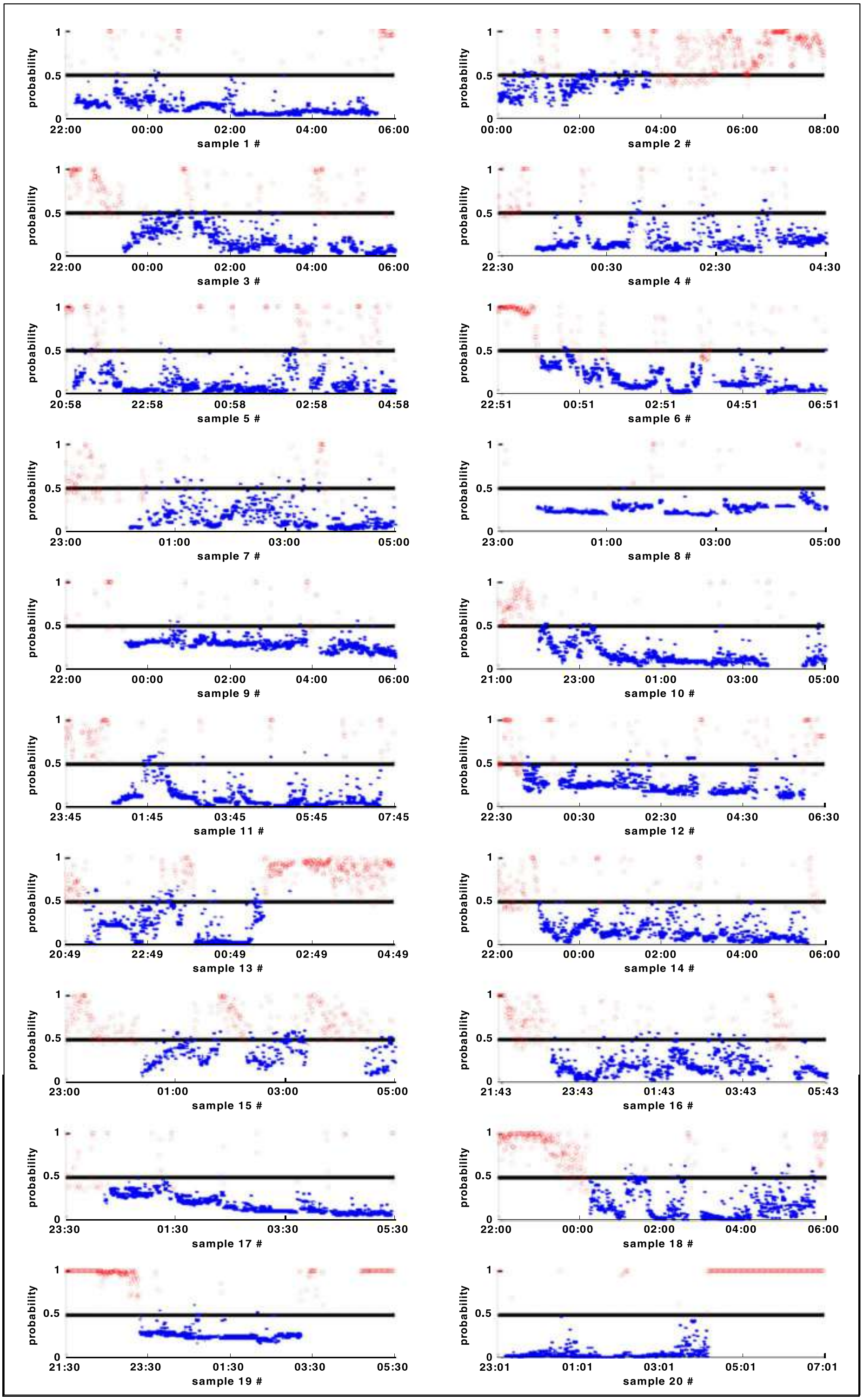


Fig. 8 Example of the feature distribution during sleep and wake with sample 1#

Table 2 Features of the data set during each iteration. Data is presented as mean and the feature value is statistic separately of sleep or wake

	All state				Sleep state				Wake state			
	HR	DTW_H	SPO2	DTW_SPO	HR	DTW_H	SPO2	DTW_SPO	HR	DTW_H	SPO2	DTW_SPO
1	66.6	0.02	89.7	0.01	68.4	0.01	93.0	0.00	62.5	0.03	82.3	0.03
2	57.3	0.03	94.1	0.04	53.9	0.01	94.5	0.01	59.0	0.05	93.8	0.05
3	71.8	0.02	95.6	0.01	69.6	0.01	95.6	0.00	76.7	0.04	95.7	0.04
4	60.4	0.04	96.0	0.01	59.1	0.02	96.1	0.00	63.2	0.09	96.0	0.04
5	78.6	0.02	95.1	0.02	76.7	0.01	94.8	0.01	82.6	0.04	95.8	0.03
6	67.8	0.02	95.3	0.01	65.1	0.01	95.2	0.00	74.6	0.06	95.4	0.02
7	65.0	0.06	96.2	0.02	63.2	0.05	96.5	0.01	69.8	0.09	95.2	0.04
8	59.3	0.02	96.3	0.01	59.2	0.01	96.5	0.00	59.4	0.04	96.0	0.04
9	61.8	0.02	90.1	0.01	62.0	0.01	91.5	0.00	61.3	0.04	87.5	0.02
1	58.1	0.04	97.1	0.01	56.8	0.03	97.2	0.00	62.5	0.09	96.7	0.04
11	64.4	0.04	93.2	0.01	62.2	0.02	93.0	0.00	71.2	0.07	93.6	0.03
1	71.3	0.02	92.6	0.04	71.0	0.00	93.4	0.02	71.9	0.06	91.1	0.07
1	73.7	0.04	93.4	0.02	72.4	0.02	90.6	0.01	74.7	0.06	95.6	0.02
1	61.9	0.03	94.6	0.04	60.3	0.01	95.2	0.01	65.2	0.06	93.2	0.10
1	57.4	0.05	95.2	0.03	56.1	0.03	94.9	0.02	58.9	0.07	95.5	0.05
1	50.6	0.01	90.9	0.01	50.2	0.01	90.5	0.01	51.4	0.02	91.8	0.03
1	73.6	0.02	97.3	0.02	72.5	0.01	97.5	0.00	77.5	0.07	96.4	0.07
1	82.2	0.03	94.6	0.02	81.3	0.00	94.2	0.00	83.5	0.05	95.2	0.04
1	52.3	0.01	89.8	0.02	50.1	0.00	95.3	0.00	53.9	0.01	85.7	0.03
2	65.1	0.01	84.3	0.02	75.9	0.01	95.0	0.02	52.7	0.02	72.0	0.02


probability
1
0.5
0
22:00 00:00 02:00 04:00 06:00
sample 1 #
00:00 02:00 04:00 06:00 08:00
sample 2 #
22:00 00:00 02:00 04:00 06:00
sample 3 #
22:30 00:30 02:30 04:30
sample 4 #
20:58 22:58 00:58 02:58 04:58
sample 5 #
22:51 00:51 02:51 04:51 06:51
sample 6 #
23:00 01:00 03:00 05:00
sample 7 #
23:00 01:00 03:00 05:00
sample 8 #
22:00 00:00 02:00 04:00 06:00
sample 9 #
21:00 23:00 01:00 03:00 05:00
sample 10 #
23:45 01:45 03:45 05:45 07:45
sample 11 #
22:30 00:30 02:30 04:30 06:30
sample 12 #
20:49 22:49 00:49 02:49 04:49
sample 13 #
22:00 00:00 02:00 04:00 06:00
sample 14 #
23:00 01:00 03:00 05:00
sample 15 #
21:43 23:43 01:43 03:43 05:43
sample 16 #
23:30 01:30 03:30 05:30
sample 17 #
22:00 00:00 02:00 04:00 06:00
sample 18 #
21:30 23:30 01:30 03:30 05:30
sample 19 #
23:01 01:01 03:01 05:01 07:01
sample 20 #

◀ **Fig. 9** DTW-based Bayesian approach result of 20 samples, the red circle is the wake state of each sample, and the blue dash point is the sleep state. The black line is the threshold value to distinguish sleep state or wake state of the Bayesian approach, the point above the line is wake, and the point below it is sleep. The training data set are selected as described in chapter 4.4, during each iteration one subset is left out used as test set and the remaining subsets are used to train the Bayesian network

Blood institute. It tests whether sleep related breathing is associated with an increased risk of coronary heart disease, stroke, all-cause mortality, and hypertension. During January 2001 to June 2003, a polysomnogram (PSG) was obtained in 3295 of the participants, and the according sleep stage result as provided by professor staff in SHHS [28].

4.1 DataSets

A total of 20 HR and SPO2 signals in PSG recordings from SHHS are available. The records usually start from midnight before the volunteers asleep to the next day morning after he/she wake up. And the sleep states results are initially provided by

Table 3 Compare of different classification methods

All samples, n=20

	DTW-based Bayesian approach				DTW method				Neural Network method			
	Pre%	Re%	Sp%	Ac%	Pre%	Re%	Sp%	Ac%	Pre%	Re%	Sp%	Ac%
1#	0.8908	0.6366	0.9654	0.8644	0.7339	0.2733	0.9561	0.7463	0.9944	0.5315	0.9987	0.8552
2#	0.8254	0.8462	0.6484	0.7794	0.6686	0.9902	0.0357	0.6682	0.2206	0.3254	0.4825	0.4337
3#	0.7872	0.5572	0.9303	0.8122	0.7227	0.2590	0.9540	0.7340	0.3993	0.6158	0.7951	0.7626
4#	0.7910	0.5128	0.9385	0.8057	0.7956	0.3993	0.9535	0.7806	0.2639	0.4914	0.7905	0.7509
5#	0.8811	0.7326	0.9541	0.8838	0.7435	0.4971	0.9203	0.7860	0.5439	0.5350	0.8704	0.7952
6#	0.8960	0.5839	0.9729	0.8616	0.7212	0.3839	0.9406	0.7814	0.4818	0.6198	0.8565	0.8146
7#	0.7299	0.5000	0.9293	0.8107	0.7212	0.3839	0.9406	0.7814	0.3878	0.6296	0.7873	0.7595
8#	0.7222	0.4354	0.9776	0.7247	0.7526	0.2535	0.9642	0.7508	0.1860	0.4455	0.7704	0.7362
9#	0.9245	0.3838	0.9829	0.7716	0.7973	0.1540	0.9787	0.6878	0.4281	0.6542	0.7856	0.7597
10#	0.7660	0.4303	0.9604	0.8376	0.5915	0.3347	0.9304	0.7924	0.4016	0.7299	0.8427	0.8284
11#	0.7568	0.5691	0.9394	0.8473	0.6591	0.1179	0.9798	0.7654	0.3190	0.3377	0.8671	0.7846
12#	0.8870	0.4604	0.9679	0.7884	0.6561	0.3636	0.8957	0.7075	0.3500	0.5291	0.7866	0.7407
13#	0.8615	0.8883	0.8176	0.8573	0.6429	0.8571	0.3920	0.6529	0.3416	0.3974	0.5878	0.5212
14#	0.8095	0.5812	0.9347	0.8204	0.7628	0.3390	0.9497	0.7523	0.4039	0.3445	0.8573	0.7449
15#	0.7673	0.6565	0.8187	0.7414	0.6355	0.7440	0.6116	0.6747	0.1382	0.3446	0.6079	0.5673
16#	0.7760	0.5527	0.9238	0.8039	0.6821	0.2934	0.9347	0.7274	0.1573	0.2781	0.7594	0.6924
17#	0.9028	0.3023	0.9905	0.8347	0.8590	0.3116	0.9850	0.8326	0.2707	0.6516	0.4678	0.5105
18#	0.8957	0.7992	0.9280	0.8718	0.9264	0.3192	0.9804	0.6919	0.5667	0.6296	0.7947	0.7454
19#	0.9916	0.9479	0.9931	0.9721	0.8100	0.4147	0.8671	0.6059	0.4686	0.7493	0.6145	0.6565
20#	0.9754	0.7576	0.9739	0.8490	0.9200	0.0461	0.9965	0.5562	0.0234	0.0241	0.6981	0.5422
Avg	0.8419	0.5917	0.9274	0.8269	0.7401	0.3868	0.8583	0.7238	0.3673	0.4932	0.7510	0.7001

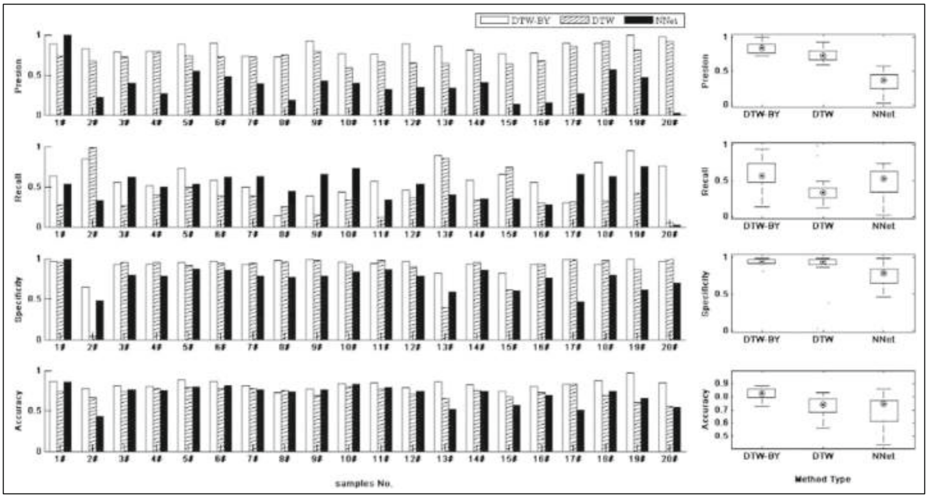


Fig. 10 Compare of different classification methods. The white bars labeled DTW-BY means the shadowed bars is the DTW-based Bayesian approach, DTW means the DTW method is presented in the black bar. The box plots on the right part give the distribution information of 20 samples among the precision, recall, specificity and accuracy

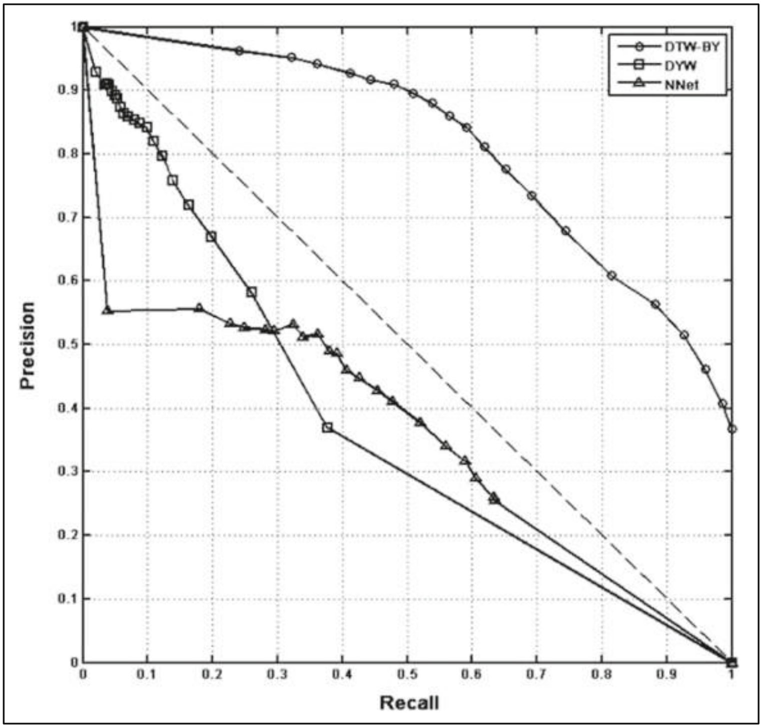


Fig. 11 Precision and recall curves

SHHS and have been revised by professor/doctors major on sleep health study. According to the standard, the signals are split into non-overlapping epoch of 30 s, and each epoch has a sleep or wake label [8, 21, 31]. A sample of the PSG records associated with the sleep states is shown in Fig. 6. The detail information of the whole data sets is listed in the following Table 1.

4.2 Preprocessing and feature generation

Before testing the Bayesian approach to the data set, we have performed some preprocess to the data set to decrease the noise and the outlier. Here we use traditional threshold filtering and smoothing algorithm to process the data set roughly [10, 32, 38]. According to the knowledge of biomedical that medicine signal often stays in the lower frequency, and then fast Fourier transform is used to cutoff the high frequency from the original frequency.

The mean value of HR and SaO2 signal during each 30 s epoch is selected as the feature, and as mentioned in chapter 2, The DTW distance between each two epoch is also selected as the feature. The preprocessing and feature generation can be illustrated in the following Fig. 7. Here the DTW distance is normalized [6, 9, 12].

In order to demonstrate the efficiency of our method, comparison of the distribution between the features above and the output of Bayesian network to discrimination sleep and wake is made [37]. Figure 8 shows the difference distribution of the features above, during sleep and wake. The sample is the same one in Fig. 7.

Due to the relatively small size of our data set, it is not appropriate to split it into separate training and test sets [7, 23]. The method is used that keep one subset used as test set and the rest subsets are used to train the Bayesian network in each iteration [22]. During each iteration, the features of the training data set are as list in Table 2.

4.3 Results and discussion

Using the mean value and the DTW distance of HR and SaO2 as the feature, the posterior distribution is used to find a point estimation of w, and that is equivalent to minimize Eq. (15).

There are many regression methods which have been discussed, thus the Laplace approximation procedure is used as in Mackay [24]. The hyper parameter α and σ valued by expert system, as α be the unite vector and α=0.01. Then according to Eq. (16), the probability $p(y^*/y,\alpha,\sigma^2)$ during each 30s epoch can be calculated, which means the probability of the wake state of this 30s epoch. Hence, the sleep vector is set as a singular point 0.5. Finally the results of the 20 samples in this DTW-based Bayesian approach are shown in Fig. 9. It can be seen our method classification efficiency of all the samples, that most of the samples have a good performance. The result of relatively high precision, recall, specificity and accuracy value is listed in the left column of Table 3.

To compare with the performances of different methods, we use the DTW method and the neural network method to give the contrast analysis, and the classification abilities of each sample are demonstrated in Table 3, which calculated respectively by DTW-based Bayesian, DTW, and neural networks method.

In order to show the different efficiency clearly, the following figure presented the results in multi-bars and box plot. We can see that most of the times, DTW-based approach has a better performance from the 20 samples statistic result as shown in Fig. 10.

The Precision and recall curves are shown in Fig. 11. The figure is generated by increasing the classified threshold p value from 0 to 1 step by 0.05, each point stand for the mean precision and recall value of 20 samples. The DTW-based Bayesian approach is perform best among the three methods.

5 Conclusion

Nowadays, with the advance of internet of things, many convenient wearable devices are used for monitoring signal during sleep. A method for classifying sleep and wake states using DTW features have been developed. The proposed method in this paper performs clustering on the features of the training data to characterize the sleep and wake states. It also introduces the use of dynamic time warping method to extract time series feature and the Bayesian classification approach with the probabilistic model to deal with uncertain attributes [13, 35].

The experiments shows evidence that the DTW features performs useful for characterize the sleep and wake states, which can be classified by the Bayesian approach proposed in this paper. The average precision of all the samples is 84.19%, and the recall is 59.17%. This means that a good performance can be obtained with fewer requirements for measuring physiological signals collected at nights [14].

We believe this work can be really contributive in medical areas. As future work, it is suggested that more examples from the PSG can be tested in this study to find out how to use this DTW-based Bayesian approach could characterize the sleep and wake states better. And it clears about what kind of devices need to be developed.

Acknowledgments This work was supported (in part) by the National Natural Science Foundation of China under Grant Nos. 71201168.

References

1. Bishop CM (1995) Neural Networks for Pattern Recognition. Clarendon Press, Oxford
2. Bishop CM, Bayesian PCA (1997) Microsoft reaserch
3. Bishop CM, Tipping ME (2004) Bayesian regression and classification, Comp Syst Sci, 190
4. CA K, MR L, TM M (2005) Practice parameters for the indications for polysomnography and related procedures: An update for 2005. Sleep 28:21
5. Devot S, Dratwa R, Naujokat E (2010) Sleep wake detection based on cardiorespiratory signals and actigraphy, 32nd Annual Int Conf IEEE EMBS
6. Dingde J, Guangmin H (2009) GARCH model-based large-scale IP traffic matrix estimation. Comm Letters 13(1):52–54
7. Dingde J, Xu Z, Li W, Chen Z (2015) Network coding-based energy-efficient multicast routing algorithm for multi-hop wireless networks. J Syst Software
8. Dingde J, Ying X, Han Y, Lv Z (2015) Collaborative multi-hop routing in cognitive wireless networks. Wireless Personal Communications

9. Dingde J, Zhengzheng X, Zhenhua C, Yang H, Hongwei X (2011) Joint time–frequency sparse estimation of large-scale network traffic. Comput Netw 55(15):3533–3547
10. Dongfang Z et al. (2014) FusionFS: toward supporting data-intensive scientific applications on extreme-scale high-performance computing systems. Big Data (Big Data), 2014 I.E. Int Conf. IEEE
11. Hedner J, Pillar G, Pittman SD, Zou D, Grote L, White DP (2004) A novel adaptive wrist actigraphy algorithm for sleep-wake assessment in sleep apnea patients. Sleep 27:10
12. Jiachen Y, Yun L, Qinggang M, Rongrong C (2015) Objective Evaluation Criteria for Stereo Camera Shooting Quality Under Different Shooting Parameters and Shooting Distances. Sensors J, IEEE 15(8):4508–4521
13. Jiachen Y et al. (2015) A low-power and portable biomedical device for respiratory monitoring with a stable power source. Sensors
14. Jiang D, Xu Z, Zhang P, Zhu T (2014) A transform domain-based anomaly detection approach to network-wide traffic. J Netw Comput Appl 40:292–306
15. Liu S, Cheng X, Fu W et al (2014) Numeric characteristics of generalized M-set with its asymptote [J]. Appl Math Comput 243:767–774
16. Liu S, Fu W, He L et al (2015) Distribution of primary additional errors in fractal encoding method [J]. Mult Tools Appl. doi:10.1007/s11042-014-2408-1
17. Liu S, Zhang Z, Qi L et al (2015) A Fractal Image Encoding Method based on Statistical Loss used in Agricultural Image Compression [J]. Mult Tools Appl. doi:10.1007/s11042-014-2446-8
18. Long X, Fonseca P, Foussier J, Haakma R, Aarts RM (2012) Using dynamic time warping for sleep and wake discrimination, Proc IEEE-EMBS Int Conf Biomed Health Informatics
19. Long X, Fonseca P, Haakma R, Aarts RM, Foussier J (2012) Time-frequency analysis of heart rate variability for sleep and wake classification, Proc 2012 I.E. 12th Int Conf Bioinformatics Bioengineering
20. Long X, Foussier J, Haakma R, Aarts RM (2014) Sleep and wake classification with actigraphy and respiratory effort using dynamic warping, IEEE J Biomed Health Informatics, 18
21. Lv Z, Tek A, Da Silva F, Empereur-Mot C, Chavent M, Baaden M (2013) Game on, science-how video game technology may help biologists tackle visualization challenges. PLoS One 8(3), e57990
22. Lv Z, Halawani A, Feng S, Li H, Shafiq Ur Réhman (2014) Multimodal hand and foot gesture interaction for handheld devices. ACM Trans Mult Comput, Communications Appl (TOMM) 11, 1s 10
23. Lv Z, Halawani A, Feng S, Shafiq ur Rehman, Li H (2015) Touch-less interactive augmented reality game on vision based wearable device. Personal Ubiquitous Comput
24. MacKay DJC (1992) The evidence framework applied to classification networks, Neural Computation, 4, 16
25. Marcos JV, Hornero R, Alvarez D, Nabney IT, Del Campo F, Zamarron C (2010) The classification of oximetry signals using Bayesian neural networks to assist in the detection of obstructive sleep apnoea syndrome. Physiol Meas 31:375–94
26. Nabney IT (2002) Netlab: Algorithms for Pattern Recogniton. Springer, London
27. Neal RM (1996) Bayesain Learning for Neural Networks. Springer, New York
28. R S, S MH, L BK, Q SF, I C, G DJ et al (2014) Sleep heart health study. Available: http://sleepdata.org/datases/shhs
29. Rechtschaffen EA, Kales A (1968) A Manual of Standardized Terminology, Techniques and Scoring System for Sleep Stages of Human Subjects. National Institutes of Health, Bethesda
30. Skolidis G, Sanguinetti G (2011) Bayesian multitask classification with Gaussian process priors. Neural Networks, IEEE Transactions on, 22(12)
31. Song Z, Huajiong J (2014) Fast log-Gabor-based nonlocal means image denoising methods. 2014 I.E. Int Conf Image Proc (ICIP). 2724–2728. IEEE
32. Tianyun Su, Wang W, Lv Z, Wu W, Li X (2016) Rapid Delaunay triangulation for random distributed point cloud data using adaptive hilbert curve. Comp Graphics
33. Wang Y, Hussein II (2011) Bayesian-Based Decision-Making for Object Search and Classification. IEEE Trans Control Syst Technol 19:9
34. Wang K et al. (2014) Next generation job management systems for extreme-scale ensemble computing. Proc 23rd Int Symp High-Performance Parallel Distributed Computing. ACM
35. Yang Y et al (2009) A contrast-sensitive reversible visible image watermarking technique. Circuits Syst Video Technol, IEEE Trans 19(5):656–667
36. Yang J et al (2014) Multiview image rectification algorithm for parallel camera arrays. J Electronic Imaging 23(3):033001–033001
37. Yang J et al (2014) Accurate and stable continuous monitoring module by mainstream capnography. J Clin Monit Comput 28(4):363–369
38. Yi W, Su Y, Agrawal G (2015) A novel approach for approximate aggregations over arrays. In Proc 27th Int Conf Sci Statistical Database Manag, ACM

Chunxiao Fu received the B.Sc. degree in Management Engineering from NUDT in 2014. She is currently a Master candidate in the School of Information Alternative and Management Science at NUDT. Her main research interests include Control systems, Model predictive control and data-based effective evaluation method.

Pengle Zhang received the B.Sc. degree in Management from NUDT in 2013. He is currently a Master candidate in the School of Information Alternative and Management Science at NUDT. His main research interests include Defense Acquisition and Alternative of Alternatives requirement modeling and data-based effective evaluation method.

Jiang Jiang received the B.Sc. degree and M.Sc. degree in alternative engineering from NUDT, PRC in 2004 and 2006, respectively. He is currently a lecture in School of Information Alternative and Management, NUDT. His research interests include multiple criteria decision making and risk analysis.

Kewei Yang received the Ph.D. degree in Management from NUDT in 2004. He is an Professor at the College of Information Alternative and Management in NUDT. His research interests focus on Intelligent Agent Simulation, Defense Acquisition and Alternative of Alternatives requirement modeling.

Zhihan Lv is a native Chinese. He is an engineer and researcher of virtual/augmented reality and multimedia major in mathematics and computer science, having plenty of work experience on virtual reality and augmented reality projects, engage in application of computer visualization and computer vision. His research application fields widely range from everyday life to traditional research fields (i.e. geography, biology, medicine). During the past years, he has completed several projects successfully on PC, Website, Smartphone and Smartglasses.