8-2016

Automatic Sleep Stages Classification

Maryam Zokaeinikoo
University of Tennessee, Knoxville, mzokaein@vols.utk.edu

To the Graduate Council:

I am submitting herewith a thesis written by Maryam Zokaeinikoo entitled "Automatic Sleep Stages Classification." I have examined the final electronic copy of this thesis for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Master of Science, with a major in Industrial Engineering.

Anahita Khojandi, Major Professor

We have read this thesis and recommend its acceptance:

Oleg Shylo, Xueping Li

Accepted for the Council:
Carolyn R. Hodges

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

Automatic Sleep Stages Classification

A Thesis Presented for the

Master of Science

Degree

The University of Tennessee, Knoxville

Maryam Zokaeinikoo

August 2016

To my parents

For their endless love, support and encouragement

And all of my friends,

Without whom none of my success would be possible

Abstract

In this thesis, we first develop an efficient automated classification algorithm for sleep stages identification. Polysomnography recordings (PSGs) from twenty subjects were used in this study and features were extracted from the time–frequency representation of the electroencephalography (EEG) signal. The classification of the extracted features was done using random forest classifier. The performance of the new approach is tested by evaluating the accuracy of each sleep stages and total accuracy. The results shows improvement in all five sleep stages compared to previous works.

Then, we present a simulation decision algorithm for switching between sleep interventions. This method improves the percentage of average amount of sleep in each stage. The results shows that sleep efficiency can be maximized by switching between intervention chains.

Table of Contents

List of Tables

List of Figures

Chapter 1

Sleep Stages Classification

Sleep is a natural part of every person's life. One-third of every individual's life is spent in the sleeping state. Many restorative functions of body including physical recreation and immune functions as well as mental restoration, memory consolidation, mood and behavior are dependent on a healthy sleep. Deprivation of sleep can lead to rising risk of serious health problems such as heart diseases, obesity, diabetes and weakness of immunity system. Sleep is non-homogeneous process and has an internal cyclical structure. Each cycle consists of different sleep stages. Rechtschaffen and Kales (R-K) introduce sleep stages based on visual observations of the patterns and signals of electroencephalography (EEG), Electrooculography (EOG), and Electromyography (EMG). This standard method of sleep evaluation is called polysomnography (PSG). Based on R-K sleeping criteria, sleep stages are defined as: awake, rapid eye movement (REM) and non-rapid eye movement (NREM) which includes stages 1, 2, 3 and 4 [15]. Sleep stages 3 and 4 are often combined together and considered as the deep sleep stage. Transitions between these stages occur in stochastic directions and at unpredictable moments. Based on this, sleep has been modeled as a semi-Markov stochastic process [20].

Due to the complicated system of PSG recordings which is costly, time consuming and uncomfortable for the patient, automatic sleep scoring system could be very

helpful. Several automatic sleep staging studies have been published in the literature which used the public dataset which is available online (physionet online dataset) [1, 3, 4, 9]. Other studies apply different datasets. Some of them use smaller datasets (less than 10 subjects) [6,10,14,18]. Datasets with more subjects have been applied by different studies [5,11,12,19]. Most of previous works use multiple channels including EEG channels, EOG and EMG channels [3, 13, 14, 19, 21]. However, there are a few works which consider only one EEG channel for prediction [1, 4, 5, 9–12]. Several methods in the literature have been conducted to extract features from channels including: discrete wavelet transform (DWT) [13,14,21]; fast fourier transform (FFT) [9,14,21] ; Welch method [7]; continuous wavelet transform (CWT) [5]; time-frequency image (TFI) [1]. Based on these methods, several features can be extracted which their numbers differs in different studies. Some researches use few features in order to train their algorithm (less than 30 features) [1,3–5,9,11,13,21]. There are other studies which extract too many features for their classification algorithm. For instance, Ozsen et al.(2013) extract 57 features from EEG, EMG, left and right EOG signals. These features are mainly statistical features of signal (mean, standard deviation, skewness, kurthosis, etc) as well as power features of sub-bands. Gunes et al.(2010) extract 258 features by applying Welch method for EEG and chin EMG signals. They use statistical measures including minimum value, maximum value, standard deviation, and mean value belonging to each feature in sleep stage dataset in order to decrease to eight features for EEG and chin EMG signals. Different algorithms have been applied to classify sleep stages. Artificial neural network (ANN) algorithm have been applied by several studies [4, 9, 13, 14, 18, 19]. Zoubek et al. [21] apply three different classifiers including two bayes rule-based classifiers (parametric and non-parametric ones) as well as multi-layer perceptron (MLP). Doroshenkov et al.(2007) use the classification approach based on hidden Markov model (HMM). Gunes et al.(2009) use C4.5 decision tree to classify sleep stages. Fraiwan et al.(2012) apply Random forest (RF) as a classifier. Bajaj and Pachori(2013) use the multiclass least squares support vector machines (MC-LS-SVM) with the radial basis function (RBF),

Mexican hat wavelet, and Morlet wavelet kernel functions to classify sleep stages of EEG signals.

Finally, after feature extraction and applying the appropriate classification algorithm, different accuracy ranges have been reported in the literature. Many researchers report only the total accuracy without going to the details of each sleep stages. Other studies combine some sleep stages together and obtain the accuracy. There are few researches which obtain the accuracy of each sleep stages. However, they report the low accuracy for the sleep stage one.

The purpose of this present study is to develop an automatic sleep staging system based on single-channel EEG for patient convenience. In this study, the dataset is acquired from physionet online dataset which consists of two EEG channels (Fpz-Cz and Pz-Oz), EOG and EMG. Each channel is divided to 10-second epochs. Seven different features are extracted from time-frequency domain of signal. The number of features used in this study is smaller than that of previous works. The novel method of structuring of features is proposed in this research which considers the history-dependent feature of sleep epochs. By using this method of structuring dataset and RF as a classifier, we succeed to obtain the higher classification accuracy for total accuracy as well as accuracy of each sleep stages (specially stage 1).

1.1 Methods

In order to classify sleep stages automatically, we need an efficient signal processing technique to extract features. First, the PSG recordings are obtained from physionet online dataset. In addition to two EEG channels (Fpz-Cz and Pz-Oz), EOG and EMG channels are selected for feature extraction. By applying the signal processing method, the features are extracted. Finally, they have been used as input variables for the classification algorithm.

1.1.1 Signal Processing

Wavelet transform (WT) is a powerful technique in signal processing for solving various real-life problems. This method analyzes non-stationary signals which frequency responses varies in time in both time and frequency . Wavelet is a small wave which its energy is concentrated in time to analyze EEG signal as a non-stationary signal. Wavelet analysis measures the frequency similarity between the signal and the original wavelet(mother wavelet). In WT, computations are done for different frequency components (scale) by shifting the time window until the wavelet reaches at the end of the signal.WT has a precise time resolution at high frequencies and good frequency resolution at low frequencies. With this feature, WT helps the analysis of non-stationary signals.

The continuous wavelet transform (CWT) of a signal, $x(t)$, is defined as follows,

$$CWT(a,b) = \int_{-\infty}^{\infty} x(t) \frac{1}{\sqrt{|a|}} \psi(\frac{(t-b)}{a}) dt. \tag{1.1}$$

The coefficients of CWT are computed within this formula by the integral of the original signal which is multiplied by a mother wavelet. The scaling parameter (a) is related to frequency. High scales correspond to low frequencies which give information of the entire signal whereas low scales (high frequencies) give detailed information in the signal. The parameter b corresponds to the location of time window which is shifted over the length of the signal. In fact, CWT measures the similarity of the frequency in the original signal and the mother wavelet. The CWT has a weak point for calculating coefficients at each scale. Because it requires expensive computational task as the matter of redundancy. The Discrete Wavelet Transform (DWT)solves this problem by operating filtering tasks.In this procedure, the signal is passed through a half band low pass filter which results removing some samples from signal. Therefore, the scales and time window shifts are chosen based on powers of two (dyadic). The DWT is defined as,

$$DWT(j,k) = \frac{1}{\sqrt{|2^j|}} \int_{-\infty}^{\infty} x(t)\psi(\frac{(t-2^j k)}{2^j})dt, \tag{1.2}$$

where a and b in the CWT are replaced by 2^j and $2^j k$, respectively. At every level of the DWT, the signal is passed through a low pass (LP) and high pass (HP) filters which results in half number of samples and half the frequency. The outputs of LP and HP at each level i are called approximation (A_i) and detail (D_i) coefficients , respectively. Figure 1.1 shows the wavelet decomposition of a signal through 3 levels of filtering. In this figure, the coefficients A_1, D_1, A_2, D_2, A_3 and D_3 are the DWT coefficients.

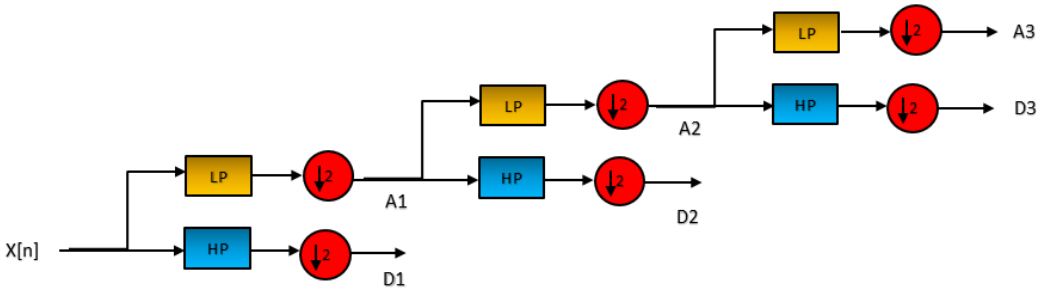

Figure 1.1: Sub-band filtering of DWT implementation

In this figure, the discrete $x(n)$ signal crosses has the sampling rate of (100Hz) which passes iteratively through HP to generate detail coefficients ($D_i[n]$) and crosses through LP to obtain approximation coefficients ($A_i[n]$). In analysis of EEG signals, the number of levels of decomposition is chosen based on the sampling rate of the original signal and the range of frequency components which are desired to be extracted from the signal. Since the range of the useful frequency information of EEG signals falls between $0-60$ Hz, usually decomposition level is set at five. Selecting inappropriate number of decomposition levels causes loss of desired information. The five level of DWT decomposition of EEG data (100 Hz) is given in Figure 2. It can be seen that the components A_5 decomposition is within the delta range ($0-3$ Hz), D_5 decomposition is within the theta range ($3-6$ Hz), D_4 decomposition is within

the alpha range $(6-12$ Hz) , D_3 decomposition is within the beta range $(12-25$ Hz) and D_2 decomposition is within the gamma range $(25-50$ Hz). Therefore, in order to extract the meaningful features from the EEG signal, D_2-D_5 detail sub-bands and A_5 approximation band are used in this study. Several successful studies related to EEG choose Daubechies wavelet as an appropriate wavelet as well as level four and level five of this function is preferred. In this paper, db5 is selected as the mother wavelet for DWT decomposition (Figure 1.2).

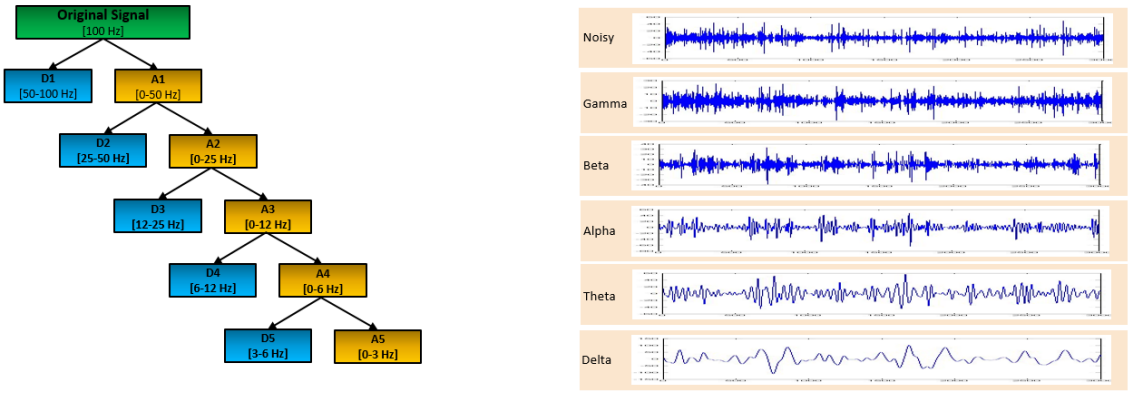

Figure 1.2: Sub-band decomposition of DWT implementation

1.1.2 Features

In this section, the features of signals are obtained based on the coefficients which are the results of DWT implementation. These coefficients are computed for every sample of the original signal $x(n)$. The coefficients of D_1, D_2, D_3, D_4, D_5 and A_5 are related to noisy, gamma, beta, alpha, theta and delta sub-bands, respectively.

Let X_i^n, $i=1,2,...,6$, be the vector of coefficients of relative sub-bands of the n^{th} epoch in a signal channel, where, D_{ij}^n and A_{5j}^n, $i=1,2,...,5$, $j=1,2,...,N$ are the

coefficients of sample j in epoch n.

$$X_1^n = (D_{11}^n, D_{12}^n, \cdots, D_{1N}^n) \tag{1.3}$$

$$X_2^n = (D_{21}^n, D_{22}^n, \cdots, D_{2N}^n) \tag{1.4}$$

$$X_3^n = (D_{31}^n, D_{32}^n, \cdots, D_{3N}^n) \tag{1.5}$$

$$X_4^n = (D_{41}^n, D_{42}^n, \cdots, D_{4N}^n) \tag{1.6}$$

$$X_5^n = (D_{51}^n, D_{52}^n, \cdots, D_{5N}^n) \tag{1.7}$$

$$X_6^n = (A_{51}^n, A_{52}^n, \cdots, A_{5N}^n) \tag{1.8}$$

Each channel is divided to t-second epochs where the number of samples in these epochs depends on the sampling rate of signal. For instance, the 10-second epoch with sampling rate of 100 Hz has $10 \times 100 = 1000$ samples.

In order to distinguish between different sub-bands, we need to quantify them as features. The wave energy vector of sub-band i of epoch n (E_i^n) is defined as the summation squared signal components,

$$E_i^n = X_i^n (X_i^n)^T. \tag{1.9}$$

After computing the energy of six sub-bands in each N-second epoch, we define total energy of n^{th} epoch (T^n) which is the summation of these six energy sub-bands,

$$T^n = \Sigma_{i=1}^6 E_i^n. \tag{1.10}$$

Another feature which is used in this study is the entropy (ξ). Entropy measures relative randomness of variables. In computer-brain systems, it is used to show the the level of chaos of the system [16]. In terms of sub-bands of EEG signal, it is defined as the following,

$$\xi^n = -\Sigma_{i=1}^6 p_i^n \log p_i^n, \tag{1.11}$$

where p_i^n is defined as the relative energy of sub-bands,

$$p_i^n = \frac{E_i^n}{T^n}, \tag{1.12}$$

Shanon entropy is another concept in signal processing. In the following expression s is the signal sub-band and s_i is the ith component of this sub-band,

$$E1(s) = -\Sigma_{i=1}^{N} s_i^2 \log s_i^2, \tag{1.13}$$

We used some statistical features as well:

- Variance, minimum, maximum, median, range, skewness, kurtosis of signal's points (7 features)
- Standard deviation and mean of signal components for each sub-band (12 features)

The method used in this thesis which leads to higher accuracy compared to previous works is coming from the idea that sleep is a stochastic process (semi-Markov process) and the history of data leads to strong supporting information to train the dataset. Our history-dependent model is defined by combining the preceding epoch of each t-second epoch as another feature and repeat this process iteratively to obtain the optimum number of previous epochs which leads to higher accuracy.

In order to evaluate the transition between epochs, we consider the succeeding epochs as well as the preceding epochs. We report our results with and without these neighboring features.

1.2 Computational Results

In this section, we implement a classification algorithm with the available public dataset at Physionet. Then we run our algorithm with all possible combinations of parameters to present a comprehensive experimental design.

1.2.1 Classification Method

The dataset which is used in this study includes signals from EEG (Fpz-Cz and Pz-Oz channels), horizontal EOG, submental chin EMG and an event marker. This dataset consists of PSG recordings of 20 healthy subjects from the study of age effects on sleep.

According to National Institute of Neurological Disorders and Stroke, a normal person spends 50% of the whole night sleep in stage 2, 20% in stage REM and the remaining 30% in other sleep stages [8]. Therefore, the distribution of the number of the epochs is not uniform among sleep stages (classes) in any PSG recordings. We have more epochs which classified as sleep stage 2 than other sleep stages. In addition, we consider 10 minutes before the person falls sleep and 10 minutes after the person becomes awake for each PSG recording. This consideration enables us to test the transition between awake stage when the person is totally conscious and non-awake sleep stages.

Many classifiers were used in the literature for the process of feature classification. However, random forest classifier used in this work which has the better performance compared to other algorithms [17]. Random forest classifier is an ensemble classifier using many decision tree models [2].

Random forest classifier consists of many tree predictors (the number of trees which is used in this study is 500) which created by selection of random samples with replacement of the training set. At each node of these tree classifiers, random number of features are selected (in this study five features) to build subsequent nodes and leafs. The same process is repeated until all trees is constructed. The majority votes from out put of these trees determines the classification output. The whole algorithm is coded in *R*.

1.2.2 Experimental Design

To obtain the optimum result, we need to test the combination of three variables including: number of levels of dependency (d), channels (Fpz-Cz (C_1), Pz-Oz (C_2) , EOG (C_3) and EMG (C_4)) and epoch length. All the experiments are done for each combination of three variables which are: preceding dependency levels (p),succeeding dependency levels (s), epoch length (t) and channel combination (c). The values of these variable are shown in table 1.1.

Table 1.1: Variables

Parameter	Levels	Values
p	2	0, 2
s	2	0, 2
t	8	1, 2, 3, 5, 6, 10, 15, 30
c	15	all combinations of 4 channels

To compare our work with other studies, we test our algorithm on only the first channel with 30-second epochs with and without dependencies of previous epochs.

Since we have imbalance dataset with different number of epochs in each classes, it is probable our algorithm ignores the class with the least epochs (class of sleep stage one). In order to avoid skewed results, we used balanced random forest by down-sampling the majority classes. In other words, for each tree in random forest classifier, we consider the class with the least number of epochs (sleep stage one) and decrease the number of epochs in other classes by sampling with replacement to have the same classes in number of epochs.

To implement the down-sampling method for random forest algorithm, we create many forests with limited number of trees. In each forest we sample the same amount of sleep stage 1 epochs for other sleep stages with replacement. Therefore, we will have balanced dataset for each forest (table 1.2). After training every forest, we combine them to have one balanced training model for testing.

Table 1.2: Balancing dataset using downsampling

Number of epochs	stage 1	stage 2	stage 3	stage R	stage W
Unbalanced dataset	2749	15110	4856	6905	4783
Balanced dataset	2749	2749	2749	2749	2749

1.2.3 Evaluation

In order to asses the predictive ability of our algorithm, we need to test it on set of of data set not used in training. K-fold cross validation technique is a common way to measure how accurately a predictive model will behave in future. Based on our data set, we have recordings of 20 subjects with two subsequent nights (except one subject with one night recording). Therefore, we can use 20-fold cross validation which each fold consists of combination of two nights recordings of each subject for testing and all other recordings for training.

The measures we use in this study are sensitivity, precision, F1-score, per-stage accuracy and overall accuracy. F1-score is the harmonic mean of sensitivity and precision. precision is the fraction of retrieved instances that are relevant. Sensitivity is the fraction of instances that are retrieved.

1.2.4 Results

In this study, the DWT method has been applied for obtaining the coefficients of each sub-bands. Then, features of epochs are extracted based on these coefficients. The data from Physionet dataset which has four channels is divided to epochs with appropriate length. Then, DWT is implemented to these epochs in order to extract the coefficients. After obtaining the vector of features which has seven elements with the results of DWT implementation, the combination of channels with different dependency levels is tested for every epoch with specific length.

We evaluate the performance of our model using 20-fold cross validation with and without neighboring epochs for the first channel(Fpz-Cz). The following tables shows the results of running 20-fold cross validation with two parameters sets:

1. $p = 2$,$s = 2$, $t = 30$ and $c = 1$.

2. $p = 0$,$s = 0$, $t = 30$ and $c = 1$.

The results of implementing 20-fold cross validation for the these parameters sets (Number of forests:25, Number of trees:20) are shown in tables 1.3 through 1.8. In addition, the details of implementing 20-fold cross validation is depicted in table 1.9 for each fold.

1.2.5 Outlier Detection

The results shows that the accuracy of each sleep stages is dependent on the dataset. We have a wide range of accuracy(especially for sleep stage 1)for some datasets. For instance, patients 11 and 17 lead to accuracy of 0.94 and 0.37 for sleep stage 1 (table 1.9). In fact, outliers play an important role in prediction. In order to recognize outliers, random forests algorithm uses proximity matrix. To construct this matrix, the algorithm puts all the data (both training and out of bag(OOB)) down to each tree. If two cases fall in the same terminal node, the related element of proximity matrix increases by one. After all, algorithm normalizes this matrix by dividing by number of trees. In other words, the (i, j) element of proximity matrix is the fraction of trees which the corresponding elements of i and j fall in the same node.

After generating proximity matrix, we can compute the outlier measure. Outliers are cases which their proximities to all other cases are relatively small and should be removed from the data.

Table 1.3: Raw Confusion Matrix for the first parameter set

	stage 1	stage 2	stage 3	stage R	stage W
stage 1	1834	301	36	383	250
stage 2	1502	12925	1227	1555	590
stage 3	14	270	5117	8	294
stage R	1334	850	39	5306	188
stage W	1132	124	112	140	3657

Table 1.4: Normalized Confusion Matrix for the first parameter set

	stage 1	stage 2	stage 3	stage R	stage W
stage 1	0.65	0.11	0.01	0.14	0.09
stage 2	0.08	0.73	0.07	0.09	0.03
stage 3	0.00	0.05	0.90	0.00	0.05
stage R	0.17	0.11	0.01	0.69	0.02
stage W	0.22	0.02	0.02	0.03	0.71

Table 1.5: Evaluation for the first parameter set

	Precision		Sensitivity		F1score		Total accuracy
	mean	min	mean	min	mean	min	
Current	0.69	0.32	0.73	0.65	0.70	0.43	0.74

Table 1.6: Raw Confusion Matrix for the second parameter set

	stage 1	stage 2	stage 3	stage R	stage W
stage 1	1486	307	31	597	383
stage 2	1690	12768	1178	1535	628
stage 3	20	282	5010	12	379
stage R	1826	816	55	4728	292
stage W	1134	214	136	164	3517

Table 1.7: Normalized Confusion Matrix for the second parameter set

	stage 1	stage 2	stage 3	stage R	stage W
stage 1	0.53	0.11	0.01	0.21	0.14
stage 2	0.09	0.72	0.07	0.09	0.04
stage 3	0.00	0.05	0.88	0.00	0.07
stage R	0.24	0.11	0.01	0.61	0.04
stage W	0.22	0.04	0.03	0.03	0.68

Table 1.8: Evaluation for the second parameter set

	Precision		Sensitivity		F1score		Total accuracy
	mean	min	mean	min	mean	min	
Current	0.65	0.24	0.68	0.53	0.65	0.33	0.70

Table 1.9: 20-fold cross validation result for the first parameter sets

	type	S_1	S_2	S_3	S_R	S_W	S_T
1	Sensitivity	0.45	0.85	0.97	0.90	0.61	0.84
2	Sensitivity	0.79	0.88	0.95	0.92	0.32	0.84
3	Sensitivity	0.60	0.82	0.95	0.93	0.62	0.82
4	Sensitivity	0.73	0.75	0.87	0.66	0.65	0.73
5	Sensitivity	0.74	0.71	0.91	0.73	0.65	0.73
6	Sensitivity	0.64	0.89	0.96	0.86	0.48	0.81
7	Sensitivity	0.83	0.89	0.98	0.86	0.49	0.85
8	Sensitivity	0.82	0.74	0.95	0.58	0.93	0.79
9	Sensitivity	0.42	0.50	0.91	0.23	0.73	0.60
10	Sensitivity	0.54	0.78	0.89	0.58	0.78	0.73
11	Sensitivity	0.94	0.60	0.81	0.79	0.46	0.66
12	Sensitivity	0.52	0.20	0.38	0.57	0.82	0.34
13	Sensitivity	0.57	0.78	1.00	0.89	0.33	0.74
14	Sensitivity	0.64	0.59	0.67	0.85	0.62	0.67
15	Sensitivity	0.53	0.87	0.97	0.82	0.94	0.88
16	Sensitivity	0.60	0.76	0.93	0.23	0.49	0.62
17	Sensitivity	0.37	0.71	0.98	0.27	0.80	0.70
18	Sensitivity	0.52	0.55	0.85	0.84	0.86	0.70
19	Sensitivity	0.61	0.84	0.88	0.76	0.87	0.81
20	Sensitivity	0.54	0.71	0.96	0.78	0.47	0.72

The average proximity is defined as the measure from case n in class $j$to the remaining of training data in the same class as [2]:

$$P(n) = \Sigma_{cl(k)=j} prox^2(n,k). \tag{1.14}$$

The raw outlier measure for case n is defined as:

$$O(n) = nsample/P(n). \tag{1.15}$$

This measure will be small if the average proximity is large. Therefore, the outlier measure and similarity of data points have the reverse relations.

In this study, we use oulier detection in random forest to improve the whole data set. We first run the random forest algorithm with activated proximity measure for every recording of our dataset(Each night of sleep for each patient). We test different values for the outlier measure (5,6,..,10). It seems that the algorithm performs better (better accuracy) when we remove the cases with outlier measure greater than six. Therefore, we keep all cases which have the outlier measure less than six in all recordings. Then we repeat the 20-fold cross validation with the new training dataset. Tables 1.10,1.11 and 1.12 show the result for the first parameter sets($p = 2$,$s = 2$, $t = 30$ and $c = 1$). The results of implementing outlier detection for the second parameters sets ($p = 0$,$s = 0$, $t = 30$ and $c = 1$)are reported in tables 1.13, 1.14 and 1.15.

Table 1.10: Raw Confusion Matrix

	stage 1	stage 2	stage 3	stage R	stage W
stage 1	1524	268	26	602	384
stage 2	1933	12742	921	1522	681
stage 3	24	401	4823	9	446
stage R	1917	787	41	4677	295
stage W	1243	211	111	158	3442

Table 1.11: Normalized Confusion Matrix

	stage 1	stage 2	stage 3	stage R	stage W
stage 1	0.54	0.10	0.01	0.21	0.14
stage 2	0.11	0.72	0.05	0.09	0.04
stage 3	0.00	0.07	0.85	0.00	0.08
stage R	0.25	0.10	0.01	0.61	0.04
stage W	0.24	0.04	0.02	0.03	0.67

Table 1.12: Evaluation

	Precision		Sensitivity		F1score		Total accuracy
	mean	min	mean	min	mean	min	
Current	0.65	0.23	0.68	0.54	0.65	0.32	0.69

Table 1.13: Raw Confusion Matrix

	stage 1	stage 2	stage 3	stage R	stage W
stage 1	1918	268	31	352	235
stage 2	1754	12951	975	1516	603
stage 3	15	361	4957	9	361
stage R	1463	853	23	5203	175
stage W	1269	112	86	120	3578

Table 1.14: Normalized Confusion Matrix

	stage 1	stage 2	stage 3	stage R	stage W
stage 1	0.68	0.10	0.01	0.13	0.08
stage 2	0.10	0.73	0.05	0.09	0.03
stage 3	0.00	0.06	0.87	0.00	0.06
stage R	0.19	0.11	0.00	0.67	0.02
stage W	0.25	0.02	0.02	0.02	0.69

Table 1.15: Evaluation

	Precision		Sensitivity		F1score		Total accuracy
	mean	min	mean	min	mean	min	
Current	0.69	0.30	0.73	0.67	0.69	0.42	0.73

1.2.6 Conclusion

In this chapter, we developed a predictive model for automatic sleep stages classification. To the best of our knowledge, this model achieved the best performance for sleep stage 1 (65%) using a single channel of EEG. In addition the number of features for training this model is less than similar studies. Therefore, we can use only one channel to predict sleep stages which is more convenient and applicable for the patient.

Chapter 2

Switching between Sleep Interventions

In this chapter, the goal is to promote the quality of whole night sleep using non-pharmacological interventions such as white noise, soft music and temperature. Smart home technologies can be beneficial for monitoring health changes and for providing interventions to sustain or improve human health. Yang and Hursch [20] argue that the Markov chain model cannot satisfy all modeling requirements of sleep stages. They maintain that if the sleep sojourn times follow an exponential distribution then the continuous time Markov process will be the appropriate model. They also assume non-homogeneity since it is well known that the pattern of sleep changes over the course of the night in normal young adults.

2.1 Problem Description

The non-homogeneous semi-Markov process is a stochastic process where the probability of going to state j after remaining in state i for $Y_n \leq y$ amount of time is

$$P(X_{n+1} = j, Y_{n+1} \leq y | X_n = i, Y_n, X_{n-1}, Y_{n-1}, ..., X_1, Y_1, X_0) = p_{ij} H_{ij}(x), \qquad (2.1)$$

where Y_0 is defined as 0, and it is assumed that the subject begins in the awake state, i.e., $X_0 = 0$. The transition probability from state i to state j at time t is defined as

$$P(X_{n+1} = j | X_n = i, X_{n-1}, ..., X_1, X_0) = p_{ij}, \tag{2.2}$$

and

$$P(Y_{n+1} \leq y | Y_n, Y_{n-1}, ..., Y_1) = H_{ij}(x), \tag{2.3}$$

is the distribution function of the sleep sojourn time at state i before transition to state j. We can estimate the transition probabilities by introducing the following integer-valued stochastic processes:

$N_i(t)$= the total number of transitions out of state i before time t,

$N_{ij}(t)$= the number of transitions from state i to state j before time t,

and the transition probability will be computed as:

$$P_{ij}(t) = N_{ij}(t)/N_i(t), \tag{2.4}$$

In order to estimate the distribution function of the sleep sojourn time, we fit various distributions to the data and finding the one that fits the best to all sojourn times. The results show that log-normal distribution can estimate better the distribution of all sojourn times (lower AIC compared to Gamma, Weibull, Normal).

Since these estimators are for a homogeneous process, Yang and Hursch then empirically separate the night's sleep into disjoint hour-long intervals. In this study, we estimate Transition probability matrix of 24 recordings during the first hour of sleep for each subject (Table 2.1).

Table 2.1: Transition probabilities between sleep stages

States	stage 1	stage 2	stage 3	stage R	stage W
stage 1	0.00	0.72	0.01	0.00	0.27
stage 2	0.18	0.00	0.69	0.05	0.08
stage 3	0.08	0.90	0.00	0	0.03
stage R	0.34	0.33	0.00	0.00	0.33
stage W	0.94	0.03	0.03	0.00	0.00

We simulate the process to obtain the average amount of time spent in each state (Table 2.2).

Table 2.2: Average percentage of time spent in each state

Chain	stage 1	stage 2	stage 3	stage R	stage W
Normal	0.14	0.50	0.28	0.01	0.07

2.2 Switching Between Two Chains With Simulation

Environment, daily experience, diet, etc. can affect the sleep pattern. These changes in subject's environment (temperature, soft music,...) are called interventions which produce another chain with different transition probabilities between sleep stages.

The objective is to match the sojourn time percentages of normal sleep every night by adjusting between interventions. Each intervention may result in a different sleep experience. We consider two separate chains with different transition probability matrices, corresponding to two interventions (Figure 2.1).

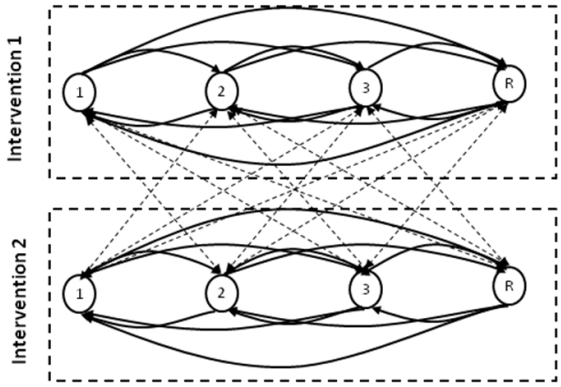

Figure 2.1: Average percentage of time spent in each state

The algorithm of switching between these two chains is described as the algorithm 1 and has been coded in *R*.

Algorithm 1 Switching Between Two Chains

Start from state 1 in chain 1
Determine which state the process will transition to next
Spend random lognormal amount of time accordingly
Make a decision whether to remain in chain 1 or switch to chain 2
Compute the expected amount of time in states of each chain by simulation
Compute the distance of each chain from the normal chain, namely d1 and d2
if $d_1 \geq d_2$ **then**
    Go to chain 2
    else
    Remain in chain 1
end if
until the time horizon fulfilled
goto *top.*

Figure 2.2 shows the schematic process of this algorithm for two chains with two states.

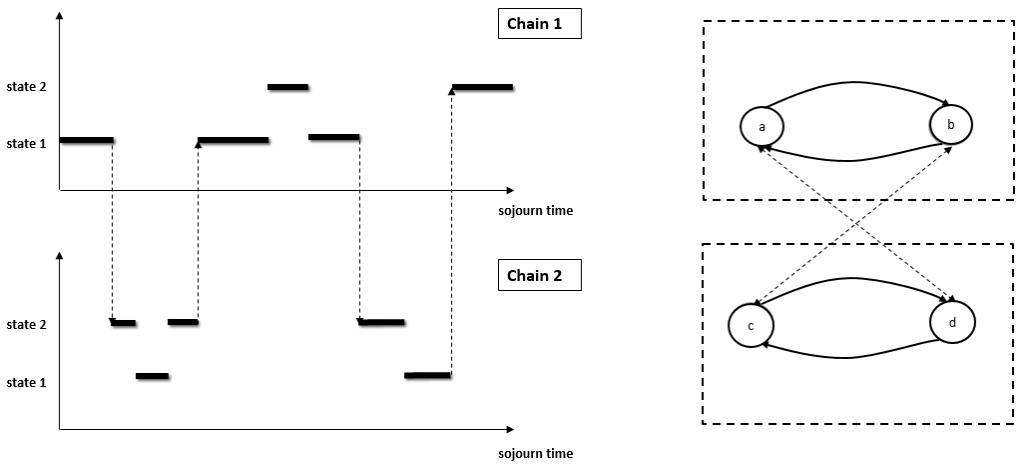

Figure 2.2: Switching between two chains with two states

Suppose we have two chains with different transition probabilities between states. After running simulation model, the average amount of sojourn time is obtained. For making improvement of switching between these two chains, we run the algorithm and the results are shown in table 2.3.

Table 2.3: Result of switching between two chains with five states

Chains	stage 1	stage 2	stage 3	stage R	stage W
Normal	0.14	0.50	0.28	0.01	0.07
Chain 1	0.19	0.47	0.28	0.01	0.05
Chain 2	0.12	0.56	0.27	0.01	0.04
Resulting chain	0.16	0.51	0.27	0.01	0.05

The average number of switches between two chains in 100 iterations is three with the following specifications:

- distance(chain1, Normal)= 0.004
- distance(chain2, Normal)= 0.005
- distance(resulting chain, Normal)= 0.001

As the result shows, we can optimize the sleep pattern by switching between these two different chains.

2.3 Conclusion

In this chapter, we developed a simulation model for making decision of sleep MDP to minimize the difference of patient's sleep pattern with a normal one by switching between interventions. Therefore, we can maximize the sleep efficiency by switching between intervention chains.

Bibliography

[1] Varun Bajaj and Ram Bilas Pachori. Automatic classification of sleep stages based on the time-frequency image of eeg signals. *Computer methods and programs in biomedicine*, 112(3):320–328, 2013. 2

[2] Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001. 9, 16

[3] LG Doroshenkov, VA Konyshev, and SV Selishchev. Classification of human sleep stages based on eeg processing using hidden markov models. *Biomedical Engineering*, 41(1):25–28, 2007. 2

[4] Farideh Ebrahimi, Mohammad Mikaeili, Edson Estrada, and Homer Nazeran. Automatic sleep stage classification based on eeg signals by using neural networks and wavelet packet coefficients. In *Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE*, pages 1151–1154. IEEE, 2008. 2

[5] Luay Fraiwan, Khaldon Lweesy, Natheer Khasawneh, Heinrich Wenz, and Hartmut Dickhaus. Automated sleep stage identification system based on time–frequency analysis of a single eeg channel and random forest classifier. *Computer methods and programs in biomedicine*, 108(1):10–19, 2012. 2

[6] S Güneş, K Polat, Ş Yosunkaya, and M Dursun. A novel data pre-processing method on automatic determining of sleep stages: K-means clustering based feature weighting. 2009. 2

[7] Salih Güneş, Kemal Polat, and Şebnem Yosunkaya. Efficient sleep stage recognition system based on eeg signal using k-means clustering based feature weighting. *Expert Systems with Applications*, 37(12):7922–7928, 2010. 2

[8] Vladimir Hachinski, Costantino Iadecola, Ron C Petersen, Monique M Breteler, David L Nyenhuis, Sandra E Black, William J Powers, Charles DeCarli, Jose G Merino, Raj N Kalaria, et al. National institute of neurological disorders and stroke–canadian stroke network vascular cognitive impairment harmonization standards. *Stroke*, 37(9):2220–2241, 2006. 9

[9] Yu-Liang Hsu, Ya-Ting Yang, Jeen-Shing Wang, and Chung-Yao Hsu. Automatic sleep stage recurrent neural classifier using energy features of eeg signals. *Neurocomputing*, 104:105–114, 2013. 2

[10] Han G Jo, Jin Y Park, Chung K Lee, Suk K An, and Sun K Yoo. Genetic fuzzy classifier for sleep stage identification. *Computers in Biology and Medicine*, 40(7):629–634, 2010. 2

[11] B Koley and D Dey. An ensemble system for automatic sleep stage classification using single channel eeg signal. *Computers in biology and medicine*, 42(12):1186–1195, 2012. 2

[12] Sheng-Fu Liang, Chin-En Kuo, Yu-Han Hu, Yu-Hsiang Pan, and Yung-Hung Wang. Automatic stage scoring of single-channel sleep eeg by using multiscale entropy and autoregressive models. *Instrumentation and Measurement, IEEE Transactions on*, 61(6):1649–1657, 2012. 2

[13] Edgar Oropesa, Hans L Cycon, and Marc Jobert. Sleep stage classification using wavelet transform and neural network. *International computer science institute*, 1999. 2

[14] Seral Özşen. Classification of sleep stages using class-dependent sequential feature selection and artificial neural network. *Neural Computing and Applications*, 23(5):1239–1250, 2013. 2

[15] Allan Rechtschaffen and Anthony Kales. A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects. 1968. 1

[16] Saeid Sanei and Jonathon A Chambers. *EEG signal processing*. John Wiley & Sons, 2013. 7

[17] Baha Şen, Musa Peker, Abdullah Çavuşoğlu, and Fatih V Çelebi. A comparative study on classification of sleep stage based on eeg signals using feature selection and classification algorithms. *Journal of medical systems*, 38(3):1–21, 2014. 9

[18] Rakesh Kumar Sinha. Artificial neural network and wavelet based automated detection of sleep spindles, rem sleep and wake states. *Journal of medical systems*, 32(4):291–299, 2008. 2

[19] M Emin Tagluk, Necmettin Sezgin, and Mehmet Akin. Estimation of sleep stages by an artificial neural network employing eeg, emg and eog. *Journal of medical systems*, 34(4):717–725, 2010. 2

[20] Mark CK Yang and Carolyn J Hursch. The use of a semi-markov model for describing sleep patterns. *Biometrics*, pages 667–676, 1973. 1, 19

[21] Lukáš Zoubek, Sylvie Charbonnier, Suzanne Lesecq, Alain Buguet, and Florian Chapotot. Feature selection for sleep/wake stages classification using data driven methods. *Biomedical Signal Processing and Control*, 2(3):171–179, 2007. 2

Vita

Maryam Zokaeinikoo was born in Qazvin, Iran, on May 26, 1987. She got the diploma in mathematics at Farzanegan high school which is, in fact, affiliated with the National Organization for Development of Exceptional Talents (NODET). After finishing high-school,she succeeded to pass the national entrance exam for undergraduate level and ranked among top 1% participants. She completed undergraduate studies in the field of industrial engineering at Khaje Nasir University of Technology, Tehran, in September 2009. She ranked 16^{th} in Iran national entrance exam for master degree which opened the path for her to continue her studies in Iran number one university, Sharif University of Technology. She received her master degree in industrial engineering in September 2012. Maryam accepted a graduate research assistantship position at The University of Tennessee, Knoxville, in the industrial and systems engineering program in April 2014. She is going to continue her education for PhD level in industrial and manufacturing engineering at the Penn State University, PA.