

SleepTight: Identifying Sleep Arousals Using Inter and Intra-Relation of Multimodal Signals

Tanuka Bhattacharjee¹, Deepan Das¹, Shah Nawaz Alam¹, Achuth Rao M V², Prasanta Kumar Ghosh², Ayush Ranjan Lohani³, Rohan Banerjee¹, Anirban Dutta Choudhury¹, Arpan Pal¹

¹ Tata Consultancy Services Ltd, Kolkata, India

² Indian Institute of Science, Bangalore, India

³ Indian Institute of Engineering Science and Technology, Shibpur, India

Abstract

Sleep arousal directly affects the quality of sleep. PhysioNet Challenge 2018 aims to correctly identify designated target arousal (non-apnea arousal) and non-arousal regions from simultaneously recorded multiple biomedical signals. Our contribution lies in a feature extraction algorithm that extracts generic and domain-specific features from different biomedical signals available in the challenge provided dataset to form a composite feature vector. 50 most significant features are selected based on Minimum Redundancy Maximum Relevance scores for final classification using multiple unbiased Random Forests. The approach is designed to produce a single label for a 20-second segment containing all channels, followed by smoothing the label time-series per subject. Our algorithm yields the median Area Under Precision-Recall Curve (AUPRC) as 0.29 on 5-fold cross-validation on the training dataset. The same value of AUPRC is maintained for the test dataset as well, thereby emphasizing the stability of the proposed algorithm. This method secured the global rank of 8 during the official phase of the challenge.

1. Introduction

Sleep plays an essential role for a healthy life. Improper and inadequate sleep may lead to a wide range of possible physical, mental and psychological disorders. Sleep occurs in five basic stages including wakefulness, non-rapid eye movement (non-REM) (stage 1, stage 2, stage 3) and rapid eye movement (REM). An arousal is a type of sleep disorder which causes a shift from deep sleep (REM) to light sleep due to some abrupt changes in the pattern of brain wave activity. Arousals can be of different types, namely spontaneous arousals, Respiratory Effort Related Arousals (RERA), bruxisms, hypoventilations, hypopneas, apneas (central, obstructive and mixed),

vocalizations, snores, Periodic Leg Movements (PLM), Cheyne-Stokes breathing or partial airway obstructions.

Automatic detection of sleep arousals is an important area of research as manual annotation of hours of various physiological signals is a tedious and time consuming job. An arousal makes some abrupt changes in brain activities that can be captured in the Electroencephalography (EEG) signals. Certain arousals also cause instantaneous variations in regular heart beat pattern of the subject. Arousals caused due to breathing problem during sleep can also be detected from the rapid fall in the blood oxygen saturation level. However, a very accurate sleep arousal detection system is an unsolved area of research till date due to the following major reasons.

1. Lack of open access dataset,
2. Difficulty in processing multiple noisy data, and
3. Insufficient number of well-established features for the detection of arousals.

PhysioNet Challenge 2018 encourages the participants to develop an algorithm to classify non-RERA-non-apnea and RERA types of arousals using a variety of physiological signals recorded from the subjects as they slept through the night. Our algorithm performs an extensive analysis on 13 different physiological signals provided by the challenge organizer, to propose a novel feature set containing discriminating markers for the two types of target arousal labels. A set of best 50 features are selected based on the Minimum Redundancy Maximum Relevance (mRMR) scores, upon which the classification task is performed using multiple Random Forest (RF) classifiers.

The rest of the paper is organized in the following manner. Section 2 provides a very brief description of the experimental dataset. A detailed description of our methodology, including our feature set, is given in Section 3. Section 4 shows our results on the training and test dataset. Section 5 lists the concluding remarks.

2. Data

A detail description of the experimental dataset provided by the PhysioNet organizer can be found in [1]. The dataset consists of the Polysomnography recordings of a total of 1,985 subjects, a part of which is shared for developing the algorithm. The dataset contains a total of 13 time-series from 6 different physiological signals including 6-channel Electroencephalography (EEG) (F3-M2, F4-M1, C3-M2, C4-M1, O1-M2, O2-M1), single channel Electrooculography (EOG) (from left eye with right ear EEG, M2 as reference), 3-channel Electromyography (EMG) (Chin, Chest and Abdomen), single-lead Electrocardiology (ECG), Respiratory Airflow and Oxygen Saturation (SaO2). Excluding SaO2, all other signals are sampled to 200 Hz and are measured in microvolts. SaO2 is measured as a percentage. For analytic convenience, SaO2 is resampled to 200 Hz.

3. Methodology

This section describes in detail the steps of the proposed methodology of non-apnea sleep arousal detection.

3.1. Data Preparation

For every subject, signals from every channel is split into segments of 20 second duration, with 10% overlap between the consecutive segments. In case of training data, the corresponding annotation time-series is replaced by a single label per 20-second segment. If the entire segment is annotated as non-arousal (nA), it is replaced by a single nA label. On the other hand, if at least 80% duration of the segment is annotated as target arousal (tA), it is replaced by a single tA label. Ratio of the number of instances in classes nA to tA is around 16:1.

We have also tried labeling segments, with non-apnea arousal throughout, as class tA' . But, in that approach, ratio of the number of instances in classes nA to tA' turns out to be greater than 20:1. Thus, our approach slightly improves the class imbalance in favor of the minority class.

3.2. Preprocessing

All channels are filtered appropriately, according to the requirements of the individual sensors. ECG signal is found to contaminate EEG, EOG and Chin EMG signals as additive noise. EOG also contaminates the frontal EEG channels similarly. To deal with this, adaptive Recursive Least Squares (RLS) filtering technique [2] is used to first remove ECG from EOG and Chin EMG, then remove both ECG and decontaminated EOG from EEG. Figure 1 shows a representative outcome of preprocessing on EEG.

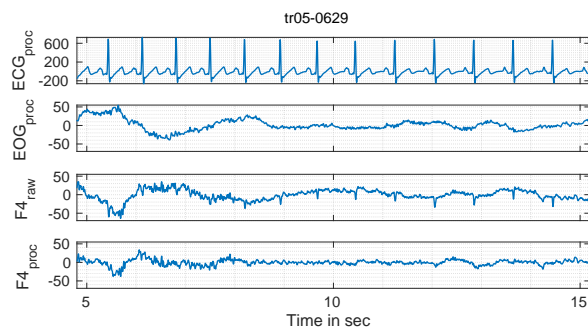


Figure 1. EEG waveform recorded from F4-M1, before ($F4_{raw}$) and after ($F4_{proc}$) adaptive RLS filtering is employed to remove ECG and EOG contamination

3.3. Feature Extraction

Discovering features capable of distinguishing the concerned classes is the most important part of any conventional approach of handling classification problems. In the proposed method, all 13 physiological signals of the given polysomnographic recordings are used to extract features which may correlate to the target-arousal/non-arousal labels. The considered features can be classified broadly into two categories, as follows.

3.3.1. Domain/Sensor Independent Generic Features

A set of 392 generic features is derived from each signal resulting in a total of ($13 \times 392 =$) 5096 features. These features include various statistical properties (e.g. mean, variance, skewness, kurtosis, Box-Pierce statistics, Hurst exponent) of the time domain signal, its Fast Fourier Transform (FFT) coefficients and Discrete Wavelet Transform (DWT) coefficients. Other spectral features like centroid, roll off, flux and different entropy measures (e.g. Shannon, Renyi, Tsallis) are also considered.

3.3.2. Domain/Sensor Specific Features

A total of 1044 features spread across the 13 signals are computed. These comprise asymmetry features from 5 standard frequency bands (e.g. delta, theta, alpha, beta and gamma) of the 6 EEG channels, morphological and heart rate variability features from ECG [3], visibility graph based features from ECG, EEG and Cardio-Respiratory Interaction (CRI) time-series [4, 5], breathing rate variability features from Chest, Abdomen EMG and Airflow [6]. Morphologies of the breathing cycles, breath-by-breath correlation, phase relation between Chest and Abdomen EMG are also explored [7, 8].

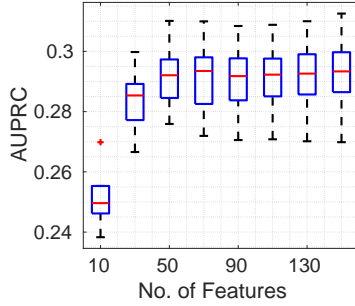


Figure 2. 5-fold performance (AUPRC) evaluation for an increasing feature set, ranked using mRMR technique, over all training subjects

3.4. Feature Selection

Feature selection helps in discarding the irrelevant and redundant features, thereby lowering both bias and variance of the proposed algorithm. All the extracted features are ranked using the mRMR technique [9] to find the most effective one. Figure 2 shows the variation of AUPRC over 5-folds of cross-validation with the increase in the number of selected features. The cross-validation partition is done in such a manner that recordings from the same subject never belong to both train and test folds.

It can be seen from Figure 2 that the best bias-variance value pair is obtained while using the top 50 features. With this many number of features, any test data would need approximately 5×10^{12} CPU instructions at most for complete execution, which is well within the prescribed quota of the challenge.

Figure 3 shows the strengths of the top 50 features selected. Among these, 34% features (including the best one) come from the domain/sensor specific category. The distribution of the selected features over the different physiological signals is depicted in Table 1. It shows that Chin EMG, Abdomen EMG, Airflow and SaO2 are the most important indicators of non-apnea arousal.

3.5. Fusion of Multiple Trained RF Models

We choose RF models as our classifiers. However, training a single classifier model using the unbalanced

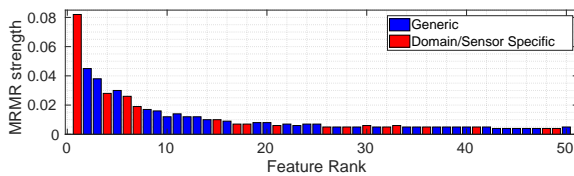


Figure 3. Strength and distribution of the selected 50 features

Table 1. Distribution of top 50 features across sensors and across generic and domain/sensor specific varieties

Physiological Signal	Generic Features Selected	Sensor-specific Features Selected	Total	
EEG	F3-M2	2	-	2
	F4-M1	1	1	2
	C3-M2	1	-	1
	C4-M1	2	-	2
	O1-M2	1	-	1
	O2-M1	-	1	1
EOG	3	-	3	
EMG	Chin	4	2	6
	Abdomen	5	3	8
	Chest	2	2	4
Airflow	6	-	6	
SaO2	6	1	7	
ECG	-	5	5	
Multi-signal	-	2	2	
Total	33	17	50	

data leads to most tested instances being classified into the majority class, nA . On the other hand, dropping major class instances in order to train the model on balanced data produces a more balanced outcome. But this means that only $1/16^{th}$ of the major class is used in training. To have the best of both worlds, we decide to take an approach using multiple trained models. 10 RF classifiers are trained. All the minority class in training, i.e. instances labeled tA , are fed to every classifier. The majority class, nA , is balanced per classifier and is mutually exclusive with the nA instances used to train any other classifier. Thus, both a balanced training and a greater coverage of majority class are ensured. While labeling a test instance, the mean of the output probabilities of belonging to target class tA , of all 10 models, is considered as the final posterior probability. The advantage of using multiple fair models,

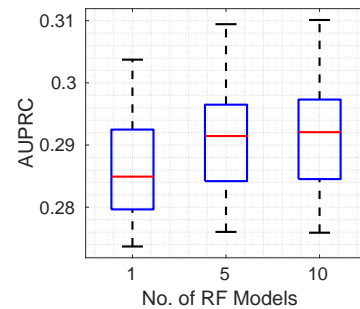


Figure 4. Variation in 5-fold AUPRC when different number of fair trained models are used and their mean probability (of belonging to target class tA) taken as output for a tested instance

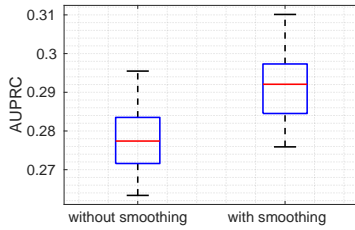


Figure 5. Variation in 5-fold AUPRC with and without smoothing of the predicted label time-series

over a single such model, is demonstrated in Figure 4.

3.6. Prediction Stitching and Smoothing

Using the above method, each 20-second test instance receives a single output label. The same is expanded to a label time-series at 200 Hz, at par with the annotation time-series provided. The 10% overlap between consecutive signal segments is handled and is expected to avoid sharp transitions between segments. The label time-series is further smoothed using a 5-second moving average filter. Performance gain from smoothing is shown in Figure 5.

4. Results

The performance of the proposed approach during 5-fold cross-validation over the training dataset provided by the challenge organizer is detailed in Table 2. It should be noted that, the gross AUPRC for all tested subjects first peaks for small number of subjects and then with increase in the number of subjects the AUPRC comes down and asymptotically reaches the final value. For example, in case of the first fold, the AUPRC peaks to 0.45 for around 20 subjects, comes down to 0.30 for 100 subjects and asymptotically approaches 0.29 as subjects increase. The AUPRC obtained on the hidden test dataset is **0.29**, which is identical to the median AUPRC of train set cross-validation. Since there is no drop in performance from train to test dataset, it can be said that the proposed algorithm is robust for the concerned arousal detection problem.

5. Conclusion

In this paper we propose an algorithm of detection of arousals from various physiological data collected during

Table 2. AUPRC of 5-folds in training set cross-validation

	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5
AUPRC	0.29	0.29	0.28	0.31	0.29

sleep. Several statistical and morphological features have been considered to derive a novel feature set for designing a robust sleep arousal classification system. The said algorithm yields median AUPRC of 0.29 on 5-fold cross-validation on the training dataset shared during the PhysioNet Challenge 2018. The same performance of 0.29 is maintained in case of the hidden test dataset. Our future work would involve proposing a hybrid classification system, by adding machine generated features (deep learning based approach) to the existing feature set. The source code is available in public under GPL license.

References

- [1] Ghassemi MM, Moody BE, Lehman LH, Song C, Li Q, Sun H, Mark RG, Westover MB, Clifford GD. You snooze, you win: the PhysioNet/Computing in Cardiology Challenge 2018. In *Computing in Cardiology*, volume 45. Maastricht, Netherlands, 2018; 1–4.
- [2] Lu G, Brittain JS, Holland P, Yianni J, Green AL, Stein JF, Aziz TZ, Wang S. Removing ECG noise from surface EMG signals using adaptive filtering. *Neuroscience Letters* 2009; 462(1):14–19.
- [3] Datta S, Puri C, Mukherjee A, Banerjee R, Choudhury AD, Singh R, Ukil A, Bandyopadhyay S, Pal A, Khandelwal S. Identifying Normal, AF and other abnormal ECG rhythms using a cascaded binary classifier. *Computing* 2017;44:1.
- [4] Long X. On the analysis and classification of sleep stages from cardiorespiratory activity. *Sleep Wake* 2015;1–232.
- [5] Diykh M, Li Y. Complex networks approach for EEG signal sleep stages classification. *Expert Systems with Applications* 2016;63:241–248.
- [6] Prabha A, Trivedi A, Kumar AA, Kumar CS. Automated system for obstructive sleep apnea detection using heart rate variability and respiratory rate variability. In *Advances in Computing, Communications and Informatics (ICACCI)*, 2017 International Conference on. IEEE, 2017; 1303–1307.
- [7] Redmond SJ, Heneghan C. Cardiorespiratory-based sleep staging in subjects with obstructive sleep apnea. *IEEE Transactions on Biomedical Engineering* 2006; 53(3):485–496.
- [8] Al-Angari HM, Sahakian AV. Automated recognition of obstructive sleep apnea syndrome using support vector machine classifier. *IEEE Transactions on Information Technology in Biomedicine* 2012;16(3):463–468.
- [9] Ding C, Peng H. Minimum redundancy feature selection from microarray gene expression data. *Journal of Bioinformatics and Computational Biology* 2005; 3(02):185–205.

Address for correspondence:

Tanuka Bhattacharjee

Embedded Systems and Robotics, Tata Consultancy Services

Building 1B, Ecospace Plot - IIF/12, New Town, Rajarhat

Kolkata - 700160, India

bhattacharjee.tanuka@tcs.com