

# Detection of Respiratory Effort-Related Arousals Using a Hidden Markov Model and Random Decision Forest

János Szalma, András Bánhalmi, Vilmos Bilicki

SZTE Department of Software Engineering, Szeged, Hungary

## Abstract

*The efficient detection of respiratory effort-related arousals requires enormous amount of data and a suitable learning model. Using a dataset taken from PhysioNet.org, windows of 20 seconds were extracted with their median aligned with the starting point of the arousals. The same amount of data was selected from non-arousal regions.*

*Features derived using these windows were reduced to 38 by using various feature selection methods. A cross-validated Random Forest(RF) was used for the evaluation.*

*The training data was processed with a 20-second sliding window and a 1 second resolution. Windows were labelled according to their temporal location in the data. This was used to train three separate RFs on different parts of the data, which provided a probability emission model. The probability values used in a Hidden Markov Model and established the most probable path with the Viterbi algorithm. Probability values were aggregated based on the Viterbi path, then smoothed and resampled to match the original sample rate. This method achieved an 0.29 score of AUPRC.*

## 1. Introduction

Everyone knows that healthy sleep is paramount to our well-being, but several things can interfere with one's ability to fall or stay asleep. Some of the more well-studied and important sleep disorders are breathing-related. These are described by abnormal breathing and have an enormous influence on sleep quality [1][2].

According to the current classification, there are four major types of sleep-related breathing disorders; these being central apnea syndromes, obstructive apnea syndromes, hypoventilation/ hypoxemia syndromes

associated with sleep and undefined/non-specific sleep disorders [2].

Respiratory effort-related arousals (RERAs) are also responsible for abnormal sleep. It is a breathing disorder characterized by an obstructive upper airway airflow reduction (which does not meet the criteria of apnea or hypopnea), associated with an increased respiratory effort that resolves itself with the appearance of arousals. Diagnostic criteria are [3]:

1. A series of respiratory cycles of increasing/ decreasing effort or flattening, recorded by nasal manometry and leading to an arousal that cannot be defined as apnea or hypopnea.
2. Duration  $\geq 10$  sec.

“Arousals are a main part of breathing-related sleep issues and they are defined as a sudden change of EEG frequency consisting of alpha and theta activity or waveforms with frequency greater than 16 Hz (but not sleep spindles) and a duration of 3-15 seconds [3].”

Sleep arousal detection was carried out by researchers [4][5], with a relatively small patient size and without distinguishing between apneas/hypopneas and RERAs.

The detection of arousals is a cornerstone of sleep study evaluations. As arousals are mostly caused by sleep-related breathing disorders, correct classifications could lead to crucial diagnoses. Arousals also have detrimental effects on human sleep cycles. Assessing the correct Respiratory Disturbance Index would mean the accurate detection of the number of RERAs, apnea, hypopneas. This index is an excellent indicator of sleep quality and also a great tool for medical personnel to do accurate evaluations on the polysomnographic data of patients [3].

## 2. Materials and methods

### 2.1 Dataset

The dataset were contributed by the Massachusetts General Hospital's (MGH) Computational Clinical Neurophysiology Laboratory (CCNL), and the Clinical Data Animation Laboratory (CDAC). The dataset includes 1,985 subjects who were monitored at an MGH sleep laboratory for the diagnosis of sleep disorders. The data was divided into balanced training (n = 994), and test sets (n = 989) [7].

This included signals like oxygen saturation, chin/chest/abdomen movement, airflow, ECG, EOG and EEG [7].

More information on the data used by this study can be found on the challenge website [7].

## 2.2 Data processing

The first step of the processing was to look at all the signal channels and determine any factor that could distort signals in some way. In our study, Matlab was used as a programming tool.

Looking at a spectrogram using a short-time Fourier transform, noise from certain frequencies were identified. The most important was the 60Hz utility frequency of the alternating mains current. A second order digital IIR notch filter was used to filter this out. Every channel excluding oxygen saturation contained the 60Hz noise. During the final data processing all signals were resampled at 60Hz, which effectively meant a low-pass 30Hz filter.

Sensors dropping or giving false measurements were also a factor that was considered. Parts where all the signals were zero and where strangely low or high spikes were present, but relative to the size of the data set these could be disregarded.

The size of the annotated training data was 135 gigabytes, so in order to be able to work with it, segmentation was necessary. 20-second-long intervals were chosen in such a way that their center point was aligned with the start of an arousal. This was motivated by other studies on related topics and the fact that arousals are described by a sudden change in certain physiological signals. These 20-second-long windows reflected that change. For a given subject's recording the same amount of windows were determined from non-RERA, non-arousal regions. This process was done with every fifth subject. The segments were assembled, hence we got a smaller data set containing the same amount of RERA and non-RERA data. This was suitable for determining valuable features.

The ratio of the 10 second part before a RERA, and the 10 second segment at the beginning of a RERA was enough to show what changes and how, when these arousals occur. Also it was helpful to look at not only the change, but the standalone features as well.

## 2.3 Feature extraction

Using all features that looked even slightly promising we ended up with 223 overall. A number of attribute selection methods were used to reduce their number. A WEKA[8] implementation of these methods was used with default settings. The performance of each individual feature selection was evaluated using a Random Decision Forest.

Table 1. Feature selection results

Attribute selection	Number of attributes	ROC-AUC
none	223	0,812
CFS	22	0,806
Correlation	30	0,762
Gain Ratio	31	0,799
Information Gain	31	0,802
One Rule	31	0,801
Union of all	68	0,814
Reduced union of all	38	0,810

The combination of these evaluation methods reduced the number of attributes to 38. This contained at least one feature for every data channel available.

Table 2. Feature set

Standalone features	Ratio of two consecutive 10 second segments	Ratio of a 3 second segment and a preceding 10 second segment
<b>Breathing-related:</b> -Mean of the absolute value of chest and abdomen. -Standard deviation of chest and abdomen. -Max-min of abdomen. -Movement of abdomen and chest.	<b>Breathing-related:</b> -Mean of the absolute value of chest, chin and airflow. -Standard deviation of abdomen. -Max-min of abdomen. -Convolution of airflow and chest/abdomen movement	<b>Breathing-related:</b> <i>none</i>

<b>Cardiovascular:</b> <i>none</i>	<b>Cardiovascular</b> : -Mean of Sao2. -Heart rate.	<b>Cardiovascular:</b> <i>none</i>
<b>Neural:</b> -Standard deviation of occipital EEG and frontal EEG. -EEG Delta band: central Hjorth mobility and energy, occipital Hjorth activity and complexity. -EEG Theta band: central and occipital energy. -EEG Alpha band: occipital energy and Hjorth mobility. -EEG Beta band: central energy and Hjorth mobility, occipital Hjorth activity	<b>Neural:</b> <i>none</i>	<b>Neural:</b> -EEG Delta band: occipital energy. -EEG Theta band: occipital energy of both (O2,O1) signals.

Using these features the training data was processed with a 20-second sliding window and a 1-second resolution. The data was labelled according to the following rules:

Table 3. Data labelling

No arousal	Beginning of RERA (first 2 seconds)	Middle of RERA	Ending of RERA (last 2 seconds)	Apnea/hypopnea arousal
1	2	3	4	-1

In the training process, segments marked with -1 were not included. Random Decision Forests (RFs) are an ensemble learning method for classification and regression that operate by constructing a multitude of decision trees during training and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees. RFs forests correct for the tendency of decision trees overfitting the training set. Significant improvements in classification accuracy resulted from growing an ensemble of trees and using their individual votes to find the most likely class. In order to grow these ensembles, often random vectors

are generated that control the growth of each individual tree. An example is bagging, where to grow each tree a random selection is made from the samples in the training set [9][10].

Three separate binary RFs were trained on different parts of the data. The first was trained on labels 1 and 2, the second on 1 and 3 and the last RF was trained using data annotated by labels 1 and 4. Because of the highly unbalanced number of labels, RFs tend to under predict classes that appeared less during the training process. This could be solved by setting a uniform prior distribution or by resampling data used for training. Undersampling data from label 1 produced a model that proved to be the best and least complex.

## 2.4 Sequence analysis

A Markov process is a process that is capable of being in more than one state and it can make transitions among these states. In a Markov process the available states and transition probabilities depend only upon what state the system is currently in. A Markov Chain is a statistical model of a system that moves sequentially from one state to another. The transition probabilities from one state to another are dependent only on the current state (not on previous states). An alternative representation that is sometimes used for Markov chains does not rely on a start or end state, instead it represents the distribution over initial states and only accepts states explicitly. Extending this idea to a new model which is a doubly embedded stochastic process with an underlying process that is not observable (hidden), it is only observable via another set of stochastic processes that produce a set of observations. It is called a Hidden Markov Model (HMM).

The transition probabilities were assigned by looking at all the unique state (label) changes in the dataset and dividing them by the total state changes of that label [11].

Using the probability scores of the RF and the state transition probability matrix the most likely path was calculated by the Viterbi algorithm(VA). VA is a recursive optimal solution to the problem of decoding the state sequence of a discrete-time finite-state Markov process [12]. After finding the most likely path of state sequences; the probability of a RERA event at a certain time frame was calculated by adding up the probability values of states 2,3,4 for areas where the most likely state was either 2,3,4. For segments that were marked by VA with 1 (no arousal), the probability state 3 was assigned.

This is reasonable as all states except 1 represent some part of the arousal. This proved to be a better method than a trivial binary classification and Viterbi

decoding. The probability values were smoothed with a moving average, which is a low-pass filter with a window length set to 24 seconds. This eliminated some noise in the time series prediction. Then the results were resampled with linear interpolation to match the original (200Hz) sample rate.

### 3. Results

Our final algorithm was only graded for its binary classification performance by challenge organizers on target arousal and non-arousal regions, measured by the area under the precision-recall curve (AUPRC). This is the gross AUPRC (i.e., for each possible value of the precision and recall are calculated for the entire test database), which is not the same as averaging the AUPRC for each record. The running time was 10% of the given quota.

The AUPRC for our final model was 0.29.

### 4. Conclusions

The vast amount of the training data set provided a perfect opportunity to find a large array of predictive features, compare the performance of machine learning methods and discover new connections between breathing-related sleep disorders and partially hidden physiological attributes.

During the study, due to time and hardware restrictions only 20-35% of the total information was processed.

Difficulties arose because of these limitation, the imperfectly labelled data and the interrater reliability for sleep scoring, which is 82% for the AASM standard used by this study [13], and this tells us that there still is no absolute agreement for sleep scoring standards.

RF proved to be helpful in showing how good a given feature set is and how reliable a learning model is.

However, for sequential data like a polysomnogram it does not have the means to enforce consecutive data to follow certain rules.

The Hidden Markov Model provided these rules and the Viterbi algorithm found the most likely sequence called the Viterbi path. Using a special aggregation on this path, better results were obtained than the default ones. Together these methods were capable of detecting RERAs and other arousals with an acceptable accuracy and could be beneficial in many similar tasks.

## 5. Acknowledgements

This project (EFOP-3.6.1-16-2016-00008) was supported by the European Union and co-financed by the European Regional Development Fund.

## References

- [1] Oswald, I. (1980). Sleep as a Restorative Process: Human Clues. *Prog Brain Res*, 53, 279-288.
- [2] Sleep – Related Breathing Disorders in Adults. AASM Task Force. *SLEEP*. 1999; 22:667–689
- [3] Sateia, M. J. (2014). International classification of sleep disorders. *Chest*, 146(5), 1387-1394.
- [4] American Academy of Sleep Medicine, R.B Berry, et al. *The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications*. American Academy of Sleep Medicine, 2015.
- [5] Mahalaxmi, U. S. B. K., & Patnaik, M. R. (2017). EEG Arousal Detection using SVM and EMD based Frequency Detection Method. *International Journal of Applied Engineering Research*, 12(18), 7663-7674.
- [6] Sugi, T., Kawana, F., & Nakamura, M. (2008). Automatic EEG Arousals Detection for Obstructive Sleep Apnea Syndrome. *IFAC Proceedings Volumes*, 41(2), 5209-5214.
- [7] Mohammad M Ghassemi, Benjamin E Moody, Li-wei H Lehman, Christopher Song, Qiao Li, Haoqi Sun, Roger G. Mark, M Brandon Westover, Gari D Clifford, You Snooze, You Win: the PhysioNet/Computing in Cardiology Challenge 2018, *Computing in Cardiology Volume 45*. Maastricht, Netherlands, 2018. pp 1-4
- [8] Eibe Frank, Mark A. Hall, and Ian H. Witten (2016). *The WEKA Workbench*. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques", Morgan Kaufmann, Fourth Edition, 2016.
- [9] Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5-32.
- [10] Cutler, A., Cutler, D. R., & Stevens, J. R. (2012). Random Forests. In *Ensemble Machine Learning* (pp. 157-175). Springer, Boston, MA.
- [11] Rabiner, L., & Juang, B. (1986). An Introduction to Hidden Markov Models. *IEEE ASSP Magazine*, 3(1), 4-16.
- [12] Forney, G. D. (1973). The Viterbi Algorithm. *Proceedings of the IEEE*, 61(3), 268-278.
- [13] Danker-hopfe, H., Anderer, P., Zeitlhofer, J., Boeck, M., Dorn, H., Gruber, G., ... & Saletu, B. (2009). Interrater Reliability for Sleep Scoring According to the Rechtschaffen & Kales and the new AASM Standard. *Journal of Sleep Research*, 18(1), 74-84.

Address for correspondence:

János Szalma  
University of Szeged, Department of Software Engineering,  
Hungary, 6720 Szeged, Dugonics tér 13.  
sz.janni@gmail.com

