Automatic Detection of Respiratory Effort Related Arousals From Polysomnographic Recordings

Ivan M Lazić, Nikša M Jakovljević, Danica M Despotović, Tatjana G Lončar-Turukalo

University of Novi Sad, Faculty of Technical Sciences, Novi Sad, Serbia

Abstract

A reliable automatic categorization of respiratory effort is paramount for sleep-disordered breathing characterization from polysomnography. A respiratory effort related arousal (RERA) is a subtle breathing obstruction associated with an arousal. For identification of RERAs we focused on: chest and abdomen EMGs, airflow, and EEG; monitoring changes in ECG and SaO2. The quality of signals was assessed to overcome sensor associated problems and sporadic individual signal losses. We evaluated an ensemble learning and a deep learning approach using the engineered feature set trained on the 994 available records. The initial ensemble model was officially scored achieving a 0.081 area under the precision-recall curve (AUPRC) on a test set, whereas for the recurrent neural network model the average AUPRC was 0.295, obtained using 10-fold cross-validation.

1. Introduction

Sleep-related breathing disorders (SBD) are increasingly common in general population [1]. The gas exchange can be disturbed totally (apnea) or partially (hypopnea), resulting in oxygen desaturation, hypercapnia, and fragmentation of sleep. Consequently, impairments associated with SBD vary from cardiovascular, cognitive to metabolic [1,2]. The increased inspiratory effort in obstructive SBD usually terminates with an arousal from sleep during which breathing is restored [3]. Even much subtler airway obstructions, associated with increased respiratory effort, without notable reduction in airflow can lead to an arousal, and thus excessive daytime sleepiness [3]. These events, named respiratory-effort related arousals (RERAs) are predominant in the upper airway resistance syndrome (UARS) [4].

In diagnosing SBD, measuring of a respiratory effort is done using different monitoring techniques. Oesophageal manometry is being considered the gold standard for this purpose. Being invasive and not routinely used, oesophageal manometry is often replaced with non-invasive techniques such as: alternation of a nasal cannula flow curve, continuous positive airway pressure, pulse transit time, and a sum from respiratory inductance plethysmography (RIP) [3,4]. As diagnosis of SBD is routinely done using overnight polysomnography, the sleep technicians use multitude of recorded signals to detect RERAs, simultaneously observing the airflow curve, electroencephalogram and RIP [2].

In this paper we investigated the possibility of an automatic detection of RERAs, using polysomnographic recordings of 994 subjects, provided as a training data set for Physionet/CinC Challenge 2018 [5]. Physiological variability, subtle airflow changes, and recording artefacts significantly hamper the reliable detection. Relying on electroencephalogram (EEG) based arousal detection, oxygen saturation (SaO2), RR interval monitoring, thoraco-abdominal electromyography (EMG), and airflow features, we applied two classification schemes: an ensemble learning approach based on logistic regression classifiers and neural network consisting of two fully connected layers and a bidirectional long short-term memory (LSTM) layer.

2. Challenge database

The Physionet/CinC Challenge 2018 training set consists of polysomnographic recording sets of 994 subjects, adding up to the overall 135GB of data. Each polysomnographic recording set contains 13 signals: 6 EEG channels, electrooculography (EOG), chin, thoracic (CHEST) and abdominal (ABD) EMG, airflow, electrocardiology (ECG), and oxygen saturation SaO2 data. All signals are digitized at 200Hz, except for SaO2 which is resampled to this frequency. The data is annotated in a manner that it classifies arousals as either: spontaneous arousals (total of 70 arousals), respiratory effort related arousals (43822), bruxisms (30), hypoventilations (4), hypopneas (56936), apneas (central - 22763, obstructive - 32547 and mixed - 2641), snores (28), periodic leg movements (36), Cheyne-Stokes breathing (3) or partial airway obstructions (11). The target arousals exclude apnea and hypopnea arousals, as

 ISSN: 2325-887X DOI: 10.22489/CinC.2018.226

much more studied and more prominent, concentrating on the rest of the arousal causes. Additional annotations for different sleeping stages are as well provided.

The test dataset consists of the same type of polysomnographic recordings from another 989 subjects.

3. Methods

3.1. Preprocessing

The numerous recording artefacts, sporadic loss of signal, and specific sensor induced noise hampered the reliable feature extraction. To overcome these issues, a preprocessing procedure was applied.

An illustrative example is given in Fig. 1 presenting the sharp changes in amplitude range of an airflow signal. To overcome these issues, energy normalization on 1-minute long window was applied. Besides, all polysomnographic signals (SaO2 is the only exception) were normalized with their median and inter quartile range to obtain similar amplitude ranges.

Mains hum was detected in respiratory airflow, ECG, and EMG (abdomen, chest) and thus low pass filtering was applied. Depending on the frequency range of interest, different low-pass FIR filters were applied, as summarized in Table 1.

It should be noted that EMG signals from the chin movement and the EOG signals were omitted from the algorithm pipeline.

3.2. Feature extraction

To detect RERA events, the features were extracted every 5s, but the duration of the analysis window depended on the signal type and varied in the range of 10-60s. All of the features used in this study are listed in Table 2.

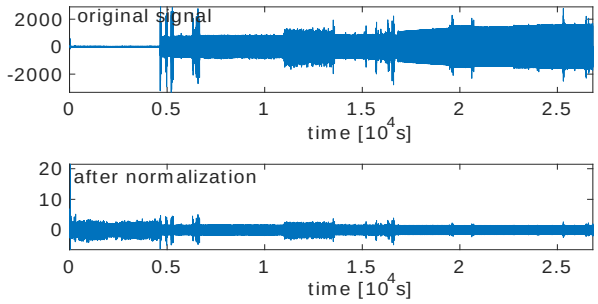

Figure 1. Example of the resulting preprocessing step on an airflow signal

Table 1: Parameters of noise removal FIR filters

Signal type	Passband freq.[Hz]	Stopband freq.[Hz]	Window function
Airflow	2	10	Blackman
ECG	30	50	Blackman
EMG	20	30	Blackman

For airflow signals the features were extracted from 10s windows, reflecting the smaller amplitude range of the airflow signal in the target arousals compared to non-arousal segments. The abdomen and chest signals were observed for the changes in amplitude and envelope [6] within a 30s window centered around the current 10s segment. EEG signals were analysed both at 60s and 10s frames, resulting in two sets of features. The 60s frames served for wakefulness detection based on wavelet features [7] extracted from a single O1-M2 channel. In 10s EEG windows we focused on a narrow alpha band [5-12Hz], proposed in [8], and used only F3-M2, C3-M2 and O1-M2 (as the same features from the other EEG channels were highly correlated). SaO2 features were extracted from both 10s and 60s window intervals to monitor the changes in the SaO2 signal usually visible 10-30s after a RERA occurs. RR series were extracted from whole ECG signals using two QRS detection algorithms [9-11] and their agreement served for additional quality control [12]. In the feature extraction step 30s windows were used to characterize heart rate variability using the recommended time and frequency domain descriptors [13].

A subset of features for some channels was accompanied with an additional feature reflecting feature set veracity, i.e. informing on the eventual signal loss. These features were added only in the official phase, to overcome the problem of invalid segments in a sequential model. These features are marked with * in Table 2.

The whole feature set, with the exception of the supplementary features, was normalized using standard z score normalization, and were randomly arranged for a 10-fold cross-validation input taking care that each patient's recording is either in training or test set.

3.3. Classification

Two different approaches were tested for the task of classifying the RERA arousals from the non-target arousals and normal sleep sequences.

For the unofficial competition phase, we have employed an ensemble approach (Fig. 2a). The classification ensemble comprised of one classifier per training subject, each built up from five logistic regression classifiers, each using distinct feature subsets. Each test subject was then classified according to these equally weighted models, and a voting mechanic

Table 2. Feature list. Features written in *italic* are features additionally added for the official phase. An asterisk (*) indicates features with the supplementary feature measuring validity

Category	Feature
AIR	average energy
AIR	position of the most prominent nonnegative side peak in correlation (MPSP)*
AIR	ratio of MPSP and energy*
AIR	number of nonnegative side peaks in correlation (NSP)*
AIR	average NSP distance*
AIR	standard deviation of NSP distances*
AIR	average NSP value*
AIR	NPS standard deviation*
AIR	90^{th} percentile
ABD,CHEST	maximum
ABD,CHEST	minimum
ABD,CHEST	standard deviation
ABD,CHEST	average peak distances*
ABD,CHEST	standard deviation of peak distances*
ABD,CHEST	average peak value*
ABD,CHEST	standard deviation of peak values*
ABD,CHEST	average peak-to-valley measurement*
ABD,CHEST	standard deviation of peak-to-valley measurements*
ABD,CHEST	*90^{th} percentile*
ABD,CHEST	average phase offset*
ABD,CHEST	*standard deviation of the phase offsets**
ABD,CHEST	the respiratory disturbance variable
ABD,CHEST	the respiratory disturbance variable from the joined signals
EEG	O1-M2 average
EEG	O1-M2 kurtosis
EEG	O1-M2 coefficients from the wavelet decomposition
EEG	*F3-M2, C3-M2, O1-M2 average power*
EEG	F3-M2, C3-M2, O1-M2 standard deviation of the power
EEG	F3-M2, C3-M2, O1-M2 average power of the 5-12Hz band
EEG	F3-M2, C3-M2, O1-M2 power ratio of the 5-12Hz band and the full band
SaO2	minimum value from 10s interval
SaO2	*average value from 10s interval*
SaO2	*standard deviation from 10s interval*
SaO2	zero cross rate from the 10s interval
SaO2	average absolute difference of 5 segments comprised of 12s intervals in a 60s region
SaO2	difference in value of the central 12s segment from the following segment
SaO2	*difference in value of the central 12s segment from the previous segment*
ECG	maximum
ECG	minimum
ECG	interquartile range
ECG	standard deviation
ECG	median absolute deviation
ECG	*maximum result likeness from the 2 QRS detectors*
ECG	average RR distance in the central 10s window*
ECG	standard deviation of the RR distances in the central 10s window*
ECG	root mean squared value of the approximate derivative of the RR distances in the central 10s window*
ECG	standard deviation of the approximate derivative of the RR distances*
ECG	*3 variants of pNN50**
ECG	*minor and major axes of the Poincaré plot obtained from the 30s interval**
ECG	power percentage of band up to 0.04Hz*
ECG	power percentage of band in range 0.04-0.15Hz*
ECG	power percentage of band in range 0.15-0.5Hz

determined the overall probability that an input sample is deemed as a target arousal.

During the official phase we employed a single classifier model based on the currently most effective sequence model - bidirectional LSTM [14] whose architecture is shown in Fig. 2 b). The input is a concatenation of features in the currently observed frame and its predecessor and successor. The goal of the layers between the input and LSTM layer was feature selection.

Dropout layers with a factor of 0.4 were applied to reduce overfitting. To accelerate training procedure after each layer minibatch normalization is applied. Output has 2 values, in order to accommodate the network structure to the implemented cross entropy loss function in the Microsoft Cognitive Toolkit (CNTK) [15]. The training was done for 21 epochs. Additional network architectures were explored, in the terms of number of hidden layers, number of cells within a layer and different activation functions, but did not yield better results for our case of inputs.

4. Results and discussion

The evaluation of the constructed models is performed by calculating the area under the precision-recall curve (AUPRC) comparing the target arousals and normal sleep sequences with the true labels. The ensemble model from

the unofficial phase scored modestly with an AUPRC of 0.081 on the test set, and a similar cross-validation AUPRC estimate of 0.105. This result indicated that due to high inter-subject variability a linear model is not able to distinguish the target arousal, and applied general voting mechanism is not adequate. The neural network model was not evaluated on the test set, yet the cross-validation AUPRC estimate of 0.295 was a considerable improvement. The neural network model performed better as the introduced non-linearities managed to detect the target arousals.

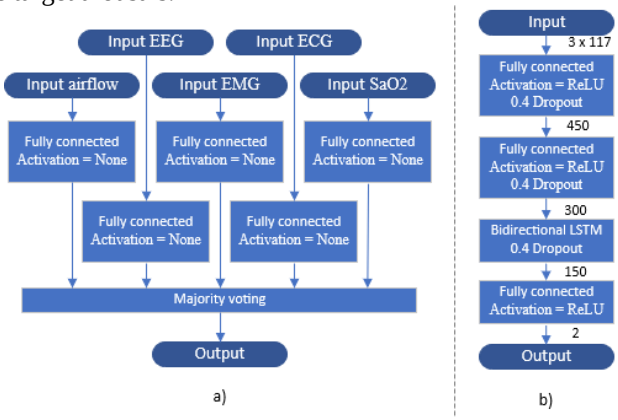

Figure 2. The proposed models for the a) unofficial and b) official phase

5. Conclusion

The goal of the 2018 Physionet challenge was to determine target arousal areas during sleep. In this paper we explored two approaches, both of which used tailored preprocessing and feature extraction. The classifier models were an ensemble based linear model with voting for decisions, and a neural network model. The final score for the initial model was 0.081 and our estimate for the second model was 0.295. Additional improvements could be made by including the omitted signals of the chin area and eye movement as well as training the network to find the optimal features.

Acknowledgements

This work was in part supported by the grants TR32035, III43002, TR32040, of the Ministry of Education, Science and Technological Development of the Republic of Serbia. We would like to thank the BioSense Institute (University of Novi Sad) for kindly permitting the usage of their computational resources.

References

[1] Jordan AS, McSharry DG, Malhotra A. Adult obstructive sleep apnoea. Lancet 2014;383(9918):736-747.

[2] American Academy of Sleep Medicine Task Force. Sleep related breathing disorders in adults: recommendations for syndrome definition and measurement techniques in clinical research. The report of an American Academy of Sleep Medicine task force. Sleep 1999;22(5):667-689.

[3] Vandenbussche N, Overeem S, Johannes P. van Dijk, Simons PJ, Pevernagie DA. Assessment of respiratory effort during sleep: esophageal pressure versus noninvasive monitoring techniques. Sleep Medicine Reviews 2015;24:28-36.

[4] Masa JF, Corral J, Martín MJ, Riesco JA, Sojo A, Hernández M, Douglas NJ. Assessment of thoracoabdominal bands to detect respiratory effort-related arousal. Eur Respir J 2003;22(4):661-667.

[5] 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. Computing in Cardiology Volume 45. Maastricht, Netherlands, 2018. pp 1-4

[6] Díaz JA, Arancibia JM, Bassi A, Vivaldi EA. Envelope analysis of the airflow signal to improve polysomnographic assessment of sleep disordered breathing. Sleep 2014;37(1):199-208.

[7] Zoubek L, Charbonnier S, Lesecq S, Buguet A, Chapotot F. Feature selection for sleep/wake stages classification using data driven methods. Biomedical Signal Processing and Control 2007;2:171-179.

[8] Oropesa E, Cycon HL, Jobert M. Sleep stage classification using wavelet transform and neural network. ICSI Technical Report TR-99-008 1999.

[9] Afonso VX, Tompkins WJ, Nguyen TQ, Luo S. ECG beat detection using filter banks. IEEE Transactions on Biomedical Engineering 1999;46:192-202.

[10] Oppenheim AV, Schafer RW, Buck JR. Discrete-time signal processing, second edition. Prentice Hall 1999:chapter 4.7.3

[11] http://www.librow.com/articles/article-13

[12] Clifford GD, Behar J, Li Q, Rezek I. Signal quality indices and data fusion for determining clinical acceptability of electrocardiograms. Physiol Meas 2012;33(9):1419-1433

[13] Task Force of The European Society of Cardiology and The North American Society of Pacing and Electrophysiology. Heart rate variability: standards of measurement, physiological interpretation, and clinical use. European Heart Journal 1996;17:354–381.

[14] Goodfellow I, Bengio Y, Courville I. Sequence modeling: recurrent and recursive nets in deep learning. MIT Press, 2016, pp. 373-420.

[15] Seide F, Agarwal A. CNTK: Microsoft's open-source deep-learning toolkit. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining 2016:2135-2135

Address for correspondence.

Ivan Lazić
Trg Dositeja Obradovica 6, 21000 Novi Sad, Serbia
ivan.lazic@uns.ac.rs