Automatic Detection of Target Regions of Respiratory Effort-Related Arousals Using Recurrent Neural Networks

Heiðar Már Þráinsson*, Hanna Ragnarsdóttir*, Guðni Fannar Kristjánsson, Bragi Marinósson, Eysteinn Finnsson, Eysteinn Gunnlaugsson, Sigurður Ægir Jónsson, Jón Skírnir Ágústsson, Halla Helgadóttir

Nox Research, Nox Medical ehf, Reykjavík, Iceland

*These authors contributed equally

Abstract

We present a method for classifying target sleep arousal regions of polysomnographies. Time- and frequency-domain features of clinical and statistical origins were derived from the polysomnography signals and the features fed into a Bidirectional Recurrent Neural Network, using Long Short-Term Memory units (BRNN-LSTM). The predictions of five recurrent neural networks, trained using different features and training sets, were averaged for each sample, to yield a more robust classifier. The prosed method was developed and validated on the PhysioNet Challenge dataset which consisted of a training set of 994 subjects and a hidden test set of 989 subjects. The classifier obtained an area under precision-recall curve (AUPRC) score of 0.489 and an area under receiver operating characteristic curve (AUROC) score of 0.912 on a random test subset. Five-fold cross-validation on the training set resulted in an average AUPRC score of 0.452 and AUROC score of 0.901.

1. Introduction

In the sleep medicine and diagnostics AASM manual, arousals are defined as interruptions of sleep lasting 3 to 15 seconds with at least 10 seconds of previous stable sleep, characterized with an abrupt shift of electroencephalography (EEG) frequency [1]. Arousals can occur spontaneously or as a result of sleep-disordered breathing or other sleep disorders [2]. Respiratory Event Related Arousals (RERA) are arousals that are caused by sequences of breaths lasting more than 10 seconds characterized by increasing respiratory effort [1]. The identification of arousals is important for the evaluation of sleep continuity because repeated sleep arousals cause sleep fragmentation [3].

Manual scoring of these events is costly due to the huge amount of data recorded per night, and difficult due to variance across patients and technicians experience [4]. Automation of the detection procedure is therefore important and different works have explored different ways of automating the process. Alvarez-Estevez and Moret-Bonillo [5] developed a method for the detection of EEG arousals using two EEG channels and electromyography (EMG). Experiments conducted on 20 patients reported a sensitivity and specificity respectively of 0.86 and 0.76 in the detection of the arousal events. Behera et al. [6] followed the study, adding more features to the input of an artificial neural network, and combining different models. Experiments conducted on 26 patients reported a sensitivity of 0.81 and a specificity of 0.88 with an error of 0.13. More recently, Isaac Fernández-Varela et al. [7] combined different various signal analysis solutions to identify relevant arousal patterns with special emphasis on robustness and artifacts tolerance. Experiments conducted on 22 patients reported precision of 0.86 and F1 score of 0.79. However, all of these methods were developed on datasets containing relatively few subjects and may not generalize well across different populations.

We propose a recurrent neural network-based approach for classifying target sleep arousal regions, using full polysomnography recordings. The algorithm was trained and tested on the PhysioNet Challenge 2018 database, which includes 1985 subjects. The results are thus based on a larger dataset than previous methods. The method was implemented using Keras 2.1.5, using Tensorflow 1.8.0 backend. The code was submitted for the OpenSource Challenge call of the PhysioNet Challenge 2018.

2. Methods

We employed a three layer neural network with two hidden layers. The first hidden layer is a BRNN and the second hidden layer is a dense neural network. Time- and frequency domain features were derived from multiple available signal including the EEG, ECG and respiratory

signals. The features were fed into the neural network, which after training outputs the probability that a given segment is a target arousal region.

2.1. Feature extraction

For each subject a variety of biometric signals, relevant to sleep studies, were recorded. EEG recordings were made in the following configurations: F3-M2, F4-M1, C3-M2, C4-M1, O1-M2, O2-M1. Additionally the left eye electrooculogram was recorded using the E1-M2 configuration. Signals relating to cardiorespiratory activity were recorded and were as follows: EMG recordings made at the chin, chest and abdomen; oxygen saturation (SaO2), airflow, and electrocardiogram (ECG). Features were extracted from all relevant signals with different signals requiring unique processing methods. All features were calculated over a 10 second sliding window with 50% overlap unless otherwise specified.

2.1.1. EEG features

For each EEG signal, various frequency and time domain features were extracted. The signals were decomposed into sub-bands using the wavelet packet decomposition (WPD). The Daubechies 4 wavelet has been shown to perform well in EEG feature extraction [8] and was used to decompose the signal down to the 4th level, resulting in sub-bands of 6.25 Hz resolution. For each sub-band, statistical features were calculated, as well as sub-band energy. Additionally, the Hjorth parameters were calculated for the signals. These parameters are Hjorth activity, mobility, and complexity and represent signal power, mean frequency, and change in frequency respectively [9].

2.1.2. Respiratory features

Features were extracted from the respiratory signals which could indicate respiratory disturbance. Statistical features calculated from the SaO2 signal indicate changes in oxygen saturation which correlate with apnea [10]. Since the characteristics of the airflow, chest, and abdomen signals vary between individuals the statistical features give information about changes in respiratory activity. Correlation between abdomen and thorax signals was calculated to detect when the two signals go out of phase, which is an indicator of obstructive apnea [11].

2.1.3. ECG features

For the ECG signal we derived various features relating to the heart rate. The QRS complexes of the ECG signals were detected using a robust R-peak finder [12]. From the locations of the R-peaks the heart rate and Heart Rate Variability (HRV) signals were calculated. Statistical features were calculated from the heart rate while more complex frequency domain features were derived from the HRV. The power spectrum of the HRV is an important indicator of the function of the nervous system and has been shown to be a good indicator of apneas [13]. The HRV signal was interpolated using cubic spline interpolation to get a signal of constant sampling frequency. A spectrogram of the HRV was then calculated using Welch's method with a sliding Hamming window. The frequency bands of interest are the very low frequency (VLF) 0.003 - 0.04 Hz, low frequency (LF) 0.04 - 0.15 Hz and high frequency (HF) 0.15 - 0.4 Hz [14]. For each band we calculated the normalized total energy, peak energy and peak frequency as well as the ratio of LF and HF power.

2.2. Classification

Recurrent neural networks, using LSTM hidden units, are powerful models for learning from sequence data since they are capable of remembering information for a long period of time. Bidirectional recurrent neural networks can further learn from both past and future states, which is important when context of the input is needed, as for detecting sleep arousals [15]. We thus considered LSTM-based BRNN model for the sleep arousal detection.

2.2.1. Data preparation

After feature extraction, the data was reshaped for the BRNN-LSTM layer into a three-dimensional array, where the three dimensions are:

- Number of training sequences, N
- Sequence length (number of time-steps), W
- Number of features of each sequence, F

By experimenting with different values for the sequence length and different positions of the label, we found $W = 20$ to result in the best performance, positioning the label at time step 11. The time-steps were composed of features extracted over a 10 second window with a 5 second overlap. Thus, each sequence considered by the classifier was $20 \cdot (10 \cdot 0.5) = 100$ seconds long, with the neural network looking 50 seconds in the past and 40 seconds in the future.

Regions in the training dataset labeled neither as normal regions nor target arousal regions were ignored, as those regions are not considered in the PhysioNet Challenge. The remaining training dataset is unbalanced, with 7% of the data being arousal regions and 93% being normal sleep regions. To achieve a more balanced training dataset we randomly removed 90% of the normal sleep regions.

2.2.2. Sequence Classifiers

We experimented with several model structures and hyperparameters. The model that performed the best was a three-layer model, where the first hidden layer is a LSTM-based BRNN layer consisting of 50 LSTM blocks, and the second hidden layer is a dense layer consisting of 50 nodes. We use a softmax activation function on the output layer to extract probabilities for classification. To combat overfitting, dropout is applied to the output of both hidden layers, and all layers have l2-kernel regulizer of strength 0.01 to further combat overfitting [16]. The neural network was trained using a batch size of 200, learning rate was reduced on plateau and early stopping was used. The loss function used was binary cross-entropy and the optimizer was Adam.

2.2.3. Ensemble Classifier

Five classifiers of the same structure as described above were trained, using different random seeds and different sets of features. The predictions of these classifiers were then averaged per sample, to create a more robust classifier and to reduce variance arising from the random initialization of the weights and the random split between train and validation set. The final ensemble classifier thus consisted of five classifiers, each trained on all the respiratory features, but with different set of two to three EEG and ECG features.

3. Results and Discussion

To evaluate the performance of our method, we perform a 5-fold cross-validation on the training dataset. During each cross validation fold, 20% of the available training data was set aside for final testing. The other 80% were used for the training and validation of the models during the cross validation. For each model the validation set was randomly selected containing 10% of the training and validation data.

According to the PhysioNet Challenge scoring system, results are reported as gross AUPRC score and AUROC score. The gross AUPRC score is the metric used to rank competitors in the PhysioNet Challenge and it is defined as: $AUPRC = \sum_j P_j(R_j - R_{j+1})$, where R_j denotes the number of arousals with predicted probability $(j/1000)$ or greater divided by total the number of arousal samples, and P_j denotes the number of arousal samples with predicted probability $(j/1000)$ or greater divided by the total number of samples with predicted probability $(j/1000)$ or greater. The cross validated scores of the individual models as well as the ensemble classifier are shown in table 1.

Table 1. Cross validated AUPRC and AUROC scores of the individual models as well the final ensemble

Model	AUPRC		AUROC	
	Mean	STD	Mean	STD
Model 1	0.432	0.037	0.893	0.0026
Model 2	0.429	0.035	0.893	0.0027
Model 3	0.426	0.038	0.891	0.0030
Model 4	0.430	0.040	0.893	0.0020
Model 5	0.428	0.032	0.895	0.0030
Averaged model	0.452	0.038	0.901	0.0030

The performance of the ensemble classifier was higher than of the individual models. The performance of the final model was further verified by the official test data with the AUPRC and AUROC score on a random subset of the test set being 0.489 and 0.912 respectively which is higher than the cross-validated score.

4. Conclusion

The problem of automatically detecting sleep arousals is not a trivial one and more work remains to be done. However, being able to effectively score arousals automatically is important, as manual scoring of arousals is time consuming and difficult. In this paper we have proposed a method for classifying target sleep arousal regions, using a BRNN-LSTM ensemble model. The method was validated on PhysioNet Challenge 2018 dataset and the results are encouraging.

Acknowledgements

This work was supported by the Icelandic Centre for Research under the Icelandic Student Innovation Fund and the Horizon 2020 SME Instrument number 733461.

References

[1] Berry RB, Albertario CL, Harding SM, Lloyd RM, Plante DT, Quan SF, Troester MM, Vaughn BV. The AASM Manual for the Scoring of Sleep and Associated Events; Rules, Terminology and Technical Specifications, version 2.5. American Academy of Sleep Medicine.

[2] Bonnet M, Carley D, Carskadon M, Easton P, Guilleminault C, Harper R, Hayes B, Hirshkowitz M, Ktonas P, Keenan S, Pressman M, Roehrs T, Smith J, Walsh J, Weber S, Westbrook P, Jordan B. EEG arousals: Scoring rules and examples. a preliminary report from the sleep disorders atlas task force of the american sleep disorder association 01 1992; 15:173–184.

[3] Mesquita J, Porée F, Carrault G, Fiz JA, Abad J, Jané R. Respiratory and spontaneous arousals in patients with sleep apnea hypopnea syndrome. In 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society. ISSN 1094-687X, 2012; 6337–6340.

[4] Bonnet MH, Doghramji K, Roehrs T, Stepanski EJ, Sheldon SH, Walters AS, Wise M, Chesson AL. The scoring of arousal in sleep: Reliability, validity, and alternatives ; 03(2):133–145. ISSN 1550-9389.

[5] Álvarez Estévez D, Moret-Bonillo V. Identification of electroencephalographic arousals in multichannel sleep recordings. IEEE Transactions on Biomedical Engineering Jan 2011;58(1):54–63. ISSN 0018-9294.

[6] Behera CK, Reddy TK, Behera L, Bhattacarya B. Artificial neural network based arousal detection from sleep electroencephalogram data. In 2014 International Conference on Computer, Communications, and Control Technology (I4CT). Sept 2014; 458–462.

[7] Moret-Bonillo V, Fernández-Varela I, Hernández-Pereira E, Alvarez-Estévez D, Perlitz V. On the automation of medical knowledge and medical decision support systems. Computers in biology and medicine ;87:77–86.

[8] Faust O, Acharya UR, Adeli H, Adeli A. Wavelet-based EEG processing for computer-aided seizure detection and epilepsy diagnosis. Seizure 2015;26:56 – 64. ISSN 1059-1311.

[9] Hjorth B. EEG analysis based on time domain properties. Electroencephalography and Clinical Neurophysiology 1970;29(3):306 – 310. ISSN 0013-4694.

[10] Alvarez D, Hornero R, Marcos JV, del Campo F. Multivariate analysis of blood oxygen saturation recordings in obstructive sleep apnea diagnosis. IEEE Transactions on Biomedical Engineering Dec 2010;57(12):2816–2824. ISSN 0018-9294.

[11] Berry RB, Budhiraja R, Gottlieb DJ, Gozal D, Iber C, Kapur VK, Marcus CL, Mehra R, Parthasarathy S, Quan SF, Redline S, Strohl KP, Ward SLD, Tangredi MM. Rules for scoring respiratory events in sleep: Update of the 2007 aasm manual for the scoring of sleep and associated events: Deliberations of the sleep apnea definitions task force of the american academy of sleep medicine. J Clin Sleep Med Oct 2012;8(5):597–619. ISSN 1550-9389. 23066376[pmid].

[12] Kathirvel P, Sabarimalai Manikandan M, Prasanna SRM, Soman KP. An efficient r-peak detection based on new nonlinear transformation and first-order gaussian differentiator. Cardiovascular Engineering and Technology Dec 2011; 2(4):408–425. ISSN 1869-4098.

[13] Penzel T, McNames J, de Chazal P, Raymond B, Murray A, Moody G. Systematic comparison of different algorithms for apnoea detection based on electrocardiogram recordings. Medical and Biological Engineering and Computing Jul 2002;40(4):402–407. ISSN 1741-0444.

[14] Shaffer F, Ginsberg JP. An overview of heart rate variability metrics and norms. Frontiers in Public Health 2017;5:258. ISSN 2296-2565.

[15] Lipton ZC. A critical review of recurrent neural networks for sequence learning. CoRR 2015;abs/1506.00019.

[16] Cogswell M, Ahmed F, Girshick R, Zitnick L, Batra D. Reducing overfitting in deep networks by decorrelating representations 11 2015;.

Address for correspondence:

Nox Medical
Katrínartún 2, 105 Reykjavík
halla@noxmedical.com