Research Article

Sleep Stage Classification Using Unsupervised Feature Learning

Martin Längkvist, Lars Karlsson, and Amy Loutfi

Center for Applied Autonomous Sensor Systems, Örebro University, 701 82 Örebro, Sweden

Correspondence should be addressed to Amy Loutfi, amy.loutfi@oru.se

Most attempts at training computers for the difficult and time-consuming task of sleep stage classification involve a feature extraction step. Due to the complexity of multimodal sleep data, the size of the feature space can grow to the extent that it is also necessary to include a feature selection step. In this paper, we propose the use of an unsupervised feature learning architecture called *deep belief nets* (DBNs) and show how to apply it to sleep data in order to eliminate the use of handmade features. Using a postprocessing step of hidden Markov model (HMM) to accurately capture sleep stage switching, we compare our results to a feature-based approach. A study of anomaly detection with the application to home environment data collection is also presented. The results using raw data with a deep architecture, such as the DBN, were comparable to a feature-based approach when validated on clinical datasets.

1. Introduction

One of the main challenges in sleep stage classification is to isolate features in multivariate time-series data which can be used to correctly identify and thereby automate the annotation process to generate sleep hypnograms. In the current absence of a set of universally applicable features, typically a two-stage process is required before training a sleep stage algorithm, namely, feature extraction and feature selection [1–9]. In other domains which share similar challenges, an alternative to using hand-tailored feature representations derived from expert knowledge is to apply unsupervised feature learning techniques, where the feature representations are learned from unlabeled data. This not only enables the discovery of new useful feature representations that a human expert might not be aware of, which in turn could lead to a better understanding of the sleep process and present a way of exploiting massive amounts of unlabeled data.

Unsupervised feature learning and in particular deep learning [10–15] propose ways for training the weight matrices in each layer in an unsupervised fashion as a preprocessing step before training the whole network. This has proven to give good results in other areas such as vision tasks [10], object recognition [16], motion capture data [17], speech recognition [18], and bacteria identification [19].

This work presents a new approach to the automatic sleep staging problem. The main focus is to learn meaningful feature representations from unlabeled sleep data. A dataset of 25 subjects consisting of electroencephalography (EEG) of brain activity, electrooculography (EOG) of eye movements, and electromyography (EMG) of skeletal muscle activity is segmented and used to train a deep belief network (DBN), using no prior knowledge. Validation of the learned representations is done by integrating a hidden Markov model (HMM) and compare classification accuracy with a feature-based approach that uses prior knowledge. The inclusion of an HMM serves the purpose of improving upon capturing a more realistic sleep stage switching, for example, hinders excessive or unlikely sleep stage transitions. It is in this manner that the knowledge from the human experts is infused into the system. Even though the classifier is trained using labeled data, the feature representations are learned from unlabeled data. The architecture of the DBN follows previous work with unsupervised feature learning for electroencephalography (EEG) event detection [20].

A secondary contribution of the proposed method leverages the information from the DBN in order to perform

anomaly detection. Particularly, in light of an increasing trend to streamline sleep diagnosis and reduce the burden on health care centers by using at home sleep monitoring technologies, anomaly detection is important in order to rapidly assess the quality of the polysomnograph data and determine if the patient requires another additional night's collection at home. In this paper, we illustrate how the DBN once trained on datasets for sleep stage classification in the lab can still be applied to data which has been collected at home to find particular anomalies such as a loose electrode.

Finally, inconsistencies between sleep labs (equipment, electrode placement), experimental setups (number of signals and categories, subject variations), and interscorer variability (80% conformance for healthy patients and even less for patients with sleep disorder [9]) make it challenging to compare sleep stage classification accuracy to previous works. Results in [2] report a best result accuracy of around 61% for classification of 5 stages from a single EEG channel using GOHMM and AR coefficients as features. Works by [8] achieved 83.7% accuracy using conditional random fields with six power spectra density features for one EEG signal on four human subjects during a 24-hour recording session and considering six stages. Works by [7] achieved 85.6% accuracy on artifact-free, two expert agreement sleep data from 47 mostly healthy subjects using 33 features with SFS feature selection and four separately trained neural networks as classifiers.

The goal of this work is not to replicate the R&K system or improve current state-of-the-art sleep stage classification but rather to explore the advantages of deep learning and the feasibility of using unsupervised feature learning applied to sleep data. Therefore, the main method of evaluation is a comparison with a feature-based shallow model. Matlab code used in this paper is available at http://aass.oru.se/~mlt.

2. Deep Belief Networks

DBN is a probabilistic generative model with deep architecture that searches the parameter space by unsupervised greedy layerwise training. Each layer consists of a restricted Boltzmann machine (RBM) with visible units, $\mathbf{v}$, and hidden units, $\mathbf{h}$. There are no visible-visible connections and no hidden-hidden connections. The visible and hidden units have a bias vector, $\mathbf{c}$ and $\mathbf{b}$, respectively. The visible and hidden units are connected by a weight matrix, $\mathbf{W}$, see Figure 1(a). A DBN is formed by stacking a user-defined number of RBMs on top of each other where the output from a lower-level RBM is the input to a higher-level RBM, see Figure 1(b). The main difference between a DBN and a multilayer perceptron is the inclusion of a bias vector for the visible units, which is used to reconstruct the input signal, which plays an important role in the way DBNs are trained.

A reconstruction of the input can be obtained from the unsupervised pretrained DBN by encoding the input to the top RBM and then decoding the state of the top RBM back to the lowest level. For a Bernoulli (visible)-Bernoulli (hidden) RBM, the probability that hidden unit h_j is activated given
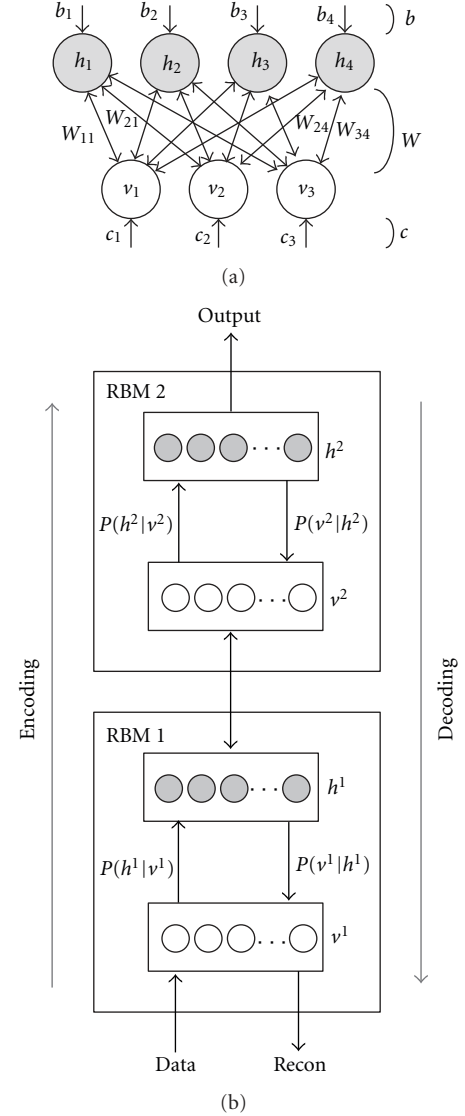

Figure 1: Graphical depiction of (a) RBM and (b) DBN.

visible vector, $\mathbf{v}$, and the probability that visible unit v_i is activated given hidden vector, $\mathbf{h}$, are given by

$$P\left(h_j \mid \mathbf{v}\right) = \frac{1}{1 + \exp^{b_j + \sum_i W_{ij} v_i}}$$

$$P(v_i \mid \mathbf{h}) = \frac{1}{1 + \exp^{c_i + \sum_j W_{ij} h_j}}. \tag{1}$$

The energy function and the joint distribution for a given visible and hidden vector are

$$E(\mathbf{v}, \mathbf{h}) = \mathbf{h}^T \mathbf{W} \mathbf{v} + \mathbf{b}^T \mathbf{h} + \mathbf{c}^T \mathbf{v}$$

$$P(\mathbf{v}, \mathbf{h}) = \frac{1}{z} \exp^{E(\mathbf{v}, \mathbf{h})}. \tag{2}$$

The parameters $\mathbf{W}$, $\mathbf{b}$, and $\mathbf{v}$ are trained to minimize the reconstruction error. An approximation of the gradient of

the log likelihood of **v** using contrastive divergence [21] gives the learning rule for RBM:

$$\frac{\partial \log P(\mathbf{v})}{\partial W_{ij}} \approx \left\langle v_i h_j \right\rangle_{\text{data}} - \left\langle v_i h_j \right\rangle_{\text{recon}}, \qquad (3)$$

where $\langle \cdot \rangle$ is the average value over all training samples. In this work, training is performed in three steps: (1) unsupervised pretraining of each layer, (2) unsupervised fine-tuning of all layers with backpropagation, and (3) supervised fine-tuning of all layers with backpropagation.

3. Experimental Setup

3.1. Automatic Sleep Stager. The five sleep stages that are at focus are awake, stage 1 (S1), stage 2 (S2), slow wave sleep (SWS), and rapid eye-movement sleep (REM). These stages come from a unified method for classifying an 8 h sleep recording introduced by Rechtschaffen and Kales (R&K) [22]. A graph that shows these five stages over an entire night is called a hypnogram, and each epoch according to the R&K system is either 20 s or 30 s. While the R&K system brings consensus on terminology, among other advantages [23], it has been criticized for a number of issues [24]. Even though the goal in this work is not to replicate the R&K system, its terminology will be used for evaluation of our architecture. Each channel of the data is divided into segments of 1 second with zero overlap, which is a much higher temporal resolution than the one practiced by the R&K system.

We compare the performance of three experimental setups as shown in Figure 2.

3.1.1. Feat-GOHMM. A Gaussian observation hidden Markov model (GOHMM) is used on 28 handmade features; see the appendix for a description of the features used. Feature selection is done by sequential backward selection (SBS), which starts with the full set of features and greedily removes a feature after each iteration step. A principal component analysis (PCA) with five principal components is used after feature selection, followed by a Gaussian mixture model (GMM) with five components. The purpose of the PCA is to reduce dimensionality, and the choice of five components was made since it captured most of the variance in the data, while still being tractable for the GMM step. Initial mean and covariance values for each GMM component are set to the mean and covariance of annotated data for each sleep stage. Finally, the output from the GMM is used as input to a hidden Markov model (HMM) [25].

3.1.2. Feat-DBN. A 2-layer DBN with 200 hidden units in both layers and a softmax classifier attached on top is used on 28 handmade features. Both layers are pretrained for 300 epochs, and the top layer is fine-tuned for 50 epochs. Initial biases of hidden units are set empirically to -4 to encouraged sparsity [26], which prevents learning trivial or uninteresting feature representations. Scaling to values between 0 and 1 is done by subtracting the mean, divided by the standard deviation, and finally adding 0.5.
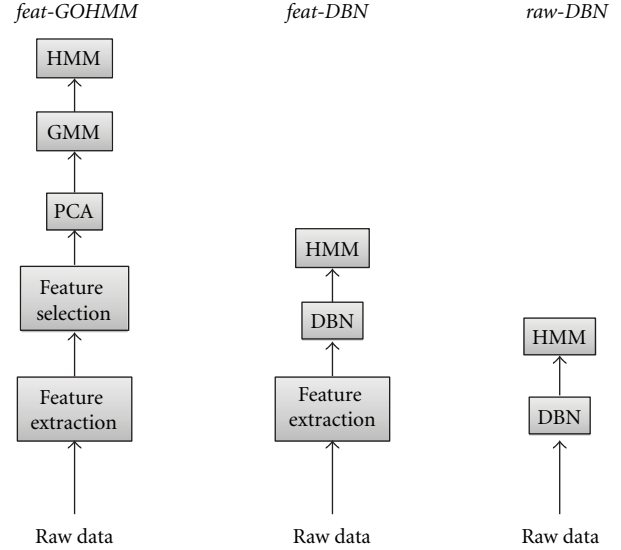

Figure 2: Overview of three setups for an automatic sleep stager used in this work. The first method, *feat-GOHMM*, is a shallow method that uses prior knowledge. The second method, *feat-DBN*, is a deep architecture that also uses prior knowledge. And, lastly, the third method, *raw-DBN*, is a deep architecture that does not use any prior knowledge. See text for more details.

3.1.3. Raw-DBN. A DBN with the same parameters as *feat-DBN* is used on preprocessed raw data. Scaling is done by saturating the signal at a saturation constant, $\text{sat}_{\text{channel}}$, then divide by $2 * \text{sat}_{\text{channel}}$, and finally adding 0.5. The saturation constant was set to $\text{sat}_{\text{EEG}} = \text{sat}_{\text{EOG}} = \pm 60 \, \mu\text{V}$ and $\text{sat}_{\text{EMG}} = \pm 40 \, \mu\text{V}$. Input consisted of the concatenation of EEG, EOG1, EOG2, and EMG. With window width, w, the visible layer becomes

$$v = \begin{bmatrix} \text{EEG}_1^{64} & \text{EOG1}_1^{64} & \text{EOG2}_1^{64} & \text{EMG}_1^{64} \\ \text{EEG}_{1+w}^{64+w} & \text{EOG1}_{1+w}^{64+w} & \text{EOG2}_{1+w}^{64+w} & \text{EMG}_{1+w}^{64+w} \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix}. \quad (4)$$

With four signals, 1 second window, and 64 samples per second, the input dimension is 256.

3.2. Anomaly Detection for Home Sleep Data. In this work, anomaly detection is evaluated by training a DBN and calculating the root mean square error (RMSE) from the reconstructed signal from the DBN and the original signal. A faulty signal in one channel often affects other channels for sleep data, such as movement artifacts, blink artifacts, and loose reference or ground electrode. Therefore, a detected fault in one channel should label all channels at that time as faulty.

Figure 3 shows data that has been collected at a healthy patient's home during sleep. All signals, except EEG2, are nonfaulty prior to a movement artifact at $t = 7$ s. This movement affected the reference electrode or the ground electrode, resulting in disturbances in all signals for the rest of the night, thereby rendering the signals unusable by a clinician. A poorly attached electrode was the cause for
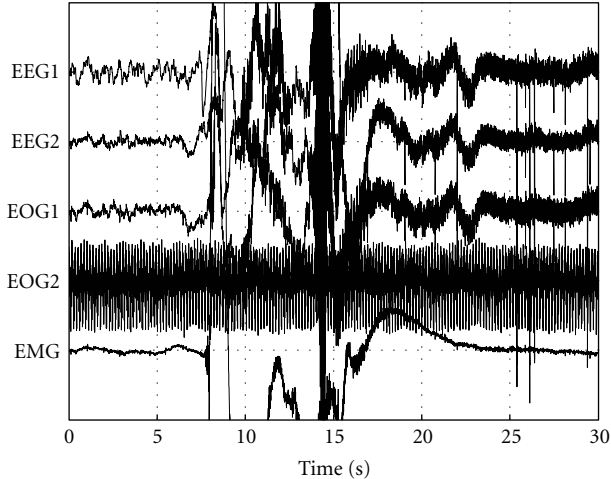
FIGURE 3: PSG data collected in a home environment. A movement occurs at $t = 7$ s resulting in one of the electrodes to be misplaced affecting EOG1 and both EEG channels. EOG2 is not properly attached resulting in a faulty signal for the entire night.

the noise in signal EEG2. Previous approaches to artifact rejection in EEG analysis range from simple thresholding on abnormal amplitude and/or frequency to more complex strategies in order to detect individual artefacts [27, 28].

4. Experimental Datasets

Two datasets are used in this work. The first consists of 25 acquisitions and is used to train and test the automatic sleep stager. The second consists of 5 acquisitions and is used to validate anomaly detection on sleep data collected at home.

4.1. Benchmark Dataset. This dataset has kindly been provided by St. Vincent's University Hospital and University College Dublin, which can be downloaded from PhysioNet [29]. The dataset consists of 25 acquisitions (21 males 4 females with average age 50, average weight 95 kg, and average height 173 cm) from subjects with suspected sleep-disordered breathing. Each acquisition consists of 2 EEG channels (C3-A2 and C4-A1), 2 EOG channels, and 1 EMG channel using 10–20 electrode placements system. Only one of the EEG channel (C3-A2) is used in this work. Sample rate is 128 Hz for EEG and 64 Hz for EOG and EMG. Average recording time is 6.9 hours. Sleep stages are divided into S1: 16.7%, S2: 33.3%, SWS: 12.7%, REM: 14.5%, awake: 22.7%, and indeterminate: 0.1%. Scoring was performed by one sleep expert.

All signals are preprocessed by notch filtering at 50 Hz in order to cancel out power line disturbances and downsampled to 64 Hz after being prefiltered with a band-pass filter of 0.3 to 32 Hz for EEG and EOG, and 10 to 32 Hz for EMG. Each epoch before and after a sleep stage switch is removed from the training set to avoid possible subsections of mislabeled data within one epoch. This resulted in 20.7% of total training samples to be removed.

A 25 leave-one-out cross-validation is performed. Training samples are randomly picked from 24 acquisitions in order to compensate for any class imbalance. A total of approximately 250000 training samples and 50000 training validation samples are used for each validation.

4.2. Home Sleep Dataset. PSG data of approximately 60 hours (5 nights) was collected at a healthy patient's home using a Embla Titanium PSG. A total of 8 electrodes were used: EEG C3, EEG C4, EOG left, EOG right, 2 electrodes for the EMG channel, reference electrode, and ground electrode. Data was collected with a sampling rate of 256 Hz, which was downsampled to match the sampling rate of the training data. The signals are preprocessed using the same method as the benchmark dataset.

5. Results

5.1. Automatic Sleep Stager. A full leave-one-out cross-validation of the 25 acquisitions is performed for the three experimental setups. The classification accuracy and confusion matrices for each setup and sleep stage are presented in Tables 1, 2, 3, and 4. Here, the performance of using a DBN based approach, either with features or using the raw data, is comparable to the *feat-GOHMM*. While the best accuracy was achieved with *feat-DBN*, followed by *raw-DBN* and lastly, *feat-GOHMM*, it is important to examine the performances individually. Figure 4 shows classification accuracy for each subject. The *raw-DBN* setup gives best, or second best, performance in the majority of the sets, with the exception of subjects 9 and 22. An examination of the performance when comparing the F_1-score for individual sleep stages indicates that S1 is the most difficult stage to classify and awake and slow wave sleep is the easiest.

For the *raw-DBN*, it is also possible to analyze the learned features. In Figure 6, the learned features for the first layer are given. Here, it can clearly be seen that both low and high frequency features for the EEG and high and low amplitude features for the EMG are included, which to some degree correspond to the features which are typically selected in handmade feature selection methods.

Some conclusions from analyzing the selected features from the SBS algorithm used in *feat-GOHMM* can be made. Fractal exponent for EEG and entropy for EOG were selected for all 25 subjects and thus proven to be valuable features. Correlation between both EOG signals was also among the top selected features, as well as delta, theta, and alpha frequencies for EEG. Frequency features for EOG and EMG were excluded early, which is in accordance to the fact that these signals do not exhibit valuable information in the frequency domain [30]. The kurtosis feature was selected more frequently when it was applied to EMG and less frequently when it was applied to EEG or EOG. Features of spectral mean for all signals, median for EMG, and standard deviation for EOG were not frequently selected. See Figure 5 for errors bars for each feature at each sleep stage.

It is worth noting that variations in the number of layers and hidden units were attempted, and it was found

TABLE 1: Classification accuracy and F_1-score for the three experimental setups.

	Accuracy (mean ± std)	F_1-score				
		Awake	S1	S2	SWS	REM
feat-GOHMM	63.9 ± 10.8	0.71	0.31	0.72	0.82	0.47
feat-DBN	72.2 ± 9.7	0.78	0.37	0.76	0.84	0.78
raw-DBN	67.4 ± 12.9	0.69	0.36	0.78	0.83	0.58
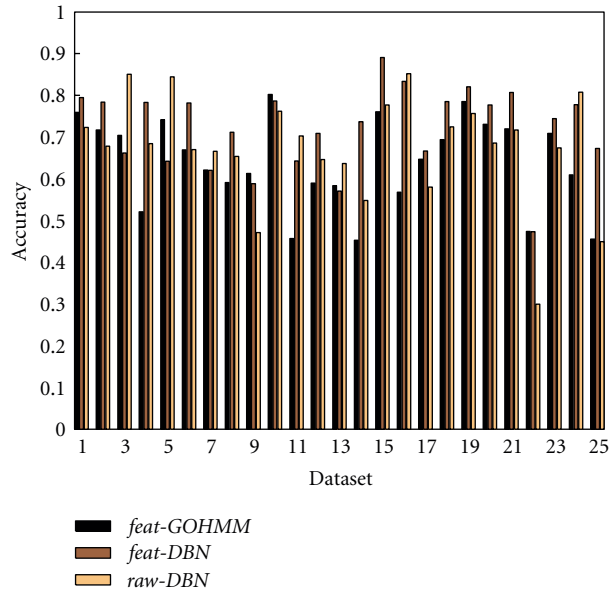
FIGURE 4: Classification accuracy for 25 testing sets for three setups.

TABLE 2: Confusion matrix for *feat-GOHMM*.

%	Classified				
	Awake	S1	S2	SWS	REM
Awake	72.5	16.8	3.0	2.5	5.2
S1	29.4	31.1	25.6	1.5	12.4
S2	2.0	8.4	71.9	7.1	10.6
SWS	1.1	1.3	9.6	87.8	0.2
REM	11.7	13.3	29.4	2.7	42.9

TABLE 3: Confusion matrix for *feat-DBN*.

%	Classified				
	Awake	S1	S2	SWS	REM
Awake	75.8	18.2	1.8	0.2	4.1
S1	26.0	37.6	25.1	0.7	10.6
S2	1.0	9.8	73.1	7.2	9.0
SWS	0.4	0.1	13.9	85.5	0.1
REM	1.9	4.0	10.3	0.1	83.7

TABLE 4: Confusion matrix for *raw-DBN*.

%	Classified				
	Awake	S1	S2	SWS	REM
Awake	68.4	13.4	2.5	0.7	15.1
S1	20.3	33.1	24.8	1.6	20.2
S2	1.0	6.3	76.5	9.1	7.1
SWS	0.1	0.0	11.1	88.8	0.0
REM	21.1	6.9	11.1	0.8	60.1

minimum. With the configuration of the three experimental setups described above and simulations performed on a Windows 7, 64-bit machine with quad-core Intel i5 3.1 GHz CPU with use of a nVIDIA GeForce GTX 470 GPU using GPUmat, simulation time for *feat-GOHMM*, *feat-DBN*, and *raw-DBN* were approximately 10 minutes, 1 hour, and 3 hours per dataset, respectively.

5.2. Anomaly Detection on Home Sleep Data. A total of five acquisitions were recorded at a patient's home during sleep and manually labeled into faulty or nonfaulty signals. A DBN with the *raw-DBN* setup was trained using the benchmark dataset. The root mean square error (RMSE) between the home sleep data and the reconstructed signal from the trained DBN for the five night runs and a close-up for night

that an increase did not significantly improve classification accuracy. Rather, an increase in either the number of layers or hidden units often resulted in a significant increase in simulation time, and therefore to maintain a reasonable training time, the layers and hidden units were kept to a
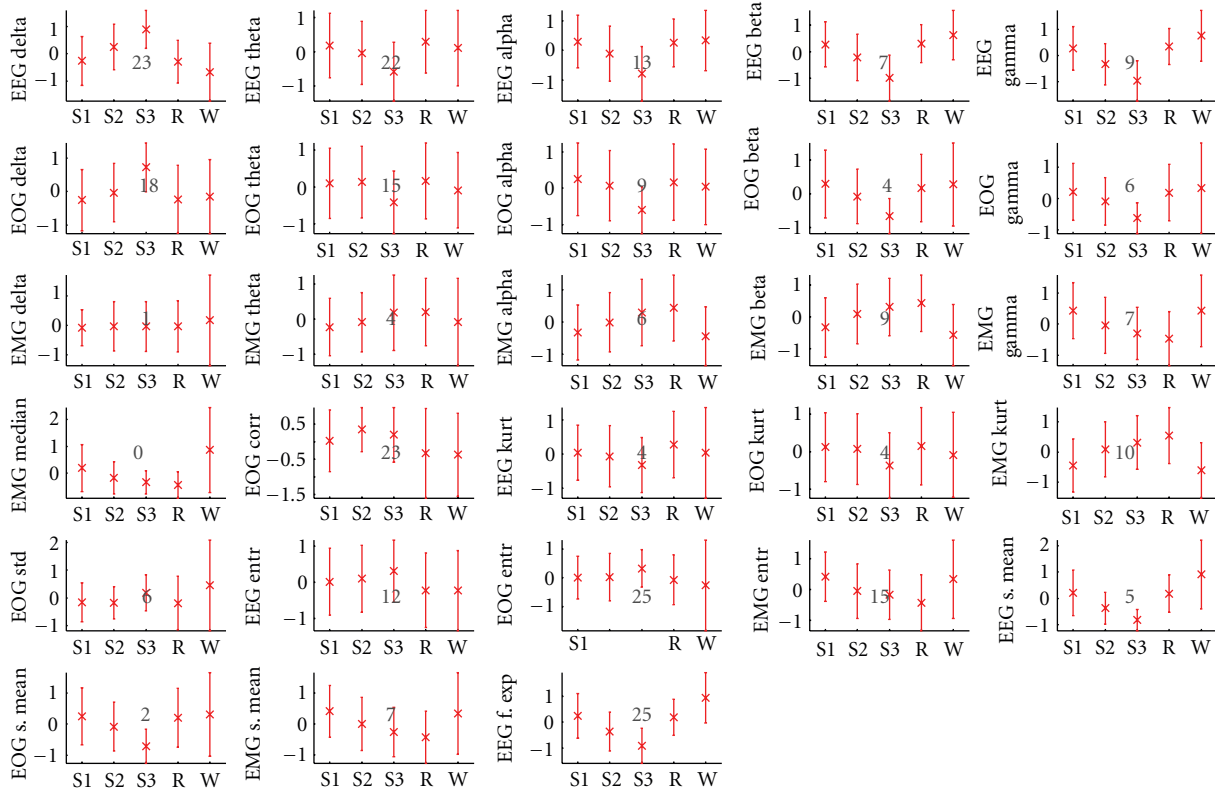
FIGURE 5: Error bar of the 28 features. Gray number in background represents how many times that feature was part of best subset from SBS algorithm (maximum is 25).

2 where an electrode falls off after around 380 minutes can be seen in Figure 7.

Interestingly, attempts on using the *feat-GOHMM* for sleep stage classification on the home sleep dataset resulted in faulty data to be misclassified as awake. This could be explained by the fact that faulty data mostly resembles signals in awake state.

6. Discussion

In this work, we have shown that an automatic sleep stager can be applied to multimodal sleep data without using any handmade features. We also compared the reconstructed signal from a trained DBN on data collected in a home environment and saw that the RMSE was large where an obvious error had occurred.

Regarding the DBN parameter selection, it was noticed that setting initial biases for the hidden units to −4 was an important parameter for achieving good accuracy. A better way of encourage sparsity is to include a sparsity penalty term in the cost objective function [31] instead of making a crude estimation of initial biases for the hidden units. For the *raw-DBN* setup, it was also crucial to train each layer with a large number of epochs and in particular the fine tuning step.

We also noticed a lower performance if sleep stages were not set to equal sizes in the training set. There was also a high variation in the accuracy between patients, even if they came from the same dataset. Since the DBN will find a generalization that best fits all training examples, a testing set that deviates from the average training set might give poor results. Since data might differs greatly between patients, a single DBN trained on general sleep data is not specialized enough. The need for a more dynamic system, especially one including the transition and emission matrices for the HMM, is made clear when comparing the hypnograms of a healthy patient and a patient with sleep disordered breathing. Further, although the HMM provides a simple solution that captures temporal properties of sleep data, it makes two critical assumptions [13]. The first one is that the next hidden state can be approximated by a state depending only on the previous state, and the second one is that observations at different time steps are conditionally independent given a state sequence. Replacing HMM with conditional random fields (CRFs) could improve accuracy but is still a simplistic temporal model that does not exploit the power of DBNs [32].

While a clear advantage of using DBN is the natural way in which it deals with anomalous data, there are some limitations to the DBN. One limitation is that correlations between signals in the input data are not well captured. This gives a feature-based approach an advantage where, for example, the correlation between both EOG channels can easily be represented with a feature. This could be solved by either representing the correlation in the input or extending the DBN to handle such correlations, such as a cRBM [33].
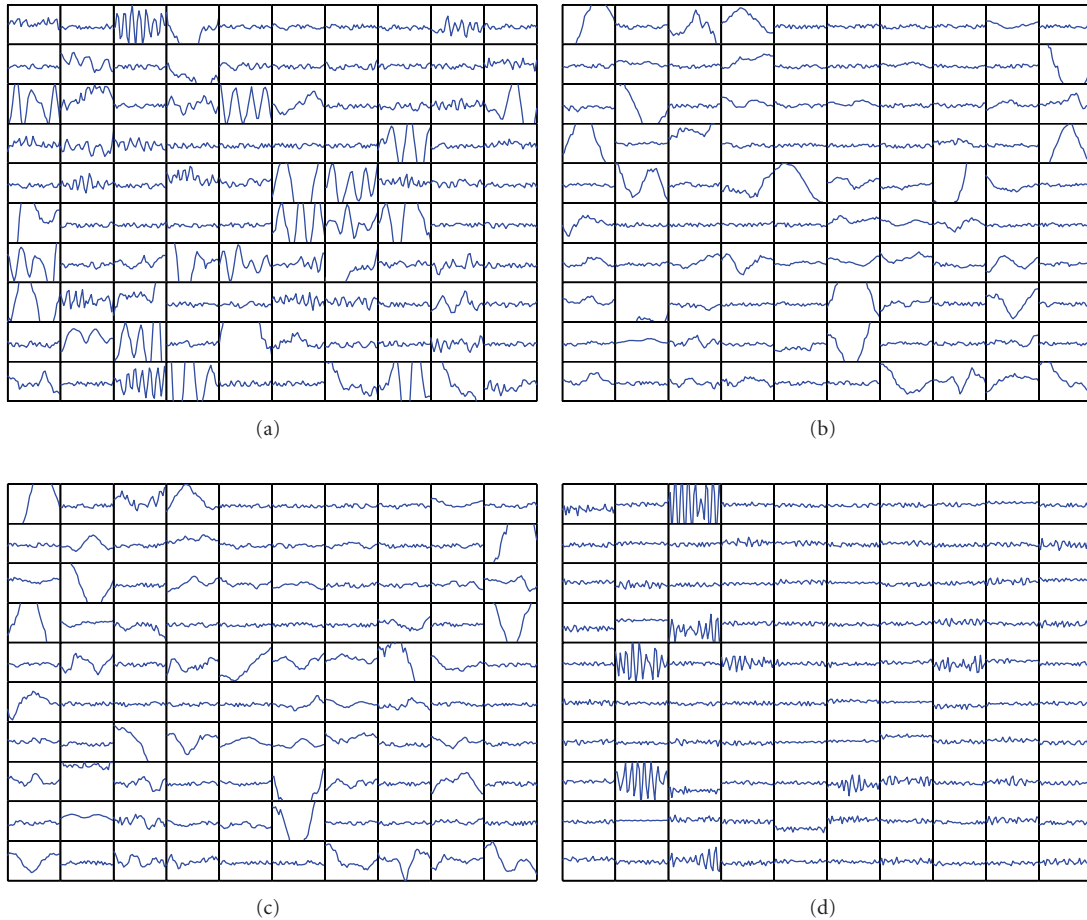
(a)

(b)

(c)

(d)

FIGURE 6: Learned features of layer 1 for (a) EEG, (b) EOG1, (c) EOG2, and (d) EMG. It can be observed that the learned features are of various amplitudes and frequencies and some resemble known sleep events such as a K-complex or blink artifacts. Only the first 100 of the 200 features are shown here.

Regarding the implemented *feat-GOHMM*, we have tried our best to get as high accuracy with the setup as possible. It is almost certain that another set of features, different feature selection algorithm, and/or another classifier could outperform our *feat-GOHMM*. However, we hope that this work illustrates the advantages of unsupervised feature learning, which not only removes the need for domain specific expert knowledge, but inherently provides tools for anomaly detection and noise redundancy.

It has been suggested for multimodal signals to train a separate DBN for each signal first and then train a top DBN with concatenated data [34]. This not only could improve classification accuracy, but also provide the ability to single out which signal contains the anomalous signal. Further, this work has explored clinical data sets in close cooperation with physicians, and future work will concentrate on the application for at home monitoring as sleep data is an area where unsupervised feature learning is a highly promising method for sleep stage classification as data is abundant and labels are costly to obtain.

Appendix

A. Features

A total of 28 features are used in this work.

Relative power for signal y in frequency band f is calculated as

$$y_{P_{\text{rel}}}(f) = \frac{y_P(f)}{\sum_{f=f_1}^{f_5} y_P(f)}, \tag{A.1}$$

where $y(f)_P$ is the sum of the absolute power in frequency band f for signal y. The five frequency bands used are delta (0.5–4 Hz), theta (4–8 Hz), alpha (8–13 Hz), beta (13–20 Hz), and gamma (20–64 Hz).

The median of the absolute value for EMG is calculated as

$$\text{EMG}_{\text{median}} = \text{median}\left(\sum_{k=1}^{N} |\text{EMG}(k)|\right). \tag{A.2}$$
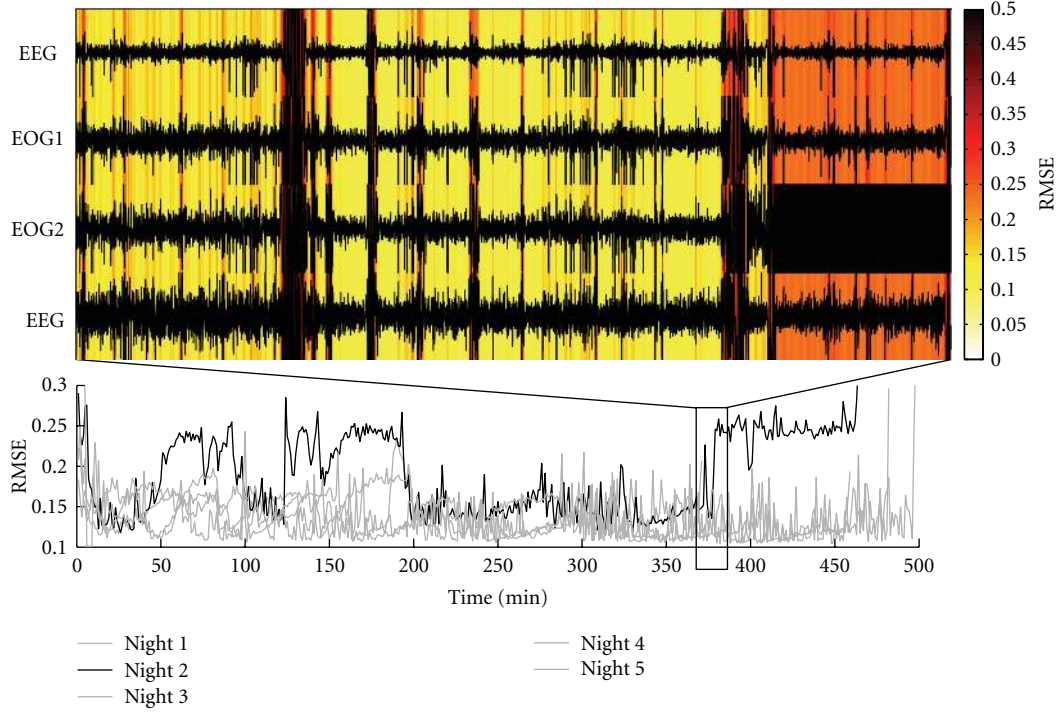
FIGURE 7: RMSE for five night runs recorded at home (bottom). Color-code of RMSE for night run 2 where the redder areas more anomalous areas of the signal. EOG2 falls off at around 380 minutes (top).

The eye correlation coefficient for the EOG is calculated as

$$\mathrm{EOG}_{\mathrm{corr}} = \frac{E\left[\left(y_1 - \mu_{y_1}\right)\left(y_2 - \mu_{y_2}\right)\right]}{\sigma_{y_1}\sigma_{y_2}}, \qquad (A.3)$$

where $y_1 = \mathrm{EOG}_{\mathrm{left}}$ and $y_2 = \mathrm{EOG}_{\mathrm{right}}$.

The entropy for a signal y is calculated as

$$y_{\mathrm{entropy}} = -\sum_{k=1}^{8} \frac{n_k}{N} \ln \frac{n_k}{N}, \qquad (A.4)$$

where N is the number of samples in signal y, and n_k is the number of samples from y that belongs to the kth bin from a histogram of y.

The kurtosis for a signal y is calculated as

$$y_{\mathrm{kurtosis}} = \frac{E[y - \mu]^4}{\sigma^4}, \qquad (A.5)$$

where μ and σ are the mean and standard deviation, respectively, for signal y.

The spectral mean for signal y is calculated as

$$y_{\mathrm{spectralmean}} = \frac{1}{F} \sum_{f=f_1}^{f_5} y(f)_{P_{\mathrm{rel}}} \cdot f, \qquad (A.6)$$

where F is the sum of the lengths of the 5 frequency bands.

Fractal exponent [35, 36] for the EEG is calculated as the negative slope of the linear fit of spectral density in the double logarithmic graph.

Normalization is performed for some features according to [37] and [30]. The absolute median for EMG is normalized by dividing with the absolute median for the whole EMG signal.
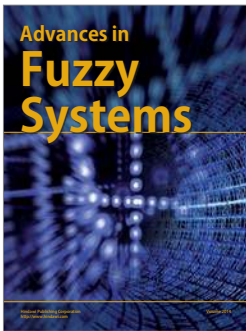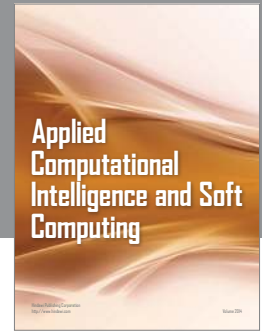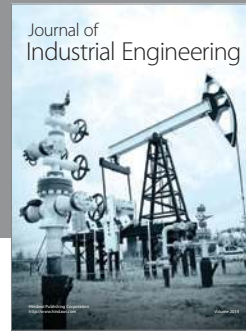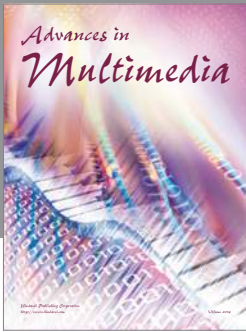
References

[1] K. Šušmákováemail and A. Krakovská, "Discrimination ability of individual measures used in sleep stages classification," *Artificial Intelligence in Medicine*, vol. 44, no. 3, pp. 261–277, 2008.

[2] A. Flexer, G. Gruber, and G. Dorffner, "A reliable probabilistic sleep stager based on a single EEG signal," *Artificial Intelligence in Medicine*, vol. 33, no. 3, pp. 199–207, 2005.

[3] L. Johnson, A. Lubin, P. Naitoh, C. Nute, and M. Austin, "Spectral analysis of the EEG of dominant and non-dominant

alpha subjects during waking and sleeping," *Electroencephalography and Clinical Neurophysiology*, vol. 26, no. 4, pp. 361–370, 1969.

[4] J. Pardey, S. Roberts, L. Tarassenko, and J. Stradling, "A new approach to the analysis of the human sleep/wakefulness continuum," *Journal of Sleep Research*, vol. 5, no. 4, pp. 201–210, 1996.

[5] N. Schaltenbrand, R. Lengelle, M. Toussaint et al., "Sleep stage scoring using the neural network model: comparison between visual and automatic analysis in normal subjects and patients," *Sleep*, vol. 19, no. 1, pp. 26–35, 1996.

[6] H. G. Jo, J. Y. Park, C. K. Lee, S. K. An, and S. K. Yoo, "Genetic fuzzy classifier for sleep stage identification," *Computers in Biology and Medicine*, vol. 40, no. 7, pp. 629–634, 2010.

[7] L. Zoubek, S. Charbonnier, S. Lesecq, A. Buguet, and F. Chapotot, "A two-steps sleep/wake stages classifier taking into account artefacts in the polysomnographic signa," in *Proceedings of the 17th World Congress, International Federation of Automatic Control (IFAC '08)*, July 2008.

[8] G. Luo and W. Min, "Subject-adaptive real-time sleep stage classification based on conditional random field," in *Proceedings of the American Medical Informatics Association Annual Symposium (AMIA '07)*, pp. 488–492, 2007.

[9] T. Penzel, K. Kesper, V. Gross, H. F. Becker, and C. Vogelmeier, "Problems in automatic sleep scoring applied to sleep apnea," in *Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (IEEE EMBS '03)*, pp. 358–361, September 2003.

[10] G. E. Hinton, S. Osindero, and Y. W. Teh, "A fast learning algorithm for deep belief nets," *Neural Computation*, vol. 18, no. 7, pp. 1527–1554, 2006.

[11] Y. Bengio, P. Lamblin, D. Popovici, and H. Larochelle, "Greedy layer-wise training of deep networks," in *Advances in Neural Information Processing Systems (NIPS '06)*, vol. 19, pp. 153–160, 2006.

[12] M. Ranzato, C. Poultney, S. Chopra, and Y. LeCun, "Efficient learning of sparse representations with an energy-based model," in *Proceedings of the Advances in Neural Information Processing Systems (NIPS '06)*, J. Platt, T. Hoffman, and B. Schölkopf, Eds., MIT Press, 2006.

[13] Y. Bengio and Y. LeCun, "Scaling learning algorithms towards AI," in *Large-Scale Kernel Machines*, L. Bottou, O. Chapelle, D. DeCoste, and J. Weston, Eds., MIT Press, 2007.

[14] Y. Bengio, "Learning deep architectures for AI," Tech. Rep. 1312, Department of IRO, Universite de Montreal, 2007.

[15] I. Arel, D. Rose, and T. Karnowski, "Deep machine learning—a new frontier in artificial intelligence research," *IEEE Computational Intelligence Magazine*, vol. 14, pp. 12–18, 2010.

[16] V. Nair and G. E. Hinton, "3-d object recognition with deep belief nets," in *Proceedings of the Advances in Neural Information Processing Systems (NIPS '06)*, 2006.

[17] G. Taylor, G. E. Hinton, and S. Roweis, "Modeling human motion using binary latent variables," in *Proceedings of the Advances in Neural Information Processing Systems*, 2007.

[18] N. Jaitly and G. E. Hinton, "Learning a better representation of speech sound waves using restricted boltzmann machines," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '11)*, 2011.

[19] M. Längkvist and A. Loutfi, "Unsupervised feature learning for electronic nose data applied to bacteria identification in blood," in *NIPS Workshop on Deep Learning and Unsupervised Feature Learning*, 2011.

[20] D. F. Wulsin, J. R. Gupta, R. Mani, J. A. Blanco, and B. Litt, "Modeling electroencephalography waveforms with semi-supervised deep belief nets: fast classification and anomaly measurement," *Journal of Neural Engineering*, vol. 8, no. 3, Article ID 036015, 2011.

[21] G. E. Hinton, "Training products of experts by minimizing contrastive divergence," *Neural Computation*, vol. 14, no. 8, pp. 1771–1800, 2002.

[22] A. Rechtschaffen and A. Kales, *A Manual of Standardized Terminology, Techniques and Scoring System for Sleep Stages of Human Subjects*, U.S. Government Printing Office, Washington DC, USA, 1968.

[23] M. Hirshkowitz, "Standing on the shoulders of giants: the Standardized Sleep Manual after 30 years," *Sleep Medicine Reviews*, vol. 4, no. 2, pp. 169–179, 2000.

[24] S. L. Himanen and J. Hasan, "Limitations of Rechtschaffen and Kales," *Sleep Medicine Reviews*, vol. 4, no. 2, pp. 149–167, 2000.

[25] L. R. Rabiner and B. H. Juang, "An introduction to hidden markov models," *IEEE ASSP Magazine*, vol. 3, no. 1, pp. 4–16, 1986.

[26] G. E. Hinton, *A Practical Guide to Training Restricted Boltzmann Machines*, 2010.

[27] S. Charbonnier, L. Zoubek, S. Lesecq, and F. Chapotot, "Self-evaluated automatic classifier as a decision-support tool for sleep/wake staging," *Computers in Biology and Medicine*, vol. 41, no. 6, pp. 380–389, 2011.

[28] A. Schlögl, C. Keinrath, D. Zimmermann, R. Scherer, R. Leeb, and G. Pfurtscheller, "A fully automated correction method of EOG artifacts in EEG recordings," *Clinical Neurophysiology*, vol. 118, no. 1, pp. 98–104, 2007.

[29] A. L. Goldberger, L. A. Amaral, L. Glass et al., "PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. E215–220, 2000.

[30] L. Zoubek, S. Charbonnier, S. Lesecq, A. Buguet, and F. Chapotot, "Feature selection for sleep/wake stages classification using data driven methods," *Biomedical Signal Processing and Control*, vol. 2, no. 3, pp. 171–179, 2007.

[31] G. Huang, H. Lee, and E. Learned-Miller, "Learning hierarchical representations for face verification with convolutional deep belief networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '12)*, 2012.

[32] D. Yu, L. Deng, I. Jang, P. Kudumakis, M. Sandler, and K. Kang, "Deep learning and its applications to signal and information processing," *IEEE Signal Processing Magazine*, vol. 28, no. 1, pp. 145–154, 2011.

[33] M. Ranzato, A. Krizhevsky, and G. E. Hinton, "Factored 3-way restricted boltzmann machines for modeling natural images," in *Proceedings of the 30th International Conference on Artificial Intelligence and Statistics*, 2010.

[34] J. Ngiam, A. Khosla, M. Kim, H. Lee, and A. Y. Ng, "Multimodal deep learning," in *Proceedings of the 28th International Conference on Machine Learning*, 2011.

[35] A. R. Osborne and A. Provenzale, "Finite correlation dimension for stochastic systems with power-law spectra," *Physica D*, vol. 35, no. 3, pp. 357–381, 1989.

[36] E. Pereda, A. Gamundi, R. Rial, and J. González, "Non-linear behaviour of human EEG: fractal exponent versus correlation dimension in awake and sleep stages," *Neuroscience Letters*, vol. 250, no. 2, pp. 91–94, 1998.

[37] T. Gasser, P. Baecher, and J. Moecks, "Transformations towards the normal distribution of broad band spectral parameters of the EEG," *Electroencephalography and Clinical Neurophysiology*, vol. 53, no. 1, pp. 119–124, 1982.