Analysis of sleep EEG activity during hypopnoea episodes by least squares support vector machine employing AR coefficients

Elif Derya Übeyli a,*, Dean Cvetkovic b, Gerard Holland c, Irena Cosic d

a TOBB Ekonomi ve Teknoloji Üniversitesi, Faculty of Engineering, Department of Electrical and Electronics Engineering, 06530 Söğütözü, Ankara, Turkey
b RMIT University, School of Electrical and Computer Engineering, GPO Box 2476V, Melbourne VIC 3001, Australia
c St. Luke’s Hospital, Sleep Centre, Sydney NSW, Australia
d RMIT University, Science, Engineering and Technology, GPO Box 2476V, Melbourne VIC 3001, Australia

A B S T R A C T

Keywords:
Least squares support vector machines
AR coefficients
Sleep apnoea hypopnoea
Electroencephalogram (EEG)

This paper presents the application of least squares support vector machines (LS-SVMs) for automatic detection of alterations in the human electroencephalogram (EEG) activities during hypopnoea episodes. The obstructive sleep apnoea hypopnoea syndrome (OSAHS) means “cessation of breath” during the sleep hours and the sufferers often experience related changes in the electrical activity of the brain and heart. Decision making was performed in two stages: feature extraction by computation of autoregressive (AR) coefficients and classification by the LS-SVMs. The EEG signals (pre and during hypopnoea) from three electrodes (C3, C4 and O2) were used as input patterns of the LS-SVMs. The performance of the LS-SVMs was evaluated in terms of training performance and classification accuracies and the results confirmed that the proposed LS-SVM has potential in detecting changes in the human EEG activity due to hypopnoea episodes.

© 2009 Elsevier Ltd. All rights reserved.

1. Introduction

Sleep is a vital physiological function and high quality sleep is essential for maintaining the good health. Sleep disorders however are among the most common disorders suffered by humans and it is rare for most people to regularly enjoy the amount of quality sleep they need. The behavioral and social causes of sleep disorders are typically the result of modern lifestyle, which are usually linked to obstructive sleep apnoea hypopnoea syndrome (OSAHS). Healthcare professionals and sleep researchers are currently looking for ways to improve the clinical diagnosis of OSAH sufferers. OSAH means “cessation of breath” (Vgontzas & Kales, 1999). The sufferer might experience related changes in the electrical activity of the brain and heart due to this cessation of breath. In sleep research, the analysis of sleep microstructure has been recognised substantially since the conventional parameters of sleep staging introduced by Rechtschaffen and Kales (1968), which covered only a limited part of sleep-related electroencephalogram (EEG) phenomena. A considerable number of computerised scoring systems have been introduced to attain a more standardised system of sleep EEG evaluation. However, the sleep EEG, in particular during the OSAH episodes often contains linear and nonlinear information as well as stochastic components (noise). The separation and evaluation of these signal components remains a problem not entirely solved. Therefore, new approaches to the detection and evaluation of sleep EEG transients during the OSAH episodes are required (Cvetkovic & Cosic, 2008, Cvetkovic, Übeyli, Holland, & Cosic, 2009).

Abnormalities in the EEG in serious sleep disorders are at times very difficult to be detected using conventional techniques. Such techniques work by transforming the mostly qualitative diagnostic criteria into a more objective quantitative signal feature classification problem. The techniques have been used to address this problem such as the analysis of EEG signals for detection of electroencephalographic changes using the autocorrelation function, frequency domain features, time frequency analysis, and model-based methods (Adeli, Zhou, & Dadmehr, 2003; Cvetkovic & Cosic, 2008; Cvetkovic, Holland, & Cosic, 2007, 2008, 2009; Hazarika, Chen, Tsai, & Sergejew, 1997; Rosso, Figliola, Creso, & Serrano, 2004; Übeyli, 2006, 2008a, 2008b, 2008c, 2009a, 2009b, 2009c, 2009d; Übeyli, Cvetkovic, & Cosic, 2008). The model-based methods (parametric) are based on modeling the data sequence \( x(n) \) as the output of a linear system characterized by a rational system. In the model-based methods, the spectrum estimation procedure consists of two steps. Given the data sequence \( x(n) \), \( 0 \leq n \leq N - 1 \), the parameters of the method are estimated. Then from these estimates, the power spectral density (PSD) estimate is computed. The model-based methods spectra have a good statistical stability for short segments of signal and have a good spectral resolution and the resolution is less dependent on the length of the
record (Kay, 1988; Kay & Marple, 1981; Proakis & Manolakis, 1996; Stoica & Moses, 1997). The results of the studies in the literature have demonstrated that the autoregressive (AR) method is one of the most promising method among the model-based methods (Kay, 1988; Kay & Marple, 1981; Proakis & Manolakis, 1996; Stoica & Moses, 1997). In this respect, in the present study the AR coefficients were used as the features representing the EEG signals. The AR coefficients were estimated by the Burg method.

A new approach based on the least squares support vector machines (LS-SVMs) employing the AR coefficients was presented for detecting any possible changes in the human EEG activity due to hypopnoea (mild case of cessation of breath) occurrences. The proposed technique involved training the LS-SVMs to classify the hypopnoea occurrences. The process involved training the LS-SVMs employing the AR coefficients was presented for classifying the hypopnoea (mild case of cessation of breath) occurrences. The proposed technique involved training the LS-SVMs to classify the hypopnoea occurrences. The process involved training the LS-SVMs employing the AR coefficients was presented for classifying the hypopnoea occurrences.

2. Data description

One human subject (age: 32, sex: male, weight: 96 kg, height: 1.76 m) was recruited for an overnight PSG recording at St. Lukes Hospital (Sydney, NSW, Australia). The subject was diagnosed with the AH1 (apnoea hypopnoea index) of 5.1.

The sleep polysomnographic (PSG) was recorded from 22:30 h until 05:00 the next day using bio–logic system and adult sleepcan vision analysis (Bio–Logic Corp., USA). The PSG data was recorded at 256 Hz sampling frequency. Surface electrodes were placed on the scalp's surface (C3, C4 and O2; 10–20 system) and referenced to bridged left and right mastoid to record the EEG activity. Two channels were used to record the eye movements, with one electrode placed 1 cm above and slightly lateral to the outer canthus of one eye and the second electrode recording the potentials from an electrode 1 cm below and slightly lateral to the outer canthus of the other eye. Both electrodes were referenced to left–right mastoid. The other electrodes recorded the electromyogram (EMG) from the muscle areas on and beneath the chin, electrocardiogram (ECG) (using lead-II across the chest area), nasal and oral airflow, snoring sounds, breathing effort (measured at the chest and abdomen), oxymetry, actigraphy recording body positioning and leg movements (right and left anterior tibialis).

The apnoea and hypopnoea events were visually scored by the sleep technician according to Chicago scoring criteria as set out in a report by the ASDA 1998 sleep stage scoring according (Rechtschaffen & Kales, 1968) from 30 s epochs. The analysis reported 64.5% sleep efficiency and 91.8% in sleep maintenance with 132 min spent in wake (W) stage, 30 min in stage 1 (S1), 125 in stage 2 (S2), 17 min in stage 3 (S3), 49 min in stage 4 (S4), 26 min in stage REM, 220 min non-REM and 3 min in movement time (Cvetkovic & Cosic, 2008; Cvetkovic et al., 2007; Cvetkovic et al., 2009).

3. Brief review of classifiers

3.1. Support vector machines (SVMs)

The SVM proposed by Vapnik (1995) has been studied extensively for classification, regression and density estimation. Fig. 1 shows the architecture of the SVM. The SVM maps the input patterns into a higher dimensional feature space through some nonlinear mapping chosen a priori. A linear decision surface is then constructed in this high dimensional feature space. Thus, SVM is a linear classifier in the parameter space, but it becomes a nonlinear classifier as a result of the nonlinear mapping of the space of the input patterns into the high dimensional feature space. Let m-dimensional training data be \( x(i = 1, \ldots, M) \) and their class labels be \( y \), where \( y = 1 \) and \( y = -1 \) for classes 1 and 2, respectively. If these input data are linearly separable in the feature space, then the following decision function can be determined:

\[
D(x) = w^T g(x) + b
\]  

(1)

where \( g(x) \) is a mapping function that maps \( x \) into the \( l \)-dimensional space, \( w \) is the \( l \)-dimensional vector and \( b \) is a scalar. To separate data linearly, the decision function satisfies the following condition:

\[
y_i(w^T g(x_i) + b) \geq 1 \quad \text{for} \quad i = 1, \ldots, M
\]  

(2)

If the problem is linearly separable in the feature space, there are an infinite number of decision functions that satisfy Eq. (2). Among them we require that the hyperplane has the largest margin between two classes. The margin is the minimum distance from the separating hyperplane to the input data and this is given by \( |D(x)|/||w|| \). Then we call the separating hyperplane with the maximum margin optimal separating hyperplane.

Assuming that the margin is \( \rho \), the following condition needs to be satisfied:

\[
y_iD(x) \geq \rho \quad \text{for} \quad i = 1, \ldots, M
\]  

(3)

The product of \( \rho \) and \( ||w|| \) is fixed:

\[
\rho ||w|| = 1
\]  

(4)

In order to obtain the optimal separating hyperplane with the maximum margin, \( w \) with the minimum \( ||w|| \) that satisfying Eq. (3) should be found. From Eq. (4), this leads to solving the following optimization problem. Minimizing

\[
\frac{1}{2}w^Tw
\]  

(5)

subject to the constraints:

\[
y_i(w^T g(x_i) + b) \geq 1 \quad \text{for} \quad i = 1, \ldots, M
\]  

(6)

When training data are not linearly separable, we introduce slack variables \( \xi_i \) into Eq. (2) as follows:

\[
y_i(w^T g(x_i) + b) \geq 1 - \xi_i, \quad \xi_i \geq 0 \quad \text{for} \quad i = 1, \ldots, M
\]  

(7)

The optimal separating hyperplane is determined so that the maximization of the margin and the minimization of the training error are achieved. Minimizing

![Fig. 1. Architecture of the SVM (N is the number of support vectors).](image-url)
subject to the constraints:
\[
y_i(w^Tg(x_i) + b) \geq 1 - \xi_i, \quad \xi_i \geq 0 \text{ for } i = 1, \ldots, M
\]  
(9)

where \( C \) is a parameter that determines the tradeoff between the maximum margin and the minimum classification error and \( p \) is 1 or 2. When \( p = 1 \), the SVM is called L1 soft margin SVM (L1-SVM), and when \( p = 2 \), L2 soft margin SVM (L2-SVM). In the conventional SVM, optimal separating hyperplane is obtained by solving the above quadratic programming problem. The kernel function enables the operations to be carried out in the input space rather than in the high-dimensional feature space. Some typical examples of kernel functions are \( K(u,v) = v^T u \) (linear SVM); \( K(u,v) = (v^T u + 1)^n \) (polynomial SVM of degree \( n \)); \( K(u,v) = \exp(-|u - v|^2/2\sigma^2) \) (radial basis function – RBF SVM); \( K(u,v) = \tanh(\nu v^T u + \theta) \) (two layer neural SVM) where \( \sigma, k, \theta \) are constants (Cortes & Vapnik, 1995; Vapnik, 1995). However, a proper kernel function for a certain problem is dependent on the specific data and till now there is no good method on how to choose a kernel function. In this study, the choice of the kernel functions was studied empirically and optimal results were achieved using RBF kernel function.

3.2. Least squares support vector machines (LS-SVMs)

In contrast to the SVM, the LS-SVM is trained by minimizing
\[
\frac{1}{2} w^T w + \frac{C}{2} \sum_{i=1}^{M} \xi_i^2
\]  
(10)

subject to the equality constraints:
\[
y_i(w^Tg(x_i) + b) = 1 - \xi_i \text{ for } i = 1, \ldots, M
\]  
(11)

In the LS-SVM, equality constraints are used instead of inequality ones employed in the conventional SVM. Therefore, the optimal solution can be obtained by solving a set of linear equations instead of solving a quadratic programming problem. To derive the dual problem of Eqs. (10) and (11), the Lagrange multipliers are introduced as follows:
\[
Q(w, b, \alpha, \xi) = \frac{1}{2} w^T w + \frac{C}{2} \sum_{i=1}^{M} \xi_i^2 - \sum_{i=1}^{M} \alpha_i (y_i w^T g(x_i) + b) - 1 + \xi_i
\]  
(12)

where \( \alpha = (\alpha_1, \ldots, \alpha_M)^T \) is Lagrange multipliers, which can be positive or negative in case of LS-SVM formulation. The conditions for optimality are derived by differentiating the above equation with respect to \( w, \xi_i, b \), and \( \alpha_i \) and equating the resulting equations to zero. The detailed information can be found in the papers presented by Suykens and Vandewalle (1999) and Tsujinishi and Abe (2003).

4. Experimental results

4.1. Computation of AR coefficients

In this study, the AR PSD estimations of the EEG signals (pre and during hypopnoea) recorded from C3, C4, O2 EEG electrodes were obtained. The PSDs describe the distribution of power with frequency. The sample AR PSDs of the EEG signals (pre and during hypopnoea) recorded from C3, C4, O2 EEG electrodes were obtained. The PSDs describe the distribution of power with frequency. The sample AR PSDs of the EEG signals (pre and during hypopnoea) recorded from C3, C4, O2 EEG electrodes were presented in Figs. 2–4. The AR equation may model spectra with narrow peaks by placing zeroes close to the unit circle. This is an important feature since narrowband spectra are quite common in practice. In addition, the estimation of parameters in the AR signal models is a well-established topic; the estimates are found by solving linear equations of the system. There fore, the AR method was employed in computation of the features. The Burg method was used in order to estimate the AR PSDs. Therefore, the Burg AR coefficients were used as the features representing the signals under study.

The selection of the model orders in the model-based spectral estimators is a critical subject. Too low order results in a smoothed estimate, while too large order causes spurious peaks and general statistical instability. In the case of the dimension of autocorrelation matrix is inappropriate and the model orders chosen incorrect, poor spectral estimates are obtained by the model-based spectral estimators. In this study, Akaike information criteria (Akaike, 1974) was taken as the base for choosing the model order. The model order was taken as 10 for the AR method. The mean values of the Burg AR coefficients of the EEG signals (pre and during hypopnoea) recorded from C3, C4, O2 EEG electrodes, which were used to train the LS-SVM, are presented in Tables 1–3. From Figs. 2–4 (AR PSDs of the signals) and Tables 1–3 (AR coefficients), one can see that the extracted features of the EEG signals (pre and during hypopnoea) are different from each other. Therefore, we decided...
that the AR coefficients can serve as useful parameters in detecting changes in the human EEG activity due to hypopnoea episodes.

4.2. Experiments for implementation of LSSVM

The LS-SVM proposed for classification of the EEG signals were implemented by using the MATLAB software package (MATLAB version 7.0 with neural networks toolbox). The architecture and the training process are important for the neural networks used in classification. The adequate functioning of neural networks depends on the sizes of the training set and test set. In this study, the 11 time series of 3725 samples for each class windowed by a rectangular window composed of 256 discrete data and then training and test sets of the LS-SVM were formed by 160 vectors (80 vectors from each class) of 11 dimensions (dimension of the extracted feature vectors). The 80 vectors (40 vectors from each class) were used for training and the 80 vectors (40 vectors from each class) were used for testing.

The LS-SVM used the RBF kernel functions. One has to assume a value for \( r \) for the implementation of the LS-SVMs with the RBF kernel functions. The optimal \( r \) can only be found by systematically varying its value in the different training sessions. The support vectors were extracted from the training data file with an assumed \( r \) value. After the support vectors have been found and LS-SVM constructed, the model was applied to 1/3 of the evaluation data set to compute the misclassification rate. The \( r \) value was varied between 0.1 and 0.6, at interval of 0.1. The \( r = 0.4 \) resulted in the minimum misclassification rate was thus chosen.

4.3. Classification errors and ROC analysis

Confusion matrix is used to display the classification results of the classifiers. In a confusion matrix, each cell contains the raw number of exemplars classified for the corresponding combination of desired and actual network outputs. The confusion matrix showing the classification results of the LS-SVM is presented in Table 4. From this matrix, one can tell the frequency with which an EEG segment (pre and during hypopnoea) is misclassified as another.

### Table 3

<table>
<thead>
<tr>
<th>AR parameters (model order = 10, EEG signals (pre and during hypopnoea) recorded from O2 EEG electrode).</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR parameters</td>
</tr>
<tr>
<td>a[1]</td>
</tr>
<tr>
<td>a[2]</td>
</tr>
<tr>
<td>a[3]</td>
</tr>
<tr>
<td>a[4]</td>
</tr>
<tr>
<td>a[5]</td>
</tr>
<tr>
<td>a[6]</td>
</tr>
<tr>
<td>a[7]</td>
</tr>
<tr>
<td>a[8]</td>
</tr>
<tr>
<td>a[9]</td>
</tr>
<tr>
<td>a[10]</td>
</tr>
<tr>
<td>( \sigma^2 )</td>
</tr>
</tbody>
</table>

4.4. Classification errors and ROC analysis

Confusion matrix is used to display the classification results of the classifiers. In a confusion matrix, each cell contains the raw number of exemplars classified for the corresponding combination of desired and actual network outputs. The confusion matrix showing the classification results of the LS-SVM is presented in Table 4. From this matrix, one can tell the frequency with which an EEG segment (pre and during hypopnoea) is misclassified as another.

### Table 4

Confusion matrix.

<table>
<thead>
<tr>
<th>Desired result</th>
<th>Output result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-hypopnoea EEG signals</td>
<td>During-hypopnoea EEG signals</td>
</tr>
<tr>
<td>37</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>39</td>
</tr>
</tbody>
</table>

### Table 5

The values of the statistical parameters.

<table>
<thead>
<tr>
<th>Statistical parameters (%)</th>
<th>Specificity</th>
<th>Sensitivity</th>
<th>Total classification accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Specficity</td>
<td>92.50</td>
<td>97.50</td>
<td>95.00</td>
</tr>
</tbody>
</table>
The test performance of the classifiers can be determined by the computation of sensitivity, specificity and total classification accuracy. The sensitivity, specificity and total classification accuracy are defined as:

**Sensitivity:** Number of true positive decisions/number of actually positive cases.

**Specificity:** Number of true negative decisions/number of actually negative cases.

**Total classification accuracy:** Number of correct decisions/total number of cases.

Table 5 presents the values of the statistical parameters (sensitivity, specificity and total classification accuracy). Receiver operating characteristic (ROC) plots provide a view of the whole spectrum of sensitivities and specificities because all possible sensitivity/specificity pairs for a particular test are graphed. The performance of a test can be evaluated by plotting a ROC curve for the test and therefore, we used ROC curves to describe the performance of a test can be evaluated by plotting a ROC curve for the test. The ROC curve, which is shown in Fig. 5, demonstrates the performance of the ROC-SVM on the test files. The classification results presented in Table 5 and Fig. 5 (classification accuracies and ROC curve) demonstrate that the ROC-SVM obtained high accuracies. The total classification accuracy of the ROC-SVM was 95.00%.

5. Conclusions

The EEG signals (pre and during hypopnoea) from three electrodes (C3, C4 and O2) were considered as a classification problem with the AR coefficients. The features defining the EEG signals were computed by the Burg AR method. The extracted features (Burg AR coefficients) were used as inputs of the LS-SVM. The classification accuracies and the ROC curves of the classifiers were used in evaluation of the performance of the classifier. The classification results and the values of statistical parameters indicated that the proposed combined AR coefficients/LS-SVM had high accuracy in discriminating the EEG signals (the total classification accuracy was 95.00%).

References


