Journal of mathematics and computer Science 14 (2015) 33 - 41

Contents list available at JMCS

Journal of Mathematics and Computer Science

Journal Homepage: www.tjmcs.com

An Expert Clinical System for Diagnosing Obstructive Sleep Apnea with Help from the XCSR Classifier

Ehsan Sadeghipour[1,*], Ahmad Hatam[2], Farzad hosseinzadeh[1,+]

[1]Sama technical and vocational training college, Islamic Azad University,
Bandar Abbas Branch, Bandar Abbas, Iran,
[*]*ehsan61.ai@gmail.com*- **(**the corresponding author**)**
[+]*fhosseinzadeh05@gmail.com*
[2]University of Hormozgan, Faculty of Power and Computer Engineering,
University Hormozgan, Bandar Abbas, Iran,
a.hatam@hormozgan.ac.ir

Article history:
Received July 2014
Accepted October 2014
Available online November 2014

Abstract

Obstructive sleep apnea is a common condition with serious neural-psychological complications and cardiovascular problems if not diagnosed and treated in time. Despite the importance of this disease in our country, it has not received much attention and there are few centers for evaluating patients suffering from it. In this article, an intelligent method is introduced for diagnosing obstructive sleep apnea that uses features extracted from changes in heart rate and respiratory signals in the ECG as input for training and testing the modified XCS classifier system. Comparison of results obtained from implementing the mentioned method with those of other methods on physionet database showed desirable performance and high accuracy of the proposed system in diagnosing obstructive sleep apnea.

Keywords: Apnea, fuzzy neural network, SVM, KNN, XCSR

1. Introduction

Obstructive sleep apnea and insomnia are both among the most prevalent types of sleep disorders having destructive effects on the health and performance of the person afflicted with them. Concurrence of these two disorders in patients causes problems in their diagnosis [1]. Obstructive sleep apnea refers to a disease in which recurrent attacks obstruct breathing passages of people for more than 10 seconds while they are asleep. This disease is more prevalent in middle age, and studies have indicated up to 4% of men and 2% of women suffer from it. Research has shown people who are afflicted with acute obstructive sleep apnea, and who have not received treatment, are four times more likely to die of all factors than people who are free of this disorder. The reason for the occurrence of this disease is that muscle tone decreases during sleep and, hence, muscles that keep the airways open relax too. In some people, this muscle relaxation leads to a narrowing of the air passages following which snoring takes place. If this muscle relaxation results in a complete blockage of air passages, breathing cessation (apnea) happens [2], [3]. In this disease, the central system in the brain controlling breathing is disrupted and breathing cessation occurs during apnea without any attempt to inhale or exhale [4]. Obstructive sleep apnea has very important complications including the sudden decline in blood oxygen when breathing stops, which will raise blood pressure and put a strain on the blood circulation system increasing the risk of cardiovascular problems too. When sleep apnea deprives the person from a pleasant night sleep, it will result in fatigue and malaise during the day. Research has shown chances for an accident to happen to a person suffering from sleep apnea are six times greater than for those who are not afflicted with it. Rising pulmonary arterial pressure, arrhythmia, ischemia, following which heart attack, stroke, and some metabolic diseases happen, are common complications of this disease [5]. Obesity, temporomandibular disorders, snoring during sleep, old age, being male, race, smoking, alcohol, etc. are risk factors for developing this disease. Monitoring breathing and chest movements, performing electromyography and oxygen saturation test, and observation of other vital signs during sleep at night, are sufficient for making a definitive diagnosis of sleep apnea. Sleep studies are carried out in laboratories equipped with modern technology based on polysomnography (PSG), with patients under supervision of specialists all night long. However, this kind of study is not convenient, is very costly, and cannot be conducted everywhere. Therefore, techniques of more accurate diagnosis of sleep apnea that involve simple measurements, have low costs, and enjoy suitable accuracy and correctness, seem to be necessary. At present, many methods are used to reduce the number of signals and to extract optimal features from these signals. For example, evaluation of apnea based on electrocardiogram signals is an optimized method for diagnosing this disease [6].

2. Wavelet packet transform

As shown in Figure 1, in a discrete wavelet transform, a signal is divided into two parts. The coefficients produced by the low-pass filter are called coarse coefficients and those produced by high pass filter are called detail coefficients.

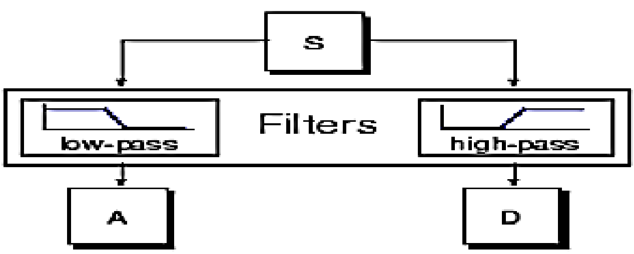

Figure 1: Discrete wavelet transform

A wavelet packet transform is a generalized wavelet decomposition in which the details can also be decomposed as the approximations. The figure below shows decomposition of a signal using the wavelet packet at level two. For example, the signal S can be shown as $AA_2 + DA_2 + D_1$ by using the wavelet packet, while the ordinary wavelet cannot display the S signal in this way

3. Various classification methods for diagnosing diseases

Task automation, and the role played by pattern recognition in realizing it, is one of the most important reasons why pattern recognition enjoys a special status in new research on diagnosing various diseases. Designing pattern recognition systems requires the type of data from systems that describes them well. Therefore, those features of diseases are considered that can be used to differentiate diseases belonging to various categories.

3.1. The KNN algorithm

The KNN algorithm, also called the nearest neighbor algorithm, is one of the instance-based learning algorithms that store only the training instances during the learning phase. To characterize the class an instance belongs to, the KNN algorithm calculates the distance between the instance and the other training instances. The most common criterion for calculating such distances is the Euclidean criterion, although other criteria such as Manhattan Minovsky are used for this purpose too. After calculating the distance, a majority vote is taken of the k nearest training instances and the label of the majority is allocated to the current test instance.

3.2. The SVM algorithm

SVM is a method for classifying linear and nonlinear data. This algorithm is one of the supervised learning methods and is used for classification and regression. Linear classification of data is the basis on which the SVM classifier works. In this method, a nonlinear mapping is used first to convert the initial data to data with higher dimensions, and a search is then made in the new dimensions to find the best separating hyperplane. This hyperplane is a decision boundary that separates records of one class from those of the other classes. Linear data classification tries to select a line with a higher confidence interval. To solve a problem, an optimum line must be found for the data by using QP methods, which are known methods for solving problems with restrictions. Prior to linear classification, the data must be mapped onto higher dimensional spaces by φ functions so that the machine can classify very complex data. Equation 1 is the relation for the hyperplane that classifies the data. Equations 2 and 3 are relations for parallel hyperplanes based on maximum margins. If we plot these simple functions, the distance between two margin planes will be equal to $\frac{2}{||w||}$:

$$w\times x - b = 0 \qquad (1)$$

$$w\times x - b = -1 \qquad (2)$$

$$w\times x - b = 1 \qquad (3)$$

In the above relations, x is the input variable, w the normal vector of the separating line, and b the y-intercept of the separating line. In this research, linear functions were used to reduce the number of calculations.

3.3. Adaptive neuro fuzzy inference systems

The fuzzy system is a system based on if-then rules that cannot be analyzed by classic probability theories. The purpose in fuzzy logic is to extract accurate results by using a set of rules defined by expert and specialist people. On the one hand, neural networks have the ability of learning and being trained and can use observed data to determine network parameters in a way that the desired output is achieved with the desired input. At the same time, neural networks do not have the ability to use human knowledge: that is, they cannot use language sentences to make inferences as fuzzy systems do. Therefore, adaptive neuro fuzzy inference networks were introduced in 1999 to achieve the learning ability of neural networks and the inference capability of fuzzy systems in the TSK fuzzy model. These networks, while having the learning ability of neural networks and the inference capability of fuzzy systems, are able to find any nonlinear model or mapping that can accurately relate the inputs to the outputs. Therefore, the ANFIS is a multi-layer neural network based on fuzzy systems with the structure (shown in Figure 1) in which the circles represent fixed nodes and the squares represent adaptive nodes. In the first layer, all the nodes are adaptive and the degrees of fuzzy membership of the inputs are the outputs of this layer.

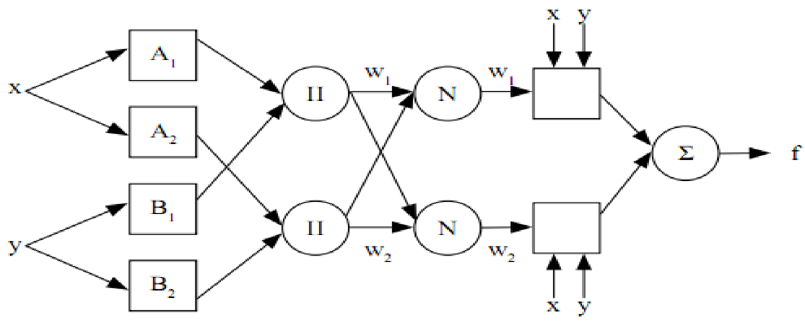

Figure 1: Adaptive neuro fuzzy inference system

3.4. The XCSR learning system

Machine learning refers to a wide range of supervised and unsupervised learning algorithms the purpose of which in data mining is to avoid exhaustive search of data and to replace this kind of time-consuming search with intelligent methods that make clustering data or modeling data behavior very simple through finding the pattern that exists in the data. During the past two decades, many methods have been introduced in the realm of data mining in which various supervised, unsupervised, or reinforcement-learning algorithms are used for purposes such as recognition and allocation of patterns. Classifier systems are among the most successful of these methods. Generally, classifier systems include a set of rules with “if-then” format each of which presents a potential solution for the target problem. A reinforcement-learning mechanism gradually evaluates this set of rules and updates it at specific time intervals with the help of a genetic algorithm. During the course of this gradual evolution, the system learns the behavior of the environment and then, in the application phase, presents suitable answers to the queries raised by the user. The first classifier system, called Learning Classifying Systems (LCS), was introduced by Holland in 1976. In this system, the value of each rule was evaluated by an index called “strength.” The strength of a rule increased in proportional to the degree of its correct answering to training instances in the framework of standards for reinforcement-learning and, at specific time intervals, an evolutionary algorithm in search (usually a genetic algorithm) was responsible for producing new rules and for omitting inefficient ones. At the end of the training stage, this set of rules had the relative ability to present acceptable solutions when faced with new queries. At the same time, successful performance of the LCS was contingent on selecting appropriate values for control parameters of the system, which depended on the type of experience the system designer had. Since the introduction of LCS, other types of classifier systems have been proposed including the Extended Classifier Systems (XCS). Prior to the introduction of XCS in 1995, these systems had very limited ability in obtaining appropriate answers but, since their introduction, classifier systems have gradually developed into more intelligent and more accurate factors and it is now believed that XCS and their improved versions can solve complex problems without requiring parameter adjustment. With the introduction of the classifier system having continuously varying variables (XCSR), some of the innate weaknesses of binary classification systems such as the inability to introduce specific intervals of variable values were mostly resolved. Now, these systems have been accepted as one of the most successful learning agents for solving data mining problems in semi-observable environments. In this method, the (limited) set of training data is used as usual for modifying the attributes of the rules (including “prediction,” “prediction error,” and “fitness.”) To do this, the following relations are used:

Updating prediction and prediction error:

If $expi < 1/\beta$ then $Pi = Pi + (R-Pi) / exp_i$, $\varepsilon i = \varepsilon i + (|R-Pi| - \varepsilon i) / exp_i$

If $expi \geq 1/\beta$ then $Pi = Pi + \beta (R-Pi)$, $\varepsilon i = \varepsilon i + \beta (|R-Pi| - \varepsilon i)$

Updating fitness:

If $\varepsilon i < \varepsilon 0$ then $ki=1$

If $\varepsilon i \geq \varepsilon 0$ then $ki = \beta (\frac{\varepsilon i}{\varepsilon 0})^{-\gamma}$

$Fi = fi + \beta [ki / \sum kj) - fi]$

In the above relations, β is the learning rate, γ the power of the accuracy rule, ε prediction error, exp the experience rule, P the prediction rule, R the reward received from the environment, k the accuracy rule, and f its fitness. The index i is the number of the rule in the rule set. In the next stage, the “random selection of residuals” i method is used to select many pairs from among strings that represent the CONDITION part of the new data, and the CONDITION part of the new data is applied on these strings of parents using the medial crossover method ii, to expand variety in the data set. In this method, the value of each conditional variable is obtained from relation 4:

$$a_i = \alpha(a_i^F) + (1-\alpha)(a_i^M) \qquad (4)$$

In the above relationship, a_i is the value of the ith conditional variable in the new data, a_i^F the value of the ith conditional variable in the first parent (father), a_i^M the value of the ith conditional variable in the second parent (mother), and α the participation coefficient of the parents that is determined adaptively. A nonlinear mapping from the conditional variable space to the ACTION space, that has been created using the available data, produces the ACTION part of the new data too. Creating variety in the available data continues until the learning cessation condition (for example, until the percentage of correct answers given by the system to the test data reaches a predetermined threshold) is satisfied with the help of the completed data [7].

4. Methodology

4.1. The data used

The data used in this research was taken from the physionet site storing nightly ECG recordings of normal people and patients afflicted with obstructive sleep apnea for the cardiology contest of the year 2000. The purpose of this contest was detecting and diagnosing apnea based on a lead from the ECG signal [8]. The electrocardiogram data was obtained with 100 Hz sampling frequency and 16- bit resolution. The data was that of women and men of 27-60 years of age weighing 53-135 kilograms. Every recording was of 8-hour duration and included information related to occurrence or nonoccurrence of apnea. This information was prepared from reports of experienced people and specialists based on breathing signals that had been simultaneously recorded.

4.2. Extracting and selecting features

The question of selecting features is one of the issues involved in machine learning and in studying problems related to pattern recognition. Attribute selection is in fact selecting attributes that have the maximum ability in predicting output [9]. Optimal subsets depend on problems we intend to solve [10]. To extract time-frequency features using packet wavelet too, the available data was analyzed to level 4. After this analysis, all the signals obtained at various levels were used to extract features. However, not all extracted features included important and required information, the volume of information presented by the discrete wavelet coefficients was very large, and it was, generally, very difficult to process the whole data and, specifically, impossible to apply it on intelligent systems. Therefore, effective features had to be selected that offered more information about the condition of the obstructive sleep apnea. For this purpose, an algorithm of feature selection called the IDE was selected in this research. To select the features using the time and frequency domain method, the threshold limit of 0.6 was used for the IDE algorithm. Three maximum wavelet coefficients at all four levels, the variance of wavelet coefficients at all four levels, and the standard deviations of wavelet coefficients at each level, were used as the input of the expert clinical system. The output of the proposed system was the differentiation of the set of instances into the two classes of apnea and non-apnea.

4.3. Differentiation using the KNN method

This method finds k nearest neighbors of the test data and allocates it to a group having the largest number of data items among the k nearest neighbors. In this method of classification, the value of k must first be determined. Mean accuracies obtained from 10 runs of the introduced KNN algorithm are listed in Table 1 below. For k=9, the mean accuracy obtained was better compared to other k values.

Table 1: Accuracy of diagnosis based on various K values

K value	5	7	9	11	13	15
Mean accuracy	76.5%	77.2%	77.5%	74.9%	75.8%	75.5%

4.4. Differentiation using the SVM method

The SVM method can only divide data into two groups and, therefore, this method was used to find a solution for the problem, as there were two groups in it. To implement this algorithm, a Gaussian radial basis function was employed as the kernel to make nonlinear classification possible too.

4.5. Differentiation using a neuro-fuzzy network

Before training the system, data was clustered using the SUBTRACTIVE method in which six Gaussian functions are considered for each input. Then the system was trained using the combination method. Input data was first mapped onto the [-1, 1] interval to obtain better results, because in this way the data is normalized and this prevents the input that has high values compared to others from becoming dominant, and allows for identical effects of the inputs at the start of network training. Of course, during the stages of training the network, the effects of the inputs will gradually change. Weights were given to the network inputs so that inputs that were more effective had greater weights and, hence, had more effects on network results and on improving the neuro-fuzzy network. To do this, the data was first normalized and the weight of each network input was then obtained using the IDE method. The values are listed in Table 2.

Table 2: Weight values of ANFIS network inputs

Input number	1	2	3	4
Weight value	0.717	0.638	1	0.839

After giving weights to the inputs, network inputs were differentiated once again and data classification was carried out for the new ANFIS network.

4.6. Differentiation using the XCSR method

One hundred sets of rules were produced completely at random by the XCSR classifier system to recognize features extracted from the IDE algorithm. They included 12 rules for conditions and one for the result of the conditions. Twenty of the 30 data items available were considered as the training data and 10 as the test data. Table 3 shows the comparison between results of the different classification methods for diagnosing the obstructive sleep apnea disease using the above- mentioned algorithms in the two sections of computational costs and the percentage of correct answers to the test data. For test results to be in identical conditions at the time the calculations were made, all systems used laptops with the specifications of CPU=2 and 4 GB RAMs.

Table 3: Efficiencies of various classifier systems in diagnosing obstructive apnea

Method	Computational cost in minutes	Percentage of correct answers
SVM	4	80.2%
ANFIS	6	87.9%
KNN	3.5	77.5%
XCSR	4	89.8%

5. Conclusions

A new method based on the XCSR classifier learning system was introduced in this research because polysomnography is costly, time-consuming, and requires a standard sleep clinic with experiences staff, and also because it is necessary to make correct diagnoses. Comparison of this method with some of the other machine learning algorithms showed it was superior to them and created this hope that doctors can use this decision support system, together with their specialty and experience, to make more accurate and better diagnosis of this disease.

References

[1] Chervin RD, "*Sleepiness, fatigue, tiredness, and lack of energy in obstructive sleep apnea*", Chest (2000), 118: 372-379.

[2] Goncalves MA, Paiva T, Ramos E, Guilleminault C, "*Obstructive sleep apnea syndrome, sleepiness and quality of life*". Chest (2004), 125: 2091-2096.

[3] World Health Organization Technical report series 894 ”*Obesity: preventing and managing the global epidemic* ’’.Geneva, World Health Organization, (2000).

[4] *www.medscape.com.*

[5] J. L. Lattimore, D. S. Celermajer, and I. Wilcox , “*Obstructive Sleep Apnea and Cardiovascular disease*”, Journal of the American College of Cardiology, (2003).

[6] Chazal, P. D., Penzel, T., and Heneghan, C., “*Automated detection of obstructive sleep apnoea atdifferent time scales using the electrocardiogram*” Physiological Measurement, (2004) pp. 967-983.

[7] M. Shariat Panahi, N. Moshtaghi Yazdani, “*An Improved XCSR Classifier System for Data Mining with Limited Training Samples*”, Global Journal of Science, Engineering and Technology, (ISSN: 2322-2441) Issue 2, (2012), pp. 52-57.

[8] *htpp://www.physionet.org/physiobank/database*

[9] Jensen R, “*Combining rought and fuzzy sets for feature selection*”, PhD. Thesis, School of informatics, Univ Edinburgh, (2005)

[10] Newman D. J., Hettich S, Blak C. L. S, Merz C., UCI “*repository of machine learning database*” Irvine, CA: University of California, Dep of Information and Computer Science archive (1998) icsuciedu/ml/database/pima +indian+database.