32nd Annual International Conference of the IEEE EMBS
Buenos Aires, Argentina, August 31 - September 4, 2010

Automatic Annotation of Actigraphy Data for Sleep Disorders Diagnosis Purposes

Alexandre Domingues, Ondrej Adamec, Teresa Paiva (MD) and J. Miguel Sanches

Abstract— **The diagnosis of Sleep disorders, highly prevalent in the western countries, typically involves sophisticated procedures and equipments that are intrusive to the patient. Wrist actigraphy, on the contrary, is a non-invasive and low cost solution to gather data which can provide valuable information in the diagnosis of these disorders. The acquired data may be used to infer the *Sleep/Wakefulness* (SW) state of the patient during the circadian cycle and detect abnormal behavioral patterns associated with these disorders . In this paper a classifier based on *Autoregressive* (AR) model coefficients, among other features, is proposed to estimate the SW state. The real data, acquired from 23 healthy subjects during fourteen days each, was segmented by expert medical personal with the help of complementary information such as light intensity and *Sleep e-Diary* information.**

Monte Carlo tests with a *Leave-One-Out Cross Validation* (LOOCV) strategy were used to assess the performance of the classifier which achieves an accuracy of 96%.

I. Introduction

Normal sleep circadian patterns are fundamental for regular and healthy conditions [1]. Sleep disorders affect both adult and young population and can be associated with diabetes, obesity, depression and cardiovascular diseases. An accurate diagnosis of this type of disorders is usually performed with *polysomnography* (PSG) which involves complex hardware, is uncomfortable to the patient and is usually done in clinical facilities. These highly constrained conditions prevent its use in a non intrusive way in normal daily life conditions.

The actigraphy data, obtained with non invasive and highly portable actigraph sensors, which reflect the motor activity of the subjects, has become a popular method in sleep studies [2], [3], [4], [5], [6] due to its ability to register behavioral data under normal life conditions.

Although some of the proposed methods to infer the SW state from actigraphy report accuracy performances ranging from 80% to 90% in certain conditions [2], [5], [4], robustness and reproductibility improvements are possible and desired [7]. The accuracy of actigraphy based methods is highly subject dependent. It was shown that in cases of young and children populations the actigraph is quite acurate [7] but in cases of adult insomnia the actigraph is prone to detect false sleep states mainly because adult people tend to remain still while awake [8].

The main weakness of actigraphy method, strongly related with its dependence on the movement of the patient, is more severe on populations that present fragmented sleep periods [9], [10]. Although actigraphy measurements may occasionally report false wakefulness states, *e.g.* in periodic limb movement disorders [11], the false sleep detections rise the major concerns [9].

In this paper *auto-regressive* (AR) coefficients, among other features, estimated from the actigraphy data, are used to discriminate the sleep and wakefulness states and automatically identify the corresponding periods of the circadian cycle of the sleep. The real data, acquired from 23 healthy subjects during fourteen days each, was segmented by trained medical staff with help of complementary information such as light intensity and *Sleep e-Diary* information. Monte Carlo tests with a *Leave-One-Out Cross Validation* (LOOCV) strategy were used to assess the performance of the classifier which achieves an accuracy of 96%.

A. State of the art

Actigraphy is a well established technique used in different scopes and approaches. In [5], four different scoring algorithms, two based on thresholding and the other two based on regression analysis are used to detect wakefulness states in three different conditions. In [4], actigraphy was used to study sleep in pregnant women and in [3] a statistical characterization of wrist actigraphy is proposed by using events detection. In this method the time between successive groups of movements and the number of movements at each fixed time measurement epoch are registered and used to detect abnormal movement patterns during the sleep state. In order to overcome the limitations of actigraphy based methods, mixed techniques are combined, such as the one described in [12] where video is combined with actigraphy to detect wakening episodes in children.

B. Paper organization

The paper is organized as follows: Section I introduces and motivates the problem and the state of the art is revised. The problem is formulated in Section II, where the used data is described and the mathematical tools and algorithms are presented. In Section III the numerical results are presented along with it's discussion. Finally, conclusions are drawn in Section IV.

This work was supported by the Portuguese Government – FCT (ISR/IST plurianual funding).

Alexandre Domingues and João Sanches are with Institute for Systems and Robotics / Instituto Superior Técnico. adomingues@gmail.com and jmrs@ist.utl.pt

Ondrej Adamec is with Institute for Systems and Robotics / Instituto Superior Técnico and Faculty of Electrical Engineering and Computer Science,VSB - TU Ostrava. ondrej.adamec@vsb.cz

Teresa Paiva is with Faculdade de Medicina da Universidade de Lisboa.

II. Problem formulation

Visual inspection of the actigraphy data, displayed in Fig. 2, clearly shows differences between the day and night periods. The goal of this paper is to discriminate these periods based on the characterization of the actigraphy data mainly based on the coefficients of AR models fitted to the data.

A. Experimental data

The data used in this paper was collected from 23 healthy subjects, data from non-healthy subjects was also available. The patients wore an actigraph from *Somnomed* on the non dominant wrist for a period of approximately 14 days. The actigraph sensor is mainly composed by a a 3D axis accelerometer and the output signal is the mean magnitude of the acceleration in each epoch, set to 1 minute in these experiments.

The data obtained from the 23 healthy patients was manually segmented into wakefulness and sleep periods by medical trained personal through visual inspection, in approximately 350 segments. The segmentation of data from healthy patients is easier and may be done by using simple criteria to identify the *sleep/wakefulness* (SW) binary state, because it is much more predictable than in subjects with sleep disorders. Periods corresponding to day time, with the obvious exceptions of naps during the day, and continued movements during the night longer than one minute were classified in the wakefulness class, $SW(t) = 0$. All other periods, with the exception of small transition intervals, were classified in the sleep class, $SW(t) = 1$, where t is the sample index. The extra information such as light intensity and *sleep e-Diary* (SeD) data was used as a complement in the classification process.

B. Auto-regressive model

In order to extract statistical properties of the data, AR models [13] were fitted to limited segments of the actigraphy data. A p-order AR model assumes that a given sample may be expressed as a linear combination of the previous p samples plus a residue

$$x(n) = \sum_{k=1}^{p} a_k x(n-k) + \epsilon[n] \tag{1}$$

From each segment, containing only wakefulness or sleep data, W overlapped windows were extracted. The size of each window (window-length), N, and the shift between them (window-step) were initially set to 150 and 60 minutes respectively but these values were later adjusted in order to achieve optimal results. A p-order AR model was fitted to each one of these windows, leading to a $p \times W$ sized matrix of coefficients per segment.

The order of the model, p, was calculated using the *Akaike's Information Criterion* (AIC) [14]

$$AIC(p) = N ln(\sigma^2) + 2p \tag{2}$$

where N is the size of the window, p the order of the model and σ^2 the prediction error variance associated with p. The optimal model order is the value of p that minimizes (2). The value of p given by AIC was also optimized to obtain an optimal compromise between error and computation time.

The set of all p estimated AR coefficients, $\mathbf{a}_\tau^k$ with $\tau \in \{s, w\}$ (where **s** and **w** refer to Sleep and Wakefulness respectively), obtained from all windows from all segments from each class/state are used to estimate the corresponding mean, $\{\mu_s, \mu_w\}$ and covariance matrices, $\{\mathbf{C}_s, \mathbf{C}_w\}$, of the sleep, ω_s and wakefulness, ω_w, classes.

C. Bayesian Estimator

The obtained $p \times W$ matrices were combined into two single matrices, containing the coefficients of the estimated AR models obtained for wakefulness and sleep periods respectively. The clouds of points of each class are described by the following multivariate Gaussian distribution functions,

$$p(\mathbf{a}|SW = 0) \sim \mathcal{N}(\mu_s, \mathbf{C}_s) \tag{3}$$

$$p(\mathbf{a}|SW = 1) \sim \mathcal{N}(\mu_w, \mathbf{C}_w) \tag{4}$$

where $\mathbf{a} = \{a_1, a_2, ..., a_p\}$ is a vector of coefficients, $\mathbf{C}_w$ and $\mathbf{C}_s$ are $p \times p$ covariances matrices and μ_w and μ_s are the centroid locations associated with the wakefulness and sleep states respectively, computed in the training step.

Given an actigraphy sample, $y(n)$ and its p previous samples, $Y(n) = \{y(n-1), y(n-2), ..., y(n-p)\}^T$, the p coefficients of a p-order AR model are estimated as follows,

$$\mathbf{a}(n) = (Y(n)Y^T(n))^{-1}Y(n)y(n) \tag{5}$$

where $\mathbf{a}(n) = \{a_1(n), a_2(n), ..., a_p(n)\}$.

The discriminative functions to classify $\mathbf{a}$, the logarithm of the joint probability of $\mathbf{a}$ and ω_τ, where $\omega_\tau \in \{\omega_s, \omega_w\}$ are the sleep and wakefulness classes respectively, are

$$g_\tau(n) = \log p(\mathbf{a}(n), \omega_\tau) = \log p(\mathbf{a}(n)|\omega_\tau) + \log p(\omega_\tau) \tag{6}$$

where

$$\begin{aligned} \log p(\mathbf{a}(n), \omega_\tau) \quad = \quad & -\frac{1}{2}(\mathbf{a}(n) - \mu_\tau)^T \mathbf{C}_\tau^{-1}(\mathbf{a}(n) - \mu_\tau) \\ & -\frac{1}{2}\log|\mathbf{C}_\tau| + C^{te} \end{aligned} \tag{7}$$

and $p(\omega_s) = p(\omega_w)$ in this paper.

The classification of each sample is based on the Bayes factor, $\gamma(n) = g_s(n) - g_w(n)$ according to

$$SW(n) = \begin{cases} 1 & \text{if } \gamma(n) \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

D. Validation

The performance assessment of the estimator was tested using a *leave-one-out cross validation* (LOOCV) approach. The LOOCV approach consists in selecting one single segment of the segmented data and use it as the validation data while the remaining segments are used as the training data. This process is repeated such that each segment is used once as validation data. The total estimation error is obtained as follows

$$\text{Total error} = \frac{FP + FN}{N} \tag{9}$$

where FP and FN are the false positives and false negatives respectively and N the total number of segments classified. The Sensitivity and Specificity of the classifier were calculated as

$$\text{Sensitivity} = \frac{TP}{TP + FN} \tag{10}$$

$$\text{Specificity} = \frac{TN}{FP + TN} \tag{11}$$

where TP and TN are the true positives (SW = 0) and true negatives (SW = 1) respectively. The error returned by the LOOCV approach was also used to find the optimal AR model order and the optimal window-length/window-step combination. These optimal values were found by running the LOOCV for several combinations of the parameter and selecting the ones that lead to higher classification rates.

Finally, the estimator was trained with all the segmented healthy data and tested with several non-segmented healthy and non-healthy patients.

III. Results

Due to the large size of the data sets and the high number of iterations in some of the methods, namely the LOOCV, the tests involve a heavy computational load. The code was implemented in MATLAB@R2009 running on a 4×2GHz quad-core computer, with $4GB$ Ram, making use of Matlab's parallel processing functionalities.

The discriminative power of the classifier can be clearly observed in Fig.1 where the coefficients of the second order AR models associated with the data from both classes are displayed.

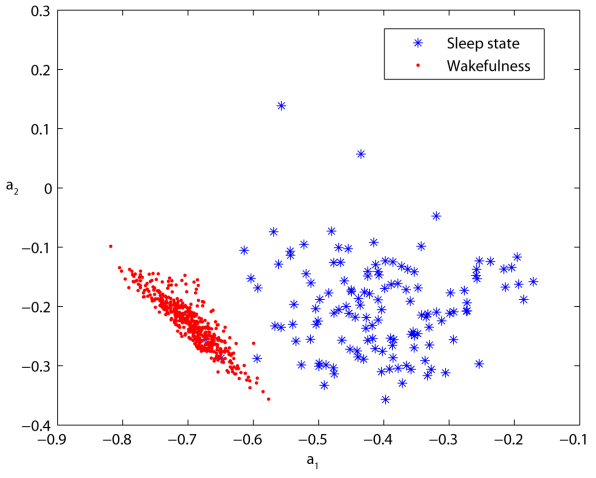

Fig. 1. Plot of α_1 vs α_2 (2^{nd} order AR model) for both Sleep and Wakefulness data segments.

The cloud of coefficients obtained for wakefulness data exhibits lower variance than the coefficients for sleep state and a small overlap between both clouds is observed. The obtained result confirms that the two states can be well represented by a multivariate normal distribution.

TABLE I

Error obtained from the LOOCV to several model orders

Model Order	Error (%)
3	3.69
4	3.67
5	3.82
10	3.86
20	4.47

TABLE II

Percentage of error obtained for several combinations of window-length, N, and window-step, M, with a 4^{th} order AR model.

N	M			
	1	10	30	50
60	5.2	5.22	5.0	5.1
80	4.34	4.27	4.21	4.14
100	3.77	3.74	3.77	3.59
200	2.69	2.62	2.51	2.43
500	2.23	2.22	2.17	2.18
1000	2.36	2.25	3.45	5.88

The Akaike's Information Criterion was used to determine a first guess of the optimal AR model order. Several simulations, for different window-length/window-step combinations, were tested, resulting in an optimal order of 20. Due to the high number of iterations of the algorithms, a model of a lower order is more convenient. To infer the effect of the model order on the final error of the Bayesian estimator, LOOCV was used to test several orders.

Table I shows the obtained error for the tested orders, it can be seen that the optimal model order obtained from AIC does not directly translates into a better result in the LOOCV. It is important to note that, while the error obtained for $p = 20$ is not the minimum error, the differences between this result and the ones with other orders are minimal. Through the rest of the tests a model order of 4 is assumed.

In order to check the influence of window-length and window-step, tests were run with different combinations of these values. Window-lengths of $N = \{60, 80, 100, 200, 500\}$ samples were tested with $1, 10, 20$ and 50 samples for window-step.

Table II shows that the minimum error value returned by LOOCV method corresponds to a window-size of 500 and a window-step of 30.

The sensitivity of the classifier is 98% and the specificity 74%. Although wakefulness state is more prone to false sleep detections this result reflects the different sizes of the two groups of data.

The algorithm was finally tested with non-segmented data. The Bayesian estimator was trained with healthy segmented data using 4^{th} order AR models to fit the data and window-size/window-step were set to 500/30. Since the non-segmented data to be classified has alternate sleep and wakefulness states, a window size of 500 minutes, approximately 8 hours, was too large, resulting into slow transitions from sleep to wakefulness state and vice-versa.

Thus, while the Bayesian estimator was trained with 500 minute data windows, the data to be estimated was segmented into windows of 100 minutes, maintaining a window-step of 30 minutes. Figure 2 shows a segment of approximately 2 nights and days of actigraphy data and the correspondent Bayes factor. It can be seen that the value of the Bayes factor is consistent with the actigraphy data although the wakefulness state is more prone to estimation errors than the sleep state.

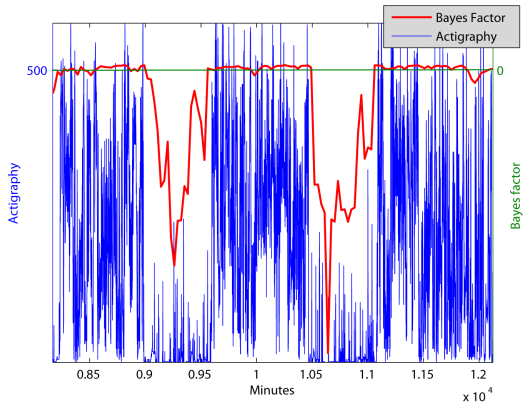

Fig. 2. Plot of Actigraphy data and Bayes factor for approximately 2 nights/days. The light green line marks the value zero, to better evaluate the Bayes factor.

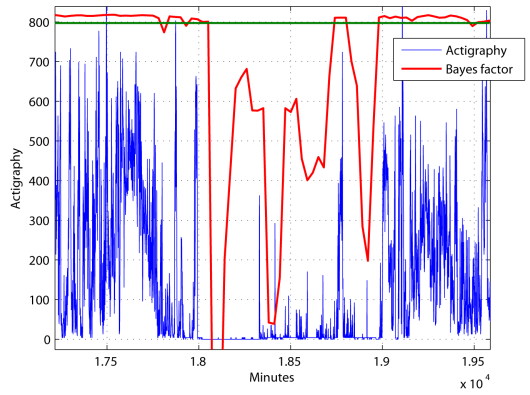

Fig. 3. Plot of Actigraphy data and Bayes factor for approximately 2 nights/days.

Figure 3 plots a segment of data from a patient with sleep disorders. The segment from the figure shows the detection of a wakening episode, that was confirmed from the SeD and actigraphy data. It can also be seen the detection of a sleep period during the theoretical wakefulness period, this detection can be either an error or a true episode.

This preliminary results suggest that a more complex analysis of the Bayes factor along the circadian cycle may provide more useful and complex information about the stages of the sleep. The method proposed in this paper is uniquely based on movement which imposes an upper bound on the accuracy for sleep/wakefulness segmentation purposes. In the future other physiological measures will be incorporated such as the *Heart Rate Variability* (HRV), body temperature, *Sleep Diary*, respiration and skin electrical conductivity.

IV. Conclusions

This paper describes a classification procedure for actigraphy data for SW state estimation and detection of abnormal temporal patterns associated with some of the most common sleep disorders. The tests performed using real data form 23 healthy subjects resulted in an overall classification rate of 96% using segmented data. The data used in this paper for testing and training was manually classified by experts with help of complementary information obtained from *sleep diary* and light intensity. While it is expected that the error increases for continuous data the ability of the method to accurately estimate the SW state with minimum error seems promissing. Especially when taking into account the possibility of adding extra physiological or external features to the classification procedure. These results are a step toward an alternative method in the diagnosis of some specific sleep disorders involving long term monitoring and normal day life conditions, mainly when movement abnormalities are present.

References

[1] L. D. Pandi-Perumal, S.R., *Sleep Disorders: Their Impact on Public Health*. Informa Healthcare, 2006.

[2] R. J. Cole, D. F. Kripke, W. Gruen, D. J. Mullaney, and J. C. Gillin, "Automatic sleep/wake identification from wrist activity," *Sleep*, vol. 15, no. 5, pp. 461–9, 1992.

[3] V. Gimeno, T. Sagales, L. Miguel, and M. Ballarin, "The statistical distribution of wrist movements during sleep," *Neuropsychobiology*, vol. 38, no. 2, pp. 108–12, 1998.

[4] G. Jean-Louis, D. F. Kripke, R. J. Cole, J. D. Assmus, and R. D. Langer, "Sleep detection with an accelerometer actigraph: comparisons with polysomnography," *Physiol Behav*, vol. 72, no. 1-2, pp. 21–8, 2001.

[5] J. Paquet, A. Kawinska, and J. Carrier, "Wake detection capacity of actigraphy during sleep," *Sleep*, vol. 30, no. 10, pp. 1362–9, 2007.

[6] P. Pedro, P. Teresa, and S. Joao, "Sleep/wakefulness state from actigraphy," 2009, 1573810 362-369.

[7] A. Sadeh and C. Acebo, "The role of actigraphy in sleep medicine," *Sleep Med Rev*, vol. 6, no. 2, pp. 113–24, 2002.

[8] K. L. Lichstein, K. C. Stone, J. Donaldson, S. D. Nau, J. P. Soeffing, D. Murray, K. W. Lester, and R. N. Aguillard, "Actigraphy validation with insomnia," *Sleep*, vol. 29, no. 2, pp. 232–9.

[9] S. Ancoli-Israel, R. Cole, C. Alessi, M. Chambers, W. Moorcroft, and C. P. Pollak, "The role of actigraphy in the study of sleep and circadian rhythms," *Sleep*, vol. 26, no. 3, pp. 342–92, 2003.

[10] B. Sivertsen, S. Omvik, O. E. Havik, S. Pallesen, B. Bjorvatn, G. H. Nielsen, S. Straume, and I. H. Nordhus, "A comparison of actigraphy and polysomnography in older adults treated for chronic primary insomnia," *Sleep*, vol. 29, no. 10, pp. 1353–8, 2006.

[11] P. J. Hauri and J. Wisbey, "Wrist actigraphy in insomnia," *Sleep*, vol. 15, no. 4, pp. 293–301, 1992.

[12] S. L. Sitnick, B. L. Goodlin-Jones, and T. F. Anders, "The use of actigraphy to study sleep disorders in preschoolers: some concerns about detection of nighttime awakenings." *Sleep*, vol. 31, no. 3, pp. 395–401, 2008.

[13] R. Takalo, H. Hytti, and H. Ihalainen, "Tutorial on univariate autoregressive spectral analysis," *J Clin Monit Comput*, vol. 19, no. 6, pp. 401–10, 2005.

[14] H. Akaike, "A new look at the statistical model identification," *IEEE Transactions on Automatic Control*, vol. 19, pp. 716–723, 1974.