01 Nov 2017

Classification of Rest and Active Periods in Actigraphy Data using PCA

Isaac W. Muns

Yogesh Lad

Ivan G. Guardiola
Missouri University of Science and Technology, guardiolai@mst.edu

Matthew S. Thimgan
Missouri University of Science and Technology, thimgan@mst.edu

Recommended Citation

I. W. Muns et al., "Classification of Rest and Active Periods in Actigraphy Data using PCA," *Procedia Computer Science*, vol. 114, pp. 275-280, Elsevier, Nov 2017.
The definitive version is available at https://doi.org/10.1016/j.procs.2017.09.041

Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 114 (2017) 275–280

Procedia Computer Science

www.elsevier.com/locate/procedia

Complex Adaptive Systems Conference with Theme: Engineering Cyber Physical Systems, CAS October 30 – November 1, 2017, Chicago, Illinois, USA

Classification of Rest and Active Periods in Actigraphy Data Using PCA

Isaac W. Muns[a], Yogesh Lad[a], Ivan G. Guardiola[a*], Matthew Thimgan[b]

[a]*Missouri S&T, Dept. of EMSE, 600 W 14th Street, Rolla, MO, 65409, USA.*
[b]*Missouri S&T, Dept. of Biological Sciences, 400 W. 11th St., Rolla, MO, 65409, USA.*

Abstract

In this paper we highlight a clustering algorithm for the purpose of identifying sleep and wake periods directly from actigraphy signals. The paper makes use of statistical Principal Component Analysis to identify periods of rest and activity. The aim of the proposed methodology is to develop a quick and efficient method to determine the sleep duration of an individual. In addition, a robust method that can identify sleep periods in the accelerometer data when duration, time of day varies by individual. A selected group of 10 individual's sensor data consisting of actigraphy from an accelerometer (3-axis), near body temperature, and lux sensors from a single GENEActiv watch worn on the non-dominant hand. The actigraphy of each individual was collected 24 hours a day for a period spanning 80 days. We highlight that a simple data preprocessing stage followed with a 2 phase clustering method provides results that align with previously validated methodologies.

Peer-review under responsibility of the scientific committee of the Complex Adaptive Systems Conference with Theme: Engineering Cyber Physical Systems.

Keywords: Clustering; Actigraphy; Principal Component Analysis

* Corresponding author. Tel.: +1-573-341-6153; fax: +1-573-341-6990.
E-mail address: guardiolai@mst.edu

1877-0509
Peer-review under responsibility of the scientific committee of the Complex Adaptive Systems Conference with Theme: Engineering Cyber Physical Systems.
10.1016/j.procs.2017.09.041

1. Introduction

"Sleep is the single most effective thing you can do to reset your brain and body[1]" Scientifically, sufficient sleep has emerged as a critical variable is our health and well-being. Inadequate sleep, erodes mental and physical well-being whilst often exacerbating health problems[2]- from obesity[3], depression[4], diabetes[5], and cardiovascular incidents[6, 7] to Alzheimer's[8] and cancer[9]. Moreover, sleep problems may go undetected or unaddressed for years. One study documented an average of 7.3 years from the onset of obvious indicators of sleep apnea to diagnosis (range 1 month to 40 years)[10]. Sleep apnea may affect 20% of the population[11], increasing the risk of stroke[11], diabetes[12], obesity[11], car accidents[13], and mortality[14]. In addition, a workforce with inadequate sleep is less efficient, costly, and less productive[15]. In professions that require high accuracy and safety levels, sleeplessness can have detrimental consequences. For example, an estimated 810,000 sleep-related collisions per year resulting in 1,400 fatalities at a cost of $16 billion dollars in the United States alone[16]. Hence, the premise that a well-rested individual will be more productive, safer, and healthier comes with a plethora of benefits, in cost reductions from healthcare insurance, to less accidents/increased safety, and higher levels of productivity. To improve their health and performance, people need to understand their unique sleep and wake patterns and how they are evolving over a lifetime or in specific situations. To address this problem, wearable sensing devices have been developed that produce informative data in a passive, non-intrusive manner. Wearables spread sleep awareness beyond clinicians and researchers to include the millions of users who rely on health applications to track their physical activity and unique sleep patterns [9]. The modern wearable devices are now embedded with a multitude of sensors which include but not limited to: ambient light sensors, accelerometers (three axis), skin temperature sensor, and ambient temperature sensors. The most common package for such a wearable device is often a watch that is worn in the non-dominant arm of the individual under study, who wears the device and collects all activity through such a sensor package during a period of days or weeks. These types of studies are often referred to as an actigraph or actigraphy study, which are considered a cost effective opportunity to conduct longitudinal, naturalistic studies of the sleep-wake system. It is this knowledge that is leading to the growth in the development of devices associated with sleep health, such as wearable sensing devices.

The analysis of actigraphy has been extensively studied over the last three decades. Since the seminal works of Cole et al.[17] and Sadeh et al.[18, 19] in the 1990s, the goal to automate the classification of sleep and wake periods directly from an accelerometer signal has been pursued and continues to gain interest within the research and industry communities. A recent review on the approaches to convert activity data to sleep/wakefulness patterns can be found in[20]. Despite the inherent interest, there is not a single algorithm that can be applied to extract meaning from the enormous amounts of data generated every day with wearable technologies. To address this problem, we employed a paired clustering approach to determine periods of 'rest' and 'activity'. For the purposes of 'proof of concept,' of the 10 individuals two were directly compared. This comparison of the 'rest' period duration was done on these two individual for 29 days using our proposed methodology and the ActiGraph algorithm that has been previously validated[21].

1.1. Data Details

Activity data were collected on 10 individuals for 80 continuous days. Each individual wore the GENEActiv-Sleep[22] watch 24 hours a day, with the exception of charging time. Charging time was done once a month and often resulted in only a gap of 2-6 hours without data. The data was collected at a fidelity of 10Hz and was down sampled to supply averages of acceleration, lux, and temperature. The watch collected the following datum for each minute: time stamp; mean x, mean y, mean z axis acceleration; mean lux; sum vector magnitudes (SVM); x axis, y axis, and z axis standard deviation; and peak lux. The SVM is calculated internally, by employing the following the equation, $SVM^g s = \sum[(x^2 + y^2 + z^2)^{\frac{1}{2}} - 1(g)]$ where x, y, z are the mean x, mean y and mean z actigraphy counts and 1(g) is subtracted. That is, (g) corresponds to a unit of gravity. Therefore, when the accelerometer is static and the earth's gravitation pull is the only acceleration, the result of this will be zero. The data is supplied as a binary file that can be downloaded from the watch. This binary file is then converted into a comma separated file (*.csv) by using the supplied software from GENEActiv. Each line of the csv file summarizes one minute of sensor readings.

2. Methodology

The rest versus active period recognition methodology is comprised of two main steps: data pre-processing and clustering. In the pre-processing step the raw actigraphy data collected is transformed into a form more amenable to clustering. Then based on the characteristics of the data the most appropriate clustering tool is applied to generate two clusters. These the two main clusters are to identify the part of the actigraphy time series, which is a rest period and similarly identify the period of activity. In the pre-processing step it is necessary to determine if the actigraphy count time-series is in a form that would lend itself to the application of well know and established clustering techniques and tools. In Figure 1, an individual's 24 hours actigraphy SVM is illustrated (1440 minutes). There is a period of low activity from minute 450-1000 (roughly 9.1hours), which can be deemed a rest period. However, there is a large number of observations during periods of activity that have low actigraphy counts, this is common as low movement does not necessarily mean that an individual is at full rest. In Figure 2, a histogram illustrates that the actigraphy signal is highly skewed with low actigraphy counts being far more common.

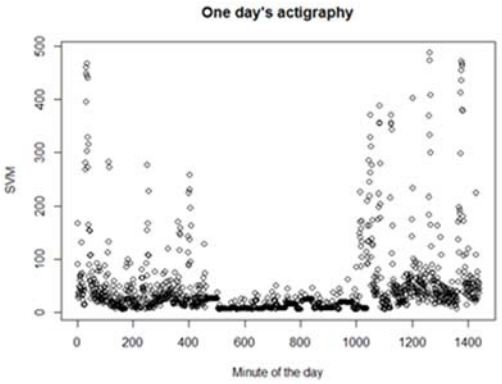

Figure 1: An entire day's actigraphy counts for one individual (SVM)

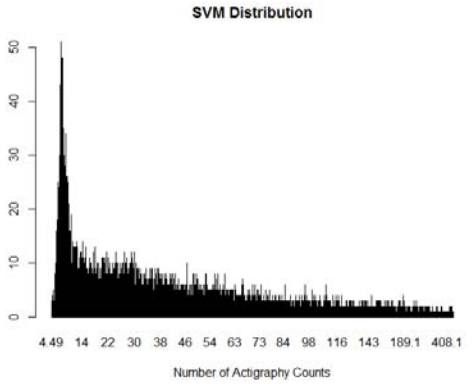

Figure 2: SVM histogram across all individuals for one day.

The data contains all sensor data, acceleration, lux, temperature and all will be combined through the use of Principal Components (PCs). Thus, since the acceleration data is skewed it is necessary to standardize the data. Thus, the accelerometer, temperature, and lux data were scaled and centered. "Centered" subtracts the mean while "scaled" divides by the standard deviation. This is necessary as we seek to incorporate the information from all sensors. In the context of PC Analysis (PCA), the longer the tail the more its influence. The primary justification for transforming our data is due to the range of values are significantly different from one sensor to another. Moreover, if the relationships between the variables analysed are not linear, the values of correlation coefficients can be lower. Thus, it is sometimes useful to transform the original variables prior to the Principal Component Analysis to "linearize" these relationships. A common approach is the use of the Box-Cox transformation, which is in the family of power transformations indexed with parameter, λ, which is estimated from the data directly. The variance is changing with time, therefore; the process is nonstationary. After applying Box-Cox with a particular value of lambda the process should look stationary. Specifically, sensor readings are centered about their respective means and divided by their respective standard deviations. Next, the Principal Component Analysis (PCA), which will be used to concisely represent the data, is applied. After this application the PCA scores are determined. These PCA score are filtered using a two sided moving average, which will require the moving window length to be determined empirically. The successful completion of the moving average filtering completes the data pre-processing step and can now be clustered. In the clustering step, the k-means clustering technique is carried out. First, an optimal number of clusters is determined by applying k-means clustering for many values of k. The premise that an individual rests for a certain portion of a 24 hour period. The 10 participants were asked to determine they duration of rest in a 24 hour period for the duration of the study. Imperially, the group reported a mean of 5.7 hours of daily sleep with a standard deviation of 2.7. The cluster closest to containing 5.7 hours of all sensor readings is deemed to be the 'rest' cluster. Duration of rest can then be determined by counting how many sensor readings fall into the rest cluster.

3. Results

All of the computation and applications were carried in the R package 3.3.3. The pre-processing step yielded a $\lambda = 0.2$ for the Box Cox transformation. The 24 hr (12pm to 12pm) time series were filtered using a 30 minute two sided moving average. Then centering, scaling, and a power transformation was applied. From a scree analysis only two principal components are needed to represent the all sensor readings, that is, more than 95% of variation in the data was contained by two PCs, which was determined via a scree plot. Next, the application of the k-means algorithm was applied on the two principal component scores. Figure 3 depicts the sample of PCA scores. It can be easily seen that when plotting the two scores they result in a non-convex set. Moreover, k-means can be paired with another algorithm to describe non-convex clusters[23]. In this case the k-means algorithm was paired with another algorithm, hierarchal clustering, to determine better clusters. Figure 5 shows the k means initialization, k=6 and was reduced to k=3. Through the employment of the gap statistic, we determined our data is best clustered into three groups. It should be noted that a single linkage hierarchal can be directly applied to the data. The purpose of the k means is to take advantage of the k mean capability to deal with large sets. Simply, the goal is to iteratively determine the tightest clusters, that is, we repeatedly pick the two clusters that are closest together and merge them. It is important in this scenario that we use the "single-link" method, in which the distance between two clusters is defined by the distance between the two closest data points we can find, one from each cluster. Figure 4 illustrates the clustering for a single day's actigraphy for one individual after the paired clustering was carried out. Next, the 'rest' cluster must be identified. On average, humans are asleep for one third of their lives. Because of this, the cluster closest to containing one third of all sensor readings is deemed to be the 'sleep' or 'rest' cluster. This results in the identification of a 'rest' cluster which was found to be cluster 2 in Figure 4 and is shown in days' time series for one individual in Figure 6. Figure 6 illustrates the rest period (cluster 2) in black and the active clusters (cluster 1 and 2) data as red in colour. Duration of sleep can then be determined by counting how many sensor readings fall into the sleep/rest cluster.

3.1. Validation of Rest Duration Periods

Through the employment of the proposed methodology, a 'rest' duration period can be determined for each 24 hour period. However, often the validation of any algorithm that seeks to score sleep/wake states is often validated through the collecting actigraphy as well as electroencephalogram (EEG)[24] simultaneously on the same individual during the same period of time. This form of validation is common and has been employed in various studies[25-28]. Here we validate by comparing our duration to a known validated algorithm produced by GENEActiv, which has been rigorously validated in [29] and [21]. Thus, the validation consisted of taking 29 days for two individuals chosen at random from the 10 participants and a continuous period of 29 days randomly chosen and duration was determined for 'rest' every 24 hour period (12pm -12pm). The GENEActiv duration was determined using their software and algorithm and the duration value was then derived using our proposed methodology. The results are illustrated in Figure 7 and Figure 8. While the duration values statistically differed according to the ANOVA both having P-value of 0.00000, the difference was on average 44 minutes and at its maximum difference were only observed to be 80 minutes apart. Our methodology generally underestimated the duration of the rest period in comparison to the GENEActiv Algorithm.

4. Conclusions

This highly automated sleep recognition methodology is very useful because it is tailored to each individual subject and set of sensors. Furthermore, its use of PCA allows it to holistically analyze all sensor information as opposed to only using acceleration. Furthermore, the results showed that although the duration by the proposed algorithm was generally lower than that of GENEActiv, the use of such a simple algorithm can provide quick and effective results. Future works should focus on the validation of the clustering algorithm using EEG data. Furthermore, an exploration of other clusters should be done as they may highlight periods prior to the 'rest' state. Overall our proposed clustering approach can be used across individual and devices with minimal computational cost. The combination of k means with single link clustering turned out to be a robust method for actigraphy. It is less sensitive to initialization, and less sensitive to the choice of parameters making it more robust across individual variability in activity and rest. K means does the brunt of the work and then once completed the more expensive hierarchical

method is applied to reap its benefits.

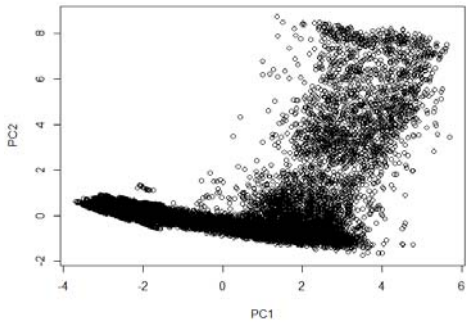

Figure 3: Pre-processed data PCA score sample

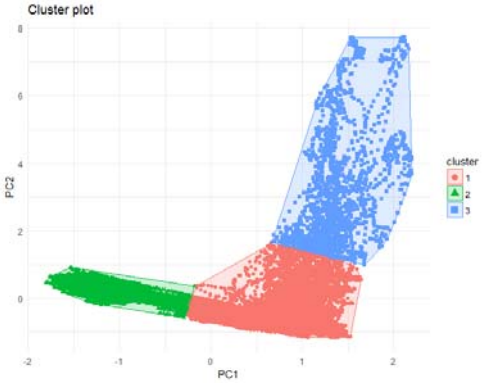

Figure 4: Optimal Clustering 3 for one individual

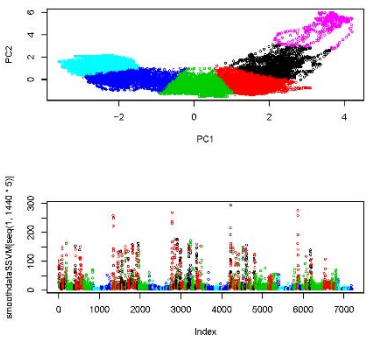

Figure 5: Six group clustering

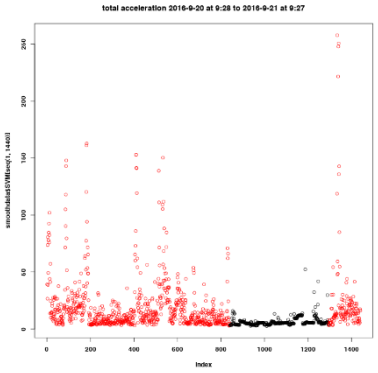

Figure 6: Rest (black) plotted with active (red) periods during a single day.

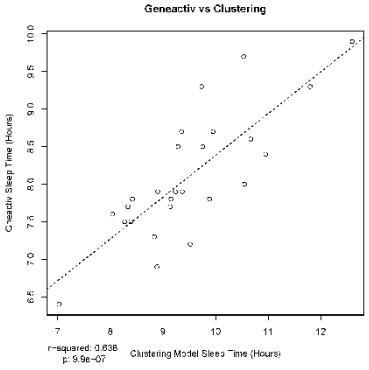

Figure 7: ANOVA result for first sample.

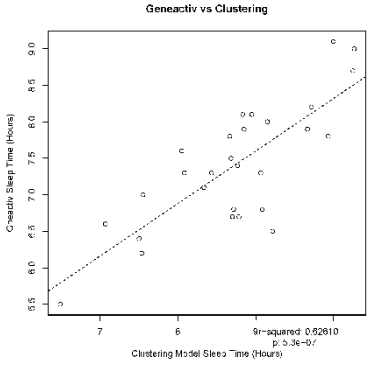

Figure 8: ANOVA result for second sample

Acknowledgements

The authors would like to extend their thanks to the all of the undergraduate students involved in the collection, analysis, and participation in this study.

References

1. Gree, P., *Sleep is the New Status Symbol* in *The New York Times*. 2017. p. ST1.
2. Buysse, D.J., *Sleep health: can we define it? Does it matter.* Sleep, 2014. **37**(1): p. 9-17.
3. Taheri, S., et al., *Short sleep duration is associated with reduced leptin, elevated ghrelin, and increased body mass index.* PLoS Med, 2004. **1**(3): p. e62.
4. Friedman, R.A., *Yes, Your Sleep Schedule Is Making You Sick*, in *The New York Times*. 2017: NYC.
5. Kadono, M., et al., *Various patterns of disrupted daily rest–activity rhythmicity associated with diabetes.* Journal of sleep research, 2016.
6. Grandner, M.A., et al., *Sleep duration, cardiovascular disease, and proinflammatory biomarkers.* Nat Sci Sleep, 2013. **5**: p. 93-107.
7. Young, T. and P. Peppard, *Sleep-disordered breathing and cardiovascular disease: epidemiologic evidence for a relationship.* Sleep, 2000. **23 Suppl 4**: p. S122-6.
8. Lim, M.M., J.R. Gerstner, and D.M. Holtzman, *The sleep-wake cycle and Alzheimer's disease: what do we know?* Neurodegener Dis Manag, 2014. **4**(5): p. 351-62.
9. Sathyanarayana, A., J. Srivastava, and L. Fernandez-Luque, *The Science of Sweet Dreams: Predicting Sleep Efficiency from Wearable Device Data.* Computer, 2017. **50**(3): p. 30-38.
10. Rahaghi, F. and R.C. Basner, *Delayed Diagnosis of Obstructive Sleep Apnea: Don't Ask, Don't Tell.* Sleep Breath, 1999. **3**(4): p. 119-124.
11. Franklin, K.A. and E. Lindberg, *Obstructive sleep apnea is a common disorder in the population-a review on the epidemiology of sleep apnea.* J Thorac Dis, 2015. **7**(8): p. 1311-22.
12. Tasali, E., B. Mokhlesi, and E. Van Cauter, *Obstructive sleep apnea and type 2 diabetes: interacting epidemics.* Chest, 2008. **133**(2): p. 496-506.
13. Teran-Santos, J., A. Jimenez-Gomez, and J. Cordero-Guevara, *The association between sleep apnea and the risk of traffic accidents. Cooperative Group Burgos-Santander.* N Engl J Med, 1999. **340**(11): p. 847-51.
14. Young, T., et al., *Burden of sleep apnea: rationale, design, and major findings of the Wisconsin Sleep Cohort study.* WMJ, 2009. **108**(5): p. 246-9.
15. Swanson, L.M., et al., *Sleep disorders and work performance: findings from the 2008 National Sleep Foundation Sleep in America poll.* J Sleep Res, 2011. **20**(3): p. 487-94.
16. Sassani, A., et al., *Reducing motor-vehicle collisions, costs, and fatalities by treating obstructive sleep apnea syndrome.* SLEEP-NEW YORK THEN WESTCHESTER-, 2004. **27**(3): p. 453-458.
17. Cole, R.J., et al., *Automatic sleep/wake identification from wrist activity.* Sleep, 1992. **15**(5): p. 461-469.
18. Sadeh, A. and C. Acebo, *The role of actigraphy in sleep medicine.* Sleep medicine reviews, 2002. **6**(2): p. 113-124.
19. Sadeh, A., et al., *The role of actigraphy in the evaluation of sleep disorders.* Sleep, 1995. **18**(4): p. 288-302.
20. Bei, B., et al., *Beyond the mean: a systematic review on the correlates of daily intraindividual variability of sleep/wake patterns.* Sleep medicine reviews, 2016. **28**: p. 108-124.
21. Esliger, D.W., et al., *Validation of the GENEA Accelerometer.* 2011.
22. GENEACtiv. *Compare Products*. 2017 [cited 2017 May 1, 2017]; GENEActiv Sensor]. Available from: https://www.geneactiv.org/products/compare-products/.
23. Wu, X., et al., *Top 10 algorithms in data mining.* Knowledge and information systems, 2008. **14**(1): p. 1-37.
24. Sadeh, A., K.M. Sharkey, and M.A. Carskadon, *Activity-Based Sleep—Wake Identification: An Empirical Test of Methodological Issues.* Sleep, 1994. **17**(3): p. 201-207.
25. Ancoli-Israel, S., et al., *The role of actigraphy in the study of sleep and circadian rhythms. American Academy of Sleep Medicine Review Paper.* Sleep, 2003. **26**(3): p. 342-392.
26. Blood, M.L., et al., *A comparison of sleep detection by wrist actigraphy, behavioral response, and polysomnography.* Sleep, 1997. **20**(6): p. 388-395.
27. Cheung, J., et al., *0778 VALIDATION OF MINUTE-TO-MINUTE SCORING FOR SLEEP AND WAKE PERIODS IN A CONSUMER WEARABLE DEVICE.* Sleep, 2017. **40**(suppl_1): p. A288-A288.
28. Hedner, J., et al., *A novel adaptive wrist actigraphy algorithm for sleep-wake assessment in sleep apnea patients.* Sleep, 2004. **27**(8): p. 1560-1566.
29. te Lindert, B.H. and E.J. Van Someren, *Sleep estimates using microelectromechanical systems (MEMS).* Sleep, 2013. **36**(5): p. 781-789.