35th Annual International Conference of the IEEE EMBS
Osaka, Japan, 3 - 7 July, 2013

Addressing the challenges of sleep/wake class imbalance in bed based non-contact actigraphic recordings of sleep

Andrew McDowell, Mark P. Donnelly, Chris D. Nugent, *Member, IEEE*, Leo Galway and Michael J. McGrath

Abstract— **Utilising strategically positioned bed-mounted accelerometers, the Passive Sleep Actigraphy platform aims to deliver a non-contact method for identifying periods of wakefulness during night-time sleep. One of the key problems in developing data driven approaches for automatic sleep monitoring is managing the inherent sleep/wake class imbalance. In the current study, actigraphy data from three participants over a period of 30 days was collected. Upon examination, it was found that only 10% contained wake data. Consequently, this resulted in classifier overfitting to the majority class (sleep), thereby impeding the ability of the Passive Sleep Actigraphy platform to correctly identify periods of wakefulness during sleep; a key measure in the identification of sleep problems. Utilising Spread Subsample and Synthetic Minority Oversampling Techniques, this paper demonstrates a potential solution to this issue, reporting improvements of up to 28% in wake detection when compared to baseline data while maintaining an overall classifier accuracy of 90%.**

I. INTRODUCTION

While the function of sleep is not fully understood, it is known to be an essential biological function [1]. Correspondingly, there are a range of medical conditions and lifestyle factors known to have a detrimental impact on the duration and quality of sleep [2]. Depending on the severity, sleep problems have been linked to numerous health related issues including a reduction in physical and mental performance and is, furthermore, considered a risk factor in a range of secondary conditions such as diabetes and depression [2]. Subsequently, due to these wide ranging consequences, it is clear that benefits may be gained from proactively identifying and monitoring sleep problems.

A. Sleep Monitoring Technologies

In recent years there has been a growing interest in the provision of a technological solution suitable for monitoring sleep at home. A popular alternative to polysomnography (PSG) achieving widespread clinical approval in recent years is sleep actigraphy [3]. This approach utilises a piezoelectric accelerometer embedded within a wrist-worn watch-like device to record a wearer's movement; the amplitude of these movements are subsequently used to determine the wearer's sleep/wake state. While not diagnostically as sensitive as PSG, comparatively, sleep actigraphy is less costly, suitable for long-term observation, minimally invasive and applicable for use at home [3]. The successful operation of sleep actigraphy, however, is reliant on user compliance. For example, users must be comfortable wearing the device at all times, particularly while sleeping. Furthermore they must remember to replace the device when removed for bathing, charging etc. For this reason, there is merit in the development of a non-contact pervasive bedroom based approach that would mitigate or remove user compliance as a risk to data acquisition.

Non-contact methods of monitoring sleep are, however, predominantly confined to research environments. Systems involving under mattress pressure sensors [4], Doppler based radar [5] and bed feet load cells [6] have all been reported. While these approaches have been somewhat effective, all require the use of application specific equipment to function. The Passive Sleep Actigraphy (PSA) platform, presented in our earlier work [7, 8] proposes an alternate approach to non-contact sleep monitoring that is based on traditional sleep actigraphy. In the current implementation, a number of strategically positioned wireless accelerometers are deployed on a mattress to record the movements of an occupant while in bed. In alignment with traditional sleep actigraphy techniques, the amplitude of these movements has been evaluated in order to determine the sleep/wake state of the bed occupant. This highlights one significant advantage of the PSA platform in that it is able to directly adopt the techniques employed by clinically validated sleep actigraphy into this potential non-contact alternative.

B. Current Work

Both the PSA Platform and non-contact approaches previously discussed have the potential for determining the sleep/wake state of an individual whilst in bed. There is, however, an inherent issue in the training of many classifiers due to the typical class distribution of sleep and wake data. This issue is highlighted in the dataset collected during our previous work, which comprised of 129 hours of bed based actigraphic data, with only 15% of this data being labelled as wake. This distribution is expected, given that an individual in bed at night will typically spend the majority of their time asleep. This imbalance, however, can lead to issues of classifier overfitting where in order to achieve the highest

*This work has been part sponsored by Intel Labs Europe (Ireland) and the Department for Education and Learning (N. Ireland) under a Co-operative Awards in Science and Technology (CAST) collaboration.

A. McDowell is with the University of Ulster, Shore Road Newtownabbey, BT37 0QB, Northern Ireland (e-mail: mcdowell-a10@email.ulster.ac.uk, phone: +44 2890 368394)

M. P. Donnelly, C. D. Nugent and L. Galway are also with the University of Ulster (email: mp.donnelly@ulster.ac.uk; cd.nugent@ulster.ac.uk; l.galway@ulster.ac.uk)

M. J. McGrath is with Intel Labs Europe, Collinstown Industrial Park, Leixlip, Co. Kildare, Ireland (email: michael.j.mcgrath@intel.com)

possible overall accuracy, the classifier will favour the majority class [9]. Within the context of the PSA platform, this is expressed by a high classification performance in identifying periods of sleep at the cost of correctly identifying wake periods.

Imbalanced data sets are not uncommon in the field of pattern recognition and machine learning. Fraud detection, for example, often has a class distribution in the order of 100 legitimate instances to one fraudulent instance [10]. Likewise, the detection of oil spills in satellite images is reported in the order of 1000:1 [11]. Indeed some classification problems must address a class imbalance in the order of 10,000:1 [11]. A further consideration when dealing with imbalanced datasets is that often the correct identification of the minority class yields the most useful information. For example the primary output of a fraud detection system is suspected fraudulent transactions, not legitimate ones. Subsequently, within this work, correctly identifying wakeful periods has a higher diagnostic value than correctly identifying periods of sleep for the purposes of identifying potential sleep problems [3].

The techniques of undersampling and oversampling are often used to address class imbalance problems [9]. Both undersampling and oversampling manipulate a given dataset until a specified class distribution has been achieved. Undersampling achieves this by removing instances of the majority class, while oversampling operates by replicating or synthesising additional instances from the minority class [9]. In this work, both undersampling and oversampling will be evaluated as potential solutions to address the sleep/wake class imbalance problem for non-contact bed-based actigraphic recordings.

II. Materials and Methods

In the current study, PSA was investigated as a non-contact method for identifying disturbed sleep by localising periods of wakefulness derived from the recorded movement of a participant whilst in bed.

A. *Hardware Deployment*

This study utilised wireless kinematic sensors with tri-axial accelerometers attached to specific points on a bed to identify the movements of the bed occupant during sleep. For the purposes of evaluating optimal accelerometer placement, five Shimmer v2.0r (Realtime Technologies, Ireland) sensor motes were deployed as summarised in Figure 1. This comprised of three mattress mounted accelerometer positions, one to the middle left, middle right and top centre of the mattress. Additionally, devices were placed centrally in the participant's duvet and pillowcase. Each unit recorded time stamped values of acceleration on the X, Y and Z axis, sampled at 51.2Hz and transmitted samples in real time via Bluetooth to a nearby laptop running a LabVIEW (National Instruments) application designed to capture the data.

B. *PSA Validation*

Originally selected for the PSA platform in [7], a Philips Respironics Actiwatch Spectrum (actiwatch) containing a tri-axial accelerometer sampling at 32Hz was employed for the purposes of validation. Data was captured and stored locally on the actiwatch in 30sec epochs pending transfer via USB to a computer running the Actiware software.

C. *Participants*

This paper presents the findings from an investigation involving three participants (one male, two female; age 21-30 years). All participants described themselves as healthy with no clinically diagnosed sleep problems and confirmed they would be sleeping in a single occupancy bed throughout the trial period. Subsequently, each participant was invited to record data for 10-nights over a 14-night period.

D. *Data Processing*

From a theoretical perspective, the actigraphy data gathered from the PSA platform contains the required information to determine a participant's sleep/wake status. Adapted from traditional sleep actigraphy techniques, our previous work implemented and evaluated a number of steps designed to extract features applicable to sleep/wake detection from the raw accelerometer signal [7, 8]. This process begins by converting the magnitude of the accelerometer signal into activity counts summarised in 30sec epochs to facilitate data synchronisation with the actiwatch. These activity counts are derived using the Time Above Threshold (TAT) method with a threshold of 0.1g and Digital Integration (DI) methods [7]. By utilising TAT and DI methods, the activity counts can be used to represent both the duration and amplitude of any movement within each epoch [3]. Next, a number of statistical features previously identified by Tilmanne *et al*. [12] and evaluated for the PSA platform in [7, 8] are extracted from both TAT and DI activity counts. These features, in addition to the TAT and DI activity counts, are time synchronised with the Sleep/Wake state reported by the Actiwatch for each epoch. The resulting baseline dataset can then be input directly into a nominated classifier or further manipulated as required.

E. *Data Undersampling and Oversampling*

When evaluating potential undersampling techniques, it is important to note that there may be instances that are important to achieving the classification objective. Therefore, there is a greater chance of optimising classification performance if discriminant instances can be identified and kept, while less discriminant instances may be marked for exclusion. Unfortunately, in this work there currently exists no benchmark on which to allocate the merit of one sleep instance over another. Subsequently, simple random selection was employed using the Spread Subsample (SS) filter within WEKA (University of Waikato, New Zealand) to achieve the desired sleep/wake class balance.

Some implementations of oversampling operate by replicating existing instances from the minority class, however, this approach has been criticised for introducing additional overfitting problems [11]. Focusing on Decision Trees (DT) given that they are implemented within the PSA platform, Chawla *et al*. [10] demonstrates that by oversampling through data replication, the effect is to isolate a smaller decision region in the feature space thus reducing the DT's ability to identify unknown instances outside the training set, effectively overfitting. A proposed solution exists in the Synthetic Minority Oversampling Technique (SMOTE), which has demonstrated substantial improvements over data replication [10, 11]. In this oversampling technique, synthetic data is generated within the scope of the minority class using a k nearest neighbour approach [10]. This effectively introduces new minority instances while preventing the overfitting issues previously discussed. Reporting good results in a number of classification modalities including DTs [10], the SMOTE oversampling method was applied to the baseline datasets.

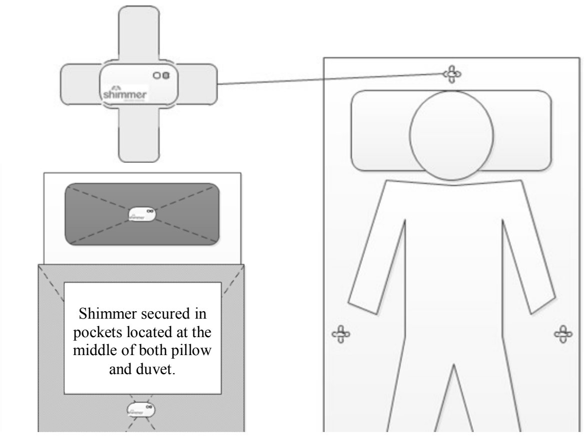

Figure 1. Standard deployment configuration for Shimmer accelerometers.

III. Results

Over the trial period, a total of 30 nights actigraphy data was collected, summarised into epochs using the features extracted from TAT and DI activity counts, then collated by accelerometer position and participant. Table I summarises this data and highlights the continuing class imbalance issue with only 11% of the data being labelled as wake.

TABLE I. Summary of sleep vs. wake data recorded

	P1		P2		P3		All	
DS	**Hrs**	**%**	**Hrs**	**%**	**Hrs**	**%**	**Hrs**	**%**
S	63.80	95.14	48.11	94.53	68.86	79.04	180.77	89.07
W	3.10	4.86	2.63	5.47	14.02	20.36	19.75	10.93
T	66.90	-	50.74	-	82.88	-	200.52	-

Note: S-Sleep, W-Wake, T-Total
Note: P1-Participant 1, P2-Participant 2, P3-Participant 3, All- Combined dataset of P1, P2 & P3

Previous work using the PSA platform [7] identified and evaluated a number of DT classification methods. Of these, the Random Forest (RF) achieved the best overall sleep/wake predictive performance. Subsequently, this same approach was adopted into the current work with sensitivity and specificity measures being calculated to determine if the RF classifier's performance may be increased through the use of SS and SMOTE techniques. Note that for this work sensitivity refers to the proportion of known sleep epochs that are positively identified as sleep. Likewise, specificity refers to the proportion of known wake epochs that are positively identified as wake.

Table II summarises the RF classification performance for each of the datasets by accelerometer position and participant. Excluding participant 3 the optimal positions are located in the pillow (PL) and middle right (MR). Note that all participants favoured sleeping on the right side of their bed, hence the MR accelerometers improved performance over the middle left (ML) position as discussed in [7,8]. Based on user experience feedback and classification performance, the MR accelerometer was nominated as the optimal position on which to evaluate SS and SMOTE sampling techniques. Table III provides a comparison of the baseline MR classification results against experiments involving SS, SMOTE and a combination of the two termed as hybrid. The impact of sampled data on classification performance was evaluated by applying a number of different scaling factors to the baseline dataset.

TABLE II. Summary of sensitivity and specificity results for the RF classifier by participant ID and accelerometer position

Position	Result Type	P1	P2	P3	All
Middle Left	Sen.	1.00	0.99	0.97	0.99
	Spe.	0.62	0.64	0.55	0.56
Middle Right	Sen.	0.99	0.99	0.97	0.99
	Spe.	0.65	0.66	0.56	0.57
Top	Sen.	1.00	0.99	0.97	0.99
	Spe.	0.61	0.63	0.54	0.53
Duvet	Sen.	0.99	0.99	0.97	0.98
	Spe	0.62	0.62	0.58	0.57
Pillow	Sen.	1.00	0.99	0.97	0.98
	Spe	0.65	0.65	0.58	0.58

TABLE III. Impact of oversampling, undersampling and a combination of both techniques on sleep/wake classification performance by participant for the MR accelerometer position.

Resample Type	Sampling Factor	Result Type	P1	P2	P3	All
N/A	Original	Sen.	0.99	0.99	0.97	0.99
		Spe.	0.65	0.64	0.57	0.57
Oversampling *Minority Class Increased By*	25%	Sen.	0.99	0.98	0.96	0.98
		Spe.	0.70	0.74	0.66	0.65
	50%	Sen.	0.99	0.99	0.96	0.98
		Spe.	0.69	0.73	0.66	0.65
	100%	Sen.	0.99	0.99	0.96	0.98
		Spe.	0.73	0.75	0.73	0.72
	200%	Sen.	0.99	0.99	0.95	0.98
		Spe	0.79	0.82	0.82	0.80
Undersampling *Majority Class Scaled By*	8:1	Sen.	0.99	0.98	0.97	0.98
		Spe.	0.69	0.74	0.56	0.59
	5:1	Sen.	0.98	0.97	0.97	0.97
		Spe.	0.71	0.77	0.57	0.63
	3:1	Sen.	0.96	0.96	0.94	0.95
		Spe.	0.74	0.79	0.63	0.68
	1:1	Sen.	0.89	0.89	0.82	0.86
		Spe	0.78	0.84	0.72	0.75
Hybrid *Mix of undersampling and oversampling*	25%	Sen.	0.99	0.99	0.96	0.98
	8:1	Spe.	0.70	0.76	0.62	0.62
	50%	Sen.	0.98	0.98	0.96	0.98
	5:1	Spe.	0.75	0.79	0.67	0.69
	100%	Sen.	0.97	0.97	0.96	0.97
	3:1	Spe.	0.80	0.83	0.75	0.78
	200%	Sen.	0.94*	0.95*	0.92*	0.93*
	1:1	Spe	0.88*	0.89*	0.85*	0.86*

Note: * best sensitivity and specificity performance achieved per dataset as indicated by accuracy

From Table III a universal improvement in specificity performance can be viewed as the class imbalance equalises. SS, however, has a corresponding decline in sensitivity performance as the scaling factor increases reducing the overall accuracy of the classifier. Overall, the hybrid approach performs the best by producing the highest overall specificity results whilst mitigating the sensitivity loss in SS.

IV. Discussion

As previously discussed, SS and SMOTE have been evaluated as potential techniques for increasing the ability of the PSA platform to correctly identify wake periods during sleep while maintaining high overall accuracy. The baseline dataset presented in this work continues to show evidence of overfitting due to class imbalance as evidenced through a high sensitivity value of up to 0.99 for the MR position and an overall sleep/wake balance of approximately 10:1.

Both SS and SMOTE produced improved specificity results as the number of sleep to wake instances is reduced. For a 1:1 sleep/wake radio, SS produced specificity values of up to 0.84, an improvement of 20% over the baseline performance. There is, however, an average decline of 12% in sensitivity performance across all datasets for this sampling factor thus reducing the overall accuracy of the PSA platform. As previously discussed, removing statistically relevant instances of sleep from the majority class may cause this. Future evaluations of undersampling for the PSA platform should consider including functionality for identifying statistically redundant sleep instances for removal. This feature selection may be implemented as an independent function; however, a wrapper approach, which utilises the predictive power of a nominated classifier to optimise the feature selection algorithm, as discussed by Guyon [13] may further improve classifier performance.

SMOTE produced improvements of up to 25% at an oversampling factor of 200%, however, with a slightly lower specificity performance of 0.82 when compared to SS. One particular advantage of oversampling is that no data is removed from the overall dataset avoiding the selection problems of SS. This is seen in Table III with only a minimal reduction of 2% in sensitivity performance. Continued evaluation of the SMOTE approach should consider further increasing the sampling factor. Chawla *et al*. [10] reports good results with SMOTE using a 500% increase in minority class size. It should, however, be noted that at higher sampling factors, SMOTE has been known to over generalise the data. Further work should consider evaluating adaptive synthetic sampling techniques such as Borderline-SMOTE [14]. In this approach, minority instances bordering the majority class that are often misclassified are reinforced through synthetic data. This mitigates the generalisation issues previously mentioned by limiting the amount of data generation that may be required.

Implementing a combination of SS and SMOTE techniques worked well and provided the best specificity results of 0.89, an increase of 28% over the baseline performance while somewhat mitigating the sensitivity loss caused by SS. At this stage the combination method would appear to be the optimal approach, however, it is important to note that its performance will be impacted by issues from both SS and SMOTE as previously discussed. This said, as discussed in [9], the subsequent benefits of tuning and combining the undersampling and oversampling approaches may be to further maximise the specificity and overall classification performance.

In conclusion, this preliminary evaluation has identified that the accurate identification of wakeful periods may be achieved by utilising undersampling and oversampling techniques. Furthermore, by employing a combination of the two methods, specificity performance can be further enhanced while mitigating the negative impact on sensitivity. Overall the results presented here are considered positive and a number of potential methods to further tune the current approach have been suggested. Moving forward, however, it is important to ensure that with the application of undersampling and oversampling techniques, the resulting datasets must remain representative of the real world data on which they are based.

References

[1] S. Ancoli-Israel and J.R. Cooke, “Prevalence and Comorbidity of Insomnia and Effect on Functioning in Elderly Populations,” *J. American Geriatric Society*, Vol. 53, No. S7, pp. S264-S271, 2005.

[2] M. Lee, A.C. Choh, E.W. Demerath, K.L. Knutson, D.L. Duren, R.J. Sherwood, S.S. Sun, WM.C. Chumlea, B. Towne, R.M. Siervogel and S.A. Czerwinski, “Sleep disturbance in relation to health-related quality of life in adults: The FELS longitudinal study,” *J. Nutrition, Health & Aging*, Vol. 13, No. 6, pp. 576-583, 2009.

[3] 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. 342-392, 2003.

[4] L. Walsh, E. Moloney and S. McLoone, “Identification of Nocturnal Movements During Sleep using the Non-Contact Under Mattress Bed Sensor,” in *Proc. 30th Annu. Int. Conf. EMBS*, Massachusetts, 2011, pp. 1660-1663.

[5] P. Chazal, N. Fox, E. O’Hare, C. Heneghan, A. Zaffaroni, P. Boyle, S. Smith, C. O’Connell and W. McNicohlas, “Sleep/wake measurement using a non-contact biomotion sensor,” *J. Sleep Res*., Vol.20, No. 1, pp. 356-366, 2009.

[6] A.M. Adami, M. Pavel, T.L. Hayes and C.M. Singer, “Detection of Movement in Bed Using Unobtrusive Load Cell Sensors,” *IEEE Trans. Information Technology in Biomedicine*, Vol. 14, No. 2, pp. 481-490, 2010.

[7] A. McDowell, M.P. Donnelly, C.D. Nugent and M.J. McGrath, “Passive Sleep Actigraphy: Evaluating a Non-contact Method of Monitoring Sleep,” *Lecture Notes in Computer Science,* No. 7251, pp. 157-164, 2012.

[8] A. McDowell, M.P. Donnelly, C.D. Nugent and M.J. McGrath, “Utilising wireless sensor networks towards establishing a method of sleep profiling,” *Int J. Computers in Healthcare,* Vol. 1, No. 4, pp. 346-363, 2012.

[9] N.V. Chawla, “Data Mining for Imbalanced Datasets: An Overview,” in *Data Mining and Knowledge Discovery Handbook*, 2nd Ed., O. Maimon and L. Rokach Ed., New York: Springer Science + Business Media, 2010, ch. 45, pp. 875-886.

[10] N.V. Chawla, K.W. Bowyer, L.O. Hall and W.P. Kegelmeyer, “SMOTE: Synthetic Minority Over-sampling Technique,” *J. Artificial Intelligence*, Vol. 16, No. 1, pp. 321-357, 2002.

[11] H. He and E.A. Garcia, “Learning from Imbalanced Data”, *IEEE Trans. Knowledge and Data Engineering*, Vol. 21, No. 9, pp. 1263-1284, 2009.

[12] J. Tilmanne, J. Urbain, M.V. Kothare, A.V. Wouwer and S.V. Kothare, “Algorithms for sleep-wake identification using actigraphy: a comparative study and new results,” *J. Sleep Res*., Vol. 18*,* No. 1, pp. 85-98, 2009.

[13] I. Guyon and A. Elisseeff, “An introduction to variable and feature selection,” *J. Machine Learning Research,* Vol. 3, pp. 1157-1182, 2003.

[14] H. Han, W.Y. Wang, and B.H. Mao, “Borderline-SMOTE: A New Over-Sampling Method in Imbalanced Data Sets Learning,” in *Proc. Int. Conf. Intelligent Computing*, pp. 878-887, 2005.