

SLEEP MONITORING DATA VALIDATION USING A LINEAR DISCRIMINANT CLASSIFIER

J. M. Perez-Macias¹, Misha Pavel^{1,2}, Ilkka Korhonen¹, Holly Jimison^{1,2}

¹Department of Signal Processing, Tampere University of Technology, Tampere, FI

²College of Computer & Information, Northeastern University, MA, USA

jose.perez-macias.ENG@ieee.org

Abstract – Sleep quality has an important impact in overall health. There are several new commercially available devices for sleep tracking. Although they have been tested in sleep laboratories, for some users, estimates are reliable whereas for others are very different. In this paper we present a framework to predict the error range of the total sleep time. We studied 23 subjects' sleep patterns using Fitbit One and Beddit Pro devices. We used Beddit Pro as our reference devices to verify Fitbit's reliability. Previous analyses found that for some users the agreement between systems was high, whereas for others was low. We used a linear discriminant to predict this behavior in the Fitbit data. 72.7% accuracy was achieved. Our results suggest that reliability of sleep estimates could be estimated without the use of a second device.

I. INTRODUCTION

Sleep is an important factor in our health and wellbeing. Poor sleep quality has been related to obesity, diabetes, heart diseases, dementia, and depression [1]. An important step in improving sleep is to increase our awareness via reliable assessment. Polysomnography (PSG) is considered the *de facto* gold reference for sleep monitoring [2]. However, its high cost and the disturbance of the usual bed environment has encouraged the development of new monitoring sleep devices in the home. The accuracy of these systems varies under a variety of contexts. Our lab has tested Beddit Pro, Fitbit One, Traxmeet, Firstbeat devices and self-reporting for comparative analysis to characterize the accuracy of the different types of sleep monitoring approaches. The objective of this paper is to study the ability to predict the reliability of Fitbit measurements alone, by using Beddit Pro as reference.

II. SUBJECTS AND SIGNALS

This study was conducted involving 23 healthy subjects (18 male and 5 females). Subject's age ranged from 21 to 31 years. They were recruited from the Tampere University of Technology in Finland by word of mouth and by advertisement in internal mailing lists. Inclusion criteria included age 18-65 years and good health. Participants who shared the bed with a partner, had a BMI > 28, or were pregnant were excluded from the trial. The trial lasted from 7 to 10 consecutive days.

Participants were instructed to carry a Fitbit One (Fitbit Inc., CA, USA) in a pocket or on the wrist during the day and to wear it in a band on the wrist and activate the *night mode* when "they decided to go to sleep" and after watching television or reading. The Beddit Pro (Beddit, Espoo, Finland) sensor strip was installed directly under the bed sheets, and the set-top box next to the bed. All volunteers agreed to participate in this research and signed a written informed consent.

III. METHODS

A. Data model

A linear model with subject-dependent parameters, $f = c_i + d_i b$, was used to estimate our ability to harmonize data from different devices. We computed a robust M-estimate regression using a Tukey biweight influence function on the total sleep time (TST) estimates of N-1 subjects [3]. The resulting regression was evaluated on the remaining subject's estimates by computing the root mean squared error (RMSE) and the coefficient of determination $R^2 = 1 - \sigma_{res}^2 / \sigma_f^2$,

where σ_{res}^2 , σ_f^2 are the residual and total variances, respectively. We divided the subjects in two groups: a first group with an error under 6% of the all-subjects total sleep time average (TST_{avg}), and a second group with the remaining subjects.

B. Features

To avoid overfitting and given our small sample, we selected three features: the median of the sleep efficiency (SE_{med}), and the standard deviations of the total sleep time and sleep efficiency (TST_{std}, SE_{std}) to spot the differences between the distribution of the amplitude values of the time series.

C. Classification

A linear discriminant [3] was used to perform the classification. We used a leave-one-out cross-validation (loo-cv) method to train and test the classifier. We normalized each set of features to zero mean and a standard deviation

of one; these values were stored applied to normalize each test measurement.

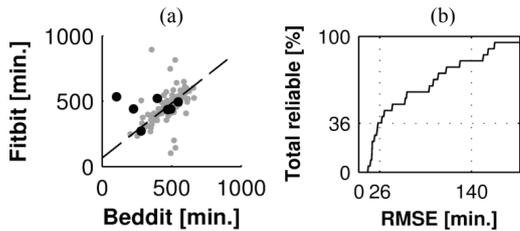


Figure 1: (a) Example of one of the fittings for TST; black dots represent the values of the tested subject and grey dots represent the values of all other subjects, (b) RMSE histogram of all nights and all subjects.

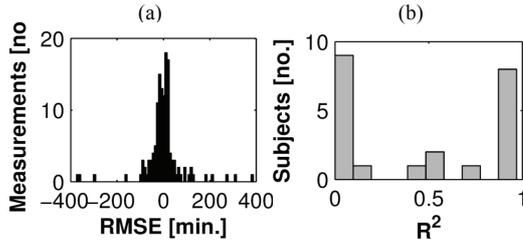


Figure 2: (a) Percentage of subjects versus RMSE. (b) R^2 histogram of the fit.

D. Statistical analysis

Patient data on clinical and features were expressed as mean \pm SD. We computed the sensitivity (Se, proportion of subject's reliable data, correctly classified), specificity (Sp, proportion of subject's unreliable data correctly classified), and accuracy (Acc, percentage of subject's correctly classified over the entire sample). The statistical analyses were performed using the software package MATLAB (version 7.0; Mathworks, Natick, MA).

IV. RESULTS

A total of 23 subjects, 5 women and 18 men (mean age 25.8 ± 2.9 years, height 172.2 ± 24.7 cm, mass 74.3 ± 12.5 kg), participated in the trial. One subject was excluded from the study due to a device malfunction. A total of 138 nights were recorded. Average Beddit-TST time was 469.9 (SD ± 96.0) min and SE was 87.8% (SD $\pm 9.2\%$). The first group (RMSE < 26 min.) consisted in 8 subjects, with an average TST 485.6 (SD ± 51.8 min.) and an average SE 90.2% (SD $\pm 4.9\%$). The second group consisted in 14 subjects, with an average TST 431.1 (SD ± 122.43 min.) and an average SE of 92.1% (SD $\pm 3.4\%$). Both Kolmorow-Smirnov and the Shapiro-Wilk tests were used to assess normality of each input feature distribution. Homoscedasticity was verified with Levene's test.

Fig. 1a displays the estimates for one subject and the M-estimate regression for the remaining N-1 subjects. Fig. 1b shows the

percentage of subjects with less than the specified RMSE. Fig. 2a shows the distribution of the errors for all nights and subjects. We visually found a fair amount of outliers (around 14%). The distribution of the R^2 is displayed in Fig. 2b. We aimed to classify those subjects under approximately 6% RMSE of the TST_{avg} (26 min.), corresponding with approximately 36% per cent of the subjects. Discriminant analysis was used to classify three different types of input patterns. Sensitivity, specificity and accuracy were 64.3%, 87.5% and 72.7%, respectively. All results were achieved following loo-cv.

IV. DISCUSSION

The aim of this study was to assess the validity of sleep-recorded data from Fitbit One using only this sensor estimates. We considered Beddit Pro as our reference to train our data. This bed sensor promises to be precise classifying sleep stages [4]. In our study, around 14% of the measurements were clearly outliers, and we used robust estimation techniques to lessen their influence. Most of them were a result of typical sleep related behaviors, e.g. lousy sleeping, watching television in bed, Fitbit misuse, etc. Using this method, we were able to identify those subjects with an error less than 6% of the TST_{avg}, achieving 72.7% of accuracy.

V. CONCLUSION

The main conclusion based on the present study is that machine-learning methods could be used to identify reliable Fitbit One measurements when compared with Beddit Pro. These conclusions are bound to some limitations: the amount of measured nights and number of subjects was limited, and the participant set comprised only of young healthy adults.

VI. REFERENCES

- [1] Mahowald, M. W. (2007). Sleep Disorders and Sleep Deprivation: An Unmet Public Health Problem. *New England Journal of Medicine*, 356(2), 199-200.
- [2] A. Chesson et al. Practice parameters for the indications for polysomnography and related procedures. In *Sleep*, volume 20, pages 406-422, 1997.
- [3] Huber, P.J. (1981). *Robust Statistics*. Wiley, New York.
- [4] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York, NY: Springer, 2006.
- [5] Paalasmaa, J., Waris, M., Toivonen, H., Leppakorpi, L., & Partinen, M. (2012). Unobtrusive online monitoring of sleep at home (pp. 3784-3788). Presented at the Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE.