Monitoring Sleep and Detecting Irregular Nights through Unconstrained Smartphone Sensing

Ke Huang,[1] Xiang Ding,[1] Jing Xu,[1] Guanling Chen,[1] Wei Ding[2]
[1]Department of Computer Science, University of Massachusetts Lowell, USA
[2]Department of Computer Science, University of Massachusetts Boston, USA
{khuang, xding, jxu, glchen}@cs.uml.edu, ding@cs.umb.edu

***Abstract*—Sleep is essential for a person's health and well-being. Recent advances of wearable devices and smartphone sensing have led to the proliferation of at-home sleep monitoring solutions for the consumer market. In this paper, we study how to monitor basic sleep behavior and how to detect irregular sleep nights, through unconstrained smartphone sensing, which can serve as an important indicator for both mental and physical health if the sleep problems persist. We first propose a supervised learning approach to predict bedtime and sleep duration with a light-weight context sensing schedule to minimize battery consumption. The proposed solution is validated through an extensive user study, and the prediction accuracy of bedtime and sleep duration significantly outperformed the state-of-art solution. In addition, we propose an unsupervised approach to detect irregular sleep nights by profiling and detecting contextual variations. The experiment results show that the proposed solution is effective in detecting irregular sleep nights. To the best of our knowledge, this is the first work that uses unconstrained smartphone sensing to detect sleep pattern changes with the benefits of reduced training efforts and improved robustness against behavior diversity.**

***Keywords*-Sleep Monitoring, Smartphone Sensing, Context Awareness, Anomaly Detection, Mobile Health.**

I. Introduction

Regardless of the age, good restorative sleep is essential to physical health and emotional well-being. Notably, maintaining a regular bedtime and sufficient sleep is a key aspect of healthy life, yet millions of people are suffering from sleep disorder or chronic sleep deprivation around the world [1]. These sleep problems are associated with the onset of a number of chronic diseases and conditions, such as diabetes, cardiovascular disease, obesity, and depression [2], [3], and even affect memory and cognitive functions [4].

In particular, a change of sleep patterns is often an important behavior indicator for potential health problems. For instance, sleep disorder is one of the most common effects of depression, which is a brain illness and can affect sleep-wake cycle to cause more irregular sleep nights. Studies have shown that sleep disturbance is a major risk factor of developing major depression [5]. Interestingly, evidences also exist that erratic sleep/wake schedules, often towards end of school semesters, are positively associated with stress and negatively associated with academic performance for adolescents [6], [7]. *In this paper, we study smartphone sensing for unconstrained sleep monitoring, which can serve as a scalable solution to address these societal challenges. We propose a supervised learning approach to predict bedtime and sleep duration, and a novel unsupervised context-profiling approach to detect irregular sleep nights that may suggest change of sleep patterns.*

There are many existing sleep assessment and monitoring solutions. The polysomnography is the accepted gold standard for sleep assessment [8], but it requires professional monitoring and expensive equipment. Recent advances of wearable devices make them accessible to people who are interested in sleep assessment at home. These devices typically have embedded accelerometer (e.g. Fitbit [9]) or ECG electrodes (e.g. Zeo [10]) and can track sleep quality based on limb movements or brain waves. The use of these devices are intrusive and cumbersome, as users need to wear them during the entire bedtime and the quality of analysis depends on how well user wears the device [11].

On the other hand, smartphones have increasingly been used as a health device since they have many embedded sensors and people frequently carry it with them [12]. There are existing mobile applications available that can track sleep quality simply using smartphones without any extra device. Users typically need to place the phone on the bed so the application can analyze body motion [13], or users need to plug in the phone and explicitly start the app to analyze ambient sound (e.g. snoring etc.) for sleep quality analysis [14], [15]. While existing wearable devices and smartphone sensing applications can provide deep analysis of sleep stages and quality for several nights, they are often not designed for long-term sleep monitoring. A recent survey shows that many people are interested in the technology of unobtrusive sleep tracking but are resistant to wear a device during sleep [16]. The smartphone-based sleep tracking applications also require explicit user cooperation (e.g. plug in, put on bed, start/end app) and are difficult to comply over a long period of time.

With the research objectives of long-term sleep monitoring and detecting irregular sleep nights (reflecting sleep disturbance), our work focuses on *unconstrained* smartphone sensing without requiring explicit user cooperation and actions. *The contributions of this paper include*:

1) We propose an unconstrained smartphone-sensing approach to predict bedtime and sleep duration through classification of context features. Our sensing strategy has minimal impact on the battery, typically consuming less than 5% smartphone energy throughout a day, and the supervised learning algorithm is validated through a user study of 18 participants over 14 weeks who have diverse sleep patterns. The experiment results show that the proposed algorithm significantly outperforms a state-of-art solution in prediction accuracy of bedtime and sleep duration.

2) Further we propose a new unsupervised context profiling approach to detect irregular sleep nights to reduce training effort and to improve robustness against diverse sleep patterns. We observed that there is a strong correlation between context profiles and sleep patterns, and the experiment results validated that the proposed method is effective detecting irregular sleep nights. *To the best of our knowledge, this is the first work that uses unconstrained smartphone sensing to detect sleep pattern changes, such as caused by elevated stress or other health problems, which is practical and suitable for long-term sleep monitoring.*

II. Related Work

We summarize relevant academia research and industry products for sleep monitoring from two aspects: 1) solutions requiring dedicated devices; and 2) smartphone-only approaches leveraging onboard sensors.

On the consumer fitness tracking market, many commercial devices are available for sleep tracking. Zeo [10] is a head-mounted sensor that can monitor brain and muscle signals to track sleep through polysomnography. While being accurate, it is also a very intrusive approach. Basis [23] is a wristwatch that packs many sensors to track steps, sleep, heart rate, perspiration, and skin temperature. Fitbit [9] and Jawbone Up [24] are another two popular wrist-worn bands for sleep monitoring. Most of these devices use actigraphy for sleep-wake assessment, which was found to be good at detecting sleep episodes but not wake episodes for healthy populations [25]. Thus some of these wearable sensors require user cooperation to explicit tell them the bedtime and waketime (e.g. by pressing a button) to provide accurate sleep analysis. A common problem of these devices is that user needs to wear the sensor to collect data, which is inconvenient for long-term sleep tracking and uncomfortable for those who already have trouble sleeping.

Rather than having users wear a monitoring device, Beddit [26] is a sensor placed under the bed sheet, for unobtrusive monitoring, with data transmitted to a nearby smartphone through Bluetooth. Sense [17] is a dedicated device placed at the bedside table, collecting environment data (noise, light, temperature, humidity and particles in the air) and communicating with another small sensor clipped to the pillow. Academic researchers have also attempted non-invasive sleep monitoring through instrumented environment. Metsis et al. combine data from pressure mattress and Kinect to provide detailed analysis of users' sleep patterns, including both posture and motion [19]. Lullaby combines many off-the-shelf sensors as a dedicated system to be placed near the bed, which provides a comprehensive recording of a person's sleep to help users to understand the relation between sleep quality and environmental conditions [18]. Hoque et al. propose to attach accelerometers to the bed mattress to infer body position and movements [27].

In addition to dedicated devices, another field of sleep related research leverages smartphone as a powerful computational and sensing tool for the purpose of sleep tracking. Behar et al. conducted a review of sleep screening apps for smartphones [11]. In particular, several smartphone apps use onboard accelerometer and microphone to provide sleep analysis by simply placing the phone on the bed, such as Sleep As Android [13], Sleep Cycle [28], and Sleepbot [29]. Using a similar approach, Gu et al. evaluate the user's sleep quality by measuring the durations of different sleep stages in a sleep process rather than recording some certain sleep-related activities [15]. These apps all require user cooperation to place their phone besides his/her head on the bed. In addition, there is a risk for the phone falls out of the bed and people may be concerned with the battery usage the radiation problem affecting their health, making it not an ideal solution for long-term sleep tracking [16]. iSleep [14] addresses some of these concerns by focusing on detecting acoustic events, rather than motion states, to predict sleep quality so the phone does not need to be placed on the bed. However, users are required to manually start/end the app and put the phone near the bed (e.g. night stand).

Recently researchers have started to focus on unconstrained sleep monitoring through smartphone sensing, taking advantage of that many people are already leaving their phone on during sleep and keeping the phone in the same room [12]. Chen et al. propose a linear regress model (BES) to infer sleep duration using smartphone sensing [20]. Currently the BES model predicts only sleep duration and it has been used to monitor students' long-term sleep trend [7]. Min et al. propose Toss 'N' Turn (TNT) with a more intense sensing schedule for sleep state and quality classification [21]. TNT achieved 35-minute and 49-minute accuracy for bedtime and sleep duration prediction, respectively. Interestingly TNT preserves phone's power by automatically switching to lower sensing intensity when the battery is low. Abdullah et al. developed a simpler method to detect bedtime and waketime of college students, who are usually heavy smartphone users, by finding the longest non-use segment (NUS) after 10pm [22].

Motivated by the objective of easy-to-use and long-term sleep monitoring, we propose a new supervised learning approach for smartphone sensing that outperformed existing solutions for predicting bedtime and sleep duration, and we further propose an unsupervised context-profiling approach to detect irregular sleep nights that has not been addressed in the existing literature.

III. Data Collection and Dataset

We developed an Android app, installed on participants' own mobile phones, to collect the sensing data and provide a user interface for participants to log their bedtime and wakeup time as ground truth for experiment evaluations. A simple questionnaire was also presented to collect participants' demographic data when the app starts at the first time. The participants were given a tutorial of the data collector app and they were asked to use their phone as usual.

A. Participants

We recruited 23 participants for a 14-week study. During the first 2 weeks, 5 participants dropped out of the study and the results we show are based on data from the other 18 participants. Among those 18 participants, the demographic data shows that 7 female joined the study, and 6 lived on campus. There were 8 undergraduate students, 8 graduate students and 2 faculty/staff participants. We had 9 participants who were 18 to 24 years old, 7 participants between 25 to 35, and the other 2 more than 35 years old. On the questionnaire, the participants were asked when they usually go to bed. There were 3, 11, 2 and 2 participants who responded that they usually go to bed between 8pm to 10pm, 10pm to midnight, after midnight, or have an irregular sleep schedule, respectively. Each participant was rewarded with a US$20 gift card every 4 weeks to keep them engaged.

B. Sensing Strategy

To minimize the energy impact on participants' phone, our collector app adopts two sensing strategies. The first sensing strategy is coarse grained that contextual data is collected every 5 minutes. And a fine grained sensing strategy tries to collect sensing data every 30 seconds. The fine grained sensing data is processed and summarized every 5 minutes. The summarized data and coarse grained sensing data compose a *context record* every 5 minutes. A context record has 4 groups of data that are temporal, spatial (WiFi signatures), phone status, and app usage. Once the context sensing started, all four categories of contextual data were collected and stored. We describe details of the context records in Section IV-A. The bedtime and wakeup time logged by participants were also stored.

The collector app uploads the context records to our server via WiFi connection on a daily basis. In order to protect participants' privacy, we used the phone's one-way hashed device ID to distinguish each participant. The MAC addresses of APs were also hashed to avoid potential inference of location. We only logged decibel sound level values without recording raw audio content.

Overall our collector app consumed little energy - often less than 5% on our testing devices under normal usage. Throughout the study, we did not receive any complains from the participants regarding the power consumption of the app.

C. Data Characterization

The dataset we collected has a total of 974 days of logged sleep episodes and more than 253,000 context records. Table I shows the statistics of the data collected from each participant. The compliance rate varied significantly, and 3 participants logged less than 30 days, while 7 participants logged more than 2 months. We also noticed the data collected from participant 8 missed illuminance values, possibly due to a broken sensor.

IV. Predicting Bedtime and Sleep Duration

If a user's bedtime and waketime, from which the sleep duration is calculated as their difference, can be accurately predicted, it is then possible to detect irregular sleep nights by analyzing the bedtime and waketime patterns. Thus in this section we focus on predicting users' bedtime and waketime using supervised learning algorithms. We first describe the details of the contextual features. Then we compare the performance of different feature combinations and classification algorithms. Finally we present experiment results and analysis of the large prediction errors on bedtime and sleep duration.

A. Features for Classification

A context record is obtained by our collector app every 5 minutes and it contains four types of context data as follows. Each context record consists of the coarse grained and processed fine grained sensing data.

1. Temporal Context: People's sleep tends to follow a circadian rhythm or body clock. Hence temporal context potentially is an important feature to learn a person's sleep pattern. The temporal data, which is coarse grained data, we collected including day of week and hour in a day for the corresponding context record.

2. Phone Status: The phone status is defined by 7 raw data items including charging status, phone's power level, environmental decibel value (dB), illuminance values (Ev), proximity value, duration of screen on, and phone movement count detected by accelerometer. We collect charging status, power level, and proximity sensing data once every 5 minutes. While both decibel value and illuminance value are collected every 30 seconds, and the min, max, average and std values are calculated every 5 minutes as processed sensing data. The decibel value is calculated by analyzing a 3-second audio from microphone and the illuminance values

Table I
PARTICIPATED DAYS AND NUMBER OF SENSED CONTEXT RECORDS COLLECTED FROM 18 PARTICIPANTS

Participant ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Participated Days	19	36	32	89	32	33	30	25	56	33	88	100	26	81	85	49	78	82
Total Records	4100	9518	8117	24828	8627	8177	7047	6486	14875	9072	22226	27115	5352	22538	22233	11522	20335	21251
Sleep Records	1629	3039	3020	9164	2944	3024	2791	2460	5142	3257	7216	9648	2441	8561	8313	5069	7582	8462

are collected within 3 seconds for every 30-second sensing period. The screen on time and movement count sensing data are the cumulate values of each 5-minute sensing period. These contextual data are intuitively relevant to sleep as people usually sleep in a dark and quiet environment. We collect phone's power level data since some people like to charge their phones during sleep time. The screen-on time and the movement count are strong indicator of whether a user is using a phone and awake.

3. Spatial Context: Intuitively a user's location associates strongly with sleep, since most of time we sleep at home. Due to concerns of privacy and power consumption, we did not use GPS to locate the phone. Instead, we used nearby WiFi access points (AP) and their signal strength as a *signature* to get a location at room level [30]. The app scans for APs for 15 seconds during each context sensing period and stores the scanning result as a location signature.

4. App Usage: App usage is another feature that may be relevant to sleep. For instance, when people go to bed they may launch the Alarm app and set the alarm. Others may like to read some news or use social apps before sleep. Our collector app checks the foreground app every second when the screen is on and the app usage statistics are stored in a context record.

B. Classification Performance of Features

1) Feature Combinations: Different context features may have different significance for learning sleep patterns. We conducted a thorough performance evaluation on different combinations, which are 15 in total, of these four context categories. We chose SVM with linear kernel as the classifier to evaluate the performance of different context combinations. A 10-fold cross-validation was performed on all 18 participants' data individually with 15 different combinations. We found that both temporal and phone status context played a significant role in the classification. The context combinations including those two context categories always outperformed the combinations without them. Another interesting finding is that the temporal context is not the dominant feature. The prediction accuracy decreased only a little bit without using the temporal context.

On the other hand, the spatial and app usage context are not significant for some participants. For some participants, the prediction performance actually decreased when combining these context with other significant features, maybe due to the curse of dimensionality. While for other participants, adding these context has a positive effect on the prediction. In summary, the temporal and phone status feature groups are consistently important across participants. Next we further study how individual features in these two groups contribute to the classification accuracy for different participants.

2) Chi-squared Test on Features: We conducted χ^2 (Chi-squared) tests on 15 individual context features for temporal and phone status feature groups, which are day of week, hour, movement count, max, min, average and std decibel values, max, min, average, and std illuminance values, charging status, power level, proximity value, and duration of screen on. The χ^2 statistic measures dependence between stochastic variables, so a transformer based on this function "weeds out" the features that are most likely independent of class and therefore irrelevant for prediction.

We conducted the χ^2 test on those 15 context for all participants The result shows that almost all context's P_{value} are significant ($p_{value} \leq 0.05$). Within these context, the p_{value} of day of week is significant for some participants while not significant for others. And the STD values of illuminance and decibel are not significant for all participants. These results suggest that people do follow a circadian rhythm for sleep and the ambient environment, like light and sound, has a strong correlation to the sleep patterns. The p_{value} of day of week, however, contributed differently across participants. Some may sleep regularly despite weekdays or weekends while others do have a different pattern. These observations are similar to findings in existing literature showing people's sleep behaviors vary significantly, and individual predictive models are preferable since a general model learned from a diverse population will perform poorly when applied to individuals [20], [21].

C. Performance of Classification Algorithms

We next evaluate the performance of three widely used classification algorithms, including SVM, Logistic Regression and Random Forest, with out dataset. We chose all four categories of context data as features, for these algorithms to classify whether the user is asleep or not at the time of each context record was collected. Based on the observations above, we constructed individual models using participants' own data. Figure 1 shows the average classification accuracy of three algorithms using 10-fold cross validation.

The performance result shows that Random Forest (RF) classifier outperformed the other two algorithms, and it

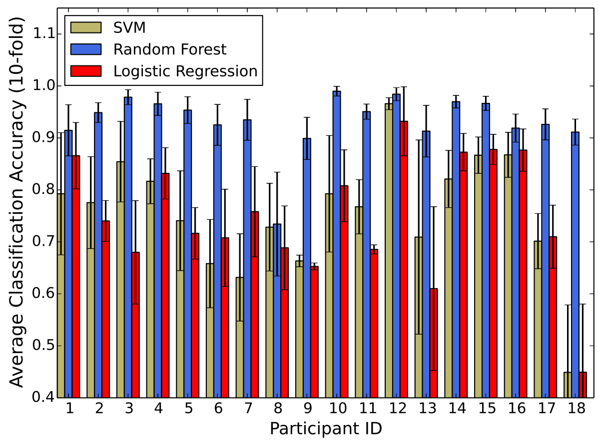

Figure 1. Comparing the prediction accuracy of three classification algorithms.

achieved more than 91% accuracy for 17 participants. For participant 8, the RF classifier achieved 74% accuracy, still better than the other two algorithms, mostly due to the lack of illuminance data. Given its supreme performance and its built-in feature selection capability, we chose RF classifier for further evaluation in the rest of the paper.

D. Prediction Performance of Bedtime & Sleep Duration

For each context record, the RF classifier classifies whether the user is asleep at the time when the context record was obtained. We then connect consecutive context records with the same classification results into asleep or awake segments. If the interval between two consecutive asleep segments is less than 30 minutes, we combined them into a single asleep segment. The gaps between asleep segments were due to misclassification, where a small number of context records were classified as awake. For example, the light went on when the participant used the bathroom at night. While technically the classifier was not wrong, we chose to ignore such cases as we are only concerned with long-term trends of bedtime and sleep duration for the purpose of this study. We then use the asleep segment as the predicted bedtime and sleep duration, and compare them with the user logged sleep data for evaluation.

Figure 2 shows the cumulative distributions of bedtime and sleep duration errors with leave-one-day-out cross-validation for all participants. The average bedtime error is ± 24.0 minutes and average sleep duration error is ± 40.7 minutes. For a typical 8-hour sleep, 40 minutes sleep duration error is about 8% of the entire sleep. We can see that 80% of bedtime errors are within ± 40 minutes and the sleep duration errors are within ± 70 minutes, respectively.

For comparison, we also implemented the longest non-use segment (NUS) method, which uses phone usage as a hint to infer bedtime and sleep duration as proposed in [22] between 10pm to 10am. With our dataset, the average

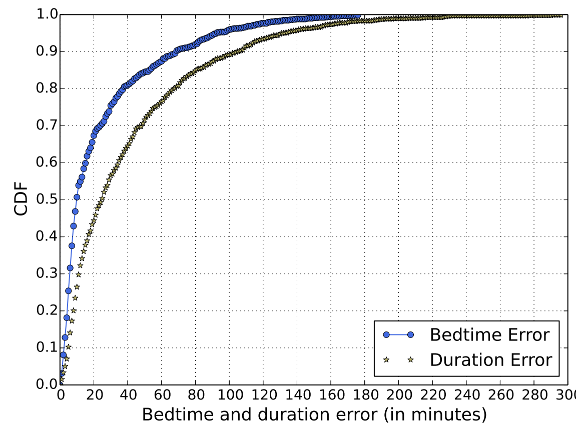

Figure 2. Overall predicted bedtime error and sleep duration error CDF.

bedtime error for the NUS method was ± 26.0 minutes and the average sleep duration error was ± 86.3 minutes. While NUS is a light-weight sleep inference method and there are optimizations to further improve its performance, it is not appropriate for all population, particularly as our participants had significant varying sleep behaviors. Figure 3 shows the comparison between average NUS predicted sleep duration and average user logged sleep duration for 8 participant who logged more than 50 nights sleep. An interesting finding is that the NUS predicted duration is always less than the logged sleep duration. This is because NUS depends on the app usage, a sleep will be segmented if a user used the phone or the phone received an notification during sleep. We also observe that the variation of NUS is larger than the logged sleeps.

E. Error Analysis and Sleep Variations

Though 80% of predicted sleeps have a small bedtime error and sleep duration error, there are still 20% of sleeps have a large bedtime error or duration error. The false positive and false negative mainly happened during a certain time before participants went to bed or after they woke up. As we can see in Figure 2, the duration error is nearly the double of the bedtime error for the same cumulative probability. It is because the waketime error is in the same scale of bedtime error, thus the sleep duration has both errors included.

We also investigated the causes of those big errors. First of all, it comes from mis-classifications before bedtime and after waketime when the context was not significantly different than that of sleep time. For example, a participant may wake up and walk out of bedroom without taking the phone (e.g. to bathroom for cleaning up or to kitchen for breakfast). The context sensed by the phone prior and after waketime are similar and thus confused the classifier. The same situation may occur around the bedtime, when the phone is in the bedroom while the participant being absent. We expect the classification will improve as users become

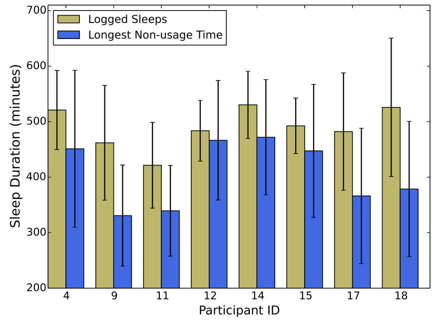

Figure 3. Comparing logged and NUS duration.

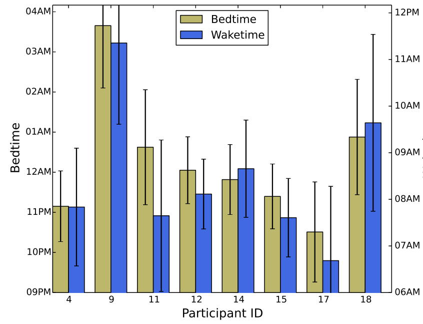

Figure 4. Sleep statistic of 8 participants.

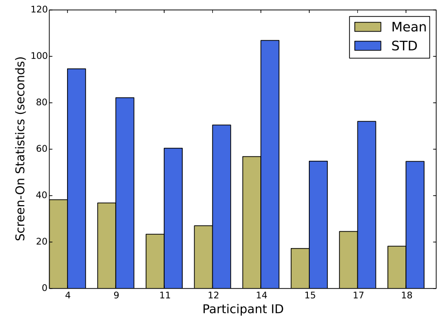

Figure 5. Screen-on statistic of 8 participants.

more engaged with the smartphone, with habitual checking before sleep and after wakeup.

Secondly, our participants had significant varying sleep behaviors. Figure 4 shows the statistics of the bedtime and waketime for the participants logged more than 50 nights sleep. We can see that the sleep patterns varied greatly across participants. Figure 5 shows the statistics of screen on time between 10PM and 10AM next morning for those participants. Again a great variance can be observed. These two figures indicate irregularity of sleep behaviors. Irregular sleeps not only reduce the prediction accuracy but also introduce big errors. For instance, participant 9 and 18 had some very irregular sleep nights, when the bedtime was in the early morning (3AM) and the sleep duration was more than 11 hours. These irregular sleep nights led to irregular context feature values, and it is thus not surprising that the classification results were not ideal and both bedtime and duration errors were large.

Thirdly, the prediction accuracy highly depends on the size of training set and regularity of sleep behavior. We chose participant 11 and 12 as examples. These two participants have similar high average prediction accuracy which is about 94%. While participant 11 has a more volatile sleep patterns than participant 12 and there are more big errors as well. Since participant 11 had more irregular sleeps, those irregular sleeps bring too much noise data into the training set and thus the big error is more than participant 12. Figure 6 shows the correlation between variance of logged bedtime and the prediction errors. We can observe that while the prediction errors remained small despite increasing bedtime variance for some cases, there are non-trivial amount of cases where the prediction errors increased with the bedtime variance.

In summary, the diverse sleep behaviors impose a significant research challenge for unconstrained sleep monitoring. While the above classification algorithm is suitable for predicting bedtime and sleep duration, we need a new solution to detect irregular sleep nights to cope with the natural sleep variations. Next we propose an unsupervised algorithm using statistical context profiling for this purpose.

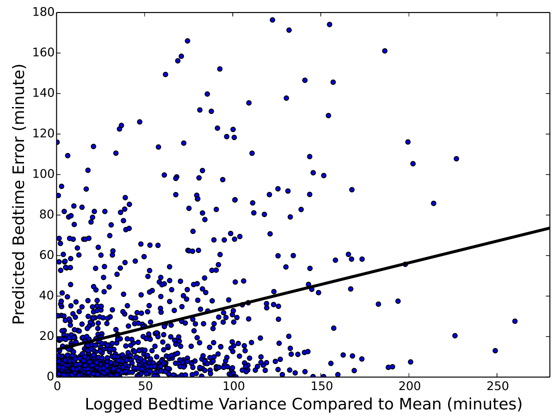

Figure 6. Correlation between variance of logged bedtime and the prediction errors (in minutes).

V. Unsupervised Irregular Sleep Detection

One of our research goals is to detect irregular sleep nights as a potential warning indicator of sleep pattern changes that may be symptoms of some health problems. A possible approach is to monitor the user's bedtime and sleep duration predicted by the supervised learning, from which the irregularity can then be detected through sequential analysis. According to the extensive analysis in Section IV, however, supervised sleep prediction suffers two problems when used to detect irregular sleep nights. First, despite good prediction accuracy of bedtime and sleep duration, large errors persist and have strong association with the sleep variance. Namely, the more "irregular" of the sleep (i.e. variance is large) is, the larger error is produced by the prediction algorithms (see Figure 6). Second, supervised learning requires users to label the ground truth for training data, and irregular sleep nights demands more labeled data to train the learning model. It is difficult for users to reliably label the data over a long period of time (e.g. 3 or more weeks). A machine learning model with insufficient training data, however, will produce more prediction errors, making it difficult to detect irregular nights based on inaccurate predictions.

To address these challenges, we propose an unsupervised approach to detect irregular sleep nights. Our observation is that context features are strongly associated with the sleep patterns, as discussed in Section IV-B. Thus to detect irregular sleep nights, instead of trying to predict bedtime and sleep duration, we directly monitor the context changes. If the context features show irregularity, we consider sleep at that night is also irregular.

Next we describe the unsupervised context profiling approach and the algorithm to detect irregular sleep nights based on context variations. Experiment results that validated our approach are also presented.

A. Irregular Sleep and Context Profiling

We take a general approach to define "irregularity" as the statistical outliers. Assuming one's sleep/wakeup time and sleep duration follow normal distributions, the data points that are τ standard deviations away from the mean can be considered as the outliers, thus irregular. The threshold τ can be obtained empirically based on problem domains. We need to learn from the user logged sleeps to define regular sleeps. Since learning samples also contain outliers (irregular sleeps) and the calculation of mean and standard deviation (STD) is sensitive to those outliers, we need to weed out logged outliers first. A very robust scale estimator is the Median Absolute Deviation (MAD) proposed by Hampel [31]. While MAD is inefficient at Gaussian distribution and it computes symmetric statistic, it cannot deal with skewness (an irregular sleep is often later than usual time and longer than usual duration). Rousseeue and Crous proposed alternatives to the MAD based on pairwise differences, and we adopted the S_n estimator [32]:

$$S_n = c * med_i\{med_j|x_i - x_j|\}$$

Where c is a constant and has value 1.1926. For a given sleep x_i, we first calculate the median of this sleep to all samples x_j, which is $med_j|x_i - x_j|$, then if $med_j|x_i - x_j|/S_n \geq \tau$, we say sleep x_i is irregular. The threshold τ is set to 2 based on the problem domain in our empirical study.

Our approach is to use context variations as an indirect indicator of the sleep regularity. From Section IV-B we can observe that the light, sound, screen-on, movement count context are highly correlated to the sleep classification across all participants. So we use these context to build context profiles. Since most people go to bed between 10PM and 2AM, and wake up between 6AM and 10AM, we segment the time into four time slots: 10PM - 12AM, 12AM - 2AM, 6AM - 8AM, and 8AM - 10AM. We used the statistical values of context in those four time slots to profile a sleep. These time periods, of course, can be adjusted for different populations. For each sleep night, the mean values of illuminance value, decibel value, screen-on seconds and movement count are calculated during the first and the fourth time slots. The movement count context is removed for the second and third time slots since the value is very close to zero in the late night and early morning time slots. The resulting 14 statistic values are used to profile a sleep. Given enough samples, a 14-dimension multivariate normal distributed context model can be built.

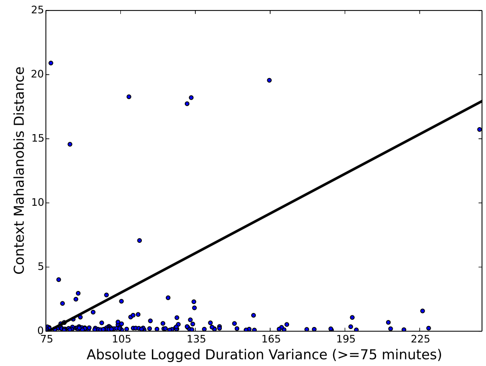

Figure 7. Correlation between logged big sleep duration variance and Mahalanobis distance.

B. Detection Algorithm

We use Mahalanobis distance [33] to detect context outliers. For multivariate normal data with mean μ and covariance matrix Σ, Mahalanobis distance $d_M(x)$ measures how many standard deviations away a data point x is from the mean:

$$d_M^2(x) = (x - \mu)^T \Sigma^{-1} (x - \mu)$$

Here, $x = (x_1, x_2, \cdots, x_n)^T$ is a 14-dimension context profile data point, $\mu = (\mu_1, \mu_2, \cdots, \mu_n)^T$ is the mean of the distribution and Σ is the covariance matrix. For a learning sample set S with size N, and each data point has dimension D, then the Mahalanobis distance satisfy the Beta distribution [34]:

$$\frac{N}{(N-1)^2} d_M^2 \sim Beta(\frac{D}{2}, \frac{(N-D-1)}{2})$$

Since the mean μ and covariance matrix Σ are very sensitive to outliers as well, we used Minimum Covariance Determinant estimator [35] to get a robust estimation of μ and Σ.

To support our intuition of using context variations to detect irregular sleep nights, we show the correlation of large sleep duration variance (e.g. irregular) and the Mahalanobis distance for those 8 participants who logged more than 50 nights sleep in Figure 7. The Mahalanobis distance between context vector and the mean is calculated on a daily basis using the robust estimation [35]. We can observe a clear correlation between the Mahalanobis distance of context profiles and the large duration variance ($\geq 75 minutes$).

The unsupervised algorithm to detect irregular sleep nights is shown in Algorithm 1. We only used two night time slots to detect irregular bedtime, two morning time slots to detect irregular waketime, and all four time slots to detect irregular sleep duration.

Algorithm 1 Detect irregular sleep using context profiling

D_c : Data set of n days context statistic vectors
D_c^i : Context statistic vector of day i (d-dimension)
$\overline{D_c^i}$: Data set of context statistic vectors without day i
MinCovDet : Minimum covariance determinant estimation for μ and Σ
τ = 99% Percent point of $Beta(\frac{d}{2}, \frac{n-d-1}{2})$
for Day i = 1 **to** n **do**
  μ, Σ = MinCovDet($\overline{D_c^i}$)
  $x_i = D_c^i$
  $d_i^2 = (x_i - \mu)^T \Sigma^{-1} (x_i - \mu)$
  if $\frac{n}{(n-1)^2} d_i^2 \geq \tau$ **then**
    Assert it is an irregular sleep
  else
    Assert it is a regular sleep
  end if
end for

For a data set with n context vector samples, and each context has d-dimension, we leverage leave-one-day-out to detect irregular sleeps. We need n ($n \geq 30$) [36] days of context statistical data to profile sleeps. Note that context profiling does not require labeled data. The mean μ and covariance matrix Σ are calculated by using the minimum covariance determinant estimation, and the Mahalanobis distance is calculated based on μ and Σ. Any distance d_i^2 that is after the 99% point of $Beta(\frac{d}{2}, \frac{n-d-1}{2})$ distribution is regarded as outlier, and we assert it as an irregular sleep.

C. Evaluation Results

In order to evaluate our approach, we first need to determine the “ground truth” of irregular sleep nights. Take the bedtime as an example, if the time a person went to sleep is 2 standard deviations away from the mean bedtime (robust estimated mean through S_n estimator [32]), we define it as an irregular bedtime. According to the 3-sigma rule [37], 95% sleeps will be defined as regular sleeps and the rest of 5% are irregular sleeps. We used n days of user logged time to define regular sleep, and those n days context statistics to profile regular sleep. The robust S_n estimation [32] is used to eliminate the effect of outliers and get a robust estimation of the mean and standard deviation of regular sleep. Most statisticians believe we need 30 or more samples [36] before the sampling distribution of the mean becomes a normal distribution. So we used at least 50 days of context data to profile sleeps. For the purpose of this evaluation, we focused on 8 participants who logged more than 50 nights of sleep.

Here we also apply leave-one-day-out cross-validation to evaluate the context profiling approach. For a n days dataset, we select $(n-1)$ days of context statistics to profile sleeps and get a robust estimation of regular sleep, then by applying Algorithm 1 we test whether the left-out day is an outlier or not, which is used to determine its regularity. Evaluations were conducted for bedtime, waketime and sleep duration, respectively. Table II presents the confusion matrix of irregular sleep detection using the unsupervised context profiling approach.

1) Irregular Bedtime Detection: We can see unsupervised context profiling approach can effectively detect irregular bedtime for most participants. For instance, it detected 14 irregular bedtime out of 15 for participant 4, and it detected 9 irregular sleeps out of 13 for participant 14. A horizontal comparison shows that participant 9, 14 and 15 have a better detection results. We can also see the recall score is higher than the precision score for most participants, which indicates that context profiling approach can detect more true irregular sleeps than false positives.

2) Irregular Waketime Detection: The irregular waketime detection only uses context profiles in time slots 6AM - 8AM and 8AM - 10AM. Table II shows that it is harder to detect irregular waketime than detecting irregular bedtime, mostly due to the phone usage behavior after users have waken up. If the context variation is not significant, for example people left their phone in the bedroom, context profiling may produce a small Mahalanobis distance and detect it as regular. As we can see, participant 9, 14 and 15 also have a better detection results than other participants. Even participant 9 and 18 have very irregular wakeup patterns, which can be seen in Figure 4, both the recall and precision scores are considerably high. This shows the advantage of detecting irregular waketime to cope with sleep variation using the proposed context profiling approach.

3) Irregular Sleep Duration Detection: The irregular sleep duration detection for all 8 participants seems to have similar prediction accuracy and recall scores except it missed 5 irregular durations for participant 11. From Table II we can see the average bedtime for participant 11 is about 1AM, which is close to the end of the second time slot, and the average waketime is around 7:30AM which is close to the start of the fourth time slot. This could be a problem for the context profiling approach since it can only capture the context variations between those 4 time slots. If the context change is close to the edges of those time slots, the variation is not significantly enough to detect irregular sleep. This could be solved by shifting time slots. We can also observe that there are less false negative than false positive, which indicates the context profiling approach can effectively detect irregular sleep durations.

In general, we can see there are more false positives (false alarms) than false negatives. This is because context variation has two directions, the value of context sensing could be either much more than average or much less than average. In both cases, they will be detected to be irregular

Table II
CONFUSION MATRIX OF IRREGULAR BEDTIME, WAKETIME AND DURATION DETECTION FOR 8 PARTICIPANTS

		4		9		11		12		14		15		17		18	
		Detected Regular	Detected Irregular	Detected Regular	Detected Irregular	Detected Regular	Detected Irregular	Detected Regular	Detected Irregular	Detected Regular	Detected Irregular	Detected Regular	Detected Irregular	Detected Regular	Detected Irregular	Detected Regular	Detected Irregular
Bedtime	True Regualr	60	14	38	11	65	10	79	12	62	6	64	6	72	4	70	5
	True Irregular	1	14	5	2	10	3	6	3	4	9	8	7	1	1	6	1
Waketime	True Regular	78	11	45	5	72	16	76	15	53	16	65	11	61	14	61	17
	True Irregular	0	0	3	3	0	0	7	2	5	3	5	4	3	0	2	2
Duration	True Regular	67	17	43	10	67	16	76	11	59	12	66	10	63	13	63	16
	True Irregular	3	2	1	2	5	0	8	5	3	7	6	3	2	1	1	2

when using Mahalanobis distance to detect outliers. Hence, how people use their smartphone has a direct effect on the detection results. Another reason the detection accuracy may deteriorate is that, the context profiling approach assumes the sensing data is normal distributed, though in reality the distribution is more likely a skewed normal distribution. For instance, the mean illuminance value and decibel value are always tend to be small during night, which truncated most of the left half of the distribution. We have tried to reshape the data to make it more normal distributed, by getting the fourth power root of the original value, it is still difficult to achieve normal distributions for all participants.

In summary, the proposed unsupervised approach worked well for most participants to detect irregular sleep nights, which reflect sleep pattern changes that are often an indicator of potential health problems.

VI. DISCUSSION AND FUTURE WORK

The accuracy of passive sleep monitoring depends on users' smartphone usage behavior. If the user does not have a habit of checking smartphone frequently or bringing the phone to the bedroom, the proposed system will not work well. However, a recent study found that some people check their phone 150 times a day,[1] and Dey et al. found smartphones are within the same room as the user 90% of time [12]. While these statistics are encouraging, it may not generalize to special population, such as elder people who are slow to adopt new technologies.

In this paper we propose to detect irregular sleep nights through context profiling. We used a statistical definition of irregularity as outliers (τ standard deviations away from the mean). The users' perceived irregularity, however, may be different than this statistical definition. For example, sleep quality may be an easier metric for people to understand, though we believe bad sleep quality will also correlate strongly to context variation. There are existing work that can estimate sleep quality through smartphone sensing, though with higher energy consumption due to faster sampling rate [21]. We plan to explore this tradeoff and study the correlation between sleep quality and context profiles.

[1] http://www.kpcb.com/insights/2013-internet-trends

VII. CONCLUSIONS

We present a smartphone-based approach for unconstrained sleep assessment, with the objective to predict basic sleep parameters (bedtime, waketime, and sleep duration) and to detect irregular sleep nights for long-term monitoring. Through a user study of 18 participants over 14 weeks, the proposed supervised prediction outperformed state-of-art methods with average bedtime and sleep duration errors about 24 and 41 minutes, respectively. In addition, we propose a new unsupervised context profiling approach to detect irregular sleep nights that are experimentally validated to achieve good precision and recall results. To the best of our knowledge, this is the first work for unconstrained smartphone sensing to achieve accurate prediction accuracy with minimal energy consumption, and to detect irregular sleep nights that requires no manual labeling.

ACKNOWLEDGMENT

This material is based upon work supported in part by the National Science Foundation under Grant No. 1016823. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the funding agencies.

REFERENCES

[1] M. H. Bonnet and D. L. Arand, "We are chronically sleep deprived," *SLEEP-NEW YORK-*, vol. 18, pp. 908–911, 1995.

[2] A. NT, W. DP, M. JE, and et al, "A prospective study of sleep duration and coronary heart disease in women," *Archives of Internal Medicine*, vol. 163, no. 2, pp. 205–209, 2003. [Online]. Available: +http://dx.doi.org/10.1001/archinte.163.2.205

[3] G. DJ, P. NM, N. AB, and et al, "Association of sleep time with diabetes mellitus and impaired glucose tolerance," *Archives of Internal Medicine*, vol. 165, no. 8, pp. 863–867, 2005. [Online]. Available: +http://dx.doi.org/10.1001/archinte.165.8.863

[4] P. Maquet, "The role of sleep in learning and memory," *Science*, vol. 294, no. 5544, pp. 1048–1052, 2001. [Online]. Available: http://www.sciencemag.org/content/294/5544/1048.abstract

[5] D. E. Ford and D. B. Kamerow, "Epidemiologic study of sleep disturbances and psychiatric disorders: an opportunity for prevention?" *Jama*, vol. 262, no. 11, pp. 1479–1484, 1989.

[6] A. R. Wolfson and M. A. Carskadon, "Understanding adolescent's sleep patterns and school performance: a critical appraisal," *Sleep medicine reviews*, vol. 7, no. 6, pp. 491–506, 2003.

[7] R. Wang, F. Chen, Z. Chen, T. Li, G. Harari, S. Tignor, X. Zhou, D. Ben-Zeev, and A. T. Campbell, "Studentlife: assessing mental health, academic performance and behavioral trends of college students using smartphones," in *Ubicomp, 2014*. ACM, 2014, pp. 3–14.

[8] R. B. Berry, R. Budhiraja, D. J. Gottlieb, D. Gozal, C. Iber, V. K. Kapur, C. L. Marcus, R. Mehra, S. Parthasarathy, S. F. Quan *et al.*, "Rules for scoring respiratory events in sleep: update of the 2007 aasm manual for the scoring of sleep and associated events," *J Clin Sleep Med*, vol. 8, no. 5, pp. 597–619, 2012.

[9] Fitbit. [Online]. Available: http://www.fitbit.com

[10] Zeo. [Online]. Available: http://www.myzeo.com

[11] J. Behar, A. Roebuck, J. S. Domingos, E. Gederi, and G. D. Clifford, "A review of current sleep screening applications for smartphones," *Physiological measurement*, vol. 34, no. 7, p. R29, 2013.

[12] A. K. Dey, K. Wac, D. Ferreira, K. Tassini, J.-H. Hong, and J. Ramos, "Getting closer: an empirical investigation of the proximity of user to their smart phones," in *Ubicomp, 2011*. ACM, 2011, pp. 163–172.

[13] Sleep as android. [Online]. Available: https://sites.google.com/site/sleepasandroid

[14] T. Hao, G. Xing, and G. Zhou, "isleep: unobtrusive sleep quality monitoring using smartphones," in *Proceedings of the ACM Conference on Embedded Networked Sensor Systems*. ACM, 2013, p. 4.

[15] W. Gu, Z. Yang, L. Shangguan, W. Sun, K. Jin, and Y. Liu, "Intelligent sleep stage mining service with smartphones," in *Ubicomp, 2014*, ser. UbiComp '14. New York, NY, USA: ACM, 2014.

[16] E. K. Choe, J. A. Kientz, S. Halko, A. Fonville, D. Sakaguchi, and N. F. Watson, "Opportunities for computing to support healthy sleep behavior," in *CHI'10 Extended Abstracts on Human Factors in Computing Systems*. ACM, 2010, pp. 3661–3666.

[17] Sense. [Online]. Available: https://hello.is

[18] M. Kay, E. K. Choe, J. Shepherd, B. Greenstein, N. Watson, S. Consolvo, and J. A. Kientz, "Lullaby: a capture & access system for understanding the sleep environment," in *Ubicomp, 2012*. ACM, 2012, pp. 226–234.

[19] V. Metsis, D. Kosmopoulos, V. Athitsos, and F. Makedon, "Non-invasive analysis of sleep patterns via multimodal sensor input," *Personal and ubiquitous computing*, vol. 18, no. 1, pp. 19–26, 2014.

[20] Z. Chen, M. Lin, F. Chen, N. D. Lane, G. Cardone, R. Wang, T. Li, Y. Chen, T. Choudhury, and A. T. Campbell, "Unobtrusive sleep monitoring using smartphones," in *Pervasive Computing Technologies for Healthcare (PervasiveHealth)*. IEEE, 2013, pp. 145–152.

[21] J.-K. Min, A. Doryab, J. Wiese, S. Amini, J. Zimmerman, and J. I. Hong, "Toss 'n' turn: Smartphone as sleep and sleep quality detector," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '14. New York, NY, USA: ACM, 2014.

[22] S. Abdullah, M. Matthews, E. L. Murnane, G. Gay, and T. Choudhury, "Towards circadian computing: "early to bed and early to rise" makes some of us unhealthy and sleep deprived," in *Ubicomp, 2014*, ser. UbiComp '14. New York, NY, USA: ACM, 2014.

[23] Basis. [Online]. Available: http://www.mybasis.com

[24] Jawbone up. [Online]. Available: https://jawbone.com/up

[25] A. Sadeh, "The role and validity of actigraphy in sleep medicine: an update," *Sleep medicine reviews*, vol. 15, no. 4, pp. 259–267, 2011.

[26] Beddit. [Online]. Available: http://www.beddit.com

[27] E. Hoque, R. F. Dickerson, and J. A. Stankovic, "Monitoring body positions and movements during sleep using wisps," in *Wireless Health 2010*. ACM, 2010, pp. 44–53.

[28] Sleep cycle. [Online]. Available: http://www.sleepcycle.com

[29] Sleepbot. [Online]. Available: http://mysleepbot.com

[30] C. Zhang, K. Huang, and G. Chen, "Mvptrack: Energy-efficient places and motion states tracking," in *MobiQuitous, 2013*. IEEE, December 2013.

[31] F. R. Hampel, "The influence curve and its role in robust estimation," *Journal of the American Statistical Association*, vol. 69, no. 346, pp. 383–393, 1974.

[32] P. J. Rousseeuw and C. Croux, "Alternatives to the median absolute deviation," *Journal of the American Statistical association*, vol. 88, no. 424, pp. 1273–1283, 1993.

[33] Mahalanobis distance. [Online]. Available: http://en.wikipedia.org/wiki/Mahalanobis_distance

[34] D. Ververidis and C. Kotropoulos, "Gaussian mixture modeling by exploiting the mahalanobis distance," *Signal Processing, IEEE Transactions on*, vol. 56, no. 7, pp. 2797–2811, 2008.

[35] P. J. Rousseeuw and K. V. Driessen, "A fast algorithm for the minimum covariance determinant estimator," *Technometrics*, vol. 41, no. 3, pp. 212–223, 1999.

[36] R. V. Hogg, E. A. Tanis, and M. J. M. Rao, *Probability and statistical inference*. Macmillan New York, 1977, vol. 993.

[37] Normal distribution. [Online]. Available: http://en.wikipedia.org/wiki/Normal_distribution