

Monitoring and Improving Personalized Sleep Quality from Long-Term Lifelogs

Wenbin Gan, Minh-Son Dao, Koji Zetsu

Big Data Integration Research Center

National Institute of Information and Communications Technology (NICT)

Tokyo, Japan

{wenbingan, dao, zetsu}@nict.go.jp

Abstract—Sleep plays a vital role in our physical, cognitive, and psychological well-being. Despite its importance, long-term monitoring of personalized sleep quality (SQ) in real-world contexts is still challenging. Many sleep researches are still developing clinically and far from accessible to the general public. Fortunately, wearables and IoT devices provide the potential to explore the sleep insights from multimodal data, and have been used in some SQ researches. However, most of these studies analyze the sleep related data and present the results in a delayed manner (i.e., today's SQ obtained from last night's data), it is still difficult for individuals to know how their sleep will be before they go to bed and how they can proactively improve it. To this end, this paper proposes a computational framework to monitor the individual SQ based on both the objective and subjective data from multiple sources, and moves a step further towards providing the personalized feedback to improve the SQ in a data-driven manner. The feedback is implemented by referring the insights from the PMDD dataset based on the discovered patterns between life events and different levels of SQ. The deep learning based personal SQ model (PerSQ), using the long-term heterogeneous data and considering the carry-over effect, achieves higher prediction performance compared with baseline models. A case study also shows reasonable results for an individual to monitor and improve the SQ in the future.

Index Terms—Sleep Quality Monitoring, Life Event Patterns, Personalized Feedback, Lifelogs

I. INTRODUCTION

Sleep is of great importance for us human beings to maintain health. This complex physiological process is a regular part of our daily routine, allows our body and mind to recharge, become refreshed and alert [1]. A good night's sleep is closely linked to a greater level of physical, cognitive, and psychological well-being [2], [3]. While getting enough sleep is definitely important, good SQ is also essential for a good night's sleep. In this paper we focus on monitoring the long-term sleep qualities of individuals using daily logs and providing the personalized lifestyle feedback to improve the sleep qualities (Research questions are shown in Figure 1).

Sleep covers nearly one third of the individual lifespan; given its significant role for human lives and its high impact on our health, a large amount of work has been done to explore the quantitative sleep measures [1], [3], the modeling of factors influencing sleep, such as physical activities [4], serious diseases [2] and lifestyle [5]. However, many sleep researches are still developing clinically and far from accessible to the general public [6]. For most people, it may be possible to

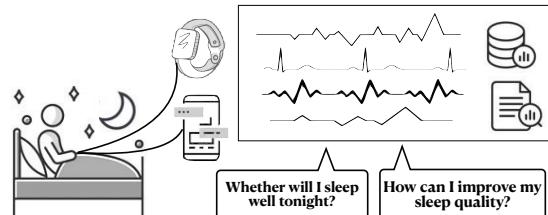


Fig. 1. Research questions in this paper.

simply self-assess the last night's sleep by observing the mental status and physical conditions after getting up, but it is still difficult to know how their sleep will be before they go to bed and how they can improve it [6], [7].

Some clinical tools for sleep monitoring, such as the Polysomnography (PSG), can provide a quantitative insight of sleep and diagnose sleep disorders [8]. PSG-based sleep studies generally perform overnight while being continually observed by a qualified technician using various medical sensors. Although such studies archive quite precise results for SQ measurement, they are expensive and unusual for an average person and thus only limited to clinical settings. As an inexpensive alternation, actigraphy has been used as a popular non-invasive method to monitor human sleep; this method uses wristwatch-like wearables to continuously record human body movements overnight and assess the SQ. With the increasingly popularity of wearables and IoT devices, the multimodal data streams and daily events can be captured using these devices [9], providing a promising way to jointly analyzing heterogeneous data sources to yield more accurate monitoring of individual SQ from a computational perspective [7], [10].

Existing work on SQ measurement is mainly conducted at long-term and per-night level in terms of the duration of the measurements lasts [11]. The former usually conducts the measurements over a relatively long time interval based on some self-rated questionnaires, such as the Pittsburgh SQ Index (PSQI) [12]. The latter performs the per-night SQ prediction on a daily basis [11]. Compared with the long-term methods, the per-night methods provide the severity of a particular sleep problem more timely. Despite this benefit, most of the per-night SQ measurement methods do not consider the inter-

dependencies between the coming tonight's sleep and the sleep conditions in the previous days (the carry-over effect), which may potentially cause the performance degradation. Besides, most of these two categories of studies analyze the sleep related data and present the results in a delayed manner (i.e., today's SQ obtained from last night's data); it will be more useful to inform users their potential SQ before they go to bed and give feedback on action taking to improve it. Moreover, the majority of existing researches build the population-based models to score user sleep without giving them the personalized feedback for improving their future SQ. Considering that individuals have varying requirements and reactions to sleep, it is necessary to monitor the individual SQ and provide the personalized lifestyle feedback to improve the SQ.

To this end, this paper proposes a holistic framework to monitor the individual SQ based on both the objective and subjective lifelog data from multiple sources, and moves a step further to provide the personalized feedback to improve the SQ in a data-driven manner. It infers the individual SQ very timely before the bedtime using the long-term data to model the carry-over effect. Moreover, it builds the connections between SQ and the everyday life events, and actively feedbacks the individuals to bring the potential SQ improvement based on these connections. The contributions of this work are threefold:

- We propose an computational framework for monitoring the individual SQ and going a step further to provide the personalized feedback to improve the SQ.
- We present the PerSQ model with high performance to monitor the individual SQ timely using the long-term data to model the carry-over effect.
- We discover the insights of patterns between life events and different levels of SQ, and use them to feedback the individuals to bring the potential SQ improvement.

II. RELATED WORK

A. *SQ Monitoring*

Sleep is a complex physiological process [1]. Although there is still a lack of an established definition for the term SQ, researchers have explored to measure SQ from both clinical and technical fields using various quantitative metrics, such as total sleep time, sleep onset latency, degree of fragmentation, total wake time, and sleep efficiency [3]. It is well acknowledged that accurate SQ monitoring is essential for the early detection of sleep deprivation and insomnia.

In general, SQ studies are mainly conducted from both objective and subjective perspectives in terms of the data source for the measurement. Sleep diaries and/or self-rated questionnaires are widely used to analyze the individual's subjective SQ; the PSQI [12] is a typical tool used in this direction. Sleep related measures are collected using the individual reported questionnaires based on the subject's perception over a long-term interval [8]. A single-item SQ scale (SQS) was newly developed in [13] and provided a simple and practical alternative for the subjective evaluation. Despite the potential

bias, this subjective method is easy to be implemented and has shown useful in various studies [1].

Meanwhile, the prevalence of clinical and commercial IoT devices provide the possibility to measure the SQ objectively. Advanced sensing techniques enable to record various signals of the human body; popular deep learning techniques offer new ideas for data analysis [14], [15]. These provide more promising solutions to explore SQ in a date-driven manner. PSG is recognized as the gold standard for SQ studies by capturing multiple channels of psychological states, but is limited to the sleep clinics [8]. Actigraphy method using unobtrusive wristbands to collect the body movement data, is an inexpensive alternative, but it measures the SQ only based on the body movement data which limits its usage in the real applications [6]. Hence many researches investigate to use the holistic data (psychological, behavioral and subjective data, etc.) on this task [1], [16], [17]. Park et al. [10] proposed a LSTM-Attention method for predicting the SQ using heterogeneous data, including sleep records, daily activities, and demographics. Bai et al. [11] developed a prototype system called SleepMiner to collect on-phone data, such as daily activity, living environment and social activity, and used a factor graph model to predict SQ. Sathyaranayana et al. [18] predicted good or poor sleep efficiency as an indicator of SQ using the physical activity wearable data during awake time. Liu et al. [19] predicted SQ based on users' physical exercise data. Liu et al. [20] used the Kinect sensor to capture the gait pattern to predict SQ, and showed the feasibility of the gait pattern for revealing SQ. Krishna et al. [21] combined features related to user surroundings and movements during sleep, and used regression models to monitor SQ and determine the apt sleep duration.

Some researchers developed IoT-based sleep monitoring system or mobile apps to monitor the SQ. SleepApp [22], SleepCoacher [23], SleepTight [24] and iSleep [8] are typical mobile-based self-experimentation systems for SQ. Reference [25]–[31] proposed various IoT based hardware and software solutions to monitor SQ based on multimodal sensors.

Different from the above researches, this paper monitors the individual SQ based on both the objective and subjective data from multiple sources, and move a step further to provide personalized feedback to improve the SQ in a data-driven manner.

B. *Life Event and SQ*

Various studies have explored the impact of various life events on sleep, for example the physical activities and the sleep context [4], [6], [7], [10]. Pandey et al. [7] and Upadhyay et al. [6] conducted N-of-1 experiments to identify the causal relationships between the daytime activities and SQ. Park et al. [10] modeled SQ from daily activities and demographics. Singh et al. [32] explored the relationship between physical activity and SQ and extracted the behavior modes that resulting in the good sleep using clustering technique. Liu et al. [19] predict the SQ only based on the users' daily exercise information. Ravichandran et al. [22] explored the evidence

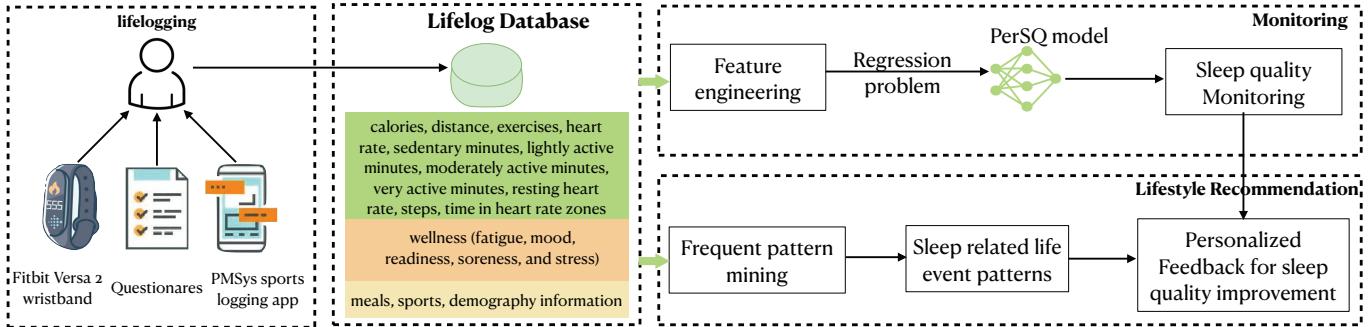


Fig. 2. Our framework for monitoring and improving SQ from life-logs.

of individual behaviors affect the SQ, and also showed how sleep affect daytime cognitive function (daytime alertness and working memory). Farajtaba et al. [33] conducted a large-scale many-to-many factor study of sleep effects and exercise, search query logs, location and aggregated social media data. All these researches show the close relations between the SQ and various life events, providing the references for this paper to choose variables in SQ monitoring.

Compared with these existing researches, this paper builds the connections between SQ and the daily activities, the subjective wellness and personal demographic data, with a focus on the timely SQ monitoring before bed time and providing feedback to improve it. Hence it intentionally avoids using the sleep related data during the current day, an unique feature to enable the real-time and make it different from existing studies.

III. OUR PROPOSED FRAMEWORK

This paper proposes a novel framework for the SQ monitoring and improvement from long-term lifelogs (Figure 2). The lifelogs are collected from three aspects: the daily activity and psychological data from the Fitbit wristband, the individual wellness data logged through PMSys app, and the daily reporting data (meals, body weight, etc.) by everyday questionnaire. Using these lifelogs, our framework conducts the SQ monitoring and improvement in a consecutive manner. After conducting feature engineering on the original data, we model the SQ monitoring as a regression problem and propose a deep learning based personal SQ model, named PerSQ, to monitor the SQ from long-term personal heterogeneous data. For improving the personal SQ, we conduct frequent pattern mining on the whole lifelogs and generate three sets of sleep patterns (i.e., low, normal and high ones). Based on the predicted personal SQ, we match the user data with the obtained sleep patterns to find the parameters to be optimized and provide feedback to the individuals for further SQ improvement.

A. Personalized Long-term SQ Monitoring

Problem Formulation For an individual user u_i , this research aims to predict his/her SQ on the current day t_m based on the objective data (i.e., daily activity da , personal demographic

information pd) and subjective data (i.e., wellness ws) collected from multiple sources in a long term. Following existing researches [3], we use the sleep efficiency as the index of SQ calculated as follows:

$$SQ = \frac{\text{minutes_Asleep}}{\text{time_in_bed}} \in [1, 100] \quad (1)$$

Let $X_m = \{da_m, pd, ws_m\}$ be the list of all data obtained for a day t_m , and SQ_m be the value of SQ on the same day. The problem of SQ monitoring is formulated as:

$$f(X_{m-t}, X_{m-t+1}, \dots, X_{m-1}, X_m) \rightarrow SQ_m \quad (2)$$

where t is the number of previous days considered, and f is the function to predict the SQ value. We consider the carry-over influence of previous days on the current day.

PerSQ Model The architecture of the proposed PerSQ model is shown in Figure 3. It mainly consists of three layers: the pre-processing layer, the recurrent layer and the output layer.

Pre-processing layer: we re-sample the input data into the day interval and encode the category variables using one-hot encoding. Considering the variables are in different units, we normalize the data into $[0,1]$ using min-max normalization. To capture the carry-over effects, we shift the long-term data using a windowing mechanism, and slide the fix-length window to generate the time-series data $X = <X_{m-t}, X_{m-t+1}, \dots, X_{m-1}, X_m>$.

Recurrent layer: Long short-term memory (LSTM) network is used in this layer to extract useful information from the sequences. We uses three LSTM networks with dropout to process the time-series data.

Output layer: Given the processed data v from the recurrent layer, we firstly map v into the target space using fully-connected layer to obtain v' . To map the final predicted value v' to the original scale (i.e., $[1, 100]$), we concatenate it with the original data X and use inverse transformation to make it in the original scale. Hence the final output for the predicted SQ is in range of 1 and 100.

$$\begin{aligned} v' &= Wv + b \\ \hat{SQ} &= \text{inverse_transform}(X, v') \in [1, 100] \end{aligned} \quad (3)$$

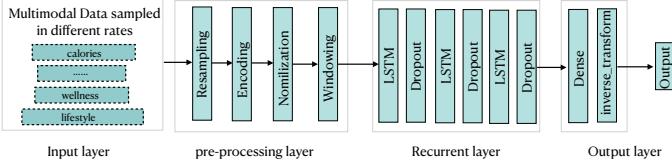


Fig. 3. The architecture of the proposed PerSQ model.

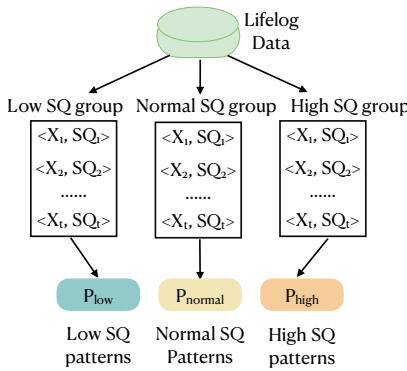


Fig. 4. Lifelog mining for obtaining three sets of SQ patterns.

B. Life Event Pattern Mining

To obtain the SQ insights from daily life events, we conduct pattern mining on the lifelogs to discover the different sets of patterns. Our hypothesis is that these patterns contain the relations of SQ and daily life events and can be used as references to provide further lifestyle feedback.

To this aim, we divide the lifelogs of all persons into three groups: the high, normal and low SQ groups based on the distribution of SQ values, as shown in Figure 4. The pattern mining process is then conducted on each group to find the frequent patterns in different degree. Specifically, we first select a group of variables, which are related with the daily activities and individual wellness, and infer the potential patterns with daily SQ. The group of variables for our dataset is shown in Figure 8. These life event factors are then thresholded to transform them from continuous to the category ones for easier working with the pattern mining method. The Apriori algorithm [34] is adopted to mine the frequent association rules from these three sets of items. Finally, three groups of frequent patterns are obtained, providing the insights for understanding the impact of daily life events on different degrees of SQ. The results of pattern mining are shown in Section V-C.

C. Personalized Lifestyle Feedback for SQ Improvement

After obtaining a user's daily data, we can predict his/her SQ using the proposed PerSQ model. Based on the predicted value (low, normal, high), personal lifestyle feedback is conducted using the life event patterns for improving their SQ.

Inspired by [7] to use condition-action rules for health recommendations, we use the obtained frequent life event patterns to generate the personal feedback. Since these event

Algorithm 1 Personalized Lifestyle Feedback

```

Input: A user's time-series data  $X = \langle X_{m-t}, X_{m-t+1}, \dots, X_{m-1}, X_m \rangle$ , the trained SQ monitoring model  $PerSQ$ , the mined three sets of frequent patterns  $P_{low}, P_{normal}, P_{high}$ .
Output: The predicted SQ value  $SQ_m$  and the feedback  $f$ .
1: initialize the set  $\mathbb{P} = \emptyset$  for parameters to be optimized;
2: Sleep quality Prediction:
3:  $SQ_m \leftarrow PerSQ(X)$ ;
4: Personalized Feedback:
5: threshold the personal data to category ones:  $X' \leftarrow X$ ;
6: if  $SQ_m$  is in low SQ group then
7:    $\mathbb{P} \leftarrow RuleMatch(X', P_{normal} \& P_{high})$ ;
8: else if  $N$  is normal and high SQ group then
9:    $\mathbb{P} \leftarrow RuleMatch(X', P_{high})$ ;
10: generate  $f$  based on  $\mathbb{P}$ ;
11: return  $SQ_m$  and  $f$ .

```

patterns contain the daily life context information linking to different degrees of SQ, they provide the optimization insights for improving SQ. We abstract and match the personal data with the event patterns using a simple rule matching strategy based on the max co-occurrence of the same contextual parameters. Meanwhile we rank the matched rules with the support count and the length of the variables in the dataset to enable the better matching results. For the low predicted SQ, we match the user data with the event patterns in the normal and high group; while for the normal and high results, we match the data with the patterns only in the high group, with the aim of potential improving the personal SQ based on the superior patterns. After obtaining the proper matched pattern, we discover the contextual parameters from the unmatched parts in the pattern and generate the feedback based on these optimizable parameters. For reference, the algorithm for this feedback process is shown in Alg. 1. Finally, the actionable multiple-item feedback (including the predicted SQ score and the suggestions) can be presented to the users for improving their personal SQ.

IV. DATASET AND PREPROCESSING

A. Dataset

The PMData¹ [35] is used for this study. It is a lifelogging dataset combining the traditional lifelogging with sports activity logging from 16 persons during 5 months (November 2019 to March 2020) using Fitbit Versa 2 smartwatch wristband, the PMSys sports logging app, and Google forms for the multimodal data collection. The objective physical and physiological data, such as the calories, steps, distance, exercise, heart rate, different kinds of activity minutes, sleep information and heart rate measurements, are recorded using Fitbit Versa 2 smartwatch. Personal wellness, including parameters like fatigue, mood, readiness, soreness and stress, is subjectively measured and logged by users through the PMSys sports logging app. Sports, meal and personal information are reported daily using the Google forms. Eventually, there are

¹PMData: <https://datasets.simula.no/pmdata/>

TABLE I
THE MULTIMODAL DATA SOURCES USED FOR THE STUDY.

Data type (Multimodalities)	Description	Mean \pm SD
Daily Activity (Source: Fitbit)		
calories	calories the person has burned per day	3053.2 ± 742.3
distance	the distance (in centimeters) moved per day	950694.3 ± 505781.7
steps	the number of steps per day	12221.6 ± 5959.0
sedentary minutes	the number of sedentary minutes per day	685.1 ± 172.2
lightly active minutes	the number of lightly active minutes per day	229.4 ± 95.7
moderately active minutes	the number of moderately active minutes per day	24.4 ± 23.0
very active minutes	the number of very active minutes per day	49.0 ± 44.4
time in heart rate zones	the number of minutes in different heart rate zones (4 types)	
wellness (Source: PMSys App)		
fatigue	1-5 scale of the subjective measurement of fatigue	2.7 ± 0.7
mood	1-5 scale of the subjective measurement of mood	3.2 ± 0.7
readiness	1-10 scale of how ready to exercise	4.9 ± 1.9
soreness	1-5 scale of the subjective measurement of soreness	2.8 ± 0.6
stress	1-5 scale of the subjective measurement of stress	2.9 ± 0.7
Personal Demographic (Source: Questionnaire)		
age	the age of a user	35.6 ± 11.4
gender	the gender of a user	male:12, female:3
A or B	whether the person has a Type A or Type B personality; getup early in the morning (Type A) and wakes up late (Type B)	A:9, B:6
sleep information (Source: Fitbit)		
minutes Asleep	the number of minutes being asleep	
time in bed	the number of minutes in bed	
sleep quality	the quality of sleep calculated by minutes Asleep / time in bed	87.8 ± 3.8

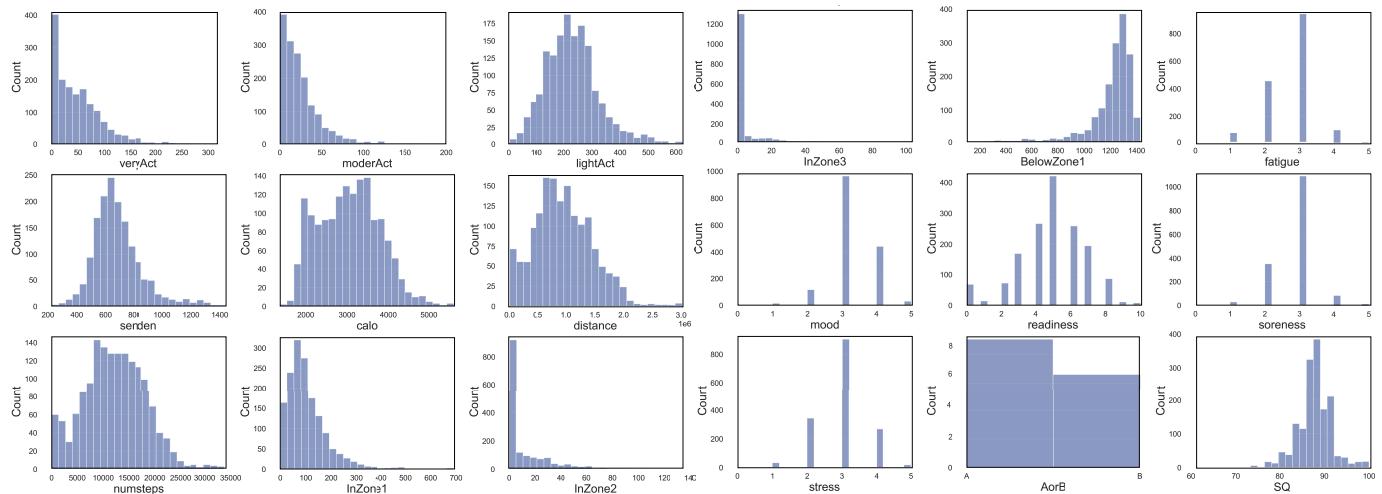


Fig. 5. The distributions of different life event variables.

2,440 activity sessions, 20,991,392 heart rate measurements, 1,836 days of sleep scores and 1,747 wellness reports in the final dataset.

As the 17 overall types of logged data have different rate of entries (per minutes, per day, per 5 seconds, etc.), we unify them into the same day interval. After checking the missing values, we find one user (ID:p12) did not have the record about light active minutes, we remove the data from this user and use the other 15 participant data in the final experiments. We select the factors that are highly related with the SQ task in previous studies, and finally get 8 factors belong to daily activity, 5 factors for the personal wellness, and 3 factors for

the personal demographic (as shown in Table I). The sleep information collected using the Fitbit are used as the ground-truth to build our model.

The distributions of these factors are shown in Figure 5. Among the 15 persons, 12 are male, the average age is 35.6 years old. Nine of them have the habit of getup early in the morning. The factors in the category of daily activity are transformed into the day interval with different distributions. For the factors in wellness, fatigue, mood, soreness and stress use the 5-point Likert scale to measure the subjective feeling of participants, while the readiness uses 10 scale. The ground-truth SQ values for all the people have an average of 87.8.

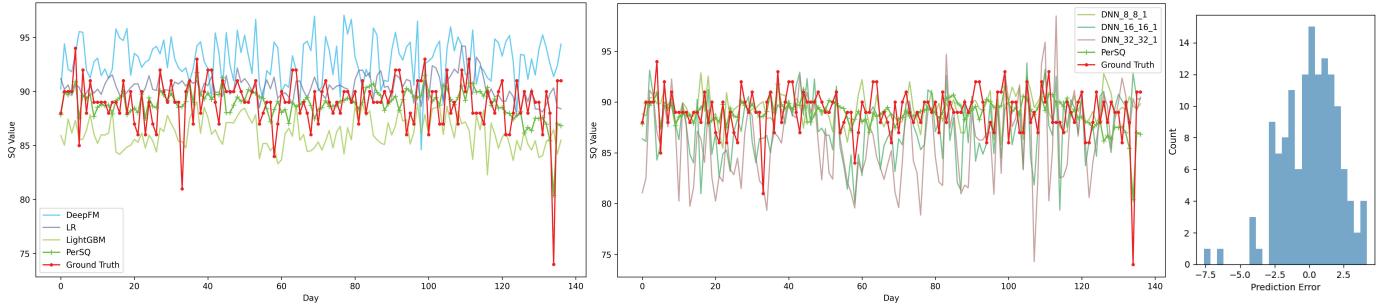


Fig. 6. (left): Comparisons of different models on SQ prediction for an individual user; (right): the distributions of prediction errors of the PerSQ model.

V. EXPERIMENTS

A. Experimental Setup

We trained the PerSQ model for SQ modeling using leave-one-out cross validation on the preprocessed dataset. Grid-search over heuristic choices of hyper-parameters was performed over choices of the numbers of sub-layers of LSTM networks and the number of days to track back. Eventually the cell units for the three consecutive sub-layers of LSTM in our model were set as $< 50, 30, 20 >$ and the model performances were compared on different considered days. For a new user who has no long-term event logs, we initialized the personalized model using the trained model at the first stage, and then updated it when more personalized data was collected to make it more tailored to the individual user.

To verify the effectiveness of the proposed PerSQ model on SQ monitoring, we compared with several competing models that have been shown to be effective in previous researches [1], [16]: the linear regression (LR) [22], the LightGBM [19], the multilayer perceptron (MLP) network [4] and the deep factorization machine (DeepFM) [36]. To fairly compare these models with the proposed model, we also take consideration of the carry-over effect by adding the SQ value in the previous day as a variable to input into these models. All the models were implemented in Python, the optimal hyperparameters were tuned to report the best results. Four criteria are employed to evaluate different facets of the regression performance: the Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Square Error (RMSE) and the R^2 . The first three tell how well a regression model can predict the value of SQ while the R^2 is a statistical measure that tells how well the model fit on the original data. Moreover, to validate the lifestyle feedback, we performed case study to show how to leverage the generated frequent patterns to improve the individual SQ.

B. Results on SQ Monitoring

The different models were evaluated by their performances in monitoring the SQ on the PMData. As shown in Table II, the proposed PerSQ model achieves the best performance on all the four criteria, with RMSE score of 2.017 and R^2 score of 0.246. The MLP regression model with 8-8-1 network architecture obtains the second best, with RMSE score of

TABLE II
PERFORMANCE COMPARISON OF DIFFERENT MODELS.

Model	MAE	MSE	RMSE	R^2
Linear regression	2.040	8.148	2.856	-0.511
LightGBM	3.217	13.871	3.724	-1.571
MLP regression(32-32-1)	4.035	26.804	5.177	-3.969
MLP regression(16-16-1)	3.032	15.551	3.944	-1.883
MLP regression(8-8-1)	1.869	7.033	2.652	-0.304
DeepFM	4.164	24.874	4.987	-3.611
PerSQ	1.591	4.068	2.017	0.246

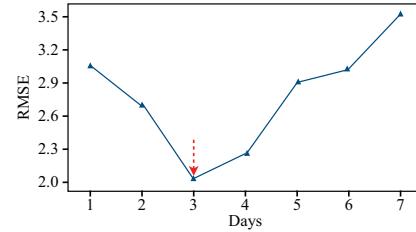


Fig. 7. The model performances considering different numbers of days.

2.652. The variations of MLP with 32-32-1 and 16-16-1 architectures do not get better results with deeper neurons, this is probably because the dataset is not big enough to train these models well. Even with the linear regression model, it performs much better than the two variations. LightGBM and DeepFM, two of the powerful models that rank in the top of various Kaggle competitions, seem to not get better performance than the linear regression model on this task. By contrast, the proposed PerSQ model takes consideration of the dependencies of previous life events on the SQ of the current day, making it perform better for SQ monitoring in a long term.

Moreover, we compared the performance of these different models for monitoring the individual SQ. As shown in Figure 6(left), the predicted SQ scores of an individual user by different models are visualized in different colors. The red line is the ground-truth, the prediction of our PerSQ model is shown in dark green. Compared with other models, the PerSQ is more closer to the ground-truth. The prediction error of the PerSQ is shown in Figure 6(right); most of the predicted errors are in range [-2.5, 2.5], making it acceptable for real applications.

TABLE III
SOME EXAMPLES OF THE LIFE EVENT PATTERNS OBTAINED FOR THREE GROUPS.

SQ level	Some examples of life event patterns	Frequency	Total number	Insights
Low (251)	('soreness_normal', 'InZone3_normal', 'numsteps_low', 'distance_low', 'veryAct_low', 'InZone2_low')	48	46	Physical Activities
	('numsteps_low', 'InZone3_normal', 'moderAct_low', 'distance_low', 'veryAct_low', 'InZone2_low')	47		
	('InZone1_low', 'InZone3_normal', 'numsteps_low', 'distance_low', 'veryAct_low', 'InZone2_low')	46		
	('InZone3_normal', 'numsteps_low', 'distance_low', 'veryAct_low', 'InZone2_low')	58		
	('soreness_normal', 'InZone3_normal', 'numsteps_low', 'distance_low', 'InZone2_low')	55		
Normal (998)	('readiness_normal', 'InZone3_normal', 'numsteps_low', 'distance_low', 'InZone2_low')	51	52	Wellness
	('soreness_normal', 'mood_normal', 'stress_normal', 'fatigue_normal')	272		
	('soreness_normal', 'mood_normal', 'InZone3_normal', 'stress_normal')	267		
	('mood_normal', 'InZone3_normal', 'stress_normal', 'AorB_A')	264		
	('soreness_normal', 'InZone3_normal', 'stress_normal', 'fatigue_normal')	256		
High (322)	('soreness_normal', 'mood_normal', 'stress_normal')	340	49	Physical Activities, lifestyle, wellness
	('soreness_normal', 'mood_normal', 'fatigue_normal')	339		
	('soreness_normal', 'numsteps_normal', 'InZone3_normal', 'distance_normal', 'veryAct_normal')	86		
	('soreness_normal', 'numsteps_normal', 'InZone3_normal', 'fatigue_normal', 'distance_normal')	82		
	('AorB_A', 'numsteps_normal', 'InZone3_normal', 'distance_normal', 'veryAct_normal')	76		
	('AorB_A', 'soreness_normal', 'numsteps_normal', 'InZone3_normal', 'distance_normal')	75		
	('InZone3_normal', 'veryAct_normal', 'numsteps_normal', 'distance_normal')	109		
	('InZone3_normal', 'stress_normal', 'numsteps_normal', 'distance_normal')	100		

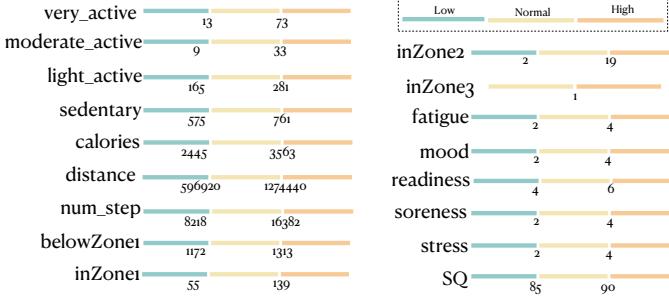


Fig. 8. The thresholds for different life event variables.

To test the influence of life events in previous days, we tested the model performance using data from different numbers of days. As shown in Figure 7, the RMSE scores of different considered days are compared. PerSQ model achieves best performance when considering three previous days; the RMSE score decreases when considering more days, and afterwards increases gradually when more than three days are considered.

C. Frequent Patterns of SQ and Life Events

To discover the insights of life event patterns with daily SQ, we threshold the considered factors to transform them from continuous to the category ones, as shown in Figure 8, based on their own distributions in Figure 5.

Based on the thresholded SQ value, we divided all the samples into three group: 251 samples in the low level, 998 in the normal level and 322 in the high group. After conducting the Apriori algorithm on each group, we obtained the frequent life event patterns in each group. Table III shows some examples of these patterns with highest frequencies. We filtered the patterns with at least 20% percent of the total number as the minimum support count. After pruning the frequent itemsets, we retained the longest and second-longest patterns and used them for future lifestyle feedback. Finally,

we got 46, 52 and 49 life event patterns for these three SQ groups.

The differences of these frequent patterns in three groups present some insights for exploring different levels of SQ. For the low group, most of the patterns are with low values of variables, and most of these variables are related with the indicators of physical activities, such as the heart beat zones, the number of steps, the distance and the very and moderate active minutes. These patterns generally show the low SQ is greatly related with low level of physical activities. For the normal SQ group, most of the variables in the patterns are related with the human wellness, such as the soreness, mood, stress and fatigue. While for the high SQ group, the variables are not only related with the physical activities but also the lifestyle (i.e., keep early hours) and wellness factors. This is consistent with existing researches that regular exercise, proper lifestyle and good wellness could help people get a high quality of sleep.

D. Case Study

To test the effectiveness of the life event patterns for providing SQ improvement feedback, we designed the feedback items in various contexts and conducted a case study to show how they works.

As shown in Table IV, we designed some example feedback in three categories based on the report in [37] to optimize the life parameters to improve the individual SQ. The feedback can be provided to users as voice or messages through the mobiles or wristwatches. Based on the designed feedback and the discovered life event patterns, we conducted a case study using the data from a user, as illustrated in Figure 9.

The piece of data is input to our model and predicted as low SQ (SQ score: 75). After thresholding the data and matching it with the life event patterns in normal and high group, we obtain the most proper pattern (pattern 1). By comparing the differences of the user data and the matched pattern, we find the parameter to optimize, i.e., the *numbersteps* and *distance*. The final feedback is then generated by providing the predicted

TABLE IV
FEEDBACK DESIGNING TO OPTIMIZE LIFE PARAMETERS FOR IMPROVING SQ.

Category	Parameter to optimize*	Feedback examples
Physical activities	numsteps	“Please try to walk more”
	distance	“Let’s go out and have a walk”
	calories	“Let’s do something to consume more calories”
	very_active or inZone3 moderate_active or inZone2 light_active or inZone1	“Recommended activities: Running, jogging, race walking/Hiking/Fast biking/ Stair climbing, etc.” “Recommended activities: Brisk walking/Ball sports/Water aerobics/Ballroom dancing, etc.” “Recommended activities: Slow walking/Slow bike riding/Stretching exercise/Gentle or chair yoga, etc.”
wellness	fatigue	“You seems tired, let’s take some rest”
	mood	“Let’s adjust our mood, how about listening to a happy music?”
	readiness	“How about taking a deep breath and making ourselves ready?”
	soreness	“You may need some relax”
lifestyle	stress	“Go easy on yourself, let’s make time for hobbies”
	AorB: B → A	“Do you want to try to keep early hours?”

*The optimization is to improve the level of parameters: low → normal → high.

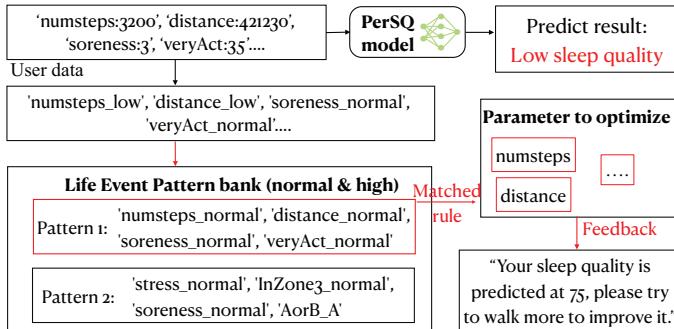


Fig. 9. Case study of personalized feedback for a user.

SQ value and also the designed feedback for these two factors referencing Table IV. An individual can know not only how his/her SQ will be on tonight but also how to further improve it based on the feedback.

VI. CONCLUSION

This paper presented a framework for monitoring and improving SQ from daily life-logs. It monitored the SQ using both the objective and subjective data from multiple sources (the Fitbit wristband, the sports logging app, and questionnaires), and move a step further to explore the insights from three groups (low, normal, and high SQ) of data to discover the potential patterns between life events and different levels of SQ. Based on these patterns, it further provided personalized feedback to the individuals to improve their SQ in a data-driven manner. To monitor SQ, it proposed the deep learning based PerSQ model using the long-term personal heterogeneous data and taking consideration of the carry-over effect of previous life events on the SQ of current day. The frequent pattern mining was conducted on the log data to discover the insights of patterns. Finally the user data was matched with these patterns to find the parameters to be optimized and provide feedback to the individuals for further SQ improvement.

We took the PMData dataset to build and test our framework. The results showed that the proposed PerSQ model

achieved the best performance for the SQ monitoring compared with the other baseline models and the best performance was obtained when considering life events in the previous three days. The insights on the patterns showed that low SQ is more likely related with the insufficient physical activities, normal SQ is more linked with personal wellness, while high SQ may requires many factors. Moreover, a case study was conducted to show the process of personalized feedback, providing information for the individuals to improve their SQ in the future.

For future work, we plan to conduct real application studies to validate the effectiveness of the feedback, and further improve the framework in a perception-action dynamic loop. Ultimately, we expect that the suggested framework will help people learn about their sleep and further gain better SQ.

REFERENCES

- [1] Fábio Mendonça, Sheikh Shanawaz Mostafa, Fernando Morgado-Dias, Antonio G Ravelo-Garcia, and Thomas Penzel. A review of approaches for sleep quality analysis. *Ieee Access*, 7:24527–24546, 2019.
- [2] Goran Medic, Micheline Wille, and Michiel EH Hemels. Short-and long-term health consequences of sleep disruption. *Nature and science of sleep*, 9:151, 2017.
- [3] Antonino Crivello, Paolo Barsocchi, Michele Girolami, and Filippo Palumbo. The meaning of sleep quality: a survey of available technologies. *IEEE access*, 7:167374–167390, 2019.
- [4] Aarti Sathyaranayana, Shafiq Joty, Luis Fernandez-Luque, Ferda Ofli, Jaideep Srivastava, Ahmed Elmagarmid, Shahrad Taheri, and Teresa Arora. Impact of physical activity on sleep: A deep learning based exploration. *arXiv preprint arXiv:1607.07034*, 2016.
- [5] Tamar Shochat. Impact of lifestyle and technology developments on sleep. *Nature and science of sleep*, 4:19, 2012.
- [6] Dhruv Deepak Upadhyay, Vaibhav Pandey, Nitish Nag, and Ramesh Jain. Personalized user modelling for sleep insight. In *Proceedings of the 1st International Workshop on Human-centric Multimedia Analysis*, pages 13–20, 2020.
- [7] Vaibhav Pandey, Dhruv Deepak Upadhyay, Nitish Nag, and Ramesh C Jain. Personalized user modelling for context-aware lifestyle recommendations to improve sleep. In *HealthRecSys@ RecSys*, pages 8–14, 2020.
- [8] Xiangmao Chang, Cheng Peng, Guoliang Xing, Tian Hao, and Gang Zhou. Isleep: A smartphone system for unobtrusive sleep quality monitoring. *ACM Transactions on Sensor Networks (TOSN)*, 16(3):1–32, 2020.
- [9] Wenbin Gan, Minh Son Dao, Koji Zettsu, and Yuan Sun. Iot-based multimodal analysis for smart education: Current status, challenges and opportunities. In *Proceedings of the 3rd ACM Workshop on Intelligent Cross-Data Analysis and Retrieval*, pages 32–40, 2022.

- [10] Sungkyu Park, Cheng-Te Li, Sungwon Han, Cheng Hsu, Sang Won Lee, and Meeyoung Cha. Learning sleep quality from daily logs. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 2421–2429, 2019.
- [11] Yin Bai, Bin Xu, Yuanchao Ma, Guodong Sun, and Yu Zhao. Will you have a good sleep tonight?: sleep quality prediction with mobile phone. In *BODYNETS*, pages 124–130, 2012.
- [12] Daniel J Buysse, Charles F Reynolds III, Timothy H Monk, Susan R Berman, and David J Kupfer. The pittsburgh sleep quality index: a new instrument for psychiatric practice and research. *Psychiatry research*, 28(2):193–213, 1989.
- [13] Ellen Snyder, Bing Cai, Carla DeMuro, Mary F Morrison, and William Ball. A new single-item sleep quality scale: results of psychometric evaluation in patients with chronic primary insomnia and depression. *Journal of Clinical Sleep Medicine*, 14(11):1849–1857, 2018.
- [14] Wenbin Gan, Yuan Sun, and Yi Sun. Knowledge interaction enhanced sequential modeling for interpretable learner knowledge diagnosis in intelligent tutoring systems. *Neurocomputing*, 488:36–53, 2022.
- [15] Wenbin Gan, Yuan Sun, and Yi Sun. Knowledge structure enhanced graph representation learning model for attentive knowledge tracing. *International Journal of Intelligent Systems*, 37(3):2012–2045, 2022.
- [16] Jiyoun Lim, Chi Yoon Jeong, Jeong Muk Lim, Seungeun Chung, Gague Kim, Kyoung Ju Noh, and Hyuntae Jeong. Assessing sleep quality using mobile emas: Opportunities, practical consideration, and challenges. *IEEE Access*, 10:2063–2076, 2022.
- [17] Thi Phuoc Van Nguyen, Do Van Nguyen, and Koji Zettsu. Models to predict sleeping quality from activities and environment: Current status, challenges and opportunities. In *Proceedings of the 2021 Workshop on Intelligent Cross-Data Analysis and Retrieval*, pages 52–56, 2021.
- [18] Aarti Sathyanarayana, Shafiq Joty, Luis Fernandez-Luque, Ferda Ofli, Jaideep Srivastava, Ahmed Elmagarmid, Teresa Arora, Shahrad Taheri, et al. Sleep quality prediction from wearable data using deep learning. *JMIR mHealth and uHealth*, 4(4):e6562, 2016.
- [19] Xiaoli Liu, Satu Tamminen, Topi Korhonen, Juha Röning, and Jukka Riekki. Prediction of sleep efficiency from big physical exercise data. In *Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers*, pages 1186–1189, 2019.
- [20] Xingyun Liu, Bingli Sun, Zhan Zhang, Yameng Wang, Haina Tang, and Tingshao Zhu. Gait can reveal sleep quality with machine learning models. *PloS one*, 14(9):e0223012, 2019.
- [21] Amrith Krishna, Madhumita Mallick, and Bivas Mitra. Sleepsensei: an automated sleep quality monitor and sleep duration estimator. In *Proceedings of the First Workshop on IoT-enabled Healthcare and Wellness Technologies and Systems*, pages 29–34, 2016.
- [22] Ruth Ravichandran, Shwetak N Patel, and Julie A Kientz. Sleepapp: Providing contextualized and actionable sleep feedback. In *Proceedings of the 14th EAI International Conference on Pervasive Computing Technologies for Healthcare*, pages 336–346, 2020.
- [23] Nediyana Daskalova, Danaë Metaxa-Kakavouli, Adrienne Tran, Nicole Nugent, Julie Boergers, John McGarry, and Jeff Huang. Sleepcoach: A personalized automated self-experimentation system for sleep recommendations. In *Proceedings of the 29th annual symposium on user interface software and technology*, pages 347–358, 2016.
- [24] Eun Kyoung Choe, Bongshin Lee, Matthew Kay, Wanda Pratt, and Julie A Kientz. Sleaptight: low-burden, self-monitoring technology for capturing and reflecting on sleep behaviors. In *Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing*, pages 121–132, 2015.
- [25] Khizra Saleem, Imran Sarwar Bajwa, Nadeem Sarwar, Waheed Anwar, and Amna Ashraf. Iot healthcare: design of smart and cost-effective sleep quality monitoring system. *Journal of Sensors*, 2020, 2020.
- [26] Stefano Milici, Antonio Lázaro, Ramon Villarino, David Girbau, and Marco Magnarosa. Wireless wearable magnetometer-based sensor for sleep quality monitoring. *IEEE Sensors Journal*, 18(5):2145–2152, 2018.
- [27] Ya-Ti Peng, Ching-Yung Lin, Ming-Ting Sun, and Carol A Landis. Multimodality sensor system for long-term sleep quality monitoring. *IEEE Transactions on Biomedical Circuits and Systems*, 1(3):217–227, 2007.
- [28] Jung-Yoon Kim, Chao-Hsien Chu, and Mi-Sun Kang. Iot-based unobtrusive sensing for sleep quality monitoring and assessment. *IEEE Sensors Journal*, 21(3):3799–3809, 2020.
- [29] Feng Zhang, Chenshu Wu, Beibei Wang, Min Wu, Daniel Bugos, Hangfang Zhang, and KJ Ray Liu. Smars: Sleep monitoring via ambient radio signals. *IEEE Transactions on Mobile Computing*, 20(1):217–231, 2019.
- [30] Feng Lin, Yan Zhuang, Chen Song, Aosen Wang, Yiran Li, Changzhan Gu, Changzhi Li, and Wenya Xu. Sleepsense: A noncontact and cost-effective sleep monitoring system. *IEEE transactions on biomedical circuits and systems*, 11(1):189–202, 2016.
- [31] Marco Laurino, Lucia Arcarisi, Nicola Carbonaro, Angelo Gemignani, Danilo Menicucci, and Alessandro Tognetti. A smart bed for non-obtrusive sleep analysis in real world context. *IEEE Access*, 8:45664–45673, 2020.
- [32] Meghna Singh, Saksham Goel, Abhiraj Mohan, Louis Kazaglis, and Jaideep Srivastava. Paris: Personalized activity recommendation for improving sleep quality. *arXiv preprint arXiv:2110.13745*, 2021.
- [33] Mehrdad Farajtabar, Emre Kiciman, Girish Nathan, and Ryen W White. Modeling behaviors and lifestyle with online and social data for predicting and analyzing sleep and exercise quality. *International Journal of Data Science and Analytics*, 8(4):367–383, 2019.
- [34] Rakesh Agrawal, Ramakrishnan Srikant, et al. Fast algorithms for mining association rules. In *Proc. 20th int. conf. very large data bases, VLDB*, volume 1215, pages 487–499. Citeseer, 1994.
- [35] Vajjira Thambawita, Steven Alexander Hicks, Hanna Borgli, Håkon Kvale Stensland, Debesh Jha, Martin Kristoffer Svensen, Svein-Arne Pettersen, Dag Johansen, Håvard Dagenborg Johansen, Susann Dahl Pettersen, et al. Pmdata: a sports logging dataset. In *Proceedings of the 11th ACM Multimedia Systems Conference*, pages 231–236, 2020.
- [36] Hufeng Guo, Ruiming Tang, Yunming Ye, Zhenguo Li, and Xiuqiang He. Deepfm: a factorization-machine based neural network for ctr prediction. *arXiv preprint arXiv:1703.04247*, 2017.
- [37] U.S. Department of Health and Human Services. Physical activity guidelines for americans, 2nd edition. 2018.