A Frequency Domain Algorithm to Identify Recurrent Sedentary Behaviors from Activity Time-Series Data

Qian He and Emmanuel O. Agu, *Member, IEEE*

***Abstract*—Sedentary behaviors such as sitting and watching TV increase the risk of many ailments including diabetes, cardiovascular disease and all-cause mortality. Fitness trackers and smartphone apps automatically track users' physical activities and sedentary behaviors to support self-monitoring. These trackers generate activity time-series data that require manual analysis to detect recurrent patterns. We propose a frequency domain algorithm for detecting recurrent sedentary patterns from activity time-series data at multiple timescales (hourly, daily, and weekly). In our experiment, subjects who exhibited recurrent sedentary behaviors yielded periodic functions with a Mean Square Error as low as 0.003817 for predicting recurrent sedentary behaviors. For subjects with no recurrent sedentary behaviors, our algorithm yielded a constant center amplitude value. Our algorithm can be used to predict future occurrences of recurrent sedentary patterns, facilitating many computer-tailored interventions.**

I. Introduction

Sedentary behaviors such as sitting, watching TV and various forms of screen-based entertainment are activities that do not increase energy expenditure substantially above the resting levels [1]. Studies show that sedentary behaviors are associated with a 112% increase in the risk of diabetes, 147% increase in cardiovascular disease, 90% increase in cardiovascular mortality, and 49% increase in all-cause mortality [2]. Determinants (causes) of sedentary behaviors include transport (e.g., sitting in cars on the way to work), leisure-time activities (e.g., surfing the web), household (e.g., child care), and occupation (e.g., desk jobs) [3], [4].

Self-monitoring, an effective health behavior change strategy for many health ailments [5] requires users to track unhealthy behaviors and try to change them. Consequently, activity trackers such as Fitbit [6], smartphone apps such as Google Fit [7] and smartwatch activity apps [8] now recognize many user activities. However, these trackers generate a time-series of user activities (and sedentary behaviors), which have to be manually inspected in order to detect recurrent patterns of misbehavior.

In this paper, we propose an algorithm for automatically detecting patterns of sedentary behaviors from activity time-series data. We focus on recurrent (as opposed to one-time) sedentary behaviors (or habits) since they can be predicted with the past sedentary behaviors and targeted with interventions. Examples of recurrent sedentary behaviors include lying on couches and playing video games after dinner every evening.

Recognized sedentary patterns can make people aware of their specific sedentary habits in order to self-correct. The user's doctor may also use these observed patterns to counsel them on how to make appropriate changes. Our approach identifies recurrent daily and weekly sedentary patterns, which are good predictors of future occurrences of these patterns. Computer-tailored interventions can be delivered at predicted re-occurrence times (e.g., through smartphones) to prevent deleterious health consequences [9]–[11].

Our approach is to propose a model that we validate using existing activity dataset gathered from 48 Dartmouth College students over the course of a 10-week term (the StudentLife dataset) [12]. Since we target recurrent (periodic) behaviors, we analyze time-series activity logs in the frequency domain to detect dominant sedentary cycles. As expected, some students followed their schedules, while others had no clear schedules. Our model worked well for students who followed a schedule, achieving a Mean Squared Error (MSE) of 0.003817 for predicting recurrent sedentary behaviors. While this paper focuses on sedentary behaviors, our models may be applied to any behavior that is detectable with activity trackers today.

II. Background and Related Work

In the health behavior change domain, many social cognitive models, such as the health belief model [13] and theory of planned behavior [14], are commonly used in the prediction of health-related behavior. While such models can assist in understanding unhealthy human behaviors [15], they are not computational in nature. Specifically, they require human interpretation of user data and cannot be used to predict future behaviors based on past behaviors.

Recurrent pattern modeling techniques have been previously used in many domains, such as hidden Markov model used in DNA analysis to infer recurrent DNA Copy Number Alterations (CNAs) from array Comparative Genomic Hybridization (aCGH) data [16], seasonal autoregressive moving averages used in financial data analysis to track business information [17], cross-correlation used in computer vision to construct optimal motion fields [18], and autocorrelation used in astrophysics to find recurrent solar activities [19].

*Research supported by NSF grant IIS **1065298.**

Qian He and Emmanuel Agu are both with the Dept of Computer Science, Worcester Polytechnic Institute, Worcester, MA 01609 USA
(e-mail: {qhe, emmanuel}@cs.wpi.edu).

III. METHODOLOGY

Our model segments a user's time into buckets (e.g. 10AM-11AM) and quantifies the percentage of time sedentary behaviors occur in each bucket. In this section, we start by describing the StudentLife dataset we tested our model with, followed by describing how we preprocess and explore the dataset. Finally, a frequency domain algorithm is developed to automatically search and evaluate recurrent sedentary patterns. The pseudocode of our algorithm is below:

Input:
Physical activity logs
Output:
A periodic function
Algorithm:
1. Convert logs to levels (percentage) of sedentariness time series;
2. Discrete Fourier transform the time series to frequency domain and get N cosine components $\boldsymbol{h}$[N];
3. Sort $\boldsymbol{h[N]}$ by amplitude in descending order
4. For $\boldsymbol{i}$ from $\boldsymbol{0}$ to $\boldsymbol{N-1}$
   Add the $\boldsymbol{h[0 \ldots i]}$ together as a function $\boldsymbol{f_i}$
   Reconstruct the time series with $\boldsymbol{f_i}$
   Calculate the $\boldsymbol{mse_i}$ of $\boldsymbol{f_i}$
5. Find the $\boldsymbol{f_k}$ with the smallest MSE, $\boldsymbol{mse_k}$

A. Data Source

In 2013 Spring, Wang *et al.* conducted a ubiquitous computing study called "StudentLife" at Dartmouth College [12] to track the mental health of students over the course of a term. Forty-eight students were given a smartphone with a data collection application installed. Over the 10-week term, the application collected automatically sensed data such as the students' activities, schedules, locations visited, smartphone usage and speech durations (audio inferences). Data that could not be automatically sensed such as mood and stress were also gathered through pop-up (EMA) questionnaires served to the students on their smartphones. The StudentLife dataset is available on the web of [12]. For our purposes, this dataset includes time series logs of automatically sensed physical activities of 48 actual students.

The physical activity logs consist of a timestamp and an activity type: *Stationary*, *Walking*, *Running*, and *Unknown*. This activity type was recognized from features extracted from the smartphone's accelerometer data with 94% accuracy every 2 seconds using a classifier (decision tree).

B. Preprocessing and Exploration

Since our goal is to identify statistically recurrent patterns of sedentary behavior, as the first step, we generate summary activity statistics for each time bucket. For each bucket, we calculate the percentage of all activities performed by the user that are sedentary ("*Stationary*"). For instance, if the time bucket is 1 hour and 1,260 of the 1,800 (1h × 60m/h × 60s/m / 2s) recognized activities in this hour are *Stationary*, then, the percentage of sedentary activities is 70%. We assume all "*Stationary*" activities within the StudentLife dataset are sedentary behaviors, which may not be true. For instance, standing still is stationary but is not a sedentary behavior. Although this assumption may be a potential limitation of our analysis, our algorithm will likely filter out standing behaviors since it is rare that a person stands still for prolonged periods at the same time (e.g., 11:30AM) every day or week. In future, we will gather our own data to address this limitation.

The bucket size in our model is the timescale we consider for prediction. For instance, if the bucket size is set to 1 hour, our model can be used to predict hourly sedentary behaviors. Similarly, for a bucket size of 6 hours, our model can predict morning (6AM-12PM), afternoon (12PM-6PM), evening (6PM-12AM), and night (12AM-6AM) intervals of sedentary behaviors. We anticipate that different subjects may exhibit sedentary behaviors at different timescales. For each subject, our algorithm searches 1 hour, 2 hours, 3 hours, 4 hours, 5 hours, and 6 hours bucket sizes to find the bucket size that best fits their recurrent sedentary behaviors.

For the data exploration purpose of this subsection, we only use 1 hour as the bucket size.

1) Daily Recurrent Pattern

Next, we use activity heat maps to visualize how much a given StudentLife subject was sedentary for each bucket of time per week. Fig. 1 is the heat map of the percentage of sedentary behaviors of Subject 10 for 1-hour buckets. Subject 10 usually became more sedentary around 10:00PM in the evening and less sedentary between 08:00AM and 09:00AM in the morning. We guess that Subject 10 went to sleep around 10:00PM and commuted to the academic buildings on campus to take classes around 08:30AM.

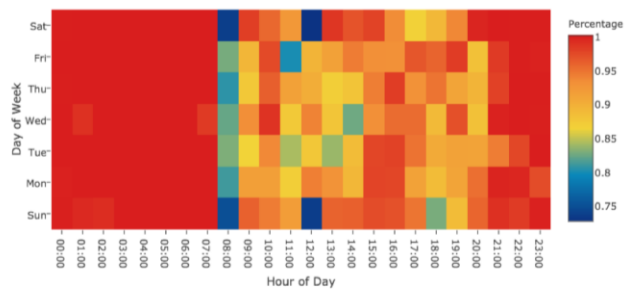

Figure 1. Subject 10's percentage of sedentary behaviors in 1-hour buckets grouped by the Day of Week and Hour of Day

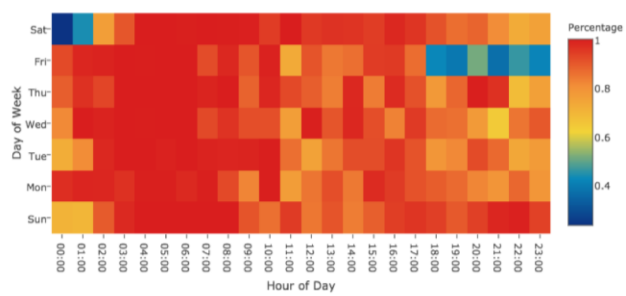

Figure 2. Subject 14's percentage of sedentary behaviors in 1-hour buckets grouped by the Day of Week and Hour of Day

Not all students followed their schedule as closely as Subject 10. For instance, recurrent sedentary patterns were not so clear for Subject 14 (e.g., the sleep and wakeup times). However, what we can tell from Subject 14's heat map (Fig. 2), s/he was very active between Friday evening and Saturday midnight. We speculate that s/he may have attended parties or exercised at gyms during this period of time. This guess can be validated from his/her geolocation logs but this is out of the scope of this study.

2) Weekly Recurrent Pattern

Modern day schedules (especially undergraduate students') follow a weekly schedule. For example, students may have classes on Mondays, Wednesdays, and Fridays at 10:00AM, and professors may have weekly project meetings, research group meetings, and meetings with their students at specific times every week. To explore this pattern, we performed a cross-correlation analysis on subjects' weekly sedentary behaviors for the same time buckets in two different weeks. The cross-correlation between two weeks is calculated with equation 1,

$$\rho = \frac{E[(P_1-\mu_1)\cdot(P_2-\mu_2)]}{\sigma_1\cdot\sigma_2} \qquad (1)$$

where $P_1 = \{p_{1,1}, p_{1,2}, p_{1,3}, \ldots, p_{1,k}\}$ and $P_2 = \{p_{2,1}, p_{2,2}, p_{2,3}, \ldots, p_{2,k}\}$ are the percentages of sedentary behaviors in each time bucket for the compared two weeks respectively, μ is the mean, and σ is the standard deviation. Fig. 3 shows the correlations of sedentary behaviors in week-vs-week pairs for Subject 10. Not surprisingly, the similarity between Subject 10's weeks was high since Subject 10 strictly followed his/her schedules based on our observation on his/her daily patterns.

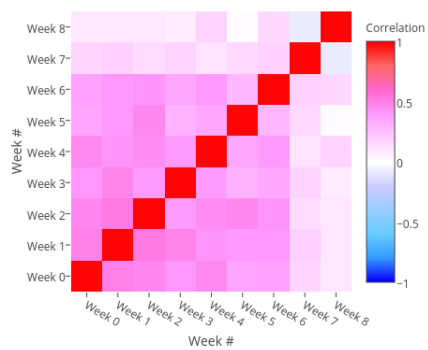

Figure 3. Week vs. Week cross-correlation of Subject 10's sedentary behaviors

C. *Procedure of Building Model*

While our prior visualizations focus on daily and weekly recurrent patterns, recurrent behaviors can occur at any timescale. We now propose a computational model that captures the recurrent patterns on all time-scales including hourly, daily and weekly recurrent patterns. In theory, one could exhaustively search for all recurrent cycles by testing all the possible timescales. However, it is not efficient.

To find recurrent patterns, we consider the percentage of sedentary behavior in each bucket as a periodic signal whose cycle depends on how often the sedentary behavior is performed. Since time cycles are periodic in nature, it is reasonable to explore the signal in frequency domain. We perform discrete Fourier transform on the signal with Cooley-Tukey algorithm to convert the signal from time domain to frequency domain [20]. After the transformation, we will get a center amplitude D and a list of decomposed components—periodic (cosine) functions, $g_1, g_2, g_3, \ldots, g_n$:

$$g_i(t) = A_i \cdot \cos(2\pi \cdot f_i \cdot t + \varphi_i) \qquad (2)$$

and with their *amplitudes* (A_i), *frequencies* (f_i), and *phases* (φ_i). Then, the original signal—time series of sedentary behaviors—can be reconstructed by summing these functions:

$$g(t,n) = \sum_{i=1}^{n} g_i(t) + D \qquad (3)$$

Because the reconstructed sedentary behavior signal may be out the possible range of percentage, 0% to 100%, we further normalize the signal by clamping:

$$h(t,n) = \min(1, max(0, g(t,n))) \qquad (4)$$

However, not all the components are useful for reconstructing the original signal. Some components may be noise. To find the best combination of components, we sort the components by amplitude in descending order ($A_1 \geq A_2 \geq A_3 \geq \ldots \geq A_n$) and find the first k components by minimizing MSE of the predicted percentage of sedentary behavior:

$$\arg\min_k(mse(k)) \qquad (5)$$

where

$$mse(k) = E\{[h(t,k) - signal(t)]^2\} \qquad (6)$$

IV. RESULT

Table I shows subjects exhibiting sedentary recurrent

TABLE I. BEST FITTED MODELS AND THEIR SUBJECTS

Subject ID	Bucket Size	MSE	1/f (Component 1)	1/f (Component 2)	1/f (Component 3)	1/f (Component 4)
12	6 hours	0.003817	24 hours			
10	4 hours	0.004320	24 hours			
2	5 hours	0.004471	76.8 hours	33.4 hours	27.9 hours	24 hours
57	5 hours	0.009150	24 hours			
51	6 hours	0.011938	168 hours	24 hours		
4	5 hours	0.012449	24 hours			

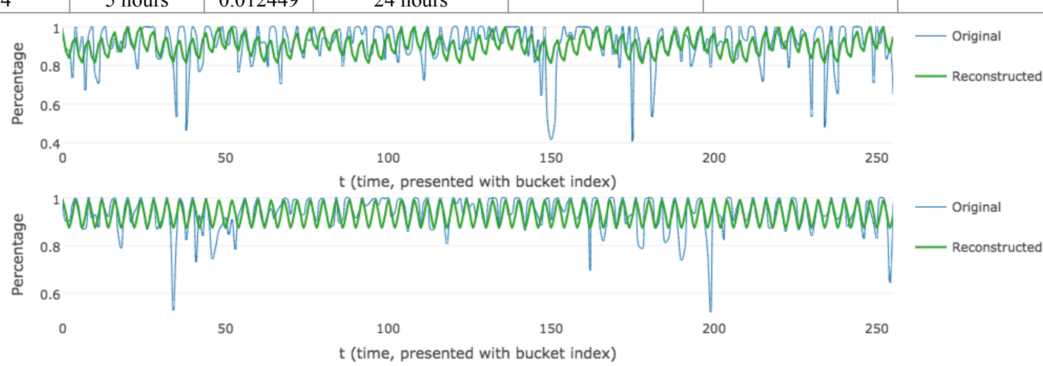

Figure 4. Reconstructed (**green**) vs. original (**blue**) sedentary signal of Subject 51 (top) and Subject 12 (bottom)

patterns and their best fitted (MSE < 0.02) recurrent model (bucket size and components). As previously mentioned, our algorithm tests 6 different bucket sizes for each subject and chooses the bucket size with the lowest MSE. For instance, for Subject 51, the bucket size with the lowest MSE is 6 hours, which implies that we can best predict Subject 51's sedentary behaviors at a timescale of 6 hours (12AM - 6AM (night), 6AM to 12PM (morning), 12PM to 6PM (afternoon), and 6PM to 12AM (evening)). The k value determined automatically by our algorithm for subject 51 using a 6-hour bucket is 2, which means 2 cosine functions can best reconstruct subject 51's sedentary signal. The 2 corresponding cycles (1/f) are 168 hours (1 week) and 24 hours (1 day), which correspond to the *daily recurrent pattern* and *weekly recurrent pattern* we have assumed in the previous section. Fig. 4 (top) compares the original sedentary signal and the reconstructed signal.

Among all the subjects, Subject 12's best fitted model has the lowest MSE (Fig. 4 bottom). The MSE of this model was 0.003817 and the distribution of *percentage differences* (or called *errors*) is shown in Fig. 5. For example, if the original sedentary behavior percentage at $t=t_i$ is 80% and in the reconstructed model calculates the percentage at $t=t_i$ is 90%, then the difference is -10 (shown on X-axis). If 20 sedentary behavior percentages have -10% difference from reconstructed ones, the count of "-10" is 20 (shown on Y-axis).

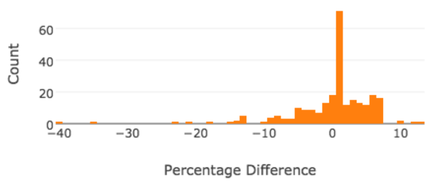

Figure 5. Distribution of differences between the original data points and reconstructed data points of Subject 12

For some subjects (e.g., Subject 24), the minimal MSE was achieved when only the center amplitude D was used and no component (cosine function) was added to reconstruct his/her sedentary behavior model. This type of results implies that they did not have recurrent sedentary behaviors, but they did have random sedentary behaviors if their *percentage differences* obey normal distribution. This conclusion can be validated by examining their physical activity logs manually, which is outside the scope of this study.

V. Conclusion and Future Work

In the StudentLife dataset, we identified students who had recurrent sedentary behaviors automatically by performing discrete Fourier analysis on their physical activity time-series and selecting components by minimizing MSE. People who had recurrent sedentary behaviors yielded periodic functions as the outcome of the model building procedure. People who did not have recurrent sedentary behaviors had constant center amplitude value as the outcome. This model building procedure can be run automatically on computer or mobile devices like smartphone. Using the periodic function generated for a specific person, we believe that we can predict his/her future sedentary behaviors by calculating the possibility of behavior at given time *t*, with his/her periodic functions. The accuracy of prediction requires further investigation.

We need to highlight that this study has several limitations. First, since all subjects in the StudentLife were all students, more investigation is required to examine how well our model works for other occupations. Second, subjects may not have carried their smartphones all the time, which may introduce erroneous activity detection (e.g., a subject left his/her phone in a locker while working out in the gym). In future, we will repeat the experiment with more diverse subject pool. Subjects will be given wearable devices, such as Fitbit trackers, to track their 24×7 physical activity. We will also develop a sitting-or-not classifier to discriminate non-sitting stationary activities, similar to the Apple Watch's sitting alert feature.

References

[1] R. R. Pate, J. R. O'Neill, and F. Lobelo, "The evolving definition of 'sedentary'.," *Exerc. Sport Sci. Rev.*, vol. 36, pp. 173–8, Oct. 2008.

[2] E. G. Wilmot, C. L. Edwardson, F. A. Achana, M. J. Davies, T. Gorely, L. J. Gray, K. Khunti, T. Yates, and S. J. H. Biddle, "Sedentary time in adults and the association with diabetes, cardiovascular disease and death: systematic review and meta-analysis," *Diabetologia*, 2012.

[3] N. Owen, T. Sugiyama, E. E. Eakin, P. A. Gardiner, M. S. Tremblay, and J. F. Sallis, "Adults' sedentary behavior determinants and interventions.," *Am. J. Prev. Med.*, vol. 41, pp. 189–96, Aug. 2011.

[4] S. F. M. Chastin, N. Fitzpatrick, M. Andrews, and N. DiCroce, "Determinants of sedentary behavior, motivation, barriers and strategies to reduce sitting time in older women: a qualitative investigation.," *Int. J. Environ. Res. Public Health*, Jan. 2014.

[5] C. Abraham and S. Michie, "A taxonomy of behavior change techniques used in interventions.," *Heal. Psychol.*, 2008.

[6] Fitbit Inc., "Fitbit." [Online]. Available: http://www.fitbit.com/.

[7] Google Inc., "Android APIs - Google Fit." [Online]. Available: https://developers.google.com/fit/android/. [Accessed: 05-May-2015].

[8] Q. He, "On11, be healthy." [Online]. Available: http://on11.mobi/.

[9] J.-F. Etter and T. V. Perneger, "Effectiveness of a Computer-Tailored Smoking Cessation Program," *Arch. Intern. Med.*, Nov. 2001.

[10] M. K. Campbell, B. M. DeVellis, V. J. Strecher, A. S. Ammerman, R. F. DeVellis, and R. S. Sandler, "Improving dietary behavior: the effectiveness of tailored messages in primary care settings.," *Am. J. Public Health*, vol. 84, no. 5, pp. 783–787, May 1994.

[11] Q. He and E. Agu, "On11: An Activity Recommendation Application to Mitigate Sedentary Lifestyle," in *Proceedings of the 2014 workshop on physical analytics - WPA '14*, 2014, pp. 3–8.

[12] 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 *Proc.of UbiComp '14 Adjunct*, 2014, pp. 3–14.

[13] V. J. Strecher and I. M. Rosenstock, "The health belief model," *Cambridge Handb. Psychol. Heal. Med.*, pp. 113–117, 1997.

[14] I. Ajzen, "The theory of planned behavior," *Organ. Behav. Hum. Decis. Process.*, vol. 50, no. 2, pp. 179–211, Dec. 1991.

[15] M. Conner and P. Norman, *Predicting Health Behaviour*. McGraw-Hill Education, 2005.

[16] S. P. Shah, W. L. Lam, R. T. Ng, and K. P. Murphy, "Modeling recurrent DNA copy number alterations in array CGH data," *Bioinformatics*, vol. 23, no. 13, pp. i450–i458, Jul. 2007.

[17] J. Z. Shan, "Methods and systems for identifying recurrent patterns," 03-Apr-2007.

[18] F. H. Schilling, P. Altena, and H. A. K. Mastebroek, "The computational measurement of apparent motion: A recurrent pattern recognition strategy as an approach to solve the correspondence problem," *Biol. Cybern.*, vol. 62, no. 6, pp. 463–473, Apr. 1990.

[19] R. S. Bogart, "Recurrence of solar activity: Evidence for active longitudes," *Sol. Phys.*, vol. 76, no. 1, pp. 155–165, Feb. 1982.

[20] J. W. Cooley and J. W. Tukey, "An algorithm for the machine calculation of complex Fourier series," *Math. Comput.*, May 1965.