Smartphone Usage Contexts and Sensable Patterns as Predictors of Future Sedentary Behaviors

Qian He[1] and Emmanuel O. Agu[2]

Abstract— **Sedentary behaviors such as prolonged occupational and leisure-time sitting are now ubiquitous in modern societies. Sedentary time is positively associated with increased risk of obesity, diabetes, cardiovascular disease, and all-cause mortality. Smartphones can sense the sedentary behaviors performed by their users, as well as the contexts (situations) in which sedentary behaviors occur. In this paper, we explore whether the contexts that can be sensed by users' smartphones can be used to predict their future sedentary behaviors reliably. We analyze data gathered in a term-long study of 49 college students in order to discover their sedentary behavior patterns and contexts strongly correlated with sedentary states. The ability to predict sedentary behaviors will facilitate more effective computer-driven interventions based on the theory of planned behavior. Using logistic regression, we are able to classify user context variables such as location, time, and app usage to predict if the user will be "very sedentary" in the next hour with a precision of 73.1% (recall of 87.7%).**

I. INTRODUCTION

Sedentary behaviors caused by screen-based occupations such as desk jobs, and leisure-time activities such as watching TV after dinner, are now ubiquitous across all ages, and have been associated with deleterious health outcomes regardless of physical activity levels [1], [2]. Prolonged sedentary time increases the Relative Risk (RR) of diabetes (112% increase), cardiovascular mortality (90% increase), and all-cause mortality (49% increase) [3]. Consequently, methods to mitigate sedentary behaviors have become a topical but challenging research problem.

Manini *et al.* suggested that environments in which people have a high exposure to sedentary behaviors could benefit from new technologies (e.g., smartphone alerts of elapsed sedentary times), environmental changes (e.g., active workstations), policies (e.g., periodic desk breaks), and norms surrounding prolonged sitting (e.g., standing meetings) [4]. In many behavior intervention studies, the sedentary behavior contexts considered primarily were TV viewing, screen-focused behaviors in domestic environments, prolonged sitting in the workplace, and time spent sitting in automobiles [1]. In 2015, 68% of American adults and 86% of young adults (18-29-year-olds) owned a smartphone [5]. As smartphones become ubiquitously owned, new types of screen-focused sedentary behaviors are increasing (e.g., watching YouTube on smartphones) while TV viewing is decreasing [6].

In addition to being a determinant (cause) of new types of sedentary behaviors, smartphones have a wide range of sensors (e.g., accelerometer, GPS, and light sensor) that can be used to sense its user's activities and contexts. For instance, smartphone sensors can be used to sense whether the user is walking, running or making phone calls, what apps are being used, as well as the locations and times of these activities (behavior contexts).

In this paper we investigate whether the contexts sensed by user's smartphone can be used to reliably predict their future sedentary behaviors. Our approach is to characterize the contexts (e.g., location types, time, app usage) passively sensed by smartphones carried by 49 students in the publicly available StudentLife dataset [7]. These contexts are then used to synthesize a predictive model. Our result shows that a Logistic Regression classifier can utilize 22 context variables such as user location, time, and app usage to predict if students will be very sedentary in the next hour with a precision of 73.1% (recall of 87.7%).

The reliable prediction of future sedentary behaviors will enable just-in-time computer-driven interventions based on the theory of planned behavior wherein modifications of a person's day/week/month at the planning stage are more likely to be successful [8]. Our overarching philosophy is that prevention of sedentary behaviors before they occur is better than the interventions provided reactively. For instance, a smartphone app that reminds its user to stand up every 30 minutes *before* he/she starts watching TV is likely to have higher compliance than one that sends the user a notification every 30 minutes *after* he/she has started watching [9].

II. RELATED WORK

A. Smartphone Usage Analysis

Since smartphones are widely owned, research around understanding how people use their smartphones has become topical. Researchers have investigated smartphone usage patterns including use frequency, app popularity, network traffic, energy consumption, and (un)installation behaviors [10]. Smartphone usage patterns that reliably predict various real-world user behaviors have been proposed by several researchers. Likamwa *et al* [11] demonstrated that the moods of smartphone users could be inferred from their smartphone usage patterns. Sarker *et al.* [12] found that *location*, *affect*, *activity type*, *stress*, *time*, and *day of the week* features of smartphone users can predict their availability. Pielot *et al.* [13] found that user boredom could be predicted from their smartphone usage with an accuracy of up to 82.9%. Our goal is to determine smartphone-sensed contexts that reliably predict future sedentary behaviors.

[1]Q. He and [2]E. O. Agu are both with the Department of Computer Science, Worcester Polytechnic Institute, Worcester, MA 01609 USA qhe, emmanuel at cs.wpi.edu

978-1-5090-1166-7/16/$31.00 ©2016 IEEE

B. Behavior Prediction

Human behaviors have some repetitive patterns that can be modeled and predicted. Pentland and Liu used Kalman filters and Markov chains to predict actions performed by automobile drivers [14]. Graph-based probabilistic predictive models such as Hidden Markov Model, Conditional Random Fields, and rule-based models have been suggested for future activity prediction research [15]. Computer vision techniques to predict human behaviors in videos have also been proposed [16].

III. DATASET USED IN OUR EXPERIMENTS

Our approach is to analyze an existing dataset in order to discover smartphone usage patterns and activity contexts that can reliably predict future sedentary behaviors. We utilized a public dataset from Dartmouth College, called "StudentLife" [7], which contains logs of 49 Dartmouth College students including smartphone-sensed behaviors (such as app usage, locations visited, and ambient sound/light) as well as their periodic responses to mental health (e.g., mood, affect, and depression) questionnaires. The StudentLife dataset provided us with real physical activity data—from which we extract sedentary behaviors—and contexts of the students over a 10-week term (spring of 2013).

A. Dataset Preprocessing

The raw StudentLife data required some pre-processing steps to facilitate analysis, including:

1) Discretizing Time into 1-Hour Buckets: The StudentLife dataset contained physical activity logs with activity labels (*Stationary*, *Walking*, *Running*, and *Unknown*) that were sampled every 2–3 seconds in 1 of every 4 minutes. Smartphone app usage was logged every 20 minutes, and location logs were sampled every 10 minutes. We discretized time such that all students' activity time series data were associated with "1-hour time buckets". For instance a bucket might represent *2013-03-01 12:00AM–01:00AM*. In total, we constructed 65,601 such 1-hour buckets for the 49 students in the StudentLife dataset.

2) Computing Sedentary Levels within Each 1-Hour Bucket: Student activities labeled as *Stationary* were mostly sedentary behaviors such as sitting and lying down. Within each 1-hour bucket, we computed each subject's *sedentary level*, which we defined as the percentage of physical activities performed in a given 1-hour bucket that had the label *Stationary*. As a numerical example, if 360 physical activities out of 450 physical activities performed by a student in a 1-hour time bucket were labeled as *Stationary*, we calculated the student's sedentary level in that time bucket as 80% ($= \frac{360}{450} \times 100\%$). In practice, coarse ranges of sedentary levels will likely be more useful than exact sedentary level values. Thus, we further discretized sedentary levels into 3 ranges: *Very Sedentary* (99.99–100%), *Sedentary* (90.52–99.99%), and *Less Sedentary* (0–90.52%). These cutoff values were chosen using an unsupervised equal-frequency binning strategy that placed an equal number of activity logs in these three bins.

3) Placing Activity Locations in Corresponding Buildings: In the Studentlife dataset, the location of student activities were represented as either GPS logs (more accurate) or SSID (identifier) logs of Wi-Fi access points closest to the student. We converted Wi-Fi SSIDs to Dartmouth College campus building names, which were more meaningful. Since the mapping of Wi-Fi access point SSIDs to building names was not publicly available, we retrieved this information in the JavaScript Object Notation (JSON) format from Dartmouth College's map service. As an example, the "*sudikoff*" SSID was mapped to building "*Sudikoff Hall*". In total, we mapped 101 SSIDs to Dartmouth College's buildings.

4) Extracting Student App Usage Information: The 49 students in the StudentLife study were given Google Nexus 4 Android phones for daily use. A data gathering app running in the background of these phones recorded various types of data. For each Android app used by the smartphone user, its Java package name (e.g., "*com.google.android.gm*") was recorded. In total, 698 unique package names were found in 1,990,510 app usage logs. In order to determine which apps corresponded with these package names, we cross-referenced the 698 unique package names with Google Play Store, Amazon AppStore, Wandoujia Market, and online searches. We found the ***app name*** (e.g., "*Gmail*"), the ***app developer*** (e.g., "*Google Inc.*"), and ***app category*** (e.g., "*Communication*") information of 480 unique apps. Apps with package names such as "*com.example.xxx*" and "*edu.dartmouth.cs65.xxx*" were considered as homework apps programmed by students. We labeled 204 such homework apps with the app name "*Study*" and categorized them as "*Experiment*" app.

IV. CONTEXTS IN WHICH SEDENTARY BEHAVIORS OCCUR

We now present smartphone-sensed contexts and app usage of students in the StudentLife dataset, which were strongly correlated with their sedentary behaviors, with our intuition on why certain patterns and correlations occurred.

A. Sedentary Times

Figure 1 is a heatmap of the average sedentary levels for all 49 students in a *"Day of Week" (Y-axis) vs. "Hour of Day" (X-axis)* grid. For instance, the cell at the bottom left corner represents that the average sedentary level of all 49 students between 12:00AM–01:00AM on Sunday was 81.60%. Students were generally more sedentary at night (10:00PM–08:00AM) than daytime (08:00AM–10:00PM), probably because they mostly slept at night. Students were also relatively more sedentary on Sunday, Monday, and Tuesday nights than other nights of the week. We speculate that the first half of a week was busier than the second half for Dartmouth College students, possibly caused by class schedules and course deadlines that pushed students to sit while studying and using their computers. Students were generally more active in the mornings (08:00AM–12:00PM) than the afternoons (12:00PM–05:00PM), probably because more courses and meetings were scheduled in the morning.

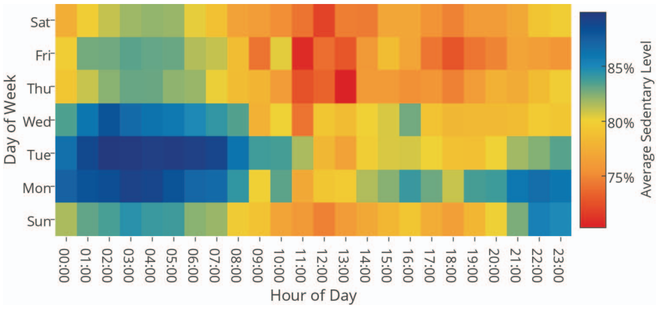

Fig. 1. Average Sedentary Level

B. Locations of Sedentary Behaviors

The locations at which sedentary behaviors occurred were scattered across the Dartmouth College campus. We analyzed the specific buildings in which *Very Sedentary* (99.99–100%) behaviors occurred. The buildings ranked 1st ("North Park Street", the graduate student apartments) and 3rd ("Mass Row Cluster", the undergraduate residence hall) were places where students sleep and relax (mostly sedentary behaviors). The building ranked 2nd highest was "Sudikoff Hall", where the Computer Science Department is located. We speculate that students are highly sedentary in this building since computer science research and study require long periods of sitting in front of computers. We further categorized the buildings into types shown in Figure 2. As expected, residential buildings and academic buildings were places where the most sedentary behaviors occurred.

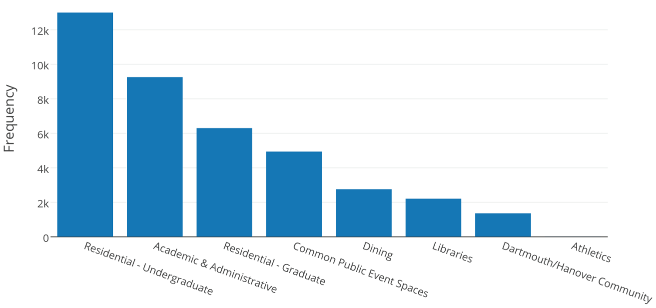

Fig. 2. Categories of Buildings Where Sedentary Behaviors Occurred

C. Sound/light Levels Correlated with Sedentary Behaviors

Different environmental ambient sound/light levels may be associated with different activities types. For instance, sleep may be associated with quiet and dark locations. Figure 3 shows the ambient sound levels associated with our three ranges of sedentary levels. In the dataset, ambient sound levels were characterized as *"Silence"* (quiet), *"Voice"* (people were talking near the smartphone), and *"Noise"* (noisy). Students tended to be more sedentary in quiet environments. Similarly, we found students were more active in bright environments, probably because students were busy with lively activities during the day (or bright places) and were more sedentary in dark places (e.g., sleeping at night).

D. Smartphone Usage Correlated with Sedentary Behaviors

Of the 480 apps used by students, the most widely installed app types were *Games* (95 apps), *Tools* (80 apps), and *Productivity apps* (51 apps). Although students installed numerous unique game apps, many were played infrequently.

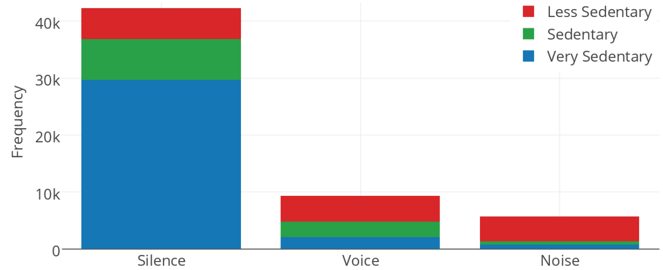

Fig. 3. Ambient Sound vs. Sedentary Behaviors

The usage of certain apps may influence sedentary levels. For instance, bank accounts are typically reviewed while sitting. Figure 4 shows the average sedentary levels when students were using different types of apps. Students were very sedentary when using *Transportation* apps (e.g., *Uber* and *Lyft*), *Finance* apps (e.g., *Bank of America*), *Photography* apps (e.g., *Picasso*, a drawing app), *Books and Reference* apps (e.g., *Google Play Books*).

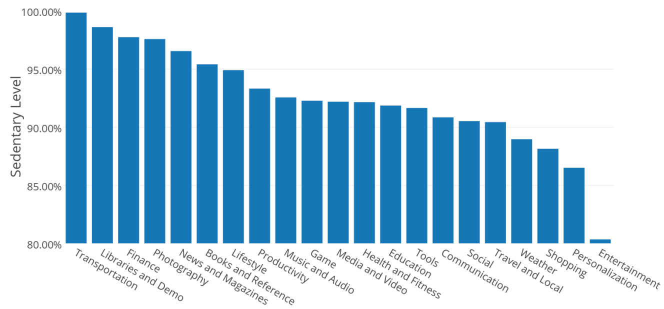

Fig. 4. Sedentary Level vs. App Category

V. PREDICTING SEDENTARY BEHAVIORS

We denote the previously explored m context variables, such as time (*Hour of Day* and *Day of Week*), location, and environment, at time t as $X_{t,1}, X_{t,2}, X_{t,3}, \ldots, X_{t,m}$. We denote a sedentary behavior that occurred at time t as Y_t, and at time $t+1$ as Y_{t+1}. We then use multivariate context variables $\{X_{t,i} \,|\, i \in [1, m]\}$ as inputs to predict a future sedentary behavior Y_{t+1}—$\langle X_{t,1}, X_{t,2}, X_{t,3}, \ldots, X_{t,m}\rangle \Rightarrow Y_{t+1}$ (Figure 5).

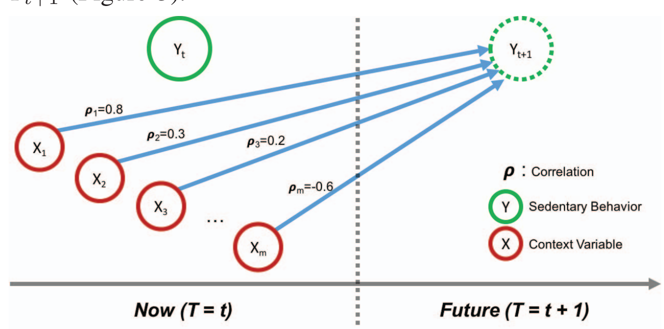

Fig. 5. Prediction of Future Sedentary Behavior Using Context Variables

The intuition behind this prediction model is that we believe a person's current behavior and context affect his/her future sedentary behavior. For instance, if a student is exercising vigorously at 8AM, he/she has high chance of being sedentary (resting) in the next hour.

A. Context Variables as Predictors

In our model, we use the following 22 smartphone-sensed context variables generated by pre-processing the StudentLife dataset to predict sedentary levels: (1) ***dayOfWeek***, day of the week (Monday to Sunday); (2) ***hourOfDay***, hour of the day (00:00 to 23:00); (3) ***activityMajor***, the type of activity with the most instances in a 1-hour time bucket; (4) ***audioMajor***, the type of audio with the most instances in a 1-hour time bucket; (5) ***radius***, the radius (in meters) of the area covers the student's geographical movement during a 1-hour time bucket; (6) ***latitude***, the average latitude of the student's location; (7) ***longitude***, the average longitude of the student's location; (8) ***travelstate***, moving or not, sensed by GPS; (9) ***wifiMajor***, the major Wi-Fi SSID connected to; (10) ***isCharging***, whether the smartphone was charging; (11) ***isLocked***, whether the smartphone was locked; (12) ***isInDark***, whether smartphone was in the dark; (13) ***isInConversation***, whether student was conversing; (14) ***appNameMajor***, the app used most frequently in a 1-hour time bucket; (15) ***appCategoryMajor***, the app category most frequently used in a 1-hour time bucket; (16) ***diningPlace***, the on-campus restaurant at which the student purchased a meal; (17) ***diningType***, the type of meal purchased by the student on campus (*Breakfast*, *Lunch*, *Supper*, or *Snack*); (18) ***courseName***, the name of course the student should be attending; (19) ***courseWifi***, whether the student should be attending a course the Wi-Fi SSID of course classroom; (20) ***afterLastDeadline***, how many seconds since the last homework was due; (21) ***beforeNextDeadline***, how many seconds before the next homework is due; (22) ***hasCalendarEvent***, whether an event is scheduled in the calendar.

B. Predictive Models

Using the results of our sedentary behavior analysis, we then attempted to predict future sedentary behaviors from smartphone-sensed contexts. Predicting sedentary levels can be considered a classification problem if the future sedentary level is treated as the class variable. We utilized two probability models: Naïve Bayes classifier, which maximizes the joint probability $P(Y_{t+1}, X_{t,1}, X_{t,2}, \ldots, X_{t,m})$, and Logistic Regression classifier, which maximizes the conditional probability $P(Y_{t+1}|X_{t,1}, X_{t,2}, \ldots, X_{t,m})$. The accuracy of using Naïve Bayes classifier and Logistic Regression classifier to predict sedentary behavior in levels (*Very Sedentary*, *Sedentary*, and *Less Sedentary*), in terms of correctly classified instances using 10-fold cross-validation, were 59.7% (with recall 64.2%) and 60.8% (with recall 62.6%) respectively. In predicting *Very Sedentary* behaviors which will likely require behavior intervention, Naïve Bayes classifier achieved a precision of 70.2% (with recall 86.0%) and Logistic Regression classifier achieved a precision of 68.6% (with recall 89.3%).

Finally, we explored using the user's current sedentary behavior Y_t in addition to their current context variables $\{X_{t,i} \,|\, i \in [1, m]\}$ to predict their future sedentary behavior Y_{t+1}. Our results improved slightly, predicting *Very Sedentary* behaviors with a precision of 73.1% (recall of 87.7%). All results are shown in Table I. Because Naïve Bayes classifier has the naïve independence assumption of context variables, which may not be true, Logistic Regression classifier generally has better accuracy.

TABLE I
RESULT FOR PREDICTING OF *"Very Sedentary"* BEHAVIOR

	Naïve Bayes		**Logistic Regression**	
	Precision	*Recall*	*Precision*	*Recall*
Without Current	70.2%	86.0%	68.6%	89.3%
With Current	69.2%	84.4%	73.1%	87.7%

VI. CONCLUSION AND FUTURE WORK

We analyzed a public dataset to identify user contexts that can be used to predict future sedentary behaviors. Our result shows that a Logistic Regression classifier can utilize smartphone-sensed context variables such as user location, time, and app usage to predict if a student will be very sedentary in the next hour with a precision of 73.1% (recall of 87.7%). We believe that the ability to predict next-hour sedentary behavior will enable more effective interventions based on the theory of planned behavior. The models used in this study are the simplest predictive models. In future work, we will study temporal patterns of contexts and explore high-order observable Markov chain and linear-chain Conditional Random Field for predicting sedentary behaviors.

REFERENCES

[1] N. Owen, T. Sugiyama, E. E. Eakin, P. A. Gardiner, M. S. Tremblay, and J. F. Sallis, "Adults' sedentary behavior determinants and interventions." *American journal of preventive medicine*, 2011.

[2] A. Biswas, P. I. Oh, G. E. Faulkner, R. R. Bajaj, M. A. Silver, M. S. Mitchell, and D. A. Alter, "Sedentary Time and Its Association With Risk for Disease Incidence, Mortality, and Hospitalization in Adults," *Annals of Internal Medicine*, 2015.

[3] 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.

[4] T. M. Manini, L. J. Carr, A. C. King, S. Marshall, T. N. Robinson, and W. J. Rejeski, "Interventions to Reduce Sedentary Behavior," *Medicine & Science in Sports & Exercise*, 2015.

[5] M. Anderson, "U.S. Technology Device Ownership 2015," 2015.

[6] M. Ingram, "The Smartphone Is Eating the Television, Nielsen Admits," 2015.

[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 *Proceedings of UbiComp '14 Adjunct*. ACM Press, 2014.

[8] I. Ajzen, "The theory of planned behavior," *Organizational Behavior and Human Decision Processes*, 1991.

[9] D. S. Bond, J. G. Thomas, H. A. Raynor, J. Moon, J. Sieling, J. Trautvetter, T. Leblond, and R. R. Wing, "B-MOBILE," *PLoS ONE*, Jun 2014.

[10] H. Li, X. Lu, X. Liu, T. Xie, K. Bian, F. X. Lin, Q. Mei, and F. Feng, "Characterizing Smartphone Usage Patterns from Millions of Android Users," in *Proceedings of IMC '15*. ACM Press, 2015.

[11] R. LiKamWa, Y. Liu, N. D. Lane, and L. Zhong, "MoodScope: Building a Mood Sensor from Smartphone Usage Patterns," in *Proceeding of MobiSys '13*. ACM Press, 2013.

[12] H. Sarker, M. Sharmin, A. A. Ali, M. M. Rahman, R. Bari, S. M. Hossain, and S. Kumar, "Assessing the availability of users to engage in just-in-time intervention in the natural environment," in *Proceedings of UbiComp '14 Adjunct*. ACM Press, 2014.

[13] M. Pielot, T. Dingler, J. S. Pedro, and N. Oliver, "When attention is not scarce - detecting boredom from mobile phone usage," in *Proceedings of UbiComp '15*. ACM Press, 2015.

[14] A. Pentland and A. Liu, "Modeling and Prediction of Human Behavior," *Neural Computation*, 1999.

[15] D. J. Cook and N. C. Krishnan, *Activity Learning, Discovering, Recognizing, and Predicting Human Behavior from Sensor Data*. John Wiley & Sons, Inc, 2015.

[16] Y. Fu, Ed., *Human Activity Recognition and Prediction*. Springer International Publishing, 2016.