

Towards Sedentary Lifestyle Prevention: An Autoregressive Model for Predicting Sedentary Behaviors

Qian He

Department of Computer Science
Worcester Polytechnic Institute
Worcester, MA 01609
Email: qhe@cs.wpi.edu

Emmanuel O. Agu

Department of Computer Science
Worcester Polytechnic Institute
Worcester, MA 01609
Email: emmanuel@cs.wpi.edu

Abstract—Sedentary lifestyles are ubiquitous in modern societies. Watching television, playing video games, and using the computer while sitting are examples of sedentary behaviors that are now common worldwide. Multiple research studies have found that the length of sedentary time is linked with an increased risk of obesity, diabetes, cardiovascular disease, and all-cause mortality. Medical researchers have identified time, daily rhythm, and past sedentary habits as three determinants (causes) of sedentary behaviors. Building on these insights, in this paper we propose an approach to automatically discover patients’ temporal patterns of sedentary behaviors from raw activity logs by using an autoregressive model with maximum entropy method. With this approach, health care providers will be able to predict patients’ deleterious sedentary behaviors as evidence to drive computer-generated preventive interventions based on behavior change theories such as the Theory of Planned Behavior.

I. INTRODUCTION

Sedentary behaviors are lifestyle activities performed during a person’s day, which do not increase energy expenditure substantially above resting levels. Examples include sleeping, sitting, lying down, watching television, and other forms of screen-based entertainment. Quantitatively, sedentary behaviors are defined as waking behaviors with an energy expenditure ≤ 1.5 METs (Metabolic Equivalent of Task) while in a sitting or reclining posture [1]. Recent study results have shown that sedentary time is associated with an increased risk of diabetes, cardiovascular disease, and cardiovascular and all-cause mortality [2], [3]. Lees and Booth [4] coined a new term—Sedentary Death Syndrome (SeDS)—to characterize disorders caused by sedentary lifestyles resulting in conditions that increase premature death.

Understanding the “determinants”—reproducible associations that are potentially causal [5]—of sedentary behavior is a key step in mitigating it. New targeted interventions could be designed by medical researchers to directly target specific determinants when they occur. For instance, workplace furniture that encourages sitting was identified as one of the determinants of sedentary behavior [6]. In cases where this determinant is identified, adjustable standing desks may be recommended as an intervention.

Sedentary behaviors have been categorized in prior work. Owen *et al* categorized the determinants of sedentary behaviors into 4 domains (*transport, occupation, household, and leisure time*) and 3 environments (settings) in which sedentary behaviors occur (*domestic, workplace, and transportation environments*) [6]. For a specific population—university students, Deliens *et al* categorized the determinants of sedentary behaviors as *individual* (intrapersonal), *social environment* (interpersonal), *physical environment* (community settings), *macro environment*, and *university characteristics* [7]. Although categories of sedentary behaviors are well defined with clear examples, the specific determinants contributing to sedentary behaviors vary from person to person. Moreover, multiple determinants can contribute simultaneously, which makes manual inference challenging even for experienced doctors. Automatic and unintrusive methods to assess the specific determinants will free health care providers from this complex and tiring task, and will lead to personally tailored interventions.

Three common individual determinants of sedentary behavior are *time* (e.g., having limited time for exercise), *daily rhythm* (e.g. students sitting in classes that follow a schedule), and *past sedentary habits* (e.g., watching TV shows every night after work) [7]. In this paper, we propose a method to automatically discover these three determinants by analyzing patients’ smartphone-sensed time-series logs of physical activity. Our method explores a user’s history of sedentary behaviors (the “past sedentary habits” determinant) to find temporal patterns (the “daily rhythm” determinant) that can be used to predict future sedentary behaviors.

The ability to predict sedentary behaviors will facilitate preventive interventions that are guided by the Theory of Planned Behavior (TPB) [8]. TPB suggests that a patient is more likely to participate in recommended interventions to reduce sedentary behaviors if such activities are included in their plans. If it is possible to predict when a user will be sedentary at some future time, suggestions on how they can modify their plans to eliminate or reduce sedentary time can be made. For instance, if a user is predicted to be

sedentary for 3 hours while watching TV shows, he/she can be sent a reminder in advance on through his/her smartphone recommending him/her to stand up and walk around during commercial (advertisement) breaks. TPB suggests that such situation-specific reminders sent in advance are more likely to be accepted by the user.

The rest of this paper is organized as follows: we provide a brief literature review of the related work in Section II and then describe our model in detail in Section III. In section IV, we test our model with a public dataset and evaluate our model predicting sedentary behaviors. Finally, in section V, we discuss the limitations of our work and future work.

II. BACKGROUND

A type of random process that is commonly observed in nature. The variable in the process is time-varying ($\{X_t | t \in T\}$), and it linearly depends on its previous values and a stochastic term (see Equation 1, where ϵ is the stochastic term and c is a constant number). This type of random process is usually modeled with autoregressive (AR) model and has been widely used in different domains where data is represented in time domain and data points have linear relations with previous data points.

$$X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \dots + \alpha_k X_{t-k} + c + \epsilon \quad (1)$$

In medical research, AR models are commonly used in electroencephalogram (EEG) analysis [9] where continuous and linearly correlated biomedical signals are analyzed in time domain. For example, in [10], Kamińska, Katarzyna, and Szelenberger analyzed overnight sleep EEG data with AR model and found the differences in coherence and propagation of EEG activity during sleep and wakefulness. In Brain-Computer Interfaces (BCIs)—a system that can recognize certain patterns in the human brain [11], AR models are used to analyze electroencephalogram (EEG), magnetoencephalography (MEG) and electrocorticography (ECoG) signals.

In finance, a stock's return is considered as a random variable with discrete values over time, yielding a time series. Linear time series analysis, such as AR models, moving-average (MA) models, and mixed AutoRegressive Moving-Average (ARMA) models, provides us a natural framework to study the dynamic structure of such series [12]. In economics, AR model is a standard tool for making economic forecasts [13].

In health behavior change research, behavior predictions are usually subjective [14], and not computational in nature. Predictions rely on subjective evaluations and causal relations (e.g., a person with good intentions to lose weight will be predicted to exercise more), which sometimes leads to wrong or missed predictions. For instance, researchers who have examined the Transtheoretical Model (TTM) [15], a popular behavior change theory, found that the processes of change were not able to predict transitions in smoking stages [16]. In this work, our goal is to predict sedentary behavior based on

user's objective historical activity data rather than subjective terms, which are hard to evaluate and often inaccurate.

III. METHODOLOGY

In this study, we assume that sedentary behaviors are correlated in time domain such that user's next behavior depends on his/her current and previous behaviors. We utilize AR model with Maximum Entropy Method (MEM) [17], [18], [19] to discover such pattern. We explore user's history of sedentary behaviors (the "past sedentary habits" determinant) and find temporal patterns (the "daily rhythm" determinant) that can be used to predict future sedentary behaviors.

The information about sedentary behavior—the necessary inputs to our approach—is simply the time when sedentary behaviors occur (the "time" determinant), denoted as t (e.g., 10:00AM), and the levels of being sedentary—the sedentary time per given period (e.g., sits 5 minutes in a one-hour period), expressed as a percentage and denoted as X_t . This sedentary level can be measured automatically with smartphone's accelerometer or fitness tracker such as Fitbit.

By utilizing AR model, we assume that a person's sedentary behavior—percent of time spent on sitting—is linearly correlated with his/her previous sedentary behaviors. This assumption will be validated through our experiments in the next section. An AR model for sedentary behavior can be defined as:

$$X_t = \sum_{i=1}^l \alpha_i X_{t-i} + \alpha_0 + \epsilon_t \quad (2)$$

where α_i are the autoregression coefficients, l is the order of AR model, and ϵ_t is the noise term. As a first step, we choose the most reasonable values for l of the AR model, which we will expound on in the experiment section. Next, the autoregression coefficients (α_i) are estimated using the series $\{x_1, x_2, x_3, \dots, x_t\}$ of sedentary behaviors from a person's historical data. A number of techniques can be used to perform coefficient estimation. In this study, we use the Maximum Entropy Method (Burg algorithm [19]). We chose the MEM method because it outperformed the LSE method in our pilot experiments.

Our AR model is not fixed and it updates along the time series. For every patient, at time t , an update-to-date AR model is built with automatically sensed and logged activity time series $\{x_{t-l+1}, x_{t-l+2}, x_{t-l+3}, \dots, x_t\}$. Then, this model is used to predict his/her sedentary level x_{t+1} at time $t+1$. Based on this predicted level of being sedentary, a computer system such as the patient's smartphone can suggest real-time health behavior interventions. For example, based on a patient's logged sedentary levels in the morning, if the up-to-date AR model predicts that his/her sedentary level will be high in the afternoon (100% of the afternoon time will be spent on sitting), then, his/her smartphone could pop up a message (intervention) to encourage him/her to stand up and take a walk every 30 minutes.

IV. EXPERIMENT & DISCUSSION

Our AR-based model for sedentary behaviors is validated on an existing dataset. In 2013 Spring, a ubiquitous computing study called “StudentLife” was conducted by Wang *et al.* at Dartmouth College to track the mental health and well-being of students over the course of a term [20]. Forty-eight students were given a smartphone with a data collection application installed. Over the 10-week term, the application collected automatic sensed data such as the students’ activities, schedules, locations visited, and smartphone usage. The StudentLife dataset is publicly available on the project’s website (<http://studentlife.cs.dartmouth.edu>). For our purposes, the StudentLife dataset presents a complete academic term of activity time series of students, including sedentary behaviors that we utilize in validating our model.

Our experiments utilized the time series logs of automatically sensed physical activities from the StudentLife dataset. The physical activity logs consist of a timestamp and an activity type: *Stationary*, *Walking*, *Running*, and *Unknown*. The StudentLife data gathering app recognized each student’s activity type every 2 seconds with 94% accuracy by using a decision tree classifier to classify features extracted from smartphone’s accelerometer data [20].

A. Data Preprocessing

This data preprocessing step is to divide time into time buckets (e.g. 1 hour, 3 hours). As previously explained, the information used by the AR model is a timestamp (t) and a sedentary level (percent of time spent on sitting, x_t). To extract this information from the raw activity logs, we need to select a “bucket” size—the length of time for calculating the sitting percentage. For example, if the bucket size is 1 hour, all the logs in time range $[t, t+1h)$ (1-hour interval) will be grouped into a bucket. For each bucket, we calculate the percentage of sedentary activities (logs with label *Stationary*) performed by the user. If there are 1,800 ($3600s \times 0.5Hz$) logs recorded in one hour and 1,260 logs have *Stationary* labels, then the sitting percentage is 70%.

One limitation of this experiment is that we assume all *Stationary* activities in the StudentLife dataset are sedentary behaviors, which may not be true. For instance, standing still is stationary, but it is not a sedentary behavior. This limitation arose because we used an existing dataset. In future work, we will gather our own data and will accurately classify such false stationary states.

B. Overview of Students’ Sedentary Behaviors

For illustrative purposes, we examine activity patterns of specific students in the StudentLife dataset. After setting bucket size to 1 hour and performing data preprocessing, subject 23’s time series of sedentary levels is shown in Figure 1. Subject 23 spent many full (100% of a bucket) hours being sedentary (sitting or sleeping). To see if Subject 23 sat a lot during the day time, we generate a heatmap with the X axis as *Hour of Day*, Y axis as *Day of Week*, and color representing the sedentary level (Figure 2). Figure 2 shows

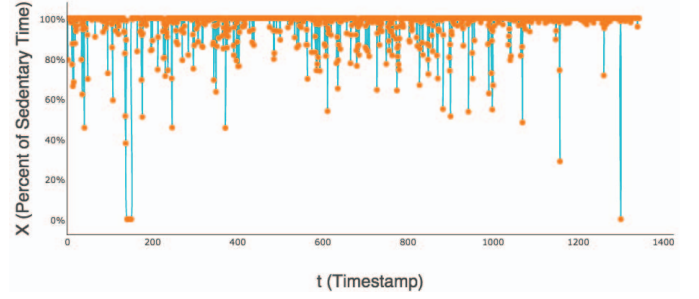


Fig. 1: Subject 23’s Time Series of Sedentary Behavior

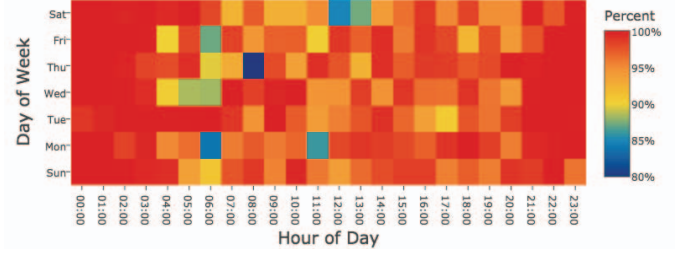


Fig. 2: Weekly View of Subject 23’s Sedentary Behavior

that Subject 23 started to be more sedentary around 9PM and resumed activeness around 5AM.

To build an AR model, we need to know: (1) How many past buckets (the order l) are linearly correlated with a future bucket (X_{t+1})? (2) How are they linearly correlated (the coefficients α_i). These two parameters are experimentally determined in the next two subsections.

C. Choosing Order l

Order l can be interpreted as the length of historical data that affects a person’s behavior. For instance, if the activity performed by a subject in the next hour (X_{t+1}) depends on their current activity, and the subject’s activities in the prior 3 hours (X_t , X_{t-1} , and X_{t-2}), the order of the AR model required to model his/her behavior is 3. Different subjects may exhibit behaviors that are best fitted with models of different orders. However, we assert that human behavior in a modern society is schedule-driven and is relatively regular, exhibiting predictable daily, weekly and monthly cycles [21]. The number of possible orders is not infinite.

To get a sense of how much the sedentary behaviors performed in a given bucket depends on the subject’s history, we calculate the lag- τ autocorrelation of time series. Lag- τ autocorrelation is calculated as such:

$$R(\tau) = \frac{\sum_{t=\tau+1}^{n-\tau} (X_t - \bar{X})(X_{t-\tau} - \bar{X})}{\sum_{t=1}^n (X_t - \bar{X})^2} \quad (3)$$

where \bar{X} is the average sedentary level and τ is the order of lag. The greater $R(\tau)$ is, the higher the correlation with order τ in time series is. Figure 3 shows lag- τ autocorrelation for different subjects. The Y axis is the subject ID and the X axis is the lag time in hours. Autocorrelation is coded in color (demonstrated with a color scale). The autocorrelation value is

positive if the time series is positively linearly self-correlated. A negative value indicates negative correlation, and a value of zero implies no correlation. The higher the absolute value is, the stronger the correlation is.

In this figure, we found three phenomena: (1) When the lag is less than 6 hours, autocorrelation is positive and high; (2) As the lag increases, autocorrelation decreases; (3) When the lag is in multiples of 24 hours (1 day), autocorrelation becomes relatively high; and (4) When the lag is 168 hours (1 week), autocorrelation is relatively high.

These phenomena imply that sedentary behavior is positively correlated with the sedentary behaviors in the prior 6 hours and sedentary behaviors may repeat on daily or weekly cycles. The first implication means that l should satisfy $6 \text{ hours} \leq l \times \text{bucket size} \leq 168 \text{ hours}$. For example, if the bucket size is 2 hours, l should be 3 at least and at most 84. The second implication means that sedentary behaviors are recurrent on fairly predictable cycles. From a medical perspective, recurrent sedentary behaviors can be anticipated and prevented automatically based on observations of subject's past behaviors.

D. Searching for Optimal Combinations of AR Parameters

In order to build and update the AR model for subjects in the StudentLife dataset, optimal parameters need to be searched automatically. Our findings in the previous section provide information on what ranges of order l should be searched, depending on the selection of bucket size. In this study, specifically, the bucket sizes that will be tested are 1 hour, 2 hour, 3 hour, and 6 hours. Different bucket sizes represent the granularity of the model and predictions (e.g. predicting sedentary behaviors in the next hour or next 6 hours).

There are 324 possible parameter combinations of $(l, \text{bucket size})$. Since the number of combinations is not large and building an AR model with MEM for a subject takes less than 5 seconds on a computer with a 2.7GHz dual-core CPU, we will perform an exhaustive search on the parameter combinations for each subject. To evaluate the goodness of fit, we use Mean Square Error (MSE) as the metric for comparison. At a given time t , we use $\{x_1, x_2, \dots, x_t\}$ to build an AR model and apply this model to predict x_{t+1} . The predicted x_{t+1} is denoted as \hat{x}_{t+1} . MSE is calculated as:

$$MSE = \frac{1}{n} \sum_{i=l+1}^n (\hat{X}_i - X_i)^2 \quad (4)$$

where \hat{X}_i is the sedentary level predicted with the up-to-date AR model built from history $x_1, x_2, x_3, \dots, x_{i-1}$ and X_i is the real sensed sedentary level. Whenever a new data point (x_i) is added to the time series, a new AR model is built and used to predict x_{i+1} . Due to page limits on this paper, we only show the top-10 best parameter combinations and their corresponding MSE values in Table I. The results show that a subject's probability of being sedentary in the next 6 hours is more predictable than in the next 1 hour. We believe that

TABLE I: Top-10 Best Parameter Combinations for AR Model

Subject ID	l	Bucket Size	MSE
23	26	6h	0.002260491
2	28	6h	0.003847917
12	8	6h	0.004364295
10	6	6h	0.005164001
19	16	6h	0.005829452
24	28	6h	0.005944031
35	19	6h	0.006409604
57	8	6h	0.008092296
39	155	1h	0.008917593
44	3	2h	0.010712386

it is because many extraneous (random) factors may affect a person's activity in the next hour, but the influence of these factors will be less significant in 6 hours. We will explore this further in future work.

E. Predicting Sedentary Behaviors using the AR(l) Model

In Figure 4, we show the 1-hour bucket AR(9) model for Subject 23. This AR(9) model uses the sedentary levels observed in the previous 9 hour-long buckets in time series to predict sedentary levels in the next hour. The predicted time series captures the trend of the real time series well ((Figure 4) with an overall MSE 0.009912744.

V. CONCLUSION

In this study, we demonstrated that people's sedentary behaviors are strongly correlated with historical sedentary behaviors in the past 6-hour window and exhibits repetitive daily and weekly patterns. We modeled these sedentary patterns using an AR model enabling prediction of future sedentary behaviors from the past behaviors. We described steps for preprocessing raw activity logs, choosing AR model order, searching activity traces for optimal AR parameters, and our evaluation of AR model using a Mean Square Error (MSE) metric. Our steps can easily be programmed to run on a computing device such as the patient's smartphone or autonomously in the cloud without doctors' help.

In conclusion, we have been able to model the three important individual determinants of sedentary behaviors using our AR-based model by: (1) Performing time series analysis (*time*); (2) Finding order l (*rhythm*); and (3) Estimating AR coefficients and building model on the historical data (past habits). Our current work is based on a single existing student dataset. In future, we hope to test our AR model on other populations.

REFERENCES

- [1] R. R. Pate, J. R. O'Neill, and F. Lobelo, "The evolving definition of "sedentary".*" Exercise and sport sciences reviews*, 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] K. I. Proper, A. S. Singh, W. van Mechelen, and M. J. M. Chinapaw, "Sedentary behaviors and health outcomes among adults: a systematic review of prospective studies." *American journal of preventive medicine*, 2011.

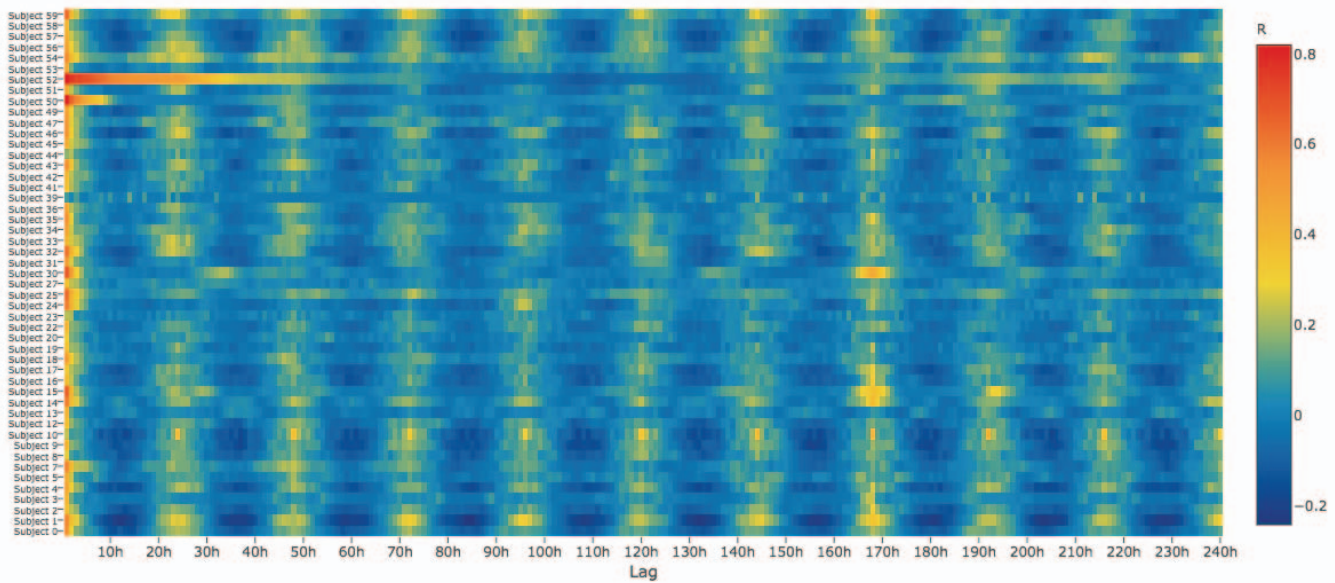


Fig. 3: Lag- τ Autocorrelation of Subjects' Sedentary Behavior Time Series with Different

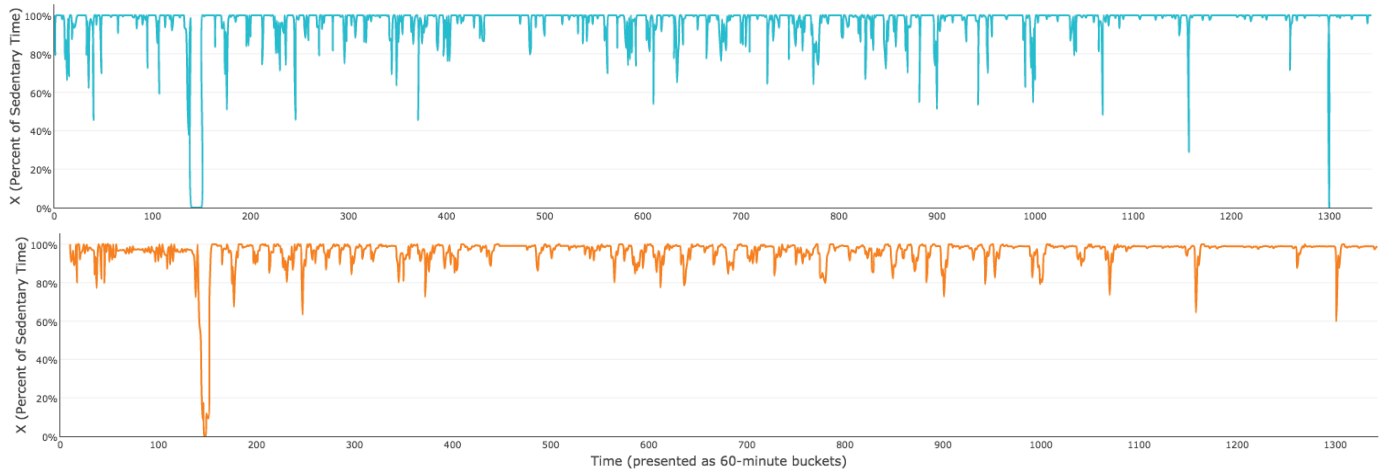


Fig. 4: Real Time Series (top) of Subject 23 and Predicted Time Series (bottom) with AR(9) Model and 60-minute Bucket

- [4] S. J. Lees and F. W. Booth, "Sedentary death syndrome," *Canadian Journal of Applied Physiology*, 2004.
- [5] J. Buckworth, *Exercise Psychology*. Human Kinetics, 2013.
- [6] 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.
- [7] T. Deliens, B. Deforche, I. De Bourdeaudhuij, and P. Clarys, "Determinants of physical activity and sedentary behaviour in university students: a qualitative study using focus group discussions," *BMC Public Health*, 2015.
- [8] I. Ajzen, "The theory of planned behavior," *Organizational Behavior and Human Decision Processes*, 1991.
- [9] L. S. Lustick, B. Saltzberg, J. K. Buckley, and R. G. Heath, "Autoregressive model for simplified computer generation of eeg correlation functions," in *Annual Conference on Engineering in Medicine and Biology. IEEE, New York, NY, USA*, 1968.
- [10] M. Kamiski, K. Blinowska, and W. Szelenberger, "Topographic analysis of coherence and propagation of eeg activity during sleep and wakefulness," *Electroencephalography and Clinical Neurophysiology*, 1997.
- [11] L. F. Nicolas-Alonso and J. Gomez-Gil, "Brain computer interfaces, a review," *Sensors*, 2012.
- [12] R. S. Tsay, *Analysis of Financial Time Series*, 2005.
- [13] R. S. Pindyck and D. L. Rubinfeld, *Econometric Models and Economic Forecasts*, 1998.
- [14] R. Schwarzer, "Modeling health behavior change: How to predict and modify the adoption and maintenance of health behaviors," *Applied Psychology*, 2008.
- [15] J. O. Prochaska and W. F. Velicer, "The transtheoretical model of health behavior change," *American Journal of Health Promotion*, 1997.
- [16] T. A. Herzog, D. B. Abrams, K. M. Emmons, L. A. Linnan, and W. G. Shadel, "Do processes of change predict smoking stage movements? a prospective analysis of the transtheoretical model," *Health Psychology*, 1999.
- [17] T. J. Ulrych and R. W. Clayton, "Time series modelling and maximum entropy," *Physics of the Earth and Planetary Interiors*, 1976.
- [18] T. J. Ulrych and T. N. Bishop, "Maximum entropy spectral analysis and autoregressive decomposition," *Reviews of Geophysics*, 1975.
- [19] J. P. Burg, "Maximum entropy spectral analysis," Ph.D. dissertation, 1975.
- [20] 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 UbiComp '14 Adjunct*, 2014.
- [21] O. o. T. A. U.S.Congress, *Biological Rhythms: Implications for the Worker*. Congress of the U.S., Office of Technology Assessment, 1991.