Probabilistic Latent Variable Modeling for Assessing Behavioral Influences on Well-Being

Ehimwenma Nosakhare
Massachusetts Institute of Technology
Cambridge, MA 02139, USA
ehinosa@mit.edu

Rosalind Picard
Massachusetts Institute of Technology
Cambridge, MA 02139, USA
picard@media.mit.edu

ABSTRACT

Health research has an increasing focus on promoting well-being and positive mental health, to *prevent* disease and to more effectively treat disorders. The availability of rich multi-modal datasets and advances in machine learning methods are now enabling data science research to begin to objectively assess well-being. However, most existing studies focus on detecting the current state or predicting the future state of well-being using stand-alone health behaviors. There is a need for methods that can handle a complex combination of health behaviors, as arise in real-world data.

In this paper, we present a framework to 1) map multi-modal messy data collected in the "wild" to meaningful feature representations of health behavior, and 2) uncover latent patterns comprising *multiple* health behaviors that best predict well-being. We show how to use supervised latent Dirichlet allocation (sLDA) to model the observed behaviors, and we apply variational inference to uncover the latent patterns. Implementing and evaluating the model on 5,397 days of data from a group of 244 college students, we find that these latent patterns are indeed predictive of self-reported stress, one of the largest components affecting well-being.

We investigate the modifiable behaviors present in these patterns and uncover some ways in which the factors work together to influence well-being. This work contributes a new method using objective data analysis to help individuals monitor their well-being using real-world measurements. Insights from this study advance scientific knowledge on how combinations of daily modifiable human behaviors relate to human well-being.

CCS CONCEPTS

• **Applied computing** → **Consumer health**; • **Computing methodologies** → *Latent variable models*.

KEYWORDS

mental health and well-being, latent variable modeling, machine learning

1 INTRODUCTION

Poor mental well-being has become a world-wide problem. The World Health Organization estimates that one in four people will be affected by a mental disorder at some point in their lives.[1] Depression has become the leading cause of disability worldwide, with an estimated 300 million people depressed[2]. Further, it has been shown that not only can depression be elicited by major diseases such as heart disease, but also chronic stress and depression can independently cause physical diseases, including heart disease [9] and diabetes [17]. Thus, there is an increasing need to foster research that promotes well-being and positive mental health.

Mental well-being is dynamic and studies have shown that major contributing factors include daily health behaviors, environmental, and social factors [4]. The growing use of personal health-related mobile applications, fitness trackers and other sensors has made it feasible to collect data to better characterize environmental and lifestyle factors driving mental health outcomes [18]. These devices are able to collect real-time information with the average user likely to produce over one million gigabytes of health-related data in his or her lifetime—the equivalent of about 300 million books.[3] Although there are still limitations, the availability of these multi-modal data sources, with advances in machine learning methods, suggest that it is possible to objectively assess how our daily behaviors and activities influence mental health and well-being.

Typically, studies investigate daily life health and behavior data to find effects of a stand-alone behavior on health and well-being [10, 24, 26]. A problem is that behaviors typically work together and do not affect well-being in isolation, with different combinations of behaviors possibly compensating for each other, and resulting in different health outcomes. To that end, understanding how complex factors – behavioral, environmental and social – work *together* could help provide insights into the groups of behaviors to target when looking to improve an individual's well-being. There is therefore a need to develop a new solution providing personalized models that are able to analyze how complex groups of behaviors contribute to the state of well-being, and provide actionable evidence-based insights to individuals who wish to improve their mental health.

In this paper, we address this complex real-world problem of extracting and identifying important patterns of health behaviors given raw multi-modal data collected in the "wild". The specific

---

[1] http://www.who.int/whr/2001/media_centre/press_release/en/ Accessed Jan. 10, 2019
[2] http://www.who.int/news-room/fact-sheets/detail/depression/ Accessed Jan. 10, 2019
[3] https://www-03.ibm.com/press/uk/en/pressrelease/46609.wss/ Accessed Dec. 15, 2018

group we address is college students, who provided the sensor and smartphone data we use; however, the framework we present is general. We present a new framework for uncovering patterns of modifiable human behaviors that influence the well-being of individuals. We posit that there are latent *patterns* responsible for the set of health behaviors observed in an individual on any given day, and that these patterns consequently contribute to that individual's well-being.

Our solution is two-fold and employs latent Dirichlet allocation (LDA), a method commonly used in summarizing topics within documents of text. First, we map the students' real-world messy data to meaningful feature representations of modifiable health behaviors (analogous to identifying words). Second, we use supervised LDA (sLDA) to model the observed behaviors, and we apply variational inference to uncover latent patterns that give rise to the observations (analogous to uncovering the important topics undergirding a set of documents). We then evaluate the learned latent patterns to see how well they predict student well-being (measured by self-reported stress labels). We find the sLDA patterns are indeed predictive of well-being, and perform significantly better than a control model that obtains patterns in an unsupervised way. We also investigate the student behaviors present in these patterns and examine how they work together to influence the well-being of each individual student. The knowledge achieved by this new framework provides insights into how multiple real-world behaviors might affect health; further, the results help identify candidate variables for future causality studies.

Our multi-modal dataset comes from a prior study that measured Sleep, Networks, Affect, Performance, Stress and Health using Objective Techniques (SNAPSHOT) continuously for 30 days in college students [20, 22]. Chronic stress leads to depression and college years are usually very high stress periods for most young adults. This trend is unfortunately on the increase–in 2017, there was a 12% increase in American college students reporting more than average to tremendous stress when compared to 2013.[4] The high stress period of college years provide a useful case study for demonstrating the new framework. Also, the student behaviors are highly heterogeneous, improving the likelihood that the health findings may apply beyond this dataset.

In Section 2, we discuss the current approaches of modeling human behavior in the context of well-being and the machine learning methods used. In Section 3, we describe the data used in this work. In Section 4 we present our framework: First, we introduce a method of extracting behaviors from complex multi-modal real-world data and binning them to ensure that they are descriptive of a participant's day. Second, we discuss how we model interpretable latent patterns of health behaviors that are predictive of affective states. Finally, Section 5 presents experiments and evaluation, Section 6 presents results, and Section 7 closes with discussion.

2 RELATED WORK

Previous work on the SNAPSHOT dataset has focused on predicting well-being as defined by self-reported stress, mood and health. Taylor et al. use multi-task learning Bayesian and neural network

approaches to predict future well-being, achieving binary prediction accuracies in the range of 72–83% [25]. Sano et al. reported that wearable sensor features such as skin temperature and skin conductance were important features in the classification of high/low stress and mental health [22]. While predicting an individual's well-being is important, it is equally important to help a person become aware of which sets of behaviors – actions that they might control or change – specifically contribute to their personal well-being.

While the studies above have focused only on predicting well-being, others have studied how certain daily behaviors affect well-being. For example, in an 8000-person study, Stubbe et al. showed that exercise participation is associated with higher levels of well-being [24]. Also, Wang et al. demonstrated that students who slept less were more likely to be depressed, and a lower perceived stress score (PSS) was correlated with higher conversation frequency during the day [26]. These studies investigate the stand-alone effects of each behavior on well-being. But these behaviors typically work together and do not affect well-being in isolation, with different combinations possibly resulting in different health outcomes. To that end, understanding how these factors work together could help provide insights into the groups of behaviors to target when looking to improve an individual's mental well-being.

Others have used machine learning models to summarize human behavior. For example, using a correlation method, Phithakkitnukoon et al. identified daily human activity patterns of eating, shopping, entertainment, and recreation from location estimates extracted from calls, messages, and Internet connections of one million users over a period of few months [15]. In addition, topic models have been used to discover patterns underlying human behavior specifically as it relates to activity recognition. For example, Farrahi et al. applied unsupervised topic models to large scale location data from smartphones to discover daily human routine activities [5]. Phung et al. used LDA to infer the sequences of places a user routinely visits [16]. LDA based models have been used to discover daily routine patterns with data from wearable sensors [8, 23]. Also, Ferrari et al. used topic models to obtain mobility behavior using location data collected from Google Latitude [6]. These works demonstrate the potential of using probabilistic models to summarize human behavior, but they only focus on mining activity patterns and do not have a direct application of improving well-being.

The current approaches to modeling human behavior either summarize routine behavior, model physical activity, or learn associations between single health behaviors and well-being. Our work differs from these approaches because we focus on learning patterns of modifiable health behaviors that are predictive of well-being as opposed to only discovering structure in activity sensor data or finding topics that summarize various activities in different settings. Our approach is important for a number of reasons. First, using *modifiable* behaviors, those behaviors that an individual has control over, is significant because the insights discovered can provide individuals with an opportunity to change their behavior and consequently improve their well-being – hopefully keeping them from needing mental health care, even as they repeatedly encounter stressful circumstances. Second, in learning patterns of health behaviors, we present a Bayesian model that creates meaningful and interpretable representations of these health behaviors.

---

[4]American College Health Assessment Spring 2013 and 2017 reports: http://www.acha-ncha.org/pubs_rpts.html. Accessed Jan. 10, 2019.

We demonstrate that these representations are useful for predicting important affective states such as stress, providing actionable insights on how individuals can improve their well-being.

3 DATA

The dataset we use in this work is from an ambulatory study by MIT and Brigham and Women's hospital [20, 22]. The SNAPSHOT study continuously monitored students for 30 days over the course of six semesters, Fall 2013 to Spring 2016, and in Spring 2017, students were monitored all semester long. We use a portion of the 30-day study where different sets of about 50 students participated each semester. So far over 250 students have participated in this study. The study gathers rich daily behavioral, well-being, smartphone, and physiological data from the students. Several modalities contribute measurements to this dataset, they include pre- and post-study questionnaires, wearable sensors (Actiwatch, Q-sensor) to collect activities and physiological data, mobile phones to log location, SMS, call and screen activities, and daily weather information. We used the following modalities from this dataset, ignoring the modalities such as physiology that are not easily modifiable by participants:

Actiwatch: This sensor, sampled at 8Hz, measures activity and sleep-wake schedule–sleep time, wake time and sleep duration, during the course of the day.

Mobile Phone: Participants installed an android application on their phones that logs the timestamps for all outgoing, incoming and missed calls as well as outgoing and incoming SMS and emails. The app also logs whenever the screen is turned on or off and the GPS location of the student.

Behavioral Surveys: Everyday during the study, participants fill out surveys in the morning and evening. Information is collected about academic, extracurricular, and exercise activities, sleep, caffeine intake, and social interaction. In addition, twice daily, the students were asked to rate on a visual analog scale of 0–100 how stressed or calm, sick or healthy, sad or happy they felt. The values from these rating scales constitute the self-report well-being labels.

4 METHODS

In this section, we discuss the modifiable behaviors that were selected from the data, and then present how to map the data to feature representations suitable for the modeling task. Next, we describe the supervised modeling framework for uncovering the latent patterns of modifiable human behaviors.

4.1 Selection of Modifiable Behaviors:

We define modifiable behaviors as actions that can be controlled by the study participants, such as sleep regularity, phone usage, etc. This information was extracted from smartphone logs, wearable sensors, GPS coordinates and behavioral surveys obtained from the participants during the 30-day study. We believe that these groups of behaviors are accessible and easy enough to be extracted from other population groups.

Varying degrees of a particular behavior can lead to different outcomes. For example, sleeping eight hours at night might lead to a participant reporting calmness, while sleep deprivation might lead to a stress report from the same participant; these are two

Table 1: Selected modifiable behaviors, the bins and total number of days the behaviors were present in the data.

Modality	Behaviors	Bins	Total num. of days
Call	Num. total outgoing calls	0, 1, 2, 3–4, ≥ 5	3911
	Num. unique outgoing calls	0, 1, 2, ≥ 3	3911
	Num. unique missed calls	3–4, ≥ 5	528
	Duration incoming calls (all day) (mins.)	0, 0–2, 2–4, 4–6, 6–12, ≥ 12	3911
	Duration outgoing calls (all day) (mins.)	0, 0–2, 2–4, 4–6, 6–12, ≥ 12	3911
Screen	Num. on/off events (5PM–Midnight)	0, 0–25, 25–50, 50–75, 75–100, 100–150, ≥ 150	5040
	Duration on/off events (all day) (hours)	0, 0–2, 2–3, 3–4, ≥ 4	5064
	Duration on/off events (Midnight–3AM) (hours)	0, 0–0.5, ≥ 0.5	5057
	Duration on/off events (5PM–Midnight) (hours)	0, 0–0.5, 0.5–1, 1–2, ≥ 2	5040
SMS	Num. unique outgoing SMS (all day)	0, 1, 2, 3–4, 5–10, ≥ 10	5071
	Num. unique outgoing SMS (5PM–Midnight)	0, 1, 2–5, ≥ 5	5071
Actiwatch sleep	Duration today's sleep (hours)	1–4, 4–6, 6–7, 7–8, 8–10, ≥ 10	4371
	Duration yesterday's sleep (hours)	1–4, 4–6, 6–7, 7–8, 8–10, ≥ 10	3999
	Duration day before yesterday's sleep (hours)	1–4, 4–6, 6–7, 7–8, 8–10, ≥ 10	3841
	Bedtime (hour of the day)	12–1AM, 1–2AM, 2–3AM, 3–4AM, 4–6AM, 6–10AM, 10AM–8PM, 8–11PM	4277
	Bedtime deviation from participant's mean (hours)	≤ -2, -1, 0, 1, ≥ 2	4371
	Weekly sleep regularity[5]	0–0.4, 0.4–0.5, 0.5–0.6, 0.6–0.7, 0.7–0.8, 0.8–0.9, 0.9–1	4363
Location	Time on campus (hours)	0, 0–1, 1–8, ≥ 8	3716
	Time indoors (hours)	0, 0–2, 3–5, 6–7, 8, 9–10, ≥ 10	3716
	Time outdoors (hours)	0, 0–1, 1–8, ≥ 8	3716
Survey Activities	Study duration (hours)	0, 0–2, 2–4, 4–6, 6–8, ≥ 8	5397
	Exercise duration (hours)	0, 0–1, 1–2, ≥ 2	5397
	Extracurricular duration (hours)	0, 0–2, 2–4, ≥ 4	5397
	Academic duration (hours)	0, 0–2, 2–3, 3–4, 4–6, ≥ 6	5397
	Naps (0/1)		1084
	All-nighter (0/1)		141
	Caffeine consumption (0/1)		2029
	Alcohol consumption (0/1)		666
Survey interaction	Negative social interaction (0/1)		555
	Positive social interaction (0/1)		1439
	Pre-sleep in-person interaction (0/1)		1961
	Pre-sleep media interaction (0/1)		1961

different behaviors and need to be treated as such. As a result, most of the behaviors were binned based on duration, time of day, or number of occurrences. Binning helps to provide an accurate summary of a participant's daily activities by allowing us to be as specific as possible when describing each day of data. This leads to a sparse representation of the data that is preferred by LDA. Also, binning allows us to capture some information about the associated numerical value of any behavior– time of day, duration, or number of occurrences. We select the number of bins for each feature by carefully examining its distribution and looking for natural breaking points. After binning, there were 134 behaviors present. See Table 1 for the list of behaviors selected and their various bins.

4.2 Feature Representation and Labels

4.2.1 "Bag-of-Behaviors" Feature Representation. After selecting the modifiable behaviors, we convert the data into input features for the model. Following the bag-of-words representation commonly used to model documents, we represent the binned modifiable behaviors in each participant's day with a binary bag-of-behaviors vector $x \in \{0, 1\}^{|V|}$, where V is the set of modifiable behaviors used in the model, 0 means that the modifiable behavior is absent in that day and 1 means that the modifiable behavior is present. The goal

---

[5]Sleep regularity (SR) measures the likelihood of an individual being in the same sleep or wake state at any two time points 24 hours apart with a 1 minute resolution, averaged across one week in the study. Similar to Phillips et al. [14], we calculate sleep regularity as $\frac{2}{T-\tau}\left(\int_0^{T-\tau} s(t)s(t+\tau)\,dt\right) - 1$, where $s(t) = 1$ during wake and $s(t) = -1$ during sleep, $\tau = 24$, and $T \leq 168$ is the total hours of sleep data per week. We note that our SR values are modified to be between the values of 0 and 1, with the most regular sleepers having sleep regularity scores close to 1.
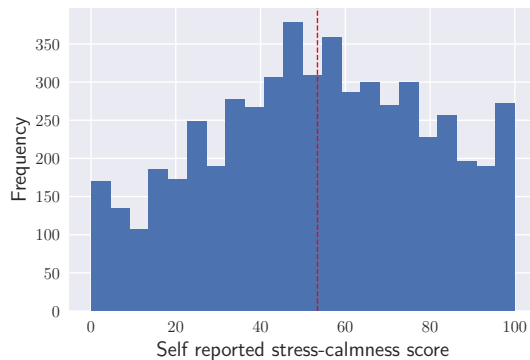
Figure 1: Frequency of self-reported stress-calmness response variables. The mean of the response variables, 53.43, is shown in the vertical *red* line.

of this work is to learn how combinations of behaviors affect self-reported well-being. So, each individual's day is represented by the group of specific behaviors present. This representation ignores the ordering of the behaviors but the binning allows us retain some of that information. We throw out days with fewer than 15 behaviors present because such days have insufficient data to train the model. We end up with a total of 5397 days of data and 115,659 observations from 224 unique participants across 6 semesters.

4.2.2 Response Variables. The participants self-reported their evening perceived stress on a 0–100 scale, with 0 indicating the highest level of stress and 100 the highest level of calmness. Figure 1 shows the frequency of these response variables in the dataset. In psychology, these self-reports are currently the gold standard for assessing well-being; however, in this work, we learn associations between these self-reports and combinations of behaviors that can be measured objectively. We thus provide evidence-based insights about behaviors that an individual might consider to modify if they wish to improve their well-being.

As a pre-processing step, we log-transformed the labels to achieve approximate normality as required by the model (see Section 4.3). These real-valued log-transformed stress labels, $y \in \mathbb{R}$, are the response variables in the supervised model. Our modeling goal is to find latent patterns that are best predictive of these real-valued labels.

We also trained the model using other kinds of response variable transformations: divided each response by 100 (scaled), centered the scaled labels at 0 by subtracting the mean (scaled-shifted), z-transformed the labels (normalized), log-transformed the scaled labels (scaled-log), and centered the log-transform at zero by subtracting the mean (log-shifted). However, none of these other response transformations performed as well as the log-transform in the prediction task. We omit their results for brevity.

4.3 Supervised Latent Dirichlet Allocation

LDA was developed to learn the probability distribution of words in collections of documents in a corpus in an unsupervised manner [3]. However, the goal of this work is to use LDA to learn patterns that maximize the likelihood (or posterior probability) of behaviors

present in the study participants' days. In addition, while unsupervised models are sufficient to summarize words in a document, they are not ideal for real-world problems such as modeling health behaviors that may influence well-being. Health behaviors are dynamic and vary over the time of day and across people. If we model health behaviors using an unsupervised method, we will likely discover broad patterns that run throughout the days. Supervision helps us identify the patterns that are consistently and specifically predictive of an outcome such as well-being.

In this work, we formulate our solution by using a supervised latent Dirichlet allocation (sLDA) model, proposed by Blei and McAuliffe [2] originally for text analysis. Each day (or feature vector) is made up of observed behaviors and is associated with a response variable (the log-transformed self-reported stress label) external to the behaviors present. The aim is to uncover the latent *patterns*, probability distributions over the set of observed health behaviors, that are best predictive of the response variable. SLDA accomplishes this goal by jointly modeling the learning of these patterns and the coefficients of the regression of the learned patterns on the response labels. By doing this, we effectively tie the behaviors expressed on a day (and the night before) to the stress response reported at the end of that day. This modeling choice is reasonable because it constrains the latent patterns to explain both the behaviors and the response variable, and ensures that the learned patterns are best predictive of the stress response.

In this approach, we assume that the same set of latent patterns generate the behaviors we observe in all the days present in the dataset, but these patterns have different proportions on different days. The different proportions give rise to different combinations of behaviors each day, and thus to different self-reported stress responses.

The goal of inference is to 1) learn the latent patterns, and 2) estimate the pattern proportions expressed each day. We use variational Expectation-Maximization to iteratively approximate these two inference tasks. The update equations used are outlined in Blei and McAuliffe [2]. Once we have the fitted model, we can predict any new participant's self-reported well-being and also learn the new day's latent pattern proportions. To aid reproducibility, we provide all of the code used in our analysis so that others may easily compare it on different datasets.[6] Full details about the methods and experimental setup are available in [11].

5 EXPERIMENT

As outlined in the previous section, sLDA jointly models the behaviors and corresponding stress response, resulting in latent patterns that are best predictive of the stress response. To test its performance, we compared sLDA to two baselines:

(1) We ran a linear regression analysis on the latent patterns learned from unsupervised LDA. This is the same as using LDA topics as prediction features. We call this the "LDA + regression" approach.
(2) We performed an L1-regularized least-squares regression (or LASSO) analysis using each day's empirical distribution over behaviors as the model covariates (i.e., normalizing the bag-of-behavior feature representations to sum to one). The

---

[6]Code available at: https://github.mit.edu/ehinosa/assessing-well-being

L_1-regularized regression is widely used in high-dimensional problems because of its feature selection property. When you have a group of correlated features, LASSO retains only one feature and sets the others in the group to zero. While this enhances the interpretability of the model, the coefficients cannot be interpreted as causal coefficients but rather as associations. Still, it provides insights into participants' behaviors. Thus, in addition to comparing its performance results with the sLDA, we are interested in examining the features selected by LASSO for the regression task.

For the sLDA analysis, we initialized the K topics to randomly perturbed uniform topics. The hyper-parameters of the response variable (δ and η) were initialized to the variance of the log-transformed stress response, and to a K-dimensional vector of zeros, respectively. Also, we fixed the pattern proportions hyperparameter, α, to $1/K$. We ran the expectation step until the relative change was below 0.001. We ran the maximization step until the relative change in the overall likelihood bound fell below 0.0001. For the LDA model, we initialized topics similarly and used the same convergence criteria and value for α.

For evaluation, we applied 5-fold cross-validation to daily feature vectors from all individuals in the dataset, and quantified the model's performance by computing binary prediction accuracies and F1 scores on the held-out-fold predictions. To convert to binary labels, true and predicted *real-valued* response labels equal to or greater than 50 were considered high stress values and values below 50 were considered low stress values. We also computed binary prediction accuracies and F1 scores using the mean of the population's self-reported stress (mean = 53.43) as the threshold. This measured how well the learned patterns predicted high versus low self-reported stress, and the F1 score combined the model's precision and recall into one metric. We report the same performance metric on the held-out-fold predictions for the LDA + Regression approach. We assessed the prediction quality of both models over different numbers of topics, 2–20.

We used the scikit-learn LASSO implementation in our analysis. We used 5-fold cross validation to select the optimal regularization parameter. We searched for this parameter in the range of 10^{-4} to 10^0 in 10 steps. We report the highest mean (± standard deviation) binary prediction accuracy and F1 score achieved on the held-out fold. We compare these values to the highest value sLDA achieved across the different numbers of topics.

6 RESULTS

The average binary prediction accuracies and F1 scores (over five runs) for both thresholds, 50 and the mean stress response (53.43), are shown in Figures 2, 3, 4 and 5, respectively. Remember, when reviewing these results, that our goal is *not* to build the best model for a general stress-recognition problem, and thus we do not compare different machine learning models. Instead, the challenge is to identify sets of *modifiable* behaviors (patterns) that are significant in helping predict stress, and evaluate if they are better than a control set (the unsupervised patterns found by LDA).

The sLDA model performed better at predicting high versus low stress when the threshold is 50 compared to when the threshold is the mean stress response of the participants. Our results

indicate that across multiple numbers of topics, the sLDA model outperformed the LDA + Regression model on all four performance metrics. It is important to note that the baseline unsupervised LDA model (LDA + Regression) is performing even worse than a random predictor. This implies that sLDA learns latent patterns that are better at predicting the self-reported stress compared to the patterns learned in an unsupervised way.
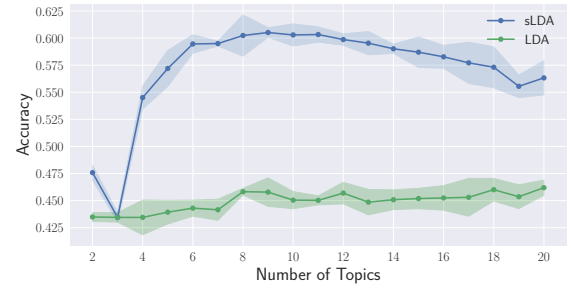

Figure 2: Average binary prediction accuracy for sLDA and LDA + Regression models across different topics. Shaded region indicates standard deviation. Threshold = 50.
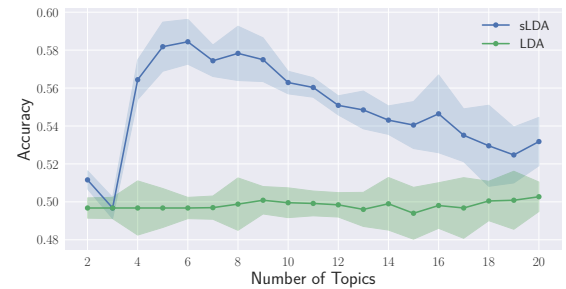

Figure 3: Average binary prediction accuracy for sLDA and LDA + Regression models across different topics. Shaded region indicates standard deviation. Threshold = 53.43.

Table 2 shows the best mean (± standard deviation) of the performance metrics LASSO achieved, and the highest value sLDA achieved on the held-out fold across the different topics and over five repetitions. For the binary prediction accuracy metric, sLDA had improvements over LASSO of 7% and 2.5%, when the thresholds were set at 50 and the mean stress response respectively. For the F1 score metric, sLDA had improvements over LASSO of 50% and 61%, when the thresholds were set at 50 and the mean stress response respectively. A Welch's t-test revealed that sLDA had a statistically significant improvement over LASSO's prediction accuracy when the binary data split threshold was set at 50. The sLDA also had statistically significant improvements over LASSO's F1 scores for both binary split thresholds. We also note the advantage that sLDA models the latent structure of the data, making its representation useful for other purposes beyond prediction.
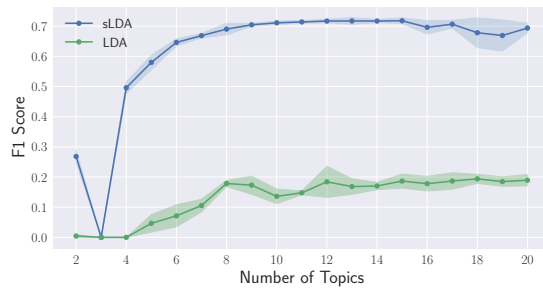
Figure 4: Average F1 score for sLDA and LDA + Regression models across different topics. Shaded region indicates standard deviation. Threshold = 50.
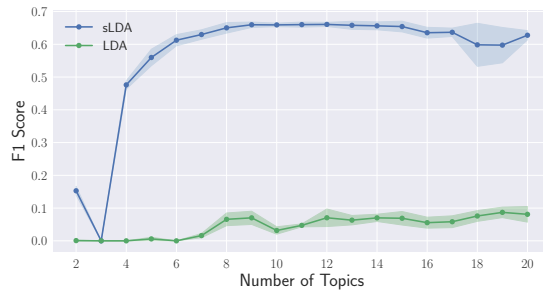

Figure 5: Average F1 score for sLDA and LDA + Regression models across different topics. Shaded region indicates standard deviation. Threshold = 53.43.

Table 2: Mean (± std) of binary accuracies and F1 scores for sLDA and LASSO. Bold entries represent a statistically significant improvement of the sLDA model over LASSO ($p < 0.05$).

	Performance Metrics			
	Binary Accuracy	Binary Accuracy	F1 Score	F1 Score
Model	(thresh. = 50)	(thresh. = mean)	(thresh. = 50)	(thresh. = mean)
sLDA	**60.5% (± 0.4)**	58.4% (± 1.2)	**0.72 (± 0.01)**	**0.66 (± 0.01)**
LASSO	56.5% (± 1.0)	57.0% (± 1.2)	0.48 (± 0.02)	0.41 (± 0.03)

6.1 LASSO

We are interested in examining the behaviors that were selected as best predictors of self-reported stress in this cohort. Figure 6 shows the non-zero model coefficients for LASSO, and Table 3 shows the behaviors that were selected in order of decreasing significance. In the study, the self-reported calmness/stress labels were on a scale of 0–100, with 100 indicating the highest level of calmness and 0 indicating the highest level stress. The stress-calm levels were forecast for each night's labels using data from before that night. This means that the model selected features that were best predictive of increasing calmness (or decreasing stress), and decreasing calmness (or increasing stress). We discuss some of the behaviors below.
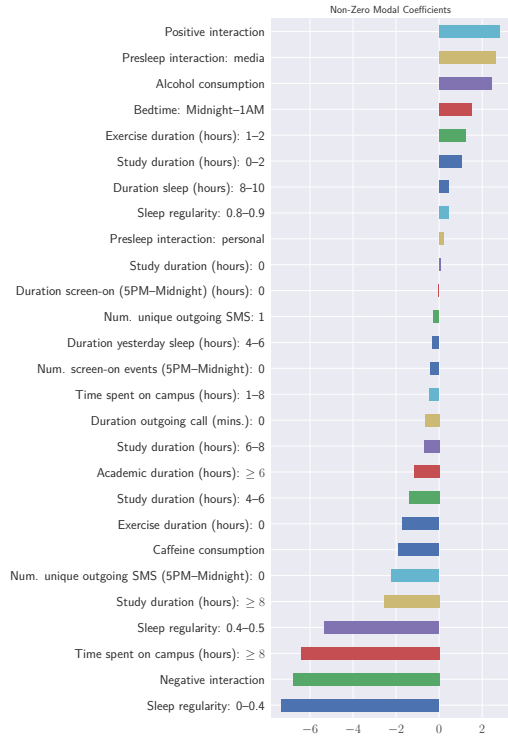

Figure 6: LASSO Non-zero coefficient estimates for self-reported stress. In the data, 100 represents the highest level of calmness and 0 represents the highest level of stress. So, the positive coefficients are positively correlated with increasing calmness.

Table 3: Behaviors selected by LASSO in decreasing order of importance.

	Best predictors of calm	Best predictors of stress
1	Positive interaction	Low sleep regularity (0–0.4)
2	Pre-sleep interaction: media	Negative interaction
3	Alcohol consumption	Spending more than 8 hours on campus
4	Bedtime around midnight and 1AM	Low sleep regularity (0.4–0.5)
5	Exercising 1–2 hours a day	No social interaction via SMS after 5PM
6	Studying 0–2 hours a day	Caffeine consumption
7	Sleep duration between 8–10 hours	No exercise
8	High sleep regularity (0.8–0.9)	Studying 4–6 hours a day
9	Pre-sleep interaction: personal	Not initiating any phone calls
10	Not studying during the day	Sleep duration the day before: 4–6 hours

Social interactions: Different types of social interactions had different effects on stress. According to the model, positive interactions during the day and personal interactions just before sleep were positively correlated with calmness reported later that evening. In contrast, social interactions reported as "negative" were best predictive of stress. Having few interactions through electronic devices such as not initiating any calls during the day or not sending any text messages after 5PM were also high predictors of stress. This is consistent with research that showed that students with low mental health, as measured by metal health composite scores (MCS), had

a lower probability of interacting with electronic media prior to bedtime [22]. Also, the StudentLife study by Wang et al. found that students who had frequent conversations in the evening were less likely to be stressed ($r = -0.386$, $p = 0.015$) [26]. In addition to in-person interactions, we see that interaction with media (which could be entertainment) is a predictor of calmness.

Bedtime, sleep duration and sleep regularity: Sleep features had high importance in stress prediction. First, going to bed between Midnight and 1AM (slightly earlier than the study participants' average of 2AM), and having about 8–10 hours of sleep was associated with a calm day. This finding aligns with results by Wang et al. that showed a strong negative association between sleep duration and perceived stress ($r = -0.355$, $p = 0.024$); students getting more sleep experienced less stress [26]. Second, sleep regularity is a predictor of stress and calm even when controlling for sleep duration [21]. In this model, high sleep regularity (0.8–0.9) was highly predictive of calmness and low sleep regularity (0–0.5) was highly predictive of stress. This is consistent with a previous study that found that sleep regularity was associated with better well-being in college students [21]. We were also interested in seeing the effects of the previous two days sleep duration on today's outcome, so we included the sleep duration of those days as behaviors in our model. From this, we see that having about 4–6 hours of sleep the previous day was predictive of today's perceived stress.

Academic activity: The model selected low study duration (0–2 hours) as a high predictor of calmness, and spending more that 8 hours on campus as a high predictor of stress. We do not know if the students were studying while on campus, but given the amount of time spent, it is highly likely they were studying.[7]

Physical activity: Having one to two hours of exercise a day was found to be a high predictor of calmness within the study participants. This is in line with several research studies that have shown that physical activity improves well-being [7, 12, 13, 27]. We also found that having no exercise was a high predictor of stress. While these associations do not show causality, it may be that the correlations we found would potentially map to "causation" in a trial designed to assess causality.

Caffeine and alcohol consumption: Caffeine and alcohol consumption had opposite effects on the prediction of calmness. While caffeine consumption was found to be a predictor of high stress, alcohol was a predictor of calmness. We note that the students typically fill out the surveys after 8PM and we do not have information on the amount of alcohol consumed by the students. Therefore, there are two possible explanations for the reason alcohol shows up as a predictor of calm. First, we may be seeing the calming effect of alcohol if the evening surveys were filled soon after taking alcohol. Secondly, in the dataset, those who reported alcohol consumption, had a 65% percent chance of doing so on the weekend (Friday to Sunday), and the weekend has been shown to be associated with better psychological well-being [19]. So we may be seeing the weekend effect mixed in with the alcohol effect.

In summary, we have seen that given the normalized bag-of-behaviors feature representation, LASSO was able to select intuitively correct features to be predictors of calmness and stress. It is

important to note that the relationships discovered are associative. As a result, further studies will be required to understand causal relationships.

6.2 SLDA

The sLDA model learned latent patterns that are meaningful representations of health behavior. Here, we further analyze these patterns.

Since we are interested in good predictive patterns, we select $K = 11$ for our discussion because it provides us with the highest binary prediction accuracy (when the threshold is set to 50). Figure 7 shows the 11-pattern sLDA model fit to our data. For each learned pattern, we highlight the seven health behaviors with the highest probability of occurrence. The pattern with the most negative coefficient pattern 1 (*bottom*), comprises highly probable behaviors such as no exercise, caffeine consumption, very little social interaction, study duration of 4–6 hours, and negative social interaction. The pattern with the most positive coefficient pattern 11 (*top*), consists of highly probable behaviors such as high sleep regularity, going to bed around the average bedtime of the study population, spending time outdoors, sufficient sleep the night before, and very little phone usage between the hours of midnight and 3AM (likely because the participant is asleep). We also see sleep regularity increasing as the patterns become more positively associated with calmness, from SR = 0.6–0.7 in pattern 7 to SR = 0.8–0.9 in pattern 11. The patterns in-between have a mix of health behaviors, which provide evidence that the "positive" behaviors start to compensate for the others as the patterns become more positively correlated with calmness. For example, in pattern 9, we see a high sleep regularity of 0.7–0.8, but late bedtime, 3–4AM, and low sleep duration for two consecutive days, 4–6 hours. It may be that the high sleep regularity helps build resilience, enabling an individual to weather a couple days of late bedtime and short sleep without the expected increase in stress.

These findings are consistent with our belief that complex combinations of health behaviors lead to different outcomes, and models that only learn prediction rules or that focus on only analysis of the impact of single behaviors are insufficient to uncover these complex latent patterns that drive real-world data.

One advantage sLDA has over LASSO is that it learns the day-pattern probability distributions over each participant's day. This gives us insight into how the patterns expressed in each person's day contribute to the self-reported perceived stress. Also, this personalized distribution over patterns illustrates how the sLDA model is able to capture the similarity in the behaviors of participants with comparable self-reported stress. Figure 8 shows the pattern proportions learned in 50 different participants randomly chosen from low, mid, and high self-reported stress levels. Participants who reported high calmness (or low stress) have the most calm pattern (pattern 11) dominantly expressed; participants with high self-reported stress have the most stress pattern (pattern 1) dominantly expressed; while participants in the mid-range of 55–65 self-reported stress have varying representations of patterns expressed in each day.

7 CONCLUSION AND DISCUSSION

We proposed a novel framework to map multi-modal data collected in the "wild" to meaningful representations of health behavior,

---

[7]A possible confounder is that some students in the study lived in a couple of dorms that were part of the geographical boundaries we defined as being "on campus"
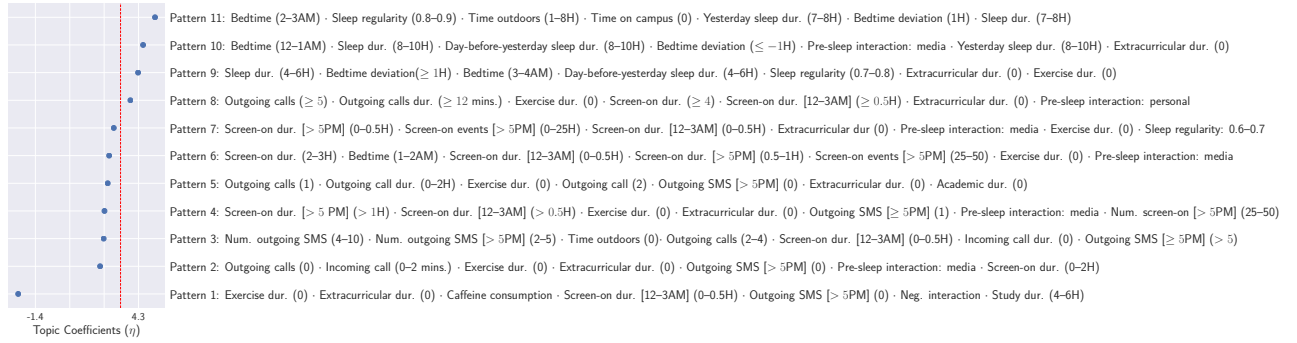
Figure 7: An 11-pattern sLDA model fit to the SNAPSHOT data (red vertical line represents mean of the positive coefficients). Topic coefficient decrease from *top* to *bottom*. In each topic, the behaviors are ordered from *left to right* starting from the most probable behavior.
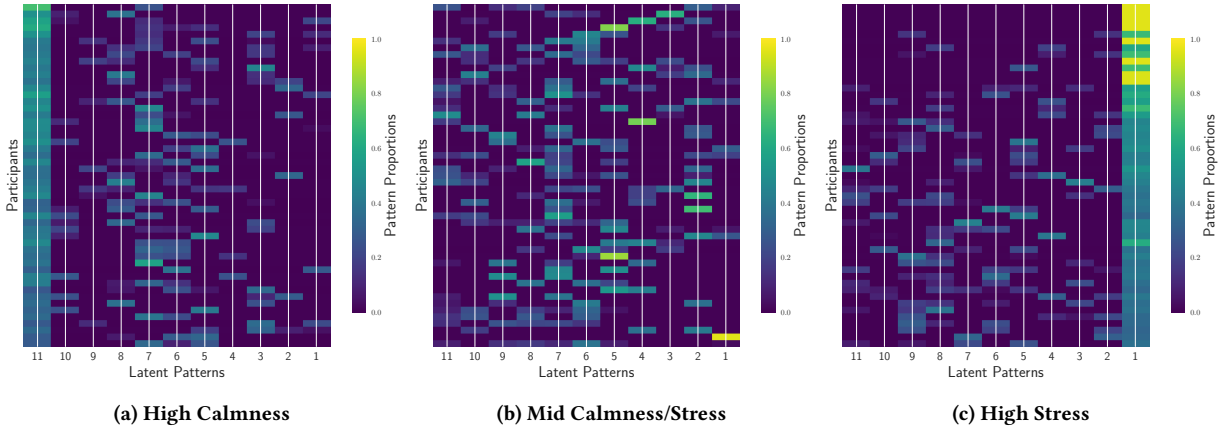

(a) High Calmness

(b) Mid Calmness/Stress

(c) High Stress

Figure 8: Learned day-pattern proportions (θ) in three different levels of the response label (y): (a) 50 participant-days with $y \geq 80$; (b) 50 participant-days with $55 \leq y < 65$; (c) 50 participant-days with $y \leq 40$. Pattern 11 is the most *calm* pattern, and pattern 1 is the most *stress* pattern.

which are then used with sLDA to learn patterns of modifiable health behaviors predictive of stress, a major factor in well-being. We compared the predictive performance of the patterns learned with sLDA to the patterns learned with an unsupervised LDA, and showed that the sLDA significantly outperforms LDA in binary stress prediction. We additionally compared the performance of sLDA to LASSO and found that the sLDA outperforms LASSO in binary stress prediction.

We analyzed the latent structure uncovered by the sLDA model and showed how specific patterns of behavior are associated with stress prediction. Previous work has shown correlations with single classes of behaviors such as modifying sleep or social interaction [21, 26], but we have shown how combinations of specific behaviors such as going to bed 2 hours earlier or later than usual, getting 7–8 vs. 4–6 hours of sleep, and having positive vs. negative social interactions, etc., are associated with student stress levels. Through this work, we have shown that the sLDA model finds informative structure, which produces patterns that are predictive of stress, and connects with an intuitive understanding of health behaviors.

This work has limitations that can be improved in future work. First, the sLDA model did not consider the inherent temporal effects of behaviors on well-being. Although each day is modeled independently, we included information about the previous two days of sleep to test the hypothesis that sleep features of past days affect today's well-being. We found two instances supporting this hypothesis: first, in the LASSO model, sleeping 4–6 hours the previous day was predictive of high stress; secondly, in the sLDA model, the most *calm* pattern had day-before-yesterday sleep duration of 7–8 hours as a highly probably behavior, and the next calm predictive pattern found going to bed one or more than two hours earlier than your personal average bedtime (bedtime deviation −1 or ≤ −2, respectively) as a highly probable behavior. With the collection of more data, we can use a supervised variant of dynamic topic models [1] to learn how time-varying behaviors affect an individual's well-being. Second, our work used self-reported stress labels as a proxy for stress measurement. However, these labels can be inconsistent across and within users. Our team has an ongoing effort to produce

objective labels measuring physiological stress from SNAPSHOT data, work that is outside the scope of this paper.

A third limitation is the model predictive accuracy. Using data from the SNAPSHOT study, other models have been able to provide more accurate prediction. In Taylor et al. for example, the Bayesian multi-task learning model and personalized neural network framework achieved higher prediction accuracies [25]. One reason for their higher prediction performance is that other features in addition to modifiable behaviors were used to learn the prediction rule; for example, they used physiology. This gave the models more information to distinguish between low or high stress days. In contrast, this paper restricted features to modifiable behaviors, like bedtime and screen time, yielding variables that individuals can change behaviorally for future causality studies. Nonetheless, our results showed that modifiable behaviors have a very significant predictive influence on a participant's stress level. Also, the Taylor et al. problem was made easier than ours by discarding data corresponding to the middle 20% of the self-reported stress labels, throwing out ambiguous "middle" values. For the work in this paper, we did not throw out the middle 20% because sLDA uses a generalized linear regression framework to model the real-values from 0–100 that were originally reported. This is valuable because we can directly predict a finer estimate of self-reported stress rather than just a binary category. For example, individuals might be interested in seeing the trends in their self-reported stress and whether or not it is gradually or suddenly increasing. Algorithms that only predict oscillating values between high and low stress will not be as useful. We converted the real-valued predictions to binary labels only for the purposes of simplifying model comparisons, and showed the gains were robust using two different thresholds.

In conclusion, the new modeling framework we present can uncover latent patterns of behaviors that are most predictive of a participant's well-being as measured by stress. Although we use data from a student population, we believe that given any set of modifiable behaviors, our framework can be applied to other population groups.

ACKNOWLEDGMENTS

REFERENCES

[1] David M Blei and John D Lafferty. 2006. Dynamic Topic Models. In *Proceedings of the 23rd International Conference on Machine Learning*. ACM, 113–120.

[2] David M Blei and Jon D Mcauliffe. 2010. Supervised Topic Models. https://arxiv.org/pdf/1003.0783.pdf

[3] David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent Dirichlet Allocation. *Journal of Machine Learning Research* 3 (2003), 993–1022.

[4] Ed Diener and Martin Seligman. 2017. Beyond Money: Progress on an Economy of Well-Being. *Psychological Science in the Public Interest* 5, 1 (2017), 1–31.

[5] Katayoun Farrahi and Daniel Gatica-Perez. 2014. A Probabilistic Approach to Mining Mobile Phone Data Sequences. *Personal and Ubiquitous Computing* 18, 1 (2014), 223–238.

[6] Laura Ferrari and Marco Mamei. 2011. Discovering Daily Routines from Google Latitude with Topic Models. In *Proceedings of the IEEE International Conference on Pervasive Computing and Communications Workshop*. 432–437.

[7] Suzanne Hudd, Jennifer Dumlao, Diane Erdmann-Sager, Daniel Murray, Emily Phan, Nicholas Soukas, and Nori Yokozuka. 2000. Stress at College: Effects on Health Habits, Health Status and Self-Esteem. *College Student J.* 34, 2 (2000).

[8] Tâm Huynh, Mario Fritz, and Bernt Schiele. 2008. Discovery of Activity Patterns Using Topic Models. In *Proceedings of the 10th ACM International Conference on Ubiquitous Computing*. 10–19.

[9] Imran Shuja Khawaja, Joseph J Westermeyer, Prashant Gajwani, and Robert E Feinstein. 2009. Depression and coronary artery disease: the association, mechanisms, and therapeutic implications. *Psychiatry (Edgmont)* 6, 1 (2009), 38.

[10] Jia Li, Yu Rong, Helen Meng, Zhihui Lu, Timothy Kwok, and Hong Cheng. 2018. TATC: Predicting Alzheimer's Disease with Actigraphy Data. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM, 509–518.

[11] Ehimwenma Nosakhare. 2018. *Probabilistic Latent Variable Modeling for Predicting Future Well-Being and Assessing Behavioral Influences on Mood, Stress and Health*. Ph.D. Dissertation. MIT.

[12] Gregory A Panza, Beth A Taylor, Paul D Thompson, C Michael White, and Linda S Pescatello. 2017. Physical Activity Intensity and Subjective Well-Being in Healthy Adults. *Journal of Health Psychology* (feb 2017).

[13] Tim Pawlowski, Paul Downward, and Simona Rasciute. 2011. Subjective Well-Being in European Countries—on the Age-Specific Impact of Physical Activity. *European Review of Aging and Physical Activity* 8, 2 (oct 2011), 93–102.

[14] Andrew J Phillips, William M Clerx, Conor S O'Brien, Akane Sano, Laura K Barger, Rosalind W Picard, Steven W Lockley, Elizabeth B Klerman, and Charles A Czeisler. 2017. Irregular Sleep/Wake Patterns are Associated with Poorer Academic Performance and Delayed Circadian and Sleep/Wake Timing. *Scientific Reports* 7, 1 (2017), 3216.

[15] Santi Phithakkitnukoon, Teerayut Horanont, Giusy Di Lorenzo, Ryosuke Shibasaki, and Carlo Ratti. 2010. Activity-Aware Map: Identifying Human Daily Activity Pattern Using Mobile Phone Data. In *International Workshop on Human Behavior Understanding*. Springer, 14–25.

[16] Dinh Phung, Brett Adams, Kha Tran, Svetha Venkatesh, and Mohan Kumar. 2009. High Accuracy Context Recovery Using Clustering Mechanisms. In *Proc. of the IEEE International Conference on Pervasive Computing and Communications*. 1–9.

[17] Frans Pouwer, Nina Kupper, and Marcel C Adriaanse. 2010. Does emotional stress cause type 2 diabetes mellitus? A review from the European Depression in Diabetes (EDID) Research Consortium. *Discovery medicine* 9, 45 (2010), 112–118.

[18] Tom Quisel, Luca Foschini, Alessio Signorini, and David C Kale. 2017. Collecting and analyzing millions of mhealth data streams. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 1971–1980.

[19] Richard M Ryan, Jessey H Bernstein, and Kirk Warren Brown. 2010. Weekends, Work, and Well-Being: Psychological Need Satisfactions and Day of the Week Effects on Mood, Vitality, and Physical Symptoms. *Journal of Social and Clinical Psychology* 29, 1 (2010), 95–122.

[20] Akane Sano. 2015. *Measuring College Students' Sleep, Stress and Mental Health with Wearable Sensors and Mobile Phones*. Ph.D. Dissertation. MIT.

[21] Akane Sano, Andrew J Phillips, Andrew W McHill, Sara Taylor, Laura K Barger, Charles A Czeisler, and Rosalind W Picard. 2017. Influence of Weekly Sleep Regularity on Self-Reported Wellbeing. *Sleep* 40, suppl_1 (2017), A67–A68.

[22] Akane Sano, Sara Taylor, Andrew W McHill, Andrew J Phillips, Laura K Barger, Elizabeth Klerman, and Rosalind Picard. 2018. Identifying Objective Physiological Markers and Modifiable Behaviors for Self-Reported Stress and Mental Health Status Using Wearable Sensors and Mobile Phones. *Journal of Medical Internet Research* 20, 6 (2018).

[23] Julia Seiter, Oliver Amft, Mirco Rossi, and Gerhard Tröster. 2014. Discovery of Activity Composites Using Topic Models: An Analysis of Unsupervised Methods. *Pervasive and Mobile Computing* 15 (2014), 215–227.

[24] Janine H Stubbe, Marleen H M de Moor, Dorret I Boomsma, and Eco J C de Geus. 2007. The Association Between Exercise Participation and Well-Being: A Co-Twin Study. *Preventive Medicine* 44, 2 (2007), 148–152.

[25] Sara Taylor, Natasha Jaques, Ehimwenma Nosakhare, Akane Sano, and Rosalind Picard. 2017. Personalized Multitask Learning for Predicting Tomorrow's Mood, Stress, and Health. *IEEE Trans. on Affective Computing* 14, 8 (June 2017), 1–1.

[26] Rui Wang, Fanglin Chen, Zhenyu Chen, Tianxing Li, Gabriella Harari, Stefanie Tignor, Xia Zhou, Dror Ben-Zeev, and Andrew T Campbell. 2014. StudentLife: Assessing Mental Health, Academic Performance and Behavioral Trends of College Students Using Smartphones. In *Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing*. 3–14.

[27] Simone Wilson-Salandy and Mary A Nies. 2012. The Effect of Physical Activity on the Stress Management, Interpersonal Relationships, and Alcohol Consumption of College Freshmen. *SAGE Open* 2, 4 (2012).