Probabilistic Latent Variable Modeling for Predicting Future Well-Being and Assessing Behavioral Influences on Mood, Stress and Health

by

Ehimwenma Nosakhare
B.S., Electrical Engineering, Howard University (2011)
S.M., Electrical Engineering and Computer Science, Massachusetts Institute of Technology (2013)

Submitted to the Department of Electrical Engineering and Computer Science in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Electrical Engineering and Computer Science at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
September 2018

© Massachusetts Institute of Technology 2018, All rights reserved.

Signature of Author: __________________________________________
Department of Electrical Engineering and Computer Science
August 31, 2018

Certified by: __________________________________________
Rosalind W. Picard
Professor of Media Arts and Sciences
Thesis Supervisor

Accepted by: __________________________________________
Leslie A. Kolodziejski
Professor of Electrical Engineering and Computer Science
Chair, Department Committee on Graduate Students
Probabilistic Latent Variable Modeling for Predicting Future Well-Being and Assessing Behavioral Influences on Mood, Stress and Health

by

Ehimwenma Nosakhare

Submitted to the Department of Electrical Engineering and Computer Science on August 31, 2018 in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Electrical Engineering and Computer Science

Abstract

In recent years, there has been a shift in the psychological research literature from an emphasis on dysfunction to a focus on well-being and positive mental health. As a result, enhancing well-being in individuals has become a viable approach to improving health, in addition to treating disorders when present. Also, the availability of rich multi-modal datasets and advances in machine learning methods have spurred an increase in research studies assessing well-being objectively. However, most of these studies tend to primarily focus on using data to estimate or detect the current state of well-being as opposed to the prediction of well-being. In addition, these studies investigate how stand-alone health behaviors and not a combination of health behaviors influence well-being. Furthermore, these studies do not provide data-backed insights and recommendations to individuals seeking to improve their well-being.

In this dissertation, we use a real-world dataset from a population of college students and interpretable machine learning methods to (1) predict future mood, stress and health, (2) uncover how combinations of health behaviors work together to influence well-being, and (3) understand how to make evidence-based recommendations to individuals looking to improve their well-being. The use of these methods contributes to the development of objective techniques that can help individuals monitor their well-being. In addition, insights from this study contribute to knowledge advancement on how combinations of daily human behaviors can affect well-being.

Thesis Supervisor: Professor Rosalind W. Picard
Title: Professor of Media Arts and Sciences
To mum and dad -
you both are the giants
on whose shoulders I stand.

And, everlastingly,
to Ayobami.
Acknowledgments

Hast thou not seen how thy desires have been granted in what He ordaineth....
To God be the glory, great things He has done!

I would like to acknowledge several people for their encouragement, support and guidance towards the completion of this thesis.

I thank my thesis supervisor Professor Rosalind Picard for her tremendous support over the past three years. Indeed, it has been my utmost pleasure learning from and working with her. She is extremely patient and kind. This work has benefited from her thoughtful guidance. I am also fortunate to have Professor Peter Szolovits and Professor John Guttag on my thesis committee. I am grateful for their timely and invaluable comments that made this work stronger.

The MIT Media Lab Consortium provided the needed funding for this work, and I could not be more grateful for it. The funding allowed me to focus on research after five semesters of consecutive teaching assistantships. I would like to thank the fellow Affective Computing research group members for making my time memorable. You all are extremely kind and I thoroughly enjoyed spending time with you these past few years. Specifically, I thank Sara, Natasha, and Akane for their collaboration in the first part of this thesis work. Also, many thanks to Javier for his open doors, helpful advice and listening ears.

I also would love to thank other members of my MIT community. Many thanks to my friends in the sidney-pacific small group who constantly prayed for me and kept my spirit up during those “middle-years” of graduate school. Christina Lee Yu, meeting you during preview weekend and being your roommate in graduate school are some of the highlights of my time here. Monica Orta, words are not enough to express my gratitude for your unwavering support, thank you. Janet Fischer, thank you for your help and for always easing me off my graduate school concerns. Fmr. Professor Joel Dawson, thank you for being my very first research mentor. You inspired me to pursue this degree.

To my friends outside of MIT, thank you for always reminding me that my research was not my life’s work, it definitely kept the journey in perspective. Bola and Ezinne, I am grateful for the countless prayers you both offered over the past two years. Monisola,
thank you for pushing me and encouraging me along the way. Bodunrin, Chioma and Tayo, Chiso, Dotun, Georgiana, Karen, Lois, Peace, you all are amazing! Thank you for your friendships. To my Pentecostal Tabernacle family, thank you for making church on Sundays a day I always look forward to.

I am indebted to my parents. Mum and dad, your constant love, prayers and support have helped me become the woman I am today. This journey would not have been possible without the sacrifices you both made for me. And mum, you can now add another degree to your name. To my sisters Esosa, Adesuwa, Nosariemen, Osato, and my brother-in-law Taiwo, you are the best! Thank you for your love and support. Also, many thanks to members of the Shobiye family.

I am especially thankful to my husband, Ayobami, who has been there for me every step of the way. Thank you for believing in me, loving, encouraging and supporting me through and through.
This doctoral thesis has been examined by a committee of the Department of Electrical Engineering and Computer Science as follows:

Professor John Guttag

Member, Thesis Committee
Professor of Electrical Engineering and Computer Science

Professor Peter Szolovits

Member, Thesis Committee
Professor of Electrical Engineering and Computer Science

Professor Rosalind W. Picard

Thesis Supervisor
Professor of Media Arts and Sciences
# Contents

<table>
<thead>
<tr>
<th>List of Figures</th>
<th>13</th>
</tr>
</thead>
<tbody>
<tr>
<td>List of Tables</td>
<td>17</td>
</tr>
</tbody>
</table>

## 1 Introduction
1.1 Well-Being Prediction ........................................... 21
1.2 Assessing Health Behaviors ...................................... 21
1.3 Study Cohort ...................................................... 21
1.4 Probabilistic Latent Variable Models ......................... 23
1.5 Contributions .................................................... 24
1.6 Organization ...................................................... 25

## 2 Background
2.1 Multi-Task Learning ................................................. 27
2.2 Probabilistic Models and Inference Methods ................... 28
  2.2.1 Posterior Inference: Variational Inference ............... 30
2.3 Dirichlet Process .................................................. 32
2.4 Latent Dirichlet Allocation ..................................... 34

## 3 Well-Being Prediction with Multi-Task Learning
3.1 Introduction ......................................................... 40
  3.1.1 Related Work .................................................. 41
3.2 Data Representation ................................................ 42
  3.2.1 Classification Labels ....................................... 42
  3.2.2 Features ..................................................... 43
  3.2.3 Pre-Study Surveys .......................................... 46
3.3 Hierarchical Bayesian Logistic Regression ....................... 47
  3.3.1 Variational Inference ...................................... 50
  3.3.2 Prediction .................................................. 51
3.4 Evaluation ........................................................ 53
  3.4.1 Experiments ................................................ 53

Draft: August 28, 2018
3.5 Results ......................................................... 54
  3.5.1 Discussion ............................................. 56
  3.6 Summary .................................................. 63

4 Assessment of Health Behaviors with Supervised Topic Models 65
  4.1 Introduction ............................................. 66
    4.1.1 Related Work ....................................... 67
    4.1.2 Our Framework ..................................... 68
  4.2 Data Representation .................................... 69
  4.3 Supervised Latent Dirichlet Allocation ..................... 73
    4.3.1 Variational Expectation-Maximization ............... 74
    4.3.2 Prediction .......................................... 76
  4.4 Evaluation .............................................. 77
  4.5 Results and Discussion .................................. 79
    4.5.1 LASSO .............................................. 82
    4.5.2 SLDA ................................................ 85
  4.6 Summary ................................................ 87

5 Recommending Health Behaviors: Two Case Studies 95
  5.1 Introduction ............................................. 95
  5.2 Related Work ........................................... 96
  5.3 Health Behavior Recommendation .......................... 97
    5.3.1 Jensen-Shannon Distance ........................... 98
  5.4 Case 1 .................................................. 99
  5.5 Case 2 ................................................ 102
  5.6 Summary ................................................ 107

6 Conclusion .................................................. 109
  6.1 Future Work ............................................. 110

Bibliography .................................................... 113

A SNAPSHOT Data .............................................. 121

B Assessment of Behavioral Influences on Self-Reported Stress 123

C Assessment of Behavioral Influences on Self-Reported Happiness 125

D Assessment of Behavioral Influences on Self-Reported Health 131
## List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Graphical model representation of a generic latent variable model.</td>
<td>29</td>
</tr>
<tr>
<td>2.2</td>
<td>Graphical model representation of the latent Dirichlet allocation.</td>
<td>35</td>
</tr>
<tr>
<td>3.1</td>
<td>Distribution of self-report labels after discarding the middle 20%.</td>
<td>43</td>
</tr>
<tr>
<td>3.2</td>
<td>Features extracted for each detected, non-artifact SCR.</td>
<td>44</td>
</tr>
<tr>
<td>3.3</td>
<td>An example of a participant’s SMS frequency over four consecutive days.</td>
<td>45</td>
</tr>
<tr>
<td>3.4</td>
<td>Percent of participants sleeping, studying, in extra-curricular activities,</td>
<td>45</td>
</tr>
<tr>
<td></td>
<td>and exercising throughout the day.</td>
<td></td>
</tr>
<tr>
<td>3.5</td>
<td>Graphical model representation of the HBLR model. Shaded nodes are</td>
<td>49</td>
</tr>
<tr>
<td></td>
<td>observed and arrows indicate dependencies between variables.</td>
<td></td>
</tr>
<tr>
<td>3.6</td>
<td>Prediction accuracies for the STL, MTL-well-being, and MTL-participants</td>
<td>55</td>
</tr>
<tr>
<td></td>
<td>approaches (* = p &lt; 0.05).</td>
<td></td>
</tr>
<tr>
<td>3.7</td>
<td>AUC for the STL, MTL-well-being, and MTL-participants approaches</td>
<td>55</td>
</tr>
<tr>
<td></td>
<td>(* = p &lt; 0.05).</td>
<td></td>
</tr>
<tr>
<td>3.8</td>
<td>Resulting soft clustering ((\phi)) learned when predicting the different</td>
<td>56</td>
</tr>
<tr>
<td></td>
<td>labels (mood, stress, and health) for the 104 participants.</td>
<td></td>
</tr>
<tr>
<td>3.9</td>
<td>Distribution of HBLR weights on the total number of screen-on events</td>
<td>57</td>
</tr>
<tr>
<td></td>
<td>(5PM–Midnight) feature for each cluster when predicting tomorrow’s mood.</td>
<td></td>
</tr>
<tr>
<td>3.10</td>
<td>Example of different weight distributions induced by the soft clustering</td>
<td>58</td>
</tr>
<tr>
<td></td>
<td>for three different participants in the mood prediction, over the same</td>
<td></td>
</tr>
<tr>
<td></td>
<td>feature in Figure 3.9.</td>
<td></td>
</tr>
<tr>
<td>3.11</td>
<td>Mean feature weights for mood clusters in HBLR model. The positive label is</td>
<td>61</td>
</tr>
<tr>
<td></td>
<td>happy and the negative label is sad.</td>
<td></td>
</tr>
<tr>
<td>3.12</td>
<td>Mean feature weights for stress clusters in HBLR model. The positive label</td>
<td>62</td>
</tr>
<tr>
<td></td>
<td>is calmness and the negative label is stress.</td>
<td></td>
</tr>
<tr>
<td>4.1</td>
<td>Distribution of self-reported stress/calmness response variables.</td>
<td>72</td>
</tr>
<tr>
<td></td>
<td>The mean of the response variables, 53.43, is shown in red.</td>
<td></td>
</tr>
</tbody>
</table>
4.2 Graphical model representation for the supervised latent Dirichlet allocation (sLDA) model. Shaded nodes are observed and arrows indicate dependencies between variables. .................................................. 75
4.3 Average binary prediction accuracy (over 5 runs, ± standard deviation) for sLDA and LDA + Regression models across different topics (threshold = 50). .................................................. 79
4.4 Average binary prediction accuracy (over 5 runs, ± standard deviation) for sLDA and LDA + Regression models across different topics (threshold = 53.43, average of the participants self-reported stress). .................. 80
4.5 Average F1 score (over 5 runs, ± standard deviation) for sLDA and LDA + Regression models across different topics (threshold = 50). .................. 80
4.6 Average F1 score (over 5 runs, ± standard deviation) for sLDA and LDA + Regression models across different topics (threshold = 53.43). .................. 81
4.7 Average correlation (over 5 runs, ± standard deviation) between the true and predicted response variables for sLDA and LDA + Regression models across different topics. .................................................. 81
4.8 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. .................................................. 91
4.9 Raster plots showing a regular sleeper, SR = 0.91 (top), and an irregular sleeper, SR = 0.68 (bottom). .................................................. 92
4.10 An 11-topic 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. .................. 93
4.11 Learned day-pattern proportions (θ) in three different levels of the response label (y): (a) 50 participant-days with y ≥ 80; (b) 50 participant-days with 55 ≤ y < 65; (c) 50 participant-days with y ≤ 40. Pattern 4 is the most calm pattern, and pattern 5 is the most stress pattern. .................. 94
5.1 Self-reported stress-calm scores of Participant A for the past 17 days of the study. .................................................. 99
5.2 Health behaviors extracted from cell-phone usage for the past 17 days. Ordered by the average of self-reported stress-calm scores on days the behavior was present (100: most calm, 0: most stress). .................. 100
5.3 Health behaviors corresponding to sleep extracted from Actiwatch for the past 17 days. Ordered by the average of self-reported stress-calm scores on days the behavior was present (100: most calm, 0: most stress). 101
5.4 Health behaviors extracted from GPS coordinates and twice daily surveys for the past 17 days. Ordered by the average of self-reported stress-calm scores on days the behavior was present (100: most calm, 0: most stress). 101
5.5 Self-reported stress-calm scores of Participant B for the past 13 days of the study. ........................................... 103
5.6 Health behaviors extracted from cell-phone usage for the past 13 days. Ordered by the average of self-reported stress-calm scores on days the behavior was present (100: most calm, 0: most stress). .......................... 104
5.7 Health behaviors corresponding to sleep extracted from Actiwatch for the past 13 days. Ordered by the average of self-reported stress-calm scores on days the behavior was present (100: most calm, 0: most stress). 105
5.8 Health behaviors extracted from GPS coordinates and twice daily surveys for the past 13 days. Ordered by the average of self-reported stress-calm scores on days the behavior was present (100: most calm, 0: most stress). 105

C.1 Distribution of self-reported sad/happiness response variables. The mean of the response variables, 60.96, is shown in red. .................. 126
C.2 LASSO Non-zero coefficient estimates for self-reported happiness. In the data, 100 represents the highest level of happiness and 0 represents the highest level of sadness. So the positive coefficients are positively correlated with increasing happiness. ........................................ 127
C.3 Average binary prediction accuracy (over 5 runs, ± standard deviation) for sLDA and LDA + Regression models across different topics (threshold = 60.96) for self-reported happiness. .......................... 128
C.4 Average F1 score (over 5 runs, ± standard deviation) for sLDA and LDA + Regression models across different topics (threshold = 60.96) for self-reported happiness. .......................... 128
C.5 Average correlation (over 5 runs, ± standard deviation) between the true and predicted response variables for sLDA and LDA + Regression models across different topics for self-reported happiness. .......................... 129

D.1 Distribution of self-reported sick/healthy response variables. The mean of the response variables, 65.00, is shown in red. .................. 132
D.2 LASSO Non-zero coefficient estimates for self-reported health. In the data, 100 represents the highest level of feelings of health and 0 represents the highest level of feelings of sickness. So the positive coefficients are positively correlated with increasing health. ........................................ 133
D.3 Average binary prediction accuracy (over 5 runs, ± standard deviation) for sLDA and LDA + Regression models across different topics (threshold = 65) for self-reported health. .......................... 134
D.4 Average F1 score (over 5 runs, ± standard deviation) for sLDA and LDA + Regression models across different topics (threshold = 65) for self-reported health. .......................... 134
D.5 Average correlation (over 5 runs, ± standard deviation) between the true and predicted response variables for sLDA and LDA + Regression models across different topics for self-reported health. .......................... 135

Draft: August 28, 2018
List of Tables

3.1 Selected 21 features and modalities 47
3.2 Prediction performance (Accuracy, AUC) of the majority classifier, STL, MTL well-being-as-task, and MTL users-as-task approaches. Bold entries represent significant improvements over the other approaches. 54
3.3 Computed pre-study measures for the HBLR mood prediction clusters. Bold entries represent significant differences from the sample average. 59
3.4 Computed pre-study measures for the HBLR stress prediction clusters. Bold entries represent significant differences from the sample average. 60
4.1 Selected modifiable behaviors, the bins and total number of days the behaviors were present in the data. 71
4.2 Mean (± standard deviation) of binary accuracies, F1 scores, and correlation coefficients for sLDA and LASSO. Bold entries represent a statistically significant improvement of the sLDA model over LASSO (p < 0.05). 82
4.3 Behaviors selected by the LASSO model in decreasing order of importance. 83
5.1 Health behaviors exhibited on Day 18 by participant A. 102
5.2 Health behaviors of two participants most similar to Participant A. Bold entries highlight possible recommendations. 103
5.3 Health behaviors exhibited on Day 14 by participant B. 106
5.4 Health behaviors of two participants most similar to Participant B. Bold entries highlight possible recommendations. 106
B.1 List of top 15 behaviors in the 11 patterns (topics) learned by sLDA for self-reported stress. Behaviors are listed in decreasing order. Patterns are listed in order of decreasing association with calmness. 123
B.2 List of top 15 behaviors in the 11 patterns (topics) learned by sLDA for self-reported stress (continued). Behaviors are listed in decreasing order. Patterns are listed in order of decreasing association with calmness. 124
C.1 Mean (± standard deviation) of binary accuracy, F1 score (threshold = mean), and correlation coefficients for sLDA and LASSO when predicting self-reported happiness. Bold entries represent a statistically significant improvement ($p < 0.05$).

C.2 List of top 15 behaviors in the 11 patterns (topics) learned by sLDA for self-reported happiness. Behaviors are listed in decreasing order. Patterns are listed in order of decreasing association with happiness.

C.3 List of top 15 behaviors in the 11 patterns (topics) learned by sLDA for self-reported happiness (continued). Behaviors are listed in decreasing order. Patterns are listed in order of decreasing association with happiness.

D.1 Mean (± standard deviation) of binary accuracy, F1 score (threshold = mean), and correlation coefficients for sLDA and LASSO when predicting self-reported health. Bold entries represent a statistically significant improvement ($p < 0.05$).

D.2 List of top 15 behaviors in the 12 patterns (topics) learned by sLDA for self-reported health. Behaviors are listed in decreasing order. Patterns are listed in order of decreasing association with health.

D.3 List of top 15 behaviors in the 12 patterns (topics) learned by sLDA for self-reported health (continued). Behaviors are listed in decreasing order. Patterns are listed in order of decreasing association with health.
There is an increasing push to improve the health and well-being of individuals domestically and globally. The growing presence and use of personal health related mobile applications, fitness trackers and other sensors has made this possible. 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.\footnote{IBM Watson Health: https://www-03.ibm.com/press/uk/en/pressrelease/46609.wss. Retrieved May 15, 2018.} As a result, it is now feasible to personalize health care to each user from the data being generated. At the individual level, this has led to a shift in health seeking behavior, as users are increasingly demanding insights from the data being generated to better understand their body systems and improve their health and well-being. Improving the health and well-being of individuals is also receiving a growing attention at the global level. For example, in September 2015, at the United Nations summit, countries gathered together to adopt “good health and well-being for all and at all ages” as one of the 17 Sustainable Development Goals to be achieved by 2030.\footnote{World Health Organization: http://www.who.int/mental_health/suicide-prevention/SDGs/en/. Retrieved May 15, 2018.}

Health and well-being combines being in both good physical and mental state, and con-
tributes to disease prevention. For instance, studies have shown associations between increased levels of well-being and decreased risk of heart disease and illness, and increased longevity [1,2]. Another study has shown that individuals with higher levels of well-being are more productive at work and tend to contribute more to the society [3].

Physical health is the most visible part of the various dimensions of health and well-being and as a result, there is a huge focus on its improvement and maintenance. But mental well-being is equally important as it includes our emotional, psychological and social well-being. Theoretically, mental well-being in individuals is defined as having a multidimensional structure comprising: feelings of self acceptance, positive relation with others, self determination, environmental mastery, purpose in life, and self growth [4]. Empirically, it is defined through prominent indicators as the state of being comfortable (no stress), healthy, and in a good mood (having feelings of positive affect and absence of negative affect) [1,3,4].

The complex and subjective nature of mental well-being makes it challenging to assess. In the literature, it is assessed empirically using self-reports [1–3,5]. However, self-reports are not always accurate and cannot be solely relied upon. This is because they can be subjective, the calibration of rating scales used is not always consistent across users, and the scales are often subject to response biases. Also, self-reports cannot be used to inform individuals of which continuous behaviors in their daily lives contribute to mental well-being or lack thereof. Consequently, the use of objective data to predict personal well-being and study healthy behaviors in individuals, and to augment self-reports, has become important. Although there are still limitations, the availability of adequate multi-modal data, and advances in psychology, neuroscience, and machine learning methods suggest that well-being in individuals can be objectively measured and predicted accurately [6–8].
1.1 Well-Being Prediction

Prediction of future perceived mood, stress and health is useful to any person who might want to adjust their routine with the intention of improving their well-being. Modeling well-being is a difficult task, and across many research efforts using sophisticated models and/or multi-modal data, classification accuracies range from 55–80% (e.g., [9–11]). Research has shown that there are personality factors that affect stress, anxiety, and vulnerability to mental illnesses like depression [12–14]. Thus, a one-size-fits-all approach to predicting well-being is likely to be prone to poor performance and low accuracies. It is therefore essential to develop prediction models that can account for individual differences while sharing data from similar users.

1.2 Assessing Health Behaviors

In addition to the future prediction of well-being, understanding the behaviors that contribute to that state of well-being is equally important. Studies have shown that some of the major factors that contribute to well-being include daily health behaviors such as exercise, adequate sleep, and positive social interactions [15]. 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.

1.3 Study Cohort

In this thesis, we focus on improving the well-being of college students. This is because college years are usually very high stress periods for most young adults, and chronic stress can contribute to depression. For example, in the Spring 2017 National College Draft: August 28, 2018
Health Assessment, about 56% of the students surveyed reported more than average to tremendous stress compared to 50% in the Spring 2013 report. In addition, there has been a decline in mental health and well-being among college students. For instance, among US young adults aged 18–25 (college years), 8.3% have had very serious thoughts of suicide. Improving the well-being of college students is therefore pertinent at this time.

To improve well-being, a good starting point is to examine an individual’s daily behavior. Lyubomirsky and colleagues suggest that an individual has control over daily intentional activities and behaviors, and that these behaviors are important drivers of psychological well-being, and have the capacity to increase the individual’s level of happiness. Thus, predicting future mood, stress and health, and assessing behavioral influences on well-being are valid approaches that can help improve the well-being of college students.

For all of the analyses in this thesis, we use data from a past study that measured Sleep, Networks, Affect, performance, Stress and Health using Objective Techniques (SNAPSHOT) on college students. The SNAPSHOT study 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. For the 30-day study, about 50 students were enrolled each semester, and 15 students were enrolled for the semester-long study. In this thesis, we use a portion of the data from the 30-day study. The study administered pre- and post- study questionnaires, and surveyed the students twice daily for 30 days. The students were given wearable sensors (Q-sensor, Actiwatch) to collect their activity and physiological data. The study also collected location patterns from the students’ mobile phones. Daily weather information was also downloaded from nearby weather

---

3American College Health Assessment S’13, S’17 reports: [http://www.acha-ncha.org/pubs_rpts.html](http://www.acha-ncha.org/pubs_rpts.html)


stations. In addition, twice daily, the students were asked to rate on a scale of 0–100 how sad or happy, stressed or calm, sick or healthy they felt. See Appendix A for more details on the study.

1.4 Probabilistic Latent Variable Models

Probabilistic latent variable models assume that there is a hidden structure or pattern that governs a stochastic data-generation process. These models are able to learn the relationship between the latent patterns and the observed variables in the data, and the patterns can be used to summarize the data or to form predictions. Examples of well studied latent variable models are Gaussian mixture models, hidden Markov models, and latent Dirichlet allocation (LDA) [20], which are used extensively in natural language processing, analysis of large scale social networks and computational biology. In Chapter 3, we use a hierarchical Bayesian logistic regression model to predict future well-being of college students in the SNAPSHOT study.

The LDA is a widely used latent variable model for summarizing documents. Its major assumption is that in generating the words in a document, a latent topic is first selected and then a word that frequently appears in that topic is generated. This process is repeated several times to form the documents in the collection. Documents have several topics represented in different proportions and the words have varying degrees of memberships in these topics. Statistical inference is used to learn these latent topics (details in section 2.2). After learning, the topics can be used to organize and browse large scale documents efficiently. Apart from document modeling, the LDA can also be applied to other domains such as image retrieval and bioinformatics. In Chapter 4, we use a supervised variant of the LDA to model health behaviors in college students, and in Chapter 5, we illustrate how the learned representations can provide actionable
insights to individuals looking to improve their well-being.

1.5 Contributions

The major contributions of this dissertation are outlined below. Since we address different problems in subsequent chapters, we provide thorough literature reviews of specific prior work in those chapters.

Predicting Future Well-being: Using Multi-task learning (MTL) with a hierarchical Bayesian logistic regression (HBLR) model, we predict future well-being without requiring an extensive history of self-reports, achieving statistically significant improvements over the single-task learning (STL) case. We demonstrate that a personalized MTL model is more effective at making accurate predictions when compared to an equivalent non-personalized MTL model. The reason for this improved accuracy is because the personalized model takes into account inter-individual differences that non-personalized MTL models do not account for. We also show that the HBLR model can uncover relevant relationships between input features and previously unseen personality traits of individuals.

Modeling Health Behaviors: We propose a framework to map multi-modal data collected in the “wild” to meaningful representations of health behavior. Using supervised latent Dirichlet allocation (sLDA), we capture latent patterns of health behaviors that are best predictive of well-being in a college student population when compared to latent patterns uncovered by an unsupervised topic model. We also illustrate how these patterns can provide actionable insights to individuals looking to improve their well-being.
The rest of the thesis is organized as follows:

1. In Chapter 2, we provide the detailed background and prior work done on multi-task learning, probabilistic generative models and variational inference helpful in understanding the rest of the dissertation.

2. In Chapter 3, we use multi-task learning to predict future well-being of college students [8, 21].

3. In Chapter 4, we use supervised LDA to learn efficient representations of health behaviors present in a cohort of college students.

4. In Chapter 5, using a case study approach, we illustrate how efficient representations of health behaviors can provide actionable insights to individuals looking to improve their well-being.

5. In Chapter 6, we conclude and provide future research directions.
Chapter 2

Background

In this chapter, we discuss some background knowledge and previous work on multi-task learning and inference techniques for probabilistic modeling that will be helpful in understanding the rest of the dissertation. We also introduce some definitions that will be used in the chapters that follow.

2.1 Multi-Task Learning

Multi-task learning (MTL) is the ability of a learning algorithm to exploit commonalities between different learning tasks in order to share statistical strength and transfer knowledge across tasks [22]. Its goal is to improve generalization performance between tasks that are sufficiently related [23,24]. MTL is beneficial when training datasets are scarce and noisy. This makes it well-suited for many real-world problems, including predicting well-being. Although it was originally proposed as a way to enforce efficient internal representations within neural networks [23], MTL is currently used across different types of models. For example, Baxter et al. proposed hierarchical Bayesian learning as an approach to MTL [25]. The general idea is to draw each task’s model parameters from a common prior distribution, thus imposing similarity constraints through the prior. An example is the Transfer-Aware Naive Bayesian model where using Bayesian
inference techniques, the model updates its parameters by decreasing the variance if the tasks are very similar [24].

MTL has also been applied in the area of affective computing. Kandemir et al. introduced treating the prediction of a single person’s affect as a task using Multi-Task Multi-Kernel Learning (MTMKL) [26]. This method was used in prior work which applied MTMKL to the SNAPSHOT dataset by treating the classification of different well-being states as related tasks [8,21,27]. MTL has also been used to treat outcomes like arousal and valence as the related tasks in a model. Xia et al. improved the performance of a deep belief network by training it to simultaneously recognize both valence and arousal from speech [28]. Another study of emotion recognition from speech found that treating different corpora, domains, or genders as related tasks in an MTL framework offered performance benefits over learning a single model for all the domains, or a separate model for each domain [29].

In chapter 3, we use MTL and a Bayesian model approach to account for inter-individual differences in the relationship between behavior, physiology, and resulting well-being. We define each task as predicting the well-being of participants in the SNAPSHOT study. The Bayesian model adapts to the specific characteristic of each person, while sharing information through a common prior placed across tasks.

### 2.2 Probabilistic Models and Inference Methods

A probabilistic latent variable model is a generative model that represents unobserved structure in data. Given latent variables $z$, and model parameters $\theta$, the observed data $x = \{x_1, \cdots, x_N\}$ are stochastically generated by a model $P(x \mid z, \theta)$. Priors that represent the initial belief about the data generating process can also be included in the model. There are different types of latent variable models including Gaussian mixture
models, hidden Markov models, and mixed-membership models such as latent Dirichlet allocation (LDA) [20]. Figure 2.1 illustrates the graphical model representation (or plate notation) of a generic latent variable model. The shaded nodes are observed variables; unshaded nodes represent hidden (latent) variables; and a directed edge from node $a$ to node $b$ indicates that node $b$ depends on node $a$. The plates show that variables are grouped together, and the value in the lower corner of the plate indicates the cardinality of the grouped variables.

The goal of any Bayesian inference task is to learn the posterior probability of the latent variables and model parameters given the observed data $P(z, \theta | x)$. This is achieved via Bayes’ rule:

$$P(z, \theta | x) = P(x | z, \theta)P(z, \theta) / P(x) = \frac{\int_z \int_\theta P(x | z, \theta)P(z, \theta) dz d\theta}{\int_z \int_\theta P(x | z, \theta)P(z, \theta) dz d\theta}$$

where, $P(z, \theta | x)$ is the posterior, $P(z, \theta)$ is the prior, $P(x | z, \theta)$ is the likelihood, and $P(x)$ is the evidence. Typically, the posterior is intractable because of the normalizing constant in the denominator. Therefore, techniques such as sampling based methods (e.g. Markov chain Monte Carlo (MCMC) [30, 31]), expectation-propagation [32], or variational inference [33] are used to approximate the posterior.
2.2.1 Posterior Inference: Variational Inference

Variational inference reduces the problem of posterior inference to an optimization problem. The basic idea behind variational inference is to choose a tractable family of variational distributions $q(z, \theta)$ to approximate the intractable posterior $p(z, \theta | x)$. To derive the approximate posterior, Jensen inequality is applied as follows:

$$\log P(x) = \log \int_{z, \theta} P(x, z, \theta) \, dz \, d\theta$$
$$= \log \int_{z, \theta} q(z, \theta) \frac{P(x, z, \theta)}{q(z, \theta)} \, dz \, d\theta$$
$$\geq \int_{z, \theta} q(z, \theta) \log \frac{P(x, z, \theta)}{q(z, \theta)} \, dz \, d\theta$$
$$= \mathbb{E}_q[\log P(x, z, \theta)] - \mathbb{E}_q[\log q(z, \theta)],$$

where the expectation is taken with respect to the variational distribution $q(z, \theta)$. This results in a lower bound of the log-evidence $\log P(x)$. To find the optimal $q(z)$, note that

$$\Delta = \log P(x) - \int_{z, \theta} q(z, \theta) \log \frac{P(x, z, \theta)}{q(z, \theta)} \, dz \, d\theta$$
$$= \int_{z, \theta} q(z, \theta) \log P(x) \, dz \, d\theta - \int_{z, \theta} q(z, \theta) \log \frac{P(x, z, \theta)}{q(z, \theta)} \, dz \, d\theta$$
$$= \int_{z, \theta} q(z, \theta) \log \frac{P(x)q(z, \theta)}{P(x, z, \theta)} \, dz \, d\theta$$
$$= \int_{z, \theta} q(z, \theta) \log \frac{q(z, \theta)}{P(z, \theta | x)} \, dz \, d\theta \equiv \text{KL}(q(z, \theta) || p(z, \theta | x))$$

This implies that, the difference between the variational distribution $q(z, \theta)$ and the posterior distribution $P(z, \theta | x)$ is the Kullback-Leibler (KL) divergence between the
two distributions. The log-evidence can therefore be re-written as

$$\log P(x) = \mathbb{E}_q[\log P(x, z, \theta)] - \mathbb{E}_q[\log q(z, \theta)] + \text{KL}(q(z, \theta) || p(z, \theta | x))$$

where $\mathcal{L}$ is called the evidence lower bound (ELBO). Since the KL-divergence is non-negative and equals 0 only if $q(z, \theta) = p(z, \theta | x)$, $\mathcal{L}(q, z, \theta)$ acts as a tight lower-bound that is equal to $\log P(x)$ when $q(z, \theta) = p(z, \theta | x)$. Therefore optimizing the ELBO is equivalent to minimizing the KL-divergence between the variational distribution and the posterior distribution. The approximate posterior that minimizes the KL-divergence is used as a proxy for the true posterior and can then make predictions.

The complexity of the variational distribution determines the complexity of the optimization. So a simplified popular choice is the mean-field variational family, where the latent variables are mutually independent and the variational distribution is fully factorized: $q(z, \theta) = (\prod_n q_n(z_n)) (\prod_m q_m(\theta_m))$. The general form for the optimal mean-field variational distributions are:

$$q^*(z_n) \propto \exp\{\mathbb{E}_{q_{-n}}[\log P(z_n, x, z_{-n}, \theta)]\}$$

$$q^*(\theta_m) \propto \exp\{\mathbb{E}_{q_{-m}}[\log P(\theta_m, x, \theta_{-m}, z)]\},$$

where $z_{-n}$ represents all of $z$ except the $n$-th dimension and $\mathbb{E}_{q_{-n}}[\cdot]$ represents taking expectations with respect to everything excluding the distribution $q_n(z_n)$, similarly for $\theta$. When the complete conditionals $P(z_n | x, z_{-n}, \theta)$ and $P(\theta_m | x, \theta_{-m}, z)$ are in the exponential family, the optimal mean-field variational distributions can be computed exactly. This results in the standard coordinate ascent variational inference (CAVI) algorithm, where both conditionals are iteratively re-estimated until convergence. In chapter 3 and 4, we use mean-field variational inference to learn the posterior probability
CHAPTER 2. BACKGROUND

\textbf{Input} : A model \( P(x, z, \theta) \), a data set \( x \)

\textbf{Output}: A variational density \( q(z, \theta) = (\prod_{n=1}^{N} q_{n}(z_{n})) \left( \prod_{m=1}^{M} q_{m}(\theta_{m}) \right) \)

\textbf{Initialize} : Variational factors \( q_{n}(z_{n}), q_{m}(\theta_{m}) \)

\textbf{while} ELBO has not converged \textbf{do}

\hspace{1em} for \( n \in \{1, \cdots, N\} \) \textbf{do}

\hspace{2em} Set \( q_{n}(z_{n}) \propto \exp\{E_{-n}[\log P(z_{n} | z_{-n}, x, \theta)]\} \)

\hspace{1em} end

\hspace{1em} for \( m \in \{1, \cdots, M\} \) \textbf{do}

\hspace{2em} Set \( q_{m}(\theta_{m}) \propto \exp\{E_{-m}[\log P(\theta_{m} | \theta_{-m}, x, z)]\} \)

\hspace{1em} end

\hspace{1em} Compute \( \text{ELBO}(q) = E_{q}[\log P(x, z, \theta)] - E_{q}[\log q(z, \theta)] \)

\hspace{1em} end

\textbf{return} : \( q(z, \theta) \)

\begin{algorithm}[h]
\caption{CAVI algorithm for mean-field variational inference}
\end{algorithm}

of the hierarchical Bayesian logistic regression model, and the supervised latent Dirichlet allocation model, respectively.

\subsection*{2.3 Dirichlet Process}
An example of a prior that is used in probabilistic models is the Dirichlet process (DP). The DP is often used in Bayesian inference to describe the prior knowledge of the distribution of random variables. It is specified by a base distribution \( G_{0} \), and concentration parameter \( \alpha, \{\alpha \in \mathbb{R} | \alpha > 0\} \). The mathematical representation of a DP is

\[ G \sim \text{DP}(\alpha, G_{0}) \]

The base distribution \( G_{0} \) is the expected value of the DP. Random variables are drawn around the base distribution, similar to how the Gaussian process draws values around its mean; it represents our prior knowledge or expectation around \( G \). The DP is very popular for its clustering inducing property, and \( \alpha \) controls the probability of creating new clusters; the higher the value of alpha, the more clusters are created.

There is no explicit form for \( G \), but it is represented by the Chinese restaurant process or...
the stick breaking representation. The Chinese restaurant process is a single parameter distribution over partitions of the integers. In this representation, data is drawn from the DP through the following process:

1. Draw $x_1 \sim G_0$

2. For $n > 1$:
   
   (a) with probability $p = \frac{\alpha}{\alpha + n - 1}$, draw $x_n \sim G_0$
   
   (b) where $x$ represents a past observation and $n_x = \text{number of past observations of } x$, with $p = \frac{n_x}{\alpha + n - 1}$, set $x_n = x$.

The stick breaking view of the Dirichlet distribution places a distribution on nonnegative $K$-dimensional vectors whose components sum to one. In this representation, the interval $(0, 1)$ is a unit-length stick that is constructed as follows:

1. Draw $v_1 \sim \text{Beta}(1, \alpha)$, and break off a fraction $v_1$ of the stick. Set $\pi_1 = v_1$ and $(1 - \pi_1)$ is the remainder of the stick.

2. For $k > 1$, draw $v_k \sim \text{Beta}(1, \alpha)$ and set $\pi_k = v_k \prod_{i=1}^{k-1} (1 - v_i)$

Sethuraman et al. shows that the resulting sequence $\{\pi_i\}$ satisfies $\sum_i \pi_i = 1$ with probability one [34]. To generate random variables $x$ from the DP, each $x_n$ is drawn from the base distribution $G_0$, and $v_k$ is simultaneously drawn from $\text{Beta}(1, \alpha)$ as described above. This results in a DP distribution

$$G = \sum_{k=1}^{\infty} \pi_k \delta_{x_k},$$

where $x_k$ and $\pi_k$ represent the location and weight of the $k$th stick. Setting $v_K$ to one instead of being drawn from the Beta distribution results in a truncated approximation to the DP-process, which is more computationally efficient in practice:

$$G = \sum_{k=1}^{K} \pi_k \delta_{x_k}.$$

There is no loss of generality with this approximation. In Chapter 3, we use the trun-
cated approximation of the DP to induce clustering of the tasks classifier weights.

2.4 Latent Dirichlet Allocation

Latent Dirichlet Allocation (LDA) or Topic model is a mixed-membership Bayesian latent variable model. It has been applied to many tasks, such as documents classification and summaries [20], relevant documents retrieval from a query [35], and recommending purchased items [36], amongst others. In document modeling, the observed variables are the words in each document, and the basic idea is that each document can be represented as a random mixture over latent topics.

For the LDA model, \( n \in \{1, \cdots, N\} \) is used to index the number of words in a document, \( d \in \{1, \cdots, D\} \) is used to index the number of documents in the corpus, and \( k \in \{1, \cdots, K\} \) is used to index the number of latent topics. The model has two latent variables: \( \theta_{1:K} \) - the topic proportions in each document, and \( \beta_{1:K} \) - the topics representing a probability distribution over all the words present in the collection of documents (or corpus). Given a Dirichlet parameter \( \alpha \), the LDA assumes the following data generation process:

1. Choose \( \theta_{1:K} \sim \text{Dir.}(\alpha) \)
2. For each word \( w_n \) in document \( d \),
   (a) Choose a topic assignment \( z_n \sim \text{Mult.}(\theta) \)
   (b) Choose a word \( w_n \sim P(w_n \mid z_n, \beta_{1:k}) \)

where \( \beta_{ij} = P(w^j = 1 \mid z^i = 1) \). Figure 2.2 illustrates the graphical model representation of the LDA.
Given the model parameter $\alpha$ and $\beta_{1:K}$, the joint distribution is given by:

$$P(\theta, z, w \mid \alpha) = P(\theta \mid \alpha) \prod_{n=1}^{N} P(z_n \mid \theta) P(w_n \mid z_n, \beta_{1:K}),$$

and the posterior distribution of the latent variables given the model parameters is:

$$P(\theta, z \mid w, \alpha, \beta_{1:K}) = \frac{P(\theta, z, w \mid \alpha, \beta_{1:K})}{P(w \mid \alpha, \beta_{1:K})},$$

where,

$$P(w \mid \alpha, \beta_{1:K}) = \int P(\theta \mid \alpha) \left( \prod_{n=1}^{N} \sum_{z_n} P(z_n \mid \theta) P(w_n \mid z_n, \beta_{1:K}) \right) d\theta$$

As explained in Section 2.2, this posterior is intractable because of the marginal distribution of words $P(w \mid \alpha, \beta_{1:K})$, and so mean-field variational inference is employed to approximate it. The fully factorized variational distribution is specified as

$$q(\theta, z \mid \gamma, \phi) = q(\theta \mid \gamma) \prod_{n=1}^{N} q(z_n \mid \phi_n),$$

where $\gamma$, a Dirichlet parameter, and $\phi_{1:N}$, multinomial parameters, are the parameters of the variational distribution. Following the derivation outlined in Section 2.2.1, update equations that optimize the document level ELBO for each word $w_n$ and each topic $k$ are:

$$\phi_{nk} \propto \beta_{kw_n} \exp\{E_q[\log(\theta_k) \mid \gamma]\}$$

$$\gamma_k = \alpha_k + \sum_{n=1}^{N} \phi_{nk}.$$

Given the estimated latent variables ($\gamma$ and $\phi$), the next step is to find the model
**Input**: Words \( \{x_{1:N}\}_{d=1}^{D} \), number of topics \( K \), parameter \( \alpha \)

**Output**: Learned parameters: \( \gamma, \phi, \beta_{1:K} \)

**Initialize**: Variational factors \( \phi_{n}, \gamma \), and model parameter \( \beta_{1:K} \)

**while** ELBO has not converged **do**

1. **for** \( d \leftarrow 1 \) to \( D \) **do**
   1. **for** \( n \leftarrow 1 \) to \( N \) **do**
      1. Update variational parameter \( \phi_{n} \)
   2. Normalize \( \phi_{n} \) to sum to 1
   **end**

2. Update variational parameter \( \gamma \)

3. Estimate model parameter \( \beta_{1:K} \)
   1. Normalize \( \beta_{k} \) to sum to 1

**end**

**return**: \( \gamma, \phi, \beta \)

**Algorithm 2**: Variational Expectation-Maximization algorithm for LDA.

Parameter \( \beta_{1:K} \) that maximizes the log-likelihood of the data:

\[
\mathcal{L}(\beta, \alpha) = \sum_{d=1}^{D} \log P(w_{d} | \alpha, \beta_{1:K}).
\]

This distribution cannot be analytically solved, and so it is estimated by the following variational Expectation-Maximization procedure:

1. (E-Step) For each document, optimize the variational parameters \( \gamma, \phi_{1:N} \) following the description above.

2. (M-Step) Find the maximum likelihood estimate that optimizes the log-likelihood.

Blei et al. showed that the M-step update for the model parameter \( \beta \) is [20]:

\[
b_{i,j} \propto \sum_{d=1}^{D} \sum_{n=1}^{N_d} \phi_{d,ni}w_{d,n}^j.
\]

In practice, the model parameter \( \alpha \) is selected via cross-validation. The variational expectation-maximization algorithm is summarized in Algorithm 2.

It is important to note that LDA is not only used to model text, it has applications to data from domains such as bioinformatics [37], image retrieval [38], and human
activity modeling [39] to name a few. In Chapter 4, we use a supervised variant of the LDA to model health behaviors. In Chapter 5, we illustrate how information from this supervised LDA can be used to recommend health behaviors to individuals looking to improve their well-being.
Well-Being Prediction with Multi-Task Learning

Prediction of perceived well-being is an important but challenging task. In this chapter, we treat well-being prediction as a Multi-Task Learning (MTL) problem. MTL is a type of transfer learning in which models are simultaneously fitted for similar tasks, while sharing data and commonalities across the tasks [23]. We define each task as the prediction of a user’s perceived well-being as measured by self-reported mood, stress and health. We apply the hierarchical Bayesian logistic regression (HBLR) model, and constrain similar tasks (people) to share a common Dirichlet prior. This approach provides the best performance when compared to simply using a logistic regression classifier (the single-task learning alternative), or when tasks are defined as predicting the population’s self-reported mood, stress, and health. We present performance results on the SNAPSHOT dataset and also glean personalized and actionable insights from the study cohort.
Well-being leads to better physical health and its absence has dire consequences for clinical health. Research has shown that happier people tend to have healthier lifestyles [40], and stress increases vulnerability to infection and illness [41]. Also, in a 29-year study, self-reported health, which is strongly related to actual health [42], was found to be the most predictive measure of reduction in mortality risk, above more objective health measures such as blood pressure readings and difficulty breathing [43]. Therefore, the ability to model and predict these measures could be beneficial in the management, treatment and prevention of both mental illness and disease. Also, predicting future well-being using data collected from smartphones and wearable sensors, and subject to privacy restrictions, could be useful to any person who might want to adjust their routine with the intention of improving their well-being.

However, accurately modeling and predicting well-being is difficult, and across many research efforts using sophisticated models or multi-modal data or both, classification accuracies still range from 55–80% [9–11]. We argue that the inability to adapt to individual differences in these methods accounts for a portion of the prediction inaccuracies. Personality factors such as extraversion and neuroticism predispose individuals to be affected by stress and anxiety, and to be vulnerable to mental illnesses like depression in varying degrees [12–14]. Therefore, one-size-fits-all approaches to predicting well-being will be limited in the accuracies they can achieve.

In this chapter, we show that by using MTL, personalized models can account for individual differences and improve prediction accuracies in college students for these well-being states: mood, stress, and self-reported health. Also, we take a more challenging approach than what is typically done in the literature when modeling well-being. Most prior research focus on detecting the current state of well-being, but here, we include a
prediction gap of at least 20 hours and predict future well-being, given only today’s data. Specifically, let $x_t$ be all the smartphone, wearable sensor, and weather data collected about a person on day $t$ (from 12:00AM to 11:59PM). Let $y_t$ be the person’s perceived well-being in the evening of day $t$ (reported after 8pm). Unlike previous works that have focused on modeling $P(y_t|x_t)$, the probability of today’s well-being given today’s data (which we refer to as well-being detection), we model $P(y_{t+1}|x_t)$, the probability of tomorrow night’s well-being given today’s data, true future prediction.

3.1.1 Related Work

Recently, there has been an increase in research studies estimating mood, stress, and health using data collected from smartphones and wearable sensors. Bogomolov et al. use data from smartphones, information about participants personalities and the weather to detect stress with an accuracy of 72% [44]. Other researchers have used smartphone monitoring to detect depressive and manic states in bipolar disorder, attaining an accuracy of 76% [11]. And research on detecting workplace stress is also growing [45].

Individuals vary widely in their level of well-being, and the idea that these variations need to be accounted for in well-being modeling has been investigated by several researchers. For instance, Koldijk et al. found that adding the participant ID as a feature to their workplace stress estimation model improved accuracy in classifying mental effort [46]. Likewise, Canzian et al. found that using location data and data from surveys to train a single SVM to classify depressive mood resulted in sensitivity and specificity values of 0.74 and 0.78, respectively, while training an independent SVM for each person resulted in values of 0.71 and 0.87 [47].

As previously mentioned, the majority of prior work has focused on well-being detec-
CHAPTER 3. WELL-BEING PREDICTION WITH MULTI-TASK LEARNING

However, a recently published paper by Suhara et al. used a Recurrent Neural Network (RNN) to forecast future mood given two weeks of mood history self-reported every day [48]. This is equivalent to learning the function \( P(y_{t+1}|y_t, y_{t-1}, \ldots, y_1; x_t, x_{t-1}, \ldots, x_1) \). The authors achieved an AUC score of 0.886 in predicting future severely depressed mood. While a significant contribution, the results show that past mood features were most effective at predicting future mood compared to other data collected. This implies that using mood history to predict future mood is a significantly easier problem. In contrast, we are able to predict tomorrow’s well-being, not just mood, given a multi-modal dataset from today \( P(y_{t+1}|x_t) \), without requiring users to consistently and manually input self-reported labels consistently for at least two weeks.

3.2 Data Representation

We first describe the data used in our well-being prediction work. The data were collected as part of the SNAPSHOT study, and the goal of this work is to use the data to predict students well-being (mood, stress, and health). See Appendix A and Sano et al. [19] for more details on the complete SNAPSHOT study.

3.2.1 Classification Labels

The participants self-reported their evening perceived well-being – mood (sad/happy), stress (stressed out/calmness), and health (sick/healthy), on a visual analog scale from 0–100; we split these scores based on the median value to create binary classification labels and discard the middle 20% of scores. We also discarded participants with less than 10 days worth of data, since their dataset is insufficient to train viable models. At the end, the dataset comprised 104 users and 1842 days. Figure 3.1 shows the distribution of the labels after the middle 20% was thrown out. Almost all participants...
3.2. Features

To predict the labels, we extracted 343 features from smartphone logs, location data, physiological sensor recordings, and behavioral surveys obtained from participants each day, and carefully extracted features as detailed in previous work [21, 49]. Here we summarize the feature types.

**Physiology:** For 24 hours each day, skin conductance (SC), skin temperature, and 3-axis acceleration were collected at 8 Hz using wrist-worn Affectiva Q sensors. SC, controlled by the sympathetic nervous system, can peak when the body experiences a “fight or flight” response. This peak is termed a skin conductance response (SCR). We denoised the SC signal using a pre-trained algorithm [50], detected SCRs, and computed features related to their amplitude, shape, and rate, as illustrated in Figure 3.2. From the skin temperature and accelerometer, we extracted measures of activity, step count, and stillness. Since physical activity has been shown to reduce stress and improve mood [51], and skin temperature is related to the body’s circadian rhythm [52], we expect these features to be relevant. In total we computed 172 physiology features.
Figure 3.2: Features extracted for each detected, non-artifact SCR.

over different daily time periods.

**Phone (call, SMS, screen):** An app on participants’ phones logged their calls, text messages (SMS), and phone’s screen on or off times. We computed features based on the timing and duration of these events and the number of unique contacts each person interacts with. This resulted in a total of 20 call, 30 SMS, and 25 screen features. An example of SMS data from a participant over four consecutive days is shown in Figure 3.3, where we see that the texting pattern on a sad day is noticeably different from a happy or an indifferent day. For both physiology and phone, we computed each feature set over four time intervals: 12–3AM, 3–10AM, 10AM–5PM, 5–11:59PM. These intervals were determined by examining density plots of the times students were most likely to be asleep (3–10AM), or in class (10AM–5PM), as shown in Figure 3.4.

**Behavioral surveys:** We computed 38 features on students’ self-reported extra-curricular and academic activities, exercising, sleeping, napping, social interactions, and alcohol and drug consumption. We also included three extrinsic variables: participant ID, the day of the week, and whether it is a school night.
Weather: Previous studies have reported on how the weather affects mood, particularly in relation to Seasonal-Affective Disorder [52, 53]. Additionally, it is well known that particular seasons of the year (i.e., winter) have higher rates of poor health. Therefore, we extracted 40 features about the weather from DarkSkys Forecast.io API [54]. These features include information about sunlight, temperature, wind, barometric pressure, and the difference between today’s weather and the rolling average.

Location: The smartphone app logged participants’ GPS coordinates throughout the day. After cleaning, interpolating, and downsampling the signal, we computed a total of 15 features including the total distance traveled, the radius of the minimal circle
enclosing the location samples, time spent on the university campus, and time spent outdoors based on wifi usage. The location coordinates were also used to learn a Gaussian mixture model (GMM), giving a probability distribution over each participant’s typical locations. We then computed features such as the log-likelihood of the location pattern for each day. Essentially, this measures the routineness of the participant’s day, which has been previously found to be negatively associated with happiness and calmness [49].

**Feature selection:** We selected features based on assessing the ANOVA F-scores between each feature and the classification label in the training data. We removed highly correlated features with the constraint that at least one feature from each of the above data sources is retained. This process gave rise to a total of 21 features, which are listed in Table 3.1.

### 3.2.3 Pre-Study Surveys

Participants completed personality and mental health questionnaires at the start of the SNAPSHOT study. These measures included Myers-Briggs and Big Five Factor personality scores, state and trait anxiety scores, the 12-item Short Form Mental Health Composite Score (MCS) and Physical Health Composite Score (PCS), Pittsburgh Sleep Quality Index (PSQI), the Perceived Stress Scale (PSS) and the participant’s GPA and BMI (see Sano et al. [19] for details on these measures). Although this data is not used in the training of the HBLR model, we posit that it is relevant to the soft clustering learned by the model.
Table 3.1: Selected 21 features and modalities

<table>
<thead>
<tr>
<th>Modality</th>
<th>Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classifier</td>
<td>Day of the week</td>
</tr>
<tr>
<td>Physiology 3–10AM</td>
<td>% mins. with &gt;= 5 SCRs (w/o artifacts)</td>
</tr>
<tr>
<td></td>
<td>Temperature weighted SCR AUC</td>
</tr>
<tr>
<td>Location</td>
<td>Time on campus</td>
</tr>
<tr>
<td></td>
<td>Log likelihood of day given previous days</td>
</tr>
<tr>
<td>Call</td>
<td>Total missed calls</td>
</tr>
<tr>
<td>SMS</td>
<td>Total incoming (midnight–3AM )</td>
</tr>
<tr>
<td></td>
<td>Number of unique contacts outgoing</td>
</tr>
<tr>
<td></td>
<td>Number unique incoming (5–11:59PM)</td>
</tr>
<tr>
<td></td>
<td>Number unique outgoing (5–11:59PM)</td>
</tr>
<tr>
<td>Screen</td>
<td>Total duration (Midnight–3AM)</td>
</tr>
<tr>
<td></td>
<td>Total number on/off events (5–11:59PM)</td>
</tr>
<tr>
<td>Survey Activities</td>
<td>Exercise duration</td>
</tr>
<tr>
<td></td>
<td>Study duration</td>
</tr>
<tr>
<td>Survey Interaction</td>
<td>Positive social interaction</td>
</tr>
<tr>
<td></td>
<td>Pre-sleep in-person interaction (T/F)</td>
</tr>
<tr>
<td>Survey Sleep</td>
<td>Number of naps</td>
</tr>
<tr>
<td></td>
<td>All-nighter (T/F)</td>
</tr>
<tr>
<td>Weather</td>
<td>Cloud cover rolling std. dev.</td>
</tr>
<tr>
<td></td>
<td>Max precipitation intensity</td>
</tr>
<tr>
<td></td>
<td>Pressure rolling std. dev.</td>
</tr>
</tbody>
</table>

3.3 Hierarchical Bayesian Logistic Regression

The non-parametric Bayesian multi-task model we adopted for our well-being prediction was originally proposed by Xue et al. [55]. In MTL, it is assumed that tasks are related to one another, and so it is ideal to identify similar tasks and allow these tasks to share information. The model identifies similarities by clustering the tasks based on the relationship between input features and the resulting self-reported well-being. The model then performs MTL by jointly learning logistic regression classifiers for each cluster of similar tasks. Unlike in Caruana’s MTL work [23] where tasks were defined as similar if they use the same features for classification, this model defines task-similarity as when two classification boundaries are close, i.e., the weight vectors.
of the two classifiers are similar.

We use \( m \in 1, \cdots, M \) to index tasks (or people), \( d \in 1, \cdots, D \) to index feature dimensions, and \( n \in 1, \cdots, N_m \) to index the number of similar tasks in each cluster. The data set of a particular task \( m \) is defined as \( \{ \mathbf{x}^{(m)}_n, y^{(m)}_n \} \) such that the vector \( \mathbf{x}^{(m)}_n \in \mathbb{R}^d \) represents the set of task-specific input features, and \( y^{(m)}_n \in \{0, 1\} \) represents the corresponding label. For task \( m \), the conditional distribution of \( y^{(m)}_n \) given \( \mathbf{x}^{(m)}_n \) (modeled via logistic regression) is defined as,

\[
P(y^{(m)}_n | w_m, \mathbf{x}^{(m)}_n) = \sigma(w_m^T \mathbf{x}^{(m)}_n)^{y^{(m)}_n} [1 - \sigma(w_m^T \mathbf{x}^{(m)}_n)]^{1 - y^{(m)}_n},
\]

where \( \sigma(x) = \frac{1}{1 + \exp(-x)} \), and \( w_m \) is the weight vector of the classifier for task \( m \).

To induce the clustering of similar tasks, a Dirichlet process prior \( \alpha \) with base distribution \( G_0 \) is placed on the parameters \( w_m \) as follows (see Section 2.3 for a discussion on the Dirichlet process):

\[
w_m | G \sim G, G | \alpha, G_0 \sim DP(\alpha, G_0), G_0 \sim N_d(\mu, \Sigma).
\]

This ensures that the weight vectors are shared by similar tasks. A one-hot encoding of each task’s cluster assignment, \( c_m = [c_{m,1}, \cdots, c_{m,\infty}] \), selects the appropriate cluster weights \( w_m \) for a given task. Using the stick-breaking representation of the Dirichlet process, the observed data \( \{ \mathbf{x}^{(m)}_n, y^{(m)}_n \} \) is from the following data-generation process:

\[
\alpha \sim \text{Gamma}(\tau_{10}, \tau_{20}) \\
v_k \sim \text{Beta}(1, \alpha) \\
\pi_k = v_k \prod_{i=1}^{k-1} (1 - v_i) \\
(c_{m,1}, \cdots, c_{m,\infty}) \sim \text{Multinomial}(1 : \pi_1, \cdots, \pi_{\infty}) \\
w^*_k \sim \text{Normal}(\mu, \Sigma) \\
w_m = \prod_{k=1}^{\infty} (w^*_k)^{c_{m,k}} \\
y^{(m)} \sim \text{Binomial}(1, \sigma(w_m^T \mathbf{x}^{(m)}))
\]

where \( \alpha, \mu \) and \( \Sigma \) are model parameters, \( \tau_{10} \) and \( \tau_{20} \) are model hyper-parameters, and \( k \in 1, \cdots, \infty \) is used to index the learned clusters. Note that since this is a non-
Figure 3.5: Graphical model representation of the HBLR model. Shaded nodes are observed and arrows indicate dependencies between variables.

parametric model, in theory, there are no restrictions on the number of clusters the model can learn. A graphical model representation for HBLR is shown in Figure 3.5.

There are a couple of reasons to choose this model over other MTL approaches for well-being prediction as seen in Taylor et al. [21]. First, because the model is able to learn its own soft clustering over people, we can directly define each task as predicting the well-being of a single person. Secondly, the soft clustering provides valuable insights into groups of people that have different relationships between their physiology, behavior, and well-being. However, the HBLR model cannot make predictions about a new person’s well-being without first receiving at least one labeled training data point from that person. Still, HBLR quickly adapts to make predictions about a new person, and the predictions improve with more data [55].
3.3.1 Variational Inference

The set of latent variables $Z$ comprises the cluster assignments for each task (or person $m$) $\{c_m\}_{m=1}^M$, the probability of a task belonging to a cluster $\{v_k\}_{k=1}^\infty$, the Dirichlet parameter $\alpha$, and the weight vector for each cluster $\{w^*_k\}_{k=1}^\infty$. The model’s hyper-parameters $\Phi$ are $\{\tau_{10}, \tau_{20}, \mu, \Sigma\}$. Given $\Phi$, to learn $Z$ from the training data $\{x^{(m)}_n, y^{(m)}_n\}$, we need to compute the posterior $P(c_m, v_k, w^*_k | \{x^{(m)}_n, y^{(m)}_n\}, \tau_{10}, \tau_{20}, \mu, \Sigma)$.

This expression is intractable and so following Xue et al. [55], mean-field variational inference is used to approximate this posterior as described in Section 2.2.1.

There are two computational complexities that arise when applying mean-field variational inference to this model. First, the logistic regression model is not in the conjugate-exponential family and so a variational method that bounds log convex functions is used to approximate $P(y^{(m)}_n | w^*_k, x^{(m)}_n)$ as derived by Jaakkola et al. [56]:

$$P(y^{(m)}_n | w^*_k, x^{(m)}_n) \geq \sigma(\xi) \exp\left(\frac{(2y - 1)w_k^T x^{(m)}_n - \xi}{2} + \rho(\xi)(x^{(m)}_n w_k w_k^T x^{(m)}_n - \xi^2)\right),$$

where $\rho(\xi) = \frac{0.5 - \sigma(\xi)}{2\xi}$ and $\xi$ is a variational parameter. The equality holds when $\xi = \pm w_k^T x^{(m)}_n$.

Secondly, for improved computational efficiency, a truncated stick-breaking representation is used for the variational distribution. We set $v_K = 1$, where $K$, the truncation level, is the last cluster the algorithms learns. At the start of the iteration, $K$ is set to the number of tasks $M$, and so there is no loss of generality with this approach. We also make an additional computational enhancement to the algorithm by removing clusters with a task-assignment probability of less than machine epsilon; this allows for faster convergence.

The fully-factorized family of variational distributions is

$$q(Z) = \left[ \prod_{m=1}^M q_{c_m}(c_m) \right] \cdot \left[ \prod_{k=1}^K q_{v_k}(v_k) \right] \cdot q_\alpha(\alpha) \cdot \left[ \prod_{k=1}^K q_{w^*_k}(w^*_k) \right].$$
This variational distribution approximates the posterior so that the KL-divergence between the variational distribution $q(Z)$ and the true posterior is minimized. Minimizing the KL-divergence is equivalent to maximizing the following variational objective (ELBO):

$$\mathcal{L} = \mathbb{E}_q[\log P(y^{(m)}_m, Z)] + H(q),$$

where $H(q)$ is the entropy of the variational distribution $q$. Following the standard results of variational inference, the closed-form updates for the variational parameters are obtained, see Xue et al. for equations [55]. The variational objective is optimized using coordinate ascent variational inference (CAVI), where each factor of the variational distribution and the variational parameter of the sigmoid function $\xi$ is iteratively re-estimated conditioned on the current estimate of all others until convergence. The algorithm is summarized in Algorithm 3.

### 3.3.2 Prediction

The equation to predict a new test sample is as follows:

$$P(y^{(m)}_\star = 1 | c_m, \{w^*_m\}_{m=1}^K, x^{(m)}_\star) = \sum_{k=1}^K c_{m,k}^{} \sigma(w^*_k^T x^{(m)}_\star),$$

where $\sigma$ is the sigmoid function of a logistic regression classifier. After integrating over the learned variational distributions on the classifier’s weights $q_{w^*_k}(w^*_k)$ and tasks’ cluster assignments $q_{c_m}(c_m)$, we get

$$P(y^{(m)}_\star = 1 | x^{(m)}_\star, \{\phi_{m,k}\}_{k=1}^K, \{\theta_k\}_{k=1}^K, \{\Gamma_k\}_{k=1}^K)$$

$$= \sum_{k=1}^K \phi_{m,k}^{} \int \sigma(w^*_k^T x^{(m)}_\star) N_d(\theta_k, \Gamma_k) dw^*_k$$

There is no analytical form for the integral above. Therefore, the prediction function uses an approximation of the integral derived in Mackay et al. [57], resulting in the
CHAPTER 3. WELL-BEING PREDICTION WITH MULTI-TASK LEARNING

Input: Well-being features \( x \), labels \( y \), hyper-parameters \( \tau_1, \tau_2, \mu, \Sigma \)

Output: Variational parameters: \( \phi, \Gamma, \theta \)

Initialize variational parameters

while not converged do

  for \( m \leftarrow 1 \) to \( M \) do

    for \( k \leftarrow 1 \) to \( K \) do

      Update variational parameter \( \phi_{m,k} \),

      for cluster assignment \( c_m \sim \text{Mult.}(1; \phi_{m,1}, \cdots, \phi_{m,K}) \)

    end

  end

  for \( m \leftarrow 1 \) to \( M \) do

    Update variational parameter \( \varphi_{1,k} \),

    for \( k \leftarrow 1 \) to \( (K - 1) \) do

      for \( i \leftarrow (k + 1) \) to \( (K - 1) \) do

        Update variational parameter \( \varphi_{2,k} \),

        for the probability of belonging to cluster \( k \), \( v_k \sim \text{Beta}(\varphi_{1,k}, \varphi_{2,k}) \)

      end

    end

  end

Update variational parameter \( \tau_1 \),

for \( k \leftarrow 1 \) to \( K - 1 \) do

  Update variational parameter \( \tau_2 \),

end

Which updates \( \alpha \sim \text{Gamma}(\tau_1, \tau_2) \)

for \( m \leftarrow 1 \) to \( M \) do

  for \( k \leftarrow 1 \) to \( K \) do

    Update variational parameters \( \theta_k, \Gamma_k \)

    for the weights of the LR classifier \( w_k^* \sim \mathcal{N}_d(\theta_k, \Gamma_k) \)

  end

end

for \( m \leftarrow 1 \) to \( M \) do

  for \( n \leftarrow 1 \) to \( N_m \) do

    for \( k \leftarrow 1 \) to \( K \) do

      Update auxiliary variable \( \xi_{m,n} \)

    end

  end

end

return: \( \phi, \Gamma, \theta \)

**Algorithm 3:** CAVI algorithm for hierarchical Bayesian logistic regression.
following:
\[
P(g^*(m) = 1 \mid x^*(m) \{\phi_{m,k}\}_{k=1}^{K}, \{\theta_k\}_{k=1}^{K}, \{\Gamma_k\}_{k=1}^{K}) \\
\approx \sum_{k=1}^{K} \phi_{m,k} \sigma \left( \frac{\theta_k^T x^*(m)}{\sqrt{1 + \pi x^*(m)^T \Gamma_k x^*(m)}} \right),
\]
where \(\pi\) is the constant approximately 3.1416 (not the parameter of the Dirichlet process).

### 3.4 Evaluation

To ascertain the performance benefits of MTL, we compared it to three approaches. First, we compared HBLR to its single task learning (STL) equivalent, logistic regression. Secondly, to determine whether personalization via MTL has a performance advantage over simply using MTL, we explored multitasking over the well-being measures. In this approach, we treated predicting mood, stress, and health as the related tasks; we call this well-being-as-tasks. Note that this well-being-as-tasks approach to MTL is similar to that taken in the affective computing literature \([26–29]\). Finally, we compared HBLR performance results to a majority-class baseline classifier (the performance that can be expected from simply predicting the most frequent label in the training data). The models were developed in python and all the source code is available online.\(^1\)

#### 3.4.1 Experiments

HBLR automatically learns soft-clustering across the participants and as stated earlier, we hypothesize that these clusters provide valuable insights into the groups of people in them. A participant \(m\) may have some degree of membership in more than one cluster, as defined by \(\phi_m\) (the vector of probabilities of person \(m\) belonging to the \(K\) clusters),

\(^1\)https://github.com/mitmedialab/personalizedmultitasklearning
Table 3.2: Prediction performance (Accuracy, AUC) of the majority classifier, STL, MTL well-being-as-task, and MTL users-as-task approaches. Bold entries represent significant improvements over the other approaches.

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Mood</th>
<th>Stress</th>
<th>Health</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>Majority class</td>
<td>50.4%, .500</td>
<td>50.7%, .500</td>
</tr>
<tr>
<td>STL</td>
<td>LR</td>
<td>56.9%, .569</td>
<td>59.4%, .594</td>
</tr>
<tr>
<td>MTL-well-being</td>
<td>HBLR</td>
<td>58.3%, .583</td>
<td>57.8%, .578</td>
</tr>
<tr>
<td>MTL-participants</td>
<td>HBLR</td>
<td>72.0%, .720</td>
<td>73.4%, .734</td>
</tr>
</tbody>
</table>

so we define a matrix $P \in \mathbb{R}^{M \times T}$, where $M$ is the number of participants or tasks and $T$ is the number of pre-study measures (such as Big Five personality, PSS, etc.) mentioned in Section 3.2.3. $P_{m,t}$ represents participant $m$’s score on measure $t$. Using $P$, we then computed a score representing the average value of each pre-study measure for each cluster, as follows:

$$Q_{k,t} = \frac{\sum_{m} P_{m,t} \phi_{m,k}}{\sum_{m} \phi_{m,k}},$$

where $Q \in \mathbb{R}^{K \times T}$ and $K$ is the number of clusters learned by the HBLR model. $Q_{k,t}$ represents a weighted average of a cluster $k$ pre-study trait $t$, where the weights are the degree of membership of each participant in that cluster.

To test whether a cluster’s $Q_{k,t}$ value is significantly different from the entire group average, we used a one-sample $t$-test, with a Bonferroni correction, to compare $Q_{k,t}$ to the values of the reported measure $t$.

3.5 Results

The prediction accuracy and area under the ROC curve (AUC) of the different approaches to well-prediction are shown in Table 3.2, and Figures 3.6 and 3.7.

As shown in Table 3.2, the accuracy obtained using traditional logistic regression STL classifier is poor, achieving accuracy values in the range of 56–60%; this is similar to...
prior work that trained STL classifiers to detect mood on a simplified version of this dataset [49]. The performance obtained when multitasking over the related outcome labels, i.e., mood, stress, and health is shown as \textit{MTL-well-being}. Evidently, this MTL approach does not significantly enhance performance when compared to the STL results, achieving values in the range of 55–59%. One reason for this could be that the outcome labels are not close enough to each other to benefit from sharing parameters. We
have therefore demonstrated that at least for this data, MTL alone is not sufficient in improving future well-being prediction.

Using MTL to account for individual differences is what makes the difference. The performance results for the personalized MTL approach is shown as MTL-participants in Table 3.2 and in Figures 3.2 and Figures 3.6. The results show that using MTL to personalize over people (MTL-participants) provides significant improvements to well-being prediction performance. The improvements in accuracy over the non-personalized MTL-well-being approach range from 13–21%. McNemar tests with a Bonferroni correction applied within each label type revealed that the predictions of the personalized model significantly outperformed ($p < 0.05$) both the STL and MTL-well-being approaches.

### 3.5.1 Discussion

Figure 3.8 shows the soft clustering learned for predicting each of the three outcome labels. Each row shows one of the 104 participant’s degree of membership in each cluster. There were 4, 3, and 17 major clusters learned in predicting mood, stress, and health, respectively.

![Figure 3.8: Resulting soft clustering ($\phi$) learned when predicting the different labels (mood, stress, and health) for the 104 participants.](image-url)
As discussed in section 3.3.1, each cluster’s weight vector $w_k^*$ is drawn from a multivariate normal distribution. Figure 3.9 shows an example of the different marginal distributions the model learned over a single feature (total number of screen-on events (5PM–Midnight)) for the four mood clusters. We note that for this feature, cluster 0 and cluster 1 have very different distributions on the logistic regression weights, while cluster 2 and 3 have similar distributions. For example, Figure 3.9 shows us that cluster 0 places a negative weight on the feature while cluster 1 places a positive weight on the same feature. Thus, when participants who belong almost exclusively to cluster 0 use their phone excessively in the evening, the model will be more likely to predict a sad day tomorrow. In contrast, the model is more likely to predict a happy day tomorrow for participants belonging almost exclusively to cluster 1 based on the same behavior.

Figure 3.9: Distribution of HBLR weights on the total number of screen-on events (5PM–Midnight) feature for each cluster when predicting tomorrow’s mood

However, because participants do not belong exclusively to one cluster, the marginal distribution over a weight parameter for a given participant can be more complex than a multivariate normal. Figure 3.10 shows an example of the weight distributions for 3
different participants. Participant 3 is almost exclusively in cluster 1, participant 5 has membership in clusters 0, 1, and 2, and participant 31 is almost exclusively in cluster 2. For Participant 5, the model constructed a bimodal distribution over the weight by combining the distributions of multiple clusters. Thus, the model is able to customize the decision boundary for each person while still clustering the participants into similar representations.

We are also interested in determining if the clusters learned by the HBLR model differ significantly in terms of the typical personality or mental health scores of the participants. Following the procedure outlined in Section 3.4.1, we computed the matrix $Q$, then conducted significance tests to determine if there were significant differences among the clusters for some of the traits. We note that the model did not use any of the data in the pre-study survey during training and so, the model does not know a priori which traits are relevant to the clusters. Below, we discuss the results of these
Table 3.3: Computed pre-study measures for the HBLR mood prediction clusters. Bold entries represent significant differences from the sample average.

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Pre-study measure</th>
<th>All participants</th>
<th>Cluster $Q_{k,m}$</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Percent happy days</td>
<td>49 ± 37</td>
<td>56</td>
<td>-1.86</td>
<td>&gt; .10</td>
</tr>
<tr>
<td></td>
<td>Judging</td>
<td>61 ± 21</td>
<td>73</td>
<td>-7.69</td>
<td>&lt; .001</td>
</tr>
<tr>
<td></td>
<td>Sensing</td>
<td>47 ± 20</td>
<td>57</td>
<td>-7.22</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>1</td>
<td>Percent happy days</td>
<td>49 ± 37</td>
<td>55</td>
<td>-1.81</td>
<td>&gt; .10</td>
</tr>
<tr>
<td></td>
<td>PSQI</td>
<td>4.7 ± 2.3</td>
<td>4.1</td>
<td>3.48</td>
<td>&lt; .01</td>
</tr>
<tr>
<td>2</td>
<td>Percent happy days</td>
<td>49 ± 37</td>
<td>41</td>
<td>2.29</td>
<td>&gt; .10</td>
</tr>
<tr>
<td></td>
<td>Agreeableness</td>
<td>50 ± 28</td>
<td>43</td>
<td>3.63</td>
<td>&lt; .01</td>
</tr>
<tr>
<td>3</td>
<td>Percent happy days</td>
<td>49 ± 37</td>
<td>78</td>
<td>-8.00</td>
<td>&lt; .001</td>
</tr>
<tr>
<td></td>
<td>Extraversion</td>
<td>49 ± 30</td>
<td>76</td>
<td>-13.1</td>
<td>&lt; .001</td>
</tr>
<tr>
<td></td>
<td>State anxiety</td>
<td>38 ± 10</td>
<td>30</td>
<td>10.9</td>
<td>&lt; .001</td>
</tr>
<tr>
<td></td>
<td>Trait anxiety</td>
<td>43 ± 10</td>
<td>36</td>
<td>9.85</td>
<td>&lt; .001</td>
</tr>
</tbody>
</table>

computations for some traits in the mood and stress clusters. We do not show the same analysis for health since it is impractical to do the analysis on the 17 different clusters.

Table 3.3 shows the relevant trait values for the mood clusters ($Q_{k,m}$) and the mean ± standard deviation for those traits computed over all participants in the dataset. According to these findings, in predicting mood, the HBLR model learned clusters that can be characterized as Cluster 0: judging and sensing personality types; Cluster 1: people with better than average sleep quality (PSQI); Cluster 2: agreeable people, and Cluster 3: happy extroverts with low state and trait anxiety. This suggests that these traits are highly relevant for predicting how a person’s mood will change given the input features. Bower et al. showed that poor sleep quality has a negative effect on mood [58], and so it is possible that the normally high sleep quality of participants in cluster 1 makes their mood sensitive to sleep disturbances. Personality, particularly extraversion and neuroticism, is a strong prediction of an individual’s emotional style [59]. Research has shown that extraversion is strongly associated with a positive emotional style [60, 61]. This could explain why cluster 3 which had a significantly higher proportion of extroverts when compared to the group average, also had higher percent of happy days.
We discover interesting results when we compare these findings to the average value of the weights learned for these clusters, as shown in Figure 3.11. The positive label is “happy” so features with positive mean weights contribute to being happy tomorrow, while features with negative mean weights contribute to being sad tomorrow. For example, the “agreeable” cluster (cluster 2) placed high weights on four social interaction features; this is consistent with research indicating that people with an agreeable personality type value social interactions with others [62]. In contrast, the “high sleep quality” cluster (cluster 1) placed negative weights on features related to SMS use in the evening. Finally, we observe that the “judging and sensing” cluster (cluster 0) has a positive association with exercise, but a negative association with time spent on campus.

Another observation is that the HBLR model is not making better predictions by simply clustering happy or sad people together, most of the clusters have a mix of peoples’ mood. As shown in Table 3.3, three of the clusters do not differ significantly from the group average in percent of happy days, although cluster 3 (extroverts with low state and trait anxiety) corresponds to particularly happy participants.

For the stress model, we show the results of the same analysis of HBLR cluster pre-study

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Pre-study measure</th>
<th>All participants</th>
<th>Cluster</th>
<th>( Q_{k,m} )</th>
<th>( t )</th>
<th>( p )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Percent calm days</td>
<td>48 ± 38</td>
<td>46</td>
<td>.492</td>
<td>&gt; .60</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Percent calm days</td>
<td>48 ± 38</td>
<td>55</td>
<td>-1.88</td>
<td>&gt; .10</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GPA</td>
<td>4.4 ± 0.61</td>
<td>4.6</td>
<td>-3.95</td>
<td>&lt; .001</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Conscientiousness</td>
<td>51 ± 28</td>
<td>58</td>
<td>-3.43</td>
<td>&lt; .01</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Percent calm days</td>
<td>48 ± 38</td>
<td>39</td>
<td>2.32</td>
<td>&gt; .10</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Extraversion</td>
<td>49 ± 30</td>
<td>58</td>
<td>-4.50</td>
<td>&lt; .001</td>
<td></td>
</tr>
<tr>
<td></td>
<td>BMI</td>
<td>24 ± 4.4</td>
<td>25</td>
<td>-4.09</td>
<td>&lt; .001</td>
<td></td>
</tr>
<tr>
<td></td>
<td>PCS</td>
<td>58 ± 4.2</td>
<td>57</td>
<td>3.77</td>
<td>&lt; .01</td>
<td></td>
</tr>
</tbody>
</table>

Table 3.4: Computed pre-study measures for the HBLR stress prediction clusters. Bold entries represent significant differences from the sample average.
Sec. 3.5. Results

The positive label is “calm” so features with positive mean weights contribute to being relaxed tomorrow, while features with negative mean weights contribute to being stressed tomorrow. Here, none of the clusters differed significantly from the group average in terms of the percentage of calm days, implying that the clusters are diverse in the representation of stressed or calm people. For cluster 0, we did not detect any significant differences from the group average. Cluster 1 however represents conscientious people with a high GPA. It is fitting that this cluster is relevant to predicting stress, since conscientious students concerned about their grades are likely to have strong stress reactions in an academic environment. As shown in Figure 3.12 this cluster places a positive weight on the “likelihood of day” feature, which is

![Figure 3.11: Mean feature weights for mood clusters in HBLR model. The positive label is happy and the negative label is sad.](image)

\[\text{Draft: August 28, 2018}\]
a measure of how routine the participants’ location patterns were that day. A higher value on this feature implies that the participant mainly travels to typical school and home locations. The stress cluster 2 represents students who are extroverted, and have slightly increased BMI and lowered physical health. In Figure 3.12, we see that cluster 2 has highly positive mean feature weights on the SMS feature consistent with the personality trait of extraversion. On the contrary, cluster 1 has highly negative weights on the social SMS features, meaning more SMS use for these participants increases the likelihood of predicting a stressful day tomorrow. One possible explanation is that these conscientious, high GPA students become stressed out by having to add-on a social life.

![Figure 3.12: Mean feature weights for stress clusters in HBLR model. The positive label is calmness and the negative label is stress.](image-url)
to their academic goals.

3.6 Summary

We present a personalized MTL model that predicts the future well-being of participants in the SNAPSHOT study. Our model uses the hierarchical Bayesian logistic regression framework that clusters similar participants and learns classification boundaries over these clusters. We predict future well-being without requiring a history of collected well-being labels for each person. We use data collected as participants go about their daily lives through surveys, wearable sensors, weather monitoring, and smartphones, and thus are relevant to use in a real-world well-being prediction system. In addition, we provide insights into the relationship between the participants in the clusters and the pre-study traits. These insights emphasize the need to train personalized models that can account for individual differences.

There are a few ways our work can be advanced. For example, an extension to the HBLR model is able to learn a new task (person) previously unseen by the model through transfer of relevant information from past tasks. This is done by sampling the learned posterior of the previous tasks using MCMC [55]. In future work, we can test the accuracy of well-being prediction on an entirely new person. This will be relevant to any real world application that aims to intervene on or prevent a decline of well-being across a broader but similar population. From a modeling perspective, we currently do not take into account the effect time has on a person’s well-being. The reason for this is that we do not have enough data to separately model 30 days of each of the 104 participants. With more data collected and with time-series models like recurrent neural networks and LSTM’s, we can build a well-being forecast model akin to that of a weather forecast.

DRAFT: AUGUST 28, 2018
DIFFERENT combinations of behavioral factors can have varying effects on an individual’s well-being. This chapter aims to provide understanding on how combinations of health behaviors affect self-reported stress in a cohort of college students in the SNAPSHOT study. Following the latent Dirichlet allocation (LDA) framework, 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. We use supervised latent Dirichlet allocation (sLDA) to model the observed behaviors, and we apply variational inference to uncover these latent patterns. We show that these latent patterns are indeed predictive of self-reported stress when compared to patterns learned in an unsupervised way. We also investigate the behavioral factors present in these patterns and uncover how these factors work together to influence the well-being of individuals.
4.1 Introduction

In recent years, there has been a shift in the psychological research literature from an emphasis on dysfunction to a focus on well-being and positive mental health [63]. As a result, enhancing well-being in individuals has become a viable approach to improving health, in addition to treating disorders when present. Well-being is a dynamic and modifiable process that can be influenced by social, economic, and environmental factors. However, individuals also influence their well-being and thus have a personal responsibility to maintain and improve it. In order to empower individuals seeking to improve their well-being, there is a need to develop personalized models that are able to understand factors that contribute to the state of well-being, and provide actionable evidence-based insights to these individuals.

As discussed in Chapter 3, measuring and predicting future well-being is a difficult task, which makes it challenging to develop interventions that can improve it. Individuals’ state of health and well-being can be determined by everyday choices, such as exercising, social interactions, or lack of sleep, and the ability to influence these choices can provide an opportunity to improve their well-being. Studies have shown the effect of some of these behaviors on well-being. For example, in an 8000-person study, Stubbe et al. showed that exercise participation is associated with higher levels of well-being [64]. 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 [65]. However, these studies investigate the standalone effects of each behavior on well-being, and it remains to be seen how varying combinations of health behaviors affect an individual’s well-being.

In this chapter, we focus on assessing how combinations of modifiable health behaviors contribute to self-reported stress among a college student population in the SNAP-
SHOT study. Similar to document modeling with topic models, we posit that there are latent patterns that generate the behaviors we observe; these patterns are probability distributions over the health behaviors. Additionally, we have real-valued labels that are the students’ best estimate of their well-being state, reported as perceived stress. We use sLDA to uncover these latent patterns by jointly modeling the observed health behaviors and the self-reported stress labels. The techniques of variational expectation-maximization are used to estimate the posterior probability of these patterns given the observed data, and we show that the inferred patterns are best predictive of the self-reported stress labels.

4.1.1 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. (and as partly discussed in Chapter 3) use MTL approaches to predict future well-being, achieving prediction accuracies in the range of 72–83% [21]. 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 [19]. While predicting an individual’s well-being is important, it is equally important to be aware of the set of behaviors that specifically contribute to that person’s well-being. Understanding these factors will do more to guide prevention and early intervention efforts.

While the studies above have focused only on predicting well-being, others have used 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 [66]. In addition, LDA or 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 [39]. Phung et al. used LDA to infer the sequences of places a user routinely visits [67]. LDA based models have been used to discover daily routine patterns with data from wearable sensors [68,69]. Also, Ferrari et al. used topic models to obtain mobility behavior using location data collected from Google Latitude [70]. 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.

### 4.1.2 Our Framework

We propose a framework for uncovering patterns of modifiable human behaviors that can influence the well-being of individuals. The current approaches to modeling human behavior either summarize routine behavior or model physical activity. 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 to modeling human behavior is important for a number of reasons. First, using modifiable behaviors is significant because the insights discovered can provide individuals with an opportunity to change their behavior and consequently improve their well-being. In Section 4.2 we introduce a method of extracting these behaviors from multi-modal data and binning them to ensure that they are descriptive of a participant’s day. Second, in learning patterns of health behaviors, we present an interpretable Bayesian model that creates meaningful and interpretable representations of these health behaviors. Representation learning can be characterized as finding repre-
sentations of data that contain useful information towards some goal [22]. In this work, our goal is to create representations that are useful for predicting affective states and helpful in recommendation tasks. Third, it is important to study how combinations of behaviors may influence well-being because well-being is not affected by stand-alone behaviors. As previously discussed, past research tends to learn associations between single health behaviors and well-being. But well-being is complex to model and it is important to examine how groups of behaviors work together to influence it.

To that end, models that learn patterns of human behaviors with a specific goal of improving well-being are important. Using multi-modal data from smartphones, surveys, and wearable sensors, we employ supervised topic models to uncover latent patterns best predictive of well-being (measured as perceived stress or calmness). We show that combining these sources of data provide us with better summaries of daily human behaviors. We also learn what set of behaviors work together to influence well-being and the changes that occur as these patterns of behaviors move from a positive to a negative association. This knowledge provides an opportunity to get insights into influential variables that can become candidates for future causal studies.

### 4.2 Data Representation

We use data collected as part of the 30-day snapshot study [19]. First, we discuss the modifiable behaviors that were selected from the data, and then we summarize the structure of the two sources of data to the model: the response variables, and the input data vector, which we call “bag-of-behaviors”.

**Modifiable Behaviors:** We define modifiable behaviors as actions that can be controlled by the study participants, such as sleep duration, phone usage, etc. Information about modifiable behaviors was extracted from smartphone logs, wearable sensors, GPS
coordinates and behavioral surveys obtained from the participants during the 30-day study. We note that 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 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 topic models. 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. See Table 4.1 for the list of behaviors selected and their various bins.

After selecting the modifiable behaviors, we convert the data into input features for the model. The two types of input to the model are:

- **Bag-of-behaviors.** 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 not present in that day and 1 means that the modifiable behavior is present. 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.

- **Response variables.** As mentioned previously, each day, 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 4.1 shows the
Table 4.1: Selected modifiable behaviors, the bins and total number of days the behaviors were present in the data.

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

distribution of these response variables in the dataset. As a pre-processing step, we log-transformed the labels to achieve approximate normality as required by the

\(^{1}\)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. \[71\], we calculate sleep regularity as

\[
\frac{\tau}{T-\tau} \left( \int_0^T s(t) s(t+\tau) \, dt \right) - 1,
\]

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.
model (see Section 4.3). These real-valued log-transformed stress labels, \( y \in \mathbb{R} \), are used as our response variables in the supervised model. The goal is to find latent patterns that are best predictive of these 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-transform of 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 the results for brevity.

![Figure 4.1: Distribution of self-reported stress/calmness response variables. The mean of the response variables, 53.43, is shown in red.](image)

---

**Draft: August 28, 2018**
4.3 Supervised Latent Dirichlet Allocation

Topic models were developed to learn the probability distributions of words in collections of documents in a corpus in an unsupervised manner [20]. However, the goal of this work is to use topic models 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. If we use an unsupervised model to model health behaviors, we will likely discover broad patterns that run throughout the days. But supervision is required if we want patterns that are predictive of an outcome such as well-being. This is because health behaviors are dynamic and vary over the time of day and across people. In this work, we use the supervised latent Dirichlet allocation (sLDA) model, proposed by Blei and McAuliffe [72], to assess health behaviors. In this application, each day has a response variable (the self-reported stress label) external to the behaviors present, and the aim is to find patterns that are best predictive of the response variable.

We use $d \in 1, \cdots D$ to index the number of days present in the dataset, $n \in 1 \cdots N$ to index the number of behaviors observed on day $d$, and $k \in 1, \cdots K$ to index the number of latent patterns. In the model, each day is represented as a collection of discrete random variables $x_{1:N}$; these are the observed behaviors. The unobserved patterns $\beta_{1:K}$ are probability distributions over the set of health behaviors observed in the dataset. We assume that the same set of $K$ latent patterns, $\beta_{1:K}$, 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 of these patterns give rise to different combinations of behaviors each day, and thus to different stress/calmness responses.

Following sLDA, we jointly model the days $x$, and the response variables $y$ to find
CHAPTER 4. ASSESSMENT OF HEALTH BEHAVIORS WITH SUPERVISED TOPIC MODELS

latent patterns that are best predictive of the stress responses of unseen days. Given the Dirichlet parameter, $\alpha$, response parameters $\eta$ and $\delta$, and a fixed latent pattern $\beta_{1:K}$, the behaviors in each day and corresponding response variable come from the following data-generative process:

$$
\theta | \alpha \sim \text{Dir.}(\alpha) \quad \text{(pattern proportions)}
$$

$$
z_n | \theta \sim \text{Mult.}(\theta) \quad \text{(pattern assignment for each behavior)}
$$

$$
x_n | z_n, \beta_{1:K} \sim \text{Mult}(\beta z_n) \quad \text{(observed behavior)}
$$

$$
y | z_{1:N}, \eta, \delta \sim \mathcal{N}(\eta^T \bar{z}, \delta) \quad \text{(response variable)}
$$

where

$$
\bar{z} := \frac{1}{N} \sum_{n=1}^{N} z_n
$$

represents the unobserved empirical pattern frequencies that occurred on day $x_n$. See Figure 4.2 for a graphical model representation of the sLDA.

The response variable $y$ is modeled as a Gaussian linear model, with regression coefficients $\eta$, and model covariates $\bar{z}$. By regressing $y$ on the empirical frequencies $\bar{z}$, we effectively tie the behaviors expressed on day $d$ to the stress response reported on 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 $y$. The mean and variance parameters of the Gaussian are $\eta^T \bar{z}$ and $\delta$ respectively.

4.3.1 Variational Expectation-Maximization

The latent variables in this model are the pattern proportions $\theta$ and each behavior’s pattern assignment $z_n$. In this model, $\beta_{1:K}$, $\eta$, and $\delta$ are constants to be estimated, and $\alpha$ is a hyper-parameter. To learn the latent variables from the training data $\{x_{d,1:N}, y_d\}_{d=1}^{D}$,
we need to approximate the posterior $P(\theta, z_{1:N} | x_{1:N}, y, \alpha, \beta_{1:K}, \eta, \delta)$, which we do via variational inference. Following Blei and McAuliffe [72], the fully factorized variational distribution that approximates the posterior for each day is specified as

$$q(\theta, z_{1:N} | \gamma, \phi_{1:N}) = q(\theta | \gamma) \prod_{n=1}^{N} q(z_n | \phi_n),$$

where $\gamma$, a K-dimensional Dirichlet parameter vector, learns the latent pattern proportions for each day $d$; and $\phi_n$, a K-dimensional categorical parameter vector, learns the assignment probabilities of behavior $n$ to each of the $K$ topics. $Z_n$ is a K-dimensional indicator vector, with the non-zero entry corresponding to the topic assigned to behavior $n$; thus

$$E[Z_n] = q(z_n) = \phi_n,$$

and

$$E[\bar{Z}] = \bar{\phi} = \frac{1}{N} \sum_{n=1}^{N} \phi_n.$$  

After learning the posterior, we estimate the model parameters $(\beta_{1:K}, \eta, \delta)$ using variational expectation maximization (EM). The parameters are chosen such that the fol-
CHAPTER 4. ASSESSMENT OF HEALTH BEHAVIORS WITH SUPERVISED TOPIC MODELS

**Input**: Behaviors $x$, labels $y$, hyper-parameter $\alpha$ and number of topics $K$

**Output**: Parameters: $\phi, \gamma, \beta, \eta, \delta$

Initialize variational parameters: $\alpha$ and $\gamma$, and model parameters: $\beta, \eta, \delta$

**while** not converged **do**

**for** $d \leftarrow 1$ to $D$ **do**

**while** not converged **do**

**for** $n \leftarrow 1$ to $N$ **do**

**for** $k \leftarrow 1$ to $K$ **do**

Update variational parameter $\phi_{n,k}$

end

Normalize $\phi_n$ to sum to 1

end

Update variational parameter $\gamma_d$

end

Estimate model parameters $\beta, \eta, \delta$

Normalize each pattern, $\beta_k$ to sum to 1

end

**return**: $\phi, \gamma, \beta, \eta, \delta$

Algorithm 4: Variational expectation-maximization algorithm for supervised latent Dirichlet allocation.

The following ELBO is optimized:

$$
\mathcal{L}(\alpha, \beta_{1:K}, \eta, \delta; \{x_{d,1:N}, y_d\}_{d=1}^D) = \sum_{d=1}^D \mathbb{E}_d[\log P(\theta_d, z_{d,1:N}, x_{d,1:N}, y_d)] + H(q_d),
$$

where $H(q_d)$ is the entropy of each day’s variational distribution $q_d$.

In the E-step, the posterior distribution for each day-response pair is estimated by updating $\gamma$ and each of the $\phi_n$. And in the M-step, the model parameters, $\beta_{1:K}, \eta, \delta$ are updated to optimize the ELBO. The update equations are outlined in Blei and McAuliffe [72]. The algorithm is summarized in Algorithm 4.

### 4.3.2 Prediction

Once we have a fitted model, the pattern proportions $\gamma_{\text{new}}$, and behavior assignment probabilities $\phi_{\text{new}} = \{\phi_n\}_{n=1}^{N_{d,\text{new}}}$, of any previously unseen day are iteratively
re-estimated conditioned on the current estimate of the model parameters $\beta, \eta, \delta$ until convergence. Then, the day’s response is predicted as follows:

$$E[Y \mid x_{1:N}, \alpha, \beta_{1:K}, \eta, \delta] \approx \eta^T E_q[\tilde{Z}] = \eta^T \tilde{\phi}_{\text{new}},$$

where the expectation is taken with respect to the variational distribution $q$. With this, the model can predict any new participant’s self-reported well-being and also learn the day’s latent patterns proportions.

### 4.4 Evaluation

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

- We ran a linear regression analysis on the $\tilde{\phi}_d$ learned from unsupervised LDA (which is described in Section 2.4). This is the same as using LDA topics as prediction features. We call this the “LDA + regression” approach.

- We performed an $L_1$-regularized least-squares regression analysis (LASSO), using each day’s empirical distribution over behaviors as the model covariates (i.e., normalizing the features to sum to one). The LASSO is widely used in high-dimensional problems because of its feature selection property. In addition to comparing its performance results with the sLDA, we are also interested in examining the features that were selected by the LASSO model for the regression task.

For the analysis, we initialized $\beta_{1:k}$ to randomly perturbed uniform topics, $\delta$ to the variance of the log-transformed stress response, and $\eta$ to a K-dimensional vector of zeros (initializing $\eta$ to a vector of ones provided comparable results). Also, we fixed $\alpha = 1/K$. In the E-step, we ran coordinate ascent variational inference (CAVI) until
the relative change in each day’s ELBO was less than 0.001. We ran the M-step until the relative change in the overall likelihood bound was less than 0.0001. For the LDA model, we initialized $\beta_{1,K}$ similarly and used the same convergence criteria and value for $\alpha$.

We applied 5-fold cross-validation to the dataset, and quantified the model’s performance in two ways. First, we computed a binary prediction accuracy on the held-out-fold predictions, which measured how well the learned patterns predicted high versus low self-reported stress. To convert to binary labels, true and predicted response variables equal to or greater than 50 were considered high stress values (label 1) and values below 50 were considered low stress values (label 0). We also computed binary prediction accuracy results using the mean of the population’s self-reported stress (mean $= 53.43$) as the threshold. Secondly, we measured the correlation between the held-out-fold real-valued prediction and true response variables. Also, we report the same performance metrics on the held-out-fold predictions for the LDA + Regression approach. We assessed the prediction quality of both models over different numbers of topics.

For the LASSO model, we used 5-fold cross validation to select the optimal regularization parameter and report the highest mean (± standard deviation) correlation and binary prediction accuracies achieved. We compare these values to the highest value sLDA achieved across the different numbers of topics.

We note that all analyses were repeated for self-reported happiness and self-reported health. We present the results for these analyses in Appendices C and D, respectively.
4.5 Results and Discussion

The average binary prediction accuracies for both thresholds, 50 and the mean stress response, are shown in Figures 4.3 and 4.4 respectively. The average F1 scores for both thresholds, 50 and the mean stress response, are shown in Figures 4.5 and 4.6 respectively. The average correlations between the real-valued true and predicted response values are shown in Figure 4.7. The sLDA model is 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. However, our results indicate that the sLDA model outperformed the LDA + Regression model on all five performance metrics. This implies that the sLDA model learns latent patterns that are better at predicting the self-report stress compared to the patterns learned in an unsupervised way.

Figure 4.3: Average binary prediction accuracy (over 5 runs, ± standard deviation) for sLDA and LDA + Regression models across different topics (threshold = 50).
**Figure 4.4:** Average binary prediction accuracy (over 5 runs, ± standard deviation) for sLDA and LDA + Regression models across different topics (threshold = 53.43, average of the participants self-reported stress).

**Figure 4.5:** Average F1 score (over 5 runs, ± standard deviation) for sLDA and LDA + Regression models across different topics (threshold = 50).
Figure 4.6: Average F1 score (over 5 runs, ± standard deviation) for sLDA and LDA + Regression models across different topics (threshold = 53.43).

Figure 4.7: Average correlation (over 5 runs, ± standard deviation) between the true and predicted response variables for sLDA and LDA + Regression models across different topics.

Table 4.2 shows the best mean (± standard deviation) of the performance metrics the LASSO achieved, and the highest value sLDA achieved across the different topics and
over five repetitions. For the binary prediction accuracy metric, sLDA had modest 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. The LASSO achieved a higher correlation value, but the difference is not statistically significant. Notwithstanding modest gains in prediction accuracy, we note the advantage that sLDA models the latent structure of the data that is useful for other purposes beyond prediction.

Table 4.2: Mean (± standard deviation) of binary accuracies, F1 scores, and correlation coefficients for sLDA and LASSO. Bold entries represent a statistically significant improvement of the sLDA model over LASSO (p < 0.05).

<table>
<thead>
<tr>
<th>Model</th>
<th>Binary Accuracy (thresh. = 50)</th>
<th>Binary Accuracy (thresh. = mean)</th>
<th>F1 Score (thresh. = 50)</th>
<th>F1 Score (thresh. = mean)</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>sLDA</td>
<td>60.5% (± 0.4)</td>
<td>58.4% (± 1.2)</td>
<td>0.72 (± 0.01)</td>
<td>0.66 (± 0.01)</td>
<td>0.26 (± 0.04)</td>
</tr>
<tr>
<td>LASSO</td>
<td>56.5% (± 1.0)</td>
<td>57.0% (± 1.2)</td>
<td>0.48 (± 0.02)</td>
<td>0.41 (± 0.03)</td>
<td>0.30 (± 0.03)</td>
</tr>
</tbody>
</table>

### 4.5.1 LASSO

As previously mentioned, the LASSO model is useful for 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 the participants in the study.

We are interested in the behaviors that were selected as best predictors of self-reported
Table 4.3: Behaviors selected by the LASSO model in decreasing order of importance.

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

stress or calmness in this cohort of students. Figure 4.8 shows the non-zero model coefficients for the LASSO model, and Table 4.3 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. 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.

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. While negative interactions 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 high predictors of stress. This is consistent with research that showed that students with low mental health, as measured by mental health composite scores (MCS), had a lower probability of interacting with electronic media prior to bedtime [19]. Also, the StudentLife study by Wang et al. found that students who
had frequent conversations in the evening were less likely to be stressed \( r = -0.386, p = 0.015 \) \[65\]. In addition to in-person interactions, we see that interaction with media (which could be entertainment) is a predictor of calmness.

**Bed time, 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 indicative of 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 \[65\]. Second, sleep regularity (SR), which captures changes in sleep timing on a day-to-day timescale while controlling for sleep duration, is a predictor of stress and calm. Figure 4.9 shows raster plots of an example of a regular and an irregular sleeper in our dataset. 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 \[73\]. We were also interested in seeing the effects of the previous two days sleep duration on today’s outcome, so we included the sleep durations 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 than 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.

**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 [74–77]. In contrast, we also found that not having any exercise was a high predictor of stress. It is difficult to speculate about cause and effect, but it is plausible that people who are stressed-out exercise less.

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, since the students may be drinking in the evening, the surveys may have been filled out soon after taking alcohol. And so we may be seeing the calming effect of 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 [78]. So we may be seeing the weekend effect mixed in with the alcohol effect.

In summary, we have seen that the LASSO model 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.

4.5.2 SLDA

The sLDA model learned latent patterns (topics) that are meaningful representations of health behavior. As a result, we would like to further analyze these patterns.

Since we are interested in good predictive patterns, we select $K = 11$ for our analysis.
because it provides us with the highest binary prediction accuracy (when the threshold is set to 50). Figure 4.10 shows $\beta$, the 11-topic sLDA model fit to our data. For each learned pattern, we highlight the seven health behaviors with the highest probability of occurrence (Appendix B has the 15 most probable behaviors in each pattern in Tables B.1 and B.2). The pattern with the most negative coefficient (bottom) has behaviors similar to the ones found to be most predictive of stress by the LASSO model: no exercise, caffeine consumption, very little social interaction, study duration of 4–6 hours, and negative interaction. The pattern with the most positive coefficient (top) has behaviors found to be most predictive of calmness by the LASSO model: high sleep regularity and going to bed around the average bedtime of the study population. We also highlight some new behaviors that were learned by the model for this calm pattern: 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). The patterns in-between have a mix of health behaviors, with the number of “positive” behaviors increasing as the pattern coefficients become more positively associated with calmness. This is consistent with our belief that varying combinations of health behaviors lead to different outcomes, and models that only learn prediction rules are insufficient to uncover these complex latent patterns.

One advantage the sLDA model has over the LASSO is that it learns the day-pattern probability distributions over each participant’s day ($\theta$, as estimated by the variational parameter $\gamma$). 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 illustrate how the sLDA model is able to capture the similarity in the behaviors of participants with comparable self-reported stress. Figure 4.11 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 4) dominantly expressed; participants with high self-reported stress have the most stress pattern (pattern 5) dominantly expressed; while participants in the mid-range of 55–65 self-reported stress have varying representations of patterns expressed in each day.

4.6 Summary

We proposed a novel approach to map multi-modal data collected in the wild to meaningful representations of health behavior. We used a supervised latent Dirichlet allocation model to learn patterns of health behavior that are predictive of stress. We compared the predictive performance of the patterns learned with sLDA to that of patterns learned with an unsupervised LDA, and we showed that the sLDA significantly outperforms the LDA in a binary prediction task, and the predicted responses are more correlated. This implies that unsupervised modeling only finds broad patterns that run through the days, but when prediction of an outcome such as well-being is needed, supervision is beneficial. We also compared the performance of sLDA to the LASSO model and found that the sLDA significantly outperforms the LASSO in a binary prediction task.

Furthermore, we analyzed the latent structure that was uncovered by the sLDA model and showed how specific behaviors are associated with stress prediction. Previous works have shown correlations with general health behaviors such as increasing or decreasing sleep duration, social interaction, or sleep regularity [65, 73], but using data from the SNAPSHOT study, we have shown how specific features such as going to bed at a particular time or a specific duration of sleep or exercise, etc., contribute to the well-being of college students. Other works have used topic models to model routine human behavior with a focus on activity recognition [39, 67, 70], but to the best of our knowledge,
no one has combined data from multiple sources when using topic models to model human behavior, while also focusing on well-being improvement.

Through this work, we have shown that the sLDA model finds informative structure, which produces patterns that are best predictive of stress, and connects with our intuitive understanding of health behaviors. However, there are limitations to consider and ways in which this work can be furthered. 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 to two hours earlier than your personal average bedtime (bedtime deviation −1 or −2) as a highly probable behavior. We believe that this effect of past days sleep behavior on today’s well-being also extends to other types of health behaviors. Thus, with the collection of more data, we can use a supervised variant of dynamic topic models [79] to capture the evolution of the latent behavioral patterns over time. This will provide us with information about how time-varying behaviors affect an individual’s well-being. Also, we can use a Bayesian tensor factorization method, like Schein et al. [80], to directly model the temporal latent factors present. Secondly, our work used self-reported stress labels as a proxy for stress measurement. However, these labels can be inconsistent across and within users. As a result, using more objective ways to measure the physiological stress in individuals could help strengthen the modeling of health behaviors related to stress and well-being.
Using data from the SNAPSHOT study, other models have been able to better approximate prediction functions. In Chapter 3 for example, the HBLR model achieved higher prediction accuracies. In addition, a personalized neural network framework used in Taylor et al. [21] achieved higher prediction accuracies. There are some reasons for these higher predictive performances. First, in both instances, other features in addition to modifiable behaviors were used to learn the prediction rule. This gave the models more information to learn how to distinguish between low or high stress days. However, for the work presented in this chapter, we included only modifiable behaviors, so that we can investigate how these behaviors work together to influence well-being and also, uncover possible candidates for future causal studies. Secondly, data corresponding to the middle 20% of the self-reported labels were thrown out in the pre-processing step of the HBLR and neural network models. This was done because the data corresponding to the middle 20% are ambiguous and may confuse any model learning to predict high and low stress, health, or happiness. Thus, throwing out this portion of data allows for an easier prediction problem. For the work in this chapter however, we do not throw out the middle 20% because we use a generalized linear regression framework to model the response variable \( y \) in order to directly predict the real-valued labels that were reported. This is valuable because we can directly predict a finer estimate of self-reported stress rather than just a binary category and these estimates can be useful for other purposes beyond binary prediction. For example, individuals might be interested in seeing the trends in their self-reported stress and whether or not it is a gradual or sudden change. Algorithms that only predict oscillating values between high and low stress may not be useful for this purpose.

Higher prediction accuracies notwithstanding and unlike the proposed sLDA model, the HBLR and neural network models do not directly (1) reveal personalized combinations of behaviors that are most influential of a participant’s well-being, and (2) create
personalized evidence-based recommendations to individuals looking to improve their well-being (see Chapter 5 for a discussion on how we can leverage insights from the sLDA model to create personalized recommendations).
### Figure 4.8: 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.
Figure 4.9: Raster plots showing a regular sleeper, SR = 0.91 (top), and an irregular sleeper, SR = 0.68 (bottom).
**Figure 4.10:** An 11-topic 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.
Figure 4.11: 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 4 is the most calm pattern, and pattern 5 is the most stress pattern.
Chapter 5

Recommending Health Behaviors:

Two Case Studies

In this chapter, we use a case study approach to illustrate how the sLDA model developed in Chapter 4 can help provide actionable insights and recommend healthy behaviors, using data from two participants in the SNAPSHOT study.

5.1 Introduction

Recommendation systems have been around for over a decade and have mostly been used in the area of e-commerce and Internet-based services. Their primary goal is to help users synthesize all the information available and deliver personalized content, either in the form of movies to watch next on MovieLens [81], relevant news articles to read on Google [82], or products to buy on Amazon [83]. In these applications, the recommendation system is able to efficiently represent user behavior and how it relates to the items of interest being recommended. It does this by predicting or estimating a user’s predilection for new items based on past preference and similarity to other users.
5.2 Related Work

There is recent and growing interest in applying recommendation system approaches to health care problems, although work done in this area is still limited. Ge et al. used a content-based method to recommend personalized healthy recipes for meals after users logged in and indicated meal and recipe preferences [84]. Chomutare et al. used a hybrid model to recommend peers to users who needed peer-support to manage chronic diseases like diabetes [85]; their approach was inspired by the “patient-like-me” concept [86]. These methods require the user to extensively rate each of the previously seen items, and then recommendations of new items are made based on user ratings.

Rabbi et al. used a multi armed bandit algorithm to make sequential decisions about which food a user should eat or which physical activity a user should perform in order to remain healthy [87]. In this work, users were required to enter data manually about their eating habits, and the application mined users’ physical activities from logged mobility patterns. The suggestions were contextual in nature and only depended on the user’s history. Farrell et al. used a heuristics based approach to make recommendations on how to use physical activity and proper meal choices to reduce weight [88]. The method used data from the user’s daily meal and physical activity logs. A major drawback of these approaches however, is that users can easily become dissatisfied with a system that repeatedly makes familiar suggestions and recommends routine behaviors. Even though personalization is important, there could be value in learning and recommending healthy behaviors from similar individuals in a population.

To that end, we would like to predict a person’s well-being and then recommend healthy behaviors that can help improve that person’s state of well-being, if needed. However, this is difficult to implement, as it would likely require the collection of a plethora of data from that individual, and a manual rating for each behavior (not just a score at
the end of day). Nevertheless, meaningful representations of a person’s health behavior learned by the sLDA model affords us an opportunity to suggest behaviors that may improve well-being. This would be based on the model’s ability to find “healthy” participant-days that are similar to the day in consideration. In this chapter, we use two case studies to illustrate how information from the sLDA model can be used to provide actionable insights to individuals looking to improve their well-being.

5.3 Health Behavior Recommendation

In making health behavior recommendations, it is important to take the following considerations into account. First, recommended behaviors should be feasible, i.e., these behaviors should only require slight modifications to an existing routine, so as not to burden an already stressed person. This would mean recommending behaviors from similar days with better outcomes. For each day, sLDA learns personalized probability distributions $\gamma_{\text{new}}$ over the latent patterns $\beta_{1:K}$, and thus is able to capture similarities in the behaviors present in participant-days. Secondly, since the stress scores are reported in the evening, it is important to recommend behaviors that can affect a future outcome, such as bedtime, length of sleep, social interactions the following day, etc. Finally, when finding similar days, it is beneficial to include data from other participants. This is because currently available data is insufficient to train an sLDA model for each person. Also, if a person only has unhealthy behaviors, it would be impossible to recommend a healthy behavior. Including behaviors from others also ensures that participants do not get bored from being recommended recycled behaviors.

As discussed in Section 4.3.2, given data from a previously unseen day, $x_{1:N}$, the fitted sLDA model is able to learn the probability distribution of the latent patterns present, $\gamma_{\text{new}}$, and the topic assignments for each behavior $n$, $\{\phi_n\}_{n=1}^W$, by iteratively
re-estimating the update equations. Once these latent variables are learned, the model can now compute the expected stress score for the day, \( E[Y \mid x_{1:N}, \alpha, \beta_{1:K}, \eta, \delta] \). Given \( \gamma^{\text{new}} \), we can find other participant-days closest to it. We do this by comparing \( \gamma^{\text{new}} \) estimated by the sLDA model to the topic distributions of other days previously seen by the model. There are several similarity metrics that can be used for this comparison, but we select the Jensen-Shannon (JS) distance.

### 5.3.1 Jensen-Shannon Distance

The Jensen-Shannon distance is derived from the Jensen-Shannon divergence. The Jensen-Shannon divergence (JSD) is a method of measuring the similarity between two probability distributions, and it is based on the KL-divergence. However, unlike the KL-divergence, JSD is symmetric and always finite. Its symmetry ensures that the similarity between distributions \( P \) and \( Q \) is the same as the similarity between distributions \( Q \) and \( P \), i.e., \( JSD(P||Q) = JSD(Q||P) \). Also, when the natural log is used JSD is bounded as follows: \( 0 \leq JSD \leq 1 \).

For two discrete distributions \( P \) and \( Q \), JSD is defined as

\[
JSD(P||Q) = \frac{1}{2} D(P||M) + \frac{1}{2} D(Q||M)
\]

where

\[
M = \frac{1}{2}(P + Q),
\]

and \( D \) is the KL-divergence

\[
D(P||Q) = \sum_i P_i \log \frac{P_i}{Q_i}.
\]

We have that

\[
JSD(P||Q) = \frac{1}{2} \sum_i \left[ P_i \log \left( \frac{P_i}{\frac{1}{2}(P_i + Q_i)} \right) + Q_i \log \left( \frac{Q_i}{\frac{1}{2}(P_i + Q_i)} \right) \right],
\]

and the JS-distance is \( \sqrt{JSD(P||Q)} \).
Sec. 5.4 Case 1

The smaller the JS-distance, the more similar two distributions are, and in our case, the more similar participant-days are. When using the JS-distance, we constrain the model to find similar participant-days with higher self-report scores. The sections that follow illustrate how we can make recommendations to two different participants in the SNAPSHOT study using sLDA and the JS-Distance.

5.4 Case 1

Participant A is on the evening of the 18th day of the study. Figure 5.1 shows the self-reported stress scores of the participant for the past 17 days. Figures 5.2, 5.3 and 5.4 show box-plots of the behaviors exhibited by participant A during the past 17 days.

![Figure 5.1](image)

**Figure 5.1:** Self-reported stress-calm scores of Participant A for the past 17 days of the study.

The health behaviors are ordered according to the average of the self-reported stress scores on the days the behaviors are exhibited. The behaviors exhibited on Day 18 are shown in Table 5.1. Given these behaviors, the sLDA model predicted that the participant is in the mid-high calm range. The true score reported by the participant
is 75. We query the day-pattern distributions used to train the model and report the two most similar participant-days with calm scores higher than Participant A in Table 5.2.

The two participants returned in the query had similar sleep duration the night before, but they went to bed earlier. Similar-participant-2 had better sleep duration three nights in a row, and on the previous night, bedtime was the same as the average bedtime during the study (no deviation). Based on the health behaviors of the similar participants, two recommendations to Participant A could be: go to bed an hour or two earlier, and also to have more social interactions the next day. From Figure 5.3, we see that the average stress scores over the past days Participant A went to bed at 2–3AM and 1–2AM were 87 and 80, respectively. Also in Figure 5.2, Participant A had an average reported score of 82.5 over the days when outgoing call duration was
Figure 5.3: Health behaviors corresponding to sleep extracted from Actiwatch for the past 17 days. Ordered by the average of self-reported stress-calm scores on days the behavior was present (100: most calm, 0: most stress).

Figure 5.4: Health behaviors extracted from GPS coordinates and twice daily surveys for the past 17 days. Ordered by the average of self-reported stress-calm scores on days the behavior was present (100: most calm, 0: most stress).
between 6–12 minutes. Therefore, these recommendations are reasonable and might improve the well-being of participant A.

### 5.5 Case 2

Participant B is on the 14th day of the study. Figure 5.5 shows the self-reported stress scores of the participant for the past 13 days. We can see that the participant reported high stress (score < 55) for most of the days. Figures 5.6, 5.7, and 5.8 show box-plots of the past behaviors exhibited by participant B. The behaviors are ordered according to the average self-reported stress scores.

The behaviors exhibited on Day 14 are shown in Table 5.3. Given these behaviors, the sLDA model predicted that the participant is in the high stress range (self-report scores < 50). The actual score reported by the participant is 21. We query the day-pattern
Table 5.2: Health behaviors of two participants most similar to Participant A. Bold entries highlight possible recommendations.

<table>
<thead>
<tr>
<th>Behaviors</th>
<th>Similar Participant 1</th>
<th>Similar Participant 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Score = 87</td>
<td>Score = 91</td>
</tr>
<tr>
<td>Total num. outgoing calls</td>
<td>≥ 5</td>
<td>3–4</td>
</tr>
<tr>
<td>Total unique num. outgoing calls</td>
<td>≥ 3</td>
<td>≥ 3</td>
</tr>
<tr>
<td>Screen-on events (5PM - Midnight)</td>
<td>25–50</td>
<td>25–50</td>
</tr>
<tr>
<td><strong>Num. outgoing SMS (5PM - Midnight)</strong></td>
<td>2–5</td>
<td>2–5</td>
</tr>
<tr>
<td>Total num. outgoing SMS</td>
<td>4–10</td>
<td>4–10</td>
</tr>
<tr>
<td>Incoming call duration</td>
<td>≥ 12 mins.</td>
<td>0</td>
</tr>
<tr>
<td><strong>Outgoing calls duration</strong></td>
<td>≥ 12 mins.</td>
<td>≥ 12 mins.</td>
</tr>
<tr>
<td>Total screen-on duration</td>
<td>2–3 hours</td>
<td>2–3 hours</td>
</tr>
<tr>
<td>Total screen-on duration (Midnight–3AM)</td>
<td>0.5 hours</td>
<td>0–0.5 hours</td>
</tr>
<tr>
<td>Total screen-on duration (5PM-Midnight)</td>
<td>0.5–1 hours</td>
<td>0.5–1 hours</td>
</tr>
<tr>
<td>Sleep duration</td>
<td>6–7 hours</td>
<td>6–7 hours</td>
</tr>
<tr>
<td><strong>Bedtime</strong></td>
<td>3AM-4AM</td>
<td>1AM-2AM</td>
</tr>
<tr>
<td>Yesterday sleep duration</td>
<td>7–8 hours</td>
<td>7–8 hours</td>
</tr>
<tr>
<td>Day-before-yesterday sleep duration</td>
<td>4–6 hours</td>
<td>8–10 hours</td>
</tr>
<tr>
<td>Bedtime deviation</td>
<td>1 hour</td>
<td>0 hour</td>
</tr>
<tr>
<td>Sleep regularity</td>
<td>0.8–0.9</td>
<td>0.7–0.8</td>
</tr>
<tr>
<td>Time on campus</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Time with indoors indication</td>
<td>6–7 hours</td>
<td>0</td>
</tr>
<tr>
<td>Time with outdoor indication</td>
<td>1–8</td>
<td>0</td>
</tr>
<tr>
<td><strong>Study duration</strong></td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Exercise duration</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Extracurricular duration</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><strong>Academic duration</strong></td>
<td>2–3</td>
<td>4–6</td>
</tr>
<tr>
<td>Type of interaction</td>
<td>Positive</td>
<td>Positive</td>
</tr>
<tr>
<td>Pre-sleep interaction</td>
<td>Personal</td>
<td>Personal and media</td>
</tr>
<tr>
<td><strong>Alcohol or Caffeine consumption</strong></td>
<td>None</td>
<td>Caffeine</td>
</tr>
</tbody>
</table>

Figure 5.5: Self-reported stress-calm scores of Participant B for the past 13 days of the study.
Figure 5.6: Health behaviors extracted from cell-phone usage for the past 13 days. Ordered by the average of self-reported stress-calm scores on days the behavior was present (100: most calm, 0: most stress).

distributions used to train the model and report the two most similar participant-days with calm scores higher than Participant B in Table 5.4.

Health behaviors such as less than 4 hours of sleep, minimum phone usage, negative social interactions, and late bedtime the night before are present in this high stress day. On the contrary, the two similar participants returned by the model have very high calm scores and exhibit behaviors such as moderate phone usage, only positive interactions, adequate sleep and going to bed earlier that their average bedtime. In order to improve well-being, it could be recommended that Participant B go to bed an hour earlier, sleep for at least 6 hours that night and keep a regular bedtime going forward. Also, avoiding negative interactions or seeking positive interactions or both could help improve well-being of participant B. Figure 5.7 shows higher average calm scores of 85 and 67 on days Participant B went to bed between 2AM–3AM, and slept for over
Figure 5.7: Health behaviors corresponding to sleep extracted from Actiwatch for the past 13 days. Ordered by the average of self-reported stress-calm scores on days the behavior was present (100: most calm, 0: most stress).

Figure 5.8: Health behaviors extracted from GPS coordinates and twice daily surveys for the past 13 days. Ordered by the average of self-reported stress-calm scores on days the behavior was present (100: most calm, 0: most stress).
Table 5.3: Health behaviors exhibited on Day 14 by participant B

<table>
<thead>
<tr>
<th>Modality</th>
<th>Behaviors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cell-phone usage</td>
<td>Screen-on events (5PM–Midnight): 0</td>
</tr>
<tr>
<td></td>
<td>Num. outgoing SMS (5PM–Midnight): 0</td>
</tr>
<tr>
<td></td>
<td>Total unique num. outgoing SMS: 1</td>
</tr>
<tr>
<td></td>
<td>Total screen-on duration: ~ 0</td>
</tr>
<tr>
<td></td>
<td>Total screen-on duration(Midnight–3AM): ~ 0</td>
</tr>
<tr>
<td></td>
<td>Total screen-on duration(5PM–Midnight): ~ 0</td>
</tr>
<tr>
<td>Actiwatch</td>
<td>Sleep duration: 0.1–4 hours</td>
</tr>
<tr>
<td></td>
<td>Bedtime: 4AM–6AM</td>
</tr>
<tr>
<td></td>
<td>Yesterday sleep duration: 4–6 hours</td>
</tr>
<tr>
<td></td>
<td>Day-before-yesterday sleep duration: 4–6 hours</td>
</tr>
<tr>
<td></td>
<td>Bedtime deviation: ≥ 2</td>
</tr>
<tr>
<td></td>
<td>Sleep regularity: 0.5–0.6</td>
</tr>
<tr>
<td>Behavioral surveys</td>
<td>Study duration: 4–6 hours</td>
</tr>
<tr>
<td></td>
<td>Exercise duration: 0</td>
</tr>
<tr>
<td></td>
<td>Extracurricular duration: 0</td>
</tr>
<tr>
<td></td>
<td>Academic duration: 4–6 hours</td>
</tr>
<tr>
<td></td>
<td>Negative Interaction</td>
</tr>
<tr>
<td></td>
<td>Positive interaction</td>
</tr>
<tr>
<td></td>
<td>Pre-sleep interaction: media and personal</td>
</tr>
</tbody>
</table>

Table 5.4: Health behaviors of two participants most similar to Participant B. Bold entries highlight possible recommendations.

<table>
<thead>
<tr>
<th>Behaviors</th>
<th>Similar Participant 1</th>
<th>Similar Participant 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Screen-on events (5PM - Midnight)</td>
<td>0–25</td>
<td>50–75</td>
</tr>
<tr>
<td>Num. outgoing SMS (5PM - Midnight)</td>
<td>1</td>
<td>2–5</td>
</tr>
<tr>
<td>Total num. outgoing SMS</td>
<td>1</td>
<td>4–10</td>
</tr>
<tr>
<td>Total screen-on duration</td>
<td>0–2 hours</td>
<td>≥ 4 hours</td>
</tr>
<tr>
<td>Total screen-on duration (Midnight–3AM)</td>
<td>0</td>
<td>≥ 0.5 hours</td>
</tr>
<tr>
<td>Total screen-on duration (5PM–Midnight)</td>
<td>0–0.5 hours</td>
<td>1–2 hours</td>
</tr>
<tr>
<td>Sleep duration</td>
<td>6–7 hours</td>
<td>≥ 10 hours</td>
</tr>
<tr>
<td>Bedtime</td>
<td>3AM–4AM</td>
<td>N/A</td>
</tr>
<tr>
<td>Yesterday sleep duration</td>
<td>4–6 hours</td>
<td>≥ 10 hours</td>
</tr>
<tr>
<td>Day-before-yesterday sleep duration</td>
<td>4–6 hours</td>
<td>8–10 hours</td>
</tr>
<tr>
<td>Bedtime deviation</td>
<td>−1 hour</td>
<td>≤ −2 hours</td>
</tr>
<tr>
<td>Sleep regularity</td>
<td>0.7–0.8</td>
<td>0.7–0.8</td>
</tr>
<tr>
<td>Time on campus</td>
<td>0</td>
<td>≥ 0–1 hour</td>
</tr>
<tr>
<td>Time with indoors indication</td>
<td>0–2 hours</td>
<td>8 hours</td>
</tr>
<tr>
<td>Time with outdoor indication</td>
<td>0</td>
<td>0–1</td>
</tr>
<tr>
<td>Study duration</td>
<td>0–2</td>
<td>2–4</td>
</tr>
<tr>
<td>Exercise duration</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Extracurricular duration</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Academic duration</td>
<td>0–2</td>
<td>3–4</td>
</tr>
<tr>
<td>Type of interaction</td>
<td>N/A</td>
<td>Positive</td>
</tr>
<tr>
<td>Pre-sleep interaction</td>
<td>Media</td>
<td>Personal and media</td>
</tr>
<tr>
<td>Alcohol or Caffeine consumption</td>
<td>None</td>
<td>None</td>
</tr>
</tbody>
</table>

Draft: August 28, 2018
eight hours, respectively. In Figure 5.8, we see an average score of 50, with a maximum calmness score of 85 on the days Participant B had positive social interactions. These recommendations could help improve Participant B’s well-being.

5.6 Summary

We have illustrated how the sLDA model can be used to provide personalized recommendations to two participants in the SNAPSHOT study: one with a mid-high calm score, and another with a high stress score. To get closer to the goal of building a real-world behavior recommendation system, future work can expand this approach in the following directions:

- **Longer Study:** We used data from a 30-day study to train the model. Recruiting more people for longer studies up to a semester for example, will provide more data that can enable us to learn better representations of health behaviors, and thus make better personalized recommendations. We note that a pilot semester-long study was run on 15 people, so a semester long study is feasible.

- **Integrated Development Environments (IDE):** Developing IDEs will make it easier to collect data from participants, run the model and make recommendations. Also, it will enable us to collect data from participants after they have followed the recommendations.

- **Data Analysis and Follow-up Study:** It is important to test the validity of the recommendations, i.e., whether or not well-being was improved, and to include any feedback into the model in an iterative process. Analyzing the data collected after recommendations are followed will provide us with valuable insight. A follow-up study that can test the causal relationships between some of the health
behaviors recommended and well-being will also strengthen the health behavior recommendations.
Chapter 6

Conclusion

This work has presented techniques to predict future well-being and assess behavioral factors that may influence the well-being of individuals. Enhancing future well-being in individuals has become a viable approach to improving health. However, past studies primarily focus on estimating or detecting the current state of an individual’s well-being state. In addition, these studies investigate how stand-alone health behaviors influence well-being. Furthermore, they do not provide data-backed insights and recommendations to individuals seeking to improve their well-being. Using multimodal data from wearable sensors, behavioral surveys, weather and mobile phones, this dissertation addresses these limitations and makes the following contributions:

- **Predict future well-being:** We employed hierarchical Bayesian logistic regression, a personalized multi-task learning framework, to predict future well-being of individuals in the SNAPSHOT study. We demonstrated statistically significant improvement of the personalized MTL model’s prediction performance over an equivalent non-personalized MTL model and a logistic regression, the STL equivalent. The personalized HBLR model had improvements over the non-personalized HBLR model and the STL model of 24–38% and 24–37%, respectively. Our results highlighted the importance of accounting for inter-individual differences when...
learning predictive rules of well-being. In addition, we uncovered relevant relationships between input features and personality traits of the individuals previously unseen by the model.

- **Assess behavioral influences on well-being:** We proposed a novel approach to map multi-modal data collected in the “wild” to meaningful representations of health behaviors. This approach was evaluated using 134 behaviors and 5397 days of data extracted from 224 unique participants in the SNAPSHOT study. We demonstrated the usefulness of the learned representations for (1) predicting well-being, and (2) creating actionable insights into behavioral patterns that may influence well-being. We learned associations between self-reported mood, stress and health, and combinations of behaviors that can be measured objectively. Finally, we illustrated how the learned behavioral patterns can be leveraged to create personalized evidence-based recommendations to individuals looking to improve their well-being.

### 6.1 Future Work

There are many exciting opportunities for this work to continue. A major limitation of the work presented in this dissertation is the small sample size of our study cohort. Collecting more data over longer periods can allow for some additional analyses. First, in Chapter 3, the HBLR model learned soft-clustering over the participants in the study. With data from a semester-long study, it will be interesting to see if the dominant cluster of an individual would change over different 30-day periods or if certain interventions can bring about a change in cluster memberships. This analysis can provide insights into the temporal dynamics of an individual’s well-being. Secondly, the HBLR model can be adapted to predict the future well-being of a novel individual who has not provided classification labels. Future work can extend the model and test the prediction accuracy.
on novel participants. This will be useful in a real world well-being prediction system with an aim of reaching a larger population. Thirdly, in Chapter 4, we model static latent representations of modifiable behaviors over 30-day time periods. With more data, we can use a supervised variant of the dynamic LDA [79] to model the time evolution of the learned behavioral patterns. For example, we can study how the latent patterns change over the course of a semester or from one season to another–fall to spring. We can also analyze how the effect of the learned patterns on well-being change over time, i.e., do certain combinations of behaviors have positive effect on calm in the fall and negative effect in the spring?

In Chapter 5, we presented two case studies that highlighted how the sLDA model can be used in a behavior recommendation model. Yet, many open questions and unsolved issues exist before a real-life health behavior recommendation system can be made available. First, given the recommended behaviors from a similar participant’s day, future work can learn how to prioritize the behaviors such that the recommendation is effective. Secondly, in order to test the validity of the recommendations, a large-scale study is needed. It would also be helpful to build an integrated development environment (IDE) that will make data collection, modeling, making recommendations, and providing feedback easier. Thirdly, since it will be up to the user to follow the recommendations provided, compliance may be an issue with such a system. Separate research on adherence will need to be carried out.

The models presented in this dissertation learned associative mechanisms between health behaviors and well-being. Initially, we set out to learn causal relationships, but we were hindered by the presence of several confounding variables in our dataset. As a result, we learned attributes of participants that are predictive of well-being, and attempted to relate our findings to known and possible causes. However, the knowledge
gathered in this work, points us to influential variables that can be tested in n-of-1 studies. Also, future work in this area should collect data on specific interventions that would allow for causal inference models to be tested.
Bibliography


*Draft: August 28, 2018*


*Draft: August 28, 2018*


[49] N. Jaques, S. Taylor, A. Azaria, A. Ghandeharioun, A. Sano, and R. Picard, “Predicting Students’ Happiness from Physiology, Phone, Mobility, and Behavioral Data,” in *Affective Computing and Intelligent Interaction*, Xi’an, China, 2015.


Appendix A

SNAPSHOT Data

The aim of the ambulatory study was to measure Sleep, Networks, Affect, Performance, Stress and Health using Objective Techniques (SNAPSHOT) [18]. Thirty days of multimodal wearable sensor, phone and self-report data were collected from undergraduate students over the course of six semesters. One of the goals is to be able to identify which behavioral factors separate participants into self-reported mood groups (healthy or sick, energy or sluggish, alert or sleepy, happiness or sad, calm or stressed). Another is to have the ability to forecast mood which is a difficult problem. So far over 250 students have participated in this study. Below are the different modalities that contribute measurements to this dataset:

**Pre Survey Questionnaire:** This is a survey filled out by each student prior to the start of the study. It includes information about the following: Pittsburgh Sleep Quality Index (PSQI), the Big Five Inventory Personality Test, the Horne Ostberg Morningness Evenness Questionnaire (MEQ), Perceived Stress Scale (PSS), State and Trait Anxiety Score, physical and mental health composite scores from SF-12, Academic performance (GPA for previous term), age, BMI, gender, ethnicity, academic major, and whether the student resides in On or Off campus housing.
Q-Sensor: Affectiva’s Q sensor provides the following measurements 24hrs a day at a sampling frequency of 8Hz: 3-axis accelerometer (Actigraphy), Skin surface temperature and Electrodermal Activity (Skin conductance).

Actiwatch: This sensor measures activity, light exposure levels and sleep-wake schedule during the course of the day and is sampled at 8Hz.

Phone: Participants installed an android application on their phones. Timestamps are collected for outgoing, incoming and missed calls as well as outgoing and incoming SMS and emails. The location of the student as well as screen-on timing is also logged by the application.

Study Surveys: During the study, participants fill out surveys in the morning and evening. Information is collected about academic, extracurricular, and exercise activities, sleep, caffeine intake, social interaction, and self reported general health, mood, alertness, tiredness and stress levels of the student.

Weather: Information about the weather is collected from Dark Sky’s forecast.io application. Specifically about sunlight, temperature, wind barometric pressure, cloud, etc for all days the students were in the study.

Post Study Questionnaire: Physical and mental health composite scores, Perceived stress scores are calculated for the student at the conclusion of the study. Also GPA for current term is reported at the end of the semester.
## Appendix B

### Assessment of Behavioral Influences on Self-Reported Stress

Table B.1: List of top 15 behaviors in the 11 patterns (topics) learned by sLDA for self-reported stress. Behaviors are listed in decreasing order. Patterns are listed in order of decreasing association with calmness.

<table>
<thead>
<tr>
<th>Topics</th>
<th>Top 15 Behaviors</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>Bedtime: 2AM–3AM, Sleep regularity: 0.8–0.9, Time spent on campus (hours): 0, Duration yesterday sleep (hours): 7–8, Time spent outdoors (hours): 1–8, Bedtime deviation (hours): 1, Time spent indoors (hours): 0, Duration sleep (hours): 7–8, Duration sleep (hours): 6–7, Bedtime deviation (hours): 0, Duration screen-on (hours): 0–2, Duration screen-on (Midnight–3AM) (hours): 0–0.5, Duration day-before-yesterday sleep (hours): 6–7, Academic duration (hours): 0, Duration day-before-yesterday sleep (hours): 7–8</td>
</tr>
<tr>
<td>9</td>
<td>Bedtime: Midnight–1AM, Duration sleep (hours): 8–10, Bedtime deviation (hours): ≤−2, Bedtime deviation (hours): -1, Pre-sleep interaction: media, Duration day-before-yesterday sleep (hours): 8–10, Extracurricular duration (hours): 0, Exercise duration (hours): 0, Duration yesterday sleep (hours): 8–10, Time spent outdoors (hours): 0, Pre-sleep interaction: personal, Sleep regularity: 0.7–0.8, Time spent on campus (hours): 0, Duration screen-on (Midnight–3AM) (hours): 0, Num. unique outgoing SMS (5PM–Midnight): 2–5,</td>
</tr>
<tr>
<td>2</td>
<td>Duration sleep (hours): 4–6, Bedtime deviation (hours): ≥2, Bedtime: 4AM–6AM, Bedtime deviation (hours): 1, Bedtime: 3AM–4AM, Extracurricular duration (hours): 0, Exercise duration (hours): 0, Duration day-before-yesterday sleep (hours): 4–6, Study duration (hours): 0, Duration screen-on (Midnight–3AM) (hours): 0–0.5, Time spent outdoors (hours): 0, Academic duration (hours): 0, Duration yesterday sleep (hours): 4–6, Duration sleep (hours): 0.1–4, Num. unique outgoing SMS (5PM–Midnight): 2–5,</td>
</tr>
<tr>
<td>1</td>
<td>Num. total outgoing calls: ≥5, Num. unique outgoing calls: ≥3, Duration outgoing call (mins.): ≥12, Duration outgoing call (mins.): ≥12, Exercise duration (hours): 0, Duration screen-on (hours): ≥4, Duration screen-on (Midnight–3AM) (hours): ≥0.5, Extracurricular duration (hours): 0, Duration screen-on (5PM–Midnight) (hours): 1–2, Duration screen-on (hours): 3–4, Pre-sleep interaction: personal, Pre-sleep interaction: media, Academic duration (hours): 0, Duration screen-on (5PM–Midnight) (hours): ≥2, Study duration (hours): 0, Duration incoming call (mins.): ≥12,</td>
</tr>
<tr>
<td>10</td>
<td>Duration screen-on (5PM–Midnight) (hours): 0–0.5, Num. screen-on events (5PM–Midnight): 0–25, Duration screen-on (5PM–Midnight) (hours): 0–0.5, Extracurricular duration (hours): 0, Num. unique outgoing SMS (5PM–Midnight): 0, Exercise duration (hours): 0, Pre-sleep interaction: media, Num. unique outgoing SMS (5PM–Midnight): 1, Num. unique outgoing SMS (5PM–Midnight): 1, Time spent outdoors (hours): 0–1, Bedtime deviation (hours): 0, Time spent outdoors (hours): 0, Sleep regularity: 0.6–0.7, Study duration (hours): 2–4,</td>
</tr>
<tr>
<td>3</td>
<td>Duration screen-on (5PM–Midnight) (hours): 0.5–1, Num. screen-on events (5PM–Midnight): 25–50, Duration screen-on (5PM–Midnight) (hours): 0.5–1, Num. screen-on events (5PM–Midnight): 0, Pre-sleep interaction: media, Exercise duration (hours): 0, Extracurricular duration (hours): 0, Duration screen-on (5PM–Midnight) (hours): 1–2, Time spent outdoors (hours): 0, Bedtime deviation (hours): -1, Study duration (hours): 0–2, Num. unique outgoing SMS (5PM–Midnight): ≥2, Time spent on campus (hours): 1–8, Bedtime deviation (hours): 0,</td>
</tr>
</tbody>
</table>
Table B.2: List of top 15 behaviors in the 11 patterns (topics) learned by sLDA for self-reported stress (continued). Behaviors are listed in decreasing order. Patterns are listed in order of decreasing association with calmness.

<table>
<thead>
<tr>
<th>Topics</th>
<th>Top 15 Behaviors</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Num. unique outgoing calls: 1, Duration outgoing call (mins.): 0–2, Num. total outgoing calls: 2, Duration incoming call (mins.): 0, Exercise duration (hours): 0, Num. unique outgoing SMS (5PM–Midnight): 0, Num. total outgoing calls: 1, Extracurricular duration (hours): 0, Academic duration (hours): 0, Num. unique outgoing SMS: 0, Pre-sleep interaction: media, Duration incoming call (mins.): 0–2, Duration screen-on (hours): 0–2, Pre-sleep interaction: personal, Duration screen-on (Midnight–3AM) (hours): 0–0.5</td>
</tr>
<tr>
<td>6</td>
<td>Duration screen-on (5PM–Midnight) (hours): 0, Num. screen-on events (5PM–Midnight): 0, Duration screen-on (Midnight–3AM) (hours): 0, Duration screen-on (hours): 0, Pre-sleep interaction: media, Exercise duration (hours): 0, Extracurricular duration (hours): 0, Num. unique outgoing SMS: 1, Bedtime deviation (hours): 0, Academic duration (hours): 0–2, Bedtime: 1AM–2AM, Study duration (hours): 0–2, Caffeine consumption, Study duration (hours): 2–4, Pre-sleep interaction: personal</td>
</tr>
<tr>
<td>7</td>
<td>Num. unique outgoing SMS: 4–10, Num. unique outgoing SMS (5PM–Midnight): 2–5, Time spent outdoors (hours): 0, Num. unique outgoing calls: 2, Num. total outgoing calls: 3–4, Duration screen-on (Midnight–3AM) (hours): 0–0.5, Duration incoming call (mins.): 0, Num. unique outgoing SMS (5PM–Midnight): ≥ 5, Extracurricular duration (hours): 0, Pre-sleep interaction: personal, Num. unique outgoing SMS: 10–15, Num. screen-on events (5PM–Midnight): 50–75, Positive interaction, Pre-sleep interaction: media, Study duration (hours): 2–4</td>
</tr>
<tr>
<td>8</td>
<td>Duration outgoing call (mins.): 0, Num. unique outgoing calls: 0, Num. total outgoing calls: 0, Exercise duration (hours): 0, Duration incoming call (mins.): 0–2, Extracurricular duration (hours): 0, Duration incoming call (mins.): 0, Pre-sleep interaction: media, Num. unique outgoing SMS (5PM–Midnight): 0, Duration screen-on (Midnight–3AM) (hours): 0–0.5, Num. unique outgoing SMS: 0, Duration screen-on (hours): 0–2, Academic duration (hours): 0, Pre-sleep interaction: personal, Positive interaction</td>
</tr>
<tr>
<td>5</td>
<td>Exercise duration (hours): 0, Extracurricular duration (hours): 0, Caffeine consumption, Duration screen-on (Midnight–3AM) (hours): 0–0.5, Num. unique outgoing SMS (5PM–Midnight): 0, Duration screen-on (hours): 0–2, Negative interaction, Pre-sleep interaction: media, Time spent outdoors (hours): 0, Study duration (hours): 4–6, Time spent on campus (hours): 0, Academic duration (hours): 0, Naps, Duration incoming call (mins.): 0, Duration sleep (hours): 4–6</td>
</tr>
</tbody>
</table>
Appendix C

Assessment of Behavioral Influences on Self-Reported Happiness

In this appendix, we show the results of the analysis in Chapter 4 on self-reported happiness. Figure C.1 shows the probability distribution of the sad/happy response labels in the dataset; 0 represents the highest level of sadness and 100 represents the highest level of happiness.

C.1 Results

SLDA versus LASSO: Table C.1 shows the best mean (± standard deviation) of the performance metrics the LASSO achieved, and the highest value sLDA achieved across the different topics and over five repetitions. For the binary prediction accuracy metric, sLDA had a modest improvement over LASSO of 2.3%, when the threshold was set at the mean happiness response (60.96). For the F1 score metric, sLDA had an improvement over LASSO of 62% when the threshold was set at the mean happiness response. A Welch’s t-test revealed that sLDA had a statistically significant improvement over LASSO’s F1 score. The LASSO achieved statistically significant higher correlation when compared to the sLDA. We note the advantage that sLDA models the latent
Figure C.1: Distribution of self-reported sad/happiness response variables. The mean of the response variables, 60.96, is shown in red.

Table C.1: Mean (± standard deviation) of binary accuracy, F1 score (threshold = mean), and correlation coefficients for sLDA and LASSO when predicting self-reported happiness. Bold entries represent a statistically significant improvement (p < 0.05).

<table>
<thead>
<tr>
<th>Model</th>
<th>Binary Accuracy</th>
<th>F1 Score</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>sLDA</td>
<td>57.6% (± 1.7)</td>
<td>0.68 (± 0.02)</td>
<td>0.21 (± 0.03)</td>
</tr>
<tr>
<td>LASSO</td>
<td>56.3% (± 1.7)</td>
<td>0.42 (± 0.01)</td>
<td>0.27 (± 0.03)</td>
</tr>
</tbody>
</table>

SLDA versus LDA: The average binary prediction accuracies when the threshold is set as the mean of self-reported happiness are shown in Figure C.3. The average F1 scores are shown in Figure C.4. The average correlations between the real-valued true and predicted response values are shown in Figure C.5.
Figure C.2: LASSO Non-zero coefficient estimates for self-reported happiness. In the data, 100 represents the highest level of happiness and 0 represents the highest level of sadness. So the positive coefficients are positively correlated with increasing happiness.

SLDA Model: The 11-Topic fit of the sLDA model for self-reported happiness is shown in Tables C.2 and C.3. The patterns are listed in decreasing association with happiness. The 15 most probable behaviors in each pattern are listed in decreasing order.
Figure C.3: Average binary prediction accuracy (over 5 runs, ± standard deviation) for sLDA and LDA + Regression models across different topics (threshold = 60.96) for self-reported happiness.

Figure C.4: Average F1 score (over 5 runs, ± standard deviation) for sLDA and LDA + Regression models across different topics (threshold = 60.96) for self-reported happiness.
Figure C.5: Average correlation (over 5 runs, ± standard deviation) between the true and predicted response variables for sLDA and LDA + Regression models across different topics for self-reported happiness.

Table C.2: List of top 15 behaviors in the 11 patterns (topics) learned by sLDA for self-reported happiness. Behaviors are listed in decreasing order. Patterns are listed in order of decreasing association with happiness.

<table>
<thead>
<tr>
<th>Topics</th>
<th>Top 15 Behaviors</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>Num. total outgoing calls: ≥ 5, Num. unique outgoing calls: ≥ 3, Duration outgoing call (mins.): ≥ 12, Time spent indoors (hours): 0, Num. unique outgoing SMS (5PM–Midnight): 2–5, Academic duration (hours): 0, Exercise duration (hours): 0, Num. unique outgoing calls: 2, Extracurricular duration (hours): 0, Duration screen-on (Midnight–3AM) (hours): 0–0.5, Caffeine consumption, Duration incoming call (mins.): 0–2, Num. total outgoing calls: 3–4, Duration outgoing call (mins.): 6–12, Time spent outdoors (hours): 1–8</td>
</tr>
<tr>
<td>3</td>
<td>Bedtime: 2AM–3AM, Pre-sleep interaction: media, Extracurricular duration (hours): 0, Sleep regularity: 0.7–0.8, Duration screen-on (5PM–Midnight) (hours): 0.5–1, Study duration (hours): 0–2, Duration sleep (hours): 7–8, Duration screen-on (hours): 2–3, Duration screen-on (Midnight–3AM) (hours): 0–0.5, Num. unique outgoing SMS (5PM–Midnight): 2–5, Duration yesterday sleep (hours): 6–7</td>
</tr>
<tr>
<td>9</td>
<td>Bedtime: 2AM–3AM, Pre-sleep interaction: media, Extracurricular duration (hours): 0, Sleep regularity: 0.7–0.8, Duration screen-on (5PM–Midnight) (hours): 0.5–1, Study duration (hours): 0–2, Duration sleep (hours): 7–8, Duration screen-on (hours): 2–3, Duration screen-on (Midnight–3AM) (hours): 0–0.5, Num. unique outgoing SMS (5PM–Midnight): 2–5, Duration yesterday sleep (hours): 6–7</td>
</tr>
<tr>
<td>2</td>
<td>Duration sleep (hours): 4–6, Bedtime deviation (hours): 1, Bedtime: 3AM–4AM, Bedtime deviation (hours): ≥ 2, Bedtime: 4AM–6AM, Duration yesterday sleep (hours): 4–6, Duration day-before-yesterday sleep (hours): 4–6, Sleep regularity: 0.7–0.8, Exercise duration (hours): 0, Extracurricular duration (hours): 0, Time spent outdoors (hours): 0, Duration screen-on (Midnight–3AM) (hours): 0–0.5, Num. unique outgoing SMS (5PM–Midnight): 2–5, Duration day-before-yesterday sleep (hours): 6–7, Sleep regularity: 0.5–0.6</td>
</tr>
<tr>
<td>0</td>
<td>Num. unique outgoing calls: 1, Duration outgoing call (mins.): 0–2, Num. total outgoing calls: 2, Duration incoming call (mins.): 0, Num. total outgoing calls: 1, Exercise duration (hours): 0, Extracurricular duration (hours): 0, Num. unique outgoing SMS (5PM–Midnight): 0, Num. unique outgoing SMS: 0, Pre-sleep interaction: media, Duration incoming call (mins.): 0–2, Duration screen-on (Midnight–3AM) (hours): 0–0.5, Duration screen-on (hours): 0, Academic duration (hours): 0, Pre-sleep interaction: personal</td>
</tr>
<tr>
<td>10</td>
<td>Duration screen-on (5PM–Midnight) (hours): 0–0.5, Num. screen-on events (5PM–Midnight): 0–25, Duration screen-on (hours): 0–2, Duration screen-on (Midnight–3AM) (hours): 0–0.5, Extracurricular duration (hours): 0, Exercise duration (hours): 0, Num. unique outgoing SMS (5PM–Midnight): 0, Pre-sleep interaction: media, Num. unique outgoing SMS: 1, Duration screen-on (Midnight–3AM) (hours): 0, Time spent on campus (hours): 0, Academic duration (hours): 0, Bedtime deviation (hours): 0, Num. unique outgoing SMS (5PM–Midnight): 1, Study duration (hours): 4–6</td>
</tr>
</tbody>
</table>
Table C.3: List of top 15 behaviors in the 11 patterns (topics) learned by sLDA for self-reported happiness (continued). Behaviors are listed in decreasing order. Patterns are listed in order of decreasing association with happiness.

<table>
<thead>
<tr>
<th>Topics</th>
<th>Top 15 Behaviors</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>Duration outgoing call (mins.): 0, Num. unique outgoing calls: 0, Num. total outgoing calls: 0, Duration incoming call (mins.): 0–2, Exercise duration (hours): 0, Duration incoming call (mins.): 0, Extracurricular duration (hours): 0, Duration screen-on (hours): 0–2, Duration screen-on (Midnight–3AM) (hours): 0–0.5, Pre-sleep interaction: media, Num. unique outgoing SMS (5PM–Midnight): 0, Num. unique outgoing SMS: 0, Academic duration (hours): 0, Pre-sleep interaction: personal, Caffeine consumption</td>
</tr>
<tr>
<td>6</td>
<td>Duration screen-on (5PM–Midnight) (hours): 0, Num. screen-on events (5PM–Midnight): 0, Duration screen-on (Midnight–3AM) (hours): 0, Duration screen-on (hours): 0, Exercise duration (hours): 0, Pre-sleep interaction: media, Num. unique outgoing SMS (5PM–Midnight): 0, Extracurricular duration (hours): 0, Academic duration (hours): 0, Caffeine consumption, Num. unique outgoing SMS: 1, Pre-sleep interaction: personal, Naps, Bedtime deviation (hours): 0, Study duration (hours): 2–4</td>
</tr>
<tr>
<td>1</td>
<td>Num. unique outgoing SMS (5PM–Midnight): 1, Exercise duration (hours): 0, Num. screen-on events (5PM–Midnight): 25–50, Duration screen-on (Midnight–3AM) (hours): 0–0.5, Num. unique outgoing SMS: 2, Duration screen-on (hours): 2–3, Time spent indoors (hours): 9–10, Num. unique outgoing SMS: 1, Time spent on campus (hours): ≥ 8, Duration screen-on (5PM–Midnight) (hours): 0.5–1, Duration screen-on (5PM–Midnight) (hours): 1–2, Time spent outdoors (hours): 0, Study duration (hours): 0, Caffeine consumption, Time spent outdoors (hours): 0–1</td>
</tr>
<tr>
<td>5</td>
<td>Extracurricular duration (hours): 0, Exercise duration (hours): 0, Negative interaction, Pre-sleep interaction: media, Duration screen-on (hours): 0–2, Caffeine consumption, Academic duration (hours): 0, Duration screen-on (Midnight–3AM) (hours): 0–0.5, Num. unique outgoing SMS (5PM–Midnight): 0, Time spent on campus (hours): 0, Duration incoming call (mins.): 0, Pre-sleep interaction: personal, Time spent outdoors (hours): 0, Duration screen-on (5PM–Midnight) (hours): 0–0.5, Study duration (hours): 2–4</td>
</tr>
</tbody>
</table>
Appendix D

Assessment of Behavioral Influences on Self-Reported Health

In this appendix, we show the results of the analysis in Chapter 4 on self-reported health. Figure D.1 shows the probability distribution of the sick/healthy response labels in the dataset; 0 represents the highest level of feelings of sickness and 100 represents the highest level of feelings of being healthy.

D.1 Results

SLDA versus LASSO: Table D.1 shows the best mean (± standard deviation) of the performance metrics the LASSO achieved, and the highest value sLDA achieved across the different topics and over five repetitions. For the binary prediction accuracy metric, sLDA had an improvement over LASSO of 10.5%, when the threshold was set at the mean self-reported health response (65). For the F1 score metric, sLDA had an improvement over LASSO of 183% when the threshold was set at the mean self-reported health response. A Welch’s t-test revealed that sLDA had a statistically significant improvement over LASSO’s binary prediction accuracy and F1 score. The LASSO achieved statistically significant higher correlation when compared to the sLDA.
Figure D.1: Distribution of self-reported sick/healthy response variables. The mean of the response variables, 65.00, is shown in red.

Table D.1: Mean (± standard deviation) of binary accuracy, F1 score (threshold = mean), and correlation coefficients for sLDA and LASSO when predicting self-reported health. Bold entries represent a statistically significant improvement (p < 0.05).

<table>
<thead>
<tr>
<th>Model</th>
<th>Binary Accuracy</th>
<th>F1 Score</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>sLDA</td>
<td>54.6% (± 1.8)</td>
<td>0.68 (± 0.02)</td>
<td>0.10 (± 0.02)</td>
</tr>
<tr>
<td>LASSO</td>
<td>49.4% (± 1.0)</td>
<td>0.24 (± 0.02)</td>
<td>0.14 (± 0.02)</td>
</tr>
</tbody>
</table>

Figure D.2 shows the features that were selected by the LASSO model in predicting self-reported health. In particular, studying for more than 8 hours a day was found to be a strong predictor of self-reported ill-health. We note that this is an associational relationship, and the direction of causality is not known. However, one possible explanation is that feelings of ill-health could be exacerbated by long hours of studying.

SLDA versus LDA: The average binary prediction accuracies for when the threshold is set as the mean self-reported response are shown in Figure D.3. The average F1
Figure D.2: LASSO Non-zero coefficient estimates for self-reported health. In the data, 100 represents the highest level of feelings of health and 0 represents the highest level of feelings of sickness. So the positive coefficients are positively correlated with increasing health.

scores are shown in Figure D.4. The average correlations between the real-valued true and predicted response values are shown in Figure D.5.

SLDA Model: The 12-Topic fit of the sLDA model for self-reported health is shown in Tables D.2 and D.3. The patterns are listed in decreasing association with happiness. The 15 most probable behaviors in each pattern are listed in decreasing order.
Figure D.3: Average binary prediction accuracy (over 5 runs, ± standard deviation) for sLDA and LDA + Regression models across different topics (threshold = 65) for self-reported health.

Figure D.4: Average F1 score (over 5 runs, ± standard deviation) for sLDA and LDA + Regression models across different topics (threshold = 65) for self-reported health.
**Figure D.5:** Average correlation (over 5 runs, ± standard deviation) between the true and predicted response variables for sLDA and LDA + Regression models across different topics for self-reported health.

**Table D.2:** List of top 15 behaviors in the 12 patterns (topics) learned by sLDA for self-reported health. Behaviors are listed in decreasing order. Patterns are listed in order of decreasing association with health.

<table>
<thead>
<tr>
<th>Topics</th>
<th>Top 15 Behaviors</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td><strong>Bedtime:</strong> Midnight–1AM, <strong>Duration sleep (hours):</strong> 8–10, <strong>Bedtime deviation (hours):</strong> ≤−2, <strong>Bedtime deviation (hours):</strong> −1, <strong>Duration day-before-yesterday sleep (hours):</strong> 8–10, <strong>Pre-sleep interaction:</strong> media, <strong>Sleep regularity:</strong> 0.7–0.8, <strong>Exercise duration (hours):</strong> 0, <strong>Time spent outdoors (hours):</strong> 0, <strong>Duration yesterday sleep (hours):</strong> 8–10, <strong>Pre-sleep interaction:</strong> personal, <strong>Extracurricular duration (hours):</strong> 0, <strong>Sleep regularity:</strong> 0.6–0.7, <strong>Study duration (hours):</strong> 2–4, <strong>Num. unique outgoing SMS (5PM–Midnight):</strong> 2–5</td>
</tr>
<tr>
<td>4</td>
<td><strong>Bedtime:</strong> 2AM–3AM, <strong>Duration day-before-yesterday sleep (hours):</strong> 7–8, <strong>Num. unique outgoing SMS (5PM–Midnight):</strong> 1, <strong>Duration sleep (hours):</strong> 6–7, <strong>Duration screen-on (Midnight–3AM) (hours):</strong> 0–0.5, <strong>Time spent indoors (hours):</strong> 0, <strong>Time spent outdoors (hours):</strong> 1–8, <strong>Time spent on campus (hours):</strong> 0, <strong>Duration screen-on (hours):</strong> 0–2, <strong>Duration yesterday sleep (hours):</strong> 7–8, <strong>Sleep regularity:</strong> 0.7–0.8, <strong>Exercise duration (hours):</strong> 0, <strong>Academic duration (hours):</strong> 0, <strong>Num. unique outgoing SMS:</strong> 1, <strong>Pre-sleep interaction:</strong> media</td>
</tr>
<tr>
<td>7</td>
<td><strong>Num. total outgoing calls:</strong> ≥5, <strong>Num. unique outgoing calls:</strong> ≥3, <strong>Duration outgoing call (mins.):</strong> ≥12, <strong>Num. unique outgoing calls:</strong> 2, <strong>Num. total outgoing calls:</strong> 3–4, <strong>Duration incoming call (mins.):</strong> 0, <strong>Exercise duration (hours):</strong> 0, <strong>Num. unique outgoing SMS (5PM–Midnight):</strong> ≥5, <strong>Num. unique outgoing SMS:</strong> 4–10, <strong>Num. unique outgoing SMS (5PM–Midnight):</strong> 2–5, <strong>Pre-sleep interaction:</strong> personal, <strong>Academic duration (hours):</strong> 0, <strong>Positive interaction, Duration incoming call (mins.):</strong> 0–2, <strong>Duration screen-on (Midnight–3AM) (hours):</strong> 0–0.5</td>
</tr>
<tr>
<td>1</td>
<td><strong>Duration screen-on (Midnight–3AM) (hours):</strong> ≥0.5, <strong>Duration screen-on (5PM–Midnight) (hours):</strong> 1–2, <strong>Duration screen-on (hours):</strong> 3–4, <strong>Duration screen-on (hours):</strong> ≥4, <strong>Extracurricular duration (hours):</strong> 0, <strong>Exercise duration (hours):</strong> 0, <strong>Duration screen-on (5PM–Midnight) (hours):</strong> ≥2, <strong>Pre-sleep interaction:</strong> media, <strong>Pre-sleep interaction:</strong> personal, <strong>Num. screen-on events (5PM–Midnight):</strong> 75–100, <strong>Num. unique outgoing SMS (5PM–Midnight):</strong> 1, <strong>Study duration (hours):</strong> 0, <strong>Num. screen-on events (5PM–Midnight):</strong> 25–50, <strong>Positive interaction, Academic duration (hours):</strong> 0</td>
</tr>
<tr>
<td>6</td>
<td><strong>Sleep regularity:</strong> 0.8–0.9, <strong>Num. unique outgoing SMS (5PM–Midnight):</strong> 2–5, <strong>Time spent outdoors (hours):</strong> 0, <strong>Bedtime deviation (hours):</strong> 0, <strong>Bedtime: 1AM–2AM, Duration sleep (hours):</strong> 7–8, <strong>Extracurricular duration (hours):</strong> 0, <strong>Time spent indoors (hours):</strong> 8, <strong>Time spent on campus (hours):</strong> 0, <strong>Duration screen-on (Midnight–3AM) (hours):</strong> 0–0.5, <strong>Duration day-before-yesterday sleep (hours):</strong> 6–7, <strong>Num. unique outgoing SMS:</strong> 4–10, <strong>Num. unique outgoing SMS:</strong> 3–4, <strong>Pre-sleep interaction:</strong> media, <strong>Duration yesterday sleep (hours):</strong> 6–7</td>
</tr>
<tr>
<td>10</td>
<td><strong>Duration screen-on (5PM–Midnight) (hours):</strong> 0–0.5, <strong>Num. screen-on events (5PM–Midnight):</strong> 0–25, <strong>Duration screen-on (hours):</strong> 0–2, <strong>Extracurricular duration (hours):</strong> 0, <strong>Num. unique outgoing SMS (5PM–Midnight):</strong> 0, <strong>Exercise duration (hours):</strong> 0, <strong>Duration screen-on (Midnight–3AM) (hours):</strong> 0, <strong>Duration screen-on (Midnight–3AM) (hours):</strong> 0–0.5, <strong>Pre-sleep interaction:</strong> media, <strong>Num. unique outgoing SMS:</strong> 1, <strong>Num. unique outgoing SMS (5PM–Midnight):</strong> 1, <strong>Academic duration (hours):</strong> 0, <strong>Num. unique outgoing SMS:</strong> 0, <strong>Naps, Time spent on campus (hours):</strong> 0</td>
</tr>
</tbody>
</table>
Table D.3: List of top 15 behaviors in the 12 patterns (topics) learned by sLDA for self-reported health (continued). Behaviors are listed in decreasing order. Patterns are listed in order of decreasing association with health.

<table>
<thead>
<tr>
<th>Topics</th>
<th>Top 15 Behaviors</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>Duration sleep (hours): 4–6, Bedtime: 3AM–4AM, Bedtime deviation (hours): ≥ 2, Bedtime: 4AM–6AM, Bedtime deviation (hours): 1, Exercise duration (hours): 0, Duration screen-on (Midnight–3AM) (hours): 0–0.5, Time spent outdoors (hours): 0, Extracurricular duration (hours): 0, Duration yesterday sleep (hours): 4–6, Duration day-before-yesterday sleep (hours): 4–6, Sleep regularity: 0.7–0.8, Num. unique outgoing SMS (5PM–Midnight): 2–5, Sleep regularity: 0.6–0.7, Sleep regularity: 0.5–0.6</td>
</tr>
<tr>
<td>0</td>
<td>Num. unique outgoing calls: 1, Duration outgoing call (mins.): 0–2, Num. total outgoing calls: 2, Duration incoming call (mins.): 0, Num. total outgoing calls: 1, Exercise duration (hours): 0, Extracurricular duration (hours): 0, Duration screen-on (Midnight–3AM) (hours): 0–0.5, Pre-sleep interaction: media, Duration incoming call (mins.): 0–2, Num. unique outgoing SMS (3PM–Midnight): 0, Academic duration (hours): 0, Pre-sleep interaction: personal, Num. unique outgoing SMS: 0, Caffeine consumption</td>
</tr>
<tr>
<td>3</td>
<td>Duration screen-on (hours): 2–3, Duration screen-on (5PM–Midnight) (hours): 0–0.5, Duration screen-on (4PM–Midnight) (hours): 1–2, Exercise duration (hours): 0, Num. screen-on events (5PM–Midnight): 25–50, Time spent outdoors (hours): 0, Study duration (hours): 0–2, Pre-sleep interaction: media, Extracurricular duration (hours): 0, Time spent indoors (hours): 9–10, Time spent indoors (hours): ≥ 10, Time spent on campus (hours): 1–8, Num. screen-on events (5PM–Midnight): 50–75, Duration day-before-yesterday sleep (hours): 4–6</td>
</tr>
<tr>
<td>8</td>
<td>Duration outgoing call (mins.): 0, Num. unique outgoing calls: 0, Num. total outgoing calls: 0, Duration incoming call (mins.): 0–2, Duration incoming call (mins.): 0, Exercise duration (hours): 0, Extracurricular duration (hours): 0, Duration screen-on (Midnight–3AM) (hours): 0–0.5, Pre-sleep interaction: media, Num. unique outgoing SMS (5PM–Midnight): 0, Duration screen-on (hours): 0–2, Num. unique outgoing SMS: 0, Academic duration (hours): 0, Caffeine consumption</td>
</tr>
<tr>
<td>11</td>
<td>Duration screen-on (Midnight–3AM) (hours): 0, Duration screen-on (hours): 0, Exercise duration (hours): 0, Num. unique outgoing SMS (5PM–Midnight): 0, Pre-sleep interaction: media, Extracurricular duration (hours): 0, Caffeine consumption, Num. unique outgoing SMS: 1, Academic duration (hours): 0, Sleep regularity: 0–0.4, Bedtime deviation (hours): 0, Duration yesterday sleep (hours): 4–6, Num. unique outgoing SMS: 0</td>
</tr>
<tr>
<td>5</td>
<td>Extracurricular duration (hours): 0, Exercise duration (hours): 0, Pre-sleep interaction: media, Time spent on campus (hours): 0, Caffeine consumption, Academic duration (hours): 0, Duration screen-on (Midnight–3AM) (hours): 0, Num. unique outgoing SMS (3PM–Midnight): 0, Pre-sleep interaction: personal, Num. unique outgoing SMS: 3–4, Num. unique outgoing SMS (5PM–Midnight): 2–5, Time spent outdoors (hours): 0, Duration screen-on (hours): 0–2, Num. screen-on events (5PM–Midnight): 25–50, Naps</td>
</tr>
</tbody>
</table>

136 Draft: August 28, 2018