Towards Future Health and Well-being: Bridging Behavior Modeling and Intervention

Xuhai Xu
xuhaixu@uw.edu
University of Washington, Seattle, USA

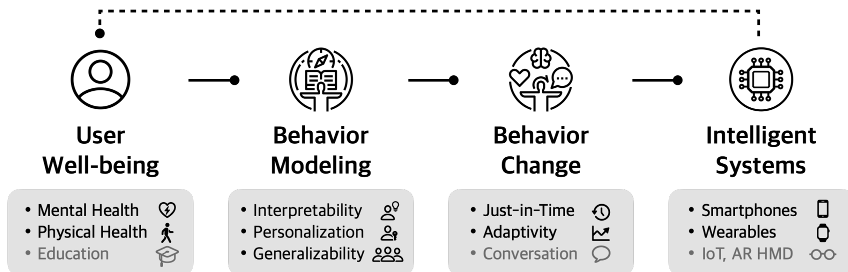

Figure 1: Research Roadmap towards Future Health and Well-being.

ABSTRACT

With the advent of always-available, ubiquitous devices with powerful passive sensing and active interaction capabilities, the opportunities to integrate AI into this ecosystem have matured, providing an unprecedented opportunity to understand and support user well-being. A wide array of research has demonstrated the potential to detect risky behaviors and address health concerns, using human-centered ML to understand longitudinal, passive behavior logs. Unfortunately, it is difficult to translate these findings into deployable applications without better approaches to providing human-understandable relationship explanations between behavior features and predictions; and generalizing to new users and new time periods. My past work has made significant headway in addressing modeling accuracy, interpretability, and robustness. Moreover, my ultimate goal is to build deployable, intelligent interventions for health and well-being that make use of succeeding ML-based behavior models. I believe that just-in-time interventions are particularly well suited to ML support. I plan to test the value of ML for providing users with a better, interpretable, and robust experience in supporting their well-being.

KEYWORDS

Behavior modeling, behavior intervention, health and well-being

ACM Reference Format:
Xuhai Xu. 2022. Towards Future Health and Well-being: Bridging Behavior Modeling and Intervention. In *The Adjunct Publication of the 35th Annual ACM Symposium on User Interface Software and Technology (UIST '22 Adjunct), October 29-November 2, 2022, Bend, OR, USA.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3526114.3558524

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).
UIST '22 Adjunct, October 29-November 2, 2022, Bend, OR, USA
© 2022 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-9321-8/22/10.
https://doi.org/10.1145/3526114.3558524

1 INTRODUCTION

As artificial-intelligent(AI)-powered devices become more embedded in our lives, they offer an unprecedented ability to passively sense our daily behavior at a high resolution, leading to a unique opportunity to model and impact our behavior to support our goals. An essential capability for these devices is to understand (via accurate and interpretable *modeling*) behavior of interest and impact (via carefully designed *intervention*) the behavior for various aspects of human well-being.

There is a growing body of research on behavior modeling via passive sensing data to predict specific behavioral outcomes [6, 9], such as detecting physical health issues [1, 7], monitoring mental health status [11, 16], and tracking education outcomes [12]. However, due to the scarcity of the behavior label, existing machine learning (ML) techniques for behavior modeling still have a set of challenges for real-life robust deployment: interpretability, personalization, and generalizability. More importantly, sensing and modeling behavior is only the first step towards an intelligent ubiquitous system. Given a well-performed model, we need appropriate ways to bring the model back to users for good. One opportunity is to bridge modeling techniques and intelligent intervention designs so that it can better support users' behavior goals. **My Ph.D. research aims to address two main human-centered challenges via AI/ML techniques (see Fig. 1):**

1) Achieving interpretable, personalized, and generalizable behavior modeling techniques.

2) Based on succeeding models, bringing intelligent, adaptive intervention experiences to users for better well-being.

Meaningful and effective support of end-user goals requires *interpretability* to achieve a transparent, trustful, and deployable system. However, most longitudinal behavior modeling work in our domain has focused on achieving higher detection or prediction accuracy, but not on interpretability (*e.g.*, depression detection [2]). Further, each individual has their own unique behavior pattern, ability, and preference, making *personalization* important for leveraging individual nuances, so that predictions and interventions can be tailored to each individual. However, longitudinal studies involving passive mobile sensing usually have very few labels as they are expensive to collect [20]. Therefore, previous work in this field has tended to use a one-model-fits-all methodology rather than a personalized model [11, 15]. Addressing these two challenges in a single dataset is the prerequisite for building deployable and generalizable modeling and intervention techniques. In real-life deployment, the behavior patterns of new users, or the same users under new contexts, could have a different distribution from the dataset a model is trained on. This requires a technique with cross-dataset *generalizability*. However, there is a lack of work in our community to address this challenge, blocking us from practical deployment.

Given an interpretable, personalized, and generalizable model, we need to leverage the model properly to bring users a better experience. One potential opportunity is to positively and adaptively intervene users' daily behavior. Recent advances in Just-in-Time Adaptive Interventions (JITAIs) in mobile health point out the need to adapt interventions to each individual based on users' contexts and status [8]. Nevertheless, most prior works have focused on manually pre-designing intervention delivery policies. The algorithmic aspect of the smart delivery of JITAIs is still in its infancy [10]. We can leverage AI/ML techniques to achieve an automatic, explainable intervention delivery process to maximize adherence and effectiveness. JITAIs can start with a set of predetermined policies, then 1) provide conversational explanation of users' behaviors and reactions towards interventions for both users themselves and stakeholders for better sense-making (*interpretability*), 2) evolve over time to better fit each individual (*personalization*), and 3) can easily work under novel contexts for new users (*generalizability*). Currently, there is no intervention delivery algorithm that can address these challenges at the same time.

My previous work improved the interpretability (Sec. 2.1) [15, 18, 21, 22], the personalization (Sec. 2.2) [16, 17], and the generalizability aspects (Sec. 2.3) [19, 23] of behavior modeling algorithms on multi-year longitudinal passive sensing datasets. Moreover, based on my exploration on just-in-time intervention techniques [24], I aim to achieve intelligent intervention delivery algorithms that are interpretable, personalized, and generalizable (Sec. 3.2). Combining these efforts, I want to build a generalizable and intelligent system that can understand and positively influence our behavior to facilitate our daily lives for better well-being.

2 RECENT WORK – INTERPRETABLE, PERSONALIZED, GENERALIZABLE MODELING ALGORITHMS

I start by addressing the three important machine learning challenges in the behavior modeling domain: interpretability (Sec. 2.1), personalization (Sec. 2.2), and generalizability (Sec. 2.3).

2.1 Interpretable Behavior Modeling

We need algorithms that can automatically extract effective behavior features to reveal the detailed relationship between activities and behavior outcomes in an accurate and interpretable manner, so that users and stakeholders can leverage the interpretation for better sense-making. For example, users can better reflect on themselves, and therapists/caregivers can better understand users' behavior patterns and model predictions. However, existing algorithms on the longitudinal behavioral data are hard to interpret, as they usually aggregate behavior features across days into a single value. The losses of information not only affect the model performance, but also limit the interpretation between these features and final behavior outcomes [2, 13]. **How can we extract fine-grained and effective behaviors features to provide interpretability and achieve an accurate model simultaneously?**

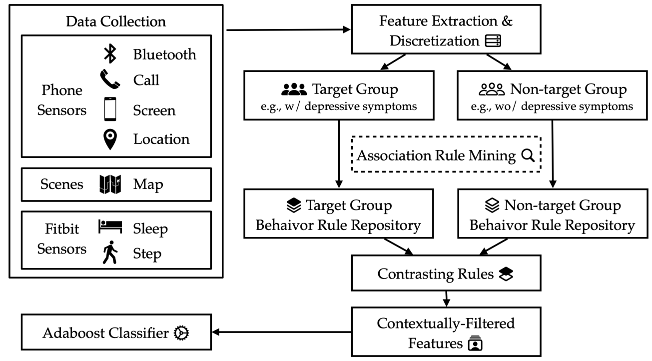

Figure 2: The framework of modeling with interpretability.

I have been involved in multiple studies as the technical lead and study coordinator collecting sensor data from college students' mobile phones and smart bands. The data contain rich information about each student's life experience, including mental health status, academic performance, physical activity, sleep pattern, *etc.* Currently, our team has collected data for at least three years, each from two institutes (University of Washington, UW 2018-2022, and Carnegie Mellon University, CMU 2017-2020). Each year 100-200 students joined the study, with around 50% students carrying over from prior years. The dataset has accumulated over one thousand student-years of behavior data, together with their subjective reports on a wide range of established questionnaires. We are actively working on releasing the UW dataset to contribute to the open-source community [23].

Using these data, I developed a new behavior modeling algorithm to optimize both modeling accuracy and interpretability (see Fig. 2). It utilizes the co-occurrence of multiple sensors' values within a short time period (6 hrs) and leverages association rule mining (ARM) to generate interpretable behavior rules and provide better contextual behavior features [15]. To validate my algorithm, I applied it to our CMU dataset to identify students' depressive symptoms. This approach significantly outperforms the state-of-the-art model by 10.2% on absolute accuracy. Moreover, the algorithm can generate behavior rules that are highly interpretable, capture behavior routines, and reveal behavior pattern differences between people with and without depressive symptoms. For example, for a rule [*long duration on campus, low number of bluetooth devices around* → *good sleep quality*], we find that it is significantly more frequent for students without depressive symptoms than those with the symptom. This indicates that students with depression tended to have a more disturbed sleep pattern. Compared to prior work, our contextualized rules can provide more details about user's sleep

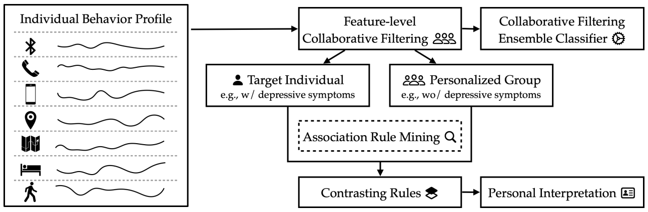

Figure 3: The framework of modeling with personalization.

behavior. These patterns can not only help users to better understand their daily routines, but also potentially assist care providers with better therapy design.

2.2 Personalization with Interpretability

However, some problems remain. Most of the algorithms and models, including my work on interpretability [15], follow a one-size-fits-all (*i.e.*, population modeling) approach that looks for common behaviors amongst all users. Such an approach de-emphasizes the important fact that individuals behave differently. Achieving personalization is important for both better model performance and more tailored individual interpretation. Nevertheless, for many passive mobile sensing studies, obtaining ground truth from end-users is expensive. For example, to get reliable labels of whether a person has depressive symptoms, users need to complete a well-established survey (*e.g.*, PHQ-4). It can only be reliably administered every two weeks [13]. Therefore, many longitudinal studies involving passive mobile sensing end up with a large amount of sensing data, but very few labels. **How can we leverage the large-scale individual behavior data for personalization to enhance the model accuracy and interpretability?**

To address this challenge, I developed a collaborative-filtering-based framework that effectively uses the study population's large volume of behavioral feature data and limited label data to enable personalization [16] (see Fig. 3). When a new user is added, the behavior relevance matrices are calculated between the new user and existing users to select users whose behaviors are strongly related to the new user to support classification (both positively and negatively); thus each new user will leverage a unique sub-population for classification, supporting personalization. The framework leverages individual users' behavioral relevance (both similarities and differences) to generate classifications. To obtain a detailed interpretation of users' behavior, the framework follows my previous work [15] and further leverages ARM to enable personalized interpretation. Specifically, ARM is applied two times, one to the target individual (*e.g.*, with depressive symptoms) and the other to the group who have differing labels (*e.g.*, without depressive symptoms). The paired rules capture the behavior differences, providing a personalized understanding of an individual's behaviors. I applied this framework to depression detection using our CMU and UW datasets. The results show that this framework further outperforms the previous work by 5.1% on absolute accuracy. Moreover, the interpretations identify both homogeneity (*e.g.*, same disturbed sleep patterns) and heterogeneity (*e.g.*, opposite social activities), providing a personalized understanding of each user's behavior. This can assist care providers in better tailoring for every individual.

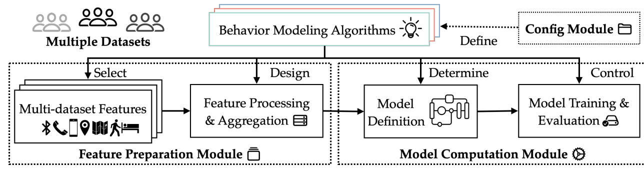

Figure 4: The overview of GLOBEM platform.

2.3 Generalizability across Datasets

Both the interpretable and personalized modeling techniques, together with most prior studies, mainly focus on building and evaluating ML models using data collected from a single population. However, for real-world longitudinal deployment, the behavior patterns of new populations, and existing populations under new contexts, could have a different distribution from the training dataset, leading to a generalizability challenge. To build a model with practical and useful deployability, it is essential to evaluate the model across multiple datasets to ensure its generalizability. **Can prior existing algorithms generalize across datasets? If not, how can we push it closer to real-life deployment by improving its generalizability across multiple datasets?**

To address the challenge, I develop the first open-source benchmark platform **GLOBEM** – short for **G**eneralization of **LO**ngitudinal **BE**havior **M**odeling – to support researchers in using, developing, and evaluating different longitudinal behavior modeling methods' cross-dataset generalizability [19] (see Fig. 4). Using depression detection as the example application, I employed four institute-year datasets (two years of data before COVID from two institutions) and closely re-implemented and evaluated nine prior depression detection algorithms (including my previous work [15, 16]), using the same cross-validation setup within each dataset. Although these models outperform the naive baseline predicting the majority (Δ = 7.1%), we observed a substantial drop between these models' performance on our datasets and their reported results in previous literature (average $\Delta = 18.8 \pm 9.8\%$), which indicates the lack of cross-dataset generalizability of these algorithms. Moreover, I also implemented eight recently popular domain generalization algorithms from the ML community and evaluated them using a leave-one-dataset-out setup. The results indicate that these methods also do not generalize well on our datasets, with barely any advantage over the naive baseline (Δ = 2.0%).

Finally, I developed a new algorithm, *Reorder*, that significantly outperforms existing methods on various cross-dataset generalization setups. Specifically, Reorder creates a new task of solving a temporal reordering puzzle in addition to the main depression detection task, thus it is forced to learn the continuity of behavior trajectories and achieve better generalization. Our results demonstrate that *Reorder* outperforms other methods by at least 3.2% on balanced accuracy and 3.4% on ROC AUC in the leave-one-dataset-out evaluation, taking one step closer to the real-life deployment.

3 CURRENT AND FUTURE WORK – FROM BEHAVIOR MODELING TO INTELLIGENT INTERVENTION

Based on my previous work, I am currently working on two directions: 1) to further improve the model performance with new users under new contexts (Sec. 3.1), 2) to leverage behavior models to create intelligent intervention systems to help users to achieve their well-being goals (Sec. 3.2).

3.1 Generalizability and Adaptation

Although the newly proposed technique Reorder has better generalizability, its overall advantage is incremental and still has great room for improvement (balanced accuracy as 55.2% in the leave-one-dataset-out setup). Our analysis indicates that the individual behavior differences introduce a more significant generalization challenge than temporal differences of the same users (*i.e.*, how users may change from one year to the next). Due to the limited amount of ground truth for each individual (common property of longitudinal behavior datasets), generalization is even more challenging. This may suggest the need to move from dataset generalization to dataset adaptation, *i.e.*, allow the model to access a small fraction of the data of new datasets or new users. **How can we enable a trained model to quickly adapt and personalize to each user by tuning its parameters on a small amount of data from the target individual?**

This idea is practical in a real-life deployment: after a model is trained on existing datasets and applied to a new user, it can first accumulate data for the first few weeks and tune the model based on the new user's behavior. Although such a design would require the new user to provide extra behavior labels, it can potentially address the individual distribution shift and achieve model personalization. This closely resonates the topic of domain adaptation and few-shot learning in the ML community [14]. There have been some recent promising advances in this direction in the behavior modeling area [3, 5], yet there is still a lack of systematic validation of existing few-shot techniques, or newly proposed adaptive algorithms on longitudinal passive sensing data streams for health and well-being applications. I am leading a group of students to extend our GLOBEM platform [19] to develop cross-dataset fast adaptation algorithms. We have investigated the taxonomy of supervised domain adaptive and few-short learning techniques, and started re-implementing representative ML techniques. We plan to propose and develop new adaptive techniques that are designed specifically for longitudinal mobile and wearable sensing data.

3.2 Intelligent Intervention

Ubiquitous devices should not only understand our behavior, but more importantly, we need to find an appropriate way to loop the models back to users. One promising direction is to provide intelligent interventions that can positively affect our behavior and help us to achieve our goals, achieving just-in-time adaptive intervention (JITAI). Nevertheless, the algorithmic aspect of the smart delivery of JITAIs is still in its infancy. Recent research uses simple multi-arm bandits to generate dynamic interventions, but they are not interpretable or personalized enough [10]. This leads to my next research question: **How to leverage the interpretability, personalization, and generalizability in behavior modeling to better design intervention techniques?**

Smartphone overuse is related to a variety of issues such as physical health and mental health [4]. Depression is one of them. Meanwhile, my works on interpretable behavior rules from our datasets illustrate that depressive symptoms are positively correlated to frequent smartphone overuse [15, 16]. Therefore, I start with smartphone overuse intervention as an initial step. Recently, we explored the application of Self-Affirmation Theory on smartphone overuse intervention in a just-in-time manner to reduce smartphone usage [24] (see Fig. 5). We developed an intervention technique that integrates two components: an in-situ typing-based unlock process to improve user engagement, and personalized, self-affirmation-based typing content to enhance effectiveness. Users first go through a list of value items and select those they think are important to themselves, resulting in a personalized value list for each individual. Then, when users tend to overuse their phones (*e.g.*, staying up late to browse video streams), they need to first type two short sentences with persuasive content designed based on Self-Affirmation Theory They can choose to quit the app, or access the app after they finish typing. Our 10-week field experiment shows that our technique could reduce app usage by over 50%, and both app opening frequency and usage duration by over 25%, all significantly outperforming baseline intervention techniques.

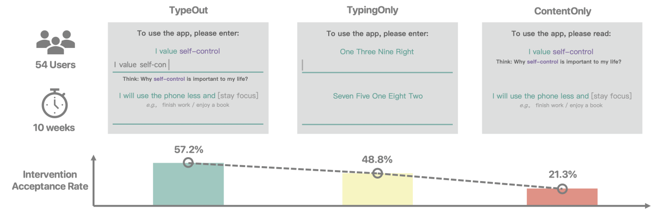

Figure 5: Overview of TypeOut Intervention Technique.

These results indicate the effectiveness of our method, and illuminate the next step of leveraging ML techniques to achieve a truly JITAI [8]. Our current technique enables a simple personalization (*i.e.*, tailored value list), with the content randomly sampled from a set of pre-designed templates and delivered whenever one of the apps pre-determined by users is launched. Both the content and timing can be further improved via better interpretability and personalized adaptivity. I am leading another group of researchers to leverage the algorithm in my previous work [15, 19] to reveal the relationship between users' routines and intervention reactions under different contexts. Based on the interpretation, I aim to achieve several goals: 1) leveraging adaptive ML techniques such as few-shot learning to determine the optimal delivery time of JITAIs for each individual, 2) exploring reinforcement learning methods such as Multi-arm Bandit or Q-learning that can determine the contents of JITAIs, and 3) providing a conversational intervention experience based on the interpretation. Combining these technical and design pieces, we can build a system that evolves based on each individual's behavior, with its effectiveness improving over time.

4 CONCLUSION

With the ecosystem of ubiquitous devices becoming more and more mature, it provides an unprecedented opportunity to understand and facilitate our daily lives for better well-being. However, interpretability and robustness are still big challenges for behavior modeling and intervention techniques. My work aims to address these challenges via better ML algorithms and interaction designs, so that I can bridge behavior modeling and intervention to advance the future technology for health and well-being.

REFERENCES

[1] Sangwon Bae, Denzil Ferreira, Brian Suffoletto, Juan C. Puyana, Ryan Kurtz, Tammy Chung, and Anind K. Dey. 2017. Detecting drinking episodes in young adults using smartphone-based sensors. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 2, Article 5 (Jun 2017), 36 pages. https://doi.org/10.1145/3090051

[2] Luca Canzian and Mirco Musolesi. 2015. Trajectories of depression: unobtrusive monitoring of depressive states by means of smartphone mobility traces analysis. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous computing*. ACM, 1293–1304.

[3] Taesik Gong, Yewon Kim, Adiba Orzikulova, Yunxin Liu, Sung Ju Hwang, Jinwoo Shin, and Sung-Ju Lee. 2021. DAPPER: Performance Estimation of Domain Adaptation in Mobile Sensing. *arXiv preprint arXiv:2111.11053* (2021).

[4] Joshua Harwood, Julian J Dooley, Adrian J Scott, and Richard Joiner. 2014. Constantly connected–The effects of smart-devices on mental health. *Computers in Human Behavior* 34 (2014), 267–272.

[5] Joy He-Yueya, Benjamin Buck, Andrew Campbell, Tanzeem Choudhury, John M Kane, Dror Ben-Zeev, and Tim Althoff. 2020. Assessing the relationship between routine and schizophrenia symptoms with passively sensed measures of behavioral stability. *NPJ schizophrenia* 6, 1 (2020), 1–8.

[6] Nicholas D Lane, Emiliano Miluzzo, Hong Lu, Daniel Peebles, Tanzeem Choudhury, and Andrew T Campbell. 2010. A survey of mobile phone sensing. *IEEE Communications magazine* 48, 9 (2010), 140–150.

[7] Jun-Ki Min, Afsaneh Doryab, Jason Wiese, Shahriyar Amini, John Zimmerman, and Jason I. Hong. 2014. Toss "n" turn: Smartphone as sleep and sleep quality detector. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) *(CHI '14)*. Association for Computing Machinery, New York, NY, USA, 477–486. https://doi.org/10.1145/2556288.2557220

[8] Inbal Nahum-Shani, Shawna N Smith, Bonnie J Spring, Linda M Collins, Katie Witkiewitz, Ambuj Tewari, and Susan A Murphy. 2018. Just-in-time adaptive interventions (JITAIs) in mobile health: key components and design principles for ongoing health behavior support. *Annals of Behavioral Medicine* 52, 6 (2018), 446–462.

[9] Charith Perera, Arkady Zaslavsky, Peter Christen, and Dimitrios Georgakopoulos. 2013. Context aware computing for the internet of things: A survey. *IEEE communications surveys & tutorials* 16, 1 (2013), 414–454.

[10] Mashfiqui Rabbi, Min Hane Aung, Mi Zhang, and Tanzeem Choudhury. 2015. MyBehavior: automatic personalized health feedback from user behaviors and preferences using smartphones. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. 707–718.

[11] Rui Wang, Fanglin Chen, Zhenyu Chen, Tianxing Li, Gabriella Harari, Stefanie Tignor, Xia Zhou, Dror Ben-Zeev, and Andrew T Campbell. 2014. StudentLife: assessing mental health, academic performance and behavioral trends of college students using smartphones. In *Proceedings of the 2014 ACM international joint conference on pervasive and ubiquitous computing*. ACM, 3–14.

[12] Rui Wang, Gabriella Harari, Peilin Hao, Xia Zhou, and Andrew T Campbell. 2015. SmartGPA: how smartphones can assess and predict academic performance of college students. In *Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing*. 295–306.

[13] Rui Wang, Weichen Wang, Alex daSilva, Jeremy F Huckins, William M Kelley, Todd F Heatherton, and Andrew T Campbell. 2018. Tracking depression dynamics in college students using mobile phone and wearable sensing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 1 (2018), 43.

[14] Yaqing Wang, Quanming Yao, James T. Kwok, and Lionel M. Ni. 2020. Generalizing from a Few Examples: A Survey on Few-shot Learning. 53, 3 (2020), 1–34. https://doi.org/10.1145/3386252

[15] Xuhai Xu, Prerna Chikersal, Afsaneh Doryab, Daniella K. Villalba, Janine M. Dutcher, Michael J. Tumminia, Tim Althoff, Sheldon Cohen, Kasey G. Creswell, J. David Creswell, Jennifer Mankoff, and Anind K. Dey. 2019. Leveraging Routine Behavior and Contextually-Filtered Features for Depression Detection among College Students. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 3 (Sept. 2019), 1–33. https://doi.org/10.1145/3351274

[16] Xuhai Xu, Prerna Chikersal, Janine M. Dutcher, Yasaman S. Sefidgar, Woosuk Seo, Michael J. Tumminia, Daniella K. Villalba, Sheldon Cohen, Kasey G. Creswell, J. David Creswell, Afsaneh Doryab, Paula S. Nurius, Eve Riskin, Anind K. Dey, and Jennifer Mankoff. 2021. Leveraging Collaborative-Filtering for Personalized Behavior Modeling: A Case Study of Depression Detection among College Students. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 1 (March 2021), 1–27. https://doi.org/10.1145/3448107

[17] Xuhai Xu, Jun Gong, Carolina Brum, Lilian Liang, Bongsoo Suh, Shivam Kumar Gupta, Yash Agarwal, Laurence Lindsey, Runchang Kang, Behrooz Shahsavari, Tu Nguyen, Heriberto Nieto, Scott E Hudson, Charlie Maalouf, Jax Seyed Mousavi, and Gierad Laput. 2022. Enabling Hand Gesture Customization on Wrist-Worn Devices. In *CHI Conference on Human Factors in Computing Systems*. ACM, New Orleans LA USA, 1–19. https://doi.org/10.1145/3491102.3501904

[18] Xuhai Xu, Ahmed Hassan Awadallah, Susan T. Dumais, Farheen Omar, Bogdan Popp, Robert Rounthwaite, and Farnaz Jahanbakhsh. 2020. Understanding User Behavior For Document Recommendation. In *Proceedings of The Web Conference 2020*. ACM, Taipei Taiwan, 3012–3018. https://doi.org/10.1145/3366423.3380071

[19] Xuhai Xu, Xin Liu, Han Zhang, Weichen Wang, Subgiya Nepal, Kevin S. Kuehn, Jeremy Huckins, Margaret E. Morris, Nurius, Paula S., Eve A. Riskin, Shwetak Patel, Tim Althoff, Andrew Campbell, Anind K. Dey, and Jennifer Mankoff. 2022. GLOBEM: Cross-Dataset Generalization of Longitudinal Human Behavior Modeling (under major revision). 1–33.

[20] Xuhai Xu, Jennifer Mankoff, and Anind K. Dey. 2021. Understanding practices and needs of researchers in human state modeling by passive mobile sensing. *CCF Transactions on Pervasive Computing and Interaction* (July 2021). https://doi.org/10.1007/s42486-021-00072-4

[21] Xuhai Xu, Ebrahim Nemati, Korosh Vatanparvar, Viswam Nathan, Tousif Ahmed, Md Mahbubur Rahman, Daniel McCaffrey, Jilong Kuang, and Jun Alex Gao. 2021. Listen2Cough: Leveraging End-to-End Deep Learning Cough Detection Model to Enhance Lung Health Assessment Using Passively Sensed Audio. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 1 (March 2021), 1–22. https://doi.org/10.1145/3448124

[22] Xuhai Xu, Chun Yu, Yuntao Wang, and Yuanchun Shi. 2020. Recognizing Unintentional Touch on Interactive Tabletop. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 1 (March 2020), 1–24. https://doi.org/10.1145/3381011

[23] Xuhai Xu, Han Zhang, Yasaman S Sefidgar, Yiyi Ren, Xin Liu, Woosuk Seo, Jennifer Brown, Kevin S. Kuehn, Mike Merrill A., Paula S. Nurius, Shwetak Patel, Tim Althoff, , Margaret E. Morris, Eve A. Riskin, Jennifer Mankoff, and Anind K. Dey. 2022. GLOBEM: Multi-Year Datasets for Longitudinal Human Behavior Modeling Generalization (under submission). *Advances in neural information processing systems* (2022).

[24] Xuhai Xu, Tianyuan Zou, Han Xiao, Yanzhang Li, Ruolin Wang, Tianyi Yuan, Yuntao Wang, Yuanchun Shi, Jennifer Mankoff, and Anind K Dey. 2022. TypeOut: Leveraging Just-in-Time Self-Affirmation for Smartphone Overuse Reduction. In *CHI Conference on Human Factors in Computing Systems*. ACM, New Orleans LA USA, 1–17. https://doi.org/10.1145/3491102.3517476