STRETCH: Stress and Behavior Modeling with Tensor Decomposition of Heterogeneous Data

Chunpai Wang
cwang25@albany.edu
University at Albany, State University of New York
Albany, New York, USA

Shaghayegh Sahebi
ssahebi@albany.edu
University at Albany, State University of New York
Albany, New York, USA

Helma Torkamaan
helma.torkamaan@uni-due.de
University of Duisburg-Essen
Germany

ABSTRACT

Stress level modeling and predictions are essential in recommending activities and interventions to individuals. While successful stress models have been proposed in the literature, there is still a missing connection between user engagement behaviors, interest in activities, and their stress levels. In this paper, we propose a novel multi-view tensor decomposition method for stress and user behavior modeling with heterogeneous data, which could provide personalized stress tracking and plausible user behavior modeling across time. To the best of our knowledge, it is the first method that could model user stress and behavior at the same time with multiple resources of data, such as stress measurement, activity rating, and engagement. Our experiments show that leveraging multiple resources of data could not only improve predictions with sparse data, but also results in discovering the underlying stress-activity patterns. We demonstrate the effectiveness of our proposed model on the dataset collected via a self-contained stress management mobile application.

CCS CONCEPTS

• **Applied computing** → **Health care information systems**; **Health informatics**; **Consumer health**.

KEYWORDS

stress management, behavior modeling, tensor decomposition

ACM Reference Format:
Chunpai Wang, Shaghayegh Sahebi, and Helma Torkamaan. 2021. STRETCH: Stress and Behavior Modeling with Tensor Decomposition of Heterogeneous Data. In *IEEE/WIC/ACM International Conference on Web Intelligence (WI-IAT '21), December 14–17, 2021, ESSENDON, VIC, Australia.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3486622.3493967

1 INTRODUCTION

Stress, as a fundamental phenomenon, has been defined as "*a particular relationship between the person and the environment that is appraised by the person as taxing or exceeding his or her resources and endangering his or her well-being*" [20, p.19]. Stress has a direct influence on the physical and psychological health of an individual, her (or his) cognition, and performance [19, 41]. As a result, the study and tracking of user stress and mood [5, 17, 22, 44, 46, 50], predicting them [1, 16, 18, 21, 22, 37, 42], and presenting proper interventions to help individuals manage them better [24, 26, 29, 36, 43], have been a growing topic of research. Particularly, studying stress level and its changes in association with the individual's activities and behaviors is needed for suggesting effective interventions to them. Ideally, considering that each individual's stress response to an activity may be different, a personalized model is needed to study such an association. For example, calling a friend may reduce the stress levels in some, while being ineffective for others. So, recommending the same one-size-fits-all activity to everyone would not be effective. Additionally, such association between an activity and stress levels can be due to multiple underlying factors, such as activity type and the coping mechanisms that it provides. Such underlying factors can be also associated with the individual's personality, e.g. calling a friend (as a social activity with the distancing coping mechanism) may be more effective in reducing the stress levels of an extrovert person compared to an introvert. Consequently, understanding the underlying processes between activities and stress levels would lead to a better choice of intervention recommendations, especially in the face of data sparsity.

Furthermore, in this paper, we argue that, an individual's preferences and interests play an additional role in their engagement to perform a recommended activity and its effectiveness on their stress levels. For example, one may prefer mindful breathing to guided imagery, although they both are individual activities with self-controlling strategy types and the same difficulty levels. Additionally, such an association can be reciprocal: not only one's activity interest and engagement can be predictive of their future stress levels, their current stress levels may be related to the choice of their next activity. While a few current literature has focused on creating personalized stress prediction models [15], personalized interventions [6], and finding the associations between stress levels and activities [24], none of the current studies have modeled a personalized association model that can be used to both recommend interesting activities to individuals and discover the underlying stress-activity patterns. More importantly, the reciprocal relationship between stress level, activity interest, and engagement has not been studied before.

In this paper, we propose a holistic model for simultaneously representing activity interest, engagement, and stress level over time for users of a stress-management mobile application. Our model is personalized, as it finds user-specific latent factors, can find the

WI-IAT '21, December 14–17, 2021, ESSENDON, VIC, Australia

ACM ISBN 978-1-4503-9115-3/21/12…$15.00
https://doi.org/10.1145/3486622.3493967

related underlying latent factors in activities, and track the stress change patterns across time. To this end, we propose a multi-view tensor and matrix factorization model that share user, activity, and stress-change latent factors. Sharing these factors across the stress, interest, and engagement views allows our model to tackle the cold-start problem; e.g., to predict stress levels for users who have never reported their stress levels, only based on their past activities. Additionally, our experiments on the real-world application dataset show that our proposed multi-view model results in better prediction performance in all three modeled data views and discovers meaningful patterns and clusters among users and activities.

2 RELATED WORK

Mental Health Recommender Systems. Health Recommender Systems is relatively new compared to other areas of recommender systems, being discussed mostly for the past five years [8, 33, 39], with isolated concept-level efforts in late 2000s [3, 35]. Mental well-being and health promotion are relatively underrepresented in the literature of recommender systems. Related work often focus on the curative aspects, such as therapy [38], patients with depression [12, 30, 48, 49], or supporting counselors for suicide preventions [34]. However, preventive measures are as important, if not more, as the curative efforts. And yet, the contributions related to health recommender systems are limited to a few instances. Design of health recommender systems for stress [6, 13, 25, 45], general well-being [24, 29], happiness [10], emotion regulation [2], and depression [12, 31, 48], are example domains of mental health promotion-related contributions. These contributions mostly use mobile devices and rely on these devices' pervasiveness and capabilities in functioning as both the tracking tool and interaction medium [6, 24, 25, 28, 30, 48, 49]. Among them, Paredes et al. [25] and Clarke et al. [6] focused specifically on stress and building a stress intervention system. They used variants of reinforcement learning on less than 20 total interventions and relied on the stress value in the objective function.

Stress Prediction Models. For detecting mood and stress, self-reported measures, such as [7], have been traditionally the standard tool for estimating the subjective user experience. However, primarily because of user fatigue and compliance issues, non-self-reported approaches, such as physiological, behavioral, or environmental sensor data have been increasingly used to predict stress, e.g., [27]. Most such not-self-reported approaches rely on wearables and physiological sensor data, e.g., [1, 4, 6, 14, 23, 37]. However, this kind of data has challenges, such as fragment issues [16] and convenience for users in general. Ultimately, the most convenient way of determining users' experience of stress is passive detection or prediction. Accordingly, various learning techniques have been investigated to predict stress. Such prediction-based research is only concerned with real-time estimations, e.g., [11]; however, stressful experiences and mood generally are time-dependent concepts. Therefore, a growing share of the related work models stress with time-based considerations, particularly using wearables and physiological sensor data, e.g., [1, 4, 6, 14, 16, 23, 37]. For example, time-based considerations are sometimes also used in non-physiological data, such as mobile sensor data namely, containing location, activity, and conversation [37, 40, 50].

Table 1: Descriptive Statistics of Experimental Dataset.

Variables	Size	Mean	SD	Min.	Med.	Max.
Users	53	-	-	-	-	-
Days	14	-	-	-	-	-
Items	49	-	-	-	-	-
Users' age	53	37.71	11.55	18	33	61
Users' coping strategy score	52	76.36	24.39	0	80	136
Activity rating records	134	3.88	1.11	1	4	5
Activity rating records per item	46	2.91	1.79	1	3	10
Average activity rating per item	46	3.87	0.77	1	4	5
Average activity records per user	33	4.06	3.43	1	3	15
Average activity rating per user	33	3.83	0.69	2.66	4	5
Stress reported	188	2.10	0.91	1.00	1.96	5
Stress reported records per user	39	4.82	3.03	1	4	11
Average stress reported per user	39	2.09	0.80	1.14	2.01	4.80
Open records	408	-	-	-	-	-
Engagement records	196	-	-	-	-	-
Open records per user	45	9.06	6.34	1	7	34
Engagement records per user	45	4.35	4.64	0	4	20
Open records per activity	49	8.33	4.37	1	8	23
Engagement records per activity	46	4.26	3.33	1	3	18

3 STRESS AND BEHAVIOR MODELING

3.1 Application and Data

We used data collected as part of a larger longitudinal study with a self-contained mobile platform and application called **PAX**. As a result of this running study, a preliminary dataset, for health promotion and mental well-being based on the random recommendation of stress coping activities, were built. This dataset contains a combination of various smartphone-based data, including self-reported age, mood, and stress data collected through experience sampling method and app usage log data. In addition, users completed a pre-study questionnaire assessing their general stress coping skills using ways of coping instrument [9] and perceived stress level over the past two weeks using instrument of [7] (CS score). Using the app, various stress-reducing activities were randomly presented to users at different points in time. Each activity has a corresponding description, labeled types, and strategies. There are five general types of activity items, including *Social Engagement*, *Physical Activities*, *Mindfulness*, *Positive Thinking*, and *Enjoyable Activities*. Additionally, there are different item coping strategy types including *Distancing*, *Seeking Social Support*, *Planful Problem Solving*, *Accepting Responsibility*, *Confront the Problem*, *Positive Reappraisal*, *Self-controlling*, and *Accepting the Problem*. Users could open the suggested activities, follow them, mark them as "done", and afterwards, provide ratings for them. They could also close the recommendation session, ignoring the suggested activities.

For experiments, we select users who use the **PAX** application frequently and have at least four records on either one view or type of data. Eventually, we end up with 53 users for our experiments. The descriptive data statistics are shown in Table 1. As shown in the table, there are only 134 rating records, 188 stress records, and 408 engagement records on 53 users, 49 items, and 14 days of usage. The data is extremely sparse, and not all users have records for all three views. Although all activities had been opened at least once by users, not all activities got engaged with and received user ratings. In addition, there are only 33 out of 53 users who ever rated at least one activity and 39 out of 53 users who ever reported their stress levels.

3.2 Problem Formulation and Assumptions

Assume we have U users, T time points, and A stress-coping activities or items. Input includes three types of data: the sequence of activities marked as "done" by each user (user engagement), sequences of stress scores calculated for each user across different time points (stress levels), and user ratings on the recommended activities (user interest). We model the activity sequences as a $U \times T \times A$ binary tensor $\mathcal{X}$, in which $x_{u,t,a} = 1$ means that the user u has interacted with suggested activity a at a time point t, and $x_{u,t,a} = 0$ means that the user u has ignored the suggested activity and has not engaged with it. Note that in some cases $x_{u,t,a}$ is missing and unknown, e.g., because a has not been suggested to u at t. We model the stress score sequence as a $U \times T$ matrix Y, and the user-activity ratings as a $U \times A$ matrix Z. Similar to $\mathcal{X}$, matrices Y and Z are sparse and noisy, include a high percentage of missing values, and a few observations.

We assume that users can be more (or less) similar to each other based on how they cope with stress, what activities they like, and what they engage in. Hence, we consider users to be represented by underlying latent features that describe such patterns. These features can be used to softly cluster the users into different groups, with similar users having closer latent features to each other. Additionally, we consider the activities to be represented by similar latent features, representing their usefulness in helping to cope with stress, their appeal to users, and the effort needed to engage with them, among other factors. We regard users' stress levels and preference in engaging with different activities (marking them as "done") to be dynamic over time. These dynamics can be more similar within some groups of users over similar activities. Finally, we regard the underlying patterns deriving user interest in activities, their engagement with the activities, and their stress level dynamics to be related. A user's stress level will not change as a result of an activity, unless the user engages in that activity. We also consider this change and the user's choice in doing an activity to be related to how much the user enjoys that activity. This results in having "interrelated patterns" among the three types of data input.

3.3 STRETCH Model

Here, we propose a novel stress, rating, and engagement prediction model according to the above assumptions: *STR*ess and behavior modeling with *T*ensor de*C*omposition of *H*eterogeneous data (*STRETCH* [1]). We represent user u's latent features as an N-dimensional feature vector $\mathbf{s}_u$. Similarly, we use an N-dimensional activity feature vector $\mathbf{q}_a$ to represent activity a's latent features. Having that user engagement with activities may change over time, we attribute this change to the dynamics in user and activity latent features. In other words, we would like to model the way user latent features evolves with activity latent features over time, with respect to how users engage in those activities. As an example of such dynamics, suppose that a group of users who are generally interested in doing social activities start engaging with the easier social activities first, and end up doing the ones that require more efforts. To

[1]The source code is provided at https://github.com/persai-lab/WIIAT2021-STRETCH

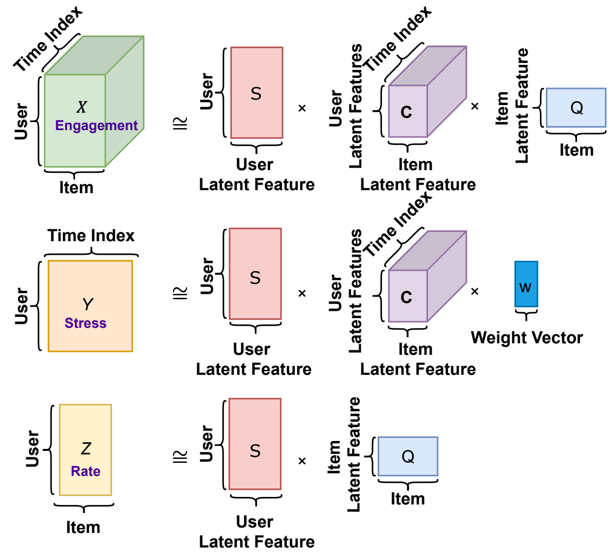


Figure 1: STRETCH Model

model these dynamics over time, we use an $N \times N \times T$ tensor C. The t^{th} slice of this tensor (time matrix C_t) shows the relationship between user latent features and item latent features at time step t. As a result, we can represent the user u's choice in engaging with activity a at time step t (estimated binary value $\hat{x}_{u,t,a}$) as a combination of user latent features, item latent features, and latent feature engagement dynamics at time point t, as in Equation 1.

$$\hat{x}_{u,t,a} \approx \sigma(\mathbf{s}_u \cdot C_t \cdot \mathbf{q}_a + b_{x_u} + b_{x_t} + b_{x_a}) \tag{1}$$

Parameters b_{x_u}, b_{x_t}, and b_{x_a} are intercepts representing user, time point, and item biases in the model, and $\sigma(.)$ represents the Sigmoid function.

At the same time, considering our last assumption of interrelated patterns, we set the same underlying user interests and activity properties to drive users' ratings of activities. Accordingly, we model the estimated rating value $\hat{z}_{u,a}$ of user u on activity a as in Equation 2, where b_{z_u} and b_{z_a} respectively represent user rating bias (e.g., if the users usually rates the items highly) and activity rating bias (e.g., activity popularity).

$$\hat{z}_{u,a} \approx \mathbf{s}_u \cdot \mathbf{q}_a + b_{z_u} + b_{z_a} \tag{2}$$

Similarly, we assume that user stress levels are related to user features, the activities they have done, and as a whole, how much each activity feature is associated with user stress levels. Hence, we estimate user stress level at time step t as a combination of user latent feature $\mathbf{s}_u$, latent feature engagement dynamics at that time C_t, and the overall association between each latent feature and stress levels $\mathbf{w}$ (Equation 3).

$$\hat{y}_{u,t} \approx \mathbf{s}_u \cdot C_t \cdot \mathbf{w} + b_{y_u} + b_{y_t} \tag{3}$$

Here, $\mathbf{w}$ is an N-dimensional vector working as a weighted average to calculate an overall value of the engaged activity's latent features for users' stress levels; b_{y_u} and b_{y_t} are biases to capture overall user stress levels and overall stress trends.

Tensor C captures the dynamics in user engagement and stress level changes. Assuming that these changes are slow and smooth over time steps, we regularize the change in C. Particularly, using a regularization constraint, we enforce the L_2 norm between consecutive time point slices in C to be small. Alternatively, assuming that the C values over time steps can vary dramatically, we do not use such a regularization term in our model. In our experiments, we will study the effect of having this regularization on model results. Figure 1 illustrates our full model. Using the same user, activity, and dynamics' latent features across different views of the data, not only is consistent with our assumptions, but also helps our model in dealing with data sparsity, noisiness, and missing user data. For example, if user data is missing in one view, e.g., if a user has not rated any activities, the learned user factors from the two other views will be transferred and used to estimate the missing data, e.g., the target user's ratings.

To learn the model parameters, the objective would be to minimize the difference between observed and estimated x, y, and z values, formulated in Equation 4. To add to the generalizability and interpretability of our model, we impose L_2 norm regularization on parameters in S, Q, $\mathbf{w}$, C_t, and the vector for all biases $\mathbf{b}$. Here $\lambda_s, \lambda_q, \lambda_w, \lambda_b$, and λ_C are regularization weight or trade-off hyperparameters, which will be tuned over the training process. Finally, we control addition of the time smoothness regularization of C using the hyperparameter η.

$$\begin{aligned}\arg\min_{S,\mathbf{C},Q,\mathbf{w},\mathbf{b}} & \sum_{x_{u,t,a}\in\Omega_X} (x_{u,t,a} - \hat{x}_{u,t,a})^2 + \sum_{y_{u,t}\in\Omega_Y} (y_{u,t} - \hat{y}_{u,t})^2 \\ & + \sum_{z_{u,a}\in\Omega_Z} (z_{u,a} - \hat{z}_{u,a})^2 + \lambda_s \sum_u \|\mathbf{s}_u\|_2^2 + \lambda_q \sum_a \|\mathbf{q}_a\|_2^2 + \\ & \lambda_w \|\mathbf{w}\|_2^2 + \lambda_b \|\mathbf{b}\|_2^2 + \lambda_C \sum_t \|C_t\|_F^2 + \eta \sum_{t=1}^{T} \|C_t - C_{t-1}\|_F^2\end{aligned} \tag{4}$$

Ω_X, Ω_Y, and Ω_Z denote all observed user records on activity engagement, stress measurement, and activity rating, respectively.

4 EXPERIMENTS

We design our experiments to investigate these research questions:
- RQ1: Can our proposed multiview model effectively predict user's sequential engagement, stress-level, and rating behavior with few and sparse observations?

- RQ 1.1. How good our proposed multiview model's fit is?
- RQ 1.2. How does each view of the data affect the prediction performance in all views?
- RQ 1.3. How does the time step smoothness assumption affect the prediction performance in all views?

- RQ2: Can our proposed multiview model provide insights and interpretations of user behavior and mental health data?

- RQ 2.1. How do activity groups found by our model associate with their types, popularity, and engagement?
- RQ 2.2. How do user groups found by our model associate with their demographics, behaviors, and stress levels?
- RQ 2.3. How do user groups found by our model associate with their activity preferences, activity characteristics, and activity popularity?

Table 2: Best Hyperparameters of Each Prediction.

	N	λ	η	learning rate	min epoch
Stress Prediction	4	0.001	0.1	0.01	4
Rate Prediction	1	0.001	0.1	0.1	4
Engagement Prediction	2	0.001	0	0.1	4

Table 3: Stress Prediction Results.

Methods	Stress Prediction	
	RMSE	MAE
Global Running Average (GRA)	$0.834 \pm 0.307^{***}$	$0.692 \pm 0.223^{***}$
User Running Average (URA)	$0.487 \pm 0.059^{*}$	$0.383 \pm 0.043^{***}$
STRETCH-SV (Stress)	$0.605 \pm 0.108^{***}$	$0.478 \pm 0.073^{***}$
STRETCH-TV (Stress+Rate)	$0.435 \pm 0.210^{**}$	$0.297 \pm 0.142^{**}$
STRETCH-TV (Stress+Engag.)	$0.428 \pm 0.175^{**}$	$0.303 \pm 0.125^{**}$
STRETCH-SV-NS (No-Smooth Stress)	$0.934 \pm 0.117^{***}$	$0.770 \pm 0.118^{***}$
STRETCH-FULL-NS (3-view+NoSmooth)	$0.487 \pm 0.245^{**}$	$0.354 \pm 0.179^{**}$
STRETCH-FULL (3-view)	$\mathbf{0.373 \pm 0.201}$	$\mathbf{0.261 \pm 0.131}$

4.1 Experimental Setup

In experiments, we would like to validate the performance of our proposed multiview model on the tasks of stress prediction, rate prediction, and engagement prediction.

Online Testing and Hyperparameter Tuning. We leverage the nested 5-fold user-stratified cross-validation for hyperparameter tuning and performance evaluation. We first split users into five folds. Three folds of users are training users, one fold is validation, and the last fold is testing. We use validation users' records for hyperparameters tuning and testing users' records for performance evaluation. To find the optimal hyperparameters for each prediction task, we apply grid search separately, and report the best hyperparameters we find in Table 2. N is the dimension of latent factors. λ is the regularization trade-off parameter. η is the smoothness weight. *min-epoch* is the minimum number of epochs to run before the early stop. As we can see, due to the sparsity of our dataset, the dimension of latent factors is small as expected.

For testing, we predict the last 50% of testing users' records in the target view, given their first 50% records. We predict each test user's stress level, activity rating, and activity engagement over time as in the real-world scenario, to avoid the information leak between time steps in different views. In other words, we do not use any user's records from any resource after the testing time point for training, and we will merge the outdated testing data before the testing time point into the training data. In addition, due to online training and testing, we leverage the prediction performance on validation data for early stopping to avoid over-fitting.

4.2 Performance Experiment Results

In this section, we answer RQ 1.1., RQ 1.2., and RQ 1.3. Specifically, we evaluate the prediction performance of our proposed model in stress level prediction, user engagement prediction, and user rating prediction. Given that stress level and rating data are numeric, we use the root mean squared error (RMSE) and mean absolute error (MAE) metrics for evaluating their predictions. To evaluate the performance of engagement prediction, we use the ROC curve (AUC) and accuracy (ACC).

Table 4: Rating Prediction Results.

Methods	Rating Prediction	
	RMSE	MAE
Global Running Average (GRA)	$1.081 \pm 0.398^{***}$	$0.880 \pm 0.379^{***}$
User Running Average (URA)	$0.859 \pm 0.330^{*}$	$0.624 \pm 0.259^{*}$
Item Running Average (IRA)	$0.921 \pm 0.193^{**}$	$0.729 \pm 0.145^{***}$
STRETCH-SV (Rate)	$1.455 \pm 0.167^{***}$	$1.234 \pm 0.154^{***}$
STRETCH-TV (Rate+Stress)	$0.794 \pm 0.290^{***}$	$0.525 \pm 0.151^{***}$
STRETCH-TV (Rate+Engag.)	0.674 ± 0.302	$0.484 \pm 0.180^{*}$
STRETCH-FULL (3-view)	$\mathbf{0.614 \pm 0.231}$	$\mathbf{0.417 \pm 0.127}$

Table 5: Engagement Prediction Results.

Methods	Engagement Prediction	
	AUC	ACC
Global Running Average (GRA)	$0.652 \pm 0.094^{***}$	$0.659 \pm 0.019^{***}$
User Running Average (URA)	$0.748 \pm 0.141^{***}$	$0.555 \pm 0.088^{***}$
Item Running Average (IRA)	$0.814 \pm 0.098^{***}$	$0.736 \pm 0.090^{***}$
STRETCH-SV (Engag.)	$0.643 \pm 0.090^{***}$	$0.708 \pm 0.061^{***}$
STRETCH-TV (Engag.+Rate)	0.955 ± 0.038	$0.890 \pm 0.051^{**}$
STRETCH-TV (Engag.+Stress)	$\mathbf{0.967 \pm 0.036}$	$\mathbf{0.918 \pm 0.034}$
STRETCH-SV-NS (No-Smooth Engag.)	$0.643 \pm 0.090^{***}$	$0.708 \pm 0.061^{***}$
STRETCH-FULL-NS (3-view + No-Smooth)	$0.925 \pm 0.039^{**}$	$0.834 \pm 0.076^{***}$
STRETCH-FULL (3-view)	$0.925 \pm 0.039^{**}$	$0.834 \pm 0.076^{***}$

To answer each of the research questions under RQ1, we compare our proposed model's performance in all three views (STRETCH-FULL) with the following **baseline methods**:

- `Global Running Average (GRA)`: uses the average over all users' observed records of target view as prediction value (for RQ 1.1.).
- `User Running Average (URA)`: uses the average of the target user's observed records of target view as prediction value (RQ 1.1.).
- `Item Running Average (IRA)`: uses the average of the target item's observed records of target view as prediction value (RQ 1.1.).
- `STRETCH-SV Single View Model`: is based on our model, but only using the single view of data to train the model (RQ 1.2.).
- `STRETCH-TV Two View Model`: is based on our model, but using two views of the data (target view with one additional view) to train the model (RQ 1.2.).
- `STRETCH-NS No-Smooth Model`: is based on our model, but does not impose the smoothing regularization term on tensor C (RQ 1.3.).

We report the average performance as well as 95% confidence interval for tasks of engagement prediction, stress prediction, and rating prediction in Tables 3, 4, and 5, respectively. We show the best performance with bold letters and use ***, **, and * to indicate if the best model is significantly outperforming others with p-values less than 0.05, 0.10, and 0.15, respectively.

RQ 1.1., Model Fit. Looking at the global running average (GRA), user running average (URA), and item running average (IRA) results, we can see that these baselines are performing relatively well in all tasks. Specifically, URA performs the best among the average baselines for stress and rating predictions and IRA performs better than GRA and URA in engagement prediction. Having more than 73% accuracy in predicting user future user engagement, 0.85 RMSE in rating prediction, and 0.48 RMSE in stress level prediction shows that these baselines are competitive and strong. For example, Table 4 shows that these approaches perform better than STRETCH-SV (the single-view model) in rating prediction. One reason for that could be the time-based patterns in the data that cannot be captured by the matrix-based STRETCH-SV model for rating prediction. IRA, GRA, and URA change over time steps as running averages and can pick up some time-based trends in the data. Comparing our proposed three-view STRETCH-FULL model with these baselines, we see that STRETCH-FULL significantly outperforms all three of IRA, GRA, and URA. This shows that our model fits the data well, and that the data has more complexities than that can be captured using an average trend. It also shows that STRETCH-FULL can do well in predicting the future values in all three views of the data.

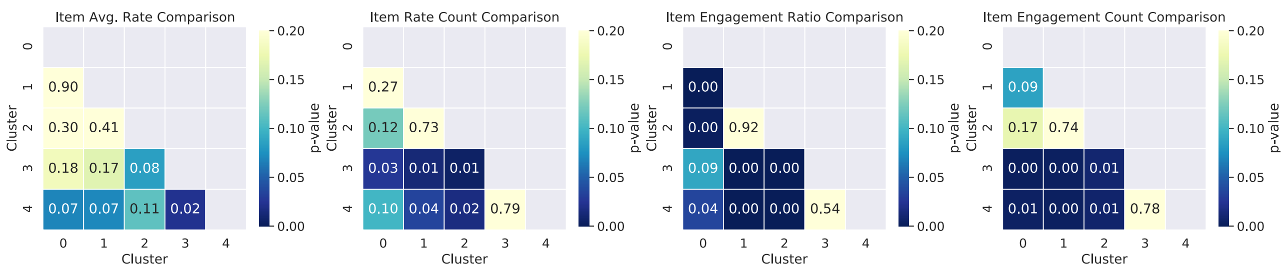
RQ 1.2., Multiview Helpfulness. To assess if modeling multiple views to the data help in the prediction performance of our proposed model, we compare STRETCH-SV (the single-view model) and STRETCH-TV (the two-view model) results with STRETCH-FULL. For stress prediction, as we could see in Table 3, the STRETCH-FULL model achieve the best prediction performance, significantly better than the other two versions of the model. Additionally, STRETCH-TV results are better than STRETCH-SV results. It shows that modeling both rating and engagement views of the data are helpful in predicting future stress levels, compared to not using them at all. The engagement data is also more helpful than the rating data in predicting users' stress levels. Furthermore, modeling both engagement and rating views coupled with stress view is even more helpful compared to adding only one of these auxiliary views.

For rating prediction (Table 4), both STRETCH-TV models perform better than STRETCH-SV, which shows that modeling two views of the data, whether adding stress or engagement modeling, results in better rating predictions than only modeling the ratings alone. Our STRETCH-FULL model achieves the lowest prediction error compared to all STRETCH-SV and STRETCH-TV models. However, for STRETCH-TV, this improvement is more significant when having the stress data as the auxiliary view. Overall, it shows that the engagement data is more beneficial than the stress-level data here. One reason could be having more user records on the engagement data, compared to stress-level records. As another potential reason, the rating data could be more related to the engagement data as two views of the user behavior, e.g., the user is likely to rate an activity after engaging with it.

The results are slightly different in the engagement prediction task. Similar to stress and rating prediction, training the model with only engagement data (STRETCH-SV) results in relatively poor predictions. With modeling any additional view (both STRETCH-TV models), the performance boosts relatively 50%, with the stress data being more useful than rating data. But, modeling all three views of the data together (STRETCH-FULL) does not help more than STRETCH-TV models. In fact, STRETCH-TV with stress data is significantly better than STRETCH-FULL in engagement predictions. It seems that using both stress-level and rating data would result in some conflicts in the learned model. However, this conflict is not large enough to completely cancel out the boost that is gained, compared to not using any of these two views (in the STRETCH-SV model). One potential reason could be the richness of engagement data compared to the other two views. Accordingly, the shared parameters learned in each of the STRETCH-TV models would fit the engagement data better, while modeling the three views together

Table 6: Descriptions of Representative Items in Each Cluster

Cluster ID	Item ID	General Type	Short Description
0	6	Positive Thinking	Don't be afraid of pride. It is OK to tell yourself how great you are or how long you waited and tried to get where you are now.
	23	Positive Thinking	Acknowledging the limitations of your abilities and things that are in your control.
1	30	Mindfulness	Grab a pen and paper and get writing.
	33	Mindfulness	Use a herb, essential oil or a mild smell like perfume, flower, or a leaf of Pelargonium to relax and focus on the problem.
2	11	Social Engagement	Choose someone of your social media contact that you know he/she is a supportive person and meet them again face to face.
	20	Social Engagement	Remember friends birthdays and reconnect.
3	25	Physical Activity	Sit at a desk all day? A few in-home or in-office exercises to strengthen your muscles.
	28	Physical Activity	Go to the hallway and go up and down the stairs
4	44	Mindfulness	Roll breathing helps you to develop full use of your lungs and to focus on the rhythm of your breathing.
	47	Mindfulness	Mindful breathing helps you take a break for a moment to unwind and concentrate on the here and now.



(a) Item Avg. Rate **(b) Item Rate Count** **(c) Item Engagement Ratio** **(d) Item Engagement Count**

Figure 2: Pairwise T-Tests on Item Cluster Attributes.

would impose more constraints on the learned parameters to fit an additional view and be less flexible to the engagement data.

RQ 1.3., Smoothing Helpfulness. To study if the smoothing assumption helps in achieving better predictions, we compare the STRETCH-FULL and STRETCH-SV models with their variants without the smoothing regularization. Since only stress-level and engagement data include time-steps, we compare the results in these two views. As Table 3 shows, STRETCH-SV and STRETCH-FULL perform better than STRETCH-SV-NS and STRETCH-FULL-NS, respectively. Hence, the smoothness regularization helps in both single view and full models for stress predictions. This approves our assumption that user stress levels do not change dramatically and, instead, change slowly in time.

However, the results are different in engagement predictions. As we can see in Table 5, the best STRETCH-SV and STRETCH-FULL models have the same results as STRETCH-SV-NS and STRETCH-FULL-NS, respectively. This, as well as Table 2, show that during the hyperparameter tuning, our model automatically selects the models with no smoothing as the best performers for engagement predictions. As a result, the smoothness assumption does not work for engagement data, showing that user engagement patterns are highly variant across consecutive time steps.

4.3 Activity Cluster Analysis

To answer **RQ 2.1.**, and investigate if our model finds meaningful activity characteristics, we perform activity cluster analysis based on the learned model parameters. Particularly, for each activity, we concatenate the learned item factors (Q_a), item rating bias (b_{z_a}), and item engagement bias (b_{x_a}) to build an item feature vector. Then, we apply the spectral clustering algorithm[47] on item feature vectors to group the activities into clusters. Using the Silhouette coefficient [32] to choose the best number of clusters, we discover five item clusters.

Table 7: Statistics of Each Item Cluster

Cluster ID	Item Average Rate		Item Rate Count		Item Engagement Ratio		Item Engagement Count	
	Mean	STD	Mean	STD	Mean	STD	Mean	STD
0	2.65	1.08	4.02	0.56	0.37	0.15	3.47	1.72
1	3.46	2.31	3.99	0.67	0.66	0.17	5.92	4.38
2	3.78	1.75	3.75	0.60	0.67	0.20	5.33	3.37
3	1.33	0.47	4.67	0.47	0.23	0.07	1.33	0.47
4	1.50	0.87	2.58	0.92	0.18	0.17	1.14	1.36

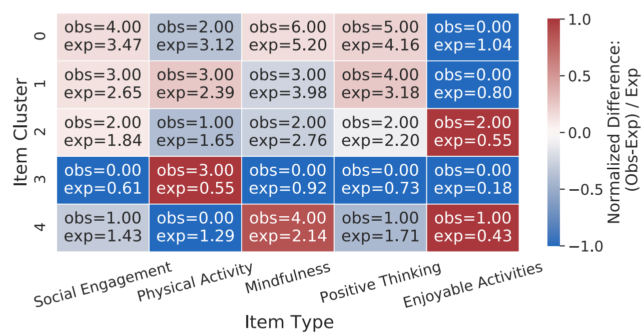
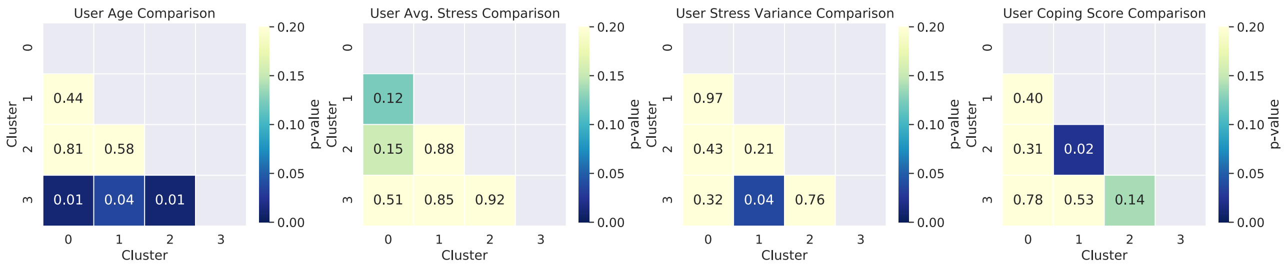


Figure 3: Chi-square Test for Item Cluster and Item Type.

To assess the validity of these clusters, we (1) study the activity descriptions and coping strategies in each cluster and (2) investigate the association between these clusters with item attributes in the data. Checking the activity descriptions in each cluster, we find that items in some clusters share inherent similarities. For example, activities in cluster-3 are mostly physical activities and in-place exercises, while cluster-4 activities are about breathing and mindfulness. Since the whole data would not fit, we show some representative activities of each cluster in Table 6. As we can see, activity 30 and 33 in cluster-1 and activity 44 and 47 in cluster-4



(a) User Age **(b) User Avg. Stress** **(c) User Stress Variance** **(d) User CS Score**

Figure 4: Heatmaps of P-Values of Pairwise T-Test on User Cluster Attributes.

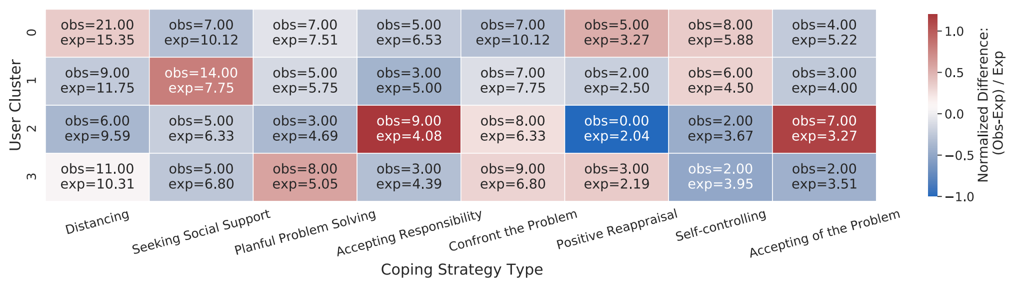


Figure 5: Chi-square Test between User Cluster and Coping Strategy Type.

are all in `Mindfulness` general type. However, activity 30 and 33 are more about solution-oriented suggestions, and activity 44 and 47 are more about breathing for relaxation and calming.

Accordingly, for assessing if activity clusters are significantly related to specific item characteristics, such as activity type or coping strategy, we perform Chi-Square tests between item clusters and these characteristics. Particularly, we test if the distribution of items in terms of activity general type and activity coping strategy are significantly different in item clusters. To this end, we calculate the number of activities with each activity general type (or activity coping strategy type) observed in each cluster. Then, we compare this value with the expected number of activities of each type in each cluster, assuming the null hypothesis (that no pattern exists between item clusters and item attributes). The Chi-Square tests calculate the p-value for this comparison. Specifically, Figure 3 shows a heatmap of normalized difference between the observed and expected number of items from each general item type in each cluster. The red color indicates that the observed value was more than the expected value, and the blue color is the reverse. The larger the difference, the darker the color is. According to the Chi-Square test, the p-value for cluster 3 is 0.0098, indicating a significant level for rejecting the null-hypothesis. As we can see in the figure, cluster 3 items are more about physical activities and less from other types, confirming our previously mentioned results. Similarly, cluster 4 items, although not significant, are more about enjoyable and mindful activities rather than physical activities. Cluster 2 items are more enjoyable than expected, and cluster 0 and 1 activities are less about enjoyable activities than expected. We omit the Chi-Square test results for activity coping strategy types, as none of the item clusters significantly rejected the null hypothesis.

At the same time, not all item clusters can be explained by item descriptions. The item feature vectors are based on many factors, such as prediction accuracy, which could be different from the semantic activity similarities. To assess this aspect, we compare the discovered clusters over several attributes, such as items' average rating, item rating counts, item engagement ratio, and item engagement counts. Each cluster's statistics are shown in Table 7. As we can see, clusters 3 and 4 have activities with relatively lower item average ratings, engagement ratios, and engagement counts. However, items in cluster 3 received more ratings than items in cluster 4. On the other hand, clusters 1 and 2 have activities with the highest item average rating, engagement ratio, and engagement counts. To achieve a deeper understanding of these results, we conduct the statistical t-test on each attribute for each pair of clusters and report the p-values with heatmaps, which are shown in Fig. 2. As we can see (part (a)), cluster 3 has a significantly lower mean of item average rating than cluster 4. However, cluster 3 has significantly fewer ratings than other clusters, except cluster 4 (part (b)). In addition, items in cluster 1 have a similar ratio of engagement and engagement count as items in cluster 2; and items in clusters 3 and 4 also share these similarities (parts (c) and (d)). However, item clusters 3 and 4 have significantly less user engagement compared with clusters 1 and 2 (part (d)).

4.4 User Cluster Analysis

Similar to item clustering, we conduct the user cluster analysis to answer **RQ 2.2.** and **RQ 2.3.**. The users are clustered based on the discovered user factors $\mathbf{s}_u$, user rating bias b_{z_u}, user stress bias b_{y_u}, and user engagement bias b_{x_u}, via the spectral clustering algorithm. Accordingly, four clusters are found among the users.

To answer **RQ 2.2.**, we study several user attributes in user clusters, such as user age, coping strategy (CS) score, average stress levels, stress level variance, average rating, rate count, engagement ratio, as well as engagement count for comparison. The first two

Table 8: Statistics of Each User Cluster

Cluster ID	User Age		User Average Stress		User Stress Variance		User CS Score	
	Mean	STD	Mean	STD	Mean	STD	Mean	STD
0	40.88	8.85	1.77	0.24	0.24	0.20	19.00	7.68
1	37.22	12.93	2.17	0.79	0.24	0.19	19.44	6.10
2	39.73	12.05	2.13	0.61	0.13	0.21	20.60	5.64
3	29.42	5.39	2.08	1.14	0.11	0.09	22.17	5.03
Cluster ID	**User Average Rate**		**User Rate Count**		**User Engagement Ratio**		**User Engagement Count**	
	Mean	STD	Mean	STD	Mean	STD	Mean	STD
0	4.15	0.93	6.38	5.15	0.93	0.07	9.14	5.28
1	3.42	0.48	1.44	2.57	0.25	0.23	3.27	5.04
2	3.81	0.59	1.73	2.14	0.28	0.25	3.08	3.27
3	4.04	0.52	2.58	1.98	0.55	0.30	4.30	2.37

attributes are measured using questionnaires, independently of our model. The mean and standard deviation of each attribute for each user cluster is shown in Table 8. The pairwise T-Test results among user clusters are shown in Fig 4.

By checking the Table 8 and Fig. 4 together, we can see that the discovered clusters can be interpreted using these attributes. First, we see that users in cluster 3 have an average age of 29 years old, which is significantly lower than the users in other clusters. Also, cluster 3 users have the highest mean coping strategy test score among all user clusters. Additionally, these users have relatively less stress variance or fluctuation compared with other clusters, specifically significantly lower than cluster 1 users, and have a medium-level engagement ratio. It indicates that cluster 3 users are young people with less stress level changes and more resilience in coping with stress compared to other users. Similarly, we can see that cluster 0 includes the older users with relatively lower stress levels, who tend to engage in more activities, rate more activities, and give higher ratings to these activities.

In addition to user attributes, we are interested to know if there is a relation between user clusters and the activity attributes that they interact with (**RQ 2.3.**). Particularly, we test if the distribution of user-engaged activities are significantly different in clusters, in terms of activity general type and activity coping strategy. First, to assess activity general types in each cluster, we calculate the number of activities with each activity general type used by users of each user cluster. We do the same thing to calculate the number of observed results based on activity coping strategy types in each cluster. Then, we use the Chi-square test on the categorical variables, user cluster and item type. Figure 6 shows the results of item general types. Clusters 1 and 3 have significantly different distributions than expected with p-values 0.047 and 0.095 and Chi-square statistics 9.64 and 7.90 respectively. We can see that users in cluster-1 who on average are 37 years old engage more with *Social Engagement* and *Mindfulness* activities and are less interested in *Physical Activity* and *Enjoyable Activity* categories. In contrary, users in cluster-3 who are 29 years old on average are more interested in engaging with *Physical Activities* and *Enjoyable Activities*, but less engaged with *Positive Thinking* and *Mindfulness*.

Performing the same analysis on coping strategy types, we see that cluster 2 has a significantly different distribution of activity coping strategies than expected (with p-value 0.028 and Chi-square

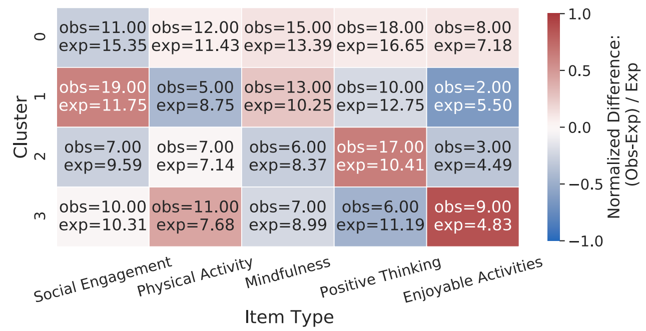


Figure 6: Chi-square Test for User Cluster and Item Type.

statistic of 15.68). The heatmap results are shown in Figure 5. As we can see, users in cluster 2, who are on average 39.73 years old, engaged more with items related to *Accepting the Responsibility*, *Accepting the Problem*, and *Confronting the Problem*, but less with items in other categories such as *Positive Reappraisal*, *Distancing*, *Planful Problem Solving*, and *Self-controlling*. As shown in Table 8, cluster-2 has relatively high average stress but with a relatively high coping strategy score.

5 CONCLUSIONS

In this paper, we introduced STRETCH, a multi-view tensor and matrix decomposition model for holistic stress-level, activity engagement, and interest predictions, that can generalize well with small, missing, and sparse data. We designed experiments to answer 6 research questions. In our experiments, we showed that the multi-view modeling of STRETCH provides improvements over single-view predictions. Particularly, we showed the reciprocal relations between stress levels, user interests, and engagements with activities: not only observing user interest and engagement in activities helped in predicting their stress levels; but also observing user stress levels helped in predicting users' interests and engagement. We also showcased the validity of STRETCH results by studying user and item clusters discovered by our model, in association with various factors. Specifically, our experiments demonstrated that the item clusters represent item general types, average item popularity, and average item desirability. Additionally, we presented interesting findings in the discovered user clusters: a cluster of younger users with less stress variation and more preference for physical and enjoyable activities, and a cluster of older users with lower stress levels and higher engagements. This study is limited in using only one dataset from one application, with smaller data size. In the future, we would like to expand this study to other datasets.

ACKNOWLEDGMENTS

This paper is based upon work supported by the National Science Foundation under Grant No. 2047500.

REFERENCES

[1] Daniel A. Adler, Vincent W.-S. Tseng, Gengmo Qi, Joseph Scarpa, Srijan Sen, and Tanzeem Choudhury. [n. d.]. Identifying Mobile Sensing Indicators of Stress-Resilience. 5, 2 ([n. d.]), 51:1–51:32. https://doi.org/10.1145/3463528

[2] Mawulolo K. Ameko, Miranda L. Beltzer, Lihua Cai, Mehdi Boukhechba, Bethany A. Teachman, and Laura E. Barnes. [n. d.]. Offline Contextual Multi-armed Bandits for Mobile Health Interventions: A Case Study on Emotion Regulation. In *Fourteenth ACM Conference on Recommender Systems* (New York, NY, USA, 2020-09-22) *(RecSys '20)*. Association for Computing Machinery, 249–258. https://doi.org/10.1145/3383313.3412244

[3] Aditya Chand, Monica Gonzalez, Julian Missig, Purin Phanichphant, and Pen Fan Sun. [n. d.]. Balance pass: service design for a healthy college lifestyle. In *CHI '06 Extended Abstracts on Human Factors in Computing Systems* (New York, NY, USA, 2006-04-21) *(CHI EA '06)*. Association for Computing Machinery, 1813–1818. https://doi.org/10.1145/1125451.1125795

[4] Chuanwu Chen, Guangyuan Liu, and Wanhui Wen. [n. d.]. The Recognition and Classification of Stress Base on Pulse Transit Time Series. In *Proceedings of the International Conference on Compute and Data Analysis* (New York, NY, USA, 2017-05-19) *(ICCDA '17)*. Association for Computing Machinery, 268–272. https://doi.org/10.1145/3093241.3093266

[5] M. Ciman and K. Wac. [n. d.]. Individuals' stress assessment using human-smartphone interaction analysis. PP, 99 ([n. d.]), 1–1. https://doi.org/10.1109/TAFFC.2016.2592504

[6] Shanice Clarke, Luis G. Jaimes, and Miguel A. Labrador. [n. d.]. mStress: A mobile recommender system for just-in-time interventions for stress. In *2017 14th IEEE Annual Consumer Communications Networking Conference (CCNC)* (2017-01). 1–5. https://doi.org/10.1109/CCNC.2017.8015367 ISSN: 2331-9860.

[7] Sheldon Cohen, Tom Kamarck, and Robin Mermelstein. [n. d.]. A Global Measure of Perceived Stress. 24, 4 ([n. d.]), 385–396. https://doi.org/10.2307/2136404 Publisher: [American Sociological Association, Sage Publications, Inc.].

[8] David Elsweiler, Bernd Ludwig, Alan Said, Hanna Schaefer, and Christoph Trattner. [n. d.]. Engendering Health with Recommender Systems. In *Proceedings of the 10th ACM Conference on Recommender Systems* (Boston, Massachusetts, USA, 2016-09-07) *(RecSys '16)*. Association for Computing Machinery, 409–410. https://doi.org/10.1145/2959100.2959203

[9] S. Folkman, R. S. Lazarus, C. Dunkel-Schetter, A. DeLongis, and R. J. Gruen. [n. d.]. Dynamics of a stressful encounter: cognitive appraisal, coping, and encounter outcomes. 50, 5 ([n. d.]), 992–1003. https://doi.org/10.1037//0022-3514.50.5.992

[10] Amelie Gyrard and Amit Sheth. [n. d.]. IAMHAPPY: Towards an IoT knowledge-based cross-domain well-being recommendation system for everyday happiness. 15 ([n. d.]), 100083. https://doi.org/10.1016/j.smhl.2019.100083

[11] Jiayuan He, Ke Li, Xiaoli Liao, P. Zhang, and N. Jiang. [n. d.]. Real-Time Detection of Acute Cognitive Stress Using a Convolutional Neural Network From Electrocardiographic Signal. ([n. d.]). https://doi.org/10.1109/ACCESS.2019.2907076

[12] Galen Chin-Lun Hung, Pei-Ching Yang, Chen-Yi Wang, and Jung-Hsien Chiang. [n. d.]. A Smartphone-Based Personalized Activity Recommender System for Patients with Depression. In *Proceedings of the 5th EAI International Conference on Wireless Mobile Communication and Healthcare* (Brussels, BEL, 2015-12-22) *(MOBIHEALTH'15)*. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 253–257. https://doi.org/10.4108/eai.14-10-2015.2261655

[13] Luis G. Jaimes, Martin Llofriu, and Andrew Raij. [n. d.]. PREVENTER, a Selection Mechanism for Just-in-Time Preventive Interventions. 7, 3 ([n. d.]), 243–257. https://doi.org/10.1109/TAFFC.2015.2490062 Conference Name: IEEE Transactions on Affective Computing.

[14] Luis G. Jaimes, Martin Llofriu, and Andrew Raij. 2014. A Stress-Free Life: Just-in-Time Interventions for Stress via Real-Time Forecasting and Intervention Adaptation. In *Proceedings of the 9th International Conference on Body Area Networks* (London, United Kingdom) *(BodyNets '14)*. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), Brussels, BEL, 197–203. https://doi.org/10.4108/icst.bodynets.2014.258237

[15] Natasha Jaques, Ognjen (Oggi) Rudovic, Sara Taylor, Akane Sano, and Rosalind Picard. 2017. Predicting Tomorrow's Mood, Health, and Stress Level using Personalized Multitask Learning and Domain Adaptation. In *Proceedings of IJCAI 2017 Workshop on Artificial Intelligence in Affective Computing (Proceedings of Machine Learning Research, Vol. 66)*, Neil Lawrence and Mark Reid (Eds.). PMLR, 17–33. http://proceedings.mlr.press/v66/jaques17a.html

[16] Jyun-Yu Jiang, Zehan Chao, Andrea L. Bertozzi, Wei Wang, Sean D. Young, and Deanna Needell. [n. d.]. Learning to Predict Human Stress Level with Incomplete Sensor Data from Wearable Devices. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management* (New York, NY, USA, 2019-11-03) *(CIKM '19)*. Association for Computing Machinery, 2773–2781. https://doi.org/10.1145/3357384.3357831

[17] Christina Kelley, Bongshin Lee, and Lauren Wilcox. [n. d.]. Self-tracking for Mental Wellness: Understanding Expert Perspectives and Student Experiences. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2017) *(CHI '17)*. ACM, 629–641. https://doi.org/10.1145/3025453.3025750

[18] Zachary King, Judith Moskowitz, Laurie Wakschlag, and Nabil Alshurafa. [n. d.]. Predicting Perceived Stress Through Mirco-EMAs and a Flexible Wearable ECG Device. In *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers* (New York, NY, USA, 2018-10-08) *(UbiComp '18)*. Association for Computing Machinery, 106–109. https://doi.org/10.1145/3267305.3267639

[19] Brigitte M. Kudielka, D. H. Hellhammer, and Stefan Wüst. [n. d.]. Why do we respond so differently? Reviewing determinants of human salivary cortisol responses to challenge. 34, 1 ([n. d.]), 2 – 18. https://doi.org/10.1016/j.psyneuen.2008.10.004

[20] Richard S Lazarus and Susan Folkman. [n. d.]. *Stress, Appraisal, and Coping*. Springer Publishing Company.

[21] Boning Li and Akane Sano. [n. d.]. Extraction and Interpretation of Deep Autoencoder-based Temporal Features from Wearables for Forecasting Personalized Mood, Health, and Stress. 4, 2 ([n. d.]), 49:1–49:26. https://doi.org/10.1145/3397318

[22] Mehrab Bin Morshed, Koustuv Saha, Richard Li, Sidney K. D'Mello, Munmun De Choudhury, Gregory D. Abowd, and Thomas Plötz. [n. d.]. Prediction of Mood Instability with Passive Sensing. 3, 3 ([n. d.]), 75:1–75:21. https://doi.org/10.1145/3351233

[23] Oscar Martinez Mozos, Virginia Sandulescu, Sally Andrews, David Ellis, Nicola Bellotto, Radu Dobrescu, and Jose Manuel Ferrandez. [n. d.]. Stress Detection Using Wearable Physiological and Sociometric Sensors. 27, 2 ([n. d.]), 1650041. https://doi.org/10.1142/S0129065716500416

[24] Ehimwenma Nosakhare and Rosalind Picard. [n. d.]. Toward Assessing and Recommending Combinations of Behaviors for Improving Health and Well-Being. 1, 1 ([n. d.]), 4:1–4:29. https://doi.org/10.1145/3368958

[25] Pablo Paredes, Ran Gilad-Bachrach, Mary Czerwinski, Asta Roseway, Kael Rowan, and Javier Hernandez. [n. d.]. PopTherapy: Coping with Stress Through Pop-culture. In *Proceedings of the 8th International Conference on Pervasive Computing Technologies for Healthcare* (ICST, Belgium, 2014) *(PervasiveHealth '14)*. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 109–117. https://doi.org/10.4108/icst.pervasivehealth.2014.255070

[26] Mashfiqui Rabbi, Predrag Klasnja, Tanzeem Choudhury, Ambuj Tewari, and Susan Murphy. [n. d.]. Optimizing mHealth Interventions with a Bandit. In *Digital Phenotyping and Mobile Sensing: New Developments in Psychoinformatics*, Harald Baumeister and Christian Montag (Eds.). Springer International Publishing, 277–291.

[27] Mahbubur Rahman, Rummana Bari, Amin Ahsan Ali, Moushumi Sharmin, Andrew Raij, Karen Hovsepian, Syed Monowar Hossain, Emre Ertin, Ashley Kennedy, David H. Epstein, Kenzie L. Preston, Michelle Jobes, J. Gayle Beck, Satish Kedia, Kenneth D. Ward, Mustafa al'Absi, and Santosh Kumar. [n. d.]. Are We There Yet? Feasibility of Continuous Stress Assessment via Wireless Physiological Sensors. 2014 ([n. d.]), 479–488. https://doi.org/10.1145/2649387.2649433

[28] Darius A. Rohani, Andrea Quemada Lopategui, Nanna Tuxen, Maria Faurholt-Jepsen, Lars V. Kessing, and Jakob E. Bardram. [n. d.]. MUBS: A Personalized Recommender System for Behavioral Activation in Mental Health. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2020-04-21) *(CHI '20)*. Association for Computing Machinery, 1–13. https://doi.org/10.1145/3313831.3376879

[29] Darius Adam Rohani, Aaron Springer, Victoria Hollis, Jakob E. Bardram, and Steve Whittaker. [n. d.]. Recommending Activities for Mental Health and Well-being: Insights from Two User Studies. ([n. d.]), 1–1. https://doi.org/10.1109/TETC.2020.2972007 Conference Name: IEEE Transactions on Emerging Topics in Computing.

[30] Darius A Rohani, Nanna Tuxen, Andrea Quemada Lopategui, Maria Faurholt-Jepsen, Lars V Kessing, and Jakob E Bardram. [n. d.]. Personalizing Mental Health: A Feasibility Study of a Mobile Behavioral Activation Tool for Depressed Patients. In *Proceedings of the 13th EAI International Conference on Pervasive Computing Technologies for Healthcare* (New York, NY, USA, 2019-05-20) *(PervasiveHealth'19)*. Association for Computing Machinery, 282–291. https://doi.org/10.1145/3329189.3329214

[31] Renata Lopes Rosa, Gisele Maria Schwartz, Wilson Vicente Ruggiero, and Demóstenes Zegarra Rodríguez. [n. d.]. A Knowledge-Based Recommendation System That Includes Sentiment Analysis and Deep Learning. 15, 4 ([n. d.]), 2124–2135. https://doi.org/10.1109/TII.2018.2867174 Conference Name: IEEE Transactions on Industrial Informatics.

[32] Peter J Rousseeuw. 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics* 20 (1987), 53–65.

[33] Alan Said, Hanna Schäfer, Helma Torkamaan, and Christoph Trattner. [n. d.]. Fifth International Workshop on Health Recommender Systems (HealthRecSys 2020). In *Fourteenth ACM Conference on Recommender Systems* (New York, NY, USA, 2020-09-22) *(RecSys '20)*. Association for Computing Machinery, 611–612. https://doi.org/10.1145/3383313.3411540

[34] Salim Salmi, Saskia Mérelle, Renske Gilissen, and Willem-Paul Brinkman. [n. d.]. Content-Based Recommender Support System for Counselors in a Suicide Prevention Chat Helpline: Design and Evaluation Study. 23, 1 ([n. d.]), e21690. https://doi.org/10.2196/21690 Company: Journal of Medical Internet Research Distributor: Journal of Medical Internet Research Institution: Journal of Medical Internet Research Label: Journal of Medical Internet Research Publisher: JMIR Publications Inc., Toronto, Canada.

[35] A. Sami, R. Nagatomi, M. Terabe, and K. Hashimoto. [n. d.]. Design of physical activity recommendation system. In *MCCSIS'08 - IADIS Multi Conference on Computer Science and Information Systems; Proceedings of Informatics 2008 and Data Mining 2008* (2008). 148–152. https://www.scopus.com/inward/record.uri?eid=2-s2.0-58449131065&partnerID=40&md5=a4072fcc00476ecb8e5a8f3d61804284

[36] Akane Sano, Paul Johns, and Mary Czerwinski. [n. d.]. HealthAware: An advice system for stress, sleep, diet and exercise. In *2015 International Conference on Affective Computing and Intelligent Interaction (ACII)* (Xi'an, China, 2015-09). IEEE, 546–552. https://doi.org/10.1109/ACII.2015.7344623

[37] Hillol Sarker, Matthew Tyburski, Md Mahbubur Rahman, Karen Hovsepian, Moushumi Sharmin, David H. Epstein, Kenzie L. Preston, C. Debra Furr-Holden, Adam Milam, Inbal Nahum-Shani, Mustafa al'Absi, and Santosh Kumar. [n. d.]. Finding Significant Stress Episodes in a Discontinuous Time Series of Rapidly Varying Mobile Sensor Data. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, 4489–4501. https://doi.org/10.1145/2858036.2858218

[38] Jessica Schroeder, Jina Suh, Chelsey Wilks, Mary Czerwinski, Sean A. Munson, James Fogarty, and Tim Althoff. [n. d.]. Data-Driven Implications for Translating Evidence-Based Psychotherapies into Technology-Delivered Interventions. ([n. d.]). https://www.microsoft.com/en-us/research/publication/data-driven-implications-for-translating-evidence-based-psychotherapies-into-technology-delivered-interventions/

[39] Hanna Schäfer, Santiago Hors-Fraile, Raghav Pavan Karumur, André Calero Valdez, Alan Said, Helma Torkamaan, Tom Ulmer, and Christoph Trattner. [n. d.]. Towards Health (Aware) Recommender Systems. In *Proceedings of the 2017 International Conference on Digital Health* (London United Kingdom, 2017-07-02). ACM, 157–161. https://doi.org/10.1145/3079452.3079499

[40] Moushumi Sharmin, Andrew Raij, David Epstien, Inbal Nahum-Shani, J. Gayle Beck, Sudip Vhaduri, Kenzie Preston, and Santosh Kumar. [n. d.]. Visualization of time-series sensor data to inform the design of just-in-time adaptive stress interventions. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (New York, NY, USA, 2015-09-07) *(UbiComp '15)*. Association for Computing Machinery, 505–516. https://doi.org/10.1145/2750858.2807537

[41] Katrin Starcke and Matthias Brand. [n. d.]. Decision making under stress: a selective review. 36, 4 ([n. d.]), 1228–1248. https://doi.org/10.1016/j.neubiorev.2012.02.003

[42] Sara Ann Taylor, Natasha Jaques, Ehimwenma Nosakhare, Akane Sano, and Rosalind Picard. [n. d.]. Personalized Multitask Learning for Predicting Tomorrow's Mood, Stress, and Health. ([n. d.]), 1–1. https://doi.org/10.1109/TAFFC.2017.2784832

[43] Helma Torkamaan and Jüergen Ziegler. [n. d.]. Integrating Behavior Change and Persuasive Design Theories into an Example Mobile Health Recom- mender System. In *Adjunct Proceedings of the 2021 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2021 ACM International Symposium on Wearable Computers (UbiComp-ISWC '21 Adjunct)* (Virtual, 2021-09). ACM. https://doi.org/10.1145/3460418.3479330

[44] Helma Torkamaan and Jürgen Ziegler. [n. d.]. Mobile Mood Tracking: An Investigation of Concise and Adaptive Measurement Instruments. 4, 4 ([n. d.]), 155:1–155:30. https://doi.org/10.1145/3432207

[45] Helma Torkamaan and Jürgen Ziegler. [n. d.]. Rating-based Preference Elicitation for Recommendation of Stress Intervention. In *Proceedings of the 27th ACM Conference on User Modeling, Adaptation and Personalization* (New York, NY, USA, 2019-06-07) *(UMAP '19)*. Association for Computing Machinery, 46–50. https://doi.org/10.1145/3320435.3324990

[46] Helma Torkamaan and Jürgen Ziegler. [n. d.]. A taxonomy of mood research and its applications in computer science. In *2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)* (2017-10). 421–426. https://doi.org/10.1109/ACII.2017.8273634 ISSN: 2156-8111.

[47] Ulrike Von Luxburg. 2007. A tutorial on spectral clustering. *Statistics and computing* 17, 4 (2007), 395–416.

[48] Fabian Wahle, Tobias Kowatsch, Elgar Fleisch, Michael Rufer, and Steffi Weidt. [n. d.]. Mobile Sensing and Support for People With Depression: A Pilot Trial in the Wild. 4, 3 ([n. d.]). https://doi.org/10.2196/mhealth.5960

[49] Shiqi Yang, Ping Zhou, Kui Duan, M. Shamim Hossain, and Mohammed F. Alhamid. [n. d.]. emHealth: Towards Emotion Health Through Depression Prediction and Intelligent Health Recommender System. 23, 2 ([n. d.]), 216–226. https://doi.org/10.1007/s11036-017-0929-3

[50] Xiao Zhang, Fuzhen Zhuang, Wenzhong Li, Haochao Ying, Hui Xiong, and Sanglu Lu. [n. d.]. Inferring Mood Instability via Smartphone Sensing: A Multi-View Learning Approach. In *Proceedings of the 27th ACM International Conference on Multimedia* (Nice, France, 2019-10-15) *(MM '19)*. Association for Computing Machinery, 1401–1409. https://doi.org/10.1145/3343031.3350957