

Learning to Predict Human Stress Level with Incomplete Sensor Data from Wearable Devices

Jyun-Yu Jiang^{†*}, Zehan Chao^{‡*}, Andrea L. Bertozzi[‡], Wei Wang[†], Sean D. Young[§] and Deanna Needell[‡]

[†]Department of Computer Science, University of California, Los Angeles

[‡]Department of Mathematics, University of California, Los Angeles

[§]University of California Institute for Prediction Technology, Departments of Emergency Medicine and Informatics, University of California, Irvine

{jyunyu, weiwang}@cs.ucla.edu, {zchao3, bertozzi, deanna}@math.ucla.edu, sdyoung@mednet.ucla.edu

ABSTRACT

Stress is a common problem in modern life that can bring both psychological and physical disorder. Wearable sensors are commonly used to study the relationship between physical records and mental status. Although sensor data generated by wearable devices provides an opportunity to identify stress in people for predictive medicine, in practice, the data are typically complicated and vague and also often fragmented. In this paper, we propose Data Completion with Diurnal Regularizers (DCDR) and Temporally Hierarchical Attention Network (THAN) to address the fragmented data issue and predict human stress level with recovered sensor data. We model fragmentation as a sparsity issue. The nuclear norm minimization method based on the low-rank assumption is first applied to derive unobserved sensor data with diurnal patterns of human behaviors. A hierarchical recurrent neural network with the attention mechanism then models temporally structural information in the reconstructed sensor data, thereby inferring the predicted stress level. Data for this study were from 75 undergraduate students (taken from a sample of a larger study) who provided sensor data from smart wristbands. They also completed weekly stress surveys as ground-truth labels about their stress levels. This survey lasted 12 weeks and the sensor records are also in this period. The experimental results demonstrate that our approach significantly outperforms conventional methods in both data completion and stress level prediction. Moreover, an in-depth analysis further shows the effectiveness and robustness of our approach.

KEYWORDS

Human stress level prediction; Sensor data completion; Hierarchical recurrent neural networks; Attention mechanism

*Equal contribution.

Permission to make digital or hard copies of all or part 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 components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

CIKM '19, November 3–7, 2019, Beijing, China

© 2019 Association for Computing Machinery.

ACM ISBN 978-1-4503-6976-3/19/11...\$15.00

<https://doi.org/10.1145/3357384.3357831>

ACM Reference Format:

Jyun-Yu Jiang, Zehan Chao, Andrea L. Bertozzi, Wei Wang, Sean D. Young and Deanna Needell. 2019. Learning to Predict Human Stress Level with Incomplete Sensor Data from Wearable Devices. In *The 28th ACM International Conference on Information and Knowledge Management (CIKM '19)*, November 3–7, 2019, Beijing, China. ACM, New York, NY, USA, 9 pages. <https://doi.org/10.1145/3357384.3357831>

1 INTRODUCTION

Psychological stress is a human reaction to perceived pressure; it can be a serious issue for modern people in their daily life.[37] In the United States, more than half of the population has stress problems [4] while 75% of college students are experiencing at least moderate stress [43]. Moreover, stress usually leads to a myriad of diseases and mental illness, such as heart attacks and depression [19]. For example, more than 25 and 17 million Americans suffer from stress ulceration and major depression [1, 45]. Therefore, understanding human stress is an essential task for healthcare and clinical research. In addition, it is also beneficial for predictive medicine to identify stressed patients earlier and prevent potential harmful effects.

To quantify human stress level, one possible solution is to exploit sensor data collected by wearable devices, such as smartwatches and smart-wristbands. More specifically, the sensors in wearable devices can obtain physiological signals from the users, thereby tracking user activities and health situations. For instance, electrocardiogram circuits and accelerometers monitor heart rates and user motions [14]. However, high-quality and fine-grained signal data are not always available for users and developers to predict human stress level. Precisely, the response speed of sensors can be extremely inefficient while some wearable devices only provide aggregate information [29]. Moreover, the average daily wear times are typically 10.9 and 8.4 hours on weekdays and weekends for smartwatch users while users typically take off their wearable devices more than three times every day [28]. Furthermore, many wearable devices only report the distribution of physical activities in a short period without detailed sensor data for high-level applications [3]. Hence, the missing or sparse data issues are a difficult hurdle in the prediction of human stress level with wearable devices.

Although missing sensor data can lead to extremely low performance of downstream applications, the redundancy between missing and existing data can be leveraged to recover unavailable

data and boost the model performance. Matrix factorization and collaborative filtering are popular approaches to recover missing data and have been applied to various domains, such as recommender systems [35] and sensor networks [10]. Particularly, sensor data can be represented by a low-rank matrix so that the missing entries can be recovered by decomposing the matrix into a product of several smaller latent matrices. For instance, Zhang et al. [55] and Wang et al. [51] apply matrix factorization to ambulatory fall detection with wearable sensor data while the latent matrices can be treated as features for human activity recognition [54]. Although matrix factorization and collaborative filtering are robust to capture short-term and universal patterns, complicated long-term and individual behaviors in sensor data can still be difficult to be considered by latent matrices [36].

The recent success of deep learning, especially recurrent neural networks (RNNs) on sequences, provides an opportunity to model user behaviors and predict human stress level with temporal sensor data. For example, RNNs have been applied to model latent transitions of words and queries for natural language processing [5, 31] and information retrieval [18, 30]. Several previous studies also exploit deep learning to model sensor data of wearable devices for downstream applications, such as human activity recognition [23] and sleeping quality analysis [46]. However, there are some drawbacks of existing deep learning based approaches for predicting human stress level. First, although RNNs can effectively model short-term sensor data, human stress is both complex and also related to long-term behavior, while only a small portion of the data is crucial for human stress level prediction. Second, human behaviors are sophisticated and sometimes have non-sequential patterns. For instance, RNNs can suffer from modeling separated but periodic patterns like diurnal behaviors without enough training data [36]. Last but not least, most of the existing studies using sensor data rely on dense raw sensor values so that the data sparsity problem can be more significant if wearable devices only provide aggregated information instead of original sensor data.

In this paper, Data Completion with Diurnal Regularizers (DCDR) and Temporally Hierarchical Attention Network (THAN) are proposed to address the limitations of using incomplete and aggregated wearable device sensor data for predicting human stress level, in the context of deep learning techniques. More precisely, we focus on addressing the data sparsity issue and modeling structural human behaviors. Based on the low-rank assumption and observations of diurnal patterns as regularization, DCDR partially recovers sensor data. THAN then follows a temporally hierarchical structure to encode daily and weekly behaviors with a hierarchical recurrent neural network and the attention mechanism for predicting human stress level with recovered sensor data. In addition, several supplementary statistical user features are also considered to personalize predictions according to different users. Here we summarize our contributions in the following.

- To the best of our knowledge, this paper is the first work to apply deep learning techniques to predict human stress level with long-term wearable device sensor data. Although longstanding sensor data are usually sparse and noisy, they are still capable of providing thorough understandings about human stress.

- We propose DCDR and THAN to deal with the data sparsity and precisely predict human stress level. More specifically, diurnal behavioral patterns can significantly benefit missing data recovery while user behaviors can be more effectively captured by exploiting temporally hierarchical structures of sensor data.
- Experiments conducted on 75 college students for 12 weeks demonstrate that our approach significantly outperforms existing methods in both data completion and human stress level prediction. The analysis of parameter sensitivity and ablation studies then indicates the robustness and effectiveness of the proposed framework. In addition, we will release our code to facilitate future research.

2 RELATED WORK

2.1 Stress Level Prediction

Traditional methods for monitoring stress level are mainly from physiological signals such as heart rate, blood pressure, and body temperature [6]. By applying machine learning techniques, researchers can predict human stress and emotion by measuring their physiological signals. A study has shown that, with the specially designed sensor recording body movement, speech volume, and pitch, a person-specific model gives 93% accuracy on binary stress prediction [41]. Personal-specific models significantly outperform general models [7, 12, 27] regardless of the machine learning model. Alternatively, the lack of personal history information motivates us to develop general models that treat all input instances in the same way.

Physiological signals also have limitations: the necessity of continuous signal measurement and the data noise restricts its broad application. Meanwhile, researchers have analyzed stress and mood predictors from broadly accessible data such as smartphone usage [38] or general wearable device data such as geographic location [12] and activities. There are many studies about predicting health-related index with large-scale accessible datasets, integrating both public health information and data science. For example, an existing study shows a significant association between negative sentiment tweet and stress level [20], inferring the possibility to monitor people's mood and stress using social media data. Several studies [7, 27] showed the potential to use wearable sensor and other integrated data sources to monitor people's stress level through a machine learning model.

In recent years, different machine learning models have been applied to predict stress and other similar health indexes. Multi-Linear Regression (MLR) and Support Vector Machine (SVM) perform well when the data appears to be linearly separable; these methods have been applied to phone call/email usage and mobility metrics to predict mood and depression [12, 38]. Random Forest (RF) and Adaboost (AB) [15, 22], classifiers with a decision tree structure, are good at detecting impactful signals and have outperforming several other baselines when predicting binary stress labels from mobile phone usage and weather [7]. The k-Nearest Neighbors (KNN) graph serves as another non-neural network classifier, used in several diagnostic applications [16, 41, 48]. In the case of Artificial Neural Network (ANN), adding a task-specific layer for each person could significantly improve the prediction performance [27].

2.2 Data Completion

For many regression and classification tasks, when data appears as a matrix with missing values, researchers would require the completed matrix instead of dropping all rows with missing data [47]. Besides the most basic strategy, data padding and data interpolation, recent studies proposed various matrix completion method for different objectives. For example, Maximum Likelihood Estimation (MLE) is a useful tool to recover the matrix with sparse missing values or errors [13]. Matrix Factorization (MF), along with its non-negative version, NMF, is another popular method which generates a low-rank matrix from sparse observations. The drawback of MF and NMF is that the algorithm will not fix the original data but only minimize the difference between the original matrix and the completed matrix. Alternatively, Nuclear Norm Minimization (NNM), inspired by the same assumption that the matrix has a low-rank structure, will keep the existing data unchanged. Instead of forcing the rank of the completed matrix to be r , NNM minimizes a convex relaxation of matrix rank. In [44], researchers proved that given the underlying matrix is low-rank and the observed entries are sampled uniformly at random, one can achieve exact recovery with high probability under mild additional assumptions by using NNM.

Based on NNM, the performance of matrix completion could be further improved by adding a specific regularizer. This regularizer is usually from the knowledge of some known pattern of the given matrix. For example, if we are recovering an image with a piece-wise smooth pattern, adding the total variation regularizer could improve the performance of the task [24]. If one assumes entries with smaller absolute value are more likely to be missing, one can then add a regularization term for unobserved entries that penalize the size of the missing values [40].

3 PROBLEM STATEMENT

We first formally define the objectives of predicting human stress level with sparse and incomplete sensor data. Let $\mathbf{U}$ and $\mathbf{S}$ be the set of m users and the set of activities that can be recorded by wearable devices. For each activity $s \in \mathbf{S}$, $M^s \in \mathbb{R}^{m \times n}$ denotes the sensor data that should be collected by wearable devices for n hours. Each entry $M_{ij}^s \in M^s$ denotes the percentage of time doing the activity s in the j -th hour for the i -th user $u_i \in \mathbf{U}$. Note that entries in M^s can be missing and unavailable because users may not always wear the devices. Here the set of recorded hours Ω for all users can be defined as:

$$\Omega = \{(i, j) \mid M_{ij}^s \text{ exists}\}. \quad (1)$$

The recorded data can then be described as a linear mapping $P_\Omega(M^s) : \mathbb{R}^{m \times n} \rightarrow \mathbb{R}^{m \times n}$, where

$$[P_\Omega(M^s)]_{ij} = \begin{cases} M_{ij}^s, & (i, j) \in \Omega \\ 0, & (i, j) \notin \Omega. \end{cases} \quad (2)$$

Each user $u_i \in \mathbf{U}$ also has a set of features $\mathbf{r}_{u_i}$ to indicate personal characteristics. The two goals of this paper are listed as follows:

- (1) **Sensor Data Completion:** Given the recorded data $P_\Omega(M^s)$, the first goal is to recover the accurate sensor data M^s . More precisely, for each activity s , we aim to infer a matrix X^s so that X^s can be as similar to M^s as possible.
- (2) **Stress Level Prediction:** For each user $u_i \in \mathbf{U}$, the next goal is to predict the stress level $c_{u_i} \in C$ of the user with the recovered sensor data X^s , where C is a predefined sets of stress levels. Specifically, a predicted level $\hat{c}_{u_i} \in C$ is inferred for the user u_i so that $\hat{c}_{u_i}$ can be likely to be c_{u_i} . In this paper, we consider both binary-class and 5-class stress level prediction.

4 PROPOSED APPROACH

In this section, we present the proposed framework with two stages, including Data Completion with Diurnal Regularizers (DCDR) and Temporally Hierarchical Attention Network (THAN), for two goals in this paper.

4.1 Data Completion with Diurnal Regularizers

To recover the sensor data, we follow the low-rank assumption [35] to define a minimization problem that can be approximately solved by the alternating direction method of multipliers (ADMM) [24, 52]. **Nuclear Norm Minimization.** Based on the low-rank assumption, the task of recovering M^s with a low rank can be formally defined as the following optimization problem:

$$\begin{aligned} X^s &= \underset{X}{\operatorname{argmin}} \operatorname{rank}(X) \\ \text{s.t. } P_\Omega(X) &= P_\Omega(M^s). \end{aligned} \quad (3)$$

Unfortunately, the problem has been proven to be NP-hard [50]. Therefore, we follow Candès and Tao [11] to apply the nuclear norm $\|X\|_*$ as a substitution and convert the original problem into a convex optimization problem of nuclear norm minimization (NNM):

$$\begin{aligned} X^s &= \underset{X}{\operatorname{argmin}} \|X\|_* \\ \text{s.t. } P_\Omega(X) &= P_\Omega(M^s). \end{aligned} \quad (4)$$

Here we denote $\|X\|_* = \sum_{k=1}^n \sigma_k(X)$ as the nuclear norm of a matrix X , where $\sigma_k(X)$ is the k -th singular value of X . The nuclear norm is connected with matrix rank since matrix rank can be interpreted as the number of non-zero singular values.

Daily and Weekly Regularizers. In addition to solving NNM, previous studies [24, 40] have demonstrated that the performance of matrix completion can be improved by specific regularizations according to the data structure. Since the sensor data are generated as a temporal sequence, we regularize the differences across different days and different weeks. Note that the regularizers should be convex so that the optimization problem can have a proper unique solution for further NNM optimization.

We first assume that people should have relatively regular diurnal activity patterns. Hence, a diurnal penalizer can regularize diurnal patterns as:

$$D_{i,j}^d(X) = \begin{cases} X_{i,j+24} - X_{i,j}, & 1 \leq j \leq n - 24, \\ 0, & j > n - 24. \end{cases} \quad (5)$$

Similar, weekly patterns can also imply a weekly penalizer as:

$$D_{i,j}^w(X) = \begin{cases} X_{i,j+168} - X_{i,j}, & 1 \leq j \leq n - 168, \\ 0, & j > n - 168. \end{cases} \quad (6)$$

The regularizer based on two patterns can then be computed as:

$$\|X\|_{Reg} = \gamma_1 \sum_{i,j} (D_{i,j}^d(X))^2 + \gamma_2 \sum_{i,j} (D_{i,j}^w(X))^2, \quad (7)$$

where γ_1 and γ_2 control the importance of two penalizers. Finally, the minimization problem becomes

$$\begin{aligned} X^S &= \underset{X}{\operatorname{argmin}} (||X||_* + ||X||_{Reg}) \\ \text{s.t. } P_\Omega(X) &= P_\Omega(M^S). \end{aligned} \quad (8)$$

Optimization with ADMM. ADMM [24, 52] is a general algorithm to solve convex optimization problems with the following form:

$$\begin{aligned} \underset{x_1, x_2}{\operatorname{argmin}} & f_1(x_1) + f_2(x_2) \\ \text{s.t. } & A_1 x_1 + A_2 x_2 = b, \end{aligned} \quad (9)$$

where f_1 and f_2 are convex functions of x_1 and x_2 , respectively; A_1 and A_2 are linear constraints. The idea is to breakdown the expression into smaller convex pieces and update x_1 , x_2 separately so that each step become easier to solve. To solve Eq. (8) through ADMM, we first follow Han et al. [24] to convert the regularizer shown in Eq. (7) into the following form:

$$||X||_{Reg} = \gamma_1 ||X - X\phi_1||_F^2 + \gamma_2 ||X - X\phi_2||_F^2, \quad (10)$$

where ϕ_1 and ϕ_2 are two auxiliary matrices defined as.

$$\phi_1 = \begin{bmatrix} O_{24, (n-24)} & O_{n-24, 24} \\ I_{n-24} & I_{24} \end{bmatrix}, \quad \phi_2 = \begin{bmatrix} O_{168, (n-168)} & O_{n-168, 168} \\ I_{n-168} & I_{168} \end{bmatrix},$$

where $O_{i,j}$ is an $i \times j$ zero sub-matrix; I_k is an identity sub-matrix of size k . Therefore, our minimization problem (8) can be reformulated as:

$$\begin{aligned} X^S &= \underset{X}{\operatorname{argmin}} (||X||_* + \gamma_1 ||X - X\phi_1||_F^2 + \gamma_2 ||X - X\phi_2||_F^2) \\ \text{s.t. } & P_\Omega(X) = P_\Omega(M^S). \end{aligned}$$

To fit our optimization problem into ADMM, we introduce an auxiliary variable Y such that $P_{\Omega^C}(Y) = P_{\Omega^C}(X)$, where Ω^C is the complement set of the observation set Ω . To optimize Eq. (4.1), two auxiliary matrices can then be computed as:

$$\begin{aligned} X_D^S &= \underset{X, Y}{\operatorname{argmin}} ||X||_* + \gamma_1 ||Y - Y\phi_1||_F^2 \\ \text{s.t. } & P_\Omega(X) = P_\Omega(M) \text{ and } P_{\Omega^C}(X) = P_{\Omega^C}(Y). \end{aligned} \quad (11)$$

$$\begin{aligned} X_W^S &= \underset{X, Y}{\operatorname{argmin}} ||X||_* + \gamma_2 ||Y - Y\phi_2||_F^2 \\ \text{s.t. } & P_\Omega(X) = P_\Omega(M) \text{ and } P_{\Omega^C}(X) = P_{\Omega^C}(Y). \end{aligned} \quad (12)$$

Therefore, the optimization problem now has the general form of ADMM as shown in Eq (9). More precisely, $f_1(X) = ||X||_*$ is a convex functions on X ; $f_2(Y) = ||Y - Y\phi_1||_F^2$ is a convex functions on Y ; $P_\Omega(X) = P_\Omega(M)$ and $P_{\Omega^C}(X) = P_{\Omega^C}(Y)$ are the two linear constraints. Based on ADMM, the augmented Lagrangian of Eq. (11) can be derived as:

$$\begin{aligned} L_t(X_D, Y_D, Z_D) &= ||X_D||_* + \gamma_1 ||Y_D - Y_D\phi_1||_F^2 \\ &\quad + \frac{t}{2} ||X_D - Y_D||_F^2. \end{aligned}$$

where Z is the dual variable and $t > 0$ is the parameter for quadratic penalty. At iteration k , ADMM first minimizes L_t over x_1 while fixing x_2 and z : $X_D^{s, k+1}$, $Y_D^{s, k+1}$, and $Z_D^{s, k+1}$ can then be respectively

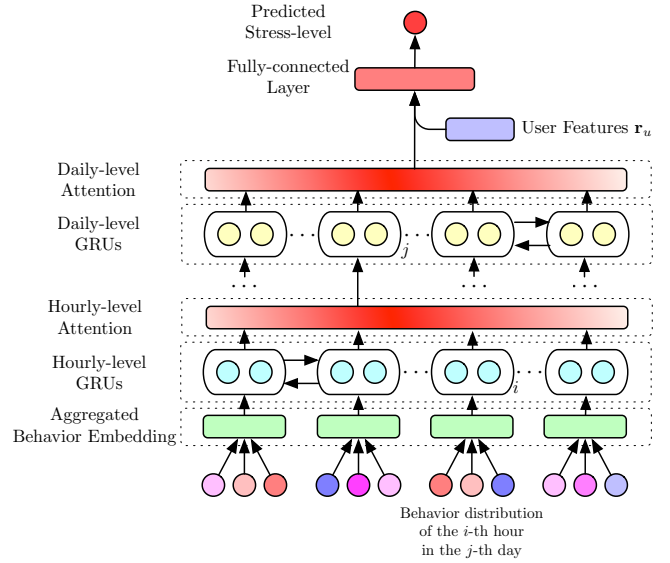


Figure 1: The overall schema of THAN.

updated with three sub-problems as follows::

$$X_D^{s, k+1} = \underset{X}{\operatorname{argmin}} L_t(X, Y_D^{s, k}, Z_D^{s, k}), \quad (13)$$

$$Y_D^{s, k+1} = \underset{Y}{\operatorname{argmin}} L_t(X_D^{s, k+1}, Y, Z_D^{s, k}), \quad (14)$$

$$Z_D^{s, k+1} = Z_D^{s, k} + t(Y_D^{s, k+1} - X_D^{s, k+1}). \quad (15)$$

The sub-problem (4.1) can be solved by singular value shrinkage [9] as:

$$X_D^{s, k+1} = \mathcal{D}_{(1-\gamma_1)/t}(Y_D^{s, k} + \frac{1}{t}Z_D^{s, k}), \quad (16)$$

where the soft-thresholding operator $\mathcal{D}$ is defined as:

$$\mathcal{D}_\tau(X) = U \mathcal{D}_\tau(\Sigma) V^*, \quad \mathcal{D}_\tau(\Sigma) = \operatorname{diag}(\{\sigma_i - \tau\}_+). \quad (17)$$

The sub-problem (4.1) can be solved by computing $\partial L_t / \partial Y$ and setting it to be zero:

$$Y_D^{k+1} = (\lambda X_D^{s, k+1} - Z_D^{s, k}) [2\gamma_1 (I - \phi_1 - \phi_1^T + \phi_1^T \phi_1) + tI]^{-1}. \quad (18)$$

The sub-problem can directly be updated with the solutions of Eq. and . The optimization process iteratively solves sub-problems and stops when $X_D^{s, k}$ converges to X_D^S . Similarly, Eq. (12) for obtaining X_W^S can be also optimized with the same approach. Finally, we complete the sensor data matrix by $X^S = (X_D^S + X_W^S)/2$.

4.2 THAN for Stress Level Prediction

To predict human stress level, we propose Temporally Hierarchical Attention Network (THAN) to exploit temporally hierarchical structures to model behaviors. To ease the discussion, we focus on the two-level temporal structure, including daily- and hourly-level, while the depth of the hierarchical structure can be simply expanded to monthly-level or replaced with other temporally hierarchical structures. Note that although the hierarchical structures have been applied in natural language processing [31, 53], this paper is the first work to apply the idea in stress level prediction.

Figure 1 shows the overall schema of the proposed THAN for stress level prediction. Each activity s has a trainable embedding vector so that user behaviors in each hour can be represented by an aggregated behavior embedding vector with the recovered sensor data M^s . The hourly-level RNN models user behaviors in each day with hourly-level attention while the daily-level RNN derives an overall sensor representation with daily-level attention. Finally, the sensor representation and user features can be applied to predict the stress level with a fully-connected hidden layer. More details are explained as follows.

Aggregated Behavior Embedding. Suppose we are predicting the stress level of the user $u \in U$. Since the wearable devices can only provide aggregated activity information for sensor data, we propose aggregated behavior embedding to represent user behaviors in a latent space. For each activity $s \in S$, we learn a d_s -dimensional behavior embedding $\mathbf{e}_s$. For the j -th hour of the i -th day, the aggregated behavior embedding $\mathbf{a}_{i,j}$ is defined as:

$$\mathbf{a}_{i,j} = \sum_{s \in S} X_{u,i,j}^s \cdot \mathbf{e}_s, \quad (19)$$

where $X_{u,i,j}^s$ is the recovered sensor data about the activity s in the j -th hour of the i -th day for the user u .

Hourly-level RNN. THAN follows a bottom-up approach for computations with bidirectional recurrent neural networks (Bi-RNNs) with attention [39]. To encode the behaviors in the i -th day, a Bi-RNN scans the sequences of corresponding aggregated behavior embeddings during both forward and backward passes. In the forward pass, the Bi-RNN generates a sequence of hidden states $[\vec{\mathbf{h}}_{i,1}, \vec{\mathbf{h}}_{i,2}, \dots, \vec{\mathbf{h}}_{i,24}]$, where $\vec{\mathbf{h}}_{i,j} = \text{RNN}(\vec{\mathbf{h}}_{i,j-1}, \mathbf{a}_{i,j})$ is a d_g -dimensional hidden states generated by a dynamic function such as LSTM [26] or GRU [17]. Here we use GRU instead of LSTM because it requires fewer parameters [32]. The backward pass then processes the sequence reversely and derives the backward hidden states $[\overleftarrow{\mathbf{h}}_{i,1}, \overleftarrow{\mathbf{h}}_{i,2}, \dots, \overleftarrow{\mathbf{h}}_{i,24}]$, where $\overleftarrow{\mathbf{h}}_{i,j} = \text{RNN}(\overleftarrow{\mathbf{h}}_{i,j+1}, \mathbf{a}_{i,j})$. The forward and backward hidden states are then concatenated as hidden representations for hours in the day as follows: $[\vec{\mathbf{h}}_{i,1}, \mathbf{h}_{i,2}, \dots, \mathbf{h}_{i,24}]$, where $\mathbf{h}_{i,j} = [\vec{\mathbf{h}}_{i,j}; \overleftarrow{\mathbf{h}}_{i,j}]$.

To estimate the importance of each hour in the day, the attention mechanism [39] is applied to extract and aggregate important hidden representations. More precisely, the importance $\alpha_{i,j}$ of $\mathbf{h}_{i,j}$ can be estimated as:

$$\alpha_{i,j} = \frac{\exp(\mathbf{z}_{i,j} \cdot \mathbf{z}^H)}{\sum_{j'} \exp(\mathbf{z}_{i,j'} \cdot \mathbf{z}^H)}, \quad (20)$$

where $\mathbf{z}_{i,j} = \tanh(\mathcal{F}^H(\mathbf{h}_{i,j}))$; $\mathcal{F}^H(\cdot)$ is a fully-connected layer; $\tanh$ is the activation for computing similarity; $\mathbf{z}^H$ is the context vector to measure the importance of each hour. Finally, the representation of the i -th day can be represented as the weighted sum of the hidden representations as follows:

$$\mathbf{a}_i = \sum_j \alpha_{i,j} \cdot \mathbf{h}_{i,j}. \quad (21)$$

Daily-level RNN. With the daily representations, the Bi-RNN is applied again to derive the overall representation for stress level

prediction. The sequences of forward and backward hidden states can be generated as $[\vec{\mathbf{h}}_1, \vec{\mathbf{h}}_2, \dots, \vec{\mathbf{h}}_L]$ and $[\overleftarrow{\mathbf{h}}_1, \overleftarrow{\mathbf{h}}_2, \dots, \overleftarrow{\mathbf{h}}_L]$, where L is the number of days for sensor data; $\vec{\mathbf{h}}_i = \text{RNN}(\vec{\mathbf{h}}_{i-1}, \mathbf{a}_i)$; $\overleftarrow{\mathbf{h}}_i = \text{RNN}(\overleftarrow{\mathbf{h}}_{i+1}, \mathbf{a}_i)$. The hidden representations for all days can then be represented as $[\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_L]$, where $\mathbf{h}_i = [\vec{\mathbf{h}}_i; \overleftarrow{\mathbf{h}}_i]$.

The importance of each day α_i in sensor day can also be estimated by the attention mechanism as follows:

$$\alpha_i = \frac{\exp(\mathbf{z}_i \cdot \mathbf{z}^D)}{\sum_{i'} \exp(\mathbf{z}_{i'} \cdot \mathbf{z}^D)}, \quad (22)$$

where $\mathbf{z}_{i,j} = \tanh(\mathcal{F}(\mathbf{h}_{i,j}))$; $\mathcal{F}(\cdot)$ is a fully-connected later

To estimate the importance of each hour in the day, the attention mechanism [39] is applied to extract and aggregate important hidden representations. More precisely, the importance $\alpha_{i,j}$ of $\mathbf{h}_{i,j}$ can be estimated as:

$$\alpha_{i,j} = \frac{\exp(\mathbf{z}_{i,j} \cdot \mathbf{z}^H)}{\sum_{j'} \exp(\mathbf{z}_{i,j'} \cdot \mathbf{z}^H)}, \quad (23)$$

where $\mathbf{z}_{i,j} = \tanh(\mathcal{F}^D(\mathbf{h}_{i,j}))$; $\mathcal{F}^D(\cdot)$ is a fully-connected layer; $\mathbf{z}^D$ is the context vector to measure the importance of each day. Finally, the overall representation of sensor data can be derived as:

$$\mathbf{a} = \sum_i \alpha_i \cdot \mathbf{h}_i. \quad (24)$$

Stress Level Prediction. The representation of sensor data $\mathbf{a}$ presents the user behaviors while the user features $\mathbf{r}_u$ provides the aspects of user characteristics. To precisely predict the stress level of the user, we apply a fully-connected hidden layer to combine the knowledge of two resources as:

$$\hat{c}_u = \underset{c}{\operatorname{argmax}} \mathcal{F}^c(\mathcal{F}^P([\mathbf{a}; \mathbf{r}_u])), \quad (25)$$

where $\mathcal{F}^P(\cdot)$ is a fully-connected layer with d_p hidden neurons; $\mathcal{F}^c(\cdot)$ computes the logit of the stress level $c \in C$ for classification.

Learning and Optimization. The task of stress level prediction can be modeled as a classification problem. Here we adopt the categorical cross-entropy [21] as the loss function for optimization. More precisely, the loss function for each training example can be written as

$$-\sum_c \mathbb{1}(c = c_u) \log(\mathcal{F}^c(\mathcal{F}^P([\mathbf{a}; \mathbf{r}_u]))), \quad (26)$$

where $\mathbb{1}(\cdot)$ is an indicator function; c_u is the ground truth of the stress level.

User Features. Any user-related information can be represented in the user feature vectors to potentially boost the prediction performance. In this paper, we adopt historical stress levels in the previous 11 weeks as user features. The mean and variance values of historical stress levels are also considered. Finally, there are 13 user features in the experiments. Note that we also evaluate the performance without using user features to demonstrate the effectiveness of both of the proposed model and user features.

Table 1: The sensor data completion performance of five methods for the activities *Sleep* and *Walk*. All improvements of DCDR against baseline methods are significant at 99% level in a paired *t*-test.

Method	<i>Sleep</i>						<i>Walk</i>					
	RMSE	MAE	Accuracy	Precision	Recall	F1	RMSE	MAE	Accuracy	Precision	Recall	F1
ITP [34]	0.2957	0.1975	0.8488	0.7249	0.7481	0.7363	0.1027	0.0613	0.6544	0.7395	0.4072	0.5252
MF [8]	0.3011	0.2010	0.8568	0.6964	0.7873	0.7390	0.1111	0.0504	0.7439	0.0191	0.6618	0.0371
NMF [42]	0.3053	0.2055	0.8514	0.6796	0.7815	0.7270	0.1113	0.0504	0.7996	0.7432	0.0127	0.0250
ADMM [52]	0.3096	0.2089	0.8433	0.5880	0.8472	0.6942	0.1025	0.0552	0.7202	0.5841	0.4301	0.4954
DCDR	0.2861	0.1754	0.8671	0.7464	0.7864	0.7659	0.1039	0.0573	0.7200	0.5999	0.4676	0.5256

5 EXPERIMENTS

In this section, we conduct extensive experiments and in-depth analysis to evaluate the proposed DCDR and THAN.

5.1 Experimental Settings

Data Collection. To obtain the datasets, we filtered data from a longitudinal study among college students, designed to collect twitter, survey, and wearable device data for 12 weeks from September 28, 2015, to December 20, 2015.

The students were asked to fill in an online survey and report their stress levels. There are five options: 1 (Not at all), 2 (Low), 3 (Average), 4 (High), and 5 (Extremely high). For the purpose of this paper, we view the survey data as "ground truth". Each of the students was invited/requested to wear a smart-wristband to record hourly aggregated information of activities, including *sleep*, *walk*, *run*, *bike*, and *moderate activities*, as the collected sensor data.

Incomplete data should at least have basic structural information, so we filtered out students with ≤ 2 survey taken and ≤ 2 hours sleeping record. This leaves 75 students and only 52% entries of the sensor data observed. Finally, we have 855 records of weekly reported stress levels with incomplete sensor data. Note that the number is not $75 * 12$ because some students did not turn in the survey every week, so that some ground truth data is missing.

Training Data for Algorithms. For the task of sensor data completion, we randomly sample 90% of observed entries for training and evaluate DCDR with the remaining observed entries. For the task of stress level prediction, we treat each week for each student as an instance with the corresponding sensor data and the stress level indicated in the survey. We then randomly sample 80% of instances as training data and evaluate THAN with the remaining instances. Finally, we have 7,0786 and 7,865 training and testing entries for sensor data completion while 684 and 171 records are applied for training and testing in stress level prediction. For the set of stress levels C , we attempt two different partitions, including binary-class and 5-class stress levels. In the binary-class setting, we consider the students who reported stress levels greater than 3 as stressed students.

Implementation Details. We implement DCDR in Matlab while THAN is implemented in Tensorflow [2]. The Adam optimizer [33] is applied to optimize THAN with a 10^{-4} initial learning rate. After parameter tuning, γ_1 and γ_2 in DCDR are set as 0.2154 and 0.0774. The number of dimensions for behavior embeddings d_s in THAN is 4 while the numbers of dimensions for GRU hidden states and context vectors of attention mechanism are 64. The number of hidden neurons in the fully-connected layer is 128.

Evaluation Metrics. For sensor data completion, we treat the problem as a regression task and consider Root Mean Squared Error (RMSE) and Mean Average Error (MAE) as the evaluation metrics. In addition, we also use a threshold 0.5 to convert the task into a classification problem while accuracy, precision, recall, and F1 score can be applied for evaluation. Stress level prediction is a classification task, so accuracy, precision, recall, and F1 score are considered as the evaluation metrics.

5.2 Sensor Data Completion

To evaluate the performance of DCDR, we compare with four conventional approaches for data completion, including interpolation (ITP) [34], matrix factorization (MF) [8], non-negative matrix (NMF) [42], and NNM [24, 52]. Note that NNM is a special case of DCDR at $\gamma_1 = \gamma_2 = 0$. Table 1 shows the performance of five methods in the task of sensor data completion for two activities. For the *Sleep* activity, baseline methods have similar and good performance in both regression and classification tasks because the low-rank assumption and data redundancy are leveraged. Our proposed DCDR significantly outperforms all of the baseline methods. This is because the regularizers capture the patterns of user behaviors instead of only modeling sensor data as a low-rank matrix. For the *Walk* activity, it is interesting that all of the methods perform worse because people usually have diurnal sleeping patterns while the walking behaviors are more random and irregular. This is also the reason that THAN does not improve ADMM by adding regularizers.

5.3 Stress Level Prediction

To evaluate the performance of THAN, we compare with seven baseline methods introduced in Section 2.1, including decision tree (DT) [49], random forest (RF) [7], AdaBoost (AB) [15, 22], k nearest neighbor (KNN) [16], support vector machine (SVM) [7], multi-linear regression (MLR) [38], and artificial neural network (ANN) [25]. Table 2 and 3 indicate the performance of eight methods in stress level prediction with binary-class and 5-class stress levels. For the baseline methods, RF performs the best because of its robustness with limited training data. The complicated methods, such as SVM and ANN, perform worse because more training data are needed without considering temporally structural information. Our proposed THAN significantly outperforms all of the baseline methods in both binary-class and 5-class stress level prediction. This is because the temporally hierarchical structures can precisely capture

Table 2: The stress level prediction performance of eight methods with binary-class stress levels. All improvements of THAN against baseline methods are significant at 99% level in a paired t -test.

Method	Accuracy	Precision	Recall	F1
DT [49]	0.6433	0.6111	0.5714	0.5906
RF [7]	0.7076	0.7288	0.5584	0.6324
AB [15, 22]	0.6784	0.6774	0.5455	0.6043
KNN [16]	0.5789	0.5352	0.4935	0.5135
SVM [7]	0.6491	0.6735	0.4286	0.5238
MLR [38]	0.5965	0.5541	0.5325	0.5430
ANN [25]	0.6608	0.6727	0.4805	0.5606
THAN	0.7661	0.6796	0.9091	0.7778

Table 3: The stress level prediction performance of eight methods with 5-class stress levels. All improvements of THAN against baseline methods are significant at 99% level in a paired t -test.

Method	Accuracy	Micro-F1	Macro-F1
DT [49]	0.2982	0.2982	0.2747
RF [7]	0.4561	0.4561	0.4631
AB [15, 22]	0.4737	0.4737	0.4389
KNN [16]	0.2982	0.2982	0.2568
SVM [7]	0.3216	0.3216	0.1861
MLR [38]	0.4737	0.4737	0.4389
ANN [25]	0.3509	0.3509	0.3013
THAN	0.4757	0.4757	0.4706

the structural patterns of user behaviors while the attention mechanism is capable of effectively seizing essential information for predicting human stress level.

5.4 Analysis and Discussions

In this section, we conduct ablation studies and parameter sensitivity analysis to demonstrate the robustness and effectiveness of our proposed approach.

Effectiveness of Sensor Data Completion. We first verify the effectiveness of recovered sensor data for predicting the stress level. Figure 2 illustrates the F1 scores of eight methods in stress level prediction with and without sensor data completion. Almost all of the methods can be improved by sensor data completion. The results show that although about 50% of sensor data are unobserved, DCDR can still appropriately recover missing data, thereby benefiting the downstream applications.

Depth of Temporally Hierarchical Structures. One of the contributions in THAN is to consider temporally hierarchical structures, so here we verify the effectiveness of this idea. Table 4 shows the performance of THAN in stress level prediction with different depths of temporally hierarchical structures. Note that depth-1 THAN degenerates to a sequential RNN with attention without considering any hierarchical structure. More specifically, depth-1 THAN only utilizes hourly RNNs when depth-2 THAN exploits both hourly and daily RNNs based on hierarchical structures. The results

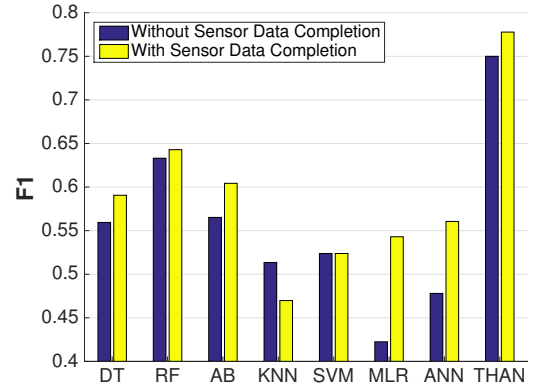


Figure 2: The F1 scores of eight methods in stress level prediction with and without sensor data completion.

Table 4: The stress level prediction performance of THAN with different depths of temporally hierarchical structures.

Depth	Acc	Prec	Rec	F1
1 (Hourly RNNs)	0.7485	0.6771	0.8442	0.7514
2 (Hourly&Daily RNNs)	0.7661	0.6796	0.9091	0.7778

Table 5: The stress level prediction performance of THAN with different features. S denotes the sensor data while U represents the user features.

Features	Accuracy	Precision	Recall	F1
S Only	0.6140	0.5902	0.4675	0.5217
U Only	0.7485	0.7361	0.6883	0.7114
S+U	0.7661	0.6796	0.9091	0.7778

show that the stress level prediction performance can be significantly boosted after considering temporally hierarchical structures. It further demonstrates that the temporally structural information in sensor data is much beneficial for understanding human stress.

Effectiveness of User Features. We propose a set of personal user features for THAN to capture individual characteristics for predicting stress level more precisely. Here we conduct an ablation study to validate the effectiveness of our proposed user features. Table 5 shows the performance of THAN in stress level prediction with different features. Note that S and U in Table 5 represent the usage of sensor data and user features. The results demonstrate that the proposed user features are effective because the historical and statistical features can directly reflect the stress level distribution for a specific user. Without personalized information, THAN with only sensor data can still achieve good performance. It indicates that sensor data and user features all are valuable for stress level prediction so that THAN can achieve the best prediction performance by simultaneously considering both of features.

Regularizers in DCDR. As shown in Section 5.2, regularizers play an important role in sensor data completion, so we conduct a parameter sensitivity analysis for the regularization weights of two

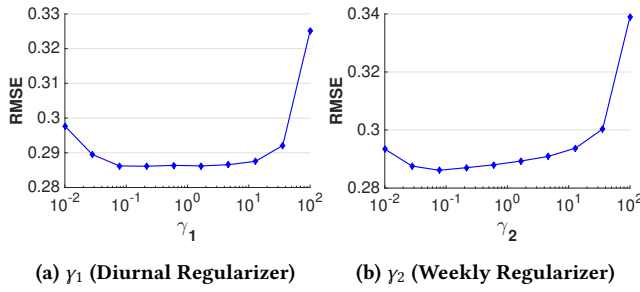


Figure 3: The RMSE scores of DCDR in sensor data completion with different weights of the parameters γ_1 and γ_2 .

regularizers. Figure 3 illustrates the RMSE scores of DCDR in sensor data completion with different weights of γ_1 and γ_2 for diurnal and weekly regularizers. The results show that the regularization weights cannot be too small or too large for satisfactory data completion, so the parameter tuning is essential. Moreover, the best parameters of γ_1 and γ_2 are 0.2154 and 0.0774. It indicates that the diurnal patterns are more effective than the weekly patterns for stress level prediction.

Parameter Sensitivity of THAN. We also conduct the parameter sensitivity analysis for three parameters of THAN, including the numbers of dimensions for behavior embeddings d_s , dimensions for GRU hidden states d_g , and hidden neurons in the last fully-connected layer d_p . Figure 4 shows the F1 scores of THAN in stress level prediction across different parameters. For behavior embeddings, it is interesting that THAN does not favor a large dimension number d_s . This can be because the number of available activities is limited so that it does not require a complicated latent space to represent each activity. For the number of dimensions for GRU hidden states d_g and the number of hidden neurons d_p , greater numbers usually lead to more satisfactory results. However, the model can start over-fitting if the numbers become too large. As a result, it demonstrates again that parameter tuning is required to achieve the best performance.

6 CONCLUSIONS

In this paper, we focus on predicting human stress level with incomplete sensor data from wearable devices. To recover missing sensor data, our model, Data Completion with Diurnal Regularizers (DCDR), exploits user diurnal and weekly behaviors patterns as two effective regularizers, thereby precisely inferring unobserved sensor data. To predict human stress level, we propose Temporally Hierarchical Attention Network (THAN) to consider temporally hierarchical structures in sensor data with hierarchical RNNs with the attention mechanism. The extensive experiments demonstrate that our approach significantly outperforms competitive baseline methods in both sensor data completion and stress level prediction. Moreover, the analysis results also show the robustness and effectiveness of our proposed models. The reasons and insights can be concluded as follows: (1) The conventional low-rank assumption is not enough for precise data completion while the patterns of diurnal behaviors can be beneficial as regularizers; (2) The hierarchical structures can leverage temporal knowledge in sensor data so

that the model does not require much training data for satisfactory prediction performance; (3) Parameter tuning is required to reach the best performance for both DCDR and THAN in sensor data completion and stress level prediction.

ACKNOWLEDGMENT

We would like to thank the anonymous reviewers for their helpful comments. The work was supported by National Institute of Allergy and Infectious Diseases (R56 AI125105; 5R01 AI 132030). We also acknowledge the support of NSF grants DGE-1829071 and DMS-1737770.

REFERENCES

- [1] 2019. Mental Health Information, Statistics of Depression. NIH National Institute of Mental Health. <https://www.nimh.nih.gov/health/statistics/major-depression.shtml>. Accessed on 22 June 2019.
- [2] Martín Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, et al. 2016. Tensorflow: A system for large-scale machine learning. In *12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16)*. 265–283.
- [3] Eirik Årsand, Miroslav Muzny, Meghan Bradway, Jan Muzik, and Gunnar Hartvigsen. 2015. Performance of the first combined smartwatch and smartphone diabetes diary application study. *Journal of diabetes science and technology* 9, 3 (2015), 556–563.
- [4] American Psychological Association et al. 2017. Stress in America: The state of our nation. *Stress in AmericaTM Survey* (2017).
- [5] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2014. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473* (2014).
- [6] Jörn Bakker, Mykola Pechenizkiy, and Natalia Sidorova. 2011. What's your current stress level? Detection of stress patterns from GSR sensor data. In *2011 IEEE 11th international conference on data mining workshops*. IEEE, 573–580.
- [7] Andrey Bogomolov, Bruno Lepri, Michela Ferron, Fabio Pianesi, and Alex Sandy Pentland. 2014. Daily stress recognition from mobile phone data, weather conditions and individual traits. In *Proceedings of the 22nd ACM international conference on Multimedia*. ACM, 477–486.
- [8] Aeron M Buchanan and Andrew W Fitzgibbon. 2005. Damped newton algorithms for matrix factorization with missing data. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, Vol. 2. IEEE, 316–322.
- [9] Jian-Feng Cai, Emmanuel J Candès, and Zuowei Shen. 2010. A singular value thresholding algorithm for matrix completion. *SIAM Journal on Optimization* 20, 4 (2010), 1956–1982.
- [10] Emmanuel J Candès and Yaniv Plan. 2010. Matrix completion with noise. *Proc. IEEE* 98, 6 (2010), 925–936.
- [11] Emmanuel J Candès and Terence Tao. 2009. The power of convex relaxation: Near-optimal matrix completion. *arXiv preprint arXiv:0903.1476* (2009).
- [12] 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.
- [13] Jinhui Chen and Jian Yang. 2013. Low-Rank Matrix Completion Based on Maximum Likelihood Estimation. In *2013 2nd IAPR Asian Conference on Pattern Recognition*. IEEE, 261–265.
- [14] Kemeng Chen, Wolfgang Fink, Janet Roveda, Richard D Lane, John Allen, and Johnny Vanuk. 2015. Wearable sensor based stress management using integrated respiratory and ECG waveforms. In *2015 IEEE 12th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*. IEEE, 1–6.
- [15] Bo Cheng. 2011. Emotion recognition from physiological signals using AdaBoost. In *International Conference on Applied Informatics and Communication*. Springer, 412–417.
- [16] Mohamed Amine Chikh, Meryem Saidi, and Nesma Settouti. 2012. Diagnosis of diabetes diseases using an artificial immune recognition system2 (AIRS2) with fuzzy k-nearest neighbor. *Journal of medical systems* 36, 5 (2012), 2721–2729.
- [17] Kyunghyun Cho, Bart Van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. 2014. Learning phrase representations using RNN encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078* (2014).
- [18] Mostafa Dehghani, Sascha Rothe, Enrique Alfonseca, and Pascal Fleury. 2017. Learning to attend, copy, and generate for session-based query suggestion. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*. ACM, 1747–1756.

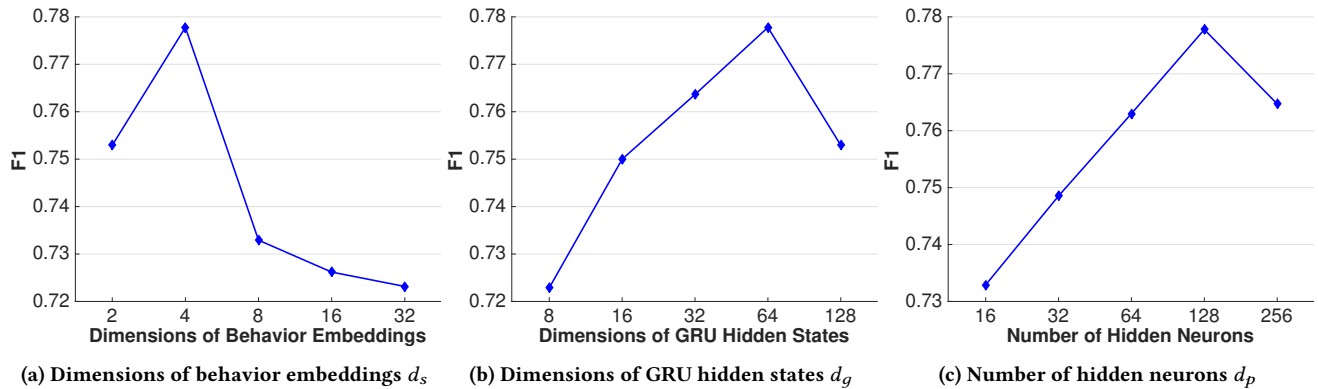


Figure 4: The F1 scores of THAN in stress level prediction across different parameters.

- [19] Susan Folkman. 2013. *Stress: appraisal and coping*. Springer.
- [20] Renee Garrett, Sam Liu, and Sean D Young. 2017. A longitudinal analysis of stress among incoming college freshmen. *Journal of American college health* 65, 5 (2017), 331–338.
- [21] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. 2016. *Deep learning*. MIT press.
- [22] Gábor Gósztoya, Róbert Busa-Fekete, and László Tóth. 2013. Detecting autism, emotions and social signals using AdaBoost. *Interspeech*.
- [23] Nils Y Hammerla, Shane Halloran, and Thomas Plötz. 2016. Deep, convolutional, and recurrent models for human activity recognition using wearables. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*. AAAI Press, 1533–1540.
- [24] Xu Han, Jiasong Wu, Lu Wang, Yang Chen, Lotfi Senhadji, and Huazhong Shu. 2014. Linear total variation approximate regularized nuclear norm optimization for matrix completion. In *Abstract and Applied Analysis*, Vol. 2014. Hindawi.
- [25] John HL Hansen and Brian D Womack. 1996. Feature analysis and neural network-based classification of speech under stress. *IEEE Transactions on Speech and Audio Processing* 4, 4 (1996), 307–313.
- [26] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural Computation* 9, 8 (1997), 1735–1780.
- [27] Natasha Jaques, Sara Taylor, Akane Sano, Rosalind Picard, et al. 2017. Predicting tomorrow's mood, health, and stress level using personalized multitask learning and domain adaptation. In *IJCAI 2017 Workshop on Artificial Intelligence in Affective Computing*. 17–33.
- [28] Hayeon Jeong, HeePyung Kim, Rihun Kim, Uichin Lee, and Yong Jeong. 2017. Smartwatch wearing behavior analysis: a longitudinal study. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3 (2017), 60.
- [29] He Jiang, Xin Chen, Shuwei Zhang, Xin Zhang, Weiqiang Kong, and Tao Zhang. 2015. Software for wearable devices: Challenges and opportunities. In *2015 IEEE 39th Annual Computer Software and Applications Conference*, Vol. 3. IEEE, 592–597.
- [30] Jyun-Yu Jiang and Wei Wang. 2018. RIN: Reformulation Inference Network for Context-Aware Query Suggestion. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*. ACM, 197–206.
- [31] Jyun-Yu Jiang, Mingyang Zhang, Cheng Li, Michael Bendersky, Nadav Golbandi, and Marc Najork. 2019. Semantic Text Matching for Long-Form Documents. In *The World Wide Web Conference*. ACM, 795–806.
- [32] Rafal Jozefowicz, Wojciech Zaremba, and Ilya Sutskever. 2015. An empirical exploration of recurrent network architectures. In *Proceedings of the 32nd International Conference on Machine Learning (ICML '15)*. 2342–2350.
- [33] Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).
- [34] Anil C Kokaram, Robin D Morris, William J Fitzgerald, and Peter JW Rayner. 1995. Interpolation of missing data in image sequences. *IEEE Transactions on Image Processing* 4, 11 (1995), 1509–1519.
- [35] Yehuda Koren, Robert Bell, and Chris Volinsky. 2009. Matrix factorization techniques for recommender systems. *Computer* 8 (2009), 30–37.
- [36] Guokun Lai, Wei-Cheng Chang, Yiming Yang, and Hanxiao Liu. 2018. Modeling long- and short-term temporal patterns with deep neural networks. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*. ACM, 95–104.
- [37] Richard S Lazarus. 2006. *Stress and emotion: A new synthesis*. Springer Publishing Company.
- [38] Robert LiKamWa, Yunxin Liu, Nicholas D Lane, and Lin Zhong. 2013. Moodscope: Building a mood sensor from smartphone usage patterns. In *Proceeding of the 11th annual international conference on Mobile systems, applications, and services*. ACM, 389–402.
- [39] Thang Luong, Hieu Pham, and Christopher D Manning. 2015. Effective Approaches to Attention-based Neural Machine Translation. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*. 1412–1421.
- [40] Denali Molitor and Deanna Needell. 2018. Matrix Completion for Structured Observations. *arXiv preprint arXiv:1801.09657* (2018).
- [41] Oscar Martinez Mozos, Virginia Sandulescu, Sally Andrews, David Ellis, Nicola Bellotto, Radu Dobrescu, and Jose Manuel Ferrandez. 2017. Stress detection using wearable physiological and sociometric sensors. *International journal of neural systems* 27, 02 (2017), 1650041.
- [42] Pentti Paatero and Unto Tapper. 1994. Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values. *Environmetrics* 5, 2 (1994), 111–126.
- [43] Emily A Pierceall and Marybelle C Keim. 2007. Stress and coping strategies among community college students. *Community College Journal of Research and Practice* 31, 9 (2007), 703–712.
- [44] Benjamin Recht, Maryam Fazel, and Pablo A Parrilo. 2010. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM review* 52, 3 (2010), 471–501.
- [45] Robert S Sandler, James E Everhart, Mark Donowitz, Elizabeth Adams, Kelly Cronin, Clifford Goodman, Eric Gemmen, Shefali Shah, Aida Avdic, and Robert Rubin. 2002. The burden of selected digestive diseases in the United States. *Gastroenterology* 122, 5 (2002), 1500–1511.
- [46] Aarti Sathyanarayana, Shafiq Joty, Luis Fernandez-Luque, Ferda Ofli, Jaideep Srivastava, Ahmed Elmagarmid, Teresa Arora, and Shahrad Taheri. 2016. Sleep quality prediction from wearable data using deep learning. *JMIR mHealth and uHealth* 4, 4 (2016), e125.
- [47] Joseph L Schafer and John W Graham. 2002. Missing data: our view of the state of the art. *Psychological methods* 7, 2 (2002), 147.
- [48] Mai Shouman, Tim Turner, and Rob Stocker. 2012. Applying k-nearest neighbour in diagnosing heart disease patients. *International Journal of Information and Education Technology* 2, 3 (2012), 220–223.
- [49] Feng-Tso Sun, Cynthia Kuo, Heng-Tze Cheng, Senaka Buthpitiya, Patricia Collins, and Martin Griss. 2010. Activity-aware mental stress detection using physiological sensors. In *International conference on Mobile computing, applications, and services*. Springer, 282–301.
- [50] Lieven Vandenbergh and Stephen Boyd. 1996. Semidefinite programming. *SIAM review* 38, 1 (1996), 49–95.
- [51] Yuxi Wang, Kaishun Wu, and Lionel M Ni. 2016. Wifall: Device-free fall detection by wireless networks. *IEEE Transactions on Mobile Computing* 16, 2 (2016), 581–594.
- [52] Junfeng Yang and Xiaoming Yuan. 2013. Linearized augmented Lagrangian and alternating direction methods for nuclear norm minimization. *Mathematics of computation* 82, 281 (2013), 301–329.
- [53] Zichao Yang, Diyi Yang, Chris Dyer, Xiaodong He, Alex Smola, and Eduard Hovy. 2016. Hierarchical attention networks for document classification. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. 1480–1489.
- [54] Mi Zhang and Alexander A Sawchuk. 2013. Human daily activity recognition with sparse representation using wearable sensors. *IEEE journal of Biomedical and Health Informatics* 17, 3 (2013), 553–560.
- [55] Tong Zhang, Jue Wang, Liang Xu, and Ping Liu. 2006. Using wearable sensor and NMF algorithm to realize ambulatory fall detection. In *International conference on natural computation*. Springer, 488–491.