Tensor Embedding: A Supervised Framework for Human Behavioral Data Mining and Prediction

Homa Hosseinmardi*  
Department of Computer and Information Science  
University of Pennsylvania  
Philadelphia, Pennsylvania, USA  
homahm@seas.upenn.edu

Amir Ghasemian  
Human Nature Lab  
Yale University  
New Haven, Connecticut, USA  
amir.ghasemian@yale.edu

Kristina Lerman  
Information Sciences Institute  
University of Southern California  
Marina Del Rey, California, USA  
lerman@isi.edu

Shrikanth Narayanan  
Information Sciences Institute  
University of Southern California  
Marina Del Rey, California, USA  
shri@isi.edu

Emilio Ferrara  
Information Sciences Institute  
University of Southern California  
Marina Del Rey, California, USA  
ferrarae@isi.edu

I. INTRODUCTION

Rapid improvements in sensor technology have made continuous, unobtrusive sensing of individuals practical, by providing temporal streams of individual physiological and psychological states, physical activity, and social and environmental contexts [6], [16], [31], [43], [44]. Such data has, in turn, created opportunities for enhanced understanding of factors contributing to mental health and well being, including in the workplace. Several past studies collected multimodal data from individuals in real-world settings in order to infer psychological and health states. For example, the 10-week StudentLife study of Dartmouth students used passive and mobile sensor data to study well being, academic performance and behavioral trends [43], [44]. SNAPSHOT, a 30-day study on MIT undergraduates, used mobile sensors and surveys to understand sleep, social interactions, affect, performance, stress and health [34]. RealityMining, a 9-month study of 75 MIT Media Laboratory students, used mobile sensor data to track the social interactions and networking [10], and the friends-and-families study collected data from 130 adult members of a young family community to study fitness intervention and social incentives [2]. In contrast, our work focuses on individuals in the workplace. Specifically, the present paper is based on a study that examines the complex interplay between individual differences, job performance, and well being in jobs with varying cognitive, affective and social demands, measured both at the workplace (and to some extent, complemented, outside the workplace). More than 50 clinical and other hospital staff were instrumented and assessed with a variety of wearable and environmental sensors during their work shift for a duration of 1 month.

Sensors data are typically collected from participants in their natural settings, continuously and over extended time periods. One challenge in multivariate time series classification/regression tasks is designing the desire features. The data is noisy, with missing values, and comes from multiple sensors, usually with different sampling rate. Aggregation of data across the time dimension may cause loss of temporal patterns [22], [45].

Feature engineering also quickly becomes burdensome, especially when there is more than one target variable to model in human behavioral studies, e.g., five personality traits, stress, depression, performance etc. Deep learning has recently been successfully used for feature extraction and classification/regression [49] from audio, images, social networks and other spatio-temporal data [26], [29], [42]. However,
these models require many samples (instances) for training the models, which is not often feasible in some longitudinal studies.

Another approach is using higher-order decomposition methods, and the latent features obtained from unsupervised factorization [5], [19], [20] can be used as features with any conventional regression method to predict target variables. However, it is unlikely that these new features will have predictive power for all the target variables of interest.

With the increasing demands of problems that involve higher-order data, classification and regression methods which predict the target variable from N-way input have been receiving increasing attention [11]–[13], [21], [23], [24], [40], [41], [48], [52], [53]. These methods have been widely applied in neural signal processing, image and video processing, and computational chemistry. However, data collected from multiple sources may consist of both higher-order tensors (e.g., heart rate time series) and matrices (e.g., demographic information such as gender, job, etc.), and coupling different dimensionality of data sources is a challenge for these methods. [1] proposed unsupervised joint decomposition of tensors and matrices when they have at least one common mode, while there is no guarantee that the extracted latent features have predictive power for different dependent variables. Other higher-order regression methods can not accept inputs in such format [50], [52]. While previous works have focused on prediction of one or few aspects of human behavior [16], with hand crafted feature engineering, we are interested in proposing an automated method to find most relevant features for a set of different prediction tasks.

In order to address this challenge, we would like to identify the relevant latent factors using a supervised higher-order decomposition method. It is possible to combine latent features from data sources of varying dimensions with supervised embedding, in which case any regression function may be used for the prediction task. As opposed to models and features that address a limited number of human dimensions, our method provides an overall estimation of predictability. Additionally, we define the importance of features for higher-order data, which may assist in the reduction of dimensions and enhancement of performance. Latent features and temporal trends that may provide insight into human behavior can facilitate the development of effective interventions.

**Contributions of this work**

In this paper, we propose the following contributions:

- A novel tensor embedding algorithm for high dimension multi-way noisy data.
- A flexible framework for fusion of data with different number of modes.
- Feature selection using components’ activation is integrated with supervised regression.
- We validate recovered latent patterns via prediction performance of 29 different tasks on two real-world multimodal behavioral datasets including our recent “in the wild” experimental study that collected bio-behavioral data from subjects in challenging cognitive, social, and effective demands in their natural hospital workplace (and outside work) settings.

### Significance of our results

Extracting insights from human-generated behavioral data is computationally challenging, as data can be noisy, incomplete, and originate from multiple sources with different biases and sampling rates. Under these conditions, feature engineering and data fusion quickly become burdensome. To deal with these challenges, we propose a supervised tensor embedding method that extracts latent features from multivariate time series of bio-behavioral data. Validation on two longitudinal human-subject studies show that the method can robustly estimate various aspects of human behavior.

## II. RELATED WORK

Tensor regression/classification algorithms can be categorized into two general settings. The first group extent the conventional supervised methods and directly fit a model on N-way data. For example, supervised tensor learning (STL) extends support vector machines (SVM) and minimax probability machines (MPM) to N-way data for classification problems [39]. Another body of work runs linear regression on tensor data $X$, proposing a variety of solutions for efficient estimation of parameters of the model [18], [21], [48], [50]. These methods have been mostly applied on image data, [18], [48].

The second group of methods perform simultaneous decomposition of independent variables and prediction of dependent variable. One of the widely used methods for higher-order regression problems is N-way Partial Least Square (NPLS) [3], a natural multi-way extension of Partial Least Square (PLS) [15], which finds the latent features with a joint canonical/parallel (CP) decomposition [20] of input and target variables [3]. Some methods [24], [51] perform joint decomposition of higher-order independent variable and higher-order dependent variable, where the dependent variable can be matrix or higher-order tensor.

In this paper, we are interested in second group of methods, because of the interpretability of extracted components over the user, time and feature dimensions. We propose a supervised decomposition, which finds the latent factors with joint CP decomposition of independent tensor and dependent target variable. While NPLS simultaneously decompose independent variable and fit a linear model for estimation of dependent variable, we are interested in supervised extraction of a set of rank-1 latent components, highly correlated with target variable, which then can be combined with any regression/classification function for subsequent estimation task. In this setting, latent components obtained from N-way data can be merged with any other features in 2-dimensional space, e.g., latent features extracted from temporal daily behavior can be joined with age, or gender of the participants in a longitudinal study.
A. Notation and Definitions

A convenient mathematical representation of multimodal data is a tensor (also known as multi-way/N-way array), $\mathcal{X} \in \mathbb{R}^{I_1 \times \cdots \times I_N}$, where $N$ is the order, or the number of ways/modes of the tensor. Table I summarizes the notation used throughout this paper. The d-mode vector product of a tensor $\mathcal{X}$ with vector $y$ is defined as: $\mathcal{Z} = \mathcal{X} \times_d y$, where $z_{i_1 i_2 \ldots i_d \ldots i_N} = \sum_{i_d=1}^{I_d} x_{i_1 i_2 \ldots i_d \ldots i_N} y_{i_d}$. The rank-$R$ CP model, will decompose the tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times \cdots \times I_N}$ into sum of rank-1 tensors, called components:

$$\mathcal{X} = \sum_{r=1}^{R} \mathcal{A}^{(1)} \circ \mathcal{A}^{(2)} \circ \ldots \circ \mathcal{A}^{(N)} = \sum_{r=1}^{R} \mathcal{A}^{(1)} \cdot \mathcal{A}^{(2)} \cdot \ldots \cdot \mathcal{A}^{(N)} = \sum_{r=1}^{R} [\mathcal{A}^{(1)}, \mathcal{A}^{(2)}, \ldots, \mathcal{A}^{(N)}] \cdot r,$$

where the outer product $\mathcal{A}^{(1)} \circ \mathcal{A}^{(2)} \circ \ldots \circ \mathcal{A}^{(N)}$ corresponds to the $r^{th}$ component of rank-$R$ estimation, and $\mathcal{A}^{(n)}$'s, $n = 1, \ldots, N$ are factor matrices. Given a tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ and vector $y$, the covariance matrix $C \in \mathbb{R}^{I_1 \times I_2}$ is defined as:

$$C = \mathcal{X}^\top y = \mathcal{X}^\top \mathcal{X},$$

where $c_{i_1 i_2} = \sum_{i_3=1}^{I_3} x_{i_1 i_2 i_3} y_{i_3}$.

B. Supervised Tensor Embedding

We are interested in finding the latent user factors of the multivariate time series in longitudinal data, such that they can be of good predictive ability of the target variables of interest. Decomposition of the collected data in an unsupervised way can help to find underlying structure, however these latent user factors may not necessarily have high correlation with all different human behavior aspects and may only explain a subset. By applying supervised decomposition for each target variable, we would like to find the latent factors which correlate with it the most. Then we can use any regression function $g(.)$ on the obtained user latent matrix $U$ to estimate $\mathbf{y}$.

Our work builds upon the idea of NPLS regression, where the algorithm constructs a model of both $\mathcal{X}$ and $\mathbf{y}$ for each component simultaneously. We are interested in extracting rank-1 models from $\mathcal{X}$, and finding the latent factors highly correlated with $\mathbf{y}$, without fitting a prediction model at the time.

We will present a multivariate time series as tensor $\mathcal{X} \in \mathbb{R}^{M \times I \times J}$ for $M$ individuals, with $I$ time units and $J$ features, and a dependent vector $\mathbf{y} \in \mathbb{R}^{M}$. Entry $x_{mij}$ of this tensor...
corresponds to the value of \( j \)th feature of \( n \)th individual at the \( t \)th time unit. The goal is finding latent feature matrices of \( \mathcal{X} \approx [U, T, F] \) in a joint decomposition with \( y \). For a rank-1 model, it means finding \( u, t \) and \( f \), when minimizing \( \sum_m \sum_i \sum_j (x_{mij} - u_m t_i f_j)^2 \) such that \( u \) has maximum correlation with \( y \), or maximizing \( \sum u_m y_m \).

For this purpose we start with building the cross-covariance matrix \( C = \langle X', y \rangle \), where \( c_{ij} = \sum_m x_{mij} y_m \), (line 7, algorithm 1). The least square solution of \( \sum_m \sum_i \sum_j (x_{mij} - u_m t_i f_j)^2 \) is \( u_m = \frac{\sum_i \sum_j x_{mij} t_i f_j}{t_i r_i f_j} \). Then \( \sum_m u_m y_m = \sum_m \sum_i \sum_j (x_{mij} t_i f_j) \), which is equal to:

\[
\max_{t, f} \sum_{i=1}^{J} \sum_{j=1}^{I} c_{ij} t_i f_j = \max_{t, f} ||t^T C f||^2
\]

The answer for this problem is the first set of normalized vectors from a singular value decomposition on \( C \):

\[
t, f = \text{SVD}(C),
\]

When \( t \) and \( f \) are extracted, we can find \( u = \mathcal{X}' \times t \times f \). From the three latent factors, we can reconstruct the rank-1 model \( \mathcal{X}_{\text{model}} \). This process will be repeated \( R \) times, when at each iteration a rank-1 component will be extracted and \( \mathcal{X}_{rem} \) will be updated \( \mathcal{X}_{rem} = \mathcal{X} - \mathcal{X}_{\text{model}} \), cf. Fig. 1(left).

**Algorithm 1** Supervised Tensor Embedding

1: input: \( \mathcal{X} \) independent tensor, \( y \) dependent variable
2: parameters: \( R \) - decomposition rank
3: output: \( U, T, F \)
4: Center \( \mathcal{X} \) and \( y \)
5: \( \mathcal{X}_{rem} = \mathcal{X} \)
6: for \( r = 1:R \) do
7: \( C = \mathcal{X}_{rem} \times t \times f \)
8: \( t, f \leftarrow \text{SVD}(C) \)
9: \( t_r \leftarrow t_r / ||t_r||, f_r \leftarrow f_r / ||f_r|| \)
10: \( u_r \leftarrow \mathcal{X}_{rem} \times t_r \times f_r \)
11: \( \mathcal{X}_{model} \leftarrow u_r \circ t_r \circ f_r \)
12: \( \mathcal{X}_{rem} = \mathcal{X}_{rem} - \mathcal{X}_{model} \)

1) **Feature Selection**: Given different target variables for the same input, not all the features are equally informative for different targets. Therefore, a proper feature selection can improve prediction performance. Although the STE model down-weights the irrelevant features, it does not discard them. As a result, it is possible that a large number of irrelevant features can still contaminate the predictions. Here, we use feature activation in the feature latent components and drop the uninformative features.

The feature activation vector defines the importance of each feature for the specific prediction task. We use the latent feature factor matrix \( F \) and extract the feature importance as

\[
FI_j = \sum_{r=1}^{R} f_{rj}^2, \quad j = 1, 2, ..., J
\]

Given the feature importance score, any desired technique can be applied for feature selection. For presented results in this paper, we have chosen filter methods and we pick top \( K \) features and discard the others.

2) **Regression model**: In previous section we introduced our supervised decomposition (STE) algorithm to obtain the user latent factors with high correlation with target variable of interest and then defined feature importance to discard irrelevant features. The latent factors can be used for exploration of active features, temporal trends and similar users given a certain target variable. Furthermore, we can apply any regression function \( g(.) \) for inference of dependent variable. To evaluate the effectiveness of supervised embedding and feature selection in higher-order data, we perform a regression task, using features extracted via STE. The parameter of our model would be number of features to keep \( K \), number of components \( R \), and parameters of the regression function \( g(.) \) cf. Algorithm 2.

**Algorithm 2** Supervised Tensor Regression (STR)

1: input: independent variable \( \mathcal{X} \), dependent variable \( y \)
2: parameters: number of features \( K \), number of factors \( R \), regression model \( g(.) \)
3: output: \( y \)
4: Compute feature importance score for each feature using STE
5: Form a reduced data tensor \( \mathcal{X}_{\text{reduced}} \) consisting of only time series whose feature importance score is among top \( K \)
6: Compute latent factors using STE
7: Use user latent matrix \( U \) in a regression model to predict outcome \( \hat{y} = g(U) \)
8: Pick \( R, K \) and parameters of model \( g(.) \) by cross-validation

**IV. RESULTS**

Our goal is to understand whether supervised decomposition can find the low-dimensional structure of daily life from wearable devices that better correlates with target behavioral constructs. We test our models along with CP plus a regression function, NPLS and PLS on two real-world datasets described below. STR, CP, and NPLS were used to model the data in a tensor form, where PLS was used on a mode-1 matricized version of the same tensor. To compare the predictability, we compare \( R^2 \) and Pearson correlation \( \rho \) obtained from each method. We use the general definition of

\[
R^2 = 1 - \frac{\sum (\hat{y}_i - y_i)^2}{\sum (y_i - \bar{y})^2},
\]

which ranges from \(-\infty \) to 1. In this definition, \( R^2 = 1 \) would be equivalent to perfect prediction, \( R^2 = 0 \) is equivalent to the same performance of estimating \( \hat{y} = \bar{y} \), i.e., the average of target variable, and negative \( R^2 < 0 \) would mean worse performance than estimating the average. Different target variables have different ranges: we use these metrics instead of RMSE to be able to compare the performance among targets within the same model too. Because we have a small number of samples, we perform a nested cross-validation, when in the
train-validation set we tune the parameter of the model and then we present the result on test set (10-fold in the inner loop and Leave-2-out in the outer loop). We have repeated each experiment 20 times and have reported mean value for $R^2$ and $\rho$. In the bar graphs standard deviation is also reported. For STE method, we test different regression functions, with and without feature selection and we present the best result. Later in Fig. 4 and Tables VI and VII, we investigate the effect of feature selection, choice of regression function, and coupling with static data. In both datasets, we have stream of data with different sampling rate, e.g., 1-min, 5-min, per day. We have create tensors with 1-day resolution and have aggregation over 6 hours time-bins. More details is provided for each dataset.

A. StudentLife Data

StudentLife is a 10-week study conducted during 2013 spring semester on 48 Dartmouth students (30 undergraduate and 18 graduate students), [43]. Psychometric data were collected from students via a pre-assessment and post-assessment Survey. GPA was also collected at the end of the semester, which will be used as a measure of students academic performance. The other surveys include Big Five Inventory (BFI) [37], Positive Affect and Negative Affect Schedule (PANAS) [46], Perceived Stress scale [8], UCLA loneliness scale [33], Depression (PHQ9) [27], Perceived Stress [7], flourishing scale [9], VR-12 as a measure of students’ health [25], and Pittsburgh Sleep Quality Index (PSQI) [4], Table II.

Using the raw sensor data collected from students, physical activity (stationary, walk, run and unknown) and audio activity (conversation, silence, audio, noise and unknown) have been inferred and made available publicly [43]. To create our tensor, each time unit comprises one day worth of data, and is divided into four time-bins, bedtime (midnight-6 am), morning (6 am-12 pm), afternoon (12 pm-6 pm), and evening (6 pm-midnight). We extract duration (minutes) of running, walking, stationary, silence, voice, noise, and dark, per time-bin in each day. Frequency and number of changes in each behavior (e.g. from walking to running) for each time-bin has been also captured. From GPS and WiFi, the number of unique locations visited, and from Bluetooth, the number of unique nearby devices per time-bin are added to the variable set. We normalize all the variables across time dimension per user to have the same range to avoid variables with large values (e.g. duration in minutes) dominate the analysis. At the end, we organize our data as tensor $X$ with $M = 46$ individuals, $I = 108$ features and $J = 63$ days. Missing values are imputed by filling them with the mean value of the time series. Numbers of samples for different targets varies from 30 to 46, as not every participant had answered all the surveys. A summary of time series available from StudentLife dataset is provided in Table III.

### Table II: StudentLife dataset dependent variables

| Dimension | GPA |
|-----------|-----|
| Performance | Personality |
|   | Neuroticism | Conscientiousness | Extraversion | Agreeableness | Openness |
| Affect | Positive Affect (PANAS) | Negative Affect (PANAS) | Stress | Depression (PHQ9) | Loneliness |
| Health | Sleep quality (PSQI) | Health (VR-12) | Flourishing |

Fig. 2: StudentLife – Prediction performance (top: $R^2$, bottom: Pearson correlation $\rho$) using STR, NPLS, PLS, CP methods.
Fig. 2 presents the results for StudentLife dataset (in both plots the y-axis is limited to -0.2). For some of the target variables, e.g. negative effect, openness and sleep quality, $R^2$ has improved significantly. Also it is worth mentioning that we are using only passively collected data for all prediction tasks, without using any EMA or self-report values by the participants as the features.

B. WorkPlace Data

WorkPlace\footnote{This is a fictitious, anonymized project name for review purpose} is an ongoing research study of workplace performance which measures physical activity and physiological state of employees (such as nurses) in a large university hospital setting. For this research, over 50 full-time employees of a large urban hospital were volunteered. Participants were 27\% (n = 14) male and 73\% (n = 47) female and ranged in age from 25 years to 65 years.

The dataset includes passively collected sensor values, psychometrics and job performance measures. Sensory data were collected from garment-based wearable sensors (OMsignal) and wristbands (Fitbit). OMsignal is a Biometric Smartwear company that produces smart under shirts and bras.

Wearable sensors have been used passively in clinical applications to serve patients who need a long-term personal care with continuous monitoring. The establishment of mobile health has extended the usage of wearable sensors into daily life that benefits individuals and society as a whole in health promotion.

The OMsignal garments include health sensors embedded into the fabric that measure biometric data in real-time and can relay this information to the participant’s smartphone. OMsignal sensor provides information such as heart rate, heart rate variability (HRV), breathing, and accelerometry (to provide sitting position, foot movement and more). Fitbit collects heart rate, steps, sleep and cardio information. Participants were asked to wear their Fitbit 24/7. However, they were instructed to wear OMsignal sensors only during their work shifts. It is worth mentioning that clinical staff in this study work a minimum of 3 days per week (in 12 hour shifts), which can be any day during weekdays or weekend. Also some belong to day shift and others to night-shift, which would be 7am-7pm, 12pm-6pm, and evening (6pm-12am), respectively.

Psychometric data were collected from participants via pre- and post study surveys. These surveys measured job performance, cognitive ability, personality, affect, and health state and are used as our ground truth in the model. More specifically, the target variables we predict in this dataset include the In-Role Behavior Scale [47], Individual Task Proficiency Scale [17], Shipley abstraction and vocabulary (cognitive abilities) [36], Big Five Inventory (BFI), Positive Affect and Negative Affect Schedule (PANAS), State-Trait Anxiety Inventory (STAI) [38], Alcohol use Disorders Identification Test (AUDIT) [35], International Physical Activity Questionnaire (IPAQ) [30], and Pittsburgh Sleep Quality Index (PSQI), TableV.

For the experiments reported in this paper, we use the pre-survey scores provided by the participants as the target variables for WorkPlace dataset. Similar to the previous dataset each time unit comprises one day’s worth of data. Fitbit sleep and cardio summary have 1-day resolution. For other variables with 5-min resolution, each day is divided into four time bins: bedtime (midnight-6 am), morning (6 am-12 pm), afternoon (12 pm-6 pm), and evening (6 pm-midnight). We extract a set of statistics such as mean, standard deviation, kurtosis, etc. from each time series with continues values, in each time-bin. For categorical features we calculated frequency of each category. We organize the data as tensor $X$ with $M = 51$ individuals, $I = 1225$ features and $J = 30$ days. Each time series is normalized to have mean zero and standard deviation equal to one. About 60\% of the sensor is missing, which we imputed by filling them with average of time series. We tested our model for 15 different target variables in comparison with PLS, NPLS and CP. Similar to StudentLife dataset, we divided the data into train, validation and test set, and performed nested cross validation and report the result on test set, cf. Fig. 3 (in both plots the y-axis is limited to -0.2).

C. Analysis

Looking at 29 different predictions across two datasets, STR has higher $R^2$ in 21 tasks than the other three methods. Some of the constructs were not predictable with any of approaches which can be due to the lack of appropriate features, or inadequate feature engineering. Also it is possible tensor tri-linear models are not suitable for modeling those constructs, as by applying the no-free-lunch idea to all sorts of scientific problems it has been shown that different type of algorithms may work well for different type of problems, [32]. The improvement in performance by STR comes from 1) supervised embedding, 2) choice of regression function or 3) feature selection. In order to understand the effect of supervised embedding, we use a simple ordinary least square
model as regression function of STR with no feature selection on the time series and compare it with NPLS. For 10 target variables out of 29 total, both $R^2$ and $\rho$ improved. As another contribution, by separating the embedding and regression steps in STR model, the latent features can be tested and paired with the most appropriate regression function to improve the performance. For example, for negative affect from StudentLife dataset and shipley abstraction from WorkPlace, we have applied Ridge and linear SVR (Support Vector Machine-Regression) as two different regression models, which we can see SVR will lead to 4% improvement in both $R^2$ and $\rho$, cf. Table VI.

Parameter selection effect has been demonstrated in Fig. 4, presenting $R^2$ and $\rho$ versus number of components and number of selected features. The best result for flourishing from StudentLife dataset is obtained at $K = 70$ and $R = 6$. Physical activity score (IPAQ) prediction reaches its best performance at around $K = 500$ and $R = 5$.

Compared to multi-way regression methods, another benefit of supervised embedding is the flexibility to add relevant static features, such as demographic, cf. Fig. 1. In the WorkPlace study, we use participant’s age, gender, commute time to work, and job related data such as nurse vs. non-nurse, shift, supervisory role, income, extra hours, having other jobs. In StudentLife, we don’t have access to such data. Next, we present results for prediction performance of two constructs from WorkPlace. First, we only use temporal data and extract latent features. In another experiment, we add the above mentioned metadata to the feature list, cf. Table VII.

One interesting property of using tensor methods for the extraction of latent features is interpretability for all modalities, e.g., activated sensors in the feature space, and temporal group behavior in the time mode. In this section, we look at the temporal patterns obtained from two different supervised

---

**Table V: WorkPlace dataset dependent variables.**

| Dimension       | Construct                        |
|-----------------|----------------------------------|
| Job Performance | In-Role Behavior                 |
|                 | Individual Task Proficiency      |
| Cognitive Ability| Shipley Abstraction              |
|                 | Shipley Vocabulary               |
| Personality     | Neuroticism                      |
|                 | Conscientiousness                |
|                 | Extraversion                     |
|                 | Agreeableness                    |
|                 | Openness                         |
| Affect          | Positive Affect                  |
|                 | Negative Affect                  |
|                 | Anxiety                          |
| Health          | Alcohol usage                    |
|                 | Physical activity                |
|                 | Sleep quality                    |

---

**Table VI: Using different regression models is present by two examples; StudentLife: negative affect, WorkPlace: shipley abstraction**

|               | Negative affect | Shipley abstraction |
|---------------|-----------------|--------------------|
|               | $R^2$ | $\rho$ | $R^2$ | $\rho$ |
| SVR           | 0.31  | 0.57  | 0.19  | 0.44  |
| Ridge         | 0.27  | 0.33  | 0.15  | 0.39  |

---

**Table VII: Coupling latent features with metadata**

|               | Alcohol usage | Agreeableness |
|---------------|---------------|---------------|
|               | $R^2$ | $\rho$ | $R^2$ | $\rho$ |
| Metadata      | 0.04  | 0.19  | 0.09  | 0.32  |
| Latent features| 0.07  | 0.27  | 0.09  | 0.33  |
| Latent features+metadata | 0.15  | 0.39  | 0.11  | 0.37  |
decomposition tasks from StudentLife. This dataset has high-
level features, e.g. conversation, activity, and information of
temporal events over the course of semester. One interesting
observation is that GPA’s first latent temporal factor increases
towards end of the semester and depression’s first latent
temporal factor start decreasing around after mid-semester.
The duration of being physically stationary in the afternoon,
evening and midnight, the duration of audio silence, and
the number of detected on-campus wifi locations are among
the top features of the first latent factor of GPA. This
inferred latent component can be an indicator of studying
in a quiet environment (the studying factor). To obtain a
better understanding of the correlation between depression and
performance, we also looked at depression latent factor. The
top activated features include running, walking during evening,
duration of conversation in the morning, and the number of on-
campus wifi locations. This latent behavior has features related
to school engagement activity. We name it as ”diminished
interest in activities”, as the temporal trend decreases over the
second half of the semester, Fig. 5, red graph. For the
students that have higher value in this latent factor, there
was a higher chance of depression. The user latent factor of ”diminished interest in activities” has a correlation of -0.9
with the studying user latent factor. It can mean that students
who grow depressive symptoms over the semester have lower
performance at the end of the semester. Recently in another
study [44], it has been observed that depression has negative
correlation with the slope of the duration of time students spent
in study places during the semester on-campus.

V. CONCLUSIONS AND FUTURE WORK

Rich multimodal data collected from wearable sensors (e.g.
Fitbit), mobile phones, online social networks, etc is be-
coming increasingly available to reconstruct digital trails and
study human behavior. In this paper, we use two real-world
datasets–WorkPlace and StudentLife–collected using passive
and mobile sensors, with the goal of inferring well being,
performance, and personality traits. We developed a learning
framework based on supervised tensor embedding to find latent
features that are highly correlated with target variables of
interest. This type of decomposition can uncover latent user
factors which are strong predictors of target variables. In
addition, in our setting static features can be coupled with
extracted user latent matrix. One limitation of our work is
that the framework captures only linear structure. Another
limitation is using prediction performance as a metric for
selection of best rank \( R \) and \( K \) (number of top features).
We plan to use kernel methods for nonlinear projection and
defining information theoretic metrics for best embedding.
Feature selection can be extended to be applied on latent
features too. Also, as WorkPlace study is an ongoing project,
we plan to implement supervised predictions of individuals’
performance and personality directly from different modalities,
such as social media activity, location, audio.

ACKNOWLEDGMENT

The authors are grateful to the TILES team for the efforts
in study design, data collection and sharing, that enable
this work. This research is based upon work supported by
the Office of the Director of National Intelligence (ODNI),
Intelligence Advanced Research Projects Activity (IARPA),
via IARPA Contract No 2017- 17042800005 and by the
Defense Advanced Research Projects Agency (DARPA grant
#D16AP00115). The views and conclusions contained herein
are those of the authors and should not be interpreted as
necessarily representing the official policies or endorsements,
either expressed or implied, of the ODNI, IARPA, DARPA or
the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon.

REFERENCES

[1] Evrim Acar, Tamara G Kolda, and Daniel M Dunlavy. All-at-once optimization for coupled matrix and tensor factorizations. arXiv preprint arXiv:1105.3422, 2011.

[2] Nadav Aharony, Wei Pan, Cory Ip, Inas Khayal, and Alex Pentland. Social fMRI: Investigating and shaping social mechanisms in the real world. Pervasive and Mobile Computing, 7(6):643–659, 2011.

[3] Rasmus Bro. Multisway calibration, multilinear pls. Analytica chimica acta, 185:1–17, 1996.

[4] D. Buysse, C. Reynolds, T. Monk, S. Berman, and D. Kupfer. The pittsburgh sleep quality index: A new instrument for psychiatric practice and research. Psychiatry Research, 28(2):193–213, 1989.

[5] J Douglas Carroll and Jih-Jie Chang. Analysis of individual differences in multidimensional scaling via an n-way generalization of “eckart-young” decomposition. Psychometrika, 35(3):283–319, 1970.

[6] Tanzeem Choudhury, Sunny Consolvo, Beverly Harrison, Jeffrey High- tower, Anthony LaMarca, Louis LeGrand, Ali Rahimi, Adam Rea, G Bondello, Bruce Hemingway, et al. The mobile sensing platform: An embedded activity recognition system. IEEE Pervasive Computing, 7(2), 2008.

[7] Sheldon Cohen, T Kamarck, R Mermelstein, et al. Perceived stress scale. Measuring stress: A guide for health and social scientists, pages 235–238, 1983.

[8] Sheldon Cohen, Tom Kamarck, and Robin Mermelstein. A global measure of perceived stress. J. of health and social behavior, pages 385–396, 1983.

[9] Ed Diener, Derrick Wirtz, William Tov, Chu Kim-Prieto, Dong-won Choi, Shigehiro Oishi, and Robert Biswas-Diener. New well-being measures: Short scales to assess flourishing and positive and negative feelings. Social Indicators Research, 97(2):143–156, 2010.

[10] Nathan Eagle and Alex Pentland. Reality mining: Sensing complex social systems. Personal and Ubiquitous Computing, 10(4):255–268, May 2006.

[11] Andrey Eliseyev and Tetiana Aksenova. Recursive n-way partial least squares for brain-computer interface. PloS one, 8(7):e69962, 2013.

[12] Andrey Eliseyev, Vincent Auboroux, Thomas Costecalde, Lilía Langar, Guillaume Charvet, Corinne Mestais, Tetiana Aksenova, and Alim-Louis Benabd. Recursive exponentially weighted n-way partial least squares regression with recursive-validation of hyper-parameters in brain-computer interface applications. Scientific reports, 7(1):16281, 2017.

[13] Andrey Eliseyev, Cecile Moro, Jean Faber, Alexander Wyss, Napoleon Torres, Corinne Mestais, Alim Louis Benabd, and Tetiana Aksenova. L1-penalized n-way pls for subset of electrodes selection in bei experiments. J. of neural engineering, 9(4):045010, 2012.

[14] Stefania Favilla, Caterina Durante, Mario Li Vigni, and Marina Cocchi. Assessing feature relevance in npls models by vip. Chemometrics and Intelligent Laboratory Systems, 129:76–86, 2013.

[15] Paul Geladi and Bruce Kowalski. Partial least-squares regression: A tutorial. Analytica chimica acta, 185:1–17, 1986.

[16] Asma Ghandhariehroui, Szymon Fedor, Lisa Sangermano, Dawn Ionescu, Jonathan Alpert, Chelsea Dale, David Sontag, and Rosalind Picard. Objective assessment of depressive symptoms with machine learning and wearable sensors data. In Proc. Int. Conf. on Affective Computing and Intelligent Interaction (ACII), San Antonio, Texas, 2017.

[17] M. Griffin, A. Neil, and S. Parker. A new model of work role performance: positive behavior in uncertain and interdependent contexts. Academy of Management Journal, 50(2):327–347, 2007.

[18] Weiwei Guo, Irene Kotis, and Ioannis Patras. Tensor learning for regression. IEEE Trans. on Image Processing, 21(2):816–827, 2012.

[19] Martin Haardt, Florian Roemer, and Giovanni Del Galdo. Higher-order svd-based subspace estimation to improve the parameter estimation accuracy in multidimensional harmonic retrieval problems. IEEE Transactions on Signal Processing, 56(7):3198–3213, 2008.

[20] Richard A Harshman. Foundations of the parafac procedure: Models and conditions for an “explanatory” multimodal factor analysis. 1970.

[21] Jarvis Haupt, Xingguo Li, and David P Woodruff. Near optimal sketching of low-rank tensor regression. arXiv preprint arXiv:1709.07093, 2017.

[22] Joyce C Ho, Joydeep Ghosh, and Jimeng Sun. Marble: high-throughput phenome-wide readout via electronic sensor mapping and nonnegative tensor factorization. In Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining, pages 115–124. ACM, 2014.

[23] Peter D Hoff. Multilinear tensor regression for longitudinal relational data. The annals of applied statistics, 9(3):1169, 2015.

[24] Ming Hou, Qibin Zhao, Brahim Chaib-draa, and Andrez Cichocki. Common and discriminative subspace kernel-based multiblock tensor partial least squares regression. 2016.

[25] SU Iqbal, A Selim W Rogers, S Qian, XS Ren A Lee, J Rothendier, D Guo, and LE Kazis. The veterans rand 12 item health survey (vor12): What is it and how it is used. Center for Health Quality, Outcomes, and Economic Research, A Health Services Research and Development Center of Excellence, VA Medical Center, Bedford, MA, USA, 2009.

[26] Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross Girshick, Sergio Guadarrama, and Trevor Darrell. Caffe: Convolutional architecture for fast feature embedding. In Proceedings of the 22nd ACM int. conf. on Multimedia, pages 675–678. ACM, 2014.

[27] Kurt Kroenke and Robert L Spitzer. The phq-9: a new depression diagnostic and severity measure. Psychiatric annals, 32(9):509–515, 1996.

[28] Max Kuhn and Kjell Johnson. Applied predictive modeling, volume 26. Springer, 2013.

[29] Quoc V Le, Will Y Zou, Serena Y Yeung, and Andrew Y Ng. Learning hierarchical invariant spatio-temporal features for action recognition with independent subspace analysis. In Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conf. on, pages 3361–3368. IEEE, 2011.

[30] R. Maddison, C. Ni Mhurchu, Y. Jiang, S. Vander Hoorn, A. Rodgers, C. Lawless, and E. Rush. International physical activity questionnaire (ipaq) and new zealand physical activity questionnaire (nzpaq): A doubly labelled water validation. Int. Journal of Behavioral Nutrition and Physical Activity, 4(6):2, 2007.

[31] Emiliano Miluzzo, Nicholas D Lane, Kristofor Fodor, Ronald Peterson, Hong Lu, Mirco Musolesi, Shane B Eiseman, Xiao Zheng, and Andrew T Campbell. Sensing meets mobile social networks: the design, implementation and evaluation of the cenceme application. In Proc. of the 6th ACM conf. on Embedded network sensor systems, pages 357–358. ACM, 2008.

[32] Marco A Muñoz, Laura Villanova, Davaasteren Baatar, and Kate Smith-Miles. Instance spaces for machine learning classification. Machine Learning, pages 1–39, 2017.

[33] Daniel W Russell. Ucla loneliness scale (version 3): Reliability, validity, and factor structure. Journal of personality assessment, 66(1):20–40, 1996.

[34] Akane Sano, Andrew J Phillips, Amy Z. Yu, Andrew W. McHill, Sara Taylor, Natasha Jaques, Charles A. Czeisler, Elizabeth B. Klerman, and Rosalind W. Picard. Recognizing academic performance, sleep quality, stress level, and mental health using personality traits, wearable sensors and mobile phones. 2015 Int. Conf. on Wearable Implantable Body Sensor Networks, pages 1–13, June 2015.

[35] J. B. Saunders, O. G. Asland, T. F. Babor, J. R. De la Fuente, and M. Grant. Development of the alcohol use disorders identification test (audit): Who collaborative project on early detection of persons with harmful alcohol consumption. Addiction, 89(6), 1993.

[36] W. C. Shipley, C. P. Gruber, T. A. Martin, and A. M. Klein. Shipley-2 Manual, Western Psychological Service, Los Angeles, CA, 2009.

[37] C. Soto and O. John. The next big five inventory (bfi-2): Developing a tutorial. Analytica chimica acta, 185:1–17, 1986.

[38] K. Taylor, Natasha Jaques, Charles A. Czeisler, Elizabeth B. Klerman, and Rosalind W Picard. Recognizing academic performance, sleep quality, stress level, and mental health using personality traits, wearable sensors and mobile phones. 2015 Int. Conf. on Wearable Implantable Sensor Networks, pages 1–13, June 2015.

[39] C. D. Spielberger, G. A. Jacobs, S. Russell, and R. S. Crane. Assessment of anger: The state-trait anger scale. In Advances in Personality Assessment, Erlbaum, Hillsdale, New Jersey, 1983.

[40] Dacheng Tao, Xuelong Li, Weiming Hu, Stephen Maybank, and Xiaodong Wu. Supervised tensor learning. IEEE Conf. on Data Mining (ICDM), pages 8–pp. IEEE, 2005.

[41] Dacheng Tao, Xuelong Li, Xiaodong Wu, Weiming Hu, and Stephen J. Maybank. Supervised tensor learning. Knowl., Inf. Syst., 13(1):1–42, September 2007.

[42] Nazgol Tavabi, Homa HosseiniMardi, Jennifer L Villatte, Andrés Abelínuk, Shrinankh Narayanam, Emilio Ferrara, and Kristina Lerman. Learning behavioral representations from wearable sensors. In Social, author.
[42] Daixin Wang, Peng Cui, and Wenwu Zhu. Structural deep network embedding. In Proceedings of the 22Nd ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining, KDD ’16, pages 1225–1234, New York, NY, USA, 2016. ACM.

[43] Rui Wang, Fanglin Chen, Zhenyu Chen, Tianxing Li, Gabriella Harari, Stefanie Tignor, Xia Zhou, Dror Ben-Zeev, and Andrew T. Campbell. Studentlife: Assessing mental health, academic performance and behavioral trends of college students using smartphones. UbiComp ’14 Proc. of the 2014 ACM Int. Joint Conf. on Pervasive and Ubiquitous Computing, pages 3–14, September 2014.

[44] Rui Wang, Weichen Wang, Alex daSilva, Jeremy F. Huckins, William M. Kelley, Todd F. Heatherton, and Andrew T. Campbell. Tracking depression dynamics in college students using mobile phone and wearable sensing. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol., 2(1):43:1–43:26, March 2018.

[45] Yichen Wang, Robert Chen, Joydeep Ghosh, Joshua C Denny, Abel Kho, You Chen, Bradley A Malin, and Juneng Sun. Rubik: Knowledge guided tensor factorization and completion for health data analytics. In Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pages 1265–1274. ACM, 2015.

[46] D. Watson and L. A. Clark. The PANAS-X: Manual for the Positive and Negative Affect Schedule - Expanded Form. University of Iowa, 1999.

[47] L. J. Williams and S. E. Anderson. Job satisfaction and organizational commitment as predictors of organizational citizenship and in-role behaviors. J. of Management, 17(3):601–617, 1991.

[48] Fei Wu, Xu Tan, Yi Yang, Dacheng Tao, Sihang Tang, and Yueting Zhuang. Supervised nonnegative tensor factorization with maximum-margin constraint. 2013.

[49] Yinchong Yang, Denis Krompass, and Volker Tresp. Tensor-train recurrent neural networks for video classification. arXiv preprint arXiv:1707.01786, 2017.

[50] Rose Yu and Yan Liu. Learning from multway data: Simple and efficient tensor regression. In Int. Conf. on Machine Learning, pages 373–381, 2016.

[51] Qibin Zhao, Cesar F Caiafa, Danilo P Mandic, Zenas C Chao, Yasuo Nagasaka, Naotaka Fujii, Liqing Zhang, and Andrzej Cichocki. Higher order partial least squares (hopls): a generalized multilinear regression method. IEEE trans. on pattern analysis and machine intelligence, 35(7):1660–1673, 2013.

[52] Qibin Zhao, Guorui Zhou, Tulay Adali, Liqing Zhang, and Andrzej Cichocki. Kernelization of tensor-based models for multiway data analysis: Processing of multidimensional structured data. IEEE Signal Processing Magazine, 30(4):137–148, 2013.

[53] Hua Zhou, Lexin Li, and Hongtu Zhu. Tensor regression with applications in neuroimaging data analysis. J. of the American Statistical Association, 108(502):540–552, 2013.