Personalized Student Stress Prediction with Deep Multi-Task Network

Anonymous Authors[1]

Abstract

With the growing popularity of wearable devices, the ability to utilize physiological data collected from these devices to predict the wearer's mental state such as mood and stress suggests great clinical applications, yet such a task is extremely challenging. In this paper, we present a general platform for personalized predictive modeling of behavioural states like students' level of stress. Through the use of Auto-encoders and Multi-Task learning we extend the prediction of stress to both sequences of passive sensor data and high-level covariates. Our model outperforms the state-of-the-art in the prediction of stress level from mobile sensor data, obtaining a 45.6% improvement in F1 score on the StudentLife dataset.

1. Introduction

Today's competitive and demanding environment often overwhelms students with assignments, tests and part-time work. A prolonged exposure to stressful academic and social environment causes cardiovascular diseases (Rozanski et al., 1999; Kario et al., 2003), alterations of the brain causing differences in memory and cognition (SJ et al., 2009), suppression of the immune system (Khansari et al., 1990), and poor academic performance (Sano et al., 2015; Trokel et al., 2000). With the efforts of researchers at various institutions several technologies for detecting stress has been accomplished. Few use heart rate and heart rate variability (Vrijkotte et al., 2000), cortisol levels (Dickerson & Kemenyr, 2004), skin conductance (Setz et al.) and speech (Adams et al., 2014). Other techniques do not depend on sensors but simply try to discover the user's stress through self-reporting tools e.g., (Rahman et al., 2014) and surveys like the Perceived Stress Scale (Cohen et al., 1983).

With the induction of high quality, robust sensors in Wearables like FitBit, Apple Watch and smartphones, efficient collection of physiological and behavioural data with reasonable accuracy has become affordable. The StudentLife study in (Wang et al., 2014) collected Sleep Patterns, Activity, Conversation, Location, Information regarding Mental Health like stress levels and much more through the StudentLife application on android smartphones.

Contemporary research such as (Sano & Picard, 2013) and (Sano et al., 2015) has leveraged similar type of data from sensors and Machine Learning to predict stress levels of students. Furthermore, using the data collected from the StudentLife application, (Wang et al., 2018) have been successful in classifying students' as depressed or not in a binary classification problem. However, the task of predicting human psychological state (e.g., stress) using passive sensing data on a multi-class classification problem remains a challenge. Lack of gold standard labeled data, noisy raw sensor data, heterogeneity in granularity and inter subject variability in behavioural and environmental patterns have stymied predictive modeling of this kind.

(Mikelsons et al., 2018) have tried to predict stress of students in the StudentLife dataset by novel feature engineering of location based features and Neural Networks. To the best of our knowledge their model is the state-of-the-art on predictive modeling of stress on this dataset and we refer this model as Location Based MultiLayer Perceptrons (Location-MLP). However, it doesn't address the challenges of inter-subject variability or heterogeneity in granularity. The model is also limited to location and few covariates as features. In this paper we introduce the **Cross-personal Activity LSTM Multitask Auto-encoder Network (CALM-Net)** which considers data as time-series and is able to identify temporal patterns contained in student data. By including these different levels of information and personalizing the predictions to students, CALM-Net can achieve an average *F1-score* of **0.594** which is an improvement of **45.6%** when compared to the Location Based MLP. CALM-Net offers the flexibility to personalize models and the ability to incorporate time-series information, which in general, can be used by researchers to improve performance for categorical prediction of psychological states of humans.

[1]Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

Preliminary work. Under review by the International Conference on Machine Learning (ICML). Do not distribute.

2. Background Information

The StudentLife study was conducted in Dartmouth college where passive sensing and survey data was collected over 10 weeks among 48 students. The data collection, which mainly comprises of sleep patterns, activity, meal counts and conversations was facilitated by the StudentLife application on a smartphone. On a daily basis the StudentLife application collected stress data on a scale of 1-5 in the form of Ecological Momentary Assessments (EMA) which are responses to questionnaires in real time. Although EMAs are self-reported and consequently noisy, they are usually a good indicator of the actual stress state of the person making it feasible but not ideal to use them as labels for supervised learning tasks.

Out of the several challenges in the predictive modeling of stress, one major challenge that makes this task formidable is that the features have disparate granularity; they are also missing at random, which could be caused by technical issues such as a sensor failure or the phone being switched off. The dataset is heterogeneous in nature since the data is collected from a variety of sources like passive sensors, surveys and self reported EMAs. Out of the many discrete sequence and covariate features in StudentLife dataset, we select the ones that suggest evidence of these being good predictors of stress in (Stults-Kolehmainen & Sinha, 2014; Sano & Picard, 2013; Trokel et al., 2000; Sano et al., 2015) etc.

Among the discrete sequence data, we use Activity and Audio which are categorical integer values, recorded by the StudentLife app as follows *0- No Activity/Silence, 1- Walking/Voice, 2- Running/Noise, 3- Unknown*. Conversation, Phone Charge, Phone Lock are all inferred as binary values. These features are recorded at a variable rate, ranging from once every 10 seconds to once every minute.

Along with the above passive sensing data our model uses inferred and recorded covariates: Day of Week, Sleep Rating, Sleep duration (all recorded or inferred as integer values) and a binary covariate "Exam Period". We use time to next deadline as a feature which is inferred by the recorded deadlines in StudentLife. We believe that as a deadline approaches the stress levels of students must increase. The features, covariates, their respective value bounds and modes are listed in appendix Table 3.

3. Methods

3.1. Problem Setup

Passive sensors in smartphones allow collection of rich discrete time-series data crucial for stress prediction. The raw and elongated time-series features such as Activity and Audio can contribute to a few thousand data points everyday and cannot be used for training 'as is' due to infrequent label samples. To deal with this we first bin the whole time-series into 1-minute bins and take the mode of the categorical inferences. This offers a compact representation of what the subject was doing in that minute, for example *'was the wearer running or having a conversation?'*. This results in 1,440 sequences per day, which is still a very long sequence to be modeled using Recurrent Neural Networks. We further address this by computing the histogram of the features in 1-hour bins yielding 24 sequences per stress label. Intuitively, we are modeling how much conversation or activity a student has undergone in an hour which led to the stress label in a consistent manner, removing any bias due to irregular sampling of sensors by containing that information in a base bin of 1-minute. This type of feature engineering can easily be extended to different datasets with similar/same type of passive sensing time-series data as the final histogram is independent of the initial binning granularity of 1 min and can accommodate even higher-resolution data.

3.2. Models

3.2.1. Location Feature Based MLP

In the work done by (Mikelsons et al., 2018), a Multilayer Perceptron (MLP) with 4 fully connected layers was employed to perform stress inference. Each fully connected layer uses the *tanh* activation followed by a Batch Normalization and Dropout layer. The input to the model is feature engineered GPS data aggregated on a daily basis. There are a total of 8 location features and 4 covariates. The location features are total distance covered, max displacement, distance entropy on 10 minutes bins, distance standard deviation, number of unique tiles visited, difference in tiles visited from the previous day, approximate area of the GPS convex hull, and number of clusters on the GPS data. Tiles are non-overlapping, consecutively partitioned squares of 50 meters on the sides, where the combined squares represent the area the student may have traversed to. Each tile is uniquely labeled and counts on tiles is one of the features used. Throughout a day, the sequence of tiles were visited, and the difference in sequence of tiles covered compared to one on the previous were computed using Levenshtein edit-distance. Finally, the covariates used are indicators of whether the day is the start of term, mid-term, end of term or a weekend.

We followed the paper in building the baseline and also achieve an F-1 score of *0.42* with their experiment setup. To the best of our knowledge this is the-state-of-the-art on the StudentLife dataset. The features engineered with this baseline was done to the best of our abilities, although we note there may have been some discrepancy with the original work in obtaining certain features.

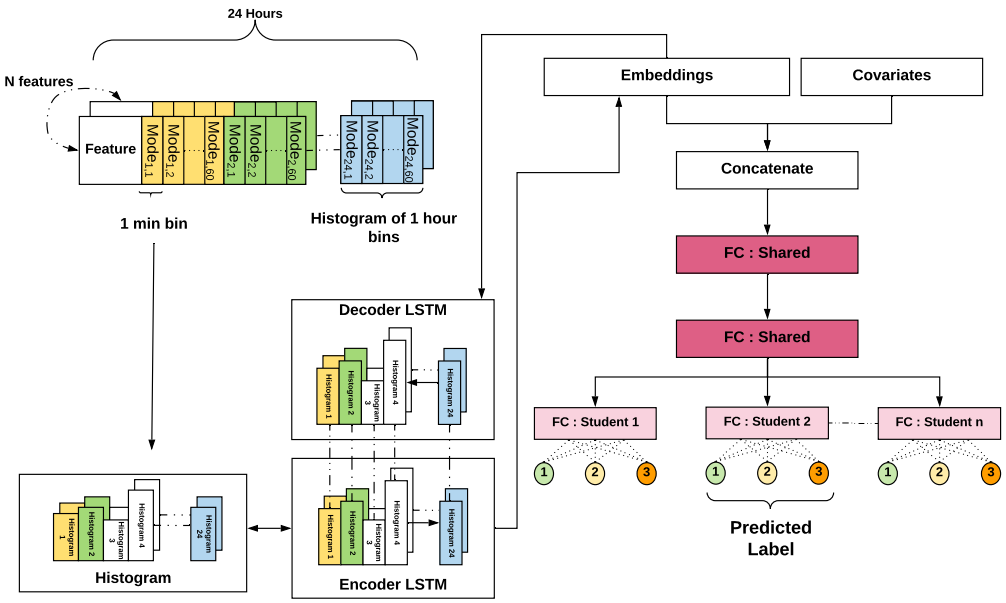

Figure 1. Cross-personal Activity LSTM Multitask Auto-encoder Network (CALM-Net).

3.2.2. LSTM

The state-of-the-art model is incapable of using time-serie data and only works on aggregates. This leads to the inability to use the information in granular passive sensing data which is ubiquitous in these kinds of datasets. To model the temporal patterns of features like Activity, Audio and Conversation we put the sequences of hourly histograms through an LSTM (Long Short-term Memory Network) (Hochreiter & Schmidhuber, 1997). Then we concatenate the last hidden state with the covariates. This concatenated output is passed through multiple layers of fully connected layers with the ReLU activation to finally obtain the class probabilities by using Weighted Categorical Cross-Entropy. This as one of our baselines and compare our final method to this. It is also a part of our next model.

3.2.3. LSTM Multi-Task Network (LM-Net)

Due to the heterogeneous nature of the dataset it is hard to incorporate information at different levels of granularity for a predictive task. Furthermore, trying to capture personal dynamics of all subjects using one model is demanding, as these dynamics are very distinct and have high inter-subject variability. Following the work done by (Jaques et al., 2017), to learn personalized models for each student, we use a Multi-Task approach which comprises of a LSTM to model sequence of histograms followed by shared fully connected layers with ReLU activation and a MLP for each student. As indicated by our experiments detailed in section 4 this approach can learn the differences between students and subsequently yield significant improvement in performance. It also gives evidence that learning a single model for all the students is unsuitable. Multi-Task learning also acts as a heavy regularizer, preventing the model to overfit for one student or the most common label. The shared layers learn common features, while the personal layers learn features that are relevant to the respective subject.

3.2.4. Cross-personal Activity LSTM Multitask Auto-encoder Network (CALM-Net)

Amongst the popular techniques for modeling time-series data, variations of RNNs like GRU and LSTM are the most popular, however people have used Auto-encoders for compression and reduction of the temporal dimensions. In (Lngkvist et al., 2014), different techniques for time-series have been summarised. Out of which we try RNN-LSTM and Auto-encoders. In CALM-Net we replace the LSTM layer with an LSTM Auto-encoder. Due to the low amount of training data available, we find it useful to reconstruct the sequence of histograms through an Encoder-Decoder pair. This also ensures that the model does not overfit to the the discrete training sequences. For calculating reconstruction error between the decoded sequences and the original sequences we used Mean Absolute Error(MAE). The final error/loss (expression in equation 1) is a weighted sum of the Reconstruction Error and Classification Error where α and β are hyperparameters.

4. Result

Due to a heavy imbalance of class labels on a scale of 1-5, we follow (Mikelsons et al., 2018), converting the five stress

Table 1. F1 scores of stress level prediction on StudentLife dataset.

Model	F1-score
Location-MLP	0.408
LSTM	0.426
Most Common Label	0.551
LM-Net	0.586
CALM-Net No covariates	0.571
CALM-Net	**0.594**

label scale to a scale of three stress labels by defining our classes as - *below median stress, median stress and above median stress*. We determined that some students present have high level of inter-subject variability which will make classification of stress extremely difficult for the current methods and selected the students who have greater than 40 labels and trained CALM-Net with a learning rate of 10^{-6} and a weight decay of 10^{-4}, the details of the data split and model configuration is given in appendix.

$$\boldsymbol{\alpha * RE + \beta * CE} \tag{1}$$

To evaluate our methods and make a fair comparison with baselines and previous state-of-the-art. We report the average F1-score achieved by each method on 5-fold cross validation. We compare against the state-of-the-art Location-MLP method with the data of a full day on which the label was reported, which potentially uses some data that is recorded after the stress label. Furthermore, we are comparing against LSTM, LSTM Multi-Task Network (LM-Net) which are explained in section 3.2. The achieved results are summarised in Table 1. As you can see, considering the data as time-series along with additional features available in the dataset improves the performance of the model as indicated by LSTM model improving upon the performance of the previous state-of-the-art model (Location-MLP). Since we have personalized models for every student, we also considered another baseline which is just predicting the most common label for the respective student as their stress state and outperforming it. From the results it is evident that personalizing the model to students can outperform state-of-the-art model (Location-MLP) and LSTM, showing the value of personalizing the methods to students. You can further see that CALM-Net with a *F1-score* of 0.594 outperforms the other models due to its ability to capture both temporal patterns and learning personalized information about the students. The detailed model configuration is given in appendix section A.2.

Moreover, since CALM-Net can learn personalized patterns for each student it can generalize better to more students. To test this hypothesis we designed an experiment where we try our model without Multi-Task heads and with Multi-Task heads on 5 ,13 and 23 students. The results of this experiment are summarised in Table 2. These results indicate that when we increase the number of students the performance of the model without Multi-Task heads will drops significantly, while the model with Multi-Task heads will achieve almost the same or better results, validating our hypothesis.

Table 2. Percent gain in F1 Score with Multi-Task Learning compared to the LSTM model on varying number of students.

Students	F1-score w/o Multi-Task	F1-score Multi-Task	% Inc
5	0.47	0.585	24.2 %
13	0.436	0.583	33.7 %
23	0.426	0.594	39.4 %

5. Discussion

CALM-Net yields superior performance as it is able to model the temporal events contributing to stress of a subject while dealing with long sequences of sensor data. The Auto-encoder prevents the model to overfit on the training sequences and provides an additional boost to the *F1-score*.

The ability of CALM-Net to incorporate granular temporal information and high-level covariates, along with an architecture which is capable of deciphering personalized patterns for each student without overfitting, contributes to its high performance. Multi-Task learning improves the performance of all evaluated models, showing that stress indicators can generally be better modeled using personalized layers.

6. Conclusion

We presented CALM-Net model for predicting stress levels in StudentLife dataset. Our models are specially designed to solve three challenges which are ubiquitous in passive sensing datasets. First, it presents a general platform to address the issue of data heterogeneity with use of LSTM Auto-encoders. Second, it is able to deal with long and irregular sequences by feature engineering and histogram of categorical inference values addressing the Multi-Resolution nature of the data which is commonplace in the field. Third, by creating personalized models for every student while leveraging information from all the students it is able to achieve a *F1-score* of **0.594**. This allows us to cope with inter-subject variability providing significant improvement upon previous state-of-the-art models and beats all our baselines for predictive modeling of stress on the StudentLife dataset with minimum feature processing. Such a model could be extended for predictive modeling of affect, or more generally, behavioural and psychological quantities that are unique across subjects but have significant inter-subject similarity.

References

Adams, P., Rabbi, M., Rahman, T., Matthews, M., Voida, A., Gay, G., Choudhury, T., and Voida, S. Towards personal stress informatics: comparing minimally invasive techniques for measuring daily stress in the wild. In *Proceedings of the 8th International Conference on Pervasive Computing Technologies for Healthcare*, pp. 72–79. ICST (Institute for Computer Sciences, Social-Informatics and , 2014.

Cohen, S., Kamarck, T., and Mermelstein, R. A global measure of perceived stress. *Journal of Health and Social Behavior*, 24:386–396, 1983.

Dickerson, S. S. and Kemenyr, M. E. Acute stressors and cortisol responses: a theoretical integration and synthesis of laboratory research. *Psychological bulletin*, 130:35591, 2004.

Hochreiter, S. and Schmidhuber, J. Long short-term memory. *Neural computation*, 9:1735–80, 12 1997. doi: 10.1162/neco.1997.9.8.1735.

Jaques, N., Rudovic, O. O., Taylor, S., Sano, A., and Picard, R. W. Predicting tomorrow's mood, health, and stress level using personalized multitask learning and domain adaptation. In *Proceedings of the 1st IJCAI Workshop on Artificial Intelligence in Affective Computing (AffComp 2017), Melbourne, Australia, August 20, 2017.*, pp. 17–33, 2017. URL http://proceedings.mlr.press/v66/jaques17a.html.

Kario, K., McEwen, B., and Pickering, T. Disasters and the heart: a review of the effects of earthquake-induced stress on cardiovascular disease. *Hypertension Res*, 26:355367, 2003.

Khansari, D., Murgo, A., and Faith, R. Effects of stress on the immune system. *Immunol Today*, 11:170175, 1990.

Lngkvist, M., Karlsson, L., and Loutfi, A. A review of unsupervised feature learning and deep learning for time-series modeling. *Pattern Recognition Letters*, 42:11 – 24, 2014. ISSN 0167-8655. doi: https://doi.org/10.1016/j.patrec.2014.01.008. URL http://www.sciencedirect.com/science/article/pii/S0167865514000221.

Mikelsons, G., Smith, M., Mehrotra, A., and Musolesi, M. Towards deep learning models for psychological state prediction using smartphone data: Challenges and opportunities. *31st Conference on Neural Information Processing Systems (NIPS) 2017*, 2, 2018.

Rahman, T., Zhang, M., Voida, S., and Choudhury, T. Towards accurate non-intrusive recollection of stress levels using mobile sensing and contextual recall. In *Proceedings of the 8th International Conference on Pervasive Computing Technologies for Healthcare*, pp. 166–169. ICST (Institute for Computer Sciences, Social-Informatics and , 2014.

Rozanski, A., Blumenthal, J., and Kaplan, J. Impact of psychological factors on the pathogenesis of cardiovascular disease and implications for therapy. *Immunol Today*, 99: 21922217, 1999.

Sano, A. and Picard, R. W. Stress recognition using wearable sensors and mobile phones. *Humaine Association Conference on Affective Computing and Intelligent Interaction*, 24:386–396, 2013.

Sano, A., Phillips, A. J., Yu, A. Y., and here full, F. Recognizing academic performance, sleep quality, stress level and mental health using personality traits, wearable sensors and mobile phones. *Draft for Body Sensor Networks 2015*, 24:386–396, 2015.

Setz, C., Arnrich, B., Schumm, J., Marca, R. L., G.Troste, and Ehler, U. Discriminating stress from cognitive load using a wearable eda device , year =.

SJ, L., BS, M., MR, G., and C, H. Effects of stress throughout the lifespan on the brain, behaviour and cognition. *Nat Rev Neurosci*, 10:434445, 2009.

Stults-Kolehmainen, M. A. and Sinha, R. The effects of stress on physical activity and exercise. *Sports Med.*, 44: 81121, 2014.

Trokel, M. T., Barnes, M. D., and Egget, D. L. Health-related variables and academic performance among first-year college students: Implications for sleep and other behaviours. *Journal of American College health*, 49:125–131, 2000.

Vrijkotte, T. G., van Doornen, L. J., and de Geus, E. J. Effects of work stress on ambulatory blood pressure, heart rate and heart rate variability. *Hypertension*, 35:880886, 2000.

Wang, R., Chen, F., Chen, Z., Li, T., Harari, G., Tignor, S., Zhou, X., Ben-Zeev, D., , and Campbell, A. T. Studentlife: Assessing mental health, academic performance and behavioral trends of college students using smartphones. *UbiComp*, 2014.

Wang, R., Wang, W., Dasilva, A., Huckins, J. F., Kelley, W. M., Heatherton, T. F., and Chambell, A. T. Tracking depression dynamics in college students using mobile phone and wearable sensing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies,*, 2, 2018.

A. Appendix

A.1. Used Features

In all, our data comprises of 23 students, totaling to 1183 data points achieving roughly equal amount of training data in (Mikelsons et al., 2018). These 1183 data points have the following label distribution - 263 below median stress, 511 median stress and 409 above median. Since student 59 has 269 labels which is on an average, four times the number of labels of other students, is removed from the training set as he/she may dominate the shared layer and skew our predictions.

The list of student IDs used for training - [4, 7, 8, 10, 14, 16, 17, 19, 22, 23, 24, 32, 33, 35, 36, 43, 44, 49, 51, 52, 53, 57, 58]

We use both time-serie features and covariates as input to our LSTM, LM-Net and CALM-Net models. The time-serie features we used are time to next label, time to next deadline, activity mode (a discrete scale of 0 to 3 indicating being sedentary to high level of activity such as running), conversation duration mode, phone charge duration mode, and phone lock duration mode. Finally the covariates used are the day of the week, sleep rating, hours slept and an indicator for an exam period.

Feature Type	Feature Name	Feature Values	Mode in dataset
Discrete Sequence	Activity	$[0, 3]$	0
	Audio	$[0, 3]$	0
	Conversation	$[0, 1]$	0
	Phone Charge	$[0, 1]$	0
	Phone Lock	$[0, 1]$	1
	Day of the Week	$[0, 6]$	N/A
	Exam Period	$[0, 1]$	0
Covariates	Time to next deadline	$[0, \infty+)$	N/A
	Sleep Rating	$[0, \infty+)$	N/A
	Sleep Duration	$[0, \infty+)$	N/A

Table 3. Here we list the features we used in our experiment

A.2. Model Configurations

Hyper-parameter	value
α	0.001
β	1
Auto-encoder embedding size	128
shared layers hidden size	256
Personal layer hidden size	64

Table 4. The configuration details of CALM-NET