Independent Study Report

University of Massachusetts Amherst
Instructors: Madalina Fiterau, Iman Deznabi
Student: Yunfei Luo
Spring 2020

May 12, 2020

Personalized Student Stress Prediction with Deep Multitask Network

Abstract

In this semester, I'm working on predicting student stress with putting students in groups, in order to explore the possibility to make the original personalized model more applicable. Though it is hard for the "grouplized" model to have better performance than the personalized model, my experiments shows that grouping students by the average stress level could have F1 score that is really close to the score achieved by the personalized model.

1. Introduction and Previous work

The previous state-of-the-art model is called Cross-Personal Activity LSTM Multitask Auto-Encoder Network (CALM-Net), see *Personalized Student Stress Prediction with Deep Multitask Network* by Abhinav Shaw, Natcha Simsiri, Iman Deznabi, Madalina Fiterau, Tauhidur Rahman.

The model is constituted by a Long Term Short Memory (LSTM) autoencoder connected by fully connected layers, then connected by sub multi-layer-perceptrons (MLP) for each students. I'm mainly working on the sub MLP part.

The main problem is that when we have a new student, without collecting some data from the student, it is hard for the personalized model to make prediction due to the lack in training data. So for the sub MLP, instead of building them for each student, we build them for each group of students. So when a new student come, instead of collecting data from the student, we just need to collect some features of the students, then we could put the student into the group that contains students who share similar characteristics with the new student, then make prediction.
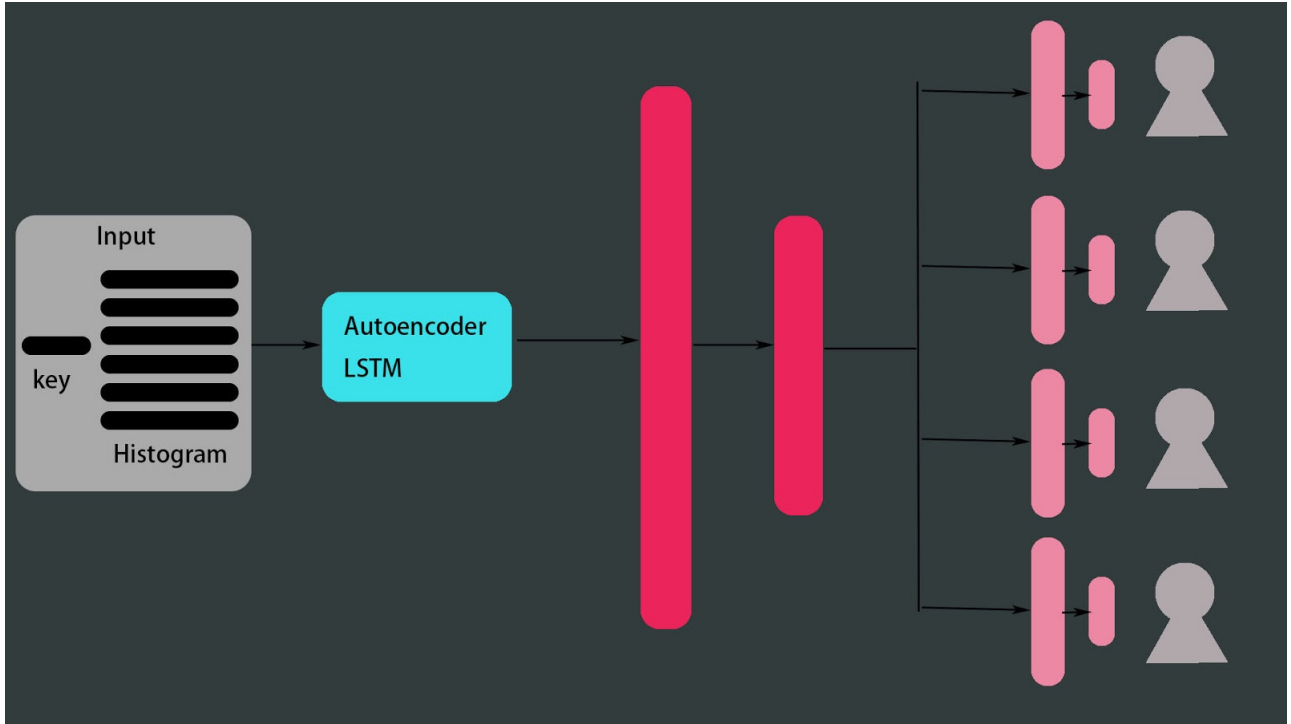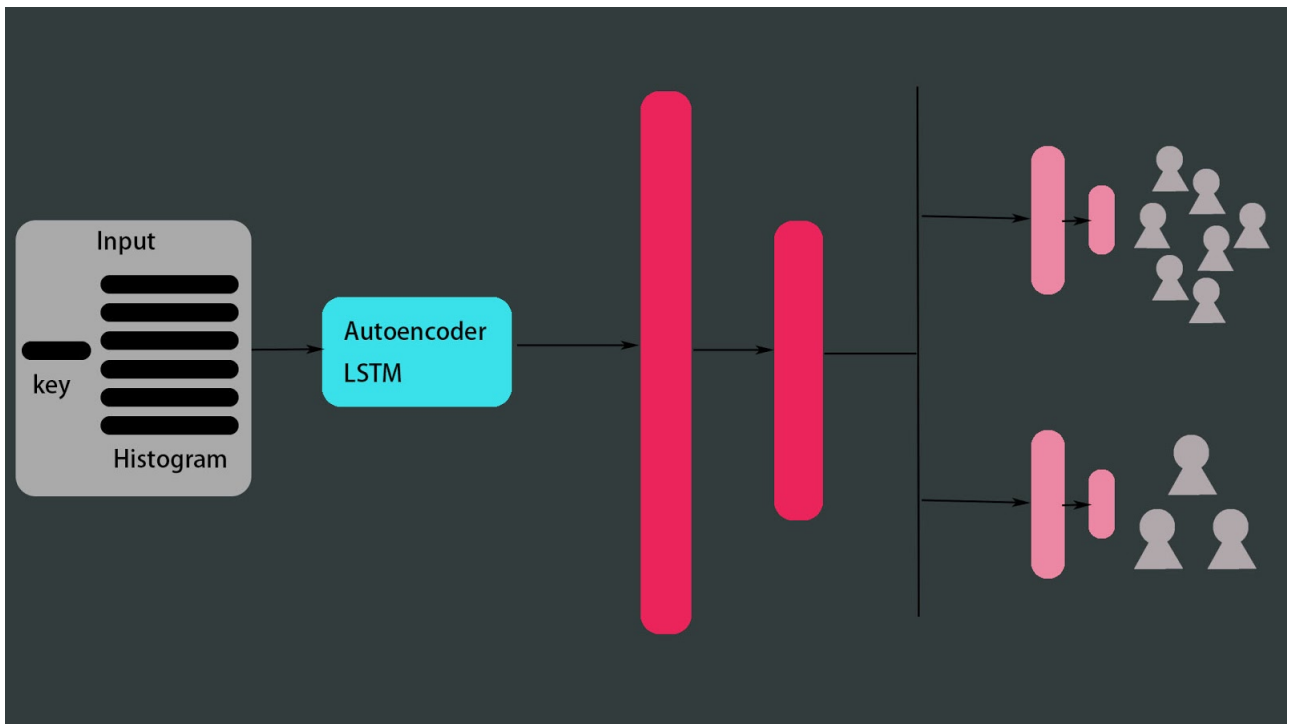
Figure 1. Personalized Model

Figure 2. Grouplized Model

2. Background Information
The StudentLife dataset was conducted in Dartmouth college where passive sensing and survey data was collected over 10 weeks among 48 students.

3. Methods
3.1. Density Based Clustering based on Dynamic Time Warping (DTW) on series of stress labels

The stress labels we have for each student form a time series data. Since the stress labels for

the students falls in different set of days during the 2 months range, using distance criteria like Euclidean and Manhattan distance is not reasonable. In order to find the similarity among students, students can be clustered based on the DTW distance, that considering the distortion of the data.

Since each student have different number of stress label recorded, the lengths of these time series data are different. So the methods based on finding centroids such K-Means and Mean-shift are not applicable. Instead, hierarchical clustering methods can be used in this case. Density-Based-Spatial-Clustering-of-Applications-with-Noise (DBSCAN) is the one I used for clustering these series data. The noise data points were greedily assigned to the existed groups that are the closest to them.
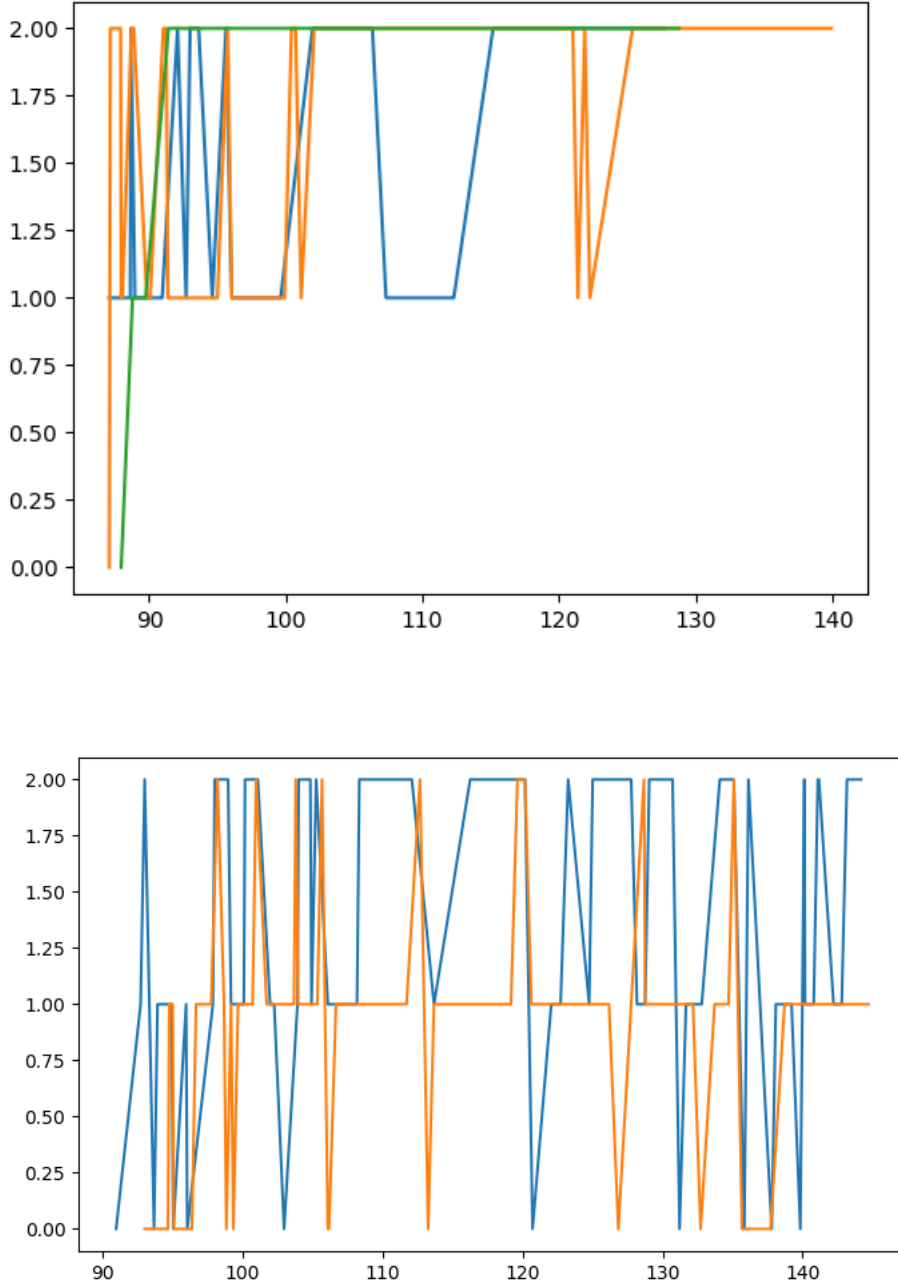

Figure 3. samples of clusters.
The x-axis is the time (days in a year + (hours in the day / 24)), the y-axis is the stress label

3.2. K-Means Clustering on aggregated stress labels, i.e. average stress labels

Considering that the density based clustering methods based on DTW is a bit complex, I simply aggregated the series data with their means to be the data point. Then the clustering becomes much easier. Though I use K-Means for clustering, it is nothing but put thresholds to split the data points.
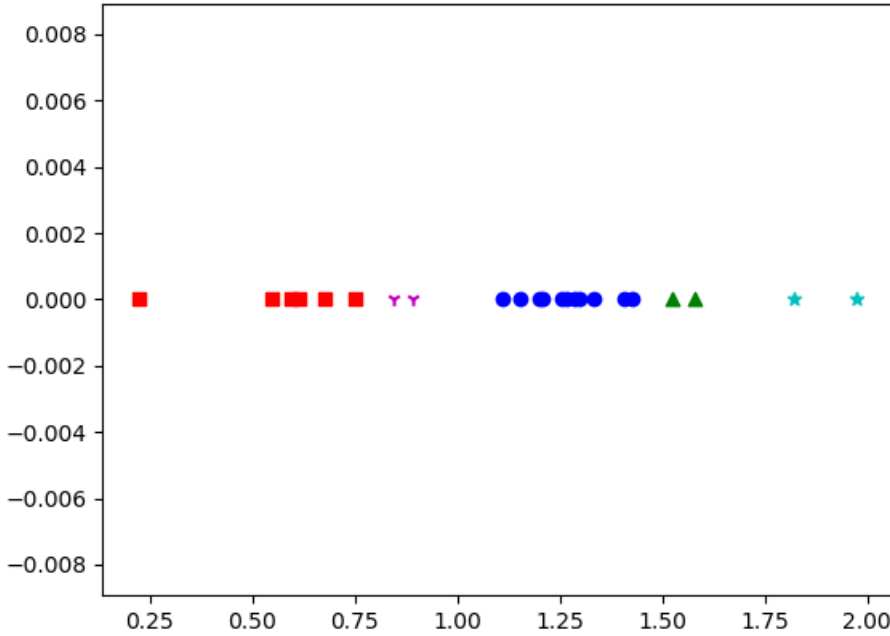

Figure 4. clusters visualize, where x-axis is the average stress labels

3.3. Density Based Clustering on chosen features from aggregated surveys score

The ideal way is to clustering students based on surveys score. So for a new student, we could first collect the survey results, then put the student in to proper cluster. For now, I had use the features: average hours slept, average deadline per week, and mode sleep rating. In order to guarantee that in the training dataset, no group contains only one student, I use the same methods in 3.1: use DBSCAN first, then put the noise data in the closest clusters.
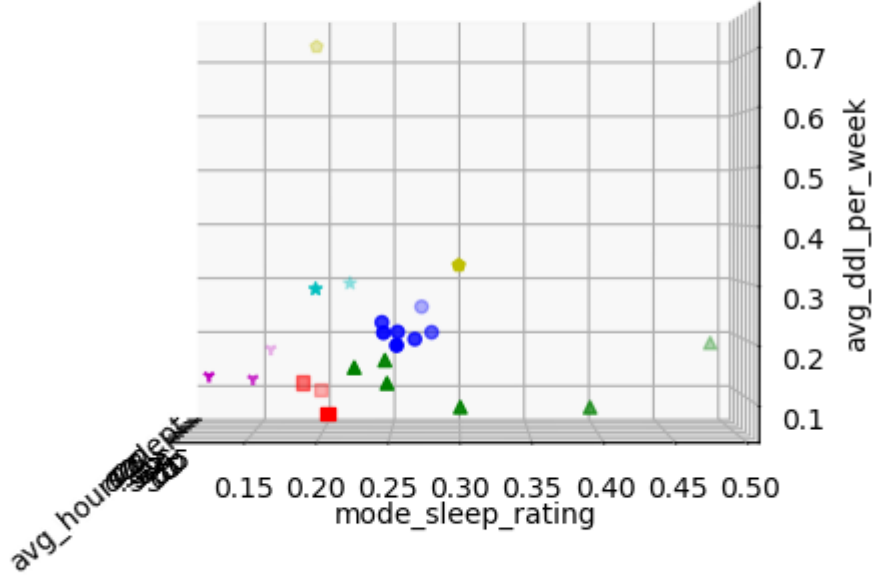
Figure 5. Clusters visualize with three chosen features from surveys

4. Experiment results:

The model evaluation method I used is 5-fold cross validation. The metrics include F1 score and AUC were averaged across the 5 splits. I've also include the results of original data in the table, i.e. each group contain exactly one student.

	Number of Groups	F1	AUC (macro)	AUC (micro)	AUC (weighted)
One_for_each (original)	23	0.594	0.679	0.702	0.680
Everyone in one group	1	0.399	0.544	0.601	0.550
avg_stress	7	0.584	0.676	0.689	0.671
Surveys	7	0.510	0.611	0.651	0.618
Surveys_normalized	6	0.515	0.615	0.651	0.616
dtw	7	0.486	0.60	0.645	0.606

Table 1. The best results that each different clustering method obtained

5. Conclusion and Future work

The results from experiments shows that using clusters based on average stress labels has better

performance. However, in order to know how well a model performed, we need to run leave one out validations, which haven't being completed. After that, we could know how the model generalized among each individual.

In addition, there are still many other surveys scores that haven't been used in clustering. In order to use these features, I need to do more work on knowing how the scores of different types of surveys are aggregated.

A. Appendix
A.1. Model Configuration
Weights for Losses: $\alpha = 0.0001$ for autoencoder reconstruction error, $\beta = 1$ for classification error
Auto-Encoder bottleneck size: 128
Shared Layer size: 256
User(Group) Layer size: 64
Epochs: 500
Gradient Descent step size: 0.000001
L2 norm regularization coeff: 0.0001
Dropout probabilities: None
Validation method: 5-fold validation, stratified splitting by the group-id/stress-label pairs