Preserving Student Mental Health

- A Data Mining Analysis on Student Stress Level

Yuan Li, Jing Wang
MSI Candidates, School of Information, University of Michigan

Abstract

For this project, we plan to build up a prediction system to indicate students mental health status based on students' daily activities. The system can help students or healthy institutes or schools to detect the potential risks of the mental issues.

Introduction of Projects

The students in college seems live in the tower of ivory and enjoy the peaceful life. But actually the students are facing different issues, such as assignments, projects, courses, the academic challenges, the uncertain future career and the peer to peer pressure. Mental health is a hot topic in every university. According to the Association for University and College Counseling Center Directors survey of counseling center directors, 95% of college counseling center directors surveyed said the number of students with significant psychological problems is a growing concern in their center or on campus. Seventy percent of directors believe that the number of students with severe psycho- logical problems on their campus has increased in the past year. But the mental problem is easy to be overlooked by students and is detected at very late phase.

Is there a solution to predict the mental health status? If we can figure out a solution to provide the students mental health in advance, it will help students and schools to support the students who have the potential possibility of mental problem to avoid or relieve the mental diseases. For this project, we are planning to build up a system to predict the students mental health status. The system can provide a score of students' mental health based on the students' daily activities. When the score is really low, it indicates the risk to cause the mental health problem.

The data set we will dive into is the students activities monitor data and the mental health sur- vey results. The whole data set includes several aspects information, for example, the sensor data which record the daily activities of the students, and the survey results which indicate the mental health status of the students. The system will com- bine the two types of data and find the connection between physical activities and mental health status.

There are many research present that physical activity can have beneficial effects on mental health. At the same time, there are many algorithms to predict the results based on the similarity. That's why we believe the system can be built up by using the proper algorithms.

Dataset

Original Dataset:
- 3.05 GB, 2,081 files
- App_usage, calendar, call_log, dinning, education, EMA, sensing, SMS, survey

Dataset used in this project:
- Education, EMA, survey combination
- 60 participants, 25 features
- Time series considered

Data clean:
- Data relatively clean and don't have much noisy
- Data missing: for survey data, use average value; for educational data, use zero if it is dominated by zero

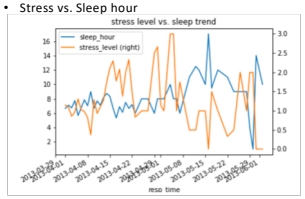
Methodology

Time Series and Behavior data mining – students study life vs. stress

- Data selection and cleaning
- Data analysis and similarity
- Features analysis
- Two baselines and evaluation
- Time series analysis
- Stress prediction based on features

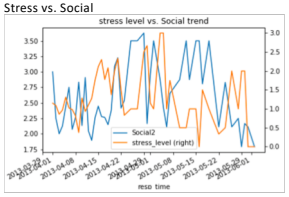
Analysis – Granger Causality

- Stress vs. Sleep hour



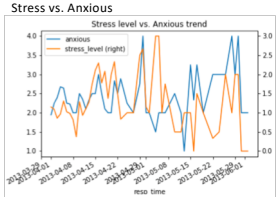
- Stress vs. Exercise Trend

- Stress vs. Social

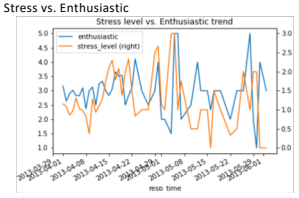


- Stress vs. Working Hours

- Stress vs. Anxious



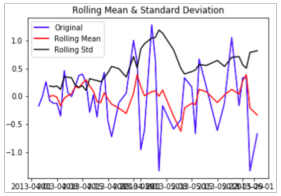
- Stress vs. Enthusiastic



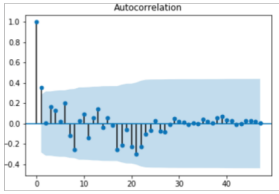
Prediction

- Time Series Prediction

- Rolling Mean and Standard Deviation



- Autocorrelation



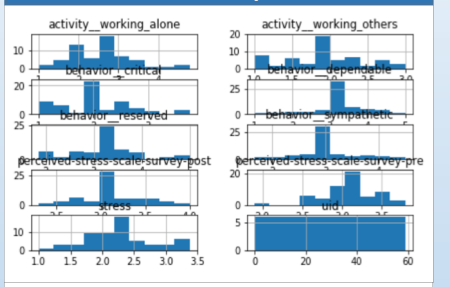
Conclusion

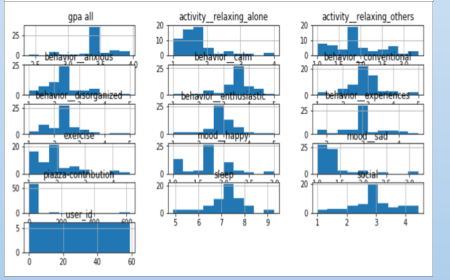
This project performs granger causality analysis on the student behavior data within a period of time, intended to explore the relationships between students' stress level and their daily behavior. Our results reveals the general pattern of students' behavior in dealing with stress -- their anxious level would peak soon after their stress level peak, and students' working hours demonstrates a leading influence on their stress scale to increase, with p-value reaching the smallest of all situations(p-value = 0.0010).

We are able to conclude that instead of sleeping hours affecting students' stress level, in fact it is their stress scale that would have a direct influence on sleeping hours, though in general sleeping hours and stress level are on the opposite direction of changing. Our analysis also unveils students' general approach to relieve stress -- they tend to exercise more and be more social to seek support from acquaintances not long after they get stressed out.

In this project, one main goal is to predict students' stress scale based on the interactions with other factors. We build a Random Forest classifier to handle that. The results show that 73% of the students are classified as "stressful", which is slightly smaller to our baseline of 75%. This is partially due to the sparsity characteristic of our data set, including only 60 sampled students yet with over 80 features that are measured.

Data Shape



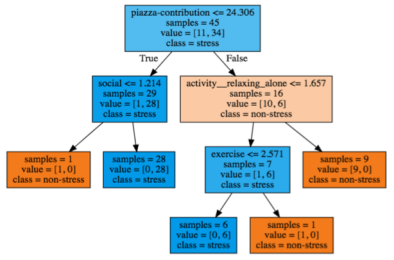


Evaluation Methods

Baseline one
- Using the mean value of feature "stress"
- The average perceived stress scale: 3.086

Baseline two
- Random guessing baseline
- Per survey result, 75% of students are perceived as stress, and about 16% of students are perceived as non-stress

Evaluation method
- Decision tree classifier
- Random Forest Classifier



Decision Tree Classifier Analysis