Refined Initial Proposal

Shichen Li
sl4bq

March 2, 2020

1 Introduction

According to the literature, individuals struggling with mental health problems face a significant amount of obstacles in their daily life. The recent advancements in m-Health field have demonstrated significant potential for alleviation or even removal of some obstacles. Social anxiety is defined as an intense fear of being judged and scrutinized by others that is beyond what would normally be expected in a situation. We are eager to know that which features might be indicative of changes in anxiety, with the goal of being able to provide immediate interventions to individuals when an increase in anxiety is detected. We aim to discover the bio-markers who can predict the changes in anxiety in this project.

2 The Decision Problem

According to the experiment done in the last semester, we focus on the questions measuring anxiety and there are 4 surveys taken in baseline, anticipatory, experience and post event periods for each experience type (solo-video watching, dyad no-evaluation conversations and dyad evaluation conversations on specific topics). Each question is scored from 1 to 5 according to the severity of the situation.

During the data preprocessing, we need to extract the data streams about the heart rate, accelerometer and gyroscope signal vector magnitude (SVM), audio data, skin conductance data, and any other motion data we have.

Now this project should be extended to something else: we need to detect the fluctuation and tendency of all measures, such as physiological ones and physical ones. The features can be different in many aspects in various types of interactions. What we need is to differentiate and try our best to predict the right symptom. Based on this, we can take the right action to treat every individual with what they want.

3 The Predictive Tool

The predictive tool is very complex, and each step needs a specific algorithm. All these algorithms are done on Python 3 platform. First, in the signal feature extraction, we need to use the sliding window algorithm to get the most predictive features, maybe from 5 to 12 items. Then we need to find the score changes from each episode to the next episode. There are 3 different score changes in each experience type. Finally, we need to use some machine learning techniques, like Random Forest Classification, T-test, ANOVA, and so forth, to build the predictive model. The final goal is to know how well each type of patient will respond to each treatment option.

4 Formalized Model

4.1 The decision makers' action set

The action depends on the symptoms of each patient, and we need to make a complete and specific list of options available to decision makers, and the choice is always unique. To find the most suitable treatment, we need to identify the most important period and activity that each subject might show the fear.

The action we take a^* can be one or up to two combinations of options from these two sets:

$\mathbb{A}_1 = Cogitivebehavioraltherapies(CBT) = promotionofexercise, exposureandsocialskillstraining, grou$

$\mathbb{A}_2 = Pharmacologicaltherapies = selectiveserotoninreuptakeinhibitors(SSRIs), selectiveserotoninnore$

4.2 The Sample Space

The sample space here is the 4 subject's data of ECG and GSR signals. Due to the extremely large size of .csv files, we cannot upload on Github, and we just list some of the data streams in each file: Heart rate, skin conductance, accelerometer, gyroscope...

4.3 The Parameter Space Θ

All possible values depend on the data stream and this is the only factor making great influence on this project.

4.4 The Prior Beliefs

The prior beliefs is based on what we already have done.

By using the random Forest classifier feature ranking our best choice predictors were the Acceleration mean, minimum acceleration values, acceleration values for 75%, the maximum acceleration values, the activity level, gyroscope entropy val- ues, the galvanic skin conductance entropy, standard deviation, and mean, the maximum values, skewness, kurtosis.

The achieved accuracy was very low for both of the classification algorithm Support Vector machines, and Random Forest Classifier, just above 30%, as Accuracy was only 31.08, and for RF between the 5 classes was 33.4%, training and testing cross validation was in 60 to 40 ratios.

We have tried to generalize a little bit as there was no change in anxiety scores, after using three classes the classifier accuracy was improved a lot, and by using only 12 features the attainable accuracy was 76.22%. After that we wanted to build personalized models, for these the classifier accuracy was high as expected, for subject 1 it was near 100%, similar results are working for participant 2 also. The support vector machine using radial basis function was not similar to the random forest, the participant results are 79.24% and 77.81% for subject 1 and 2 respectively, the rest of the participant does not have changed sores to generate any classifier model. The reason for using only top 12 features were the results from most prominent features for random forest classifier.

4.5 The Payoff function

The loss function is focusing on the social anxiety scores for next hour of each subject, and we need to calculate the score changes after treatment, and use the product of treatment.

4.6 The Payoff function

The utility function is 1 when it proves to be true, and -1 when it proves to be false.