NeuroFlow Data Team Take-Home Project

Ming Zeng (ming.zeng26@gmail.com)

February 2020

1 what insights can you draw?

I first randomly selected a patient in the dataset. Then, I created his/her personal dataframe with the four metrics (mood, sleep, anticipatoryStress, ruminationStress) which he/she scored at different time. In this process, the missing values have been imputed by the mode in each column. Plotted the four metrics over time in the Figure 1.

Figure 1: Therapy Progress over Time

We can see that there is no monotonous upwards or downwards of the effectiveness of the therapy for this patient. Considering the findings that patients with insomnia attended one CBT session via the internet per week for 6 weeks as a course. I resampled the data with the frequency of 42 days (6 week) in order to see the progress of every course of the therapy. We can see in the Figure 2:

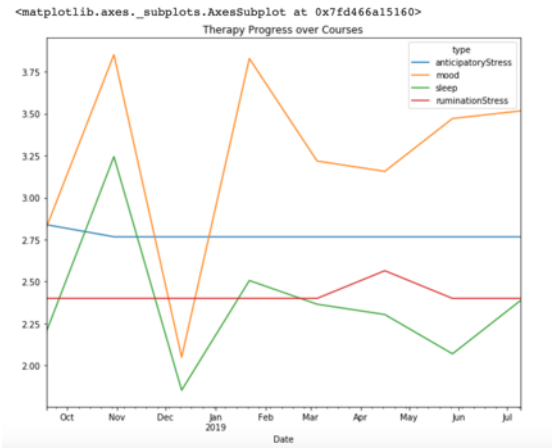

Figure 2: Therapy Progress over Courses

Although we cannot see an apparent improvement of the four metrics in the courses of the therapy. We find an interesting phenomenon that the scores of mood and sleep experienced a dramatically decrease in November and December. I therefore want to examine whether this change is mainly caused by SAD or it is a real reflection of the patient. By decomposing the two series we can see in the Figure 3:

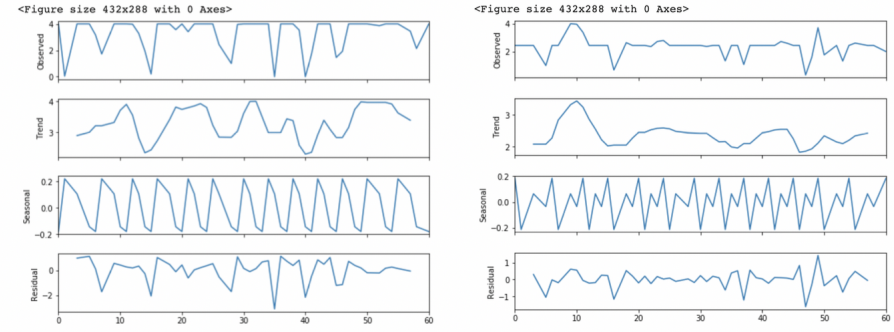

Figure 3: Seasonal Decomposition of Mood and Sleep

There has a seasonality existed in the mood and sleep respectively. We need to rule out the seasonality when we attempt to analyze the progress.

Until now, we have plotted the progress of a random selected patient. This methodology can be applied to the entire dataset.

	type	anticipatoryStress	mood	ruminationStress	sleep
	type				
anticipatoryStress		1.000000	-0.197735	-0.501365	-0.022452
mood		-0.197735	1.000000	-0.205882	-0.011066
ruminationStress		-0.501365	-0.205882	1.000000	-0.048487
sleep		-0.022452	-0.011066	-0.048487	1.000000

Figure 4: Correlations

From Figure 4, we can conclude that sleep is negative correlated with the other factors. The ruminationStress has the maximal negative relationship with sleep.

2 What assumptions have you made about the data?

In order to make this dataset useful and reliable, these assumptions shuld be satisfied:

- Good completion rates.
- The patients are well-diversified. If the sample are formed by a certain type of individuals, like mid-age teachers, the conclusion about the effectiveness of the therapy could have bias.
- The patients are telling their real feelings. Otherwise, the data is unreliable.
- No discrepancies in our data. That means our dataset have proper designed data entry forms, no human error in data entry, no deliberate errors, etc.
- No other therapies regarding the same issues are taking at this time.

3 What are 2-3 additional pieces of information that would be important to collect?

- Sociodemographic factors. Each patient's age, sex, marital status, education, employment, and personal health history should be included in this dataset.

- Objective sleep outcomes. The objective sleep outcomes will give us a robust measurement of the effectiveness of the therapy.
- Other factors, such as how many alcohol or coffee the patient consumed.

Reference

1. Thakur, V.K., Wong, J.Y., Randall, J.R. et al. An evaluation of large group cognitive behaviour therapy with mindfulness (CBTm) classes. BMC Psychiatry 19, 132 (2019). https://doi.org/10.1186/s12888-019-2124-5;

2. Colin A. Espie, PhD, Simon D. Kyle, PhD, Chris Williams, MD, Jason C. Ong, PhD, Neil J. Douglas, MD, DSc, Peter Hames, MA Oxon, June S.L. Brown, PhD, A Randomized, Placebo-Controlled Trial of Online Cognitive Behavioral Therapy for Chronic Insomnia Disorder Delivered via an Automated Media-Rich Web Application, Sleep, Volume 35, Issue 6, 1 June 2012, Pages 769–781, https://doi.org/10.5665/sleep.1872