CS711: GAME THEORY AND MECHANISM DESIGN

Project Report: Modelling Depression Among University Students using a Game Theoretic Approach

Group Members : Gagesh Madaan (20111406), Muchamad Taufiq Anwar (20111270), Sri Madhan M (20111420), Siddharth Goel (180761), Harsika Diksha (20111021).
Project Mentor: Utsav Singh

1 ABSTRACT

In this Project, we try to prove a hypothesis that a university/college student surrounded by neighbouring students who have a more positive life cycle/habits is less likely to get depressed as compared to one who has more positive life cycle/habits as compared to most of his neighbours. We take help of Game theory to design a game that a student plays in daily university life in the quest to achieve the Strongly Dominant Strategy in this Game. We will take help of a well contained student related data set, and random lattice based simulation to prove the hypothesis and the results would be validated using the Law of Large Numbers. We get some plausible results/conclusions from this kind of a modelling that keeping a healthy amount of depressed students arranged closer to each other may result in increase in the final no of depressed students.

2 INTRODUCTION TO THE PROBLEM

Certainly college/university is a place, no matter how much big in size, always remains a tightly knit environment where students engage among themselves and in the process, they learn things from each other. With immense pressure from academics and other things, there may be situations when some students fall prey to the trap of serious Depression and tend to deteriorate their positive habits like sleep, activity, exercise etc. Most of the times unrevealed, there may be very less times when students actually reveal their disturbed state of mind to the outer world.

Our Game-Theoretic Model tries to fit this tightly knit environment into a lattice based simulation combined with some randomness. The model uses a Game Theory concept based Extensive Form Game that a university/college student plays to maintain their positive habits even when lot of pressure is there. This game considers these habits as a strategy and the payoffs are the reward for maintaining these habits even when times are difficult. The model takes these students as players of their own game and put them in a simulated environment(which we call as lattice) and we try to prove a hypothesis that a student's depression gets highly affected by the people who are around him/her.

2.1 Related work

We take many main ideas like Lattice Simulation, Game Theory Game from [1]. It discusses Depression among working professionals taking the Game of formal education. The other important difference in this work's approach and our approach is the way we setup our hypothesis and the way we find the final deciding threshold. [2] uses Game Theory to assess some patient's depression state by making them play some of the most common simple games like Prisoner's Dilemma and by noting down if they give a very irrational response. [3] helped us to identify the parameters/habits/strategies of the game. [4] is the parent study of the dataset that we used. This paper also tries to study depression among college students but not by using Game Theory but by correlating parameters and machine learning models. [5] is similar to [4] and [3] and uses a large sample of 200 students for their analysis.

2.2 Brief overview of the report

The next section introduces the StudentLife dataset that we used and what kind of data it contains and how we are retrieving our game parameters from this dataset. Fifth section discuses the Game in details and the mathematical objective associated with our hypothesis. Sixth section discusses how we compute the deciding factors of our simulation. Seventh section discusses the insights and conclusions that we get from our simulation based hypothesis.

3 StudentLife Dataset

[dataset] gives in detail information of the StudentLife dataset. It was part of the study of [studentlife] that took place at DartMouth College,New Hampshire, United States. It uses passive and automatic sensing data from the phones of a class of 48 Dartmouth students(8 seniors, 14 juniors, 6 sophomores, 2 freshmen, 3 Ph.D students, 1 second-year Masters student, and 13 first-year Masters students) over a 10 week term to assess their mental health (e.g., depression, loneliness, stress), academic performance (grades across all their classes, term GPA and cumulative GPA) and behavioral trends(sleep, exercise, conversation etc).
Comprises of over 53 GB of continuous data, 32,000 self-reports, and pre-post surveys and contains data related to many parameters that we were interested in.

3.1 Parameters used for the Game Model

There are 3 types of data we are using from the dataset,

1. EMA
2. Sensing
3. Survey

EMA data contains set of questionnaires such as Sleep, Social, Activity, Events, Exercise, Behaviour and each set contains multiple questions. Sensing data contains sensor data of Activity, Phone Lock, Light, Conversation. Survey data contains standard PHQ-9 depression questionnaire to screen the depression level of the students. EMA and Survey are using **Likert Scale**(number of items depends on the types of question), where as Sensing data uses timestamps along with the sensor data.

3.2 Calculation of Payoff and Depression score

Based on the type of question, weight-age were given to the each of the options in the EMA and Survey questions. Within each questionnaire set, weights are summed up and averaged over each student(finding average because, students took the questionnaires multiple times). Averaged score in PHQ-9 Survey data is used as it is, where as each questionnaire set in EMA data are then normalised between 0 and 1. From Sensing data (Activity, Phone Lock, Light), we derived the Sleep behaviour based on the condition that *"student is stationary, phone is locked and phone is in dark environment"*. Now we have 2 different sensing data, Sleep and Conversation. Students are given score as 1 if their sleeping behaviour is good, else score 0. Similarly, students are given score as 1 if they have consistent phone conversation to people they know, else score 0.

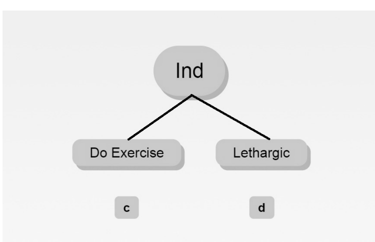
Then total payoff is calculated as,

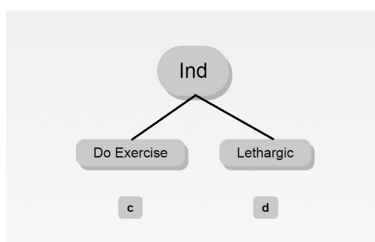
$$\begin{aligned}\text{Total Payoff} = \; & \text{EMA}_{Sleep} + \text{EMA}_{Social} + \text{EMA}_{Activity} + \text{EMA}_{Events} + \text{EMA}_{Exercise} + \text{EMA}_{Behaviour} \\ & + \text{Sensing}_{Sleep} + \text{Sensing}_{Conversation}\end{aligned}$$

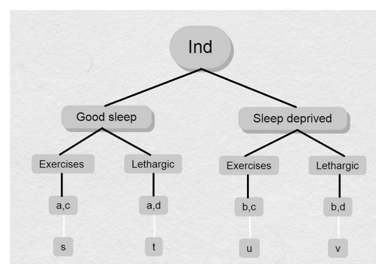
4 THE GAME

As we have discussed earlier, there are 7 parameters based on which we are determining number of individuals affected by depression. We have used all 7 parameters to find the total payoff of any player.

An individual can play multiple sub games simultaneously. We have considered choosing each parameter as a game, which will reward a player with certain payoff. After playing all those 7 games, an individual will have 7 payoff values. We add all those payoffs to find a final score of that player. Please refer to the images below.







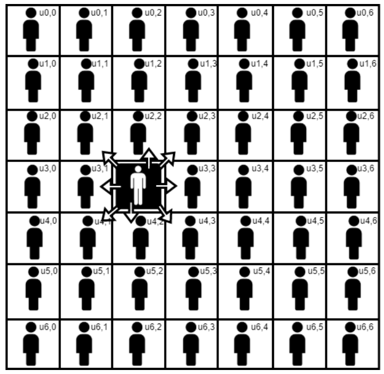
Similarly we can define game for other parameters (Social, Activity, Behaviour, Events, Conversation) also.The payoff values (a,b,c,d...etc) are normalized values and the individual can be awarded with either a or b but not both for 1st game and likewise for further games. It is reasonable to assume that sound sleep is better than being sleep deprived; therefore, a>b.

A game with 7 strategies will give 128 payoffs and representing all parameters is not feasible, so we have tried to show the complete EFG game using 2 parameters, say sleep and exercise data, shown in the image 4.

An individual can choose the strategies as Good sleep or Sleep deprived. After choosing this, he can now choose between to do Exercise or to be lethargic. Now according to strategy he choose, he will be awarded with the score, s, t , u or v. We can definitely conclude few of the points after looking at the figure 4,

- s>t,u,v and u>v, but we cannot make any inequality function for t and u as these values depends on the state of players.
- **v** is *strictly dominated* strategy and no player wants to play that combination of strategies.

- Every player would likely to deviate towards the left subtree in order to achieve higher payoff.



After every individual have completed playing the game, we will have in total of 49 payoffs. We have represented the entire population in a square lattice, n*n matrix. In this lattice top bottom edges as well as right left edges are connected. An individual maintains social contact with 8 neighbors. The similar model is used in finding the spread of contagious disease. These 49 payoffs are randomly distributed in the lattice, as shown in the figure.

5 THE FORMAL MATHEMATICAL REPRESENTATION

5.0.1 SET OF USERS(U)

$$U = \{u_{0,0}, u_{0,1}, \ldots, u_{0,n-1}, u_{1,0} \ldots u_{1,n-1} \ldots\ldots u_{n-1,0} \ldots u_{n-1,n-1}\}$$

5.1 SET OF NEIGHBOURS($N(u_{i,j})$)

$$\begin{aligned} \mathrm{N}(u_{ij}) = \{u_{xy} \mid x \in \{(i-1) \bmod n,\ i,\ (i+1) \bmod n\} \wedge \\ y \in \{(j-1) \bmod n,\ j,\ (j+1) \bmod n\}\} - \{u_{ij}\} \end{aligned}$$

where σ(REAL NUMBER) and δ(NATURAL NUMBER) are our parameters and $p(u_{i,j})$ is the payoff function for user $u_{i,j}$. If looked carefully, δ is the number of neighbours who has less payoff than the users and σ is the real number value that tells how much less this payoff should be.

5.2 LIKELY TO BE DEPRESSED FUNCTION($D(u_{i,j},\sigma,\delta)$)

$$D(u_{i,j},\sigma,\delta) \ = \ I\Big(\sum_{u_{m,n} \in N(u_{i,j})} I(p(u_{i,j}) - p(u_{m,n}) > \sigma) > \delta \ \Big)$$

5.3 Count of people who are likely to get depressed($C(U, \sigma, \delta)$)

$$C(U,\sigma,\delta) = \sum_{u_{i,j} \in U} D(u_{i,j},\sigma,\delta)$$

In reference to our Random Allotment Experiment, Number of Persons Likely to Get Depressed becomes a Random Variable whose value will be determined as an outcome of the very same experiment, we use a subscript "Random" to denote it. We also have the actual Number of Persons Depressed from our dataset. We use subscript "Actual" to denote it: C_{Random}, C_{Actual}

5.4 FINAL OBJECTIVE((σ^*, δ^*))

$$(\sigma^*,\delta^*) = arg_{\sigma,\delta}\ E[C_{Random}(U,\sigma,\delta)] - C_{Actual} \approx 0$$

6 (σ^*, δ^*) DETERMINATION

The determination of the σ and δ values were done by doing repeated experiment by randomly distributing the users(this is the random experiment that we mentioned about in the last section) in to the lattice. For each of the σ and δ value pair, we did 100 random experiments and calculate the average number of depressed student and also the standard deviation. The result of the experiment is shown in Table 7.1. From the experiment, we observed that the model is very sensitive to the σ value with lower σ will result on higher number of depressed students. On the other hand, higher δ values will result on fewer number depressed students. The best experiment result was found for $\sigma = 1.71$ and $\delta = 4$ which gave an average number of depressed students of 5.03. You can refer **??** from appendix B.

7 INSIGHTS AND CONCLUSION

Our proposed hypothesis is represented by a lattice of randomly assigned positions to the users $u_{i,j}$ with different calculated payoffs, each having eight neighbors. We analyze the user's neighborhood by comparing the payoff with every neighbor as per the algorithm mentioned in (Section-6). We found out that as we randomly assign positions to different users, we get different answers depending on their positions. We hypothesize that, if similar score clusters are put together, there will be fewer number of depressed people due to the positive surroundings.

If similar payoff people hold the positions in a scattered fashion far from each other the number of negative people would increase hence getting a larger number of negative people. i.e.

$$C_{Random}(U,\sigma,\delta) > C_{Actual}$$

On the contrary, if they are put in closer vicinity, there would be lesser number of depressed people. i.e.

$$C_{Random}(U,\sigma,\delta) < C_{Actual}$$

This is a reference to the graphs (C_{Random} vs Iteration) and (Average of depressed people by Number of Iteration) respectively **[C]**.

Extending on the idea “You are the average of the five people you spend the most time with” - Jim Rohn. We analyzed our experiment over 1000 iterations, we found that the variation is as expected, as shown in the plot (C_{Random} vs Iteration). We analyzed our experiment over 1000 iterations, we found that the variation is as expected, as shown in this plot (C_{Random} vs Iteration). For a set of forty-nine people we get the range of peaks for a different position on the lattice, still, with a high density of people clustered around five for the lattice threshold calculated for this experiment.

We randomized our experiment for 1000 iterations putting people at different lattice points, and we got an intriguing result which follows the L.L.N. (Law of Large Numbers [LLN]). The Law describes the result of performing the same experiment a large number of times. According to the law, the average of the results obtained from a large number of trials should be close to the expected value and will tend to become closer to the expected value as more trials are performed.

In the plot (Average of depressed people by Number of Iteration), we can see that plot is converging very close to the theoretical value for the lattice threshold. This resulted in a constant number of depressed people after a large number of iterations, resulting in the validation of our hypothesis. So we are concluding that regular randomization is good for social welfare to a certain extent. People sharing similar interests will get clustered together as time passes. Which in turn uplifts the already uplifted but cast down the dispirited part of the crowd, causing loss of enthusiasm and hope. While these explanations are certainly plausible, they develop isolation from strategic considerations such as selective incentives or community enforcement that have dominated the analysis of collective action in the rational choice tradition.

CONTRIBUTION

Siddharth Goel - Data cleaning, Prepossessing of parameters, Lattice formulation, Insights.
Harsika Diksha - Game formulation for calculating the payoffs.
Taufiq Anwar - Data cleaning, Prepossessing of parameters, Lattice formation.
Sri Madhan M - Data cleaning, Prepossessing of parameters, Analysing the Dataset.
Gagesh Madaan - Data cleaning, Prepossessing of parameters, Game formulation and mathematical model.

8 References

[1] R. Loula1 and L. H. A. Monteiro, A Game Theory-Based Model for Predicting Depression due to Frustration in Competitive Environments. - https://www.hindawi.com/journals/cmmm/2020/3573267/

[2] Yun Wang, Liu-Qing Yang, Shu Li, and Yuan Zhou - Game Theory Paradigm: A New Tool for Investigating Social Dysfunction in Major Depressive Disorders - https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4569817/

[3] Mei G, Xu W, Li L, Zhao Z, Li H, Liu W, Jiao Y - The Role of Campus Data in representing Depression among College Students - https://mental.jmir.org/2020/1/e12503/

[4] Rui Wang, Fanglin Chen, Zhenyu Chen, Tianxing Li, Gabriella Harari,Stefanie Tignor, Xia Zhou, Dror Ben-Zeev, and Andrew T. Campbell, StudentLife: Assessing Mental Health, Academic Performance and Behavioral Trends of College Students using Smartphones - https://www.andrew.cmu.edu/user/fanglinc/pubs/ubicomp14-studentlife.pdf

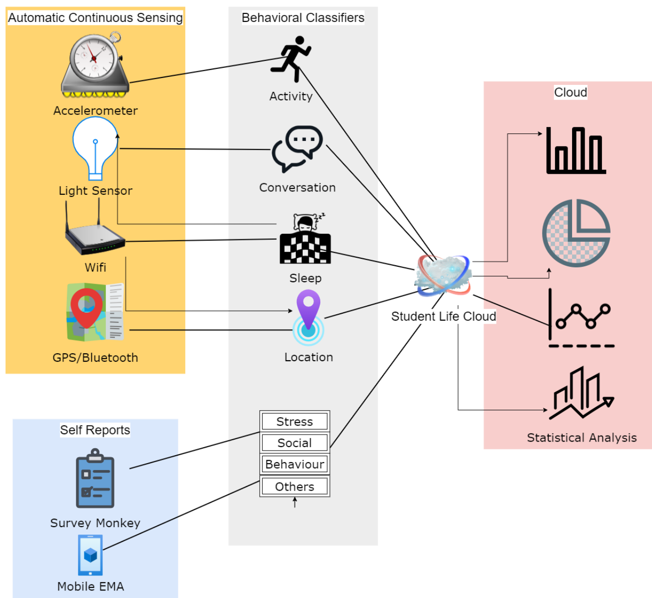
[5] "Trockel MT, Barnes MD, Egget DL", "Health-related variables and academic performance among first year college students implications for sleep and other behaviours - https://pubmed.ncbi.nlm.nih.gov/11125640/

[6] StudentLife Study - https://studentlife.cs.dartmouth.edu

[7] LAW OF LARGE NUMBERS - https://en.wikipedia.org/wiki/Law_of_large_numbers

Appendices

A StudentLife Architecture

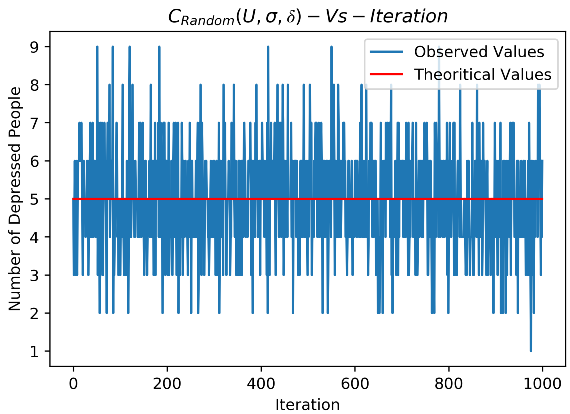


B Experiment result with different σ and δ values

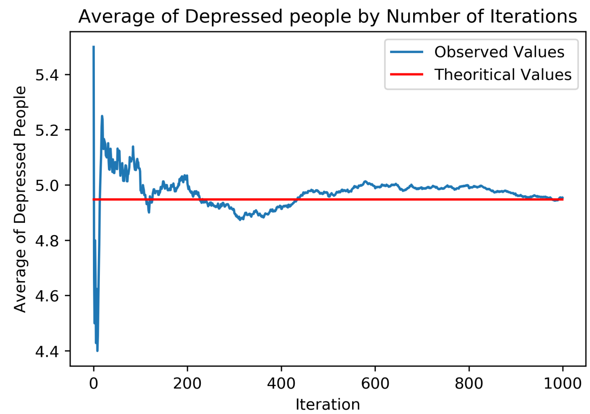
Table A.1: Experiment result with different σ and δ values

σ	δ	**mean**	**stdev**
2	3	5.88	1.328
2	4	3.51	1.176
2	5	1.92	1.051
1.5	3	9.51	1.573
1.5	4	6.67	1.627
1.75	4	4.75	1.373
1.7	4	5.41	1.264
1.71	4	5.016	1.134
1.72	4	5.03	1.218

C GRAPHS



$C_{Random}(U, \sigma, \delta) - Vs - Iteration$
Observed Values
Theoritical Values
Number of Depressed People
Iteration



Average of Depressed people by Number of Iterations
Observed Values
Theoritical Values
Average of Depressed People
Iteration