Predicting the Level of Stress using Mobile Phone Data

Saman Rahbar
Nathan Yang
Rifat Zahan

Graduate Students
Department of Computer Science
University of Saskatchewan
Saskatoon, SK, CANADA

April 12, 2017

UNIVERSITY OF
SASKATCHEWAN

Motivation

- Frequent and chronic stress are associated with immune disorder, depression and different types of chronic diseases.

- Monitoring stress at work and studying the associated factors have become quite popular.

- Massive use of smartphone has created a solid ground to study human behavior.

PC: Stress and consequences source:parkipourlavie. Retrieved from: Stress Management Induction, Ludovic MOLLUSO

Objective

The objective of this study is to predict stress level among university students by using smartphone data.

Research Questions

Research Question 1

What factors are significantly associated with stress level among university students?

Research Question 2

What would be the best model to predict stress level with smartphone and survey data (e.g., transportation method used, time spent on campus, types of food purchased, involvement in physical activities, etc.)?

Operationalization

WiFi Data

Accuracies $\leq$ 50 m were considered to calculate dwell time on campus.

Response Rate

At least 5, 10, 12, 15 responses from each participants were considered for further analyses.

Operationalization: continued...

Survey Data

- **Stress Score**: Daily median stress scores (0 - 10; 0: no stress at all, 10: extremely stressed) of each individuals were considered for analysis.
- **Transportation Mode**
    - Bus
    - Car/Carpool
    - Walk/Run
- **Exercise Type**
    - Exercise/No Exercise group
    - People, who considered walk/run as exercise were eliminated from the study.
    - Physical activity, which happened 10 hours ago were not considered.
- **Food Choice**
    - Fast Food Consumption
    - No-Fast Food Consumption

Methodology

Association

Chi-square test can be used to assess the association between stress level and several other covariates.
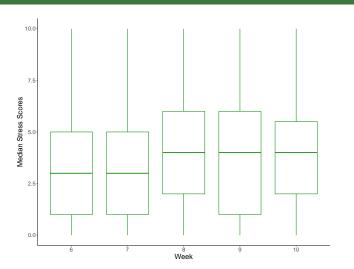
Prediction

Ordinal logistic regression model can be used to predict the stress level in terms of other covarites.

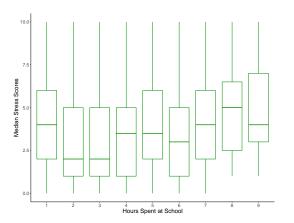| Quartiles	Stress Score
0%	0.0
25%	1.0
50%	3.0
75%	5.5
100%	10.0

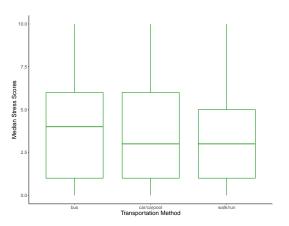Figure 1: Quantiles of stress levels among students; sample size, n = 751 from 86 participants

Figure 2: Distribution of median stress scores among different weeks.

Week	Frequency
Week 6	157
Week 7	186
Week 8	130
Week 9	210
Week 10	68

Figure 3: Distribution of median stress scores among different hours spent at school.

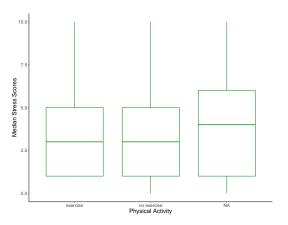| Transportation Mode	Frequency
Bus	284
Car/Carpool	170
Walk/Run	297

Figure 4: Distribution of median stress scores among different transportation methods used to come to school.

Key Background Results: continued...

Figure 5: Distribution of stress scores among different food groups.

Fast Food Consumption	Frequency
Yes	72
No	116
NA's	563

Physical Activity	Frequency
Yes	57
No	131
NA's	563

Figure 6: Distribution of median stress scores among physical activity groups.

Operationalization Results

Total sample size, n = 751 from 86 participants

Covariate	χ^2- test statistic	Degrees of Freedom	p- value
Week Number	77.2	72	0.3165
Time Spent at School	256.2	198	0.0033**
Transportation Method used to Come to School	74.4	54	0.0342**
Involvement with Physical Activity	19.2	18	0.3796
Types of Food Purchased	32.4	36	0.6416

Figure 7: Pearson's χ^2 test statistic to assess the association between stress level and different variables.

Results: Ordinal Logistic Regression Model

Covariates	Odds Ratio (OR)	95% Confidence Interval (CI)	p-value
Transportation Mode			
Bus (Reference Cat.)	-	-	-
Walk/Run	0.68	[0.49, 0.94]	<0.05
Car/Carpool	0.84	[0.58, 1.23]	>0.10
Time Spent at School			
One Hour (Reference Cat.)	-	-	-
Two Hours	0.98	[0.56, 1.72]	>0.10
Three Hours	1.35	[0.80, 2.30]	>0.10
Four Hours	1.98	[1.16, 3.40]	<0.05
Five Hours	1.50	[0.86, 2.61]	>0.10
Six Hours	2.29	[1.32, 3.98]	<0.01
Seven Hours	3.68	[1.88, 7.21]	<0.01
Eight Hours or More	3.42	[1.80, 6.24]	<0.01

Total sample size, n = 620 from 58 participants; at least 5 responses from each individuals

Figure 8: Summary results from Ordinal Logistic Regression Model (ordered stress score is the dependent variable).

Discussion

Answer to Research Questions

- **Transportation method** that were chosen to come to school and the **amount of time spent on campus** show significant association with stress level.
- **Stress level** increases with **hours spent on campus**.
- The study found that those, who come to school either by foot or by car were less stressed, compared to those, who came to school by bus.

Some Comments

- Using response rate of 5/individuals was able to give more information compared to other response rates (e.g., 1, 10, 12, 15 responses/individuals).

Limitations of the Study

- Inclusion of **physical activity** or **food consumption** in the model imposes some limitations in the number of data accessible to use, consequently they were not considered in the final model

Future Work

- **Model validation** or diagnostic checking will be done.

- Literature shows that **involvement in physical activity** are significantly associated with stress level. Different operationalization should be used to look for such association.

- **Weather condition** may have some effects on stress level; therefore, it will be considered for future analysis.

- Modern **Machine Learning** algorithm should be implemented to classify the stress level.

- **App usage data** will be incorporated as possible covariate(s) while predicting stress level.

Conclusion

Mobile phone data exhibits that different types of **transportation** used and **time spent at school** have significant effect on stress level among university students.