Analysis of College Students' psychological Anxiety and Its Causes under COVID-19

Meixue Wu, Hong Zhao*, Yun Guo
College of Computer Science
Nankai University
Tianjin, China
meixuew@mail.nankai.edu.cn, {zhaoh, guoyun}@nankai.edu.cn

***Abstract*—Due to the global outbreak of COVID-19, universities have postponed the commencement of 2020 spring semester and adopted online learning when students and staff are self-quarantined at home. In this paper, SAS(Self-Rating Anxiety Scale) was used to evaluate and analyze students' anxiety in one of the Top 20 universities in China. A total of 3,611 valid questionnaires from 34 provincial administrative regions and Taiwan in China were collected. FP-Growth was used to mine association rules for the scores of each item in SAS. K-Means was adopted to divide all samples into three categories and analyze the scoring characteristics of each SAS item in each category. Based on BiLSTM, a Chinese sentiment analysis model was constructed to analyze the text data. The results show that the average scores of the forward scoring items were relatively high. For the anxious samples, there were fewer arms and legs tingling, arms and legs trembling and syncope. The main causes of anxiety among college students included online learning and epidemic diseases.**

Keywords—Self-Rating Anxiety Scale, correlation analysis, sentiment analysis

I. INTRODUCTION

Affected by COVID-19, the Chinese Ministry of Education issued a notice to postpone the start of spring semester in 2020, colleges and universities across the country began to offer online classes. In other words, students did not return to school but started a new semester through online teaching at home. In this context, the mental health of college students is a problem worthy of attention, and anxiety is the psychological problem that is prone to occur during the epidemic.

Anxiety disorder is a kind of negative emotional experience that occurs when the individual perceives that they are unable to deal with the threat stimulus. The degree of anxiety can be evaluated by using the Self-Rating Anxiety Scale (SAS). SAS was originally developed by W. K. Zung from the School of Medicine of Duke University in 1971 to evaluate the subjective feelings of individuals with anxiety disorders[1]. SAS is a simple assessment tool widely used to analyze subjective anxiety symptoms of adults or adolescents, and it has the functions of screening and diagnosing anxiety levels.

In recent years, there have been plenty of literatures using data analysis techniques to study the mental health of different groups. Husain et al. [2] used Random Forest to predict generalized anxiety disorder of female and reported the accuracy of their prediction results. The established prediction model is expected to provide an effective method for screening generalized anxiety disorder in Malaysian women. Kim et al. [3] collected more than three months' sensor data from 20 elderly people, and used four commonly used classification models including neural network, C4.5 decision tree, Bayesian network and support vector machine to detect their severity of depression. The results showed that it had a good detection effect on both normal people and people with mild depression, with an accuracy of 96%. Tyulyupo et al. [4] developed a special mental health questionnaire and collected the answers of about 200 adolescents aged from 12 to 17 years at 11 rural schools. They used machine learning and data mining methods to automatically assess the happiness according to the respondents' answers, with an accuracy of no less than 90%. Arunditi et al. [5], using four anxiety and depression measurement scales, investigated the relationship between cyberbullying victims in children, and adolescents aged from 10 to 17 years in acute psychiatric inpatients, and found associations between depression, anxiety, self-esteem, self-concept and cyberbullying. Jia et al. [6] conducted a survey to investigate the anxiety status of college students in SARS epidemic area and the influence of coping styles and personality characteristics on their anxiety during the SARS epidemic. The conclusion is that SARS is indeed a stress event that causes anxiety among college students in the epidemic area, but mature coping style, rich psychological knowledge and so on can reduce the psychological anxiety level.

There have been a lot of data analysis on the psychological state of teenagers, college students and patients with various diseases. There're also studies on the anxiety level of college students in certain emergency outbreaks such as SARS. However, it is rarely seen a report on analysing SAS scores of students who will study online at home under emergency outbreak. In the context that COVID-19 have caused a delay of commencement of school year, this paper focuses on the potential association among scores in each item of SAS and the causes of anxiety.

II. METHODOLOGY

A. Data Collection

Undergraduates from one of China's Top20 universities, which started online learning on Feb. 17, were selected as research subjects. In order to study the anxiety state of students before they started online learning, online questionnaires were distributed via social media on February 15th, the deadline to submit the questionnaire was 10 am on February 17th, and finally a total of 3611 valid questionnaires were collected. The questionnaire contents include SAS measurement items, gender, major, grade, and an open question. The data we finally collected through the questionnaire are as follows:

- Scores of SAS

SAS contains 20 items that reflect the subjective feelings of anxiety, and each item is scored on a Likert-type scale from level 1 to 4 according to the frequency of symptom occurrence, where level 1 is a little of the time; level 2 is some of the time; level 3 is good part of the time; level 4 is most of the time. Fifteen of the twenty items are stated with negative words, and the scores corresponding to the above level 1 to 4 are 1, 2, 3 and 4 respectively, in the following we call them as forward scoring items. The other 5 items (No. 5, 9, 13, 17, 19) are stated with positive words and scored reversely, the scores corresponding to the above level 1 to 4 are 4, 3, 2, 1 respectively, in the following we call them as backward scoring items. The scores of 20 items were added up to get rough points (X), after the conversion that multiply rough points by constant 1.25 and take its integer part, we finally get the standard points (Y). In other words, Y is given by:

$$Y = \lfloor 1.25 * X \rfloor \quad (1)$$

According to the results of Chinese norm, the threshold value of SAS standard score is 50 points, among which 50 ~ 59 is mild anxiety, 60 ~ 69 is moderate anxiety, and 69 above is severe anxiety .

- Students' identity characteristics

This part includes participants' gender (male/ female), major (liberal arts/ science) and grade (freshman/ sophomore/ junior/ senior).

- Text data

An open question (Do you have anything else to express about your current physical or psychological situation during this COVID-19 outbreak?)

B. Association Analysis

Association rules mining is a method that frequently applied to find possible associations or connections that lying behind data. FP-Growth algorithm is an effective algorithm to find frequent patterns in data set. Using the Apriori principle, the algorithm only scans data set twice and improves its efficiency considerably [7]. The algorithm we have adopted runs as following steps:

Step 1: For each frequent item, construct its conditional projection database and projection FP-Tree.

Step 2: This process is repeated for each newly constructed FP-Tree until the newly constructed FP-Tree is either empty or contains only one path.

Step 3: When the constructed FP-Tree is empty, its prefix is the frequent pattern. When only one path is included, the frequent pattern is obtained by enumerating all possible combinations and then concatenating with the prefix of FP-Tree.

Related terms:

- Association rule: When both X and Y are itemsets, itemset X and itemset Y have the following relationship for transaction set S: both X and Y are subsets of S, and the intersection of X and Y is empty. An association rule looks like X→Y, where X and Y called antecedent and subsequent of association rule respectively.
- Confidence: Then the confidence of X→Y is the probability of Y occurring simultaneously in the event that X occurs.
- Support: The support of X→Y is the probability of X and Y appear at the same time.
- Strong association rule: Set the minimum confidence and minimum support of association rules. If the confidence and support of rules are not less than the minimum confidence and minimum support, they are called strong association rules.

Among all the collected responses of SAS, score of 1 has taken a relatively large percentage, thus it is not conducive to mining association rules among other scores (2, 3, 4), plus considering the scores from samples that have shown more obvious anxiety characteristics should be paid more attention. Therefore, in this study, FP-Growth algorithm was used to mine the association rules only among the responses from 557 anxiety sample.

C. K-Means Clustering

Cluster analysis can classify data, the data with high similarity can be classified into the same category, while the data with low similarity can be classified into different categories [8]. The measurement of the degree of similarity is the distance between two data elements. K-Means is an iterative cluster analysis algorithm. It's steps are as follows:

Step 1: Select an initial cluster center for each cluster;

Step 2: Allocate the sample set to the nearest cluster according to the minimum distance principle;

Step 3: Update the cluster center using the sample mean of each cluster;

Step 4: Repeat step 2 and step 3 until the cluster center no longer changes;

Step 5: Output the final cluster center and k cluster divisions.

D. Chinese Sentiment Analysis

This study builds a Chinese sentiment analysis model based on BiLSTM (Bi-directional Long Short-Term Memory), which is composed of forward LSTM (Long Short-Term Memory) and backward LSTM. LSTM is well suited for modeling contextual data, such as text data, because of its unique structure characteristics.

As the training set based on existing Chinese text analysis tools has nothing to do with education or psychology, the accuracy when applied directly on our dataset is relatively low. Therefore, we manually annotated the text data collected in part of the questionnaire. The BiLSTM network was used for training, and then the trained model was used to conduct Chinese sentiment analysis on the rest of text data.

III. RESULT

A. Overview of SAS Results

The results of SAS are shown in Table I. In a total of 3,611 students, 84.57 percent were not anxious, 9.91 percent were mildly anxious, and 5.11 percent were moderately and severely anxious.

TABLE I. DEMOGRAPHIC FREQUENCY

Factor		Sum	No Anxiety (%)	Mild Anxiety (%)	Moderate and Severe Anxiety (%)
Gender	*Male*	1454	1234 (84.87%)	123 (8.46%)	97 (6.67%)
	Female	2157	1820 (84.38%)	235 (10.89%)	102 (4.73%)
Major	*Liberal Arts*	1865	1610 (86.33%)	168 (9.01%)	87 (4.67%)
	Sience	1746	1444 (82.70%)	190 (10.88%)	112 (6.41%)
Grade	*freshman*	1754	1510 (86.09%)	153 (8.72%)	91 (5.19%)
	sophomore	1087	895 (82.34%)	129 (11.87%)	63 (5.80%)
	junior	654	556 (85.02%)	59 (9.02%)	39 (5.96%)
	senior	116	93 (80.17%)	17 (14.66%)	6 (5.17%)
Total		3611	3054 (84.57%)	358 (9.91%)	199 (5.11%)

B. *Analysis of SAS Items*

The mean and standard deviation of SAS scores for 20 items are shown in Table II. The average scores of the backward scoring items were larger than others, all of which were greater than or equal to 2.00. Among them, the item with the highest mean was hidrosis, with 645 students (17.9% of the total number) scoring 4 (most of the time). Arms and legs tingling, arms and legs trembling, and syncope were all less than 1.15 on average.

TABLE II. MEAN AND STANDARD DEVIATION OF SAS ITEMS

Item	Mean ± Standard Deviation
Anxiety	1.89±0.84
Fear	1.59±0.77
Panic	1.82±0.85
Crazy feeling	1.35±0.69
*Unfortunately hunch	2.20±0.89
Arms and legs trembling	1.12±0.41
Headache	1.50±0.76
Fatigue	1.80±0.85
*Can't sit still	2.28±0.93
Palpitations	1.32±0.60
Dizziness	1.25±0.56
Syncope	1.12±0.40
*Dyspnea	2.00±1.25
Arms and legs tingling	1.11±0.39
Stomachache and indigestion	1.33±0.66
Urinate frequently	1.44±0.69
*Hidrosis	2.42±1.00
Flush on face	1.28±0.55
*Sleep disorders	2.15±0.96
Nightmare	1.46±0.72

[a.] * is the backward scoring item.

C. *Association Analysis*

Since the number of each scores (1, 2, 3, 4) in anxiety samples was relatively average, give a low minimum support. Strong association rules with a minimum confidence of 0.65 and a minimum support of 0.20 were set. A total of 31 association rules were mined, as shown in Table III. Taking the first association rule as an example, the confidence was 0.81, meaning that 81% of the students who had a little of the time with stomachache and indigestion (1 point) also had a little of the time with arms and legs tingling (1 point); the support was 0.30 means that 30% of the students in anxiety samples encountered stomachache and indigestion for a little of the time, as well as arms and legs tingling for a little of the time. Among the first 26 association rules, the subsequents relates to arms and legs tingling (1 point), arms and legs trembling (1 point), and syncope (1 point). At the same time, in the subsection of this study "B. Analysis of SAS items", the overall mean of the three items(arms and legs tingling , arms and legs trembling, syncope) is the smallest among the twenty items. The samples with a score of 1 in these three items account for 63.6%, 59.8%, and 63.9% of the total number of anxiety samples, so even for anxiety samples, the majority have reported no or little time of occurrences of the three items. The last five association rules show that 72% of the students with good part of the time headache (3 points) were accompanied by good part of the time fatigue (3 points). 66%, 66%, 66%, and 65% of students who some of the time did not have unfortunate hunch (3 points), some of the time palpitations(2 points), some of the time had stomachache and indigestion (2 points), and some of the time had urinate frequently (2 points), respectively, were accompanied by some of the time with can sit still (3 points).

TABLE III. THE RESULTS OF ASSOCIATION RESULTS MINING

Number	Association rules	Confidence	Support
1	1 Stomachache and indigestion-> 1 Arms and legs tingling	0.81	0.30
2	4 Hidrosis->1 Arms and legs tingling	0.65	0.25
3	3 Fear->1 Arms and legs tingling	0.65	0.26
4	1 Syncope->1 Arms and legs tingling	0.74	0.45
5	1 Palpitations-> 1 Arms and legs tingling	0.76	0.22
6	1 Dizziness->1 Arms and legs tingling	0.78	0.27
7	4 *Can't sit still-> 1 Arms and legs tingling	0.67	0.23
8	1 Arms and legs trembling-> 1 Arms and legs tingling	0.77	0.44
9	1 Urinate frequently-> 1 Arms and legs tingling	0.81	0.29
10	1 Flush on face-> 1 Arms and legs tingling	0.76	0.30
11	1 Nightmare->1Arms and legs tingling	0.77	0.20
12	1 Syncope->1Arms and legs trembling	0.68	0.42
13	1 Palpitations-> 1 Arms and legs trembling	0.82	0.24

14	1Dizziness->1Arms and legs trembling	0.74	0.25
15	1 Urinate frequently->1 Arms and legs trembling	0.71	0.25
16	1 Stomachache and indigestion->1 Arms and legs trembling	0.65	0.24
17	1 Flush on face->1 Arms and legs trembling	0.68	0.27
18	4 Sleep disorders->1 Syncope	0.65	0.22
19	4 *Can't sit still->1 Syncope	0.67	0.23
20	4 Hidrosis->1 Syncope	0.66	0.25
21	3 Fear->1 Syncope	0.69	0.27
22	1 Palpitations->1 Syncope	0.74	0.21
23	1 Dizziness->1 Syncope	0.89	0.30
24	1 Urinate frequently->1 Syncope	0.74	0.26
25	1 Flush on face->1 Syncope	0.75	0.30
26	1 Arms and legs tingling->1 Syncope	0.75	0.45
27	3Headache->3Fatigue	0.72	0.22
28	3 *Unfortunate hunch->3 *Can't sit still	0.66	0.35
29	2 Palpitations ->3 *Can't sit still	0.66	0.21
30	2 Stomachache and indigestion->3 *Can't sit still	0.66	0.21
31	2 Urinate frequently->3 *Can't sit still	0.65	0.23

[b] * is the backward scoring item.

D. Cluster Analysis

The K-Means algorithm was used to perform cluster analysis on the scores of 20 items, take K=3. The clustering results are shown in Fig. 1.

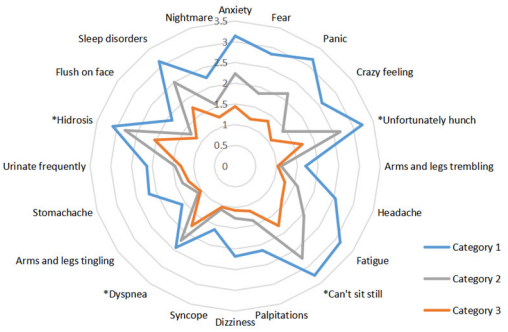

Fig. 1. The results of cluster analysis

Overall, the relative high and low of 20 items' scores remain consistent among the three different categories of students. The scores in category 1 were higher than those of the other two categories, especially the scores of fatigue, afraid, panic and crazy feeling. In category 2 and category 3, the scores of syncope, arms and legs trembling, arms and legs tingling, flush on face, and urination frequent were basically the same. Category 2 were higher than category 3 on the scores of other items. Each category accounted for 9.8%, 36.3% and 53.9% of the total sample size, respectively.

E. Text Analysis

In the responses to open question (Do you have anything else to express about your current physical or psychological situation during this COVID-19 outbreak?) in 3611 questionnaires, a total of 654 valid answers were collected. In this study, 200 answers were manually marked in the sentiment analysis, and the labels were divided into positive and negative directions. The Chinese text word segmentation tool "JIEBA" was used to construct the data set which containing 200 manually marked answers, and 80% of the data set was divided into the train set and 20% into the test set. Python3.6 was used to build the BiLSTM network. The learning rate was set at 0.01 and the number of iterations was 100. The constructed model were applied to the remaining 454 answers for sentiment analysis. In the total dataset of 654 answers, 409 were marked as positive and 225 were negative.

The topics of the answers were summarized, among which positive topics include: looking forward to go back to school; hope the epidemic will end soon; while negative topics include: operational inconvenience and concern to the online learning; anxiety about the epidemic; worry about papers and exams.

IV. Conclusions And Discussion

This study used the SAS to analyze the anxiety situation of Chinese undergraduate students before they adopted online learning at home during the COVID -19 pandemic. In the analysis process, quantitative analysis was used to analyze the numerical data, and sentiment analysis was used to analyze the text data.

According to the analysis of the scores of each SAS items, the average scores of the forward scoring items were relatively high, which may be related to the way that questionnaire questions are described.

Association analysis were conducted on the 20 SAS items of the anxiety sample, the results showed that even for the anxiety sample, most students did not have or rarely encountered arms and legs tingling, arms and legs trembling, and syncope. This is probably caused by the fact that the tested population is college students aged about twenty years old, so even for the anxious group, such physical reactions were rarely encountered. Anxiety samples who some of the time did not have unfortunate hunch, some of the time had stomachache and indigestion, and some of the time urinate frequently were most likely accompanied by some of the time can sit still.

Cluster analysis was conducted on the individual scores of 20 SAS items, and all the students were divided into three categories. The relative high and low in the scores of 20 items among three categories of students were basically the same, but the anxiety levels were different.

The results of text analysis showed that under this special background, 62.54% of the students expressed a positive attitude, and they had the desire to go back to school, but 37.46% of the students had some concerns about online learning, and some of them were worried that the original study plan could not be completed on time.

In this study, the SAS items socores of students were analyzed under the special background of the outbreak of COVID-19, which have caused the delay of commencement of 2020 spring semester. In addition to the analysis of numerical data, text data was collected for the establishment of Chinese sentiment analysis model. In the future, the sample range can be expanded to include a larger number of students from multiple universities, and the questionnaire content can be enriched to collect data on other aspects of psychological states such as depression. Meanwhile, the behavioral data of

students can be collected, and the mental health status of students can be modeled by machine learning.

ACKNOWLEDGMENT

The authors gratefully acknowledge the research funding from the following programs:

- The Key Cultivation Project of Education Achievement Award of Tianjin (PYGJ-018)
- Teaching Reform of Nankai University(NKJG2020150，NKJG2020225)

REFERENCES

[1] W W, Zung, “A rating instrument for anxiety disorders,” Psychosomatics, vol. 12, no. 6, pp. 371-379, 1971.

[2] W. Husain, L. K. Xin, N. A. Rashid and N. Jothi, "Predicting Generalized Anxiety Disorder among women using random forest approach," 2016 3rd International Conference on Computer and Information Sciences (ICCOINS), Kuala Lumpur, pp. 37-42, 2016.

[3] Kim, Jungyoon and Liu, Na and Tan, Hwee-Xian and Chu, Chao, “Unobtrusive Monitoring to Detect Depression for Elderly with Chronic Illnesses,” IEEE Sensors Journal, pp. 1-1, 2017.

[4] S. V. Tyulyupo, A. A. Andrakhanov, B. A. Dashieva and A. V. Tyryshkin, "Adolescents Psychological Well-Being Estimation Based on a Data Mining Algorithm," 2018 IEEE 13th International Scientific and Technical Conference on Computer Sciences and Information Technologies (CSIT), Lviv, pp. 475-478, 2018.

[5] B, Arunditi, “The Relationship between Social Media Bullying and the Adolescent's Self Report of Emotional Health: A Study of Urban Youth on an Acute Inpatient Psychiatric Unit,” Journal of Depression and Anxiety, vol. 04, 2015.

[6] Jia Ning, Lu Zhongyi, “Study on anxiety of college students in SARS epidemic area,” Journal of Hebei Normal University. vol. 04, pp. 740–741, 2003.

[7] M. Hossain, A. H. M. S. Sattar and M. K. Paul, "Market Basket Analysis Using Apriori and FP Growth Algorithm," 2019 22nd International Conference on Computer and Information Technology (ICCIT), Dhaka, Bangladesh, pp. 1-6, 2019.

[8] Yu Qilin. “Optimization of initial clustering center selection for K-means algorithm,” Journal of Computer System Applications, vol. 26, pp. 170-174, 2017.

APPENDIX: ONLINE SURVEY

A. Self-Rating Anxiety Scale

The following 20 items have the same options:

1. A Little Of The Time
2. Some Of The Time
3. Good Part Of The Time
4. Most Of The Time

* refer to backward scoring item.

Q1. I feel more nervous and anxious than usual.

Q2. I feel afraid for no reason at all.

Q3. I get upset easily or feel panicky.

Q4. I feel like I’m falling apart and going to pieces.

Q5*. I feel that everything is all right and nothing bad will happen.

Q6. My arms and legs shake and tremble.

Q7. I am bothered by headaches neck and back pain.

Q8. I feel weak and get tired easily.

Q9*. I feel calm and can sit still easily.

Q10. I can feel my heart beating fast.

Q11. I am bothered by dizzy spells.

Q12. I have fainting spells or feel like it.

Q13*. I can breathe in and out easily.

Q14. I get numbness and tingling in my fingers and toes.

Q15. I am bothered by stomach aches or indigestion.

Q16. I have to empty my bladder often.

Q17*. My hands are usually dry and warm.

Q18. My face gets hot and blushes.

Q19*. I fall asleep easily and get a good night’s rest.

Q20. I have nightmares.

B. Student identity information selection

Q21. Your gender.
1. Male 2. Female
Q22. Your grade.
1. Freshman 2. Sophomore 3. Junior 4. Senior
Q23. Your major.
1. Science 2. Liberal arts

C. Open-ended question

Q24. Do you have anything else to express about your current physical or psychological situation during this COVID-19 outbreak?