Correlation Analysis of Internet Addiction with Daily Behavior: A Data-Driven Method

Xingxu Zhang
Changchun University of Science and Technology
No.7186 Weixing Road, Changchun, Jilin Province, China
zhangxingxu1105@gmail.com

Yong Ma
Changchun University of Science and Technology
No.7186 Weixing Road, Changchun, Jilin Province, China
my15754315108@gmail.com

Xiaoqiang Di*
Changchun University of Science and Technology
No.7186 Weixing Road, Changchun, Jilin Province, China
dixiaoqiang@cust.edu.cn

Huamin Yang
Changchun University of Science and Technology
No.7186 Weixing Road, Changchun, Jilin Province, China
yhm@cust.edu.cn

Binbin Cai
Changchun University of Science and Technology
No.7186 Weixing Road, Changchun, Jilin Province, China
cbb@cust.edu.cn

ABSTRACT
Internet addiction refers to excessive Internet use in daily life. Its negative impact on college students calls for timely discovery of its reasons and correct guidance. However, present research methods are mainly based on the questionnaire, which can be affected by non-randomly selected samples and a low response rate. Thanks to the development of the smart campus, students' behavior can be recorded as data, thus whether Internet addiction is correlated with daily behavior can be analyzed quantitatively. In this paper, we extracted five features and proposed an Internet Addiction Rating Scale (IARS) to quantify students' Internet addiction level based on the Internet login data, and found that Internet addiction is positively correlated with consumption amount, negatively correlated with consumption speed and academic performance, but has little relationship with self-discipline. Our work throws some light on the relationship between students' Internet addiction and daily behavior through unobtrusive data analysis, helping college management staff detect abnormal learning status timely and reasonably.

CCS Concepts
• **Information systems→Information systems applications →Data Mining.**

Keywords
Internet Addiction; Campus Data Mining; Correlation Analysis.

1. INTRODUCTION
Internet has a profound influence on daily life. However, many studies have suggested addictive use of Internet. Internet Addiction (IA) refers to the inability to control one's use of Internet, leading to psychological, social, school and/or work difficulties [1][2]. Due to the multiple psychosocial stressors and inadequate ability to find proper personal resources or experience to cope with it, college students are often vulnerable to IA. Chi et al. found that students with poorer parent-child relationships, higher levels of depression, and lower levels of psychosocial competence are more likely to be Internet-addicted [3]. On the other hand, Baturay et al. revealed that game addiction, neglecting daily chores and bad relationship with professors are all significantly correlated with Internet Addiction [4]. Therefore, over the past few years, Internet addiction of college students has become a serious worldwide concern [5].

At present, studies related to the Internet addiction are mainly concentrated on psychological fields, where collection and analysis of data are mostly based on questionnaires or self-reports. According to Blachnio et al. [6], self-reported assessment and non-randomly selected samples are the two main hindrances of questionnaire methods. In recent years, thanks to the rapid development of smart campus, the consumption data, Internet access data as well as students' academic performance data are all available for quantitative analysis of their Internet addiction level.

Among previous related researches, Young's Internet Addiction Test is immensely popular and widely used, which consists of 20 items and each item is rated on the Likert scale [6]. In this paper, partially inspired by that, we first extracted five measures from students' Internet login data to represent their Internet addiction level. Then, data for each of the five measures were sorted and clustered into five groups respectively, with each measure in each cluster given a score of 1 to 5 based on data range. After that, for each student, the scores of each measure were added up to obtain students' Internet Addiction Level (IAL). Next, correlation of students' IAL with four features were analyzed, i.e., average money consumption per meal, average times for using up 100-yuan, GPA and the earliest time for using smart-card on average. The first two measures are used for representing the consumption level, GPA for academic level, and the last one as self-discipline level. All the data is collected unobtrusively, so time and resources are saved, and the potential casual answering and low response rate of questionnaires can be avoided. In addition, the related participants cover all of undergraduates in the university, so results will be more accurate. It is found that IAL is negatively correlated with academic performance, positively correlated to

ICBDT 2020, September 18–20, 2020, Qingdao, China

ACM ISBN 978-1-4503-8785-9/20/09…$15.00
https://doi.org/10.1145/3422713.3422716

consumption amount, but negatively correlated to consumption speed, yet has no significant relationship with students' first consumption time. These results are partially in line with previous conclusions from psychological field, which prove the effectiveness of our method.

The rest of this paper is organized as follows. Related work on Internet addiction and campus data mining is introduced in Section 2. Our proposed Internet Addiction Rating Scale is described in Section 3. Correlation analysis is performed in Section 4 using the measure we proposed. Section 5 concludes the paper.

2. RELATED WORK

2.1 Previous Works on Internet Addiction

Previous studies have shown the relationships between Internet addiction and psychological traits. Researchers have used measures like the Online Cognition Scale [7], the Internet Related Problem Scale [8], or the Problematic Internet Use Questionnaire [9]. By utilizing these methods, researchers have found that compulsive Internet use lowers adolescents' or college students' life satisfaction and academic performance [10]. Also, Internet addiction has been identified to be linked to academic stress and negative emotions [11]. However, such studies are limited by self-report assessment method and non-randomly sampled data, as well as low response rate. This facilitates the use of data mining method on such a topic.

2.2 Previous Works on Campus Data Mining

Recent development of smart campus makes the daily records of students' smart card increasingly important in mining their behavior, especially its correlation with academic performance. Cao et al. [12] proposed orderliness and diligence based on behavioral records, finding that regularity of campus daily life is significantly correlated with orderliness. However, Internet access data were not taken into account, and it is still to know how surfing the Internet will affect students' academic performance. Wang et al. [13] concentrated on the analysis of consumption records, water usage, dormitory electricity, Internet usage, etc. and found a changing trend in students' learning career as well as different living habits of different genders. Although Internet usage data are considered, the correlation of different behaviors were not studied. Peng et al. [14] considered students' Internet addiction as a hidden variable affecting their daily time online and other behavior. However, excessive online time may just be one aspect of Internet addiction. Therefore, in this paper, we extracted other features from the Internet login data, trying to give a more comprehensive representation of students' Internet addiction level (IAL).

3. METHODS

3.1 Basic Assumption

Internet addiction is basically a psychological disorder, and many factors account for this phenomenon. Therefore, it is difficult to propose a fully comprehensive measure for it. Here, from the perspective of Internet login data, we propose the following three assumptions:

Assumption 1: Internet-addicted students use more-than-average Internet downflow traffic.

Downflow traffic refers to the amount of data used for downloading online resources on the Internet. College students with high IAL may spend more time watching online TV programs, TV series and various kinds of movies, since these behaviors are characterized by large amount of livestreaming data downloaded to the computer.

Assumption 2: Internet-addicted students spend more time online, both at the weekends and on the workdays.

Internet downflow traffic can only partially demonstrate the Internet addiction level. Some students do not like watching online videos, yet they may enjoy playing online games, chatting with friends for long hours, etc. In addition, the online time for weekends and workdays may also differ. Therefore, the above two online time are also a reflection of Internet addiction.

Assumption 3: Internet-addicted students have a much earlier login time and much later logout time than normal students.

Login time refers to the time that students log into the campus network and start to use it. Similarly, logout time represents the time that students log out from the network and stop using. Students suffering from Internet addiction may lose control of their own behavior and unconsciously login once they get a chance. Therefore, they may have more login times than others, and even login very early in the morning or logout very late in the night. Such features are also expressions of Internet addiction.

3.2 Data Description

The dataset used in this paper is exported and integrated from the smart card system of a university in China. Since some data involves privacy, information like student ID is anonymized before use. It consists of three parts, with the time range from Jan.1st, 2018 and Dec.31st, 2018, including students' consumption records in the canteen (1603540 entries), the GPA for the spring and autumn semesters (291880 entries), and students' Internet login records (5150025 entries). The students selected were enrolled in 2015, 2016 or 2017. Not all students have all of the three records, so after combining different data tables and removing error data, there are records of 4188 students remained for experiment.

3.3 Measures of Internet Addiction

Based on the above three assumptions, the following five features are extracted from the Internet login data to quantitatively measure students' IAL.

3.3.1 Total downflow

Each entry of the original Internet login data contains a feature recording the amount of downflow traffic during this login. Therefore, the total downflow is considered. Specifically, for each student, for each day in each of the 52 weeks in the year, if there are records on that day, the downflow traffic of each login is added up.

3.3.2 Total online time on weekends and workdays

Similarly, using the total online time, we divided the original Internet login data into two groups. For each student, for each day in each of the 52 weeks in the year, his/her online time on Saturday and Sunday is added up as total weekend online time, while the online time from Monday to Friday is summed up as total workday online time.

3.3.3 Average earliest login time and average latest logout time

Generally, Internet-addicted students may have a greater number of login times. However, the unstable and ever-changing Internet environment may cause accidental connection lost. When this occurs, the reconnection will also be recorded as one entry, so the number of logins cannot directly reflect the Internet addiction level. Instead, we choose the average earliest login time and average latest logout time for each student in a year as the last two features.

3.4 Internet Addiction Rating Scale

Internet Addiction Test (IAT) is a reliable and valid measure of addictive use of Internet, developed by Dr. Kimberly Young [15]. It consists of 20 items that measures mild, moderate and severe level of Internet addiction, and each item is a question to be answered using a scale ranging from 0, 1, 2…to 5. A larger number corresponds to a higher IAL, and each question reflects the influence of Internet on a person's daily life. For example, "How often do you find that you stay on-line longer than you intended?", "How often do your grades or school work suffers because of the amount of time you spend on-line?", "How often do you lose sleep due to late-night log-ins?", etc.

Totaling up the scores for each item, the higher the score, the greater level of addiction is. Specifically, a person with 20-49 points is an average online user, 50-79 points means experiencing occasional or frequent problems because of the Internet, and 80-100 points means that Internet usage is causing significant problems in the life.

Inspired by such an idea, we proposed an Internet Addiction Rating Scale (IARS) based on the features we extracted previously. Each feature is considered as a question. The scales used are similar to IAT, which are 1 to 5, but the scores are rated based on clustering the data for each feature. The basic rating procedure is presented as follows:

Input data: Data for five IAL-related measures of each student, including total downflow, total online time on weekends and workdays, average earliest login time and average latest logout time.

Output data: The score of each measure for each student.

(1) Extract anonymized student ID array from input data.

(2) For each of the five types of measures, sort the data of that measure ascendingly.

(3) Cluster the sorted data of each measure into 5 groups using K-means according to their respective data range.

(4) For each student i, each measure j and each group k, compare each feature of student i with the data range of group k of measure j. If data of measure j falls into group k, then set the score of measure j of student i as k.

(5) Output the score of each measure for each student.

After such a process, each student will be given a respective score for each IAL-related measure. Table 1 is a sample of the processed results. The entire processed result is called the Internet Addiction Rating Scale (IARS), where fw, wd, wk, early and late represent total downflow, total weekend online time, total workday online time, average earliest login time and average latest logout time respectively. ID is the anonymized student number, and the last column shows the total Internet Addiction score of each student, which is a representation of students' IAL for subsequent analysis.

Table 1. The Internet Addiction Rating Scale

ID	fw	wd	wk	early	late	total
A	1	1	1	3	3	9
B	1	1	1	2	2	7
C	4	3	1	5	5	18
D	2	1	1	3	3	10
E	3	2	1	3	3	12

3.5 Measures of Other Behavioral Features

With the help of smart card system, many behaviors of students can be recorded as data for analysis. In this paper, we selected features related to students' consumption level, academic performance and self-discipline to analyze their relationship with IAL.

3.5.1 Consumption Level

Since breakfast is served at a low price in our university, its consumption cannot fully reflect students' economic status. Thus, we calculated average lunch and dinner consumption in the canteen during the year. In addition, we consider that students in financial hardship tend to spend money slower, so we also calculated the average number of times that each student uses up 100 yuan during the year. These two measures are combined to denote students' consumption level.

3.5.2 Academic Performance Level

The time range selected for this paper contains two semesters, autumn and spring. For each student, we figured out the average grade for all the courses attended for each semester, and then calculated the average score of these two semesters to express students' academic performance, which is named as GPA as follows:

$$GPA = \frac{\sum_{i=1}^{n} grade_i}{n} \tag{1}$$

Here, $grade_i$ is the score of each course of the two semesters, n is the total number of courses in the two semesters.

3.5.3 Self-discipline

According to Nalwa et al., Internet dependents cannot fall asleep due to late-night logons or often delay work to spend time online [16]. Therefore, we consider that students with Internet addiction are inclined to get up much later, leading to a much later time of their first consumption at the canteen. Based on such intuition, we figured out the average time of first consumption record in the year to show whether a student is self-disciplined.

3.6 Final Feature Score Table

After quantifying measures for students' Internet addiction level, consumption level, academic performance and self-discipline, a feature score array for each student is obtained. Table 2 shows the feature table of the same students we used to describe Internet addiction level.

Here, Net represents students' IAL, Grade shows their academic performance level, Spd and Amt indicates their consumption level and Time indicates their self-discipline level. Experimental results of correlation will be analyzed in the following section in detail.

Table 2. Final Feature Table for Students

ID	Net	Grade	Spd	Amt	Time
A	9	80.709	6.659	15.017	7:51
B	7	82.600	10.084	9.916	9:33
C	18	87.272	11.561	8.650	9:42
D	10	87.836	8.118	12.318	8:05
E	12	88.431	10.434	9.584	9:33

4. CORRELATION ANALYSIS

To exploit the relations between students' IAL and their behavioral features, we perform correlation analysis using the Pearson correlation coefficient r. Calculating r between different pairs of feature values, a correlation matrix is obtained in Table 3.

Table 3. Pearson r Between Different Behavioral Features

	Grade	Net	Amt	Spd	Time
Grade	1	-0.870*	-0.808	0.859*	-0.730
Net	-0.870*	1	0.889*	-0.852*	0.580
Amt	-0.808	0.889*	1	-0.615	0.692
Spd	0.859*	-0.852*	-0.615	1	-0.632
Time	-0.730	0.580	0.692	-0.632	1

*Correlation is significant at the 0.05 level.

In Table 3, 2-tailed test of significance is used, where correlation is significant at the 0.05 level. Values in the second row show that IAL is correlated with academic performance, consumption speed and consumption amount with the statistical significance. This indicates the effectiveness of our extracted features and calculation. Below is the detailed correlation analysis for the behavioral features.

4.1 Correlation of IAL with Academic Level

Firstly, the relationship between IAL and academic performance is analyzed, where binned statistics are used to aggregate all the data points. The whole dataset is divided into 6 bins, each of which contains the same number of data points. Seen from Table 3, IAL is negatively correlated with academic performance, with a Pearson r of -0.870. This means that the more a student becomes Internet addicted, the lower GPA he (she) will obtain. To further explore this phenomenon, Figure 1 is plotted to reflect the changing trend of GPA and IAL.

As is shown in Fig.1, with the increase of IAL, students' GPA generally follows a downward trend. This proved our conclusion from the Pearson r that excessive use of Internet has a negative effect on academic performance, which is in consistent with the conclusions from previous studies in psychological field [17]. However, a little difference emerges when the Internet Addiction level is between about 5 and 6, where GPA rises from 78 to 80. After repeated thinking, the following reason may account for this:

Students who seldom use Internet may not enjoy the merits of the Internet like providing online videos and courses. These courses serve as supplements of in-class learning materials, providing extra knowledge for students to learn freely by interest. Therefore, this new form of learning actually widens students' vision and increasing their academic level. In conclusion, proper Internet use is helpful for students' learning, but excessive use has the opposite effect.

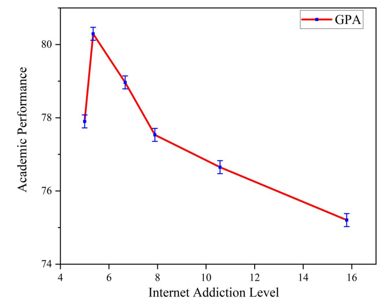

Figure 1. Correlation of IAL with Academic performance.

4.2 Correlation of IAL with Consumption

In this section, the relationship between students' IAL and their consumption level are analyzed. The consumption level constitutes two features: consumption amount and consumption speed. Pearson coefficients for these two features are 0.889 ($p < 0.05$) and -0.852 ($p < 0.05$), which shows the statistical significance.

Judging from the two coefficients, IAL is positively correlated to consumption amount, but negatively correlated to consumption speed. This means that the more a student is Internet-dependent, the more money he/she will spend for a meal on average, but the fewer times he/she will use to spend up 100 yuan. The following two reasons may account for this phenomenon.

Firstly, Internet addicted students may be liberal with their money, without the need of worrying about their living expenses. Secondly, they may suffer from some psychological disorders. For example, they cannot control their appetite for spending money, or make consumption lavishly because of loneliness or painful emotions. The root cause for these needs further research in the future.

We also studied how the consumption level varies with the increase of IAL in Fig.2. As is shown in Fig.2, when the Internet addiction level increases, the consumption amount generally increases, but some fluctuation arises when the level rises from 5 to about 6. This is because the students selected for this research come from different campuses, whose canteen serves food with a slightly different price, resulting in a slight difference of the average consumption amount. However, after that, the consumption amount goes ascendingly, suggesting a positive correlation with the IAL.

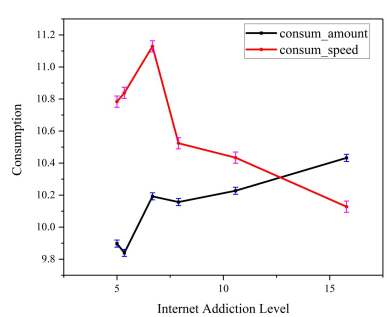

Figure 2. Correlation of IAL with consumption level.

On the other hand, the consumption speed firstly increases with the rising of IAL. At this time, we can see the average consumption amount also rises, and both of these two values are approximately between 10 and 11. This is also in accordance with our definition for consumption speed--the product of the amount and speed basically approaches 100. In other words, the consumption amount is approximately inverse proportional to the consumption speed. Therefore, their changing trend are opposite. To the best of our knowledge, this is the first work on the correlation between Internet addiction and consumption, so such findings will lay some foundations for future work in both data science and psychological field.

4.3 Correlation of IAL with Self-discipline

In previous sections, we hypothesized that Internet addicted students prone to get up much earlier than those who are not. To prove this assumption, we studied how the IAL correlated with the first consumption time on average, which indirectly reflects the time of getting up.

Seen from the correlation matrix, Pearson coefficient for these two features is 0.580, suggesting only a weak correlation, and the corresponding p-value is 0.22 without statistical significance. Therefore, we cannot conclude that IAL is significantly correlated to students' first consumption time. In other words, the first consumption time itself is not the only indicator for self-discipline. It is more of a comprehensive psychological factor than a quantitative measure. In future, we will make further research on this issue, and try to extract more behavioral features to find how self-discipline and orderliness are correlated with Internet addiction.

5. CONCLUSION

In this paper, we carried out correlation analysis between students' Internet addiction level with academic performance, average consumption amount, average consumption speed, and their self-discipline level. Internet addiction level is a comprehensive indicator and hard to quantify, so we selected five measures indicating its degree based on previous psychological research, i.e., total downflow traffic, total online time on weekends, total online time on workdays, average earliest login time and average latest logout time. Then, we proposed an Internet Addiction Rating Scale (IARS) to calculate each student's Internet Addiction Level (IAL) based on the five measures. Next, each IAL score is correlated with students' GPA, average consumption amount, average consumption speed and time of first consumption. Correlation analysis shows that IAL is negatively correlated with academic performance, positively correlated to consumption amount, but negatively correlated to consumption speed. Besides, self-discipline has no significant correlation with IAL. Our analysis results are consistent with some previous findings in the psychological field, which verify the effectiveness of our proposed measures and correlation results.

We consider that the measure we extracted for self-discipline may not proper enough to represent it, and records from campus forums and search logs are also required to get more accurate results, since students may study online courses for hours, and such behavior can not merely be detected by the measures we selected. In the future, we will dive deeper into how Internet addiction is related with self-discipline based on behavior data and try to use students' geographical behavior trace as well as their social network. In addition, other features related to students' daily behavior will be considered to study how Internet addiction influence students' campus life.

6. ACKNOWLEDGMENTS

This work is supported by the Higher Education Research Project of Jilin Province (Grant No. ZD18027).

7. REFERENCES

[1] Marcantonio, and M.Spada, "An overview of problematic Internet use," *Addictive Behaviors,* vol. 39, no. 1, pp. 3-6, 2014.

[2] K. C. Man, and S. P. M. Law, "Factor structure for Young's Internet Addiction Test: A confirmatory study," *Computers in Human Behavior,* vol. 24, no. 6, pp. 2597-2619, 2008.

[3] X. Chi, L. Lin, and P. Zhang, "Internet Addiction Among College Students in China: Prevalence and Psychosocial Correlates," *Cyberpsychology Behavior Social Networking,* vol. 19, no. 9, pp. 567-573, 2016.

[4] M. H. Baturay, and S. Toker, "Internet addiction among college students: Some causes and effects," *Education information technologies,* vol. 24, no. 5, pp. 2863-2885, 2019.

[5] W. Peng, D. Li, D. Li, J. Jia, Y. Wang, and W. Sun, "School Disconnectedness and Adolescent Internet Addiction: Mediation by Self-Esteem and Moderation by Emotional Intelligence," *Computers in Human Behavior*, 2019.

[6] A. Błachnio, A. Przepiórka, O. Gorbaniuk, M. Benvenuti, A. M. Ciobanu, E. Senol-Durak, M. Durak, M. N. Giannakos, E. Mazzoni, I. O. J. C. B. Pappas, and S. Networking, "Cultural Correlates of Internet Addiction," *Cyberpsychology Behavior Social Networking,* vol. 22, no. 4, pp. 258-263, 2019.

[7] R. A. Davis, G. L. Flett, and A. Besser, "Validation of a New Scale for Measuring Problematic Internet Use: Implications for Pre-employment Screening," *Cyberpsychology & Behavior,* vol. 5, no. 4, pp. 331-345, 2002.

[8] L. Armstrong, J. G. Phillips, and L. L. Saling, "Potential determinants of heavier internet usage," *International Journal of Human Computer Studies,* vol. 53, no. 4, pp. 537-550, 2000.

[9] Z. Demetrovics, B. Szeredi, and S. B. R. M. Rózsa, "The three-factor model of Internet addiction: The development of the Problematic Internet Use Questionnaire," *Behavior Research Methods,* vol. 40, no. 2, pp. 563-574, 2008.

[10] A. Dhir, S. Chen, and M. Nieminen, "Psychometric Validation of the Compulsive Internet Use Scale," *Social Science Computer Review,* vol. 34, no. 2, pp. 197-214, 2016.

[11] S. Jun, and E. Choi, "Academic stress and Internet addiction from general strain theory framework," *Computers in Human Behavior,* vol. 49, no. aug., pp. 282-287, 2015.

[12] Y. Cao, J. Gao, D. Lian, Z. Rong, J. Shi, Q. Wang, Y. Wu, H. Yao, and T. Zhou, "Orderliness predicts academic performance: behavioural analysis on campus lifestyle," *Journal of the Royal Society Interface,* vol. 15, no. 146, 2018.

[13] B. Wang, K. Deng, W. Wei, S. Zhang, W. Zhou, and Y. Shui, "Full Cycle Campus Life of College Students: A Big Data Case in China." In 2018 IEEE International Conference on Big Data and Smart Computing (BigComp). IEEE, 2018, pp. 507-512.

[14] W. Peng, X. Zhang, and X. Li, "Using behavior data to predict the internet addiction of college students." In International Conference on Web Information Systems and Applications. Springer, 2019, pp. 151-162.

[15] K. S. Young, "Internet addiction: symptoms, evaluation and treatment," *Innovations in clinical practice: A source book,* vol. 17, no. 17, pp. 351-352, 1999.

[16] K. Nalwa, and A. P. Anand, "Internet Addiction in Students: A Cause of Concern," *Cyberpsychology & Behavior,* vol. 6, no. 6, pp. p.653-656, 2003.

[17] R. W. Kubey, M. J. Lavin, and J. R. Barrows, "Internet Use and Collegiate Academic Performance Decrements: Early Findings," Journal of Communication, vol. 51, no. 2, pp. 366-382, 2001.