The Relationship between Internet Usage Behavior and Health Problems

Rozita Jamili Oskouei
Motilal Nehru National Institute of Technology
Allahabad, UP, India

Rahul Sharma
Motilal Nehru National Institute of Technology
Allahabad, UP, India

ABSTRACT

World Wide Web (WWW) has huge number of Web pages with various types of information and has become one of the main resources of knowledge nowadays. Since more than two billion people currently use Internet and Social network Websites, therefore exploration of different effects of Internet usage on various aspects of life of these users is very essential for both business owners and academic advisors.

Outlier detection is an area of research with a long history which has applications in many fields. However, there is no specific study to identify outliers in students' community based on their Internet usage behaviors.

Major contributions of this research are:

- The summary of Internet usage behavior of students (such as the average [time spent on Internet, number of visited Web pages] and the most popular visited Websites) is presented.

- A group of users whose usage behavior is perceptible differences with others is identified and introduced as Outliers.

- The relationships between Internet usage behavior and medical problems in both normal users and Outliers have been studied.

General Terms

Medical Problems, Internet Usage patterns

Keywords

Internet Usage Behavior (IUB), category of visited Websites, Social Networking (SN), Academic Performance, outliers.

1. INTRODUCTION

World Wide Web (WWW) has huge number of Web pages with various types of information and it has become one of the main resources of knowledge nowadays. It is estimated, based on several surveys [22~25], that there are approximately 172 million active Websites, 550 billion documents on Internet and around two billion Internet users. The growing use of computing and communication technology has both good and bad effects [2 ~ 7]. Kraut and his colleagues [2, 3] claimed that greater use of the Internet was associated with negative effects on individuals. Also, many quantitative studies confirmed that loneliness was associated with increase on Internet usage [4 ~ 7].

Extracting the users' behavior and its variation on different situation or conditions is desirable for business as well as academic researchers. In academic environments, the main goal is improving students' knowledge and academic performance. Since students use the Internet and it has become a main resource of information for them and each student spends few minutes per day in Internet to visit different Websites, therefore it is very important to know what is the effect of this usage on their personal life as well as their academic performance and activities [9 and 10].

This study attempts to show briefly the usage patterns of students (as users) and identify the group of users whose usage pattern has significant difference with others and named them as Outliers. Finally explore the relationships between students' health problems and variation on their Internet usage behaviors during period of illness such as the changes which might happen on the time spent on Internet or category of visited Websites specially Social Network (SN) and Entertainment (EN) Websites. In the other word, determine whether the disease has caused any changes in users' Internet usage patterns, or not?

This paper is organized as follows: Section 2nd, covers related works. Section 3rd presents our data collections and pre-processing steps. Section 4th presents users' gender based Internet usage behaviors and the changes which are happening during the period of illness. Section 5th introduces Outliers and our proposed method for identifying Outliers among students and comparison between Outliers' behaviors with other students. Section 6th concludes the paper by summarizing our results and future issues.

2. RELATED WORKS

Web usage mining aims to discover interesting and frequent user access patterns from web usage data. In this section, we survey recent studies undertaken to model individual and group behavior of Internet users and to evaluate their differential usage pattern for different services and effects of this usage of their different activities or behaviors.

Several researchers [11 ~ 16] have attempted to explore users Internet usage behaviors. Liccardi [16] studied the role of social networks in students' learning experiences and concluded that social networking place a positive role in students' learning experiences. Some of the researchers [17 ~ 19] studied the impacts of Internet usages on students' academic performances. Awais [18] explored the impact of Internet usage on students' academic performance. Several researchers [20 ~ 27] studied gender based effects of information technology in particular on education. Jo Sander [21] revealed that woman is significantly underrepresented in Information and Communication Technology in most of countries. Kim and Chang [23] study has concluded that computer usage have differential effects on academic performance of users from the immigrant and gender groups. Fisher [24] reported that users of different ages are having different behaviors and feelings even though they use similar software. Braten [27] showed, males reported higher levels of participation in Internet-based communication activities than females.

Several research efforts [7 and 28~ 31] have been directed to extract the relationships between Internet usage behaviors and users' different activities and loneliness. Erdoğan [7] studied the Internet attitudes and loneliness of students on Turkish

adolescents. Result was revealed that Turkish adolescents' loneliness was associated with both increased Internet usage and Internet attitudes. Tosun & Lajunen [28] have found that psychoticism individuals tend to use the Internet as a substitution for face-to-face communication. The other article [29] investigated the relationship between loneliness and the use of Internet for entertainment, and found that lonely people tend to use Internet for entertainment. Figl [31] found those students who do not have a supporting network generally know only fewer friends.

3. DATA COLLECTION AND PREPROCESSING

Motilal Nehru National Institute of Technology (MNNIT) Allahabad has a decentralized computing environment. Each academic and administrative department/section has own computing facilities in addition to a computer center, which is a central facility. The institute has approximately 1800 computing nodes distributed all over campus and connected through optical fiber backbone including hostels and residential areas. The medical center of Motilal Nehru National Institute of Technology (MNNIT) provides services all working days from 7:30 A.M. to 6.00 P.M.

This research is made based on the data collected from:

- Proxy Server access log files of Motilal Nehru National Institute of Technology (MNNIT) Allahabad, India for a period of 30 months continually. The computer center also provided user identification information, which includes user-id, full-name and department in one text file.

- The data collected from dean (Academic Affairs) office including general and academic information of students such as Registration-Number, Program, Branch, Gender and Cumulative Performance Index (CPI). The total number of enrolled students was 5210 undertaken different programs.

- The medical history of all students which is collected from medical center of MNNIT for the period of 30 months.

Each record of the access log files has 11 fields for each visited Web page by one individual user. For our analysis, we have selected three fields of each record: User-Id, Date & Time of connection and the URL of the visited Web page. It may be noted that we created Virtual-User-Id for each student to hide real identity of the students to protect their identity.

For statistical analysis, we used Rapid-Miner and R-Miner tools and all inconsistent or missing values are removed during the pre-processing step. Minute is used as a unit of time duration to measure the time spent by one individual user on each Website during a day. This unit of measurement implies that all durations less than a minute have been rounded to one minute.

4. GENDER BASED INTERNET USAGE BEHAVIOURS

This investigation includes three steps: data cleaning, data abstraction, extracting Web access behavior pattern for various purpose such as identifying outliers, the comparison between Internet usage behaviors in general days and the period of illness for both outliers and other students.

The gender-wise Internet usage behaviors of users during semester such as the average time spent on Internet, average number of visited Web pages per day and the category of visited Websites by each user are shown on Table 1. Based on the results of this table, the majority of time spent on Internet belongs to female users. Whereas, the average number of visited Web pages per day by male users was more than female students. The other analysis shows that this behavior depends on different condition or on various periods are changed. For example during examination time or a day before examination the number of users and average time spent on Internet per day are changed [8 and 9].

Table 1 . Gender Wise Internet Usage Behaviors

Gender Wise Internet Usage Behaviors	*Female*	*male*
Average-Time-Spent (Minutes)	108	90
Average Number of Visited Webpages	121	350

In this investigation, for study the effects of medical problem on Internet usage behavior of students, first the number of patients in every month is extracted and then the outliers in students community based on their medical history are identified and finally the category of visited Websites and academic performance of these outliers are discovered. Looking on the number of patients in every months of a year shows that the maximum number of patients belongs to the month of August which is the first month of academic year. Our study shows that the maximum number of these patients (76%) belongs to the students of first semester. Since the number of these patients is decreased on the next months of the academic year, we made an interview with these students and asked them to explain their feelings and health conditions during first semester. Based on the results of these conversations, 81% of these students (who were undertaken the first semester) had stress and tension in the first and second months of joining to new environment and independently living without parental care and maximum after three months, 78% of these students were adopted with new academic and social life and their number of visits to medical

center was decreased. However, 3% of those students could not adopt with new academic environment or they had serious medical problems. We named this group of students as Outliers. Rest of this paper presents comparison between the behaviors of these students specially their academic performance with other students.

5. VARIATION ON USAGE PATTERNS DURING THE PERIOD OF ILLNESS

The disease or period of illness has been characterized in three categories in this study. The day of illnesses and refer to medical center, the day before referring to the medical center (or the day before the illness) and the day after refer to doctor and checkup (or recovery period). For extracting any changes on actual behaviors of users due to medical problems, we made Table 2. It is evident from the above table that, the average time spent on Internet and the average number of visited Web pages per day are decreased during a day before illness and illness day. But the average time spent on Internet in a day after illnesses (the period of recovery) is increased.

Further, the categories of Websites which have been visited by these users during these periods are extracted.

Table 2. Gender wise Changes in Internet Usage Behaviors in the Period of Illness

Gender Wise Internet Usage Behaviors		Day before the illness	Day illnesses	A day after illnesses (Recovery)
Average-Time-Spent (Minutes)	Female	30	60	179
	Male	42	20	130
Average Number of Visited Webpages (daily base)	Female	20	30	99
	Male	41	55	200

Table 3 compares the category and percentage of Websites visited by users on different days, including period of illnesses, normal days during a semester. It is interesting to note that:

- The percentage of time spent on SN websites during a day illness decreases and maximum decreasing belongs to a day before referring to medial center.

- The other interesting point is related to AC Websites usages. During the period of illness, the percentages of time spent on AC (Academic) Websites are increased and number of visited Web pages on these category of Websites are decreased specially a day before referring which are increased during illness period. Therefore we can conclude that, during illness period students spent lesser time in Internet and maximum time spent belongs to academic Websites which might be related to their current course or class homework or requirements. Further, it is interesting to note that, the time spent on EN and SN Websites are decreased during period of illness.

Table 3. The category of the Websites visited by users on different days (hits=number of visited Web pages)

Category of Visited Websites Comparison	AC(Academic)		NAC(Non AC)		SN		EN	
	Time	Hits	Time	Hits	Time	Hits	Time	Hits
A Day Before illness day	53%	30%	47%	70%	3%	47%	18%	8%
Day illnesses	56%	51%	44%	49%	17%	32%	5%	38%
Recovery Day	60%	54%	40%	46%	5%	46%	13%	32%
General Days	39%	42%	61%	58%	19%	34%	21%	23%

5. OUTLIERS

An outlier is generally considered to be a data point that is far outside the norm for a variable or population [104 ~ 111]. We decided to take all students with continues referring to medical center every week per months during semester. 9% of outliers based on medical problems were female and 91% were male students. We analyzed their CPI and Internet usage behaviors. 123 students out of 2980 visitors visited more than 3 days per week medical center during semester, therefore these students identified as Outliers in our study. 85 out of these 123 students were Regular users of Internet. The academic performance (Cumulative Performance Index [CPI]), average time spent and category of visited Websites by these outliers are shown on Table 4 and Table 5.

The range of CPI is zero to 10 and the students with CPI less than 4 are considered as outliers based on academic performance.

Table 4. Outliers Based on Medical Problems vs. Their Internet Usage and Academic Performance

Outliers	CPI	Average Time Spent
Regular Users (All are Male)	5/85 [8.5<=CPI<9.7]	41/85 [T<5min]
	5/85 [8<=CPI<8.5]	23/85 [10<=T<30]
	7/85 [7.5<=CPI<8]	5/85[360<=T<420]
	8/85 [7<=CPI<7.5]	5/85 [420<=T<465]
	20/85[6<=CPI<7]	11/85 [T>=465]
	33/85[5=<CPI<6]	
	5/85 [2.5<CPI<5]	
	2/85 [2.5>=CPI]	

Table 5. Category of Visited Websites by Outliers Based on Medical Problems

Website Categories	Sub-Categories Usages	Average Time Spent	%
Academic	Computer/search engine	10 minutes	15%
NAC(Entertainment)	ART/Media	210 minutes	58%
NAC(Social Networks)	Society/SN/Blog & Email	90 minutes	25%
NAC(Other categories)	Other categories (ex. Business)	8 minutes	2%

From the results of Table 4 it is evident that, almost half of this population 48%[41 out of 85] had spent less than 5 minutes on Internet and they visited Websites category belonging to ART/Media. Only 13% [11 out of 85] of this population had average time spent on Internet more than 465 minutes per day. Other conclusion from Table 5 is regarding maximum time spent on Internet which is belongs to NAC (Non-Academic Websites) subcategory of EN (Entertainments). It is interesting to note, there is only 1% female Regular user among these group of outliers based on their medical history and 78% of these outliers belongs to students undertaken first and second semesters. Our study is shown that 81% of these outliers from third semester become same as other normal group of students. Based on the results of discussions with these students, the stress and tension made their abnormal behaviors.

6. CONCLUSION AND FUTURE WORKS

Based on results of this study , it is interesting to note that, the majority of students with medical problems, belongs to the students of first semester and within these students, male students had more problems compare to female. Moreover, the number of patients on first months of each semester was more than other months. In other words, we can conclude fresher students whose are starting for alone life and far from parental cares mostly getting medical problem which might be psychological related problems in the first months of joining to college and living in campus life. It is interesting to note that, only 1% of female outliers based on their medical problems were regular users of Internet and all of these female outliers belong to the group of students who had weak academic performances.

However, this research work is just a primary study for extracting the effects of illness on students Internet usage behaviors. For generalization of this research work we need to collect data from different colleges all-around of the country and analyze separately psychological problems effects and other medical problems effects for making strong relationships which can be help us for many purposes such as predictions the mental health of students.

7. REFERENCES

[1] Greengard, S. 2011. Living in a digital world. *Communications of the ACM*,*54*(10), 17-19.

[2] Kraut, R., Patterson, M., Lundmark, V., Kiesler, S., Mukophadhyay, T., & Scherlis, W. 1998. Internet paradox: A social technology that reduces social involvement and psychological well-being?. *American psychologist*, *53*(9), 1017.

[3] Kraut, R., Kiesler, S., Boneva, B., Cummings, J., Helgeson, V., & Crawford, A. 2002. Internet paradox revisited. *Journal of social issues*, *58*(1), 49-74.

[4] Lavin, M., Marvin, K., Mclarney, A., Nola, V., & Scott, L. 1999. Sensation seeking and collegiate vulnerability to Internet dependence.*CyberPsychology & Behavior*, *2*(5), 425-430.

[5] Nie, N. H., & Erbring, L. 2000. Internet and society. *Stanford Institute for the Quantitative Study of Society*.

[6] Stoll, C. 1995. Silicon Snake Oil. New York: Anchor Books.

[7] Turkle, S. 1996. Virtuality and its discontents: Searching for community in cyberspace. The American Prospect, 24, 50-57.

[8] Erdoğan, Y. 2008. Exploring the relationships among internet usage, internet attitudes and loneliness of Turkish adolescents. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, *2*(2), 11-20.

[9] Ceyhan, A. A., & Ceyhan, E. 2008. Loneliness, depression, and computer self-efficacy as predictors of problematic Internet use. *CyberPsychology & Behavior*, *11*(6), 699-701.

[10] Guadagno, R. E., Okdie, B. M., & Eno, C. A. 2008. Who blogs? Personality predictors of blogging. *Computers in Human Behavior*, *24*(5), 1993-2004.

[11] Tosun, L. P., & Lajunen, T. 2009. Why do young adults develop a passion for Internet activities? The associations among personality, revealing "true self" on the Internet, and passion for the Internet. *CyberPsychology & Behavior*, *12*(4), 401-406.

[12] Oskouei, R. J., & Chaudhary, B. D. 2010, April. Internet Usage Pattern by Female Students: A Case Study. In *Information Technology: New Generations (ITNG), 2010 Seventh International Conference on* (pp. 1247-1250). IEEE.

[13] Oskouei, R. J. 2010, July. Identifying students' behaviors related to internet usage patterns. In *Technology for Education (T4E), 2010 International Conference on* (pp. 232-233). IEEE.

[14] Shaw, L. H., & Gant, L. M. 2002. In defense of the Internet: The relationship between Internet communication and depression, loneliness, self-esteem, and perceived social support. *CyberPsychology & Behavior*, *5*(2), 157-171.

[15] Tang, H. L. Personality, Social Network Sites, And Leisure Activities–A Conceptual Exploration.

[16] Stepanikova, I., Nie, N. H., & He, X. 2010. Time on the Internet at home, loneliness, and life satisfaction: Evidence from panel time-diary data.*Computers in Human Behavior*, *26*(3), 329-338.

[17] Kirschner, P. A., & Karpinski, A. C. 2010. Facebook® and academic performance. *Computers in Human Behavior*, *26*(6), 1237-1245.

[18] Correa, T., Hinsley, A. W., & De Zuniga, H. G. 2010. Who interacts on the Web?: The intersection of users' personality and social media use. *Computers in Human Behavior*, *26*(2), 247-253.

[19] Stevens, J. P. 1984. Outliers and influential data points in regression analysis.*Psychological Bulletin*, *95*(2), 334.