An Empirical Research of Human Behavior Dynamics in Network Course Learning

Yan Cheng and Yan Zeng

***Abstract*—Learning is a kind of important human behavior to acquire knowledge. This paper discussed temporal characteristics of network course learning on behavior dynamics. Firstly, the students behavior data of the network course learning for 8-weeks are collected from online learning platform. Then, the work used the Maximum Likelihood Estimation(MLE) method for estimating the power exponent of learning behavior interval time distribution, and introduced Kolmogorov-Smirnov(KS) method to test power-law hypothesis .The empirical research results show that: both in the group and individual level, learning behavior time interval obey characteristics of power-law distribution. Underlying these, This thesis combined with the learning psychology, environment and other factors explained the statistical characteristics, and provided some suggestions for teaching management.**

***Index Terms*—Empirical research, behavior dynamics, behavior interval time distribution, power-law.**

I. Introduction

Human behavior tends to have highly complex objective laws. By discussing the behavior of different driving force, and then to predict or monitoring behavior, it has important research and practical significance.

Throughout the 21st century education scientific research development situation, the focus of research has gradually transferred to the exploration learners deep thinking and behaviors. Human society as a typical complex system, deep understanding of characteristics of the virtual network society, is one of the focus problems of modern scientific research, and often the foundation is understand user behavior of statistical feature of time [1]. Due to the complexity and diversity of human behavior, to the early 21st century, most of the related research work, based on individual data, questionnaires and other forms, tend to qualitative analysis ,people is generally believed that human behavior is uniform, silent for a long time and short time eruption is rare and often can be ignored [2]. 2005, however, A.L. Barabási [3] by studying the statistical data on the user's E-mail time interval, the analysis results show that the adjacent time interval between two events with high heterogeneity, obey power-law distribution.

Manuscript received September 2, 2015; revised October 21, 2015. This work was supported by National Natural Science Foundation of China (NSFC, Grant No. 61262080), Jiangxi Province Science and Technology Support Major Project (20151BBE50121), and 2011 collaborative innovation center "Based on Big Data of Jiangxi Province Teacher Quality Monitoring, Evaluation and Service".

Yan Cheng is with Tongji University. She is also with Jiangxi Normal University, Nanchang, Jiangxi, China (e-mail: chyan88888@jxnu.edu.cn).

Yan Zeng is with Jiangxi Normal University, Nanchang, Jiangxi, China (e-mail: ningmoxi@163.com).

Human behavior could keep silent in a long time, and may also be high frequency in a short time. Not long after, Vazque *et al*. [4] made further study, and put forward human behavior can be divided into two general classes: the power exponents is 1 and 1. 5, including the empirical data of the mail communication model show that the power exponent is 1.5. Since then, a large number of scholars at home and abroad have carried out quantitative research on human behavior in real life. It is proved that the time interval of human behavior is a power law, and the power exponent is between 1 and 3 [2].

At present, the empirical research as the preliminary stage of quantitative understanding human behavior motivation has important research necessity. This article from the perspective of behavioral dynamics, explored the regular pattern of network learning behavior time interval, and combined with the learning environment and psychological factors to explain results. On the one hand, the paper lay the foundation of exploring the network learning behavior driven mechanism, seek dynamic model to promote the virtuous cycle of learning and teaching at late stage. On the other hand, it can also be a complement to the empirical and modeling studies of human dynamics theory.

II. Research Method

A lot of researchers have found that such as web browsing [1], [5], blog comments [6], market trading [7], human communication behavior [8]-[10], logistics transportation [11] and other human behavior all obey the power law relationship .Therefore, many researchers studied the method to determine whether a distribution is a power law distribution and an estimate of the power exponent.

In mathematics, the random variable X obeys the power law distribution which means that when X takes a certain number x, the probability is proportional to a power of x:

$$P\left(X = x\right) \propto x^{-\alpha} \tag{1}$$

The α is called power exponent or scale parameter.

When the X represents the continuous random variables, the power-law distribution can be described by the probability density $P(X)$ as follows:

$$P(x)dx = P(x <= X < x + dx) \propto x^{-\alpha}\, dx \tag{2}$$

Obviously, Refer to eq.(2),if take logarithm on both sides of the equation at the same time, the probability density function of power-law distribution is approximately fitted to a

straight line with a slope of -α in the double logarithmic plot. Therefore, in the early, the researchers generally used a simple graphical method to test whether a random variable X follows power-law distribution. By defining the probability density function of the random variable, if the distribution is a straight line in the double logarithmic plot, the X obeys the power law distribution. And they often using least squares method to estimate out the slope of the line α (not including minus) [12].

However, due to the traditional least squares fitting curve is calculated through the principle of the minimum distance square points to the fitting line to select the optimal curve, but the exponential distribution is the long tail distribution, value of X in the tail may very disperse. So if using the standard linear regression method, we would get several different solution, in the end, the optimal fitting effect has certain subjectivity [13]. Clauset *et al.* [14] used the maximum likelihood method (MLE) to estimate the parameters of the power law distribution. Compared to the results obtained by the least square estimation based on linear regression, the accuracy of the results is significantly improved. According to the maximum likelihood estimation, can get parameter equation is as follows:

$$\alpha = 1 + n\left[\sum_{i=1}^{n} \ln \frac{x_i}{x_{min}}\right]^{-1} \tag{3}$$

x_{min} is the minimum values conform to the power-law distribution in x.

At present, the researchers generally recognize paper [14] proposed the method to examine cumulative distribution function under the double logarithmic coordinate .Computing time interval of the cumulative distribution events, can effective use of all the data.

Cumulative distribution of power-law distribution in statistical physics CDF is still obey the power law, only the power minus one:

$$P(X >= x) = \frac{\alpha - 1}{x_{min}^{-\alpha+1}} \int_x^{\infty} X^{-\alpha} dX = \left(\frac{x}{x_{min}}\right)^{-\alpha+1} \tag{4}$$

The ideal of Kolmogorov-Smirnov (KS) test is compare the actual data and theoretical data of the cumulative distribution, calculate the maximum deviation value D, then test deviation value significance. If there is a small gap between them, infer the samples meet the theoretical distribution. In this paper we assumed the learners' learning behavior online time interval follows power-law distribution, use the maximum likelihood estimation method to calculate the power law index which was proposed by Clauset A, Shalizi C R, Newman M E J [12], and through the KS test the power-law hypothesis.

The key parts of empirical research:

First, we processed the data, input time interval.

Then, we let each time interval as x_{min}, based on MLE method to estimate x_{min}, calculate each x_{min} corresponding α and D values.

Finally, through the KS test, select the minimum value of D the corresponding minimum α, was given the best fitting α and corresponding x_{min}.

III. Empirical Research

A. Data Sources and Data Processing

The research data of this paper comes from the experimental platform--Moodle online learning system (http://happyclub825.oicp.net/moodle/), which is constructed by the Jiangxi normal university computer information engineering college professor Yan Cheng group. The platform was formally launched in June 2012, and it has been used in the course of "Computer Culture Knowledge Base" and "C++ Program Design" as the experimental course, including curriculum, discussion community, online testing, and other modules. For more targeted, we selected the 88 students learning "c ++ programming" (course number is 1)'s behavior data of 2014 electronic business class in our school. From the "mdl.log" file of the learning platform. The behavior information of educational administrators and a few visitors were removed, and the 16365 records that behavior of online learning of the course was obtained from July 5, 2015 to May 5, 2015 in eight weeks (60 days). After further processed errors (such as the action failed etc.), a total of 16145 records.

Fig. 1. The homepage of online learning platform.

B. Analyzed the Time Interval Distribution of Group Behavior

Observational study of group behavior can reflect the group's behavior if there is some objective law. Therefore, the study on the dynamics of human behavior is often pay attention to the distribution of the time interval between the similar events. In this paper, we analyze the time interval of the student's learning behavior (Don't distinguish their id).

Fig. 2 shows the overall situation of students accessed to the platform with time (date). It can be seen, the overall trend is: to gradually rise, reached a peak value and then decreased gradually, after a low access rates during the period, will appear "burst", presented in this trend cycle. And, from Fig. 3, There are short time intervals, but the long time interval is also very obvious. In order to describe the time trend of the study of the student's log in platform more clearly, according to the experimental process design, Fig. 4 is time interval of the cumulative distribution network course learning behavior

in log-log coordinate (dotted line is based on MLE, the optimal curve obtained by KS test), the exact solution of the power exponent is 1.66, the x_{min} is 11 minutes, the check list can be checked by KS. At the group level, the time interval of the students learning a course is not uniform.

TABLE I: TIME INTERVAL FITTING THE PARAMETERS OF THE INDIVIDUAL BEHAVIOR

Learners	Frequency	Date Range	α	x_{min}
Lilingling	560	2015/5/7 21:08:19~2015/6/16 11:47:16	1.5	3
Zoujiaming	418	2015/5/20 10:50:04~2015/6/17 10:50:29	1.5	4
Luoyuanyuan	398	2015/5/10 23:05:08~2015/6/16 11:46:19	1.5	5
Wangrong	396	2015/5/20 10:54:00~2015/6/17 11:05:05	1.5	5
Yuanke	383	2015/5/7 12:57:12~2015/6/17 10:54:31	1.5	4
Mengting	361	2015/5/7 14:16:11~2015/7/2 17:35:21	1.5	6
Denghonghong	351	2015/5/10 21:50:11~2015/6/16 11:50:31	1.5	4
Mengxiaoqing	350	2015/5/11 19:08:02~2015/6/16 11:52:22	1.5	4
Guyeping	343	2015/5/11 17:47:11~2015/6/16 13:36::23	1.5	4
Caoningyan	342	2015/5/11 17:46:02~2015/6/18 13:17:23	1.5	5
Liuhui	320	2015/5/7 22:01:56~ 2015/6/16 11:47:12	1.5	3
Wanyiping	287	2015/5/20 10:50:21~2015/6/17 11:07:08	1.5	5
Duanxiaofang	287	2015/5/7 20:58:32~2015/6/16 16:34:38	1.5	7
Jiayue	195	2015/5/7 23:18:53~2015/6/16 23:43:43	1.5	4
Wanglianxiang	132	2015/5/20 10:57:10~2015/6/17 11:15:03	1.5	5
Zhuyuanfeng	125	2015/5/20 10:51:30~2015/6/17 13:26:33	1.5	6
Chengqigan	122	2015/5/11 19:16:09~2015/6/16 11:45:16	1.5	7
Linbinxi	119	2015/5/26 10:56:33~2015/6/16 11:45:28	1.5	7
Fushuhao	107	2015/5/31 12:13:50~2015/6/16 11:44:34	1.5	8
Zhudaxiong	106	2015/5/20 11:06:26~2015/6/17 11:20:16	1.5	6

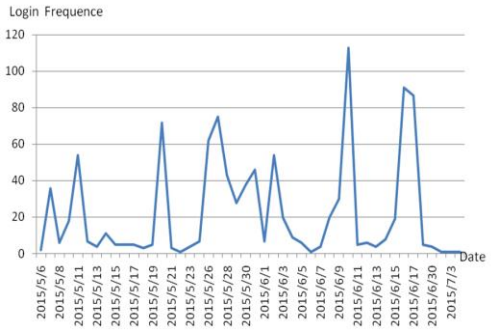

Fig. 2. The overall situation of students accessed to the platform.

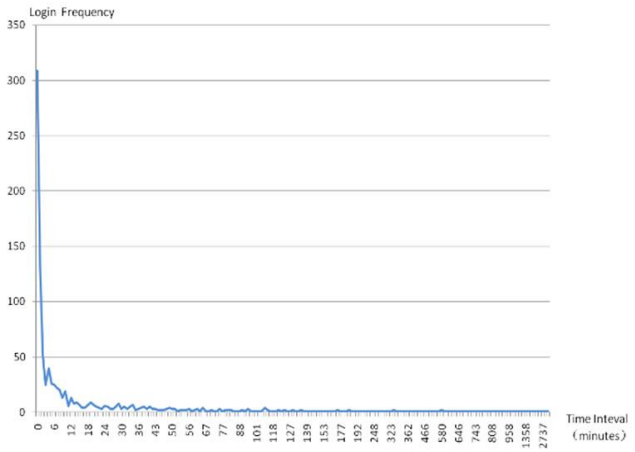

Fig. 3. The time interval frequency distribution of the students' group access platform.

C. Analyzed the Time Interval Distribution of Individual Behavior

The group is made up of individual; therefore, it is necessary to investigate whether the behavior of the individual is really the cause of the time interval of the group behavior. The frequency of individual's log is much less than that of the group, and often the individual learning time interval is shorter, so the choice of seconds for the unit is more reasonable.

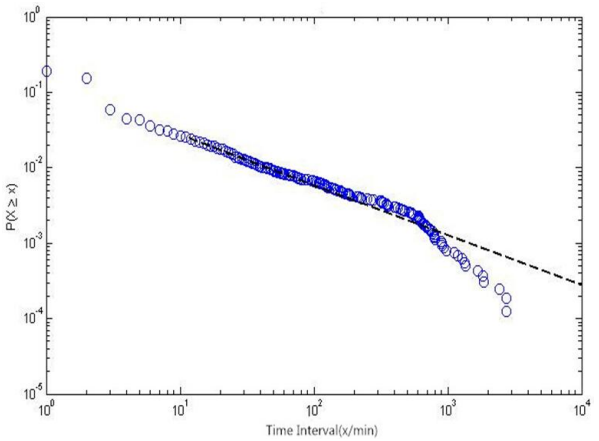

Fig. 4. Time interval of the cumulative distribution network course learning behavior in log-log plot.

According to the design of the experimental process, the statistical results show that the time interval distribution of the learning behavior of the individual still exhibits a power law characteristic. The data obtained in Table I(part) shows that different individuals, although the frequency of access is different, but the time interval distribution based on the power exponent of 1.5. Fig. 5 may give us one reason to explain the statistical characteristic. The majority of students access to the test module, nevertheless, the test is generally done after a

certain knowledge point, so although the time to visit the platform is free, the learning of the knowledge point is mainly through the teachers' classroom teaching, even if the individual complete the time point of the network course learning is different, the law is similar. The Vazquez based on the distribution of time interval of the celebrity communication fitted to the power index of the situation similar reasons, it is likely that the students will take the test or other pressure, supervise the "cramming", for after a long time of rest then be engaged in intensive .The situation of communication behavior in daily life is: both sides may be a long time no contact, because something get closely, and afterwards resumed calm cycle. There is a certain similarity between the time and the behavior of the two, so it is possible to have the same power exponent. Moreover, by the value of x_{min} can be found in the high activity of the individual, the next action is generally shorter, and the low activity of the relatively long time interval, which is consistent with the actual situation.

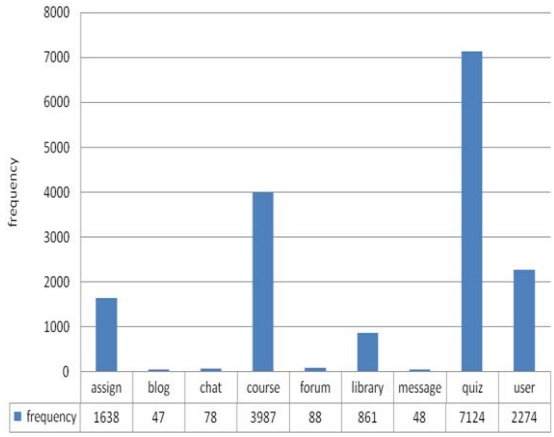

Fig. 5. Students accessed modules frequency statistics.

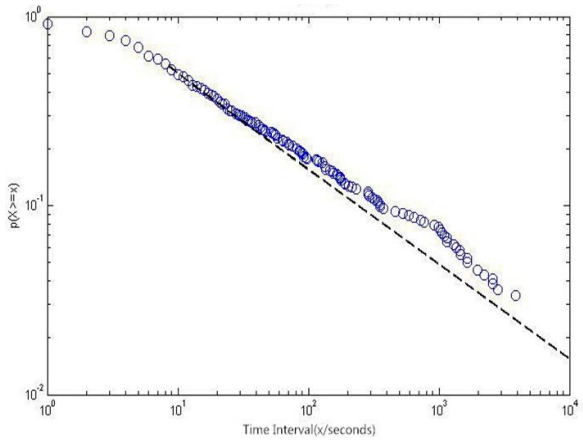

Fig. 6. Cumulative distribution function of the time interval on the log_log plot of Zoujiaming.

Because of the space is limited, Fig. 6, Fig. 7 respectively the cumulative distribution of two students' network course learning time interval. As shown in the diagram, the cumulative distribution of each individual is parallel to the dotted line, which is the same as the power exponent calculated in the table. And, from the view of the graph, only the first half of the fitting line of the cumulative probability is in good agreement with the original distribution, and then the half part is deviated from the original distribution. The main reason for this situation is that the MLE method is based on the probability distribution of the curve fitting, which is to ensure that more points are distributed in the fitting line as shown in the graph, therefore, the result is reasonable. At the same time, it is obvious that the low activity of the individual time interval distribution is poor, there are more cases of long time interval.

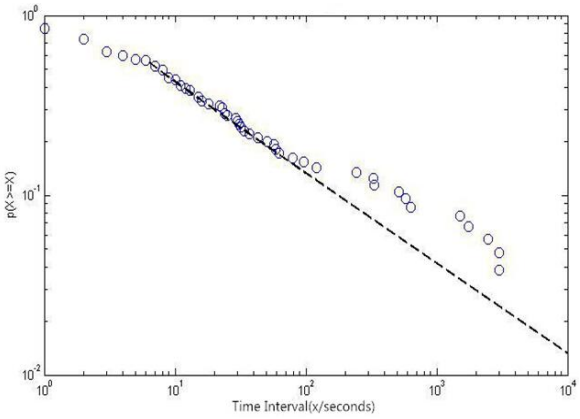

Fig. 7. Cumulative distribution function of the time interval on the log_log plot of Fushuhao.

IV. Conclusion

In this paper, we discussed the statistical properties of the time interval between the learners, which can provide an example for the study of human behavior. The empirical results are as follows:

1) Both the learning behavior time interval in group and individual level are obey power-law distribution, but most of individual level power index is 1.5, the power of the group rate index is relatively high at 1.66.The causes of this rule may be: In the study of group behavior of time series is made by a single individual behavior, and therefore different behavior between each other filling is likely to shorten the time of crowd behavior. In the plot, data points will be more close to the y axis, fitting graphics more steep, so power exponent is even greater.
2) At the individual level, power exponent is very unified, the reason maybe is the phenomenon of students' individual "cramming" too prominent. Although there is no rule stipulate when individual independent access platform , students from same class, spare time distribution is very similar, thus it is possible to subject to the time interval of online learning is very similar, too.
3) According to the study the situation of students in different modules, teachers need to on the basis of law in study behavior time interval , at the appropriate time (such as before testing) to strengthen the students' online tutoring, and by updating the resources of the modules and other ways to attract students to learn more.
4) Studies on human behavior not only contribute to a profound understanding of the complexity of socio-economic system and human beings, but also have broad prospects in practice [15]. Such as A. J. Jara, Y. Bocchi, D. Genoud [16], X. J. Chun, R. R. Liu [17]. Behind human behavior may be more essential mechanism still remain to be discovered, such as the network learning behavior rule is driven by some certain factor. The correlation of time interval of learning

behavior can also be done in the future. Quantitative understanding online learners' learning behavior and learning psychology, may provide new basis for updating teaching management method.

REFERENCES

[1] G. S. Zhao, N. Zhang, and T. Zhou, "Scale of web browsing behavior study," *Journal of Statistics and Decision*, vol. 1, pp. 18-19, 2009.

[2] C. Fan, J. L. Guo, X. P. Han, and B. H. Wang, "A review of research on human dynamics," *Complex Systems and Complexity Science*, vol. 8, no. 2, pp. 1-17, 2011.

[3] A. L. Barabási, "The origin of bursts and heavy tails in human dynamics," *Nature*, vol. 435, no. 7039, pp. 207-211, 2005.

[4] A. Vazquez, J. G. Oliveira *et al.*, "Modeling bursts and heavy tails in human dynamic," *Phys. Rev. E.*, vol. 73, no. 3, pp. 036127-036127, 2005.

[5] Z. Dczso, E. Almaas, A. Lukacs *et al.*, "Dynamics of information access on the web," *Phys. Rev. E.*, vol. 73, no. 6, 2006.

[6] J. L. Guo, "Empirical study and modeling of human behavior dynamics of comments on blog posts," *Application Research of Computers*, vol. 28, no. 4, pp. 1422-1424, 2011.

[7] T. Kaizoji and M. Kaizoyi, "Power law for the calm-time interval of price change," *Physica A: Statistical Mechanics and its Applications*, vol. 336, pp. 563-570, 2004.

[8] W. Hong, X. P. Hang, T. Zhou *et al.*, "Heavy-tailed statistic in short-message communications," *Chin. Phys. Lett.*, vol. 26, 2009.

[9] Z. D. Zhao, H. Xia, M. S. Shang, and T. Zhou, "Empirical analysis on the human dynamics of a large-scale short message communication system," *Chin. Phys. Lett.*, vol. 28, no. 6, June 2011.

[10] W. J. Wang, N. Yuan, L. Pan, P. F. Jiao, W. D. Dai, G. X. Xue, and D. Liu, "Temporal patterns of emergency calls of a metropolitan city in China," *Physica A*, vol. 436, pp. 846-855, Oct. 2015.

[11] Q. Wang and J. L. Guo, "Human dynamics scaling characteristics for aerial inbound logistics operation," *Physica A: Statistical Mechanics and Its Applications*, vol. 389, pp. 2127-2133, 2010.

[12] L. A. Adamic. (1999). Zipf, power-law and pareto — A ranking tutorial. [Online]. Available: http://www.hpl.hp.com/research/idl/papers/ranking/ranking.html.

[13] P. Han, G. F. Lu, and D. B. Wang, "Zipf's law based on maximum likelihood estimation method validation," *Journal of Theory and Practice of Intelligence*, vol. 11, pp. 6-11, 2012.

[14] A. Clausct, C. R. Shalizi, and M. E. J. Newman, "Power law distribution in empirical data," *SIAM Review*, vol. 51, no. 4, pp. 661-703, 2009.

[15] L. Wang and S. C. Qu, "The empirical results and modeling of human dynamics," in *Proc. 2010 2nd International Conference on Information Engineering and Computer Science (ICIECS)*, Dec. 25-26, 2010.

[16] A. J. Jara, Y. Bocchi, and D. Genoud, "Determining human dynamics through the internet of things," in *Proc. 2013 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)*, vol. 3, pp. 109-113, Nov. 17-20, 2013.

[17] X. J. Chun and R. R. Liu, "Improve the algorithmic performance of collaborative filtering by using the interevent time distribution of human behaviors," *Physica A*, vol. 436, pp. 236–245, May 2015.

Yan Cheng was in February, 1976 in Wu Yuan City. She got the bachelor's degree in 1998, the master's degree in 2005, and the Ph.D. degree in 2010 from Tongji University in Shanghai and she is now a postdoctor of Tongji University in 2015, majored in computer science, engaged in e-learning, virtual learning community and educational technology innovation. From July 2013 to March 2014, she went to the California University as a visiting scholar. Cheng's main research interests include intelligent computer aided education, educational data mining, virtual learning community and e-learning.

Currently she is a postdoctor of Tongji University and a professor in Jiangxi Normal University in Nanchang, Jiangxi province. As a researcher, currently she is the leader of 2 national NSFC (Natural Science Foundation of China) Plan Project and more than 5 provincial projects on IT and education innovations. She is the author of more than 30 scientific papers, won Guanghua Ph.D. scholarship, 3 talks in the global intelligent automation conference, 1 national doctoral academic best paper award. She published 1 academic monograph (Beijing, Science Press, 2014) on educational data mining, 5 computer professional teaching materials.

Dr. Cheng is the senior member of the Systems Engineering Association of China, the review expert of Computer Application Study, one member of Chinese Intelligent Automation, the consulting evaluation expert of small and medium-sized enterprise in Jiangxi province. Dr. Cheng is also the direction leader of the ministry of education "software engineering" discipline and the expert in education informationization.

Yan Zeng was born in September 1990. She is studying the master of computer science and technology of Jiangxi Normal University. And her research direction is e-learning and behavior analysis. As the team members of tutor, she actively participates in a number of researches, interested in complex system and network learning behavior.