Prediction of Posting Behavior of Online UGC Communities Users Based on Poisson Regression

Xingyu CHEN[a], Yu You LING[a], Xianqi HU[a], Xiuchen LIN[a] and Liang ZHU[b1]
[a] *College of Management, Shenzhen University, China*
[b] *College of Economics, Shenzhen University*

Abstract. As a new form of Internet service, User Generated Content (UGC) community has gained popularity these years, and also attracts academic attention. The posting behavior of users is an important index to measure whether the users are active or not. Identifying the potential active users in UGC Communities helps the companies make proper management strategies, which renders the prediction of users' posting behavior important. This study used Poisson regression to examine the effect of users' past behavior and status on users' later behavior of posting. The data for analysis were extracted from ten thousand users on Qiushibaike throughout a window of three months, which included the number of posts and the number of likes, comments, and times they were featured on the home page by administrators. The results indicate that users' past comments have a negative effect on users' later posting behavior; users' past posts have a positive effect on users' later posting behavior; users' past likes and hot posts they got have a positive effect on users' later posting behavior.

Keywords. UGC Online Community, Poisson Regression, Community Participation

Introduction

With the introduction of Web 2.0 15 years ago, online communities catering to User-Generated Content (UGC) is already a concept that marketers, consumers and scholars are familiar with. UGC is presented in many formats including video, still imagery, text, and even audio. These various content formats have become the focus of social media platforms big and small. In academia, UGC has also become a pertinent topic of research and has accumulated numerous discoveries. In this paper, User-Generated Content is classified as visual and audio web content that facilitates the sharing of personal opinions and knowledge. Through the acts of commenting, sharing, and liking such content, users interact with other members of the community [1][2][3]. Platforms like Facebook, Twitter, Instagram, Weibo, Tik-Tok, QiushiBaike have become household names, representing a group of UGC Communities that have garnered a significant and active user base.

UGC Communities have the ability to concurrently provide the utility of learning, self-expression, and entertainment for large amounts of users. At the same time, while

---

[1] Corresponding Author, Email: lzhu111@szu.edu.cn.

users engage in such activities, the success of UGC Communities is directly correlated with its frequency, and increasing this very frequency is a perpetual business challenge for these platforms [4][5] （Therefore, engaging users and motivating them to be active on UGC Communities is recognized as the key for platform success [6][7] Encouraging higher amounts of activity in the user base is a question all platforms administrators constantly ponder over. Scholars have also recognized the motivations of UGC Community users as an important topic of research. Some scholars have identified two forms of user motivations, namely Proactive and Passive. Proactive is classified as the type of motivation behind users that generate large amounts of content, while Passive is classified as the type of motivation behind users who mainly view or consume content [8].

Existing research have discussed these Proactive and Passive motivations, and conclude that other factors like Platform Culture, User Profile, Platform Design, and User Sentiment towards the platform are also major factors influencing their frequency of platform activity [9][10][11][12][13][14]. While the research mentioned previously has contributed discoveries in understanding what motivates and attracts users to UGC Communities, an unexplored avenue would be how can we better understand future activity of users through analyzing the history of their activity (liking, sharing and etc.) on the platform. Therefore, this paper discusses the correlations and relationships between users past activity and future activity, and the psychological motivations behind these relationships.

This research is conducted on Chinese social media platform Qiushibaike, using data collected from 10th May to 1st August of year 2016. Data points collected are total posts, total likes, total amount of comments and total times featured by community administrators. The data is analyzed using Poisson Regression and results conclude that user activity history is correlated with future activity patterns. Our research has discovered that the amount of comments a user received is inversely correlated with the rate of future posts, and received likes, past posting frequencies are positively correlated with future posting frequencies.

1. Literature Review

Since their beginnings 15 years ago, UGC Communities have been embraced globally as one of the most popular online services available. As UGC Communities continuously introduce new capabilities, questions that pertain to the definition of UGC Communities, and the drivers and motivations behind user engagement and participation are key to understanding the UGC market and industry.

1.1. Definition of UGC

The earliest definition of a UGC platform was made by Howard Rheingold in 1993, which defined UGC Communities as a web platform that allows multiple users to share their opinions, thoughts and sentiments, to create simultaneous online social interactions [15]. Preece saw UGC Communities from an administrative point of view, emphasizing the need to serve as a platform for practical activity, and a clear definition of its utility will be needed to guide users [16]. According to Preece, an online community is made up of, users or people socially interacting and sharing a purpose, of

policies to guide these interactions, and of computer systems that facilitate all these interactions. In Social Sciences, communities are defined by the significance of the members' experiences and what it means to them, rather than structures or appearances. Cohen introduced a concept of symbolic communities, defined as one that does not exist physically, but is constructed through common meanings, norms, and cultures [17]. Blanchard and Markus defined shared emotional connections and a sense of belonging, which in simpler terms meant a sense of community as a distinctive feature of communities [18][19]. The similarities in the discoveries of Preece, Cohen, Blanchard and Markus show that UGC Communities, aided by computer systems, also facilitate these social interactions of shared meaning and culture. Even with abundant existing research on UGC Communities, its definition is often evolving and in question. In this paper, a UGC Community is defined as an online community that by providing a virtual space, facilitates the sharing of information and personal opinions. Through receiving appraisal by liking, and through interaction by commenting or sharing, users engage in social interactions and form relationships [1][2][3].

1.2. UGC Community Participation

Members of UGC Communities are mainly classified into two categories of Lurkers and Diffusers [20]. Diffusers are proactive users who contribute majority of the content, while Lurkers view or consume these content, which serve them the utility of entertainment, or learning without the need to generated any content [20]. Lurkers provide large amounts of web traffic and click-through rates. However, a UGC platform, even with a large user base, might not succeed commercially if majority of the users are Lurkers [21].

All UGC Community platforms identify User Engagement and Activity as keys to success, and are often used as a metric to measure their probabilities of success. Websites that have been abandoned by their users are often classified as "Ghost Towns" [20]. In order to measure the success of UGC Community platforms, scholars have identified reliable metrics. According to Iriberri & Leroy [22], the most generalizable metrics that project a UGC platform's probability of success are content generation rate and the quality of relationships between users. Content Generation Rate is positively correlated with the Quality of Relationships between users, scoring high on both metrics often meant success for the UGC platform. Therefore, proactive users generating content will be key to a platform's success.

Based on first party data from a UGC Community Platform, with the data points of Total Likes, Total Posts, Total Comments received, and Total Times Featured by community administrators this paper discusses the question of how User Activity History affects Future Activity in the UGC Community.

2. Data and Methods

The Data used in this paper originated from UGC Community Platform Qiushibaike, which is ranked highly popular in China. Founded in 2005, it has accumulated 20 Million users and 10 Billion user-generated posts. It launched its mobile application version in 2012. Users contribute personally generated content in both imagery and text to participate. They can also view the content of other users and through liking and commenting, interact with other members of the UGC Community. We have used first

party data collected between 10th May to 1st August 2016. Data points include Total Likes, Total Comments and Total Posts cumulated at 24:00 daily.

This research seeks to understand how user data dating from Day T-1 will affect the same variables of the same user at a given day T. Therefore, the regression model can be written as:

$P_t = \alpha_0 + \alpha_1 P_{(t-1)} + \alpha_2 〚AV〛_{(t-1)} + \alpha_3 〚AC〛_{(t-1)} + \alpha_4 〚AH〛_{(t-1)} + \varepsilon_t$,

where P_t refers to Total Posts at time T; $P_{(t-1)}$, $〚AV〛_{(t-1)}$, $〚AC〛_{(t-1)}$, and $〚AH〛_{(t-1)}$ are Total Posts, Total Likes, Total Comments, and Total times a post was featured by administrators of the UGC Community at time T-1, respectively.

2.1. Descriptive Analysis

We have used descriptive analysis, for a comprehensive understanding of the data. Data of 10,000 users dated from 10th May to 8th August 2016, show that the average number of Total Posts per user is 52.409, with the lowest at 0 and highest at 4494 (Table 1). The average number of Total Likes Received are 23481.538, with the lowest at 0, highest at 4004899. The average number of Total Comments Received is 137.098, with the lowest at 0 and highest at 42986. Lastly, the average number of Total Times Featured is 1.128, with the lowest at 0 and highest at 352. Results show that, dating from T-1, Total Likes Received has a higher mean than Total Comments Received. Total Comments Received has a higher mean than Total Posts. The act of Liking requires the least effort, while Posting requires moderate effort and Commenting requires the most. Our Data reveals that majority of the users in this UGC Community are less proactive.

Table 1. Descriptive Analysis of Variables.

	Mean	Min	Median	Max	SD
APt	52.409	0	18	4494	156.162
AVt	23481.538	0	2730	4004899	129302.298
ACt	137.098	0	0	42986	972.371
AHt	1.128	0	0	352	9.004

2.2 Data Model

Total Posts of the day would be the dependent variable, numbering from 0, 1, 2, 3….n. Being numbered variables, we used Poisson Regression as the basis of our experiment's model. Firstly, analyzing the data with a model that can accommodate all our variables. Checking for overall significance and likelihood of generalizability, we arrive at a tiny P factor, meaning this model is significant. Therefore, out of 4 designated independent variables, at least 1 will have a significant correlation with user

posting activity in the near future. Further analysis using Z inspection, we conclude that apart from average Total Comments Received, the remaining independent variables are all highly correlated with the dependent variable. Therefore, a history of high Total Posts, Total Likes Received, and Total Times Featured, are indicators of high future posting activity. Table 2 displays our data and conclusions.

Table 2. Model Estimated Results

	Estimate	**Std. Error**	**z value**	**Pr(>|z|)**
(Intercept)	-3.612e+00	6.507e-03	-555.10	<2e-16
APt	7.747e-04	1.267e-05	61.17	<2e-16
AVt	8.972e-07	1.207e-08	74.34	<2e-16
ACt	-3.528e-05	1.893e-06	-18.64	<2e-16
AHt	1.629e-02	1.893e-06	89.33	<2e-16

Therefore p<0.001

3. Results and Conclusion

Posting behavior is an important metric that defines how proactive an individual. Existing research has discussed the motivations behind posting behavior, and provided relevant implications based on their conclusions. However, they have not studied the relationships between behavioral history and future user activity in UGC Communities. Using the history of relevant data points (which in the case of this research are the Total Posts, Total Likes, Total Received Comments and Total Times Featured) managers of UGC Community platforms can leverage such data to understand and predict the future activity of users.

Data shows that in UGC Communities, Total Comments (〖AC〗_(t-1)) Received are positively correlated with future posting activity. Total Likes Received are positively correlated with future posting activity. Total Times Featured by the platform are positively correlated with future posting activity. The data also shows that posting activity of users are consistent throughout the study, therefore users who have historically high posting behavior are likely to post more and generate more content in the near future. Users with historically low posting rate are unlikely to generate more content, since receiving comments and likes are seen as a form of appraisal by community members and have been proven to reinforce future posting activity, a low post count would mean there are less opportunities for an individual to receive such encouragement from the community which leads to low future posting activity in the near future.

In addition, Total Comments Received is negatively correlated with an individual user's future posting activity. A high number of Total Comments Received has led to an individual exhibiting low posting activity in the near future, and a low number of Total Comments Received has led to high posting activity in the near future. It is

possible that the social interactions required to reply to these comments have exhausted the attention span of the individual user, and the fatigued attention leads to low future posting activity.

Finally, the 3 dependent variables although positively correlated with the independent variable, exhibit positive correlation at different degrees. Total Times Featured has a lesser effect on future posting activity than the other independent variables. Being featured is an occasional activity, and it is possible that users have learnt to pay more attention to appraisal by community members (Total Likes, Total Comments) instead of appraisal by administrators. Total comments have moderate effects, and it is the Total Likes that exhibit the strongest positive correlation on future posting activity.

3.1. Practical Implications

The results show that historical user data to a very high extent, can be used to accurately predict future user behavior. Therefore, it is important for UGC Community platforms to allocate adequate resources for the data collection and analysis of historical user activity. Administrators, through design of the platform, can allow being featured to be a much more special event, making it a much stronger reinforcement for future posting activity. While received likes and comments continue to motivate future posting activity, administrators can further implement immersive interactive features or prominent mechanisms to like or comment, so as to engage users on an even higher level.

3.2. Limitations and Future Research

This paper focused on understanding how user behavior history affects future behavior. We used the data points of Total Likes Received, Total Comments Received, Total Times Featured, and Total Posts. Results conclude that Total Posts are positively correlated with Total Posts in the near future. Total Likes Received are positively correlated with Total Posts in the near future. Total Comments Received are negatively correlated with Total Posts in the near future. Although our research has arrived at discoveries of the behavioral patterns of users in a UGC Community, we have yet to uncover the psychological motivations behind such patterns, and it is an important topic that calls for further research.

Acknowledgement

References

[1] H. Li and Y. Liu, Understanding post-adoption behaviors of e-service users in the context of online travel services. *Information & Management*, 2014, 51(8), pp. 1043-1052.
[2] H. Lin, W. Fan and P.Y. Chau, Determinants of users' continuance of social networking sites: a self-regulation perspective. *Information & Management*, 2014, 51(5), pp. 595-603.

[3] U. Pfeil, R. Arjan and P. Zaphiris, Age differences in online social networking – A study of user profiles and the social capital divide among teenagers and older users in MySpace, *Computers in Human Behavior*, 2009, 25(3), pp. 643-654.

[4] E.P.S. Baumer, M. Sueyoshi and B. Tomlinson, Bloggers and readers blogging together: Collaborative co-creation of political blogs. *Computer Supported Cooperative Work (CSCW): An International Journal*, 2011, Vol. 20, No. 1–2, pp. 1–36.

[5] A. Velasquez, R. Wash, C. Lampe and T. Bjornrud, Latent users in an online user-generated content community. *Computer Supported Cooperative Work*, 2013, 23(1), pp. 21–50.

[6] J. Koh and Y.G. Kim, Knowledge sharing in virtual communities: An e-business perspective. *Expert Systems with Applications*, 2004, 26(2), pp. 155–166.

[7] J. Koh, Y.G. Kim, B. Butler and G.W. Bock, Encouraging participation in virtual communities. *Communications of the ACM*, 2007, 50(2), pp. 69–73.

[8] J. Preece, B. Nonnecke and D. Andrews, The top five reasons for lurking: Improving community experiences for everyone. *Computers in Human Behavior*, 2004, 20(2), pp. 201–223.

[9] Y. Du, Modeling the behavior of lurkers in online communities using intentional agents. *2006 International Conference on Computational Inteligence for Modelling Control and Automation and International Conference on Intelligent Agents Web Technologies and International Commerce (CIMCA'06)*, 2006, DOI: 10.1109/CIMCA.2006.149.

[10] Y.-W. Fan, C.-C. Wu and L.-C. Chiang, Knowledge sharing in virtual community: The comparison between contributors and lurkers, *The 9th international conference on electronic business*, Macau, 2009, pp. 662–668.

[11] G. Leshed, Posters, lurkers, and in between: A multidimensional model of online community participation patterns. In G. Salvendy (Ed.), *Proceedings of the 11th HIC international conference*. Las Vegas: Lawrence Erlbaum Associates Inc., 2005.

[12] R.B. Nonnecke, *Lurking in email-based discussion lists*. PhD thesis, South Bank University, 2000.

[13] B. Nonnecke and J. Preece, Why lurkers lurk. *Americas conference on information systems*, 2001, pp. 1–10.

[14] S.J. Tedjamulia, D.L. Dean, D.R. Olsen and C.C. Albrecht, Motivating content contributions to online communities: Toward a more comprehensive theory. *Proceedings of the IEEE 38th annual Hawaii international conference on System sciences, HICSS'05*. 2005, DOI: 10.1109/HICSS.2005.444.

[15] H. Rheingold, *The virtual community. Homesteading on the electronic frontier*. MIT Press, London, 1993.

[16] J. Preece, *Online communities: Designing usability, supporting sociability*. Wiley, New York, 2000.

[17] A. Cohen, *The symbolic construction of community*. Ellis Horwood Ltd., London, 1985.

[18] A. Blanchard and M.L. Markus, Sense of virtual community – Maintaining the experience of belonging. *Proceedings of the 35th annual Hawaii international conference on system sciences (HICSS)*, 2002, pp. 3566–3575.

[19] A. Blanchard and M.L. Markus, The experienced sense of virtual community: Characteristics and processes. *The Data Base for Advances in Information Systems*, 2004, Vol. 35, No. 1, pp. 65–79.

[20] J. Preece and D. Maloney-Krichmar, Online communities. In J. Jacko and A. Sears (eds.), *Handbook of Human–Computer Interaction*, Lawrence Erlbaum Associates, Mahwah, 2003, pp. 596–620.

[21] C. Ridings, D. Gefen and B. Arinze, Psychological Barriers: Lurker and Poster Motivation and Behavior in Online Communities. *Communications of the Association for Information Systems*, 2006, 18, https://doi.org/10.17705/1CAIS.01816.

[22] A. Iriberri and G. Leroy, A life-cycle perspective on online community success. *ACM Computing Surveys (CSUR)*, 2009. 41(2), 11.