Exploring User’s Classification Model for UGC and Its Influence Factors on Contribution in Social Media

Yuxiang Zhao  Kewen Wu  Qinghua Zhu  Fei Meng
Department of Information Management, Nanjing University, Nanjing, China

Abstract
This paper aims at exploring the user’s contribution in different social media sites, and further examining what factors will influence user’s motivation to participate in UGC. After a brief literature review, the authors collect data from 8 social media sites and analyze the quantitative relationship between the users and contents, and then a classification model based on contribution is proposed. Furthermore, two-factor theory is used to study the hygiene and motivational factors for UGC by surveying the three groups of users in the classification model. Finally, the preliminary results are presented, and the implications of this study are highlighted.

Keywords: UGC, social media, user’s classification model, motivation studies, two-factor theory, contribution behavior

1. Introduction
Social media refers to those websites which build on Web 2.0 technologies to provide space for in-depth social interaction, communication, and collaboration. User-Generated Content (UGC), also known as User-Created Content (UCC) or Consumer-Generated Media (CGM), is a prevailing mode for creation and contribution of information resources in social media sites, such as Blogs, Wikis, Podcasting, online reviews, Social Networking Services (SNS), etc[9]. In 2006, Time Magazine selected you, especially those people who contribute to social media, as its esteemed Person of the Year. Meanwhile, some challenges and threats cannot be overlooked, such as the low entry barrier, the ever-decreasing cycle of technology reproduction, and the fierce competition in digital content industry. The question thus arises: How to attract user’s participation and motivate them to share and contribute more contents in social media. The question is of great importance to today’s media, internet and E-business researchers and practitioners. However, there are few studies that have systematically addressed them. This paper attempts to fill this research gap by exploring the user’s classification model based on the extent of user’s contribution, and further examines what factors will influence user’s motivation to participate in UGC with the consideration of different user groups.

2. Brief Literature Review
Generally, UGC can be split into two parts, i.e. users and the contents produced by them. If we take a closer look at the users mentioned here, it is natural to view it as a combination of consumer and producer from a marketing perspective. Consumer refers to the individuals who only read or watch but never post anything. Some studies also call them lurkers [11]. Producer encompasses creation and pub-
lication of one’s personal contents such as text, images, audio, and video. Furthermore, some studies go deep into the classification of users in online environment. Preece proposed a reader-to-leader framework by literature review, in which more color was added to the concept of user, including reader, contributor, collaborator, and leader [12]. Porter’s Funnel Model also identified four types of users, i.e. interested readers, first-time users, regular users, and passionate users based on the extent of user’s participation [10]. Nakakoji defined users’ roles in OSS communities as 8 groups, included passive user, reader, bug reporter, bug fixer, peripheral developer, active developer, core member, project leader [7]. Mao used the data mining techniques to illustrate the user’s categories in online community as reader, learner, responder, contributor, and leader [2].

Although few papers can be retrieved by the literal combining of “UGC” and “influence factors”, many studies have high relevance with our research topic actually. Some researchers pay attention to the motivation of user’s participation. Benkler stated that user’s motivations in UGC are not driven by monetary reward (at least not in any direct way), and much emphasis should be placed on social motivations, such as the desire for attention, reputation, credibility, etc [1]. Some researchers in information field have also studied the motivational factors for socio-technical systems in order to facilitate user’s adoption towards ICTs [16]. Some researchers investigated the motivations for blogging [14], tagging [6], rating[3] and video sharing[5]; while others explored why mass users participate in wiki collaboration[8] and OSS campaign[13]. The extant researches indicate that many factors have significant influence on users’ motivations to share, contribute, and collaborate in online setting, such as perceived enjoyment and fun, curiosity, self-efficacy, altruism, reciprocity, rewards, social identification and recognition, and sense of belonging, to name a few.

3. Classification Model for UGC in Social Media

3.1. Data Collection

In order to encourage the contribution of content, many social media sites present the contributor’s username (real name or pseudonym) together with the content metadata. Given the various types and granularities of UGC, we mainly collect the user’s ID and its contribution number. Two criteria were set up: 1) the site is representative enough of a specific UGC type, for example, text based contents, video or audio based contents, etc; 2) the available of data, although many social media sites offer the API, there are some limitations when actually using it. Therefore, we developed a scrapping tool named Unified UGC Crawler System in JAVA. It can easily add new crawling and filtering rules based on the difference of target data. In addition, multitask program can be paralleled and worked independently. In this study, we selected 8 social media sites for data collection and analysis, including Flickr, Delicious, YouKu (a Chinese video publication site), YouTube, Amazon Book Reviews, Fan Fiction (literary publication), CiteULike and Yahoo! Answers. Data were collected between 09/01/2009 and 10/31/2009.

3.2. Data Analysis

A simple descriptive statistical analysis is conducted to reveal the distribution condition of UGC. We analyze the random variable that representing the number of items produced by each user, and analytically test if the distribution is symmetric around a central value or if it presented a long tail. Some values are selected for
this analysis, such as mean, quartiles, minimal, maximal and Standard Deviation (SD). Those values are presented in Table 1 for each of the 8 empirical data sets. As it can be seen, the values of quartiles present a very asymmetric distribution for all the data sets. The mean values also provide evidence for the inequality. In the cases of Yahoo! Answer and YouTube, the mean value is far larger than the upper quartile value. Moreover, we also notice that there are many lurkers in some sites, such as YouTube. The counted number of ID is 30486, and among them, 17017 users have contributed at least one content, which means that there are more than 13469 users are lurkers. The comparison of the mean and quartiles shows that the distribution of contribution is very unbalanced, which illustrates that there is no so called normal distribution in UGC.

The empirical data could be represented through its cumulative mass function (CDF), which can be used to estimate the differences in contribution from different groups of users. The CDF value is taken at fixed proportions of the number of sources. These values reflect which proportion of contents has been contributed by the corresponding segment of users. Table 2 shows the cumulative contribution of different user segments at 0.1%, 1%, 10%, 20%, 40% and 80% of the total users. An interesting finding is that there is no such thing as an average user in social media sites. In the case of Yahoo! Answer, the 1% users almost contribute half of the content (47.40%), and the contribution from the 20% users’ amount to 93.78% of the total contents, which is fairly higher than what Pareto rule tells. In the case of CiteUlike and YouTube, the 80/20 rule is a good rule-of-thumb to illustrate the distribution of the contribution of UGC. Very similarly, the 40% users contributed about 90% contents, which indicate that there are many inactive users in those social media sites. Furthermore, among the 8 sites, the contribution from the 80% users nearly amount to 100% contents, which reflect a zero contribution from the rest 20% users. Meanwhile, we also notice that there exists “fat-belly” in some sites, such as Flickr, Delicious, YouKu, and Fan Fiction, which means except the prolific users, a group of middle users also continually contributed to the social media.

| Sites      | User | Contrib. | Q1   | Q2   | Q3   | Min | Max | Mean | SD  |
|------------|------|----------|------|------|------|-----|-----|------|-----|
| Flickr     | 7133 | 11109497 | 154  | 482  | 1551 | 5   | 264099 | 1557.48 | 4993.742 |
| Delicious  | 33725| 35913367 | 218  | 560  | 1256 | 1   | 39033 | 1064.89 | 1617.793 |
| YouKu      | 2736 | 84656    | 4    | 12   | 27   | 1   | 685   | 30.94  | 72.743   |
| YouTube    | 17017| 2673257  | 5    | 19   | 91   | 1   | 26125 | 157.09 | 568.085  |
| Amazon     | 9988 | 1753374  | 49   | 89   | 177  | 2   | 20603 | 175.55 | 363.088  |
| Fan Fiction| 31295| 300651   | 3    | 6    | 11   | 1   | 848   | 9.61   | 15.437   |
| CiteUlike  | 5385 | 816935   | 5    | 30   | 124  | 1   | 35810 | 151.71 | 719.699  |
| Yahoo! Answer | 29839| 5380690 | 3    | 11   | 50   | 1   | 98101 | 180.32 | 1209.151 |

Table 1: Simple Descriptive Statistical Results of Collected Data for UGC.
As it shows in Table 2, although there are some differences in the user’s contribution among the 8 social media sites, similarities can still be found according to the statistics. Firstly, in all cases, the 10% users contributed 40% contents above, and in some particular cases, such as Yahoo! Answer and YouTube, even the 0.1% or 1% users have a marked contribution. In view of the generalization, we define the 10% users as core contributors. Secondly, the statistics also show that the 40% users contributed at least 70% contents. If we eliminate the contribution from the 10% users, 30% of the total contribution still comes from this user group, which indicates that the users in this category play an active role in contribution, thus we call them active contributors (10%-40%). Thirdly, the contribution from 40%-80% users varied greatly from 5% to 25% around, which reflects some fluctuation and randomness in this group, thus we define it as peripheral participants (40%-80%). Finally, as what mentioned above, the rest 20% users contributed very limited contents, so we call them lurkers (the rest).

As shown in Figure 1, the paper proposes a four-level user classification model based on the different extent of contribution. The model has a contribution to the concept of user segmentation in social media because it is constructed on the basis of various types and granularities of UGC, which to some extent, can reflect some common grounds of user’s contribution structure. The model can roughly be categorized as lurkers, peripheral participants, active contributors, and core contributors, from the outer to inner ring. The emoticons in the model have some metaphors. The eyes indicate that those users just keep an eye on the contents produced by others without any real contribution. The face indicates that those users have their mouths and can do some easy or elementally contributions such as rating, voting, and comment. The face with hands means that they are more powerful and active contributors in the social media community. The whole body of people shows that they play a leading role in the contribution, and of course they are the core participants in social media sites. It can be clearly seen from the model that the number of people is decreasing from outer to the inner. There are also some positive and negative evolving trends in our model, which indi-
cates that users can change their roles of participation through their effective interaction, contribution and collaboration in the communities and the evolvement can move in either a linear or a non-linear fashion.

Fig. 1: Classification Model for UGC Based on Contribution in Social Media.

4. Influence Factors for UGC Based on Contribution Model

4.1. Theoretical Basis

For this study, the researchers adopt the dual structure of Herzberg’s two-factor theory [4]. Herzberg’s theory is based on the description of situations when professionals such as engineers and accountants felt exceptionally good or bad about their jobs. Some researchers argue that there are several useful aspects of Herzberg’s theory for website studies. We assume that the distinction between satisfaction and dissatisfaction as two dimensions rather than two values of the same dimension. Thus, relating factors to perceived satisfaction and dissatisfaction is a way of examining the factors in the online environment. Zhang has examined the two-factor model as a powerful theory to study the design and evaluation of websites [17]. In this study, the researchers propose a two-factor model for UGC that is analogous to Herzberg’s two-factor theory in the work place. Accordingly, there should be basically two types of influence factors on UGC: hygiene and motivator.

4.2. Research Approach

Few existing studies provide either theoretical frameworks or heuristics of examining the relative importance among various UGC influence factors. Thus, we
decided to use an inductive grounded theory approach. This approach starts on the collected data from related literature review and interview. The researchers developed the initial themes or codes by extracting meaningful words or phrases from a subset of literature or users’ answers. An iterative process refines and validates a code schema. In this study, we used the code schema we have developed in our previous research [15], and selected the most relevant factors as in Table 3.

Second, a survey was conducted to investigate users’ attitudes and intentions towards the hygiene and motivational factors in UGC. In this paper, we mainly focused on the core contributors, active contributors and peripheral participants because they have more knowledge and experience on sharing and contribution in social media sites. A questionnaire was developed based on Table 3, and we used -1, 0, 1 to indicate the hygiene factors, not sure and motivator factors. We explained the two-factor theory in particular in the questionnaire, and listed some related examples. The survey started from 12/01/2009 to 01/31/2010. The well-known Chinese video sharing and publication site Youku was selected as the empirical platform, which has a large number of users and enjoys a high reputation in Chinese social media community. We handed out the questionnaires to those users with active ID. Finally, we received 210 valid responses in total and 70 for each group respectively.

4.3. Preliminary Findings

After calculating users’ judgments on motivational and hygiene factors from three user groups, we illustrate them in Figure 2-4 at the threshold of 20 in order to achieve stable results.

The preliminary results indicate that there are certain factors that are perceived equally important among three user groups, such as External Rewards as a
commonly recognized motivation factor, and Information Architecture, Ease of Use, Human-Computer Interaction, and Information Privacy & Security as commonly recognized hygiene factors; meanwhile, there are other factors that are regarded as extremely important for one user group and not significant for another. For example, some factors played opposite roles in different groups of users’ viewpoints, such as Usability, Interface Design, Social Interaction, and Sense of Belonging. Moreover, some factors may not appear in some users’ viewpoints while demonstrated in other users’ judgments, such as Challenge, Altruism, and Responsibility & Obligation. Therefore, the hygiene and motivational factors are not absolute concepts but changing with different uses.

5. Conclusions

Recently, the focus of internet economy is changing from the attention to the creation, thus UGC is playing an increasingly important role in this transition. This paper proposes a contribution model based on user’s participation, and then explores the hygiene and motivational factors in three segments of users. The preliminary results indicate that there is no average contributor in social media sites, and different groups of users do have various judgments towards the hygiene and motivational factors on UGC. The study provides empirical evidences for website designers, managers, and service providers about what factors are more important to focus on when dealing with different groups of users, thus leading to a better understanding on the incentive mechanism towards different target users.

Acknowledgment

We are grateful for the financial support from the National Social Science Foundation in China (Grand No.10ATQ004), Ministry of Education, Humanities and Social Sciences Council in China (Grand No. 09YJA870014) and Bureau of Education, Jiangsu Province, China (Grand No: CX09B_027R and CX10B_022R).

References

[1] Benkler, Y., H. Nissenbaum, "Commons-based peer production and virtue", Journal of Political Philosophy (14) 4, pp. 394-419,2006.
[2] Bo Mao, Wenwen You, "Classifying model for virtual community members", Journal of Tsinghua University (Science and Technology) (S1) 46, pp. 1070-1073,2006. (in Chinese)
[3] Dellarocas, C., N. Awad, and M. Zhang, "Using online ratings as a proxy of word-of-mouth in motion picture revenue forecasting", Working Paper, University of Maryland, 2005.
[4] Herzberg, F., "One more time: How do you motivate employees?", Harvard Business Review, January-February, 53-62,1968.
[5] Molyneaux, H., O’Donnell, S., Gibson, K., Singer, J, "Exploring the Gender Divide on YouTube: An Analysis of the Creation and Reception of Vlogs", American Communication Journal (2) 10, pp. 212-220,2008.
[6] Morgan Ames, Mor Naaman, "Why We Tag: Motivations for Annotation in Mobile and Online Media", Proceedings of the SIGCHI conference on Human factors in computing systems, pp. 971-980, 2007.
[7] Nakakoji, K., Y. Yamamoto, Y. Nishinaka, K. Kishida, and Y. Ye, "Evolution Patterns of Open-Source Software Systems and Communities", Proceedings of International Workshop on Principles of Software Evolu-
Nov, Oded, "What motivates Wikipedians", Communications of the ACM, (11) 50, pp. 60-64, 2007.
[9] http://www.oecd.org/document/40/0,3343,en_2649_34223_39428_481_1_1_1,00.html.2007.
[10] Porter, J, "Designing for the Social Web", Berkeley, CA: New Riders, 2008.
[11] Preece, J., B. Nonnecke, and D. Andrews, "The top five reasons for lurking: Improving community experiences for everyone", Computers in Human Behavior (2) 20, pp. 201-223, 2004.
[12] Preece, Jennifer, Shneiderman, Ben, "The Reader-to-Leader Framework: Motivating Technology-Mediated Social Participation", AIS Transactions on Human-Computer Interaction (1), pp. 13-32, 2009.
[13] Yunwen Ye, Kouichi Kishida, "Toward an Understanding of the Motivation of Open Source Software Developers", Proceedings of International Conference on Software Engineering (ICSE 2003), Portland, OR, May 3-10, 2003.
[14] Yuxiang Zhao, Qinghua Zhu, "Blog acceptance model: An empirical study on exploring users’ acceptance and continual usage of blogs", Chinese Journal of Library and Information Science (3) 2, pp. 44-61, 2009.
[15] Yuxiang Zhao, Qinghua Zhu, "Exploring the motivations to produce User-Generated Content in Web 2.0", Journal of Library Science in China (in Chinese) (5) 29, pp. 107-116, 2009.
[16] Zhang, P, "Motivational affordances: Reasons for ICT Design and USE", Communications of the ACM (61) 11, pp. 145-147, 2008.

Zhang, Ping, Gisela von Dran, "Satisfiers and dissatisfies: a two-factor model for website design and evaluation", Journal of the American Society for Information Science (JASSIS), (51) 14, pp. 1253-1268, 2000.