

User Contribution Measurement in Online Forum with Fraud Immunity

Guo-Ying WANG¹ and Shen-Ming QU²

¹ Information Engineering College, Zhejiang A&F University
Hangzhou, 311300, China

² Computing Center, Henan University
Kaifeng, 475004, China

Abstract

It's very important to reward the contributive users of online forums, for that almost all contents are provided by users in such forums. There should be some rewards for contributive users, and rewards should be proportional to the contributions. So the determination and measurement of user contributions are needed in online forums. At the same time, some users may do some fake contributions to obtain more rewards. In this paper, we analyzed possible frauds in online forum, examined features of each kind of fraud, and proposed some fraud-tolerant parameters according to the features of frauds. Results of our experiment show that almost 81% users in the examined online forum have fraudulent activities and pure advertising users can be discovered according to the fraud-tolerant parameters we considered. On the other hand, the experiment results also show that the biggest count of fraud type detected is with the parameter minimum intervals of posts from the same users, and followed by the parameter minimum length of posts. While minimum average rate value of post after specified rates is the parameter that was used for least times. Based on the idea of this paper, frauds of user contributions could be discriminated well, and user contributions can be measured quantitatively and fraud-tolerantly, which provides a basis for online forums to reward users in various ways.

Keywords: *contribution measurement, fraud immunity, online forum.*

1. Introduction

Success of an online forum depends on two key aspects: the forum infrastructure and the contents. The contents are all provided by forum users. To encourage users to provide more valuable contributions, the forums usually give some rewards to their contributive users.

A good contribution-reward mechanism can motivate users to provide more valuable contents. Obviously, the rewards should be proportional to the contributions. But currently most forums do not possess a good approach to measure the contributions of their users. On the other hand,

some users do some fake contributions to cheat more rewards. How to discriminate the fraud from contributions is another problem.

The purpose of this paper is to model user contributions with some features of user actions, measure user contributions in a quantitative way, and provide a basis for online forums to award their users in various ways.

The rest of the paper proceeds as follows. Section 2 describes some existing technologies of measuring and motivating user contributions in different environments. Section 3 lists the possible frauds in online forums. Section 4 gives the fraud-tolerant parameters according to each feature of frauds, and then describes the fraud-tolerant user contributions measurement method. Section 5 describes a user contribution measuring experiment and its results. Section 6 concludes the paper.

2. Related Works

Some researches have been done about user contribution motivation and measurement.

Cheng [1] introduced hierarchical membership levels (gold, silver and bronze) to motivate user contributions in P2P communities, and five contribution relevant factors are used to measure user contributions.

Abtoy [2] proposed a model that supports monitoring the quality of content according to its life on the Web. The model emphasized on the prioritizing information and users both.

One simple way of measuring user contribution is allowing users to rate the quality of contents provided by other users. A similar method is a member-controlled reward mechanism, that is, users who posted questions

could rate the quality of other users' answers [3].

Klamma [4] evaluated user contributions with 5 factors such as postings, replies to posts, post ratings, replies to their posts and post ratings received. Chai [5] proposed a model of user contributions measurement with 16 factors, and user contribution score are calculated by summing up the values of 16 factors with different weights individually. This method didn't consider the possible frauds, and a user with many junk posts or advertisement posts may be considered more contributive.

Shi [6] studied the patterns of user participation behavior, and the feature factors that influence such behavior on different forum datasets, found that users' community joining behaviors display some strong regularities, built social selection models, Bipartite Markov Random Field (BiMRF), quantitatively evaluated the prediction performance of those feature factors and their relationships, and showed that we show that some features carry supplementary information, and the effectiveness of different features vary in different forums.

Wikipedia is a website in which all contents are provided and edited by users from the whole world voluntarily. Adler [7] considered the problem of measuring user contributions to versioned, collaborative bodies of information, such as wikis, considered and compared various alternative criteria that take into account the quality of a contribution, in addition to the quantity, and analyze how the criteria differ in the way they rank authors according to their contributions, proposed to adopt total edit longevity as a measure of author contribution. Edit longevity is resistant to simple attacks, since edits are counted towards an author's contribution only if other authors accept the contribution.

3. Possible Frauds in Forums

In some online forum, many particular actions of users, such as postings and replies to posts, are considered as contribution. In order to be rewarded as contributive ones or for other purposes, some users may posts as many topics as possible, even useless topics. Table 1 lists some fraud types of fraud usually occurred.

3.1 Junk posts

Junk posts are useless and have no contribution to a forum, which is interesting to few users. In worst case, junk posts may makes users leave the forum if where is full of junk posts. Junk posts have some common features as follows.

Table 1: Fraud Types in Online Forums

<i>Code</i>	<i>Fraud Types</i>
T1	Junk posts
T2	Advertisement posts
T3	Advertisement messages
T4	Dummy replies
T5	Dummy rates

Short length: Most junk posts are created by users who just pursued the counts he posted, are always very short and even zero length.

Short interval: Junk posts by the same user are usually posted with short intervals so as to achieve a large number of posts in short time.

Few views, few replies: Because of the useless of junk posts, they can't incur others users' interest and obtain few view.

Low rated or Judging replies: As to such junk posts, common users surely rated them low or give judging replies such as "junk", "useless", and so on.

3.2 Advertisement posts

Some users join forums in order to post advertisements. Such posts are similar to junk posts: they have show interval, few views and replies and are low rated.

Short interval in different subforums: To achieve as many viewers as possible, the advertiser users usually post their advertisements in several even all subforums. The advertisements are prepared well, and just copied and pasted when posting, and the post interval is short in many cases.

Not empty and the same contents: Advertisement posts are usually copied and pasted using a prepared template, so they are not empty and have the same contents.

Few views, few replies: Few users are fascinated by advertisement posts and give replies.

Low rated or Judging replies: Advertisement posts usually are low rated and even receive some replies that give some advertising judgments.

3.3 Advertisement messages

Users can send personal messages to other users in most forums. Sending messages means user's active level and is

also considered as a feature of user contribution. Some advertisement users do not post advertisements in forum, but send advertisements as personal message to other users. Such messages can be featured as following.

Sent to many users: Advertisement messages are usually sent to as many users as possible to achieve more viewers.

Short interval: The advertiser users tend to send advertisement messages with short intervals.

Not empty and the same contents: A advertisement sent to many users is not empty and with the same content.

3.4 Dummy replies

To achieve more user contribution score or pretend not to be a junk or advertisement post, a user may forge some dummy replies to a topic he/she posted. These dummy replies may be produced by the same account, other accounts of the same user or accounts of the user's friends. Dummy replies can not be distinguished easily.

Dummy replies can be featured as many replies from a few users (may including topic owner). When a user want to forge replies, only few dummy replies is useless to forge contribution. So the amount of dummy replies of a topic is usually not a small number. But the involved user accounts are from few users.

3.5 Dummy rates

To influence the rate result, a user may also forge dummy rates, while which can be identified as a very few rates deviate from dominate rates if there are many rates.

4. User Contribution Measurement

In this section we will describe our method of fraud-tolerant user contribution measurement.

4.1 Common features

Here we use the features listed in Chai's paper [5], which are shown in table 2. Using these features, we can only measure user contribution without the consideration of frauds in the contributions.

4.2 Fraud-tolerant parameter

Now we introduce fraud-tolerant capability into our user contribution measurement method, and some parameters are introduced. As is shown in table 3, in which the denotation of each fraud is the same to that in section 3.

Table 2: Common Features

<i>Code</i>	<i>Name</i>	<i>Weight</i>
F1	# of posts created by a user	1
F2	# of voting polls created by a user	4
F3	# of votes cast by a user	1
F4	# of questions asked by a user	1.5
F5	# of questions answered by a user	2
F6	# of topics created by a user	1.5
F7	# of sticky topics created by a user	4
F8	# of topics that the user has provided the first reply	2
F9	# of responses received user topics	1.5
F10	# of views received for user topics	0.1
F11	# of personal messages sent	0.1
F12	# of personal messages received	0.2
F13	# of topic update notifications	0.1
F14	# of board update notifications	0.1
F15	# of quality posts created	3
F16	Frequency of user posts	3

Threshold of post's minimum length (P1): The parameter is used to resolve the "short length" problem of junk posts. It is not used when a user do a post, but used when calculated a user's contribution, that is, subtracting the counts of posts whose length is shorter than this threshold from the total posts by the user.

Threshold of post's minimum interval (P2): This is according to "shot interval" feature of junk and advertisement posts. A post is considered as junk or advertisement posts and is omitted when its post time and the former post's time by the same user is too close to the threshold.

Threshold of post's minimum views in its first X hours (P3): This is used to check posts with "few views" to discriminate junk posts or advertisement posts.

Threshold of post's minimum replies per Y views (P4): This is used to discriminate junk posts or advertisement posts with "few replies", based on we considered that normal topics may be replied at least once every Y views.

Threshold of post's minimum average rate value after Z rates (P5): This is according to the "low rated" feature of junk posts or advertisement posts. Here we considered that first Z rates can't represent the real average rates for the small samples count.

Table 3: Fraud-Tolerant Parameters

Code	Fraud-Tolerant Parameter	Frauds type	Fraud feature
P1	Minimum length of post	T1	Short length
P2	Minimum interval of post	T1, T2	Short interval
P3	Minimum views of post in the first X hours	T1, T2	Few views
P4	Minimum replies of post per Y views	T1, T2	Few replies
P5	Minimum average rate value of post after Z rates	T1, T2	Low rated
P6	Maximum proportion of filter-matched replies	T1, T2	Judging replies
P7	Maximum users be sent messages by a user continually without responses	T3	Sent to many users
P8	Maximum messages sent by a user per minute	T3	Short interval
P9	Maximum message sent by a user with same contents	T3	The same contents
P10	Maximum replies per user	T4	Many replies from a few users
P11	Minimum proportion of a rate value	T5	Few rates deviate from dominate rates

Threshold of post's maximum proportion of filter-matched replies (P6): When too many replies contain specified words such as "junk", "useless", "advertisement", this post is considered as a junk or advertisement post and should not be counted in.

Threshold of maximum users be sent messages by a user continually without responses (P7): This is used to discriminate advertisement messages on the basis of "sent to many users" feature. When a user sends too many messages continually without any response, the user is considered as sending advertisement.

Threshold of maximum messages sent by a user per second (P8): This is used to discriminate advertisement messages on the basis of "short interval" feature.

Threshold of maximum message sent by a user with same contents (P9): This is used for the "the same contents" feature of advertisement messages.

Threshold of post's maximum replies per user (P10): This is used to distinguish dummy replies.

Threshold of post's minimum proportion of a rate value (P11): This is used to find dummy rates, which is on the basis of that the proper rate of post should not be given by only a very small proportion of replies.

4.3 Fraud-tolerant user contribution measurement

In Chai's paper [4], the user contribution score (*UCS*) is calculated by summing up all features with different weight, as is shown in equation 1, in which *u* means a user, *m* means the count of features, *f_{iu}* means the *i*th feature value of user *u*, *w_i* means weight of the *i*th feature.

$$UCS_u = \sum_{i=1}^m W_i f_{iu} \quad (1)$$

According the fraud-tolerant parameters, we describe the fraud-tolerant user contribution score (*FUCS*) as equation 2, in which *f'_{iu}* means the result of *f_{iu}* subtracts count of posts or messages that were consider as frauds using corresponding threshold of fraud-tolerant parameters in table 2.

$$FUCS_u = \sum_{i=1}^m W_i f'_{iu} \quad (2)$$

5. Experimental Evaluation

We examined the contributions of 200 random users in an online forum in one month. In this forum, contributive users are rewarded with virtual coins. Coins can be used to ask for help in the forum, and users are granted different titles representing different levels according to the count of coins. So some users would like to post more topics or give meaningless replies to gain more coins.

Using the features and weights in table 2, we computed the *UCS*s of these users according to formula (1). Then according to the fraud-tolerant parameters in table 3 and values of each parameter in table 4, the *FUCS*s were computed using formula (2). Results are shown in figure 1.

Table 4: Experimental Values of Fraud-Tolerant Parameters

Parameters code	Values
P1	50 chars
P2	60 seconds
P3	20 views (X=1)
P4	1 replies (Y=10)
P5	C level (Z=10)
P6	40%
P7	20
P8	20
P9	5
P10	5
P11	5%

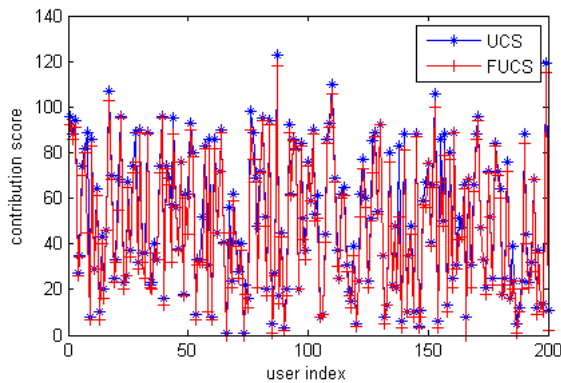


Fig. 1 UCSs and FUCSs of 200 randomly chosen users in an online forum during one month.

From the results of experiment we found that almost 81% users got a bigger UCS than FUCS, as is shown in figure 1, the reason of which is that users may do some frauds unconsciously for the allurements of rewards. A few users got very low UCS and even zero FUCS, which may be considered as a discriminative result of pure advertising users.

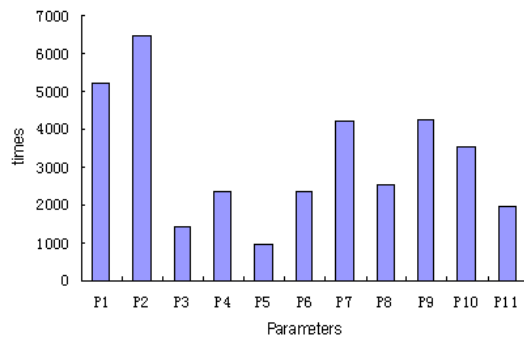


Fig. 2 Times of frauds detected using each parameter in an online forum during one month

Figure 2 shows the result of frauds detected with all parameters. As can be seen that the biggest count of frauds is detected using the parameter P2, minimum intervals of posts from the same users, and followed by the parameter P1, minimum length of posts, while P5 is the parameter that is used for least times.

6. Conclusions

In this paper, several possible frauds of user contribution in online forum are examined, and according to which, corresponding parameters are considered in the measurement of user contribution. These fraud-tolerant

parameters can discriminate frauds from real user contribution. The result of this paper can be used in all kinds of online forum to build fair and effective contribution-reward mechanisms.

Acknowledgments

This work is supported by Zhejiang Provincial Natural Science Foundation of China (LY12F02016), and Henan provincial Natural Science Foundation of China (112300410009; 112300410125).

References

- [1] R. Cheng and J. Vassileva, "User Motivation and Persuasion Strategy for Peer-to-peer Communities", in Proc. of the 38th Annual Hawaii International Conference on System Sciences (HICSS' 05), 2005, pp. 193a.
- [2] A. Abtoy, N. Aknin, B. Sbihi, A. El Moussaoui and K.E. El Kadiri, "Content validation as a tool for new pertinent Web 2.0 Blogs", International Journal of Computer Science Issues, Vol. 9, Issue 3, No 3, May 2012, pp. 146-151.
- [3] C. Wiertz and K. Ruyter, "Beyond the Call of Duty: Why Customers Contribute to Firm-hosted Commercial Online Communities", Journal of Organization Studies, vol. 28, no. 3, 2007, pp. 347-376.
- [4] R. Klamma, M. A. Chatti, E. Duval, H. Hummel, E. T. Hvannberg, M. Kravcik, E. Law, A. Naeve, and P. Scott, "Social Software for Lifelong Learning", Journal of Educational Technology & Society, vol.10, no. 3, 2007, pp. 72-83.
- [5] K. Chai, V. Potdar and E. Chang. "User Contribution Measurement Model for Web-based Discussion Forums", In Proc. of 3rd IEEE International Conference on Digital Ecosystems and Technologies, June 2009, pp. 347-352.
- [6] X. Shi, J. Zhu, R. Cai, and L. Zhang. "User Grouping Behavior in Online Forums", in Proc. of the 15th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD 2009), 2009, pp.777-786.
- [7] B. Thomas Adler, L. Alfaro, I. Pye, V. Raman, "Measuring Author Contributions to the Wikipedia", Technical Report UCSC-SOE-08-08, School of Engineering, University of California, Santa Cruz, CA, USA. May 2008.

First Author Guo-Ying WANG received the B.S. degree in computer science from Beijing JiaoTong University, Beijing, China, in 1999 and M.S. degree in computer science from Guangxi University, Nanning, China, in 2004. In the same year, He joined the faculty of the Computer Science Department, Information Engineering College, Zhejiang A&F University, where he is currently a lecturer. His research interests include computer networks, peer-to-peer networks, wireless ad-hoc and sensor networks, and mobile networks.

Second Author Shen-Ming QU received the B.Eng. degree from Hebei University and the M.S. degree from Henan University. Currently, he is a lecturer in Computing Center, Henan University. His research interests include Computer Network, information retrieval and computer vision.