Adaptive Reward Mechanism for Sustainable Online Learning Community

Ran Cheng and Julita Vassileva
Computer Science Department, University of Saskatchewan,
Saskatoon, SK, S7N 5A9 Canada
rac740@mail.usask.ca , jiv@cs.usask.ca

Abstract. Abundance of user contributions does not necessarily indicate sustainability of an online community. On the contrary, excessive contributions in the systems may result in information overload and user withdrawal. We propose a user- and community- adaptive reward mechanism aiming to regulate the quantity of the contributions and encourage users to moderate the quality of contributions themselves. The mechanism has been applied and evaluated in an online community supporting undergraduate students to share course-related web-resources.

Introduction

The proliferation of online communities may lead people to the conclusion that the development of custom-made communities for particular purpose, for example, to support a class, is straightforward. Unfortunately, this is not the case. Although software providing basic community infrastructure is readily available, it is not enough to ensure that the community will "take off" and become sustained. For example, the multi-agent-based synchronous private discussion component of the I-Help system [1] did not enjoy much usage by students and was abandoned in favor of the more traditional asynchronous public discussion forum [2]. A critical mass of user participation was missing in the private discussion forum since the students did not stay constantly logging in the system.

Comtella[*] [3] is a small-scale peer-to-peer online community developed at the MADMUC lab at University of Saskatchewan for sharing academic papers and class-related web-resources among students. Comtella, just like I-Help, depends on a critical mass of participation both in terms of quantity and quality of contributions. Our previous work [4, 5] addressed the problem of motivating students to bring new resources in the system. To achieve a sustainable critical amount of participation, this paper proposes a new adaptive reward mechanism to encourage users to rate contributions thus ensuring decentralized community moderation. The mechanism adapts to the current needs of the community in terms of the number of contributions and also to the individual trends/preferences in the type of contributions of each individual member.

1. Previous work

The problem of ensuring user participation is very important for all online communities [6]. The "critical mass" hypothesis proposed by Hiltz and Turoff [7] states that a certain number of active users have to be reached for a virtual community to be sustained. Our experience with Comtella confirms this hypothesis. In order to stimulate users to make contributions we looked into Social Psychology, specifically in the theories of discrete emotions and of social comparison. We proposed, implemented and evaluated in a case study [5] a motivational approach based on hierarchical memberships in the community (gold, silver, and bronze), awarded to users depending on the quantity of their contributions to the community. The different memberships implied different privileges and prestige in the community. While the case study of using the system to support an Ethics and IT class showed that the motivational strategy effectively increased the number of user contributions, it also seemed to motivate a small number of users to game the system to achieve higher membership levels. They shared many resources that were of poor quality or unrelated to the topic. This made it hard for users to find good resources in the system, resulting in the decreased level of participation in the last week of the study and disappointment reflected in negative user comments in the post-experiment questionnaire.

Our observations mirror the phenomenon called “information overload” [8], which has arisen in some other online communities. It makes users feel swamped by a mass of unwanted information. Jones and Rafaeli [9] found that the users’ most common response is to end their participation in the community, both as contributors and as consumers. Therefore, to create a self-maintaining community, it is necessary to avoid the information overload by controlling the overall number of contributions in the system, motivating users to contribute high-quality resources and simultaneously inhibiting the contribution of poor-quality resources. Therefore, a mechanism of measuring the quality of user contributions is needed.

It is difficult to measure the value of user contributions accurately since quality measures are mostly subjective. Centralized moderation is feasible only for small and narrowly focused communities, where members have very similar evaluation criteria. Therefore, decentralized mechanisms for quality measurement are necessary. One way of evaluating the quality of resources used in online communities like Slashdot [10] is through explicit user ratings. The mechanism has two merits. Firstly, it distributes the task of evaluating resources among the large user pool, thereby making achievable a job that would otherwise have been overwhelming. Besides, the final ratings of resources are more unbiased since they are computed based on ratings from many users. However, a study of the Slashdot rating mechanism [11] showed that some deserving comments may receive insufficient attention and end up with an unfair score, especially the ones with lower initial rating and those contributed late in the discussion. Therefore the timeliness of making a contribution is important and a motivational mechanism should encourage early contributions. This is especially relevant in a class-supporting system like Comtella, or I-Help, since the discussion topic typically change on a weekly basis according to the class curriculum. When the topic is new, it is important to have more contributions, but later it is important to have more ratings to help users cope with the information overload. The needs of the community change in time. Therefore, a motivational mechanism needs to adapt to the dynamic needs of the community and encourage users to contribute early.

The Slashdot study [11] also showed that comments starting their life at a low initial rating have a lower chance to be viewed and rated and are more likely to end up with unfair score. In Slashdot, the initial rating depends on the “karma” of the user who made the comment. The user’s “karma” is his/her reputation for contributing high-quality comments, measured by the ratings his/her previous comments collected. In this way, good comments made by new users or the users who haven’t contributed highly rated comments so far tend not to receive a deserving attention and to collect sufficient ratings to raise the “karma” level of their contributor. This causes a feedback loop resulting in the Matthew effect [12] or “the rich get richer”. A fair rating mechanism should give all contributions an equal chance at start.

A challenge in systems that rely on decentralized moderation is to ensure that there are enough user ratings. MovieLens tried to motivate users to rate movies by sending them email-invitations [13]. The results showed that users seemed to be influenced more by personalized messages emphasizing the uniqueness of their contributions and by messages that state a clear goal (e.g. number of movies the user should rate). While this approach is questionable as a long-term solution because the effect of receiving email will likely wear off, it is interesting that personalization seems important and that setting specific goals are more persuasive than general appeals. To stimulate users to rate resources constantly, persistent incentives are necessary.

Our previous case study showed that different people had different contribution patterns. Some contribute many, but average (or even poor-quality) resources, while some contribute few, but very good ones. An adaptive motivational mechanism should encourage the users of the second category to contribute more resources unless the quality of their contributions starts to drop and inhibit the contributions from the users of the first category unless the users improve the quality of their contributions. The motivational mechanism should make users regard the quality and the quantity of their contributions equally.

Based on the discussion above, a collaborative rating system is introduced into the Comtella system, through which users can rate the resources in the community. The adaptive reward mechanism is designed based on the quality data from user ratings.

2. Collaborative rating

The Comtella rating mechanism is inspired from the Slashdot moderation system. In order to have a broader source of ratings, all the users can rate others' contributions by awarding them points (either +1 or -1). However, the users with higher membership levels receive more points to give out and are thus more influential in the community. To ensure that contributions have an equal chance to be read and rated initially, the initial rating for every new contribution is zero regardless of its providers' membership level or the quality of his/her previous contributions. In the end, the final rating for the contribution is the sum of all the ratings it has obtained. The summative rating for each contribution is displayed in the list of search results (Fig.1).

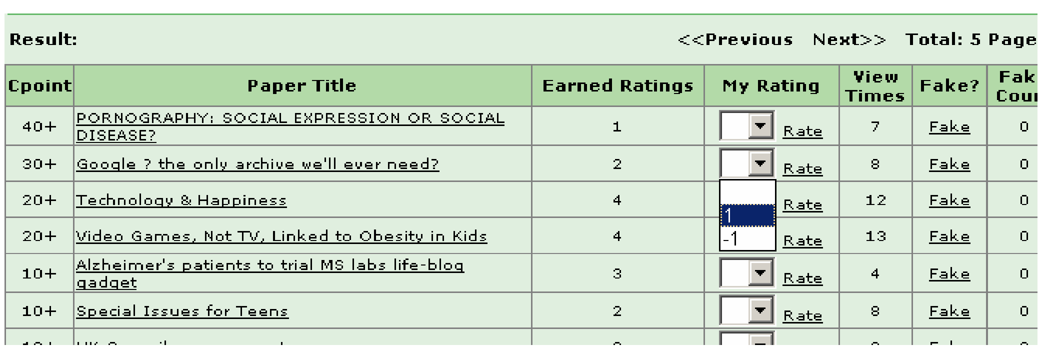

Fig. 1. A segment of a search result list

As a persistent incentive for users to rate contributions, a virtual currency is introduced, called "*c-point*". Whenever a user rates an article, he/she is awarded a certain number of c-points, depending on his/her reputation of giving high-quality ratings. The user can use the earned *c-points* to increase the initial visibility of his/her postings in the search result list. Most users desire that their contributions appear in salient positions, e.g. in the first place or among the top 10, because in those positions they will have a better chance to be read and rated. The Comtella search facility displays all the contributions matching a query in a sorted list according to the number of *c-points* allocated by the contributors (Fig.1). Unlike the mechanism in Slashdot, this one allows the user flexibility to invest c-point in a particular posting.

3. Community model, individual model and adaptive rewards

The adaptive reward mechanism is introduced as an improvement of the mechanism of hierarchical memberships [5]. The basic idea is to adapt the rewards of particular forms of participation for individual users and displaying personalized messages to them depending on their current status and reputations and the current need of the community, thereby influencing and directing the users' behaviors of contributing.

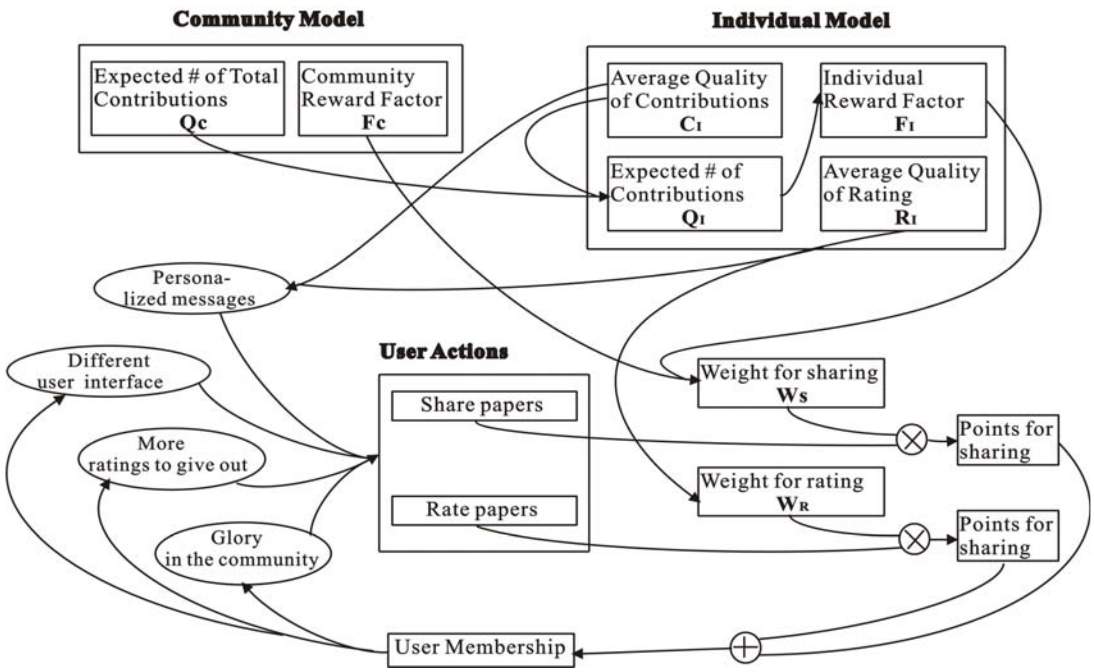

Fig. 2. An overview of adaptive reward motivation mechanism

Fig.2 presents an overview of the mechanism. The community model is used to describe the current phase of the whole community. It includes the expected sum of user contributions for current topic (Q_c) and the community reward factor (F_c). For each week, when a new discussion topic is introduced, Q_c is set by the community administrator (e.g. the instructor of the course) for the new topic, depending on his/her knowledge of certain features of the topic (e.g. how interesting it is expected to be for the users, how much materials are available) and the users' potential ability (e.g. how much time and energy they can devote, depending on their coursework, exams, etc.). F_c reflects the extent to which new contributions are useful for the whole community. Generally, new contributions are useful as soon as possible after a topic has been announced or opened. Therefore, F_c has its maximum value when a new topic discussion begins and decreases gradually with the time according to a function depicted in Fig.3.

Each user has an individual model that keeps statistical evaluations of his/her previous contributions and ratings and contains the data describing his/her current status. The average quality of a user's contributions (C_I) is defined in a straightforward way as the average summative rating of all the resources he/she has shared so far.

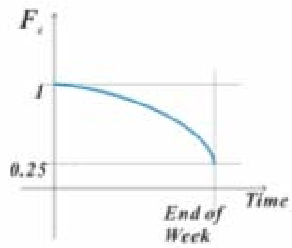

Fig. 3. The change of the community reward factor (F_C)

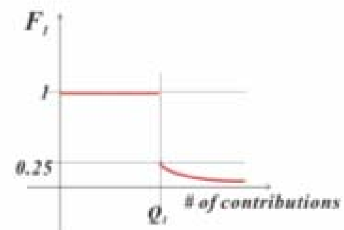

Fig. 4. The change of the individual reward factor (F_I)

However, the quality of user ratings can not be defined so easily, since they are by nature subjective. The average of all the ratings awarded to a given resource reflects the community criteria for quality and is more unbiased. Therefore, we chose to measure the quality of each rating for a given resource by the difference between the rating and the average rating that this resource has received so far. The quality equals to the reciprocal of the difference. Accordingly, the average quality of a user's ratings (R_I) equals to the average of the quality values of all the ratings he/she has made. Since this method can be skewed if users intentionally rate close to the average rating of the resource, the average rating should not be shown to the users directly.

The expected number of contributions of each user (Q_I) is a fraction of the total number of contributions that the community is expected to make for the topic, Q_c. The users with higher C_I will get a larger Q_I. If details are ignored, formula (1) can demonstrate how Q_c is distributed among users.

$$Q_I \approx Q_C \bullet \frac{C_I}{\sum C_I} \qquad \textbf{(1)}$$

The individual reward factor (F_I) defines the extent to which the user's contributions are being rewarded. F_I is a function that is a constant value as long as the number of the user's contributions is less than or equal to his/her Q_I. When the number exceeds the expectation, F_I drops to one fourth of the constant value instantaneously and keeps decreasing with the increment of the users' contributions (Fig.4)

Varying weights $W_i(t)$ for particular forms of participation are applied to compute the value of users' contributions and determine their membership levels, which are associated with different rewards and privileges. If we represent with $t=(1,2,3 \ldots T_i)$ the sequence of the contributions in each kind, the overall evaluation of a user's contributions (V) is calculated through formula (2).

$$V = \sum_{i=1}^{n}\left[\sum_{t=1}^{T_i} W_i(t)\right] \qquad \textbf{(2)}$$

The weights are adaptable to the states of the users' individual model and the community model at the current time. They, as well as the personalized messages, are conveyed to the users to influence their contribution patterns individually. The adaptive weight for sharing resources (W_S) is calculated through formula (3). Here W_{s0} is a constant, which is the initial value of the weight.

$$W_S = W_{S0} \bullet F_C \bullet F_I \qquad \textbf{(3)}$$

W_S is equal to W_{s0} when a new disucssion begins and the number of the user's contributions have not reached his/her expected value Q_I. After that, it decreases gently with time. Whenever the number of the user's contributions goes beyond his/her Q_I, W_s sharply decreases to one fourth of its original value and continues to decrease with the accumulation of the user's contributions and time.

It can be seen that W_S inherits the features of both reward factors, F_c and F_I. In this way, a user who shares many papers but does not pay enough regard to their quality gets a low C_I and a small Q_I and therefore, little reward for his/her subsequent contributions. Thus the personalized message to the user would be to contribute less in next period but improve the quality. This situation continues until the user finally improves his/her reputation for sharing. On the other hand, if a user tends to share a small number of good resources, he/she obtains a high C_I and a large Q_I. Potentially he/she will be able to earn more rewards by sharing more resources, and this continues until the quality of the user's contribution drops. For both kinds of users, early contributions always earn more points. Hence, W_S is able to restrict the quantity of user contributions, inhibit the contributions of poor quality, elicit good ones and stimulate users to share early in the discussion period.

The adaptive weight for giving ratings is proportional to the average quality of the users' previous ratings (R_I). The users who have gained a good reputation in making ratings get higher weight for their subsequent ratings, which stimulates them to rate more papers. However, those with poor R_I will not get much reward for rating articles. They have to rate less and improve the quality of their ratings to win their reputation back and this would be the suggestion of the personalized message.

4. Case study

To evaluate the effectiveness of the adaptive reward mechanism, a case study was launched in the course on "Ethics and Information Technology" offered by the Department of Computer Science at University of Saskatchewan in the second term 2004/2005 (Jan.-Apr. 2005). The study was carried out for eight weeks and the topic was updated weekly according to the curriculum. Thirty-two 4th-year students were the participants, who were encouraged to share web-articles related to the discussion topic using Comtella. The students were evenly divided into two groups: one group using the system with all the features of the proposed mechanism, including the functions of rating articles, earning and investing *c-points*, adaptive weights, personalized messages, etc. (test group / Comtella 1) and the other using the system with only the rating function (control group / Comtella 2). Since there might be some cultural and gender-based differences in the users' initial predisposition for participation, the assignment of users to groups was based on having equal proportion of Canadian to foreign and male to female students in each group. To avoid the effects that the contribution patterns of one group could have impact on the other group, the two groups inhabited two completely separated online communities, but shared the same classes, followed the same schedule, curriculum and coursework.

After the evaluation, post-experiment questionnaires were distributed to the participants to collect feedback about their experiences. The data from the questionnaires and the two systems were analyzed and contrasted to answer the following questions.

- Did the users in the test group (Comtella 1) give more ratings?

The data over the eight weeks suggested that the answer to this question was clearly positive since the number of ratings given in Comtella 1 was consistently (over each week) higher than that in Comtella 2. Throughout the eight weeks, the total number of ratings in Comtella 1 was 1065 and in Comtella 2 was 594. This clearly shows that the motivational mechanism with *c-points* and the associated rewards showed sustained effectiveness in stimulating users to rate articles.

- If more ratings was given in test group than in control group, did the summative ratings in test group reflect the quality of the contributions better?

Although we did not look into each article to evaluate its quality, we asked users about their attitude to the summative rating for their contributions. 56% of the users (9 users) in Comtella 1 felt that the final summative ratings could fairly reflect the quality of their contributions, while in Comtella 2, only 25% (4 users) thought so. This result shows that the increment of the quantity of user ratings can improve the accuracy of quality evaluation based on collaborative rating.

- Did the users in the test group tend to share resources earlier in the week?

According to the data collected in the eight weeks, the answer to this question is also positive. The users in Comtella 1 shared higher percentage of their contributions (71.3%) in the first three days of the week than the users in Comtella 2 did (60.6%) and the difference between the two groups was significant in each week (ranging between 7% and 14%).

- Did the users in the test group (Comtella 1) share the number of resources that was expected from them?

In the questionnaires, half of the users (8 users out 16) indicated they tended to share the number of resources that was expected from them. We calculated for each user the average difference between the actual shared number and the expected number over eight weeks and found that for half of the users the average difference was less than 2, which means these users contributed according to the expected number. Although the two groups of 8 users did not totally overlap, the results show that about half of the users were persuaded to share resources in or close to the number that was expected from them.

- Is there a significant difference with respect to the total number of contributions between the test and the control group?

The difference in the total number of contributions in the two groups is not significant (613 in Comtella 1 versus 587 in Comtella 2). The standard deviations of individual user contributions in the two systems are large, although in Comtella 1 it is slightly smaller than in Comtella 2 (30.18 versus 32.1). In Comtella 2 the top user is responsible for 21% of all the contributions, while the top user in Comtella 1 is responsible for 18% of the contributions. In both systems there was one user who didn't contribute at all.

- What is the user's perception with respect to cognitive overload and quality of contributions in each group?

Nine users in Comtella 1 and six users in Comtella 2 indicated in the questionnaire that they had to spend a lot of time time filtering out uninteresting posts, which means the effect of information overload emerged in both systems. As for the quality of the articles in both systems, we asked the users to give the rough percentages of the articles of high, medium and low quality in their own system. The data in Table 1 are the averages of users' estimations, which shows that their attitude towards the quality of the articles in their communities is basically neutral. It is hard to compare the degrees of informaiton overload and the quality of contributions in the two groups based on these data because the users in each group had experiences only in one system and there might have been ordering effects, in terms of different cognitive limits and criteria of quality evaluation among the students in the two groups. We plan to invite three experts to evaluate the articles in both systems to clarify their differences in terms of informaiton overload and the quality of contributions.

Table 1. Percentages of the articles of high, medium and low quality

Quality	High	Mediun	Low
Comtella 1	24.1%	46.3%	29.6%
Comtella 2	28.5%	42.3%	29.2%

5. Discussion and Conclusions

Designing incentives into the software to ensure that online communities are sustainable has been recognized as one of the most challenging and important problems in the area of social computing. We propose a dynamic, adaptive mechanism for rewarding contributions in an educational online community which takes into account the current needs of the community (e.g. more new contributions, versus more ratings, depending on the time since the topic is introduced and the current number of contributions) and the user's personal style of contributing (e.g. fewer but higher-quality contributions versus many mediocre ones). The hypothesis is that such a mechanism will stimulate users to contribute when and what is most useful for the community at the moment, thus achieving a level of activity that makes the community sustainable.

A study to test the effectiveness of the proposed mechanism was launched in a fourth-year undergraduate class with 32 students. Currently, the data collected from the participants are being processing and analyzing. We have obtained some results, which show that the mechanism is able to encourage users to rate resources, motivate them to contribute early in the discussion and persuade at least half of them to contribute resouces in a specified number, thereby controling the amount of information in the community. The preception of the users about the quality of shared resources did not seem to improve and the presence of more ratings didn't seem to make it easier to find good resources. More research is needed to find the reason for this. The first step is to see if the quality was really lower, or the users' expectations have become higher due to the availability of more ratings. We are confident that the adaptive reward mechanism can improve the quality of contributions because it encourages the users who have a good reputation for sharing high-quality resources to share more and inhibit the contributions from the users who does not have a good repuation..

References

[1] J. Vassileva, J. Greer, G. McCalla, R. Deters, D. Zapata, C. Mudgal, S. Grant: A Multi-Agent Approach to the Design of Peer-Help Environments, in Proceedings of AIED'99, Le Mans, France, July, 1999, 38-45.

[2] J. Greer, G. McCalla, J. Vassileva, R. Deters, S. Bull and L. Kettel: Lessons Learned in Deploying a Multi-Agent Learning Support System: The I-Help Experience, in Proceedings of AIED'2001, San Antonio, 2001, 410-421.

[3] J. Vassileva, R. Cheng, L. Sun and W. Han: Stimulating User Participation in a File-Sharing P2P System Supporting University Classes, P2P Journal, July 2004.

[4] H. Bretzke and J. Vassileva: Motivating Cooperation in Peer to Peer Networks, User Modeling UM03, Johnstown, PA, 2003, Springer Verlag LNCS 2702, 218-227.

[5] R. Cheng and J. Vassileva: User Motivation and Persuasion Strategy for Peer-to-peer Communities, in Proceedings of HICSS'38 (Mini-track on Online Communities in the Digital Economy), Hawaii, 2005.

[6] P. S. Dodds, R. Muhamad and D. J. Watts: An experimental study of search in global social networks, Science 8 August 2003, 301: 827-829.

[7] S. R. Hiltz and M. Turoff: The network nation: Human communication via computer, Addison-Wesley Publishing Company, Inc., London, 1978.

[8] D. Shenk: Data smog: Surviving the information glut. HarperCollins, New York, 1997.

[9] Q. Jones and S. Rafaeli: User Population and User Contributions to Virtual Publics: A Systems Model, in Proceedings of the international ACM SIGGROUP conference on supporting group work, Phoenix, Arizona, 1999, 239-248.

[10] S. Johnson: Emergence: The Connected Lives of Ants, Brains, Cities, and Software, Publisher: Scribner, 2001, 152-162.

[11] C. Lampe and P. Resnick: Slash(dot) and Burn: Distributed Moderation in a Large Online Conversation Space, in Proceedings of CHI'2004, Vienna, Austria, Apr. 24–29, 2004.

[12] R. Merton and H. A. Zuckerman: The Matthew Effect in Science: the Reward and Communication Systems of Science are Considered, Science 199, 3810, 1968, 55-63.

[13] G. Beenen, K. Ling, X. Wang, K. Chang, D. Frankowski, P. Resnick and R. E. Kraut: Using Social Psychology to Motivate Contributions to Online Communities, in Proceedings of CSCW'04, Chicago, Illinois, Nov. 6–10, 2004.