Weighted collaborative rating system for the interactive Web

Gjorgji Kjosev

***Abstract*—This paper presents a novel weighted collaborative rating method suitable for on-line systems that provide user-generated content. The user may vote positively or negatively on the content generated by other users. The system weights the vote cast by the user according to the user's total contribution using the algorithms described in this paper. Using a data-set from a real system where the votes were unweighted, the effects of the new system on the content rating are measured and presented in this paper. The difference between the ratings produced by the unweighted and the weighted system and the advantages of the weighted system are discussed.**

***Index Terms*—collaborative rating, computer-mediated communication, rating system, moderation system**

I. Introduction

TODAY, the interactive Web is flooded with user-generated content. User-driven news channels such as Slashdot and Digg often receive thousands of content submissions per day[4]. The received content is of varying quality and accuracy, and determining those parameters is vitally important to the overall quality of the website. Owners of on-line systems that provide user-generated content often struggle to evaluate the quality of the contributions because of the volume of submissions. The need for flexible and automated rating systems has been growing and many automated and semi-automated systems have been employed to make this process easier. Solutions range from earlier moderator-based systems on newsgroups and forums to systems based on voting to complex systems including meta-moderation (rating the rating) or reputation[3]. This paper presents a flexible solution by combining ideas from multiple systems into one unified system.

II. Existing rating systems

Various methods exists to implement evaluation of user generated content.

Traditional systems such as bulletin boards or forums often use a manual, moderator based system. The moderators are people usually selected by the bulletin board administrator. At a technical level, the moderators are able to delete or edit all content on the forum they have been assigned to. Their role may include guiding or stimulating the discussion and removing or editing spam, off-topic and inappropriate content.[1].

The disadvantage of such systems are multiple, including the constant struggle of administrators to find individuals that are suitable and willing to moderate. Additionally moderators may be unable to deal with the amount of content that some systems are able to deliver, thereby increasing the possibility that they might make an error in judgment.

Furthermore users may be dissatisfied with the decision of a moderator, but there is little that they can do - mostly complain to another moderator or to the board administrator. Often users may complain publicly on the forum, creating another off-topic post that the same moderator may safely delete while complying with the rules, further increasing dissatisfaction.

Another type of systems use collaborative moderation. Such a system is the one used on Slashdot. As stories are published by the editors, randomly selected positive contributors are given moderation points. They may use these points to give a score to the story comments. They are not allowed to send comments to the stories where they moderate comments: if they choose to send a comment to such a story all the moderation points they've given will be invalidated. Additionally, certain users are given meta-moderation points. They may check the moderation scores given by the previously mentioned random moderators and evaluate the messages they moderated as over-rated, good, or under-rated. Finally, users are able to use various filters on the content to get only desired messages. [3]

Slashdot users are in general satisfied with this moderation system. However this system is very complex. There are 5 possible types of moderation: "Interesting", "Insightful", "Informative", "Troll" and "Flame-bait", of which the last 2 might be completely confusing to new users. While very expressive, this moderation system might be considered too complicated by some users.

Digg's system is much more democratic. In this system all users are equal and are able to rate all content with a positive or negative rating - to digg or bury content. Certain thresholds are implemented in Digg's system - if the content rating drops below a threshold, its hidden from most users unless explicitly requested by the user to be shown.

While this systems abandons the traditional moderator element, it doesn't recognize valued contributors very well. As such, the system may be easily swayed by both users and automated software in any desired direction and may not be appropriate for environments where an expert's opinion is more valued that the opinion of most users.

Similar to Digg's system is the one used on the programmer board on StackOverflow. [5] The system is expanded to account for and value the expert's opinion. Users may acquire reputation points which determine their ability to participate in various moderation elements. As users gather more reputation points, their privileges increase in predetermined steps.

The system presented in this paper attempts to implement most of the elements of the previously described systems. At its heart lies a distributed, rating-based system similar to the ones on Digg and StackOverflow where the users may vote positively or negatively on a submission. The main idea is inspired by Google's PageRank, which recognizes the need to evaluate the votes (links) coming from one page differently according to the PageRank of the page itself.[2].

III. Detailed Description

The basic goal of the system is to weight the votes of the users according to their contributions to the system. As users gradually accumulate positive or negative rating for their submited content, their influence in the voting process should gradually change. Users have two possible ways to vote for a content item: a positive or a negative vote.

The system assumes that users who contribute high-quality content have the ability (and will) to vote on other user's content i.e. it assumes that the ability to create content is proportional to the ability to grade the content's quality. This may not be the case for all systems, however, there are many cases where this is appropriate. For example, news sites often receive submissions that were found interesting by the users - if what a user evaluated as an interesting news story received good grades, the user's evaluation about the quality of other news stories is also accurate. In conclusion, that user should have bigger influence in the grading process.

Another assumption of this system that the content can be rated objectively. Objectivity is provided to the system by giving the system a small set of exemplary objective users as moderators. The system should gradually find other objective users automatically as votes from users accumulate over time. The exemplary users can easily become regular users or become inactive after the initial period without impeding the way the system works. Optionally those users may remain marked as exemplary and continue to serve a function similar to the moderators on classic bulletin boards.

In order to measure the voter's contributions to the system we use the votes received by the content that the voter has created. The total contribution of the voter is the sum of votes received for his content. We normalize the measured contribution of the voter, which is multiplied with the chosen vote and added to the content item's total score.

Additionally, the weighted vote is also added to the score of the creator of the content item. The creator's voting influence is now changed. This voting influence will apply for all future votes made by the content item's creator.

The following is a mathematical description of the basic system.

We denote the initial vote of the voter given for the content item as V_{V_C}, the influence of the voter as M_V, the rating of the creator's item as R_{I_C} and the influence of the content item's creator as M_C .

When the voter casts the vote, the rating of the content item is changed as follows:

$$R_{I_C} \leftarrow R_{I_C} + M_V V_{V_C}$$

Where the voter's voting influence M_V is calculated using the formula

$$M_V = M_{\min} + (1 - M_{\min}) \frac{R_V}{R_{\max}}$$

In this formula R_V is the voter's rating and $R_{\max}$ is the normalization score. The normalization score must be at least equal to the rating of the best user of the system.

The minimum influence $M_{\min}$ determines the system's optimism about new users. The initial ratings of all the users in the system are zero until they receive votes on the content they have created. Therefore a minimum score is needed to start up ("bootstrap") the system. Additionally this constant determines the "amount of democracy" in the system. Setting it to 1 will make the system function as standard democracy with no vote weighting. Setting it to 0.75 will put the weighting in the range 0.5 for the least influential user to 1.0 for the most influential user, thereby making the most influential user have twice the influence compared to the least influential.

The content creator's rating R_C is also updated accordingly

$$R_C \leftarrow R_C + M_V V_{V_C}$$

which also changes his future voting influence.

In the initial version of the algorithm the normalization score was the maximum score of the users in the system

$$R_{\max} = \max(R_U)$$

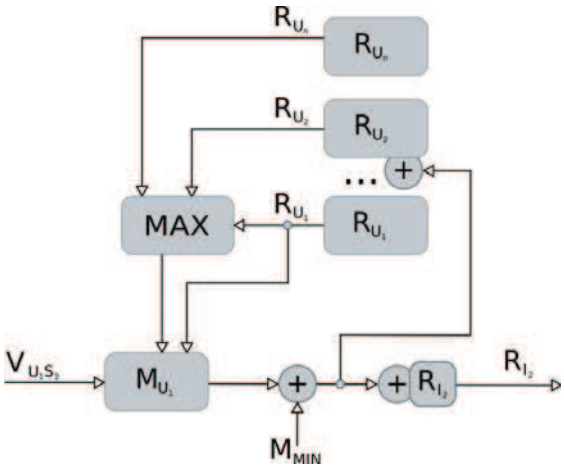

Figure 1. Initial system. The user U_1casts a vote on the content item S_2 created by the user U_2. User influence is controlled by user rating and the maximum of user ratings. Minimum influence is added and the final vote is applied to the content item and the content item's creator.

However, this caused several side effects on the system. The system had infinite memory for the user's ratings. Older users that were active in the past but may or may not be active at the present kept their high accumulated rating which is hard to achieve by new users. Additionally at bootstrap $R_{\max}$ was unstable - as soon as the first rating was given, one user had maximum influence while many other had minimum influence. To avoid these shortcomings $R_{\max}$ was made a constant. When

a user reaches a rating above $R_{\max}$, all the user's ratings are updated:

$$R_U \leftarrow K_N R_U$$

where K_N is the attenuation constant, $0 < K_N < 1$

This way all ratings are attenuated when a sufficient amount of votes accumulate for the most valuable user. Over time the absolute value of the ratings of inactive users will continue to decrease, giving room for new, active users to take the lead. Additionally the bootstrap period is now stable, with all ratings having close to minimum influence and growing slowly.

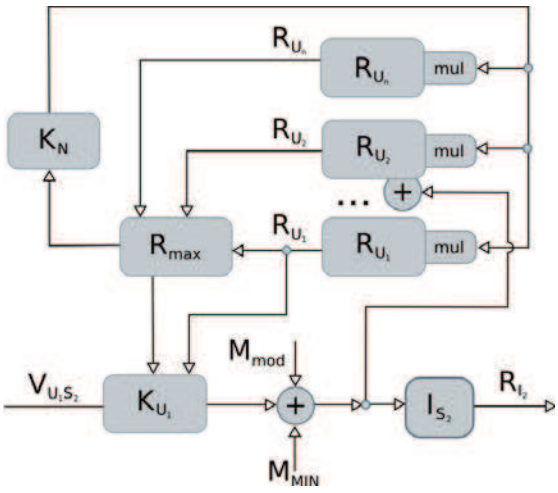

Figure 2. The system with the attenuation constant and the voting signals. The user U_1casts a vote on the content item S_2 created by the user U_2

It was easy to incorporate traditional "moderators" in the system. The voting influence formula was altered for the moderators to become:

$$M_V = M_{\min} + (1 - M_{\min}) \frac{R_V}{R_{\max}} + M_{mod}$$

where M_{mod} is additional moderator influence. This constant allows us to manually change the influence of some users and set exemplary users.

However the moderator influence is much higher than in traditional systems. By giving their vote, they're also able to shape the system by providing a reference point to the system about what to expect from its users. In essence with this modification the system should reward the properties of the exemplary users with higher influence. This significantly reduces the moderators efforts compared to traditional systems. The new system finds exemplary users automatically by using the samples from the moderators. These additional automatically detected moderators help the original moderators with the rating process and they amplify and enrich the influence of the original moderators.

To prevent abuse by the users with negative influence another step in the system was introduced:

$$M_V = \max\left(M_{\min} + (1 - M_{\min}) \frac{R_V}{R_{\max}}, 0\right)$$

This prevents users with negative ratings from inverting their votes to have the desired influence in the system. All the users with negative rating have influence 0 and cannot participate in the voting process.

The final calculated score of a content item S_I is always normalized with the positive sum of influences of all the voters. The content item ratings are between 1 (all users voted positively for the item) and -1 (all users voted negatively for the item), applied only when a custom-set minimum amount of votes are received.

$$S_I = \frac{\sum V_{V_C} M_V}{\sum M_V}$$

IV. Evaluation

The system was evaluated multiple times during development. Shortcomings were identified and new elements were added to the system to compensate accordingly. The evaluations during and after development will be described in this section.

The dataset used in the evaluation was extracted from a real on-line forum. The original forum had a democratic voting system where users could give positive or negative rating to messages sent by other users. The dataset contains approximately 80 000 messages and 450 000 votes. The measured properties were stability of the average user influence and the differences of the rating to the original democratic system and to a moderators-based system.

The initial system had a quick transitioning period during which the maximum user influence was established. However the maximum user influence varied quickly and was unstable, as were the user influences of individual users. This is not the desired property of this system.

The minimum influence was set to 0.05. Influence of 0.8, which is close to the maximum influence 1.0 was reached after only 120 votes. In the next 450 votes this influence was never reached so it is safe to assume that the maximum influence dropped below 0.5 after this transitioning period.

Quick ascend to maximum influence is undesired in this system and is considered as instability. Additionally it contradicts a system goal: users should gradually earn their influence. Finally, its harder to control the system because the initial users will inevitably set the reference point for the other users. We want the ascent of the system to maximum influence to be more gradual and the moderators to control the reference point.

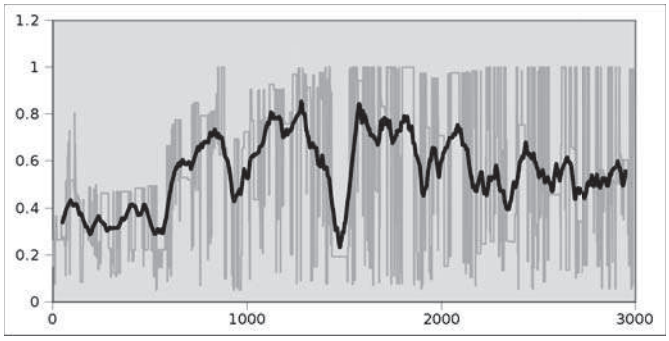

Figure 3. Initial system: Influence of the first 3000 ratings. The bold line represents a moving average of length 100

Figure 3 displays the first 3000 ratings - the complete transition of the system to stable state. After the first 2000 ratings, the system's average influence slowly converges to 0.5. Before those ratings the system's average influence fluctuates.

The changes after the constant R_{MAX} solution was introduced are shown in Figure 4

The system changes significantly less with this solution. The initial 500 ratings only slowly increase the maximum influence from 0.5 to around 0.12 as votes accumulate in the system.

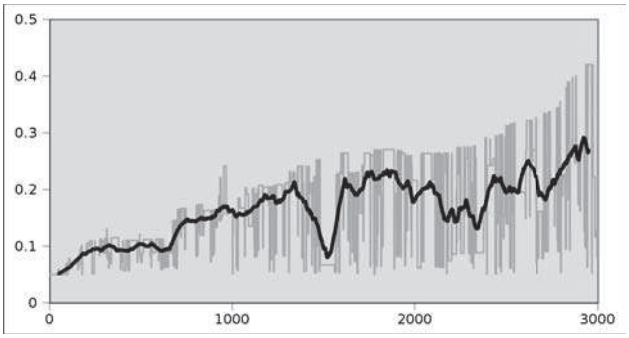

Figure 4. Constant R_{MAX} solution: Influence of the first 3000 ratings. The bold line represents a moving average of length 100

With the constant R_{MAX} solution as shown in Figure 4, both the maximum and the average influence slowly ascend with the number of ratings. The system achieved a stable state of average influence 0.3 and maximum influence of 1.0 after 8000 ratings.

The moderators ratings had a much larger influence for the first 3000 votes compared to the initial system.

After the final version of the system was designed, the difference between this system and the democratic systems (like Digg or the one used in the comments on YouTube was measured. The comparisons were made with resolution of 5% difference between the ratings.

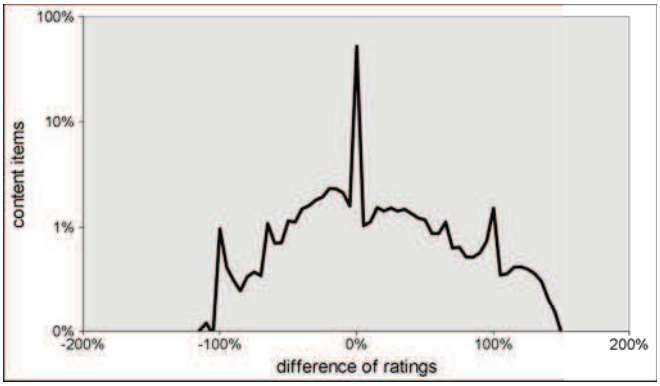

Figure 5. Difference between a democratic system and the final system. The x-axis represents the amount of difference in the rating between democracy and the new system. The y-axis represents the logarithm of the percentage of content items. Given a difference in ratings rounded to 5%, the graph shows the percentage of content items which have that difference.

Figure 5 shows the new system without moderators. With the new system as a reference point a few observations can be made:

- Democracy tends to overvalue the rating of content items. Most of the differences are positive, which indicates that ratings in the democratic system are higher on average.
- There is a large spike at 0%, which means that the systems agree on the rating of most items.
- The average absolute difference of the ratings is 25%

The following experiment was performed in order to measure if the new system amplifies the influence of the moderators. Content items which received votes from at least one moderator were isolated from the dataset. The selected items were rated using 3 different systems:

- Pure democratic system, where the influence of all users is equal and constant.
- Pure moderator-based system, where the influence of all users is zero and the influence of all moderators is the same
- Passive version of the new system, where the moderator's vote has the same influence as a normal vote on the content's rating, but significantly alters the voting influence of the content's creator.

The passive version of the system was constructed to isolate and measure the effect of user influence and vote weighting. The assumption was that even though the moderators vote on the content like regular users, their effect on the user influence will push the system significantly towards their reference point. To test this assumption, two differences between the ratings were calculated.

- The difference between the democratic system and the moderators-based system
- The difference between the passive version of the new system and the moderators-based system

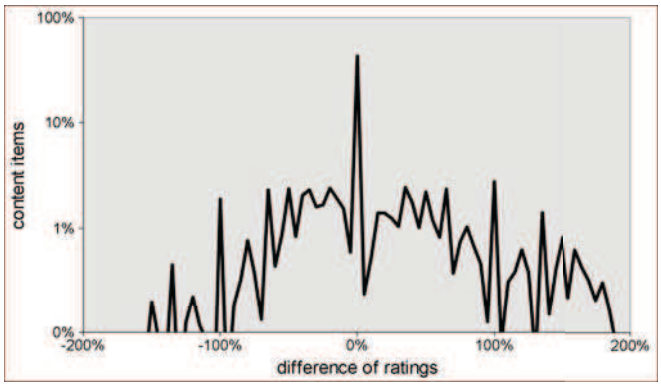

Figure 6. Difference between a democratic and a moderator-based system. The x-axis represents the amount of difference in the rating between democracy and the new system. The y-axis represents the logarithm of the percentage of content items. Given a difference in ratings rounded to 5%, the graph shows the percentage of content items which have that difference.

As shown in Figure 6, there is a significant difference in the content ratings between a democratic system and a moderators-based system. The average absolute difference of the ratings is 35%

According to the results shown in Figure 7, difference between the passive version of the new system and a moderator-based system is significantly smaller compared to the difference between a democratic and a moderator-based system.

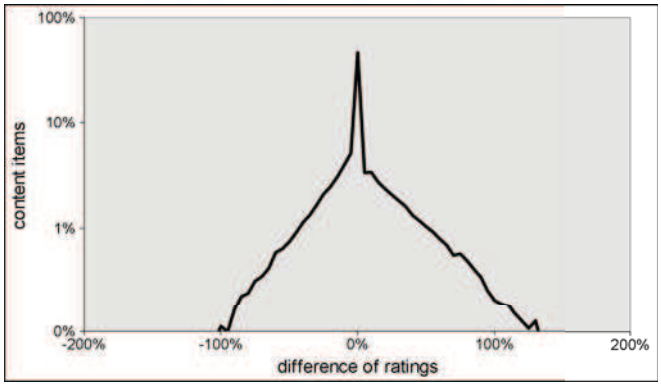

Figure 7. Difference between the passive version of the system and a moderator-based system. The x-axis represents the amount of difference in the rating between democracy and the new system. The y-axis represents the logarithm of the percentage of content items. Given a difference in ratings rounded to 5%, the graph shows the percentage of content items which have that difference.

Even though the same rules of influence were applied to both moderators and regular users regarding to votes on the content, the passive version of the new system managed to approximate the moderator system very well. The average absolute difference of the ratings is 17%

V. Conclusion

In this paper a novel rating system for the interactive web was presented. The new system is a blend of existing systems: democratic and moderator-based. Moderators are introduced in the system as influential users. The system also allows users to gain rating points and automatically gives a bigger influence to users with a higher rating. Additionally, the system helps traditional moderators do their job by finding users similar to them, increasing the moderators influence while reducing the amount of rating that the moderators need to do. Finally it automatically adjusts to transfer influence to the currently active users, preventing stagnation of the rating system when older users become inactive and stop contributing votes or content.

Work on this system is far from complete. Some of the constants used in the equations were hand tuned to achieve the desired system properties. Constants such as R_{MAX} should be automatically calculated in a production system. Depending on the average number of votes received by each content item the average growth rate of the user's rating may be slower or faster than optimal for the predetermined constant R_{MAX}. One of the ideas is to calculate R_{MAX} as maximum amount of rating points received by a user in a limited past period adjust it to the amount of voting activity in the system.

Another shortcoming that should be addressed is that the system functions with only one rating. There may be varying levels of expertise among the users on different topics. The system would function better if it has separate ratings for various content types or topics. This can be addressed by adding tagging to the system and remembering separate user rating points for each tag. Tags may be extracted automatically with text summarization algorithms.

Statistical studies of user opinions about the new system should be conducted to measure the benefits of the new system. Finally, the test implementation is still only a conceptual one. Future work should include the implementation of a production version of the system.

References

[1] T. Anderson and H. Kanuka. On-line forums: New platforms for professional development and group collaboration. *Journal of Computer-Mediated Communication*, 3(3), 1997.

[2] D. Austin. How Google finds your needle in the web's haystack. *American Mathematical Society Feature Column*, 2006.

[3] C. Lampe and P. Resnick. Slash (dot) and burn: distributed moderation in a large online conversation space. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 543–550. ACM New York, NY, USA, 2004.

[4] Kevin Rose. About digg - a couple of updates. *about.digg.com*, February 2007.

[5] J. Spolsky. Stackoverflow. com. *Joel on Software (Apr. 2008). Web: http://www. joelonsoftware. com/items/2008/04/16. html*.

Gjorgji Kjosev is a software developer at CreationPal and a graduate student at the university Ss. Cyril and Methodius, Faculty of electrical engineering and information technologies. He received a BS in computer engineering from the University Ss. Cyril and Methodius. His research interests include collaborative systems, web technologies and natural language processing. Contact him at gorgi.kosev@gmail.com