Proposal of a Ranking Method for Comments in Social Media using Ratings of Comment Posters

Keiichi Endo[1]

[1] Ehime University

Received: 16 December 2012 Accepted: 3 January 2013 Published: 15 January 2013

Abstract

Many social media adopt a ranking method in which comments are ranked in the order of the number of ratings attached to each comment. However, this method has the disadvantage of ratings being concentrated on comments posted at an early stage. Even if there are high-quality comments posted later, most of them are buried without being noticed. This paper proposes a ranking method that considers not only the ratings for each comment but also the previous ratings the comment poster has received. The effectiveness of the proposed method is evaluated through a simulation. We demonstrate that with the proposed method, high-quality comments are displayed in the higher positions regardless of the posting period.

Index terms— ranking method, comment, social media, rating, comment poster.

1 Introduction

n recent years, there has been explosive growth in the field of social media. Common examples include social networking services such as Facebook [1], video sites such as YouTube [2], social news websites such as Digg [3] and Yahoo! News [4], and shopping sites such as Amazon [5]. One major characteristic of these sites is that they allow users to post comments and provide ratings. Research on these trends and their effects is flourishing [6,7]. With social media, a wide variety of communities have been formed and the actions of their users are influenced by the information provided by other users. For example, in the case of Amazon, users: 1. Visit the site and browse the product lineup, 2. View the comments and ratings for the products, 3. Purchase a product after researching comments and ratings of the products, and 4. In turn provide comments and ratings for the product. These types of actions are seen in various social media [8]. It can be argued that the comments and ratings of other users have a greater influence on a user's decision than the product itself. In other words, on the Web, comments and ratings are extremely important elements. Therefore, social media sites provide content ranking on the basis of the comments and ratings attached to their content.

Author ? : Ehime University, Japan.

As the comments are written by ordinary users, some of them are suitable for reference by a large number of users, and some are not. In order to show users high-quality comments, social media attach ratings to not only the content but also the comments themselves, creating a comment ranking system based on ratings. With this ranking system, higher quality comments are displayed at a higher rank and viewed by a larger number of users.

Many social media adopt a ranking method in which comments are ranked in the order of the number of ratings attached to each comment, but this method has the disadvantage of ratings being concentrated on comments posted at an early stage (right after the content is created). This is evident in the comment ranking of Yahoo! News, where the posting times of many top comments for an article are close to the time the article was published. Even if there are high-quality comments posted later, most of them are buried without being noticed. Improving the reliability of comment ranking is vital to distinguish high-quality comments and to ensure that they are viewed by more people. Therefore, this study proposes a ranking method that considers not only the ratings for each comment but also the previous ratings the comment poster has received. The effectiveness of the proposed method is evaluated through a simulation.

This paper is organized as follows. In section II, we outline several researches related to comment ranking. In section III, we describe existing methods of comment ranking, which are later compared with the proposed methods in the simulation. Section IV explains the proposed methods and section V examines the simulation results. Finally, section VI concludes this paper.

2 II.

3 RELATED WORK

Chiao-Fang et al. [9] have attempted to resolve the problems of comment ranking using regression analysis. Using a model for the analysis with support vector regression (SVR), characteristics such as volume of information and characters are filtered from the comment text data and ranked according to the normalized discounted cumulative gain (NDCG). In addition, analysis is performed on the basis of not only Year each characteristic but also a combination of several characteristics. SVR enhances support vector machine (SVM) learning to deal with the issue of regression [10]. SVM is one of the learning models that contrive to output highly discriminatory features in relation to unlearned data. NDCG is an index that rates compatibility with related items through several steps. The results show that learning ranking models using SVR have higher compatibility than existing methods such as random ranking. In addition, a ranking method known as boosted ranking has been proposed. This method calculates the average and standard deviation of the number of ratings for comments whose order of posting is the same among the comments attached to the entire content and uses them to revise the ranking. For example, if a comment is the tenth one posted for a certain content item and a higher number of ratings have been collected than the average number of ratings for the tenth posted comments for all content items, this comment is judged to be of good quality and moved to a higher rank. Conversely, if it has a lower number of ratings than the average number, it is moved to a lower rank.

Onkar et al. [11] have developed a ranking method involving dynamic learning that considers comment rankings as a collection of objects and optimizes the edges that exist between the objects using Hodge analysis. The edge relationships between comment nodes (objects) are expressed using a matrix and the ranking is achieved by resolving the optimization problem defined from this. Compared with existing methods using objects, the calculation time is greatly improved and the method has a high level of compatibility.

Using NDCG, Xuanhui et al. [12] have evaluated the compatibility of rankings achieved on the basis of indices such as comment length, time passed since the post, and the ratio of positive ratings to the total ratings. Furthermore, the results of testing each index using Kendall's rank correlation coefficient showed that rankings created on the basis of the ratio of positive ratings achieved the highest level of compatibility.

Martin et al. [13] have proposed the similarityreduced explicit semantic analysis method. This method identifies comments that are most closely related to the article content, from the comments attached to an article. Adriano et al. [14] have proposed a comment selection method that employs automatic machine learning to pick out high-quality comments from a group of comments.

The above studies increased the reliability of rankings mainly by analyzing the content (text data) of the comments. In contrast, our study aims to improve the ranking by using the previous ratings of the comment poster. The proposed method can be applied to not only text comments but also comments made in the form of images, voice, or video.

4 III.

5 EXISTING METHODS

In this section, we describe two existing ranking methods that will later be compared with the proposed methods in the simulation.

6 a) Ranking method based on the ratings for comments

The ranking method based on the ratings for comments is used in various services such as the comment system of Yahoo! News and customer reviews of Amazon. The way the rankings are created differs according to the service, but the mechanism is basically that comments that have collected a large number of positive ratings are displayed at higher ranks. However, this method has an issue in that ratings are concentrated on comments posted at an early stage and high-quality comments posted at a later stage get buried without attracting ratings. This is because as there are more opportunities for comments posted at an early stage to be displayed at a higher rank, there are also more opportunities for them to be rated. As it is difficult for comments posted at a later stage to be displayed, the number of times they are viewed by users is fewer, and hence, there are fewer opportunities for them to be rated.

Many services allow users to attach either positive or negative ratings. However, this study only deals with positive ratings, and the more ratings the comment has, the higher in rank it will be displayed. The previously described differential in rating opportunities depending on the posting period is a problem unrelated to whether negative ratings are dealt with or not.

7 b) Boosted Ranking

The boosted ranking method [9] makes improvements in relation to the issues discussed in the previous subsection. This ranking method uses the average and standard deviation of the number of ratings for comments posted in the same order (among comments posted for the entire content) to revise the ranking. In concrete terms, the rating value for a comment is calculated according to the following formula:

With the boosted ranking method, high-quality comments from a later posting period can be pushed up higher in the ranking. However, as there are fewer opportunities to rank the comments from a later posting period even if they are high-quality comments, it remains where is the number of ratings for the comment, and and are the average and standard deviation of the number of ratings, respectively, for all comments whose order is the same as that of the posted comment. The comments are then displayed in the order of their rating values.

8 ?? ??? ??

IV.

9 PROPOSED METHODS

In this section, we describe the proposed ranking methods. Hereafter, a user is referred to as an agent. a) Ranking method based on the rated ratio of agents With the existing ranking methods explained in the previous section, as ratings are concentrated on the comments posted at an early stage, comments that are not of high quality may be displayed at a higher rank. This is because as there are many opportunities for comments posted at an early stage to be rated, even comments that are not of high quality can attract a large number of ratings. As there are fewer opportunities to view comments posted at a later stage even if they are high-quality comments, it is difficult for them to attract a large number of ratings. We propose a ranking method in which by reflecting the previous ratings of the agent, comments posted by "superior" agents with high ratings are displayed at a higher rank even if they are posted at a later stage. In this method, an agent is evaluated on the basis of the rated ratio of the agent, which is obtained by dividing the total number of ratings obtained on all previous comments posted by this agent by the total number of times those comments are viewed.

With this ranking method, as the ranking is created on the basis of not the ratings obtained by each comment but the ratings of the agent who posted the comment, it is possible to display the comments posted by superior agents regardless of the posting period. However, this method has a disadvantage that lowquality comments posted by an agent with high ratings continue to be displayed at a higher rank. b) Ranking method based on the rated ratio of agents and comments

In this subsection, we propose a ranking method that considers not only the rated ratio of the agent posting the comment but also the rated ratio of each comment. This ranking method does not order comments on the basis of a specific rating value but rather calculates the ranking position of a comment when it is posted or obtains a rating, and places the comment in that position.

The initial ranking position of a comment ( ) when it is posted is obtained using the following formula:

Here, represents the total number of comments attached to the content at the point before the comment was posted (namely, the ranking position of the lowest ranked comment) and is the rated ratio of the agent posting the comment. Furthermore, is a non-negative constant defined in advance.

In this paper, this is set to 0.2. The posted comment is placed in the ranking position obtained using the above formula and all comments that were at position or below are dropped by one position (Figure 1). In this way, as comments posted by superior agents are displayed at a comparatively higher rank immediately after being posted, there are sufficient opportunities to rate them even if they are posted at a later stage. The new ranking position for the comment that has obtained the rating is then calculated using the following formula and the comment is moved to that position (Figure 2).

10 Global Journal of Computer Science and Technology

Volume XIII Issue III Version I ?? cap = max[1, ??? bottom ? ??(1 ? ???c omment)?],

?? eval?? eval = max??? cap , ??? cap + ??? current ? ?? cap ? ? ???1 ? ???a gent ???, ?? cap

With this ranking method, the higher the rated ratio of the agent posting a comment is, the easier it is for the comment to be displayed at a higher rank. However, as a ranking position cap is set on the basis of the rated ratio of the comment itself, a low-quality comment posted by a superior agent is prevented from being continually displayed at a higher rank.© 2013 Global Journals Inc. (US) ? ?

where is a non-negative constant defined in advance and expresses the rated ratio of the comment itself.

11 ?? (? 1)

???c omment where is a non-negative constant defined in advance and represents the current position of the comment.

12 ?? (? 1)

?? current

V.

13 SIMULATION

To evaluate the effectiveness of the proposed ranking method, we performed a simulation using a program created in C++. In this section, we explain the simulation conditions and then present our observations based on the results.

14 a) Simulation Conditions

The simulation in this study first generates 30 content items (equivalent to articles in the case of a news site). In its initial state, no comments are attached to those content items. The simulation then generates 300 agents and randomly sets agent parameters in the range to those agents. The higher this agent parameter, the better the agent and the greater the probability of a high-quality comment being posted.

In this simulation, time units are referred to as "turns." A content browse interval is set at random between 1 and 10 turns for each agent. Each agent browses the contents, attaches ratings to the comments for the contents, and posts comments in the following procedure every time the content browse interval passes.

1. The agent randomly selects a content item to browse and views the comments attached to the content. At this time, the comment at a ranking position of has a probability of being viewed. The agent attaches a rating to the viewed comment with a probability of set for comment. 2. The agent posts a comment with a probability set randomly in advance within the range . For the comment, the comment parameter equivalent to the probability of the comment obtaining a rating is set randomly within the range . Here, is the agent parameter for the agent posting the comment. However, with a certain probability (referred to as an exceptional posting probability), the comment parameter is set at random within the range ,regardless of . The exceptional posting probability is set to a fixed value through a simulation and when it is set to a positive value, a superior agent may post low-quality comments. 3. The agent then moves to another randomly selected content item with a probability randomly set for each agent in advance within the range and repeats the procedure from step 1. When the agent decides ranking method based on the ratings for comments is referred to as Simple, the boosted ranking method is referred to as Boost, the ranking method based on the rated ratio of agents is referred to as Proposed-A, and the ranking method based on the rated ratio of agents and comments is referred to as Proposed-AC. The simulation is performed ten times under the same condition and the average of the results is plotted in a graph. ?? 0.99 ?? ?? comment [0, 1] ?? comment [?? agent ? 0.2, ?? agent + 0.2] ?? agent [0, 1] ?? agent [0, 1]

not to move, the agent terminates the procedure in the current turn.

The simulation is performed according to the procedure above until 600 turns have been completed.

15 b) Results and observations

In this section, we present the results of the simulation and make some observations. Hereafter, the In Figure 3, we compare each method in regard to the average value of the comment parameter set for the comments in each of the top 50 positions for all the ?? = 0.2 G ?? content items. In the case of Proposed-AC, the parameter values for and are 0.2 and 0.6, respectively. Proposed-AC demonstrates the highest values for the average of comment parameter in the top ranking positions, which implies that Proposed-AC is successful in displaying high-quality comments at the top.

Figure 4 shows a graph for Proposed-AC, comparing the average of comment parameter where is fixed at 0.2 and is varied. Even where is varied, no significant differences emerge in the top four rankings, but from the fifth position, the differences begin to increase. The average of comment parameter for the first position is the highest when . Figure **??** shows a graph for Proposed-AC, comparing the average of comment parameter where is fixed at 0.6 and is varied. From these results, we can see that the value of has a major influence on the quality of comments displayed in the top ranked positions. The average of comment parameter for ranking positions 1 and 2 are the highest when , but from the third position, the best results are seen when

. From this, we can see that it is better to set when we emphasize the quality of the comments on the first and second positions, and when the aim is to generally display high-quality comments at higher rankings from the third position.

Figure **??** shows a graph for Proposed-A and Proposed-AC, comparing cases where the exceptional posting probability (EX) is set at 0.2 and 0.0. In the case of Proposed-A, the average of comment parameter for the top rankings is much lower when EX is set at 0.2 than when it is set at 0.0. This is because with Proposed-A, low-quality comments posted by superior agents continue to be displayed in the top ranking positions. Since Proposed-AC considers the rated ratio of each comment in addition to the rated ratio of the agent, the quality of the top comments does not decrease even where EX is set at 0.2. Figure **??** demonstrates the comment distribution with the posting period for Proposed-AC, where EX is set at 0.2. The 600 turns in the simulation are divided into three, with the comments posted within the first 200 turns being referred to as "early," those posted during the next 200 turns as "middle," and the final 200 turns as "late." From this figure, we can see that the ranking order for Proposed-AC is significantly independent of the comment posting period and high-quality comments are displayed in the top ranking positions even when they are posted at a later period.

Figure ?? demonstrates the comment distribution with the posting period for Simple, where EX is set at 0.2.In the case of Simple, high-quality comments posted at a later period linger around the lower rankings. Most comments in the higher rankings are those posted at an early stage.

16 Global Journal of Computer Science and Technology

Volume XIII Issue III Version I

17 CONCLUSION

As the ranking order for existing ranking methods is significantly dependent on the comment G Figure **??** compares the four ranking methods in terms of the distribution of comments posted during the late period. EX is set at 0.2. With Simple and Boost, most comments posted during the late period stay in the lower ranked positions. In contrast, with Proposed-A and Proposed-AC, high-quality comments posted in the late period are displayed in the higher ranked positions. Since Proposed-A only uses the rated ratio for the agent, low-quality comments are also displayed in the higher ranks. With Proposed-AC, the rated ratio of the comments is also considered, and hence, only highquality comments are displayed in the higher rank positions.

posting period, there is an issue in that the higher ranked positions contain a mixture of high-and lowquality comments. To resolve this issue, this study has proposed a ranking method based on the previous ratings of the agent posting the comment. Furthermore, we have proposed a ranking method that also considers the rating of the comment itself as well as the agent rating. We have demonstrated that with the proposed method, high-quality comments are displayed in the higher positions regardless of the posting period. We have also demonstrated that by considering the ratings of both the agent and the comment, it is possible to prevent lower quality comments posted by superior agents from being continually displayed in higher ranking positions.

In the future, we plan to create a web application using the proposed method, and thereby examine its practicability.

18 References Références Referencias

Figure 1: I

[1]

[1] G Proposal of a Ranking Method for Comments in Social Media Using Ratings of Comment Posters © 2013 Global Journals Inc. (US)

2

Figure 2: Global 2 G

1

Figure 3: Figure 1 :

2

Figure 4: Figure 2 :

Figure 5: G

3

Figure 6: Figure 3 :4

4

Figure 7: Figure 4 :

56789

Figure 8: Figure 5 :Figure 6 :GFigure 7 :Figure 8 :Figure 9 :

Figure 9: 6

[Yahoo and News] , ! Yahoo , News . http://headlines.yahoo.co.jp/

[Amazon] http://www.amazon.com/ *Amazon*,

[Veloso et al. ()] 'Automatic Moderation of Comments in a Large Online Journalistic Environment'. Adriano Veloso , Meira WagnerJr , Tiago Macambira , Dorgibal Guedes , Helio Almeida . *Proceedings of International Conference on Weblogs and Social Media*, (International Conference on Weblogs and Social Media) 2007.

[Digg] http://digg.com/ *Digg*,

[Facebook] http://www.facebook.com/ *Facebook*,

[Cristiansanescu-Niculescu-Mizil et al. ()] 'How Opinions Are Received by Online Communities: A Case Study on Amazon.com Helpfulness Votes'. Gueorgi Cristiansanescu-Niculescu-Mizil , Jon Kossinets , Lillian Kleinberg , Lee . *Proceedings of the 18th International Conference on World Wide Web*, (the 18th International Conference on World Wide Web) 2009.

[Potthast et al. ()] 'Information Retrieval in the Comment sphere'. Martin Potthast , Benno Stein , Fabian Loose , Steffen Becker . *ACM Transactions on Intelligent Systems and Technology* 2012. 3 (4) .

[Wang et al. ()] 'Model News Relatedness through User Comments'. Xuanhui Wang , Jiang Bian , Yi Chang , Belle Tseng . *Proceedings of the 21st International Conference on World Wide Web*, (the 21st International Conference on World Wide Web) 2012.

[Dalal et al. ()] 'Multi-Objective Ranking of Comments on Web'. Onkar Dalal , H Srinivasan , Subhajit Sengamedu , Sanyal . *Proceedings of the 21st International Conference on World Wide Web*, (the 21st International Conference on World Wide Web) 2012.

[Hsu et al. ()] 'Ranking Comments on the Social Web'. Chiao-Fang Hsu , Elham Khabiri , James Caverlee . *Proceedings of International Conference on Computational Science and Engineering*, (International Conference on Computational Science and Engineering) 2009.

[Shmueli et al. ()] Erez Shmueli , Amit Kagian , Yehuda Koren , Ronny Lempel . *Proceedings of the 21st International Conference on World Wide Web*, (the 21st International Conference on World Wide Web) 2012. (Care to Comment? Recommendations for Commenting on News Stories)

[Lerman ()] 'Social Networks and Social Information Filtering on Digg'. Kristina Lerman . *Proceedings of International Conference on Weblogs and Social Media*, (International Conference on Weblogs and Social Media) 2007.

[Drucker et al. ()] 'Support Vector Regression Machines'. Harris Drucker , Chris J C Burges , Kinda Kaufman , Alex Smola , Vladimir Vapnik . *Neural Information Processing Systems*, 1997. 9 p. .

[YouTube] http://www.youtube.com/ *YouTube*,