A Ranking Algorithm for Mitigating the Influence of Contrived Ratings on Review Sites

Keiichi Endo*, Ryoki Horio, Dai Okano

Graduate School of Science and Engineering, Ehime University, Matsuyama 790-8577, Japan

Copyright ©2015 by authors, all rights reserved. Authors agree that this article remains permanently open access under the terms of the Creative Commons Attribution License 4.0 International License

Abstract In this paper, we propose a ranking algorithm for mitigating the influence of contrived ratings by spammers that try to manipulate rankings on review sites. We set a credibility level for each user, which is determined using the Pearson correlation coefficient of the rating given to an object by the user and estimated quality of the object. The estimated quality is the average value of the ratings weighted with the credibility level of raters. We propose a method that uses a provisional estimated quality calculated by eliminating the rating of a user when calculating that user's credibility level. Furthermore, we propose a method that considers the number of ratings given to an object when calculating the estimated quality of the object. Moreover, we demonstrate the superiority of the proposed methods by conducting a comparative experiment using an actual data set.

Keywords Ranking Algorithm, Review Sites, Spammers

1 Introduction

Recent years have seen the publishing of a diverse range of rankings created on the basis of ratings given to movies, bars and restaurants, products, etc. on review sites. User-participation-type review sites rank objects in terms of estimated quality on the basis of user-provided ratings [1]. Examples of such review sites include Netflix [2], MovieLens [3], and Amazon [4]. The rankings published on these review sites are used for information filtering to enable users to select high-quality objects.

The simplest type of ranking algorithm for finding estimated object quality uses the arithmetic mean of the ratings given to an object to judge its estimated quality. However, this method treats all user ratings equally, resulting in it being easily skewed by spammers that try to manipulate rankings using contrived ratings unrelated to the actual object quality. Therefore, it is inappropriate for user-participant-type review sites with users who contrive ratings [5].

The correlation-based algorithm proposed by Zhou et al. [6] aims to resolve this problem by setting a credibility level for each user. The user credibility level is determined using the Pearson correlation coefficient of the rating given to an object and its estimated quality, wherein the estimated quality is considered the average value of the ratings weighted with the credibility level. Accordingly, ratings from users with a

high credibility level are prioritized. This ranking algorithm is considered effective against various types of spammers [7–9].

Liao et al. [10] have proposed an iterative algorithm with reputation redistribution (IARR) as an improved correlation algorithm, focusing on the maximum value of the credibility level of users rating an object. IARR is said to be more robust against spammers than the correlation-based algorithm; however, it tends to provide excessively high credibility to users with the most monopolistic ratings, i.e., ratings for objects not rated by other users. This results in rankings essentially being determined by the ratings of only users with excessively high credibility levels.

To solve this problem, we propose a method that uses a provisional estimated quality calculated by eliminating the rating of a user when calculating that user's credibility level. Furthermore, since it appears that objects rated multiple times generally tend to be of a higher quality, we propose a method that considers the number of ratings when calculating the estimated object quality. Moreover, we demonstrate the superiority of the proposed methods by conducting a comparative experiment with the existing method using an actual data set.

We explain the existing and the proposed ranking algorithms in Sections 2 and 3, respectively. In Section 4, we outline the experiment conducted to evaluate the performance of the ranking algorithms, show the experimental results, and provide our observations on those results. We summarize our study and discuss future plans in Section 5.

2 Existing Methods

In this section, we explain three existing methods: the arithmetic mean method, the correlation-based algorithm, and IARR. We explain these methods for calculating estimated quality Q_α of object α ($\in O$) with the group of all objects O and the group of all users U . If we determine the estimated object quality, then we can rank objects in the order of estimated quality.

2.1 Arithmetic mean method

The arithmetic mean method is the simplest algorithm for ranking objects on the basis of quality. With this method, the estimated quality is the arithmetic mean of the ratings given to the object. We let U_α , $|U_\alpha|$, and $r_{i\alpha}$ denote the set of users rating object α , the number of users rating object

α , and the rating given by $i (\in U)$ to object α , respectively. Then, the estimated quality Q_α of object α can be calculated using formula (1).

$$Q_\alpha = \frac{1}{|U_\alpha|} \sum_{i \in U_\alpha} r_{i\alpha}. \quad (1)$$

This does not require complicated calculation, but it enables the estimated quality to be easily manipulated.

2.2 Correlation-based algorithm

The correlation-based algorithm is based on iterative refinement. It sets the credibility level for each user, repeatedly updates that level, and calculates the estimated quality until that quality converges [11,12]. The estimated quality in this algorithm is the average of user credibility level-weighted ratings. Furthermore, user credibility is obtained using a Pearson correlation coefficient of the rating that a user gives to an object and its estimated quality. The user credibility level increases or decreases, respectively, upon giving a rating close to or distant from an object's estimated quality.

The procedure for calculating the estimated object quality is explained below. First, the initial value of the credibility level R_i for user i is set by using formula (2).

$$R_i = \frac{|O_i|}{|O|}, \quad (2)$$

where $|O_i|$ is the number of objects rated by user i and $|O|$ is the total number of objects.

Furthermore, the estimated quality Q_α of object α is provided by formula (3).

$$Q_\alpha = \frac{\sum_{i \in U_\alpha} R_i r_{i\alpha}}{\sum_{i \in U_\alpha} R_i}. \quad (3)$$

Moreover, the Pearson coefficient C_i of the rating given to an object by user i and the estimated quality of that object is obtained using formula (4).

$$C_i = \frac{1}{|O_i|} \sum_{\alpha \in O_i} \left(\frac{r_{i\alpha} - \mu(r_i)}{\sigma(r_i)} \right) \left(\frac{Q_\alpha - \mu(Q_{O_i})}{\sigma(Q_{O_i})} \right). \quad (4)$$

Here, $\mu(r_i)$ and $\sigma(r_i)$ are the mean and standard deviation of the ratings given by user i , respectively. Furthermore, $\mu(Q_{O_i})$ and $\sigma(Q_{O_i})$ are the mean and standard deviation of the estimated quality of objects rated by user i , respectively. C_i satisfies $-1 \leq C_i \leq 1$, and if this value is large, we can say that there is a strong positive correlation between the rating given by user i and the estimated quality.

The credibility level R_i of user i based on C_i obtained as described above is updated using formula (5).

$$R_i = \begin{cases} C_i & (C_i \geq 0), \\ 0 & (C_i < 0). \end{cases} \quad (5)$$

The final estimated object quality is obtained by repeatedly evaluating formulas (3), (4), and (5) until the difference Δ obtained from formula (6) becomes sufficiently low.

$$\Delta = \frac{1}{|O|} \sum_{\alpha \in O} (Q_\alpha - Q'_\alpha)^2, \quad (6)$$

where Q'_α denotes the estimated quality of object α calculated in the previous iteration. Herein, it is considered to have converged when $\Delta \leq 10^{-6}$, as in [6].

2.3 IARR

IARR is a ranking algorithm used to improve the correlation-based algorithm. It was proposed with the aim of resolving the problems that occur when the number of ratings is low and higher priority is assigned to the ratings of users with a high credibility level. IARR performs iterative calculations similar to that of the correlation-based algorithm until the estimated quality converges.

The specific improvements are as follows. First, the maximum value of the credibility level of users rating an object is used in calculating the estimated quality of that object. This solves the problem of an object obtaining a low number of high ratings, resulting in the high estimated quality despite the credibility level of all users giving these ratings being low. With IARR, the estimated quality Q_α of object α is calculated using formula (7).

$$Q_\alpha = \left(\max_{i \in U_\alpha} R_i \right) \cdot \frac{\sum_{i \in U_\alpha} R_i r_{i\alpha}}{\sum_{i \in U_\alpha} R_i}. \quad (7)$$

Subsequently, the credibility level of the user with a low number of ratings is controlled to ensure that an object's estimated quality rated by a user just happening to be close to the evaluation given by that user does not unduly raise the user's credibility level despite that user having a low number of ratings. With IARR, the degree of correlation C_i for user i is calculated using formula (8). This multiplies the Pearson correlation coefficient of the rating the user i gives to the object and its estimated quality by a coefficient that considers the number of ratings.

$$C_i = \frac{\log |O_i|}{\max_{j \in U} \{\log |O_j|\}} \frac{1}{|O_i|} \sum_{\alpha \in O_i} \left(\frac{r_{i\alpha} - \mu(r_i)}{\sigma(r_i)} \right) \left(\frac{Q_\alpha - \mu(Q_{O_i})}{\sigma(Q_{O_i})} \right). \quad (8)$$

In the correlation-based algorithm, as shown in formula (5), where the Pearson correlation coefficient of the rating and the estimated quality is non-negative, this value is used intact as the credibility level; however, to increase the accuracy of the estimated quality ranking, IARR focuses on the ratings of users with a high degree of correlation by considering the credibility level as the correlation level to the power of θ (> 1). In the experiment described in Section 4, θ is set to 5. The relationship between the credibility level R_i of user i and degree of correlation C_i is expressed in formula (9).

$$R_i = \begin{cases} C_i^\theta & (C_i \geq 0), \\ 0 & (C_i < 0). \end{cases} \quad (9)$$

The initial values for each user's credibility level are the same as those in formula (2). Furthermore, changes in estimated quality are calculated using formula (6), and the evaluation of formulas (7), (8), and (9) is repeated until it reaches $\Delta \leq 10^{-6}$.

Figure 1 shows an example of ratings wherein each user can rate an object in stages from 1 to 5. Table 1 shows the transition of estimated object quality, difference Δ , and

Figure 1. Example of ratings given to objects by users.

Table 1. IARR calculation process.

	Q_δ	Q_ϵ	Q_ζ	Δ
1	0.833333	2.16667	1.16667	1.12500
2	0.217637	0.622351	0.342918	0.573759
3	0.221417	0.640539	0.368702	0.0001683
4	0.221892	0.645532	0.382257	0.0000348
5	0.221149	0.644986	0.388576	0.0000068
6	0.220277	0.643146	0.391376	0.0000020
7	0.219622	0.641547	0.392627	0.0000008

	R_j	R_k	R_l
0	1.00000	1.00000	1.00000
1	0.0832420	0.145166	0.0614749
2	0.0968704	0.147944	0.0490900
3	0.106403	0.148050	0.0409169
4	0.112409	0.147248	0.0359494
5	0.115951	0.146421	0.0330772
6	0.117964	0.145821	0.0314618
7	0.119086	0.145444	0.0305665

user credibility when performing iterative calculation using IARR. In this case, the calculation terminates after seven iterations. For a user k with the maximum credibility, $R_k/R_j \approx 1.23$, $R_k/R_l \approx 4.63$. Therefore, the rating given by user k , compared to user j , l , has 1.23 times and 4.63 times the weight, respectively.

3 Proposed Method

In this section, we first explain the issues with IARR. Figure 2 adds objects α , β , and γ , which have monopolistic ratings by user d , to the example shown in Figure 1. The transition of estimated object quality, difference Δ , and user credibility in IARR calculation process is shown in Table 2. If we focus on the R_d column, we observe that the credibility of user d is maintained at an extremely high level through the iterative calculation. We assume that this is because user d has more monopolistic ratings than other users. As user d 's ratings have approximately 30,520 times the weight of user

Figure 2. Example of ratings when there are objects that have a monopolistic rating.

Table 2. IARR calculation process when there are objects that have a monopolistic rating.

	Q_α	Q_β	Q_γ	Q_δ	Q_ϵ	Q_ζ	Δ
1	4.16667	4.16667	4.16667	1.48810	3.21429	1.16667	10.9984
2	3.99237	3.99237	3.99237	1.53405	2.59028	0.213179	0.231962
3	4.90645	4.90645	4.90645	1.95207	3.06506	0.137327	0.485425
4	4.98158	4.98158	4.98158	1.98628	3.09602	0.119569	0.0032291
5	4.98654	4.98654	4.98654	1.98865	3.09737	0.117552	0.0000143
6	4.98702	4.98702	4.98702	1.98888	3.09749	0.117341	0.0000001

	R_d	R_j	R_k	R_l
0	0.833333	0.500000	0.500000	0.500000
1	0.798474	0.0175009	0.0813810	0.135725
2	0.981291	0.0001975	0.0119943	0.125852
3	0.996315	0.0000415	0.0071162	0.112760
4	0.997309	0.0000335	0.0066731	0.111163
5	0.997405	0.0000328	0.0066280	0.110996
6	0.997415	0.0000327	0.0066234	0.110978

j 's, 151 times the weight of k 's and nine times the weight of user l 's, the ratings of users other than user d are largely ignored. Furthermore, we observe that the estimated quality Q_ζ of object ζ , which was not rated by user d , becomes extremely low compared to the estimated quality of other objects. This is because when the object's estimated quality is calculated using IARR, the maximum credibility level of the user rating that object is used.

As shown in this example, wherein there is a user with an extremely high credibility level compared to other users (referred to here as a dictator), the ranking can be easily manipulated by the dictator regardless of whether the dictator can accurately rate the object's quality. Furthermore, a problem exists in that even users with nonsensical ratings can increase their credibility levels by selecting and rating objects with low numbers of raters.

Therefore, we resolve this issue in the proposed method Proposed-E using a provisional estimated quality calculated from other users' ratings when calculating the user correlation level. Consequently, we can prevent the credibility level of users rating objects with a low number of raters multiple times from being easily raised.

In the proposed method Proposed-E, the correlation level

Table 3. Proposed-E calculation process when there are objects that have a monopolistic rating.

	Q_α	Q_β	Q_γ	Q_δ	Q_ϵ	Q_ζ	Δ
1	4.16667	4.16667	4.16667	1.48810	3.21429	1.16667	10.9984
2	0.0158110	0.0158110	0.0158110	0.173931	0.602464	0.324117	10.1579
3	0.0158110	0.0158110	0.0158110	0.207153	0.601993	0.356809	0.0003620
4	0.0158110	0.0158110	0.0158110	0.201495	0.614462	0.361919	0.0000356
5	0.0158110	0.0158110	0.0158110	0.199040	0.593842	0.354799	0.0000803
6	0.0158110	0.0158110	0.0158110	0.197790	0.594067	0.354675	0.0000003

	R_d	R_j	R_k	R_l
0	0.833333	0.500000	0.500000	0.500000
1	0.0031620	0.0224630	0.146687	0.0078330
2	0.0031620	0.0951880	0.138817	0.0362400
3	0.0031620	0.0737310	0.143162	0.0215530
4	0.0031620	0.0843160	0.137296	0.0247250
5	0.0031620	0.0812840	0.137541	0.0227650
6	0.0031620	0.0827520	0.136439	0.0231910

C_i of user i is calculated as follows. First, using formula (10), the provisional estimated quality of object α is calculated with the rating given by user i excluded.

$$E_{i\alpha} = \begin{cases} \frac{\sum_{j \in U_\alpha - \{i\}} R_j r_{j\alpha}}{\sum_{j \in U_\alpha - \{i\}} R_j} & (|U_\alpha| \geq 2), \\ 0 & (|U_\alpha| = 1). \end{cases} \quad (10)$$

Then, correlation level C_i is calculated using formula (11).

$$C_i = \frac{\log |O_i|}{\max_{j \in U} \{\log |O_j|\}} \frac{1}{|T_i|} \sum_{\alpha \in T_i} \left(\frac{r_{i\alpha} - \mu(r_i)}{\sigma(r_i)} \right) \left(\frac{E_{i\alpha} - \mu(E_{T_i})}{\sigma(E_{T_i})} \right). \quad (11)$$

Here, T_i is the group of objects rated by other users from among the objects rated by i , as expressed in formula (12).

$$T_i = \{\alpha \mid \alpha \in O_i \text{ and } |U_\alpha| \geq 2\}. \quad (12)$$

Furthermore, $\mu(E_{T_i})$ and $\sigma(E_{T_i})$ are the mean and standard deviation of provisional estimated quality for objects belonging to group T_i , calculated after excluding the rating from i , as expressed in formulas (13) and (14), respectively.

$$\mu(E_{T_i}) = \frac{1}{|T_i|} \sum_{\alpha \in T_i} E_{i\alpha}. \quad (13)$$

$$\sigma(E_{T_i}) = \sqrt{\frac{1}{|T_i|} \sum_{\alpha \in T_i} (E_{i\alpha} - \mu(E_{T_i}))^2}. \quad (14)$$

The initial credibility level, object estimated quality, and updated credibility level are calculated as shown in formulas (2), (7), and (9), respectively.

The calculation applying Proposed-E to the rating example in Figure 2 is shown in Table 3. We can confirm that this restricts the credibility level of user d , who had become a dictator in IARR, to a low level, and the estimated quality of object ζ not rated by user d is also improved.

Subsequently, we explain another proposed method (referred to as Proposed-R) for calculating the estimated quality by considering the number of raters of the object. When calculating the estimated quality of the object using IARR, the maximum value for the credibility level of users rating the object is used; however, if there is a user with a high credibility level, there is the problem that even where the number of

Table 4. Characteristics of the two data sets.

	$ U $	$ O $	$\max U_\alpha $	$\max O_i $	Sparsity (η)	b
dataset1	2500	15061	1356	5841	0.015605	361
dataset2	2500	14732	1295	2862	0.015078	356

Figure 3. Comparison of AUC in various ranking algorithms.

users giving a rating is low, the estimated quality will become high. Therefore, in Proposed-R, the estimated quality is calculated according to formula (15), considering the number of object raters rather than the maximum value of credibility level.

$$Q_\alpha = \frac{\log |U_\alpha|}{\max_{\beta \in O} \{\log |U_\beta|\}} \frac{\sum_{i \in U_\alpha} R_i r_{i\alpha}}{\sum_{i \in U_\alpha} R_i}. \quad (15)$$

The initial credibility level, correlation level, and updated credibility level are calculated by using formulas (2), (8), and (9), respectively.

Finally, we propose Proposed-ER, a ranking algorithm that combines Proposed-E and Proposed-R. This method calculates user correlation level and estimated quality similarly to Proposed-E and Proposed-R, respectively.

The initial credibility level and updated credibility level are calculated according to formulas (2) and (9), respectively.

In all the methods proposed above, convergence is judged using the same conditions as for the correlation-based algorithm.

4 Experiments and Observations

In this section, we provide an overview of the experiment using actual data sets and indices for performance evaluation of the ranking algorithms. We also provide observations on the experimental results.

Figure 4. Relationship between the number of object raters and estimated quality when applying IARR to dataset1.

Figure 5. Relationship between the number of object raters and estimated quality when applying Proposed-ER to dataset1.

4.1 Data sets and evaluation index

For the experiment, we used the data set of Netflix, which provides DVD rentals and video streaming services in the US. This data set includes rating data for 17,770 movies, with approximately 100 million ratings in total. Rather than using all of the data, we used two partial data sets of 2,500 randomly filtered users who contributed ratings 20 times or more. This follows MovieLens, a movie rating service that requires at least 20 movie ratings. Moreover, no user is included in both data sets, with all ratings ranging from 1 to 5. The characteristics of the two datasets, dataset1 and dataset2, are shown in Table 4. Here, Sparsity (η) refers to rating density, with the total number of ratings being obtained from $\eta|U||O|$. Furthermore, the number of benchmark objects is denoted by b . The benchmark objects are of high quality and are used for performance evaluation, as described later. Herein, we used movies from 1927 to 2010 that have been nominated for Academy Awards as the benchmark objects [13].

To evaluate ranking algorithm performance, we used AUC

Figure 6. Distribution of user credibility level when applying IARR to dataset1.

Figure 7. Distribution of user credibility level when applying Proposed-ER to dataset1.

(area under the curve) [14]. When one benchmark object and one normal object are randomly selected, AUC equals the probability that the benchmark object is ranked higher by the evaluated ranking algorithm. The higher the AUC is, the higher the ranking of high quality movies will be. The AUC maximum value is 1.0, in which case all of the benchmark objects are ranked higher than normal objects.

4.2 Results and observations

The AUC values resulting from applying the six ranking algorithms, i.e., arithmetic mean method, correlation-based algorithm, IARR, Proposed-E, Proposed-R, and Proposed-ER, to dataset1 and dataset2 are shown in Figure 3. First, with regard to the AUC values for dataset1, Proposed-ER had the highest value, followed by IARR. Proposed-E and Proposed-R were both lower than IARR in terms of AUC; however, when combined as Proposed-ER, the AUC value was higher than that of IARR. The correlation-based algorithm had virtually the same value as the arithmetic mean method and a far lower value than the method in which im-

Figure 8. Relationship between the number of user monopolistic ratings and credibility level when applying IARR to dataset1.

Figure 9. Relationship between the number of user monopolistic ratings and credibility level when applying Proposed-ER to dataset1.

Figure 10. Relationship between the number of object raters and estimated quality when applying IARR to dataset2.

Figure 11. Relationship between the number of object raters and estimated quality when applying Proposed-ER to dataset2.

provements had been made to the correlation-based algorithm.

Next, with regard to dataset2, AUC was highest for Proposed-ER, followed by Proposed-R. As in dataset1, Proposed-ER has a higher AUC than Proposed-E or Proposed-R. Furthermore, the AUC of IARR was the lowest, and much lower than that of the arithmetic mean method or correlation-based algorithm. Thus, we can say that combining Proposed-E and Proposed-R is effective, and that Proposed-ER, which is a combination of the two, is superior to the existing methods.

The experiment described above demonstrated that with IARR, the AUC greatly fluctuates on the basis of the data set, but a high AUC can be obtained regardless of the data set with Proposed-ER. We discuss the causes of this result hereafter.

The relationships between the number of object raters and the estimated quality in the case where IARR and Proposed-ER are applied to dataset1 are shown in Figures 4 and 5, respectively. The user credibility level distributions are shown in Figures 6 and 7, and the relationships between the number of user monopolistic ratings and credibility level are shown

in Figures 8 and 9.

First, if we concentrate on Figure 4, we observe that with IARR, the object estimated quality is distributed over six levels from 0 to 5. As shown in Figure 6, this results from there being a dictator – one user with an excessively high credibility level. Objects for which the estimated quality is roughly 1 to 5 were rated by the dictator, and the estimated quality is determined almost entirely by the dictator's rating. Objects for which the estimated quality is close to zero were not rated by the dictator. From Figure 8, we can see that the credibility level of the user with the highest number of monopolistic ratings is pronounced, and that the user in question is the dictator.

Next, if we concentrate on Figure 5, we can see that as Proposed-ER considers the number of raters of an object when calculating estimated quality, it succeeds in placing a high share of benchmark objects into the higher rankings. Furthermore, we observe that the problem wherein the estimated quality is virtually dependent on the ratings of one user does not arise. This is because, as shown in Figure 7, there is no user present with a pronounced credibility level.

Figure 12. Distribution of user credibility level when applying IARR to dataset2.

Figure 14. Relationship between the number of user monopolistic ratings and credibility level when applying IARR to dataset2.

Figure 13. Distribution of user credibility level when applying Proposed-ER to dataset2.

Figure 15. Relationship between the number of user monopolistic ratings and credibility level when applying Proposed-ER to dataset2.

Moreover, from Figure 9, we observe that with Proposed-ER, the user credibility level is not dependent on the number of monopolistic ratings.

The relationships between the number of object raters and the estimated quality in the case where IARR and Proposed-ER are applied to dataset2 are shown in Figures 10 and 11, respectively. The distributions of user credibility level are shown in Figures 12 and 13, and the relationships between the number of user monopolistic ratings and credibility level are shown in Figures 14 and 15. If we focus on Figure 10, we observe that as in the case of dataset1, the estimated quality is distributed in six rows. However, in contrast to dataset1, there are many benchmark objects with an estimated quality close to zero. Furthermore, there are many benchmark objects among the objects for which the estimated quality is approximately 0.6, those that the dictator rated the lowest. It appears that these are the causes of the AUC value being low when IARR is applied to dataset2.

In contrast, it appears in dataset1 that it is because the dictator is a user who can provide a comparatively accurate rating of the quality of the object that the AUC is higher even

if IARR is applied. That is to say, with IARR, the quality of the ranking can be affected by one user. When using the proposed method Proposed-ER, a high-quality ranking can always be provided, because the dictator will not appear.

5 Conclusion

In this study, we improved the existing IARR ranking algorithm in two ways. First, when calculating the correlation level of the user, we used a provisional estimated quality calculated from other users' ratings. Moreover, the number of object raters was considered when calculating estimated quality. Through the experiment using actual data sets, we demonstrated that with these improvements, we could resolve the problems found in IARR and provide high-quality rankings. We conclude that the proposed method, similarly to IARR and the correlation-based algorithm on which it is based, is not easily influenced by spammers' contrived ratings.

Furthermore, the improvements implemented in the pro-

posed method result in it being difficult to inappropriately manipulate rankings by rating objects with a low number of raters to raise the credibility level. Additionally, as objects with a high number of raters occupy the higher positions of rankings created with the proposed method, it is difficult for a small number of spammers to influence the ratings of objects in the higher positions of the ranking.

Future issues to be considered include verification of the usability of the proposed method when actually managing systems providing rankings based on user evaluation.

REFERENCES

-
- [1] C. Dellarocas: Immunizing Online Reputation Reporting Systems against Unfair Ratings and Discriminatory Behavior, Proceedings of the 2nd ACM Conference on Electronic Commerce (EC '00), pp. 150-157 (2000).
 - [2] Netflix: <https://www.netflix.com/>
 - [3] MovieLens: <https://movielens.org/>
 - [4] Amazon: <http://www.amazon.co.jp/>
 - [5] G. Wang, S. Xie, B. Liu, and P. S. Yu: Review Graph Based Online Store Review Spammer Detection, Proceedings of IEEE 11th International Conference on Data Mining (ICDM 2011), pp. 1242-1247 (2011).
 - [6] Y.-B. Zhou, T. Lei, and T. Zhou: A Robust Ranking Algorithm to Spamming, EPL (Europhysics Letters), Vol. 94, No. 4, Article 48002 (2011).
 - [7] E.-P. Lim, V.-A. Nguyen, N. Jindal, B. Liu, and H. W. Lauw: Detecting Product Review Spammers Using Rating Behaviors, Proceedings of the 19th ACM International Conference on Information and Knowledge Management (CIKM '10), pp. 939-948 (2010).
 - [8] K. Lee, J. Caverlee, and S. Webb: Uncovering Social Spammers: Social Honeypots + Machine Learning, Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 435-442 (2010).
 - [9] C. Grier, K. Thomas, V. Paxson, and M. Zhang: @Spam: the Underground on 140 Characters or Less, Proceedings of the 17th ACM Conference on Computer and Communications Security (CCS '10), pp. 27-37 (2010).
 - [10] H. Liao, A. Zeng, R. Xiao, Z.-M. Ren, D.-B. Chen, and Y.-C. Zhang: Ranking Reputation and Quality in Online Rating Systems, PLOS ONE, Vol. 9, No. 5, Article e97146 (2014).
 - [11] Y.-K. Yu, Y.-C. Zhang, P. Laureti, and L. Moret: Decoding Information from Noisy, Redundant, and Intentionally Distorted Sources, Physica A: Statistical Mechanics and its Applications, Vol. 371, No. 2, pp. 732-744 (2006).
 - [12] P. Laureti, L. Moret, Y.-C. Zhang, and Y.-K. Yu: Information Filtering via Iterative Refinement, EPL (Europhysics Letters), Vol. 75, No. 6, pp. 1006-1012 (2006).
 - [13] filmsite: <http://www.filmsite.org/>
 - [14] J. A. Hanley and B. J. McNeil: A Method of Comparing the Areas under Receiver Operating Characteristic Curves Derived from the Same Cases, Radiology, Vol. 148, No. 3, pp. 839-843 (1983).