DOCUMENT RESUME

ED 459 844 IR 058 380

AUTHOR	Riggs, Tracy; Wilensky, Robert
TITLE	An Algorithm for Automated Rating of Reviewers.
SPONS AGENCY	National Science Foundation, Arlington, VA.
PUB DATE	2001-06-00
NOTE	8p.; In: Proceedings of the ACM/IEEE-CS Joint Conference on Digital Libraries (1st, Roanoke, Virginia, June 24-28, 2001). For entire proceedings, see IR 058 348.
CONTRACT	CA98-17353
AVAILABLE FROM	Association for Computing Machinery, 1515 Broadway, New York NY 10036. Tel: 800-342-6626 (Toll Free); Tel: 212-626-0500; e-mail: acmhelp@acm.org. For full text: http://www1.acm.org/pubs/contents/proceedings/dl/379437/.
PUB TYPE	Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*Data Processing; Electronic Journals; *Electronic Publishing; Information Dissemination; Information Processing; Information Systems; *Peer Evaluation; *Scholarly Communication; *Scholarly Journals
IDENTIFIERS	*Filters; *Reviewers

ABSTRACT

The current system for scholarly information dissemination may be amenable to significant improvement. In particular, going from the current system of journal publication to one of self-distributed documents offers significant cost and timeliness advantages. A major concern with such alternatives is how to provide the value currently afforded by the peer review system. This paper proposes a mechanism that could plausibly supply such value. In the peer review system, papers are judged meritorious if good reviewers give them good reviews. In its place, the report proposes a collaborative filtering algorithm that automatically rates reviewers, and incorporates the quality of the reviewer into the metric of merit for the paper. Such a system seems to provide all the benefits of the current peer review system, while at the same time being much more flexible. The researchers have implemented a number of parameterized variations of this algorithm, and tested them on data available from a quite different application. Initial experiments suggest that the algorithm is in fact ranking reviewers reasonably. (Contains 19 references.) (Author/AEF)

An Algorithm for Automated Rating of Reviewers

Tracy Riggs
Robert Wilensky
Division of Computer Science
UC Berkeley
Berkeley, CA 94720
{tracyr,wilensky}@CS.Berkeley.EDU

ABSTRACT

The current system for scholarly information dissemination may be amenable to significant improvement. In particular, going from the current system of journal publication to one of self-distributed documents offers significant cost and timeliness advantages. A major concern with such alternatives is how to provide the value currently afforded by the peer review system.

Here we propose a mechanism that could plausibly supply such value. In the peer review system, papers are judged meritorious if good reviewers give them good reviews. In its place, we propose a collaborative filtering algorithm which automatically rates reviewers, and incorporates the quality of the reviewer into the metric of merit for the paper. Such a system seems to provide all the benefits of the current peer review system, while at the same time being much more flexible.

We have implemented a number of parameterized variations of this algorithm, and tested them on data available from a quite different application. Our initial experiments suggest that the algorithm is in fact ranking reviewers reasonably.

Keywords

collaborative filtering, recommender systems, electronic publishing

1. MOTIVATION AND BACKGROUND

One of the goals of our project [5] is to explore how technology may be exploited to enable alternative models of dissemination of scholarly information. In the traditional system for scholarly information dissemination, with its roots in paper-based documents, authors submit their papers to journals or conferences, where they are reviewed. The papers deemed worthy are then published in journal volumes or conference proceedings, perhaps with some modifications suggested by reviewers or editors. The volumes are then made available to readers. To a first approximation, members of the same scholarly community write, review, and read these papers, with publishers insuring quality control, distribution and, perhaps, editing services. By most accounts (see for example [17]), journal subscription costs are rising exponentially. Hence, the current scholarly publication system in effect manages to sell content produced by scholarly communities back to those communities at exponentially increasing cost. In addition, it imposes long delays, as papers are made available only at the end of the often lengthy review and distribution process. These and other problems with peer review have been addressed by several researchers, for example those in the medical community [16].

Digitalization per se provides no significant improvement. Digitization can improve the distribution process slightly, but the bulk of the delay is in the reviewing, not in the shipping. Moreover it cannot alter the basic cost structure of the system: Publishers are economically motivated, and hence will license electronic versions of journals in accordance with some charging model that lets them recover at least the same amount. Thus, a significant improvement, we maintain, requires a radical change in the basic structure of scholarly information dissemination.

Fortunately, such a radical change is possible, because the originators of scholarly content are not economically motivated. Writers of research papers receive no direct financial incentive for publishing in journals and conferences.[1] Instead, they generally seek wide circulation and recognition of the merit of their views, discoveries, etc., for other reasons, be they altruistic, or in hopes of obtaining academic rewards, such as tenure, promotion, and the esteem of one's peers.

Abstractly, we can characterize information dissemination generally as "Publication = Distribution + Filtering". Filtering is the function provided by the review system (or, more generally, by whatever mechanism a publisher uses to decide what to publish); distribution is the circulation or availability of the content subsequently. Thus the current system filters first, and distributes later, a sequence that makes sense when distribution costs are relatively high, as they are in paper-based systems. Technology, of course, makes distribution simple and cheap–most authors can post

JCDL'01, June 24-28, 2001, Roanoke, Virginia, USA.
Copyright 2001 ACM 1-58113-345-6/01/0006 ...$5.00.

[1] Of course, other, related forms of scholarship, e.g., authoring textbooks or popular publications, generally have an appreciable financial motive. While the work presented here is still applicable in these contexts, going to a self-distributed system doesn't obviously provide cost benefits if the user is charged.

a paper on web server at little or no cost, thus making it available to the world. Indeed, the scarcest resource is generally attention, so it makes economic sense to distribute first, and filter later [19].

The problem, then, is to provide a filtering mechanism that is at least as good as that provided by the current peer review system, but which can operate in the context of self-distributed publications.

There are several parts to this problem. One is that reviewers sometimes make detailed comments on submitted papers, and these comments are often relayed back to authors. Providing finer-grain capabilities for distributed annotation of electronic documents would further this goal, and we have been developing the document technology to do so, which is described elsewhere [13].

While such comments are an important part of the scholarly process, the reviewer's primary task is to indicate some rating of the merit of the submitted work. There is considerable variation of rating systems from one venue to the next, but all have the following components: Reviewers rate submissions along some scale or scales; the highest ranked submissions are published, with editors adjudicating mixed reviews or other controversies.

Thus a system in which the rating of articles is made available to readers would allow the readers to select papers that have the highest aggregate review ratings. Such a system should in theory provide the equivalent of the current peer review system, although, of course, it would be much more flexible. Readers might be interested in seeing papers in which reviewers disagree, or in looking at papers that might be quite good, but beneath the arbitrary cutoff of a journal or conference proceeding. Such functionality is not easily accommodated in the traditional journal system, but could easily be done in the system we propose. In other words, the current peer review system should be approximated by an appropriate collaborative filtering system, which would also be capable of offering additional value.

So far as we are able to discern, the primary value that journals claim to provide is quality control, in the form of the quality of the reviewers that they use. It may indeed be the case that the better journals manage to secure the services of better reviewers (and, perhaps, authors then self-select their publications, so that better journals receive better submissions as well). The bottom line, then, is judging the reviewers. That is, readers want not the papers with the best reviews, but the papers deemed best by the best reviewers. Thus, we need a collaborative filtering system that will automatically provide simultaneous quality filtering of both papers and reviewers.

There has been considerable work on collaborative filtering and recommender systems [1, 4]. The vast majority of this work relies on the principle of finding users with similar affinities. For example, Tapestry [9] provided a filtering system for e-mail messages; the GroupLens [10] project initially targeted Usenet and movies, more recently extending its scope to include general information filtering algorithms; PHOAKS [18] supports recommending and annotating Usenet messages; and Siteseer [14] and Fab [7] perform filtering on World Wide Web pages. Affinity-based recommender systems are also gaining popularity in E-commerce [15].

In contrast, we are attempting to perform collaborative *quality* filtering, based on the principle of finding the most reliable users. We would categorize as collaborative quality filtering work such as [12], which supports automatically assessing reputations in the context of E-commerce transactions. However, the problem of automatically rating reviewers seems unaddressed. Here we provide an algorithm to do so. The algorithm is applicable to any collaborative filtering scenario in which reviewers rate items along some scale. Indeed, our initial tests of the algorithm are in quite a different domain, because of the availability of reviews and reviewer ratings.

2. ALGORITHM

The general idea of the algorithm is that good reviewers are those whose reviews predict the ultimate consensus review of an item. We assume that the average rating of an item is the closest measure we can obtain to the true "value" of that item. Thus, any reviewer who consistently ranks items near their ultimate average can be considered to be a reliable reviewer.

2.1 Basic Algorithm

The basic algorithm is quite simple. First, we assume that each reviewer's rating, and each item's rating, can be translated into a score between 0 and 1. The following is a general outline of the algorithm:

```
while (not converged)
   compute item rating as weighted average
   compute reviewer score based on how close
       to average reviewer rates items
```

This iterative algorithm is similar to the web-searching algorithm proposed by Kleinberg [11]. In Kleinberg's algorithm, web pages are labeled as *hubs* and/or *authorities*. The hubs are pages which point to many authorities, and the authorities are pages that are pointed to by many hubs. An iterative algorithm is used to compute hubs and authorities, in which the "hub score" and "authority score" of a set of pages are alternately computed until convergence is reached. (A related idea is to apply Kleinberg's algorithm to research papers by using the citation graph; this feature is offered by Citeseer [3].) Note that in Kleinberg's algorithm, if the hub score of page X increases, then that increases the contribution of X to the authority scores of all the pages to which X points. The algorithm proposed here is similar–if the score of reviewer Y goes up, then that increases the contribution of Y to the ratings of items that Y has reviewed.

Kleinberg's algorithm does not apply directly here, as Kleinberg deals with a symmetric matrix of items versus items, whereas we have a set of reviewers versus a set of items. Furthermore, a reviewer should not benefit from giving an item a high score, but rather should benefit from giving an item a score that is close to the item's weighted average. This leads to a nonlinear iterative algorithm, as opposed to Kleinberg's linear algorithm. Kleinberg offers a proof that his algorithm will always converge. We do not yet have such a proof for our algorithm (and doubt that a straightforward proof exists), but in practice, we have found that it converges rapidly.

A related idea has been proposed separately by Canny [8]. Canny's algorithm is based on a consensus model; a score is assigned to each reviewer based on how closely that reviewer's ratings vector correlates with the vectors of other reviewers. Our algorithm is based on a similar idea, but also incorporates the item averages in the iteration.

The formula for computing the item rating is simple; it is just a weighted average:

$$a_j = \frac{\sum_{i \in R_j} w_i r_{ij}}{\sum_{i \in R_j} w_i}$$

where a_j is the rating for item j, R_j is the set of reviewers that have reviewed j, w_i is the score of reviewer i, and r_{ij} is the rating that reviewer i gave to item j.

Computing the reviewer scores is slightly more involved. We compute the average difference between r_{ij} and a_j, which is simply the Manhattan distance divided by the number of items reviewed.

$$w_i = 1 - \frac{\sum_{j \in S_i} |a_j - r_{ij}|}{n_i}$$

Here S_i is the set of items that user i has reviewed, and n_i is the cardinality of S_i.

2.2 Additional Factors

There are a number of additional factors that one may or may not want to incorporate in the algorithm. We define three of these factors as α, β, and γ, which are incorporated into the formula for calculating reviewer score as follows:

$$w_i = \alpha_i \left[1 - \frac{\sum_{j \in S_i} \gamma_{ij} \beta_j |a_j - r_{ij}|}{\sum_{j \in S_i} \gamma_{ij} \beta_j} \right]$$

We now discuss each of these factors in turn.

2.2.1 Number of items reviewed

If a reviewer has rated one item close to the average, it would seem unwise to conclude that he or she deserves to be ranked among the top reviewers. Instead, we might want to discount inexperience (or lack of data). The factor α is to compensate for such a lack of data, and is defined as:

$$\alpha_i = 1 - \frac{1}{n_i}$$

2.2.2 Number of reviews of an item

Another consideration is the number of reviews available for an item. If a reviewer rates an item that has very few reviews, then, without any adjustment, that review will greatly influence the overall rating of the item, and, consequently, suggest that the reviewer is highly reliable. In contrast, a review of an item that has been reviewed by many reviewers will not influence the score of that item much, and hence, have a much smaller effect on the subsequent assessment of the reviewer, despite the fact that the reviewer provided the same value in both cases. Thus, the factor β is used to give more weight to those items that have received more reviews:

$$\beta_j = 1 - \frac{1}{m_j}$$

Here m_j is the number of reviews that item j has received.

2.2.3 Time of review

Consider the time at which a reviewer rates an item with respect to other reviewers. If a reviewer is an early reviewer, and is close to the subsequent average, then that reviewer has in fact predicted the average. In contrast, if a reviewer has available to him or her the benefit of many previous reviews, that reviewer could influenced by those reviews, a concept known as "herding" [6]. It is reasonable, then, to give more credit for reviews of an item for which fewer reviews are available than reviews for which more reviews are available. Doing so is the point of the factor γ, defined as:

$$\gamma_{ij} = 1 - \frac{t_{ij}}{m'_j}$$

where m'_j is the number of available reviews and t_{ij} is the *rank* of reviewer i with respect to item j (the number of reviews that were available when reviewer i rated item j, plus one). Here we assume that a reviewer for which no reviews are available has a rank of 1; the last reviewer of an item has a rank of m'_j (when all previous reviews are available, $m'_j = m_j$).

2.3 Exploiting Undue Influence

A reviewer could attempt to unduly influence the system as follows: He rates many items in which he has no great interest at their known average, to eventually obtain a high reviewer rating; then he rates a few items in which he has a great interest as he desires, in an attempt to have a greater influence on their average. Such spoofing could be used to advance "cliques", or groups of people that would like to promote each other's work.

The parameter γ, which uses the rank order of the reviewer's rating, could alleviate this problem somewhat. However, the parameters α and β could exacerbate it.

Our belief is that this vulnerability is in fact an intrinsic problem of peer review, rather than a problem with the algorithm per se. Indeed, traditional peer review processes try to filter potential reviewers for conflicts of interest in a variety of ways: asking reviewers to name their students and advisors, or presenting papers to be reviewed without authorship in evidence. Of course, each of these measures can be implemented in our collaborative filtering scenario. However, both in the traditional case and in our proposal, such measures will be at best superficial. Indeed, it is hard to discern the difference, in terms of the patterns of reviews, between cliques of malicious spoofers and affinity groups of scholars with deeply held differences of opinion.

We believe our algorithm is not exceptionally vulnerable to this intrinsic problem, and may indeed provide some help. For example, with a data base of reviews available, it may be possible to automatically detect spoofers, or affinity groups of scholars, and adjust the weighting of a review in accordance with such an affinity. We leave this problem for future work.

2.4 A Variation on the Algorithm: Assessing Reviewer Expertise

An additional aspect we may incorporate into the algorithm is that a reviewer may have multiple areas of interest, but may not necessarily have the same level of knowledge in all areas. Thus, the reviewer may be more skilled at judging papers in one research area than another. We have proposed an enhancement to the algorithm that accounts for this detail.

There are various ways to approach the addition of this feature. One possibility is to categorize all of the documents and to give the reviewer a score for each category. However, classifying documents in this manner is limiting, as papers generally overlap several categories. We chose a method based on using pairwise similarity among documents. Two documents can be compared to one another, for example by computing the cosine of the angle between the word vectors of the documents, thus resulting in a similarity measure between them.

In the enhanced algorithm, the rating for each reviewer is a vector rather than a scalar, so a reviewer has a different score for each item - a measure of his or her "expertise" on rating that item.

$$a_j = \frac{\sum_{i \in R_j} w_{ij} r_{ij}}{\sum_{i \in R_j} w_{ij}}$$

Here w_{ij} is the reviewer i's expertise rating for item j. In this variation of the algorithm, we compute a weight for each reviewer-item pair. The idea is based on the following principle: if a reviewer has rated many items similar to item j and has given those items accurate ratings, then he or she has a high level of "expertise" on item j. Let s_{jk} be the similarity of items j and k. Then we compute w_{ij} for reviewer i and item j as follows:

$$w_{ij} = \frac{\sum_{k \in S_i} s_{jk}(1 - |a_k - r_{ik}|)}{\sum_{k \in S_i} s_{jk}}$$

The additional factors discussed in the paper may also be incorporated in the enhanced algorithm. The following equation incorporates these factors:

$$w_i = \frac{\alpha_i \sum_{k \in S_i} \gamma_{ik} \beta_k s_{jk}(1 - |a_k - r_{ik}|)}{\sum_{k \in S_i} \gamma_{ik} \beta_k s_{jk}}$$

3. AN EXPERIMENT

The basic algorithm and its parameterized variations were tested on data gathered from Epinions.com, a web site designed for consumers to share product reviews with other consumers. The Epinions.com data was chosen for several reasons: (1) the data are usable for testing the algorithm because members give items numerical ratings, (2) it is a popular website and therefore contains a large amount of data, and (3) the Epinions.com assessment of member "reliability" may be used as a metric by which to measure the performance of the algorithm.

We have not yet tested the variation of the algorithm that includes assessment of reviewer expertise, but intend to conduct a similar experiment using Epinions.com data. The items on Epinions are arranged in a taxonomy, allowing us to use item proximity in the graph as a similarity measure.

3.1 About Epinions.com

Members of Epinions.com submit reviews for any item in a finite set of items maintained by Epinions.com. The member rates the item using a score of 1 to 5 (5 being the best) and also offers a written review. Other members may then rate the review in terms of whether or not they would recommend the review to others. Furthermore, a member may read several reviews by another member and then decide to either "trust" or "distrust" that member. The result is that some members end up being "highly trusted" or have "highly recommended" opinions, while others are "not trusted" or have "not recommended" opinions.

3.2 Metrics

The following metrics were used for assessing the algorithm's performance:

- The number of members that trust a reviewer. The more a reviewer is trusted, the more reliable we can expect her reviews to be. However, a reviewer who has written more reviews can be expected to have more trusters, simply by virtue of being more visible in the community. Therefore, the number of trusters is normalized by dividing by the number of reviews written.

- The average "recommendation level" of a reviewer's reviews. We assign a score to each possible rating of a reviewer's review: "highly recommended" = 3, "recommended" = 2, "somewhat recommended" = 1, and "not recommended" = 0. If we take all of a reviewer's reviews and average the numerical value assigned to them, that should be a reasonable measure of the reliability of that reviewer.

4. RESULTS

We have run the algorithm on a set of 100,000 reviewers from the Epinions.com community. Figure 1 shows the results of the algorithm measured against the number of "trusters", and Figure 2 shows the results measured against the average recommendation level.

The graphs may be interpreted as follows: each point on the horizontal axis represents the group of reviewers who fell into a given score range. For instance, if a reviewer's score was 0.64, the reviewer is included in the group 0.6-0.8. Within each group, the average number of trusters per review (Figure 1) and the average recommendation level (Figure 2) were computed. The five bars within each group correspond to five different variations of the algorithm. For the bars labeled "none", $\alpha = 1$, $\beta = 1$, and $\gamma = 1$. For the bars labeled "α", $\gamma = 1$, $\beta = 1$, and α is computed as described above, and so forth. For each variation of the algorithm, a single-factor ANOVA test showed that the five groups were significantly different at a 99% confidence interval ($p < 0.01$).

In general, the graph shows that the ratings given by our algorithm tend to increase as the ratings given by the Epinions.com metrics increase. The most dramatic results are seen when the parameter α is used. This is unsurprising, because we would expect to see high reliability among the active members of Epinions.com. Interestingly, the algorithm appears to correlate with Epinions.com data even when no additional factors are used. The factors β and γ have a less dramatic effect.

5. DISCUSSION

We believe that these initial results suggest that (some variations of) our proposed algorithm provides a plausible way to automatically assess the reliability of reviewers, and

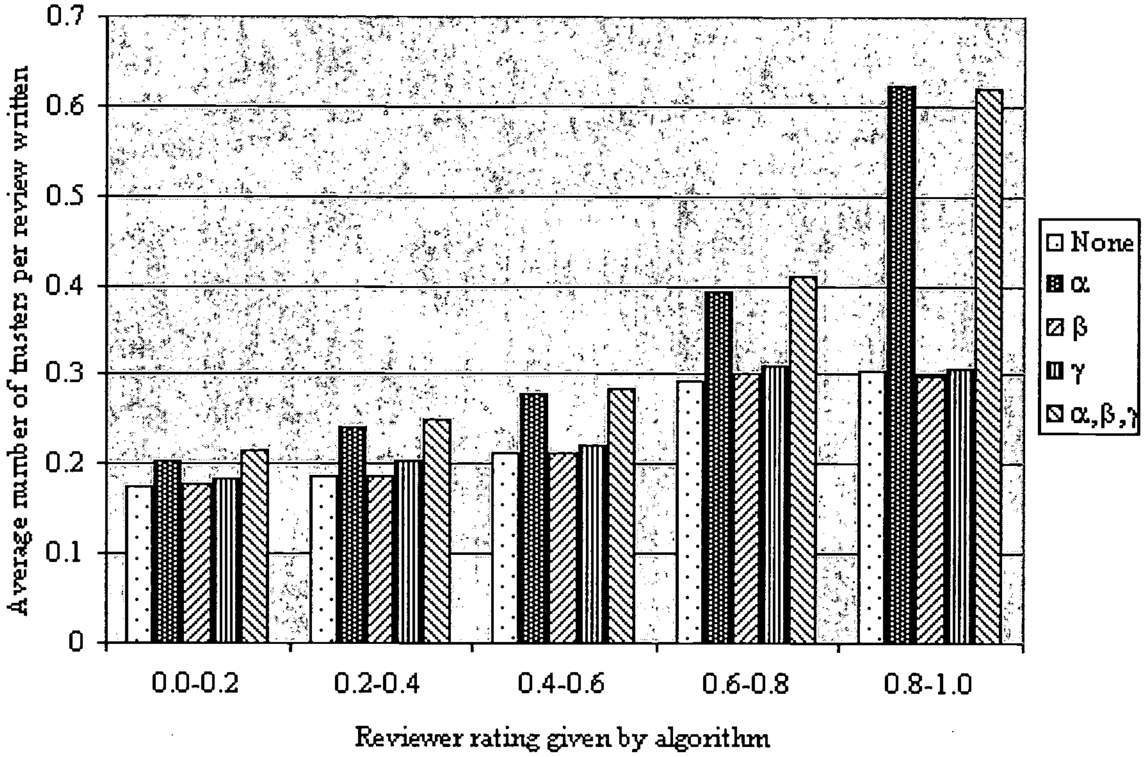

Figure 1: Epinions Trusters vs. Algorithm Rating

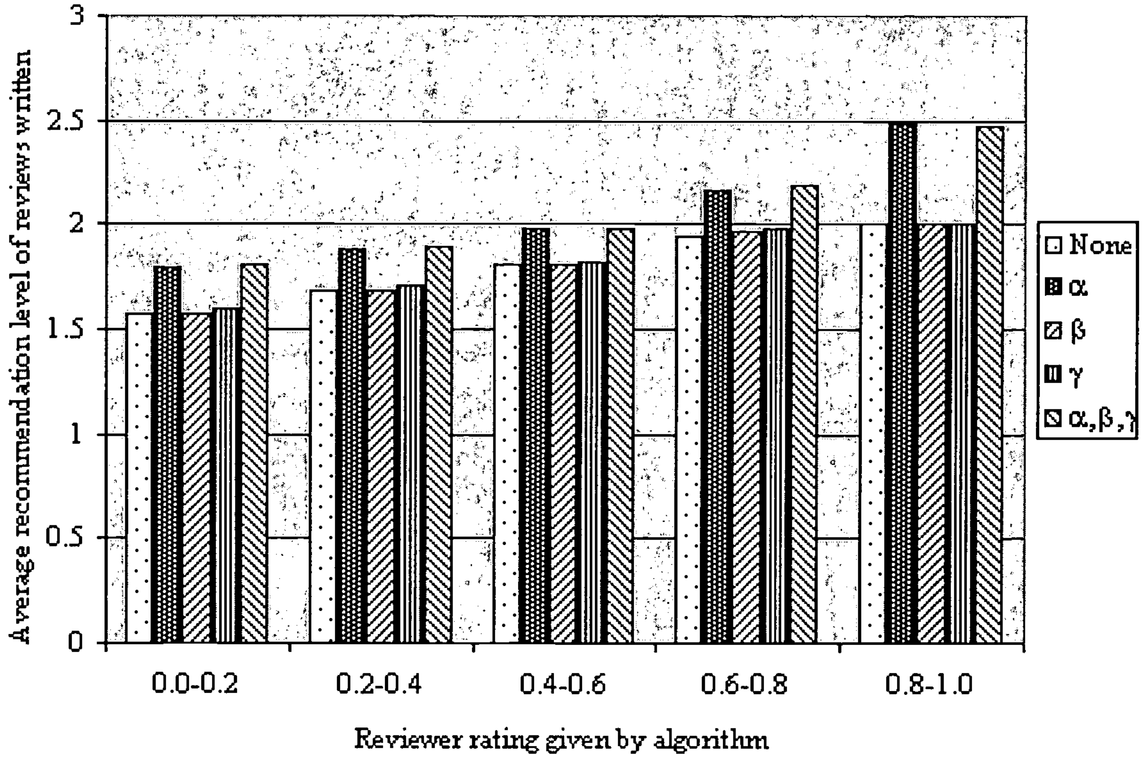

Figure 2: Epinions Recommendation Level vs. Algorithm Rating

hence, may serve the purpose of its design, namely, to supply the value of peer review in a self-distributed system of scholarly information dissemination. We must be tentative about our conclusions, of course, since we intend the algorithm to be used for scholarly information dissemination, and it is difficult to judge its efficacy using the consumer-oriented Epinions.com data. One reason that the Epinions.com metrics are not entirely ideal is that members are generally rated on the *quality* of their written review, rather than on the *accuracy* of their numeric rating. We expect there to be some correlation between the two, but have no way to verify this conjecture. However, we suggest that the Epinions.com data provide a reasonable starting point. A more definitive test would involve deploying the algorithm in a context for which it was designed, which we plan to do.

6. FUTURE WORK

One feature of the current algorithm is that it conflates confidence with quality. Specifically, one of our parameters discounts the rating of a reviewer based on the number of reviews he or she has done; another discounts the rating based on the number of reviews contributing to the item rating. However, the reviewer's quality may be excellent to begin with; it is only our confidence in his or her work that is increasing. Thus, separating the assessed quality of a reviewer from the system's confidence in that quality may be desirable, although we are uncertain that doing so will affect the algorithm's bottom line.

We mentioned above that current review systems often ask reviewers to rate papers along more than one dimension. The algorithm described here could easily be applied to multi-dimensional reviewing strategies, simply by applying it independently to each rating dimension. Indeed, it would be particularly useful, at least in some fields, to rate papers along a "correctness" dimension and an "importance" dimension, as an interesting theory may ultimately turn out to be false, but still be important, and indeed, highly referenced by discrediting work, and skeptical reviewers would have a means to express a "positive disagreement", i.e., lower correctness but high importance. Of course, rating papers along multiple dimensions also opens the possibility of rating reviewers along these same dimensions.

Separating out an importance and a correctness dimension allows for another, substantial addition to the algorithm: This is to regress on author citations. That is, the number of citations to a work is some measure of the importance of the work. Thus, reviewers whose previously "highly important"-rated articles ended up with large numbers of citations should also be considered good reviewers (insofar as importance is concerned).

Along this line, there are many other parameters one may suspect are correlated with the reliability of a review, such as the length of commentary, the institution with which a reviewer is affiliated, and so forth. With such a rating system in place, we might be able to find out if our intuitions about such items have empirical merit.

There are a number of variations of the algorithm that may be worth exploring:

- Use a different measure of distance from the average (a change to the "basic algorithm"). Superlinear distance measures will have the effect of penalizing one big "error" in a review more than the same about of error distributed over many reviews. We believe doing so is undesirable. But perhaps some other measure would prove valuable.
- Use different values for the parameters α, β, and γ. For instance, α is a such a major factor in the reviewer rating, we may wish to reduce its influence.
- There are other factors that could be used and have not been mentioned in this paper. For example, one might try to solicit a "degree of confidence" from the reviewer, i.e., a self-rating of the reviewer's own confidence in his or her review. This would be helpful if the reviewer wanted to spend only a short amount of time on the paper, or questioned his own expertise, etc. (This could not be tested using an Epinions.com metric.)

As mentioned above, it may be possible to automatically detect spoofers, or affinity groups of scholars. For example, reviews by reviewers that give each other mutually excessively positive reviews could be discounted. Alternatively, one could use affinity groups for paper recommendations. In this use, the proposed system would act more like other collaborative filtering systems, in which users simply use reviewers that they like to filter for them.

We also proposed a variation on the algorithm that includes an enhancement for assessing the reviewer's expertise in a given research area. This variation has not yet been tested, but we are currently in the process of experimenting with the algorithm using the Epinions.com data.

Of course, the algorithm should be tested in a real system where it can be judged by actual users. Performing such a test is an important future step. There are many practical and sociological issues that need to be addressed to deploy such an algorithm in a realistic context. One is to motivate individuals to review papers, and to review them accurately. While we do not believe the sociology of reviewing is well-understood, we believe that practices found effective in both traditional reviewing and other collaborative filtering work can be applied here. For example, [6] suggests that keeping early reviews unavailable is effective in both soliciting subsequent reviews and preventing "herding".

We suggest that a collaborative filtering scheme such as we propose may not only provide the same value as supplied by peer review, but may ultimately provide additional value. Journal editors believe they know who the good reviewers are, but such knowledge is apparently largely anecdotal–perhaps algorithms such as this one will provide a more objective assessment. Similarly, the value of reviews can be tracked, and one's prowess as a reviewer measured, so that the rewards currently associated with such activity may be better calibrated. Finally, entirely new motivational schemes are possible. For example, The Berkeley Electronic Press [2] has established an "authors and reviewers' bank", in which authors must review other authors' papers in order to receive reviews for their own.

The inclination to use a system such as we propose is likely to vary from discipline to discipline. In some fields, authors are very careful not to leak results pending very careful reviews; obviously, such scholarly communities would be less interested in self-distribution and quality filtering. It is an interesting challenge to see if mechanisms such as the ones we propose can be applied to disciplines with such different sociologies.

7. ACKNOWLEDGEMENTS

Paul Resnick pointed out to us the interesting difficulties of reviewer motivation. Hal Varian made valuable comments on a previous version of this paper.

This research was supported by the Digital Libraries Initiative, under grant NSF CA98-17353.

8. REFERENCES

[1] ACM-SIGIR 1999 Workshop on Recommender Systems: Algorithms and Evaluation. http://www.csee.umbc.edu/ ian/sigir99-rec/summary.html.

[2] The Berkeley Electronic Press. http://www.bepress.com/.

[3] CiteSeer. http://citeseer.nj.nec.com.

[4] Collaborative filtering. http://www.sims.berkeley.edu/resources/collab/.

[5] The UC Berkeley Digital Library Project. http://elib.cs.berkeley.edu.

[6] C. Avery, P. Resnick, and R. Zeckhauser. The Market for Evaluations. *American Economic Review*, 89(3):564–584, 1999.

[7] M. Balabanovic and Y. Shoham. Fab: Content-Based Collaborative Recommendation. *Communications of the ACM*, 40(3):66–72, March 1999.

[8] J. Canny. Personal communication with the authors.

[9] D. Goldberg, D. Nichols, B. M. Oki, and D. Terry. Using Collaborative Filtering to Weave an Information Tapestry. *Communications of the ACM*, 35(12):61–70, December 1992.

[10] J. L. Herlocker, J. A. Konstant, A. Brochers, and J. Riedl. An Algorithmic Framework for Performing Collaborative Filtering. In *Proceedings of the 1999 Conference on Research and Development in Information Retrieval.* ACM-SIGIR, August 1999.

[11] J. M. Kleinberg. Authoritative Sources in a Hyperlinked Environment. In *Proceedings of the ACM-SIAM Symposium on Discrete Algorithms*, 1998.

[12] G. Z. A. Moukas and P. Maes. Collaborative Reputation Mechanisms in Electronic Marketplaces. In *Proceedings of the 32nd Hawaii International Conference on System Sciences*, 1999.

[13] T. A. Phelps and R. Wilensky. Multivalent Documents: Anywhere, Anytime, Any Type, Every Way User-Improvable Digital Documents. *Communications of the ACM*, 43(6), June 2000.

[14] J. Rucker and M. J. Polanco. Siteseer: Personalized Navigation for the Web. *Communications of the ACM*, 40(3):73–75, March 1997.

[15] J. B. Schafer, J. Konstan, and J. Riedl. Recommender Systems in E-Commerce. In *Proceedings of the ACM Conference on Electronic Commerce*, November 1999.

[16] R. Smith. Opening up BMJ peer review. *BMJ*, 318:23–27, 1999.

[17] C. Tenopir and D. W. King. Trends in scientific scholarly journal publishing in the United States. *Journal of Scholarly Publishing*, 28:135–170, 1997.

[18] L. Terveen, W. Hill, B. Amento, D. McDonald, and J. Creter. PHOAKS: A System for Sharing Recommendations. *Communications of the ACM*, 40(3):59–62, March 1997.

[19] H. R. Varian. The Future of Electronic Journals. *Technology and Scholarly Communication*, 1999.

NOTICE

REPRODUCTION BASIS

☒ This document is covered by a signed "Reproduction Release (Blanket) form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.

□ This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").