Detecting Reviewer Bias through Web-Based Association Mining

Jessica Staddon
PARC
staddon@parc.com

Richard Chow
PARC
rchow@parc.com

ABSTRACT

Online retailers and content distributors benefit from an active community that shares credible reviews and recommendations. Today, the most popular approach to encouraging credibility in these communities is self-regulation; community members rate reviews according to their accuracy and usefulness, thus helping to weed out reviews that are inaccurate. This self-regulation, while powerful, is limited by its insularity. Community members generally base their assessments on a reviewer's comments and actions only within the community. This ignores relationships the reviewer has outside the community that may be quite relevant to evaluating the reviewer's comments; for example, a relationship between an author and reviewer. We present a simple method for mining the Web to detect many such associations. Our method, together with self-regulation, provides for more comprehensive detection of bias in reviews by alerting the user to the potential for an undisclosed relationship between a reviewer and author. We provide preliminary results using book reviews in `Amazon.com` demonstrating that our approach is a high-precision method for detecting strong relationships between reviewers and authors that may contribute to reviewer bias.

Categories and Subject Descriptors

K.6.5 [**Computing Milieux**]: Management of Computing and Information Systems—*Security and Protection*

General Terms

Security

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
WICOW'08, October 30, 2008, Napa Valley, California, USA.
Copyright 2008 ACM 978-1-60558-259-7/08/10 ...$5.00.

Keywords

Bias, reputation, trust, association rule mining.

1. INTRODUCTION

Online retailers encourage their user communities to contribute reviews and advice because such communities draw users and translate into increased sales [8, 5, 20]. However, these communities cannot be left unattended. Without some supervision there is a risk that the value of the community's content will be diluted by biased or otherwise erroneous posts (see, for example, [24]) and that the content provider will be vulnerable to lawsuits [26, 15]. A popular approach to mediating this risk is self-regulation in the form of ratings and comments. Examples of this include `Amazon.com`'s "Was this review helpful to you?" feature, "Karma" on `Slashdot.com` and "diggs" on `Digg.com`. These ratings are generally based on the content of the associated review, and possibly, the reviewer's history in the community, as this is the only information that is readily available to the community. That is, the self-regulation system as it commonly exists today does not support the discovery of information external to the community. Such information can be valuable when detecting the potential for bias on the part of a reviewer.

The lack of tools to assist the community in self-regulation means that relevant external relationships are hard to detect. For example, the `Amazon.com` community was slow to discover that a reviewer of obscure technical books had a career as a seller of the same books [17]. In addition, professional relationships and friendships are not easily detected, yet may be quite useful when weighing reviews.

We introduce an approach to assessing the validity of online reviews that aims to bring the broader context of the reviewer into the online community. Our approach uses simple Web-mining algorithms to discover association rules that impact bias. We apply association rule mining algorithms used in [6] to discover keywords associated with sensitive topics, to detect external (and Web-represented) relationships between users of online communities. In particular, we look for association rules between book reviewers and the authors of the books they review. An association rule of the form $Reviewer\ A \rightarrow Author\ B$ reflects that $Author\ B$ is frequently mentioned in Web documents that also mention

Reviewer A. Clearly, co-occurence of names in a single document is insufficient evidence of a relationship. However, when this co-occurence is repeated across a large number of documents that are also a significant fraction of the documents containing either one of the names, we believe that it is compelling evidence of a relationship. Indeed, this simple idea of looking for co-occurence of terms on the Web has been shown to be powerful in a broad array of information retrieval applications including search query interpretation (e.g. [11]) and indexing and annotation [7, 9], as well as the fundamental natural language problem of identifying semantic patterns [4, 19, 25].

We demonstrate our approach using review data extracted from `Amazon.com`. In particular, we mine the reviews associated with 300 books categorized as "cryptography" by Amazon for associations between the reviewers and the authors. Our results indicate that the approach offers high precision (few false positives). There is an inevitable limitation to the recall of our approach due to the fact that relationships not reflected strongly on the Web are not detected by our algorithm. However, for authors and reviewers with a strong Web presence, it performs well. These results demonstrate our detection tool can enhance the accuracy of self-regulation in online communities. In addition, the increasing availability of search engines with little or no query limits (see, for example, [16, 27]) means the approach can be used on a broad scale.

OVERVIEW. In Section 2 we discuss related work. Our model and algorithm are described in Section 3. Experimental results are in Section 4 and we conclude in Section 5.

2. RELATED WORK

Our approach to detecting potential bias relies on an algorithm for mining Web-based association rules introduced in [6]. We discuss association rule mining more in Section 3.

As mentioned earlier, the most prevalent approach to evaluating reviews relies on the user community to detect bias and other fallacies in reviews. Complementing this, there are some services (e.g. eModeration [10]) that provide manual evaluation of user contributions as a for-fee service.

Our bias detection technique is a natural component of a reputation system (see, for example, [22, 12, 21]. These systems generally rely on the actions of the community members to distribute trust information accurately. Our technique is complementary as it gives community members additional input on which to base their trust evaluation.

In [1, 23] and [13], automated approaches to measuring reputation are proposed. In [1], reputation is based on the stability of contributed content over time. In [23] and [13], reviewer reputation is correlated with the accuracy of reviews over time, where a review is more accurate the closer it is to the consensus view. Hence, all these approaches require multiple reviews to be effective. In contrast we offer an approach that can be used to evaluate a single isolated review.

Figure 1: The confidence of the association Bapna $\Rightarrow$ Shmueli is 0.24.

Finally, we note that our approach is similar in spirit to the Wikiscanner tool for assessing the credibility of Wikipedia content [14], as that tool also brings external information (namely authorship via IP address) into the credibility evaluation process.

3. MODEL AND APPROACH

We detect potential bias through *association rules* [2, 3]. In our experimental setting, an association rule is an implication of the form $A \Rightarrow B$, where A is the name of a reviewer of a book authored by B or B is the name of a reviewer of a book authored by A. Recall that an association rule, $A \Rightarrow B$, is said to have high confidence if $\Pr(B|A)$ is large, and large support if $\Pr(A \wedge B)$ is large.

Our algorithm takes as input the creators of a product (e.g. authors of a book) and reviewers of that product. Let the function $R(\cdot)$ output the set of reviewers associated with a given creator, that is, $B \in R(A)$ if and only if B has reviewed a product created by A. We use the algorithm of [6] to look for associations between all pairs in the set $P = \{(A, B)|A \in R(B) \, or \, B \in R(A)\}$. In particular, for $(A, B) \in P$ we take the following steps:

1. Issue a search engine query: "A", and retrieve the number of returned hits, n_A.

2. Issue a search engine query: "B", and retrieve the number of returned hits, n_B.

3. Issue a search engine query: "A" "B", and retrieve the number of returned hits, $n_{A \wedge B}$.

Association Evidence	Number of Pairs
Co-authors on a published work	11
Acknowledgment in a published work	1
Academic advisor and advisee	1
Interviewer and interviewee in a published article	2
Co-organizers of a conference	1
Co-members of standards committee	2
Speakers at the same conference	2

Table 1: The evidence found for the 20 identified associations.

Based on this, we estimate the confidence of the association $A \Rightarrow B$ using the same mechanism as in [6], namely:

$$\text{Confidence}(A \Rightarrow B) \approx n_{A \wedge B}/n_A$$

where n_A and $n_{A \wedge B}$ are the number hits returned by the search engine on the query "A" and the number returned for the query, "A" "B", respectively.

If $A \in R(B)$ then we say A has *potential bias* if at least one of $A \Rightarrow B$ and $B \Rightarrow A$ is a high-confidence association rule as defined above. Of course, a high-confidence association rule is not definitive evidence of bias, however we believe it is useful information when evaluating the credibility of the review.

We illustrate our approach to detecting associations with a concrete example. In Figure 1, we show a review of the book, "Data Mining for Business Intelligence". A Google query for the reviewer's name, Ravi Bapna, yields 3,130 results, while a query for the reviewer, Ravi Bapna, together with the first author, Galit Shmueli, yields 760 results.[1] Hence, the association {Ravi Bapna} $\Rightarrow$ {Galit Shmueli} has a confidence of 0.24, and indeed the first page of results for the latter query reveals they have authored papers together.

4. EXPERIMENTAL RESULTS

To evaluate our approach we considered the top 300 books (ordered by relevance) categorized as "cryptography" by
`Amazon.com`. We chose the area of cryptography for two reasons. First, given that the cryptography community is relatively small, our suspicion was that it would lead to more instances of associations between reviewers and authors than other categories (for example, the New York Times best seller list). Second, we have some familiarity with the community through program committee service and conference publications and so we have a better ability to evaluate the accuracy of our algorithm in detecting associations between people than we would in other communities.

In the first phase of our experiment, we manually scraped the author and reviewer names from the top 300 cryptography books. We scraped books with at most 20 reviews, since books with more reviews are unlikely to be dominated by reviews written by associates of the authors, and hence our algorithm is less needed. We scraped any name that contained both a first and last name (i.e. not just the names that `Amazon.com` refers to as "real names" because they are linked to a credit card). Middle names and initials were generally included, although we omitted anything in quotes (e.g. for the reviewer John Matlock "Gunny", we just scraped John Matlock). This phase reduced the pool of 300 to 64, as many books had either no reviews or only anonymous or single-name reviewers.

In the second phase of the experiment we issued Google queries for each author name, each reviewer name and each pair of author-reviewer names, and recorded the number of hits returned. There were an average of 1.44 authors per book, an average of 3.45 reviewers per book, and 305 author-reviewer pairs, leading to a total of 620 Google queries in total for the 64 books.[2]

We also manually reviewed the first page of hits for each pair looking for evidence of an association between the reviewer and author to serve as our "ground truth" when evaluating the output of our algorithm. Of course, our ground truth is imperfect as there certainly could be associations not represented in the first page of hits, however we did not find any associations that we were aware of a priori (using our existing knowledge of the community) for which that was the case. Through this manual process we discovered 20 associations. The reasons for the association ranged in strength, with co-authorship being strong evidence of an association, and presentations at the same conference being much weaker (since it does not imply that the speakers had any interaction at the conference, and so does not prove they know each other). We summarize the discovered associations in Table 1 below.

To evaluate the accuracy of our approach we consider all author-reviewer pairs with a minimum number of Google hits (or, "support") of 10. We make this restriction because the hits of pairs with lower support were dominated by links to the `Amazon.com` review written by the reviewer about the author's book, and thus aren't evidence of a relationship between the reviewer and author.[3]

For each association involving a pair with minimum support of 10, we calculate the confidence of the association as defined in Section 3. We set thresholds for minimum confidence and calculate the precision and

[1]Google results are from 8/13/2008.

[2]The Google queries were mostly issued on 7/17/2008.

[3]Another approach to filtering these results would be to require Google to not return hits in the `Amazon.com` domain.

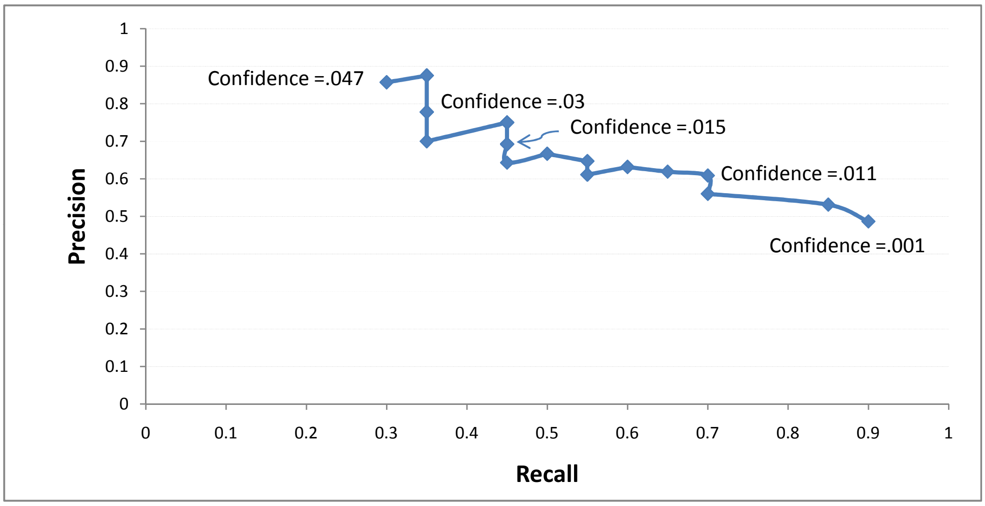

Figure 2: The recall and precision for various minimum confidence values, based on all author-reviewer pairs with support at least 10 (i.e. at least 10 Google hits).

recall of our algorithm with respect to the 20 true associations that we identified manually. That is, for a confidence threshold of c, precision is the fraction of the associations with confidence at least c that are true associations, and the recall is the fraction of the true associations that have confidence at least c. For example, if $c = 0$, and the support threshold is 10, then our algorithm produces 44 pairs including 18 of the true associations, and so precision is 0.409 and recall is 0.9.

We plot the (recall, precision) points in Figure 2. For the high precision region, the graph indicates a roughly 2:1 relationship between precision and recall. The tradeoff stems from the fact that many relationships aren't represented well on the Web, and thus requiring high-confidence in the association causes our algorithm to miss some true associations. This underscores that the best use case for this approach is as a complement to other approaches to assessing reputation/bias.

5. CONCLUSION AND FUTURE WORK

We've introduced a simple, high precision tool for detecting associations between people as an indicator of potential bias in online reviews. With the increasing availability of search engines with little or no bounds on queries [16, 27] our approach can be easily integrated into existing self-regulating communities to give the participants additional context when assessing the credibility of user contributions.

In our initial experiments, we take a simple, generic approach that appears quite accurate. That said, a general search engine query may not be the best approach for certain communities. For example, for academically oriented communities like cryptography, Google Scholar might yield improved results as a function of association confidence.

Another direction for improvement is through increased analysis of the search results. For example, various forms of linguistic analysis of the hits can be used to estimate the likelihood that the hit is really relating the author and reviewer. Such an analysis can be costly (e.g. in the case of deep linguistic parsing) but statistical attributes, like proximity, might prove quite effective while only incurring modest cost.

Acknowledgements

The authors are grateful to the anonymous reviewers, Philippe Golle and Markus Jakobsson for helpful comments on an earlier draft of this paper.

6. REFERENCES

[1] B. Adler and L. de Alfaro. A content-driven reputation system for the wikipedia. *WWW 2007.*

[2] R. Agrawal and R. Srikant. Fast algorithms for mining association rules. *Proceedings of the 20th VLDB Conference*, Santiago, Chile, 1994.

[3] R. Agrawal, H. Mannila, R. Srikant, H. Toivonen and A. Verkamo. Fast discovery of association

rules, *Advances in Knowledge Discovery and Data Mining*, AAAI/MIT Press, Cambridge, MA.

[4] D. Bollegala, Y. Matsuo and M. Ishizuka. Measuring semantic similarity between words using Web search engines. *WWW 2007.*

[5] J. Chevalier and D. Mayzlin. The Effect of Word of Mouth on Sales: Online Book Reviews. August 6, 2003. Yale SOM Working Paper No's. ES-28 & MK-15.

[6] R. Chow, P. Golle and J. Staddon. Detecting privacy leaks using corpus-based association rules. In ACM SIGKDD 2008.

[7] P. Cimiano and S. Staab. Learning by Googling, *ACM SIGKDD Explorations Newsletter*, 2004.

[8] comScore and The Kelsey Group. November 29,2007.

[9] M. Dowman, V. Tablan, H. Cunningham and B. Popov. Web-Assisted Annotation, Semantic Indexing and Search of Television and Radio News. *WWW 2005.*

[10] eModeration. `http://www.emoderation.com`

[11] N. Glance. Community search assistant. *The 6th International Conference on Intelligent User Interfaces*, 2001.

[12] R. Guha, R. Kumar, P. Raghavan and A. Tomkins. Propagation of trust and distrust. *WWW 2004.*

[13] H. Lauw, E-P.Lim and K. Wang. Bias and controversy: beyond statistical deviation. *KDD 2006.*

[14] V. Griffith. WikiScanner. `http://virgil.gr`

[15] A. Gome. Web two point uh-oh! “It’s my shout” Blog on Smart Company. March, 2007.

[16] Hakia. `http://www.hakia.com/`

[17] V. Mickunas. Amazon.com book reviewer shakeout. Dayton Daily News. April 14, 2007.

[18] A. Mowshowitz and A. Kawaguchi. Bias on the Web. Communications of the ACM (CACM), September 2002.

[19] P. Nakov and M. Hearst. Solving relational similarity problems using the Web as a corpus. *Proceedings of ACL/HLT 2008.*

[20] PowerReviews. Social Shopping Study 2007. `http://www.powerreviews.com`

[21] D. Quercia, S. Hailes and L. Capra. Lightweight distributed trust propagation. ICDM 2007.

[22] P. Resnick, R. Zeckhauser, E. Friedman, K. Kuwabara. Reputation Systems. Communications of the ACM, 2000.

[23] T. Riggs and R. Wilensky. An algorithm for automated rating of reviewers. *JCDL 2001.*

[24] G. Sandoval. Digg continues to battle phony stories. CNET.com, December 18, 2006.

[25] P. Turney. Expressing implicit semantic relations without supervision. *Proceedings of ACL 2006.*

[26] R. Verkaik. Google sued over defamatory posts found on Web search. The Independent. June, 2007.

[27] Yahoo! Search BOSS. `http://developer.yahoo.com/search/boss/`