Association for Information Systems
AIS Electronic Library (AISeL)

PACIS 2012 Proceedings

Pacific Asia Conference on Information Systems (PACIS)

7-15-2012

Predicting The Helpfulness Of Online Product Reviewers: A Data Mining Approach

Han-Wei Hsiao
Department of Information Management, National University of Kaohsiung, Kaohsiung, Taiwan, R.O.C, hanwei@nuk.edu.tw

Chih-Ping Wei
Department of Information Management, National Taiwan University, Taipei, Taiwan, R.O.C, cpwei@im.ntu.edu.tw

Yi-Cheng Ku
Department of Computer Science and Information Management, Providence University, Taichung, Taiwan, R.O.C, ycku@pu.edu.tw

Luisa Angelica Chen Ng
Institute of Technology Management, National Tsing Hua University, Hsinchu, Taiwan, R.O.C

PREDICTING THE HELPFULNESS OF ONLINE PRODUCT REVIEWERS: A DATA MINING APPROACH

Han-Wei Hsiao, Department of Information Management, National University of Kaohsiung, Kaohsiung, Taiwan, R.O.C., hanwei@nuk.edu.tw

Chih-Ping Wei, Department of Information Management, National Taiwan University, Taipei, Taiwan, R.O.C., cpwei@im.ntu.edu.tw

Yi-Cheng Ku, Department of Computer Science and Information Management, Providence University, Taichung, Taiwan, R.O.C., ycku@pu.edu.tw

Luisa Angelica Chen Ng, Institute of Technology Management, National Tsing Hua University, Hsinchu, Taiwan, R.O.C.

Abstract

The purpose of this study is to propose a data mining approach to predict the helpfulness scores of online product reviewers. Such prediction can facilitate consumers to judge whether to believe or disbelieve reviews written by different reviewers and can help e-stores or third-party product review websites to target and retain quality reviewers. In this study, we identify eight independent variables from the perspectives of reviewers' review behavior and trust network to predict the helpfulness scores for these reviewers. We adopt M5 and SVM Regression as our underlying learning algorithms. Our empirical evaluation results on the basis of two product categories (i.e., Car and Computer) suggest that our proposed helpfulness prediction technique can predict the helpfulness scores of online product reviewers.

Keywords: Product review, helpfulness of reviewers, word-of-mouth, data mining, supervised learning, M5, SVM.

1 INTRODUCTION

Many e-stores and third-party product review websites have established reputation systems to facilitate consumers' purchase decision making. Reputation system not only allows customers to rate or evaluate a specific product, i.e. product review, but it allows customers to rate a review contributed by other users, i.e. helpfulness vote. In this study, we use the term reviewers when referring to customers who contribute product reviews and the term readers when referring to customers who read or comment product reviews written by other customers. While the huge product reviews are being generated daily, most consumers recognize product reviews by credible reviewers as helpful references for their purchasing decision (Li et al. 2011). In particular, product reviews contributed by consumers are more understandable and credible than those written by experts, because consumers tend to consult experience-oriented product information instead of product-oriented product information (Park et al. 2007; Li et al. 2011). However, since a reputation system is open to all customers, consumers also face a very challenging issue; that is, whether reviews provided by a specific reviewer are helpful or not (Mudambi and Schuff 2010). Therefore how to predict the helpfulness of product reviews pertaining to a specific reviewer is an important issue. In particular, retaining reviewers who usually and consistently make helpful reviews is an important marketing strategy to e-stores or third-party product review websites, because helpful reviews may improve customer perception of the usefulness and social presence of the website (Kumar and Benbasat 2006).

A website can use reputation system to estimate the review helpfulness of a reviewer by aggregating (e.g., averaging) readers' evaluations on reviews pertaining to the same reviewer. While this practice appears to be promising and requires straightforward computation to arrive average helpfulness scores for a reviewer, using this user-driven approach to estimate the review helpfulness for a specific reviewer (named focal reviewer from here) may incur several limitations. First, to have a comprehensive coverage, this approach requires at least one reader's helpfulness vote on a review pertaining to the focal reviewer. If none of the reviews written by the focal reviewer ever receives an evaluation from readers, the focal reviewer's average helpfulness of his/her reviews cannot be derived. However, the distribution of review evaluations is often sparse (Guha et al., 2004). Some reviewers may receive many user-evaluations on their reviews, whereas many reviewers may not have any user-evaluation on their reviews. Such sparsity phenomenon significantly limits the applicability of the described user-driven approach to reputation estimation.

Second, user-evaluations on reviews may not be reliable due to well-known malicious and inflated evaluation behaviors. For example, some readers may intentionally give unfavorable evaluations to some reviewers regardless of the quality of their reviews. In contrast, if a reviewer attempts to boost his or her helpfulness scores dishonorably, he or she may create "fake" readers and use these fake readers to give positive user-evaluations to his or her reviews. In both cases, the user-driven approach may not arrive at reliable helpfulness scores for reviewers in online websites.

Accordingly, we proposed a data mining approach to predict the average helpfulness of reviews contributed by a reviewer (i.e., the helpfulness of the reviewer) to address the aforementioned limitations of the user-driven approach. Our approach will consider the review behavior of the focal reviewer and his/her associated trust network as independent variables to predict the helpfulness (i.e., the dependent variable) of the focal reviewer. Furthermore, we consider the helpfulness score of a reviewer product-category dependent. That is, we assume that a reviewer's performance in a product category is different from that in other product categories because his or her expertise levels (or other quality factors relevant to reviews) in different product categories are not identical. Correspondingly, our data mining approach predicts the average helpfulness score of the focal reviewer for a specific product category.

The remainder of the paper is organized as follows. Section 2 reviews the literature relevant to this study. Section 3 details the variables and data mining techniques employed for helpfulness prediction.

In Section 4, we describe our data collection and evaluation design and then discuss important empirical evaluation results. Finally, we conclude in Section 5, with a summary and discussion of some future research directions.

2 LITERATURE REVIEW

A helpful product review was defined as "*a peer-generated product evaluation that facilitates the consumer's purchase decision process* (Mudambi and Schuff 2010, p. 186)." In particular, product reviews are positively associated with sales (Clemons et al. 2006). Since product reviews on a website are attractive to consumers, many commercial websites, such as Amazon.com and Epinions.com, have an incentive to provide more helpful reviews to enhance customer stickiness. Therefore, how to determine the helpfulness of a review and the average helpfulness of reviews written by a reviewer has been an important research issue. There are two streams to predict the helpfulness of a review in prior studies. The first stream predicts the helpfulness of a review by the characteristics of review content. For example, Mudambi and Schuff (2010) adopted review extremity (star rating) and review depth (word count) to predict the helpfulness of a product review. Cao et al. (2011) employed a text mining approach to examine the relationships between the characteristics of review texts and helpfulness votes. They found that the semantic characteristics are associated with helpfulness votes reviews receive. Forman et al. (2008) found that moderate book reviews were less helpful than extreme book reviews.

The second stream predicts the helpfulness of a product review by the characteristics of its contributing reviewer. For example, Li et al. (2011) adopted the source of product review to predict the product review source credibility, a factor of helpfulness of product review. Prior studies revealed that identity-relevant information about reviewers shapes readers' judgment of product reviews (Forman et al. 2008; Connors et al. 2011). In addition, the subjective assessment of the helpfulness of a review could be influenced by the average helpfulness of the reviews pertaining to a focal reviewer. Evidently, these prior studies focus on predicting the helpfulness of a product review rather than on predicting the average helpfulness of a reviewer. Practically, it is also important to estimate the helpfulness of a reviewer, especially from the perspective of retaining quality reviewers by e-stores or third-party product review websites. Hence, this study will concentrate on predicting the average helpfulness of a reviewer's product reviews.

Most prior studies investigated the characteristics of a reviewer from the perspective of reviewers' behavior. For example, the study by Riggs and Wilensky (2001) explored how technology has been exploited to enable alternative models of dissemination of scholarly information (i.e., journals that are written for a specialized audience by experts in a subject area). The main focus on their research is "*to use collaborative filtering algorithm allowing automatically to rate reviewers, and also incorporates the quality of the reviewers into the metric of merit for that pape*r" (Riggs and Wilensky 2001). They consider three factors (i.e. *number of items reviewed, number of reviews of an item, and time of review*) to predict a reviewer's reputation. However, the average helpfulness of a reviewer is a subjective evaluation by other readers. Thus, we argue that it is essential and comprehensive to include factors related to reviewers' trustworthiness perceived by readers when predicting their average helpfulness. Therefore, this study will predict the helpfulness of product reviewers by their review behavior and associated trust network.

3 OUR HELPFULNESS PREDICTION TECHNIQUE

In this section, we describe our proposed helpfulness prediction technique. Specifically, we first detail the variables used for predicting the helpfulness of reviewers. Subsequently, we provide a brief overview of the data mining techniques that we employed for the target prediction task.

3.1 Variables for Helpfulness Prediction

In an online community, two important concepts relate to the helpfulness of a reviewer. The first one the reviewer's expertise level on the issues he/she comments on or the knowledge he or she shares in the community. Readers judge the quality of reviews contributed by a reviewer in the community to assess the expertise level of the reviewer. Thus readers will form their perceptions on the credibility of the reviewer. This credibility perception will become the baseline to assess the helpfulness of new reviews contributed by the same reviewer.

Prior research suggests that the expertise level of a reviewer is likely to be reflected in his or her review behavior (Ku et al. 2012; Riggs and Wilensky 2001). Thus, in this study, we consider four variables related to a reviewer's review behavior to predict the helpfulness of this focal reviewer with respect to a target product category.

- Number of reviews in the target category

This variable denotes the number of reviews written by the focal reviewer in the target category. A greater number of reviews written by the focal reviewer in the target category suggest that he or she is an active reviewer of the target category. Correspondingly, when the number of reviews contributed by a reviewer in the target product category is higher, the helpfulness of the focal reviewer is expected to be greater (Ku et al. 2012; Lu et al. 2010).

- Degree of review focus on the target category

This variable is important because it specifically measures the percentage of reviews by the focal reviewer in the target product category. That is, the degree of review focus on the target category = $\frac{\text{number of reviews in the target category}}{\text{total number of reviews}}$. Similar to the number of reviews in the target category, a reviewer with a higher degree of review focus on the target category implies that this reviewer is a more active reviewer of the target category than of other product categories (Huang et al. 2010).

- Average product rating on the target category

The average product rating on the target category is the average of individual ratings provided by the focal reviewer on products in the target category. It is expected that a more helpful reviewer of the target category should be a more critical reviewer because of his or her higher expertise/knowledge level on the target product category. Thus, his or her average product rating would be lower than that of a less helpful reviewer (Chen and Xie 2008).

- Variance of product ratings on the target category

This variable refers to the variance of individual ratings provided by the focal reviewer on products in the target category. As with the average product rating, a more helpful reviewer of the target category should be a more critical reviewer and, thus, the variance of his or her ratings would be larger than that of a less helpful reviewer.

The second important concept related to the helpfulness of reviewers in an online community is trust (Ku et al. 2012). Trust means the establishment of the relationship between two parties. This relationship is built on the basis of the belief and confidence without expecting any action in return. In other words, a seller is trustworthy if what the seller has done and how he or she has done it in the past are perceived by buyers as positive. In turn, buyers may develop trust relationships with the seller. That is, a reliable seller (or a helpful member in an online community) is generally a trustworthy one.

Web trust networks have been adopted by many online opinion-sharing communities, including Epinions (http://www.epinions.com) and FilmTrust (http://trust.mindswap.org/ FilmTrust/). The members of these communities can express their trust beliefs toward other members directly by setting trust relationships. These trust relationships reveal reviewers' trustworthiness perceived by readers. In this study, we identify and extract four variables from the web trust network pertaining to a focal reviewer to predict the helpfulness of this reviewer. Before describing the four trust-related variables, we first define the trustor and the trustee involved in a trust relationship. According to Mayer et al. (1995), if member *A* trusts member *B* (or *B* is trusted by *A*), *A* is the trustor of *B* and *B* is the trustee of *A*.

- Number of trustors

The number of trustors of the focal reviewer is the number of members trusting the focal reviewer. In this study, we also refer the number of trustors of a member as his or her trust intensity. Conceivably, if the trust intensity of the focal reviewer is high, he or she should be a more trustworthy reviewer (Ku et al. 2012). In this case, the helpfulness of the focal reviewer also should be higher.

- Number of trustees

The number of trustees of a focal member denotes the number of members trusted by the focal reviewer.

- Average trust intensity of trustors

This variable measures the average trustworthiness of all trustors of the focal reviewer (i.e., average trust intensity of the members who trust the focal reviewer). Prior research (Ku et al. 2007) empirically shows that people who are trusted by more trustworthy members tend to be more trustworthy. As a result, if the focal member’s average trust intensity of trustors is high, the helpfulness of his or her reviews should also be high.

- Average trust intensity of trustees

This variable measures the average trustworthiness of all trustees of the focal reviewer (i.e., average trust intensity of the members trusted by the focal reviewer). Prior research also suggests that people who trust more trustworthy members tend to be more trustworthy, whereas those who trust less trustworthy members are less trustworthy (Ku et al. 2007). Accordingly, if the average trust intensity of the trustees who are trusted by the focal member is high, the helpfulness of his or her reviews should also be high.

3.2 Investigated Data Mining Techniques

This study adopts the data mining approach to predict the helpfulness of online product reviewers. The use of a data mining (specifically, supervised learning) technique for prediction purposes essentially constructs an automated prediction model that captures important relationships among a set of input (independent) variables and a dependent variable, which has a numeric (continuous) value in this study. Particularly, a selected supervised learning technique uses a set of training instances (each with known values for independent and dependent variables) to construct a prediction model. The resulting model is then used to predict the dependent variable of a new (unseen) instance on the basis of its input variable values. Different supervised learning algorithms have been investigated and empirically tested in various application contexts, including linear regression, artificial neural network, model-tree–based regression, and SVM for regression.

In this study, we investigate two supervised learning algorithms for the target helpfulness prediction task. Particularly, we employ and examine M5 (a model-tree-based regression technique) and SVM for regression in this study. M5 is a model-tree–based regression technique whose prediction analysis combines a conventional decision tree and linear regression functions (Quinlan, 1992). A model tree resembles a decision tree structurally but has linear regression functions at its leaf nodes rather than the discrete (output) classes common to conventional decision-tree induction algorithms. On the other hand, Support Vector Machine (SVM) is a novel supervised learning machine, first introduced by Vapnik (1995). It is based on the Structural Risk Minimization principle from the computational learning theory. SVM uses a set of training instances to construct a regression function for classification or prediction purposes (Vapnik, 1995; Smola and Schölkopf, 2004).

4 EMPIRICAL EVALUATION

In this section, we report on our empirical evaluation of our proposed helpfulness prediction technique. In the following subsections, we detail our data collection and evaluation design. Subsequently, we discuss several important empirical evaluation results.

4.1 Data Collection

Our datasets were retrieved from epinions.com, a popular third-party product review website. Because this study focuses on predicting the helpfulness of a focal reviewer in a specific product category, we need to define who the reviewers of the specific product category are. In this study, we consider that a member is a reviewer of a target product category if he or she has ever contributed at least three reviews on that category. Furthermore, we need to determine the helpfulness scores (i.e., the value for the dependent variable) of reviewers in each given product category. On the epinions.com website, members can express their evaluations on the quality (specifically, helpfulness) of product reviews contributed by other members. As a result, we simply take the average of all user-evaluations on the reviews provided by a reviewer in the given product category as the helpfulness score of that reviewer in the target category. Specifically, in the epinions.com website, a user-evaluation on the helpfulness of a product review can be “Very Helpful,” “Helpful,” “Somewhat Helpful,” and “Not Helpful.” We convert this ordinal scale of helpfulness into the numerical scale as follows: “Very Helpful” = 3, “Helpful” = 2, “Somewhat Helpful” = 1 and “Not Helpful” = -2. We then take the macro-averaged helpfulness score given by all members on all product reviews made by the target member in the product category under discussion. The macro-averaged helpfulness score means that we take the average helpfulness score given to each product review made by the focal reviewer in the target product category and then take the average across all product reviews made by this reviewer in the target product category. The resultant average is considered as the helpfulness score of the focal reviewer in the target product category.

In this study, we focus on two product categories, including Car and Computer. Table 1 shows the statistical summary of our collected data sets for the two product categories.

Product Category	Number of Reviewers	Helpfulness Score			
		Minimum	Maximum	Mean	StdDev
Car	651	0.444	3	2.434	0.476
Computer	734	0	3	2.490	0.445

Table 1. Statistics Summary of Our Data Sets

4.2 Evaluation Design

In this study, we used the Weka open-source machine learning software (accessible at http://www.cs.waikato.ac.nz/ml/weka/) to construct the helpfulness prediction systems based on M5 and SVM for Regression. A tenfold cross-validation strategy is used to estimate the effectiveness of our proposed helpfulness prediction technique. That is, we randomly divide each data set into ten mutually exclusive subsets of equal size. We then designate each subset as the testing data set while the others serve as the training data set. We then assess the effectiveness of the proposed technique for this tenfold cross-validation process by averaging the effectiveness of these ten individual trials. In this study, we measure the effectiveness of our proposed helpfulness prediction technique on the basis of correlation coefficient and mean absolute error (MAE) between the actual helpfulness scores and the predicted ones. Furthermore, to minimize potential biases that may result from the randomized folding process, we perform this tenfold cross-validation process ten times and estimate the overall effectiveness by averaging the performance estimates obtained from the 10 cross-validation processes.

4.3 Results and Discussions

As we show in Table 2, the correlation coefficients using M5 for helpfulness prediction are 0.365 and 0.420 for the two product categories, respectively (i.e. Car and Computer). The correlation coefficients attained by SVM Regression are 0.365 for Car category and 0.324 for Computer category, respectively. Our findings reveal M5 achieves higher correlation coefficients, outperforming SVM Regression, in computer category.

Table 3 shows the comparative evaluation results of M5 and SVM for Regression, using MAE as the evaluation criterion. As Table 3 illustrates, the MAE attained by M5 is 0.347 for the Car category and 0.308 for the Computer category. On the other hand, the MAE of SVM Regression is 0.340 for the Car category and 0.324 for the Computer category. M5 outperforms SVM in Computer category, while SVM for Regression attains lower MAE than M5 in the Car category. Overall, taking both evaluation criteria together, M5 appears to be more effective than SVM Regression does for helpfulness prediction in computer category.

Product Category	M5	SVM Regression
Car	0.365	0.365
Computer	0.420	0.324

Table 2. Comparative Evaluation Results (Correlation Coefficient)

Product Category	M5	SVM Regression
Car	0.347	0.340
Computer	0.308	0.324

Table 3. Comparative Evaluation Results (Using MAE Metric)

5 CONCLUSION

This study aims at developing a data mining approach to predict the helpfulness of online product reviewers. Our prediction model not only can facilitate consumers to judge whether to believe or disbelieve reviews written by different reviewers, but also can help e-stores or third-party product review websites to target and retain quality reviewers. In this study, we identify eight independent variables related to a focal reviewer's review behavior and his or her associated trust network. Both M5 and SVM Regression are applied as our underlying learning algorithms to predict the helpfulness score of the focal reviewer. Our empirical evaluation results on the basis of two product categories suggest that our proposed technique can predict the helpfulness scores for online product reviewers.

Future research directions along the line of this research include the following. First, as mentioned, although we recognized that some user-evaluations may not be reliable, we have difficulties to identify and remove unreliable user-evaluations. As a result, we took the macro-average of all user-evaluations on the reviews provided by a reviewer in a given product category as the "true' helpfulness score of the focal reviewer in the target category. This naïve approach may introduce noises into our datasets. Future research should be devoted to controlling noises in datasets for empirical evaluation purposes. Second, this study did not analyze the content of reviews. Future research should include additional variables associated with review content to enhance the precision of this proposed model.

References

Cao, Q., Duan, W., and Gan, Q. (2011). Exploring determinants of voting for the ‘helpfulness’ of online user reviews: a text mining approach. *Decision Support Systems*, 50 (2), 511-521.

Chen, Y. and Xie, J. (2008). Online consumer review: word-of-mouth as a new element of marketing communication mix. *Management Science*, 54 (3), 477–497.

Clemons, E., Gao, G., and Hitt, L. (2006). When online reviews meet hyperdifferentiation: A study of the craft beer industry. *Journal of Management Information Systems*, 23 (2), 149-171.

Connors, L., Mudambi, S.M., and Schuff, D. (2011). Is it the review or the reviewer? A multi-method approach to determine the antecedents of online review helpfulness. In *Proceedings of the 44th Hawaii International Conference on System Sciences*.

Forman, C., Ghose, A., and Wiesenfeld, B. (2008). Examining the relationship between reviews and sales: The role of reviewer identity disclosure in electronic markets. *Information Systems Research*, 19 (3), 291-313.

Guha, R., Kumar, R., Raghavan, P., and Tomkins, A. (2004). Propagation of trust and distrust, In *Proceedings of the 13th International World Wide Web Conference*, pp. 403-412.

Huang, S., Shen, D., Feng, W., Baudin, C., and Zhang, Y. (2010). Promote product reviews of high quality on e-commerce site. *Pacific Asia Journal of the Association for Information Systems*, 2 (3), 51–71.

Ku, Y.C., Hsiao, H.W., and Wei, C.P. (2007). Analysis of web trust networks. In *Proceedings of the Second International Conference on Innovative Computing, Information and Control (ICICIC 2007)*, Kumamoto, Japan.

Ku, Y.C., Wei, C.P., and Hsiao, H.W. (2012). To whom should I listen? Finding reputable reviewers in opinion-sharing communities. Decision Support Systems, doi:10.1016/j.dss.2012.03.003.

Kumar, N. and Benbasat, I. (2006). The influence of recommendations on consumer reviews on evaluations of websites. *Information Systems Research*, 17 (4), 425-439.

Li, M.X., Huang, L., Tan, C.H., and Wei, K.K. (2011). Assessing the helpfulness of online product review: A progressive experimental approach. In *Proceedings of the PACIS 2011*.

Lu, Y., Tsaparas, P., Ntoulas, A. and Polanyi, L. (2010). Exploiting social context for review quality prediction. In *Proceedings of the 19th International Conference on World Wide Web.*

Mayer, R. C., Davis, J. H., and Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of Management Review*, 20 (3), 709-734.

Mudambi, S.M. and Schuff, D. (2010). What makes a helpful online review? A study of customer reviews on amazon.com. *MIS Quarterly,* 34 (1), 185-200.

Park, D.-H., Lee, J., and Han, I. (2007). The effect of online consumer reviews on consumer purchasing intention: The moderating role of involvement. *International Journal of Electronic Commerce,* 11 (4), 125-148.

Quinlan, J.R. (1992). Learning with continuous classes. In *Proceedings of 5th Australian Joint Conference on Artificial Intelligence*, World Scientific, Singapore, pp. 343-348.

Riggs, T. and Wilensky R. (2001). An algorithm for automated rating of reviewers. In *Proceedings of the First ACM/IEEE-CS Joint Conference on Digital libraries (JCDL)*, pp. 381-387.

Smola, A. J. and Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and Computing*, 14 (3), 199-222.

Vapnik, V. (1995). *The Nature of Statistical Learning Theory*. Springer-Verlag, New York, N.Y.