

A Helpfulness Modeling Framework for Electronic Word-of-Mouth on Consumer-Opinion Platforms

Richong Zhang and Thomas Tran
School of Information Technology and Engineering
University of Ottawa

Electronic Word-of-Mouth (eWOM) is growing exponentially with the rapid development of Electronic Commerce. As a result, consumers are increasingly crowded by a huge amount of eWOM contents and therefore there is a need to automatically recommend eWOM contents that are helpful to them. Existing helpfulness assessment approaches that deterministically estimate the helpfulness of eWOM contents lack a generative formulation and are limited to the training set that has been voted by many readers. This paper presents a rigorous probabilistic framework for inferring the “helpfulness” of eWOM contents which can build a “helpfulness” model from a low number of votes on eWOM contents. Furthermore, we introduce a measurement, “helpfulness” bias, as the benchmark for the “helpfulness” of eWOM documents. We also propose a model that exploits the graphical model and Expectation-Maximization algorithm, under this probabilistic framework, to demonstrate the versatility of our framework. Our algorithm is compared experimentally to other existing helpfulness discovering algorithms and the experimental results show that our framework can effectively model the helpfulness of eWOM contents better than other approaches, and therefore, indicate the capability of our framework to recommend helpful eWOMs to the potential consumers.

Categories and Subject Descriptors: H.4.0 [Information Systems Applications]: General; I.2.1 [Artificial Intelligence]: Applications and Expert Systems

General Terms: Algorithm, Measurement

Additional Key Words and Phrases: Ranking, Recommender Systems, Online Product Reviews

1. INTRODUCTION

The word-of-mouth (WOM) has long been shown to be an important source for consumers to make purchasing decisions. Due to the rapid development of internet business, the number of electronic commerce web sites has grown exponentially in recent years. Electronic word-of-mouth has emerged as a new source for consumers to refer to while studies have shown that eWOM affects consumer behaviors [Hu et al. 2006] and is helpful for decision-making on purchases [Park et al. 2007].

As mentioned in [Robert M. Schindler 2005], consumers seek for online eWOM as

Author’s address: R. Zhang and T. Tran, 800 King Edward Avenue, Ottawa, Ontario K1N 6N5, Canada.

Email address: {rzhan025,ttran}@site.uottawa.ca

Permission to make digital/hard copy of all or part of this material without fee for personal or classroom use provided that the copies are not made or distributed for profit or commercial advantage, the ACM copyright/server notice, the title of the publication, and its date appear, and notice is given that copying is by permission of the ACM, Inc. To copy otherwise, to republish, to post on servers, or to redistribute to lists requires prior specific permission and/or a fee.

© 20YY ACM 0000-0000/20YY/0000-0001 \$5.00

Information Input to specify purchase decisions and eWOM could help consumer decide whether a product should be purchased. According to a survey with the responses of 1,000 online shoppers [Palmer 2009], 47 percent of consumers surveyed look to onsite consumer reviews when making a purchasing decision. In another survey [Nielsen-Online 2008], 71 percent consumers agree that consumer reviews make them more comfortable with buying the right product. Furthermore, 81 percent online holiday shoppers read online customer reviews according to [Nielsen-Online 2008].

One of the characteristic features of an electronic shopping environment is that there exists a vast amount of information available when making purchase decisions [Parra and Ruiz 2009]. The User-Generated Contents (UGC) is more and more prevalent in most of the electronic commerce web sites, like Amazon.com, that provide not only products and services, but also online reviews or eWOM messages, composed by previous purchasers. By taking the advantages of the increasing availability of rich information, consumers enjoy a greater number of experiences than ever before. Still, the quality of the online reviews varies significantly. Especially with the explosive growth of the number of web blogs, forum posts, and online reviews, potential consumers have to spend an immense amount of time retrieving reviews that assist them in better understanding products.

Considering the large amount of available online product reviews, too much information may overburden any consumer and due to the unavailability of the helpfulness of eWOM messages, consumers need to search across these contents for ones that may be helpful in their purchasing decisions. In [Duan et al. 2008], authors pointed out that consumers need to find out the trustable experiences from all kinds of eWOM contents. Currently, many opinion sharing platforms provide a function to let readers vote for helpful or not about a review or opinion post. Potential consumers can make use of this information to estimate the helpfulness of the review and decide to read it or not. As the vote collection process takes time and newly posted reviews always get fewer votes, a system that can easily discover helpful online reviews is much needed to be hired to reduce the time of acquiring helpful reviews.

A number of helpfulness assessment approaches have been developed to model the helpfulness value of eWOM messages [Kim et al. 2006; Liu et al. 2008; Zhang and Varadarajan 2006]. These helpfulness assessment approaches utilize the positive vote fraction as the helpfulness benchmark and focus on applying different machine learning tools to learn a helpfulness function which is a mapping $Score : \mathcal{D} \mapsto \mathbb{R}$ where \mathcal{D} is the space of all review documents and \mathbb{R} represents the set of real numbers. The major limitation of the conventional approaches is the utilization of the positive vote fraction based benchmarking methodologies. Particularly when only a limited number of votes are available, the positive vote fraction is unable to significantly represent the true helpfulness of review documents.

Aware of such a limitation, most existing approaches solely attempt to analyze the eWOM messages for which a relatively large number of votes are available. However, even when a large number of voters' opinions are available for each eWOM messages, the same helpfulness value of positive vote fraction can come from different sizes of voter population and this fraction fails to capture the confidence of

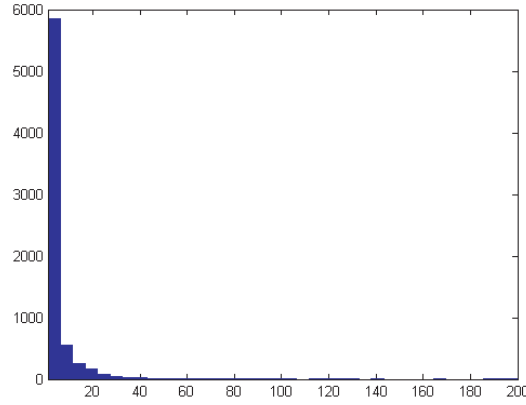


Fig. 1. Number of Votes in the Review Document Data Set

the helpfulness estimates. Furthermore, in reality, most of online consumer reviews receive very few votes (see Fig.1 for the histogram of the number of votes on a set of review documents taken from Amazon.com). Since most online consumer reviews are only associated with a small number of voters, the applicability of the existing approaches is significantly limited to a small part of available information.

In this paper, we present a more robust formulation of review helpfulness with precise mathematical meaning, and a more principled probabilistic helpfulness assessment framework for inferring the helpfulness distribution of eWOM message using the distribution of helpfulness conditioned on the feature of the document to characterize the dependency of the helpfulness and the review. Furthermore, we introduce a new quantitative helpfulness benchmark to capture the uncertainty of the helpfulness estimates and a helpfulness bias formulation based on which our model can rank user-generated reviews by the obtained helpfulness distribution.

Moreover, there are currently no standard evaluation metrics to judge the performance of review helpfulness assessment algorithms. In this paper, we also discuss the evaluation aspects of the review helpfulness modeling framework and propose a uniformed evaluation measurement for assessing the success of helpfulness prediction models. Experiments on review datasets of two product categories from Amazon.com suggest that our algorithm in fact outperforms the previously reported prediction algorithms and show that the algorithm can effectively predict the helpfulness distribution of eWOM and recommended the most helpful contents to potential consumers.

The remainder of this paper is organized as follows. Section 2 introduces our framework to estimate the helpfulness distribution of a review and how to evaluate the performance of a helpfulness assessing model. Section 3 delivers a graphical model based approach under our framework. Section 4 presents experimental evaluation of our framework. Finally, Section 5 will discuss and conclude the paper.

2. RELATED WORK

An increasing number of literatures illustrate the effect of online product reviews and eWOM on the product sales and consumer's behavior. Park et al. [2007] find that the quality of reviews has a positive effect on product sales as consumers purchase intentions increases with the quantity of product reviews. Hu et al. [2008] mentioned that consumers not only considered review ratings, but also the contextual information like reviewer's reputation. They also find that the impact of online reviews on sales diminishes over time.

Park and Kim [2008] investigated the relationship between different types of reviews and consumers. They find that consumer concerns vary at each stage of the product life cycle, and suggest marketers develop different strategies for different types of consumers. Lee et al. [2008] examined the effect of the quality of negative online reviews on product attitude and discover that high-involvement consumers consider the quality of negative reviews and low-involvement consumers tend to conform to other reviewer attitudes regardless of the quality.

In [Vermeulen and Seegers 2009], authors studied the effect of online hotel reviews on consumer consideration and conclude that positive reviews have a positive impact on consumer behavior. In [Park and Lee 2008], authors analyzed the two roles of the online consumer reviews (an informant and a recommender). When information overload occurs, it suggests that low-involvement consumers mainly focus on the perceived popularity and high-involvement consumers focus on the product information of reviews. Based on the findings of these researches, it can be confirmed that eWOM plays an important role in the purchase decision-making process of consumers as consumers are willing to consider another persons' experiences before they decide to make a purchase.

Some works have been done in the area of review mining and summarizing. Zhuang et al. [2006] mined and summarized the movie reviews based on a multi-knowledge approach that included WordNet, statistical analysis and movie knowledge. Hu and Liu [2004] summarized product reviews by mining opinion features. Mayzlin and Chevalier [2003] examined the effect of consumer reviews on relative book sales on Amazon and Barnes and Noble web sites. Hatzivassiloglou and McKeeown [1997] proposed a method to predict the semantic orientation of adjectives by a supervised learning algorithm. Turney [2002] presented an unsupervised learning algorithm to classify reviews as recommended or not recommended by analyzing the semantic orientation based on mutual information. In [Yu and Hatzivassiloglou 2003], authors proposed a classification approach to separate sentences as positive or negative. In [Pang et al. 2002], authors classified movie reviews as positive or negative by several machine learning methods that were Naive Bayes, Maximum Entropy and Support Vector Machines and they also used different features like unigram, bigram, position and a combination of these features. The results showed that unigram presence feature was the most effective and the SVM performed the best for sentiment classification. These studies focus on sentiment classification and opinion mining for eWOM, however, only a few researches consider the usefulness of eWOM contents.

Kim et al. [2006] developed an SVM-based method to assess review helpfulness, where review lengths, unigrams and product ratings are taken as the discriminating

features. Weimer et al. [2007] developed an algorithm to assess the quality of posts in web forums using a variety of features including surface, lexical and syntactic features, as well as some forum-specific features and certain similarity features. In a follow-up, Weimer [2007] extended the method into three data sets and found that the SVM classification performed the best.

Liu et al. [2008] presented a nonlinear regression model for the helpfulness prediction. Three groups of factors that might affect the value of helpfulness were analyzed and the model was build upon on these three groups of factors. The results of applying their model showed that the performance was better than the SVM regression model. Zhang and Varadarajan [2006] incorporated a diverse set of features to build a regression model to predict the utility of online product reviews. As we discussed in Section 1, these modeling approaches lack principles and can only learn helpfulness from a small part of available eWOMs. Our study focuses on proposing a generalized framework for analyzing the helpfulness of eWOM and helping consumers find the most helpful ones more efficiently.

3. PROBABILISTIC FRAMEWORK OF REVIEW HELPFULNESS

3.1 Probabilistic Framework

In this section, we introduce a probabilistic framework for the inference of review helpfulness. Throughout this paper, a random variable will be denoted by a bold-font capitalized letter, for example \mathbf{V} , \mathbf{D} ; and any value the random variable may assume will be denoted by its corresponding lower-cased letter, for example v , d . The distribution (either probability mass function or probability density function) of a random variable, say, \mathbf{X} , will be denoted by $p_{\mathbf{X}}$. Following the above mentioned notational convention, when $p_{\mathbf{X}}$ is treated as a function, we also write as $p_{\mathbf{X}}(x)$ when necessary, namely using the lower-cased letter “ x ” as the argument of the function.

We will often encounter a collection of random variables, say, $\{\mathbf{X}_k : k \in K\}$ for some index set K . In this case, we may write the set of random variables collectively as \mathbf{X}_K , and its corresponding (vector) value as x_K .

3.1.1 Probabilistic formulation of helpfulness. We will use \mathcal{D} to denote the space of all eWOMs and use \mathcal{W} to denote the population of readers who may vote on the documents in \mathcal{D} . We argue that the “helpfulness” of any document $d \in \mathcal{D}$ is not only a property of document d itself, but it, in fact, also depends on \mathcal{W} . Specifically, a given document d may be more helpful with respect to one population \mathcal{W} than it is with respect to another population \mathcal{W}' .

To rigorously define “helpfulness”, we characterize \mathcal{W} as a pair $(\mathcal{R}, \mathcal{P})$, where \mathcal{R} is a family of functions mapping \mathcal{D} to $\{0, 1\}$ and \mathcal{P} is a probability measure on \mathcal{R} . Here, each function $r \in \mathcal{R}$ specifies a rule of voting, whereby $r(d) = 1$ indicates that the document d will be voted positively (ie, as HELPFUL) under rule r . In this setting, we define the helpfulness of a document d , denoted by $\alpha(d)$, as the probability that document d will be voted positively by a random reader in \mathcal{W} , namely,

$$\alpha(d) := \Pr[\mathbf{R}(d) = 1],$$

where \mathbf{R} is drawn at random from probability space $(\mathcal{R}, \mathcal{P})$. Equivalently, the helpfulness of document d can be characterized by the conditional probability dis-

tribution $p_{\mathbf{V}|\mathbf{D}}(v|d)$, where \mathbf{D} is a random document from \mathcal{D} and \mathbf{V} is the $\{0, 1\}$ -valued opinion of a random voter from $(\mathcal{R}, \mathcal{P})$. We are particularly concerned with the probability measure on \mathcal{D} from which the random document \mathbf{D} is chosen since \mathbf{D} is always conditioned upon throughout the paper. It is then clear that $p_{\mathbf{V}|\mathbf{D}}(v = 1|d) = \alpha(d)$ and $p_{\mathbf{V}|\mathbf{D}}(v = 0|d) = 1 - \alpha(d)$.

3.1.2 The helpfulness inference problem. Under this formulation of helpfulness, the objective of inferring the helpfulness is to learn $\alpha(d)$, or $p_{\mathbf{V}|\mathbf{D}}(v|d)$ for every $d \in \mathcal{D}$, based on a sample of documents drawn from \mathcal{D} and a collection of votes on these documents.

Let \mathbf{D}_I be a random set of documents drawn from \mathcal{D} , where each element in I is the index of a document in \mathbf{D}_I . For each $i \in I$, let $\Gamma(i) \subset \mathcal{W}$ denote the set of readers who have voted on document i . Denote by $\Gamma := \cup_{i \in I} \Gamma(i)$ the set of all voters that have voted on at least one document in I .

In the helpfulness inference problem, we wish to determine $p_{\mathbf{V}|\mathbf{D}}(v|d)$ when \mathbf{D}_I is given as some d_I and \mathbf{V}_Γ is given as some v_Γ .

3.1.3 Probabilistic formulation of helpfulness inference. We now present a probabilistic framework of this problem. Essential to this framework is the creation of a model that describes the relationship between documents, their helpfulness and the voter opinion. Here, a *model* is a sensible family $\Theta_{\mathbf{V}|\mathbf{D}}$ of conditional distributions of \mathbf{V} given \mathbf{D} . The objective of helpfulness inference is to select a distribution $p_{\mathbf{V}|\mathbf{D}}$ from the family $\Theta_{\mathbf{V}|\mathbf{D}}$ such that $p_{\mathbf{V}_\Gamma|\mathbf{D}_I}(v_\Gamma|d_I)$ is maximized. That is, the problem translates to determine:

$$\begin{aligned} p_{\mathbf{V}|\mathbf{D}}^* &= \operatorname{argmax}_{p_{\mathbf{V}|\mathbf{D}} \in \Theta_{\mathbf{V}|\mathbf{D}}} \log p_{\mathbf{V}|\mathbf{D}}(v_\Gamma|d_I) \\ &= \operatorname{argmax}_{p_{\mathbf{V}|\mathbf{D}} \in \Theta_{\mathbf{V}|\mathbf{D}}} \log \prod_{i \in I} p_{\mathbf{V}|\mathbf{D}}(v_{\Gamma(i)}|d_i), \end{aligned} \quad (1)$$

where the latter equality is due to the fact that given document i , voters opinion on the document is independent on their opinions on any other documents.

Furthermore, we assume voters in each $\Gamma(i)$ are chosen independently, giving rise to:

$$p_{\mathbf{V}|\mathbf{D}}(v_{\Gamma(i)}|d_i) = \prod_{j \in \Gamma(i)} p_{\mathbf{V}|\mathbf{D}}(v_j|d_i) \quad (2)$$

Applying our definition of helpfulness, when a document D is drawn from \mathcal{D} (under any probability measure) and a random voter V is chosen from \mathcal{W} under \mathcal{P} , it is easy to see that random variables D , H and V form Markov chain $D - H - V$, where

$$H := \alpha(D). \quad (3)$$

Specifically,

$$p_{\mathbf{V}|\mathbf{D}, \mathbf{H}}(v|d, h) = p_{\mathbf{V}|\mathbf{H}}(v|h) = h^{[v=1]}(1-h)^{[v=0]}, \quad (4)$$

where $[P]$ (known as Iverson's convention) for any Boolean proposition P evaluates to 1 if P is true, and to 0 otherwise.

Putting together Eq.(1), Eq.(2) and Eq.(4), we have:

$$\begin{aligned} p_{\mathbf{V}|\mathbf{D}}^* &= \operatorname{argmax}_{p_{\mathbf{V}|\mathbf{D}} \in \Theta_{\mathbf{V}|\mathbf{D}}} \log \prod_{i \in I} \prod_{j \in \Gamma(i)} \int p_{\mathbf{V}, \mathbf{H}|\mathbf{D}}(v_j, h_i | d_i) dh_i \\ &= \operatorname{argmax}_{p_{\mathbf{V}|\mathbf{D}} \in \Theta_{\mathbf{V}|\mathbf{D}}} \left[\log \prod_{i \in I} \prod_{j \in \Gamma(i)} \int h_i^{[v_j=1]} (1 - h_i)^{[v_j=0]} p_{\mathbf{H}|\mathbf{D}}(h_i | d_i) dh_i \right], \quad (5) \end{aligned}$$

where the final equality follows from:

$$\begin{aligned} p_{\mathbf{V}|\mathbf{D}}(v|d) &= \int p_{\mathbf{V}, \mathbf{H}|\mathbf{D}}(v, h | d) dh \\ &= \int p_{\mathbf{V}|\mathbf{H}}(v|h) p_{\mathbf{H}|\mathbf{D}}(h|d) dh \\ &= \int h^{[v=1]} (1 - h)^{[v=0]} p_{\mathbf{H}|\mathbf{D}}(h|d) dh. \quad (6) \end{aligned}$$

We note Eq.(6) also implies that determining $p_{\mathbf{V}|\mathbf{D}}^*(v|d)$ boils down to determine $p_{\mathbf{H}|\mathbf{D}}^*(h|d)$ from a family $\Theta_{\mathbf{H}|\mathbf{D}}$ of conditional distribution of \mathbf{H} on \mathbf{D} which are sensible and may induce the family $\Theta_{\mathbf{V}|\mathbf{D}}$ under Eq.(6). Then we reformulate the helpfulness inference problem as finding:

$$p_{\mathbf{V}|\mathbf{D}}^* = \operatorname{argmax}_{p_{\mathbf{H}|\mathbf{D}} \in \Theta_{\mathbf{H}|\mathbf{D}}} \left[\log \prod_{i \in I} \prod_{j \in \Gamma(i)} \int h_i^{[v_j=1]} (1 - h_i)^{[v_j=0]} p_{\mathbf{H}|\mathbf{D}}(h_i | d_i) dh_i \right]. \quad (7)$$

for some properly defined family $\Theta_{\mathbf{H}|\mathbf{D}}$.

It is worth noting that H depends on D functionally (as given in (3)). The nature of the problem is that neither \mathcal{V} nor \mathcal{P} is known. This lack of knowledge necessarily implies that an appropriate dependency model of H on D is to assume that H depends on D probabilistically. This need is further amplified when a numerous amount of votes are not available for some documents.

3.1.4 Helpfulness Ranking. Once helpfulness distribution $p_{\mathbf{H}|\mathbf{D}}(h_i | d_i)$ in the above mentioned framework is determined, a question which naturally arises is how to rank document according to $p_{\mathbf{H}|\mathbf{D}}(h_i | d_i)$. We define the probability of a review being helpful is greater than 0.5 as a ranking metric:

$$r_i := p_{\mathbf{H}|\mathbf{D}}(h_i > 0.5 | d_i) = \int_{0.5}^1 p_{\mathbf{H}|\mathbf{D}}(h_i | d_i) dh_i. \quad (8)$$

We label this metric as *helpfulness bias*, which describes the helpfulness distribution bias towards 0 or 1. Intuitively, this conditional probability, r_i , can be chosen as a quantitative metric to measure the helpfulness of eWOM.

Suppose the “Oracle” is accessible to the learning machine, the learning machine should take $p_{\mathbf{H}|\mathbf{V}}(h_i | v_{\Gamma(i)})$ as the “best guess” for the helpfulness distribution; yet, an assumption we may make here is that without access to the “Oracle”, $p_{\mathbf{H}}(h_i)$ is uniformly distributed. The $p_{\mathbf{H}|\mathbf{V}}(h_i | v_{\Gamma(i)})$ is the estimated density of h_i under

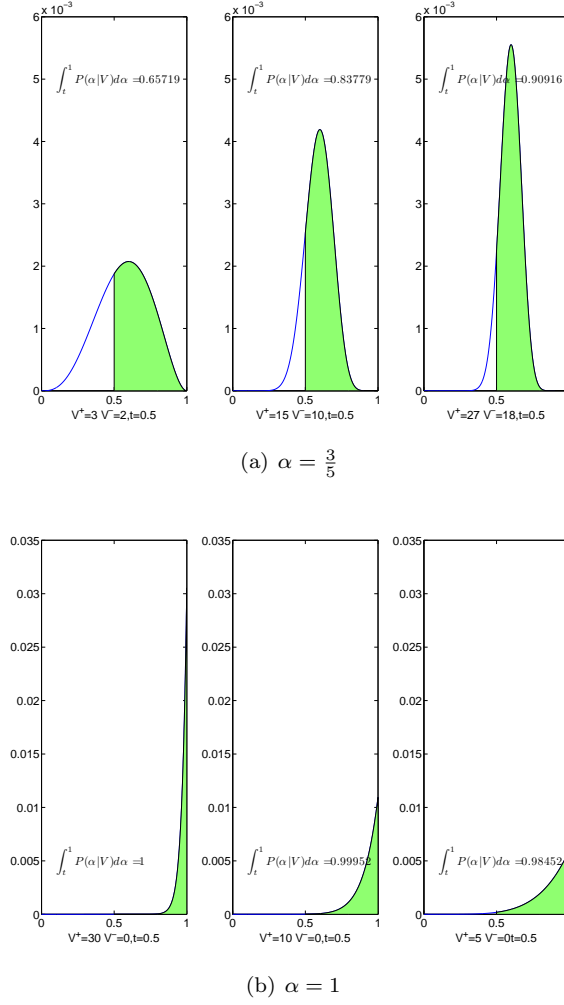


Fig. 2. The probability density function of the review helpfulness, V_i^+ and V_i^- denote the number of positive and negative votes for each eWOM in the data set. (a) Positive vote fraction is $\frac{3}{5}$. (b) Positive vote fraction is 1.

learning with the “Oracle” and it is clear that:

$$\begin{aligned}
 p_{\mathbf{H}|\mathbf{V}}(h_i|v_{\Gamma(i)}) &= \frac{p_{\mathbf{V}|\mathbf{H}}(v_{\Gamma(i)}|h_i)}{p_{\mathbf{V}}(v_{\Gamma(i)})} \\
 &= \frac{p_{\mathbf{V}|\mathbf{H}}(v_{\Gamma(i)}|h_i)}{\int p_{\mathbf{V}|\mathbf{H}}(v_{\Gamma(i)}|h_i)dh_i}.
 \end{aligned} \tag{9}$$

Fig. 2 describes the conditional probability density function $p_{\mathbf{H}|\mathbf{V}}(h_i|v_{\Gamma(i)})$ given by six different $v_{\Gamma(i)}$ examples. It can be observed that with the same helpfulness value, the degree of certainty (the width of the distribution) is completely different.

When a small number of voters are given, the distribution of helpfulness is highly uncertain. When large sufficient voters voted on a document, we expect the helpfulness distribution is narrow and highly concentrated on the value of positive vote fraction.

The helpfulness bias of the “Oracle”, with the prior knowledge of voter opinion, can be formulated as the integral of the posterior density over $[0.5, 1]$, namely,

$$p_{\mathbf{H}|\mathbf{D}}(h_i > 0.5|d_i) = \int_{0.5}^1 \prod_{j \in \Gamma(i)} h_i^{v_{\Gamma(i)}^j} (1 - h_i)^{1-v_{\Gamma(i)}^j} dh_i. \quad (10)$$

The helpfulness bias of the “Oracle” equals the area under the probability curve bounded by 0.5 and 1 (as indicated by the shaded area in Fig.2 which shows the helpfulness distribution in different circumstance). Fig. 2(a) demonstrates the helpfulness values for eWOM that hold the same positive fraction of $\frac{3}{5}$. It can be observed that the helpfulness distributions of eWOM vary significantly although the positive vote fractions are same. Fig. 2(b) demonstrates the helpfulness values for eWOM that hold the same positive fraction of 1. Conventional helpfulness formulation will fail to capture the difference between the helpfulness of those eWOMs and declare that they are the same helpful under this circumstance. In considering the problem of limitation of the positive vote fraction based benchmarking methodology, we advocate the helpfulness bias of “Oracle” to serve as the learning target for machine learning algorithms and regard it as a benchmark to compare against. With the proposed helpfulness bias as the ranking metric, the probability of the eWOM being helpful is identified and is ordered by this probability.

3.1.5 Feature as sufficient statistics for helpfulness inference. Since this function characterizing the dependency of helpfulness on an eWOM d_i is difficult to be determined in a learning framework, instead of considering helpfulness as depending on the eWOM functionally, we may consider helpfulness as depending on a set of extractable features of the review *probabilistically*. Let \mathcal{F} denote the set of all features relevant to voter opinion. That is, there is a function β mapping \mathcal{D} onto \mathcal{F} such that for every $d \in \mathcal{D}$ and every $v \in \mathcal{V}$, $v(d) = u_v(\beta(d))$ for some function u_v on \mathcal{F} . It then follows that for a randomly drawn document \mathbf{D} from \mathcal{D} (under any probability measure), random variables \mathbf{D} , \mathbf{F} and \mathbf{H} form a Markov chain, where $\mathbf{F} := \beta(\mathbf{D})$ namely, given the feature $\beta(\mathbf{D})$ of \mathbf{D} , the helpfulness of \mathbf{D} is independent of the document \mathbf{D} itself.

We can regard features as a sufficient statistic of the document, i.e.

$$p_{\mathbf{H}|\mathbf{D}}(h_i|d_i) = p_{\mathbf{H}|\mathbf{F},\mathbf{D}}(h_i|f_i, D_i) = p_{\mathbf{H}|\mathbf{F}}(h_i|f_i).$$

The objective of determining the helpfulness of the eWOM then can be formulated as finding:

$$p_{\mathbf{V}|\mathbf{F}}^* = \underset{p_{\mathbf{H}|\mathbf{F}} \in \Theta_{\mathbf{H}|\mathbf{F}}}{\operatorname{argmax}} \log \left[\prod_{i \in I} \prod_{j \in \Gamma(i)} \int h_i^{[v_j=1]} (1 - h_i)^{[v_j=0]} p_{\mathbf{H}|\mathbf{F}}(h_i|f_i) dh_i \right]. \quad (11)$$

In other words, we seek to determine $p_{\mathbf{H}|\mathbf{F}}$ from properly defined family of $\Theta_{\mathbf{H}|\mathbf{F}}$ which maximize the logarithm of $p_{\mathbf{V}|\mathbf{F}}(v_{\Gamma}|f_I)$.

3.1.6 Evaluation of Model and Estimated Parameters. Under our proposed helpfulness framework, we can design different algorithms/models to find the probability distribution of voters' opinions given a feature set. It is necessary to introduce a criterion to evaluate the goodness-of-fit for these algorithms/models. As defined in Eq.(1), the logarithm of the likelihood of observing all votes collectively given the features of all documents can be seen as how the model fit the data. Therefore, the objective of building a helpfulness model in our framework is to choose parameter values that achieve the best representation of the observed data. In other words, the criterion for choosing parameters from candidate distribution parameters is to find the one that best fit the training set.

For the "Oracle" learning machine, which has the prior knowledge of voters' opinions, $p_{\mathbf{H}|\mathbf{V}}(h_i|v_{\Gamma(i)})$ is the best possible estimate for $p_{\mathbf{H}|\mathbf{F}}(h_i|f_i)$. Such, the goodness-of-fit for the "Oracle" learning machine can be formulated as:

$$\begin{aligned} p_{\mathbf{V}|\mathbf{F}}(v_{\Gamma}|f_I) &= \prod_{i \in I} p_{\mathbf{V}|\mathbf{H}}(v_{\Gamma(i)}|h_i = \alpha_i^*) \\ &= \prod_{i \in I} \prod_{j \in \Gamma(i)} (\alpha_i^*)^{[v_j=1]} (1 - \alpha_i^*)^{[v_j=0]}, \end{aligned} \quad (12)$$

Accordingly, the difference between the logarithm of $p_{\mathbf{H}|\mathbf{V}}(h_i|v_{\Gamma(i)})$ of our algorithm and the Oracle estimation indicates the fitness of the algorithm for eWOM documents.

The log-likelihood can also assist us to measure the goodness-of-fit for different models. The log-likelihood function is the logarithm of the model formulation. In our case, the fitness of a model/algorithm can be formulated as $\log p_{\mathbf{V}|\mathbf{F}}(v_{\Gamma}|f_I)$, where $p_{\mathbf{V}|\mathbf{F}}(v_{\Gamma}|f_I)$ is the resulted marginal probability distribution learned by an algorithm.

For the other helpfulness evaluation algorithms/models not under our probabilistic framework which generate point estimate of helpfulness values, the likelihood of observing voters' opinions given a set of estimated helpfulness value $\bar{\alpha}_i, i \in I$ is

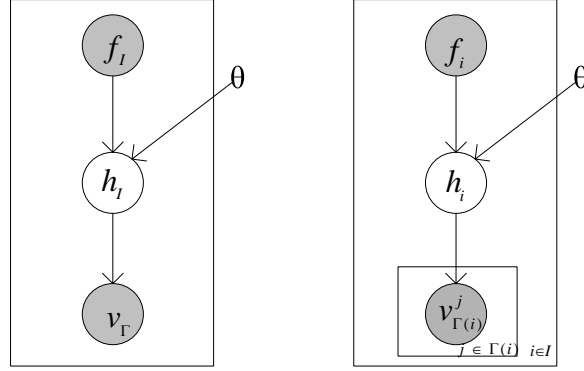
$$\begin{aligned} p_{\mathbf{V}|\mathbf{F}}(v_{\Gamma}|f_I) &= \prod_{i \in I} p_{\mathbf{V}|\mathbf{F}}(v_{\Gamma(i)}|h_i = \bar{\alpha}_i) \\ &= \prod_{i \in I} \prod_{j \in \Gamma(i)} \bar{\alpha}_i^{[v_{\Gamma(i)}^j=1]} (1 - \bar{\alpha}_i)^{[v_{\Gamma(i)}^j=0]}. \end{aligned} \quad (13)$$

4. HELPFULNESS DISTRIBUTION DISCOVERING MODEL

The algorithm under our framework proposed in this paper makes use of a graphical model to model the dependency between the random variables of our model and Expectation Maximization (EM) algorithm to learn the parameters.

4.1 A Graphical Model Fitting with Our Framework

Fig.3 provides the graphical representations of the eWOM helpfulness model. f_I denotes the feature set of eWOMs, h_I denotes the set of helpfulness, and v_{Γ} denotes the votes submitted by the readers. We suppose that each feature appears independently in an eWOM document. Then, h_i can be calculated by the following



(a) Overall Graphical Model Representation (b) Detailed Graphical Model Representation

Fig. 3. Graphical Model Representation

linear summation model:

$$h_i = f_i A + z, \quad (14)$$

where the variances of the Gaussian noise term $z \sim \mathcal{N}(0, \sigma^2)$.

Let us denote the parameters A and σ^2 by θ . Then the helpfulness distribution can be written in the form of $p_{\mathbf{H}|\mathbf{F}}(h_i|f_i; \theta)$. Consider the generative process for an eWOM containing features f_i , the probability density function of h_i is:

$$p_{\mathbf{H}|\mathbf{F}}(h_i|f_i; \theta) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(h_i - f_i A)^2}{2\sigma^2}\right). \quad (15)$$

From the graphical model specified in Fig. 3(b), it can be easily discovered that:

$$p_{\mathbf{H}|\mathbf{V}}(h_i, v_{\Gamma(i)}|f_i) = p_{\mathbf{V}|\mathbf{H}}(v_{\Gamma(i)}|h_i) p_{\mathbf{H}|\mathbf{F}}(h_i|f_i)$$

and

$$p_{\mathbf{V}|\mathbf{F}}(v_{\Gamma(i)}|f_i) = \int p_{\mathbf{V}|\mathbf{H}}(v_{\Gamma(i)}|h_i) p_{\mathbf{H}|\mathbf{F}}(h_i|f_i) d_{h_i}.$$

By the integration over the continuous hidden variable h_I , we obtain the marginal distribution in the form of

$$p_{\mathbf{V}|\mathbf{F}}(v_{\Gamma}|f_I) = \int \prod_{i \in I} p_{\mathbf{H}|\mathbf{F}}(h_i|f_i; \theta) \prod_{j \in \Gamma(i)} p_{\mathbf{V}|\mathbf{H}}(v_j|h_i) d_{h_I}, \quad (16)$$

Derived from Fig.3(a) and Fig.3(b), and taking the product of the marginal probabilities of the helpfulness of single document, we obtain the probability of observing all documents' helpfulness collectively given the features of all document and a set of parameters:

$$p_{\mathbf{H}|\mathbf{F}}(h_I|f_I; \theta) = \prod_{i \in I} p_{\mathbf{H}|\mathbf{F}}(h_i|f_i; \theta). \quad (17)$$

Given the helpfulness of h_I , the distribution of $v_{\Gamma(i)}$ is given by:

$$p_{\mathbf{V}|\mathbf{H}}(v_{\Gamma}|h_I) = \prod_{i \in I} p_{\mathbf{V}|\mathbf{H}}(v_{\Gamma(i)}|h_i). \quad (18)$$

Considering the conditional joint distribution of votes v_{Γ} and helpfulness h_I , we have:

$$\begin{aligned} p_{\mathbf{V},\mathbf{H}|\mathbf{F}}(v_{\Gamma}, h_I|f_I; \theta) &= p_{\mathbf{H}|\mathbf{F}}(h_I|f_I; \theta) p_{\mathbf{V}|\mathbf{H},\mathbf{F}}(v_{\Gamma}|h_I, f_I; \theta) \\ &= p_{\mathbf{H}|\mathbf{F}}(h_I|f_I; \theta) p_{\mathbf{V}|\mathbf{H}}(v_{\Gamma}|h_I) \\ &= \left[\prod_{i \in I} p_{\mathbf{H}|\mathbf{F}}(h_i|f_i; \theta) \right] \left[\prod_{i \in I} p_{\mathbf{V}|\mathbf{H}}(v_{\Gamma(i)}|h_i) \right] \\ &= \prod_{i \in I} p_{\mathbf{H}|\mathbf{F}}(h_i|f_i; \theta) p_{\mathbf{V}|\mathbf{H}}(v_{\Gamma(i)}|h_i). \end{aligned} \quad (19)$$

Based on Eq. (17), Eq. (18) and Eq. (19), for a given value of parameters θ and features f_I , the distribution of v_{Γ} is then obtained by integrating over h_I :

$$\begin{aligned} p_{\mathbf{V}|\mathbf{F}}(v_{\Gamma}|f_I; \theta) &= \int p_{\mathbf{V},\mathbf{H}|\mathbf{F}}(v_{\Gamma}, h_I|f_I; \theta) dh_I \\ &= \int \prod_{i \in I} p_{\mathbf{H}|\mathbf{F}}(h_i|f_i; \theta) p_{\mathbf{V}|\mathbf{H}}(v_{\Gamma(i)}|h_i) dh_I \\ &= \prod_{i \in I} \int p_{\mathbf{H}|\mathbf{F}}(h_i|f_i; \theta) p_{\mathbf{V}|\mathbf{H}}(v_{\Gamma(i)}|h_i) dh_i \end{aligned} \quad (20)$$

Our goal in adapting the feature vectors is to obtain a model of the helpfulness distribution of eWOMs. As can be seen from the discussion above, the optimization problem leads to the following maximization problem:

$$\begin{aligned} \hat{\theta} &= \operatorname{argmax}_{\theta} p_{\mathbf{V}|\mathbf{F}}(v_{\Gamma}|f_I; \theta) \\ &= \operatorname{argmax}_{\theta} \log p_{\mathbf{V}|\mathbf{F}}(v_{\Gamma}|f_I; \theta) \\ &= \operatorname{argmax}_{\theta} \sum_{i \in I} \log \int p_{\mathbf{H}|\mathbf{F}}(h_i|f_i; \theta) p_{\mathbf{V}|\mathbf{H}}(v_{\Gamma(i)}|h_i) dh_i \end{aligned} \quad (21)$$

4.2 Learning Algorithm

We make use of Expectation Maximization(EM) algorithm[Dempster et al. 1977] to solve the optimization problem to find the $\hat{\theta}$ which maximizes $p_{\mathbf{V}|\mathbf{F}}(v_{\Gamma}|f_I; \theta)$ for our proposed model.

The following statements show how to do the maximization of estimating the parameters.

E-Step:

$$\begin{aligned}
q^{t+1}(h_I) &= p_{\mathbf{H}|\mathbf{F},\mathbf{V}}(h_I|f_I, v_{\Gamma}; \theta^t) \\
&= \prod_{i \in I} p_{\mathbf{H}|\mathbf{F},\mathbf{V}}(h_i|f_i, v_{\Gamma(i)}; \theta^t)
\end{aligned} \tag{22}$$

We denote $p_{\mathbf{H}|\mathbf{F},\mathbf{V}}(h_i|f_i, v_{\Gamma(i)}; \theta^t) := q_i^t(h_i)$. Then, $q^{t+1}(h_I)$ is completely specified by $\{q_i^t(h_i) : i \in I\}$.

$$\begin{aligned}
q_i^{t+1}(h_i) &= \frac{p_{\mathbf{H},\mathbf{V}|\mathbf{F}}(h_i, v_{\Gamma(i)}|f_i; \theta^t)}{p_{\mathbf{V}|\mathbf{F}}(v_{\Gamma(i)}|f_i; \theta^t)} \\
&\propto p_{\mathbf{H},\mathbf{V}|\mathbf{F}}(h_i, v_{\Gamma(i)}|f_i; \theta^t) \\
&= p_{\mathbf{H}|\mathbf{F}}(h_i|f_i; \theta^t) p_{\mathbf{V}|\mathbf{H}}(v_{\Gamma(i)}|h_i)
\end{aligned} \tag{23}$$

M-step:

$$\begin{aligned}
\theta^{t+1} &= \operatorname{argmax}_{\theta} \int q^{t+1}(h_I) \log p_{\mathbf{V},\mathbf{H}|\mathbf{F}}(v_{\Gamma}, h_I|f_I; \theta) dh_I \\
&= \operatorname{argmax}_{\theta} \sum_{i \in I} \int q_i^{t+1}(h_i) (h_i - f_i A^T) dh_i.
\end{aligned} \tag{24}$$

In summary, the procedure for our EM algorithm consists of the following steps:

- (1) Initialize $\theta^t = \theta^0$, $\theta^{t-1} = \theta^0$, where θ^0 is the initial estimate of the parameter and the superscript represents the index of iteration. For helpfulness prediction, we set $\theta^0 = 1$
- (2) Calculate $q_i^t(h_i)$ by following Eq.(23) with respect to θ^{t-1} .
- (3) Update $q_i^t(h_i)$ by Eq.(22).
- (4) Update θ^t by Eq.(24).
- (5) Set $t = t + 1$. If convergence condition is not satisfied, go to step 2.

Note that in the M-step of the algorithm, we fixed σ for each iteration and then σ can be deducted during calculating the maximum θ . For an evaluation at algorithm level, we choose the most probable σ for each eWOM, namely, to find a $\hat{\theta}$ to maximize the log-likelihood of $p_{\mathbf{V}|\mathbf{F}}(v_{\Gamma}|v_I; \theta)$. We predefine a σ candidate set of $\{0.1, 0.2, 0.3, \dots, 1.0\}$ and iteratively go through this algorithm. As a result, 10 pairs of σ, A is learned by the EM algorithm.

As discussed in , the parameter σ can be globally optimized by 3.1.6. This forces the width of the distribution to be the same for every document. Therefore, we choose the σ and its corresponding A which maximize the probability of observing the training data. When a new eWOM content arrives, the parameter θ which maximizes $\log p_{\mathbf{V}|\mathbf{F}}(v_{\Gamma(i)}|f_i; \theta)$ is chosen to calculate the helpfulness distribution of this eWOM content. Afterwards, the helpfulness bias of eWOM contents will be calculated and eWOM contents are ranked according to their corresponding helpfulness bias.

5. EXPERIMENTAL RESULTS

This section shows the experimental results of the experimental results by applying our framework to the real eWOM set, in form of online product reviews, obtained from Amazon.com.

5.1 Data Set

Our experiments explicitly focus on the categories of electronic products and books. We crawled 1002 HDTV reviews, 1492 Camera reviews, and 2408 book reviews from Amazon.com¹. These reviews have been evaluated by at least 10 consumers as helpful or not helpful. We utilized the bag-of-word approach to build the language model. Each feature is a non-stop stemmed word and the value of this feature is a boolean value of the occurrence of the word on the review. We make use of the vocabulary of “word” or “terms” as the feature set of the model to build a document-term matrix. Each the matrix has been normalized to zero-mean.

There are more than 15,000 identical terms in each of the eWOM corpus, so the term-document matrix of the document set is sparse and we need to reduce the document dimensions. After normalizing the document-term matrix, we performed principal component analysis (PCA) [Jolliffe 2002] to reduce the dimension of term-document matrix. Therefore the top 200 components, which dominate about 70 percent of the total variance, are selected in the later on experimentation. We compute the principle eigenvectors of the covariance matrix for each training set and project the original document-term vector to a 200-dimensional space.

5.2 Results and Analysis

To validate the ranking performance of our model, we compare our probabilistic model with Support Vector Regression (SVR) [Burges 1998], Artificial Neural Network (ANN) [Bishop 1996], and Linear Regression.

The ANN is one of the most popular machine-learning algorithms to solve modeling and predicting problems. We implement a three-layer error back-propagation (BP) ANN in this study. The number of neurons in its hidden layer is chosen to be 10. Each node utilizes sigmoid transfer function. The output node makes use of log-sigmoid transfer function. The training progress is set to be stopped after 1000 iterations of learning, or when present error value less than 0.001 is reached.

The SVR is also a popular machine-learning algorithm successfully used for assessing review helpfulness [Kim et al. 2006; Zhang and Varadarajan 2006]. The SVR is applied in this study to compare the effectiveness with our probabilistic model. A linear regression model has also been used for scoring the utility of reviews in [Zhang and Varadarajan 2006]. Basically, linear regression is to find parameters which minimize the sum of the square deviations of all observed data. We applied multiple linear regression model to predict the helpfulness of reviews and examined the ranking performance of this technique.

For all of these models, we use the PCA projected review data as the training data and test data. The learning target for these algorithms is the helpfulness bias of the “Oracle”. The helpfulness predictions from these models are evaluated by

¹The complete data set is available at our web set at <http://www.site.uottawa.ca/~rzhan025/helpfulness.html>.

computing and comparing the Spearman’s Rank Correlation Coefficient between the predicted rankings and the “Oracle” rankings.

In order to investigate the effectiveness of models towards eWOMs with different voters population sizes, we separate each categories of documents into two groups according to the number of voters associated with each document. Group 2 contains all eMOMs and the corresponding voter opinions. We randomly choose 5 opinions for each eWOM message from group 2 to form group 1.

To obtain a fair comparison, we perform 10-fold cross-validation on our probabilistic model, Support Vector Regression (SVR) ², ANN, and linear regression algorithm and compare the cross-validation performances.

5.2.1 Goodness-of-Fit Tests. Fig.4 demonstrates the goodness-of-fit of the “Oracle” (straight line) and our model (curve) with varying σ from 0.1 to 1 with the training data. As described in our proposed EM algorithm, the σ which maximizes the likelihood of model and the corresponding parameter set θ is chosen to be the final parameter to build our helpfulness model. We make use of a Gaussian distribution to model the relationship between h_i and f_i . With this Gaussian distribution, if a document is associated with a small number of votes, the σ will be greater than the one with large number of votes. If σ is selected as a large number, the distribution of $p_{\mathbf{H}|\mathbf{F}}(h_i|f_i;\theta)$ will run out of the range of (0,1). Therefore, in practice, one more consideration in choosing appropriate σ is desired. The value of σ can not be too big, so that most of the Gaussian distribution stays between (0,1). In addition, based on the knowledge of the training data set (as illustrated in Fig.4), σ can not be too small to maximize the likelihood of our model. We examine our model with three categories of eWOM contents by varying σ from 0.1 to 1 and find that a value of $\sigma = 0.2$ is a satisfactory choice that meets these requirements and is used for the following experimentation.

Table I demonstrates the difference of the loglikelihood between the “Oracle”, our algorithm, SVR, linear regression and ANN with the testing data. It can be seen that the goodness-of-fit of our model consistently outperforms SVR, ANN and linear regression for reviews for all the three categories and both groups of eWOM contents. This result indicates that our model fits the observed data better than other algorithms no matter the available voter population for the training purpose is big or small. The log-likelihood of the prediction results maybe not perfect for evaluating the effectiveness of non-probabilistic algorithms. Therefore, we introduce the rank correlation coefficient to evaluate the ranking performance of different algorithms, probabilistic and non-probabilistic.

5.2.2 Ranking Performance. In order to compare the obtained helpfulness bias from different learning algorithms with the helpfulness bias of “Oracle”, we evaluate the prediction accuracy of helpful rankings in terms of Spearman’s ranking correlation coefficient, which is widely used for comparing two ordinal correlations

²We use the LibSVM[Chang and Lin 2001] tool to perform the Support Vector Regression. The parameters of the SVR, C and g, were chosen by applying a 10-fold cross-validation and a grid search on a logarithmic scale.

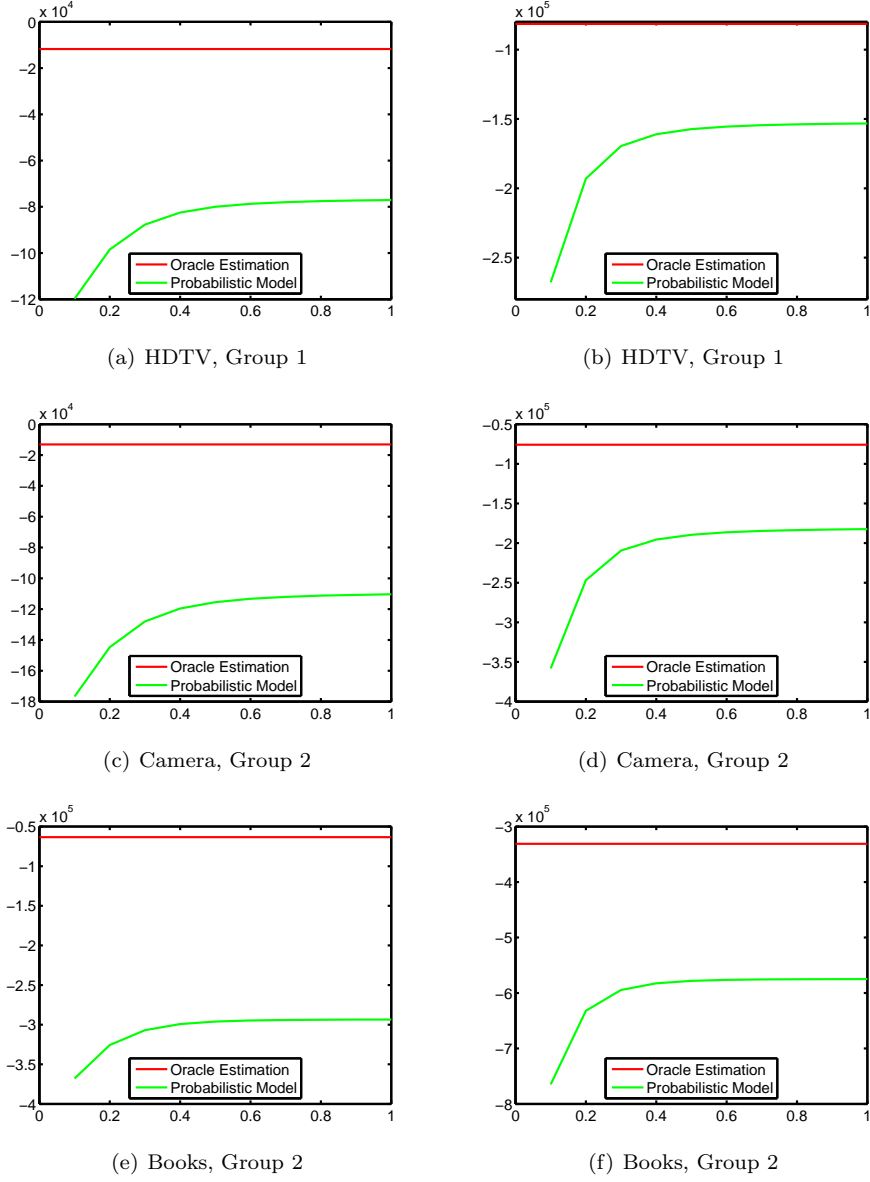


Fig. 4. The goodness-of-fit of our model and “Oracle” on the training data.

between two random variables. It is defined by the following formula:

$$r = 1 - \frac{6 \sum_i^n (x_i - y_i)^2}{n(n^2 - 1)}. \quad (25)$$

The ranking performance of different approaches is analyzed by 10-fold cross-validation. Table II shows the experimental results of our probabilistic model, SVR, ANN, and Linear Regression algorithm on HDTV, camera and book reviews.

Table I. The goodness-of-fit of our probabilistic model, SVM Regression, ANN and linear regression for the testing data (10-fold cross-validation). Group 1 and group 2 contain review documents which have been voted by 5 voters and more than 10 voters respectively.

Online Reviews	Algorithm	Group 1	Group 2
HDTV	Oracle	-9096	-9096
	Probabilistic Model	-14155	-13013
	SVR	-19194	-18757
	BP-ANN	-19960	-22841
	Linear Regression	-15602	-16072
Camera	Oracle	-8513	-8513
	Probabilistic Model	-16880	-15723
	SVR	-19584	-19101
	BP-ANN	-17541	-18774
	Linear Regression	-19656	-19574
Books	Oracle	-36814	-36814
	Probabilistic Model	-43777	-43621
	SVR	-47447	-47076
	BP-ANN	-52223	-127160
	Linear Regression	-54467	-57468

Table II. Spearman's rank correlation of our probabilistic model, SVM Regression, ANN and linear regression (10-fold cross-validation). Group 1 and group 2 contain review documents which have been voted by 5 voters and more than 10 voters respectively.

Online Reviews	Algorithm	Group 1	Group 2
HDTV	Probabilistic Model	0.62	0.65
	SVR	0.56	0.58
	BP-ANN	0.54	0.57
	Linear Regression	0.50	0.54
Camera	Probabilistic Model	0.47	0.49
	SVR	0.41	0.40
	BP-ANN	0.44	0.46
	Linear Regression	0.40	0.40
Books	Probabilistic Model	0.50	0.51
	SVR	0.42	0.47
	BP-ANN	0.39	0.45
	Linear Regression	0.42	0.46

The results demonstrate that the effectiveness of our probabilistic model and indicate that our model consistently outperforms SVR, ANN and linear regression for reviews with any voter population size.

For the reviews from all three review categories, compared with the other existing approaches, a significant improvement of the Spearman's rank correlation coefficient is obtained by our model. In each group of data, the results show a close correlation between the "Oracle" and our probabilistic model. For instance, the correlation coefficients of our model and the "Oracle" for each group of reviews are 0.62 and 0.65 respectively for HDTV reviews, 0.47 and 0.49 for camera reviews, and 0.50 and 0.51 for book reviews. This result shows that our probabilistic model has a significant performance advantage over SVR, ANN and linear regression.

Furthermore, as the rank correlation coefficients listed are averaged values of the 10 runs of experimentations, we perform t-test to evaluate the difference between the 10 resulted ranking correlations by our algorithm and other algorithms' ranking correlations of 10 runs. We find that the difference is significant at the 0.05 probabilistic level. These results strongly indicate that our model can consistently learn the helpfulness model under any voter population size and outperforms than SVM Regression, ANN and linear regression.

As we expected, SVR, ANN and Linear Regression with the reviews that are voted by more than 10 voters for all the two types of reviews performs better than the reviews voted by 5 voters. This is because that the confidence of the point prediction is low when voter population size is small. When more votes are available for an eWOM, the benchmark helpfulness value, which is used for training the point estimate model, is more certain and the ranking performance is improved by the increased certainty of data. Whereas our probabilistic approach, that estimates the distribution of the voter opinion and estimate the helpfulness bias based on the estimated helpfulness distribution, is less likely to be affected by the shortage of voter opinions.

6. DISCUSSION AND CONCLUSION

Today's online customers are impatient and demanding, while wanting to ensure the best purchasing decisions. Meanwhile, they are not willing to spend much time and effort for their purchases. The available eWOMs provide an immense amount of information that assists consumers in making purchasing decisions. Consumers are willing to read others consumers' experiences and opinions. However, there is no formal format for the eWOM content available on the Internet. These contents are free-styled and consist of unstructured text and information. This fact makes consumers unable to find useful content easily as they have to spend more time surfing for helpful opinions and experiences.

The helpful eWOMs recommendation framework proposed in this paper can surely assist to make purchasing decisions. An online store providing a review filtering system will definitely help potential consumers to find there interested product reviews and reduce the purchase decision making time. An online community that incorporates this review recommendation system can significantly increase the usability and attract more users. One important and interesting extension of our model would be to integrate eWOMs from different sources or online communities. This would provide possible consumers more opportunities to benefit from other people's experience.

The main contribution of this paper is a novel probabilistic framework for model-

ing the helpfulness of eWOMs. This probabilistic framework provides a theoretical interpretation and a mathematical estimation technique to model the helpfulness distribution of online eWOM. We have utilized the graphical modeling and EM algorithm to build a helpfulness assessment model under the proposed framework. Both of the mathematical inference and experimental results validate our framework. Our framework is clearly explained by the helpfulness inference and the experimental results show that a model under our framework can efficiently rank online reviews as our algorithm outperforms the existing helpfulness assessment approaches by a clear margin.

Previous studies simply make use of the positive vote fraction as the benchmark to evaluate the helpfulness of eWOMs. As this benchmark hardly represents the true helpfulness value when only a few voters have voted on an eWOM, existing approaches only take ones which have been voted by a large number (such as >10) of readers. However, in reality, most of eWOMs are only associated with a small number of available voter opinions and the votes need time to be collected. The available data for training the helpfulness model will be limited to a small amount of eWOMs.

Furthermore, the traditional helpfulness definition will treat the fraction of $\frac{9}{10}$, $\frac{18}{20}$, and $\frac{27}{30}$ as the same helpfulness value whereas the confidence of these helpfulness estimates varies a lot depending on voter population size. Our approach solves this problem by estimating the helpfulness distribution instead of a single value. We proposed a helpfulness bias function as “benchmarking” to justify the helpfulness of an eWOM document from the available voter opinions and evaluate the true helpfulness value from the helpfulness distribution. The experimental results show that our framework can effectively model the helpfulness of eWOM, and it will help consumers reduce the processing effort in making purchase decisions.

Although in our algorithm we only take the unigram features and a Gaussian distribution into account to simulate the helpfulness distribution, our framework is capable of making use of any features and any other distributions. We envision two main future research directions that will possibly improve the efficiency of helpfulness modeling algorithms under our framework. First, to examine the possibility of incorporating other factors, which may affect the quality of an eWOM document, to improve the precision of recommendations, e.g., when the eWOM content is published, how consumer rate the product, and which emotional terms and product features are used. Second, to apply other distributions or other combination of distributions, to model the relations between the features and helpfulness value to improve the ability and accuracy of the helpfulness distribution prediction.

REFERENCES

- BISHOP, C. M. 1996. *Neural Networks for Pattern Recognition*, 1 ed. Oxford University Press.
- BURGES, C. J. C. 1998. A tutorial on support vector machines for pattern recognition. *Data Min. Knowl. Discov.* 2, 2, 121–167.
- CHANG, C. C. AND LIN, C. J. 2001. *LIBSVM: a library for support vector machines*.
- DEMPSTER, A. P., LAIRD, N. M., AND RUBIN, D. B. 1977. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)* 39, 1, 1–38.
- DUAN, W., GU, B., AND WHINSTON, A. B. 2008. Do online reviews matter? – an empirical inves-

- tigation of panel data. *Decision Support Systems* 45, 4, 1007 – 1016. Information Technology and Systems in the Internet-Era.
- HATZIVASSILOGLOU, V. AND MCKEOWN, K. R. 1997. Predicting the semantic orientation of adjectives. In *Proceedings of the eighth conference on European chapter of the Association for Computational Linguistics*. Association for Computational Linguistics, Morristown, NJ, USA, 174–181.
- HU, M. AND LIU, B. 2004. Mining and summarizing customer reviews. In *KDD '04: Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, New York, NY, USA, 168–177.
- HU, N., LIU, L., AND ZHANG, J. J. 2008. Do online reviews affect product sales? the role of reviewer characteristics and temporal effects. *Information Technology and Management*.
- HU, N., PAVLOU, P. A., AND ZHANG, J. 2006. Can online reviews reveal a product's true quality?: empirical findings and analytical modeling of online word-of-mouth communication. In *EC '06: Proceedings of the 7th ACM conference on Electronic commerce*. ACM, New York, NY, USA, 324–330.
- JOLLIFFE, I. T. 2002. *Principal Component Analysis*, 2nd ed. Springer.
- KIM, S.-M., PANTEL, P., CHKLOVSKI, T., AND PENNACCHIOTTI, M. 2006. Automatically assessing review helpfulness. In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, Sydney, Australia, 423–430.
- LEE, J., PARK, D.-H., AND HAN, I. 2008. The effect of negative online consumer reviews on product attitude: An information processing view. *Electronic Commerce Research and Applications* 7, 3, 341 – 352. Special Section: New Research from the 2006 International Conference on Electronic Commerce - Evaluating new technological innovations for successful business on the internet, Eighth International Conference on Electronic Commerce (ICEC).
- LIU, Y., HUANG, X., AN, A., AND YU, X. 2008. Modeling and predicting the helpfulness of online reviews. *Data Mining, 2008. ICDM '08. Eighth IEEE International Conference on*, 443–452.
- MAYZLIN, D. AND CHEVALIER, J. 2003. The effect of word of mouth on sales: Online book reviews.
- NIELSEN-ONLINE, H. S. 2008. 81 percent of online holiday shoppers read online customer reviews, according to nielsen online.
- PALMER, A. 2009. Web shoppers trust customer reviews more than friends. Online.
- PANG, B., LEE, L., AND VAITHYANATHAN, S. 2002. Thumbs up?: sentiment classification using machine learning techniques. In *EMNLP '02: Proceedings of the ACL-02 conference on Empirical methods in natural language processing*. Association for Computational Linguistics, Morristown, NJ, USA, 79–86.
- PARK, D.-H. AND KIM, S. 2008. The effects of consumer knowledge on message processing of electronic word-of-mouth via online consumer reviews. *Electronic Commerce Research and Applications* 7, 4, 399 – 410.
- PARK, D.-H. AND LEE, J. 2008. ewom overload and its effect on consumer behavioral intention depending on consumer involvement. *Electronic Commerce Research and Applications* 7, 4, 386 – 398.
- PARK, D.-H., LEE, J., AND HAN, I. 2007. The effect of on-line consumer reviews on consumer purchasing intention: The moderating role of involvement. *Int. J. Electron. Commerce* 11, 4, 125–148.
- PARRA, J. F. AND RUIZ, S. 2009. Consideration sets in online shopping environments: the effects of search tool and information load. *Electron. Commer. Rec. Appl.* 8, 5, 252–262.
- ROBERT M. SCHINDLER, B. B. 2005. *Online Consumer Psychology: Understanding and Influencing Consumer Behavior in the Virtual World*. Lawrence Erlbaum Associates, Inc.
- TURNER, P. 2002. Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews. 417–424.
- VERMEULEN, I. E. AND SEEGER, D. 2009. Tried and tested: The impact of online hotel reviews on consumer consideration. *Tourism Management* 30, 1, 123 – 127.
- WEIMER, MARKUS; GUREVYCH, I. 2007. Predicting the perceived quality of web forum posts. *Proceedings of the Conference on Recent Advances in Natural Language Processing (RANLP)*.
- ACM Journal Name, Vol. V, No. N, Month 20YY.

- WEIMER, M., GUREVYCH, I., AND MÜHLHÄUSER, M. 2007. Automatically assessing the post quality in online discussions on software. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics Companion Volume Proceedings of the Demo and Poster Sessions*. Association for Computational Linguistics, Prague, Czech Republic, 125–128.
- YU, H. AND HATZIVASSILOPOULOS, V. 2003. Towards answering opinion questions: separating facts from opinions and identifying the polarity of opinion sentences. In *Proceedings of the 2003 conference on Empirical methods in natural language processing*. Association for Computational Linguistics, Morristown, NJ, USA, 129–136.
- ZHANG, Z. AND VARADARAJAN, B. 2006. Utility scoring of product reviews. In *CIKM '06: Proceedings of the 15th ACM international conference on Information and knowledge management*. ACM, New York, NY, USA, 51–57.
- ZHUANG, L., JING, F., AND ZHU, X.-Y. 2006. Movie review mining and summarization. In *CIKM '06: Proceedings of the 15th ACM international conference on Information and knowledge management*. ACM, New York, NY, USA, 43–50.

Received April 1, 2010; revised September 1, 2010; accepted September 16, 2010.