Analyzing and Comparing opinions on the Web mining Consumer Reviews

Saranya.N[1], Prof. K. Anbarasu[2], Devi Ramalingam[3], Ramya. S[4]
[1,3,4]PG SCHOLAR, [2] Associate Professor
[1,2,3,4] Computer Science and Engineering, Selvam College of Technology, Namakkal, India

Abstract— Product reviews posted at online shopping sites plays a major role in improving performance of various enterprises. To assess the performance, the posted reviews must be of good quality. The good quality is judged by using certain criteria (rules) to be satisfied. The criteria (rules) should be applied on the online reviews or the documents collected based upon reviews. Thus, it is considered to be very difficult for decision-maker with an efficient post processing step in order to reduce the number of rules. This project proposes a new classification based interactive approach to prune and filter discovered rules to eliminate low-quality reviews. The proposed approach to enhance opinion summarization is done in a two-stage framework which is (1) discriminates low quality reviews from high-quality ones and (2) enhances the task of opinion summarization by detecting and filtering low quality reviews. For the sentiment factor, we propose Sentiment PLSA (S-PLSA), in which a review is considered as a document generated by a number of hidden sentiment factors, in order to capture the complex nature of sentiments. Training an S-PLSA model enables us to obtain a succinct summary of the sentiment information embedded in the reviews.

Keywords—Review mining, sentiment analysis, prediction.

I. INTRODUCTION

Due to the introduction of web several users posting online reviews has become an increasingly popular way for people to share with other users their opinions and sentiments toward products and services and interact with others through blogs and forums. Several websites provide facilities for the users to publish their opinions in their websites such as Amazon (www.Amazon.com) . Reviews are considered to be prevalent in blogposts, social networkingsites (www.facebook.com) as well as dedicated review websites such as opinions (www.eopinion.com). This online review is considered to be a valuable tool for firms who can use them to monitor customer attitude towards product in real-time and change their manufacture rate and strategies correspondingly.

A large no of studies is exploring the relationship between the sales performance of products and their reviews [1], [2], [3], [4].The feeling of public towards the product can be very well seen on the reviews and due to this factor reviews are considered to be a very good factor for improving performance . In this paper, the knowledge can be extracted from the reviews i.e., review mining by developing the corresponding models and algorithms. By using Such models and algorithms can be used to effectively predict the future sales of products, which can provide knowledge to the stakeholders about the status of their products.

We start with modeling sentiments in reviews, which presents unique challenges that cannot be easily addressed by conventional text mining methods. The classification of reviews in to positive or negative does not exactly reciprocate the feelings of each people

If the opinion is positive then it is a gain to the firm or if it is negative it is a loss to the firm. In order to model the multifaceted nature of sentiments, we view the sentiments embedded in reviews as an outcome of the joint contribution of a number of hidden factors, and propose a novel approach to sentiment mining based on Probabilistic Latent Semantic Analysis (PLSA), which we call Sentiment PLSA (S-PLSA). In difference to traditional PLSA [6], S-PLSA focuses on sentiments rather than topics. Therefore, instead of taking a vanilla "bag-of-words" approach and considering all the words (modulo stop words) present in the blogs, we focus primarily on the words that are sentiment related. To this end, we adopt in our study the appraisal words extracted from the lexicon constructed by Whitelaw et al. [7]. Despite the seemingly lower word coverage (compared to using "bag of words"), decent performance has been reported when using appraisal words in sentiment classification of movie reviews [7]. In S PLSA, appraisal words are exploited to compose the feature vectors for review which are then used to infer the hidden sentiment factors.

The second factor to be considered is taking statistics of past sales performance. It can be done by using Autoregressive (AR) model, which has been widely used for many time series analysis problems, especially in econometric contexts [8]. Combining this AR model with sentiment information mined from the reviews, we propose a new model for product sales prediction called the Autoregressive Sentiment Aware (ARSA) model. Extensive experiments show that the ARSA model provides superior predication performance compared to using the AR model alone, confirming our expectation that sentiments play an important role in predicting future sales performance.

Since each online reviews varies from the other in quality and the product it is posted, thus, carry different predictive power, it should not treated equally when making the prediction. It is tested with the case where quality indicators are available (e.g., in the form of user ratings) and where it is not available. Our focus is on the latter case, for which we develop a model that is able to automatically predict the quality of a review based on its syntactical characteristics. The quality factor is then incorporated into the ARSA model, resulting in an Autoregressive Sentiment and Quality Aware (ARSQA) model for sales prediction. In summary, we make the following contributions: Using the shopping application as a case study, we approach the problem of predicting sales performance using online reviews as a domain-driven task, and identify the important the factors involved in generating prediction. We model sentiments.

Propose a new model for product sales prediction called the Autoregressive Sentiment Aware (ARSA) model. Extensive experiments show that the ARSA model provides superior predication performance compared to using the AR model alone, confirming our expectation that sentiments play an important role in predicting future sales performance.

Since online reviews are of varying quality and, thus, carry different predictive power, we should not treat them equally in producing the prediction. This motivates our study of the quality factor in sales prediction. We consider both cases in which quality indicators are readily available (e.g., in the form of user ratings), and cases in which they are not. Our focus is on the latter case, for which we develop a model that is able to automatically predict the quality of a review based on its syntactical characteristics. The quality factor is then incorporated into the ARSA model, resulting in an Autoregressive Sentiment and Quality Aware (ARSQA) model for sales prediction. In summary, we make the following contributions: Using the movie domain as a case study, we approach the problem of predicting sales performance using online reviews as a domain-driven task, and identify the important the factors involved in generating prediction. We model sentiments in reviews as the joint outcome of some hidden factors, answering the call for a model that can handle the complex nature of sentiments. We propose the S-PLSA model, which through the use of appraisal groups, provides a probabilistic framework to analyze sentiments in reviews. We develop a model for predicting the quality of reviews in the absence of readily available quality indicators. We propose the ARSA and ARSQA models for product sales prediction, which reflects the effect of sentiments, and past sales performance (and in the case of ARSQA, the quality of reviews) on future sales performance. Their effectiveness is confirmed by experiments. We discuss how actionable knowledge can be derived through utilizing the proposed models, explaining the practical impact of the proposed approach.

II. DESCRIPTION

A. Domain-Driven Data Mining (D3M)

In the past few years, domain-driven data mining has emerged as an important new paradigm for knowledge discovery by being a forum for sharing findings, insight, experience [9], [10]. Motivated by the significant gap between the academic goals of many current KDD
(Knowledge driven data mining) methods and the real-life business goals, D3 advocates the shift from data centered hidden pattern mining to domain-driven Actionable Knowledge Discovery (AKD). The work presented in this paper can be considered as an effort along this direction in that 1) we aim to deliver actionable knowledge by making predictions of sales performance to provee how KDD outweighs domain driven data mining 2) in developing the prediction model, we try to integrate multiple types of intelligence, including human intelligence, domain intelligence, data intelligence and network intelligence (Web intelligence) so as to identity the future challenges.

B. Review Mining

With the rapid growth of online reviews, review mining has attracted a great deal of attention. Early work in this area was primarily focused on determining the semantic orientation of reviews. Among them, some of the studies attempt to learn a positive/negative classifier at the document level. Pang et al. [11] employ three machine learning approaches (Naive Bayes, Maximum Entropy, and Support Vector Machine) to label the polarity of IMDB movie reviews. In follow-up work, they propose to first extract the subjective portion of text with a graph min-cut algorithm, and then feed them into the sentiment classifier [12]. Instead of applying the straightforward frequency-based bag-of-words feature selection methods ,Whitelaw et al. [7] defined the concept of “adjectival appraisal groups” headed by an appraising adjective and optionally modified by words like “not” or “very.” Each appraisal group was further assigned four types of features: attitude, orientation, graduation, and polarity. They report good classification accuracy using the appraisal groups. They also show that the classification accuracy can be further boosted when they are combined with standard “bag-of-words” features. We use the same words and phrases from the appraisal groups to compute the reviews’ feature vectors, as we also believe that such adjective appraisal words play a vital role in sentiment mining and need to be distinguished from other words. Our way of using these appraisal groups is different from that in [7].

C. Economic Impact of Online Reviews

Whereas marketing plays an important role for each product especially for newly released product So that it can reach them and it should be done in such a way so that each and every customer would try to buy them Therefore, online product reviews can be very valuable to the vendors in that they can be used to monitor consumer opinions toward their products in real time, and adjust their manufacturing, servicing, and marketing strategies accordingly.

Academics have also recognized the impact of online reviews on business intelligence, and have produced some important results in this area. Among them, some studies attempt to answer the question of whether the polarity and the volume of reviews available online have a measurable and significant effect on actual customer purchasing [18],[19], [20], [5], [1]. To this end, most studies use some form of hedonic regression [21] to analyze the significance of different features to certain function, e.g., measuring the utility to the consumer. Various economic functions have been utilized in examining revenue growth, stock trading volume change as well as the bidding price variation on commercial websites, such as Amazon and eBay.

Foutz and Jank [22], [23] also exploit the wisdom of crowds to predict the box office performance of movies. The work presented in this paper differs from theirs in three ways. First, we use online reviews as a source of network intelligence to understand the sentiments of the public, whereas their approach uses virtual stock markets (prediction markets) as an aggregated measure of public sentiments and expectations. Second, we use a parametric regression model to capture the temporal relationships, whereas their approach uses nonparametric functional shape analysis to extract the important features in the shapes across various trading histories and then uses these key features to produce forecasts. Third, the prediction of our model is ongoing as time progresses and more reviews are posted, whereas their approach is limited to forecasting the box office performance in the release week.

D. Ignoring Spam Review

There is also no reported study on the trustworthiness of opinions in reviews. Due to the fact that there is no quality control, anyone can write anything on the Web. This results in many low quality reviews, and worse still review spam. . There are generally three types of spam reviews

1) Type 1 (untruthful opinions):

Those that deliberately mislead readers or opinion mining systems by giving undeserving positive reviews to some target objects in order to promote the objects (which we call hyper spam) and/or by giving unjust or malicious negative reviews to some other objects in order to damage their reputation (which we call defaming spam). Untruthful reviews are also commonly known as fake reviews or bogus reviews. They have become an intense discussion topic in blogs and forums.

2) Type 2 (reviews on brands only):

Those that do not comment on the products in reviews specifically for the products but only the brands, the manufacturers or the sellers of the products. Although they may be useful, we consider them as spam because they are not targeted at the specific products and are often biased.

3) *Type 3 (non-reviews):*

Those that are non-reviews, which have two main sub-types: (1) advertisements and (2) other irrelevant reviews containing no opinions (e.g., questions, answers, and random texts.

E. Unifying Collaborative and Content-Based Filtering

In Collaborative filtering forms the correlation between the users who has similar votes and form a group between the users and this is done to predict unbounded votes. Collaborative filtering can be model based or memory based.

III. MEMORY-BASED COLLABORATIVE FILTERING ALGORITHMS

Memory-based algorithms utilize the entire user-item database to generate a prediction. These systems employ statistical techniques to and a set of users, known as neighbors, that have a history of agreeing with the target user (i.e., they either rate different items similarly or they tend to buy similar set of items). Once a neighborhood of users is formed, these systems use different algorithms to combine the preferences of neighbors to produce a prediction or top-N recommendation for the active user. The techniques, also known as nearest neighbor or user-based collaborative filtering, are more popular and widely used in practice.

IV. MODEL-BASED COLLABORATIVE FILTERING ALGORITHMS

Model-based collaborative filtering algorithms provide item recommendation by first developing a model of user ratings. Algorithms in this category take a probabilistic approach and envision the collaborative filtering process as expected value of a user prediction, given his/her ratings on other items. The model building process is performed by different machine learning algorithms such as Bayesian network, clustering, and rule-based approaches. The Bayesian network model formulates a probabilistic model for collaborative filtering problem. Clustering model treats collaborative filtering as a classification problem and works by clustering similar users in same class and estimating the probability that a particular user is in a particular class C, and from there computes the conditional probability of ratings. The content based filtering rely on textual (content) description of users and it is mainly used for books, web, newspaper Various CF algorithms ranging from typical nearest neighbor methods [27] to more complex probabilistic-based methods [28], [29] have been designed to identify users of similar interests. A few variations and hybrid methods that combine both content information and collaborative filtering have also been proposed to solve the cold-start problem [30], [31]. Similar to recommender systems, our work also accounts for textual contents and peer votes in those reviews to effectively construct and evaluate the prediction model; however, one of the objectives of this study is to investigate the quality of movie reviews, which is different from the above work.

A. Time Series Analysis

Time series analysis is the analysis of data collected over time which can be weekly, monthly, yearly.

Time series analysis is a well-established field with a large body of the literature [8]. Three broad classes of models that are most widely used are autoregressive models, the Moving Average (MA) models, and the integrated (I) models. Moving average can be Ordinary Moving Averages or Exponentially Weighted Moving Averages Ordinary Moving Averages For a "span" of k periods

$x=x_1+x_2+x_3+\ldots.x_n/k$

Where seasonal effects are expected, it is standard to use k = number of periods per cycle.

Exponentially Weighted Moving Averages These weight observations less heavily as one moves back in time from the current period. They are typically computed "recursively" as

x = exponentially weighted moving average time t

$=wy+(1-w)\,x-1$

In real applications, they are often combined to produce models like Autoregressive Moving Average (ARMA) and Autoregressive Integrated Moving Average

(ARIMA) models. When the time series data are vector valued, these models can be extended to their vectored versions, which constitutes multivariate time series analysis. Our method utilizes times series analysis in that we use the AR model to capture the temporal relationship in the box office data. The main difference between our work and conventional time series analysis is that while in conventional time series analysis the data series are generally directly observed, in our work some data (such as the sentiments embedded in the reviews, and the quality of reviews) are not directly observable and are inferred from the underlying data (reviews). Our focus is to identify appropriate methods to "recover" those latent data so that accurate prediction can be made. This paper is built upon our previous work on the predictive power of sentiments [3]. In particular, we have extended the ARSA model proposed in [3], and incorporated the important factor of review quality into the model. We have also considered how to predict the quality of reviews using content features, and conducted more experiments to evaluate the effectiveness of the proposed models.

V. PROPOSED MODELS S-PLSA: A PROBABILISTIC APPROACH TO SENTIMENT MINING

In this section, we propose a probabilistic approach to analyzing sentiments in reviews, which will serve as the basis for predicting sales performance.

A. *Feature Selection*

We first consider the problem of feature selection, i.e., how to represent a given review as an input to the mining algorithms. The traditional way to do this is to compute the (relative) frequencies of various words in a given document (review) and use the resulting multidimensional feature vector as the representation of the document. For some features, we divide products and reviews into three types based on their average ratings (rating scale: 1 to 5):

Good (rating ≥ 4), *bad* (rating ≤ 2.5) and *Average*, otherwise

Instead of using the frequencies of all the words appearing the reviews, we choose to focus on the set containing 2,030 appraisal words extracted from the lexicon constructed by Whitelaw et al. [7], and construct feature vectors based on their frequencies in reviews. The rationale behind this is that for sentiment analysis, sentiments oriented words, such as "good" or "bad," are more indicative than other words [7]. It is noted in [7] that ". . . the appraisal taxonomies used in this work are general purpose, and were not developed specifically for sentiment analysis or movie review classification." Therefore, we consider the appraisal groups developed by Whitelaw et al. a good fit to our problem and the same lexicon can be applied to other domains as well.

B. *Sentiment PLSA*

Mining opinions and sentiments present unique challenges that cannot be handled easily by traditional text mining algorithms. This is mainly because the opinions and sentiments, which are usually written in natural languages, are often expressed in subtle and complex ways. Moreover, sentiments are often multifaceted, and can differ from one another in a variety of ways, including polarity, orientation, graduation, and so on. Therefore, it would be too simplistic to just classify the sentiments expressed in a review as either positive or negative. For the purpose of sales prediction, a model that can extract the sentiments in a more accurate way is needed. To this end, we propose a probabilistic model called Sentiment Probabilistic Latent Semantic Analysis (S-PLSA), in which a review can be considered as being generated under the influence of a number of hidden sentiment factors. The use of hidden factors allows us to accommodate the intricate nature of sentiments, with each hidden factor focusing on one specific aspect of the sentiments. The use of a probabilistic generative model, on the other hand, enables us to deal with sentiment analysis in a principled way. In its traditional form, PLSA [6] assumes that there are a set of hidden semantic factors or aspects in the documents, and models the relationship among these factors, documents, and words under a probabilistic framework. With its high flexibility and solid statistical foundations, PLSA has been widely used in many areas, including information retrieval, Web usage mining, and collaborative filtering. Nonetheless, to the best of our knowledge, we are the first to model sentiments and opinions as a mixture of hidden factors and use PLSA for sentiment mining.

S-PLSA is a latent variable model for co-occurrence data. S-PLSA models the co-occurrence probability as a mixture of conditionally independent multinomial distributions

$$\Pr(b, w) = \sum_{z \in Z} \Pr(z)\Pr(w|z)\Pr(b|z)$$

VI. PERFORMANCE ANALYSIS

A. *Incorporating Review Quality*

In the SPEC, we define four categories of review quality which represent different values of the reviews to users' purchase decision: "best review", "good review", "fair review", and "bad review". A generic description of the SPEC is as follows: A *best* review must be a rather complete and detailed comment on a product. It presents several aspects of a product and provides convincing opinions with enough evidence. Usually a best review could be taken as the main reference that users only need to read before making their purchase decision on a certain product. The first review in A *good* review is a relatively complete comment on a product, but not with as much supporting evidence as necessary. It could be used as a strong and influential reference, but not as the only recommendation.

To make a better estimation of future revenue, we proposed a method that can quantitatively evaluate the review quality; we developed the ARSQA model which explicitly incorporates the review quality factor into the ARSA model. To verify the effectiveness of these methods, we first evaluate the performance of our quality prediction model, and then compare the ARSQA model with the ARSA model.

To verify the effectiveness of using ε-Support Vector Regression to predict review quality, we compare it with the conventional Linear Regression (LR) model. To this end, we first formulate feature vectors in the same way, and feed them into two approximators, respectively. We

then compare their performance in terms of the squared correlation coefficient r2 and mean squared error σ2. Let us denote the approximate' output for review i is μi, and the true helpfulness value is μi. We have
Squared correlation coefficient

$$r^2 = \frac{\left(\sum_{i=1}^{n}(\mu_i - \bar{\mu})(\hat{\mu_i} - \bar{\hat{\mu}})\right)^2}{\sum_{i=1}^{n}(\mu_i - \bar{\mu})^2 \sum_{i=1}^{n}(\hat{\mu_i} - \bar{\hat{\mu}})^2}.$$

Mean squared error

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(\mu_i - \hat{\mu_i})^2.$$

As shown in Table 1, ε-Support Vector Regression demonstrates clear advantage over LR irrespective of the two evaluation metrics. This might be because linear approximator is not reliable if the true relation between the inputs and the output is nonlinear, although it may enjoy the benefit of being straightforward with a lower computational cost.

Approximator	r^2	σ^2
Linear regression	0.0865	0.0722
ϵ-SVR	0.2698	0.0612

Table .1 Performance Comparisons of Two Approximations

In this study, fig .1, we have exploited the writing style to help assess the quality of a review. In fact, we have also considered other possible factors that may affect the helpfulness values, including the length of the review, the polarity of the review, the number of responses the review received, the subjectivity of the review, and the average rating of all reviews on the movie. Some of these factors have been studied in the previous literature to measure the quality of product reviews, e.g., reviews on digital cameras and MP3 players posted on commercial websites such as Amazon and Ebay [2], [36]. Using a set of experiments, we investigate if they are effective indicators of review quality in the movie domain.

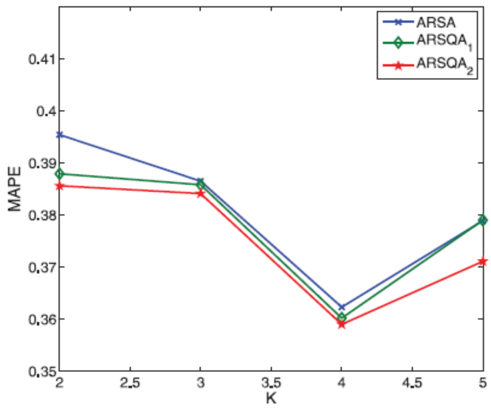

Fig.1.Incorporating quality

VII. CONCLUSION AND FUTURE WORK

This problem as a domain-driven task, and managed to synthesize human intelligence (e.g., identifying important characteristics of movie reviews), domain intelligence (e.g., the knowledge of the "seasonality" of box office revenues), and network intelligence (e.g., online reviews posted by moviegoers). The outcome of the proposed models leads to actionable knowledge that we can readily employed by decision makers. Further considered the role of review quality in sales performance prediction, and proposed a model to predict the quality rating of a review. The accuracy and effectiveness of the proposed models have been confirmed by the experiments on two movie data sets. The accuracy and effectiveness of the proposed models have been confirmed by the experiments on two movie data sets. Equipped with the proposed models, companies will be able to better harness the predictive power of reviews and conduct businesses in a more effective way. Also note that the ARSA and ARSQA models are general frameworks for sales performance prediction, and would certainly benefit from the development of more sophisticated models for sentiment analysis and quality prediction.

REFERENCES

[1] D. Gruhl, R. Guha, R. Kumar, J. Novak, and A. Tomkins, "The Predictive Power of Online Chatter," Proc. 11th ACM SIGKDD Int'l Conf. Knowledge Discovery in Data Mining (KDD), pp. 78-87, 2005.
[2] A. Ghose and P.G. Ipeirotis, "Designing Novel Review Ranking Systems: Predicting the Usefulness and Impact of Reviews,"Proc. Ninth Int'l Conf. Electronic Commerce (ICEC), pp. 303-310, 2007.
[3] Y. Liu, X. Huang, A. An, and X. Yu, "ARSA: A Sentiment-Aware Model for PredictingSales Performance Using Blogs," Proc. 30th Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR), pp. 607-614, 2007.
[4] Y. Liu, X. Yu, X. Huang, and A. An, "Blog Data Mining: The Predictive Power of Sentiments," Data Mining for Business Applications, pp. 183-195, Springer, 2009.
[5] D. Gruhl, R. Guha, D. Liben-Nowell, and A. Tomkins, "Information Diffusion through Blogspace," Proc. 13th Int'l Conf. World Wide Web (WWW), pp. 491-501, 2004.
[6] T. Hofmann, "Probabilistic Latent Semantic Analysis," Proc. Uncertainty in Artificial Intelligence (UAI), 1999.
[7] C.Whitelaw, N. Garg, and S. Argamon, "Using Appraisal Groupsfor Sentiment Analysis," Proc. 14th ACM Int'l Conf. InformationandKnowledgeManagement (CIKM), pp. 625-631, 2005.
[8] W. Enders, Applied Econometric Time Series, second ed. Wiley, 2004.
[9] L. Cao, C. Zhang, Q. Yang, D. Bell, M. Vlachos, B. Taneri, E. Keogh, P.S. Yu, N. Zhong, M.Z. Ashrafi, D. Taniar, E. Dubossarsky, and W. Graco, "Domain-Driven, Actionable Knowledge Discovery," IEEE Intelligent Systems, vol. 22, no. 4, pp. 78-88, July/Aug. 2007.

[10] L. Cao, Y. Zhao, H. Zhang, D. Luo, C. Zhang, and E.K. Park, "Flexible Frameworks for Actionable Knowledge Discovery," IEEE Trans. Knowledge and Data Eng., vol. 22, no. 9, pp. 1299- 1312, Sept. 2009.
[11] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs Up? Sentiment Classification Using Machine Learning Techniques,"Proc. ACL-02 Conf. Empirical Methods in Natural Language Processing (EMNLP)
[12] B. Pang and L. Lee, "A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts," Proc. 42nd Ann. Meeting on Assoc. For Computational Linguistics (ACL), pp. 271-278, 2004.
[13] J. Kamps and M. Marx, "Words with Attitude," Proc. First Int'l Conf. Global WordNet, pp. 332-341, 2002.
[14] P.D. Turney, "Thumbs Up or Thumbs Down?: Semantic Orientation Applied to Unsupervised Classification of Reviews," Proc. 40th Ann. Meeting on Assoc. For Computational Linguistics (ACL), pp. 417-424, 2001.
[15] B. Pang and L. Lee, "Seeing Stars: Exploiting Class Relationships for Sentiment Categorization with Respect to Rating Scales," Proc. 43rd Ann. Meeting on Assoc. For Computational Linguistics (ACL), pp. 115-124, 2005.
[16] Z. Zhang and B. Varadarajan, "Utility Scoring of Product Reviews," Proc. 15th ACM Int'l Conf. Information and Knowledge Management (CIKM), pp. 51-57, 2006.
[17] B. Liu, M. Hu, and J. Cheng, "Opinion Observer: Analyzing and Comparing Opinions on the Web," Proc. 14th Int'l Conf. World Wide Web (WWW), pp. 342-351, 2005.
[18] N. Archak, A. Ghose, and P.G. Ipeirotis, "Show Me the Money!: Deriving the Pricing Power of Product Features by Mining Consumer Reviews," Proc. 13th ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining (KDD), pp. 56- 65, 2007.
[19] J.A. Chevalier and D. Mayzlin, "The Effect of Word of Mouth on Sales: Online Book Reviews," J. Marketing esearch, vol. 43, no. 3,pp. 345-354, Aug. 2006.
[20] C. Dellarocas, X.M. Zhang, and N.F. Awad, "Exploring the Value of Online Product Ratings in Revenue Forecasting: The Case of Motion Pictures," J. Interactive Marketing, vol. 21, no. 4, pp. 23-45, 2007.
[21] S. Rosen, "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition," J. Political Economy, vol. 82, no. 1, pp. 34-55, 1974.

