Show me the Money! Deriving the Pricing Power of Product Features by Mining Consumer Reviews

Nikolay Archak
narchak@stern.nyu.edu

Anindya Ghose
aghose@stern.nyu.edu

Panagiotis G. Ipeirotis
panos@nyu.edu

Department of Information, Operations, and Management Sciences
Leonard N. Stern School of Business, New York University

ABSTRACT

The increasing pervasiveness of the Internet has dramatically changed the way that consumers shop for goods. Consumer-generated product reviews have become a valuable source of information for customers, who read the reviews and decide whether to buy the product based on the information provided. In this paper, we use techniques that decompose the reviews into segments that evaluate the individual characteristics of a product (e.g., image quality and battery life for a digital camera). Then, as a major contribution of this paper, we adapt methods from the econometrics literature, specifically the hedonic regression concept, to estimate: (a) the weight that customers place on each individual product feature, (b) the implicit evaluation score that customers assign to each feature, and (c) how these evaluations affect the revenue for a given product. Towards this goal, we develop a novel hybrid technique combining text mining and econometrics that models consumer product reviews as elements in a tensor product of feature and evaluation spaces. We then impute the quantitative impact of consumer reviews on product demand as a linear functional from this tensor product space. We demonstrate how to use a low-dimension approximation of this functional to significantly reduce the number of model parameters, while still providing good experimental results. We evaluate our technique using a data set from Amazon.com consisting of sales data and the related consumer reviews posted over a 15-month period for 242 products. Our experimental evaluation shows that we can extract actionable business intelligence from the data and better understand the customer preferences and actions. We also show that the textual portion of the reviews can improve product sales prediction compared to a baseline technique that simply relies on numeric data.

Categories and Subject Descriptors

I.2.7 [**Artificial Intelligence**]: Natural Language Processing—*text analysis*; H.2.4 [**Database Management**]: Systems—*Textual databases*; H.2.8 [**Database Applications**]: Data mining

KDD'07, August 12–15, 2007, San Jose, California, USA.

General Terms

Algorithms, Measurement, Economics, Experimentation

Keywords

consumer reviews, e-commerce, econometrics, electronic commerce, electronic markets, hedonic analysis, Internet, opinion mining, product review, sentiment analysis, text mining, user-generated content

1. INTRODUCTION

Consumer product reviews are now widely recognized to have a significant impact on consumer buying decisions [6]. Moreover, prior research on consumer decision making has established that consumer-generated product information on the Internet attracts more product interest than vendor information [2]. In contrast to product descriptions provided by vendors, consumer reviews are, by construction, more user-oriented: in a review, customers describe a product in terms of usage scenarios and evaluate the product from a user's perspective [4]. Despite the subjectivity of consumer evaluations in the reviews, such evaluations are often considered more credible and trustworthy by customers than traditional sources of information [2].

The rapid growth of the number of consumer reviews on the Web gave birth to several interesting opinion mining problems. Early work in this area was targeted primarily at evaluating the polarity of the reviews: review sentiments were classified as positive or negative by looking for occurrences of specific sentiment phrases. Different sources of sentiment phrases were proposed, including manually constructed dictionaries [7], WordNet [15], and search engine hit counts [28]. Machine learning methods were also applied to sentiment-based classification of consumer reviews [22]; all methods performed relatively well but failed to achieve high accuracy that is typical for topic-based document classification. Results suggested that sentiment classification of consumer reviews is complicated since consumers may provide a mixed review, praising some aspects of a product but criticizing others. Such heterogeneity stimulated additional research on identifying product features on which consumers expressed their opinions [9, 15–17, 26, 27]. After identifying the product features, it is then possible to use identification techniques to extract consumer opinions about each feature [3, 8, 15, 23].

Ultimately, though, we want to identify not only the opinions of the customers, but also want to examine the importance of these opinions. What features do customers value most? What is the relative importance of battery life vs.

image quality in a digital camera? Customers may praise or criticize the zoom capabilities of a digital camera, but these discussions may not really affect their decision to buy the product. Current work in opinion mining has not focused on capturing such behavior. Furthermore, current opinion mining systems cannot capture reliably the pragmatic meaning of the customer evaluations. For example, is "*good battery life*" better than "*nice battery life*"? How can we define an *objective* measure for ranking evaluations?

Towards answering questions of this nature, we propose utilizing the economic context in which the opinions are evaluated to estimate both the intensity and the polarity of the opinion. In particular, we investigate how product feature evaluations contained in consumer reviews affect product demand. Then, by tracing the respective changes in demand, we derive both the weight of the different product features, and the pragmatic "meaning" of these evaluations, providing actionable intelligence to the manufacturers that are trying to understand consumer behavior. We illustrate the intuition behind our approach with this (simplified) example:

EXAMPLE 1.1. *Consider two reviews r_A and r_B for two similar digital cameras A and B on Amazon.com: review r_A states that the "*lenses*" of camera A are "*excellent*", while review r_B states that the "*lenses*" of camera B are "*good*". To understand both the polarity and the strength of these evaluations, we observe the changes in sales associated with these reviews, all other things being equal. We observe that the evaluation "*excellent lenses*" increases sales by 5%, while the evaluation "*good lenses*" causes sales to drop by 1%. Therefore, we assign the score +5% to the evaluation "*excellent*" and the score -1% to the evaluation "*good.*" In a similar manner, we also observe that the evaluations "*excellent battery life*" and "*good battery life*" cause a respective change in sales of +2.5% and -0.5%. By comparing the sales changes associated with the evaluation of "*battery life*" with the respective changes associated with "*lenses,*" we infer that the weight of the "*lenses*" feature is twice as high as the weight of "*battery life.*"* □

Our approach is conceptually similar to the *hedonic regressions* that are commonly used in econometrics to identify the weight of individual features in determining the overall price of a product. For example, hedonic regressions are used to identify the marginal value of an extra megapixel in digital cameras.[1] As an important research contribution, we show how to incorporate in a hedonic-like framework *qualitative* features, such as "ease of use" or "image quality," that are not directly measurable and are ignored in existing economics and marketing research. Towards adapting a hedonic-like framework, we model qualitative consumer opinions as quantitative evaluation scores and then model consumer reviews as elements in a tensor product of feature space and evaluation space. To study the impact of consumer reviews on the product demand, we represent review impact as a linear functional in this tensor product. Finally, we show how to use a rank constraint for this functional to significantly reduce the number of model parameters, while still providing good experimental results. Our experimental evaluation over a real data set of 242 products monitored over a 15-month period on Amazon.com shows the validity of our approach and provides significant insights on the behavior of consumers who buy products online.

The rest of the paper is organized as follows. Section 2 gives the background for the paper. Section 3 presents our hybrid model that combines econometrics and text mining for analyzing consumer reviews using product demand. Section 4 discusses our experimental setting and results, obtained over two big product categories. Finally, Section 5 discusses related work and Section 6 provides further discussion and concludes the paper.

2. BACKGROUND

In this section, we give the necessary background for this paper. In Section 2.1, we give a brief description of hedonic regressions, which can be considered as the microeconomic framework in which our model is embedded in. Then, Section 2.2 describes existing techniques for product feature identification from consumer reviews and Section 2.3 briefly discusses existing approaches for identifying consumer opinions from product reviews.

2.1 Hedonic Regressions

The hedonic model assumes that differentiated goods can be described by vectors of objectively measured features and the consumer's valuation of a good can be decomposed into implicit values of each product feature [24]. Hedonic models are designed to estimate the value that different product aspects contribute to a consumer's utility. Implicit in the hedonic price framework is the assumption that a particular product can be viewed as consisting of various bundles of a small number of characteristics or basic attributes [1]. For instance, a backpacking tent can be decomposed to characteristics such as weight (w), capacity (c), and pole material (p), and the tent utility can be represented as a function $u(w, c, p, \ldots)$. Various restrictions can be put on the utility function. For example, it is often assumed that the utility function is monotonic, increasing, and concave to satisfy the "law of diminishing marginal utility" [25]. In brief, the hedonic hypothesis is based on the notion that heterogeneous goods are an aggregation of individual characteristics. It is important to point out that not all product categories are consistent with the assumptions of the hedonic framework. For example, while consumer appliances fit well into the hedonic model, products such as movies or books do not have clear utilitarian characteristics and are not typically classified as hedonic products. In order to avoid potential complexities arising from applying the model to non-hedonic products, in this paper we consider hedonic goods only.

The primary weakness of the existing hedonic models is the need to identify manually product features and measurement scales for them. In traditional hedonic regressions (i.e., those used by the BLS[2] to calculate the CPI) the decision of what features to include and how to measure them is made by individuals. This can lead to a bias from subjective judgments of these individuals. Furthermore, features that cannot be easily measured (e.g., image quality, style) are commonly ignored.

2.2 Product Feature Identification

In an orthogonal direction to economic research, the problem of identifying product features has been studied extensively in the last few years in the data mining and natural language processing communities. Many techniques use

[1] Hedonic regressions are also commonly used by the U.S. Bureau of Labor Statistics (BLS) to take into account product quality adjustments when recalculating the consumer price index (CPI).

[2] U.S. Bureau of Labor Statistics.

a part-of-speech tagger to annotate each review word with its part-of-speech (POS), identifying whether the word is a noun, an adjective, a verb and so on. Nouns and noun phrases are popular candidates for product features, though other constructs (like verb phrases) can be used as well. Alternative techniques search for statistical patterns in the text, for example, words and phrases that appear frequently in the reviews [8]. For example, Hu and Liu [16] use association rule mining to find frequent n-grams[3] in consumer product reviews and use these n-grams as candidate features. Hybrid methods are also developed by combining both approaches, where a POS-tagger is used as a preprocessing step before applying association mining algorithm to discover frequent nouns and noun phrases [19, 23]. An alternative technique for discovering product features that are not explicitly mentioned in the review is to use a classifier [9] that determines whether a particular feature is discussed (implicitly) in the review or not.

2.3 Mining Consumer Opinions

Of course, identifying product features *per se* is not the end goal. The important goal is to understand what the customer's opinion is about each of the identified product features. An increasingly popular trend in opinion mining is in combining feature mining and sentiment identification techniques to extract consumer opinions in form of feature-based summaries [3, 8, 15, 23]. Though technical details may differ, the proposed algorithms usually consist of three basic steps. In the first step, a feature mining technique is used to identify product features. In the second step, the algorithms extract sentences that give (positive or negative) opinions for a product feature. Finally, a summary is produced using the discovered information. Unfortunately, the existing approaches could not provide reasonable *quantitative* evaluations of product features. In most cases, the evaluation of a product feature was done in a binary scale (positive or negative). It is also possible to have a counting scale that computes the number of positive and negative opinion sentences for a particular feature; such counts can be used for feature-based comparison of two products (see, for example [19]). Such a comparison tool is undoubtedly useful for consumers using an online shopping environment. Unfortunately, such techniques fail to identify the strength of the underlying evaluations, and do not show the importance of the underlying feature in the consumer's purchasing process. In this paper we show how to address these issues by taking into consideration the economic context into which the opinions are being evaluated.

3. ECONOMETRIC OPINION ANALYSIS

In this section, we describe our econometric approach for modeling and analyzing consumer reviews. In Section 3.1, we present our approach for identifying the opinions of consumers about product features. Then, in Section 3.2, we show how to represent consumer reviews as elements of a tensor product space and in Section 3.3, we describe our econometric approach for measuring the weight of the individual product features, and the implicit evaluation score that a consumer review assigns to each feature.

3.1 Identifying Customer Opinions

In the first step of our approach, we need to identify the product features mentioned in the reviews and the respective consumer evaluations. Recall that in the hedonic model each product can be characterized by a vector of its features $X = (ef_1, \ldots, ef_n)$, where each vector element ef_i represents the quality level of the corresponding feature. In our current work, we rely on existing approaches to identify the product features in the text. Specifically, we assume that each of the n features can be expressed by a noun, chosen from the set of all nouns that appear in the reviews. For example, for a digital camera dimension 1 might be "*lens*", dimension 2 might be "*size*", dimension 3 might be "*shutter*".[4] Furthermore, to avoid overfitting, we only considered as features the nouns that appeared frequently in our data set. (We provide more details in Section 4.1.)

After identifying the product features, we need to identify the consumer evaluation of product feature quality. We rely on the observation [15, 29] that consumers typically use adjectives, such as "*bad*," "*good*," and "*amazing*" to evaluate the quality of a product characteristic. Therefore, we use a syntactic dependency parser to identify the adjectives that modify a noun that we have identified as product feature.

The result of this technique is a set of *noun-adjective* pairs that correspond to pairs of *product features* and their respective *evaluations*. We refer to these pairs as *opinion phrases*. It should be noted that such opinion phrases cannot capture all opinion information contained in consumer reviews. We would like to emphasize though that the goal of this paper is not to develop new techniques for identifying product features or phrases that evaluate such features.[5] Rather, we are interested in measuring the weight that customers place in each product feature and the implicit evaluation score (polarity and strength) associated with each feature evaluation. The core contributions of this paper are described next.

Var	Dim	Description
$\mathscr{F}$		feature space
$\mathscr{E}$		evaluation space
$\mathscr{R}$		review space
$\mathscr{V}$		basis of review space
D_{kt}	$\mathbb{R}$	demand for the product k at time t
α	$\mathbb{R}$	constant term
β	$\mathbb{R}$	price elasticity of demand
p_{kt}	$\mathbb{R}$	price for the product k at time t
$\mathbf{W_{kt}}$	$\mathbb{R}^{mn}$	consumer review component, see Eq. 1
γ	$\mathbb{R}^m$	vector of evaluation weights
δ	$\mathbb{R}^n$	vector of feature weights, $\|\|\delta\|\|_2 = 1$

Table 1: Notation and symbol descriptions

3.2 Structuring the Opinion Phrase Space

Existing consumer review mining approaches tend to consider extracted product features and opinions as simple sets

[3] An n-gram is a sequence of n consecutive words.

[4] Some researchers have noted that many technical terms are compound nouns [20, 26]. In particular, compound terms like "battery life", "image quality" and "memory card" are rated by end users as more helpful for choosing products than single noun features ("battery", "image(s)", "memory") [20]. Nevertheless, we observed that many compound terms without loss of information can be represented by a single noun only. For example, if the mining tool finds word "*battery*" in a digital camera review, we observed sure that "*battery life*" is the feature being evaluated. For this reason we didn't consider compound nouns as candidate product features and restricted ourselves to simple nouns only.

[5] Of course, our approach can clearly benefit from orthogonal advances in the topics of product feature identification, and in the extraction of the respective evaluation phrases.

and no algebraic structure is imposed on these sets. In order to use the concepts similar to hedonic regressions, we need to represent the product demand as a function from the space of consumer reviews. Such representation assumes a vector space structure on the set of feature opinions. To construct such structure, we build two vector spaces: one for product features and one for feature evaluations.

We model multiple sets of n product features as elements of a vector space over $\mathbb{R}$ with basis $f_1, \ldots, f_n$. We denote this *feature space* as $\mathscr{F}, (\dim \mathscr{F} = n)$. In the same way, we define a space of evaluations as a vector space over $\mathbb{R}$ with basis $e_1, e_2, \ldots, e_m$. We denote this *evaluation space* as $\mathscr{E}(\dim \mathscr{E} = m)$. Intuitively, when we represent a review r in the feature space $\mathscr{F}$, the representation contains the *weights* that customers assign to each product feature. Similarly, when we represent a review r in the evaluation space $\mathscr{E}$, the representation contains the *implicit evaluation scores* that the review assigns to a product feature (notice that the evaluations in the space $\mathscr{E}$ are not bound, yet, to any specific feature).

Now, based on the algebraic structures in the feature and the evaluation space, we can represent the opinion phrases and whole consumer reviews as elements of the *review space* $\mathscr{R}$ that we define as the *tensor product of the evaluation and feature spaces*:

$$\mathscr{R} = \mathscr{F} \otimes \mathscr{E}$$

Note that the set of opinion phrases $f_i \otimes e_j$ form a basis of the space $\mathscr{R}$ and we denote this basis as $\mathscr{V}$.

Now, each review can be represented as a vector in the review space $\mathscr{R}$: we only need to determine the weight that we assign to each dimension of $\mathscr{R}$ to represent the review. There are several different ways to determine the weight of an opinion phrase in a text corpus (e.g., *tf.idf* weights). In our work, we use a standard term frequency measure that discounts influence of longer reviews and calculates the weight of the opinion phrase *phrase* in review *rev* for product *prod* as:

$$w(phrase, rev, prod) = \frac{N(phrase, rev, prod) + s}{\sum_{y \in \mathscr{V}} (N(y, rev, prod) + s)} \quad (1)$$

where $N(y, r, p)$ is the number of occurrences of the opinion phrase y in the consumer review r for product p, and s is a "smoothing" constant.[6] Notice that using our convention, the sum of the weights of the opinion phrases within each review is equal to one, discounting the influence of longer reviews.

EXAMPLE 3.1. *Consider the following review for a digital camera: "*The camera is of high quality and relatively easy to use. The lens are fantastic! I have been able to use the LCD viewfinder for some fantastic shots... To summarize, this is a very high quality product.*" This review can be represented by the following element of the tensor product, assuming $s = 0$:*

$$\begin{aligned} 0.4 &\cdot (quality \otimes high) + \\ 0.2 &\cdot (use \otimes easy) + \\ 0.2 &\cdot (lens \otimes fantastic) + \\ 0.2 &\cdot (shots \otimes fantastic) \end{aligned}$$

Notice that each opinion phrase dimension has a weight coefficient determining its relative importance in the review. Since the opinion phrase quality-high *appears twice in the review, the weight of this dimension is 0.4, in contrast to all the other opinion phrases that appear only once and have weight equal to 0.2.* □

So far, we have discussed how to represent reviews in an algebraic form. While we could achieve a similar result by simply defining directly a space of opinion phrases and then represent each review as a vector in this space, we will see next that our tensor space approach has significant analytic advantages over this simpler approach. Specifically, we will see that the tensor space approach will allow us to estimate naturally the weight of each product feature and the implicit score of each evaluation, using relatively small training sets and avoiding the problems of data sparsity and of overfitting. Next, we describe our approach in detail.

3.3 Econometric model of product reviews

A simple econometric technique for modeling product demand as a function of product characteristics and its price, is the following simple linear model:

$$\ln(D_{kt}) = a_k + \beta \ln(p_{kt}) + \varepsilon_{kt}. \quad (2)$$

where D_{kt} is the demand for the product k at time t, p_{kt} is its price at time t, β is the price elasticity, and a_k is a product specific constant term which captures the unobserved product heterogeneity, such as differences in product characteristics and brand equity. The variable ε_{kt} represents a random disturbance factor which is usually assumed to be normally distributed, i.e., $\varepsilon_{kt} \sim N(0, \sigma^2)$. This is a classical ordinary-least-squares (OLS) regression with product level fixed effects and the parameters can be estimated using standard panel data methods [31].

A potential drawback of such a model is that it can not be used to evaluate separately different product characteristics because it mixes all product features into the single term a_k. Another limitation is time invariant nature of the product specific effect. Though the technical characteristics of the product typically stay the same during the product's life cycle, its popularity may change as a result of consumer evaluations of the product.

We extend the model using our original assumption that qualitative consumer reviews correspond to some quantitative evaluations of product characteristics. Such evaluations cannot be measured directly but can be learned from product demand fluctuations, if a sufficiently large data set is available. We replace the product specific effect a_k by a sum of two components:

$$a_k = \alpha + \Psi(\mathbf{W_{kt}}). \quad (3)$$

where we have

- a consumer review component $\mathbf{W_{kt}} \in \mathscr{R}$ that captures opinions contained in consumer product reviews for product k available at time t, including all reviews before time t, and
- a time and product invariant constant term α.

The vector $\mathbf{W_{kt}}$ (for fixed k and t) contains the weights of all opinion phrases that appeared in the consumer reviews for product k available at time t. (Essentially, we take the vectors that represent the consumer reviews posted by time t and "collapse" them into a single vector.) For simplicity,

[6] We set $s = 0.01$ in our paper

we compute these weights by averaging[7] the weights for each opinion phrase across all reviews posted until time t.

In Equation 3, we represented the impact of a set of consumer reviews on the product demand as some generic functional $\Psi : \mathscr{R} \to \mathbb{R}$. Similar to hedonic regressions, we assume that the functional form is linear and formally this can be written as $\Psi \in \mathscr{R}^*$ where $\mathscr{R}^*$ is the dual space of $\mathscr{R}$ (space of linear functionals). A linear functional from a tensor product of two vector spaces is just a bilinear form of two parameters, so Ψ is just a bilinear form of features and evaluations. Any linear functional can be written in the basis representation:

$$\Psi(\mathbf{W_{kt}}) = \sum_{phrase \in \mathscr{V}} \psi(x) \cdot w(phrase, reviews_t, product_k) = \sum_{i=1}^{n}\sum_{j=1}^{m} \psi(f_i \otimes e_j) \cdot w((f_i \otimes e_j), reviews_t, product_k).$$

where $\psi(x) = \psi(f_i \otimes e_j)$ is the value of the functional on the basis vector $f_i \otimes e_j$. Intuitively, the value of $\psi(x)$ is the influence of the opinion phrase $\langle f_i, e_j \rangle$, which in turn is a function of the weight of feature f_i (when evaluated by e_j) and of the implicit score that e_j assigns to f_i. Continuing the example from Section 3.2, we can write the value of the functional on the sample consumer review as:

$$0.4 \cdot \psi(quality \otimes high) + 0.2 \cdot \psi(use \otimes easy) + 0.2 \cdot \psi(lens \otimes fantastic) + 0.2 \cdot \psi(shots \otimes fantastic)$$

Using Equations 2 and 3, we have our extended linear model:

$$\ln(D_{kt}) = \alpha + \beta \ln(p_{kt}) + \Psi(\mathbf{W_{kt}}) + \varepsilon_{kt}$$

Unfortunately, this model has a very large number of parameters and would require a very large training set of product reviews to estimate. In the case where we have n product features and m evaluation words for N products with similar features, the number of model parameters will be $n \cdot m + N$. Since, we rarely have more than 30 reviews for a product, model overfitting is definitely a problem.

To alleviate this problem, we reduce the model dimension by placing a rank constraint on the matrix Ψ. In other words, we are not going to search for any linear function but only for those with low matrix ranks. The approach is based on the following observation.

THEOREM 3.1. **(Singular Value Decomposition)** *Any linear functional $\Psi \in \mathscr{R}^*$ such that $\operatorname{rank}\Psi = \operatorname{rank}(\psi_{ij}) \le p$ can be represented as $\sum_{i=1}^{p} \Delta_i \otimes \Gamma_i$ where $\Delta_i \in \mathscr{F}^*$ and $\Gamma_i \in \mathscr{E}^*$ and $||\Delta_i||_2 = 1$.*

COROLLARY 3.1. *Any linear functional $\Psi \in \mathscr{R}^*$ such that $\operatorname{rank}(\Psi) = \operatorname{rank}(\psi_{ij}) = 1$ can be represented as $\Delta \otimes \Gamma$ where $\Delta \in \mathscr{F}^*$ and $\Gamma \in \mathscr{E}^*$ and $||\Delta||_2 = 1$.*

Corollary 3.1 shows how to model features and evaluations as independent multiplicative components. We can replace a complex functional from $\mathscr{R}^*$ by a product $\Delta \otimes \Gamma$ of two simple functionals $\Delta \in \mathscr{F}^*$ and $\Gamma \in \mathscr{E}^*$. In other words, we assume that each coefficient $\psi(x)$ can be decomposed as a product of the feature component (defining importance of the feature) and the evaluation component (determining relative weight of the evaluation). For example:

$$\psi(shots \otimes fantastic) = \gamma(shots) \cdot \delta(fantastic)$$

The resulting approach is similar to ANOVA decomposition, which is frequently used to reduce model complexity for multidimensional splines and generalized additive models [14]. Our model, though, operates in different context because features and evaluations do not represent independent dimensions of consumer reviews: each evaluation word occurrence describes some feature and each feature word occurrence relates to some evaluation.

One limitation of this rank-1 approximation is the restriction that each adjective has one meaning, independently of the feature that it evaluates. While this restriction might seem too strict, we observed that it worked well for our setting. We should note though that our approach allows flexibility in the approximation approach. If we would like to assign two potential meanings for each adjective, we can use a rank-2 approximation, allowing two meanings per adjective and two potential weights for each feature (e.g., negative evaluations may have high influence when evaluating a feature, but positive evaluations may have low weight). In this case, the functional ψ would simulate a mixture model and have the form $\psi = \gamma_1\delta_1 + \gamma_2\delta_2$. (Of course, estimating such a model would require a larger number of training examples.) Similarly, we can use even higher rank approximations, allowing more meanings per adjective, but we would need more training data to avoid overfitting.

Using the rank-1 approximation of the tensor product functional, we rewrite the model of Equation 3.3 as:

$$\ln(D_{kt}) = \alpha + \beta \cdot p_{kt} + \gamma^T \cdot \mathbf{W_{kt}} \cdot \delta + \varepsilon_{kt}. \qquad (4)$$

where γ is a vector that contains n elements corresponding to the weight of each product feature, and δ is a vector that contains the implicit score that each adjective assigns to a product feature. The new model has only $m + n + 1 + N$ parameters instead of original $n \cdot m + 1 + N$, but this is achieved by sacrificing the linearity of the original model. Despite the loss of linearity, we can still search for the maximum likelihood estimate (MLE) of the model parameters. Assuming that ε_{kt} are independent and identically normally distributed residuals, the ML estimate is the same as the minimum of the least-squares fit function:

$$(\alpha^*, \beta^*, \gamma^*, \delta^*) = \operatorname{argmin} \left(\sum_{k,t} \left(\ln(D_{kt}) - \alpha - \beta \cdot p_{kt} - \gamma^T \cdot \mathbf{W_{kt}} \cdot \delta \right)^2 \right)$$

One of possible approaches to parameter estimation is to use the Newton-Rhapson method, or its variations, to search for the minimum. Unfortunately the Newton-Rhapson algorithm convergence suffers from numerical difficulties, in particular, ill-conditioning of the Hessian in the neighborhood of the solution. To solve the problem, we propose a different iterative algorithm. Our algorithm is based on the observation that if one of the vectors γ or δ is fixed, the Equation 4 represents a linear model. The steps of the algorithm are described below:

1. Set δ to a vector of initial feature weights.
2. Minimize the fit function by choosing the optimal evaluation weights (γ) assuming that the feature weights

[7] Instead of giving equal weight to each review, we also experimented with weighting schemes involving a time-discount factor (older reviews have smaller weights) and a "usefulness" factor. In the latter case, opinion weights were multiplied by a factor $(h+1)/(t+1)$ where h and t are the number of helpful and total votes for the parent review. In both cases, the results were qualitatively similar to those of the simpler model.

(δ) are fixed. If δ is fixed, the equation is just a usual linear model and γ can be estimated by ordinary-least-squares (OLS) or generalized-least-squares (GLS) estimators [12].

3. Minimize the fit function by choosing the optimal feature weights (δ) assuming that the evaluation weights (γ) are fixed. Again this can be done by using OLS or GLS estimator.

4. Repeat Step 2 and 3, until the algorithm converges.

Note that the algorithm described above (and the Newton-Rhapson algorithm as well) may converge to some local minimum of the fit function. To compensate for such potential errors, we repeat the execution of the algorithm with multiple, randomly generated, initial starting points.

Summary: In this section, we have presented our approach for estimating the weights of the different product features and the implicit scores that the adjectives assign to the product features. We have presented how to use a low-rank approximation of the review space to estimate these values, and presented an efficient algorithm for estimating the parameters of our model. Next, we present the experimental evaluation of our approach.

4. EXPERIMENTAL EVALUATION

In this section, we validate our techniques on a data set covering 242 products drawn from the product categories "*Audio & Video*" and *"Camera & Photo"* sold on Amazon. We first describe our data set (Section 4.1). Then, in Section 4.2, we describe how we run our text mining algorithm to identify the product features and the evaluations. In Section 4.3, we present the exact setup of our econometric-based technique, and, in Section 4.4 we discuss the experimental findings that show that our techniques can provide actionable business intelligence and can give useful insights on how customers behave when shopping online.

4.1 Data

We gathered data on a set of products using publicly available information at Amazon.com. The data set covered two different product categories: "Camera & Photo" (115 products) and "Audio & Video" (127 products). During a 15-month period (from March 2005 to May 2006), we have been collecting daily price and sales rank information for the products in our data set, using the API provided by Amazon Web Services. In the "Camera & Photo" category we had 31,233 observations and in the "Audio & Video" category we had 35,143 observations. Each observation contains the collection date, the product ID, the price on Amazon.com (which includes a possible Amazon discount), the suggested retail price, the sales rank of the product, and the average product rating according to the posted consumer reviews. Additionally, we used Amazon Web Services to collect the full set of reviews for each product. Each product review has a numerical rating on a scale of one to five stars, the date the review was posted and the actual text posted by the reviewer. We collected 1,955 reviews in the "Camera & Photo" category and 2,580 reviews in the "Audio & Video" category, overlapping with our sales rank observations. Thus, each product had about 20 reviews on average. Figures 1 and 2 show distributions of the number of consumer reviews per product in the "Audio & Video" and "Camera & Photo" categories. A few outlier products, with more than 50 reviews, are not displayed on the figures. For example, in the "Audio & Video" category Apple 1GB iPod Shuffle had 271 consumer reviews.

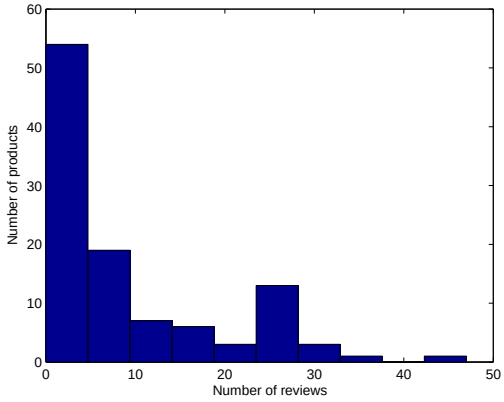

Figure 1: Distribution of the number of consumer reviews per product (Audio & Video)

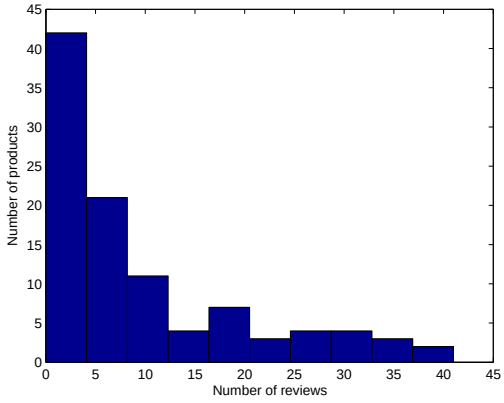

Figure 2: Distribution of the number of consumer reviews per product (Camera & Photo)

4.2 Selecting Feature and Evaluation Words

In order to determine the set of product features and evaluation words to use, we applied a simple approach consisting of three steps. First, we used the part-of-speech tagger developed by the Stanford NLP Group, to analyze all reviews and assign a part of speech tag to each word. Second, a set of frequent nouns was extracted from the tagged review corpus. The second step tends to produce a significant number of infrequent, non-relevant items. Fortunately, the set of frequent nouns was relatively small (about 800 items per category). Since, our work is not focused on product feature identification from product review, we decided to quickly process the list of frequent items, and manually select a subset of approximately 30 nouns to use as product features. For example, in the "Camera & Photo" category the set of features included "battery/batteries," "screen/lcd/display," "software," "viewfinder," "lens/lenses," and so on. To identify the set of evaluations, we used the syntactic dependency parser from the Stanford NLP toolkit, and we extracted the adjectives that evaluated the selected product features. The

set of evaluations contained popular adjectives describing the level of satisfaction with each particular feature. Words like "amazing," "bad," "great," "exceptional," and "outstanding" appeared frequently in our data set. We kept the list of 30 most frequent adjectives to create our evaluation space. We used the same set of evaluation words for both product categories, but, of course, the set of features was different in each category.

4.3 Experimental Setup

In order to evaluate our techniques using data from Amazon, we had to modify slightly the model of Section 3.3. Specifically, Amazon.com doesn't report the demand for the products available on the web site. Instead, Amazon.com reports a sales rank figure for each product, which ranks the demand for a product *relative to other products in its category*. Prior research in economics and in marketing (for instance, [5, 11]) has associated these sales ranks with demand levels for products such as software and electronics. The association is based on the experimentally observed fact that the distribution of demand in terms of sales rank has a Pareto distribution (i.e., a power law) [5]. Based on this observation, it is possible to convert sales ranks into demand levels using the following Pareto relationship:

$$\ln(D) = a + b \cdot \ln(S) \tag{5}$$

where D is the unobserved product demand, S is its observed sales rank, and $a > 0$, $b < 0$ are industry-specific parameters. Therefore, we can use the log of product sales rank on Amazon.com as a proxy of the log of product demand. We also modified the model of Section 3.3 to include both the suggested retail price (P_1) and the price on Amazon.com (P_2) since its natural to expect the prices will influence product demand, besides word-of-mouth. Furthermore, we included the review rating variable (R), which represented the average numeric rating of the product as given by the reviews.

Our model predicts the product sales rank based on its price, average review rating, and a set of consumer reviews. Note that as defined before, the review-level variables are the ones published before a given observation date.[8] Using our model, we analyze how different feature-evaluation combinations affect sales, after controlling for product price. The transformed equation is given below:

$$\begin{aligned}\ln(S_{kt}) = \alpha + \beta_1 \cdot R_{kt} + \beta_2 \cdot \ln(P_{1kt}) + \beta_3 \cdot \ln(P_{2kt}) + \\ \sum_{i=1}^{m}\sum_{j=1}^{n} W_{ktij} \cdot \gamma_i \cdot \delta_j + \varepsilon_{kt} \\ = \alpha + \mathbf{y_{kt}} \cdot \beta + \gamma^T \cdot \mathbf{W_{kt}} \cdot \delta + \varepsilon_{kt}\end{aligned} \tag{6}$$

In the above equation, $\mathbf{W_{kt}}$ is the "review matrix" (see Section 3.3) and W_{ktij} was calculated using Equation 1. We found that due to sparsity of the review data if we use a large number of features and evaluations, our model tends to overfit the noise in the training data set. To alleviate this concern, we added the regularization constraint $\left(\sum_{i=1}^{m} \gamma_i^2\right)\left(\sum_{j=1}^{n} \delta_j^2\right) \leq \lambda$, which transformed this linear model to a "ridge regression" [14]. Note that due to negligible correlation[9] between the list and retail prices, multicollinearity is not a concern in our model. Our results are also robust (and similar to the reported ones) if we were to only use the retail price in Equation 6.

4.4 Experimental Results

Predicting Future Sales: After obtaining the review matrix, we used 10-fold cross-validation to test the model's performance and obtain reliable coefficient estimates. The original sample was partitioned into 10 sub-samples of products.[10] Of the 10 sub-samples, a single sub-sample was retained as validation data for testing the model, and the remaining 9 sub-samples were used as training data. The cross-validation process was repeated 10 times and each of the 10 sub-samples was used exactly once as validation data.

For our experiments, we run our regression on the training data and derived the coefficients for each of the regressors. After deriving the coefficients from the training set, we tested how well the derived regressions predicts the sales rank for the products in the test set. We compared performance for predicting sales rank, using the Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) metrics, averaged over all the observations in the test set. We compared our model with the same model that has no consumer review text available (but including all other variables). We observed a 5% improvement in RMSE and 3% improvement in MAE. To test for statistical significance we used the Wilcoxon[11] signed rank test ($p < 0.001$). We should note that the 5% improvement is a significant achievement, given the high volatility of Amazon sales rank information, the fact that the numeric rating information is already taken into account (both models contained average rating of consumer reviews as a variable) and the fact that we are measuring errors in log-scale. This result indicates that the text of the reviews contains information that influences the behavior of the consumers, and that the numeric ratings alone cannot capture the information in the text.

Estimating Product Feature Weights and Evaluation Scores: The other significant contribution of our work is the ability to identify the product feature weights and the evaluation scores associated with the adjectives, within the context of an electronic market.

Tables 2 and 3 present the average feature and evaluation coefficients that were obtained for the 49 digital cameras[12] from the "Camera & Photo" category. The coefficients were averaged across cross-validation runs to produce reliable estimates and empirical standard error values. We should also note that we excluded the rating variable from the regression, in order to obtain feature and evaluation coefficients that are not detrended with respect to numeric review rat-

[8] We should clarify that ours is not a hedonic demand estimation model. In hedonic demand estimations, the first step involves computing individual equilibrium trait prices based on estimates of a hedonic price function to find the market price of a trait in the bundle. The second step is to estimate an inverse demand or marginal bid function using the trait prices as the dependent variable. Typically two-stage least squares (2SLS) regressions are employed with the supplier traits being appropriate instruments for the endogenous variables in the marginal bid function. Our approach is nested in a framework that is similar in notion to a hedonic demand model but very different in implementation.

[9] This happens because in the two product categories that we analyze, the retail price on Amazon is not a fixed discount off the list price, unlike in the case of other products such as books where publishers typically give retailers a fixed discount off the list price for a wide variety of books.

[10] We did not split observations for the same product due to time-series effects within a product over time.

[11] We used a non-parametric test to avoid problems with non-normality of the underlying distributions.

[12] We separated the results for point-and-shoot from those for SLR cameras and camcorders to avoid possible heterogeneity issues. Due to space constraints we do not list the results for these products and for the "Audio & Video" category. The results were qualitatively similar.

Feature	Weight	Std.Err.
camera*	0.810	0.091
quality*	0.484	0.106
battery*	0.192	0.048
resolution*	0.129	0.018
size	0.096	0.063
color	0.086	0.052
photos	0.074	0.040
lens	0.046	0.033
screen	0.037	0.037

Table 2: Feature weights for the *Camera & Photo* category. Higher weight values signify higher importance of each attribute. (The $*$ symbol indicates statistical significance at the 5% confidence level.)

Evaluation	Score	Std.Err.
great*	-2.460	0.353
good*	-1.693	0.211
best*	-0.914	0.154
excellent*	-0.442	0.180
perfect*	-0.433	0.146
nice	-0.006	0.051
decent	0.001	0.056
fantastic	0.085	0.050
bad*	0.206	0.038
amazing*	0.220	0.094
fine*	0.258	0.101
poor*	0.345	0.066

Table 3: Some evaluation scores for the *Camera & Photo* category. Higher scores mean increase in sales rank and are therefore *negative*. Lower values are better. (The $*$ symbol indicates statistical significance at the 5% confidence level.)

Phrase	Effect	Phrase	Effect
great camera	-0.4235	excellent photos	0.0040
good camera	-0.1128	nice size	0.0045
great quality	-0.0931	decent photos	0.0062
good quality	-0.0385	fantastic photos	0.0066
great battery	-0.0138	amazing resolution	0.0069
great size	-0.0060	amazing photos	0.0073
great photos	-0.0060	fine photos	0.0075
great resolution	-0.0052	excellent battery	0.0089
good battery	-0.0051	decent battery	0.0139
great lens	-0.0037	amazing battery	0.0164
good size	-0.0027	fine battery	0.0168
great color	-0.0023	best quality	0.0170
good photos	-0.0022	excellent quality	0.0507
good resolution	-0.0017	nice quality	0.0817
good lens	-0.0016	decent quality	0.0822
great screen	-0.0012	fantastic quality	0.0882
good color	-0.0004	amazing quality	0.0979
good screen	-0.0004	poor quality	0.1067
nice screen	0.0014	best camera	0.2026
excellent lens	0.0020	excellent camera	0.3936
excellent color	0.0027	perfect camera	0.3973
perfect size	0.0027	nice camera	0.5703
nice lens	0.0032	decent camera	0.5731
decent lens	0.0032	fantastic camera	0.6071
fantastic lens	0.0035	bad camera	0.6547
amazing lens	0.0038	amazing camera	0.6619
fine lens	0.0039	fine camera	0.6770

Table 4: Partial Effects for the *Camera & Photo* product category: A negative sign signifies decrease in sales rank and means higher sales.

ings. Note that higher values of evaluation coefficients signify increase of the sales rank and, therefore, have negative impact on product sales.

We should note that making conclusions based on the individual model coefficients may be incorrect, due to model complexity and interaction effects between coefficients. So, in addition to the raw coefficients, we also calculated the *partial effects* for frequent opinion phrases. The partial effect of a opinion phrase $x = e \otimes f$ is defined as $\gamma^T \cdot \bar{\mathbf{W}}^{\mathbf{x}}_{\mathbf{kt}} \cdot \delta - \gamma^T \cdot \bar{\mathbf{W}}_{\mathbf{kt}} \cdot \delta$ where $\bar{\mathbf{W}}_{\mathbf{kt}}$ is the "average review" obtained by averaging all consumer reviews in this product category with proper normalization and $\bar{\mathbf{W}}^{\mathbf{x}}_{\mathbf{kt}}$ is the "average review" where all evaluations of the feature f were replaced by the evaluation e. For example, to calculate a partial effect of the "solid lens" phrase ($f = lens$, $e = solid$) we take the "average digital camera review" and replace all evaluations of the "lens" feature to "solid". This enables us to simply calculate the difference between the modified review score and the average review score. The calculated partial effects for the "Camera & Photo" product category are presented in Table 4. Remember, that a negative sign on an opinion phrase denotes the fact that there is an increase in sales (since sales rank on Amazon is inversely proportional to demand) when that "opinion feature" is present in the content of the product review.

We observed that seemingly positive evaluations like "decent quality" or "nice/fine camera" end up hurting sales. Even though this may seem counterintuitive, it actually reflects the nature of an online marketplace: most of the positive evaluations contain superlatives, and a mere "decent" or "fine" is actually interpreted by the buyers as a lukewarm, slightly negative evaluation. Furthermore, we observed only a small number of unambiguously negative evaluations, such as "bad" and "horrible": products that get bad reviews, tend to disappear from the market quickly, and do not get many further negative evaluations. In general, the reviews that appear on Amazon are positive, especially for products with large number of posted reviews. Finally, a strange observations is that the evaluations "best camera," "excellent camera," "amazing camera," "perfect camera," and so on, have a negative effect on demand. This puzzling result may be caused by the fact that most of the reviews praising the camera in general do not usually have many details about the product features of the camera. Customers discount such reviews, and potentially treat them as bogus: a review that simply says that a camera is the "best camera" in the market, without going into other details does not provide useful information to the customer and has negative effect on sales.

Evaluation Conclusions: We have presented our experimental evaluation of our technique by presenting results on a real data set from Amazon.com. Our results show that we can identify the features that are important to consumers and that we can derive the implicit evaluation scores for each adjective, in an *objective* and *context-aware* manner. Deriving the polarity and strength of an evaluation was generally considered a hard problem, but by evaluating an opinion within an economic framework, we showed that we can solve the problem in a natural manner. Furthermore, our technique takes into consideration the peculiarities of the environment in which the opinion is evaluated. Weak positive opinions like *nice* and *decent* are actually evaluated in a negative manner in electronic markets, since customers are used to see only strong positive opinions, and a mere *nice* or *decent* is a bad signal to the buyers.

5. RELATED WORK

Our paper adds to a growing literature on sentiment analysis. Similar to almost any other sentiment mining technique, the first step involves selecting the set of features to use (see Section 2.2). Our approach to feature selection is very close to the one presented by Hu and Liu [15]. Hu and Liu used a POS-tagger and association miner to find frequent itemsets, which are then treated as candidate features. For each instance of a candidate feature in a sentence, the nearby adjective (if there is such) was treated as the *effective opinion*. Note that *effective opinions* used by Hu and Liu are direct counterparts of evaluation words in our study. A major difference of our approach is that whereas Hu and Liu were interested in the effectiveness of feature extraction and high recall values, we are concerned about gathering a small set of major features and evaluation words. In order to ensure that gathered features reflect hedonic characteristics of the products, we performed manual post-processing of frequent itemsets. Contrary to Hu and Liu, who performed additional research to identify infrequent features, we intentionally discarded such features to obtain robust estimates of model coefficients. It should also be noted that while Hu and Liu used WordNet to identify polarity of opinion words, our model evaluates polarity and strength of each evaluation directly from the regression on product demand.

Our research was inspired by previous studies about opinion strength analysis. Popescu and Etzioni [23] presented OPINE, an unsupervised information extraction system capable of identifying product features, identifying user opinions regarding product features, determining polarity of the opinions and ranking the opinions based on their strength. Unfortunately, so far, no paper was published explaining how OPINE solves the last task (opinion ranking). Evaluating strength of a sentiment quantitatively is even more challenging task than simple ranking and there were just a few papers on the topic published thus far. Wilson, Wiebe and Hwa [30] presented a supervised learning approach able to distinguish between weak, medium and strong subjectivity of the opinion. Unfortunately, this technique requires a manually annotated corpus of opinions. Human annotators might not able to distinguish strength of opinions on a finer-grained scale or estimate economic impact of each particular opinion. In a close stream of research, Pang and Lee [21] studied the rating-inference problem. They augmented SVM classifier with a metric labeling concept and applied the altered classifier to infer the author's numerical rating for Rotten Tomatoes movie reviews.

We also add to an emerging stream of literature that combines economic methods with text mining [7, 10, 18]. For example, Das and Chen [7] examined bulletin boards on Yahoo! Finance to extract the sentiment of individual investors about technology companies. They have shown that the aggregate tech sector sentiment predicts well the stock index movement, even though the sentiment cannot predict well the individual stock movements. There has also been related work on studying connections between online content such as blogs, bulletin boards and consumer reviews, and consumer behavior, in particular purchase decisions. Gruhl et al. [13] analyzed the correlation between online mentions of a product and sales of that product. Using sales rank information for more than 2,000 books from Amazon.com, Gruhl et al. demonstrated that, even though sales rank motion might be difficult to predict in general, online chatter can be used to successfully predict *spikes* in the sales rank.

To the best of our knowledge, our study is the first to use econometric approaches for analyzing the strength and polarity of consumer review opinions. Contrary to Gruhl et al. who used blogs as the indicator of spikes in the sales rank, our research is based on the premise that changes in sales rank, can be partially explained by the impact of consumer reviews. We draw on the results of Chevalier and Mayzlin [6] who have shown that online book ratings affect product sales by examining the relationship between relative market shares (sales rank) and consumer reviews across the two leading online booksellers, Amazon.com and BarnesandNoble.com. Comparing the sales and reviews of a given book across the two sites, allows to control for external factors that can affect the sales and word-of-mouth of both retailers (for example, release of a new Harry Potter movie may result in a temporary increase of the corresponding book sales). While our research is similar to the study by Chevalier and Mayzlin, it does differ in several significant ways. While the model in [6] included only numerical rating information (such as average star rating and fraction of 1 star and 5 star reviews), our approach employs text mining to extract consumer opinions about different product features. Contrary to the work in [6], we were not able to use multiple retailers in our experimental study. In order to alleviate possible concerns from the influence of external factors or measurement errors on our coefficient estimates, we used 10-fold cross validation of the model to precisely infer product-specific characteristics. We found that our results are robust across products.

6. CONCLUSIONS AND FUTURE WORK

We have presented a novel method for mining consumer reviews that combines existing text mining approaches with econometric techniques. The major novelty of our technique is that it allows an economic-aware analysis of the consumer reviews, identifying the weight that customers place on individual product features and the polarity and strength of the underlying evaluations. By using product demand as the objective function, we derive a context-aware interpretation of opinions and we show how customers interpret the posted comments and how they affect their choices. We have presented examples, where opinions that would be interpreted as positive by existing opinion mining systems, are actually negative within the context of electronic markets. The source code with our implementation, together with the data set used in this paper are available from http://economining.stern.nyu.edu.

Using a unique data set from a leading online retailer, Amazon, we have demonstrated the value of using economic data and econometric modeling for a quantitative interpretation of consumer reviews on the Internet. Our results can be used by manufacturers to determine which features contribute most to the demand for their product. Such information can also help manufacturers facilitate changes in product design over the course of a product's life cycle as well as help retailers decide on which features to promote and highlight in advertisements and in-store displays.

We believe that there is rich potential for future research. From the econometric point of view, future work can focus on pure price-hedonic regressions that highlight which feature of the product, both implicit and explicit, constitutes what proportion of the product price. Such an analysis can enable consumers to figure out exactly how much they are paying for each product attribute within a class of products.

It would also be interesting to examine how traditional hedonic studies can be enhanced when they incorporate qualitative features, and not just directly measurable parameters. From the text mining point of view, advances in algorithms for product feature identification, and for extraction of the associated evaluation phrases, can definitely improve the results of our technique. We plan to examine the limits of these improvements by analyzing manually product reviews and extracting the features that are discussed together with their evaluations. Another interesting direction that we plan to explore is the results of our technique when using higher rank approximations of the tensor space. Right now, our rank-1 approximation allows a single score for each evaluation and a single weight for a product feature. A natural extension is to examine how many different meanings we can allow for each phrase, without running into overfitting problems.

Overall, we believe that the interaction of economics research with data mining research can benefit tremendously both fields. Economic approaches can offer natural solutions to problems that seemed too hard to solve in a vacuum (e.g., determining the strength of an opinion). Similarly, data mining approaches can improve the current state of the art in empirical economics, where the focus has traditionally been on relatively smaller data sets.

Acknowledgments

We thank Rhong Zheng for assistance in data collection. This work was partially supported by a Microsoft Live Labs Search Award, a Microsoft Virtual Earth Award, and by NSF grants IIS-0643847 and IIS-0643846. Any opinions, findings, and conclusions expressed in this material are those of the authors and do not necessarily reflect the views of the Microsoft Corporation or of the National Science Foundation.

References

[1] Berndt, E. R. *The Practice of Econometrics: Classic and Contemporary*. Addison-Wesley, 1996.

[2] Bickart, B., and Schindler, R. M. Internet forums as influential sources of consumer information. *Journal of Interactive Marketing 15*, 3 (2001), 31–40.

[3] Carenini, G., Ng, R. T., and Zwart, E. Extracting knowledge from evaluative text. In *K-CAP'05: Proceedings of the 3rd International Conference on Knowledge Capture* (2005), pp. 11–18.

[4] Chen, Y., and Xie, J. Online consumer review: A strategic analysis of an emerging type of word-of-mouth. University of Arizona, Working Paper, 2004.

[5] Chevalier, J. A., and Goolsbee, A. Measuring prices and price competition online: Amazon.com and BarnesandNoble.com. *Quantitative Marketing and Economics 1*, 2 (2003), 203–222.

[6] Chevalier, J. A., and Mayzlin, D. The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research 43*, 3 (Aug. 2006), 345–354.

[7] Das, S. R., and Chen, M. Yahoo! for Amazon: Sentiment extraction from small talk on the web. Working Paper, Santa Clara University. Available at http://scumis.scu.edu/~srdas/chat.pdf, 2006.

[8] Dave, K., Lawrence, S., and Pennock, D. M. Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. In *Proceedings of the 12th International World Wide Web Conference (WWW12)* (2003), pp. 519–528.

[9] Ghani, R., Probst, K., Liu, Y., Krema, M., and Fano, A. Text mining for product attribute extraction. *SIGKDD Explorations 1*, 8 (June 2006), 41–48.

[10] Ghose, A., Ipeirotis, P. G., and Sundararajan, A. Opinion mining using econometrics: A case study on reputation systems,. In *Proceedings of the 44th Annual Meeting of the Association for Computational Linguistics (ACL 2007)* (2007).

[11] Ghose, A., and Sundararajan, A. Evaluating pricing strategy using ecommerce data: Evidence and estimation challenges. *Statistical Science 21*, 2 (2006), 131–142.

[12] Greene, W. H. *Econometric Analysis*, 5th ed. Prentice Hall, 2002.

[13] Gruhl, D., Guha, R., Kumar, R., Novak, J., and Tomkins, A. The predictive power of online chatter. In *Proceedings of the Eleventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2005)* (2005), pp. 78–87.

[14] Hastie, T., Tibshirani, R., and Friedman, J. H. *The Elements of Statistical Learning*. Springer Verlag, Aug. 2001.

[15] Hu, M., and Liu, B. Mining and summarizing customer reviews. In *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2004)* (2004), pp. 168–177.

[16] Hu, M., and Liu, B. Mining opinion features in customer reviews. In *Proceeding of the 2004 AAAI Spring Symposium Series: Semantic Web Services* (2004), pp. 755–760.

[17] Lee, T. Use-centric mining of customer reviews. In *Workshop on Information Technology and Systems* (2004).

[18] Lewitt, S., and Syverson, C. Market distortions when agents are better informed: The value of information in real estate transactions. Working Paper, University of Chicago, 2005.

[19] Liu, B., Hu, M., and Cheng, J. Opinion observer: Analyzing and comparing opinions on the Web. In *Proceedings of the 14th International World Wide Web Conference (WWW 2005)* (2005), pp. 342–351.

[20] Nakagawa, H., and Mori, T. A simple but powerful automatic term extraction method. In *COMPUTERM 2002: Second International Workshop on Computational Terminology* (2002), pp. 1–7.

[21] Pang, B., and Lee, L. Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL 2005)* (2005).

[22] Pang, B., Lee, L., and Vaithyanathan, S. Thumbs up? Sentiment classification using machine learning techniques. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2002)* (2002).

[23] Popescu, A.-M., and Etzioni, O. Extracting product features and opinions from reviews. In *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing (HLT/EMNLP 2005)* (2005), pp. 339–346.

[24] Rosen, S. Hedonic prices and implicit markets: Product differentiation in pure competition. *The Journal of Political Economy 82*, 1 (Jan.-Feb. 1974), 34–55.

[25] Samuelson, P. A., and Nordhaus, W. D. *Economics*, 18th ed. McGraw-Hill/Irwin, 2004.

[26] Scaffidi, C. Application of a probability-based algorithm to extraction of product features from online reviews. Tech. Rep. CMU-ISRI-06-111, Institute for Software Research, School of Computer Science, Carnegie Mellon University, June 2006.

[27] Snyder, B., and Barzilay, R. Multiple aspect ranking using the good grief algorithm. In *Proceedings of the Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics (HLT-NAACL 2007)* (2007).

[28] Turney, P. D. Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL 2002)* (2002), pp. 417–424.

[29] Turney, P. D., and Littman, M. L. Measuring praise and criticism: Inference of semantic orientation from association. *ACM Transactions on Information Systems 21*, 4 (Dec. 2003), 315–346.

[30] Wilson, T., Wiebe, J., and Hwa, R. Recognizing strong and weak opinion clauses. *Computational Intelligence 22*, 2 (May 2006), 73–99.

[31] Wooldridge, J. M. *Econometric Analysis of Cross Section and Panel Data*. The MIT Press, 2001.