Controversy is Marketing: 
Mining Sentiments in Social Media

Zhu Zhang
Department of Management Information Systems
University of Arizona
zhuzhang@email.arizona.edu

Xin Li
Department of Information Systems
City University of Hong Kong
xin.li@cityu.edu.hk

Abstract
We propose a set of novel quantitative measures for controversy derived from statistical properties of textual social media. We empirically establish strong positive correlations between social media controversy and sales performance across multiple datasets. The power of the newly proposed measures is further illustrated in a linear regression model for predicting product sales.

1. Introduction
Wherever there are people, there are diverse opinions on a given subject matter. With the proliferation of web-based e-commerce sites and social media sites, people publish, exchange, and spread subjective evaluations of products in unprecedented volume and speed. Such writing artifacts are typically filled with mixture of positive and negative sentiments, often referred to as controversy.

From a marketing perspective, controversy campaigns have become a key ingredient of the game, either in proactive or passive manner for the players.

Some marketers explicitly take control of such processes. Filmmaker Michael Moore, for example, says on his Web site that his new documentary, “SiCKO,” “will expose the health-care industry's greed and control over America's political processes.” Mr. Moore's work, and the backers of “SiCKO” hope that the new movie will stir up emotions and help generate the kind of buzz that made his last movie, “Fahrenheit 9/11,” both a topic of national debate and an unprecedented blockbuster in the documentary genre. “Fahrenheit 9/11” had a budget of $6 million and grossed more than $100 million in the U.S. alone. Mr. Moore's formula is simple: Pick a divisive topic and goad opponents into a public debate before the movie opens.

Another example is the exposure that “Sheetz - Made to Order” has gotten from the recent “Crispy Frickin' Sandwich” controversy. A lot of people had never heard of Sheetz before seeing the ad, especially those who don't live in the region where all of its stores are located. However, now it seems that the whole world knows its name thanks to the company’s controversial ad. It may be quite surprising that this ad has caused such a stir considering how “controversial” our pop culture and society itself have become, but Sheetz knew exactly what it was doing. Ruffling a few feathers always gets exposure, which means branding. Now many that live outside of the Sheetz region know what Sheetz is and will remember the name if they visit the region. And everyone in the region that is already familiar with Sheetz will have that Crispy Frickin' Sandwich on their mind and just may want to check out the meal that has everyone riled up.

Of course, not every company markets their products with the intention of proactively creating controversy. In the cases of the classical examples such as Microsoft Windows and Apple MP3tunes, which are highly controversial yet highly successful businesswise, it was not clear whether Microsoft and Apple planned for the controversy campaigns. On the other hand, regardless of the manufacturers’ intention, product-related controversy does pervasively exist, in particular in various forms of social media, e.g., consumer-generated reviews, blogs, forums, etc.

In this paper, we aim at computational modeling of controversy in social media, and subsequently investigating the correlation between social media controversy and product sales performance. We only consider technical controversy reflected in diverse consumer opinions, instead of issues at the ethic level.

2. Related work
In this section, we review existing literature in three areas relevant to the current study: marketing, social media, and text sentiment mining.
2.1. Marketing

Several prior marketing literatures have studied the effect of opposing opinions or Word-Of-Mouth (WOM).

In the marketplace, consumers often receive negative brand information in addition to advertising. [1] explored whether traditional models accurately predict ad processing and response when consumers integrate advertising with negative WOM communications about the brand. A test was conducted using four experimental groups: ad only, negative WOM only, advertising then negative WOM, and negative WOM then advertising. Results show that (a) advertising mitigates the detrimental cognitive effects of negative WOM communication (when the ad is processed first) and the detrimental affective and conative effects (when the ad is processed last), (b) integrating ad content with negative WOM communication causes significant changes in the message processing of both, (c) negative WOM communication significantly reduces the perceived credibility of advertising as well as brand attitudes and purchase intentions, and (d) the effect of attitude toward the ad on brand attitude becomes nonsignificant when subjects process both types of information. In short, negative WOM mostly exhibits a destructive effect on brand attitude.

In a slightly different context, negative opinions have been shown to play a somewhat constructive role. When controversial technological innovations are introduced, experts influence the diffusion rate and extent of acceptance of the innovations by acting as negative or positive opinion leaders. [2] studied Prosthodontists and their selection of opinion leaders, drawing on data from national and local samples of professionals to describe the use of non-precious alloys as substitutes for gold in dental restorations. One of the important observations was that anti-innovation experts (negative opinion leaders) eventually reinforce and legitimize rather than sway innovation adoption decisions, by hardening the information arteries.

Split views and findings have been seen on the issue of agent evaluation. On one hand, some studies (e.g., [3]) on word-of-mouth communications between consumers and agents have generally found a negativity effect for judgments of products, whereby consumers consider negative evaluations of a product as more informative than positive evaluations as inputs to product judgment. In an accessibility-diagonosticity perspective of persuasion, [4] argues along the same line. In particular, it points out that negative information tends to be more diagnostic or informative than positive or neutral information. Consequently, negative attribute information is often weighted heavily in consumer judgment. In contrast, [5] and [6] showed a positivity effect for judgments involving agents, whereby consumers consider their own previously loved (compared to hated) alternatives to be more diagnostic to agents about their tastes, and hence more useful as a basis for future agent advice. A more recent study [7] showed that moderation of the false consensus effect, whereby individuals overestimate the extent to which others share their opinions, is driven by the availability of countervalence attributes, that is, disliked attributes in liked alternatives and liked attributes in disliked alternatives. This may suggest the onset of the “mixture” view of opposing opinions, which is the spirit of our study. From a persuasion point of view, our belief is that a convincing argument is not necessarily a mono-color picture, but instead a meaningful “bag” of positive and negative reflections.

2.2. Social Media

Social media, as defined by Wikipedia (which by itself is a type of social media), is an umbrella term that defines the various activities integrating technologies, social interactions, and the construction of words, pictures, videos and audio. Social media can take many different forms, including Internet forums, message boards, weblogs, wikis, podcasts, pictures and videos. Examples of social media applications include Google Groups (reference, social networking), Wikipedia (reference), MySpace (social networking), Facebook (social networking), Last.fm (personal music), YouTube (social networking and video sharing), Second Life (virtual reality), Flickr (photo sharing), Twitter (social networking and microblogging), Epinions (opinion site), Digg (social news), and Upcoming.org (social event calendar).

In sociology, it has been commonly agreed upon that social media have changed our everyday life significantly [8, 9, 10, 11]. Within the computing-related community, various conferences, such as the new International Conference on Weblogs and Social Media (ICWSM), have been introduced to serve as forums for researchers interested in computational modeling of social media.
2.3. Mining Subjectivity and Sentiments in Text

Generally speaking, subjectivity in natural language refers to aspects of language used to express opinions, evaluations, speculations, and emotions. Subjective language can be learned and generalized from text corpora [12]. Subjectivity analysis has been performed at the sentence level [13] and phrase level [14], where the task is to distinguish subjective text units from objective ones. Text subjectivity analysis has also been shown to be useful for other natural language processing and information retrieval applications, such as information extraction [15] and multi-perspective question answering [16].

The importance of acquiring lexical clues for subjectivity analysis has been recognized; a number of reusable resources have been created, which are likely to be used by the proposed study. One of the important ones is SentiWordNet [17], to be used in our study.

Given a subjective text unit, sentiment analysis is concerned with the polarity and intensity of opinions found in text [18, 19, 20, 21]. A closely related area is affect analysis which attempts to analyze emotions instead of sentiments. Sample affect classes include happiness, sadness, anger, horror, etc. [22, 23]. Sentiment and affect analysis have many similarities, including the techniques and application domains. They are directly associated because sentiments and affects can and are often influenced by each other.

Sentiment analysis has been applied to numerous domains. Review analysis has been performed on movie, product, and music reviews [18, 19]. Web Discourse analysis evaluates various social media genres and assesses sentiments about specific issues/topics including abortion, gun control, and politics [24]. Many recent studies have analyzed sentiment and affect intensities in blogs [44] and USENET forums [25]. Sentiment analysis has also been applied to news articles [26] and political speeches [27].

In the context of mining product reviews, a product usually has multiple features, and they may be evaluated differently by customers. Shifting from classification to extraction, [28] introduced OPINE, an unsupervised information extraction system which mines reviews in order to build a model of important product features, their evaluation by reviewers, and their relative quality across products. In a similar effort, [29] proposed a framework for analyzing and comparing consumer opinions of competing products. A prototype system called Opinion Observer was also implemented.

3. Research Gaps and Research Questions

We identified the following research gaps based on the review of existing literature:
- The role of negative information/opinion has always been studied and presented as a competitive counterpart of positive information, with different effects on the dependent variable. There lacks a holistic view of the diversity of consumer opinions.
- There exists no quantitative measure of controversy in social media.
- There lacks a systematic study of correlation between social media controversy and product sales.
- Consequently, there is a lack of prediction models involving (particularly text-based) controversy measures

Based on the research gaps above, we intend to investigate the following research questions:
- Can we computationally measure social media controversy?
- Is there any correlation between social media controversy and product sales performance? If yes, what is the nature of the correlation?
- How can highly correlated variables contribute to the prediction of product sales performance?

4. Research Design

In this section, we first discuss the design of variables, including novel measures for text-mediated controversy. Correlation analysis and regression analysis will follow.

4.1. Variables

All variables are conceptualized in the context of online e-commerce websites, and further operationalized (whenever appropriate) with the expectation that empirical data will be collected from such sources.

Given a collection of products $P = \{p_1, ..., p_n\}$, and assume each product $p_i$ is rated and reviewed by $C_i$ customers. Therefore each $p_i$ is associated with two sets:
- A rating set $R_i = \{r_{i1}, ..., r_{iC_i}\}$, where each $r_{ij}$ is a positive real number, typically between 1 and 5 as in many e-commerce websites.
A review set \( V_i = \{ v_{i1}, \ldots, v_{ij}, \ldots, v_{iC_i} \} \), where each \( v_{ij} \) is a piece of evaluative text reflecting a consumer’s opinions on the given product, which can be null if the consumer chooses not to provide it. Such product reviews are again widely available in many e-commerce websites.

The items in the two sets correspond to each other pairwise, i.e. each rating \( r_{ij} \) corresponds to a review \( v_{ij} \).

### 4.1.1 Dependent variable.

The purpose of this study is to understand the connection between social media controversy and product sales performance. Therefore the dependent variable should serve as the proxy of the latter. The most realistic yet available approximation of such measures is product sales rank, as published by some e-commerce sites. Due to its ranking nature (“1” means the best sales performance; larger numbers indicate weaker sales performance), we take the opposite number of the actual sales rank (i.e., negating it) and still call it SalesRank in our study. Such a non-critical manipulation is only for the convenience of analysis and conformance to natural instinct, i.e., the higher the modified SalesRank, the stronger the product sales. The small comfort to be gained is that, in the later experiment section we do not have to say awkward sentences like “the correlation measure is negative, yet it means variable \( x \) has a positive influence on \( y \).”

### 4.1.2 Numeric-rating-based Independent Variables.

We define the following measures based on numeric ratings.

**Average Consumer Rating (AvgRating)**

This is what we use to measure the overall consumer reaction to a product, at a very coarse level. Given a product \( p_i \),

\[
\text{AvgRating}_i = \frac{1}{C_i} \sum_{j=1}^{C_i} r_{ij}
\]

**Standard Deviation in Consumer Rating (StdDevRating)**

This is used to quantify the variance of consumer opinions reflected in the numerical ratings only. Given a product \( p_i \),

\[
\text{StdDevRating}_i = \sqrt{\frac{\sum_{j=1}^{C_i} (r_{ij} - \text{AvgRating}_i)^2}{C_i}}
\]

### 4.1.3 Review-based variables.

In this category, we dive into the consumer-written reviews and define text-based measures.

To set the stage, an English word can be objective or subjective; if it is subjective, it can carry a positive or negative polarity, with certain intensity. For example, “fantastic” is a strong positive word, and “questionable” is a moderate negative word. And certainly there are a lot of words that are neutral in terms of polarity, e.g., “white.”

Formally, given a word \( w \), it is associated with two scores, a positivity score \( pscore \) and a negativity score \( nscore \), such that \( 0 \leq pscore, nscore \leq 1 \) and \( 0 \leq pscore + nscore \leq 1 \). If both scores are 0, the word is a neutral one. An excellent lexicon, SentiWordNet [17], encodes such linguistic knowledge.

A data mapping process is illustrated in Figure 1 through a simple example sentence. Notice that in the PSV, each word is accompanied by its \( pscore \) or \( nscore \), depending on its polarity. The purpose of such data transformations is to reveal the underlying sentiment distribution in the review text without involving any sophisticated natural language processing at this early stage of the study.

**Coarse text controversy measures**

When computing measures in this family, we concatenate all reviews (i.e., all members of \( V_i \)) for product \( p_i \) into a single piece of text, \( T_i \).

**Positivity Dominance (PosDom)**

We use it to measure the dominance of word positivity, defined as a ratio between collective word positivity and collective word negativity in \( T_i \).

\[
\text{PosDom}_i = \frac{\sum_{w_1 \in T_i} pscore_k}{\sum_{w_1 \in T_i} nscore_k}
\]

**Product-level Sentiment Entropy (SentEnt)**

Viewing each symbol in \( MS_i \) as a microstate, we can measure the entropy of the corresponding probability distribution \( PD_i \):

\[
\text{SentEnt}_i = \text{Entropy}(PD_i)
\]

It is named sentiment entropy since each microstate symbolizes one of 3 possible sentiment states:
0: the word is neutral (pscore=nscore=0) or almost neutral (pscore-nscore<0.1).
1: the word is positive (pscore-nscore≥0.1)
-1: the word is negative (nscore-pscore≥0.1)

By looking at the distribution of polarity states, this measure quantifies the amount of sentiment “uncertainty” in $T_i$.

**Sentiment Compactness (SentComp)**

In a similar line, we try to measure the sentiment compactness of $T_i$. We feed $MS_i$ into existing compression algorithms, and measure the size of the sequence before and after compression. The compactness is defined as a compression ratio:

$$\text{SentComp}_i = \frac{\text{size}(MS_i \text{ before compression})}{\text{size}(MS_i \text{ after compression})}$$

If the input sequence $MS_i$ has relatively less information (i.e., more compact), it is going to be more friendly to compression algorithms, thus yielding a high compression ratio. This explains why $\text{SentComp}$ is a good approximation of sentiment compactness.

In reality, we apply five existing compression algorithms (Zlib, Bzip2, LZV1, PPMd, LZF), and the resultant measures are named correspondingly as CompZlib, CompBzip2, CompLZV1, CompPPMd, and CompLZF.

**Finer-grained text controversy measures**

Instead of treating all review text for product $p_i$ as an entirety ($T_i$), we aim at quantifying the amount of controversy, or disagreement, by considering each individual review in the review set $V_i$. Roughly, we intend to use the textual information to recapture notions similar to $\text{StdDevRating}$.

In probability theory and information theory, The Kullback–Leibler divergence (abbreviated as KL divergence, also known as information divergence, information gain, or relative entropy) is a non-symmetric measure of the difference between two probability distributions. In our case, we use KL divergence to measure the difference between the underlying sentiment distributions of two pieces of review text.

Given two reviews $v_1$ and $v_2$, which are mapped into two microstate sequences $MS_1$ and $MS_2$, then translated into two probability distributions $PD_1$ and $PD_2$. We can define the symmetrised KL divergence between the two PDs:

$$\text{SymKL}(PD_1, PD_2) = KL(PD_1 \| PD_2) + KL(PD_2 \| PD_1)$$

On top of this, now we propose two text-based controversy measures:

**Average KL Divergence (AvgKL)**

For each review $v_j^i$ in $V_i$, based on the process described in Figure 1, there is a corresponding $PD_j^i$. We can define $\text{AvgKL}$ as the average pairwise symmetrised KL between all members in this set.

$$\text{AvgKL}_i = \frac{\sum_{j=1}^{C_i} \sum_{k=1, k \neq j}^{C_i} \text{SymKL}(PD_j^i, PD_k^i)}{C_i(C_i-1)}$$
Generalized Jensen-Shannon Divergence ($gJSdivergence$)

In this case, we try to capture the average distance between each review and an "average" review.

$$gJSdivergence = \sum_{i=1}^{C_i} \text{SymKL}(PD_i, \text{AvgPD})$$

Where AvgPD is the average of all $PD_i$'s.

### 4.2 Correlation Analysis

To understand the connection between the various independent variables above and the dependent variable, and in particular, the effect of social media controversy on product sales, we perform correlation analysis by correlating every independent variable against SalesRank.

Given the ranking nature of the dependent variable, SalesRank, we measure two types of popular rank correlation: Spearman’s $\rho$ and Kendall’s $\tau$.

### 4.3 Regression Analysis

Once we establish the correlations between each independent variable and SalesRank, we would like to explore the possibility of integrating the high-correlation variables into a sales prediction model, and investigate the prediction performance gain.

We compare two models:

- The baseline model: $SalesRank = f(\text{AvgRating})$
- The simple linear regression model: $SalesRank = g(f_1, ..., f_6, ..., f_m)$, where $f_i$ through $f_m$ are variables selected based on the correlation analysis.

The hypothesis is that the linear regression model will outperform the baseline model in terms of prediction power.

### 5. Experiments and Results

In this section, we first describe the data collection process and the resultant data sets; then we present the correlation and regression results, each followed by observations and discussions.

#### 5.1 Data Collection

We choose Amazon.com, a popular e-commerce site, as our data source, due to its offering of consumer ratings, reviews, and product sales information. A further technical reason is the availability of an API, Amazon Web Services (http://aws.amazon.com/).

By accessing the corresponding browse nodes (information available at http://www.browsenodes.com/) in Amazon product hierarchy, we downloaded four data sets. Their characteristics are summarized in Table 1. When performing correlation and regression analyses below, we exclude products with fewer than 10 reviews.

#### 5.2 Correlation Analysis Results

Table 2 through 5 summarizes the results of the correlation analysis, with all the correlation numbers greater than 0.1 highlighted to intuitively illustrate the effectiveness of the corresponding measures.

### Table 1 Data Sets

<table>
<thead>
<tr>
<th>Category</th>
<th>#Products</th>
<th>#Reviews</th>
<th>#Customers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Camera</td>
<td>3990</td>
<td>83005</td>
<td>78026</td>
</tr>
<tr>
<td>TV</td>
<td>1765</td>
<td>24495</td>
<td>22611</td>
</tr>
<tr>
<td>Notebook</td>
<td>4001</td>
<td>182220</td>
<td>172050</td>
</tr>
<tr>
<td>CompSciBook</td>
<td>4692</td>
<td>135005</td>
<td>110991</td>
</tr>
</tbody>
</table>

### Table 2 Correlation Results: Camera

<table>
<thead>
<tr>
<th>Measure</th>
<th>Spearman</th>
<th>Kendall</th>
</tr>
</thead>
<tbody>
<tr>
<td>AvgRating</td>
<td>0.165</td>
<td>0.117</td>
</tr>
<tr>
<td>StdDevRating</td>
<td>-0.136</td>
<td>-0.096</td>
</tr>
<tr>
<td>PosDom</td>
<td>0.176</td>
<td>0.121</td>
</tr>
<tr>
<td>SentEnt</td>
<td>0.009</td>
<td>0.007</td>
</tr>
<tr>
<td>CompZlib</td>
<td>0.107</td>
<td>0.072</td>
</tr>
<tr>
<td>CompBzip2</td>
<td>0.152</td>
<td>0.103</td>
</tr>
<tr>
<td>CompLZV1</td>
<td>0.053</td>
<td>0.035</td>
</tr>
<tr>
<td>CompPPMd</td>
<td>0.158</td>
<td>0.106</td>
</tr>
<tr>
<td>CompLZF</td>
<td>0.059</td>
<td>0.039</td>
</tr>
<tr>
<td>AvgKL</td>
<td>0.140</td>
<td>0.096</td>
</tr>
<tr>
<td>gJSdivergence</td>
<td>0.161</td>
<td>0.111</td>
</tr>
</tbody>
</table>

### Table 3 Correlation Results: CompSciBook

<table>
<thead>
<tr>
<th>Measure</th>
<th>Spearman</th>
<th>Kendall</th>
</tr>
</thead>
<tbody>
<tr>
<td>AvgRating</td>
<td>0.071</td>
<td>0.047</td>
</tr>
</tbody>
</table>
Here are some observations on the variables:

Firstly but not surprisingly, Average Consumer Rating ($AvgRating$), as a numerical summary of overall consumer opinions, is positively correlated with product sales performance.

Secondly, Sentiment Compactness ($SentComp$) and Positivity Dominance ($PosDom$) of product reviews are overall positive indicators of strong sales. The two measures essentially characterize the dominance of positive opinions in review text. In this sense, their behavior is similar to $AvgRating$, which is the numerical counterpart. Sentiment Entropy ($SentEnt$) measured at the coarse level did not turn out to be significantly correlated with sales.

When looking at numerical ratings only, controversy is a bad thing for product sales. This is manifested in the negative correlation between $StdDevRating$ and $SalesRank$. On the other hand, one of the text-based controversy measures, $gJSdivergence$, exhibits strong positive correlation with sales. These two seemingly contradicting observations are in fact justifiable with some extra thoughts. As suggested in previous marketing literature [4], negative WOM tends to be diagnostic and oriented towards product attributes. Review text is the niche place for such nuance information, which is quite possibly “lost in translation” when the consumer is asked to provide only a numerical rating, an overly succinct summary opinion. While a large variation in numerical ratings tends to hurt product reputation, negative attribute-level sentiments embody useful information for potential consumers; controversy at finer levels attracts market attention and potentially yields strong sales. Hence, “controversy is marketing,” taken with a grain of salt. An interesting analogy is the academic refereeing process. We all agree that a numerical rating has to be accompanied by a detailed textual review to capture the subtleties in the evaluation, although sometimes we do err on the side of oversimplifying the process and overly relying on the numerical rating alone. A more important yet “controversial” point to raise is that: a paper rated with a “5” and a “1” is probably more valuable than one rated with two “3”s. A controversy-inviting paper can be scientifically more intriguing, just as a controversy-inviting product may attract more potential consumers. A bold move would be to simply get rid of the mediocre “3” rating in the paper review system.

Some more observations on datasets:

- The camera and notebook datasets behave similarly; the correlations are moderate, whenever applicable.
- The TV dataset exemplifies high correlations, relative to all three others. For some reason, the market reaction to TVs is more sensitive to WOM compared to other products, which is an interesting observation that might suggest future cross-market-sector research.
- On the other extreme, the CompSciBook dataset sees the lowest correlations. It is possibly due to the academic/technical nature of the products, which involve less sentiment ingredient (scientists are more “rational?”).
By scrutinizing the correlation results presented above, we noticed that following variables are highly correlated with SalesRank: AvgRating, StdDevRating, PosDom, CompBzip2 (all compression algorithms behave similarly; this one was selected because it has consistently high correlations in all four data sets), and gJSdivergence.

We feed these variables into a simple linear regression model for SalesRank prediction. By conducting standard cross validation experiments, we obtained the results described in Table 6. The univariate prediction model based only on AvgRating is used as the baseline for comparison, and the baseline performance measures are presented in parentheses in the same table. As we can see, the multiple regression model incorporating newly proposed measures (all of which are significant in the regression model, except PosDom) consistently outperforms the baseline.

<table>
<thead>
<tr>
<th>Collection</th>
<th>Spearman</th>
<th>Kendall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Camera</td>
<td>.222 (.165)</td>
<td>.153 (.117)</td>
</tr>
<tr>
<td>CompSciBook</td>
<td>.136 (.071)</td>
<td>.091 (.047)</td>
</tr>
<tr>
<td>Notebook</td>
<td>.364 (.183)</td>
<td>.248 (.122)</td>
</tr>
<tr>
<td>TV</td>
<td>.410 (.361)</td>
<td>.276 (.242)</td>
</tr>
</tbody>
</table>

6. Conclusion and Future Work

In this study, upon proposing a set of innovative text-based measures, we found strong positive correlations between social media controversy and sales performance across multiple datasets. We further illustrated the power of the correlation by feeding newly proposed variables into a simple linear regression model for sales prediction and observing performance superiority to a baseline model.

Here are some thoughts on potential issues and future research directions:

- At the technical level, the current set of text-based measures are not only innovative but also simple, in the sense that they only exploit shallow lexical information in the review text and transform such into approximated sentiment distributions. We are yet to find out, however, whether more sophisticated NLP techniques would be helpful in better capturing the nuance, attribute-level controversy hidden in text, given the light shed by our initial findings.

- Another potential improvement on the technical end is to reformulate the regression problem when illustrating the power of multiple text-based variables. Given the ranking nature of the dependent variable, SalesRank (at least as it is available now), the statistical inference task should aim at approximating the ranking function, to which classical regression techniques are not the ideal approach. In fact, “learning to rank” has emerged as an active and growing area of research both in information retrieval and machine learning [30].

- Product reviews are accumulated over time. At the methodological level, it is our plan to introduce the temporal dimension and construct a panel data model, to further justify the effectiveness of our content-based measures.

- Theory development is missing in the current research. It is important to ground our computational research in existing marketing theories or to use our findings to enrich them. Specifically, we will focus on aligning our study with previous work on Word-of-Mouth in marketing literature, e.g., the importance of WOM in helping consumers find products matching their preferences [31].

- Many hidden patterns, upon discovery, are prone to gaming behaviors. For example, the famous PageRank algorithm behind Google has been a consistent target for spoofing and spamming. Similarly, our findings, once published, are susceptible to potential exploitations and manipulations by various online parties. We just have to be mindful of such possibilities for both caveats and opportunities.

- Now, let’s turn to the pragmatic business level consideration. As our intuition and our initial findings tell us, controversy is not always helpful for sales. It can hurt your business. In the “Crispy Frickin’ Sandwich” example, we have to consider just how many people are turned off by the ad; how many would've-been customers Sheetz has lost because of this. We might guess that the Sheetz management did not think the ad would be offensive to too many people and was willing to alienate a few to bring in more. Shining negative light on your business is generally not desirable, and should certainly be accompanied by risk assessment and damage control mechanisms. If you intend to spark controversy with your marketing, it is important to consider your target audience. Who is most likely to buy the product you are selling? If you can appeal
to the target, the controversy campaign is more likely to succeed. But if your target is broad, controversy may not be the best route to take, due to the risk of alienating part of the demographic you are trying to reach. Again, this comes down to risk assessment. Sheetz no doubt considered that a chicken sandwich would appeal to a pretty broad spectrum of people, but felt that the campaign would gain them more customers than it would cost them.

The very topic of controversy-driven marketing is potentially controversial itself. Personal ethics come into play and there will always be disagreement. It calls for more significant effort to study how such strategies have impacted, or will impact, businesses. Whatever we do in life, and whatever we consider doing, ethics should be taken into account. Otherwise, we are likely to come up with wrong decisions.

7. References


Conference on Language Resources and Evaluation.


