Paper 1839-2018

Measuring Response to the Ultimate Driving Machine: Consumer Sentiments and Brand Advertising

Amit K Ghosh, Cleveland State University; Goutam Chakraborty, Oklahoma State University; Praveen Kotekal, Oklahoma State University

ABSTRACT

Increasingly, customers use social media and other Internet-based applications (e.g., review sites) to voice their opinions and to express their sentiments about brands. These reviews profoundly influence brand performance, either *directly* (by affecting consumer behavior) or *indirectly (*by generating positive or negative word-of-mouth through online social networks.) We present a methodology that can be used to collect textual reviews from popular brand review sites and discussion boards. These reviews are then analyzed to uncover brand sentiments. Strategic implications are discussed as we explore the influence of actual product experience and brand expectations (that are shaped by advertising) on brand sentiments.

In order to demonstrate the utility of our methodology, we first collected data using customized Python code from Edmunds.com – a popular review site, where car owners post detailed evaluations of the cars they have purchased. To get a long-term perspective of brand strengths and weaknesses, data was collected for car models released from 2012 to 2017 from several brands. These detailed, unstructured, and textual reviews were then analyzed using the best practices of text mining and sentiment analysis. Finally, we explored the extent to which a brand's advertising campaign impacts consumer sentiments and overall brand assessment.

INTRODUCTION

The explosion of unstructured and qualitative customer data that is available from the Internet has created vast challenges in the field of marketing analytics (Che, Safran, & Peng, 2013) (Labrecque, 2014). Increasingly, customers are using Social media and other Internet-based applications (e.g., review sites) to voice their opinions and express their sentiments about brands (Trainor, Andzulis, Rapp, & Agnihotri, 2014). Past research suggests that such opinions and discussions have a profound influence on brand performance either *directly* by impacting behavior (Hinz, Skiera, Barrot, & Becker, 2011), or *indirectly* by generating positive or negative word-of-mouth in social networks (Elsner, Heil, & Sinha, 2010) (Yadav, et al., 2013). It is evident from the extant research that such deep consumer insights have considerable strategic value for firms.

However, a majority of the tools that are being currently used to obtain and analyze data from multiple social media platforms, blogs, and discussion boards can be classified as social media channel *reporting* tools (Thiel, Kötter, Berthold, Silipo, & Winters, 2012). The visual scorecards and dashboards that these tool create, provide a good overview or big picture (based on a specific selection criteria) of channel performance. Unfortunately, these tools are not designed to help us understand customer sentiments based on a string of terms that are used by customers in a specific context. Thus, reporting tools are not suitable to uncover deep insights into customer needs or to identify specific concerns related to a brand (Thiel, Kötter, Berthold, Silipo, & Winters, 2012).

In this paper, we propose a methodology that can be used to collect consumer reviews from popular brand review sites and discussion boards. These reviews are then analyzed to reveal customer sentiments at the attribute level of each brand. Sentiment analysis (SA) is performed by identifying consumer sentiment that relate to determinant brand attributes and by considering the valence (positive or negative) of these customer sentiments. Thus, our paper is designed to help brand managers understand the specific strengths and weaknesses of a brand based on a thorough analysis of consumer sentiments.

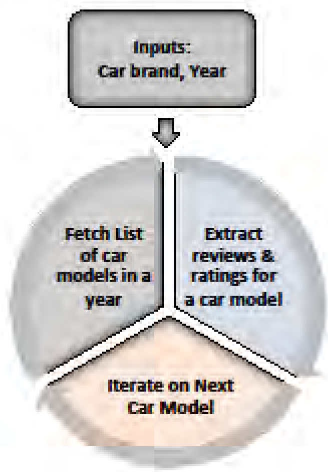
We also explore whether brand sentiments are influenced by brand advertising. Spending on television advertising in the US continues to grow and is expected to be exceed $75 billion in 2017. Brand ads are often designed to differentiate brands in terms of attributes that brands possess. These ads create positive expectations about specific brand attributes in the minds of consumers and thereby promote brand preference and choice. For instance, Subaru ads positions their brand as a safe car. It is, therefore, conceivable that safety is the primary reason consumers buy a Subaru. In this paper we investigate whether such attribute-related brand expectations encourage discussions on review boards and thereby influence brand sentiments and brand preferences. Thus, the potential influence of brand ads on consumer sentiments is also explored in this study.

The next five sections discusses each step of our methodology. Since the input to a SA system is a corpus of documents from the web, a brand review site is selected in the first step. Following that, we discuss text preparation, generation of text topics, and attribute-level sentiment analysis. Since the primary objective of SA is to convert unstructured text into meaningful information, the strategic implication of the results are discussed. We conclude by outlining some research limitations and possible research extensions.

SELECTION OF BRAND REVIEW SITE

The first step was to extract data from Edmunds.com website. We designed a tool based on Python code that can collect data for a specific brand of car from the website and then return the data in a JavaScript Object Notation (JSON) format. The tool also iterates through all the models that are available for a certain brand of car within a certain time period. The data extraction process appears in Figure 1.

Figure 1: Data Extraction Process



The Dealers API of Edmunds.com website was then accessed by using the JSON package in Python. Using this tool, we extracted textual reviews along with numerical ratings for car models released from 2012 till 2017 for four brands – Chevrolet, Honda, Subaru, and Toyota. Since the research purpose was to investigate sentiments at the brand levels (e.g., Toyota) and not at the model level (e.g., Corolla), data on each model within a certain brand was ignored. A total of 2,176 reviews were extracted that included 696 reviews of Chevy, 429 reviews of Honda, 257 reviews of Subaru, and 794 reviews of Toyota. The excel file with all the information was then imported to SAS Enterprise Miner for further analysis.

TEXT PARSING AND FILTERING

The purpose of text parsing was to break up sentences that appeared in the textual reviews to terms that relate to the attributes of cars. Terms could be single words or a string of words. The text parsing and filtering node of SAS Text Miner was used for this purpose. The objective was to retain terms that specifically relate to the attributes of a car and to eliminate articles and connectors that have a high frequency but do not actually describe any particular feature. These terms were excluded from the analysis by creating a list of stop words. The stop list also included the names of different car models, so that brand sentiments could be analyzed at the level of brands and not at the model level. Besides stop words, a custom dictionary was used to detect synonyms, correct spellings, and to detect multi-word terms. To suit to our analysis goals, we ignored parts of speech like 'Abbr', 'Aux', 'Conj', 'Det', 'Interj', 'Num', 'Part', 'Pref', 'Prep', 'Pron', and 'Prop.' Entities like 'Address', 'Currency', 'Date', 'Location', 'Measure', 'Percent', 'Phone', 'Timeperiod' and attributes like 'Punct', 'Mixed', 'Num' were also not considered in the analysis.

An important objective in textual analysis is to differentiate documents so that they can be meaningfully classified into groups that reflect text topics. Past research in the field of information retrieval suggests that term frequencies, or the number of times terms appears in a document collection, are generally not considered to be good discriminators (Chakraborty, Pagalu, & Garla, 2013). The terms with the largest frequency of occurrence appear in Figure 2. It is evident that these terms are unlikely to help us differentiate across documents since they do not relate to car attributes. Inverse document frequency (IDF) was used as term weight as suggested by (Salton & Buckley, 1988). IDF calculates term weight as the inverse of the frequency of occurrence of a term in documents. A log transformation was used to stabilize the variability across frequencies. Thus, terms with high frequencies were considered to be less important discriminators while terms with lower frequencies were considered to be more important.

Figure 2: Terms with highest Frequencies

Term	Role	Attribute	Freq	# Docs	Keep	Parent/Child Status	Parent ID	Rank for Variable numdocs
+ be	...Verb	Alpha	8083	1756	N	+	25091	1
+ have	...Verb	Alpha	4391	1484	N	+	24980	2
+ car	...Noun	Alpha	4053	1355	Y	+	9743	3
not	...Adv	Alpha	3481	1334	N		25000	4
+ do	...Verb	Alpha	2020	983	N	+	25143	5
+ get	...Verb	Alpha	1833	976	N	+	24926	6
+ drive	...Verb	Alpha	1252	779	Y	+	15206	7
s	...Noun	Alpha	1362	727	N		24923	8
very	...Adv	Alpha	1204	724	N		24957	9
+ mile	...Noun	Alpha	1170	692	Y	+	15281	10
+ good	...Adj	Alpha	1010	678	Y	+	1424	11
+ buy	...Verb	Alpha	865	634	Y	+	749	12
+ much	...Adj	Alpha	856	547	N	+	25159	13
no	...Adv	Alpha	838	546	N		25172	14
+ go	...Verb	Alpha	761	545	N	+	24946	15

Figure 3 displays the terms with the lowest frequencies. While these terms relate to car attributes, they are unlikely to be of value from a strategic perspective. Thus, unless a term appeared in at least four documents, it was excluded from the analysis.

Figure 3: Terms with lowest frequencies

zero maintenanc...	Noun Group	Alpha	1	1	Y		13035	6918
zero mechanical...	Noun Group	Alpha	1	1	Y		4258	6918
+ zero mechanic...	Noun Group	Alpha	1	1	Y	+	2423	6918
+ zero other com...	Noun Group	Alpha	1	1	Y	+	1176	6918
+ zero unexpect...	Noun Group	Alpha	1	1	Y	+	290	6918
zero warranty	...Noun Group	Alpha	1	1	Y		1057	6918
zip	...Noun	Alpha	1	1	Y		17262	6918
zippin	...Noun	Alpha	1	1	Y		6812	6918
zippy little car	...Noun Group	Alpha	1	1	Y		1883	6918
+ zombie	...Noun	Alpha	1	1	Y	+	7060	6918
+ zone	...Verb	Alpha	1	1	Y	+	12699	6918
zone climate co...	Noun Group	Alpha	1	1	Y		11229	6918
zoom	...Noun	Alpha	1	1	Y		4699	6918
zx	...Noun	Alpha	1	1	Y		11958	6918
+ zx car	...Noun Group	Alpha	1	1	Y	+	17363	6918

GENERATION OF TEXT TOPICS

This paper focuses on only five attribute level sentiments – all other attributes are ignored. Therefore, we wanted to determine the extent to which these five attributes actually explain overall sentiment about a brand of car. For this task, we used the numerical ratings of overall preference that consumers had expressed to serve as a proxy for overall sentiment and the attribute-level ratings as proxies for attribute-level sentiments. All the numeric ratings are collected from brand evaluators using a five point rating scale. Regression analysis was then performed, where the attribute level ratings were specified as independent variables and the overall rating was specified as a dependent variable. If the attribute-level numerical reviews could explain a large portion of the variability in the overall ratings, then one could speculate that these attributes would be strategically meaningful to focus on. For this analysis, data was aggregated across all the brands.

Table 1 presents the results of the regression – please see column for overall analysis which reports the results that are aggregated across brands. The adjusted r-square was 0.72 and the p-values for all the attributes were significant at the 0.01 level or lower. The directionality of all the parameter estimates were positive suggesting that positive attribute ratings contributed to positive overall assessment. All of these statistics suggest that the five selected attributes explain a large part of the variance in the overall sentiments towards a brand. Thus, using the five selected attributes to define text topics would be strategically and methodologically meaningful.

Table 1: Validity of the five attributes on consumer sentiments

	Overall	Chevy	Honda	Toyota	Subaru
Adjusted R-square	0.72	0.77	0.65	0.77	0.77
Standardized Beta					
Comfort rating	0.20	0.20	0.26	0.19	0.20
Technology rating	0.19	0.15	0.23	0.22	0.13
Safety Rating	0.13	0.15	0.05*	0.15	0.20
Interior Rating	0.18	0.17	0.14	0.15	0.23
Performance rating	0.35	0.38	0.35	0.33	0.30

* Parameter estimate is not significant at the 0.05 level. All the other parameter estimates are significant at the 0.001 level.

The next step was to create text topics. The objective was to group the primary terms in a manner that reflected sentiments related to the specific attributes that we study in these paper. It was determined that a completely automated system for generation of text topics would not be desirable for several reasons. First, an examination of the terms (generated in the last step) revealed that several terms were related to issues such as price and good value. While price and value are important attributes, there are no numerical ratings for these in the Edmunds.com website. Thus, they are not the focus of the study in this paper. The attributes under consideration were comfort, quality of technology, safety, interior quality, and performance primarily because numerical ratings were available for these factors. Second, a single core term such as “seats” could be classified under multiple attributes depending on the context or the other terms that were associated to it. A heated seat, the material with which it was made, or the contour of the seat could contribute to comfort perceptions. The same term could also influence consumer sentiments about how good the interior looked.

Due to these issues, we decided to create five custom topics that reflected the five attributes under consideration. A thorough investigation of these five features were conducted to identify terms that would reflect attribute-level sentiments when consumers were evaluating cars. Based on this domain-specific knowledge, 593 terms were identified. Out of these, 213 terms were related to performance, 123 terms to comfort, 102 terms to safety, 71 terms to technology, and 84 terms to interior. As mentioned before, a certain term could relate to multiple text topics. These terms were specified as inputs to the topic selection node of SAS Enterprise Miner. The algorithm used the input list of terms and the concept maps (or the association across terms) to determine whether a certain sections of the review reflected sentiments about a specific text topic. A single review could express sentiments about multiple text topic. For instance, a review might be positive about the performance of a car but might express apprehension about the technological features of the same car. However, a review that has none of the user input terms, would not be evaluated.

The results of text topic generation node is summarized in Table 2. Terms related to performance seem to be the most dominant followed by terms related to comfort. In many cases, multiple performance related sentiments are expressed in a single review. In contrast, safety related sentiments are the least likely to appear in reviews. The top three issues or terms that are associated with each text topic also appears in this table. It is noteworthy that “seats” contribute to perceptions related to comfort and to interior, while “camera” influences views about safety as well as about technology.

Table 2: Text Topic Selection

Text Topic	Number of			Important Issues		
	Terms	Documents	Frequency			
Performance	54	2,377	3,093	mileage	Acceleration & engine	transmission
Comfort	34	1,155	1,601	seat	Cruise control	Voice control
Interior	21	998	1,447	seat	airbag	Climate control

Technology	23	825	972	Speaker & sound system	Bluetooth	Video & camera
Safety	29	644	737	camera	Traction control	Warnings & blind spots

ASSESSING BRAND SENTIMENTS

The final step was to determine the polarity or valence of the sentiments. Document level analysis or whether the entire document can be classified as a positive or negative sentiment is not recommended since consumers express multiple sentiments related to multiple text topics in a single review as shown in Table 2. Furthermore, consumers could express positive sentiments about one attribute and negative sentiments towards others. In this paper, sentiment analysis was performed at the level of a brand (or object) as well as at the level of a text topic. Thus, our methodology would allow consumers to express positive sentiment about the performance of a certain brand of car but indicate apprehension regarding its safety features.

We used a supervised sentiment classification system according to the best practices of sentiment analysis. The numeric ratings of the five attributes that were collected from the Edmunds.com website, were used to determine the valence of the sentiments in accordance to the procedure suggested by (Chakraborty, Pagalu, & Garla, 2013). If a certain text topic) was detected in the textual review for a certain consumer, a numeric rating of four or five (on a five point rating scale) for the same attribute would indicate a positive sentiment. However, a rating of three or below would indicate a negative sentiment. Past research has shown that this naïve Bayes classification system works well for supervised classification tasks (Pang & Lee, 2008). The median was used to impute attribute ratings that had missing values in order for us to complete the classification task. The supervised sentiment analysis was performed on all the 2,156 textual reviews.

Figure 4 shows the end results of sentiment analysis for a consumer with a mixed review. The transmission-related problems, coupled with a negative rating for performance, leads to a negative performance sentiment. The textual review also expresses concern related to technological aspects such as Bluetooth, touch screen, and voice and results in a negative assessment of technology. However, “plenty of seats” results in a positive sentiment about the interior of the car.

Figure 4: Supervised Sentiment Analysis

Textual review:

“I need this car for haul my growing family around. I knew the good and bad before purchasing & for a while I loved it because it is very functional. Until the transmission (6-speed) gave temperature warning in the middle of a camping trip. I will be taking it in soon to get it checked out. The goods: plenty of seats, MPG is awesome compare to other vehicle in class & all the technology is good even for an EX model. The bads: infotainment screen is useless in direct sunlight, touch screen is very hard to use while the car is moving, armrest is uncomfortable, Bluetooth takes SIX buttons push to change connected phone, voice command is none functional, very hard to jump start because engine cover & placement of the battery. Mother of all problems, the transmission overheats!!!”

Topics identified & Valence of SENTIMENTS:

Performance (Term= “Transmission” & “Transmission Overheats”)

Negative numerical rating

Technology (Term = “Bluetooth” & “Touchscreen” & “voice”)

Negative numerical rating

Interior (Terms= “plenty of seats”)

Positive numerical rating

DISCUSSION AND MANAGERIAL IMPLICATIONS

In this paper, we propose a methodology that can be used to evaluate post-purchase brand sentiments. As mentioned before, past literature suggests that these sentiments greatly influence current as well as future consumer decisions. Our paper provides valuable insights for both managers and analysts that we now discuss.

Consumers are likely to have different expectation for different brands. Ads for Subaru emphasize safety; Hondas and Toyotas are known for reliability and great gas mileage; Chevy ads focus on performance and comfort. If the brand ads prompted consumers to discuss these differentiating features in reviews, the number of safety related reviews would be dominant in reviews of Subaru. In comparison, Chevy reviews should have fewer safety related reviews. In order to test this hypothesis, we compared the proportion of times a text topic was mentioned across all the four brands. These statistics were similar across all of the brands. Performance-related sentiments were the most prevalent and occurred in over 39% of the documents as reported in Table 2. In contrast, safety related issues appeared in just over 10% of the documents. Our results suggest that expectations created by brand ads do not seem to influence review topics. Consumers tend to discuss topics that they are more likely to observe through user experience. Consumers can feel the acceleration and the smoothness of transmission. They can observe the economic benefit of gas mileage. In contrast, the effectiveness of safety features are likely to be observed rarely – probably in cases where there was an accident or where an accident was averted. Thus, while people might buy different brands for different reasons, they are more likely to speak about the more observable attributes.

We also investigated whether prevalence of certain attributes imply that they are more important determinants in the overall assessment of a brand. For instance, since most sentiments are related to performance, is performance the most important driver of overall brand assessment? For this purpose, we performed a regression analysis of attribute-level ratings on overall ratings for each brand. The standardized regression coefficients for the brand specific analyses appear in Table 1. The results display a remarkable consistency between beta weights across all the brands. For every brand, the weights for performance were the highest and those for safety were the lowest. The regression weights for the attributes were also correlated with the frequency of times a text topic was mentioned in documents – see Table 2. Thus, our research suggests that the frequency of a certain text topic (e.g., performance) influences overall brand assessment. Expectations created by brand ads do not.

This paper also demonstrates the benefit of using sentiment analysis compared to relying solely on numerical ratings to assess brands. Sentiment analysis provides valuable details to brand managers and market analysts. By comparing assessment across brands on key attributes, it allows them to determine the strengths and weaknesses of brands. A text topic (e.g., performance) can be decomposed to various terms such as mileage and acceleration. Thus, specific brand strengths and weaknesses (related to performance) can identified by analyzing the frequency of these terms. These insights can then be used to either differentiate the brand by promoting its strengths or by taking appropriate corrective action to address weaknesses that are uncovered.

CONCLUSION

In this paper, we propose a methodology that can be used to collect textual data that are expressed in popular brand review sites and discussion boards. The Dealer API of Edmunds.com was accessed by using a JSON package in Python and 2176 textual reviews were collected for four major brands. In addition, numerical ratings for five attributes and for overall preference was collected. We purposely collected data on post-purchase reviews since these are read by consumers who are planning to buy a car and are interested in learning about the actual experience of users.

The data was imported to SAS Enterprise Miner for further analysis. Terms that specifically related to the five attributes were retained. A custom user-defined text topic generation node was used to identify terms that related to each of the five attributes. Finally, a supervised sentiment analysis methodology that utilized numerical ratings on the five attributes was used to determine the valence or polarity of the sentiments.

Several key insights were uncovered in this paper. Results suggest that consumer reviews are likely to focus on more observable attributes and “feel good” attributes such as performance. Safety-related concerns are least likely to be discussed. Furthermore, though consumers might have brand-related expectations related to attributes based on brand ads, these expectations do not automatically become topics of reviews. The frequencies of text topics are also important determinants of overall brand sentiment. Our paper allows marketers to get an in-depth understanding of the problems and prospects related to a brand by focusing on terms within a certain text topic. In summary, our methodology allows brand managers to actively manage brands based on a thorough and continual assessment of brand sentiments.

While a large portion (65-77%) of the variance in overall brand evaluations can be explained by the five attributes, future research could focus on expanding this list. Text topics could be expanded by including factors such as price, value, and predicted reliability. However, unsupervised sentiment analysis would be required to determine the valence of these perceptions. The list of brands should be increased to determine whether the insights can be generalized. Sentiment analysis is a topic of increasing importance to marketers. Currently, we are collecting data from four other brands to improve the generalizability of the study results. Although this paper is a modest attempt to investigate consumer sentiments, we hope that it will lead to more discussion and research in this vital area of marketing.

REFERENCES

Chakraborty, G., Pagalu, M., & Garla, S. (2013). *Text Mining & Analysis: Practical Methods. Examples, and Case Studies Using SAS. SAS Institute Inc., Cary, NC.* Cary, NC.: SAS Institute Inc.

Che, D., Safran, M., & Peng, Z. (2013). From Big Data to Big Data Mining: Challenges, Issues, and Opportunities. *International Conference on Database Systems.* Springer.

Elsner, M. K., Heil, O. P., & Sinha, A. R. (2010). *How social networks influence the popularity of user-generated content. Special Report, 10-206.* Marketing Science Institute.

Hinz, O., Skiera, B., Barrot, C., & Becker, J. U. (2011). Seeding strategies for viral marketing: an empirical comparison. *Journal of Marketing, 75*(4), 55-71.

Labrecque, L. I. (2014). Fostering consumer-brand relationships in social media environments: the role of parasocial interaction. *Journal of Interactive Marketing, 28*(2), 134-148.

Pang, B., & Lee, L. (2008). Using very simple Statistics for Review Search: An Exploration. *COLING 2008: Companion Volume – Posters & Demonstrations*, (pp. 75-78).

Salton, G., & Buckley, C. (1988). Term Weighting Approaches in Automatic Text Retrieval. *Journal of Documentation, 28*(5), 513-523.

Thiel, K., Kötter, T., Berthold, M., Silipo, R., & Winters, P. (2012). *Creating usable customer intelligence from Social Media Data: Network analytics meets text mining.* KNIME.com.

Trainor, K. J., Andzulis, J., Rapp, A., & Agnihotri, R. (2014). Social Media technology usage and customer relationship performance: A capabilities-based examination of social CRM. *Journal of Business Research, 67*, 1201-1208.

Yadav, M. S., de Valck, K., Hennig-Thu, T., Hennig-Thurau, T., Hoffman, D. L., & Spann, M. (2013). Yadav, Manjit S., Kristine de Valck, Thorsten Hennig-Thurau, Donna L. Hoffman, & Martin Spann. (2013). Social commerce: a contingency framework for assessing marketing potential. Journal of Interactive Marketing 27(4): 311-323. *Journal of Interactive Marketing, 27*(4), 311-323.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Amit K. Ghosh
Cleveland State University
(216)-687-4770
a.ghosh@csuohio.edu