Journal of Advertising

ISSN: 0091-3367 (Print) 1557-7805 (Online) Journal homepage: https://www.tandfonline.com/loi/ujoa20

Consumer Responses to Femvertising: A Data-Mining Case of Dove's "Campaign for Real Beauty" on YouTube

Yang Feng, Huan Chen & Li He

To cite this article: Yang Feng, Huan Chen & Li He (2019) Consumer Responses to Femvertising: A Data-Mining Case of Dove's "Campaign for Real Beauty" on YouTube, Journal of Advertising, 48:3, 292-301, DOI: 10.1080/00913367.2019.1602858

To link to this article: https://doi.org/10.1080/00913367.2019.1602858

Journal of Advertising, 48: 292–301

ISSN: 0091-3367 print / 1557-7805 online
DOI: 10.1080/00913367.2019.1602858

Routledge
Taylor & Francis Group

Check for updates

Research Note

Consumer Responses to Femvertising: A Data-Mining Case of Dove's "Campaign for Real Beauty" on YouTube

Yang Feng
San Diego State University, San Diego, California, USA

Huan Chen
University of Florida, Gainesville, Florida, USA

Li He
Qualcomm Technologies, Inc., San Diego, California, USA

Adopting a combination of qualitative textual analysis, human-based content analysis, and machine learning–based data mining, we propose a procedure to analyze user-generated content (UGC) on social media using Dove's "Campaign for Real Beauty" as a case for demonstration. We provide a guideline to explicate all six steps of the analysis procedure: topic identification through qualitative textual analysis, generation of labeled data through human coding, data preprocessing, evaluation of machine learning–based classifiers, topic classification of unlabeled data, and conducting research. The study has important methodological implications for advertising scholars and practitioners.

Unlike traditional media, such as TV, radio, magazines, and newspapers, social media open the door for public response alongside an advertisement (Knoll 2016; Walther et al. 2010). For example, as the largest video-sharing site on the Internet, YouTube enables users to participate in conversations around video ads posted on the platform (Khan 2017). Instead of consuming content posted by a single broadcaster, YouTube users can also observe comments on the video posted by other users (Walther et al. 2010). Comments on YouTube exemplify user-generated content (UGC), which is defined as "all publicly (or partially publicly) available online information initiated and/or created by end-users, as opposed to by media professionals" (Knoll 2016, p. 284). The openness of these participatory technologies has the potential to enhance or to diminish the intended effects of central advertising messages by juxtaposing consonant or contradictory opinions of users with the central messages that a brand intends to convey (Walther et al. 2010; Liu, Burns, and Hou 2017; Knoll 2016). Although about 20% of social media users actively engage in creating UGC, many less-active social media users are the receivers of UGC, and their brand attitudes and purchase intentions may be affected by being exposed to UGC (Jin and Phua 2014; Knoll 2016), highlighting the importance of studying UGC within the context of advertising displayed on social media.

Knoll (2016) noted that although a multitude of data can be accessed in the form of digital texts, when it comes to advertising displayed on social media, quite a few studies employed various forms of automated content analyses to examine UGC. It can be assumed that this kind of data analysis is likely to grow in future studies due to its cost and time efficiency. For instance, Campbell et al. (2011) examined consumer conversations around consumer-generated advertising on YouTube and developed a typology of consumer-generated ad conversations composed of two dimensions. The first dimension divides consumer comments into conceptual and emotive comments, while the second

Address correspondence to Yang Feng, School of Journalism and Media Studies, College of Professional Studies and Fine Arts, San Diego State University, 5500 Campanile Drive, San Diego, CA 92182, USA. E-mail: yfeng@sdsu.edu

Yang Feng (PhD, Southern Illinois University Carbondale) is an assistant professor of advertising, School of Journalism and Media Studies, College of Professional Studies and Fine Arts, San Diego State University.

Huan Chen (PhD, University of Tennessee) is an assistant professor of advertising, Department of Advertising, College of Journalism and Communications, University of Florida.

Li He (MS, University of Nebraska–Lincoln) is a senior engineer, Qualcomm Technologies, Inc.

dimension divides consumer comments into collaborative and oppositionary comments. These two dimensions inspire the identification of four types of consumer comments on YouTube: inquiry, laudation, debate, and flame. However, because they used a qualitative textual analysis tool called Leximancer, Campbell et al.'s (2011) study did not classify the conversation type of each comment.

To overcome the limitations of traditional content analysis and qualitative textual analysis tools such as Leximancer, we propose a six-step analysis procedure for opinion mining on UGC that effectively integrates qualitative textual analysis, human-based content analysis, and machine learning–based data mining. There are three major strengths of this analytic procedure: First, it can generate coding categories from scratch. Second, it can classify large numbers of text data with relatively high accuracy. And third, it leverages the advantages of qualitative textual analysis, quantitative content analysis, and machine learning algorithms.

Thus, the purpose of this study is twofold. First, we aim to provide a step-by-step guideline for the proposed analysis procedure. Second, for illustrative purposes, we apply the proposed analysis procedure to analyze 20,419 unique YouTube comments on a femvertising campaign—Dove's "Campaign for Real Beauty." The reason we selected this campaign is that, as a pioneer of femvertising that embraces feminist empowerment to boost brand consumption (Akestam, Rosengren, and Dahlen 2017), Dove's "Campaign for Real Beauty" went viral and generated rich conversational contexts on YouTube (Bissell and Rask 2010; Cinelli and Yang 2016; Millard 2009). As such, this campaign provides us with rich textual context to illustrate how to apply the proposed procedure to analyze UGC. To showcase some sample UGC-related issues that the proposed procedure can address, we developed two research questions regarding Dove's "Campaign for Real Beauty":

RQ1: Based on the proposed analysis procedure, what topics do consumers discuss around Dove's "Campaign for Real Beauty"?

RQ2: Is there any relationship between the topics of top-ranked comments and those of the overall viewer comments?

The machine learning component of the proposed procedure is built on supervised text classification, which refers to a group of statistical machine learning techniques that attempts to attach a label to a particular set of documents (e.g., unlabeled comments) based on predefined labels assigned to other similar sets of documents (e.g., labeled comments) (Kalita 2015; Okazaki et al. 2014). To perform supervised text classification, researchers need to prepare predefined categories (or labels) and then to assign some sets of documents to each of the predefined categories so that these unlabeled documents become labeled data. Similar to human beings who learn from past experiences, the machine needs to learn the classification logic from the labeled data and then can use the same logic to classify unlabeled data. Using supervised text classification, prior research classified unlabeled online text into predefined coding categories, such as assigning each brand tweet to one of the six predefined categories (i.e., sharing, information, opinion, question, reply, and exclude) (Okazaki et al. 2014) and allocating each online customer review to one of the two predefined sentiment categories (i.e., positive and negative) (Ye, Zhang, and Law 2009). However, prior studies on supervised text classification did not specify how to establish predefined coding categories. In our case, to fulfil the goal of assigning each Dove "Campaign"-related YouTube comment into a topic, we need to first prepare a predefined coding scheme and then use it to classify the topics of some comments to generate labeled comments. To create a predefined coding scheme, in our case, we needed to first review femvertising literature to familiarize ourselves with current conversations regarding Dove's "Campaign for Real Beauty" and then develop our coding scheme through qualitative textual analysis.

LITERATURE REVIEW

The Ambivalence of Femvertising

Feminist advertising, termed *femvertising*, is a growing marketing trend utilized by large brands, such as Dove, that appropriates feminist values and female empowerment to encourage brand consumption (Akestam, Rosengren, and Dahlen 2017). Dove was a forerunner for this advertising strategy, launching their "Campaign for Real Beauty" in 2004. In addition, a series of videos featuring real models—people who are not professional models and do not look like typical fashion models—was launched as part of the "Real Beauty" project (Millard 2009). An important aspect of this campaign is that it capitalizes on the influence of social media where it goes viral (Gill and Elias 2014). The brand was credited for being groundbreaking for using models in different sizes and with different skin colors to showcase that all women are beautiful (Bissell and Rask 2010). However, this campaign sparks feelings of both skepticism and joy.

On one hand, exposure to models who look like ordinary women in Dove's "Campaign for Real Beauty" helps reduce viewers' perceived discrepancy between themselves and the models, which further enhances their positive attitudes toward both the ad and the brand (Bissell and Rask 2010; Cinelli and Yang 2016). On the other hand, the ability of corporations to capitalize on feminism indicates

TABLE 1
Comments of YouTube Videos on Dove "Real Beauty" Campaign

<table>
<tr><th>Video Title</th><th>URL Link</th><th>Year</th><th>Number of Views</th><th>Number of Posted Comments As of August 18, 2018</th><th>Number of Retrieved Comments</th><th>Number of Comments in Final Analysis</th></tr>
<tr><td>Dove Real Beauty Sketches: You're More Beautiful Than You Think (3 min)</td><td>https://www.youtube.com/watch?v=XpaOjMXyJGk&t=88s</td><td>2013</td><td>68,172,368</td><td>11,404</td><td>11,306</td><td>8,559</td></tr>
<tr><td>Dove Evolution</td><td>https://www.youtube.com/watch?v=iYhCn0jf46U</td><td>2006</td><td>19,460,807</td><td>3,693</td><td>3,693</td><td>3,195</td></tr>
<tr><td>Dove | Beauty on Your Own Terms #MyBeautyMySay</td><td>https://www.youtube.com/watch?v=_XOa7zVqxA4&t=5s</td><td>2016</td><td>13,087,787</td><td>778</td><td>667</td><td>357</td></tr>
<tr><td>Dove Choose Beautiful | Women All over the World Make a Choice</td><td>https://www.youtube.com/watch?v=7DdM-4siaQw&t=156s</td><td>2015</td><td>10,138,980</td><td>5,524</td><td>4,718</td><td>3,369</td></tr>
<tr><td>Dove Real Beauty Sketches | You're More Beautiful Than You Think (6 min)</td><td>https://www.youtube.com/watch?v=litXW91UauE&t=48s</td><td>2013</td><td>8,640,780</td><td>3,840</td><td>3,645</td><td>2,921</td></tr>
<tr><td>Beauty Pressure</td><td>https://www.youtube.com/watch?v=Ei6JvK0W60I</td><td>2007</td><td>4,981,750</td><td>1,453</td><td>1,375</td><td>1,136</td></tr>
<tr><td>Dove Pro-Age Campaign</td><td>https://www.youtube.com/watch?v=vilUhBhNnQc</td><td>2007</td><td>3,327,560</td><td>686</td><td>685</td><td>571</td></tr>
<tr><td>Dove Legacy | A Girl's Beauty Confidence Starts with You . . . (3 min)</td><td>https://www.youtube.com/watch?v=Pqknd1ohhT4&t=15s</td><td>2014</td><td>3,184,430</td><td>367</td><td>316</td><td>168</td></tr>
<tr><td>Dove | Real Beauty Productions | Meet Diana</td><td>https://www.youtube.com/watch?v=poIrZELfEME&t=2s</td><td>2017</td><td>2,463,511</td><td>372</td><td>358</td><td>179</td></tr>
</table>

the need to be skeptical of commercialized social movement (Johnston and Taylor 2008). For instance, Dove's "Campaign for Real Beauty" promotes women's self-esteem through brand building and serves dual goals: to make women feel more beautiful and to sell more Dove beauty products. In this vein, Dove appropriates feminist values and reconstructs feminism as an ideal that is, arguably, achieved through consuming Dove products (Johnston and Taylor 2008; Millard 2009; Taylor, Johnston, and Whitehead 2016).

Prior research noted that consumers' reactions to femvertising are contingent on the extent to which the company's character, as revealed by its proposed social cause (e.g., women empowerment), and its own reputation are perceived to be congruent (Abitbol and Sternadori 2018; Sen and Bhattacharya 2001). For instance, Dove is criticized of being hypocritical because Dove and Axe (a brand of male grooming products that uses ads featuring sexy female models) have the same parent company, Unilever (Millard 2009). This skepticism toward Dove's "Campaign for Real Beauty" can be explained by the lack of fit between the proposed social cause (e.g., redefining beauty standards) and Unilever's company image (e.g., a company that promotes stereotypical women images). In addition, consumers may attend to other issues not directly related to the proposed social cause. For instance, Campbell et al. (2011) analyzed consumer conversations around YouTube video ads and found that some viewers inquired about the composer of the video ad's music in the comments.

Reception theory, which emphasizes the audience's active role in interpreting mediated messages (Calhoun 2002), sheds light on the different feelings consumers have toward Dove's "Campaign for Real Beauty." For instance, cultural studies theorist Stuart Hall (1980) was among the main proponents of reception theory and noted that audiences can interpret messages in three ways: preferred, negotiated, and oppositional. In the context of advertising, when consumers interpret advertising messages using the reference codes constructed by the advertiser, they adopt the *preferred* way of interpretation (e.g., "Thank you to [D]ove for delivering this important message: You don't have to be young or thin to be beautiful"). When consumers interpret advertising messages within the dominant meaning framework but reserve the right to appropriate the interpretation to adapt to their own life circumstances, they utilize the *negotiated* way of interpretation (e.g., "I love Dove's campaign for real beauty, but not much this one … bodies of most elder women just DON'T look like these in the commercial"). When consumers interpret advertising messages using alternative reference codes contrary to the dominant meaning fabricated by the advertiser, they use the *oppositional* way of interpretation (e.g., "Dove is owned by Unilever. Unilever also owns Axe. Now try to tell me they're pro-women").

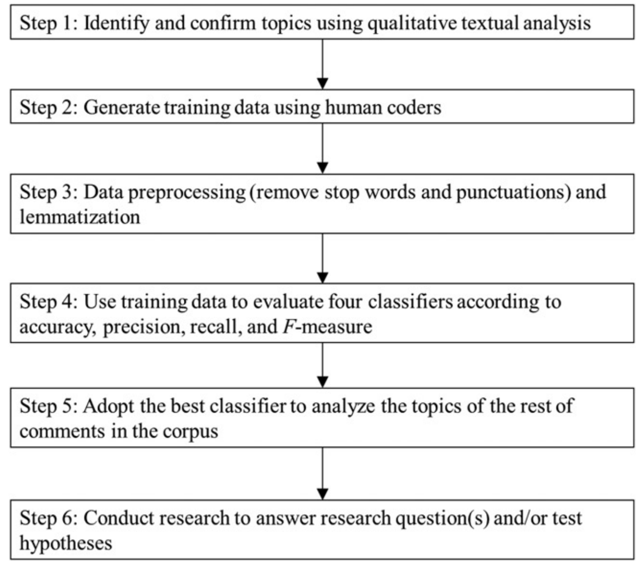

FIG. 1. Data analysis procedure.

The Role of Top-Ranked Comments on YouTube

YouTube viewers tend to post comments about a video; however, they seldom respond to others' comments or click "View Replies" to a comment (Eberle, Berens, and Li 2013; Halpern and Gibbs 2013; Walther et al. 2010). Many times, users browse only the first few comments (top comments) listed below the video (Walther et al. 2010). Thus, it seems that the content of a comment is more influential than that of a reply (Walther et al. 2010). Regarding the potential influence of top comments on viewers' reactions to a YouTube video, the social identity model of deindividuation effects (SIDE) provides a theoretical basis. According to SIDE, visual anonymity propels communicators to experience depersonalization, which refers to the situation when communicators are not able to discern individuals' interpersonal differences (Reicher, Spears, and Postmes 1995). Furthermore, depersonalization leads people to identify with others based on assumed similarity with the undifferentiated social group's members (Reicher, Spears, and Postmes 1995). In the context of YouTube, comments are not associated with pictures or visual icons depicting their contributors (Walther et al. 2010). Regarding visual anonymity on YouTube, SIDE predicts that viewers' attitudes are affected by online comments because they conform to anonymous sources as a result of depersonalization (Halpern and Gibbs 2013; Walther et al. 2010). Because users usually browse the first few comments (top comments) listed below each video, the top comments under a

TABLE 2
Initial Codes

Name	Example
Advertising criticism	Sure, you're beautiful, now buy our soap. I hate these manipulative ads.
Appreciation	Thank You, to the company of Dove. You just made my day even better.
Beauty criteria	as inspiring as this is… i truly believe beauty is in fact from within.. and whats on the outside is just the cherry on top per say;)
Comparison with model	I KNOW, I wear blue contacts and I still can't get them to look as good as hers do!
Discussion of broad issues	Dove does not stereotype. Or talks about how something as natural as a complexion makes any difference to your performance or lives.
Cynical comments	You are all beautiful! Just not as beautiful as me ;)
Dove's response	We're glad to see that this video touched you! Thanks for watching!
Emotional responses	This video just warms my heart, so emotional and earthbound! Can't forgot life-touching :)
Finish watching ad	me too!! at first i didn't realize it, but before i knew it, i had watched the whole thing.
Marketing sophistication	Marketing at its finest.
Negative comments	This commercial bothers me.
Neutral comments	Make-up makes a girl look beautiful, but photoshop is done for model/actress.
Positive comments	This is so fantastic.
Other view versus self-view	How we view ourselves is totally different how other people view us.
Self-confidence	I agree with your comment wholeheartedly, 101%. A teen version would be of great value, as too many kids are lost and have low self esteem.
Video music	Is anyone who know what is this music
Unrelated comments	whoa! super random, but i totally went to middle school with you hahaha. go highlanders!

femvertising video have the potential to influence viewers' perceptions of femvertising; therefore, it is necessary to examine the relationship between the topics of top comments and those of the overall viewer comments.

METHOD

Data Collection and Sample Size

To answer the research questions, we first searched YouTube using keywords, such as "Dove Real Beauty Campaign," and then sorted all videos based on number of comments. Next, we used a third-party scraper tool, the YouTube Comment Scraper (http://ytcomments.klostermann.ca/), to retrieve YouTube comments. This scraper was written in JavaScript and enabled us to overcome the limits of the YouTube application programming interface (API) and to collect the corpus of 26,820 users' comments on nine YouTube videos (these videos generated the largest number of comments during the time of collection) from Dove's "Campaign for Real Beauty." To use the scraper, we copied and pasted the uniform resource locator (URL) of each video clip in the "scrape" box, and then we saved the downloaded comments as a CSV file. We selected videos with the largest number of comments for two main reasons. First, if a video generates a large number of comments, it means that the video has significantly stirred discussion among viewers and then may serve as a typical case. Second, videos with the largest number of comments can generate enough data for analysis. In sum, a total sample of 26,820 comments (including replies) was collected out of the 28,117 comments posted below the videos on August 18, 2018. The missed 1,297 comments may be due to the privacy settings of some YouTube users that disallow data scraping (see Table 1). After removing non-English comments, replies, and comments in nonrecognizable characters, we included 21,605 comments. Further, we removed comments posted by Dove because in this study we focused on viewers' responses to the campaign. As a result, our final sample included 20,419 unique comments.

The Analysis Procedure

To analyze the topics of YouTube comments, we propose a six-step procedure (see Figure 1) that integrates qualitative textual analysis, quantitative content analysis, and machine learning–based data mining. We modified

TABLE 3
Automated Classification Results for Topics

Classifier	Overall Accuracy	Category	Precision	Recall	*F* Measure
NB	.67	Ad skepticism	.25	.14	.18
		Beauty definition	.68	.96	.79
		Praise	.66	.92	.77
		Discussion of broad issues	0	.00	.00
		Other	1.00	.23	.37
LR	.76	Ad skepticism	.25	.14	.18
		Beauty definition	.83	.92	.87
		Praise	.75	.83	.79
		Discussion of broad issues	.50	.20	.29
		Other	.81	.77	.79
SVM	.38	Ad skepticism	0	0	0
		Beauty definition	0	0	0
		Praise	.38	1.00	.55
		Discussion of broad issues	0	0	0
		Other	0	0	0
KNN	.53	Ad skepticism	.36	.36	.36
		Beauty definition	.51	.47	.49
		Praise	.59	.81	.68
		Discussion of broad issues	.39	.45	.42
		Other	.70	.26	.38

Note. NB = naive Bayes; LR = logistic regression; SVM = support vector machine; KNN = *k*-nearest neighbors.

the data mining approach proposed by Okazaki and colleagues (2014) in two ways. First, we incorporated literature review and qualitative textual analysis into the procedure to develop the coding scheme. This step is very important because researchers may not have a ready-to-use coding scheme on hand. Second, in addition to the three performance metrics (i.e., precision, recall, and *F* measure) that Okazaki and colleagues (2014) adopted to evaluate each machine learning–based classifier, we also examined the overall accuracy of each classifier. This was done to ensure that we had an understanding of the overall performance of each classifier. In this vein, this proposed procedure leverages the strengths of qualitative textual analysis, quantitative content analysis, and machine learning-based data mining. In the sections that follow, we explicate how to use the procedure step by step.

Step 1: Identify and Confirm Topics

Our review of femvertising literature sheds light on the current topics widely discussed regarding Dove's "Real Beauty" campaign. To develop a coding scheme, a qualitative textual analysis was conducted in NVivo. First, after calculating sample size with 95% confidence level (CI) and 3% CI, we randomly selected a sample of 953 comments from the corpus. In the first stage, one of the researchers immersed herself into the 953 comments to obtain a general sense of the whole. Then, she started open coding to identify initial codes. To do so, she carefully read each comment, looked for repeated words and similar terms, and grouped comments based on closeness of meanings. During this stage, 16 initial categories, such as advertising criticism, beauty criteria, self-confidence, and appreciation, were generated (see Table 2). After that, she reread all comments and paid close attention to relationships and internal logics among the 16 categories. Using the analytic induction method (Taylor, Hoy, and Haley 1996), she condensed the 16 codes into five potential themes. With ongoing analyzing and reviewing, five themes—advertising skepticism, definition of beauty, praise of advertising, discussion of broad issues, and other—were refined and confirmed (see Appendix for the coding scheme). Two measures were used to ensure the quality of the analysis: (1) peer review, during which a qualitative researcher who was familiar with the topic analyzed the data and reviewed the emerging themes; and (2) external audit, during which a qualitative researcher

TABLE 4
Topics in Top Comments of Each Video

Video	Topic	*n* (%)
Age	Definition of beauty	3 (50.00%)
	Ad skepticism	2 (33.33%)
	other	1 (16.67%)
Choice	Definition of beauty	4 (66.67%)
	Praise of ad	1 (16.67%)
	Other	1 (16.67%)
Diana	Praise of ad	3 (50.00%)
	Other	3 (50.00%)
Evolution	Definition of beauty	3 (50.00%)
	Praise of ad	2 (33.33%)
	Other	1 (16.67%)
Legacy	Definition of beauty	2 (33.33%)
	Other	2 (33.33%)
	Discussion of broad issues	1 (16.67%)
	Praise of ad	1 (16.67%)
Pressure	Praise of ad	4 (66.67%)
	Ad skepticism	1 (16.67%)
	Other	1 (16.67%)
Sketches	Praise of ad	3 (50.00%)
	Ad skepticism	2 (33.33%)
	Definition of beauty	1 (16.67%)
Sketches (6 min)	Definition of beauty	3 (50.00%)
	Praise of ad	2 (33.33%)
	Other	1 (16.67%)
Terms	Praise of ad	3 (50.00%)
	Definition of beauty	2 (33.33%)
	Ad skepticism	1 (16.67%)

who was not familiar with the topic examined the process and the product of the account, assessing accuracy (Creswell 2013).

Step 2: Generate Taining Data Set Using Human Coders

In this stage, we aim to generate a training data set to be used for machine learning–based data mining. Two human coders (to reduce bias, neither coder was involved in generating the coding scheme in Step 1) conducted a quantitative content analysis of the same 953 comments to examine the topics using the generated coding scheme in the Appendix. Before the formal coding process, 20 comments were randomly selected and coded by two coders for training purpose. The two coders discussed their interpretations of inconsistent coding results until a consensus was reached. Then, they independently coded 200 comments to check intercoder reliability. Intercoder reliability for topics (Cohen's kappa = .87) reached satisfactory level. Next, the two coders compared their coding results and had a discussion to resolve discrepancies. Finally, they split the rest of 733 comments and coded them in terms of topics. Thus, the two human coders labeled all 953 comments.

Step 3: Data Preprocessing and Lemmatization

To prepare for machine learning–based data mining, we cleaned the data using four steps. First, we eliminated stop words, such as *that*, *is*, and *for*, because these are common words in YouTube comments but they add little value to machine learning models (Liu, Burns, and Hou 2017). Also, we removed special characters and punctuations in the comments to avoid any distortions. Second, we performed lemmatization on words to reduce derivationally related forms of a word (e.g., *walked*, *walks*, *walking*) to a common base form (e.g., *walk*) so that the machine could treat all the derivationally related forms of a word as the same word. Third, we used a Python package (Scikit-learn) to randomly split the labeled training data set of 953 comments into training data (90%) and testing data (10%) so that the machine could learn the logic of human-labeled training data and use the logic to classify testing data without considering the original human-labeled results attached to the testing data. Then we could compare the classification results of testing data from the machine with those human-labeled results attached to the testing data to check the accuracy of the machine classification. Fourth, we adopted TF-IDF, a numerical statistic that is used to reflect how important a word is to a comment in the total sample, to select the most important words from the training data set and to convert these important words in both training and testing data into a matrix of tokens that the machine can read (Kalita 2015).

Step 4: Evaluation of Four Classifiers

In this stage, we use the training data that were classified (labeled) by human coders to train four machine learning classifiers to perform topic classification on the testing data: naive Bayes (NB; Forman 2003), logistic regression (LR; Cheng and Hullermeier 2009), support vector machine (SVM; Tong and Koller 2001), and *k*-nearest neighbors (KNN; Kwon and Lee 2003). We used the four classifiers to make predictions for the testing data in terms of topics and compared the predictions with the human-labeled results to measure the performances of the four classifiers through four metrics: accuracy, precision, recall, and *F* measure. Accuracy refers to the percentage of correctly predicted results among all sample comments. Precision measures the situation when a comment that is not coded as one category by human coders (coding result) is classified as that category by a machine learning classifier (predicted result). Recall gauges the situation when a comment is correctly classified according to its coding result. The *F* measure is a metric that balances precision and recall (Okazaki et al. 2014). Table 3 shows evaluation results of the four classifiers regarding topics, and LR turned out to be the best classifier among the four in this research.

Step 5: Analyze the Rest of the Comments in the Corpus Using LR

At this stage, we performed standard opinion mining on the rest of the comments in the corpus ($n = 19,466$) using LR as well as the other aforementioned three classifiers. As a result, the four classifiers made predictions for each comment regarding its topic. To further confirm that LR is the best classifier in this case, we invited a human coder, who was not familiar with the performance metrics of the four classifiers, to classify the topics of 20 randomly selected comments using the coding scheme in the Appendix. Next, the human coder compared her coding results with the predicted results of the four classifiers and confirmed that LR is the best classifier among the four. Therefore, when we analyzed the topics of the corpus, we adopted the results from LR for the 19,446 comments and combined them with those human-labeled results of the 953 comments.

Step 6: Conduct Research to Answer Research Questions

In the final stage, we performed several frequency analyses on the corpus to answer the two research questions. In the following section, we present the results.

RESULTS

Results for Research Question 1

Results from frequency analysis indicated that the two most frequent topics discussed in the comments were definition of beauty (35.71%) and praise of ad (35.33%), followed by other (15.43%) and ad skepticism (10.53%). The least frequent topic was discussion of broad issues (2.99%).

Among the corpus of 20,419 comments, 1,676 comments were “liked” by other video viewers at least once. Previous research noted that the number of “likes” provides social context for a social media–based message and serves as a signal of the message’s popularity and importance (Alhabash et al. 2013; Curran, Graham, and Temple 2011). In particular, a “like” indicates an individual’s positive attitude toward a message (Alhabash et al. 2013). In this vein, we analyzed the topics of the 1,676 comments that were “liked” at least once. On average, comments that discussed definition of beauty generated the largest number of likes ($M = 116.18$), followed by praise of ad ($M = 77.71$), other ($M = 56.39$), and discussion of broad issues ($M = 38.69$). In contrast, comments that focused on ad skepticism generated the least number of likes ($M = 35.70$).

Results for Research Question 2

To answer research question 2, we analyzed the top six comments listed below each of the nine videos from Dove’s “Campaign for Real Beauty.” Results indicated that definition of beauty tended to be the most frequently discussed topic in the top comments across the nine videos, followed by praise of ad (see Table 4). Notably, this pattern is compatible with the distribution of topics in the corpus, because definition of beauty (35.71%) and praise of ad (35.33%) were the two most frequent topics in the corpus, and these two topics together appeared in the majority of the total comments (71.04%).

DISCUSSION

In this study, using big data mining, we present a procedure that automatically predicts topics of YouTube comments based on human-coded results. Designed to tackle the challenge of big data in textual formats, this procedure effectively employs qualitative textual analysis, human-based content analysis, and machine learning–based data mining. Using 20,419 unique YouTube comments, this article illustrates how to follow the procedure to transform big data into femvertising insights and to answer two femvertising-related research questions. Our study has important methodological implications.

We provide detailed steps for identifying topics of YouTube comments from five categories. This classification technique is very useful in helping scholars and brand managers understand consumers’ overall attitudes toward a campaign as expressed on social media. This article provides the steps for labeling each comment. As a technical note for practical big data analysis, we recommend using Scikit-learn, a Python package to write the code for each classifier. This package is a simple and an efficient tool for data mining and analysis and is accessible to everyone. To achieve high classification accuracy, we recommend human coders go through the training data set multiple times to ensure the coding results are consistent and reliable. In this vein, when the machine learns the coding pattern of the training data set, it tends to make accurate predictions.

Limitations and Future Research Direction

Although we endeavor to maximize the quality of our work with appropriate analysis procedures and techniques, the proposed analysis procedure bears limitations. First, although the prediction accuracy of LR is quite high (.76), future researchers may give human coders intensive training sessions to prepare a refined training set to enhance the classification accuracy of machine learning algorithms. Malthouse and Li (2017) noted, "Social media is the world's largest focus group and can provide insights on what consumers think and feel about a brand" (p. 230). Once the prediction accuracy of machine learning classification remains high, machine learning approach can be a reliable tool for advertisers and marketers to uncover insights from online consumer conversations and to craft better advertising messages. Second, due to the nature of content analysis, we can identify only a similarity between the pattern of topic distribution in top comments and that in the corpus; however, we cannot determine the causal relationship between the two. This can be a limitation of using big data research, and future research may conduct a follow-up experiment to further test the influence of top comments on YouTube viewers' opinions toward femvertising. Third, in this research, we used a famous example of femvertising—Dove's "Campaign for Real Beauty"—to illustrate how to use the proposed procedure to analyze UGC on social media. Future research may examine other popular femvertising examples, such as the "Like a Girl" campaign by Always, to showcase how to use the proposed six-step analysis procedure. Moreover, to study a new subject, future research may need to review literature on the subject to get familiar with current conversations regarding the subject and then conduct qualitative analysis to generate a predefined coding scheme to be used for supervised text classification. Next, they can apply the proposed six-step analysis procedure. Fourth, because we cannot identify the gender of each commenter, we are unsure whether the positive online opinions toward Dove's "Campaign for Real Beauty" reflect its target consumers' opinions. Future research may conduct survey studies using a nationally representative sample to offset the limitation of using a big data approach. Fifth, because one of the goals is to showcase some sample UGC-related issues that the proposed procedure can address, we developed two research questions. Once the topics of the UGC are identified by using the proposed procedure, future research can develop more questions to address further issues, such as whether there is any relation between time of posting and type of topics. Finally, future research may adopt the proposed analysis procedure to contrast UGC on different social media platforms, such as YouTube versus Twitter, because prior research noted characteristics of social media channels may influence the content of UGC (Smith, Fischer, and Chen 2012). To fulfill this goal, after identifying possible themes from extant literature, future researchers need to examine UGC on different platforms through qualitative textual analysis to develop an inclusive coding scheme.

CONCLUSION

In this research, we proposed a six-step analysis procedure that combines qualitative textual analysis, human-based content analysis, and machine learning–based data mining. We showcased how the proposed analysis procedure successfully classified the topics of all 20,419 unique YouTube comments on Dove's "Campaign for Real Beauty" with satisfactory reliability. Advertising and marketing practitioners can consider adopting the proposed six-step analysis procedure to analyze the content of UGC on social media and to draw valuable consumer insights.

ORCID

Yang Feng http://orcid.org/0000-0003-1973-0014
Huan Chen http://orcid.org/0000-0003-2240-7381
Li He http://orcid.org/0000-0002-6681-0933

REFERENCES

Abitbol, Alan, and Miglena Sternadori (2018), "Championing Women's Empowerment As a Catalyst for Purchase Intentions: Testing the Mediating Roles of OPRs and Brand Loyalty in the Context of Femvertising," *International Journal of Strategic Communication*, 13 (1), 1–20.

Akestam, Nina, Sara Rosengren, and Micael Dahlen (2017), "Advertising 'Like a Girl': Toward a Better Understanding of 'Femvertising' and Its Effects," *Psychology and Marketing*, 34 (8), 795–806.

Alhabash, Saleem, Anna R. McAlister, Amy Hagerstrom, Elizabeth Taylor Quilliam, Nora Rifon, and Jef I. Richards (2013), "Between Likes and Shares: Effects of Emotional Appeal and Virality on the Persuasiveness of Anticyberbullying Messages on Facebook," *Cyberpsychology, Behavior, and Social Networking*, 16 (3), 175–82.

Bissell, Kimberly, and Amy Rask (2010), "Real Women on Real Beauty: Self-Discrepancy, Internalization of the Thin Ideal, and Perceptions of Attractiveness and Thinness in Dove's Campaign for Real Beauty," *International Journal of Advertising*, 29 (4), 643–68.

Calhoun, Craig (2002), *Dictionary of the Social Sciences*, Oxford, UK: Oxford University Press.

Campbell, Colin, Leyland F. Pitt, Michael Parent, and Pierre R. Berthon (2011), "Understanding Consumer Conversations around Ads in a Web 2.0 World," *Journal of Advertising*, 40 (1), 87–102.

Cheng, Weiwei, and Eyke Hullermeier (2009), "Combining Instance-Based Learning and Logistic Regression for Multilabel Classification," *Machine Learning*, 76 (2–3), 211–25.

Cinelli, Melissa D., and Lifeng Yang (2016), "The Role of Implicit Theories in Evaluations of 'Plus-Size' Advertising," *Journal of Advertising*, 45 (4), 472–81.

Creswell, John (2013), *Qualitative Inquiry and Research Design: Choosing among Five Approaches*, 3rd ed., Los Angeles, CA: Sage.

Curran, Kevin, Sarah Graham, and Christopher Temple (2011), "Advertising on Facebook," *International Journal of E-Business Development*, 1 (1), 26–33.

Eberle, David, Guido Berens, and Ting Li (2013), "The Impact of Interactive Corporate Social Responsibility Communication on Corporate Reputation," *Journal of Business Ethics*, 118 (4), 731–46.

Forman, George (2003), "An Extensive Empirical Study of Feature Selection Metrics for Text Classification," *Journal of Machine Learning Research*, 3, 1289–305.

Gill, Rosalind, and Ana Sofia Elias (2014), "'Awaken Your Incredible': Love Your Body Discourses and Postfeminist Contradictions," *International Journal of Media and Cultural Politics*, 10 (2), 179–88.

Hall, Stuart (1980), "Encoding/Decoding," in *Culture, Media, Language*, Stuart Hall, Dorothy Hobson, Andrew Love, and Paul Willis, eds., London, UK: Hutchinson, 128–38.

Halpern, Daniel, and Jennifer Gibbs (2013), "Social Media as a Catalyst for Online Deliberation? Exploring the Affordances of Facebook and YouTube for Political Expression," *Computers in Human Behavior*, 29 (3), 1159–68.

Jin, Seung-A Annie, and Joe Phua (2014), "Following Celebrities' Tweets about Brands: The Impact of Twitter-Based Electronic Word-of-Mouth on Consumers' Source Credibility Perception, Buying Intention, and Social Identification with Celebrities," *Journal of Advertising*, 43 (2), 941–66.

Johnston, Josee, and Judith Taylor (2008), "Feminist Consumerism and Fat Activists: A Comparative Study of Grassroots Activism and the Dove Real Beauty Campaign," *Signs*, 33 (4), 941–66.

Kalita, Deepshikha (2015), "Supervised and Unsupervised Document Classification: A Survey," *International Journal of Computer Science and Information Technologies*, 6 (2), 1971–74.

Khan, M. Laeeq (2017), "Social Media Engagement: What Motivates User Participation and Consumption on YouTube?," *Computers in Human Behavior*, 66, 236–47.

Knoll, Johannes (2016), "Advertising in Social Media: A Review of Empirical Evidence," *International Journal of Advertising*, 35 (2), 266–300.

Kwon, Oh-Woog, and Jong-Hyeok Lee (2003), "Text Categorization Based on *k*-Nearest Neighbor Approach for Web Site Classification," *Information Processing and Management*, 39 (1), 25–44.

Liu, Xia, Alvin C. Burns, and Yingjiang Hou (2017), "An Investigation of Brand-Related User-Generated Content on Twitter," *Journal of Advertising*, 46 (2), 236–47.

Malthouse, Edward C., and Hairong Li (2017), "Opportunities for and Pitfalls of Using Big Data in Advertising Research," *Journal of Advertising*, 46 (2), 227–35.

Millard, Jennifer (2009), "Performing Beauty: Dove's 'Real Beauty' Campaign," *Symbolic Interaction*, 22 (2), 146–68.

Okazaki, Shintaro, Ana Maria Diaz-Martin, Rozano Suplet, and Hector D. Menendez (2014), "How to Mine Brand Tweets: Procedural Guidelines and Pretest," *International Journal of Market Research*, 56 (4), 467–88.

Reicher, S.D., Russell Spears, and Tom Postmes (1995), "A Social Identity Model of Deindividuation Phenomena," *European Review of Social Psychology*, 6 (1), 161–98.

Sen, Sankar, and C.B. Bhattacharya (2001), "Does Doing Good Always Lead to Doing Better? Consumer Reactions to Corporate Social Responsibility," *Journal of Marketing Research*, 38 (2), 225–43.

Smith, Andrew N., Eileen Fischer, and Yongjian Chen (2012), "How Does Brand-Related User-Generated Content Differ across YouTube, Facebook, and Twitter?," *Journal of Interactive Marketing*, 26 (2), 102–13.

Taylor, Judith, Josee Johnston, and Krista Whitehead (2016), "A Corporation in Feminist Clothing? Young Women Discuss the Dove 'Real Beauty' Campaign," *Critical Sociology*, 42 (1), 123–44.

Taylor, Ronald E., Mariea Grubbs Hoy, and Eric Haley (1996), "How French Advertising Professionals Develop Creative Strategy," *Journal of Advertising*, 14 (1), 1–13.

Tong, Simon, and Daphne Koller (2001), "Support Vector Machine Active Learning with Applications to Text Classification," *Journal of Machine Learning Research*, 2, 45–66.

Walther, Joseph B., David DeAndrea, Jinsuk Kim, and James C. Anthony (2010), "The Influence of Online Comments on Perceptions of Antimarijuana Public Service Announcements on YouTube," *Human Communication Research*, 36 (4), 469–92.

Ye, Qiang, Ziqiong Zhang, and Rob Law (2009), "Sentiment Classification of Online Reviews to Travel Destinations by Supervised Machine Learning Approaches," *Expert Systems with Applications*, 36 (3), 6527–35.

APPENDIX: CODING SCHEME FOR TOPICS

1. *Advertising skepticism*—a comment criticizes the video for various reasons.
2. *Definition of beauty*—a comment discusses the definition of beauty from different perspectives.
3. *Praise of advertising*—a comment evaluates the video in positive ways.
4. *Discussion of broad issues*—a comment does not just discuss the video but addresses some societal and cultural issues, such as stereotype, gender, race, and so forth.
5. *Other*—a comment discusses minor issues of the video, such as the background music of the video, or other issues that are not related to the video.