Understanding Brand Image from Consumer-generated Hashtags

Keeyeon Ki-cheon Park
Hye-jin Kim

Follow this and additional works at: https://amj.kma.re.kr/journal

Part of the Marketing Commons

Recommended Citation
DOI: 10.15830/amj.2020.22.3.71
Available at: https://amj.kma.re.kr/journal/vol22/iss3/4

This Article is brought to you for free and open access by Asia Marketing Journal. It has been accepted for inclusion in Asia Marketing Journal by an authorized editor of Asia Marketing Journal.
Understanding Brand Image from Consumer-generated Hashtags

Keeyeon Ki-cheon Park*
Hye-jin Kim**

Social media has emerged as a major hub of engagement between brands and consumers in recent years, and allows user-generated content to serve as a powerful means of encouraging communication between the sides. However, it is challenging to negotiate user-generated content owing to its lack of structure and the enormous amount generated. This study focuses on the hashtag, a metadata tag that reflects customers’ brand perception through social media platforms. Online users share their knowledge and impressions using a wide variety of hashtags. We examine hashtags that co-occur with particular branded hashtags on the social media platform, Instagram, to derive insights about brand perception. We apply text mining technology and network analysis to identify the perceptions of brand images among consumers on the site, where this helps distinguish among the diverse personalities of the brands. This study contributes to highlighting the value of hashtags in constructing brand personality in the context of online marketing.

Keywords: hashtag, text mining, brand image, marketing research

I. Statement of intended contribution

This study yields insights into consumers’ use of hashtags and its impact on brands. We demonstrate how user-generated hashtags contribute to constructing brand personality in a market and examine changing perceptions of them among consumers. Comparing to the previous literature, our proposed method is relatively inexpensive by reducing the level of difficulty and subjectivity of text mining analysis. Therefore, our approach provides a simple and systematic way to monitor brand-associated information from massive amounts

* Ph. D. Student, Department of Business and Technology Management, Korea Advanced Institute of Science and Technology (parkky88@kaist.ac.kr)
** Assistant Professor, Department of Business and Technology Management, Korea Advanced Institute of Science and Technology (hyeinkim@kaist.ac.kr), Corresponding Author
of data.

From a practical perspective, the value of the hashtag is highlighted in the context of social media marketing strategies. The hashtag can be a valuable proxy that helps understand how consumers perceive a particular brand on social media platforms. This can help develop future marketing strategies that enhance competitive advantage. In addition, analyzing hashtags is relatively simple, so it is suitable for practitioners who struggle to collect and analyze the copious amount of social media content.

II. Introduction

As social media has become a crucial means of engagement between brands and consumers, the presence of major companies on social media has increased dramatically. User-generated content (UGC) in particular is a key medium that makes such engagement more effective. People use social media not only to search for information on products and services, but also to share their reviews through UGC. From the company’s perspective, UGC on such social media platforms as Twitter, Facebook, and Instagram allows companies to connect with new users, communicate with their followers, and thus enhance opportunities to increase sales. For example, a recent survey showed that 79% of Instagram users explore information about a new product through posts on the site, and 30% have made a purchase after discovering a new product on the site (Gotter 2020; Gallegos 2020). This indicates how UGC influences people’s trust in products, which in turn is connected with their purchasing decisions. Therefore, even the world’s big brands now use UGC as part of their marketing strategy (Netzer et al. 2012).

The above trend highlights the importance of investigating UGC in marketing research. The relevant literature has examined UGC to understand how consumers perceive brands, or the products and services offered by companies (Ghiassi, Skinner, and Zimbra 2013; Mostafa 2013; Liu 2006). Despite the wealth of information available in UGC, most studies have focused on its volume and valence because the massive amount of available data creates such difficulties as tracking and quantifying unstructured data. Netzer et al. (2012), however, have used a different approach to utilize UGC data by constructing networks to show the competitive market dynamics among brands. Nonetheless, the difficulty of analyzing the vast amount of unstructured data persists, Nam and Kannan (2014) instead used social tagging networks to demonstrate the informational value of social tags as a measure of brand performance and used it to understand the valuation of firms. While they used an online social bookmarking website to collect social tags, this paper focuses
on the hashtag, a representative metadata tag used on social media.

The hashtag is a means of engagement between online consumers and firms in social media. Firms create and promote their branded hashtags, and people subsequently share posts related to the brand and express their perceptions of it. Firms thus gain recognition for their brand while their understanding of its perception among consumers improves. Furthermore, they can follow the current trends and social issues through the real-time analysis of hashtags. Hashtag use is particularly encouraged in Instagram. Unlike other platforms such as Twitter and Facebook, there are fewer restrictions on hashtags in Instagram, which allow users to share more diverse information by using them. However, the previous literature has yielded limited insights into consumers’ use of hashtags and its impact on brands. The objective of this study is to use hashtags on Instagram to construct a method that measures how consumers perceive brands on social media. We constructed a measurement structure that involves calculating the term frequency – inverse document frequency (TF-IDF), establishing dimensions of brand personality, similarity calculation, brand image analysis, and competitive environment analysis. We first identified representative hashtags for each brand, and then measured semantic similarities between them and five dimensions of brand personality. The similarity results provided information not only on brand image, but also became links in subsequent network analysis. Then, we built networks for five brand personality dimensions to understand competitive environments.

The remainder of this paper is structured as follows. We review the literature on the hashtag, brand image, and text mining in Section 2, and introduce the data and the method used to conduct the analysis in Section 3. Section 4 details our analysis results, and Section 5 describes the limitations of this study and the implications of its results for research in the area.

III. Literature Review

The development of Information Technology and ubiquity of smartphones have allowed users to be constantly present on social media, prompting firms to adopt new approaches by using information on social media platforms for such diverse purposes as promotion, marketing, profile formation, and image creation (eMarketer 2010; Weinberg 2009). One of the key sources in social media platforms is hashtag. Hashtag marketing began when Twitter initiated its “Promoted Trends” in 2010 that allowed brands to increase brand awareness by their campaign hashtags (Twitter 2013). Then, hashtags have grown in popularity, so several social media platforms, such as Instagram and Facebook, subsequently joined in to incorporate hashtags.
Hashtags’ functions can be divided into categorization and communication (Nam and Kannan 2014). The former involves classifying and explaining the contents of posts, the latter is typically used to share information and express opinions on the contents. The classification function allows users to identify contents related to topics of their interest, then it can be implemented in advertising campaigns (Jackson and Foucault Welles 2015). Firms directly create branded hashtags to represent themselves and promote marketing campaigns. Branded hashtags provide people an easier way to search and collect information that enhances customer participation and attention. So, many industries including sports and fashion started using hashtags for consumer communication and advertising (Blaszka et al. 2012; Jones 2014; Shin, Chae, and Ko 2018). Even TV advertising uses hashtags to increase consumer engagement (Stathopoulou et al. 2017).

On the other hand, the expression function can help firms capture the voice of customers. Customers express their experience, impression, and emotions toward brands through UGC, including a variety of descriptive hashtags (Strohmaier, Körner, and Kern 2010). Thus, hashtags can clue firms in on what they are talking about. Since each person has a different knowledge structure, heterogeneity arises in hashtags (Nam and Kannan 2014), which in turn generates a variety of topics. The dynamic information of hashtags also comes from the unlimited possibilities, because there is no specific guideline that indicates how to summarize the general idea of the content in hashtags (Zain and Selamat 2019). For example, online users can create a hashtag with numerous keywords or they can generate trendy words that are not included in the dictionary. This unrestricted environment helps learn current issues by monitoring hashtags. However, few studies have examined the informal value of hashtags. Nam et al. (2018) analyzed the tags of a social bookmarking website to understand trending topics related to brands. On the other hand, this study’s focus is to uncover consumers’ brand image perception for individual brands using hashtags. We examine hashtags to show which brands share similar images and thus are in competition.

The brand image is a consumer’s perception of a particular brand as reflected in the overall impression store in his/her memory (Keller 1993). Brand image is essential for structuring brand equity, which impacts sales and profits. An affirmative image can strengthen a firm’s advantage over competitors, and can enhance customer satisfaction, loyalty, and intention to purchase (Zeithaml, Bitner, and Gremler 1996; Aaker 1996). Therefore, it is important to understand what makes a brand favorable. Brand perception comes from overall experience, use of the product, and online word-of-mouth recommendations in the consumer’s memory.
Hashtag, a major source of online word-of-mouth, can be a key component of consumer perception and a powerful tool to measure the brand image online.

To derive insights from unstructured text data such as hashtags, many studies have used text mining technology (Netzer et al. 2012; Akiva et al. 2008; Dave, Lawrence, and Pennock 2003). In particular, machine learning and statistics have been used to facilitate the analysis of unstructured data (Hotho, Nürnberger, and Paaß 2005). In marketing, for example, Lee and Bradlow (2011) analyzed the market structure by extracting product attributes and brand positions from online customer reviews. Netzer et al. (2012) used text mining and semantic network analysis to build a market structure. However, these studies used relatively complicated processing to handle vast amounts of unstructured data. In this study, we only analyze hashtags developed by consumers that can represent the entire content of posts with implicit meaning, which simplifies text mining while yielding performance similar to methods proposed in the literature.

IV. Methodology

4.1 Data

We chose Instagram for this study because it encourages hashtag use through laxer rules on their use than those of other social media platforms. Instagram has more than 700 million active users who share around 40 billion photos and videos by using hashtags. The form of the content, images or videos appended with short keywords, leads to more engagement between firms and people than does text-based content. For example, brands on Instagram have rates of engagement 10 times higher than those on Facebook (Gallegos 2020). In addition, compared with hashtags on other platform, those on Instagram account for the greatest part of text in content because users prefer to employ them as a form of expression instead of long texts. They may implicit comprehensive content in the hashtags. Therefore, we analyzed Instagram hashtags that contain implicit and dynamic information to study the brand image on social media.

For this study, we selected one of the most competitive markets on Instagram, the beauty industry (Health 2020). This industry has been redefined through visual social media platforms, such as Instagram and YouTube. UGC in particular can enhance the conversion rates for beauty brands (Mitic 2016). For example, 95% of beauty brands share posts on Instagram for greater engagement and interaction with consumers (Digidat 2015). This indicates that nearly all brands communicate their stories with consumers, share their values, and establish their images through Instagram posts. We chose
the 24 most popular beauty brands on Instagram for this study.

To collect hashtags relevant to a given brand, we gathered all hashtags co-occurring with the relevant branded hashtags in Instagram posts during December 2017. For example, if a post contained “#mac” in the text, we assumed that the relevant posts contained semantic contents regarding the brand Mac (see Figure 1). We then collected all other hashtags in the posts. The total number of Instagram posts hence collected was 88,793,851. Such irrelevant information as duplicate posts, commercials, and non-hashtag posts were removed from the dataset. We eventually obtained 1,406,389 unique hashtags describing brands. As they had already been broken down into one word, tokenizing was unnecessary during preprocessing.

4.2 Method

Figure 2 shows our method. It consists of six steps from data collection to building competitive structures. Step one is the data collection and preprocessing step, described above. In step two, we calculated the TF-IDF of each hashtag.
to derive the representative hashtags best describing a given brand. TF-IDF identifies how relevant a given word is to a specific document by calculating the relative frequency of its use in the document compared with the inverse percentage of words over the entire document body (Ramos 2003). Higher TF-IDF values reflect more frequently mentioned attributes. Intuitively, the value increases if the number of times a particular word appears in the document increases, but it is offset if the number of documents that contain that word increases (e.g., words like “the,” “this,” “what,” which don’t mean much).

The TF is a numerical value of the frequency of a given word in a document while the IDF is the reciprocal of this value, representing the number of documents in which the word has been mentioned. Assuming D is the entire document, w is a word, and individual documents belong to the entire sets of documents \((d \in D)\), the frequency of the word can be calculated as follows:

\[
w_d = f_{w,d} \times \log \left( \frac{|D|}{f_{w,D}} \right)
\]

where \(f_{w,d}\) indicates the number of times word w appears in document d, \(|D|\) indicates the size of the corpus, and \(f_{w,D}\) represents the number of times w appears in D.

In our dataset, each post was considered a document and all posts on each brand were combined into one document. For significant results, we removed all stop words in the documents. After calculating TF-IDF values for words in each document to analyze its content, we extracted 100 words with the highest TF-IDF values for each brand. Figure 3 shows an example of hashtags with the highest TF-IDF values for four brands. They
were the mostly mentioned words that consumers used to describe each brand, so those hashtags were considered representative.

In step three, we constructed brand personality dimensions following Aaker’s (1997) framework to describe how customers perceive a given brand (see Figure 4). Aaker’s brand personality framework consists of five dimensions: sincerity, excitement, competence, sophistication, and ruggedness. For each dimension, we constructed a dictionary of words that can represent the meaning, such as synonyms.

In step four, we measured the semantic similarity between each dictionary word representing brand personality and each hashtag using pre-trained fastText word embedding, an open-source library for learning text classification and representation learned by Facebook. Word
embedding is a technique in machine learning to represent words by an array of numbers, a vector. The word vectors then can capture hidden information about a language, like word analogies or semantic. In case of fastText word embedding, words are represented as 300-dimensional vectors, so in our analysis we transformed each word accordingly into vectors. Then, we used cosine similarity to measure how similar the words are. Using the dictionary word-hashtag similarity calculations, we calculated the mean of similarity values for the five dimensions for each brand, which were used to build a bipartite graph in the next step. For each brand, we created links for the bipartite graph with the five dimensions if the calculated similarity value is above the mean. The bipartite graph visually shows which brand personality dimension each brand is connected in consumers’ minds.

In the final step, we built a competitive structure for each brand personality dimension. Using the mean similarity value on each dimension for each brand, we selected the brands that have higher mean similarity values than average for each dimension. Those brands were considered to share similar brand personality, so we connected them in a network. The link weight, which suggests how close the link between two brands are, was determined by the number of co-occurrences in posts, i.e., the number of times two brands were mentioned together in a consumer’s post. This is based on the rationale that when brands are co-mentioned in consumers’ post, consumers perceive them as close competitors.

V. Results

The notable observation in our data were that the hashtags generally consisted of informal markup of languages without any rules or exact definitions, including slang words, cultural words, and obscure words. They did not have precise definitions or meaning but rather followed current language trends. This indicates a new linguistic ecosystem that social media has created. For instance, #potd is an acronym for “photo of the day”, “picture of the day”, or “poll of the day.” It has no literal definition as a word but was one of the most commonly used keywords in our dataset, with a high TF-IDF. Also, online users do not always create hashtags with single keyword. They usually combine numerous words without a space into a single hashtag. For example, #powerofgentle combines three words, “power”, “of”, and “gentle.” These characteristics of hashtags differentiates them with other social tags that use clear keywords for search function, such as the ones used by Nam et al. (2018).

Table 1 describes the results of the average similarity values between five brand personality dimensions and hashtags for the 24 beauty
brands in terms of TF-IDF. A high mean similarity value showed a higher volume of hashtags relevant to the given dimension of brand personality. Garnier has the highest mean similarity value in sincerity dimension, 0.175. Consumers consider this brand the most genuine, honest, wholesome, and cheerful. Dove has the highest similarity values in two dimensions, excitement and sophistication. Dove has differentiated itself by advocating the campaigns for confidence in women and self-esteem in young girls. Their campaigns may derive youthful and feminine traits from consumers’ minds. And our results showed Schwarzkof and Nivea have the highest mean similarity values in competence and ruggedness, respectively. Figure 5 shows the bipartite network, connecting brands with brand personality.

<table>
<thead>
<tr>
<th></th>
<th>Sincerity</th>
<th>Excitement</th>
<th>Competence</th>
<th>Sophistication</th>
<th>Ruggedness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avon</td>
<td>.142</td>
<td>.160</td>
<td>.127</td>
<td>.151</td>
<td>.139</td>
</tr>
<tr>
<td>Axe</td>
<td>.159</td>
<td>.171</td>
<td>.146</td>
<td>.156</td>
<td>.138</td>
</tr>
<tr>
<td>Chanel</td>
<td>.163</td>
<td>.172</td>
<td>.155</td>
<td>.164</td>
<td>.152</td>
</tr>
<tr>
<td>Christian Dior</td>
<td>.166</td>
<td>.176</td>
<td>.142</td>
<td>.170</td>
<td>.143</td>
</tr>
<tr>
<td>Clarins</td>
<td>.151</td>
<td>.164</td>
<td>.138</td>
<td>.157</td>
<td>.141</td>
</tr>
<tr>
<td>Clinique</td>
<td>.153</td>
<td>.174</td>
<td>.139</td>
<td>.166</td>
<td>.145</td>
</tr>
<tr>
<td>Dove</td>
<td>.168</td>
<td>.184</td>
<td>.151</td>
<td>.181</td>
<td>.153</td>
</tr>
<tr>
<td>Estée Lauder</td>
<td>.150</td>
<td>.173</td>
<td>.141</td>
<td>.161</td>
<td>.142</td>
</tr>
<tr>
<td>Garnier</td>
<td>.175</td>
<td>.181</td>
<td>.154</td>
<td>.173</td>
<td>.153</td>
</tr>
<tr>
<td>Gillette</td>
<td>.146</td>
<td>.158</td>
<td>.132</td>
<td>.156</td>
<td>.142</td>
</tr>
<tr>
<td>Head and Shoulders</td>
<td>.167</td>
<td>.182</td>
<td>.145</td>
<td>.165</td>
<td>.152</td>
</tr>
<tr>
<td>Lancôme</td>
<td>.174</td>
<td>.183</td>
<td>.152</td>
<td>.175</td>
<td>.157</td>
</tr>
<tr>
<td>L’Occitane</td>
<td>.160</td>
<td>.166</td>
<td>.135</td>
<td>.164</td>
<td>.138</td>
</tr>
<tr>
<td>Loreal</td>
<td>.166</td>
<td>.167</td>
<td>.142</td>
<td>.163</td>
<td>.140</td>
</tr>
<tr>
<td>Mac</td>
<td>.151</td>
<td>.156</td>
<td>.141</td>
<td>.155</td>
<td>.141</td>
</tr>
<tr>
<td>Maybelline</td>
<td>.157</td>
<td>.173</td>
<td>.156</td>
<td>.161</td>
<td>.150</td>
</tr>
<tr>
<td>Neutrogena</td>
<td>.166</td>
<td>.173</td>
<td>.148</td>
<td>.171</td>
<td>.158</td>
</tr>
<tr>
<td>Nivea</td>
<td>.169</td>
<td>.180</td>
<td>.148</td>
<td>.169</td>
<td>.161</td>
</tr>
<tr>
<td>Olay</td>
<td>.159</td>
<td>.161</td>
<td>.149</td>
<td>.158</td>
<td>.147</td>
</tr>
<tr>
<td>Pantene</td>
<td>.167</td>
<td>.180</td>
<td>.147</td>
<td>.168</td>
<td>.147</td>
</tr>
<tr>
<td>Rexona</td>
<td>.166</td>
<td>.175</td>
<td>.147</td>
<td>.157</td>
<td>.141</td>
</tr>
<tr>
<td>Schwarzkopf</td>
<td>.157</td>
<td>.168</td>
<td>.159</td>
<td>.158</td>
<td>.151</td>
</tr>
<tr>
<td>Shiseido</td>
<td>.160</td>
<td>.173</td>
<td>.140</td>
<td>.168</td>
<td>.152</td>
</tr>
<tr>
<td>The Body Shop</td>
<td>.141</td>
<td>.147</td>
<td>.133</td>
<td>.142</td>
<td>.131</td>
</tr>
</tbody>
</table>

(Table 1) Mean similarity results
dimensions. Interestingly, most beauty brands have lower mean similarity values on both competence and ruggedness dimensions. This result indicates the characteristic of the beauty industry. This industry has products that are generally used to supplement the appearance of the human body, so the brands in the industry promote glamorous, young, cheerful, and feminine features of their products, rather than intelligent, rugged, and tough features. Therefore, consumers perceive those traits from the beauty brands in our results as well.

The networks that explain competitive structure in each brand personality dimension is presented in Figure 6. These networks show brands that are perceived to have similar brand images by online customers. We can obtain the dynamics of competitive structure such as the number of competitors, and relative strength of each competitor from our network results. The brands that have links in the networks are those sharing the same brand personality in consumers’ minds. The link color indicates how close competitors they are. The darker a link is, the closer they are in competition. For example, the sincerity dimension network shows that Christian Dior and Chanel are the closest competitors in consumers’ minds. In the competence dimension, three brands, Lancôme, Chanel, and Maybelline, are relatively close competitors. In all dimensions, Lancôme and Chanel are in the close competition that shares similar brand image. In other words,
Figure 6: Competitive structures on each brand personality dimension
online consumers perceive similar values for these two brands in the market.

VI. Conclusion

This study examined the value of hashtags. We showed how hashtags contribute to constructing brand personality in a market, and examined changing perceptions of them among consumers. Our results showed that hashtags can help users comprehend the complex and dynamic information associated with a brand.

We developed a scale to measure brand image. Previously proposed measurement scales tended to be complicated because they used a large volume of unstructured text data. However, our proposed method uses condensed keywords, hashtags, and its performance was verified here. The scale is relatively simple and inexpensive, which reduces the level of difficulty of text mining analysis. Furthermore, it can reduce subjectivity in defining brand image based on textual information. In defining the textual pattern, previously proposed approaches used at least some level of human judgment, for example, to subjectively assign the topic representation of a group of words in topic modeling. However, we constructed the representative dimensions first, and then linked them with the considered brands based on the calculated similarities between the dimensions and the descriptive texts concerning the brands. Therefore, our approach provides a simple and systematic way to monitor brand-associated information from massive amounts of data.

From a practical perspective, the value of the hashtag was highlighted in the context of online marketing strategies. The hashtag can be a valuable proxy that helps understand how consumers perceive a particular brand on social media platforms. Marketing managers should thus examine hashtags to identify their value in brand positioning and building competitive structures. This can help develop future marketing strategies that enhance competitive advantage. In addition, analyzing hashtags is relatively simple, because of which our proposed method is suitable for practitioners who struggle to collect and analyze the copious amount of social media content.

However, there are limitations to this study. First, a sufficient number of brand-descriptive hashtags may not always be available for all brands. If a brand does not have sufficient engagement in consumer posts, enough data might not be available to obtain the required information. Second, the interpretation of a hashtag implies interpreting trends in society. Social media channels do not adhere to traditional rules of language. Social media users may associate new meanings with old words. Thus, interpretation can be challenging. Third, we analyzed only the 24 most popular beauty brands in this study, which resulted in only
two communities in the results. To increase the generalizability of the results of this study, future work should consider a larger number of brands.

<Received September 26, 2020>
<Accepted October 27, 2020>

References


Cosmetic Products [@k.paaark]. (2020, October 27). #Mac #studentkit #eyeshadow #lippencil #eyebrows #powderblush #makeup #products #makeupfashion #loveyourself #instagram [Photograph]. Instagram. https://www.instagram.com/p/CG1d1kglrn0/.


