Using text analytics to measure an effect of topics and sentiments on social-media engagement: Focusing on Facebook fan page of Toyota

Heuiju Chun¹, Byung-Hak Leem² and Hyesun Suh³

Abstract
In this study we investigate whether Facebook fan-page posting types and topics have a significant effect on engagement. More specifically, the media type and content theme of posting on Facebook are examined to see whether or not there was a difference between content topics. In order to achieve this goal, we set hypotheses as follows: (1) the media types of posting have a significant effect on engagement; (2) the topics and sentiment polarity of posting have a significant effect on engagement. We tested these hypotheses using research procedures as follows: (1) collection and preprocessing of social-media data, including posting types, comments, and reactions on Facebook fan pages, (2) topic modeling of fan-page postings using R and SAS, (3) testing hypotheses using a negative binomial regression model, and (4) implications and insights for social-media marketing. Topic modeling applying to textual data and sentiment analysis were conducted. After that, in order to find the factors to affect the number of Facebook fan-page engagements, the negative binomial regression model including post type, topic, sentiment, reactions of “love,” “haha,” and their interaction as exploratory variables was considered. Finally, the results show that post type is the most influential factor to affect social-media engagement, and content topics, sentiments of posts and comments also have significant effects on it.

Keywords
Topic model, sentiment analysis, LDA, GLM, Facebook fan page, engagement, negative binomial regression

Date received: 21 September 2020; accepted: 6 April 2021

Introduction
With the development of WEB 2.0, social-media marketing is becoming increasingly attractive. From a marketer’s point of view, social media have a high potential that is hard to ignore today. Social media have become a leading digital communications channel for sharing information with customers and providing them with up-to-the-minute information, such as on products, services, and events.

Social media are now an important marketing tool for most businesses. According to a survey by CMO (Chief Marketing Officers), about 12% of the marketing budget at the end of 2019 was spent on social media, an amount expected to increase by 20% over the next 5 years,¹ because 88% of adults aged 18–29 use social media, and since 2000, the Millennial generation has relied on social media to make purchasing decisions. Thus, indicators of social-media use are an important concern and source for

¹Department of Statistics and Information, Dongduk Women’s University, Seoul, Korea
²Department of International Marketing, Busan University of Foreign Studies, Busan, Korea
³Department of Computer Science Engineering, Daejin University, Pocheon, Korea

Corresponding author:
Hyesun Suh, Department of Computer Science Engineering, Daejin University, Pocheon 11159, Korea.
Email: jako403@daejin.ac.kr

Creative Commons CC BY: This article is distributed under the terms of the Creative Commons Attribution 4.0 License (https://creativecommons.org/licenses/by/4.0/) which permits any use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (https://us.sagepub.com/en-us/nam/open-access-at-sage).
marketing strategy. The most important one of the indicators is engagement. Companies are interested in social-media engagement because they can provide social proof of business, cost-effective advertising, and brand awareness by means of social media. For this reason, companies build their brand fan pages on Facebook and engage in marketing strategies that increase engagement by means of interaction with consumers.

Scholars, on the other hand, have identified various factors that affect social media and online engagement in a multifaceted viewpoint. One type of study is based on use and gratification theory (UGT) and social-exchange theory (SET). UGT proposes that social-media engagement depends on the customer's values or utility of, such as information value, hedonic value, or social value.2 Thus, fan pages on social media should deliver these experiential values to customers.

Wang and Liu3 found that individuals’ social interactions on social media based on social-exchange theory (SET) increase social capital and then support individuals' online engagement. Dolan et al.4 investigate the relationships between content types (informational, remunerative, entertaining, and relational) and the four types of engagement behavior (consuming, liking, sharing, and commenting). They highlight strategic development of social-media content to increase engagement among consumers. Therefore, the posts on fan pages should include informational, remunerative, entertaining, and relational contents.

Another is the social-media engagement (SME) theory5 based on value co-creation theory.6 The central point of SME theory is that higher user engagement leads to greater usage of the social-media platform. Usage is defined as the frequency of a user’s contribution, retrieval, and/or exploration of content within a social-media site.7,8 The more frequently users take part in a variety of activities, the more valuable the social-platform drives into the co-creation of value between firms and users.7,8 The value co-creation consists of co-production and value-in-use.9 Therefore, in order to increase customer engagement by means of value co-creation, the posts on the fan page must be co-produced with the customer and worthy of use by the customer. Recent studies have shown that online users tend to select information that adhere to their system of beliefs, ignore information that does not, and join groups that share a common narrative.10 On the other hand,11 are studied light on alternative news media ecosystems that are believed to have influenced opinions and beliefs by false and/or biased news reporting during the 2016 US President Elections. Existing research has explored how engagement affects the type of posts on social media, but has not considered the content of posts that reflect the three theories described above in a quantitative and qualitative study. Therefore, this study provides a model that can improve engagement by extending into not only posting type but also posting contents of fan pages.

The social-media marketing is possible based on big data analysis. To assess Big Data Analysis value, a conceptual model is proposed based on a resources-based view and dynamic capabilities theories.12 To empirically test this model, the study addresses a survey to a wide range of 500 European firms and their IT and business executives.13 Mikalef et al.14 proposed that big data analysis capability enable firms to generate insight that can help strengthen their dynamic capabilities, which in turn positively impact incremental and radical innovation capabilities. Such structured or unstructured big data analysis can be used to innovate or improve corporate business processes by identifying the factors of service failure and recovery for consumers.

Therefore, we intend to provide suggestions to increase social-media engagement by means of topic modeling and sentiment analysis of fan-page posts on Facebook based on bid data analysis. The subsequent sections of the paper are organized as follows: in the “Literature review and hypotheses” section, we suggest our research model based on the theoretical literature reviews on the subject. In the “Text mining and analytics” and “Hypothesis test” sections, we demonstrate empirically the research method and process. In this section, we extract topics by means of text analysis of Facebook fan-page postings and analyze the relationship between these topics and engagement. Finally, in the “Discussion and concluding remarks” section, we conclude by summarizing the results of the empirical analysis in the “Hypothesis test” section and presenting discussions and future research work.

Literature review and hypotheses

Social media analytics

Social media are an Internet-based application of WEB 2.0 that allows users to create and exchange content directly.15,16 Users can upload their photos, videos, and texts via social media using these Internet and Web-based technologies to share their ideas, feelings, opinions, and experiences with other users.15,17 Companies build emotional relationships with customers based on the interaction characteristics of these social media and use social media as an integrated marketing tool to manage the content, time, and frequency of information shared with customers.18

As a way to leverage the potential of social media, companies are increasingly using social media online forums, instant messaging services, and mobile smart platforms to communicate and collaborate on an enormous scale of personal and group trust.17,19 The second way is for companies to establish brand fan pages to interact and communicate with the brand community. Companies can post branded content on their pages, such as videos, brand messages, content, and other coupons that can be shared with fans by means of these brand fan pages. Consumers can click on the “Like” button in their brand fan page and become a
fan and interact with other consumers in the brand community. The positive or negative consumer footprint of such a product or service on brand fan pages can have a significant effect on the organization. For this reason, social-media network analysis and research on the consumer footprints in social media are attracting attention today.\textsuperscript{20}

**Topic modeling**

Topic modeling is a research method used in text mining. The most popular topic modeling technique is Latent Dirichlet Allocation (LDA), which is a generative probability distribution model for finding latent meaningful topics in the literature literature.\textsuperscript{21,22} Most of the data used in topic modeling consists of various types of unstructured text data, such as websites and online advertising,\textsuperscript{23} social-media publishing, and online product reviews.\textsuperscript{24} There are other data, such as images,\textsuperscript{25} purchase records,\textsuperscript{26} mobile app usage history,\textsuperscript{27} and traces of Internet search.\textsuperscript{28,29} Reisenbichler and Reutterer\textsuperscript{30} found that topic modeling is used in online text consumer review and service research, sales/retail business, social media, image and cross media, research on marketing literature, and public relations in marketing.

Topic modeling consists of three stages. First is the collection of related data. The second is preprocessing of collected text data, and the third is topic modeling using machine learning techniques. Data preprocessing consists of four steps: removal of special characters and tokenization from words, removal of abbreviations, removal of stop words, and word-document matrix creation.

The two-dimensional matrix structure of the vector-space representation, which is the final result of the preprocessing step, is used as input to the LDA algorithm of topic modeling. The topic modeling using the LDA model is an iterative algorithm that follows the following steps: 1. Start with parameters ($\alpha$, $\beta$), 2. Initialize topic assignments at random, 3. Repeatedly sampling topics for each word in each document, and 4. After all iterations are completed, calculate the results and evaluate the final model.

**Sentiment analysis**

Sentiment analysis, called opinion mining, refers to expressing the body of a document with positive or negative opinions or polarity, just as topic modeling summarizes the content of the document. The interest in sentiment analysis has increased with the amount and value of online text. Shoppers routinely read the posted reviews before choosing a product, hotel, or restaurant. Better reviews generate higher profits. For example, Luca\textsuperscript{31} found that adding another star, a score for Yelp, earns 5–9\% more in the restaurant.

The most direct approach to sentiment analysis relies on a dictionary that assigns a positive or negative value to a predefined list of words. The frequency of words in these dictionaries determines the emotions you assign to the document. The disadvantage of this algorithm, which uses emotional dictionaries to calculate the frequency of positive or negative words, is that positive words cannot be evaluated for satirical expression. Thus, without a good understanding of how words are used, a good dictionary is not enough to judge the feelings of a text. To make better judgments, learning software is required that understands text. Therefore, a recent state-of-the-art classifier for sentiment analysis uses deep learning, a predictive model built on large-scale neural networks.\textsuperscript{32,33}

**Social-media engagement**

The term “engagement” has recently emerged as a study of concepts, such as customer brand engagement and online brand engagement, in marketing.\textsuperscript{34} The concept of customer engagement has been understood as a multidimensional concept consisting of various combinations of emotional, cognitive, and behavioral factors rather than a single dimension.\textsuperscript{34,35} In particular, consumer engagement behavior is driven by customer voluntary motivations beyond product purchase, such as product, brand word of mouth, referral, other customer assistance, blogging, and writing reviews.

Social-media engagement, one of online or digital engagement, is defined as active involvement or participation in media, and is more specific to media than other engagements are. Social-media engagements have introduced consumer behavior (likes, comments, and shares) in posting of fan pages and recent empathy emotions such as love, haha, sad, and angry. The social-media engagement in this study is defined by the consumers’ reaction to postings and the sum of sympathy activities. Therefore, engagement of the brand fan page of a company can measure the effectiveness of the posting to show the quality of posted content. Thus, how effectively a fan engages their audience in posting is the most influential factor in winning social-media marketing over competitors.

To date, research identifying factors that influence consumer engagement has been conducted based on the Use and Gratification Theory of Klepek,\textsuperscript{2} who verified that the Hedonic and Social values are determinants of customer engagement on the social network of Facebook. Luarn et al.\textsuperscript{36} demonstrated that the media and content type of posts exert a significant effect on user online engagement. Irena Pletikosa and Florian Michahelles\textsuperscript{37} show that there is a different effect of the analyzed factors, such as content type and media type of fan-page posting over individual engagement measures. Mariani et al.\textsuperscript{38} suggest that engagement is positively affected by posting visual content (namely, photos) and posting during the weekends, and is negatively affected by evening posting. Le\textsuperscript{39} suggests that content type and discussion topic partially influence some factors of the online engagement metrics.
The usefulness of social media today depends on how information is shared with consumers. This information is presented by the posting that affects “Like,” “Comments,” or “Sharing” on Facebook. Facebook offers three types of postings: Text, Photo, and Video. These types make consumers experience in a variety of ways. Consumers passively read text-based postings, but actively participate in multimedia postings of photos and videos to increase engagement. Therefore, the more dynamic the social-media posting, the more likes, comments, replies, and empathic responses. In social media, text has a static property, and photos and videos have a dynamic property. The media type of posts, such as Text, Photo, and Video, affects social-media engagement on Facebook. Therefore, we set a hypothesis that the more dynamic the post type, the higher the social media participation.

**Hypothesis 1.** The higher the posting dynamics of the fan page, the higher the engagement will be.

And the content type, such as information, hedonic and social value, also affects social-media engagement. But the content types, not by UGT or SET, but by social media technical properties or social interactions, will influence the social-media engagement. Cvikiki et al. argued that the interaction of social media fans can be increased according to the category of posts, such as information, statements, advertisements, and announcements. In addition, McLachlan argued that posting as a topic suitable for customers can increase social media engagement. Therefore, we set the hypothesis that user participation differs according to the topic of the post.

**Hypothesis 2.** The engagement of the fan page posts will depend on topics.

Positive or negative interaction by means of Facebook fan-pages between firm and customer results in value co-creation or co-destruction, respectively. The positive polarity of posting content can increase the positive participation of consumers and contribute to the co-creation of value, whereas the negative polarity can increase the negative participation of consumer and contribute to the co-destruction of value. Thus, the degree of consumer engagement in social media depends on the sentiment polarity of post content. Therefore, we set the following hypothesis considering the sentiment polarity of post content.

**Hypothesis 3.** The positive or negative polarity of sentiment affects the engagement.

**Text mining and analytics**

This chapter describes the process of collecting and analyzing data generated directly by all activities of the business and consumers in the Facebook fan page, and then interpreting the results. This study will analyze Toyota, a global company that is actively engaging in marketing campaigns on Facebook fan pages in the United States. This study is conducted by a three-step process of analyzing Toyota's Facebook fan page in the US: data extraction and preprocessing, topic modeling, and sentiment analysis are presented in Figure 1.

**Data extraction and preprocessing**

There are many ways to extract Facebook data, but in this study data were extracted using Netvizz software, which is commonly used. The Facebook fan-page data were extracted from September 1, 2017, to August 31, 2018. The data extracted from this fan page were posting data such as videos, photos, and text of postings, response activity data,
such as comments, reactions, and sharing, and finally emotional reaction data, such as “love,” “haha,” “wow,” “sad” and “angry.”

In this study, we did data preprocessing, topic modeling, and sentiment analysis using R, and SAS software. As shown in Figure 2, preprocessing extracts prototypes of words by means of tokenization, and removes stop words and meaningless words, such as spam. We constructed and processed a specific word, a synonym, and an exception word dictionary.

**Topic modeling**

When applying LDA to textual data, we must choose the number of topics to be generated. According to Westerlund, et al., the optimal number of topics should be found by trial and error to avoid overlap and to ensure the interpretability of the topic. According to Calheiros et al., fewer topics help to avoid overlap between topics and help to ensure structural validity. Maier et al. suggested that it is practical to choose a number of possible topics that is consistent with the theoretical concept, the research context, and the purpose of the study. It is when the topics are divided exclusively that the best interpretable topic becomes the most optimal number of topics, and the words that appear in the top 10 do not overlap with each other. In this study, k-means clustering and Elbow method with distance between documents by words were used to find optimal number of topics, and k = 5 of LDA was chosen as the optimal number of topics, as shown in Figure 3.

The latent topic is chosen in the documents by finding the Dirichlet distribution parameters α for generating the topic distribution of the document and β for generating the word distribution of the topic along with the number of topics, K. According to Griffiths and Steyvers, the choices of α and β can have a significant effect on the LDA outcomes. They recommended $\alpha = 50/k$ (k: number of topics) and $\beta = 0.1$. The reason for using a small value for $\beta$ was to make more topics to represent a specific area of research. In this study, we set a topic probability distribution for each document with an $\alpha$ of 10 and $\beta$ of 0.1, and 5 topics. The topic distribution graph by 20 words and word clouds in each topic are shown at Figure 4 and Figure 5.

The main words of topic 1 are “door,” “call,” “custom,” “feature,” “lease,” “care,” “time” and “cover,” which seem to be related to car warranty service. Topic 2 and topic 5 are associated with a marketing campaign whose main word is “letsgoplaces.” Topic 2’s words are related to the 2019 new model, but those of topic 5 with the 2017 and 2018 models. The main words of Topic 3 are “truck,” “guy,” “engine,” “mechanic,” “tundra” and “care,” which are close to words related to “trucks.” Words of topic 4 are composed of “sales” such as “option,” “dealer,” “protection,” “tex,” and “replace.”

**Sentiment analysis**

This study used R software for analysis and did sentiment analysis with a general-purpose lexicon of Bing from Bing Liu and collaborators. Sentiment analysis assigns a polarity (positive, negative, neutral) to a document or to each topic. Figure 6 shows the main positive and negative sentiment words extracted in this study. Positive words are “love,” “protection,” “helped,” “won,” “beautiful,” and “support,” whereas negative word are “issue,” “limited,” “miss,” “lost,” “refuse,” “failed,” and “broke.” Therefore, it can be seen that the five topics were well classified.

Figure 7 and Table 1 show that topic 1 and topic 3 have more negative documents than positive documents, whereas topics 2, 4, and 5 are vice versa. That is, customers...
Figure 4. Frequency of top 20 words in 5 topics.
were unsatisfied with warranty service and trucks, but satisfied with the marketing campaign and sales.

**Hypothesis test**

The number of engagements of the fan-page posts, which is a dependent variable of the statistical model is a count data with a non-negative integer. Poisson regression model can be considered as an appropriate statistical model. However, Poisson distribution should have the same mean and variance, but this data has over-dispersion problem because mean and variance of the number of engagements of the fan-page posts are 1065.65 and 10,615,781.80. The negative binomial regression model, which is a generalized linear model (GLM), was considered as an appropriate statistical model to resolve over-dispersion. Link function, where was used in this model, and the parameters were estimated by the maximum likelihood estimating method. We considered “post type,” “topic,” “sentiment” and reaction variables, such as “love,” “haha,” “wow,” “sad,” and “angry,” as independent variables to find factors to affect the number of engagements in the statistical model, and used a backward selection method to select the best fitted model based on AIC and BIC. The chosen negative binomial regression model is as follows.

\[
\log \mu = 5.441 - 1.277(\text{link}) - 0.012\text{PT(photo)} \\
- 4.490\text{PT(status)} - 0.001\text{topic}(1) - 0.109\text{topic}(2) \\
- 1.269\text{topic}(3) + 0.047\text{topic}(4) + 0.121\text{sentiment} \\
+ 0.013\text{love} + 0.0158\text{haha} - 0.001\text{love} \times \text{haha}
\]

where PT(link), PT(photo), PT(status), topic(1), topic(2), topic(3), topic(4), love, and haha are indicator variables having values of 0 or 1. Table 2 shows the goodness of fit about the finally selected model and model validation criteria, since deviance/df is close to 1 and AIC is the smallest of all considered models.

Table 3 shows the chosen significant variables in the generalized linear model at statistically significant level. According to Table 3, the most significant variable is reaction “love,” post type, reaction “haha,” topic, and sentiment are significant. Also interaction of reaction “love” and “haha” is very significant.

Table 4 shows that there is much difference in the dynamics of each type of Facebook fan page: status, photo, and video. Photos showed more dynamics than status, and
videos showed more dynamics than photos. Videos and photos have almost similar effects on engagement of the Facebook fan page. Video is $89.11(=\exp (4.4899))$ times as status, and photo is $88.08$ times as status. Therefore, hypothesis 1 was accepted. For topics, topic 4 is the most effective, and next are topic 5, topic 1, topic 2, and topic 3 in order. Therefore, this shows that the topics have different effects on engagement. Topics related to sales and marketing

Figure 6. Positive and negative sentiment words.

Figure 7. Positive and negative sentiment analysis in each.
campaigns can increase customer engagement in a Facebook fan page. Therefore, hypothesis 2 was accepted.

**Table 1.** Frequency of positive and negative documents in each topic.

| Sentiment | Topic | Negative | Neutral | Positive | Total |
|-----------|-------|----------|---------|----------|-------|
|           | 1     | 20 (54.1)| 7 (18.9)| 10 (27.0)| 37 (100)|
|           | 2     | 12 (18.5)| 12 (18.5)| 41 (63.1)| 65 (100)|
|           | 3     | 15 (55.6)| 6 (22.2)| 6 (22.2)| 27 (100)|
|           | 4     | 13 (25.0)| 7 (13.5)| 32 (61.5)| 52 (100)|
|           | 5     | 10 (19.6)| 2 (3.9)| 39 (76.5)| 51 (100)|
| Total     | 70 (30.2)| 34 (14.7)| 128 (55.2)| 232 (100)|

Note: Chi-square = 40.2, p value < 0.001.

**Table 2.** Goodness of fit and model evaluation.

| Statistics            | df | Deviance | Value/df |
|-----------------------|----|----------|----------|
| Deviance              | 220| 263.7637 | 1.1989   |
| Pearson Chi-Square    | 220| 124.0770 | 0.5640   |
| Log likelihood        |    | 1676190.1|          |
| AIC                   |    | 2662.0   |          |
| BIC                   |    | 2706.7715|          |

Note: AIC (Akaike) and BIC (Bayesian Information) are criteria to select the best model.

**Table 3.** Significant factors to impact on the engagement.

| LR Statistics For Type 3 Analysis | Source | DF | $\chi^2$ | Pr > ChiSq |
|-----------------------------------|--------|----|----------|------------|
| post type                         | 3      | 168.34 | <.0001   |
| Topic                             | 4      | 13.98  | 0.0073   |
| sentiment                         | 1      | 4.08   | 0.0434   |
| Reaction “love”                   | 1      | 73.52  | <.0001   |
| Reaction “haha”                   | 1      | 56.18  | <.0001   |
| Reaction “love” $\times$ “haha”   | 1      | 40.05  | <.0001   |

For sentiment, sentiment stands for the difference between positive and negative words in each document. Engagement of a Facebook fan page increases by 1.128($=\exp(0.1206)$) times when sentiment increase by one. When emotional reaction “love” increases by one, engagement of the Facebook fan page increases by 1.013($=\exp(0.0128)$) times, whereas it increases by 1.016($=\exp(0.01580)$) times when reaction “haha” increases by one. However, the estimate of interaction of reaction “love” and “haha” has a negative value, which means that neither “love” nor “haha” has a positive effect on engagement of a Facebook fan page simultaneously. Therefore, it shows that the engagement differs depending on the post’s sentiment and emotional reaction level. Therefore, Hypothesis 3 was accepted.

**Table 4.** The parameter estimates of the negative binomial regression model.

| Parameter | DF | Estimate | S.E | Wald $\chi^2$ | Pr > ChiSq |
|-----------|----|----------|-----|---------------|------------|
| Intercept | 1  | 5.4410   | 0.2332| 544.51        | <.0001     |
| post type |    |          |     |               |            |
| link      | 1  | -1.2769  | 0.6351| 4.04          | 0.0444     |
| photo     | 1  | -0.0117  | 0.2050| 0.00          | 0.9545     |
| status    | 1  | -4.4899  | 0.2996| 224.57        | <.0001     |
| video     | 0  | 0.0000   | 0.0000| —             | —          |
| topic     |    |          |     |               |            |
| 1         | 1  | -0.0014  | 0.2978| 0.00          | 0.9963     |
| 2         | 1  | -0.1087  | 0.2335| 0.22          | 0.6417     |
| 3         | 1  | -1.2685  | 0.3495| 13.17         | 0.0003     |
| 4         | 1  | 0.0466   | 0.2442| 0.04          | 0.8486     |
| 5         | 0  | 0.0000   | —     | —             | —          |
| sentiment |    |          |     |               |            |
| reaction “love” | 1 | 0.0128 | 0.0020| 41.66 | <.0001 |
| reaction “haha” | 1 | 0.1580 | 0.0235| 45.39 | <.0001 |
| reaction “Love” $\times$ “haha” | 1 | -0.0007 | 0.0001| 33.88 | <.0001 |
| Dispersion | 1  | 1.3720   | 0.1254| —             | —          |

**Discussion and concluding remarks**

This paper attempts to investigate the effects of Facebook fan-page posting types, topics, and sentiments on social-media engagement. For this purpose, we first extracted unstructured posting data from Facebook Toyota fan-page, categorized types of fan-page postings based on previous research, extracted topics and sentiments of fan-page postings using R, and tested our hypotheses using negative binomial regression model. Tested and validated hypotheses are as follows.

First, there is a difference in the dynamics of each type of Facebook fan page: status, photo, and video. Videos and photos showed more dynamics than did status. And the more dynamic is the posting of the fan page, the more the engagement is affected (H1). Second, there is a difference in the kind of topic of a Facebook fan page. The effects of topic on engagement are as follows. Among the topic effects, topic 4 related to sales is most effective for the engagement of a Facebook fan page. Topic 5, associated with a marketing campaign, is the second most effective on it. Thus, the kinds of topic are an influential factor for the engagement. Therefore, Hypothesis 3 was accepted.
engagement of a Facebook fan page (H2). Third, the engagement of a Facebook fan page increases when sentiment arousal increases, and the positive or negative valance of sentiment also affects the engagement (H3).

We propose the following implications empirically and academically by means of this study. First, in order to activate Facebook fan pages and help with marketing, it is necessary to clarify the criteria for posting type, kinds of topics, and users’ sentiment. In a highly dynamic behavior, the use of videos or photos rather than status is a way to increase engagement. The second is the fan page posting type, which allows us to identify the type of customer. In other words, customers who use videos or photos rather than customers who use status can be seen as more engaged customers. Third, this study can also suggest ways to organize a Facebook fan page on the enterprise side. We can also propose whether to construct a fan page with high dynamics by identifying differences between users’ behavior or to make dynamics low. Finally, we propose a mix method of quantitative and qualitative research by means of topic modeling and sentiment analysis. In fact, the type of posting is important, but the content that affects posting quality is also an important factor in improving engagement. Topics and sentiments of posts which mean quality are taken into consideration in this study.

Despite these significant implications, this study has some limitations. First, research on the effect of sentence level and image level on engagement by means of content analysis of posting will be needed in the future. Also, the positive and negative images and precision of the sentences in the posting of the fan page can have a great influence on the social-media engagement. In addition, the expression of these sentences and images may differ between users’ behavior. Therefore, sentiment analysis that reflects these contents is required in the future.

**Declaration of conflicting interests**

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

**Funding**

The author(s) received no financial support for the research, authorship, and/or publication of this article.

**ORCID iD**

Byung-Hak Leem https://orcid.org/0000-0003-2584-0674

Hyesun Suh https://orcid.org/0000-0001-9469-7707

**References**

1. American Marketing Association. The CMO survey: highlights and insights reports, cmosurvey.org (2019, accessed 17 March 2020).

2. Klepek M. Determinants of customer engagement on Facebook brand pages. *Scientific Papers of the University of Pardubice*. Series D. Faculty of Economics and Administration; Pardibuce Iss 2018; 44: 127–138.

3. Wang X and Liu Z. Online engagement in social media: a cross-cultural comparison. *Comput Hum Behav* 2019; 97: 137–150.

4. Dolan R, Conduit J, Frethley-Bentham C, et al. Social media engagement behavior: a framework for engaging customers through social media content. *Eur J Mark* 2019; 53: 2213–2243.

5. Gangi PMD and Wasko MM. Social media engagement theory: exploring the influence of user engagement on social media usage. *J Organ End User Comput* 2016; 28: 53–73.

6. Prahalad CK and Ramaswamy V. Co-creation experiences: the next practice in value creation. *J Interact Mark* 2004; 18: 5–14.

7. Kankanhalli A, Tan BCY and Wei K. Contributing knowledge to electronic knowledge repositories: an empirical investigation. *MIS Q* 2005; 29: 113–143.

8. Bernoff J and Li C. Harnessing the power of the oh-so-social web. *MIT Sloan Manag Rev* 2008; 49: 36–42.

9. Ranjan KR and Read S. Value co-creation: concept and measurement. *J Acad Mark Sci* 2016; 44: 290–315.

10. Schmidt AL, Peruzzi A, Scala A, et al. Measuring social response to different journalistic techniques on Facebook. *Hum Soc Sci Commun* 2020; 7: 1–7.

11. Bhatt S, Joglekar S, Bano S, et al. Illuminating an ecosystem of partisan websites. In: *WWW ’18: Companion proceedings of the web conference 2018*, Lyon, France, April 2018, pp. 545–554. Switzerland: Republic and Canton of Geneva.

12. Mikalef P and Krogstie J. Examining the interplay between big data analytics and contextual factors in driving process innovation capabilities. *Eur J Inform Syst* 2020; 29: 260–287.

13. Corte-Real N, Oliveira T and Ruivo P. Assessing business value of big data analytics in European firms. *J Bus Res* 2017; 70: 379–390.

14. Mikalef P, Boura M, Lekakos G, et al. Big data analytics capabilities and innovation: the mediating role of dynamic capabilities and moderating effect of the environment. *Brit J Manag* 2019; 30: 272–298.

15. Turban E, Strauss J and Lai L. Social commerce—an IS and marketing perspective. LA: Springer, 2015.

16. Karimov FP, Brengman M and Van Hove L. The effect of website design dimensions on initial trust: a synthesis of the empirical literature. *J Electron Commer Res* 2011; 12: 272–301.

17. Lai LSL and Turban E. Groups formation and operations in the Web 2.0 environment and social networks. *Group Decis Negot* 2008; 17: 387–402.

18. Gonzalez-Lafayse L and Lapassouse-Madrid C. Facebook and sustainable development: a case study of a French supermarket chain. *Int J Retail Distrib Manag* 2016; 5: 560–582.

19. Hinchcliffe D. Profitably running an online business in the Web 2.0 era. Luettavissa, http://Web2. Social computing mag azine.com/runningonlinebusinessprofitablyintheWeb20era.htm (2006, accessed 10 April 2020).

20. Lai LSL and To WM. Content analysis of social media: a grounded theory approach. *J Electron Commer Res* 2015; 16: 138–152.
21. Blei DM, Ng AY and Jordan MI. Latent dirichlet allocation. J Mach Learn Res 2003; 3: 993–1022.
22. Blei DM. Probabilistic topic models. Commun ACM 2012; 55: 77–84.
23. Le DT, Nguyen CT, Coltech QT, et al. Matching and ranking with hidden topics towards online contextual advertising. In: 2008 IEEE/WIC/ACM international conference on web intelligence and intelligent agent technology, Sydney, NSW, Australia, 9–12 December 2008, pp. 888–891. IEEE.
24. Tirullina S and Tellis GJ. Mining marketing meaning from

25. Cao D, Ji R, Lin D, et al. Visual sentiment topic model based

26. Jacobs BJD, Donkers B and Fok D. Model-based purchase predictions for large assortments. Mark Sci 2016; 35: 389–404.
27. Do TMT and Gatica-Perez D. By their apps you shall understand them: mining large-scale patterns of mobile phone usage. In: Proceedings of the 9th international conference on mobile and ubiquitous multimedia, Limassol, Cyprus, December 2010, Article No. 27, pp. 1–10. New York, NY.
28. Schröder N, Falke A, Hruschka H, et al. Analyzing browsing and purchasing across multiple websites based on latent dirichlet Allocation. In: ALLDATA 2017: The third international conference on big data, small data, linked data and open data, Venice, Italy, 23–27 April 2017, pp. 40–44.
29. Trusov M, Ma L and Jamal Z. Crumbs of the cookie. User profiling in customer-base analysis and behavioral targeting. Mark Sci 2016; 35: 341–537.
30. Reisenbichler M and Reutterer T. Topic modeling in marketing: recent advances and research opportunities. J Bus Econ 2019; 89: 327–356.
31. Luca M. Reviews, reputation, and revenue: the case of yelp. Com. Harvard Bus School NOM Unit Work Paper 2016; 12: 1–39.
32. Radford A, Jozefowicz R and Sutskever I. Learning to generate reviews and discovering sentiment. arXiv:1704.01444, 6 April 2017, Cornell University.
33. Stine R. Sentiment analysis. Ann Rev Stat Appl 2019; 6: 287–308.
34. Brodie RJ, Hollebeek LD, Jurić B, et al. Customer engagement: conceptual domain, fundamental propositions, and implications for research. J Serv Res 2011; 14: 252–271.
35. Vasalou A, Joinson AN and Courvoisier D. Cultural differences, experience with social networks and the nature of “true commitment” in Facebook. Int J Hum Comput Stud 2010; 68: 719–728.
36. Luarn P, Lin Y and Chiu Y. Influence of Facebook brand-page posts on online engagement. Online Inform Rev 2015; 39: 505–519.
37. Pletikosa C and Michalheselles F. Online engagement factors on Facebook brand pages. Soc Netw Anal Min 2013; 3: 843–861.
38. Mariani MM, Mura M and Felice MD. The determinants of Facebook social engagement for national tourism organizations’ Facebook pages: a quantitative approach. J Destin Mark Manag 2018; 8: 312–325.
39. Le TD. Influence of WOM and content type on online engagement in consumption communities. Online Inform Rev 2018; 42: 161–175.
40. Carboni JL and Maxwell SP. Effective social media engagement for nonprofits: What matters? J Nonprof Public Aff 2015; 1: 18–28.
41. Cvijikj IP, Spigeler ED and Michahelles F. The effect of post type, category and posting day on user interaction level on Facebook. In: 2011 IEEE third international conference on privacy, security, risk and trust and 2011 IEEE third international conference on social computing, Boston, MA, USA, 9–11 October 2011, pp. 9–11. IEEE.
42. McLachlan S. How to increase social media engagement: a guide for marketers, https://blog.hootsuite.com/social-media-engagement (2020, accessed 10 December 2020)
43. Dolan R, Conduit J, Fahy J, et al. Social media engagement behaviour: a uses and gratifications perspective. J Strateg Marketing 2019; 24: 261–277.
44. Maier M, Waldherr A, Miltner P, et al. Applying LDA topic modeling in communication research: toward a valid and reliable methodology. Comput Method Commun Sci 2018; 12: 93–118.
45. Westerland M, Leminen S and Rajahonka MA. Topic modelling analysis of living labs research. Technol Innov Manag Rev 2018; 8: 40–51.
46. Calheiros AC, Moro S and Rita P. Sentiment classification of consumer-generated online reviews using topic modeling. J Hosp Mark Manag 2017; 26: 675–693.
47. Schmiedel T, Muller O and Brocke JV. Topic modeling as a strategy of inquiry in organizational research: a tutorial with an application example on organizational culture. Organ Res Method 2018; 22: 941–968.
48. Griffiths TL and Steyvers M. Finding scientific topics. Proc Natl Acad Sci USA 2004; 101: 5228–5235.
49. Hu M. and Liu B. Sentiment analysis: mining sentiments, opinions, and emotions. Cambridge: Cambridge University Press, 2015.