IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

SPECIAL SECTION ON ADVANCED DATA ANALYTICS FOR LARGE-SCALE COMPLEX DATA ENVIRONMENTS

Received October 30, 2017, accepted November 28, 2017, date of publication December 6, 2017,
date of current version February 14, 2018.

Digital Object Identifier 10.1109/ACCESS.2017.2780046

Identifying Product Opportunities Using Social Media Mining: Application of Topic Modeling and Chance Discovery Theory

NAMUK KO[1], BYEONGKI JEONG[1], SUNGCHUL CHOI[2], AND JANGHYEOK YOON[1]
[1]Department of Industrial Engineering, Konkuk University, Seoul 05029, South Korea
[2]Department of Industrial and Management Engineering, Gachon University, Seongnam-si 13120, South Korea

Corresponding author: Janghyeok Yoon (janghyoon@konkuk.ac.kr)

This work was supported by Konkuk University in 2014.

ABSTRACT As an emerging voice of the customer (VOC) containing feedback, such as opinions and expectations about products, social media data have the potential use for product improvement and new product development. However, most prior studies have focused on determining customer concerns, while neglecting to incorporate them into a systematic approach to identify product opportunities. In response, this paper suggests an approach to identify product opportunities from customer reviews in social media. This approach employs topic modeling to identify the product topics discussed by customers from large-scale review posts related to a given product. A keygraph is then constructed based on the co-occurrences among the topics contained in each post. The chance discovery theory is then applied to generate new product opportunities from the chance nodes obtained from the keygraph. Our approach contributes to the systematic ideation process for product opportunity analysis based on large-scale and real-time VOC.

INDEX TERMS Chance discovery theory, product opportunity, social media mining, topic modeling, voice of the customer.

I. INTRODUCTION

In the recent competitive business environment, various firms have been attempting to continue improving their product business to cope with rapidly changing market trends and customer needs. In this regard, the ability to identify product opportunities, which are defined as a chance to develop new products or improve current products, is considered to be most essential for the sustainable growth of product-based firms [1]. To effectively identify such opportunities, in product planning processes, the voice of the customer (VOC) (which includes customer expectations and opinions about a product) is generally considered the primary prerequisite [2]. Collection and in-depth analysis of the VOC of products can help firms determine product development directions in a more practical and reasonable way [3], thereby eventually enabling firms to build a customer relationship that cannot be easily copied by their competing firms.

In the literature, the VOC is defined as a statement of customer needs and desires or an explanation of customer preferences and aversions [4], [5]. VOCs are a useful material that includes clear customer feedback. Therefore, involving customers to product development processes is necessary in the process of setting the direction of developing new product concepts and improving the current products provided to customers [6]. In the customary approach, the VOCs of the product under study are usually collected through direct customer contact, including well-designed customer interviews, online surveys or contextual inquiry [7]. Recently, such voices have been used to understand customers' needs and have been incorporated into user-centric product/service development approaches combined with techniques, including quality function deployment [8], [9], conjoint analysis [10], morphology analysis [11], the Kano model [12], [13], and Fuzzy analysis [14], [15].

Unfortunately, the majority of recent VOCs have tended to be distributed online rather than delivered directly to product firms, as information and communication technologies advance and internet populations grow explosively. Massive VOCs have recently been accumulated and shared through various social media. Social media refers to computer-mediated technologies that allow the creation and sharing of user-generated content based on the concept of Web 2.0, such

as social networking services, microblogging, photo sharing, and instant messaging [16]. Social media has enabled an individual to communicate with innumerable other people about products and the firms that provide them [17]. Therefore, the growth of social media during the last decade has revolutionized the way that individuals and industries conduct their business [18]. In addition, it was revealed that over 75% of all internet users use social media by joining SNSs, reading and posting on blogs, or writing reviews on shopping sites [16].

Regarding the applicability of social media data, in previous studies it was suggested that firms can benefit from using both traditional marketing methods and big data contained in social media [19] and the network properties of the customers obtained from SNSs have an effect on customer monetary value in the sale of digital products [20]. In fact, many firms have attempted to utilize social media as a tool to enhance the ability to listen and correspond to changing customer needs, with the aim of securing their business performance and identifying business opportunities in the market [21], [22]. For these reasons, various studies have been carried out using social media data, including online product reviews and microblogs. Also, methods have been proposed for market structure analysis [23], significant topics and their customer sentiment analysis [24]–[26], consumer brand sentiments [27], product concept generation [28], product recommendation [29], tourist flow analysis [30], individual activity pattern analysis [31] and integration of virtual community members into new product development [32].

Despite the contributions made by the above-mentioned studies, they have some limitations in terms of new product opportunity analysis. First, most studies using social media data have focused primarily on identifying the topics of a given product and their current trends, such as sentiment and opinion. As discussed previously, social media data include a considerable amount of customer feedback and have the potential for product development. However, surprisingly, further approaches for product opportunity ideation were not successfully addressed. In addition, some of the previous studies center only on analyzing the primary product topics or keywords, overlooking to identify groundbreaking product opportunities. In particular, innovative and fresh ideas for new product opportunities can be obtained from infrequent but relatively significant events in the process and context in which customers use a product [33]. Considering these limitations, a systematic approach is required to identify potential product chance topics containing customer feedback from massive social media data and to generate and analyze practical product opportunities based on the chance topics.

Therefore, an approach to identify product opportunities through social media mining is proposed in this paper based on topic modeling and chance discovery theory. Topic modeling is a probabilistic generative model used to identify latent topics on text-based documents [34], while chance discovery is a network-based method used to discover infrequent but relatively significant events and situations, called breaking points, from textual information [35]. By employing these two analytical methods, this approach derives potential product opportunities from the breaking topics underlying online customer reviews of a specific product. Specifically, topic modeling is used in a step that identifies what product topics are concerned by massive customers and chance discovery is used in a step that locates breaking topics and ideates about product opportunities from the breaking topics and their subnetwork. Therefore, this approach to product opportunity identification using social media mining involves the following steps: 1) gathering online customer reviews related to a specific product, 2) extracting the product's discussion topics from the online customer reviews by applying topic modeling, 3) constructing a keygraph based on the co-occurrences between pairs of the discussion topics in the online reviews, and 4) generating product opportunities based on chance discovery analysis that uses breaking topics and their neighboring topics in the keygraph. To illustrate the operation of our product opportunity analysis approach, we apply it to the online posts of the Samsung Galaxy Note 5.

The contribution of this study is two-fold. First, this approach will contribute to the systematic identification of new product opportunities from large-scale and real-time customer feedback in social media while being a useful aid for monitoring rapidly changing customer needs and relationships among these needs. Second, this approach is a tool for product opportunity ideation that will have a synergetic effect when incorporated into the product planning process.

The organization of this study is as follows. We present the theoretical background, followed by our proposed approach and its application to identify product opportunities from social media data. We then present conclusions and suggestions for further research topics.

II. THEORETICAL BACKGROUND

Our approach for product opportunities is based on topic modeling and the chance discovery theory; this section briefly overviews this theoretical background.

A. TOPIC MODELING

A topic model is a statistical model used to identify the latent topics that occur in a collection of text documents; topic modeling is therefore often considered to be a text-mining tool for the discovery of hidden semantic structures in a text body [36]. Among various topic models, latent Dirichlet allocation (LDA) is used in the present study, which is a generative model that allows sets of observations to be explained by unobserved groups [34]. LDA helps identify the topical features of documents by postulating that documents are described by a topic distribution and that each topic is made up of a distribution of words.

LDA follows a generative process for a corpus D consisting of K topics and M documents, each of length N (Figure 1) [34]. α and β are Dirichlet priors on the per-document topic

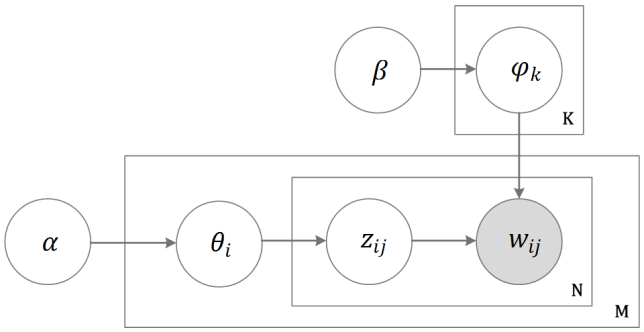


FIGURE 1. Concept of LDA-based topic modeling, redrawn from [36].

- ✓ Choose $\theta_i \sim Dir(\alpha)$, where $i \in \{1, \ldots, M\}$
- ✓ Choose $\varphi_k \sim Dir(\beta)$, where $k \in \{1, \ldots, K\}$
- ✓ For each word position i, j, where $j \in \{1, \ldots, N_i\}$, and $i \in \{1, \ldots, M\}$
  - Choose a topic $z_{ij} \sim Multinomial(\theta_i)$.
  - Choose a word $w_{ij} \sim Multinomial(\varphi_{z_{ij}})$.

distribution and the per-topic word distribution, respectively, θ_i is the distribution over topics for document i, φ_k is the distribution over words for topic k, z_{ij} is the topic for the jth word in document i, and w_{ij} is the specific word.

Due to the usability of LDA, it has been widely applied in studies dealing with large-scale corpus. In prior studies in which LDA was applied, a novel multi-corpus LDA technique was proposed to filter web spam [37], a hierarchical Bayesian model was proposed to analyze the topical relationship between news and social media [38], an automated method was proposed to compare the human and automated coding of newspaper articles [39], a novel method was proposed to recommend news articles that are appropriate to the location by reflecting the geographical context of users [40], and a satisfaction analysis method was proposed to identify key dimensions of customer service voiced by hotel visitors [41].

In this study, LDA-based topic modeling is applied to the online review data to identify product features currently being discussed by product customers. These product features are then used to generate product opportunities based on chance discovery theory.

B. CHANCE DISCOVERY THEORY

Chance discovery theory, which was initially proposed by Ohsawa [35], is a relatively new research field as an extension of textual knowledge discovery. This theory is based on a network-based analysis that discovers uncertain but relatively significant events and situations, called chances or breaking points in the theory, from text-based data; these breaking events and situations can serve as a clue to generate new ideas [33]. The advantage of chance discovery theory is its ability to evaluate the importance of data from two perspectives: term frequency and association links [42]. In analyzing networks obtained from textual data, chance discovery theory focuses on the nodes that are rare but correlated strongly with other nodes [43]. Therefore, this theory can uncover the unexamined but potentially significant situation behind textual data.

An important output of chance discovery analysis is keygraphs. A keygraph is a network visual that displays key concepts and their relationships behind textual data, providing analysts with a comprehensive understanding of knowledge contained in all of the documents under study. As a variant of network analysis, keygraphs extract essential events and their causal structures in textual documents, thereby allowing analysts to understand the meaningful sequence of a specific event by connecting events closely located to the event (Figure 2). A keygraph is generated from a document that contains sentences which in turn consist of words. The following two steps describes how keygraphs are constructed from textual documents [44], [45].

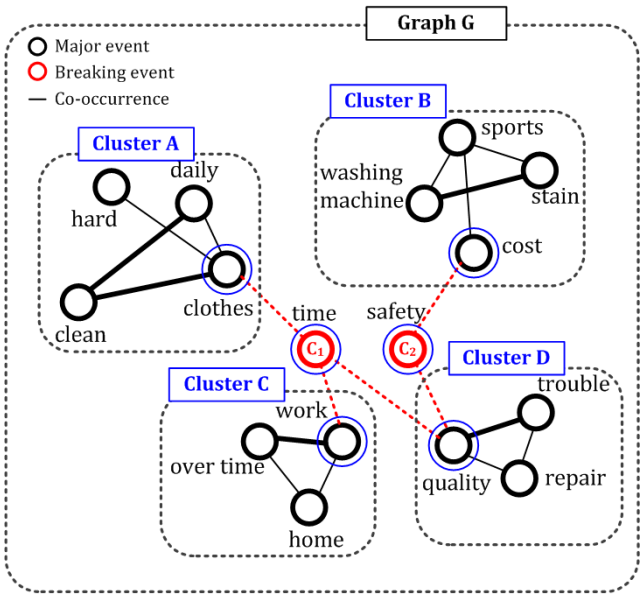


FIGURE 2. Keygraph example.

1) CONSTRUCTING THE CLUSTERS COMPOSED OF HIGH FREQUENCY ITEMS AND THEIR PAIRS

In the first step, high frequency items are extracted, i.e. terms and words. Items in a document are sorted by frequency of their occurrence. Those items with high occurrence frequency are represented as black nodes in graph G. The item-pairs that co-occur in the same sentences are then identified, and the item-pairs are sorted by their occurrence frequency. These item-pairs are represented as black solid lines in graph G, thereby constructing clusters. To measure the extent to which two high-frequency items, I_i and I_j, co-occur, the Jaccard co-efficient can be used:

$$J\left(I_i, I_j\right) = \frac{Freq(I_i \cap I_j)}{Freq(I_i \cup I_j)} \tag{1}$$

where $Freq(I_i \cup I_j)$ is the frequency of either item I_i or I_j occurring in sentences, and $Freq(I_i \cap I_j)$ is the frequency of the two items co-occurring in the same sentences.

2) IDENTIFYING CHANCE ITEMS THAT STRONGLY CO-OCCUR WITH CLUSTERS

In this second step, key items are extracted using the tightness between item I and cluster C, and the tightness measure can be defined as:

$$Key\,(I) = 1 - \prod_{C \subset \mathrm{G}} [1 - J(I, C)] \tag{2}$$

where $J(I, C)$ is the Jaccard value between item I and cluster C and thus it represents a co-occurrence extent between item I and all words in cluster C.

This key value of an item in a set of documents is used to identify chance items. The chance items with a high key value are added as nodes in red only if they are not already in the current graph G. The value of chance links between pairs of each chance item and other existing high frequency items is then computed using Eq. (1), thereby adding chance links with a high value to graph G; these links are represented as dotted lines in red. Finally, the subgraphs that are composed of a chance item and its adjacent nodes are used to develop customer scenarios and ideate about their consequential opportunities.

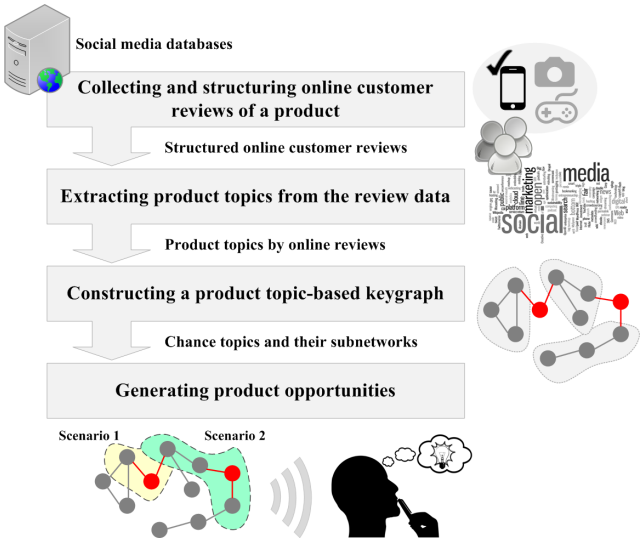


FIGURE 3. Overall procedure.

III. PROPOSED APPROACH

This section describes our approach for identifying product opportunities based on social media mining (Figure 3). This approach is composed of 1) collecting online customer reviews of a product from social media, 2) extracting product topics from the review data using topic modeling, 3) constructing a product topic-based keygraph and identifying chance topics, and 4) generating potential product opportunities by analyzing scenarios based on the subnetworks related to the chance topics. The following sections explain the steps involved in this approach in detail.

A. COLLECTING AND STRUCTURING ONLINE CUSTOMER REVIEWS OF A PRODUCT FROM SOCIAL MEDIA

The first step in our approach involves collecting online customer reviews of a given product as the material for analysis. In collecting such review data, analysts should consider two factors: the type of product and the type of online data. First, the products applicable to this approach are the high-tech products that contain various functions, components, and accessories. The majority of recent online product reviews in social media are usually generated for high-tech products because these products, which rapidly evolve and have short lifecycles, can receive a variety of feedback related to their functions, components, and accessories; as the number of product reviews that are used increases, our textual analysis becomes more effective. Second, the social media data applicable to this approach is the data generated by product customers. For example, news data may not be suitable for our approach because most purveyors of news value neutrality and objectivity, excluding subjective opinions by product customers. Therefore, product customer-generated social media data, such as blogs and online community postings, are the appropriate material for our approach.

Once the product for analysis and its data source are selected, large-scale online customer reviews related to a given product should be collected. To this end, various techniques can be used for social media data collection, such as web crawling and open application programming interfaces (APIs) provided by Twitter, Facebook or blog services. Finally, the online customer reviews are stored in the form of electronic files, such as text file and excel files, for our computational analysis.

Keywords are then extracted from each online customer review and the review is then structured as a document-keyword vector. Generally, an online customer review, as a document, has multiple sentences, which in turn are composed of keywords, such as single words and phrases. Such keywords can be extracted from documents using natural language processing (NLP) tools, such as AlchemyAPI (http://www.alchemyapi.com/), Aylien (http://aylien.com/text -api), and TextRazor (https://www.textrazor.com/docs/rest). However, some keywords extracted should be excluded from the keyword list, because they may be irrelevant for textual analysis; pronouns (e.g. 'you', 'he', 'this', 'that'), conjunctions ('and', 'or', 'but'), articles ('a', 'an', 'the'), emoticons ('^^', ':-)', ':-D'), onomatopoeic words ('haha', 'blah'), and meaningless words ('system', 'process', 'product'). By excluding such irrelevant keywords from the keyword list, a set of valid keywords is finally produced in this step. Then, each of the documents is represented as a document-keyword vector, or an array composed of keywords and their occurrence frequency in the document. All documents are transformed into document-keyword vectors and a document-keyword matrix can finally be prepared for the input for product topic extraction in the next step.

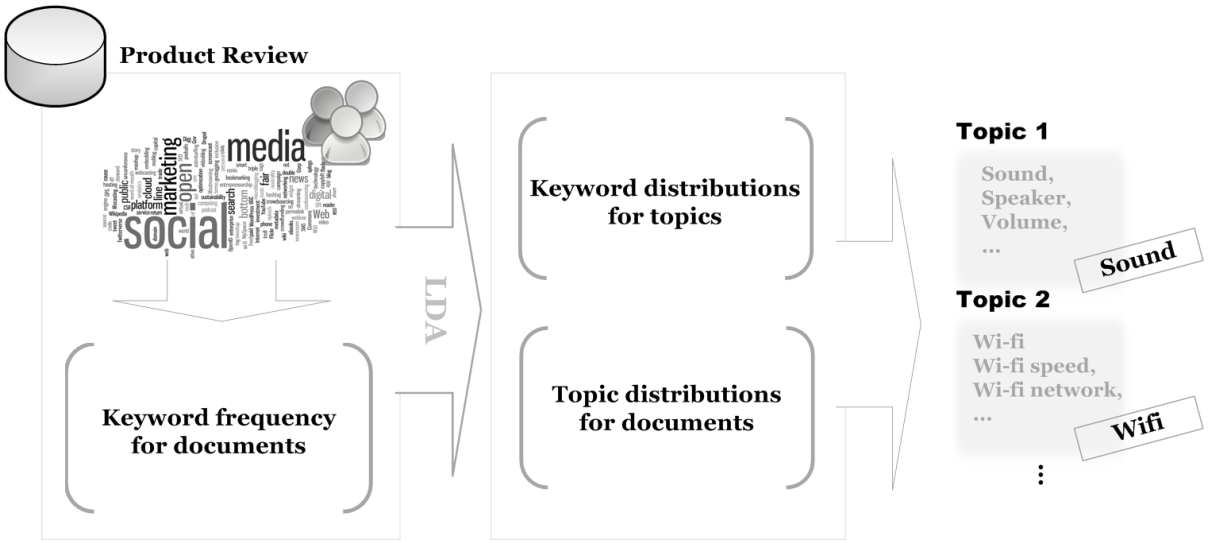


FIGURE 4. Topic extraction process.

B. EXTRACTING PRODUCT TOPICS FROM PRODUCT REVIEW DATA

In this step, LDA-based topic modeling is employed to identify product topics that are being discussed by product customers in social media (Figure 4). Our LDA-based topic modeling requires two inputs: a document-keyword matrix and the number of topics. The document-keyword matrix in the previous step can be used as the input matrix for topic modeling, while an appropriate number of topics should be determined for topic modeling. Several techniques can be used to select an appropriate number of topics, including the elbow method, information criterion method, and information theoretic method. Of these methods, the elbow method is used in this study, in which an optimal number of topics is determined by calculating the average cosine similarity value between pairs of topic-word distribution vectors outputted by topic modeling [46].

The application of topic modeling to the document-keyword matrix produces two outputs: a topic-keyword distribution matrix and a document-topic distribution matrix. Each row vector of the topic-keyword distribution matrix shows how a topic is constructed by its main keywords and their contribution probability. Therefore, the labeling process of a topic can be conducted based on these keywords and their probability of contribution to the topic; for example, if a topic has main keywords, including 'sound', 'sound volume', 'headphone', and 'speaker', this topic can be named 'sound' or 'product sound.' Topic labels in this study are considered as product topics, because they are the subjects currently discussed by customer review data in social media. Each row vector of the document-topic distribution matrix describes how a document is constructed by its topics and their contribution probability. In particular, this document-topic distribution matrix will be preprocessed and used as the input for keygraph construction and chance item discovery in the next step.

C. CONSTRUCTING PRODUCT-TOPIC BASED KEYGRAPHS

In this step, a keygraph is generated that visualizes an overall landscape of the relationship among product topics and shows the breaking points of the graph (Figure 5). As previously mentioned, the keygraph generation process requires two parameters for its inputs: the co-occurrence data of item-pairs and the number of chance items. In this step, the co-occurrence between pairs of product topics can be identified by analyzing the document-topic distribution matrix produced in the previous step. Each row vector of the document-topic distribution matrix shows the degree to which a document belongs to all product topics in terms of probability. To determine whether or not a document belongs to specific topics, a threshold value α is used; if the probability of the contribution of a product topic to its corresponding document is larger than α, the product topic is considered to be contained in the document. For example, if two product topics both appear three times in the same documents, their co-occurrence value is 3. Next, the number of chance items is the number of breaking points that will be displayed in the generated keygraph.

Different keygraphs can be outputted by different combinations of the threshold value and the number of chance items. However, the objective of our keygraph application is to trigger product opportunity ideation. It is therefore recommended that various keygraphs are generated for a wide range of product opportunities based on various combinations of the threshold value and the number of chance items. Various tools can be used to ensure the keygraph generation process is efficient, such as Java-based API (http://keygraph.codeplex.com) and Polaris (http://www.chokkan.org/software/-polaris).

A keygraph describes various major topics related to the product under study and displays breaking topics that are relatively infrequent but critically connected to the major topics. Therefore, in the next step, a subnetwork is constructed

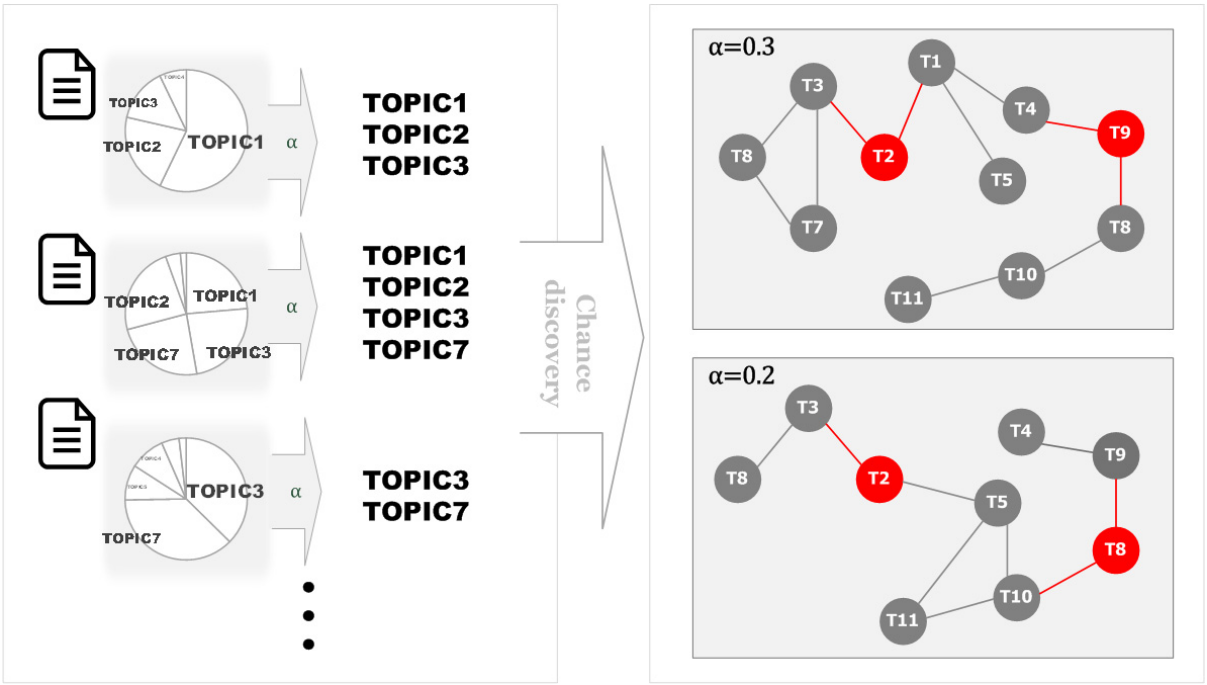


FIGURE 5. Keygraph construction process.

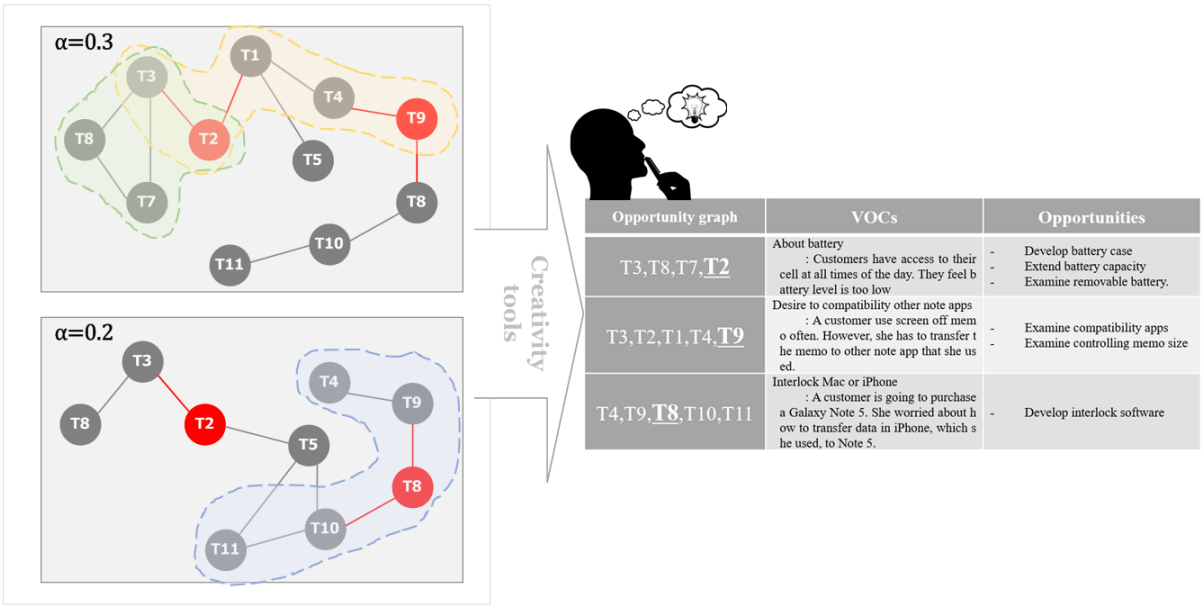


FIGURE 6. Product opportunities generation process.

from a breaking product topic and its adjacent major topics in a keygraph; the subnetwork can be used as a scenario that describes a situation for product opportunities.

D. GENERATING PRODUCT OPPORTUNITIES

In this step, scenarios are constructed for product opportunities based on opportunity graphs (Figure 6). An opportunity graph is composed of one or more breaking product topics and the neighboring major topics that are directly connected or closely located to the breaking product topics. In addition, an opportunity graph is used to describe the improvement situations from a customer's viewpoint; this graph thus acts as a guide to identify the opportunities that are being implicitly discussed by customers in large-scale customer product reviews.

Each topic is related to various documents and each document contains customer opinions of a given product. Therefore, generating product opportunities is a process of idea

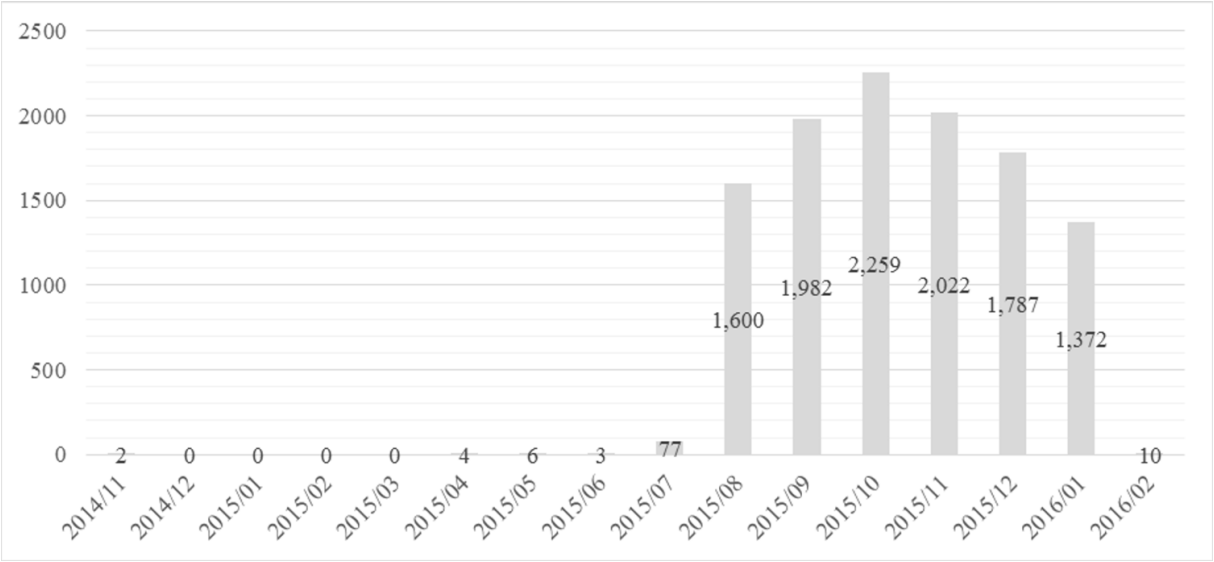


FIGURE 7. The number of documents per month.

generation using customer opinions and relationships among topics. However, if analysts only use opportunity graphs, it is difficult to generate meaningful opportunities, and the process results are influenced subjectively by the analysts. Thus, various creativity tools or supporting tools could be used to support the systematic process to generate chances. For example, the Substitute, Combine, Adapt, Modify, Put to another use, Eliminate, and Reverse (SCAMPER) [47] tool helps analysts think dynamically to create various ideas focusing on the seven features of specific products, services, or items [48]. The five general attribute types of TRIZ (Russian acronym of the theory of inventive problem solving), which are 'change', 'increase', 'decrease', 'stabilize', and 'measure', have been widely used to search for inventive problem solving [49]. These terms support the change of a description to an attribute of the system or component part.

IV. EMPIRICAL STUDY: CASE STUDY OF THE GALAXY NOTE 5

An illustrative example of the proposed approach using the Galaxy Note 5 is presented in this study. The Galaxy Note 5 is an android-based smartphone launched in the United States on August 13, 2015 by SAMSUNG. Our approach involves identifying product opportunities from various topics and the relationships among the topics. This means that products containing complex and various topics rather than simple and few products are suitable for our research. The Galaxy Note 5 is one of the most popular smartphones, and it is possible to acquire a variety of feedback and topics from customers because the smart device has various functions, components, peripheral devices, and accessories. Therefore, we believe that the Galaxy Note 5 is suitable for illustrating the applicability of our approach.

A. SOCIAL MEDIA DATA COLLECTION

Among the various social media available for collecting customer reviews, we used Reddit (https://www.reddit.com). Reddit is a social news aggregation and discussion website which has over 240 million unique users. The website is organized by areas of interest called ''subreddits'', and the number of subreddits exceeds 800,000. Subreddits are composed of not only comprehensive categories such as 'News', 'Science', Food', and 'Art', but also concrete categories which are specific products, services or items; Galaxy note 5 is among these concrete categories (https://www.reddit.com/r/galaxynote5). Therefore, we gathered the device reviews from subreddit on Galaxy Note 5 up to early February 2016 (Figure 7). From the result, a total of 23,613 textual reviews were collected, which included 2,255 posts and 21,358 comments. Since the product was launched in August 2015, a large amount of review data has been collected, the number peaking in October 2015.

We then extracted keywords from the textual data using the NLP tool, which is Alchemy API (http://www.alchemyapi.com/). This tool is a commercial API that provides various text analysis services such as keyword extraction, entity extraction, sentiment analysis, and language detection. After the total number of 32,656 keywords was extracted by Alchemy API, we removed the irrelevant words such as pronouns, conjunctions, articles, emoticons, onomatopoeic words, and meaningless words. We also excluded words that appeared only in one document because the words do not affect the relationship among other documents or topics. Finally, by identifying each keyword individually, we selected 3,549 final keywords and 11,124 textual review data that contained the final keywords.

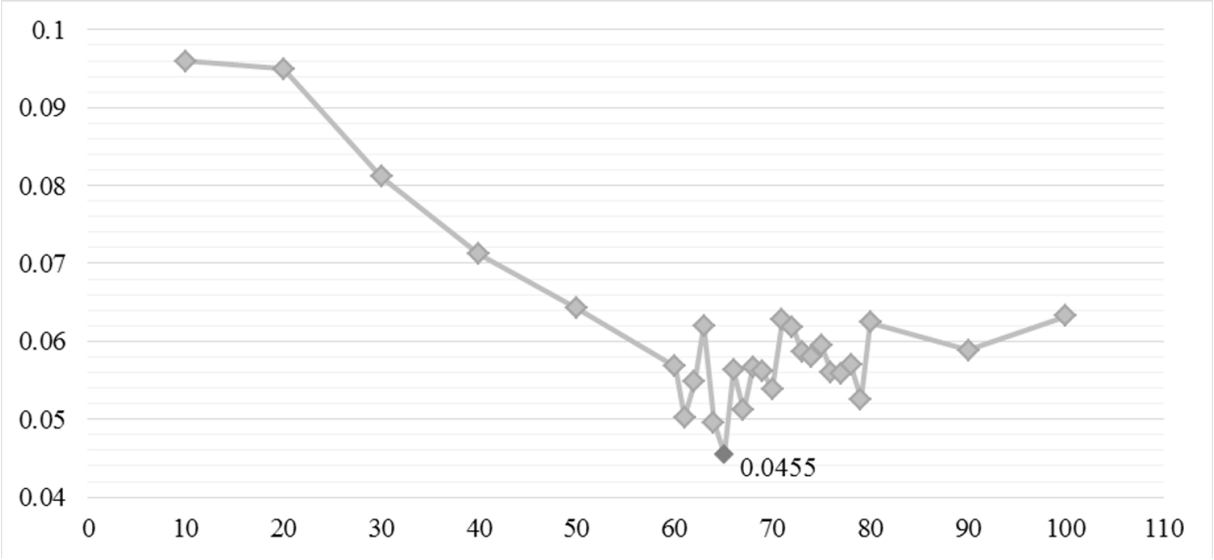


FIGURE 8. Average cosine similarity per the number of topics.

B. EXTRACTING TOPICS AND CONSTRUCTING KEYGRAPHS

Using the gathered reviews and keywords, we constructed a document-keyword matrix with term frequency values; the matrix was an input parameter for LDA. As mentioned above, LDA requires determining one extra parameter, or the number of topics. In this study, we used the elbow method, which is a method of interpretation and validation of consistency, to determine the appropriate number of topics. The optimal number of topics was determined by calculating the lowest average cosine similarity value between pairs of topic-word distribution vectors outputted by topic modeling. This means that the distance of the topics is the farthest; in other words, the topics are clearly separated from each other. The result of calculating the topic similarities showed the lowest similarity value of 0.0455 from 65 topics (Figure 8). Therefore, we selected 65 topics as the optimal parameter, and we extracted a topic-keyword distribution matrix and a document-topic distribution matrix.

For effective topic analysis, we defined a name for each topic. The naming process of a topic can be conducted based on the topic-keyword distribution matrix. Because each row vector of the matrix indicates how a topic is composed with its main keywords and their contribution probability, it is possible to name the topic using each vector. For example, a topic with the main keywords of 'fingerprint', 'finger print', 'finger print scanner', and 'fingerprint sensor' was named 'Fingerprint', and a topic with the main keywords of 'charge', 'cable', 'charger', and 'fast charge' was named 'Charge cable'. In this way, we defined the names of all the topics, 10 of which are shown in Table 1.

The document-topic distribution matrix shows how a document is constructed by its topics and their contribution probability, and the matrix used as the input for keygraph construction and chance topic discovery. The matrix can provide the co-occurrence between pairs of product topics by each row vector of the matrix which shows the number of relationships between a document and all product topics in terms of probability. However, because the matrix is a distribution matrix, a document has relationships with all of the topics. Thus, we should delete the weak relationships by using a threshold value α. To calculate the proper threshold value, we applied a technique which determines a threshold value based on network similarities calculations by a vector space model [50]. This technique is used to identify an optimized cut-off value that determines potential connectivity between any two nodes in a 1-mode network. We modified the methodology for application to the 2-mode matrix (which is the document-topic distribution matrix) by normalizing (min-max normalization) each row vector. We then utilized the modified method for calculating the proper threshold value. We then determined the threshold α as 0.19; in other words, topics with a probability of the contribution of the normalized document-topic distribution matrix to its corresponding document of less than 0.19 were excluded. The results of product topics acceding to each document are listed in Table 2.

In this study, Polaris, which is a free keygraph generation tool, was used to construct keygraphs. By constructing numerous keygraphs, we can effectively analyze product opportunities from a variety of perspectives. A well-organized keygraph helps analysts intuitively grasp potential chances expressed as breaking points from a given product. After undergoing several trial-and-error processes, we found a keygraph that clearly describes the relationships among product topics (Figure 9).

In Figure 9, high frequency topics were represented as nodes on the graph, among which chance topics were expressed in red. It can be seen that the graph was formed

TABLE 1. Example of topics and their main keywords.

Topic	Keywords(Probability) - Top 10
Fingerprint	fingerprint(0.1710), finger print(0.171), finger prints(0.083), fingerprints (0.083), sensor(0.029), finger print scanner(0.026), fingerprint scanner (0.026), fingerprint sensor(0.015), finger print reader(0.012), fingerprint reader(0.012)
Wifi	wifi(0.288), Wi-Fi(0.286), network(0.015), Speed(0.015), speeds(0.01), WiFi network(0.008), wifi speed(0.007), hotspot(0.006), airplane mode (0.006), wifi connect(0.006)
SD card	SD(0.13), SD-Card(0.094), SD card(0.093), car(0.087), removable battery (0.049), MicroSD(0.036), micro SD(0.036), SD cards(0.033), card slot (0.028), SD card slot(0.023)
Update	update.(0.344), update(0.343), software(0.053), software update(0.025), security(0.012), updating(0.006), software updates(0.006), Marshmallow update(0.005), security update(0.004), Marshmallow(0.003)
Charge cable	charge(0.221), cable(0.153), charger(0.082), fast charge(0.076), USB-C (0.041), cables(0.031), fast charger(0.031), USB cable(0.03), Charges (0.024), charging(0.023)
Write on screen	pen(0.296), pen(0.286), write(0.047), font(0.018), OTA(0.009), replacement(0.008), Screen write(0.005), screenwrite(0.005), OS(0.004), Applications(0.003)
Location	OS(0.469), map(0.05), GPS(0.043), maps(0.016), BT(0.014), tv(0.009), UI(0.005), Google Maps(0.005), insurance(0.003), SD(0.003)
Camera	camera(0.314), picture(0.156), pictures(0.087), API(0.032), cameras (0.026), camera app(0.015), shutter(0.009), Capture(0.006), front-facing camera(0.005), photograph(0.005)
E-mail	email(0.193), e-mail(0.193), emails(0.043), OS(0.032), calls(0.026), ROM(0.025), mail app(0.016), Email app(0.013), voicemail(0.012), voice mail(0.012)
S-pen	pen(0.196), pen.(0.192), Spen(0.135), S-Pen(0.13), pens(0.054), SPens (0.032), S-Pens(0.032), sensor(0.009), pen work(0.007), S-Pen work (0.005)

TABLE 2. Example of product topics for partial documents.

# of documents	Product topics
Doc #1	Galaxy Note 5, Hardware Spec, Video
Doc #2	SD card, OS feature, Internal storage, Battery
Doc #3	Charge cable, Charging
Doc #4	Sound, Device connection
Doc #5	Multi-tasking, SMS, Video, Theme
Doc #6	VR, Sound, OTA, Game, SIM
Doc #7	Touch pen, Wifi, write on screen, SMS, E-mail, Music App, Chat

around a range of topics such as 'UI', 'Peripheral', 'Battery', 'Stylus', 'Screen resolution', and 'Chat', while other isolated topics such as 'Camera', 'Design', 'SIM', and 'Music App' seem to have relatively independent relationships, although they showed high frequencies. The total number of chance topics was 10, including 'Internal storage',

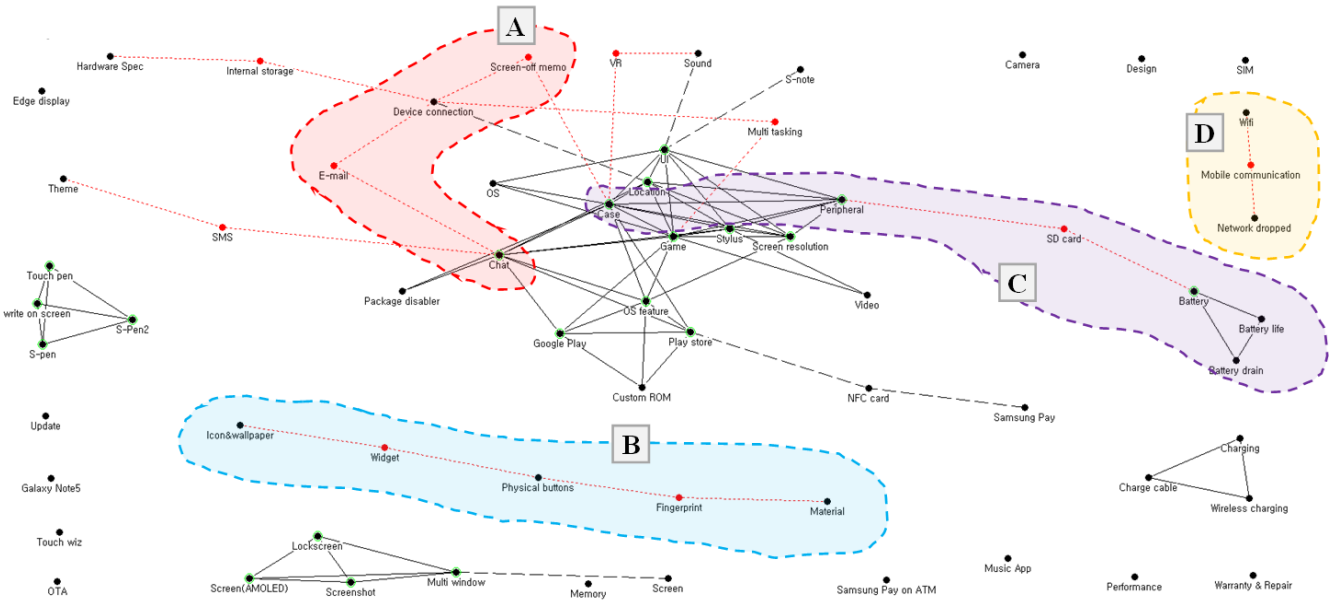


FIGURE 9. Keygraph and opportunity graphs related to Galaxy Note 5.

'SMS', 'E-mail', 'Screen-off memo', 'VR', 'Multi-tasking', 'SD card', 'Mobile communication', 'Widget', and 'Finger print'. Unexpectedly, 'Battery' and 'Charging', which are certainly interesting topics for smartphone customers, did not appear as chances because it seems that customers' interests were overly related to the topics. The topics were inconsistent with the chances defined as uncertain but relatively significant events and situations in the chance discovery theory. In fact, the interesting topics were considered to be essential problems rather than product opportunities.

The generated keygraph is a visual used to summarize all the reviews of Galaxy Note 5; it thus allows experts to efficiently analyze the massive number of customer reviews that cannot be treated by human experts alone. Therefore, experts can generate scenarios that are likely to be product opportunities by synthetically considering the chance topics and their neighboring major topics.

C. GENERATING PRODUCT OPPORTUNITIES

In this step, scenarios are generated for opportunities of Galaxy Note 5 based on opportunity graphs. Opportunity graphs, which are composed of one or more breaking topics and their neighboring major topics, were found in a keygraph. In the generated keygraph (Figure 9), only four opportunity graphs were shown, although other areas will exist depending on the perspective being analyzed. The opportunity graph A was composed of two chance topics ('Screen-off memo', 'E-mail') and their two neighbor topics ('Device connection', 'Chat'), the opportunity graph B was composed of two breaking topics ('Widget', 'Fingerprint') and their three related topics ('Icon & wallpaper', 'Physical buttons', 'Material'), the opportunity graph C was composed of a chance topic 'SD card' and its neighbor topics ('Case', 'Peripheral', 'Battery', 'Battery life', 'Battery drain'), and the opportunity graph D was composed of a chance topic ('Mobile communication') and its neighbor topics ('Wifi', 'Network dropped'). The opportunity graphs account for the improvements from a customer's viewpoint, so the graphs serve as a support tool for generating scenarios for opportunities of Galaxy Note 5.

The scenario generating process is an idea development process using Galaxy Note 5 reviews. To create meaningful opportunities, we can use a variety of creativity tools that are supported for eliciting various opportunities. Among them, we use SCAMPER, which consists of seven methods: Substitute, Combine, Adapt, Modify, Put to another use, Eliminate, and Reverse. This tool is well known as an excellent method for creating novel ideas by assembling existing products or services; it will thus serve as a guideline for analysts to generate product opportunities in this study.

By generalizing possible opportunities from the scenarios, we could acquire new Galaxy Note 5 opportunities (Table 3). From opportunity graph A and the mnemonic terms "adapt" and "put to other use", we ideated a scenario of "at the meeting, person, who wrote something through the screen-off memo, wants to email the contents". As an opportunity for the scenario, "automatic email service" can be provided and adapted handwriting or pattern recognition can be drawn. If a message is written through the screen off memo, the e-mail is written and a pattern such as a circle is drawn, we could then send an email to the address. This opportunity will be a chance for not only smartphone users but also tablet users. With the opportunity graph B and mnemonic terms "combine" and "substitute", we composed the scenario "users who use smartphones want to protect their photos or diaries from others. They will be glad if they can lock

TABLE 3. Selected scenarios and opportunities of Galaxy Note 5.

Areas	Neighbor topics	Chance topics	Mnemonic terms	Scenarios and concepts
A	Device connection, Chat	**E-mail, Screen-off memo**	Adapt Put to other uses	Scenario: at the meeting, person, who wrote something through the screen-off memo, wants to email the written contents. Opportunity: automatic email service can be provided adapting handwriting or pattern recognition.
B	Icon & wallpaper, Physical buttons, Material	**Widget, Fingerprint**	Combine, Substitute	Scenario: users who use smartphones want to protect their photos or diaries from others. They will be glad if they can lock only the specific app they want. Opportunity: Application or widget security service can be provided combining fingerprint recognition or password.
C	Case, Peripheral, Battery, Battery life, Battery drain	**SD card**	Combine, Modify	Scenario: people who like to take pictures with their smartphones have often experienced their devices storage is full. They want to get more storage without deleting anything. Opportunity: a storage expansion case can be suggested to deliver more storage for their smartphone.
D	Wifi, Network dropped	**Mobile communication**	Combine, Eliminate	Scenario: people often experience technical problems. For example, their smartphones are not connected to the mobile network. They do a factory reset but the same problem still exist. Opportunity: a systematically organized solution or guideline can be suggested to solve the difficult problems in case of network connection failure.

only the specific apps they want''. To solve this, we were able to generate an ''application or widget security service'' that can combine fingerprint recognition or password. Actually, an app is available that can lock up the specific application, although a fee is involved. Furthermore, although SAMSUNG already has a secure platform, Knox, it may be difficult for non-experts. Non-experts simply want to protect their personal data or application easily. By using the opportunity graph C and mnemonic terms ''combine'' and ''modify'', we could constructed a scenario of ''people who like to take pictures with their smartphones have often experienced their devices storage is full. They want to get more storage without deleting anything''. As a result, ''a storage expansion case'' can be suggested to deliver more storage for the smartphone. Because Galaxy Note 5 does not have SD card slots, customers cannot expand the storage of their smartphone. SAMSUNG previously developed a wireless charging battery pack to address the battery problem; however, they neglected the storage issue. For opportunity graph D and mnemonic terms ''combine'' and ''eliminate'', we proposed a scenario of ''people often experience technical problems related network connection. For example, their smartphones are not connected to the mobile network. They do a factory reset but the same problem still exist''. For this scenario, ''a systematically organized solution or guideline'' can be suggested to solve the difficult problems faced by customers. While Network drop, Disconnects from mobile network, or Hand-over problems are very difficult to solve by customers only, the problems are more frequent than we expect. These problems could be related to the reliability of the product, and may be caused by various factors such as firmware, network component problems, network quality, conflict with third party application programs, and hardware problems. However, no systematic solution is available for these problems, which means customers suffer from the inconvenience of losing the data of their smartphone through the factory reset. Therefore, providing guidance to address these issues would increase customer satisfaction with Galaxy Note 5.

V. DISCUSSION AND CONCLUSIONS

VOCs are known to be an excellent material for the process of product development. Recently, as customer voices have accumulated and been shared through various social media, many firms have attempted to utilize social media as a tool to improve their competitiveness. To this end, various social media data analyses have been conducted. However, these studies mostly centered on identifying current product trends and their relevant sentiment or opinion. In addition, while they focused only on the customers' main interests, they did not lead to exploring potential product opportunities. Therefore, an approach was proposed in this paper to identify the opportunities of a specific product through social media mining by combining topic modeling and chance discovery theory. In the steps of this approach, we used topic modeling to find what product topics customers are interested in and chance discovery theory to create new

product opportunities that may exist around breaking product topics.

To generate opportunities of a specific product, we collected large-scale online customer reviews of the product. As internet users communicate increasingly through social media, various customer opinions on a wide variety of products have accumulated on social media in real time. This means that if product-based companies can extract customer feedback on a specific product on social media, they will be instrumental in developing product opportunities; web crawling and open APIs allow analysts to collect data from social media, and NLP helps to extract keywords from the textual gathered data. Next, we attempted to generate product opportunities based on the topics in which customers were interested, by using LDA-based topic modeling. Examining all of the collected data through manual work would not be realistic. Thus, in this study, LDA was used, as it is an outstanding topic model used to identify latent topics on textual data. We found the appropriate number of topics and named each topic based on the topics' recognized characteristics. The outputs of LDA were used as input for the chance discovery. To support the scenario ideation process for product opportunities, the chance discovery theory was used. Through this theory, we could find chance topics, which were relatively rare but significant, from large-scale review data. Frequently mentioned topics are important, but might not be opportunities, and important but rare topics can be a clue to generate product opportunities. To identify such topics, called breaking topics, we constructed keygraphs that visualize the overall relationships among the topics. Using chance topics and their connected major topics, called an opportunity graph here, we can generate scenarios for the product opportunities; however, the results are over affected by the qualitative analysis of the experts. Therefore, we utilize various creativity tools that support the systematic process to generate meaningful chances.

Our approach can contribute to the systematic analysis of product opportunities from large-scale and real-time textual data from social media. To generate product opportunities, they should be analyzed from a variety of perspectives, considering the various parts of the product. In the process, product developers require a great deal of time and many resources. This study can thus become an expert support tool to generate practical product opportunities by reducing the burden on the time-consuming tasks that product developers may go through. Furthermore, as product cycles shorten and the rate of customer feedback increases, product-based companies must cope with rapidly changing market trends and customer requirements for developing sustainable improvement of their product business. Because most of the processes in this study were systemized, customer feedback can be accommodated in a short period of time and in real time. This means that the method of this paper can be used to monitor changes in customer interests and relationships among them. Finally, if a company integrates its own product R&D system and analyzes not only social media data, but also information of high-quality customers who interact with the company itself, they can derive results that will be more synergistic.

Despite the contributions of this study, several areas need further research. First, the process of generating scenarios based on the opportunity graph depends on subjective intervention by experts. To minimize this subjective dependency, we attempt to use a guideline based on SCAMPER which is an effective tool to support creative activities. In a future topic, more detailed and effective product opportunities will be generated with reduced subjectivity, if we additionally analyze semantic structures of documents belonging to a topic such as subject-action-object (SAO) structures, which are useful for semantic analysis from textual data [51]. Second, the approach is suitable for products with various functions. Products with few functions could have few customer topics, which is not appropriate for applying the process of this study. Therefore, in a further study, more advanced topic modeling or algorithms capable of extracting detailed topics for the products with a small number of functions should be applied. Third, although this study focused on deriving product opportunities from customer reviews, future research will need to consider a way to filter out particularly novel opportunities by scoring the competitiveness or likelihood of the chances. Fourth, this study generates product chances with a creativity method that identifies meaningful combinations and their product scenario from opportunity graphs. In some sense, identification of chance combinations from the opportunity graphs can be related existed research such as recommendation methods and link prediction analysis from a network. Therefore, an interesting and necessary research topic in the future will be to analyze and compare the results by various methods. Finally, only Reddit was used in this study to collect the review data of a specific product. If we can extract product review data from other renowned social media such as Twitter and Facebook, of which there are many internet users, we will be able to devise a wider range of more meaningful product opportunities.

REFERENCES

[1] W. Seo, J. Yoon, H. Park, B.-Y. Coh, J.-M. Lee, and O.-J. Kwon, "Product opportunity identification based on internal capabilities using text mining and association rule mining," *Technol. Forecasting Social Change*, vol. 105, pp. 94–104, Apr. 2016.

[2] A. W. Joshi and S. Sharma, "Customer knowledge development: Antecedents and impact on new product performance," *J. Marketing*, vol. 68, no. 4, pp. 47–59, 2004.

[3] K. Goffin and C. New, "Customer support and new product development—An exploratory study," *Int. J. Oper. Prod. Manage.*, vol. 21, no. 3, pp. 275–301, 2001.

[4] A. Griffin and J. R. Hauser, "The voice of the customer," *Marketing Sci.*, vol. 12, no. 1, pp. 1–27, 1993.

[5] E. Roman, *Voice-of-the-Customer Marketing*. New York, NY, USA: McGraw-Hill, 2011.

[6] X. Zhang and R. Chen, "Examining the mechanism of the value co-creation with customers," *Int. J. Prod. Econ.*, vol. 116, no. 2, pp. 242–250, 2008.

[7] J. Teixeira, L. Patrício, N. J. Nunes, L. Nóbrega, R. P. Fisk, and L. Constantine, "Customer experience modeling: From customer experience to service design," *J. Service Manage.*, vol. 23, no. 3, pp. 362–376, 2012.

[8] H.-S. Park and S. J. Noh, "Enhancement of Web design quality through the QFD approach," *Total Quality Manage.*, vol. 13, no. 3, pp. 393–401, 2002.
[9] T. Sakao, "A QFD-centred design methodology for environmentally conscious product design," *Int. J. Prod. Res.*, vol. 45, nos. 18–19, pp. 4143–4162, 2007.
[10] S. H. Min, H. Y. Kim, Y. J. Kwon, and S. Y. Sohn, "Conjoint analysis for improving the e-book reader in the Korean market," *Expert Syst. Appl.*, vol. 38, no. 10, pp. 12923–12929, 2011.
[11] K. Im and H. Cho, "A systematic approach for developing a new business model using morphological analysis and integrated fuzzy approach," *Expert Syst. Appl.*, vol. 40, no. 11, pp. 4463–4477, 2013.
[12] A. M. M. S. Ullah and J. Tamaki, "Analysis of Kano-model-based customer needs for product development," *Syst. Eng.*, vol. 14, no. 2, pp. 154–172, 2011.
[13] H. Raharjo, A. C. Brombacher, T. N. Goh, and B. Bergman, "On integrating Kano's model dynamics into QFD for multiple product design," *Quality Rel. Eng. Int.*, vol. 26, no. 4, pp. 351–363, 2010.
[14] X.-X. Shen, M. Xie, and K.-C. Tan, "Listening to the future voice of the customer using fuzzy trend analysis in QFD," *Quality Eng.*, vol. 13, no. 3, pp. 419–425, 2001.
[15] C. Kahraman, T. Ertay, and G. Büyüözkan, "A fuzzy optimization model for QFD planning process using analytic network approach," *Eur. J. Oper. Res.*, vol. 171, no. 2, pp. 390–411, 2006.
[16] A. M. Kaplan and M. Haenlein, "Users of the world, unite! The challenges and opportunities of social media," *Bus. Horizons*, vol. 53, no. 1, pp. 59–68, 2010.
[17] W. G. Mangold and D. J. Faulds, "Social media: The new hybrid element of the promotion mix," *Bus. Horizons*, vol. 52, no. 4, pp. 357–365, 2009.
[18] R. Zafarani, M. A. Abbasi, and H. Liu, *Social Media Mining: An Introduction*. Cambridge, U.K.: Cambridge Univ. Press, 2014.
[19] Z. Xu, G. L. Frankwick, and E. Ramirez, "Effects of big data analytics and traditional marketing analytics on new product success: A knowledge fusion perspective," *J. Bus. Res.*, vol. 69, no. 5, pp. 1562–1566, 2016.
[20] Y.-H. Joo, S. Kim, and S.-J. Yang, "Valuing customers for social network services," *J. Bus. Res.*, vol. 64, no. 11, pp. 1239–1244, 2011.
[21] J. Gallaugher and S. Ransbotham, "Social media and customer dialog management at Starbucks," *MIS Quart. Executive*, vol. 9, no. 4, pp. 197–212, 2010.
[22] E. C. Malthouse, M. Haenlein, B. Skiera, B. Wege, and M. Zhang, "Managing customer relationships in the social media era: Introducing the social CRM house," *J. Interact. Marketing*, vol. 27, no. 4, pp. 270–280, 2013.
[23] K. Chen, G. Kou, J. Shang, and Y. Chen, "Visualizing market structure through online product reviews: Integrate topic modeling, TOPSIS, and multi-dimensional scaling approaches," *Electron. Commerce Res. Appl.*, vol. 14, no. 1, pp. 58–74, 2015.
[24] F. Misopoulos, M. Mitic, A. Kapoulas, and C. Karapiperis, "Uncovering customer service experiences with Twitter: The case of airline industry," *Manage. Decision*, vol. 52, no. 4, pp. 705–723, 2014.
[25] F. H. Khan, S. Bashir, and U. Qamar, "TOM: Twitter opinion mining framework using hybrid classification scheme," *Decision Support Syst.*, vol. 57, pp. 245–257, Jan. 2014.
[26] S. Tuarob and C. S. Tucker, "Quantifying product favorability and extracting notable product features using large scale social media data," *J. Comput. Inf. Sci. Eng.*, vol. 15, no. 3, p. 031003, 2015.
[27] M. M. Mostafa, "More than words: Social networks' text mining for consumer brand sentiments," *Expert Syst. Appl.*, vol. 40, no. 10, pp. 4241–4251, 2013.
[28] Y. Park and S. Lee, "How to design and utilize online customer center to support new product concept generation," *Expert Syst. Appl.*, vol. 38, no. 8, pp. 10638–10647, 2011.
[29] W. X. Zhao, S. Li, Y. He, L. Wang, J.-R. Wen, and X. Li, "Exploring demographic information in social media for product recommendation," *Knowl. Inf. Syst.*, vol. 49, no. 1, pp. 61–89, 2015.
[30] A. Chua *et al.*, "Mapping Cilento: Using geotagged social media data to characterize tourist flows in southern Italy," *Tourism Manage.*, vol. 57, pp. 295–310, Dec. 2016.
[31] Q. Huang and D. W. S. Wong, "Activity patterns, socioeconomic status and urban spatial structure: What can social media data tell us?" *Int. J. Geograph. Inf. Sci.*, vol. 30, no. 9, pp. 1873–1898, 2016.
[32] J. Füller, M. Bartl, H. Ernst, and H. Mühlbacher, "Community based innovation: How to integrate members of virtual communities into new product development," *Electron. Commerce Res.*, vol. 6, no. 1, pp. 57–73, 2006.
[33] Y. Ohsawa, "Chance discovery: The current states of art," in *Chance Discoveries in Real World Decision Making*. Berlin, Germany: Springer, 2006, pp. 3–20.
[34] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation," *J. Mach. Learn. Res.*, vol. 3, pp. 993–1022, Mar. 2003.
[35] Y. Ohsawa, "Chance discoveries for making decisions in complex real world," *New Generat. Comput.*, vol. 20, no. 2, pp. 143–163, 2002.
[36] D. M. Blei and J. D. Lafferty, "Topic models," in *Text Mining: Classification, Clustering, and Applications*, vol. 10. New York, NY, USA: CRC Press, 2009, p. 34.
[37] I. Biró, J. Szabó, and A. A. Benczúr, "Latent Dirichlet allocation in Web spam filtering," in *Proc. 4th Int. Workshop Adversarial Inf. Retr. Web*, 2008, pp. 29–32.
[38] T. Hua *et al.*, "Topical analysis of interactions between news and social media," in *Proc. 30th AAAI Conf. Artif. Intell.*, 2016, pp. 2964–2971.
[39] B. D. Blair, C. M. Weible, T. Heikkila, and D. Evensen, "Comparing human and automated coding of news articles on hydraulic fracturing in New York and Pennsylvania," *Soc. Natural Resour.*, vol. 29, no. 7, pp. 880–884, 2016.
[40] J.-W. Son, A. Kim, and S.-B. Park, "A location-based news article recommendation with explicit localized semantic analysis," in *Proc. 36th Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, 2013, pp. 293–302.
[41] Y. Guo, S. J. Barnes, and Q. Jia, "Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent Dirichlet allocation," *Tourism Manage.*, vol. 59, pp. 467–483, Apr. 2017.
[42] L.-C. Chen, T.-J. Yu, and C.-J. Hsieh, "KeyGraph-based chance discovery for exploring the development of e-commerce topics," *Scientometrics*, vol. 95, no. 1, pp. 257–275, 2013.
[43] H. Wang and Y. Ohsawa, "Idea discovery: A scenario-based systematic approach for decision making in market innovation," *Expert Syst. Appl.*, vol. 40, no. 2, pp. 429–438, 2013.
[44] Y. Ohsawa, "*KeyGraph*: Visualized structure among event clusters," in *Chance Discovery*. Berlin, Germany: Springer, 2003, pp. 262–275.
[45] Y. Ohsawa, "Data crystallization: Chance discovery extended for dealing with unobservable events," *New Mat. Natural Comput.*, vol. 1, pp. 373–392, Nov. 2005.
[46] S. Liu and C. Chen, "The differences between latent topics in abstracts and citation contexts of citing papers," *J. Amer. Soc. Inf. Sci. Technol.*, vol. 64, no. 3, pp. 627–639, 2013.
[47] B. Eberle, *Scamper on: Games for Imagination Development*. Austin, TX, USA: Prufrock Press, 1996.
[48] B. Molina, J. Monteiro-Leitner, M. T. Garrett, and S. T. Gladding, "Making the connection interweaving multicultural creative arts through the power of group counseling interventions," *J. Creativity Mental Health*, vol. 1, no. 2, pp. 5–15, 2005.
[49] D. Mann, *Hands on Systematic Innovation*. kortrijk, Belgium: Creax press, 2002.
[50] P.-C. Lee, H.-N. Su, and T.-Y. Chan, "Assessment of ontology-based knowledge network formation by vector-space model," *Scientometrics*, vol. 85, no. 3, pp. 689–703, 2010.
[51] J. Yoon and K. Kim, "Identifying rapidly evolving technological trends for R&D planning using SAO-based semantic patent networks," *Scientometrics*, vol. 88, no. 1, pp. 213–228, 2011.

NAMUK KO was born in Jeju, South Korea, in 1991. He received the B.S. and M.S. degrees in industrial engineering from Konkuk University, Seoul, South Korea, in 2016 and 2017, respectively.

His research topics include patent intelligence, technology opportunity discovery, and social media mining for product and technology planning.

Mr. Ko is a member of the Korean Institute of Industrial Engineers.

BYEONGKI JEONG was born in Gyeonggi, South Korea, in 1993. He received the B.S. degree in industrial engineering from Konkuk University, Seoul, South Korea, in 2017.

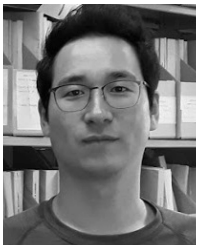
His research topics include technology intelligence, patent mining, technology opportunity identification, and social media mining for future planning.

Mr. Jeong is a member of the Korean Institute of Industrial Engineers.

SUNGCHUL CHOI was born in Busan, South Korea, in 1982. He received the B.S degree in management from Handong Global University and the Ph.D. degree in industrial management engineering from the Pohang University of Science Technology, South Korea, in 2006 and 2012, respectively.

He was a Technology Strategy Manager with the CTO Office, Samsung Advanced Institute of Technology. He has conducted research on patent analysis, technology roadmapping, and strategic planning, using a text mining approach. He is currently an Assistant Professor with the Department of Industrial and Management Engineering, Gachon University. He has been involved in applying machine learning and deep learning approaches to patent big data and has been conducting various studies related to NLP and artificial intelligence.

Prof. Choi is a member of the Korean Institute of Industrial Engineers.



JANGHYEOK YOON was born in Daegu, South Korea, in 1979. He received the B.S., M.S., and Ph.D. degrees in industrial engineering from the Pohang University of Science Technology, Pohang, South Korea, in 2002, 2004, and 2011, respectively.

He was an IT Consultant and an Application Engineer for LG CNS, an affiliate of the LG Group, from 2004 to 2007, and was an Associate Researcher with the Korea Institute of Intellectual Property, a Korean Public Research Institute, from 2011 to 2012. Since 2012, he has been an Associate Professor with the Department of Industrial Engineering, Konkuk University, Seoul, South Korea. He has authored over 50 articles, and holds ten patents related to patent mining-based technology intelligence and opportunity discovery systems. His research interests are on technology and service intelligence, which covers the analytical and multidisciplinary methodologies for technology, service (business model) and future planning, and thus they are based on the combined application of theories and analytics in various fields, such as industrial engineering, business administration, and computer science.

Prof. Yoon is a member of the Korean Institute of Industrial Engineers and has been an Editor of the *Journal of Intellectual Property*, since 2014.