Identifying product opportunities using topic modeling and sentiment analysis of social media data

Byeongki Jeong
Department of Industrial Engineering
Konkuk University, Seoul, Korea
Email: wjd958@konkuk.ac.kr

Janghyeok Yoon †
Department of Industrial Engineering
Konkuk University, Seoul, Korea
Email: janghyoon@konkuk.ac.kr

Abstract. Detecting customer needs is the most prerequisite for new product development and current product improvement. In this study, we propose a novel approach to identify product opportunities by applying topic modeling and sentiment analysis. First, this approach defines major contextual features of a product in need of product planning by applying topic modeling to social web data of the product. Second, it computes customer satisfaction for each feature by keyword(or key phrase)-level sentiment analysis. Finally, this approach quantifies the opportunity level of each feature through the opportunity algorithm based on the concepts of importance and satisfaction. This approach will contribute to monitoring customer needs in real time and identifying product development concepts.

Keywords: Product opportunity; Topic modeling; Sentiment analysis; Opportunity algorithm; Web data

1. INTRODUCTION

Detecting new opportunities for a product is one of the most important things for firms' sustainable growth (Park 2005). A firm can increase customers' satisfaction level by providing a product that complements new or insufficient functionalities (Fornell, Johnson et al. 1996). But, it is difficult to define customer needs because many of the customer needs tend to be hidden or implicit (Kärkkäinen, Piippo et al. 2001). In addition, defining customer needs and thereby extracting new product opportunities is not an easy task (Isaksson, Larsson et al. 2009).

For these reasons, this paper proposes a new approach that identifies product opportunities by the opportunity algorithm based on topic modeling and sentiment analysis. The proposed method defines product features from large-scale social media data by topic modeling; in this study, each product feature corresponds to a subject that customers express on social media data and the stock quantity of each product feature is used to identify the importance of its corresponding product feature from a customer perspective. Subsequently, the approach determines the satisfaction level of each product feature by analyzing the sentiment expressed within the product feature. Lastly, the method quantifies and visualizes the opportunity levels of the product features by the opportunity algorithm consisting of the importance and satisfaction values.

This study will contribute to advancing customer-centered product planning. Our method quantifies the process for product opportunities by identifying product features and their importance and satisfaction from large-scale social media data. In addition, our method for new opportunity identification has the potential to be applied to not only products but also services, product-service system, and technology development.

2. THEORETICAL BACKGROUND

Our approach has three theoretical backgrounds: opportunity algorithm, topic modeling and sentiment analysis. They are explained in detail in this section.

2.1 Opportunity algorithm

Opportunity algorithm is a method for quantifying and

measuring innovation opportunity. It was proposed along with Outcome-driven innovation (ODI) by Ulwick (Ulwick 2005). This algorithm uses importance and satisfaction to compute opportunities on a scale of 0 to 10 from customer perspectives (Equation 1). The opportunity landscape map can be drawn by components with their importance and satisfaction. Components on the opportunity landscape map are divided into served-right, overserved and underserved (Figure 1). Components in served-right make appropriate satisfaction compared to their importance to the customer. Overserved components mean higher satisfaction than their importance. Lastly, underserved components provide lower satisfaction than their importance. These underserved components can be understood as innovation opportunities.

In this study, we use the opportunity algorithm to identify development opportunities for product features from customer perspectives. In particular, we will find product development opportunities with not only a visual by opportunity landscape map but also opportunity score computation.

$$\text{Opportunity} = \text{Importance} + \text{Max}(\text{Importance} - \text{Satisfaction}, 0) \quad (1)$$

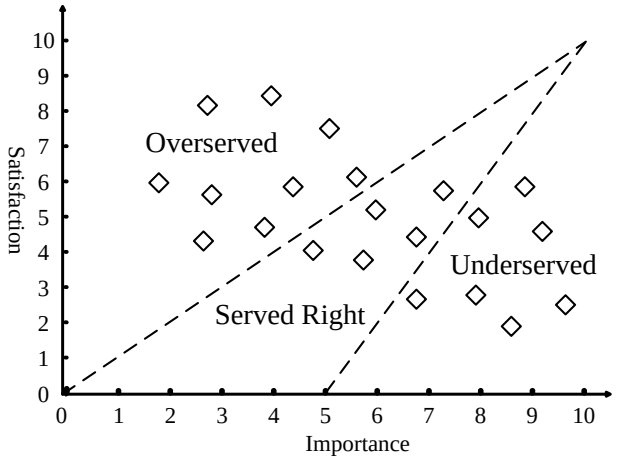


Figure 1: Schematic of opportunity landscape maps

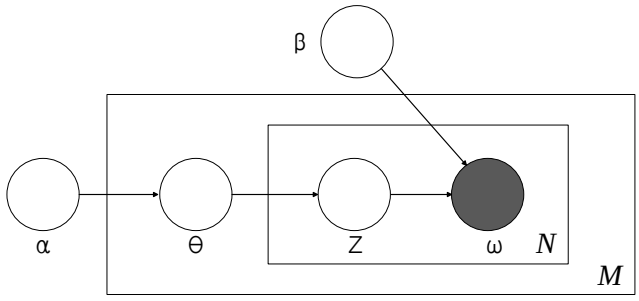
2.2 Topic modeling

Text documents consist of words. Topics spoken in documents are expressed by a combination of words and documents can belong to multiple topics. In particular, documents can share similar topics and they are expressed by the occurrence of similar words. The topic modeling is a method to inference topics of text documents by words related to context. Especially, it can be expressed by the probability that how many expressions each topic in documents. In addition, documents can be classified by the inference result of topics.

Some prior topic modeling algorithms have been proposed. Among the algorithms, this study uses Latent Dirichlet Allocation (LDA) topic modeling algorithm. The LDA topic modeling has developed from Probabilistic Latent Semantic Indexing (pLSI) topic modeling (Blei, Ng et al. 2003). The pLSI topic modeling algorithm does not have a probability inference model of document-level. The LDA is also known to have the highest performance among vairous topic modeling algorithms (Chiru, Rebedea et al. 2014).

The procedure of LDA model is composed of three steps (Figure 2). First, the words dictionary was built by keywords in a corpus. Then, the model infers the probability of terms for topics using the dirichlet distribution. As a result, we can know the contribution of each term to make up topics. Lastly, the model infers topics probability in documents with topic-term distribution. This topics-documents distribution as the final output of LDA is used for document classification.

In this study, the LDA topic modeling is used to define product features and the importance of the features from customer perspectives. This is because product features are originated from the social media data made by users (or customers). Then, the stock of each product feature (or topic) indicates the importance of the product feature from customer perspectives.



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

Figure 2: Concepts of LDA-based topic modeling

2.3 Sentiment analysis

The sentiment analysis is one of the major research fields of Natural language process (NLP). Sentiment analysis is also called opinion mining or polarity detection. Sentiment analysis is a detection method for sentiment (positive or negative) in speech. Sentiment analysis detects not only the polarity but also the degree of sentiment.

The sentiment analysis has three major ways. First, the

rule-based analysis is detecting terms' sequence in speech. Some rule-based sentiment analysis studies use an ontology to analyze the sentiment of speech (Wei and Gulla 2010). Second, the lexicon-based analysis uses a dictionary for sentiment words likes SentiWordNet (Baccianella, Esuli et al. 2010). The lexicon-based sentiment analysis is easy to use. However, it has a limitation, in that it measures only sentence-level or document-level sentiment. Lastly, the deep learning-based analysis this study adopts measures word-level sentiment scores. The deep learning-based sentiment analysis uses a deep neural network model to analyze sentiments of speech. The deep learning-based analysis is also known as the highest performance method to analyze sentiment among various methods (Dasgupta and Ng 2009).

To measure customers' opinions about various features of the target product, this study uses the deep learning-based analysis for keyword-level sentiment analysis to large-scale text data.

3. METHODOLOGY

This study proposes a new method to identify product development opportunities using three theoretical backgrounds: opportunity algorithm, topic modeling and sentiment analysis. The features of the target product and their importance will be defined by topic modeling from large-scale web data. Then, we compute customers' satisfaction of the features using the keyword-level sentiment analysis. Finally, the opportunity score of each feature is identified by opportunity algorithm. The specific overall procedure is below (Figure 3.).

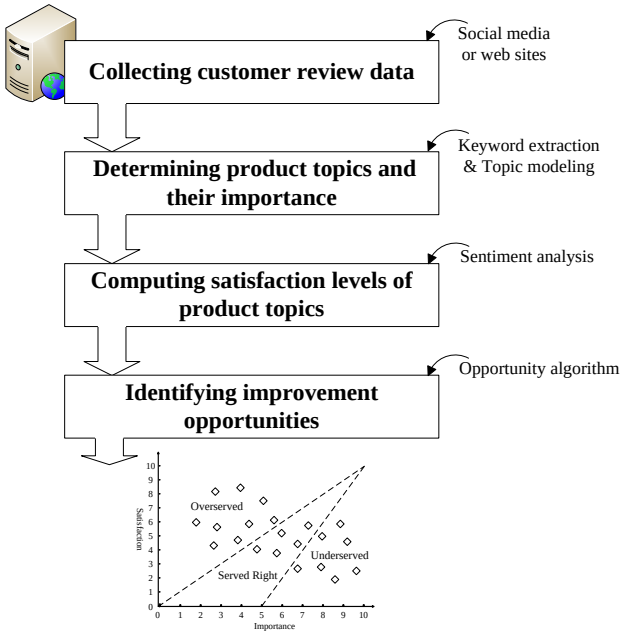


Figure 3: Overview of the proposed method

3.1 Data gathering and preprocessing

First, we collect large-scale web data related to the target product. Because this approach will compute customer satisfactions using sentiment analysis, we mainly collect social media data or social web data posted from customers. Then, we extract keywords (or key-phrase) related to the target product from the web data. Keywords must have relation to the target product and thus keywords are used to extract the features of product through topic modeling. Therefore, we select keywords that have relation to the target product.

3.2 Defining features of product and computing importance of features

We execute LDA topic modeling using the web data and selected keywords. The LDA topic modeling algorithm has three major inputs. Three major inputs are corpus, word dictionary and the number of topics. The corpus is large-scale web data and the word dictionary consists of selected keywords. Lastly, the number of topics is decided by a prior study that is the lowest number of average cosine similarity of pairs of topics (Wang, Liu et al. 2014).

Features of the target product were defined by LDA topic modeling and they are actually the features which customers are interested in about the target product. This is because those features are defined from web data originated from customers. The importance of each product feature also can be quantified by analyzing how many the product feature occurred in user mentions (Song, Lee et al. 2015). Therefore, the contribution stock as a summation of document contributions for a feature indicates the importance of the feature from customer perspectives in the corpus (Figure 4.).

Then, we use normalized contribution stocks on a scale of 0 to 10 for the importance of product features used in opportunity algorithm. The computed score is the relative importance of the features of the target product. The formula about transforming contribution stock to the importance for $Feature_i$ is Equation 2.

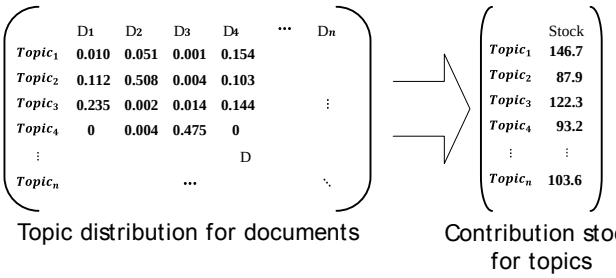


Figure 4: Concepts of computing the importance of product features

$$\text{Importance}_i = 10 \times \frac{Stock_i - Stock_{Min}}{Stock_{Max} - Stock_{Min}} \quad (2)$$

3.3 Computing customer satisfaction

In this step, we define customers' satisfaction of features of the target product to use identifying product development opportunities. First, we compute weighted sentiment scores for features by a multiple of the topics-documents contribution matrix the result of LDA topic modeling and average sentiment scores of keywords the result of sentiment analysis. Because the same keyword has different sentiment score in various customer reviews then we use average sentiment score for representative value of the keywords. Then, we standardize weighted sentiment scores of each feature. The positive area of standardized distribution is satisfaction for each feature (Figure 5.). We use positive area on a scale of 0 to 10 for opportunity algorithm. The formula about the transform weighted sentiment score to the customer satisfaction for $Feature_i$ is Equation 3.

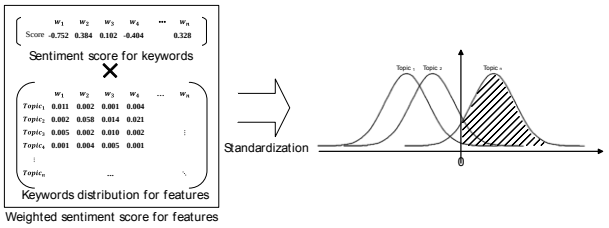


Figure 5: Concepts of computing customer satisfaction

$$\text{Satisfaction}_i = 10 \times \text{P}(\text{z} \geq 0) \quad \because \text{z} \sim \text{N}(\overline{X}_i, {\sigma_i}^2) \quad (3)$$

3.4 Finding product development opportunity

Importance and satisfaction scores of each feature on a scale 0 to 10 can make up the opportunity landscape map. Through the landscape map, we can find features that have highly opportunities than other features with separate the map into three areas overserved underserved and served right. Features of the target product closer to the underserved area have a high probability for product development opportunities. But specific opportunity scores of features of the target product were computed by the opportunity algorithm.

In this approach, we can know not only highly opportunity features but also development directions of features via weighted sentiment scores for features. Each feature includes many keywords that have a contribution to the feature and weighted by sentiment scores. Therefore, we can find both satisfaction and dissatisfaction points for each feature. Finally, we can set product development plans to increase the satisfaction or decrease the dissatisfaction.

4. CASE STUDY: Samsung galaxy note 5

We propose an approach to identify product development opportunities using LDA topic modeling and sentiment analysis. This approach is useful to be applied to multifunctional products because the topic modeling can define various features of a given product. In addition, this approach requires large-scale comments that may contain various customer needs. Smartphones are a representative product which is multifunctional and contains various customer opinions in social media. For this reason, our case study is conducted to identify product development opportunities of Samsung galaxy note 5 (SGN5).

4.1 Data gathering and preprocessing

This case study used posts and comments data of Reddit (https://www.reddit.com/). Reddit, one of the major social web sites in the United States, is made up with lots of subreddits, including SGN5. Therefore, we were able to gather the customer review data related to our target product with minimal noise. 23,613 customer reviews (2,255 posts and 21,358 comments) was collected during the period between 05 Dec 2014 and 31 Jan 2016; the release date of SGN5 in the United States is 21 Aug 2015. Then, we analyzed the data using AlchemyAPI for natural language processing, thereby extracting an initial set of 32,656 keywords. By eliminating irrelevant or noise terms from the set, we selected 3,539 keywords related to SGN5. Finally, for this case study we used 23,613 customer reviews and 3,549 keywords.

4.2 Defining product features and computing their importance

We executed the topic modeling using customer reviews and keywords. The number of topics that used for topic modeling was 65, because it can be decided by the lowest number of average cosine similarity between all pairs of topics. The cosine similarity is a measure to identify the similarity between two discrete vectors and its value ranges between 0 and 1. The average cosine similarity for all pairs of topics decreased until 60, and it was found to have the lowest value when the number of topics is 65. Then, the similarity value showed a trend of a rebound from 70 (Figure 6).

The 65 features of SGN5 were named by their major contributing keywords (and key phrases). Then, the stock quantity of each feature can be computed from documents' belongness for the feature (Table 1). The stock quantity value ranged 202.2198 (maximum stock quantity) and 165.9974 (minimum stock quantity). The average value of stocks was 171.1228. Then, the stock quantity values for product features were normalized and transformed into the

importance values for the product features by a factor of ten. Therefore, the importance value for each product feature was transformed into a value between 0 and 10.

The most important feature of SGN5 was found to be 'Samsung Pay (Importance: 10)' and the least important one was 'Accessory (Importance: 0)'. In fact, 'Samsung Pay' using Magnetic Secure Transmission (MST) is a new feature provided by not only SGN5 but also all the other new Samsung smartphones. It supports payment with a virtual credit card without Near Field Communication (NFC); similar functions are 'Apple Pay' and 'Android Pay'. Therefore, it reflects a customer concern for new technology.

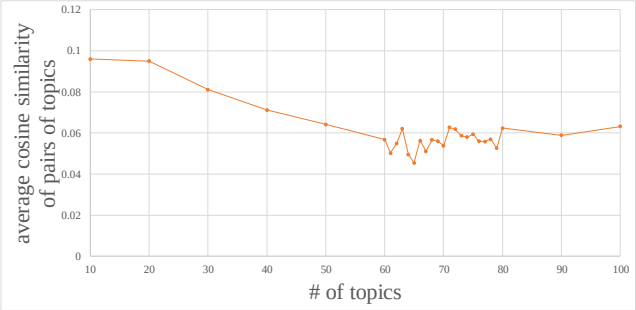


Figure 6: The average cosine similarity of pairs of topics

Table 1: Part of contributing keywords and the importance of product features

Features	Major contribution keywords	Stock	Importance
Design	Design(0.118), Material(0.029), Material design(0.014), OS(0.009)	166.8327	0.2306
Fingerprint	Fingerprint(0.171), Sensor(0.029), Finger print scanner(0.026), Smart lock(0.011)	169.6524	1.0090
Calling	Calling(0.079), Battery(0.049), VoLTE(0.037), Wi-Fi calling(0.036),	167.2703	0.3514
Optimization	OS(0.125), UI(0.071), Optimization(0.024), App optimization(0.009)	167.2323	0.3409
Wi-Fi	Wifi(0.288), Network(0.015), Speed(0.015), Hotspot(0.006)	173.6667	2.1173
UX	UI(0.227), OS(0.189), Bug(0.045), Smooth(0.029), Software(0.029)	167.4583	0.4033
Custom Rom	Flash(0.107), ROM(0.059), Knox(0.034), Recovery(0.03)	169.4675	0.9580
Edge display	Edge(0.355) Curve(0.044), Protection(0.016), Screen(0.005)	167.1840	0.3276
Expandability	SD(0.130), Removable battery(0.049), MicroSD(0.036), IR blaster(0.016)	179.3538	3.6873
Software update	Update(0.344), Software(0.053) Security(0.012), Marshmallow update(0.005)	182.7272	4.6186

4.3 Computing customer satisfaction

In this step, customer satisfaction of product features is computed by sentiment analysis. First, average sentiment score was computed for 3,549 keywords (Table 2.). Through the average sentiment score, the battery had many negative customer opinions. The battery has two major product features (topics) in SGN5. One feature is related to battery life; for example, its keywords are 'removable battery (-0.2453)', 'non-removable battery (-0.5415)' and 'fixed battery (-0.2839)'. The other feature relates to removable battery; its keywords are 'battery life (-0.5910)', 'battery size (-0.856)' and 'good battery life (0.5437)'

Table 2: Part of average sentiment scores for keywords of SGN5

Keywords	Average sentiment score
battery	-0.2334
great upgrade	0.7703
fast autofocus	0
removable battery	-0.2453
wireless charging capabilities	-0.4315
16MP camera	0.2657
SD card	-0.2963
AMOLED	0.3885
design	0.2438
UI	-0.2208
Screen size	0.1937
fingerprint scanner	-0.0155

The average sentiment score of all keywords was multiplied by the features-keywords matrix that is the result of LDA topic modeling. As a result, a weighted sentiment score matrix for features was generated. A row of weighted sentiment score matrix is a weighted sentiment score vector of a product feature. We use a normal distribution and compute $p(z \geq 0)$s to standardize each sentiment score vector. Finally, we were able to obtain satisfaction scores for 65 features, ranging between 0 and 10 (Table 3.).

Table 3: Part of distribution and satisfaction of features

Features	$E(Feature_i)$	$Var(Feature_i)$	Satisfaction
Design	-0.00002224	0.00000183	5.1441
Fingerprint	-0.00003417	0.00000093	2.6905
Calling	-0.00004204	0.00000065	0.5420
Optimization	-0.00002808	0.00000061	2.6288
Wi-Fi	-0.00003770	0.00000120	2.8175
UX	-0.00003114	0.00000126	3.6903
Custom Rom	-0.00002524	0.00000038	1.9508
Edge display	-0.00001147	0.00000013	3.2024
Expandability	-0.00005756	0.00000105	0
Software update	-0.00002588	0.00000097	3.8742

The average of satisfaction scores was 4.0361. In particular, the 'Samsung Pay' had the highest satisfaction in SGN5. It also had the highest importance feature in our analysis result. Therefore, the 'Samsung Pay' as a new feature was found to have the highest importance and highest satisfaction despite its new provision to customers. On the other hand, the 'Expandability' recorded the least satisfaction among the product features of SGN5. According to our best knowledge, this is because some features, such as external memory card, removable battery, and IR blaster, have been eliminated in SGN5; those product features were provided by previous versions of SGN5, such as Samsung galaxy note 4.

4.4 Finding product development opportunities

Finally, product development opportunities are quantified using the importance and satisfaction scores for product features. To grasp an overall trend of product opportunities, we generated an opportunity landscape map for SGN5 using the distributions of importance and satisfaction of the product features (Figure 7.). Two orange lines separate the areas of Over-served (Top), Served right (Middle), Under-served (Bottom). They are drawn from the average of importance (1.4150) and satisfaction (4.0361). Total 48 features of 65 features are found to be in 'Served right' area, 11 features in 'Over-served' area and 5 features in 'Under-served' area. In particular, 5 underserved features are 'Fast charge', 'Detect pes2', 'Charge cable', 'Screen glass' and 'Expandability'. Also, the 1 product feature ('Samsung Pay') has highest values both importance and satisfaction. In a simple sense, one can conclude that these 6 features have a higher opportunity than the others.

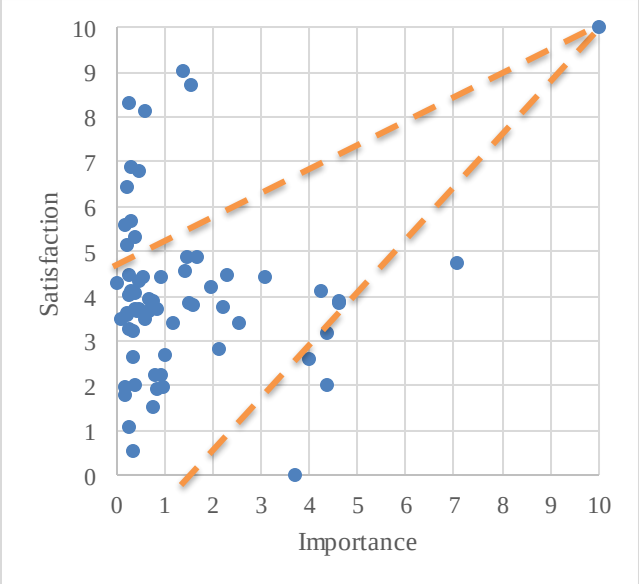


Figure 7: The opportunity landscape map of SGN5

Actually, our analysis found that those 6 product features have a higher opportunity, compared to other features (Table 4.). For the product features with a high opportunity score, specific development directions of the product can be decided by using contributing keywords and their sentiment score for the features (Table 5.). Directions can be decided dynamically, but in this study, we guess the directions to improve negative keywords for each feature. In particular, the 'Samsung Pay (Opportunity: 10)' is the highest opportunity feature in SGN5. It has both the highest importance and satisfaction, although it is a new feature that has not been provided before. But, it has some negative points. The negative points of 'Samsung Pay' are NFC support because most of the keywords that have negative sentiment are related to NFC. 'Samsung Pay' does not support NFC payment yet. It supports MST only. Therefore, 'Samsung Pay' could consider supporting NFC payment. The second highest opportunity feature is 'Fast charge (Opportunity: 9.3706)'; the importance and satisfaction of 'Fast charge' are 7.0598 and 4.7492. SGN5 currently contains a device charging technology that is twice or three-times faster than before. 'Fast charge' was the second highest importance, but it seems not to provide enough satisfaction to customers. Our detail analysis found that major negative keywords for 'Fast charge' are related to wireless charging and chargers. For this reason, 'Fast charge' could be improved by enhancing wireless charging and charging support devices. The third highest opportunity feature is 'Expandability'. It involves removable parts of SGN5 to assist usage. The most representative parts are removable battery and external storage. What 'Expandability' has the third highest opportunity indicates that customers feel some discomfort related to Micro SD use, Replaceable battery and IR blaster, compared to SGN4.

Table 4: Top 10 product features with a high opportunity score

Features	Importance	Satisfaction	Opportunity
Samsung Pay	10	10	10
Fast charge	7.0598	4.7492	9.3706
Expandability	3.6873	0	7.3746
Charge cable	4.3514	2.0019	6.7009
Detect pen2	4.3805	3.1571	5.6038
Pen out	4.6130	3.8344	5.3916
Software update	4.6186	3.8742	5.3630
Screen glass	3.9796	2.6008	5.3585
Battery life	4.2233	4.1161	4.3305
Screenshot	3.0895	4.4210	3.0895

Table 5: Part of keywords contribution weighted by sentiment

Keywords	Samsung Pay	Keywords	Fast charge	Keywords	Expandability
NFC	-0.00238071	charge	-0.04903734	SD	-0.03627358
NFC Pay	-0.00125675	charger	-0.03531872	SD card	-0.02764752
Google wallet	-0.00074977	wireless Charger	-0.00471534	car	-0.02519669
android pay	-0.00065243	Samsung Wireless Charger	-0.00089303	micro SD	-0.01419630
support Samsung Pay	-0.00036844	Charges	-0.00079399	MicroSD	-0.01246854
NFC payment	-0.00032814	car charger	-0.00066825	removable battery	-0.01191136
NFC terminals	-0.00020921	fast wireless chargers.	-0.00054525	SD card slot.	-0.01189686
Pay apps	-0.00020510	blue light	-0.00053208	SD-Card	-0.01055452
S-Pay	-0.00016297	charger work	-0.00047055	SD cards	-0.00956093
sPay	-0.00015923	wall charger	-0.00027726	SD card slot	-0.00647964

5. CONCLUSION

Customer needs from the voice of customer are the most important thing to development products (Kaulio 1998). Recently, the voice of customer appears on various media (Arnold, Barrow et al. 2007). In particular, website and social media are becoming an important source for gathering frank and true opinions from customers (Malthouse, Haenlein et al. 2013).

This study proposed a new approach to identify product development opportunities from web data including social media. The proposed approach uses the opportunity algorithm, topic modeling and the sentiment analysis to identify product development opportunities. First, we gather customer review data related to a target product from website or social media. Second, the features of the product and their importance were defined by the LDA topic modeling and keywords extraction. Then, the satisfaction of features was computed by sentiment analysis. Lastly, product development opportunities that consist of importance and satisfaction are computed. We applied the approach to SGN5 and were able to identify product features with high opportunity and formulate product development directions using the features' contributing keywords and their weighted sentiment score.

Our approach presents a quantifying method for product development opportunities based on web data. Therefore, this approach is expected to facilitate customer-centered product planning and development. In addition, this approach using social media data has the potential to be applied to not only products but also service, product-service systems to identify their development opportunities.

ACKNOWLEDGMENTS

This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Science, ICT & Future Planning (No. 2015R1A1A1A05027889)

REFERENCES

Arnold, J., A. Barrow, et al. (2007). Co-Design in Virtual Space: Including Everyday People in Product Design. National Education Symposium of the Industrial Designers Society of America. San Francisco, CA.

Baccianella, S., A. Esuli, et al. (2010). SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining. LREC.

Blei, D. M., A. Y. Ng, et al. (2003). "Latent dirichlet allocation." the Journal of machine Learning research **3**: 993-1022.

Chiru, C.-G., T. Rebedea, et al. (2014). Comparison between LSA-LDA-Lexical Chains. WEBIST (2).

Dasgupta, S. and V. Ng (2009). Mine the easy, classify the hard: a semi-supervised approach to automatic sentiment classification. Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 2-Volume 2, Association for Computational Linguistics.

Fornell, C., M. D. Johnson, et al. (1996). "The American customer satisfaction index: nature, purpose, and findings." the Journal of Marketing: 7-18.

Isaksson, O., T. C. Larsson, et al. (2009). "Development of product-service systems: challenges and opportunities for the manufacturing firm." Journal of Engineering Design **20**(4): 329-348.

Kärkkäinen, H., P. Piippo, et al. (2001). "Assessment of hidden and future customer needs in Finnish business-to-business companies." R&D Management **31**(4): 391-407.

Kaulio, M. A. (1998). "Customer, consumer and user involvement in product development: A framework and a review of selected methods." Total Quality Management **9**(1): 141-149.

Malthouse, E. C., M. Haenlein, et al. (2013). "Managing customer relationships in the social media era: introducing the social CRM house." Journal of Interactive Marketing **27**(4): 270-280.

Park, J. S. (2005). "Opportunity recognition and product innovation in entrepreneurial hi-tech start-ups: a new perspective and supporting case study." Technovation **25**(7): 739-752.

Song, B., C. Lee, et al. (2015). "Diagnosing service quality using customer reviews: an index approach based on sentiment and gap analyses." Service Business: 1-24.

Ulwick, A. W. (2005). What customers want: using outcome-driven innovation to create breakthrough products and services., McGraw-Hill New York.

Wang, B., S. Liu, et al. (2014). "Identifying technological topics and institution-topic distribution probability for patent competitive intelligence analysis: a case study in LTE technology." Scientometrics **101**(1): 685-704.

Wei, W. and J. A. Gulla (2010). Sentiment learning on product reviews via sentiment ontology tree. Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics.