IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 26 June 2023, accepted 10 July 2023, date of publication 14 July 2023, date of current version 19 July 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3295452

APPLIED RESEARCH

Context-Aware Customer Needs Identification by Linguistic Pattern Mining Based on Online Product Reviews

JIHO LEE[1], BYEONGKI JEONG[2], JANGHYEOK YOON[1], AND CHIE HOON SONG[3]
[1]Department of Industrial Engineering, Konkuk University, Seoul 05029, Republic of Korea
[2]Optimization and Analytics Office, SK Innovation, Seoul 03188, Republic of Korea
[3]Department of Management of Technology, Gyeongsang National University, Jinju 52828, Republic of Korea

Corresponding author: Chie Hoon Song (chsong01@gnu.ac.kr)

This work was supported in part by the Human Resources Program in Energy Technology of the Korea Institute of Energy Technology Evaluation and Planning (KETEP), the Ministry of Trade, Industry and Energy (MOTIE) of the Republic of Korea under Grant 20204010600220; and in part by the Framework of International Cooperation Program managed by the National Research Foundation of Korea under Grant NRF-2023K2A9A2A22000118.

ABSTRACT In the age of digital economy, customers actively share their experiences and issues about products via online product reviews. Mining potential product improvement ideas from customer needs could provide valuable insights into new functionality expected by the markets. Numerous studies have attempted to identify customer needs using these reviews, but they paid less attention to the customer's specific context in which the product was used. This study provides a novel approach for identifying customer needs based on both context information and product functions of target products. The context information and product functions are derived from online product reviews through linguistic pattern mining, whereby the customer needs are determined by the combination of extracted context information and product functions using a semantic embedding method and a clustering approach. A case study on the Amazon-Echo series was conducted to verify the applicability of the proposed approach. Consequently, we identified 1430 different customer needs, which could be used as an input for improving product design. This study is one of the first attempts to integrate context information for identifying customer needs. The proposed approach can be useful in the idea creation process for future product planning and is expected to add new empirical perspective for the e-commerce industry.

INDEX TERMS Context-awareness, customer needs, linguistic pattern, sentiment analysis, social media mining, context information.

I. INTRODUCTION

Online product reviews are an important source of information that can significantly influence customer choices and product sales [1]. They represent one of the most common ways for customers to exchange their experiences and expectations regarding a product [2]. Potential customers could reduce the risk of purchasing an unwanted product through processing of information provided in online reviews. Furthermore, online reviews include context information about a product or service that potential customers could use to evaluate the specific use cases. From the perspective of product manufacturers, the variety of information can be an expressive source of innovative ideas for identifying customer needs [3]. In this context, a profound understanding of customer needs based on online reviews is critical to providing successful innovations on the market and sustaining competitive edges. In various industries, text mining approach can be used to evaluate the impact of user reviews on business performance [4]. If companies wish to retain their existing customers and acquire more, the ability to translate latent customer needs into unique products and customer experiences is more vital than ever [5].

The associate editor coordinating the review of this manuscript and approving it for publication was Massimo Cafaro.

Various studies have shown that online product reviews can play an important role in supporting customer's purchase decisions and implementing management actions. Accordingly, various studies have extracted customer needs, such as unmet product features, customer complaints, and customer satisfactions for generating new product ideas and supporting future product planning. For example, Zhang et al. [6] have collected online product reviews of Huawei phones to define product features and provided suggestions for product feature improvement through sentiment analysis. Zhao et al. [7] have analyzed online hotel reviews for predicting customer satisfaction through the mining of technical attributes in online reviews.

Despite the contributions of previous studies, limitations may arise from insufficient consideration of context information in terms of analyzing customer needs. Typically, context information refers to any information that can be used to characterize the situation of entities (i.e. place, social situation or object) relevant to the interaction between a user and a product [8]. Green et al. [9] have noted that customer needs differ depending on the surrounding context, even if the product function remains unchanged. Accordingly, customer needs could be also considered from an alternative perspective, which involves around contextual information about product use. Ignoring the context information may make the functional improvement of a product difficult. Therefore, it is necessary to simultaneously consider the context information and product functions for identifying customer needs from a holistic perspective.

To address these challenges, this study proposes a systematic approach for identifying context-aware customer needs from online product reviews. It uses linguistic pattern mining of online product reviews to extract context information and product functions. For example, in the review "*I want to play music quietly while I sleep but the sound is too loud*", the product function "*play music*" and context information "*while I sleep*" can be extracted. The customer need can be identified with a combination of these two extracted elements. The proposed approach consists of following four steps: (1) collecting online product reviews via web scraping and selecting negative reviews based on sentiment analysis; (2) extracting context information and product function expressions using linguistic pattern mining; (3) defining context information and product function clusters based on word embedding and clustering; and (4) identifying customer needs, which are expressed as a combination of context information clusters and product function clusters. To demonstrate the validity of the proposed approach, a case study using the Amazon Echo series smart speakers was conducted.

The contributions of this study are threefold: First, this approach allows us to understand customer needs by considering both product function and context information. Therefore, latent customer needs for a product can be better understood. Second, this approach is an early attempt to extract context information and product functions from online product reviews for identifying customer needs. Hence, this study can serve as a valuable reference point for subsequent studies that aim to analyze both context information and product functions. Third, due to its data-driven and large-scale nature, the findings can assist the idea generation phase in the product improvement process.

The remainder of this paper is organized as follows. Section II includes literature review on customer needs analysis, context-aware approach, and linguistic pattern mining. Section III describes the proposed analytical framework and applied methodological approaches in detail. Section IV presents a case study for illustrating the applicability of the proposed approach. Section V discusses the applicability of the proposed approach in devising product development strategies. Finally, Section VI summarizes the major findings and concludes with an outlook on possible future research as well as theoretical and managerial contributions.

TABLE 1. Previous studies using online reviews to identify customer needs.

Author	Method	Findings
Büschken, J. et al. [15]	Sentence extraction method	Sentence-based topics related to customer experience
Lee, T.Y. et al. [16]	Algorithm for automatically eliciting product attributes	Attributes and attribute dimensions related customer needs
Netzer, O. et al. [17]	Text mining and a semantic network analysis	Customer needs related to sedan cars and diabetes drugs
Schweidel, D.A. et al. [18]	Joint model of sentiment and venue format choice	Underlying brand sentiment of the customer
Archak, N. et al. [19]	Text vectorization method	Relative preference for product features of the customer
Çallı, L. [20]	Topic modeling and The SHApley Additive exPlanations method	The 11 topics related to service quality of mobile banking service

II. LITERATURE REVIEW

A. CUSTOMER NEEDS ANALYSIS USING ONLINE REVIEWS

The ability to accurately identify and respond to customer needs is a primary concern for businesses, as customer needs form the basis for future product innovation strategy [10], [11]. Scholars have used customer complaints and customer reviews under the scope of various research disciplines, such as marketing, branding and consumer behavior [12], [13]. To better cope with customer needs, various methods are actively being investigated. Especially, the exploitation of online products reviews, which could directly transmit customers' voices about products, has received much interest [14] (Table 1).

Previous studies based on online product reviews for identifying customer needs can be broadly classified into two categories: product-feature-based and customer-sentiment-based studies. Product-feature-based studies define customer needs through product features mentioned by customers, such as physical characteristics and product specifications in online product reviews [15], [16], [17]. For example, Rai [21] have presented a method for extracting key product attributes and defining customer needs with a ranking system. The attribute terms are tagged as frequently mentioned nouns, and noun phrases were extracted using a lexical database. Based on a case study of electronics products, product properties preferred by customers, such as weight, cost, and waterproofing, could be prioritized. Furthermore, Zhan et al. [22] presented a method for summarizing online product reviews for extracting customer needs across multiple reviews. Their work has addressed the challenge of information overload that product developers face.

On the contrary, customer sentiment-based studies define customer needs through mining of customer opinions, which are expressed as emotional words or sentiment scores about product features in online product reviews [18], [19], [20]. For example, Hu and Liu [23] have presented a text summarization method for extracting opinion words from online product reviews. To this end, online product reviews on digital cameras were analyzed and grouped into positive and negative customer opinions. Customer opinions such as "horrible," "incredible," and "easy" were identified, which could contribute to quicker processing of large amount of online product reviews. Furthermore, Zhou et al. [24] have applied sentiment analysis to user-generated online product reviews and identified the number of customer needs with regard to product line planning.

Although prior studies have confirmed the suitability of online product reviews for defining customer needs, they mainly extracted the physical characteristics of the product to define customer needs. In other words, existing studies tend to not take context information into account when defining customer needs, thus demonstrating a potential limitation. Context information can reveal additional information on the applied setting of a product, such as when, where, and how a product is being used. Apart from that, the concept of product function, which represents the state of the product features, has been not applied in obtaining customer needs. For example, a previous study defined customer needs based on product features, such as battery and sound, by examining online product reviews related to a Bluetooth speaker. However, there were insufficient explanations regarding why the battery constitutes customer needs (e.g. rapid consumption in cold environments) and why the sound is customer needs (e.g. volume control). This makes it difficult to determine methods for accurately revealing customer needs. Therefore, this study proposes an approach for understanding customer needs based on the combination of context information and product function using the linguistic pattern mining of online product reviews.

B. CONTEXT-AWARE APPROACH FOR DERIVING CUSTOMER NEEDS

In addition to factual information on a specific topic, text documents (e.g. online product reviews or comments on social platforms) contain context information [25]. Context information refers to surrounding information, such as location, time, mood, and other environmental factors present within [26]. It can help explain the circumstances and settings under which a certain product is used [27]. The explicit consideration of context information can help reveal many interesting patterns that would otherwise remain hidden [28]. There was a common consensus that incorporating context information contributed to improved analysis performance for customer understanding. For example, Sarker et al. [29] collected context information such as temporal, spatial, or social information from smartphone log data. They predicted individuals' smartphone usage based on context information, and the introduced multi-dimensional context-based analysis was able to predict smartphone usage patterns more effectively. Boppana and Sandhya [30] presented a context-aware restaurant recommender model based on deep recurrent neural networks and the presented model showed an accuracy of 99.6%. Rabatel et al. [31] incorporated context information, such as age and gender of customers, for analyzing sequential patterns of purchase history. In this way, the decision makers can better adapt their strategy according to different customer types. Kavitha and Ravikumar [32] investigated context awareness in the healthcare field, and identified challenges and opportunities related to advancing context-aware applications.

However, only few studies have thus far used context information to study customer needs through social media mining. Hence, we propose a novel context-aware approach for understanding customer needs using context information derived from online product reviews. Therefore, customer needs can be identified through the combination of product function and context information, and a more specific product improvement directions can be suggested.

C. LINGUISTIC PATTERN MINING

Linguistic pattern mining of human language involves the extraction of information using the grammatical or structural features of the language [33]. Depending on the purpose, significant variations in patterns can be gleaned from the text [34]. In general, linguistic patterns are extracted by tagging parts of speech (POS), which involves labeling words in a corpus according to the context and word definition [35].

Linguistic pattern mining has been applied to a variety of research applications, such as the question answering system [36], sentiment analysis [37], and customer-aspect extraction [38]. In particular, mining of linguistic pattern has been useful in processing patent documents to perform a property-function network analysis [39]. In this regard, a property represents a specific characteristic of the patented product, and its function is the useful action of that product. The "adjective + noun" pairs are identified as properties, whereas "verb + noun" pairs are referred to as functions. For finding word pairs grammatically related to the forms "adjective + noun" or "verb + noun," Stanford typed

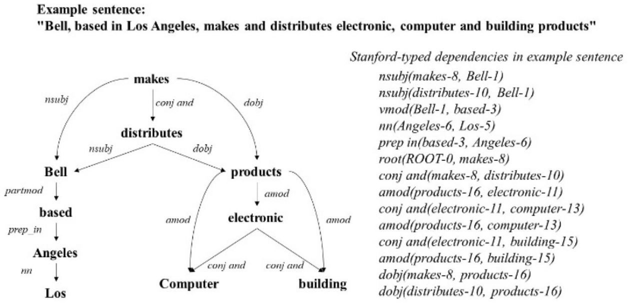

FIGURE 1. **Graphical example of Stanford typed dependencies for the sentence *"Bell, based in Los Angeles, makes and distributes electronic, computer and building products"*. (Own representation based on Yoon et al. [39].)**

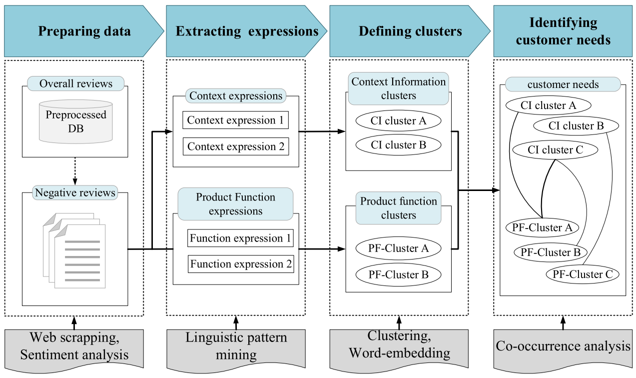

FIGURE 2. **Overview of the analysis framework.**

dependencies (SDs) can be adopted [40]. These dependencies capture grammatical relationships in a sentence and can be used by individuals who have no special linguistic expertise (Fig. 1).

For example, SDs have been applied in the domain of user experience-based product design (UX design) using online product reviews [41]. For each online product review, all possible phrases were extracted as initial candidate segments, and the structures related to the UX design were filtered out to discover customer requirements. Furthermore, linguistic pattern mining has been applied to tackle various decision-making issues, such as technology planning, detecting technological changes, and locating R&D partners [42], [43], [44].

Although linguistic patterns have been widely applied to extract various types of information included in text documents, limited attention was paid to analyzing customer needs based on contextual information and incorporating them into product improvement strategy. Hence, this study proposes a context-mining approach for discovering customer needs, which are defined as a combination of context information and product function.

III. METHODS

This section describes the overall analysis framework for discovering customer needs using context mining (Fig. 2).

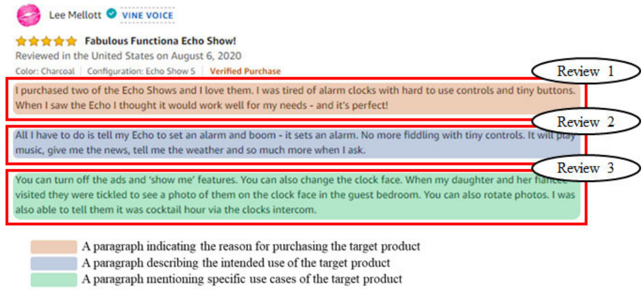

FIGURE 3. **Exemplary review paragraph structure.**

First, online product reviews were collected for the target product and negative reviews were identified using sentiment analysis. Second, context and product function expressions were extracted from the negative reviews. Third, context information and product function clusters were defined using clustering and word embedding techniques. Finally, customer needs are defined based on the combination of context information and product function. In the following, each step will be explained in a more detailed fashion.

A. DATA COLLECTION AND PRE-PROCESSING

In this step, online product reviews of a target product were collected for determining and analyzing customer needs based on context information. This can be achieved through the application of web scraping technique, such as Selenium library. In previous studies, online product reviews have been analyzed in sentence units, but this study analyzes each review paragraph as the analysis object [45], [46]. This is because paragraphs have a storyline with multiple relevant sentences explaining the subject matter in greater detail [47]. For example, the first, second, and third paragraphs in Fig. 3 refer to the rationale for purchasing, intended use, and specific use of the target product. Although an online product review was published by a single customer, each paragraph details a unique experience with different sentiment scores. For this reason, the collected reviews were divided into paragraphs and stored accordingly. Therefore, "review" in this methodology indicates an individual "paragraph" within an online product review (Fig. 3).

After structuring the reviews into paragraph units, the sentiment scores were determined to obtain negative reviews using the sentiment analysis toolkit. In this study, the Google Cloud Natural Language Processing (Google NLP) API (https://cloud.google.com/natural-language) was used. Google NLP has been used in various studies to extract sentiment scores, which range between -1 (*most negative*) and 1 (*most positive*) [48]. A negative sentiment score indicates dissatisfaction with the product. If the sentiment score was below zero, we labeled the concerned reviews as negative. Although positive online product reviews could also entail product information [49], a past study has shown that negative reviews tend to contain more detailed information on product attributes requiring improvement than positive online reviews [50]. Moreover, negative online reviews are considered more credible, altruistic, and influential than positive reviews [51], [52]. Therefore, this study considers negative online reviews to discover customer needs.

B. EXTRACTING CONTEXT EXPRESSIONS AND PRODUCT FUNCTION EXPRESSIONS

This section describes This subsection describes the extraction of context and product function expressions from online product reviews using linguistic patterns. The context expressions refer to a specific situation in which the customer uses the product, such as time, location, and social environment, whereas the product function refers to the functions provided by the target product. Because the customer behavior depends on the provided function, product functions used in this study represent both the function provided by the product itself and the behavior of customers using the product, such as "*make calls*", "*play video*", and "*search the Internet*."

Specifically, the "verb + object" represents the product functions, and the preposition can indicate the place and time associated with this behavior. Examples of prepositions are "in", "at", "on", "of" and "to". Time can also be indicated by conjunctions. For example, in following sentences "*Amazon echo show can find recipes in my kitchen*" and "*this is perfect when I'm cooking,*" the product function is understood through the verb "*find*" and its object "*recipes*". The preposition "*in*" and the noun phrase "*my kitchen*" indicate the location of product use. "*I'm cooking*" is the dependent clause for the conjunction "*when*", which reveals the timing of a product use. Accordingly, the context and product function expressions can be extracted through verbs, prepositions, and conjunctions. This study defines "preposition + phrase" and "conjunction + clause" as context expressions and "verb + phrase/clause" as product function expressions, which can be both extracted using a dependency relationship.

A dependency relationship describes a grammatical relation between a governor word and a dependent word in a sentence [53], [54]. This study relies on SDs, which were originally developed as a practical representation of English syntax. Because the online product reviews are written in English, the use of SDs is more effective than other dependency relations, such as universal dependency, attribution relation, or grammatical-relations annotation [55]. There are currently 50 pre-defined SDs. Among them, *mark*, *pcomp*, and *pobj* relationships are used to extract context expressions. *Mark* is assigned to a word that subordinates a finite clause to another clause and extracts the head of clauses with dependencies on conjunctions, such as "*while*," "*as*," and "*when*." Furthermore, *pcomp* and *pobj* are assigned to words that appear after the preposition, and they are used to extract the head of phrases dependent on prepositions, such as "*on*," "*in*," and "*at*."

To extract a product function expression, we extended the dependency types suggested by a previous study [39]. The *ccomp*, *dobj*, and *xcomp* dependency relationships are used, where *ccomp* and *xcomp* are assigned to verbs with a dependent clause, capable of extracting the verb of the dependent clause and the object of the verb. For extracting a verb and its direct object, *dobj* is assigned to a verb with a noun phrase as the object (Table 2).

TABLE 2. Different stanford typed dependencies used for extracting context and product function expressions [38].

Type	Stanford typed dependency (SD)	Description and example
Context expression	mark	The word introducing a finite clause subordinate to another clause.
		"*I wanted to be able to listen to music while taking a shower*" → (while taking a shower)
	pcomp	The head of a clause following the preposition
		"*I enjoy using this for household reminders*" → (for household reminders)
	pobj	The head of a noun phrase following the preposition
		"*I sat on the chair*" → (on the chair)
Product function expression	ccomp	The dependent clause with an internal subject whose function is similar to that of an object of the verb or adjective
		"*It's amazing that it recognizes my voice*" → (recognizes my voice)
	dobj	The noun phrase that is the (accusative) object of the verb
		"*It cannot play video from other very common sources*" → (play video)
	xcomp	The word that is an open clausal complement of the verb or adjective
		"*It suggests me a recipe*" → (suggests a recipe)

From the first example in Table 2, terms that are dependent on extracted relationships are jointly retrieved for a more specific context expression. Hence, "*while taking a shower*" is extracted using the *mark* relationship. In terms of product function expressions, "*my*", which has a dependency relationship with "*voice*", is extracted along with "*recognizes voice*" through *ccomp*. Thus, "*recognizes my voice*" is extracted (see Table 2). Algorithm 1 shows the steps involved in extracting context and product function expressions.

Through the extracted context and product function expressions, it is possible to have a deeper understanding of where customers use the product and how the product is used. However, in online product reviews, some phrases, such as "*in the home*" and "*in the house*," despite having the same meaning, are expressed differently. Therefore, the context and product function expressions with the same meaning need to be integrated through a clustering method.

Algorithm 1 Context and Product Function Expressions Extraction

Input: Review phrase, R
Output: Context expression, c, and product function expression, f, for each R
1. **Initialize** list of context = []
2. **Initialize** list of product function = []
3. **for** in R **do**
   |Check Stanford Dependency of
4. **for** word in **do**
5. Check context expression in :
   if *mark* or *pcomp* or *pobj* in,
   append word + child of words to context expression
6. Check product function expression in :
   if *ccomp* or *dobj* or *xcomp* in,
   append word + head of words to product function expression
   end for
   Reset context expression and product function expression
   end for

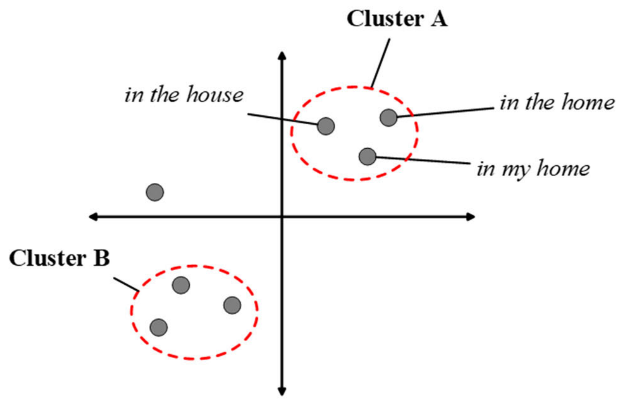

FIGURE 4. Exemplary word embedding and clustering. (Note: Although expressed in a two-dimensional space, the result of word embedding is a high-dimensional vector space with hundreds of dimensions.)

C. DEFINING CONTEXT INFORMATION CLUSTERS AND PRODUCT FUNCTION CLUSTERS

This step groups together extracted expressions with similar meanings. Product function expressions, such as "*play music*," "*play the music*," and "*play the song*," are integrated into the same cluster. For this purpose, word embedding and clustering methods are used. In word embedding, the given words are represented in a high-dimensional vector space. Similar extracted expressions are clustered together near the vector space, while dissimilar extracted expressions are placed away from each other in the vector space. For example, the extracted context expressions, such as "*in the house*," "*in the home*", and "*in my home*", when expressed in a two-dimensional space through word embedding, appear close to each other (Fig. 4).

For the word embedding, this study uses bidirectional encoder representations from transformers (BERT), an NLP technique developed by Google. The BERT model is a general-purpose model that performs well in all NLP fields [56]. The BERT model was pre-trained using a corpus of 3.3 billion words (800 million words from Book Corpus and 2.5 billion words from English Wikipedia). Unlike other language models, the BERT model is trained using word pieces, and the model can handle Out-Of-Vocabulary. Therefore, the proposed approach can be applied to other online reviews without additional training. In addition, the BERT model is suitable for integrating similar expressions because of its prominent performance and non-domain constraints [57].

The extracted context and product function expressions are represented in a vector space of 768 dimensions, respectively. Hence, it is necessary to group similar expressions using a clustering method. This study uses the k-means clustering method, which assembles expressions based on the distance between adjacent embedded expressions [58]. The k-means clustering proceeds by iteratively assigning entities (Note: Entities are context expressions and product function expressions) to clusters and updating the cluster centroids based on the mean of the assigned entities [59]. The initialization step randomly selects k data points from the dataset as the initial cluster centroids. Next, each entity in the dataset is assigned to the nearest cluster centroid based on a distance metric (e.g. Euclidean distance). This assignment is done by minimizing the variance of distance differences among clusters. Since the distance is calculated using similarity scores, the semantic similarity between entities can be obtained and the similar entities can be clustered as the same cluster. The k-means has several advantages and disadvantages. While it is an easily implementable algorithm capable of assigning every data point to a cluster and handling large datasets in a computationally efficient, it is limited in sensitivity to the initial placement of centroids, specifying the number of clusters (k) in advance and sensitivity to outliers. The optimal number of k was determined by applying a gap statistic algorithm, calculating the distance between adjacent elements in a cluster [60]. If the data are well clustered, the gap statistic value is high. Therefore, the k with the highest gap static value is chosen as the ideal k. Consequently, words that share the same meaning but have different expression forms can be considered as the same. In the next step, customer needs are defined using co-occurrence frequency of context information clusters and product function clusters.

D. IDENTIFYING CUSTOMER NEEDS

In this subsection, a combination of context information cluster and product function cluster is used to define customer needs. Customer needs are represented as a combination of each cluster appearing together in the same online product reviews (Fig. 5). For example, if the customer needs are defined through a combination of the context information cluster "*on_the_bed*" and the product function cluster "*play_a_game*", these clusters co-occur within a single online product review. Hence, the customer needs can be understood as follows: *a customer facing an issue while playing a game on the bed.*

In Fig. 5, the links between the context information cluster (CI cluster) and product function cluster (PF cluster) indicate

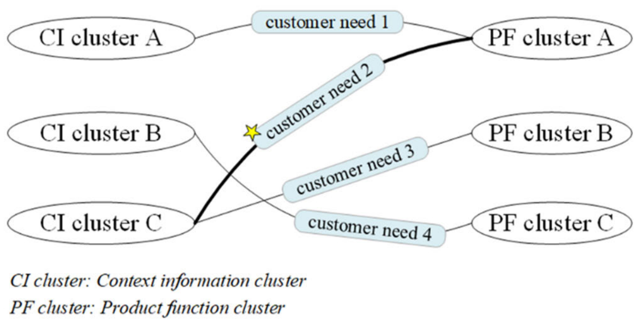

FIGURE 5. Defining customer needs using co-occurrence of context information cluster and product function cluster.

that these clusters co-occur in the same review. The link weight is proportional to the frequency of co-occurrences. Hence, the higher the co-occurrence frequency, the more important is the identified customer need for assisting the product improvement decision. The most frequently occurring customer needs have to be addressed with priority. Accordingly, "customer need 2" in Fig. 5 has a higher priority than "customer need 3", as the co-occurrence frequency of customer need 2 is more significant. Next, the effectiveness of the proposed approach is verified through a case study of the Amazon Echo series.

IV. CASE STUDY OF AMAZON ECHO

Amazon Echo is a smart speaker developed by Amazon that can connect to Alexa, a virtual assistant service. This device provides various voice-assisted services, such as playing music, setting alarms, and searching for information. With the development of artificial intelligence, the range of functions performed by Amazon Echo has become increasingly diverse. In a similar sense, customers' motives for using this speaker are becoming increasingly diversified. This implies that more attention must be paid to the analysis of underlying customer needs. Accordingly, we describe the applicability of the proposed approach to Amazon's flagship smart speaker series (Amazon Echo, Amazon Echo Show, and Amazon Echo Dot), which are collectively referred to as Amazon Echo to make suggestions for product improvement decisions.

A. COLLECTING AND PRE-PROCESSING OF ONLINE REVIEWS

A total of 11553 online product reviews, which were available at the time of conducting this study, were collected between January 1 and May 31, 2020, from the Amazon online shopping website. The collected online reviews were pre-processed by selecting reviews with one paragraph (excluding one-sentence or one-word reviews) or dividing reviews with more than two paragraphs into individual paragraph units. Hence, the pre-processed reviews refer to a set of single-paragraph derived online product reviews. For the sake of simplicity, we refer to them as "reviews" in the following case study. To determine negative reviews, sentiment scores were measured. The sentiment score was calculated using the Google NLP API. Of the 11553 reviews, 2105 reviews had a sentiment score lower than 0, and the remaining 9448 reviews had positive sentiment scores. As a result, the proposed approach was validated using 2105 negative reviews.

B. EXTRACTION OF CONTEXT AND PRODUCT FUNCTION INFORMATION

The context and product function expressions were extracted from the 2105 negative reviews using the Stanford parser. This study used Stanza (https://stanfordnlp.github.io/stanza), a Python package of the Stanford parser. Using Algorithm 1 shown in section III-B, a total of 32709 context expressions and 32346 product function expressions were obtained. However, some of the extracted expressions, such as "to this" "loves it" and "got this", were rather vague and did not cover either context information or production function. Hence, stopword processing had to be performed to remove irrelevant words.

After removal, 6747 context expressions and 4679 product function expressions remained. In Table 3, the underlined passages represent product function expression, while the bold and italicized passages describe context expressions. In case of review ID 1, only product function expressions were extracted. The expressions "*hear the music*" "*control the lights*" and "*ask prompts*" indicate different customer behavior and could be regarded as one of many product functions provided by Amazon Echo. On the contrary, in case of review ID 2, only context expressions were extracted. The context expression "*as a speaker*" highlights listening to music through speakers as one of the circumstances in which Amazon Echo device was employed.

C. GENERATION OF CONTEXT INFORMATION AND PRODUCT FUNCTION CLUSTERS

Although the product function expressions in review IDs 1 and 4, "*hear the music*" and "*playing music*" have similar meanings, they were expressed differently. Therefore, they should be consolidated and assigned to the same cluster. In this step, we first input the extracted product function expressions and context expressions to each vector space. Next, k-means clustering has been used to group similar expressions. To position the extracted context and product function expressions in individual vector spaces, word embedding was performed. Because the extracted expressions comprise multiple rather than single words, we relied on the sentence BERT model rather than the general BERT model. Using k-means clustering, similar expressions are placed close to each other, while dissimilar expressions are placed far apart in 768-dimensional vector space. The parameter k, which indicates the number of clusters in the k-means clustering algorithm, is determined using the gap statistic algorithm. The optimal choice of K is given by k for which the gap between observed data and reference data reaches a maximum. Fig. 6 shows the gap statistic values for the context and product function expressions.

TABLE 3. Examples of extracted context and product function expressions.

ID	Review	Type	Expressions
1	Plus, u can hear the music which u ask, good friend and control the lights but u have to clear ask for prompts	Product function	hear the music control the lights ask prompts
2	Bad *as a speaker* for music since there's no bass but the sound is clear and very loud	Context	as a speaker
3	I like this product a lot but when I used it *for my alarm clock* it wasn't loud enough to wake me or the birds. turned up volume the next time and it went back to halfway again. I have to use something my older Alexa	Context	for my alarm clock
		Product function	turned volume
4	I tried using it *in my dorm* and it had a hard time playing music. it continued to stop and start the songs I tried to play.	Context	in my dorm
		Product function	playing music start the songs
5	Trouble is it's not working well *with my insignia fire tv*. can't turn tv on or off, says the launcher doesn't support it when our echo dot could do it no problem, also can't control video playback as well. otherwise not bad.	Context	with my insignia fire tv
		Product function	turn tv control playback

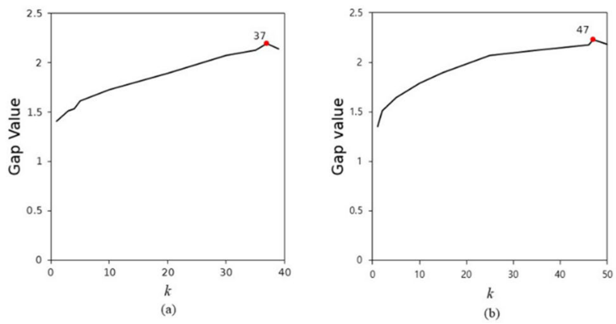

FIGURE 6. Detected optimal number of K for k-means clustering: (a) context expressions; (b) product function expressions. (Note: The red dot indicates the optimal value.)

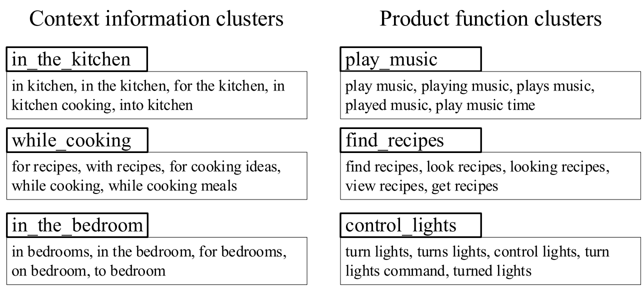

FIGURE 7. Exemplary representation of context information and product function clusters.

The highest gap statistic values were found when the k values were 37 and 47, respectively, and the cluster groups were formed accordingly.

After clustering, a unique label needs to be assigned to each cluster. The k-means clustering method groups values close to the centroid of each cluster, and these values describe the cluster composition. Therefore, each cluster was labeled using the top five expressions closest to each centroid. The entire clustering results of context and product function expressions can be found in Appendices 5 and 6. Some of the exemplary clustering results of context and product function expressions are demonstrated in Fig. 7, respectively.

In Fig. 7, The listed expressions are the top five expressions closest to the center point within each cluster. For example, context expressions such as *"in kitchen"*, *"in the kitchen"*, and *"for the kitchen"* are defined as *"in_the_kitchen"*. In product function clusters, product function expressions such as "play music", "playing music", and "plays music" are defined as *"play_music"*. In the next subsection, customer needs should be defined through a combination of 37 defined context information clusters and 47 product function clusters.

D. IDENTIFICATION OF CUSTOMER NEEDS

In the final step, the customer needs are defined through a combination of 37 context information clusters and 47 product function clusters. These combinations are calculated based on the co-occurrence analysis referred in section III-D. In total, 1739 different customer needs combinations can arise from 37 context information and 47 product function clusters. However, because the context and product function expressions do not necessarily appear together in online product reviews, only 1430 customer needs remained after calculating the co-occurrence frequency of customer needs. These identified 1430 customer needs can be interpreted as potential customer needs requiring improvement. Nevertheless, to reduce the overall complexity, this study limits the further examination to the top 10 most mentioned customer needs (Table 4).

In Table 4, some of the context information clusters represent situations customers face, such as *"while_cooking"*, *"with_tv"*, and *"as_an_intercom_system"*. Moreover, there is also context information cluster indicating the location of product usage, such as *"in_the_kitchen"*. The various product functions used in the context include *"find_recipes"*, *"connect_doorbell"*, and *"play_music"*. The identified customer needs can be used as a reference for developing product improvement strategies. For example, a customer need, which is characterized by the combination of *"with_ring_doorbell"* and *"connect_doorbell"* could indicate that the product might have a problem of connecting to third-party products (*ring_doorbell* refers to a wired doorbell device). Accordingly, a suggestion can be made that an improved product version should support compatibility with other devices. In the discussion, some specific customer needs have been selected to analyze them in more detail and propose product improvement strategies to show the advancement of the proposed approach.

TABLE 4. Examples of extracted context and product function expressions.

Rank	Co-occurrence frequency	Customer needs	
		Context information clusters	Product function clusters
1	130	while_cooking	find_recipes
2	95	with_ring_doorbell	connect_doorbell
3	76	with_tv	control_tv
4	75	in_the_kitchen	help_cooking
5	73	for_video_calls	play_videos
6	72	in_the_kitchen	find_recipes
7	66	while_cooking	help_cooking
8	65	for_an_alarm	set_alarm
9	62	with_kids	communicate_kids
10	55	as_an_intercom_system	connect_Wi-Fi

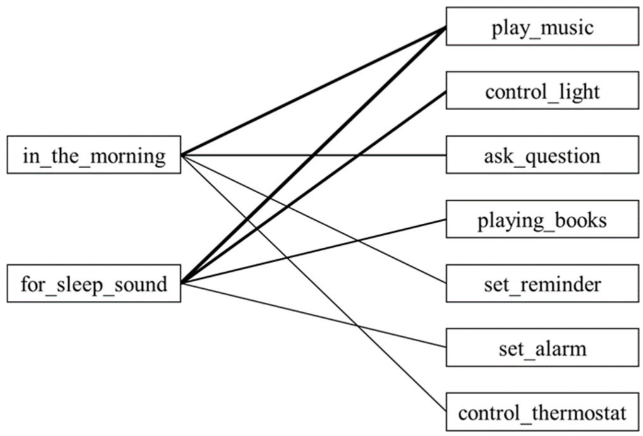

FIGURE 8. Exemplary of customer needs oriented from the context information and product function clusters.

V. DISCUSSION

A. NOVELTY OF PROPOSED CONTEXT-BASED CUSTOMER NEEDS IDENTIFICATION APPROACH

Given the nature of the customer analysis, introducing novel concepts to analyze customer needs is essential for identifying business opportunities through customers' eyes. To this end, this study proposed a novel approach for identifying customer needs based on context information and product functions. The main difference between the proposed work and prior works lies in the joint consideration of context information and product functions from online product reviews. Moreover, the current study uniquely combines linguistic pattern mining, word embedding, and sentiment analysis into a single research framework. Due to this reason, there is no comparable standard baseline approach, which can serve as an appropriate measure to evaluate the superiority of the proposed approach in a quantitative manner. Nevertheless, the inclusion of context information is significantly important, which can be backed up by the theory of Jobs-To-Be-Done (JTBD) [61]. From the JTBD, a job describes a shorthand for what a customer really seeks to accomplish in a given circumstance [62]. Here, jobs are not dependent on product functions; rather they are dependent on the context in which customers are exposed. For example, in the context of applying Amazon Echo within the kitchen, a customer's job is to prepare the meals, not staring at the display of Amazon Echo to identify recipes. Based on the proposed approach, we can infer what customers really want to accomplish in a specific context, extending the dimension of data-driven customer understanding. As the focus of this study primarily lies on the conceptual development of merging context information and product functions from online product reviews, the identified customer needs are expected to improve customer understanding.

B. APPLICABILITY OF PROPOSED CONTEXT-BASED CUSTOMER NEEDS IDENTIFICATION APPROACH

In the product development stage (e.g. product design, product planning and product specification), a clear understanding of customers is crucial to minimize the market risk [63]. Although customer uncertainty is inevitable in the marketplace, a profound understanding of customer needs can contribute to reduced uncertainty. In this perspective, the customer needs defined by the combination of the context information and product function clusters can be used to thoroughly understand the customers. In the following, we discuss the effectiveness of the proposed approach by suggesting concrete directions for potential product improvement. A closer look has been taken at some specific context-oriented customer needs. The identified customer needs can be expressed as a bipartite network composed of context information clusters on the left and product function clusters on the right (Fig. 8).

In Fig. 8, the link weight between context information and product function clusters indicates the degree of co-occurrence. In other words, the thicker the link, the more often the co-occurrence exists. For example, the context information cluster *"in_the_morning"* appears more frequently with the product function cluster *"play_music"*. This could imply that many customers are dissatisfied with listening to music in the morning. Additionally, product function cluster *"play_music"* was also mentioned most frequently with context information cluster *"for_sleep_sound"*. Therefore, the future product improvement direction could focus on the music playback function in the morning and bedtime. Next, the context information cluster *"in_the_morning"* appears more frequently with product function clusters such as *"ask_question"* and *"control_thermostat"*. This suggests that customers are dissatisfied when asked about weather and news in the morning, and they expressed dissatisfaction with the process of adjusting the ambient condition, such as a thermostat. Therefore, product planners might consider developing product functions that can check whether a consumer woke up in the morning, automatically play answers to frequently asked questions (such as weather information), or automatically adjust the ambient conditions, such as temperature and curtains.

In case of *"for_sleep_sound"*, it was frequently mentioned with product functions such as *"control_light"*,

TABLE 5. List of 37 context information clusters.

Cluster size	Label	Context expressions
482	in_every_room	in rooms, in the rooms, for each room, in each room, for every room
461	in_the_kitchen	in kitchen, in the kitchen, for the kitchen, in kitchen cooking, in to kitchen
407	in_the_household	in the household, in home office, on home, in households, for home
333	while_cooking	for recipes, with recipes, for cooking ideas, while cooking, while cooking meals
282	in_the_bedroom	in bedrooms, in the bedroom, for bedrooms, on bedroom, to bedroom
276	for_an_alarm_clock	as alarm clock, for alarm clock, with calendar, for calendar, to replace clock
270	in_the_bathroom	in the bathroom, in bathroom, for bathroom, for the bathroom, in the shower
236	with_family_members	for family, of family, with the family, in family, for family members
232	as_a_Bluetooth_speaker	as a Bluetooth speaker, of Bluetooth speakers, to the Bluetooth, for the Bluetooth speaker, with a Bluetooth speaker
220	for_video_calls	for videos, of video, to video call, on a video, for video watching
209	with_phone_apps	on the app, with app, in the app, to some apps, of the apps
209	with_kids	for kids, to children, for child, to kids, for the kids
205	in_the_garage	in the garage, in the warehouse, in garage, in house, with garage
170	in_the_morning	in the morning, at am, when I say morning, for the morning, in the am
168	on_background	in the background, with the background, of backgrounds, to have background, to backgrounds
159	for_an_alarm	for an alarms, for an alarm, to ring alarm, with alarm, for alarms
153	with_ring_doorbell	with ring doorbell, with doorbell, to use doorbell, to connect doorbell, to see door
151	for_weather_reports	for weather reports, with weather information, to know about weather, with the weather, for weather station
149	as_a_speaker	to a speaker, with the speaker, to speakers, to the speaker, as a speaker set
147	as_a_photo_album	of pictures, of photo shots, with photos, with camera, with pictures
137	with_tv	to watch TV, with TV, when watching TV, while watching TV, to watch TV
136	while_cleaning_house	while cleaning house, to clean house, while I cleaned house, to house, to the house
135	while_at_work	to work, at work, to go work, while you work, in work
134	with_smartphone	on phone, on the phone, with phone, for cellphone, with the phone
132	when_i_ask_something	for questions, to questions, to ask questions, when asked, when asked question
128	for_the_news	for news, to get news, for information, for the news, to get information

TABLE 5. *(Continued.)* List of 37 context information clusters.

121	for_nighttime	at night, of the night, for overnight, in the dark, in the evening
105	as_christmas_gifts	as a Christmas gift, for a Christmas present, for a Christmas gift, as Christmas gift, for X'mas gift
104	in_the_living_room	in living room, in the living room, for living room, on living room, for a living room
95	as_an_intercom_system	for intercom, as intercoms, as an intercom, to work intercom, of an intercom system
94	for_sleep_sounds	to go sleep, for sleep, when going sleep, while sleeping, to sleep night
94	at_any_time	for anything, with anything, for anything, at any time
91	in_the_basement	at the bottom, on the bottom, on the floor, for downstairs, in the basement
90	for_grocery_reminders	for shopping lists, on grocery list, with groceries, for grocery lists, to grocery lists
73	with_ring_devices	with ring, to ring, to get ring, with ring components, to have ring
66	with_firestick	with fire TV, on fire TV, with a fire TV, to firestick, to fire TV stick
66	with_friends	for a friend, with a friend, to a friend, of friends, to friend

"playing_books," and "set_alarm". These results could imply that customers want to use Amazon Echo as a sleep aid. Therefore, in the process of establishing a product improvement strategy, product managers can consider adding sleep assistance services, such as adjusting the light intensity through identifying customers' sleep patterns or automatically setting an alarm.

Overall, the potential improvement directions can be specifically recognized for each individual context. Through the novel approach presented in this study, it is expected that product or R&D managers are capable of identifying customer needs in more detail and reflect them in their future product planning.

VI. CONCLUSION

This study proposed a context-aware approach for the identification of customer needs from online product reviews. It differs from other studies in that it combines context information with the product function for deriving customer needs. The improved insights are particularly useful for addressing new feature development from a customer's perspective. Based on a real-life case, this study showed that the proposed approach is capable of assisting the idea generation phase in the course of product improvement. The following paragraphs pinpoint the individual contributions of this study in more detail, which are characterized by the following theoretical and managerial contributions.

With regard to its theoretical contributions, this study extends the concept of context information to the domain of customer understanding. To the best of our knowledge,

TABLE 6. List of 47 product function clusters.

Cluster size	Label	Product function expressions
441	play_music	play music, playing music, plays music, played music, play music time
402	listen_music	listening songs, listen music, listening genres, listening playlist it, listen music dot
282	search_music	reports music, shows words music, finds music playlist, playing playlist, plays music demand
227	control_lights	turn lights, turns lights, control lights, turn lights command, turned lights
189	asking_songs	asking songs, get songs, find songs, recognize songs, asked songs
165	find_recipes	find recipes, look recipes, looking recipes, view recipes, get recipes
148	ask_question	ask question, asking questions, get questions, ask question part
147	connect_device	connect devices, connect system, connect other, connecting call, connecting line
141	check_weather	check weather, find weather, ask weather, checking weather, see weather
126	make_calls	make calls, have calls, get calls, calling feature, call phone
120	show_pictures	shows pictures, see photo, see pictures, watching pictures, view cameras
120	hear_command	hear command, hear all, hear one, hear me, hear thing
115	play_videos	play videos, watch YouTube, showing videos, viewing videos, watch videos
103	set_reminder	set reminder, allows routines, check calendar, program routines, programmed day
103	ask_everything	ask everything, ask anything, ask something, ask thing, asking thing
96	help_cooking	makes menu, prepare dinner, cook recipes, helps cooking, cook book
89	control_tv	control TV, watch TV, linked TV, said TV, play TV
87	set_alarm	set alarm, have alarm, setting alarm addition, listen alarms, playing alarms
85	check_time	check time, tells me time, checking time, gets time, discover time
78	connect_doorbell	connect doorbell, connecting doorbell, get doorbell, see doorbell, works doorbell
77	answer_question	answer question, answer things, respond questions time, responds questions, asking answers
77	see_daughter	see daughter, talk daughter, seeing daughter, chat mother, call daughter
76	talk_machine	talk machine, speak English, communicate Spanish, walkie talkie device, set Echo English
76	connect_internet	connect internet, have internet, requires internet, checked internet, searching internet
73	drive_system	start car, drive you, drive system, navigating show, tuning car
66	communicate_kids	communicate kids, talk children, call kids, see kids, watch kids
66	communicate_family	communicate family, call family, see family time, make calls family, chatting family
59	listening_book	listening books, listened e-book, listen library speaker, play audiobooks, listen books tape
55	tell_joke	tells jokes, gives jokes, play jokes, provide jokes, ask jokes
55	turn_devices	turn devices, turn camera, turn screen, turn phone, turn video
55	ask_news	get news, tells news, provide me news, catching news, check news
54	control_device	control device, control all, control product, control voice, controlling devices
52	connect_Wi-Fi	connect wifi, connect wi-fi, turn wifi, joined wifi, hooked it wifi
52	stream_radio	streaming radio, turning radio channel, talk radio, replace radio, listen radio sites
50	etc.	-
50	customize_device	customize screen, customize size, adjust scree, adjust preferences, make adjustments
49	change_volume	change volume, turned volume, changing volume, select volume, set volume
48	play_movie	play movie, playing movie, shows movies, watch movie, watching movies
46	communicate_father	communicate father, receive messages dad, call dad, see dad, chat father
43	connect_spotify	play Spotify, listen Spotify, set Spotify, streaming Spotify, joining Spotify
42	request_something	request something, ask announcement, ask device, requesting station, interpret requests
40	playing_books	playing books, reading books, check library, access books, ordered book
40	communicate_grandparents	see grandparents, see grandchildren, get tune grandparents, calls grandpa, called grandma show
39	control_thermostat	control thermostat, check temperature, ask temperature, connected thermostat, work thermostat
32	ask_google	ask google, checking google, asking google, switched google, moving google
30	stream_things	stream devices, stream stations, stream phone, stream TV, stream live
13	reconnect_device	reconnect cord, reconnect camera, reconnect power, reconnect network, reconnects internet

the approach proposed in this study represents one of the first systematic methods for defining customer needs using context information in online reviews. Although previous studies extracted keywords to define customer needs based on product problems and features, they did not extract contextual information from online reviews using linguistic patterns. The results of this study can serve as a basis for forthcoming studies on mining context information from online reviews. Moreover, this study proposed dependency types for extracting context information from online product reviews. Hence, this research fills the gap in the literature on the topic of overlooking context information for product development and service quality improvement from online product reviews. The inclusion of context information extends the information value from online product reviews, making them more applicable for the customer experience management and deepening understanding of customer needs. The main strength of the proposed approach is the combination of linguistic pattern mining, natural language processing, and clustering methods into an integrated research framework. We expect that this study can pave new ways for further research related to better understanding customer needs based on mixed methods.

Accordingly, following three managerial implications have been derived from the analysis findings. (1) One of the serious problems faced by product manufacturers is how to consider the interaction of their product with the environment and behavior of their customers [64]. The proposed approach could provide a remedy to this issue, as product manufacturers can get improved insights into how, when, and where their products are used based on context information. Subsequently, product manufacturers can better understand which product functions customers perceive as uncomfortable features, and in which environment the customers may encounter difficulties. Compared with traditional product feature-based studies, the proposed approach has the advantage of better addressing customer needs. (2) Market research is an essential activity for businesses in the e-commerce industry to understand the competitors and their strategies. The proposed approach allows for the identification of product use scenarios through online product reviews, which were previously done through surveys or expert-based scenario planning. This implies that the analysis process can be more easily scaled up, allowing to generate a data-driven insight into the product complaints. Moreover, this can significantly reduce the analysis time by automating the process compared to survey-based methods, which require a higher degree of human intervention for processing information. (3) The defined customer needs using the proposed approach can help product designers in the idea generation phase. Since the proposed approach can capture the specific context in which the product is used, the product designers can leverage the use cases for future product planning.

Despite its contributions, this study is subject to some limitations. First, the proposed context-based approach could not be compared with prior studies and could not be evaluated using common performance measures, because there were no studies that explicitly rely on context information for gaining customer understanding. Therefore, additional research is required to discuss the identified context-based customer needs with the real voices of customers such as surveys, and product log data. Second, although the proposed context mining approach can be applied to other user generated contents, such as social media, blogs, and online communities, the applicability was not tested. This opens a potential future research avenue for subsequent studies. Other types of text documents might be used as an input source to make context-aware recommendation of product improvement. Finally, the linguistic patterns used in this study are based on English language grammar and cannot be applied to other languages. Considering that customer needs may vary across cultures, further research on context mining using additional language models is necessary.

APPENDIX A

See Tables 5 and 6.

REFERENCES

[1] M. M. Ali, M. B. Doumbouya, T. Louge, R. Rai, and M. H. Karray, "Ontology-based approach to extract product's design features from online customers' reviews," *Comput. Ind.*, vol. 116, Apr. 2020, Art. no. 103175.

[2] M. Zhang, B. Fan, N. Zhang, W. Wang, and W. Fan, "Mining product innovation ideas from online reviews," *Inf. Process. Manage.*, vol. 58, no. 1, Jan. 2021, Art. no. 102389.

[3] J. Qi and Z. S. Y. Zhang Jeon Zhou, "Mining customer requirements from online reviews: A product improvement perspective," *Inf. Manage.*, vol. 53, no. 8, pp. 951–963, 2016.

[4] H. Hwangbo and J. Kim, "A text mining approach for sustainable performance in the film industry," *Sustainability*, vol. 11, no. 11, p. 3207, Jun. 2019.

[5] A. Timoshenko and J. R. Hauser, "Identifying customer needs from user-generated content," *Marketing Sci.*, vol. 38, no. 1, pp. 1–20, 2019.

[6] H. Zhang, H. Rao, and J. Feng, "Product innovation based on online review data mining: A case study of Huawei phones," *Electron. Commerce Res.*, vol. 18, no. 1, pp. 3–22, 2018.

[7] Y. Zhao and X. M. Xu Wang, "Predicting overall customer satisfaction: Big data evidence from hotel online textual reviews," *Int. J. Hospitality Manage.*, vol. 76, pp. 111–121, Jan. 2019.

[8] A. Carrera-Rivera, F. Larrinaga, and G. Lasa, "Context-awareness for the design of smart-product service systems: Literature review," *Comput. Ind.*, vol. 142, Nov. 2022, Art. no. 103730.

[9] M. G. Green, J. Tan, J. S. Linsey, C. C. Seepersad, and K. L. Wood, "Effects of product usage context on consumer product preferences," in *Proc. 17th Int. Conf. Design Theory Methodol.*, vol. 5A, Jan. 2005, pp. 171–185.

[10] J. Mikulić and D. Prebežac, "A critical review of techniques for classifying quality attributes in the Kano model," *Manag. Service Quality, Int. J.*, vol. 21, no. 1, pp. 46–66, Jan. 2011.

[11] D. Kilroy, G. Healy, and S. Caton, "Using machine learning to improve lead times in the identification of emerging customer needs," *IEEE Access*, vol. 10, pp. 37774–37795, 2022.

[12] L. Çallı and F. Çallı, "Understanding airline passengers during COVID-19 outbreak to improve service quality: Topic modeling approach to complaints with latent Dirichlet allocation algorithm," *Transp. Res. Rec.*, vol. 2677, no. 4, pp. 656–673, 2023.

[13] A. Poushneh and R. Rajabi, "Can reviews predict reviewers' numerical ratings? The underlying mechanisms of customers' decisions to rate products using latent Dirichlet allocation (LDA)," *J. Consum. Marketing*, vol. 39, no. 2, pp. 230–241, Mar. 2022.

[14] E. L. Heikka, "Constructing customer knowledge in k nowledge-intensive customer relationships," *Knowl. Process Manage.*, vol. 27, no. 4, pp. 251–261, 2020.

[15] J. Büschken and G. M. Allenby, "Sentence-based text analysis for customer reviews," *Marketing Sci.*, vol. 35, no. 6, pp. 953–975, 2016.

[16] T. Y. Lee and E. T. Bradlow, "Automated marketing research using online customer reviews," *J. Marketing Res.*, vol. 48, no. 5, pp. 881–889, 2011.

[17] O. Netzer, R. Feldman, J. Goldenberg, and M. Fresko, "Mine your own business: Market-structure surveillance through text mining," *Marketing Sci.*, vol. 31, no. 3, pp. 521–543, 2012.

[18] D. A. Schweidel and W. W. Moe, "Listening in on social media: A joint model of sentiment and venue format choice," *J. Marketing Res.*, vol. 51, no. 4, pp. 387–402, 2014.

[19] N. Archak, A. Ghose, and P. G. Ipeirotis, "Deriving the pricing power of product features by mining consumer reviews," *Manage. Sci.*, vol. 57, no. 8, pp. 1485–1509, 2011.

[20] L. Çallı, "Exploring mobile banking adoption and service quality features through user-generated content: The application of a topic modeling approach to Google Play Store reviews," *Int. J. Bank Marketing*, vol. 41, no. 2, pp. 428–454, 2023.

[21] R. Rai, "Identifying key product attributes and their importance levels from online customer reviews," in *Proc. 38th Design Autom. Conf., A B*, vol. 3, Aug. 2012, pp. 533–540.

[22] J. Zhan, H. T. Loh, and Y. Liu, "Gather customer concerns from online product reviews—A text summarization approach," *Expert Syst. Appl.*, vol. 36, no. 2, pp. 2107–2115, 2009.

[23] M. Hu and B. Liu, "Mining opinion features in customer reviews," in *Proc. AAAI*, 2004, pp. 755–760.

[24] F. Zhou, J. R. Jiao, X. J. Yang, and B. Lei, "Augmenting feature model through customer preference mining by hybrid sentiment analysis," *Expert Syst. Appl.*, vol. 89, pp. 306–317, Dec. 2017.

[25] Q. Mei and C. Zhai, "A mixture model for contextual text mining," in *Proc. 12th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2006, pp. 649–655.

[26] G. M. Kapitsaki, G. Kalaitzidou, C. Mettouris, A. P. Achilleos, and G. A. Papadopoulos, "Identifying context information in datasets," in *Proc. Int. Interdiscipl. Conf. Modeling Using Context*. Cham, Switzerland: Springer, 2015, pp. 214–225.

[27] X. Liu, J. Zhang, and C. Yan, "Towards context-aware collaborative filtering by learning context-aware latent representations," *Knowl.-Based Syst.*, vol. 199, Jul. 2020, Art. no. 105988.

[28] J. M. Pérez, F. M. Luque, D. Zayat, M. Kondratzky, A. Moro, P. S. Serrati, J. Zajac, P. Miguel, N. Debandi, A. Gravano, and V. Cotik, "Assessing the impact of contextual information in hate speech detection," *IEEE Access*, vol. 11, pp. 30575–30590, 2023.

[29] I. H. Sarker, A. S. M. Kayes, and P. Watters, "Effectiveness analysis of machine learning classification models for predicting personalized context-aware smartphone usage," *J. Big Data*, vol. 6, no. 1, pp. 1–28, Dec. 2019.

[30] V. Boppana and P. Sandhya, "Web crawling based context aware recommender system using optimized deep recurrent neural network," *J. Big Data*, vol. 8, no. 1, pp. 1–24, Dec. 2021.

[31] J. Rabatel, S. Bringay, and P. Poncelet, "Mining sequential patterns: A context-aware approach," in *Advances in Knowledge Discovery and Management*. Cham, Switzerland: Springer, 2013, pp. 23–41.

[32] D. Kavitha and S. Ravikumar, "IoT and context-aware learning-based optimal neural network model for real-time health monitoring," *Trans. Emerg. Telecommun. Technol.*, vol. 32, no. 1, 2021, Art. no. e4132.

[33] U. Athira and S. M. Thampi, "Linguistic feature based filtering mechanism for recommending posts in a social networking group," *IEEE Access*, vol. 6, pp. 4470–4484, 2018.

[34] R. Schwartz, R. Reichart, and A. Rappoport, "Symmetric pattern based word embeddings for improved word similarity prediction," in *Proc. 19th Conf. Comput. Natural Lang. Learn.*, 2015, pp. 258–267.

[35] M. Shamsfard, "Lexico-syntactic and semantic patterns for extracting knowledge from Persian texts," *IJCSE Int. J. Comput. Sci. Eng.*, vol. 2, no. 6, pp. 2190–2196, 2010.

[36] G. G. Lee, "SiteQ: Engineering high performance QA system using Lexico-semantic pattern matching and shallow NLP," in *Proc. TREC*, 2001, pp. 1–10.

[37] F. Koto and M. Adriani, "The use of POS sequence for analyzing sentence pattern in Twitter sentiment analysis," in *Proc. IEEE 29th Int. Conf. Adv. Inf. Netw. Appl. Workshops*, Mar. 2015, pp. 547–551.

[38] W. Maharani, D. H. Widyantoro, and M. L. Khodra, "Aspect extraction in customer reviews using syntactic pattern," *Proc. Comput. Sci.*, vol. 59, pp. 244–253, Jan. 2015.

[39] J. Yoon and S. K. C. Kim, "Invention property-function network analysis of patents: A case of silicon-based thin film solar cells," *Scientometrics*, vol. 86, no. 3, pp. 687–703, 2011.

[40] M.-C. D. Marneffe and C. D. Manning, "Stanford typed dependencies manual," Stanford Univ., Stanford, CA, USA, Tech. Rep. Version 3.5.2, 2008.

[41] B. Yang, Y. Liu, Y. Liang, and M. Tang, "Exploiting user experience from online customer reviews for product design," *Int. J. Inf. Manage.*, vol. 46, pp. 173–186, Jun. 2019.

[42] S. Choi, H. Park, D. Kang, J. Y. Lee, and K. Kim, "An SAO-based text mining approach to building a technology tree for technology planning," *Expert Syst. Appl.*, vol. 39, no. 13, pp. 11443–11455, 2012.

[43] J. Guo, X. Wang, Q. Li, and D. Zhu, "Subject-action-object-based morphology analysis for determining the direction of technological change," *Technol. Forecasting Social Change*, vol. 105, pp. 27–40, Apr. 2016.

[44] X. Wang, Y. Fu, Y. Huang, Y. Liu, and D. Zhu, "Identifying R&D partners using SAO analysis: A case study of dye-sensitised solar cells," *Int. J. Technol. Manage.*, vol. 81, nos. 1–2, pp. 70–93, 2019.

[45] H. Wang and P. L. J. N. Yin Zheng Liu, "Sentiment classification of online reviews: Using sentence-based language model," *J. Exp. Theor. Artif. Intell.*, vol. 26, no. 1, pp. 13–31, 2014.

[46] Y. Wang, X. Lu, and Y. Tan, "Impact of product attributes on customer satisfaction: An analysis of online reviews for washing machines," *Electron. Commerce Res. Appl.*, vol. 29, pp. 1–11, May 2018.

[47] R. Jayashree, K. Srikantamurthy, and B. S. Anami, "Suitability of Naïve Bayesian methods for paragraph level text classification in the Kannada language using dimensionality reduction technique," *Int. J. Artif. Intell. Appl.*, vol. 4, no. 5, pp. 121–131, Sep. 2013.

[48] J. Hotchkiss, E. Ridderman, and W. Buftin, "Overall US hospice quality according to decedent caregivers-natural language processing and sentiment analysis of 3389 online caregiver reviews," *Amer. J. Hospice Palliative Med.*, 2023, doi: 10.1177/10499091231185593.

[49] F. Zhu and X. Zhang, "Impact of online consumer reviews on sales: The moderating role of product and consumer characteristics," *J. Marketing*, vol. 74, no. 2, pp. 133–148, 2010.

[50] N. Hu, J. Zhang, and P. A. Pavlou, "Overcoming the J-shaped distribution of product reviews," *Commun. ACM*, vol. 52, no. 10, pp. 144–147, 2009.

[51] A. Papathanassis and F. Knolle, "Exploring the adoption and processing of online holiday reviews: A grounded theory approach," *Tourism Manage.*, vol. 32, no. 2, pp. 215–224, 2011.

[52] N. Hu, T. Zhang, B. Gao, and I. Bose, "What do hotel customers complain about? Text analysis using structural topic model," *Tourism Manage.*, vol. 72, pp. 417–426, Jun. 2019.

[53] Z. W. Ras, K. A. Tarnowska, J. Kuang, L. Daniel, and D. Fowler, "User friendly NPS-based recommender system for driving business revenue," in *Proc. Int. Joint Conf. Rough Sets*. Cham, Switzerland: Springer, 2017, pp. 34–48.

[54] S. Poria, E. Cambria, G. Winterstein, and G. B. Huang, "Sentic patterns: Dependency-based rules for concept-level sentiment analysis," *Knowl.-Based Syst.*, vol. 69, pp. 45–63, Oct. 2014.

[55] M.-C. de Marneffe and C. D. Manning, "The Stanford typed dependencies representation," in *Proc. Coling, Workshop Cross-Framework Cross-Domain Parser Eval. (CrossParser)*, 2008, pp. 1–8.

[56] A. Vaswani, "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 1–11.

[57] J. Briskilal and C. N. Subalalitha, "An ensemble model for classifying idioms and literal texts using BERT and RoBERTa," *Inf. Process. Manage.*, vol. 59, no. 1, Jan. 2022, Art. no. 102756.

[58] M. Capó, A. Pérez, and J. A. Lozano, "An efficient approximation to the K-means clustering for massive data," *Knowl.-Based Syst.*, vol. 117, pp. 56–69, Feb. 2017.

[59] M. Alhawarat and M. Hegazi, "Revisiting K-means and topic modeling, a comparison study to cluster Arabic documents," *IEEE Access*, vol. 6, pp. 42740–42749, 2018.

[60] L. Mohammadi, Z. Einalou, H. Hosseinzadeh, and M. Dadgostar, "Cursor movement detection in brain-computer-interface systems using the K-means clustering method and LSVM," *J. Big Data*, vol. 8, no. 1, pp. 1–15, Dec. 2021.

[61] C. M. Christensen, T. Hall, K. Dillon, and D. S. Duncan, "Know your customers' jobs to be done," *Harvard Bus. Rev.*, vol. 94, no. 9, pp. 54–62, 2016.

[62] S. Hankammer, S. Brenk, H. Fabry, A. Nordemann, and F. T. Piller, "Towards circular business models: Identifying consumer needs based on the jobs-to-be-done theory," *J. Cleaner Prod.*, vol. 231, pp. 341–358, Sep. 2019.

[63] E. Enkel, J. Perez-Freije, and O. Gassmann, "Minimizing market risks through customer integration in new product development: Learning from bad practice," *Creativity Innov. Manage.*, vol. 14, no. 4, pp. 425–437, 2005.

[64] A. Mahboob, S. Husung, C. Weber, A. Liebal, and H. Krömker, "An approach for building product use-case scenarios in different virtual reality systems," in *Proc. 38th Comput. Inf. Eng. Conf.*, vol. 1B, Aug. 2018, Art. no. V01BT02A047.

JIHO LEE was born in Seoul, South Korea, in 1994. He received the bachelor's degree in industrial engineering and computer science from Konkuk University, Seoul, in 2019, where he is currently pursuing the Ph.D. degree in industrial engineering. His research interests include machine learning-based prediction/decision systems, computational customer analysis, computational patent analysis, and natural language processing for business intelligence.

BYEONGKI JEONG was born in Gyeonggi, South Korea, in 1993. He received the bachelor's and Ph.D. degrees in industrial engineering from Konkuk University, Seoul, South Korea, in 2017 and 2021, respectively. He is currently a Data Scientist at Optimization and Analytics Office, SK Innovation, Seoul. He is interested in the business transformation with artificial intelligence, which include prognostics and health management, advanced process control, business opportunity discovery, and product planning.

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 and Technology, Pohang, South Korea, in 2002, 2004, and 2011, respectively. He is currently a Professor with the Department of Industrial Engineering, Konkuk University, Seoul, South Korea. His research interests include technology intelligence-related topics, including technology forecasting, technology opportunity identification, technology road mapping, technology convergence, and business intelligence, including social media mining.

CHIE HOON SONG received the B.S. degree in chemistry, the M.S. degree in business chemistry, and the Ph.D. degree in business management in natural sciences from the University of Münster, Münster, Germany, in 2010, 2012, and 2015, respectively. He is currently an Assistant Professor with the Department of Management of Technology, Gyeongsang National University, Jinju, South Korea. His research interests include developing analytical frameworks for improving decision-making processes based on data mining and discovering new innovation opportunities through analysis of converging technologies and markets.