

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.DOI

Contextual Meaning-based Approach to Fine-grained Online Product Review Analysis for Product Design

KYUNGHOON PARK¹, SEYOUNG PARK², AND JUNEGAK JOUNG³

¹Department of Industrial Engineering and Management, POSETECH, Pohang 790-784 Republic of Korea (e-mail: pkh8820@postech.ac.kr)

²Department of Industrial and Enterprise Systems Engineering, University of Illinois Urbana-Champaign, Urbana, IL, 61801, USA (e-mail: seyoung7@illinois.edu)

³School of Interdisciplinary Industrial Studies, Hanyang University, Seoul 04763, Republic of Korea (e-mail: june30@hanyang.ac.kr)

Corresponding author: Junegak JounG (e-mail: june30@hanyang.ac.kr; ORCID: <https://orcid.org/0000-0003-3595-3349>)

ABSTRACT Customers share their opinions about a product through online reviews. Companies incorporate customer opinions into product design to increase customer satisfaction and market success. Many studies have attempted to analyze customer opinions on specific product features. However, these studies do not attempt to understand the actual intent of customers when mentioning a product feature in the context of a review. To overcome this limitation, this study develops a contextual meaning-based approach that considers the contextual meanings of product features in reviews. This approach enables a deeper understanding of the intent behind the target features of a product. First, a large language model-based word embedding model and clustering algorithms are introduced to divide product features into sub-features based on contextual meanings. Second, a new method is developed for creating a contextual word map to interpret the clustering results. Third, a sentiment analysis is performed to evaluate customer satisfaction for each sub-feature using BERT. A case study of a television product was conducted to demonstrate the applicability of the developed approach. The results showed that, unlike in other studies, several sub-features were identified from four targeted TV features based on their contextual meanings. Furthermore, customer satisfaction was evaluated for each sub-feature. This study is the first attempt at providing a fine-grained online product review analysis based on unsupervised learning to clarify meanings according to the context. The developed approach can be useful for determining detailed directions for improvement in the product design process and is expected to offer a new perspective.

INDEX TERMS bidirectional encoder representations from transformers (BERT), contextual meaning, customer opinion, online product review, polysemy word, product design, product feature, sentiment analysis, sub-feature identification, television (TV).

I. INTRODUCTION

An in-depth understanding of customer requirements is imperative for product design [1]. Companies continuously collect customer opinions to discern their needs and aim to incorporate them into their product designs. The adoption of a customer-centric design approach offers several benefits. First, accurately identifying and addressing customer needs and expectations can significantly enhance the likelihood of the success of a product [2]. Second, the continuous collection and analysis of customer feedback facilitate both rapid product improvements and enhanced adaptability in a constantly evolving market landscape [3]. When a product aligns closely with the actual needs and expectations of the

customers, it leads to increased customer satisfaction. This satisfaction can cascade further into positive word-of-mouth and heightened brand loyalty [4]. Finally, comprehension of customer needs can mitigate the risks associated with unnecessary design modifications or redesigns [5].

In traditional studies, customer needs for a product are determined through direct interactions with customers using methods like surveys and interviews [6]–[7]. However, these traditional methods of collecting customer opinions are time-consuming, costly, and frequently limited in scope and sample size [8]. Consequently, recent methods are shifting toward harnessing online reviews, which are abundant and easily accessible sources. Online reviews contain customers'

experiences, perceptions, and voices about products. Their utilization enables businesses to extract customer opinions and understand their needs [9]. Several product design studies have attempted to extract customer opinions from online reviews [10]–[11]. Existing studies typically focus on identifying opinion targets and analyzing the customer satisfaction associated with them. Identifying the opinion target refers to extracting the key product features or subjects that customers are interested in from their reviews about a product. Customer satisfaction regarding the opinion target is determined by sentiment analysis. For example, [12] utilized text mining techniques to extract key product features such as performance, power, fuel consumption, exterior, and safety from vehicle product reviews. [13] used deep neural networks to determine whether customer opinion targets were present in Google Glass product reviews and to analyze the sentiment associated with these targets.

Although previous studies provided valuable contributions, they had a limitation. The level of identifying customer opinion targets was not sufficiently fine-grained. In previous studies, the most fine-grained level of identifying opinion targets involved extracting keywords related to product features. However, these word-level identification can be further refined since customers often express opinions on various aspects of a single product feature. Thus, to understand the specific aspects of the product features that customers actually intend to mention, the contextual meaning information of these feature words in the review sentences should be considered. For example, when analyzing real review sentences about a television (TV) and assuming the opinion target as the product feature “screen”, many previous studies ignored the contextual meaning of “screen” in the sentences. In the three example sentences below, the term “screen” can not be further subdivided or analyzed at its various aspects level by earlier research methods. In other words, the specific aspects of the “screen” that customers intend to discuss are not considered. However, a close examination of these sentences reveals that the customer mentions of “screen” can be further refined at the feature level.

- (1) *The new TV seems to have white spots on the top of the screen.*
- (2) *I'm thinking of buying a cheap modern flat-screen TV.*
- (3) *The TV's home screen is filled with a lot of stuff that I'm never going to watch/use.*

In each sentence, “screen” is used with different contextual meanings by the customers. In (1), a white spot at the top of the screen is discussed with reference to screen quality. In (2), the flat physical feature of the screen and its price are discussed. In (3), the usability of the home screen is mentioned. As can be seen, even when discussing the same product feature, “screen”, the aspects of the “screen” that the customers intend to talk about differ slightly and need to be clearly distinguished for fine-grained analysis.

To address these limitations, this study developed a contextual meaning-based approach that identifies sub-features of a target product feature based on their contextual mean-

ings in online product reviews. The proposed methodology comprises four stages: (1) Collecting online reviews by web scraping and preprocessing the dataset using the NLTK toolkit. (2) Clustering the sub-features of the target feature based on the semantic similarities of their contextual meanings using bidirectional encoder representations from transformers (BERT)-based contextual word embedding with the affinity propagation (AP) algorithm. (3) Assigning a name to the contextual meaning of each identified sub-feature using a contextual word map. (4) Conducting sentiment analysis of each sub-feature using a fine-tuned BERT model. A case study was conducted on TV products to demonstrate the validity of the developed approach.

The main contributions of this study are threefold: First, this study proposes a fine-grained approach for opinion target identification at the sub-feature level, delving deeper than previous feature-level analyses. The proposed approach identifies aspects of product features from reviews and further categorizes them into sub-features. By doing so, customer opinions can be better understood at a more granular level. This also allows for a closer alignment with customer needs, thereby providing benefits in the product design phase. Second, the proposed approach emphasizes the importance of the contextual meaning of product feature words found within review sentences. Gaining a clear understanding of these contextual meanings provides deeper insight into the customer's authentic intent when mentioning a product feature. Therefore, this enhances the comprehension of customer's actual needs. Third, the proposed approach employs pre-trained large language models (LLMs) and unsupervised clustering techniques to identify sub-features based on their contextual meanings. The advantage of this unsupervised approach is twofold. First, it negates the necessity for labeled data, resulting in significant cost savings. Second, there is no need to predefine sub-feature classes, offering a pathway for a more customer-centric analysis of online product reviews at the sub-feature level.

The remainder of this paper is organized as follows. Section II includes a literature review of related studies on identifying targets of customer opinion based on online product reviews for product design. Section III describes the framework of the developed approach and the methods used. Section IV presents a case study of the application of the developed approach to TV products. Section V discusses the novelty and the applicability of the developed approach. Section VI presents the conclusion of this study, highlights its limitations, and suggests directions for future research.

II. LITERATURE REVIEW

Identifying the targets of opinion that customers intend to discuss and show main interest in within reviews is crucial to clearly understand customer needs [15], [31]. Therefore, many product design studies have been conducted to identify the opinion targets from online product reviews [32], [33]. In previous studies, four representative methods including term frequency-, linguistic structure analysis-, topic modeling-,

and deep learning-based methods have been employed to identify targets of customer opinion in product reviews. First, some studies employed term frequency-based techniques to extract keywords related to the product as opinion targets [12], [14], [22]. These term frequency-based techniques emphasize the prominence of words that appear more frequently in the reviews, assuming that such recurrence indicates their importance. In [14], [22], term frequency analysis in online reviews was conducted to identify keywords as opinion targets. In [12], the term frequency-inverse document frequency (TF-IDF) method was used to extract prioritized words related to product features. Second, other studies employed linguistic structure analysis-based methods [11], [17], [18], [26]. These methods emphasize the syntactic and grammatical relationships within the review text, considering the structural patterns to be indicative of the underlying meanings or sentiment. In [11], part of speech (POS) tagging was performed on review sentences to extract nouns as opinion targets. In [17], a dependency parser was employed to identify terms related to product features. In [18], linguistic pattern mining was used to extract both the product function and the specific environmental context related to it. In [26], dependency parsing, co-reference resolution, and named entity recognition were utilized to extract significant phrases from reviews. Third, several studies employed latent Dirichlet allocation (LDA)-based topic modeling methods for identifying latent topics as opinion targets from online product reviews [19], [23], [40]. These LDA-based topic modeling methods primarily focus on uncovering latent topics within a collection of review texts, allowing for a more comprehensive understanding of the underlying topics discussed in online product reviews [41]. Lastly, additional studies employed deep-learning model based methods to detect the presence of opinion targets within sentences [13]. These deep-learning-based methods primarily address classification problems and require labeled data. In [13], product reviews for Google Glass were manually labeled after being collected from Twitter. Using this labeled data, a deep neural network-based architecture was used to determine whether a review sentence expressed an opinion.

However, only a few previous studies have attempted to consider the aspects of the product feature as customer opinion targets. Although customers provide opinions on various aspects of a specific product feature, existing studies have not recognized and distinguished these aspects. Existing studies primarily extracted opinion targets at the keyword-, phrase-, and latent topic-levels. Keyword-level identification focused on extracting words related to the product from reviews [11], [12], [15], [16], [17], [20], [22], [13]. Whereas this keyword-level identification represented the most fine-grained analysis in previous studies, the analysis ended at identifying the words related to the product feature terms. The analysis did not further discern which specific aspects of each product feature were being discussed. Phrase-level identification focused on extracting expressions or phrases from review sentences that contained the customer's opinion [14], [18], [21], [24].

This phrase-level identification allowed for the most intuitive verification of customer opinions. However, it only captured the apparent information from the expressed phrases in the review. Additional analysis is needed to determine which aspects of the product itself or its features the extracted sentence refers to. Latent topic-level identification focused on uncovering hidden topics that customers discussed in the reviews [19], [23], [40]. This latent topic-level identification allowed for a relatively easy understanding of the inherent topics in large-scale reviews. However, these topics typically pertained to one of the product's features, and did not go into more specific details or further subdivide the product features. In summary, the results of identifying the target of customer opinions in previous studies, such as keywords, phrases, and latent topics, were related to product features, but they did not further identify and analyze the aspects of the product feature.

Therefore, this study proposes an approach to analyze aspects of product features. By utilizing LLMs-based contextual word embeddings, the approach identifies these aspects from reviews based on the contextual meanings of product feature words. Subsequently, the identified aspects are categorized into sub-features using clustering algorithms to group similar contextual meanings. Thus, the diverse aspects of the product feature can be captured and identified as sub-features according to the meanings used by customers in their reviews. Consequently, this provides the basis for online product review analysis at the sub-feature level.

III. METHOD

The overall process is shown in Fig. 1. Online product reviews serve as the inputs, and the outputs are sub-features of the target product feature and their sentiments. The proposed approach consists of three crucial stages after the collection of online reviews and text data preprocessing. The overall process is automated as much as possible after the target product features are determined in advance. Human intervention is only required in stage 2 to name the sub-features based on the contextual word map. The following describes the three crucial stages.

Stage 1: Product sub-feature identification:

The sub-features of the target product feature are identified based on the contextual meanings of the words for the target product feature in the review sentences. For this purpose, an LLM, specifically utilizing the BERT model, is used to perform contextual word embedding for the nouns in all review sentences. The AP algorithm is used to cluster the obtained contextual word embeddings of the target product feature words based on the similarity of their contextual meanings. The AP algorithm is an unsupervised learning method that automatically forms clusters from data without requiring both a predetermined cluster number and centroids of clusters. The product sub-features are automatically

identified after setting the hyperparameters of both BERT model and AP algorithm.

Stage 2: Naming identified product sub-feature:

To understand the contextual meanings of the identified sub-features and name the sub-features, a contextual word map is proposed to support. This map is composed of high-priority nouns that co-occur in review sentences where the representative words for each cluster appear. This map is used to interpret the contextual meaning of each cluster. Human intervention is only required to interpret the contextual meaning of the target product feature using the contextual word map and name a sub-feature (each cluster) based on the interpreted contextual meaning.

Stage 3: Sentiment analysis for each product sub-feature:

After identifying and naming the sub-features of the target product features, the fine-tuned BERT model is used to perform sentiment analysis for each sub-feature. The inputs of the model are sentences containing words within the cluster (sub-feature), and the output is the average sentiment score of the cluster (sub-feature), ranging from 1 to 5. This score indicates the degree of customer satisfaction or dissatisfaction, with 1 representing extreme dissatisfaction and 5 representing extreme satisfaction. The sentiment of a sub-feature is automatically measured after setting the hyperparameters of the fine-tuned BERT model.

A. DATA COLLECTION AND PREPROCESSING

A substantial number of online reviews pertaining to the target product were required to implement the proposed approach. These reviews were collected by web scraping from Amazon, BestBuy, and Walmart using Google Drive. The relevant information extracted from the reviews included the review title, content, product name (ASIN number), review date, and brand. Duplicate reviews are eliminated by comparing the extracted titles, dates, and product names. The review titles and contents are combined for further analysis. After collecting the reviews, standard text data preprocessing is performed [39]. The reviews are segmented into sentences using punctuation marks, such as full stops (“.”), question marks (“?”), exclamation points (“!”), and semicolons (“;”). Uppercase letters in the reviews are converted to lowercase (e.g., “Remotes” is transformed to “remotes”). Stop words, emoji, and emoticons, such as “is,” “an,” “the,” “in,” “it,” “:-),” “:-(:,” “;-),” “U+1F602,” “U+2764 U+FE0F,” and “U+1F64F” are removed and the remaining words are lemmatized (e.g., remotes is transformed to its root form remote). Consequently, the reviews are divided into sentences, with each sentence structured by lemmatized words. For example, a sentence segmented from a review like “*The Remotes are GREAT! :-)*” is transformed to “*remote great*” after the preprocessing.

The target product feature is predetermined by researchers

or practitioners prior to the initiation of the proposed method. The target feature denotes a specific attribute of the product that researchers or practitioners seek to further analyze to gain insights into customer satisfaction. The predetermined target feature is a noun because nouns typically best represent the attributes of a product or service in a product/service review analysis [11]. This target feature noun serves as a seed feature word.

B. PRODUCT SUB-FEATURE IDENTIFICATION

1) contextual word embedding

The BERT model is utilized to perform contextual word embedding among pretrained LLMs. As a pretrained deep learning model developed by Google, BERT can be utilized for advanced NLP tasks because it understands the contextual meanings of words with a high degree of accuracy. This is because of the two characteristics of BERT: bidirectionality and a transformer architecture. Unlike previously developed models that learn the context in one direction, BERT learns bidirectionally, allowing it to grasp the context more effectively. Additionally, the transformer architecture of BERT with its attention mechanism enables it to understand the relations and dependencies between words in a text [25]. Consequently, BERT can distinguish the contextual meaning of the same word based on the different sentences in which the word appears. Thus, it can generate different word embedding vectors for the same word depending on the contextual meaning. For example, in the following three sentences contain the word “screen”:

- (1) *The screen is big.*
- (2) *The screen has high resolution.*
- (3) *The screen is so clear.*

BERT can understand the contextual meaning of the word “screen” in each sentence. The contexts of sentences (1) and (2) are different. In this case, the word embedding vectors generated by BERT for “screen” in these two sentences will be different, given the dissimilar contexts of “size” versus “quality of resolution”. By contrast, the contexts of sentences (2) and (3) are similar. Therefore, the word embedding vectors generated by BERT for “screen” in these two sentences will be similar, owing to the similar contexts of “clarity” versus “quality of resolution”.

To perform contextual word embedding for target feature nouns, each sentence containing these nouns is used as an input for the BERT model. The outputs of the model are high-dimensional word embedding vectors for each word in the input sentences. Subsequently, only the word embedding vectors for the target feature nouns are extracted for further analysis. Through this process, word embedding vectors for the target feature nouns are obtained from all sentences in which they appear. In our approach, this is called contextual word embedding, which represents word embedding vectors that consider the context of the sentence. Fig. 2 shows an example of the process and result of contextual word embedding for one of the target feature nouns. The symbol denotation for process of the contextual word embedding

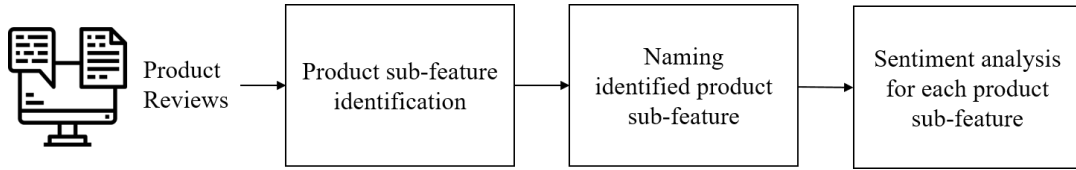


FIGURE 1. Overall process of proposed approach.

is described in Table 1. As shown in equation 1, all S_{wi}

Symbol	Description
w	a target feature noun.
j	The total number of sentences where w appears.
S_{wi}	The i -th sentence out of all the j sentences.
w_i	The noun, w from the i -th sentence out of all j sentences.
Emb_{w_i}	The word embedding vector of w_i .
$\mathcal{B}(\cdot)$	The function of the BERT model

TABLE 1. Symbol denotation

are used individually as an input for the BERT model. The outputs, represented as $\mathcal{B}(S_{wi})$, consist of word embedding vectors for every word in S_{wi} . From these outputs, only the embedding vector Emb_{w_i} for w_i is selected.

$$\text{Emb}_{w_i} = \mathcal{B}(S_{wi})[w_i] \quad (1)$$

For all j sentences of w , all embedding vectors Emb_{w_i} are obtained by:

$$\begin{aligned} &\text{Emb}_{w_1}, \text{Emb}_{w_2}, \dots, \text{Emb}_{w_j} \\ &= \mathcal{B}(S_{w1})[w_1], \mathcal{B}(S_{w2})[w_2], \dots, \mathcal{B}(S_{wj})[w_j] \end{aligned} \quad (2)$$

Each word embedding vector, Emb_{w_i} has different values depending on the context of the input sentence, S_{wi} . In the proposed method, w_i is designated as "polysemy word", which refers to words that are identical but appear in different sentences. This designation as polysemy word is made to emphasize the unique contextual characteristics determined by their respective sentences of appearance. Since the contextual meaning of w_i can vary slightly depending on the sentence, polysemy word w_i is indexed according to the sentence in which it appears. Consequently, polysemy word w_i is expressed as the same word w in each sentence, but its meaning and embedding vector may vary from one sentence to another. In the example presented in Fig. 2, the noun "screen" is assumed to be a target feature noun and appear in a total of j sentences. w corresponds to "screen", and the appearance of "screen" in the i -th sentence is indexed as the polysemy word $screen_i$ by the sentence number. The contextual word embedding vector of $screen_i$ is obtained as Emb_{screen_i} by using a BERT-based word embedding. The values of Emb_{screen_i} are expected to vary based on the contextual meaning of "screen" in the sentences where it appears.

2) Product sub-feature clustering based on contextual word embedding

Sub-features refer to the possible meanings of the target feature nouns, which are discerned based on the contextual meaning of these nouns. These sub-features are identified using the AP clustering algorithm based on the contextual word embeddings of the target feature noun, w :

$$\{\text{Emb}_{w_1}, \text{Emb}_{w_2}, \dots, \text{Emb}_{w_j}\} \quad (3)$$

These embeddings are then grouped into k distinct clusters:

$$\mathcal{C} = \{\mathcal{C}_1, \mathcal{C}_2, \dots, \mathcal{C}_k\} \quad (4)$$

Each cluster $\mathcal{C}_i$ (for $1 \leq i \leq k$) captures a distinct contextual meaning of the target product feature noun, and represents a unique sub-feature. The number of sub-features, denoted as k , is automatically determined by the AP algorithm. The AP algorithm is a centroid-based clustering algorithm that identifies clusters in data without predetermining the number of clusters and specifying the initial centroid data [27]:

Given:

- $\mathbf{x}_i$: data point i , which corresponds to a contextual word embedding vector of w_i , Emb_{w_i} in our study.
- $r(i, k)$: "responsibility" of data point $\mathbf{x}_i$ to candidate exemplar $\mathbf{x}_k$.
- $a(i, k)$: "availability" of candidate exemplar $\mathbf{x}_k$ for data point $\mathbf{x}_i$.

- 1) The "responsibility" $r(i, k)$ is updated based on accumulated evidence for $\mathbf{x}_k$ to be an exemplar for $\mathbf{x}_i$, taking into account other potential exemplars.

$$r(i, k) \leftarrow s(\mathbf{x}_i, \mathbf{x}_k) - \max_{k' \neq k} \{s(\mathbf{x}_i, \mathbf{x}_{k'}) + a(i, k')\} \quad (5)$$

Where $s(\mathbf{x}_i, \mathbf{x}_k)$ is the cosine similarity between $\mathbf{x}_i$ and $\mathbf{x}_k$.

- 2) The "availability" $a(i, k)$ is updated based on evidence that $\mathbf{x}_i$ should pick $\mathbf{x}_k$ as its exemplar, considering support from other data points.

$$a(i, k) \leftarrow \min\{0, r(k, k) + \sum_{i' \neq i, k} \max(0, r(i', k))\}, \quad (6)$$

For $i = k$:

$$a(k, k) \leftarrow \sum_{i' \neq k} \max(0, r(i', k)) \quad (7)$$

Contextual word embedding

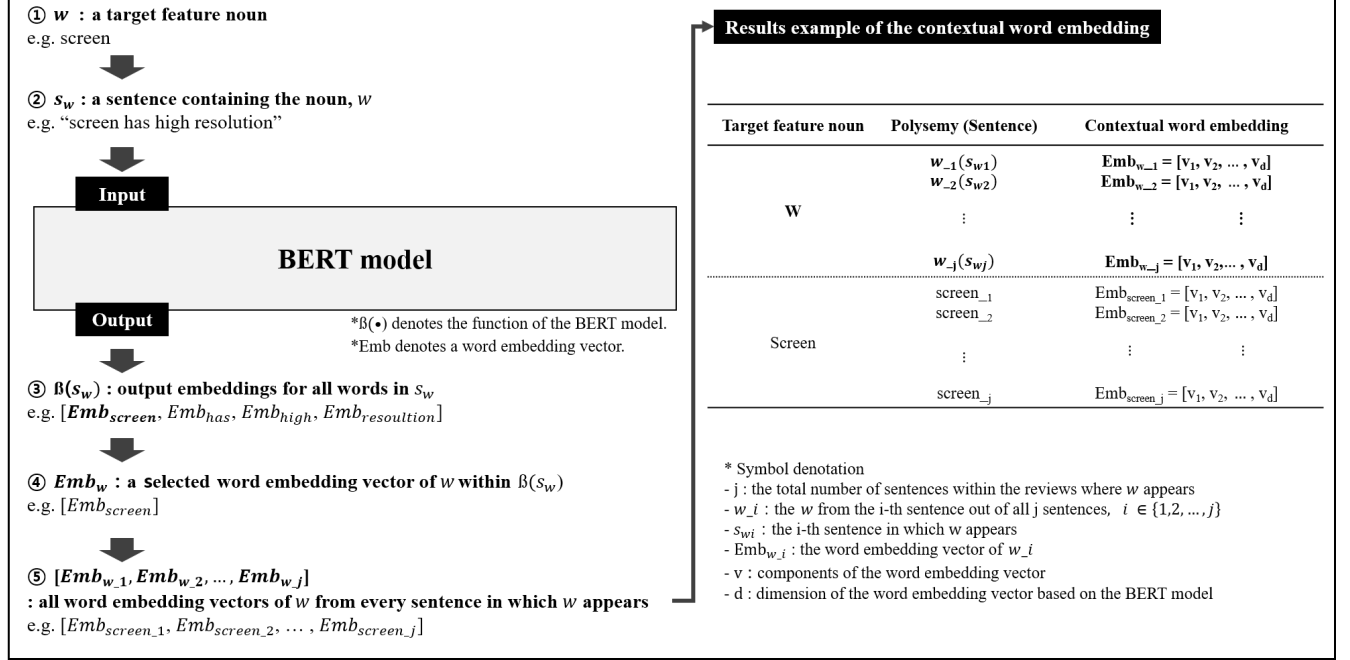


FIGURE 2. Process and result example of contextual word embedding.

- 3) When convergence is reached, exemplars can be determined where:

$$\text{Exemplar}(i) = \arg \max_k \{r(i, k) + a(i, k)\}. \quad (8)$$

An exemplar represents the central data point of a cluster, serving as a representative for other data points within the same cluster.

The Davies–Bouldin index (DBI) is utilized to evaluate clustering results. DBI evaluates clustering results based on the internal cohesion of each cluster and the discernibility between other clusters [28]. The DBI is defined as:

$$\text{DBI} = \frac{1}{N} \sum_{i=1}^N \max_{j \neq i} \left(\frac{\sigma_i + \sigma_j}{d(e_i, e_j)} \right) \quad (9)$$

where:

- N is the number of clusters.
- σ_i is the average distance of all points in cluster i to the exemplar e_i of cluster i .
- σ_j is the average distance of all points in cluster j to the exemplar e_j of cluster j .
- $d(e_i, e_j)$ is the distance between the exemplars e_i and e_j of clusters i and j .

A smaller DBI value indicates better clustering performance, which means that the clusters are cohesive internally and well separated between clusters. To evaluate the clustering quality of the seed feature, the DBI for each individual cluster is computed, and the clustering quality is evaluated by the mean of these DBI values. However, determining the quality

of clustering based strictly on the absolute value of the DBI can be challenging. This is because the DBI values can vary depending on various factors including the characteristics and distribution of the data as well as its dimensions. Therefore, a relative comparison of the clustering analysis results is necessary. This method requires identifying sub-features through an accurate understanding of the contextual meanings of the features; hence, a relatively high DBI makes it difficult to identify sub-features. Consequently, a threshold value for the DBI must be set. If the average DBI value for the seed feature exceeds this threshold, the clustering result is inferred to be unreliable, indicating that the sub-features are not well identified.

C. NAMING PRODUCT SUB-FEATURE

In this section, a contextual word map is newly proposed to name the sub-features of the seed feature. Interpreting the results and directly naming each cluster corresponding to a sub-feature is challenging. This is because the clusters are based on polysemy words, which are words with the same spelling appearing in different sentences. Thus, a contextual word map is proposed to understand the contextual meaning of each cluster and assign names to the cluster based on this contextual meaning. The construction of a contextual word map requires two steps. First, representative polysemy words within each cluster are selected. Second, the nouns within the sentences where the representative polysemy words appear are extracted and ordered according to their respective TF-IDF scores within each cluster. Based on these high TF-IDF-

scoring nouns, a contextual word map is constructed for each cluster. This map is then used as a reference to assign a descriptive name to each cluster (sub-feature). The overall process for constructing a contextual word map is shown in Fig. 3.

1) Selection of representative polysemy words within each cluster

The method for selecting representative polysemy words for each cluster utilizes the results of the AP algorithm. Since the AP algorithm is centroid-based, each cluster has a single exemplar polysemy word with the maximum cumulative cosine similarity to all other words within that cluster. However, relying solely on this exemplar to discern the contextual meaning is impractical. Therefore, within each cluster, the top- k other polysemy words with the highest cosine similarity to the exemplar are selected to provide a more comprehensive understanding of the cluster's contextual meaning and ensure a reliable interpretation. The selected words are denoted as "representative polysemy words" for each cluster. The top- k representative polysemy words for cluster $\mathcal{C}$ are obtained by:

$$\{RP_1, RP_2, \dots, RP_k\}. \quad (10)$$

Where:

- e is the exemplar of cluster $\mathcal{C}$.
- RP_i is the i^{th} representative polysemy word in cluster $\mathcal{C}$, computed as:

$$RP_i = \arg \max_{RP \in \mathcal{C} \setminus \{e, RP_1, \dots, RP_{i-1}\}} \cos(\mathbf{RP}, \mathbf{e}) \quad (11)$$

for $i = 1, \dots, k$

- $\cos(\mathbf{RP}, \mathbf{e})$ measures the cosine similarity between word embedding vector $\mathbf{RP}$ and exemplar's embedding vector $\mathbf{e}$.

To determine the value of k , the number of words related to latent topics is referenced, which is considered when evaluating results from LDA-based topic modeling [34]–[38]. In LDA-based topic modeling research, semantic coherence and discrimination between latent topics have been evaluated. This evaluation involves comparing words with a high probabilistic association with each latent topic. In most related studies, the top-10 [34]–[35], 15 [36], 20 [37], and 30 words [38] have been used for this assessment. Based on these studies, to effectively distinguish the dissimilarity or contextual meaning between clusters, the number of representative polysemy words, k , can be selected from 10, 15, 20, or 30 in our method. When the value of k is small, a few polysemy words represent a cluster and a few sentences are considered, which increases the likelihood of misinterpreting the meaning of the cluster. However, as the value of k increases, the number of representative words and sentences considered increases, making it more challenging to distinguish the contextual meanings between clusters. Therefore, setting an appropriate value for k is essential.

2) Extraction and prioritization of nouns for a contextual word map

Based on the representative polysemy words selected from the clusters using the above method, nouns are extracted from the sentences in which each representative polysemy word occurs. As previously mentioned in section III-A, nouns are POS that effectively represent the attributes of a product. Additionally, similar nouns tend to appear in similar contexts. A noun corpus is created using the nouns extracted from all clusters using the representative polysemy words. To prioritize the extracted nouns that best represent the contextual meaning of each cluster, the TF-IDF scores of the extracted nouns in each cluster are calculated using the noun corpus and the sets of the extracted nouns for each cluster:

$$\text{TF-IDF}(n, c) = \text{TF}(n, c) \times \text{IDF}(n, C) \quad (12)$$

where:

- $\text{TF}(n, c)$ is the term frequency of noun n in cluster c , defined as:

$$\text{TF}(n, c) = \frac{\text{count}(n, c)}{\text{total number of nouns in } c} \quad (13)$$

- $\text{IDF}(n, C)$ is the inverse document frequency of noun n , defined as:

$$\text{IDF}(n, c) = \log \left(\frac{|C|}{|c \in C : n \in c| + 1} \right) \quad (14)$$

- n is a specific noun.
- c is a specific cluster.
- $\text{count}(n, c)$ is the number of times n appears in c .
- C is the set of all clusters.
- $|c \in C : n \in c|$ is the number of clusters where the noun n appears. The addition of 1 ensures that there is no division by zero if n does not appear in any cluster.

Nouns with high TF-IDF scores have a low frequency of occurrence in other clusters but are frequently observed within their own clusters. Consequently, these nouns are high-priority for their respective clusters and can be considered effective in representing the clusters.

3) Construction of contextual word maps for sub-feature naming

To visually and intuitively understand the contextual meanings inherent in each cluster, a contextual word map is constructed for each cluster. This map comprises high-priority nouns that often co-occur with the target product feature word in review sentences. When the feature word carries a specific contextual meaning, the map effectively showcases this context. Consequently, the contextual meaning of each cluster can be discerned by referencing this contextual word embedding map. Based on the identified contextual meanings, names are assigned to each cluster, which represents a sub-feature.

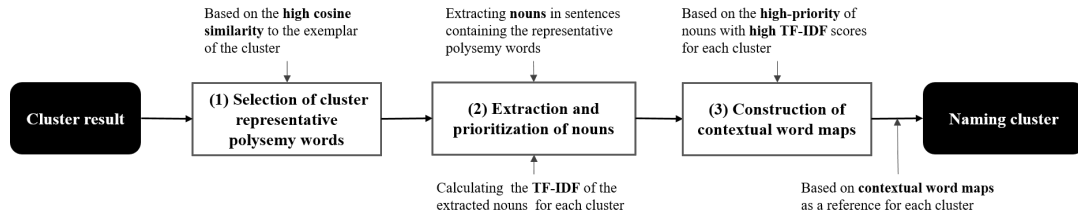


FIGURE 3. Overall process for constructing contextual word map.

D. SENTIMENT ANALYSIS FOR EACH PRODUCT SUB-FEATURE

A BERT-based sentiment analyzer is used to determine the sentiment of each product sub-feature in the reviews. Specifically, a BERT-base-multilingual-uncased-sentiment model is employed. This model is a pretrained and fine-tuned BERT model provided by nlptown from Hugging Face's Transformer Library. This model is based on the BERT-base-multilingual-uncased-sentiment model and has been fine-tuned to be directly applicable to sentiment analysis of product reviews. For English, it has been fine-tuned using 1,500,000 reviews as the training data. The underlying BERT-base-multilingual-uncased-sentiment model has a structure similar to that of the BERT model of contextual word embedding in section III-B1 and is an extended version that can be applied to various languages. The BERT-base-multilingual-uncased-sentiment model accepts sentences (tokens) as the inputs and estimates a sentiment score ranging from 1 to 5 as the output. A score closer to 1 signifies a more negative sentiment, whereas a score closer to 5 indicates a more positive sentiment.

The process of utilizing the BERT-base-multilingual-uncased-sentiment model to analyze the sentiment of a target product sub-feature is as follows: First, all sentences that contain all polysemy words belonging to each cluster are collected. These sentences are then fed into the model as the inputs. After estimating the sentiment for each sentence containing a polysemy word within the cluster by using the model, the sentiment of the sub-feature is determined based on the mean of the cluster sentiment. Consequently, the sentiment of the identified product sub-feature is assessed.

IV. CASE STUDY

A case study was conducted on TVs to demonstrate the proposed approach. The selection of TVs as the product domain was motivated by the following reasons:

- 1) Obtaining large-scale online customer reviews related to TV products is straightforward. Numerous reviews for a wide range of TV products can be easily obtained from online platforms like Amazon, BestBuy, and Walmart as mentioned in section III-A.
- 2) Terms related to TVs in reviews can be used from diverse viewpoints or contexts as the TV market evolves. TVs have transcended their traditional role of channel

content delivery to integrate with the internet, leading to the rise of smart TVs. Given these changes, customers may use identical TV-related terms in different ways or contexts. For example, "screen" was traditionally a key feature in TV reviews, predominantly in relation to picture quality. However, mentions of "screen" have expanded to include its functionality when interfacing with other devices, as TVs evolved into smart TVs. Conversely, the term "application" began appearing in reviews with the rise of smart TVs. Customers discuss "applications" from various aspects, including installation, operation, and permissions. Clarifying its various contextual usages is essential to accurately understanding the customer needs regarding "application". As evidenced by these examples, where a single TV feature term is discussed from various perspectives due to market changes, TVs were deemed suitable as the product for this study's case study.

The TV market offers a diverse range of products, varying in size, functionality, and price. Therefore, based on the most popular and commonly sold smart TV criteria, mid-range TVs with a size of 50 in or less, priced under \$800, and equipped with smart functions such as Internet connectivity and application usage were chosen. Also, as target TV features (seed features), four TV features were selected to identify customer opinions on the sub-features that interest customers in the current smart TV landscape. The TV features of screen, remote control (remote), cable, and application (app) were determined as the target product features. These represent the key components of smart TVs that have been mainly addressed from the customer perspective in previous studies targeting smart TVs [29]–[30]. In the case study, the sub-features of these four target features were identified based on their contextual meanings in review sentences. Subsequently, customer satisfaction regarding these sub-features was assessed through a sentiment analysis. Through the case study, customer opinions about multiple aspects of the TV features were obtained at the sub-feature level.

A. COLLECTING AND PREPROCESSING REVIEW DATA

Online customer reviews for TV, filtered by size, price, and functionality, were web-scraped from Amazon.com by executing Google Drive and the Python Library Selenium. The information extracted from the reviews included the review

title, content, date, brand, and product name. After removing duplicate reviews by comparing the extracted information, 16,248 reviews were obtained for 36 products from 6 brands. The release years of the products ranged from 2016 to 2022, encompassing notable advancements in smart TV technology and accommodating a typical product lifecycle. Reviews from April 2020 to March 2023 were extracted to incorporate the recent user trends and heightened demand for home entertainment systems during the COVID-19 pandemic while ensuring freshness of the data. Examining the distribution by brand, 7,086 reviews were collected for 13 Samsung products, 2,862 reviews for 4 TCL products, 2,121 reviews for 4 Hisense products, 2,098 reviews for 5 VIZIO products, 1,542 reviews for 8 LG products, and 539 reviews for 2 Sony products.

The NLTK package in Python was executed for data preprocessing, which included several steps as described in section III-A. First, reviews were divided into sentences using punctuation. All capitalized words were converted to lowercase words. The stop words, emojis, and emoticons were removed from the review sentences. Finally, lemmatization was conducted for each word in the review sentences. Consequently, the reviews were transformed into sentences composed of lemmatized words.

B. IDENTIFYING PRODUCT SUB-FEATURE

All sentences containing the four target feature nouns were extracted from the entire set of review sentences. "screen" appeared in 2,477 review sentences, "remote" in 3,802, "app" in 2,283, and "cable" in 1,043. Additionally, an index was assigned based on the sentence in which the noun appeared. These indexed nouns were "polysemy words" as mentioned in section III-B1. In case of "screen", the polysemy words included [screen_1, screen_2, ..., screen_2477], with each number denoting a sentence occurrence of "screen".

Contextual word embedding for each polysemy word was conducted using a BERT-base-uncased model, which is a version of the pretrained BERT model [25]. This model was trained using only lowercase letters and had a model size of 12 layers, 768 hidden units, 12 attention heads, and 110 million parameters. The model was pretrained using a dataset capable of processing multiple languages. For English, this model used the BookCorpus dataset, which consists of 11,038 unpublished books and English Wikipedia, excluding lists, tables, and headers. The BERT-base-uncased model took a sentence containing a polysemy word as the input and processed each word as a single token. Embedding vectors were generated for all words in a sentence; however, only the embedding vector of the polysemy word of interest was extracted, as described in equation 1.

After obtaining the contextual word embeddings, polysemy words for each target product feature with similar contexts were clustered using the AP algorithm. The hyperparameters for the AP algorithm were as follows: the damping value was set to 0.8, the maximum number of iterations was 2000, and the similarity matrix was based on

TABLE 2. Results of identifying sub-features of target product feature

Target feature	DBI	Cluster	Number of polysemy words	Exemplar polysemy word & Review sentence
Remote	3.3094	1	403	remote_246 "TV doesn't respond when I press the buttons on remote."
		2	231	remote_1089 "Cable remote is cool"
		3	3,168	remote_3234 "Remote's voice recognition doesn't work."
Screen	2.4762	1	863	screen_623 "If you are considering buying a large screen TV, believe me, you'll want to buy this one."
		2	515	screen_1515 "If the lights are turned on, they reflect like crazy, creating a rainbow effect on the screen."
		3	1,118	screen_2290 "Screen casting to TV devices has been a bad experience - it only works about 50% of the time and often takes an eternity to connect."
App	2.3926	1	2,019	app_496 "They strongly encourage you to get the app for your phone which costs \$20, but I found it to have little use especially since I can use the iPhone mirroring feature."
		2	264	app_697 "It actually instructs to cast the Xfinity app from another device and to use a Roku, which is a little disappointing."
Cable	7.2136	3		

the cosine similarity. Consequently, 3, 3, 2, and 3 clusters for screen, remote, app, and cable, respectively, were obtained automatically, confirming that each TV feature was divided into three, three, two, and three sub-features.

The mean DBI values for the clustering results of each target product feature were computed as 2.4762, 3.3094, 2.3238, and 7.2136 for screen, remote, app, and cable, respectively. The DBI value for cable demonstrated relatively higher measurements than the clustering results for the other target product features and exceeded a threshold value of 5. This finding suggested that the clustering for cable was unsuccessful, and consequently, the sub-feature for cable could not be effectively separated.

The summary of the results and the original forms of the actual review sentences in which the exemplar polysemy words appeared were presented in Table 2. Each target product feature was successfully divided into sub-features (excluding cable). However, examining the clustering results showed that each cluster consisted of polysemy words with

similar contexts appearing in different sentences. Interpreting or confirming the meanings of the sub-features or clusters based solely on polysemy words was difficult. Contextual word maps were constructed to understand the contextual meaning of these clusters in the following section IV-C.

C. ASSIGNING NAMES TO IDENTIFIED PRODUCT SUB-FEATURES

To construct the contextual word map, the top-30 representative polysemy words within each cluster that had the highest cosine similarity values to the exemplar polysemy word were selected. Subsequently, nouns were extracted from review sentences containing the selected representative polysemy words. The TF-IDF values of these nouns were calculated for each cluster. The 30 nouns with the highest TF-IDF values were prioritized and extracted for each cluster. A contextual word map was constructed based on these high-priority nouns (Fig. 4). In the map, a large font size implied a high TF-IDF score, indicating a high priority in explaining the cluster.

The contextual meaning of the identified sub-features for each target product feature was determined based on the combination of nouns in the contextual word map for each cluster. The “screen” feature was further divided into “appearance”, “resolution”, and “compatibility”. The “remote” feature was clustered into “TV remote usability”, “cable remote usability”, and “compatibility”. The “app” feature was divided into “setup” and “accessibility”. In the first cluster of “screen”, high-priority nouns like size, crack, inch, and matte had high TF-IDF values, which were interpreted as pertaining to the appearance of the screen. Thus, this cluster and sub-feature were named “Appearance”. In the second cluster, high-priority nouns like line, spot, strip, and dot had high TF-IDF values and were interpreted as being related to the quality or resolution of the screen. Therefore, this cluster and sub-feature were named “Resolution.” In the third cluster, high-priority nouns such as monitor, home, pc, phone, and smartcast had high TF-IDF values and were interpreted as pertaining to compatibility with other devices or screen operations. As such, this cluster and sub-feature were named “Compatibility”. In the first cluster of “remote”, high-priority nouns like button, volume, and sound had high TF-IDF values, which were interpreted as pertaining to the usability of the TV remote. Thus, this cluster and sub-feature were named “TV remote usability”. In the second cluster, high-priority nouns such as cable, channel, and dongle had high TF-IDF values, which were interpreted as pertaining to the usability of cable services or cable remotes. Thus, this cluster and sub-feature were named “Cable remote usability”. In the third cluster, high-priority nouns such as monitor, feature, cable, computer, and desktop had high TF-IDF values, which were interpreted as being related to the interoperability with other devices. Thus, this cluster and sub-feature were called “Compatibility”. In the first cluster of “app”, high-priority nouns like download, load, setting, and install had high TF-IDF values, which were interpreted as pertaining to

TABLE 3. Sentiment scores of all sub-features of target product features

Target feature	sub-feature	Sentiment score	Average sentiment
Screen	Appearance	2.654	2.374
	Resolution	2.1352	
	Compatibility	2.1932	
Remote	TV remote usability	3.0819	3.1756
	Cable remote usability	3.3636	
	Compatibility	3.0814	
App	Setup	3.1	3.0557
	Accessibility	3.0114	
Cable			3.319

the installation and setting up of apps on the TV. Thus, this cluster and sub-feature were named “Setup”. In the second cluster of “app”, high-priority nouns like phone, issue, access, and names of specific apps had high TF-IDF values, which were interpreted as being related to the accessibility of apps in TV. Thus, this cluster and sub-feature were named “Accessibility”.

D. DETERMINING SENTIMENTS OF PRODUCT SUB-FEATURE

In this section, the sentiment of each identified sub-feature was determined by executing the BERT-base-multilingual-sentiment model. Sentences containing the polysemy words within each cluster were used as the inputs to the model. As summarized in Table 3, the sentiment score at the feature level was calculated using the mean sentiment score of the sub-features. Among the feature levels, “screen” had a sentiment score of 2.374, implying that customer satisfaction was slightly negative. For other features such as “remote” and “app”, customer satisfaction was close to neutral. For “cable”, no divided sub-features were found. However, customer satisfaction at the cable feature level was estimated based on review sentences containing “cable” and found to be neutral. At the sub-feature level, all sub-features of “screen” had low sentiment scores. Customer satisfaction with resolution and compatibility was relatively low. These aspects should be prioritized when designing the TV product. The sub-feature “cable remote usability” of “remote” had a slightly high sentiment score, implying that customer satisfaction was relatively high. Through the case study, it was confirmed that the proposed approach effectively identifies sub-features of the target product feature using real TV review data. Additionally, customer opinions at the sub-feature level, combining both the sub-feature and its associated sentiment, were successfully extracted.

V. DISCUSSION

A. NOVELTY OF THE PROPOSED SUB-FEATURE IDENTIFICATION APPROACH

This study introduced a new approach for identifying sub-features of product features. The main distinction from prior research lies in the more fine-grained categorization that considers various aspects of a product feature. A key point of this categorization is recognizing and differentiating the

contextual meaning of aspects of product features within sentences to accurately identify sub-features. To underscore our methodology’s novelty, a qualitative comparison of the proposed approach is conducted by examining the real review sentences from section IV and contrasting these results with earlier studies. A quantitative comparison is challenging due to:

- For qualitative comparison, three earlier studies on online product review analysis are utilized [42], [33], [32]. Each study applied distinct methodologies: [42] used word2vec with a supervised CNN for TV reviews, [33] combined word2vec with the AP algorithm for Fitbit reviews, and [32] employed an LDA-based approach for smartphone reviews. Notably, these studies primarily reported results at the feature level. In the results of [42], features such as "connection", "netflix", and "brightness" were extracted for TVs. Similarly, [33] extracted features like "battery", "band", and "design" for Fitbit, and [32] identified features like "screen", "camera", and "battery" for smartphones. In contrast to previous

studies, our proposed method offers a refined categorization based on the contextual meaning of these aspects. For instance, the feature "netflix", extracted from [42], can relate to either the content or its operation on the TV according to the contexts. Similarly, the "battery" feature, extracted from both [33] and [32], can refer to either battery duration or charging speed.

For further clarification, some real review sentences from section IV in our study are considered. Regarding the product feature "screen", the following real sentences are provided:

- (1) "Poor picture quality side, unless directly front screen."
- (2) "The new TV seems to have white spots on the top of the screen. "
- (3) "I'm thinking of buying a cheap modern flat-screen TV."
- (4) "The screen is not working well by the remote."

From these, our approach identified the sub-features of the "screen" based on its contextual meaning within each sentence. Specifically, the sub-features identified were: (1) resolution, (2) resolution, (3) appearance, and (4) compatibility. However, methodologies in [42], [33], and [32] often restrict their analysis to identifying primary features without extracting finer contextual nuances. Crucially, these methodologies do not capture the aspects of "screen" nor do they leverage the contextual meaning of feature words. In [42], the analysis predominantly focuses on the mere presence or absence of pre-labeled feature words observed in reviews, such as "picture quality", "screen", and "remote". There is not a deeper exploration into the various contexts surrounding these terms. In [33], terms like "picture quality" and "flat-screen" are clustered together by the methodology of word2vec and

AP clustering used in [33]. These terms represent different aspects of "screen" but are grouped due to their lexical similarity. This approach misses the nuanced differentiation that our study offers, which is based on contextual meanings. For [32], topics incorporating the term "screen" are identified. However, this approach does not effectively differentiate between the various attributes of "screen." Attributes such as resolution, compatibility, or size tend to be conflated into a single topic due to the characteristics of LDA used in [32]. This contrasts with the granularity achieved by our method.

Through the comparison with three previous studies, the novel discriminative capability of the proposed methodology has been demonstrated. Whereas prior methods largely focus on identifying primary features, the proposed approach allows for discernment of sub-features, deeply grounded in the contextual meanings within each sentence. This approach paves the way for a closer alignment with the intentions of customers and enables a more accurate discernment of their needs.

B. APPLICABILITY OF THE PROPOSED APPROACH FOR ONLINE PRODUCT REVIEW ANALYSIS AT THE SUB-FEATURE LEVEL

This section discusses the practical application of the proposed approach. The proposed approach identifies sub-features and analyzes sentiments to discern early-stage customer needs in product design. This provides insight into the diverse aspects of primary product features from the customer's viewpoint. Additionally, this approach can provide guidance on which sub-features to focus on for technological development and resource allocation, based on customer satisfaction with product sub-features.

Moreover, the proposed methodology enables a comparative analysis of the sub-features of product features by brand. Customer satisfaction with the sub-features of product features can be identified by the brand. This allows for understanding the aspects of product features that are more or less competitive in the products of each brand. It also highlights the sub-features that should be focused on to boost overall customer satisfaction with a product. The TV reviews collected in the case study were divided by brand and analyzed using the proposed methodology. Comparative analysis was performed on relative customer satisfaction with the sub-features of the "screen," "remote," and "app" target product features in six brands, i.e., Samsung, LG, VIZIO, TCL, Sony, and Hisense. The sentiment scores for the eight sub-features across the various brands are shown in Fig. 5. For Samsung, although the sentiment score for *screen_resolution* is lower than the average, there is no significant difference. Overall, a relatively high level of satisfaction is shown across all sub-features. For LG, the sentiment scores for the two sub-features of the app feature, namely *app_setup* and *app_accessibility*, are lower than those of the other brands. In particular, *app_accessibility*, which refers to issues such as the inability to use specific apps or poor operation, has low satisfaction. VIZIO shows sentiment scores

lower than the average in terms of *screen_appearance* and *screen_compatibility*, indicating a low level of satisfaction with the design and size of the screen and its interaction with other devices (e.g., laptop connectivity). TCL has a relatively lower sentiment score for *screen_resolution* than the other brands, signifying lower satisfaction with the quality or condition of the screen. For Sony, a lower-than-average sentiment score was obtained for *TV_remote_usability*, indicating that the usability of the TV remote included with the purchase is unsatisfactory. For Hisense, sentiment scores lower than the average are obtained across all sub-features, except for *app_setup*, which has a sentiment score similar to the average. This indicates that overall satisfaction with the seven sub-features of Hisense products is low.

This study confirms the possibility of a comparative analysis of customer satisfaction across different sub-features by brand. Even within the same feature, satisfaction with a specific sub-feature may be relatively lower than that with other sub-features when compared to other brands. For example, for TCL, lower satisfaction is achieved with the *screen_resolution* aspect of the screen feature. This implies that, in future TV design processes, a greater emphasis should be placed on enhancing resolution among the various screen aspects. Consequently, assessing customer satisfaction with sub-features can assist in identifying aspects of a feature that should be enhanced to boost competitiveness and increase customer satisfaction.

VI. CONCLUSION

This paper proposes an approach for identifying and performing sentiment analysis of sub-features of predefined target product features from online product reviews. This study is the first attempt at automatically divide the sub-features of a target feature based on the contexts of the review sentence in an unsupervised manner. It employs pretrained LLM-based contextual noun-word embedding to incorporate contextual information into word embeddings. These contextual word embedding vectors are used as the basis for executing the AP algorithm, thereby clustering sub-features with similar contextual meanings. This enables the division of various inherent sub-features of a given feature from a customer perspective. Second, this study introduces a new method for constructing a contextual word map that can be used as a reference to discern the contextual meanings of sub-features. A contextual word map is created using nouns with high priority, allowing understanding and naming each cluster, or sub-feature, and its contextual meaning. Finally, a fine-tuned BERT model is used to perform sentiment analysis for each sub-feature and estimate customer satisfaction with each sub-feature.

This study has several limitations that suggest directions for future research. The proposed methodology was applied only to TV products and four target product features. Future studies can use this method to test various products and their corresponding target product features. Second, the proposed method can be expanded to address not only nouns that corre-

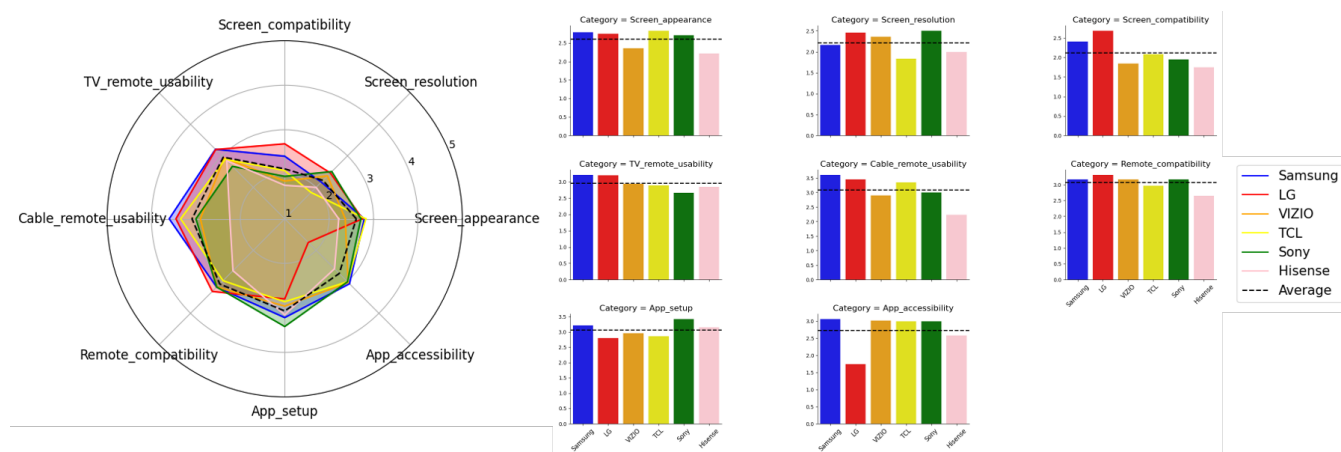


FIGURE 5. Sentiment scores of sub-features by TV brands.

spond to a target feature but also other words that have similar contexts. This expansion is needed when considering only sentences in which a noun corresponding to the target feature appears, which may limit the amount of data and potentially lead to significant information loss. Third, to maximize the automation of the entire methodology and minimize human intervention, future studies should develop a contextual word map from a reference for naming sub-features to an actual naming method. If the sub-features can be directly named using the contextual word map, the proposed methodology with become automated once the hyperparameters are determined.

ACKNOWLEDGMENTS

This research was supported by the research fund of Hanyang University (HY-202300000001721) and the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2021R111A1A01044552).

REFERENCES

- [1] Cooper, R. G., & Kleinschmidt, E. J., "New products: what separates winners from losers?," *J Prod Innov Manage*, vol. 4, no. 3, pp. 169-184, Sep. 1987, doi: 10.1016/0737-6782(87)90002-6.
- [2] Füller, J., Matzler, K., & Hoppe, M. (2008). "Brand community members as a source of innovation," *J Prod Innov Manage*, vol. 25, no. 6, pp. 608-619, Nov. 2008, doi: 10.1111/j.1540-5885.2008.00325.x.
- [3] Piller, F. T., & Walcher, D., "Toolkits for idea competitions: a novel method to integrate users in new product development," *R D Manag.*, vol. 36, no. 3, pp. 307-318, Jun. 2006, doi: 10.1111/j.1467-9310.2006.00432.x.
- [4] Gruen, T. W., Osmonbekov, T., & Czaplewski, A. J., "eWOM: The impact of customer-to-customer online know-how exchange on customer value and loyalty," *Int. J. Bus. Res.*, vol. 59, no. 4, pp. 449-456, Apr. 2006, doi: 10.1016/j.jbusres.2005.10.004.
- [5] Kahn, K. B., "Market orientation, interdepartmental integration, and product development performance," *J Prod Innov Manage*, vol. 18, no. 5, pp. 314-323, 2001, doi:10.1016/S0737-6782(01)00101-1.
- [6] Griffin, A., & Hauser, J. R., "The Voice of the Customer," *Mark. Sci.*, vol. 12 no. 1, pp. 1-27, Feb. 1993, doi: 10.1287/mksc.12.1.1.
- [7] Magnusson, P. R., "Exploring the Contributions of Involving Ordinary Users in Ideation of Technology-Based Services," *J Prod Innov Manage*, vol. 26, no. 5, pp. 578-593, Sep.2009, doi: 10.1111/j.1540-5885.2009.00684.x.
- [8] Mudambi, & Schuff, "Research Note: What Makes a Helpful Online Review? A Study of Customer Reviews on Amazon.com," *MIS Q*, vol. 34 no. 1, pp. 185, 2010, doi: 10.2307/20721420.
- [9] Xu, X., Wang, X., Li, Y., & Haghighi, M., "Business intelligence in online customer textual reviews: Understanding consumer perceptions and influential factors," *Int J Inf Manage*, vol. 37, no. 6, pp. 673-683, 2017, doi: 10.1016/j.ijinfomgt.2017.06.004.
- [10] Jin, J., Liu, Y., Ji, P., & Liu, H., "Understanding big consumer opinion data for market-driven product design," *Int. J. Prod. Res.*, vol. 54, no. 10, pp. 3019-3041, 2016, doi: 10.1080/00207543.2016.1154208.
- [11] Jiang, H., Kwong, C. K., Okudan Kremer, G. E., & Park, W.-Y., "Dynamic modelling of customer preferences for product design using DENFIS and opinion mining," *Adv. Eng. Inform.*, vol. 42, 2019, doi: 10.1016/j.aei.2019.100969.
- [12] Sun, H., Guo, W., Shao, H., & Rong, B., "Dynamical mining of ever-changing user requirements: A product design and improvement perspective," *Adv. Eng. Inform.*, vol. 46, 2020, doi: 10.1016/j.aei.2020.101117.
- [13] Gozuacik, N., Sakar, C. O., & Ozcan, S., "Social media-based opinion retrieval for product analysis using multi-task deep neural networks," *Expert Syst. Appl.*, vol. 183, 2021, doi: 10.1016/j.eswa.2021.115388.
- [14] Zhan, J., Loh, H. T., & Liu, Y., "Gather customer concerns from online product reviews - A text summarization approach," *Expert Syst. Appl.*, vol. 36, no. 2 PART 1, pp. 2107-2115, 2009, doi: 10.1016/j.eswa.2007.12.039.
- [15] Kpiebaareh, M. Y., Wu, W.-P., Agyemang, B., Haruna, C. R., & Lawrence, T., "A Generic Graph-Based Method for Flexible Aspect-Opinion Analysis of Complex Product Customer Feedback," *Appl. Sci.*, vol. 13, no. 3, 2022, doi: 10.3390/info13030118.
- [16] Ng, C. Y., & Law, K. M. Y., "Investigating consumer preferences on product designs by analyzing opinions from social networks using evidential reasoning," *Comput Ind Eng*, vol. 139, 2020, doi: 10.1016/j.cie.2019.106180.
- [17] Jin, J., Jia, D., & Chen, K., "Mining online reviews with a Kansei-integrated Kano model for innovative product design," *Int. J. Prod. Res.*, vol. 60, no. 22, pp. 6708-6727, 2022, doi: 10.1080/00207543.2021.1949641.
- [18] Lee, J., Jeong, B., Yoon, J., & Song, C. H., "Context-Aware Customer Needs Identification by Linguistic Pattern Mining Based on Online Product Reviews," *IEEE Access*, vol. 11, pp. 71859-71872, 2023, doi: 10.1109/ACCESS.2023.3295452.
- [19] Du, Y., Liu, D., & Duan, H., "A textual data-driven method to identify and prioritise user preferences based on regret/rejoicing perception for smart and connected products," *Int. J. Prod. Res.*, vol. 60, no. 13, pp. 4176-4196, 2022, doi: 10.1080/00207543.2021.2023776.
- [20] Chen, Y., Zhao, Y., Qin, B., & Liu, T., "Product aspect clustering by incorporating background knowledge for opinion mining," *PLoS One*, vol. 11, no. 8, 2016, doi: 10.1371/journal.pone.0159901.
- [21] Raghupathi, D., Yannou, B., Farel, R., & Poirson, E., "Customer sentiment appraisal from user-generated product reviews: a domain independent heuristic algorithm," *Int. J. Interact. Des. Manuf.*, vol. 9, no. 3, pp. 201-211, 2015, doi: 10.1007/s12008-015-0273-4.

- [22] Kim, H.-S., & Noh, Y., "Elicitation of design factors through big data analysis of online customer reviews for washing machines," *J. Mech. Sci. Technol.*, vol. 33, no. 6, pp. 2785–2795, 2019, doi: 10.1007/s12206-019-0525-5.
- [23] Ko, N., Jeong, B., Choi, S., & Yoon, J., "Identifying Product Opportunities Using Social Media Mining: Application of Topic Modeling and Chance Discovery Theory," *IEEE Access*, vol. 6, pp. 1680–1693, 2017, doi: 10.1109/ACCESS.2017.2780046.
- [24] Vo, A.-D., Nguyen, Q.-P., & Ock, C.-Y., "Opinion-Aspect Relations in Cognizing Customer Feelings via Reviews," *IEEE Access*, vol. 6, pp. 5415–5426, 2018, doi: 10.1109/ACCESS.2018.2797224.
- [25] Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K., "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics, Hum. Lang. Technol. (NAACL-HLT)*, vol. 1. Minneapolis, MN, USA: Association for Computational Linguistics, Jun. 2019, pp. 4171–4186, doi: 10.18653/v1/N19-1423.
- [26] Wu, X.-Y., Hong, Z.-X., Feng, Y.-X., Li, M.-D., Lou, S.-H., & Tan, J.-R., "A semantic analysis-driven customer requirements mining method for product conceptual design," *Sci. Rep.*, vol. 12, no. 1, 2022, doi: 10.1038/s41598-022-14396-3.
- [27] Frey, B. J., & Dueck, D., "Clustering by passing messages between data points," *Science*, vol. 315, no. 5814, pp. 972–976, 2007, doi: 10.1126/science.1136800.
- [28] Davies, D. L., & Bouldin, D. W., "A Cluster Separation Measure," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 1, no. 2, pp. 224–227, 1979, doi: 10.1109/TPAMI.1979.4766909.
- [29] Guo, F., Tian, X., Hu, M., & Zhang, Z., "Affective Design of Smart TV Navigation Interface Considering the Diversity of User Needs," *INT J HUM-COMPUT INT*, 2023, doi: 10.1080/10447318.2023.2179217.
- [30] Shahin, A., & Ebrahimi, S., "Revising the interrelationship matrix of house of quality by the Kano model," *TQM J.*, vol. 33, no. 4, pp. 804–822, 2020, doi: 10.1108/TQM-11-2019-0267.
- [31] Joung, J., & Kim, H. M., "Automated Keyword Filtering in Latent Dirichlet Allocation for Identifying Product Attributes From Online Reviews," *J MECH DESIGN*, vol. 143, no. 8, 2021, doi: 10.1115/1.4048960.
- [32] Joung, J., & Kim, H. M., "Approach for Importance–Performance Analysis of Product Attributes From Online Reviews," *J MECH DESIGN*, vol. 143, no. 8, 2021, doi: 10.1115/1.4049865.
- [33] Joung, J., & Kim, H. M., "Explainable neural network-based approach to Kano categorisation of product features from online reviews," *Int. J. Prod. Res.*, vol. 60, no. 23, pp. 7053–7073, 2022, doi: 10.1080/00207543.2021.2000656.
- [34] Aletras, N., & Stevenson, M., "Measuring the Similarity between Automatically Generated Topics," *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics*, vol. 2, pp. 22–27, 2014, doi: 10.3115/v1/E14-4005.
- [35] Lau, J. H., Newman, D., & Baldwin, T., "Machine Reading Tea Leaves: Automatically Evaluating Topic Coherence and Topic Model Quality," *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics*, pp. 530–539, 2014, doi: 10.3115/v1/E14-1056.
- [36] Mimno, D., Wallach, H. M., Talley, E., Leenders, M., & McCallum, A., "Optimizing semantic coherence in topic models," *EMNLP 2011 - Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference*, pp. 262–272, 2011.
- [37] Dieng, A. B., Ruiz, F. J. R., & Blei, D. M., "Topic Modeling in Embedding Spaces," *Trans. Assoc. Comput. Linguistics*, vol. 8, pp. 439–453, 2020, doi: 10.1162/tacl_a_00325.
- [38] Koltcov, S., Ignatenko, V., & Koltsova, O., "Estimating Topic Modeling Performance with Sharma–Mittal Entropy," *Entropy*, vol. 21, no. 7, pp. 660, 2019, doi: 10.3390/e21070660.
- [39] Boyd-Graber, J., Mimno, D., & Newman, D., "Care and feeding of topic models: Problems, diagnostics, and improvements," *Handbook of Mixed Membership Models and Their Applications*, 2014, doi: 10.1201/b17520.
- [40] Huang, L., Dou, Z., Hu, Y., & Huang, R., "Textual Analysis for Online Reviews: A Polymerization Topic Sentiment Model," *IEEE Access*, vol. 7, pp. 91940–91945, 2019, doi: 10.1109/ACCESS.2019.2920091.
- [41] Blei, D. M., Ng, A. Y., & Jordan, M. I., "Latent Dirichlet allocation," *J Mach Learn Res*, vol. 3, no. 4–5, pp. 993–1022, 2003.
- [42] Lee, Y., "Extraction of Competitive Factors in a Competitor Analysis Using an Explainable Neural Network," *Neural Process Lett*, vol. 53, no. 3, pp. 1979–1994, 2021, doi: 10.1007/s11063-021-10499-6.



learning, and natural language processing.

KYUNGHOON PARK was born in Ulsan, South Korea, in 1991. He received his BS degree in Life Science from Pohang University of Science and Technology (POSTECH), Pohang, South Korea, and a double major in Industrial Management and Engineering in 2015. He is currently pursuing a Ph.D. degree in Industrial Management and Engineering from POSTECH Graduate School. His research interests include customer data mining for business intelligence, machine learning, deep



SEYOUNG PARK is currently a postdoctoral research associate at the Department of Industrial and Enterprise Systems Engineering at the University of Illinois Urbana-Champaign (UIUC). She received her BS degree in Electrical Engineering from POSTECH and her MS and Ph.D. degrees in Industrial Engineering from UIUC. Her main research interests include data mining, machine learning applications, and data-driven product design.



JUNEGAK JOUNG is currently an assistant professor in the school of interdisciplinary industrial studies at Hanyang University. He received BS and Ph.D. degrees in Industrial and Management Engineering at POSTECH. His main research interests include product and service quality management and interpretable machine learning.

...