Pattern Analysis and Applications (2020) 23:1767–1777
https://doi.org/10.1007/s10044-020-00878-5

INDUSTRIAL AND COMMERCIAL APPLICATION

Extraction and prioritization of product attributes using an explainable neural network

Younghoon Lee[1] · Jungmin Park[2] · Sungzoon Cho[3]

Received: 5 January 2019 / Accepted: 30 March 2020 / Published online: 8 May 2020

Abstract
Identification of product attributes is an important matter in real-world business environments because customers generally make purchase decisions based on their evaluation of the attributes of the product. Numerous studies on product attribute extraction have been performed on the basis of user-generated textual reviews. However, most of them focused only on the attribute extraction process itself and not on the relative importance of the extracted attributes, which are critical information that can be utilized for the promotion or development of specification sheets. Thus, in this study, we focused on the development of an attribute set for a product by considering the relative importance of the extracted attributes. First, we extracted the aspects by utilizing convolutional neural network-based approaches and transfer learning. Second, we propose a novel approach, consisting of variants of the Gradient-weighted class activation mapping (Grad-CAM) algorithm, one of the explainable neural network frameworks, to capture the importance score of each extracted aspect. Using a sentimental prediction model, we calculated the weight of each aspect that affects the sentiment decision. We verified the performance of our proposed method by comparing the similarity of the product attributes that it extracted and their relative importance with the product attributes that customers consider to be the most important and by comparing the attributes used to develop the specification sheet of an existing major commercial site.

Keywords Attribute extraction · Attribute prioritization · Grad-CAM · Explainable neural network · Convolutional neural network · Transfer learning

1 Introduction

Customers generally make purchase decisions based on their evaluation and knowledge of the attributes of the product of interest [8, 29]. Thus, product developers or marketers are frequently interested in identifying the product attributes that customers consider to be most important when they evaluate and/or purchase a product [3]. The important product attributes that have been identified can then be used for product promotion. Another example is the specification sheet (Fig. 4), which is a list that describes the specifications of a product in a commercial site. By identifying the significant product attributes, marketers can effectively select the appropriate data to be included in the specification sheet.

Traditional studies on developing product attributes are costly and are heuristically performed by those who already have comprehensive domain knowledge and background information of the product or its industry. Moreover, these studies are mostly based on user surveys or focus group interviews with a few participants. Thus, they are likely

✉ Sungzoon Cho
zoon@snu.ac.kr

Younghoon Lee
yhoon.lee@seoultech.ac.kr

Jungmin Park
arena.park@lge.com

1 Department of Industrial and Systems Engineering, Seoul National University of Science and Technology, 232 Gongneung-ro, Nowon-gu, Seoul 01811, Korea

2 Idea Lab, Corporate Business Strategy Office, LG Electronics, 128 Yeouidaero, Yeongdeungpo-gu, Seoul 07336, Korea

3 Department of Industrial Engineering, Institute for Industrial Systems Innovation, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul 08826, Korea

to skew the data and their data collection process can be time-consuming.

Recently, given the increasing importance and wide availability of user-generated textual reviews, numerous studies on product attribute extraction are being performed on the basis of those reviews [27]. However, most of the previous studies on this subject focused only on the extraction of the product aspects by considering them as product attributes and not on the relative importance of the extracted aspects, which are critical information that can be utilized for the promotion or development of product specification sheets. For example, the sentence "I love the touchscreen of this, but the battery life is too short." contains two aspects [25], namely *touchscreen* and *battery life*. However, the relative importance of these aspects cannot be captured by the approaches used by the previous studies.

A recent study lightly touched on the subject of relative importance [1]; however, it only utilized an intuitive frequency-based approach. A problem with this approach is that if product attributes are organized only on the basis of their frequency, then extremely general aspects, such as *device* or *smartphone*, that are not highly relevant are ineffectively considered as the most important attributes because of their higher frequency of occurrence in textual reviews than that of other aspects.

Thus, the present study focused on the development of an attribute set for a product by considering, for the first time, the relative importance of the extracted attributes. We selected the smartphone as the target product because it is the most frequently purchased electronic device. Moreover, we utilized thousands of customer reviews collected from commercial and review sites by LG Electronics. With their understanding of relative importance, product developers or marketers should be able to determine the product attributes that customers consider to be the most important and then focus on them.

Our proposed method consists of two phases. In the first phase, we extract the aspects by utilizing convolutional neural network (CNN)-based approaches [24], which have exhibited state-of-the-art performances. Furthermore, we apply the off-the-shelf concept, one of the transfer learning approaches, to the aspect extraction to improve the performance of the method, considering in particular the latest smartphone improvements.

In the second phase, we propose a novel approach, consisting of variants of the gradient-weighted class activation mapping (Grad-CAM) algorithm, one of the explainable neural network frameworks, to capture the importance score of each extracted aspect. Using a sentimental prediction model of each textual review, we calculate the weight of each aspect that affects the sentiment decision. We found that the aspects identified to have a significant impact on the customer sentiment decision are considered to be the most important attributes by customers in their evaluation of a product. This is the first study that utilized explainable machine learning algorithms in the aspect extraction process and in the calculation of the relative importance of the extracted aspects. Furthermore, we applied the off-the-shelf feature concept again for the construction of the sentiment classification model to improve the overall efficiency of our proposed method.

We verified the performance of our proposed method in two ways. First, we compared the similarity of our extracted product attributes and their relative importance with the product attributes that customers consider as the most important purchasing factors. The dataset of the important purchasing factors was acquired from the Marketing Department of LG Electronics, and it contains the results of the customer surveys conducted internally by the company.

Second, we verified the effectiveness of our proposed method in choosing the attributes used to develop the specification sheet of an existing major commercial site. For this experiment, we conducted a survey of the satisfaction score of each extracted attribute and the overall score of the product in cooperation with LG Electronics. We then constructed a multiple regression model to examine whether the satisfaction score of the extracted attributes explained well the overall satisfaction score of the product. Subsequently, we compared the coefficients of determination of these models with an existing specification sheet of a commercial site to compare their completeness and effectiveness.

The remainder of this paper is structured as follows. Section 2 discusses the various studies on product attribute extraction, explainable machine learning, and transfer learning method. Section 3 describes the proposed method, which utilizes an aspect extraction method and an explainable neural network. Section 4 presents the data description and experimental results of the performance evaluation of our proposed method. Finally, Sect. 5 provides the conclusions and discussions as well as the directions for future work.

2 Literature review

2.1 Previous studies on product attribute extraction

The approaches used in previous works on aspect extraction can be categorized into supervised and unsupervised [1, 13, 14, 20, 21, 39]. However, our discussion here focuses on supervised approaches, which are utilized in our method. Supervised learning methods are mostly based on standard sequence labeling approaches, such as conditional random field (CRF) and hidden Markov model (HMM). Huang et al. proposed to treat product feature extraction as a sequence labeling task and employed a discriminative learning model using CRF [5]. In comparison, Choi et al. applied a hierarchical parameter sharing technique using a CRF for a fine-grained opinion analysis, combined with detection of the boundaries of opinion expressions [2]. Moreover, Yang et al. proposed a joint inference model that leverages the knowledge from predictors by optimizing the subtasks of an opinion [40]; many of the other studies were also based on HMM [7, 16, 34].

Meanwhile, Jin et al. proposed to extract highly specific product-related entities based on lexicalized HMMs [11]. Furthermore, a few domain knowledge-based methods [6, 35] have been utilized in supervised approaches. CNN-based approaches [15, 24] have recently been suggested, and they have shown state-of-the-art performance compared to those used in the previous studies. The authors of those studies utilized Amazon embeddings for word representation and constructed a seven-layer CNN architecture. The present study basically utilizes this CNN structure in the first phase and introduces variations to address the limitations of the approaches used in the previous studies by considering the rapidly changing environment of the smartphone industry.

However, these previous studies mostly focused only on the extraction of aspects and not on the relative importance of the extracted aspects. Although a recent study lightly touched on the subject of the relative importance of the extracted aspects [1], it was intuitively based on the frequency of each aspect in the textual review. Thus, in thus study, we focused on deriving the relative importance of the extracted aspects by utilizing an explainable neural network.

2.2 Explainable algorithms

An explainable algorithm concept is proposed to explain how machine learning algorithms arrive at a specific decision in contrast with the black-box characteristic of existing machine learning algorithms. Especially in the image processing field, there are a number of studies that explain or define the context of an image by utilizing matrix factorization [18], collaborative embedding [19], or metric learning [17].

In this study, we introduced variations to the Grad-CAM algorithm [30], which is one of the explainable machine learning algorithms derived from image classification methods, to calculate the relative importance of the extracted aspects. Grad-CAM uses the gradient information that flows into the last convolutional layer of a CNN to understand the importance of each neuron for a decision of interest (Fig. 1). Similarly, we constructed a sentiment classification model that utilizes the concept of the Grad-CAM algorithm to capture the importance of each aspect.

Additionally, we utilized other explainable machine learning algorithms as the baselines for comparison with our proposed method in the experiments. These algorithms included a sequence model based on attention mechanism [36] and local interpretable model-agnostic explanations (LIME) [28]. The attention mechanism allows a decoder to consider different parts of a source sentence at each step of the output generation. Then, the model learns how to generate a context vector for each output time step and what to focus on based on the input sentence and what it has produced [38]. Moreover, LIME is an algorithm that can authentically explain the prediction of any classifier by approximating it locally with an interpretable model.

2.3 Transfer learning

In both phases of our proposed method, we utilize the transfer learning concept to improve its performance and overall efficiency in handling the latest released products. The aim of transfer learning is to learn an objective predictive function for a target task using not only the target domain but also other source domains and source tasks [12]. In a classification task in one domain of interest, we only have sufficient training data of another domain of interest, which may be in a different feature space or follow a different data distribution [33]. In such cases, transfer learning, if performed successfully, would significantly improve the performance of the learning by avoiding expensive data-labeling efforts [23]. Herein, we utilized the off-the-shelf feature approach, one of the known transfer learning methods. In this approach, we use the

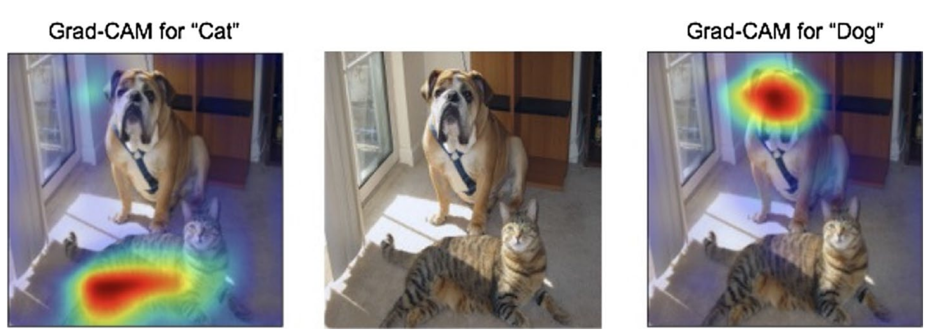

Fig. 1 Examples of Grad-CAM images

outputs of one or more layers of a network trained on a different task as generic feature detectors and train a new shallow model on the basis of these features for the target data [4, 32].

3 Method

As mentioned previous, our proposed method is composed of two phases: (1) attribute extraction by using a CNN-based approach and transfer learning, and (2) calculation of the relative importance of the extracted attributes by applying the variants of the Grad-CAM algorithm with a sentiment classification model. Additionally, we performed minor refinements such as attribute clustering (Fig. 2).

3.1 Extraction of product aspects

For the first phase, we utilized a CNN approach, which is a state-of-the-art supervised approach, to extract the attributes, following the study of Poria et al [24]. We also used another useful approach, namely transfer learning, to capture the attributes describing the latest improvements to a smartphone, which is one of the most rapidly changing products and whose specification data easily become outdated.

We first embedded all the customer reviews in a 300-dimensional vector space before the CNN model was constructed by utilizing the word2vec architecture [22]. Amazon and smartphone review datasets collected from LG Electronics were used for the word embedding task. We also added linguistic features to further improve the performance of the proposed method. Most product evaluation terms are either nouns or groups of nouns. We therefore utilized parts-of-speech (POS) tags. Specifically, we used six basic POS features of the Stanford tagger: noun, verb, adjective, adverb, preposition, and conjunction. These were encoded as a six-dimensional binary vector. Thus, for each word, the final feature vector was 306-dimensional.

We constructed and trained the CNN (Fig. 2) after the word embedding tasks using the existing datasets of SemEval 2014 [31] and Qui et al. [26]. We inputted each word with a window size of 5 into the CNN because the features of an aspect term depended on its context words.

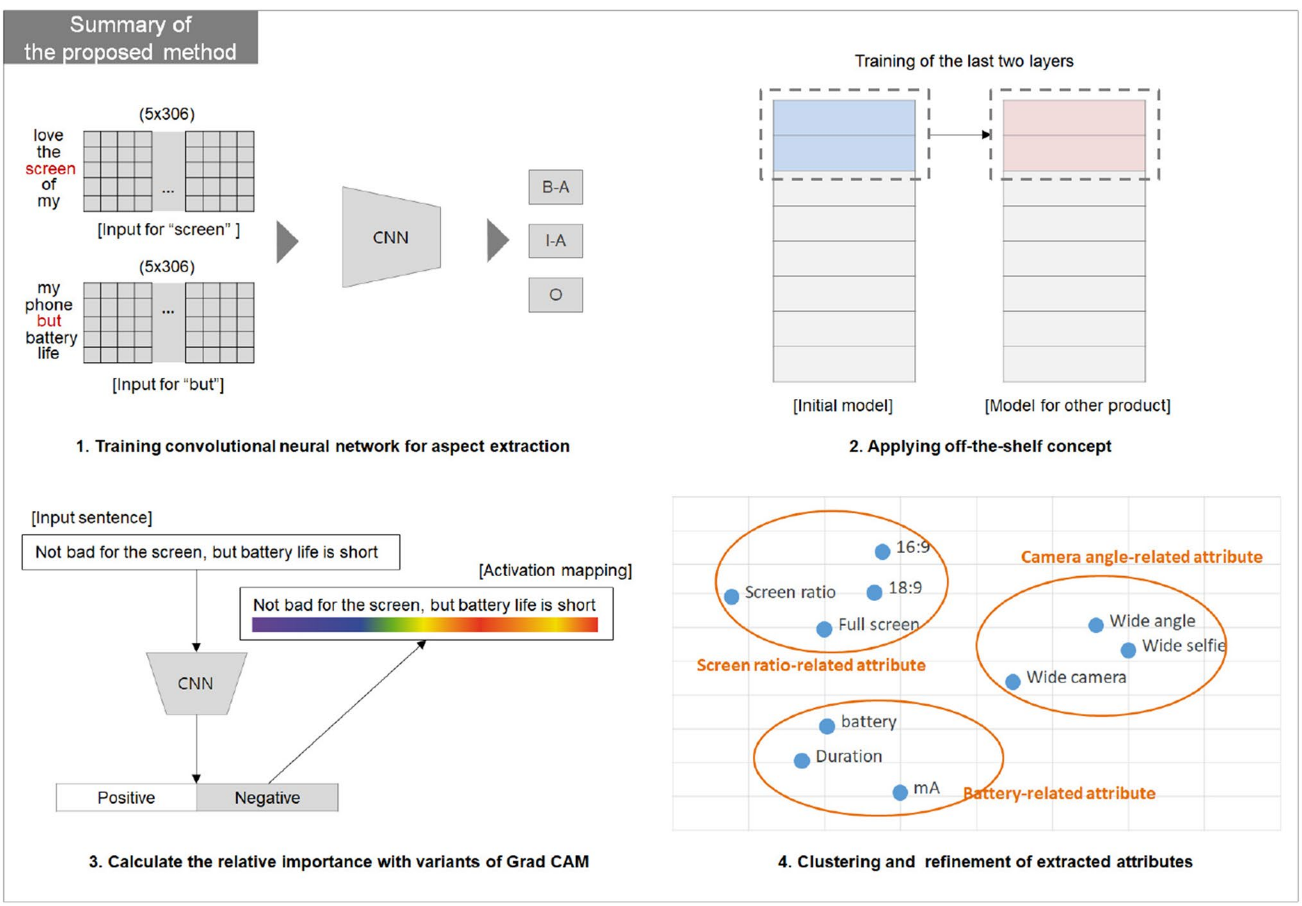

Fig. 2 Summary of our proposed method

Table 1 Structure of the convolutional neural network for aspect extraction

Layer	Number of feature maps	Size of the filter
First layer	100	3×3
Second layer	50	$2 \times 2 \times 100$
Third layer	25	$2 \times 2 \times 50$

The network contained one input layer, three convolutional layers, three max-pooling layers, and two fully connected layers with a softmax output. The convolutional layers were constructed as described in Table 1, and the stride in each convolutional layer was 1 because we wanted to tag each word.

The pool size we used in the max-pooling layer was 2×2. The output of each convolutional layer was computed using a hyperbolic tangent. The other parameters of the CNN were based on those used in a previous study [24]. Additionally, we used regularization with dropout on the penultimate layer, with the L2-norm constraint of the weight vectors having 50 epochs.

We applied the off-the-shelf feature concept after training the basic convolutional network. We maintained the weights of the convolutional layer of a previous model and retrained only the last two fully connected layers with respect to each product, such as V10, G5, V20, G6, and V30. The dataset used to train the CNN and the off-the-shelf approaches is described in the Experiments section. We then extracted the attribute keyword from the entire review dataset of each product with the trained model. Our smartphone review dataset contained 1000 reviews between September 23, 2014, and July 23, 2018; the aspect keywords in this dataset were labeled by domain experts from the Mobile Communication Department of LG Electronics.

All the abovementioned datasets were labeled using a widely used coding scheme for representing sequences. In this example, the first word of each aspect starts with a *B-A* tag. The *I-A* tag denotes the continuation of the aspect, whereas *O* is used to tag a word that is not an aspect.

3.2 Calculation of the relative importance of the extracted aspects

Without the prioritization of the extracted aspects, as mentioned previously, there would be numerous limitations in their use. Moreover, a simple prioritization approach based on simple frequencies causes a bias in which extremely general aspects are considered as the most important aspects. Thus, we provide a novel approach to calculate the relative importance of the extracted aspects using the variants of the Grad-CAM algorithm.

Fig. 3 Example of weight visualization

We assumed that the aspect that has a significant impact on the overall product sentiment also has relatively more importance than the other attributes. Thus, we utilized the weight of each aspect that affects the overall product sentiment as the importance score of each product attribute.

First, we constructed a sentiment classification model by utilizing a CNN. To improve the overall efficiency of our proposed method, we reused part of the aspect extraction model described in the previous section for the sentiment classification model. We retained the parameters of the filter used in each convolution layer and only retrained the final two layers for the sentiment classification model.

Second, we added the weighted layer like in the case with Grad-CAM to calculate the weight of each aspect that affects the sentiment decision, as shown in Fig. 3.

In detail, our proposed architecture is based on a fundamental assumption that the relative importance weight Y^c for a particular class c can be written as a linear combination of its global average-pooled last convolutional layer feature maps A^k, as in the following equation:

$$Y^c = \sum_k w_c^k \sum_i A_i^k$$

and class c is equal to $w_k^c = \partial Y^c / \partial A_i^k$ where i is a sequential location of a word in a sentence. However, this formulation makes the weights w_c^k independent of the positions i of a particular activation map A^k. Thus, we worked around this limitation by taking the global average pool of the partial derivatives, as in the following equation, which is the same as in the original Grad-CAM approach:

$$w_k^c = \sum_i \frac{\partial Y^c}{\partial A_i^k}$$

Furthermore, we added up the weights of all the aspects for a complete text review to understand the importance of each aspect. Additionally, the weights of the aspects in each review text were normalized by dividing the sum of the weights of all the aspects to remove the bias caused by the different lengths of the textual reviews.

We then sorted the attributes by the order of the importance score to reveal the relative importance of each attribute. We easily selected a relevant attribute from the limited number of attributes by sorting them.

3.3 Evaluation factor clustering and refinement

Furthermore, we conducted additional minor refinements to achieve a better performance. Our observations of the extracted attribution factors showed that there were many typographical errors, incorrect expressions, or different words that meant the same because the user review data were extremely unstructured texts. Thus, herein, we applied a clustering technique to assign synonymous words standing for an extracted attribution factor in the same cluster.

We clustered the words on the basis of the embedding vector of the extracted factors calculated in the first step using the spherical k-means method [41] to make the silhouette index the lowest. Cosine dissimilarity $1 - \cos(x, y)$ is the distance measure used in the spherical k-means method. The proposed method is affected by the number of clusters. Hence, several values were tested, with the number of clusters varying from 100 to 500 in increments of 50. Our experiment results showed that there were no significant improvements when the number of clusters exceeded 200 and, therefore, we set the number of cluster to 300 in this study.

If a cluster included an aspect that appeared more than 20 times in a textual review, all aspects were assigned to the same cluster. Moreover, if a cluster included more than two aspects appearing more than 20 times, the aspects were assigned to their respective cluster. Otherwise, all aspects in a cluster were ignored in our consideration. The proposed method is affected by the value of the threshold for frequency. Thus, several values were tested, with the threshold varying from 10 to 50 in increments of 1. Our experiment results showed that there were no significant improvements when the number of clusters exceeded 15.

In the clustering result, aspects such as *'screen ratio'*, *'16:9'*, *'18:9,'* and *'full screen'* were assigned to the same cluster. Table 2 provides examples of the extracted keywords belonging to the same cluster.

We converted the indirect expression of a user in a review comment into an appropriate wording representing each attribute after the clustering task. For instance, we converted *'fast'* and *'speed'* into *'Processor'* and *'Clearance'*, respectively, and *'Screen color'* into *'display type'* and *'glass type.'* The appropriate wording refers to aspects in the specification sheet of a commercial site.

Table 2 Examples of keywords in the same cluster

Attribute	Synonymous extracted keywords
Screen ratio	*screen ratio*, *16:9*, *18:9*, *full vision*
Design	*design*, *look*, *LG Signature*, *appearance*
OS version	*OS*, *N OS*, *Nougat*, *Android*
User interface	*User interface*, *UX*, *UX4.0*, *GUI*

4 Experiments

4.1 Experiment setup

We verified the performance of our proposed method with two experiments. In the first one, we calculated the similarity between the attributes of our prioritized product and the results of a real survey conducted internally by LG Electronics to identify the product attributes considered by real customers as the most important purchasing factors.

The survey results, utilized as an answer set, consisted of product attributes ordered by the importance of the purchasing factors considered as the most significant by costumers. To compare the order of the product attributes in the answer set and in the result of our proposed method, we measured the results with a normalized discounted cumulative gain (NDCG), which is one of the most well-known evaluation measures in information retrieval for ranking systems [9, 10]. NDCG allows each retrieved result to have a graded relevance, whereas most traditional ranking measures only allow a binary relevance. In addition, it associates a discount function with a rank, whereas many other measures uniformly weigh all the positions [37].

In detail, at a particular rank position p, the naive cumulative gain (CG), which is the predecessor of the discounted cumulative gain (DCG) and does not include the position of a result in the consideration of the usefulness of a result set, is defined as $\mathrm{CG}_p = \sum_{i=1}^{p} rel_i$. Moreover, DCG at a particular rank position p is defined as $\mathrm{DCG}_p = rel_1 + \sum_{i=2}^{p} rel_i/\log_2(i+1)$, which indicates that highly relevant documents appearing lower in a search result list should be penalized because the graded relevance value is reduced logarithmically proportional to the position of the result.

Comparing a search engine's performance from one query to the next cannot be consistently achieved using the DCG alone, and, therefore the CG at each position for a chosen value of p should be normalized across queries. This is done by sorting all relevant documents in the corpus by their relative relevance. Thus, NDCG is computed as $\mathrm{NDCG}_p = \mathrm{DCG}_p/\mathrm{IDCG}_p$, where $\mathrm{IDCG}_p = \sum_{i=1}^{|\mathrm{REL}_p|}(2^{rel_i} - 1)/\log_2(i+1)$ and REL_p represents the list of relevant documents in the text up to position p.

We measured the NDCG value of the top 30 extracted attributes and then compared the results with those of other baselines, as presented in Table 3. We assigned a relevance weight on a scale from 1 to 10 per three attributes based on the discussion with the domain experts at LG Electronics, and the weights were reduced from 1.0 to near 0.0 using a logarithm function. For instance, the attributes in the answer set were assigned the values of [10, 10, 10, 9, 9, 9, 8, 8,...,

Table 3 Baselines utilized in the first experiment

No.	Extraction method	Prioritization method
1	CNN-based + Transfer learning [15]	LSTM attention
2	CNN-based + Transfer learning [15]	LIME [28]
3	CNN-based [24]	LSTM attention
4	CNN-based [24]	LIME [28]
5	Dlirec [34]	LSTM attention
6	Dlirec [34]	LIME [28]
7	Hierarchical CRF [5]	LSTM attention
8	Hierarchical CRF [5]	LIME [28]
9	Hierarchical CRF [5]	Frequency-based approach
10	Lexicalized HMM [11]	LSTM attention
11	Lexicalized HMM [11]	LIME [28]
12	Lexicalized HMM [11]	Frequency-based approach

2, 1, 1, 1] and the weights were reduced to [1.0, 1.0, 1.0, 0.6309, 0.6309, 0.6309, 0.5, 0.5, 0.5, 0.4307, ...]. Obviously, the NDCG value increases when the largest number is listed first.

In the second experiment, we compared the effectiveness of our proposed method with those of existing major commercial sites on the development of specification sheets. In detail, we verified that the proposed method can extract the specialized factor of LG V30. The extraction of the specialized factor of each product is one of the most important considerations in the smartphone industry, which is the most rapidly changing industry.

For the experiment, we conducted two five-point Likert scale-based user surveys with 40 participants and used the following points: (1) Level of influence each attribute in the specification sheet exerts on their purchase intention of LG V30 and (2) satisfaction of the overall product and of each attribute in the specification sheet for LG V30. We then constructed a multiple regression model of the satisfaction of the overall product and of each attribute. Subsequently, we compared the coefficients of determination (R^2). The regression model demonstrated the completeness of the composition of the attribute set in the specification sheet.

In detail, $\bar{y}$ is the mean of the observed data: $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$. Then, the variability of the dataset can be measured using three sums-of-squares formulas: (1) The total sum of squares (proportional to the variance of the data): $SS_{tot} = \sum_i (y_i - \bar{y})^2$. (2) The regression sum of squares, also called the explained sum of squares: $SS_{reg} = \sum_i (f_i - \bar{y})^2$. (3) The sum of squares of the residuals, also called the residual sum of squares: $SS_{res} = \sum_i (y_i - f_i)^2$, where f represents a fitted value. R^2 is calculated as $1 - SS_{res}/SS_{tot}$.

In the experiment, we set three different numbers of attributes (i.e., 17, 30, and 44), which corresponded to the minimum, average, and maximum numbers of attributes of the previous specification sheet.

NETWORK	Technology	GSM / HSPA / LTE
LAUNCH	Announced	2017, August
	Status	Available. Released 2017, September
BODY	Dimensions	151.7 x 75.4 x 7.3 mm (5.97 x 2.97 x 0.29 in)
	Weight	158 g (5.57 oz)
	Build	Front/back glass (Gorilla Glass 5), aluminum frame
	SIM	Single SIM (Nano-SIM) or Hybrid Dual SIM (Nano-SIM, dual stand-by)
		- MIL-STD-810G compliant - IP68 certified - dust/water proof over 1.5 meter and 30 minutes
DISPLAY	Type	P-OLED capacitive touchscreen, 16M colors
	Size	6.0 inches, 92.9 cm^2 (~81.2% screen-to-body ratio)
	Resolution	1440 x 2880 pixels, 18:9 ratio (~537 ppi density)
	Multitouch	Yes
	Protection	Corning Gorilla Glass 5
		- Dolby Vision/HDR10 compliant - Always-on display - LG UX 6.0+
PLATFORM	OS	Android 7.1.2 (Nougat)
	Chipset	Qualcomm MSM8998 Snapdragon 835
	CPU	Octa-core (4x2.45 GHz Kryo & 4x1.9 GHz Kryo)
	GPU	Adreno 540
MEMORY	Card slot	microSD, up to 256 GB (dedicated slot) - single-SIM model microSD, up to 256 GB (uses SIM 2 slot) - dual-SIM model
	Internal	128 GB, 4 GB RAM (V30+) or 64 GB, 4 GB RAM (V30)
CAMERA	Primary	Dual: 16 MP (f/1.6, 1 µm, 3-axis OIS, PDAF) + 13 MP (f/1.9, no AF), phase detection & laser autofocus, LED flash, check quality
	Features	Geo-tagging, touch focus, face detection, HDR, panorama
	Video	2160p@30fps, 1080p@30/60fps, 720p@120fps, 24-bit/192kHz stereo sound rec., HDR video, check quality
	Secondary	5 MP (f/2.2, 1/5", 1.12 µm)
SOUND	Alert types	Vibration; MP3, WAV ringtones
	Loudspeaker	Yes
	3.5mm jack	Yes
		- 32-bit/192kHz audio - B&O Play certified - 24-bit/48kHz audio recording - Active noise cancellation with dedicated mic

Fig. 4 Specification sheet of LG V30. *Source*: GSMArena

4.2 Data description

For the first experiment, we acquired the survey results from LG Electronics. These consisted of the product attributes ordered by the importance of the purchasing factors considered as the most significant by customers. Such surveys are conducted periodically by the company for each of its smartphone devices, such as G4, V10, G5, V20, G6, and V30.

Furthermore, the specification sheet (Fig. 4), addressed in the second experiment, is a list describing the specifications or properties of a product on a commercial site, such as Amazon.com. The specification sheet contains information that comes uppermost to the minds of customers when they want to obtain information about a product, particularly when buying electronic devices. Thus, the selection of the attribute contained in the specification sheet is an important task, considering the ubiquity of specification sheets.

Moreover, below, we present the attributes of the specification sheets presented on commercial or review sites for a smartphone product. We examined the websites of five major companies or portals, namely, Amazon, Best Buy, GSMArena, CNET, and PhoneArena, and the official LG website. The attributes are presented in the order in which these websites are listed.

Amazon (17) Screen size, Display type, Color spectrum, Resolution, Glass type, Network, Storage, RAM, SD slot, First rear camera resolution, Second rear camera resolution, Front camera resolution, OS version, Processor, Battery, Wireless charging, In the box

Best Buy (19) Processor, OS version, Network, Screen size, Screen ratio, Resolution, Display type, First rear camera resolution, Second rear camera resolution, Front camera resolution, Camera angle, Network, Storage, SD slot, Mobile hotspot, QSlide, QuickMemo, Water resistant, Warranty

GSMArena (30) OS version, Dimensions, Weight, Materials, Fingerprint, Water resistant, Dust resistant, Colors, Screen size, Resolution, Pixel density, Display type, Glass type, Sensor, First rear camera resolution, Second rear camera resolution, Front camera resolution, Camera angle, Camera feature, Camcorder resolution, Processor, Storage, RAM, SD slot, Battery, Wireless charging, speaker, Microphone, Network, Voice feature

CNET (34) Weight, Color, Network, Form factor, OS version, User interface, Intelligent assistant, SIM Card, Sensor, Materials, Water resistant, Dust resistant, Messaging, Processor, Wireless interface, Resolution, Pixel density, Screen size, Screen features, Screen ratio, Audio codec, Video codec, Memory, SD card, Battery, Wireless charging, Camera feature, Security, RAM, First rear camera resolution, Second rear camera resolution, Front camera resolution, Warranty, Dimensions

PhoneArena (36) Network, Dimensions, Weight, Materials, Glass type, SIM card, Display type, Screen size, Resolution, Screen ratio, Multi-touch, Display feature, User interface, OS version, Processor (CPU), Processor (GPU), Memory, SD card, SIM card, First rear camera resolution, Second rear camera resolution, Front camera resolution, Video resolution, Speaker, Earphone jack, Network, GPS, NFC, Radio, USB, Sensor, Messaging, Browser, Battery, Colors, Test results

Official site (44) Screen size, Display type, Pixel density, Screen ratio, Camera feature, System features, Display features, First rear camera resolution, Second rear camera resolution, Front camera resolution, Front camera angle, Rear camera angle, Camera feature, Video resolution, Video feature, Voice recording feature hardware, Voice recording feature, Hi-Fi, DAC, Material, Fingerprint, Dimensions, Weight, Water resistant, Shock resistant, Glass type, Security features, Productivity features, Convenience features, Entertainment features, Connectivity features, OS version, User interface, Processor, Battery, Network, Fast charging, USB, Memory, Micro SD, RAM, Earphone jack, Accessory

As shown above, many differences exist between each site. For instance, Best Buy and GSMArena do not contain the *User interface* attribute and only Amazon contains the *In the box* attribute. The LG Electronics official site and PhoneArena contain more than double the number of attributes contained on Amazon. Thus, to create an effective specification sheet, we conclude that it is relevant to study and select reasonable attributes that influence the purchase intention of a customer.

4.3 Experiment results

4.3.1 Experiment for the comparison between extraction and prioritization

Table 4 lists the NDCG results of the comparison between the extracted product attributes and the answer set acquired from the user survey results conducted by LG Electronics. As mentioned previously, we tested the performances of our proposed method and a few baselines on each of the following LG products: V10, G5, V20, G6, and V30.

Although the NDCG results varied per product, our proposed method outperformed the other methods on all the considered products. In the aspect of prioritizing method, our proposed variants of the Grad-CAM approach yielded better results than those obtained by other explainable machine learning approaches such as long short-term

Table 4 Performance of the attribute extraction and prioritization (NDCG)

Method	V10	G5	V20	G6	V30
Proposed method	0.9273	0.9046	0.9215	0.9171	0.9013
CNN+Transfer learning and LSTM Attention	0.9046	0.8920	0.9103	0.9018	0.8876
CNN+Transfer learning and LIME	0.8844	0.8803	0.8961	0.8916	0.8803
CNN and LSTM Attention	0.8749	0.8806	0.8713	0.8703	0.8894
CNN and LIME	0.8561	0.8692	0.8461	0.8394	0.8467
Dlirec and LSTM Attention	0.8379	0.8503	0.8307	0.8128	0.8368
Dlirec and LIME	0.8276	0.8346	0.8149	0.8017	0.8176
Hierarchical CRF and LSTM Attention	0.8284	0.8390	0.9051	0.7948	0.8013
Hierarchical CRF and LIME	0.8013	0.8164	0.7913	0.7813	0.7851
Hierarchical CRF and Frequency	0.7304	0.7491	0.7144	0.7024	0.7107
Lexicalized HMM and LSTM Attention	0.7356	0.7327	0.7263	0.7491	0.7318
Lexicalized HMM and LIME	0.7216	0.7276	0.7137	0.7374	0.7104
Lexicalized HMM and Frequency	0.6456	0.6394	0.6214	0.6307	0.6216

memory (LSTM) attention and LIME. By examining the detailed results, we conclude that LSTM attention showed inconsistent weight calculations on the basis of the length of each textual review, and, thus, it caused bias in the overall weight calculation.

The LIME approach was more appropriate for binary decisions for each aspect, but not for weight calculation. Nonetheless, clearly, all the explainable machine learning-based approaches outperformed the simple frequency-based approaches. On the basis of the experiment results, we conclude that the explainable machine learning-based approaches provided an effective weight score in the calculation of the relative importance of the product attributes. Moreover, we also conclude that the frequency-based approaches caused general aspects to be irrelevantly ranked highest. Furthermore, the naive CNN-based approach and the combination of the CNN-based method and transfer learning outperformed the other methods such as the CRF and HMM-based approaches on the aspect extraction problem, as verified in a previous study [24].

4.3.2 Experiment on the development of an effective specification sheet

Table 5 summarizes the extracted attributes obtained by the proposed method for the LG V30 product. The major key features of the product, such as video features (Cine Video mode) and camera lens (Crystal Clear Lens), were effectively extracted by the proposed method mostly because the transfer learning approach was applied. These features also had a relevant slope coefficient, β, in the regression model for the satisfaction score.

As presented in Table 6, our proposed method shows a higher influence score and larger coefficients of determination than the values for the existing specification sheet for all the corresponding number of attributes. Thus, our proposed method effectively reflects the interest of a customer and identifies the essential element influencing the purchasing intention of a customer. We also considered the recent improvements to LG V30 by utilizing the transfer learning approach.

Therefore, the method used for the existing specification sheet is less effective than the proposed method even though it was developed by domain experts who have extensive background knowledge of the smartphone industry.

Table 5 Examples of extracted attributes

Attribute	β	Attribute	β	Attribute	β
Hi-Fi	0.1019	*Voice features*	0.0846	*Camera angle*	0.0785
AI features	0.0743	*Camera lens*	0.0716	*Display type*	0.0673
Video features	0.0654	*Finger print*	0.0584	*Water resistance*	0.0519

5 Conclusion

This study proposed an advanced method based on explainable machine learning algorithms for extracting and prioritizing the product attributes that can effectively reflect the purchasing intention of a user. Most of the previous studies focused only on the aspect extraction without considering the relative importance of the extracted attributes or utilized only simple frequency-based prioritization approaches.

Thus, we proposed a novel method to prioritize the extracted attributes using variants of the Grad-CAM algorithm to understand the importance score of each extracted aspect. We focused on understanding the relative importance of an attribute for the first time to allow product developers to consider only the attribute deemed most important by customers. Using a sentimental prediction model of each textual review, we calculated the weight of each aspect affecting the sentiment decision. By identifying the aspects having a significant effect on the customer sentiment decision, we found that they are the ones considered most important by customers when evaluating a product. Additionally, we applied the off-the-shelf feature concept to the construction of the sentiment classification model to improve the overall efficiency of our method.

Table 6 Result of the effectiveness comparison

Source	Average influence score			R^2		
Specification sheet	Minimum	Average	Maximum	Minimum	Average	Maximum
Proposed method	4.13	4.01	3.84	0.6236	0.5329	0.4829
Amazon	3.94			0.5219		
Best Buy	3.61			0.4917		
GSMArena	3.59			0.4532		
CNET	3.46			0.4048		
PhoneArena	3.33			0.4129		
Official site	3.42			0.4483		

The experiments showed that the performance of the product attribute prioritization of our proposed method outperformed those of the other previous approaches. This was found by comparing the NDCG values with the answer set acquired from LG Electronics. Moreover, our method exhibited a better performance on the development of a specification sheet than those used by the existing commercial sites for all the corresponding number of attributes. Thus, our proposed method effectively reflected the interest of customers and the essential elements influencing their purchasing intention, using numerous real customer reviews. Furthermore, we effectively considered the recent improvements to LG V30 by utilizing the transfer learning approach.

Future studies could extend the scope of prioritization research based on this study of extracting the product attributes. These studies could involve the structure of a specification sheet, including the order of the attributes or attribute grouping and the appropriate number of attributes for a more effective specification sheet. The future studies could verify the effectiveness of the development of a specification sheet by a comparison with a previous specification sheet based on real sales volume or page views. Moreover, these studies could be expected to help in the widespread application of the proposed method for developing specification sheets in various tasks arising in the context of real-world business environments.

Acknowledgements This study was supported by Research Program funded by the SeoulTech.

Compliance with ethical standards

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest to declare.

References

1. Bing L, Wong TL, Lam W (2016) Unsupervised extraction of popular product attributes from e-commerce web sites by considering customer reviews. ACM Trans Internet Technol (TOIT) 16(2):12
2. Choi Y, Cardie C (2010) Hierarchical sequential learning for extracting opinions and their attributes. In: Proceedings of the ACL 2010 conference short papers. Association for Computational Linguistics, pp 269–274
3. Fuchs C, Prandelli E, Schreier M (2010) The psychological effects of empowerment strategies on consumers' product demand. J Mark 74(1):65–79
4. van Ginneken B, Setio AA, Jacobs C, Ciompi F (2015) Off-the-shelf convolutional neural network features for pulmonary nodule detection in computed tomography scans. In: IEEE 12th international symposium on biomedical imaging (ISBI). IEEE, pp 286–289
5. Huang S, Liu X, Peng X, Niu Z (2012) Fine-grained product features extraction and categorization in reviews opinion mining. In: IEEE 12th international conference on data mining workshops (ICDMW). IEEE, pp 680–686
6. Jagarlamudi J, Daumé III H, Udupa R (2012) Incorporating lexical priors into topic models. In: Proceedings of the 13th conference of the European chapter of the association for computational linguistics. Association for Computational Linguistics, pp 204–213
7. Jakob N, Gurevych I (2010) Extracting opinion targets in a single- and cross-domain setting with conditional random fields. In: Proceedings of the 2010 conference on empirical methods in natural language processing. Association for Computational Linguistics, pp 1035–1045
8. Jamal A, Goode M (2001) Consumers' product evaluation: a study of the primary evaluative criteria in the precious jewellery market in the UK. J Consum Behav 1(2):140–155
9. Järvelin K, Kekäläinen J (2000) IR evaluation methods for retrieving highly relevant documents. In: Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval. ACM, pp 41–48
10. Järvelin K, Kekäläinen J (2002) Cumulated gain-based evaluation of IR techniques. ACM Trans Inf Syst (TOIS) 20(4):422–446
11. Jin W, Ho HH, Srihari RK (2009) A novel lexicalized hmm-based learning framework for web opinion mining. In: Proceedings of the 26th annual international conference on machine learning, pp 465–472
12. Kaboli M (2017) A review of transfer learning algorithms. Ph.D. thesis, Technische Universität München
13. Kangale A, Kumar SK, Naeem MA, Williams M, Tiwari MK (2016) Mining consumer reviews to generate ratings of different product attributes while producing feature-based review-summary. Int J Syst Sci 47(13):3272–3286
14. Lee TY, Bradlow ET (2011) Automated marketing research using online customer reviews. J Mark Res 48(5):881–894
15. Lee Y, Chung M, Cho S, Choi J (2019) Extraction of product evaluation factors with a convolutional neural network and transfer learning. Neural Proc Lett 50:149–164
16. Li F, Han C, Huang M, Zhu X, Xia YJ, Zhang S, Yu H (2010) Structure-aware review mining and summarization. In: Proceedings of the 23rd international conference on computational linguistics. Association for Computational Linguistics, pp 653–661
17. Li Z, Tang J (2015) Weakly supervised deep metric learning for community-contributed image retrieval. IEEE Trans Multimed 17(11):1989–1999
18. Li Z, Tang J (2016) Weakly supervised deep matrix factorization for social image understanding. IEEE Trans Image Proc 26(1):276–288
19. Li Z, Tang J, Mei T (2018) Deep collaborative embedding for social image understanding. IEEE Trans Pattern Anal Mach Intell 41:2070–2083
20. Lim S, Tucker CS (2016) A bayesian sampling method for product feature extraction from large-scale textual data. J Mech Des 138(6):061403
21. Liu Q, Gao Z, Liu B, Zhang Y (2015) Automated rule selection for aspect extraction in opinion mining. IJCAI 15:1291–1297
22. Mikolov T, Chen K, Corrado G, Dean J (2013) Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781
23. Pan SJ, Yang Q (2010) A survey on transfer learning. IEEE Trans Knowl Data Eng 22(10):1345–1359
24. Poria S, Cambria E, Gelbukh A (2016) Aspect extraction for opinion mining with a deep convolutional neural network. Knowl Based Syst 108:42–49
25. Poria S, Cambria E, Ku LW, Gui C, Gelbukh A (2014) A rule-based approach to aspect extraction from product reviews. In: Proceedings of the second workshop on natural language processing for social media (SocialNLP), pp 28–37

26. Qiu G, Liu B, Bu J, Chen C (2011) Opinion word expansion and target extraction through double propagation. Comput Linguist 37(1):9–27
27. Quan C, Ren F (2014) Unsupervised product feature extraction for feature-oriented opinion determination. Inf Sci 272:16–28
28. Ribeiro MT, Singh S, Guestrin C (2016) Why should I trust you? Explaining the predictions of any classifier. In: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. ACM, pp 1135–1144
29. Samiee S (1994) Customer evaluation of products in a global market. J Int Bus Stud 25(3):579–604
30. Selvaraju RR, Das A, Vedantam R, Cogswell M, Parikh D, Batra D (2016) Grad-cam: why did you say that? Visual explanations from deep networks via gradient-based localization. CoRR. abs/1610.02391, 7
31. SemEval-2014: Semantic evaluation exercises. In: International workshop on semantic evaluation (SemEval-2014), Dublin, Ireland. https://alt.qcri.org/semeval2014
32. Sharif Razavian A, Azizpour H, Sullivan J, Carlsson S (2014) CNN features off-the-shelf: an astounding baseline for recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pp 806–813
33. Socher R, Ganjoo M, Manning CD, Ng A (2013) Zero-shot learning through cross-modal transfer. In: Advances in neural information processing systems, pp 935–943
34. Toh Z, Wang W (2014) DLIREC: aspect term extraction and term polarity classification system. In: Proceedings of the 8th international workshop on semantic evaluation (SemEval 2014), pp 235–240
35. Wang T, Cai Y, Leung H, Lau RY, Li Q, Min H (2014) Productaspect extraction supervised with online domain knowledge. Knowl Based Syst 71:86–100
36. Wang Y, Huang M, Zhao L et al (2016) Attention-based LSTM for aspect-level sentiment classification. In: Proceedings of the 2016 conference on empirical methods in natural language processing, pp 606–615
37. Wang Y, Wang L, Li Y, He D, Liu TY (2013) A theoretical analysis of NDCG type ranking measures. In: Conference on learning theory, pp 25–54
38. Xu K, Ba J, Kiros R, Cho K, Courville A, Salakhudinov R, Zemel R, Bengio Y (2015) Show, attend and tell: neural image caption generation with visual attention. In: International conference on machine learning, pp 2048–2057
39. Xu X, Wang X, Li Y, Haghighi M (2017) Business intelligence in online customer textual reviews: understanding consumer perceptions and influential factors. Int J Inf Manag 37(6):673–683
40. Yang B, Cardie C (2013) Joint inference for fine-grained opinion extraction. ACL 1:1640–1649
41. Zhong S (2005) Efficient online spherical k-means clustering. In: Proceedings of IEEE international joint conference on neural networks, vol 5. IEEE, pp 3180–3185