Extracting principal smartness dimensions of smart speakers using topic modeling and sentiment analysis

Wonbin Park
Department of Industrial and Systems Engineering
Korea Advanced Institute of Science and Technology
Daejeon, Republic of Korea
wonbinp@kaist.ac.kr

Seungwoo Choi
Department of Industrial and Systems Engineering
Korea Advanced Institute of Science and Technology
Daejeon, Republic of Korea
sw.choi@kaist.ac.kr

Mun Yong Yi
Department of Industrial and Systems Engineering
Korea Advanced Institute of Science and Technology
Daejeon, Republic of Korea
munyi@kaist.ac.kr

Abstract— **Although the smart speaker market has experienced massive growth in recent years, there is a lack of research on what consumers really consider important for so called "smart" speakers, which are supposedly distinguished from traditional products. Therefore, this study aims at identifying key smartness dimensions that are related to the satisfaction of smart speaker users. First, a total of seven topics were extracted from the Amazon's review data through Latent Dirichlet Allocation, and the topics were mapped, through group discussion, to three smartness elements defined from the literature review. Then, sentiment scores of each topic were calculated using SentiWordNet, which were then used as variables to develop star rating classifiers. The feature importance of the classifiers revealed that the connectivity issue is the most influential factor in determining the customer satisfaction of smart speakers. The next important topics are sound quality and its use as a media player. The study findings have direct practical implications on smart speaker development.**

Keywords - Smart speakers, Smartness, Sentiment analysis, Topic modeling

I. Introduction

Sales of smart speakers, each of which includes a built-in virtual artificial intelligence assistant that can interact with human voice in addition to the traditional speaker functionality, have been exploding in recent years. People who have experienced a smart product have a strong motivation to buy it again [1], and the characteristics of smart products that are differentiated from ordinary products are called smartness [5], which is believed to underlie the explosion.

However, users seem not so excited about some of the smartness features despite the popularity. For example, out of a multitude of smart speaker features, only three (music, information, and automation) were found primarily used with little exploration of new domains over time [2]. There has been a considerable lack of research on how smart speaker functions influence users' satisfaction. Therefore, for the sustainable growth of the smart speaker market, it is very important to understand the smartness dimensions of the smart speaker that users really value more.

This study aims at identifying key smartness dimensions that significantly affect the satisfaction of smart speaker users. The present study was conducted through two stages. The first was to identify the smartness elements of smart speakers that users perceive important using actual users' review data. The second was to find out which of the smartness elements had a higher impact on users' star ratings. More specifically, as shown in the Fig. 1, the first stage was conducted to discover major topics in the Amazon's review data using Latent Dirichlet Allocation (LDA), and to map these topics to the smartness elements defined through literature review. The mapped smartness elements were then used to interpret subsequent analysis results. In the second step, a lexicon-based approach was used to extract sentiment scores for each topic. Then, star rating prediction models were created to check whether the identified topics were meaningful over and above the sentiment scores in predicting users' star rating. In the end, we distinguish which topics have a greater influence on the star rating through the comparison of feature importance.

II. Realted Work

There have been several attempts to define smartness dimensions of some smart products and to assess their impacts. Rijsdijk and Hultink (2009) defined five smartness dimensions and surveyed how they influence the innovation attributes of vacuum cleaners, mowers, washing machines, refrigerators, and digital cameras [3]. Lee and Shin (2018) surveyed how the same five smartness dimensions affect smartphone customer satisfaction [4]. Rhiu and Yun (2018) defined the five smartness dimensions: Autonomy, Adaptability, Multi-functionality, Connectivity, and Personalization through literature review and analyzed how these dimensions affect smartphone user satisfaction using social media data [5].

Methodologies for finding patterns of words in electronic documents made of text data have been studied for a long time. The methodologies based on hierarchical probabilistic models are called topic models. One of the most popular topic models is Latent Dirichlet Allocation (LDA) and the basic assumption of LDA is that all documents are represented by topic distribution and all topics by word distribution.

2375-9356/20/$31.00 ©2020 IEEE
DOI 10.1109/BigComp48618.2020.00-60

Sentiment analysis, also known as opinion mining, uses natural language processing techniques to extract and analyze sentiments from text data. For sentiment analysis, sentiment classification groups the sentiment direction of a given text into two or more classes. When extracting a sentiment of a word, the dictionary-based approach, which is one of the lexicon-based approaches, is to use a dictionary in which words are labeled for their sentiments.

III. METHOD

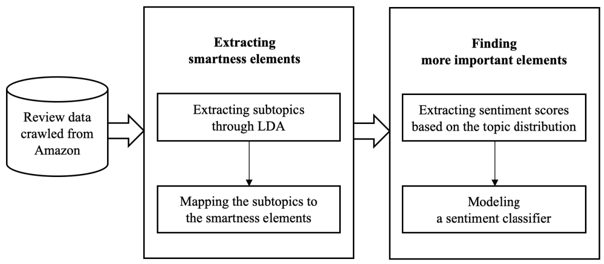

Figure 1. Framework of our approach

A. Extracting Smartness Dimensions

After crawling the Amazon's review data for smart speaker products, we performed a topic modeling through LDA. In order to increase the semantic coherence of abstract topics to be extracted, we performed LDA using only nouns [7]. The parameters were set to default, out-of-the-box settings (α = 0.1, β = 0.1, 3000 iterations). We also used a topic coherence [8] measure to select an optimal number of topics. Although many attempts have been made to create a measure that can evaluate topic models, it is still difficult to evaluate accurately as it is an unsupervised process [6]. Therefore, it is very helpful to reflect human's evaluation in terms of interpreting the model results. We recruited eight panel members and have them perform three tasks through discussion to evaluate the extracted topics: cleansing the sets of words, naming the abstract topics, and mapping the topics to smartness elements. In all of the tasks, review samples sorted by the topic probabilities were provided and the panel members were equipped with a search system that can search every word on the reviews so that they were not to be biased by the review samples. Unrelated words that received a majority vote were removed [6] and then topic distribution was calculated from the cleaned sets of words. A consensus on the name of each topic was reached through discussion. The panel members mapped the topics to the five smartness elements: Autonomy, Adaptability, Multi- functionality, Connectivity, and Personalization, defined from the literature review.

B. Extracting Sentiment Scores

As a preparatory step to compare and analyze the influence of the extracted topics on customer satisfaction, sentiment scores of each topic were extracted and used as the variables for sentiment classification. Based on the topic probabilities calculated from the cleaned sets of words, the positive and negative scores of the topics were extracted using SentiWordNet, which has been one of the most popular sentiment dictionaries. We assumed that the smallest meaningful unit of text in a review is a sentence. Thus, we first calculated the sentiment scores of each topic at a sentence level and then merged them back at a document level as shown in Figure 2.

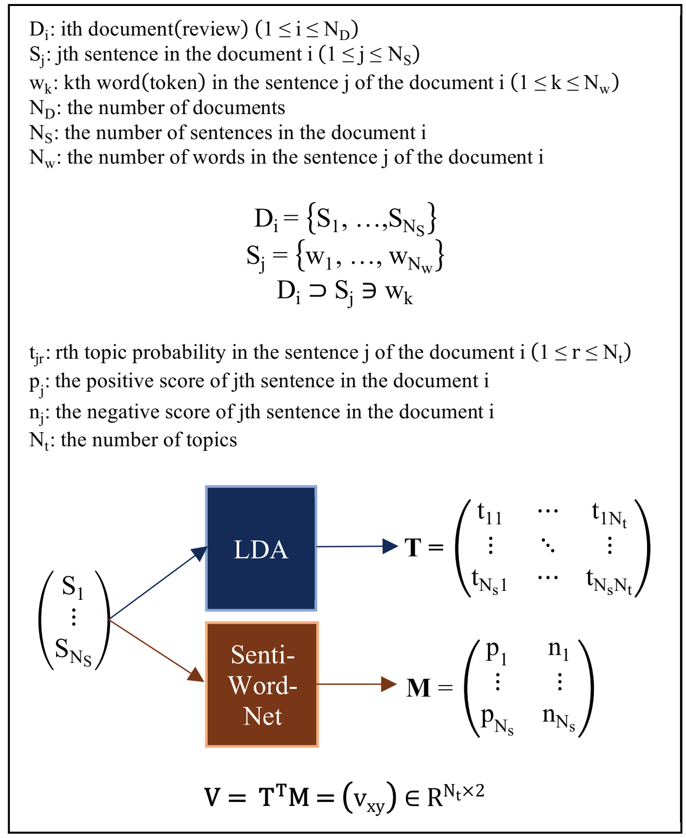

Figure 2. Sentiment score extraction

To prevent too many subscripts, we used S_j instead of S_j^i, w_k instead of w_k^{ij}, N_S instead of N_S^i and so forth. Matrix **T** represents the topic distribution of sentences and matrix **M** represents positive and negative scores of sentences. Matrix **V**, calculated from the dot product of the matrices **T** and **M**, represents positive and negative scores of each topic at a document level. The elements v_{xy} are the variables to be used when modeling a sentiment classifier.

C. Modeling Sentiment Classifiers

We performed modeling with the sentiment scores as independent variables and the user's star rating as a target variable. Amazon reviews have 1 to 5 star ratings and the problem was set as a binary classification by grouping 1 and 2 into a lower rating label and 4 and 5 into a high rating label. The control group was when the overall positive and negative scores were used as the independent variables without considering the topic distribution. The treatment group was when the positive and negative scores of each topic were separately considered as the independent variables along with the control group's variables. We used Support Vector

TABLE I. EXTRACTED TOPICS AND MAPPED SMARTNESS DIMENSIONS

Words of LDA Result	Example Review	Agreed Topic	Freq	Mapped Smartness	Freq
'echo', 'dot', 'sound', 'speaker', 'quality', 'generation', 'gen', 'volume', 'bluetooth', 'price', …	"I really like the improvement in sound … The look is improved over the gen 2 …"	Comparing Different Models (Sound quality, Appearance)	28%	Multi-functionality	63%
'alexa', 'time', 'question', 'weather', 'day', 'list', 'fun', 'alarm', 'news', 'timer', 'answer', …	"They use it to set an alarm or timer, check the weather …"	Smart Personal Assistant	20%		
'home', 'voice', 'alexa', 'google', 'command', 'skill', 'device', 'word', 'control', 'ability', 'response', …	"It can remember and hold a conversation like Google Home …"	Functionality Comparison between Brands (Voice interaction competence)	12%		
'light', 'plug', 'hub', 'smart', 'thermostat', 'bulb', 'outlet', 'switch', 'hue', 'door', …	"I Use it to control lighting in my house …"	Home Automation	3%		
'amazon', 'device', 'product', 'phone', 'app', 'problem', 'issue', 'time', 'work', 'everything', 'wifi', …	"I was never able to connect this to WiFi via Android Alexa app …"	Connectivity Problems	15%	Connectivity	15%
'music', 'room', 'house', 'song', 'feature', 'great', 'bedroom', 'family', 'love', 'radio', 'kitchen', 'play', …	"Got to play music in the kitchen while cooking …"	Media Player	18%	Personalization	22%
'gift', 'christmas', 'son', 'daughter', 'bought', 'husband', 'mother', 'year', 'birthday', 'present', …	"For Christmas gifts, I purchased Echo Dots for my husband …"	Gift	4%		

Machine (SVM), Random Forest (RF), XGBoost (XGB), Neural Network (NN) as learning algorithms. After modeling, we analyzed the feature importance of the classifiers to find more important smartness dimensions.

IV. EXPERIEMNTS

A. Dataset

We crawled 52,654 reviews of Amazon Echo products from 2016 to 2019. Ninety percent of them were used for topic modeling and the other ten percent were used as a test set for the sentiment classification.

B. Extracted Smartness Dimensions

As shown in the Table 1, a total of seven topics were extracted through LDA: 'Comparing Different Models (Sound Quality, Appearance)' which is about how much sound quality and appearance have improved over previous versions of the product; 'Smart Personal Assistant' which is about the personal assistant functions such as reporting weather, setting alarms and timers, listening to news, and answering questions; 'Functionality Comparison between Brands (Voice Interaction Competence)' which is about comparing different brands mainly on the voice interaction competence of virtual assistants; 'Home Automation' which is about the functions to use in a house such as turning on and off lights, adjusting a thermostat and locking doors; 'Connectivity Problems' which is about the troubles that occur when connecting to the Internet and other devices; 'Media Player' which is about using smart speakers as a media player; and 'Gift' which is about using smart speakers as a gift.

The extracted topics were mapped to the smartness categories of Multi-functionality (how many intelligent functions are provided), Connectivity (how much well compatible with other devices), and Personalization (how the provided content fits the user's needs). The most frequently appeared smartness element in the reviews was Multi-functionality, and its top three most frequent subtopics were Comparing Different Models (Sound Quality, Appearance), Smart Personal Assistant, and Media Player.

C. Performance of Sentiment Classfiers

All of the sentiment classifiers showed higher accuracy when the topic distribution was considered in addition to the overall sentiment scores, as shown in Table 2, indicating that the identified topics are helpful over and above the sentiment scores in predicting users' star rating. Therefore, we extracted feature importance from all the algorithms in order to compare the influences of the topics and find relatively more important topics.

TABLE II. ACCURACY OF SENTIMENT CLASSIFIERS

	Document Sentiments	Topic Sentiments
SVC	75.08%	75.33%
NN	73.49%	76.34%
RF	74.44%	76.50%
XGB	75.21%	76.95%

D. Feature Importacne

While the tree-based algorithms, RF and XGB, can determine feature importance in the modeling process, there is no built-in method to derive feature importance from SVM and NN, which is known as a black box. Hence we used a more intuitive way that is called permutation importance method for the two. Figure 3 shows the feature importance list extracted from each algorithm, and Table 3 combines all of them to rank topics. As shown in Table 3, Connectivity is the most important variable that determines customer satisfaction of smart speakers. The next four important variables are Sound Quality and Appearance, Media Player, Smart Personal Assistant and Voice Interaction Competence. In contrast, Gift and Home Automation are the variables that have less influence on customer satisfaction of smart speakers. Based

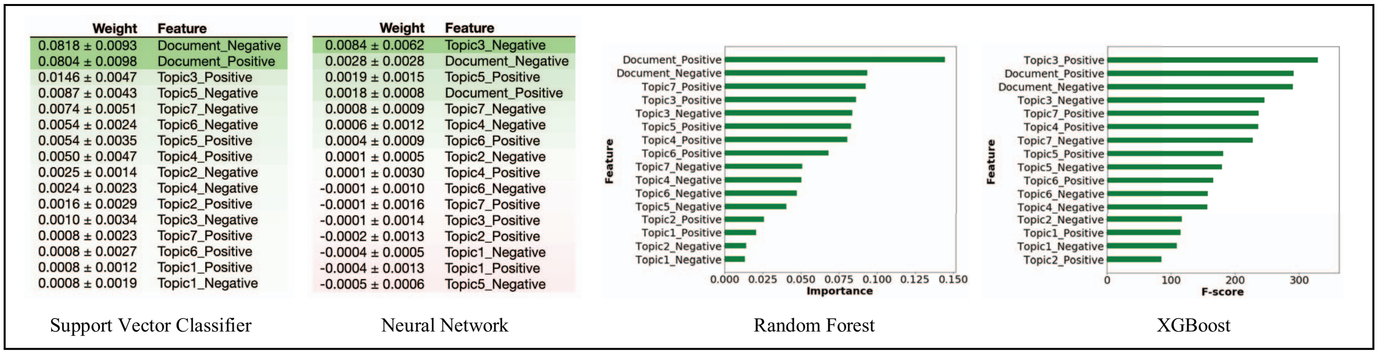

Figure 3. Feature importance in each algorithm

TABLE III. IMPORTANCE RANK OF TOPICS

Importance Rank	Topic
1	Connectivity Problems (Topic 3)
2	Comparing Different Models: Sound Quality, Appearance (Topic 7)
3	Media Player (Topic 5)
4	Smart Personal Assistant (Topic 4)
5	Functionality Comparison between Brands: Voice Interaction Competence (Topic 6)
6	Gift (Topic 2)
7	Home Automation (Topic 1)

on the results in Tables 1 and 3, discussions of the smartness dimensions are presented below.

a) Connectivity (Connectivity Problems): Although Connectivity has been mentioned not very frequently in the reviews, it is the most influential factor in determining customer satisfaction of smart speakers. In smartphones, connectivity problems are primary causes of unsatisfactory user experience [5]. Especially in the case of smart speakers, a connectivity issue can be very fatal because there are more than 60,000 devices that can be linked together. Therefore, in order for smart speakers to function as a hub for smart homes, connectivity problems must be effectively handled.

b) Multi-functionality (Sound Quality) & Personalization (Media Player): Sound Quality and Media Player are mentioned frequently in the reviews and are important variables that affect customer satisfaction. In a study for a long-term usage pattern of smart speakers, only three functions were generally used and the most requested command per day was about music [2]. Therefore, although a smart speaker has a wide variety of functions, the sound quality and its role as a media player, which are the most basic functionality and original purpose of a speaker, are still keys to user satisfaction.

c) Multi-functionality (Smart Personal Assistant, Home Automation): Smart Personal Assistant was frequently mentioned and found to have a high impact on customer satisfaction; Home Automation was little mentioned with a low impact on customer satisfaction. In other words, users tend to be more interested in personal assistant functions such as weather, alarm, timer, reminder, Q&A, and news, rather than in house automation functions such as light, thermostat, switch, and door.

V. CONCLUSION

This study aimed at determining what smartness dimensions of “smart” speakers are important for customer satisfaction. First, we identified an initial set of smartness elements by employing a Latent Dirichlet Allocation modeling using the Amazon’s review data. Second, the sentiment scores for the extracted topics were calculated and star rating prediction modeling was performed based on the extracted scores. Based on the feature importance results, key dimensions related to customer satisfaction were identified. These findings provide insights on what aspects of smart speakers consumers consider critical.

REFERENCES

[1] López, Tomás Sánchez, et al. "Taxonomy, technology and applications of smart objects." Information Systems Frontiers 13.2 (2011): 281-300.

[2] Bentley, Frank, et al. "Understanding the Long-Term Use of Smart Speaker Assistants." Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 2.3 (2018): 91.

[3] Rijsdijk, Serge A., and Erik Jan Hultink. "How today's consumers perceive tomorrow's smart products." Journal of Product Innovation Management 26.1 (2009): 24-42.

[4] Lee, Won-jun, and Seungjae Shin. "Effects of product smartness on satisfaction: focused on the perceived characteristics of smartphones." Journal of theoretical and applied electronic commerce research 13.2 (2018): 1-14.

[5] Rhiu, Ilsun, and Myung Hwan Yun. "Exploring user experience of smartphones in social media: A mixed-method analysis." International Journal of Human–Computer Interaction 34.10 (2018): 960-969.

[6] Chang, Jonathan, et al. "Reading tea leaves: How humans interpret topic models." Advances in neural information processing systems. 2009.

[7] Martin, Fiona, and Mark Johnson. "More efficient topic modelling through a noun only approach." Proceedings of the Australasian Language Technology Association Workshop 2015.

[8] Röder, Michael, Andreas Both, and Alexander Hinneburg. "Exploring the space of topic coherence measures." Proceedings of the eighth ACM international conference on Web search and data mining. ACM, 2015.