Journal of Industrial Information Technology and Application *ISSN(Online): 2586-0852*
Vol.6 No.4

Identifying and prioritizing the product attributes based on sentiment analysis and weighted formal concept analysis

Yujin Yang[1), Heejung Lee[2),)*

[1) Dept. of Industrial Data Engineering, Hanyang University, Seoul, Korea
[2) Dept. of Interdisciplinary Industrial Studies, Hanyang University, Seoul, Korea

Abstract. Identifying the relative strengths and weaknesses of the products from the customers' feedback can help companies improve their product quality and competitive strength. In the digital age, it becomes increasingly important to analyze the customer's online shopping reviews and incorporate them into product design concepts. While there is a great deal of research in this area, aspect-level sentiment analysis considering product attributes is one of the most challenging works. This study provides a new practical approach that employs sentiment analysis and weighted formal concept analysis to analyze online reviews. We conducted a case study for hairdryer brands; 17 attributes were collected and evaluated for the product quality of each brand, and Weighted Formal Concept Analysis prioritized the product attributes according to positive and negative viewpoints. The results are expected to provide helpful insights for hairdryer product designers.

Keywords: sentiment analysis, weighted formal concept analysis, hairdryer

1. Introduction

As the online market grows, the number of reviews written by customers increases. Analyzing the strengths and weaknesses through data can help companies improve product quality and competitiveness. Therefore, understanding customers reviews become essential. Sentiment analysis is a field of natural language processing that classifies positives and negatives in texts [1]. The previous studies investigating the sentiment analysis usually have studied the simple classification of positive and negative, which are helpful when product designers check the product reviews. Still, it is challenging to apply them to the product design process directly. On the other hand,

* Corresponding author: stdream@hanyang.ac.kr
Received: Feb 7, 2022; Accepted: Feb 25, 2022; Published: Dec 30, 2022

attribute unit analysis can make product designers understand consumer opinions in more detail and are highly useful for business. For this reason, conducting the sentiment analysis at the attribute level is critical [2]. In this paper, based on WFCA (Weighted Formal Concept Analysis) [3], we estimate positive and negative values at the attribute level to extract business information. It will be helpful to product designers for understanding customers' needs. Approaches for sentiment analysis include machine learning and FCA (Formal Concept Analysis) [4]. Machine learning-based sentiment analysis research is actively conducted at domestic and foreign research trends. While the FCA-based sentiment analysis approach is being conducted overseas, research is relatively few in Korea. This study has the difference from previous research in that it approaches using WFCA. Existing sentiment analysis studies classified data into positive and negative. In more advanced sentiment analysis, there is a way to organize positive words with +1 weight and negative comments with -1 weight. Furthermore, not just positive and negative classifications, it has been conducted to studies that dealt with various emotions. For example, if "performance is good," it is classified as positive, but it is difficult to find which attribute of the product is good. Therefore, there is a limit to directly utilizing the analysis results. To solve this problem, this study investigates the positives and negatives, focusing on the properties of products. The approach in this paper has a unique difference from previous studies. To extract more meaningful results, We conduct attribute-level emotion analysis. By doing so, it can improve the competitiveness of the product.

2. Backgrounds

1) Sentiment Analysis

The definition of opinion is a broad concept that involves information related to sentiment. The first goal of sentiment analysis is finding all opinions in given documents. The more advanced goals are explanation of the opinion and modifying the meaning of opinion. Generally, there are three levels of analysis related to Sentiment analysis: document-level analysis, Sentence-level analysis, and aspect-level analysis. First of all, document-level sentiment analysis is the most widely investigated topic. It aims to classify an opinion document as positive and negative opinions. The sentence-level analysis aims to classify each sentence as negative, positive, and neutral opinion. Because document-level and sentence-level analyses consider each document and sentiment as a whole, those do not consider opinion or sentiment targets. Differ from the document and sentence-level analysis, Aspect-level analysis concentrate on the complete sentiment analysis problem (classifying sentiments and extracting sentiment or opinion targets (entities and aspects)) [5].

Even the merits of aspect-level analysis have not been widely investigated because aspect-level sentiment analysis is one of the most challenging works. For this reason, we try to fill this gap.

2) Formal Concept Analysis

Formal concept analysis (FCA) is a mathematical method for data analysis based on order and lattice theory with applications in knowledge discovery and representation, data and text mining, software engineering, and so on [4, 6, 7]. FCA uses the simple data structure called a formal context as an input, and the most straightforward format for a formal context is a cross-table as shown in Table 2. More precisely, a formal context (G, M, R) consists of two sets, G and M, and a binary relation $R \subseteq G \times M$. The elements of G are called the objects and those of M attributes of (G, M, R). For two sets $A \subseteq G$ and $B \subseteq M$, the derivation operators $(\cdot)'$ are defined as $A' := \{ m \in M \mid (g, m) \in R$ for all $g \in A\}$ and dually $B' := \{ g \in G \mid (g, m) \in R$ for all $m \in B\}$. In other words, A' is the set of attributes that are common to the objects in A, while B' is the set of objects which have all attributes in B. For example, {g2, g3}' = {m2, m3}, {m1, m2, m3}' = {g3} for the formal context in Table 2. We can define (A, B) as a formal concept of (G, M, R) if and only if $A \subseteq G$, $B \subseteq M$, A' = B and B' = A. Set A is called the extent of the formal concept (A, B) and set B is called its intent. According to this definition, a formal concept has two parts: the extent consists of objects with all attributes of the intent. In contrast, the intent contains those attributes that all objects in the extent have in common.

3. Proposed Methods

The proposed study consists of two steps as shown in Fig. 1. The first step is to identify product attributes from online review data, calculate the frequency of positive and negative words for each product attribute, and prepare for WFCA analysis. In the second step, the priority for each positive and negative of the product attribute is obtained through WFCA. The final result is expressed in a dashboard so that the product designer can use it for product design.

1) Identifying Product Attributes

This step aims to discriminate the positive and negative of the product from the review data of the online shopping mall. First, we need to define the opinion schema: (e, a, s, r, u), where e: entity, a: aspect, s: sentiment words, r: reason words, u: usage. Opinion schema defines the elements essential for sentences for sentiment analysis. A predefined dictionary is required to perform sentiment analysis based on the above schema. For the sentiment dictionary, the KNU sentiment dictionary was used, and for the attribute dictionary, a dictionary built by a product planning expert was employed. To analyze the weighted FCA from the collected data after analysis, we define the formal context as shown in table I.

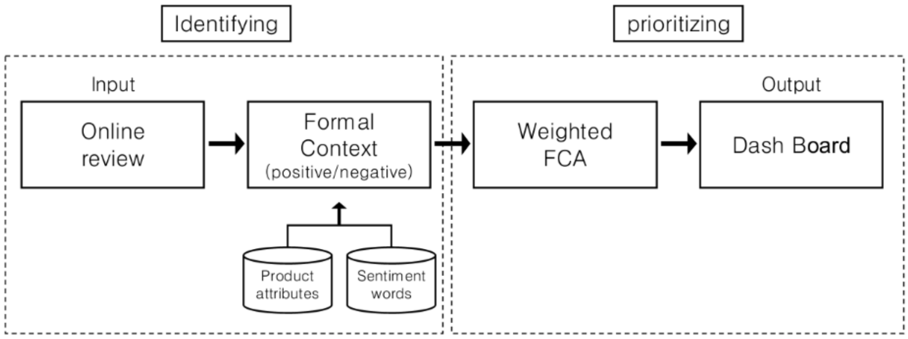

Fig 1. Research Framework

TABLE I. FORMAL CONTEXT FOR WFCA

Positive / Negative	Attributes 1	…	Attributes k
Brand 1	N11	…	N1k
…	…	…	…
Brand m	Nm1	…	Nmk

Nij : Number of times positive/negative of attribute j for brand i mentioned

2) Prioritizing Product Attributes

In general, FCA uses binary object-attribute relations to constitute a formal context of the problem. It means all attributes are considered to be equally important. To specify the preference of product attributes, we applied WFCA to introduce some weights in the formal context [3]. Weighted Formal Concept Analysis considers the relationship between objects, properties, and weights. The weight can be assigned by calculating the importance of the attribute. When estimating the significance of an attribute, four aspects are assumed. The four aspects are relevance, commonly-shared, conformance, and contribution. In this paper, we prioritize product attributes using WFCA. The formula for calculating Sa(mi), attribute importance for product attribute mi is shown as Equation (1):

$$S_a(m_i) = \sum_{C_k} \sum_{g_j \in C_k.extent} \left(\frac{g_j(m_i)}{|C_k.intent|}\right), \; for\ m_i \in C_k.intent \qquad (1)$$

where Ck.intent and Ck.extent are the intent and extent of formal concept Ck, respectively, |Ck.intent| is the size or number of objects in Ck.intent, and g(m) is zero when relevant intent or extent is an empty set. In Equation (1), Sa(mi) is proportional to the following three factors: the number of concepts that have the attribute mi as its intent (“relevance viewpoint”), the size of each concept’s extent (“commonly-shared viewpoint"), and the weighted value of the object-attribute relation ("conformance viewpoint"), while Sa(mi) is inversely proportional to the size of each concept's intent ("contribution viewpoint").

4. Case Study

Data were crawled reviews of '전문가용 드라이기' in Coupang, and ten brands were randomly selected among the crawled brands. For each brand hairdryer, 100 reviews were also randomly selected, and the study was conducted with a total of 1000 review data. When counting positive and negative, pre-constructed sentiment and attribute dictionaries were used.

First, the morphemes were analyzed using the Twitter morpheme analyzer (Okt Python library) for the sentence, sentiment dictionary, and attribute dictionary of the hairdryer review to be used in the study. Then, if the result of morphological analysis of the sentence of the review is in the attribute dictionary, the sentence is classified as a sentence with aspect. The classified sentences were organized as positive and negative based on the data defined in the emotional dictionary. The analysis results were derived in two ways: positive [Table II] and negative [Table III] of the following formal contexts

TABLE II. FORMAL CONTEXT - POSITIVE

search keyword : 전문가용 드라이기

		속성																
	긍정	가격 및 혜택	내구성	디지털센서	무선기능	바람	발생성분	브랜드	부가기능	소음진동	손잡이형태	안전기능	외관 디자인	위생	온도	조작감	코드	패키지
드라이기명	CKI	29	0	0	0	31	10	0	0	10	0	0	5	3	0	0	2	3
	JMW	2	30	3	36	65	0	0	22	1	9	20	68	0	5	3	0	30
	바비리스	12	14	21	12	78	0	0	7	5	1	7	20	0	26	21	0	6
	엘라	31	2	1	1	50	0	0	0	8	0	0	9	1	1	1	0	6
	예스뷰티	3	37	5	16	46	2	21	6	8	1	7	54	0	7	4	2	9
	유닉스	22	21	1	19	47	0	2	6	1	0	6	26	0	9	1	6	5
	조아스	28	3	6	4	57	0	0	4	2	0	4	35	0	24	6	0	28
	파테크	14	4	2	18	44	0	0	12	1	8	12	7	0	2	2	14	3
	피닉스	18	4	0	0	32	0	0	0	4	0	3	5	0	0	0	3	4
	한일전자	0	0	4	10	92	0	0	0	1	10	0	0	0	4	4	0	4

TABLE III. FORMAL CONTEXT - NEGATIVE

		속성																
	부정	가격 및 혜택	내구성	디지털센서	무선기능	바람	발생성분	브랜드	부가기능	소음진동	손잡이형태	안전기능	외관 디자인	위생	온도	조작감	코드	패키지
드라이기명	CKI	17	0	1	0	73	0	0	0	31	5	0	3	8	1	1	1	3
	JMW	0	20	1	0	18	0	0	0	11	2	0	20	0	1	1	0	1
	바비리스	1	30	0	3	31	0	0	11	27	2	11	34	0	5	0	0	0
	엘라	14	0	0	12	18	0	0	1	22	1	8	0	2	3	0	0	0
	예스뷰티	0	19	10	6	15	0	0	0	21	0	0	23	0	12	9	0	1
	유닉스	6	7	0	8	7	0	0	1	2	0	1	7	0	0	0	4	0
	조아스	2	3	0	0	20	0	0	0	37	0	0	3	0	8	0	0	6
	파테크	0	11	11	0	41	0	0	0	19	8	0	11	0	11	11	2	0
	피닉스	4	10	0	2	23	0	0	0	14	0	4	10	0	1	0	0	0
	한일전자	0	0	1	2	92	0	0	0	19	1	2	0	0	1	1	0	3

When the weight [Table IV] is calculated for the attributes, it can be seen that the 'wind' attribute item occupies the most significant weight in the positive evaluation of the hairdryer. As for the negative weight, the 'wind' and 'noise and vibration' attribute items are prominent.

TABLE IV. ATTRIBUTE PROIROTY

Product Attributes	weight for positive	weight for negative	Product Attributes	weight for positive	weight for negative
가격 및 혜택	136.60	83.33	손잡이형태	5.79	26.65
내구성	84.62	146.56	안전기능	24.78	37.32
디지털센서	21.92	29.12	온도	36.90	120.81
무선기능	63.83	58.22	외관 디자인	224.72	310.27
바람	700.84	1602.13	위생	1.13	2.62
발생성분	3.09	0.00	조작감	21.11	27.92
부가기능	12.73	6.49	코드	11.98	5.08
브랜드	2.96	0.00	패키지	131.91	14.59
소음진동	56.77	993.34	-	-	-

The attribute of the dryer, which received the most opinions from consumers, was the '바람' attribute, which was found in both positive and negative weights. The attribute that received the most opinions after '바람' in the positive weight was '외관 디자인' attribute. Same as in the positive weighted part, the volume of "외관 디자인" is considerable in the negative weighted part. As there were many opinions about '외관 디자인,' it suggests that product planners need to pay attention to the attribute. For the '소음진동' attribute, it can be seen that there are many opinions in the negative weights, and relatively few opinions in the positive weights. This can be interpreted as the need to improve the '소음진동' properties. The '패키지' attribute has a higher positive weight than the negative weight, which can be interpreted as being easy to get a favorable image if the package design is good. In summary Fig. 2 shows the dashboard for the importance of positive and negative factors for product attributes.

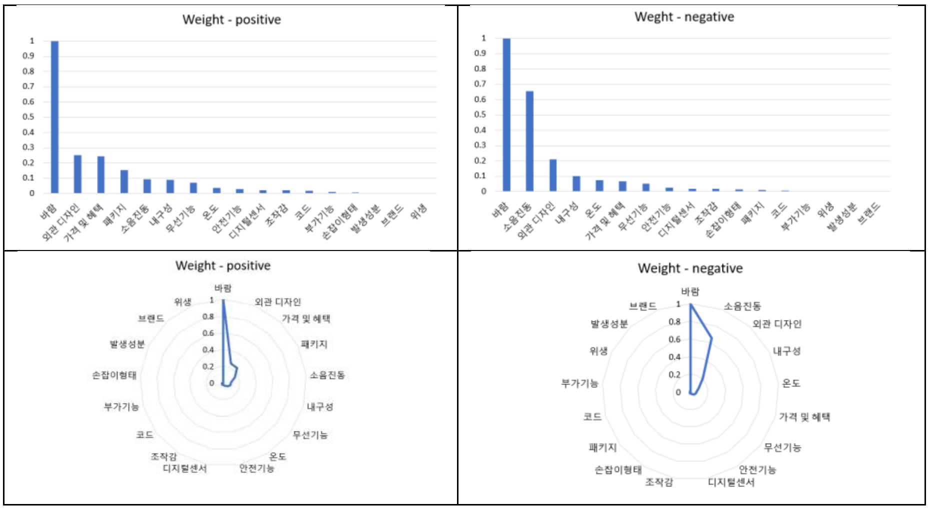

Fig 2. Dash board

5. Conclusion

In this study, sentiment analysis was performed on online shopping mall reviews in consideration of product attributes. It was attempted to obtain meaningful results by expressing them based on WFCA. Product planning managers need to manage negative opinions more subdivided. As such, by approaching sentiment analysis with attribute weights, a new perspective could be obtained.

Acknowledgment

This work was supported by the Big data-based Service Development Program (S2929815) funded by the Small and Medium Business Administration(SMBA, Korea)

References

[1] M. Birjali, K. Mohammed, B. Abderrahim, "A comprehensive survey on sentiment analysis: Approaches, challenges and trends," Knowledge-Based Systems 226: 107134, 2021

[2] X. Guo, W. Yu, X. Wang, "An overview on fine-grained text sentiment analysis: Survey and Challenges," Journal of Physics: Conference Series. 1757(1), 2021.

[3] S. Kwon, T. Kim, H. Suh, H. Lee, "Identifying the mobile application repertoire based on weighted formal concept analysis," Expert Systems with Applications 173: 114678, 2021

[4] R. Wille, (1982). "Restructuring lattice theory: An approach based on hierarchies of concepts," Ordered Sets, Springer, Dordrecht, pp. 445–470, 1982.

[5] B. Liu, "Sentiment analysis: Mining opinions, sentiments, and emotions," Cambridge university press, 2020.

[6] G. Stumme, R. Taouil, Y. Bastide, N. Pasquier, L. Lakhal, "Computing iceberg concept lattices with TITANIC," Data & knowledge engineering, 42(2), pp. 189-222, 2002.

[7] U. Priss, "Formal concept analysis in information science," Arist, 40(1), pp. 521-543, 2006.