TEXT MINING-BASED MAPPING FOR KANO QUALITY FACTOR

HaeKyung Lee, Myung Soo Cha and Taioun Kim*

Department of Industrial and Management Engineering
Kyungsung University
309 Suyeong-ro, Namgu, Busan 48434, Korea
{ hshklee72; mscha }@ks.ac.kr; *Corresponding author: twkim@ks.ac.kr

ABSTRACT. *The relationship between product quality/function and customer satisfaction has been considered as an important point for the new product development. The seminal paper by Kano was the first to thoroughly address the non-linear relationship between product performance/function and customer satisfaction. In the Kano model, two dimensional evaluation table method has been adopted for determining customer satisfaction about product and service. This research proposes combining text mining and Kano model for identifying customer satisfaction about hotel service. Based on Kano mapping model, 3,791 hotel review data are collected from TripAdvisor. The review data are analyzed using Morpheme Analysis and Emotion Analysis. Kano quality factor is derived from positive and negative sentiment.*
Keywords: Kano, Text mining, Big data analysis, Morpheme analysis, Emotion analysis

1. **Introduction.** Identifying customer needs correctly and providing product and service what the customer wants are crucial from the quality assurance point of view. In the Hertzberg's theory of motivation (also called 'two-factor theory'), the factors that make people feel satisfied are all belonging to work content, and matters that make people feel dissatisfied with their work are all work environment or relationship. The former is called as the incentive factor, and the latter is called the health factor. Kano et al. extrapolated Herzberg's two-factor theory and developed a model to identify the quality attributes that influence customer satisfaction, and suggested a non-linear relationship between them [1].

The Kano model can be used to identify and classify the quality attributes based on their impact on customer satisfaction into must-be, attractive, one-dimensional, reverse, and indifferent.

Reviewing the relevant research literature of the Kano model is conducive to discovering the development trend and mainstream direction of the theoretical research. Meng and Dong [2] performed the relevant research literature of Kano model through the visualization software and carried out comparative analysis, presented the development context of the current Kano model research, exploring the research direction that Kano model can expand in the future.

In the online text analysis, the opinion mining and the sentiment analysis of the network comment are important. By excavating the comment information of the online products, the real demand of the customers can be obtained; however, it does not reflect the differences among the requirements. Combining the data mining technology with the Kano model, the feature theme of the product was discovered by establishing the comment mining model and analyzing the sentiment of the comment through machine learning [3].

Based on Kano model, the relationship between product/service function and customer satisfaction is either linear or nonlinear. Especially, it is a competitive advantage to pro-

vide product and service with 'Attractive Quality' in this highly competitive environment. Thus, it is highly required to adopt the Kano model in the process of new product development and service during life cycle. Also, it can be applied to determining the importance of customer requirement factor in the Quality Function Deployment (QFD).

In the Kano model, two dimensional evaluation table method has been adopted for determining customer satisfaction about product and service. This method is based on positive and negative questionnaire about a product function and service [4]. This kind of survey methodology requires much time and cost, and it does not show exactly what the customer wants. Especially, in case of brand new product, it is more difficult to identify customer needs exactly because even the customer does not have any ideas about the new product.

Recently, as many customers share various usage experience about new product and service through Internet comment and Social Network Service (SNS), they purchase new product and service based on these media comment and review. Thus, suppliers of product and service utilize customer reviews to improve the quality of product and service. Big data analysis about customer reviews has been widely adopted in this area. Customer reviews have been adopted for evaluating customer satisfaction by Kano model [5].

Text mining and web crawling method are adopted for affective word collection to tackle the issue of the financial and time cost [6]. Web crawling enables the collection of large amounts of information in a short period without requiring much effort, while the text mining technique allows the quick extraction of meaningful affective words from a large amount of text data.

This research proposes combining text mining and Kano model for identifying customer satisfaction about hotel service. The paper is composed of the following chapters. Section 2 reviews theoretical background of Kano model. Section 3 proposes analysis model for determining Kano quality factors. Section 4 presents implementation results and discussion. Section 5 shows conclusion and future research directions.

2. **Kano Model & Evaluation Method.** In Kano model, there are five quality attributes, which are composed of x-axis (requirement fulfillment) and y-axis (customer satisfaction) shown in Figure 1. According to the degree of requirement fulfillment, the
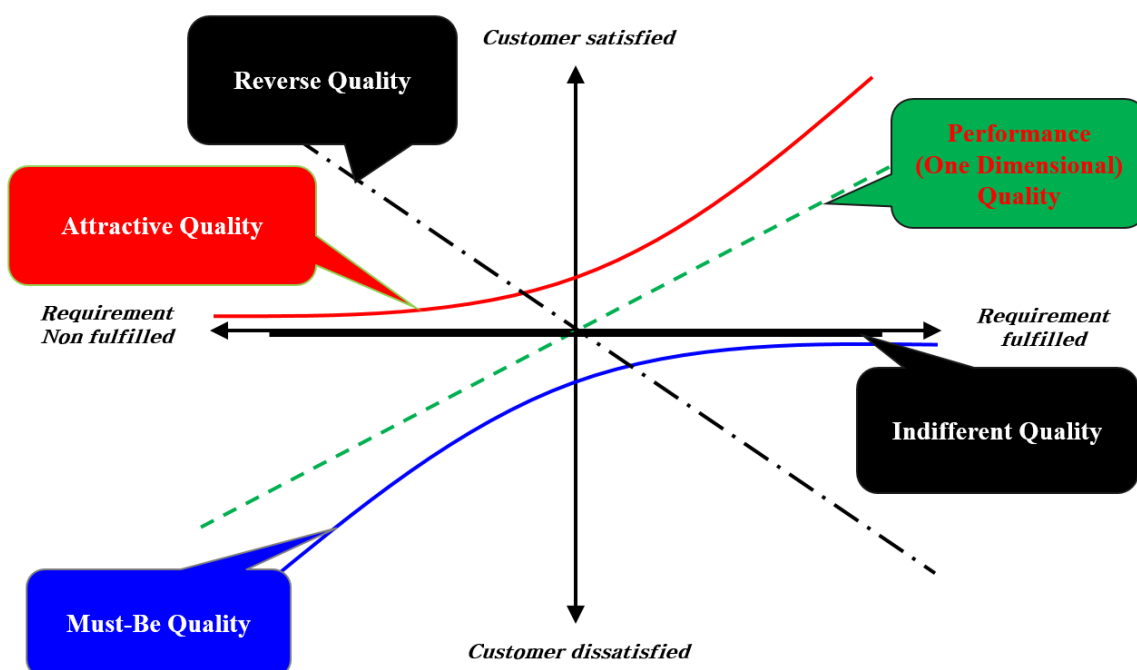

FIGURE 1. Kano model

satisfaction increases linearly, which is called 'One Dimensional Quality'. 'Attractive Quality' is meant satisfaction increases exponentially according to requirement fulfillment even though there is not serious dissatisfaction for un-fulfillment of requirement. 'Must-Be' quality means that these are the requirements that the customers expect and are taken for granted. When done well, customers are just neutral, but when done poorly, customers are very dissatisfied.

'Must-Be Quality' is the minimum required quality level in order to enter the market. In the competitive market, companies try to maintain a competitive edge by developing function and service for the 'Attractive Quality'. 'Reverse Quality' refers to a high degree of achievement resulting in dissatisfaction. For example, some customers prefer the basic model of a product and will be dissatisfied if a product has too many extra features.

When we analyze survey response using Kano model, many quality attributes can be resulted in 'Indifferent Factor'. This implies that some attributes which are meaningful tend to be classified as indifferent attributes for the customer satisfaction. In order to tackle this problem, a modified Kano model is proposed by reducing the indifferent factor. The modified Kano model can be robust for the survey response. A survey is performed for the quality attributes of smart phone [7].

Companies need to understand customer's feelings about product and service in order to improve customer's satisfaction about product and service using Kano model. For this purpose, Walden proposed two dimensional evaluation table of Kano model [4]. In this model, after positive and negative question is asked for the same function, the answer is applied to the format in Table 1. Then, the customer perception about the product function and service can be identified. For example, if the customer response is good for the provided function, but the response is neutral even the function is not supported, that function can be classified as 'Attractive Quality'.

TABLE 1. Two dimensional evaluation table of Kano model

Customer Requirements		DYSFUNCTIONAL Question				
		1. Like	2. Expect	3. Neutral	4. Accept	5. Dislike
FUNCTIONAL Question	1. Like	Q	A	A	A	O
	2. Expect	R	I	I	I	M
	3. Neutral	R	I	I	I	M
	4. Accept	R	I	I	I	M
	5. Dislike	R	R	R	R	Q

✳ A: Attractive, M: Must-Be, R: Reverse, O: One Dimensional, I: Indifferent, Q: Questionable Result

The limitation of Kano model is its sensitivity of deciding one of five factors. Sometimes, we want to know how much degree the customer is satisfied or dissatisfied about the product or service. This helps us how to increase the customer's satisfaction or resolve the customer's dissatisfaction. To tackle this issue, Timko proposed satisfaction coefficient and dissatisfaction coefficient [8]. Satisfaction coefficient implies the ratio of satisfied customers when a specific quality element of a product or service is provided to the customers. Dissatisfaction coefficient implies the ratio of dissatisfied customers when a specific quality element of a product or service is not included in the product or system. This paper adopted Timko's coefficient in order to evaluate Kano's quality characteristics based on multiple questionnaire.

3. **Kano Quality Factor Determination Framework.**

3.1. **User review valuation method.** In order to determine Kano quality characteristics based on user review, it is most important to choose valuable review among many

reviews. Among many reviews in the Internet and SNS, if the professional reviewer's comments are excluded, real customer's review refers only satisfied or unsatisfied function and service. As customers do not evaluate the overall functions and services, the original Timko coefficient is not well adopted for the Kano quality evaluation. Thus, it is highly required to sort the reviews which is worth evaluation. This paper proposes a method to sort valuable reviews which are worth evaluation. Empirical research is conducted based on the method.

Method: The consistency of Meta Data and Emotion Analysis results.

In this method, customers are considered to write reviews consistently where the star points and Emotion Analysis results coincide each other. However, it is not guaranteed the customer has written review for the whole function and service. In order to overcome the limitation, a measure is adopted by multiplying number of sentences in the reviews by emotional points in each sentence. Only reviews which exceed some limit are considered as valuable reviews to evaluate.

3.2. **Kano mapping model & application.** The mapping and analysis process using Kano and mining technique are as the following.

Step 1. Among the reviews which are gathered through Web crawling, valuable reviews are selected using the method.

Step 2. For the chosen reviews, main topic is analyzed based on 'Topic Modeling' method.

Step 3. Using morphological analysis, each sentence is tagged grammatically.

Step 4. After adverb is extracted, and the sentence is evaluated whether it is positive or negative using 'Emotion Analysis'. For example, when the customers are '*highly*' satisfied with the function 'A', and they are neutral for the deficit of function 'A', it can be classified as 'Attractive Quality'. The framework of Kano mapping model is shown in Figure 2.
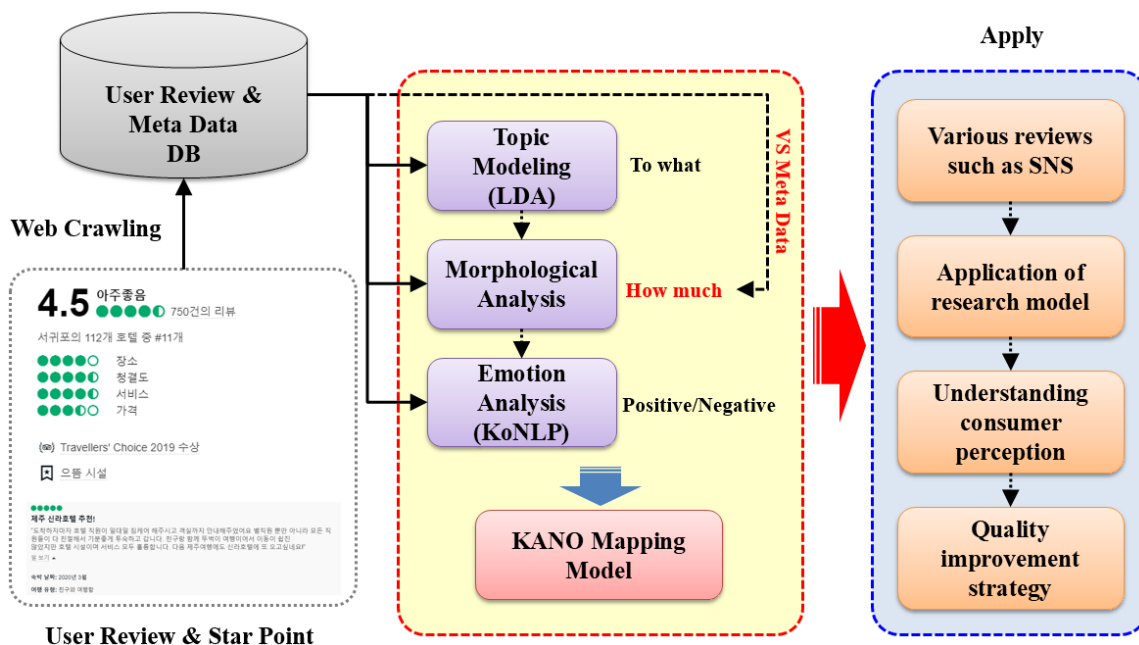

FIGURE 2. Kano mapping model

4. **Implementation.** Following the above-mentioned Kano mapping model, hotel review data are collected from TripAdvisor (www.tripadvisor.com). A web crawling program is developed to collect hotel review data based on Python. The collected review data are saved in the MySQL server. In order to evaluate the validity of collected data, hotel

TABLE 2. Review frequency ratio by meta data

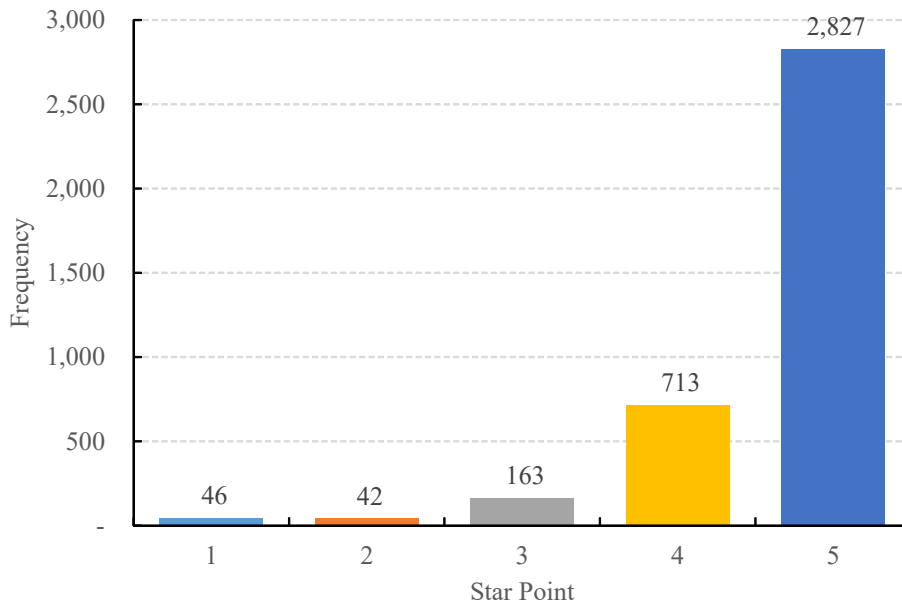| Star Point	Frequency	Ratio
1	46	1.2%
2	42	1.1%
3	163	4.3%
4	713	18.8%
5	2,827	74.6%
Total	3,791	100.0%

FIGURE 3. Meta data histogram

TABLE 3. Analysis of emotion and consistency with star ratings

Emotion Analysis	Star Point					Total
	1	2	3	4	5	
NA	12	14	56	202	708	992
Positive	3	5	41	277	1,206	1,532
Neutral	7	6	46	193	835	1,087
Negative	24	17	20	41	78	180
Total	46	42	163	713	2,827	3,791

review data of 3,791 reviews are collected. The distribution of review data is summarized in Table 2 and Figure 3. It is shown that positive reviews with five stars (*****) occupy 74.6%, four stars with 18.8%, in which the total of five and four stars reaches 93.4%.

Next step is to compare the results of Meta Data and Emotion Analysis in order to sort a meaningful review data. For the Emotion Analysis, emotion dictionary is adopted from a previous research [9]. The result from Emotion Analysis is designed to be either 'Positive', 'Neutral', or 'Negative'. The review with five stars in the Meta Data tends to be 'Positive' from the Emotion Analysis with high probability. The review with one star in the Meta Data tends to be 'Negative' from the Emotion Analysis with high probability. The missing values in the Emotion Analysis are proved to be incorrect sentence resulted from translation process. As some original reviews were written in foreign language, they were translated into Korean language using automatic translator.

Further analysis is conducted using Emotion Analysis. For two groups in which Emotion Analysis is positive and star points are level 4 and 5, and Emotion Analysis is negative and star points are level 1 and 2, the Morpheme Analysis is conducted. This paper adopted KoNLP for Morpheme Analysis [10]. Morpheme is classified as nine Taggers, among which high frequency noun is chosen, then synonym analysis is performed and the result is summarized in Figure 4.

(a) Positive reviews

	noun	n
	<chr>	<int>
1	Hotel	837
2	Clean	515
3	Ulsan	454
4	Employees	365
5	Kindness	279
6	Guest Room	277
7	Location*	238
8	Usage	195
9	Visit	193
10	Service	193
11	Next	188
12	Facility	186
13	Satisfaction	175
14	Breakfast	174
15	Price*	159

VS

(b) Negative reviews

	noun	n
	<chr>	<int>
1	Hotel	39
2	Food	37
3	Guest Room	10
4	Clean	10
5	Service	9
6	Employees	8
7	Kindness	8
8	Facility	7
9	Bathroom	5
10	Internet*	5

(c) Derived quality factor

Quality Factor	Noun
One-Dimensional Quality	Clean Kindness Service Breakfast
Must-be Quality	Internet
Attractive Quality	price

FIGURE 4. Topic modeling: (a) includes positive & star points 4 and 5; (b) includes negative & star points 1 and 2; (c) indicates derived quality factor from the mapping model

Based on the 'Topic Modeling', customer satisfies with good 'proximity' (location), but neutral for bad proximity, this can be classified as 'Attractive Quality'. In case of 'Ulsan' city in Korea, as Ulsan is industrial complex rather than tourist place, the needs of business tour are high. Location proximity seems to be evaluated as 'Attractive Quality'. For the 'Price' factor, though customers satisfy with low price, they are neutral for high price, this can be also 'Attractive Quality'. For the 'Internet Service' factor, they are highly dissatisfied with no service, but they are neutral with Internet service, this can be 'Must-Be Quality'.

"A picture is worth a thousand words" is a well-known English adage. Complex and sometimes multiple ideas can be conveyed by a single still image. Thus, data visualizations give businesses a valuable way to communicate important information at a glance. However, if your raw data is text-based, what can you do? 'Word cloud analysis' is a good solution. The results from Topic Modeling can be visualized using 'Word cloud analysis'. The result of positive review case (Figure 4(a)) and negative review case (Figure 4(b)) can be displayed as Figure 5. We can catch the underlying review comments in the picture at a glance.

5. **Conclusions.** In this research, a new method is presented for identifying customer perception about product function or service based on the text mining analysis of customer review. Some meaningful results are found which we can classify hotel service factor into five different factors of Kano quality model.

It was found that the customer review was mainly focused on his/her personal experience for the service rather than an overall review or evaluation for the product or service. This resulted in the limitation of quantification of 'Topic Modeling' method. In order

(a) Word cloud analysis from positive sentiment

(b) Word cloud analysis from negative sentiment

FIGURE 5. Visualization from 'Word cloud analysis' (This is derived from positive sentiment and negative sentiment corresponding to Figure 4.)

to overcome this limitation, another method of choosing customer reviews with valuable results can be devised and the selected review data would be applied for this model.

REFERENCES

[1] N. Kano, N. Seraku, F. Takahashi and S. Tsuji, Attractive quality and must-be quality, *The Journal of the Japanese Society for Quality Control*, vol.14, no.2, pp.39-48, 1984.

[2] Q. L. Meng and J. Dong, Future direction and visual analysis of Kano model: A literature review, *Journal of Service Science and Management*, vol.11, pp.399-413, 2018.

[3] H. Song, C. Chen and Q. Yu, Research on Kano model based on online comment data mining, *IEEE 3rd International Conference on Big Data Analysis (ICBDA)*, 2018.

[4] D. Walden, Kano's methods for understanding customer-defined quality, *Center for Quality of Management Journal*, vol.2, no.4, pp.13-14, 1993.

[5] H. Min, J. Yun and Y. Geum, Analyzing dynamic change in customer requirements: An approach using review-based Kano analysis, *Sustainability*, vol.10, no.3, https://doi.org/10.3390/su10030746, 2018.

[6] J. Ahn and K. Kim, Extracting keyword-related affective words using the text mining technique, *ICIC Express Letters*, vol.12, no.12, pp.1249-1257, 2018.

[7] T. Kim, Analysis of quality characteristics of smart phone using modified Kano model, *Journal of the Society of Korea Industrial and Systems Engineering*, vol.35, no.1, pp.57-65, 2012.

[8] M. Timko, An experiment in continuous analysis: Kano's methods for understanding customer-defined quality, *Center for Quality of Management Journal*, vol.2, no.4, pp.17-20, 1993.

[9] *DILAB (Data Mining & Artificial Intelligence Laboratory)*, https://github.com/park1200656/Knu SentiLex, Accesses in July, 2020.

[10] *KoNLP (Korean Natural Language Processing)*, https://cran.r-project.org/web/packages/KoNLP/index.html, Accesses in July, 2020.