*ICIC Express Letters*
*Part B: Applications*
Volume **12**, Number **8**, August **2021**

ICIC International ©2021 ISSN 2185-2766
pp. 707–714

# REVIEW-BASED CONTROL CHARTS FOR SERVICE QUALITY MONITORING: A BRIEF REVIEW AND FUTURE DIRECTIONS

Suah Kim, Sohyun Park and Minjung Kwak*

Department of Industrial and Information Systems Engineering
Soongsil University
369 Sangdo-ro, Dongjak-gu, Seoul 06978, Korea
{ suak210; sohyunpark }@soongsil.ac.kr; *Corresponding author: mkwak@ssu.ac.kr

Received December 2020; accepted February 2021

Abstract. *Review-based control charts that combine online review mining with traditional control charts are emerging as a new tool for quality monitoring. Although previous studies have provided successful examples of the review-based control chart, the natural characteristics of user-generated reviews and their impact on the chart have received less attention so far. After briefly surveying existing approaches, this paper presents a case study of two hotels in Korea to demonstrate the potential issues and challenges faced by review-based control charts because of the characteristics of online customer reviews. The results show that a small, inconsistent number of reviews and their uncontrollable contents poses the main challenge for the review-based control chart. Future research directions are discussed based on the results.*

**Keywords:** Service quality, Statistical process control, Control chart, Online review, Opinion mining, Sentiment analysis, Text mining, Hospitality management

1. **Introduction.** In today's competitive environment, the service quality perceived by customers and the resulting customer satisfaction are critical success factors of a firm [1]. Therefore, continuous efforts have been made in academia and industries to develop more advanced tools for service monitoring and management. Online review mining is one such tool that is receiving increasing interest in various service industries. It analyzes online customer reviews using natural language processing and extracts peoples' sentiments and opinions from written text [2,3]. Through online review mining, firms can better understand customer requirements, identify the service attributes that are important for customer satisfaction, and determine whether their customers are satisfied with certain service attributes [3-6].

Combining online review mining with traditional control charts is a recent attempt to visualize and monitor the perceived service quality [7-11]. In this technique, online review mining first measures the service quality based on the customers' sentiment toward various service attributes; then, a control chart displays the results (usually the number or the proportion of complaints in a given period) in time order. Such review-based control charts provide valuable information about the average service quality and variation in its level, help monitor the service quality over time, and highlight any need for improvement. They can complement traditional survey-based monitoring tools (e.g., SERVQUAL) that require considerable time and money [9].

Starting with Lo [8], previous studies have suggested various versions of review-based control charts adopting different combinations of review mining and control chart techniques [8-11]. They have shown that the review-based control chart can be a promising solution for service monitoring. However, the natural characteristics of user-generated reviews and their impact on the application and operation of control charts have received

DOI: 10.24507/icicelb.12.08.707

less attention so far. This study aims to investigate and highlight the potential issues regarding the review characteristics. After a brief survey of existing review-based control charts, this paper presents an empirical case study of two hotels in Seoul, Korea to demonstrate the problems that may arise when employing the review-based control chart in the real world. The case study illustrates that a small, inconsistent number of reviews and their uncontrollable contents pose the main challenge for review-based control charts.

The rest of the paper is organized as follows. Section 2 reviews recent literature. Section 3 presents the case study and demonstrates the potential and limitations of existing approaches. Section 4 concludes the paper with future research directions.

2. **Literature Review.** A control chart is a statistical process control tool to detect whether a manufacturing or business process is in a state of control. Control charts have been widely used in a variety of fields for quality monitoring, and many studies have been conducted to improve control charts and their applications (e.g., [12,13]). Recently, several studies have proposed review-based control charts that create a control chart (usually a P- or an exponentially weighted moving average (EWMA) chart) based on the results of online review mining. Table 1 summarizes the recent studies on review-based control charts including this paper.

Table 1. Summary of the research on review-based control chart

| Ref. | Data source | Monitoring level | Monitoring cycle | Review mining | Sentiment analysis | Monitoring measure | Chart type |
|---|---|---|---|---|---|---|---|
| [8] | Customer messages | Service (as a whole) | Week | SVM | No | Rate of complaints | P-chart |
| [9] | Complaints | Topic (attribute) | Week | LSA | No | Eigenvector component value | EWMA chart |
| [10] | Complaints | Topic | Week | LSA | No | Rate of complaints | P-chart |
| [11] | Reviews | Service | Day | SRJST | Positive Negative | Topic-sentiment distribution | SRJST chart |
| This paper | Reviews | Topic | Month | Lexicon-based analysis | Positive Neutral Negative | Rate of complaints | P-chart, EWMA chart |

Lo [8], who first attempted a review-based control chart, classified customer messages into complaints and non-complaints using support vector machine (SVM) and plotted a P-chart displaying the rate of complaints. The P-charts issue a warning signal if the rate of complaints exceeds the usual level.

Ashton et al. [9,10] combined the EWMA and P-charts with the latent semantic analysis (LSA) technique. They collected complaints from customers who requested a refund and identified several complaint topics using LSA. For each topic, they suggested an EWMA chart plotting the topic's eigenvector component value [9] or a P-chart displaying the referring proportion of the topic [10]. An underlying assumption of these works is that the customer comments are all negative. Sentiment classification was not incorporated.

Liang and Wang [11] proposed the sequential reverse joint sentiment-topic (SRJST) chart, a new control chart that can detect shifts in topic-sentiment combinations in a corpus (a group of reviews). A review mining method SRJST was proposed to estimate the joint topic-sentiment distribution per day. The charting statistics monitor the estimated daily joint distribution and assess if it is similar to the expected reference distribution.

Previous studies have shown that the review-based control chart can be an effective method for service monitoring. However, they have focused on proposing charting methods and paid relatively less attention to determining the effect of the online reviews' natural characteristics on the application of review-based control charts. This paper presents a case study of two hotels in Korea to demonstrate the potential issues and challenges

due to the nature of online customer reviews. With the case study, this paper details the achievements and limitations of existing approaches and suggests future research directions.

3. **Case Illustration.** This section presents an empirical study of two hotels in Seoul, Korea: Hotels A and B. Online customer reviews written in Korean from January 2018 to April 2020 were crawled from the website of TripAdvisor.com and used for the case study.

3.1. **Methodology.** Creating a review-based control chart is a two-step approach: 1) mining online customer reviews, and 2) plotting a control chart. Figure 1 shows the overall flow of the approach. In this study, review mining was conducted following the lexicon-based method from the authors' prior research [4,5]. For a predefined set of service attributes (furniture and appliances, bathroom condition, room condition, dining, service, facilities, conveniences, and surroundings; eight attributes in total), the method identifies whether a review refers to a certain attribute, and if so, what is the sentiment for the attribute.

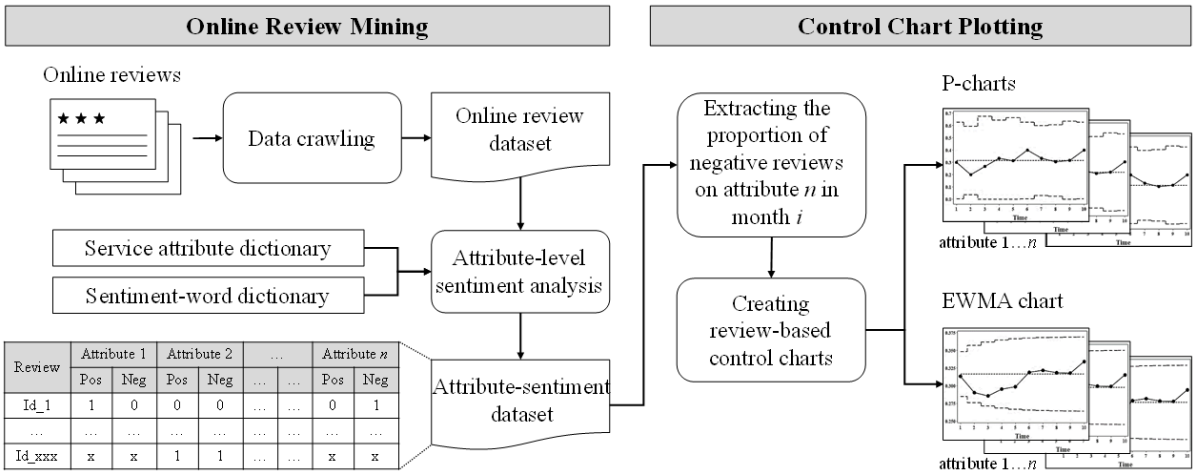



FIGURE 1. Creating a review-based control chart at an attribute level

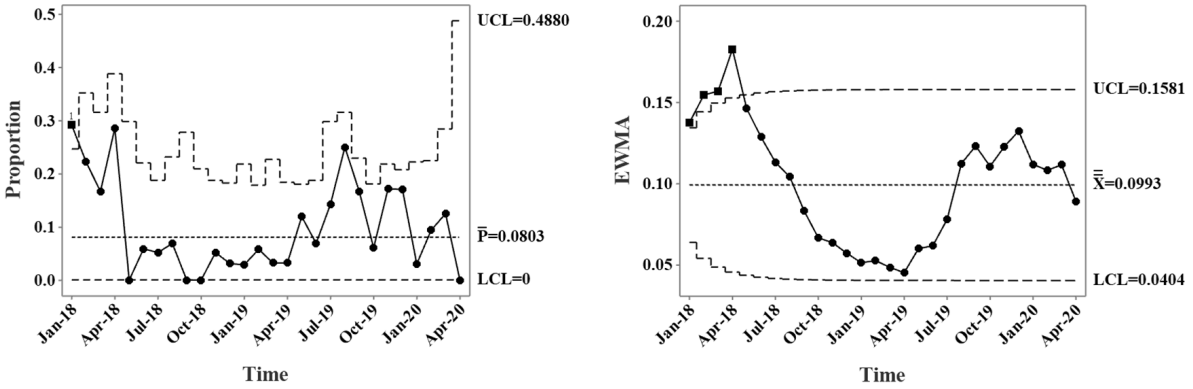
The results from the review mining are visualized using P- and EWMA charts. A P-chart is a control chart for the fraction nonconforming. In this study, the fraction nonconforming represents the rate of complaints for a given service attribute in a sample. In this study, the fraction nonconforming is calculated using Equation (1). Here, $p_{ni}$ denotes the rate of complaints for attribute $n$ in the month $i$. The sample size (denominator) is the number of reviews in a month and thus, varies from time to time.

$$p_{ni} = \frac{\#\text{ of reviews with negative sentiments for attribute } n \text{ in month } i}{\#\text{ of reviews posted in month } i} \tag{1}$$

The EWMA chart is a time-weighted control chart that considers not only the current sample observation but also the entire previous samples. It charts the EWMA of all sample observations. A constant $\lambda$ ($0 < \lambda \leq 1$) determines the weight given to each point. The EWMA chart is more effective than the P-chart in detecting small process shifts. In this study, the charting value $x_{n,i}$ denotes the EWMA of $p_{ni}$s, and is calculated as shown in Equation (2).

$$x_{n,i} = \lambda p_{ni} + (1 - \lambda) x_{n,i-1} \tag{2}$$

3.2. **Example of successful application.** Hotel A illustrates how review-based control charts can effectively assist in service monitoring. Figure 2 shows some of the results from Hotel A. A total of 971 reviews were collected and classified by month. The average number of reviews per month was 34.




(a) P-chart (room condition) (b) EWMA chart (room condition)

FIGURE 2. Review-based control charts for monitoring the service quality of Hotel A

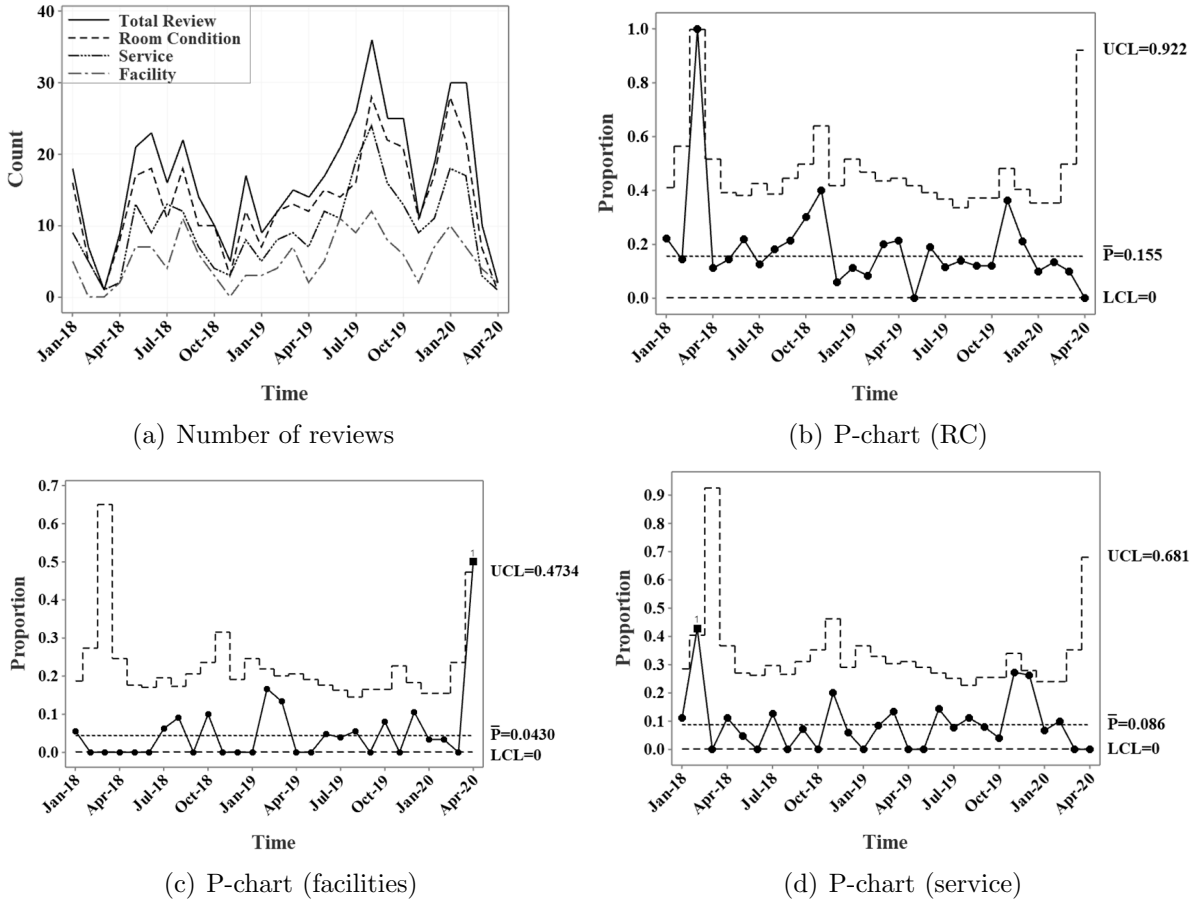
Figures 2(a) and 2(b) show the P- and EWMA charts, respectively, for the room condition of Hotel A. They provide an effective overview of the service quality at the attribute level. The average performance and its trend over time are easily observable. There was one out-of-control point in Figure 2(a) in January 2018, and there were four out-of-control points in Figure 2(b) from January to April 2018. The results imply that there might have been a special cause for the negative feedback from the hotel guests on the room condition. Hotel managers can conduct a detailed analysis to determine the reasons for the feedback and take necessary actions. The detailed analysis of the months showed that most of the complaints had come from the "oldness" of the room.

3.3. **Example of difficulties.** In contrast to the example of Hotel A, the case study of Hotel B reveals some difficulties in applying review-based control charts that are especially due to the uncontrollable quantity and quality of online customer reviews. Figure 3 shows some of the case study results of Hotel B: the review count and the P-charts.

Figure 3(a) displays the total number of reviews per month (black solid line) and the number of reviews mentioning a particular attribute per month (dotted lines). Here, lines for three service attributes, i.e., room condition (RC), service, and facility, were illustrated as an example. Figure 3(a) exhibits the natural characteristics of online customer reviews, i.e., inconsistent quantity and quality. The number of reviews greatly varies, and more importantly, there is a chance that an extremely small number of reviews are posted. For instance, Hotel B has one review in March 2018 and two reviews in April 2020.

In terms of quality, the review content cannot be controlled and customers usually do not mention every attribute in their reviews. Only the attributes they think important or worth sharing the information about tend to be referred. Accordingly, the number of reviews mentioning a particular attribute is likely to decrease from the total number of reviews, which means an increased probability of only few or zero reviews per month.

Such a small number of reviews complicates the application of control charts in several aspects. First, the performance of the control limit is not guaranteed under the small number of reviews. A small sample size increases the upper control limit and the control charts become less reliable. According to [14], the rule of thumb for the minimum sample size of the P-chart is at least 0.5 divided by the mean $p_{ni}$, i.e., 4 (RC) to 12 (facilities) in this case.




(a) Number of reviews

(b) P-chart (RC)

(c) P-chart (facilities)

(d) P-chart (service)

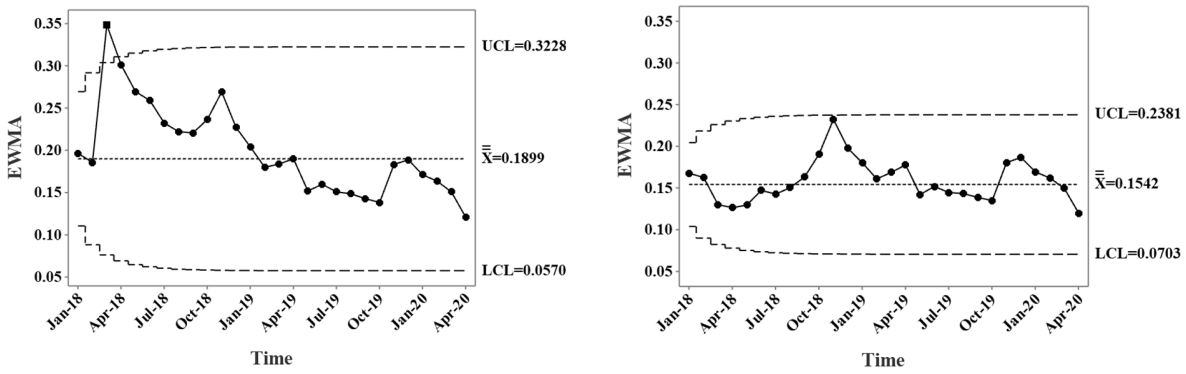
FIGURE 3. Review count and the P-chart examples of Hotel B

Second, when the $p_{ni}$ is determined by only a few review(s), the relative weight given to a single review increases in the $p_{ni}$ calculation. This can be a serious problem, considering the possibility that the single review is an aberration (e.g., false review).

Third, even if the absolute difference in the number of reviews is small, the difference in $p_{ni}$ values can be exaggerated, resulting in misconception about the service quality. Taking the P-chart in Figure 3(c) as an example, the numbers of negative reviews in March 2018 and April 2020 were 0 (out of 1 review) and 1 (out of 2 reviews), respectively, and their $p_{ni}$ values were 0 and 0.5, respectively. Given those $p_{ni}$ values, one may conclude that the service quality in March 2018 was much better than in April 2020. However, considering that the actual difference between the two time points was just one negative review, the appropriateness of the evaluation is questionable.

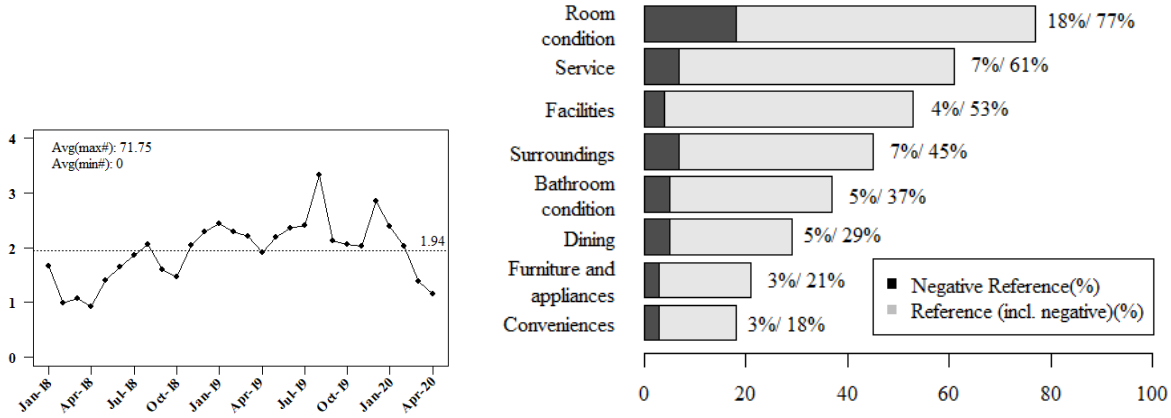
Fourth, the charting value $p_{ni}$ also raises a concern regarding the result interpretation. Figures 3(b)-3(d) show that the average $p_{ni}$s of room condition, service, and facilities are 0.155, 0.086, and 0.043, respectively. One may conclude that the performance of facilities is better than that of service or room condition. However, this can be a misjudgment. As shown in Equation (1), the sample size (i.e., denominator) is the number of entire reviews in a month, and thus, it is the same for all attributes; but as some attributes are more important to customers, a negative reference of those (i.e., numerator) tends to increase naturally in attribute importance. In other words, a smaller $p_{ni}$ does not necessarily mean fewer complaints or better performance. It can also indicate lower attribute importance.

The EWMA chart also has the same issues as the P-chart. A small sample size and the risk of being affected by aberrations are more critical in the EWMA chart, as one review can influence other charting values afterward. This means that the entire chart can be easily swayed by a single review. Figures 4(a) and 4(b) provide an example. Figure




(a) EWMA chart (RC, $p_{\text{mar-18}} = 1$) (b) EWMA chart (RC, $p_{\text{mar-18}} = 0$)

FIGURE 4. EWMA charts for the room condition (RC) of Hotel B




(a) Avg. number of reviews (b) Avg. % of negative reference/reference

FIGURE 5. The quantity and quality characteristics of the online reviews in hotel industry

4(a) shows the EWMA chart when the only review in March 2018 was a negative one. By contrast, Figure 4(b) assumes that the review was a positive one. There is only one difference, but great differences in control limits and charting values are observable.

Similarly, the EWMA chart needs careful consideration when dealing with missing values. A rule for substituting the missing values should be chosen carefully as the rule affects not only a single charting value but also the entire chart and its performance.

One may suspect that Hotel B is a special case experiencing the aforementioned problems. However, Figure 5 shows that the small, inconsistent number of reviews and uncontrollable, unequal reference to the service attributes are not rare in the real world. Figure 5(a) shows the average number of reviews per month for the 3-5 star hotels in Seoul, Busan, and Jeju, Korea (351 hotels in total). The average number of monthly reviews is 1.94; 85% of the hotels had only two or fewer reviews per month. Figure 5(b) shows the rate of the reviews mentioning each service attribute, and the rate of negative reviews for the attribute. The average rate of reference ranges from 18% to 77%; if considering only the negative reviews, the average rate drops to 3-18%. This implies that the sample size drops even more at the attribute level, and the aforementioned problems are expected to be common in other hotels as well.

4. **Discussion and Conclusions.** Review-based control charts are emerging as a new tool for service monitoring and improvement. However, applying a traditional control chart to user-generated reviews in its current form can entail various problems.

In the review-based control chart, it is almost impossible to control the number and content of the reviews. The fact that not all attributes appear in a review complicates the problem. The sample size drops at the attribute level, and missing values increase. As the impact of a single aberration increases and the control limits become questionable, the risks of false or no alarm increase. One may consider changing the sampling interval (e.g., from every month to every quarter) to get enough samples, but a loss of performance is unavoidable as a process shift cannot be detected on time. Therefore, unless a sufficiently large amount of good-quality reviews can be collected continuously, the performance and usefulness of the review-based control charts are likely to face challenges.

The interpretation of a P-chart also raises concerns. The $p_{ni}$ values of important attributes tend to be greater than those of other attributes, regardless of their actual performance. Such a difference of $p_{ni}$ can lead to a misjudgment of an attribute's quality.

The review-based control chart requires a new approach that can address the challenge of uncontrollable characteristics of customer reviews. A modified control chart should be developed in the future. Adopting alternative quality measures, such as the sentiment score (intensity of sentiment), the ratio between positive and negative reviews, or the interval between complaints, can be an option worth considering. Changing the chart type according to the new measures and setting the control limits in different ways can also be considered.

**Acknowledgment.** This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (Ministry of Science and ICT, MSIT) (No. NRF-2019R1F1A1041099).

## REFERENCES

[1] M. I. Rapert and B. M. Wren, Service quality as a competitive opportunity, *J. Serv. Mark.*, vol.12, no.3, pp.223-235, 1998.

[2] E. Boon, M. Bonera and A. Bigi, Measuring hotel service quality from online consumer reviews: A proposed method, *Information & Communication Technologies in Tourism*, pp.367-379, 2013.

[3] G. Park and M. Kwak, The life cycle of online smartphone reviews: Investigating dynamic change in customer opinion using sentiment analysis, *ICIC Express Letters, Part B: Applications*, vol.11, no.5, pp.509-516, 2020.

[4] J. Choi, S. Park, S. Kim, G. Lee and M. Kwak, Key attributes of hotel service quality in Korea: An analysis based on online review mining, *Journal of Tourism Management Research*, vol.84, no.5, pp.1069-1093, 2018.

[5] M. Kwak, J. Choi and S. Park, Sentiment analysis of online reviews for evaluating customer satisfaction on hotel service attributes, *Journal of Tourism Management Research*, vol.90, no.4, 2019.

[6] S. Park, M. Kwak and J. Choi, Comparing competitive advantages of hotel services using opinion mining of online customer reviews: A study focusing on 5-star hotels in Seoul, Korea, *ICIC Express Letters, Part B: Applications*, vol.11, no.8, pp.743-749, 2020.

[7] R. G. Aykroyd, V. Leiva and F. Ruggeri, Recent developments of control charts, identification of big data sources and future trends of current research, *Technol. Forecast. Soc. Change.*, vol.144, pp.221-232, 2019.

[8] S. Lo, Web service quality control based on text mining using support vector machine, *Expert Syst. Appl.*, vol.34, no.1, pp.603-610, 2008.

[9] T. Ashton, N. Evangelopoulos and V. R. Prybutok, Exponentially weighted moving average control charts for monitoring customer service quality comments, *Int. J. Serv. Stand.*, vol.8, no.3, pp.230-246, 2013.

[10] T. Ashton, N. Evangelopoulos and V. R. Prybutok, Quantitative quality control from qualitative data: Control charts with latent semantic analysis, *Qual. Quant.*, vol.49, no.3, pp.1081-1099, 2015.

[11] Q. Liang and K. Wang, Monitoring of user-generated reviews via a sequential reverse joint sentiment-topic model, *Qual. Reliab. Eng. Int.*, vol.35, no.4, pp.1180-1199, 2019.

[12] K. S. Chen, T. C. Chang, K. J. Wang and C. T. Huang, Developing control charts in monitoring service quality based on the number of customer complaints, *Total Qual. Manag. Bus. Excell.*, vol.26, no.6, pp.675-689, 2015.

[13] S. F. Yang and T. A. Jiang, Service quality variation monitoring using the interquartile range control chart, *Qual. Technol. Quant. Manag.*, vol.16, no.5, pp.613-627, 2019.
[14] Minitab 19 Support, *Data Considerations for P Chart*, https://support.minitab.com/en-us/minitab/19/help-and-how-to/quality-and-process-improvement/control-charts/how-to/attributes-charts/p-chart/before-you-start/data-considerations/, 2020.