Doctoral Dissertation

# Development of customer review analytic methods for customer-centric service improvement

Juram Kim

Department of Management Engineering

Graduate School of UNIST

2020

# Development of customer review analytic methods for customer-centric service improvement

Juram Kim

Department of Management Engineering

Graduate School of UNIST

# Development of
# customer review analytic methods
# for customer-centric service improvement

A dissertation
submitted to the Graduate School of UNIST
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Juram Kim

06/21/2020 of submission

Approved by

Advisor
Chiehyeon Lim

# Development of
# customer review analytic methods
# for customer-centric service improvement

Juram Kim

This certifies that the dissertation of Juram Kim is approved.

06/21/2020

Signature

Advisor: Chiehyeon Lim

Signature

Han-Gyun Woo: Thesis Committee Member #1

Signature

Sungil Kim: Thesis Committee Member #2

Signature

임성훈

Sunghoon Lim: Thesis Committee Member #3

Signature

Youngjung Geum: Thesis Committee Member #4;

# Abstract

The concept of service engineering has emerged in terms of the design and development of innovative services, adopting more technical-methodological approaches and utilizing existing engineering know-how in the area of product development. In contrast to the extensive body of literature regarding the development of service engineering approaches for use in the pre-launch stages, relatively little attention has been paid to the post-launch stages. A major research gap remains in the literature as to how to facilitate service improvement from the perspectives of customers and value creation. Recognizing the value of customer reviews as a comprehensive source of the customers' voice, this dissertation develops models and methods for customer review analytics for customer-centric continuous service improvement. The dissertation comprises three issue papers, each concerning a critical issue in continuous service improvement.

The first issue paper focuses on the dynamics of services given that different improvement strategies should be implemented along with different life cycle stages. This study develops a stochastic service life cycle analysis to gauge where a service is in its life cycle and to provide forecasts about its future prospects. The number of customer reviews is used to customer-attention-based service maturity and a hidden Markov model is employed to estimate the probability of a service being at a certain stage in its life cycle. Based on this, we also develop three indicators to represent the future prospects of a service's life cycle progression. The main advantages of the proposed approach lie in its ability to model different shapes of life cycles without any supplementary information and to examine a wide range of services within acceptable limits of time and cost. Furthermore, a case study of the mobile game services in the Apple App Store is presented.

The second issue paper focuses on correcting the existing defects for continuous service improvement. This study develops an integrated approach of sentiment analysis and statistical process control to monitoring customer complaints and detecting potential service failures. The sentiment analysis enables the systematic identification of customer complaints from customer review data. Statistical process control facilitates the early detection of significant customer complaints that may lead to service failures. The integration of these methods enables the detection of customer complaints from customer reviews and prevents service failures that might occur if customer complaints are not resolved properly. Additionally, we presented a case study regarding the mobile game services in the Apple App Store is presented.

The third issue paper focuses on the establishment of long-term strategies for continuous service improvement. This study presents an approach to identifying competitors and benchmarks and developing service benchmarking guidelines based on customer perception. The proposed approach includes (1) topic modeling to identify service attributes from customer reviews; (2) index and

sentiment analysis to measure the importance of service attributes and the focal company's performance in the same attribute; (3) clustering and the technique for order of preference by similarity to ideal solution (TOPSIS) to select competitors and best practices as benchmarks based on customer perception; and finally (4) importance-performance competitor analysis to develop a strategic action plan. The proposed approach enables the quick identification of key aspects of the focal company's service and its competitors' services and assessment of the performance of the focal company and benchmarks within acceptable time and cost limits. A case study of hotel services in Bangkok confirms that the proposed approach is valuable as a complementary tool for customer-oriented service benchmarking.

The contributions and utilities of this dissertation are three-fold. First, from a theoretical perspective, this dissertation contributes to service engineering research by developing quantitative models and methods for use in the post-launch stage. Second, from a methodological perspective, the dissertation extends prior expert- and customer survey-approaches to data-driven approaches, thereby enabling the quick analysis of the focal company's and competitors' services. Finally, from a practical standpoint, we automated the proposed approaches to allow non-specialists who are unfamiliar with customer review data and complex algorithms to benefit from the analysis results. Although further testing of a wider range of services is required, the systematic processes and quantitative outcomes of the proposed approaches offer a substantial contribution to both current research and future practice and service as a starting point for the development of more generalized models.

# Contents

# List of Figures

# List of Tables

# 1 Introduction

## 1.1 Background and motivation

As innovation cycles become shorter and service systems become more complex, the systematic development and management of services have become more important as innovation cycles become shorter and service systems become more complex (Maglio and Spohrer, 2008). The concept of service engineering has emerged in this context to design and develop innovative services, adopting more technical-methodological approaches and utilizing existing engineering know-how in the area of product development (Bullinger et al., 2003). Service engineering facilitates the entire process of development and management of services ranging from service design in the pre-launch stages to service improvement in the post-launch stages.

A variety of models, methods, and tools for service engineering have been presented for different contexts including the identification of new service opportunities (Lee and Lee, 2015), the generation of new service ideas (Kuusisto et al., 2013), and the evaluation and selection of new service concepts (Lee et al., 2012). However, in contrast to the extensive body of literature regarding the development of service engineering approaches for use in pre-launch stages, relatively limited attention has been paid to post-launch stages. Although some studies have attempted to develop service engineering approaches for use in post-launch stages (Kim and Lee, 2017), a major research gap remains in existing literature as to how to facilitate service improvement processes from the perspective of customers and value creation (Wemmerlov, 1990).

Service improvement, which aims to correct existing defects while establishing long-term strategies to secure sustainable competitiveness (Schweitzer and Aurich, 2010) – is a crucial activity in the post-launch stage for sustainable businesses. Given the distinctive characteristics of services, the nature of service improvement differs from that of product improvement, as follows. First, the characteristics of services, namely intangibility and simultaneity, make it difficult to improve service quality since the results of services cannot be observed. Second, with respect to the first issue, service is a perspective on value creation and that value creation is best understood from the lens of the customer (Parasuraman et al., 1988; Edvardsson et al., 2005), since customers are active participants in service delivery processes. In this context, the term service quality has been used interchangeably with customer satisfaction (Johnston, 2005), and customer satisfaction has been considered as the key factor that should be considered in service improvement (Aguwa et al., 2012; Johnston, 2005). Finally, the modification of services is more frequent and relatively easier than that of products since services are more closely related to business processes than products.

Regarding the use of methods for service improvement, the dominant approaches rely on expert knowledge or experience and customer surveys (Chen et al., 2014; Park et al., 2015). While the results of previous studies have proved quite useful for different purposes, such expert-centric and survey-based approaches have become extremely time-consuming and labor-intensive (Groves, 2006). Moreover, the quality and reliability of analytical processes and results strongly depend on expert panels (e.g., expert selection and panel size) (Bi et al., 2019), survey questionnaires (e.g., contents, complexity, and length) and the willingness of respondents to participate (Groves, 2006). Although some prior studies have presented engineering-centric approaches to service improvement using quantitative data and scientific methods such as data envelopment analysis (DEA) (Lee and Kim, 2014) and multi-criteria decision-making methods (Fu et al., 2011; Min, 2010), the scope of these analyses and their potential implications were limited as they focused on financial or operational indicators from the perspective of service providers rather than that of customers (Shamma and Hassan, 2013)

These drawbacks have necessitated the development of new service engineering approaches to service improvement. Three main issues are central to this problem and need to be addressed. First, different service improvement strategies should be implemented along with different life cycle stages. At its core, this issue involves the dynamics of services, i.e., a service is introduced to a market, grows in popularity, and is then removed as demand drops. To perform meaningful analysis and derive recommendations for actions regarding service improvement, it would be useful to measure the dynamics of services and then take this as a basis for developing service improvement strategies. Second, in contrast to tangible products, services have distinctive features such as intangibility and heterogeneity, which create special aspects and issues for service improvement. It has been widely accepted that service is a perspective on value creation and that value creation is best understood from the lens of the customer based on the value in use (Edvardsson et al., 2005). However, prior studies have tended to draw from existing approaches in the product development domain from a provider-oriented perspective, and little effort has been made to incorporate customer-centric aspects when developing service improvement strategies. Hence, any approaches developed need to place customer perception and value creation in the center of service improvement, taking into accounts the distinctive characteristics of service (Wemmerlov, 1990). Finally, service improvement is not a one-time activity but a continuing process (Eccles and Durand, 1998). Some improvement activities such as service failure detection and recovery should be performed proactively. Moreover, some improvement activities to establish mid- and long-term plans may require the analysis and assessment of a wide range of competitors' services. This is especially important in recent business environments wherein the number of services (or competitors) is increasing dramatically, service systems are increasingly complex and customer needs are becoming diversely segmented. Therefore, it is crucial to secure the applicable quantitative data and develop a systematic framework, enabling the quick analysis of the key aspects of services to facilitate decision making with acceptable limits of time and cost.

Researchers have begun to utilize customer reviews as a source of the comprehensive voice of the customer (VoC) to capture customer attention, complaints, requirements, and satisfaction for the purpose of service improvement (Hu et al., 2019). There has been a significant increase in attempts to develop models and methods for purpose-specific customer review analytics. While previous studies have proved quite useful for identifying service attributes and measuring service performance, they are subject to certain limitations, as follows. First, previous studies do not provide implications on the types of service improvement strategies to be implemented. In order to be able to perform meaningful analysis and derive recommendations for actions regarding service improvement, it is useful to measure the maturity and/or dynamics of services and take them as a basis for developing service improvement strategies. Second, previous studies are limited to retrospective approaches that develop service improvement strategies based on past performance and consequences at a certain point in time. Lastly, previous studies are limited to examining financial and operational aspects in the benchmarking process especially when selecting competitors (or best practices) as benchmarks.

To tackle these problems, this dissertation constructs a framework for continuous service improvement. The proposed framework enables to understand the current life cycle stages of a service in its life cycle, defects that must be fixed for short-term improvement, and strategies should be established for long-term improvement. It is designed to be executed in three discrete steps: understanding necessary improvement strategies; correcting existing defects; and establishing long-term improvement strategies. To support each step, three research issues were defined as follows.

The first issue is concerned with the dynamics of services given that different service improvement strategies should be offered along with different life cycle stages. Despite the apparent potential of life cycle analysis for application as a systematic tool for use in the service sector, the analytical framework and procedures for service life cycle analysis have rarely been discussed in previous research. Although some studies have presented the concept of service life cycles, the implications of these studies cannot be directly linked to service improvement, since they are designed to support the planning of new service development based on case-specific conceptual frameworks. This dissertation extends previous expert-centric conceptual service life cycle analysis for use in the pre-launch stage to customer-centric data-driven service life cycle analysis for use in the post-launch stage.

The second issue is associated with service failure monitoring as a tool to correct existing defects for service improvement. Recognizing the utility of statistical process control (SPC) in service sectors, several attempts have been made in previous research to integrate SPC into service quality management (Wood, 1994). However, the scope of analysis and potential implications were limited in that they only focused on monitoring and managing the physical aspects of service operations, such as waiting and response time, from the perspective of service providers rather than that of customers. This dissertation extends previous provider-oriented static assessments (i.e., diagnosing and assessing service quality at a certain point in time) to customer-oriented dynamic assessments (i.e., monitoring service failures based on customer complaints).

The third issue deals with service benchmarking as a tool for establishing long-term service improvement strategies. Previous survey-based approaches might miss unexpected but important service attributes that are difficult to identify in the design of survey questionnaires. Furthermore, the quality and reliability of the analysis results strongly depend on the contents, complexity, and length of survey questionnaires as well as the willingness of respondents to participate (Groves, 2006). Although customer review-based approaches have enabled the identification of service attributes to be considered and the evaluation of service performance, they are limited to only examining the financial and operational aspects in the benchmarking process. This is especially true for the selection of competitors (or best practices) as benchmarks. This dissertation extends previous financial and operational indicator-based approaches to customer-centric approaches.

These drawbacks provide the underlying motivation of individual issues papers and are fully addressed in this dissertation. The systematic processes and quantitative outcomes of the proposed approaches developed in this dissertation are expected to offer a substantial contribution to both current research and future practice. Research background

## 1.2 Purpose

This dissertation aims to develop models and methods for the purpose of customer review analytics for customer-centric continuous service improvement. The dissertation consists of three issue papers, each of which is concerned with a critical problem for continuous service improvement, as shown in Figure 1.

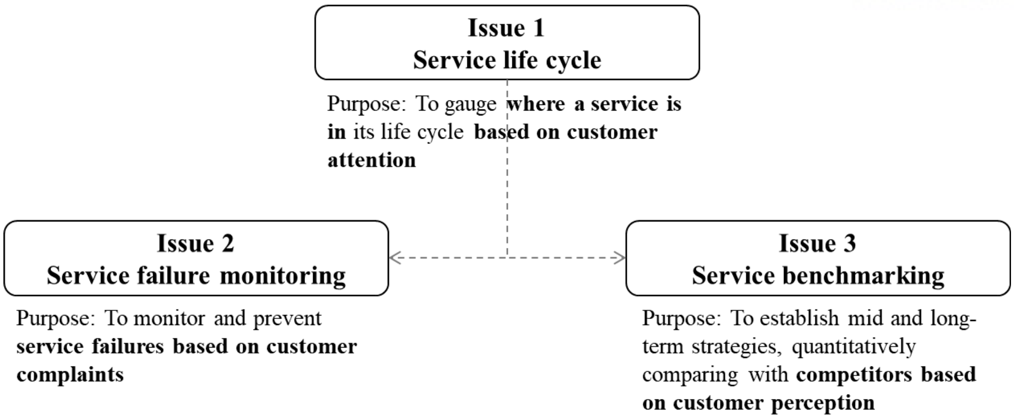



Figure 1. Research issues in this dissertation

The first paper is concerned with the dynamics of services given that different service improvement strategies should be established along with different life cycle stages. This study proposes a stochastic service life cycle analysis to gauge where a service is in its life cycle and to give forecasts about its future prospects. The number of customer reviews is employed to measure customer attention-based service maturity and a hidden Markov model (HMM) is used to estimate the probability of a service being at a certain stage of its life cycle progression. The main advantage of the proposed approach lies in its ability to model different shapes of life cycles without any supplementary information and to examine a wide range of services within acceptable time and cost limits.

The second paper is associated with service failure monitoring as a tool for correcting existing defects for service improvement. This study develops an integrated approach of sentiment analysis and SPC analysis to monitor customer complaints and to detect potential service failures. The sentiment analysis enables systematic identification of the level of customer satisfaction from customer review data. The SPC allows early detection of significant customer complaints and prevents the relevant service failures. The integration of these methods enables the detection of customer complaints from customer reviews and prevents potential service failures that may arise from the improper resolution of customer complaints.

The third paper deals with service benchmarking as a tool for establishing long-term service improvement strategies. This study presents an approach to identifying competitors and benchmarks and developing service benchmarking guidelines based on customer perception. The proposed approach includes (1) topic modeling to identify service attributes from customer reviews; (2) index and sentiment analysis to measure a service attribute's importance and the focal company's performance in the same attribute; (3) clustering and TOPSIS to select competitors and best practices as benchmarks from the perspective of customers; and finally (4) importance-performance competitor analysis to develop a strategic action plan. The proposed approach enables the quick identification of THE key aspects of the focal company's and competitors' services and assessment of the performance of the focal company and benchmarks within acceptable time and cost limits.

### 1.3 Scope and framework

This dissertation comprises three issue papers regarding customer-centric continuous service improvement in the post-launch stage. The theoretical and methodological scope of this dissertation is shown in Figure 2. From a theoretical perspective, given that different service improvement strategies should be implemented along with different life cycle stages, the first paper on service life cycle analysis guides organizations toward understanding necessary service improvement strategies. The second and third papers develop methods for continuous service improvement, focusing on the maturity stage of service life cycles. Specifically, the second paper develops a method for service failure monitoring to understand existing defects for service improvement, and the third paper develops a method for service benchmarking to aid long-term strategy development for service improvement.

From a methodological perspective, recognizing the value of customer reviews as a source of the comprehensive VoC, this dissertation extends previous provider-centric qualitative approaches to customer-centric quantitative approaches. Moreover, the dissertation adopts and integrates various approaches, such as natural language processing, machine learning, multi-criteria decision making, and information visualization to facilitate continuous service improvement within acceptable time and cost limits. We emphasize the systematic process of the developed approach in terms of inputs, throughputs, and outputs, allowing those unfamiliar with complex algorithms to benefit from the research results.

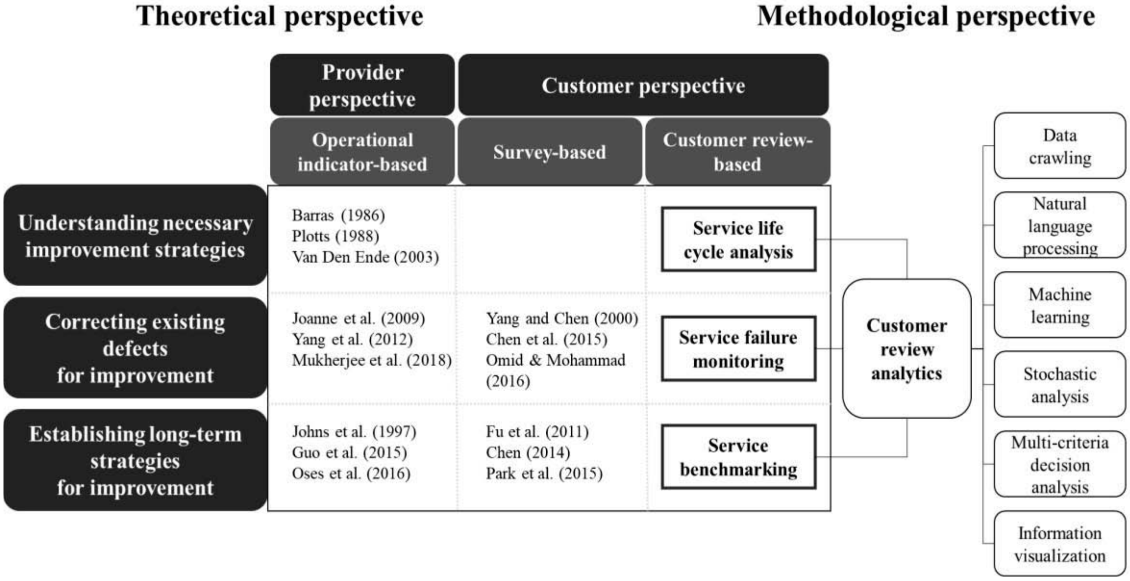



Figure 2. Theoretical and methodological scope of the dissertation

## 1.4 Outline

The remainder of this dissertation is organized as follows: Chapter 2 presents the research background regarding service improvement and customer review analytics. This chapter summarizes the results of previous studies and the differences between previous methods and the proposed approaches. Chapters 3, 4, and 5 form the main body of this dissertation as shown in Figure 3. Each chapter includes the introduction section, the background section, and summary and discussions section to clarify the necessity, motivation, and implications of the issue papers. Finally, Chapter 6 concludes with the contribution and limitations of this dissertation and suggests direction for future research.

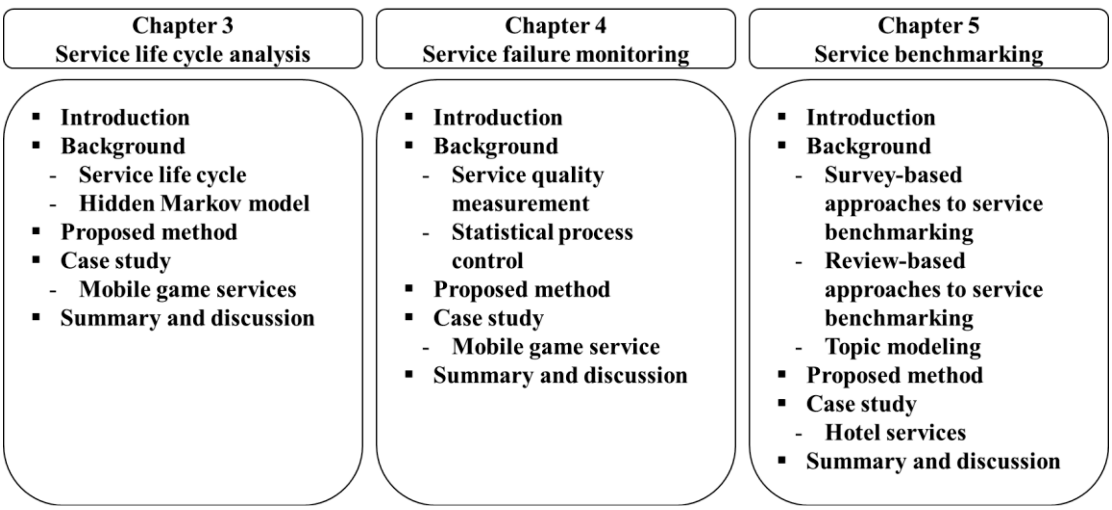



Figure 3. Structure of main bodies of the dissertation

# 2 Literature reviews

## 2.1 Service improvement

The general need for continuous service improvement is not controversial in theory and practice since every service system is created based on incomplete information and thus fallible. Moreover, firms need to respond to customers' changing and evolving needs to secure sustainable advantage and to establish long-term strategies (Schweitzer and Aurich, 2010). In this context, virtually all service companies in today's competitive and dynamic marketplace recognize continuous service improvement as the most crucial activity for differentiation and competitive advantage (Dickson, 2015; Lages and Piercy, 2012).

Given the distinctive characteristics of services, the nature of service improvement differs from that of product improvement, as sated in the introduction. As far as the models and methods for service improvement are concerned, the literature shows a trend towards extending previous qualitative approaches to quantitative approaches. The results of major qualitative studies are summarized as follows. Kingman-Brundage (1989) mapped the activities in a service process from the perspective of customers for redesigning the service process and reducing service failures. Ballantyne (1991) proposed the use of a fishbone diagram that represents a work process with potential cause and effect relationships to prevent service failures. Cantwell et al. (1997) identified problems and causes influencing the efficiency of operating rooms through brainstorming and fishbone diagrams to improve the quality of medical services. Similarly, Duckett and Nijssen-Jordan (2012) convened an expert workshop for addressing the problems of long waits in emergency departments to improve the quality of medical services.

Regarding the quantitative approaches to prioritizing service attributes to be improved, Chen (2016) employed the quality function deployment (QFD) to transform customer demands to management techniques and utilized the failure mode and effect analysis (FMEA) to determine improvement priorities of customer demands. Liou et al., (2011) applied modified VIKOR (in Serbian: *VIseKriterijumska Optimizacija I Kompromisno Resenje,* meaning multi-criteria optimization and compromise solution) method that to determine improvement priority of service attributes. Focusing on competitor analysis and benchmarking, Park et al. (2015) suggested a dual quality function deployment (QFD) by relocating the benchmarking matrix to the main body of the original QFD for comparative assessment of the activities of two firms (i.e., focal company and its competitor). Johns et al. (1997) employed data envelopment analysis (DEA) to develop the benchmarking guidelines on improving the productivity of hotel services. Specifically, four input variables (i.e. the number of rooms, total labor hours, food and beverage costs, and utility cost) and three output variables (i.e. the number of rooms sold, total cover served in the restaurant, and beverage revenue) were used. Lee and Kim (2014) proposed an integrated approach of DEA and SERVPERF for identifying whom to benchmark and examining what degree service quality should be improved. Specifically, they used the five dimension values of SERVPERF models as outputs of the DEA to provide a single measure of overall service quality and benchmarking guidelines for inefficient decision-making units (i.e., service units). Albayrak (2015) proposed an importance-performance competitor analysis that considers two types of gaps: the one is the gap between the importance of service attributes and the focal company's performance in the same service attribute, and the other is the gap between focal and competing companies' performance scores. Hemmington et al. (2018) developed an extended importance-performance analysis (IPA) to determine improvement priorities of service attributes based on the linear relationship between competitors' value and own value.

However, while prior studies have proved quite useful for different purposes, they rely heavily on customer survey data, thereby time-consuming and labor-intensive (Ziegler et al., 2008). Furthermore, the quality and reliability of analysis results strongly depend on the contents, complexity, and length of survey questionnaires as well as the willingness of respondents to participate (Groves, 2006). These drawbacks have led to a significant increase in attempts to use quantitative data and scientific methods. Among others, one attempt is to analyze customer review data which will be discussed in the next section.

### 2.2 Customer review analytics

The increased availability of online platforms and the introduction of intelligent computational algorithms have meant that service improvement no longer depends solely on expert knowledge and customer surveys, but can exploit a great deal of information that can be collected on the web. Among others, customer review data as a source of the comprehensive VoC has received increasing attention from researchers and industrial practitioners (Hu et al., 2019). Customer reviews data has advantages in that it is publicly available, easily collected, low cost, spontaneous, and insightful, but also simple to monitor and manage (Guo et al., 2017). Moreover, the use of customer reviews for service improvement is applicable to a wide range of services as many firms operate online platforms and forums to interact with customers (Bickart and Schindler, 2001). If properly analyzed, this database can provide organizations with rich and credible insight into customer perceptions and opinions on their services.

Customer review data can be obtained from many different sources, including online forums and platforms and social network services. Although the structure and content of customer reviews may differ across data sources, they generally provide information on overall ratings and review contents as well as basic information such as customer profiles and review dates. In particular, overall ratings are used as a proxy for overall service quality or customer satisfaction while contents are used to identify service attributes and to measure the service performance.

Early research on customer review analytics has led by researchers in disciplines such as computer science and applied statistics (Dave et al., 2003). For instance, Hu and Liu (2004) measured the sentiment scores of customer reviews through part-of-speech tagger by extracting the nearby adjectives of the product- and service-related features to summarize customer reviews. Popescu and Etzioni (2007) introduced an unsupervised information-extraction system that improves Hu and Liu (2004)'s feature-based summaries of customer reviews. Quan et al., (2015) proposed a self-aggregation based topic model that aggregates short texts based on general topic affinity of texts rather than particular heuristics. Xiong et al. (2018) proposed a word-pair sentiment-topic model to detect sentiments and topics simultaneously from the contents of customer reviews. Vinodhini and Chandrasekaran (2016) compared the performance of neural network-based sentiment classification methods, probabilistic neural networks (PNNs), and a homogeneous ensemble of PNNs and concluded that PNNs outperform in classifying the sentiment of customer reviews. Similarly, Hartmann et al. (2019) compared the performance of different methods for sentiment analysis across 41 social media datasets and concluded that random forests and naïve Bayes perform best among ten algorithms (five lexicon-based algorithms and five machine-learning algorithms). However, prior studies have used customer reviews as a data source of short texts and thus lack important details about value creation and implications for service improvement.

Recognizing its value, there has been a significant increase in attempts to align customer review analytics with the issues associated with service improvement. The results of major studies are summarized as follows. First, focusing on the contents of customer reviews, Kang and Park (2014) proposed a framework for measuring customer satisfaction by combining sentiment analysis and VIKOR. They assessed customer satisfaction for each service attribute via a sentiment analysis of the contents of customer reviews and measured the degree of customer satisfaction using VIKOR. Similarly, Song et al. (2016) proposed an analytical framework and procedure for analyzing customer reviews tailored to diagnosing service quality. They conducted a sentiment analysis of the content of customer reviews to capture customer perceptions and expectations at the service-feature level. Gitto and Macuso (2017) classified customer reviews into two classes (i.e. aviation related reviews and non-aviation related reviews) and compared the proportion of positive, negative, and neutral reviews in each class to improve airport services. Gao et al. (2018) identified the perceived quality of a focal firm's and its competitors' service via sentiment analysis. Min et al. (2018) classified customer reviews into two classes (i.e., positive and supplementation-required reviews) and plotted the two types of customer reviews to investigate Kano model dynamics, showing that customer requirements for certain functions change over time.

Second, some research has employed both the overall ratings and review contents to identify service attributes and measure the importance and performance of each attribute. Duan et al. (2013) employed LDA to classify customer reviews of hotel services into five SERVQUAL dimensions, namely, tangibles, reliability, responsiveness, assurance, and empathy. They found that the tangibles dimension has the strongest impact on service quality, as measured by overall ratings. Similarly, James et al. (2017) used LDA to identify six core topics, that is, staff and timeliness, physician compassion, experience, family, surgery, and diagnosis, from patient reviews of healthcare services. They identified physician compassion (i.e., whether the physician is perceived to be kind and makes the patient feel comfortable) as having the greatest effect on service quality. Guo et al. (2017) utilized LDA to identify key attributes of hotel services from the contents of customer reviews and a stepwise regression analysis to estimate the relative importance of the attributes using the sub-ratings on five dimensions (i.e., location, cleanliness, room experience, service quality, and value) as independent variables and the overall rating as the dependent variable. Luo and Tang (2019) applied the Latent Aspect Rating Analysis (LARA) to identify service attributes in accommodation sharing platform services (e.g., Airbnb) and to measure the importance of these attributes to reviewers' inclusive satisfaction (i.e. overall rating).

.

# 3 Service Life Cycle Analysis based on Customer Attention

## 3.1 Introduction

The systematic development and management of services have become strategically more important as innovation cycles become shorter and service systems become more complex (Maglio and Spohrer, 2008; Song et al., 2013). However, while there have been various discussions about scientific models, methods, and tools for use in the fuzzy front end of new service development processes (Kuusisto, Kuusisteo and Yli-Vitala, 2013; Lee and Lee, 2015; Lee, Kim and Park, 2010; Lee et al., 2012; Lee et al., 2013; Lee et al., 2009; Son et al., 2015), prior literature has not so far paid enough attention to the operational back end, especially for the post-launch review stage. There still remains a major question for decision-makers: how to support coordinated and systematic management of service after its launch using quantitative data and scientific methods.

At its core is the dynamics of services –service is introduced to a market, grows in popularity, and is then removed as demand drops. The desire for sustainable businesses in practice has thus resulted in a need to implement different business strategies in different stages of service maturity. We suggest that a life cycle analysis, which was originally based upon a biological analogy, could be a good solution for guiding organizations towards building effective post-launch strategies as a service becomes mature. For decades, life cycle analysis has been regarded as a useful tool to assist decision making in operational, marketing, and innovation issues (Aitken, Childerhouse, and Towill, 2003; Cox, 1967; Cunningham, 1969; Fuglsang, Sundbo and Sørensen, 2011; Hammer, 1981; Kotler, 1965; Scheuing, 1969; Sundbo, 1997), and indeed is in intensive use in the product and technology sector under the label ‘product life cycle analysis’ (Scheuing, 1969; Rink and Swan, 1979) and ‘technology life cycle analysis’ (Gao et al., 2013; Haupt et al., 2007). However, despite its apparent potential for application as a systematic tool for use in the service sector, the analytical framework and procedures for service life cycle analysis have rarely been scrutinized. Although a couple of studies presented the concept of stakeholder-driven service life cycles (Gu and Lago, 2007) and life cycles of mobile networks and relevant services (Van Den Ende, 2003), the implications of these studies cannot be directly linked to post-launch service strategies, since they are designed to support the planning of new service development based on case-specific conceptual frameworks.

These drawbacks necessitate the development of a new method for service life cycle analysis, so that such analysis can more adequately inform decision making. Three main issues need to be addressed to fully attain the benefits of life cycle analysis of the post-launch review stage of new service development processes. First, in terms of the scope of the analysis, life cycle analysis is a tool, which focuses on three major functional aspects: planning, monitoring, and control (Hammer, 1981).

However, previous studies on service life cycle analysis have focused only on the planning aspect, and thus cannot facilitate stage-customized decision-making as a service becomes mature. Hence, any approach that is proposed needs to take the monitoring and control aspects into account, gauging where service is in its life cycle and giving forecasts about its future prospects – which are crucial in setting up the post-launch strategies for the service. Second, with respect to the idiosyncratic aspects of service maturity, different services present different rules of evolution. The classic bell-shaped curve is the most common pattern in the product and technology life cycle literature (Gao et al., 2013; Haupt et al., 2007; Rink and Swan, 1979), but very few services follow such a prescriptive cycle in reality. Although many other patterns – such as a cycle-recycle, rapid penetration, and innovative maturity – have been identified in the product sector (Rink and Swan, 1979), these cannot easily be used for service life cycle analysis due to the lack of theoretical understanding of patterns of service life cycles and the fundamental differences in nature between products and services (e.g. degree of tangibility and perishability). Thus, any approach that is proposed needs to model the different patterns of service life cycles flexibly. Last but not least, from a practical standpoint, expert-centric approaches are time-consuming and costly, and moreover, have difficulty in gathering the expert group and cultivating knowledge and experience in many different fields (Haupt et al., 2007). The current competitive marketplace accentuates the importance and necessity of quick analysis of in-company and competitors' services (Lee and Lee, 2015; Lee et al., 2013). Empirical experience has also shown that procedures and methods become established more easily in practice if they can be supported by information technology (Bullinger et al., 2003). Therefore, any approach that is proposed needs to secure applicable quantitative data and to employ computer-supported engineering-centric methodologies so as to allow for the speedy analysis of a wide range of services at acceptable levels of time and cost.

Considering these factors, we propose a stochastic service life cycle analysis to gauge where a service is in its life cycle, and to give forecasts about its future prospects. As a data source, we employ customer reviews to measure the customer-oriented service maturity for the following reasons. Firstly, previous studies have found that customer reviews are significantly related to customer attention (Yang and Fang, 2004) as well as service performance such as customer satisfaction (Kang and Park, 2014) and service quality (Bickart and Schindler, 2001), and so can provide insight into the implications of customer-focused post-launch strategies that are of particular importance in the service sector (Alam and Perry, 2002). In particular, previous empirical studies have identified that the number of customer reviews for an item is a good proxy for the item's popularity or customer attention to the item, in that popular items are discussed more frequently than less popular items (Chevalier and Mayzlin, 2006; Duan et al., 2008; Zhu and Zhang, 2010). Secondly, it can provide information about a service's entire life, from the launch of the service to its extinction. Finally, it can be applicable to a wide range of services, as nowadays many service firms are operating channels to interact with customers (Verhoef, 2003). In this context, the proposed approach is based on the premise that the significant changes in the number of customer reviews can provide valuable information on the stages of a service's life cycle progression. Specifically, HMM, which is an unsupervised machine learning technique based on a doubly stochastic process, is used to estimate the probability of a service being at a certain stage of its life cycle from the number of customer reviews. The HMM is considered the most appropriate method for analyzing the dynamic behavior of a system (i.e. service) with unobserved states (i.e. life cycle stages). Taken together, our method incorporates the three issues stated above into service life cycle analysis. Furthermore, a software system is developed to automate our method, allowing even those who are unfamiliar with customer review analysis and complex stochastic models to benefit from the research results. It is expected that the proposed approach can assist the formation of post-launch service strategies such as pricing and promotion, and further serve as a starting point for developing more generic models. Table 1 summarizes the difference between previous research and the proposed method.

Table 1. Comparisons of previous research and the proposed method

| Factor | Previous research | Proposed method |
|---|---|---|
| Objective | Service planning | Service monitoring and controlling |
| Perspective | Service developers or providers | Customers |
| Approach | Expert-based approaches | Stochastic approaches |
| Data | Experts' judgments | Customer reviews |
| Methods | Case-based conceptual analysis | HMM which is an unsupervised machine learning technique |
| Outputs and implication | General guidelines for new service development | Stage-customized post-launch service strategies based on the service's life cycle progression and its future prospects |

This chapter is structured as follows. Section 3.2 presents the background of service life cycle analysis and HMMs. The research framework is explained in Section 3.3, and then illustrated with a case study of mobile game services in the Apple App Store in Section 3.4. Finally, Section 3.5 offers our conclusions.

## 3.2 Background

### 3.2.1 Service life cycle

A considerable amount of literature has been published on life cycle analysis to support decision making in operational, marketing, and innovation issues at the various levels (e.g. product, industry, and technology). Of these, product life cycle analysis has been at the center of attention since its first introduction by Dean (1950). Although variations exist in the literature, this method generally deals with the unit sales curve for a product, extending from the time it is first placed on the market until it is removed (Bass, 1969; Bauer and Fischer, 2000; Kotler, 1965; Scheuing, 1969). Following this, a variety of issues and suggestions – such as the determinants of product life cycles (Scheuing, 1969), patterns of product life cycles (Bauer and Fischer, 2000), forecasting models of unit sales curves (Bass, 1969), role of product life cycles in the strategy formulation (Aitken et al., 2003; Scheuing, 1969) – have been presented.

However, contrary to the extensive body of literature on product life cycle analysis, the analytical framework and procedures for service life cycle analysis have rarely been investigated; due mainly to the inherent characteristics of services such as intangibility and insubstantiality. We can summarize the major studies' results as follows. Barras (1996) built the foundation of the service innovation life cycle theory, suggesting the concept of reverse product cycles to describe the innovation process in the service sector. He divided the service life cycle into improved efficiency, improved quality, and new services stages according to the change of innovation frequency over time. Potts (1988) defined a service as follow-up management of a product, and presented the service life cycle that comprises rapid growth, transition, maturity, and end of life stages, in accordance with changes in product sales over time. Van Den Ende (2003) divided the life cycle of mobile networks and relevant services into transitional/mature and fluid phases, and developed a framework for the governance mode in different phases of the life cycle based on the degree of uncertainty and urgency involved in the networks and service development processes. Finally, Gu and Lago (2007) proposed a stakeholder-driven service life cycle model which consists of design time, run time, and change time for service-oriented architecture to represent the activities associated with the stakeholders and interactions and cooperation between them.

While previous studies took the first step in the research on service life cycle analysis, they focused only on the planning of new service development or service innovations at the macro level. As a result, the implications of these studies lie mainly in the general guidelines for new service development, and thus cannot be directly linked to the post-launch strategies for a specific service. These drawbacks provide our underlying motivation and are fully addressed in our proposed approach by considering the issues regarding the scope of analysis, idiosyncratic aspects of service maturity, and practicality, as stated in the preceding section.

### 3.2.2 Hidden Markov model

A hidden Markov model (HMM) is an unsupervised machine learning technique used to examine processes that are not fully observable. This method is based upon a doubly stochastic process in which an underlying process (i.e. state sequence) is not directly observable, but can be observed via another process (i.e. observation sequence). Because of these characteristics, the HMM has been widely used in research areas such as machine failure detection (Tai, Ching, and Chan, 2009), human activity recognition (Wong and Stamp, 2006), DNA recognition (Churchill, 1989), protein folding (Stigler, Ziegler, Gieseke, Gebhardt, and Rief, 2011), and technology growth analysis (Lee, Lee, and Yoon, 2011; Lee, Lee, and Yoon, 2012).

A graphical representation of the HMM is represented in Figure 4. In the figure, each state variable depicted with an oval takes a discrete state value, $S = \{s_1, s_2, \ldots, s_N\}$, while each observation variable, represented with a rectangle, takes either a discrete or continuous value, $V = \{v_1, v_2, \ldots, v_M\}$. The directed arcs indicate the dependency relationships between the states, and between the states and observations. Specifically, the states are related through a Markov process, and each state has a probability distribution over the possible observations. Hence, the observation sequence generated by an HMM gives information about the state sequence.

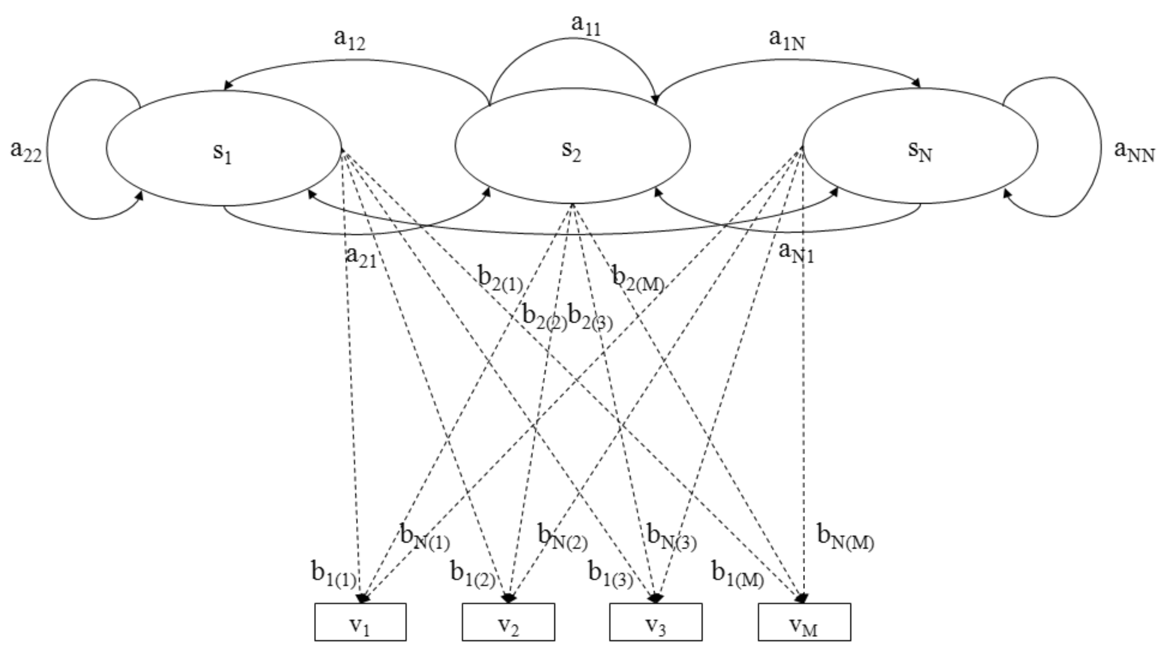



Figure 4. Graphical representation of the HMM

Three probability distributions – such as state transition probabilities (**A**), observation probabilities (**B**), and initial state probabilities (**Π**) – need to be estimated to build an HMM (Rabiner, 1989). Firstly, the state transition probability matrix **A** = [$a_{ij}$] represents the probabilities of a system moving from state $s_i$ to state $s_j$. This is generally based on the first-order Markov property where the state $q_{t+1}$ depends only on the state $q_t$, and is defined as:

$$\mathbf{A} = [a_{ij}] = P(q_{t+1}=s_j \mid q_t=s_i); \qquad i, j = 1, \ldots, N \text{------------------------------------------Eq. (1)}$$

where $q_t$ is the state at time $t$, and $N$ is the number of states. Here, the optimal number of states is determined by using qualitative analysis or such quantitative tools as Bayesian information criterion (BIC), according to the circumstances. Secondly, the observation probability matrix **B** = [$b_j(m)$] indicates the probabilities for the observation vector $v_m$ given the hidden state $s_j$, and is defined as:

$$\mathbf{B} = b_j(m)= P(o_t=v_m \mid q_t=s_j); \qquad j = 1, \ldots, N \text{--------------------------------------------Eq. (2)}$$

where $o_t$ is the observation at time $t$. Finally, the initial state probability vector **Π** denotes the probabilities that the first state in the sequence is $s_i$, and is defined as:

$$\mathbf{\Pi} = \pi_i= P(q_1=s_i); \qquad i = 1, \ldots, N \text{-----------------------------------------------------Eq. (3)}$$

where $\pi_i$ indicates the initial probability that the first state in the sequence is $s_i$.

The advantages of HMM for service life cycle analysis are three-fold. First, this method is an unsupervised technique, which can model different shapes of life cycles without any supplementary information (e.g. patterns of life cycles such as S-shaped curves). Second, this method can provide the objective information about the current life cycle stage of a service and its future prospects based on the theory of stochastic processes, which is well established. Finally, the HMM can serve as an effective tool for service engineering, enabling the quick analysis of a wide range of services. In this context, we employ the HMM based on customer reviews to examine the progression of a service's life cycle.

### 3.3 Proposed method

Service life cycle analysis should 1) provide customer-oriented information, 2) offer objective information based on systematic methods, and 3) be applicable to individual services to facilitate the formation of post-launch strategies. However, as noted above, previous studies on service life cycle analysis focus only on the planning of new service development using case-specific conceptual frameworks, and therefore do not meet any of these requirements. From a methodological perspective, curve-fitting techniques could be useful to address these issues, but suffer from drawbacks regarding the necessity of making assumptions about pre-determined growth curves. Moreover, the implications cannot be directly linked to stage-customized post-launch strategies.

Recent years have witnessed a significant increase in attempts to apply engineering-centric approaches in the service sector (Bullinger et al., 2003). However, there is little theoretical understanding of, and methodological investigation into, the ways of measuring service life cycles. In practice, it is almost impossible to obtain the historical information about the life cycle progression of similar, earlier services. Moreover, neither the process of a service's life cycle progression nor the stage it has reached can be ascertained fully since service life cycles cannot be observed directly – what we can observe is some proxy indicators such as sales and customer feedback. In this context, unsupervised methods represent the true prospects of a service's life cycle progression better than supervised methods where the values of dependent variables must be known for a sufficiently large part of the data. Among others, the HMM is considered the most appropriate method for analyzing the dynamic behavior of a system with unobserved states.

Considering these points, we developed a stochastic service life cycle analysis using the HMM based on the customer reviews. Our approach is designed to be executed in three discrete steps, as depicted in Figure 5. Firstly, we obtain the customer reviews from the web, which are generally public and visible to other customers as well as developers. Secondly, the HMM parameters including stage transition probability and observation probability are estimated to identify the stages of a service's life cycle progression based on the number of customer reviews. Here, it should be noted that the proposed approach is not limited to the use of the number of customer reviews, but can incorporate multiple indicators that may indicate a service's life cycle stage such as sentiment (Song et al., 2016), customer ratings (Zhang et al., 2010), and depth/length (Mudambi and Schuff, 2010) of the reviews. Finally, the future prospects of a service's life cycle progression are investigated based on three indicators: life cycle progression, life cycle regression, and service life time.

### 3.3.1 Step 1: Data collection and pre-processing

In fiercely competitive service marketplaces, customer management is considered a key factor for organizational sustainability (Storey and Easingwood, 1999). As one of the effort for this, companies operate the online platforms via which customers can talk about their experiences of services, and share them with others (Akehurst 2009; Witell, Kristensson, Gustafsson, and Löfgren, 2011). Although the structure and contents of reviews may differ across the platforms companies operate, they generally include information such as customer profiles, customer ratings, review contents, and review dates.

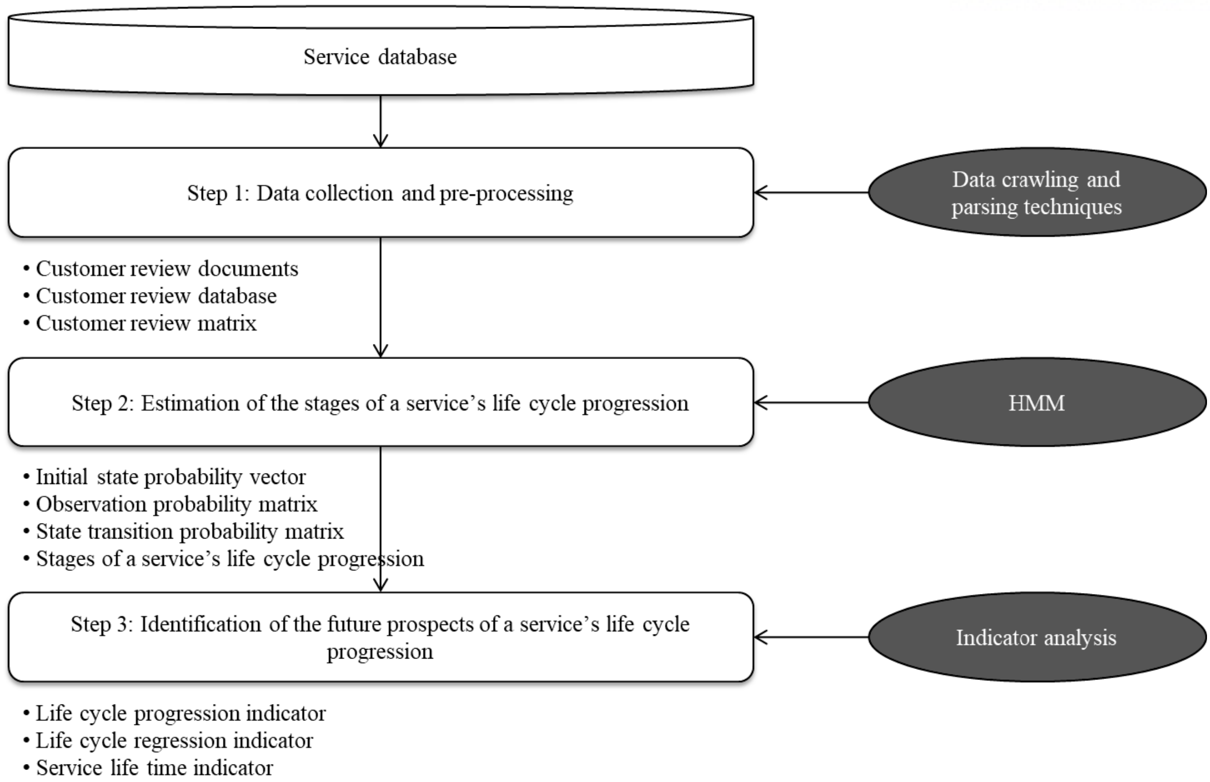



Figure 5. Overall process of the proposed approach

In this study, the number of customer reviews in a unit time is selected as a proxy of customer-oriented service maturity for service life cycle analysis.

Once a focal service is selected, customer reviews are collected from the web via data crawling techniques. The documents collected at this stage are merely expressed in text format so they need to be pre-processed via data parsing techniques. They are parsed based on the structure of the documents, and then transformed into a customer review matrix, as exemplified in Table 2. In the table, $ST_i$ and $LD_i$ indicate the title and launch date of the $i$th service while $RC_{i,j}$ denotes the review count of the $i$th service for the $j$th time unit after its launch. The sequence of review counts of a service is used as an input of the HMM to identify the stages of a service's life cycle.

Table 2. Form of customer review matrix

| Service Title | Launch date | Time $t_1$ | Time $t_2$ | … | Time $t_{m-1}$ | Time $t_m$ |
|---|---|---|---|---|---|---|
| $ST_1$ | $LD_1$ | $RC_{1,1}$ | $RC_{1,2}$ | … | $RC_{1,m-1}$ | $RC_{1,m}$ |
| $ST_2$ | $LD_2$ | $RC_{2,1}$ | $RC_{2,2}$ | … | $RC_{2,m-1}$ | $RC_{2,m}$ |
| … | … | … | … | … | … | … |
| $ST_{n-1}$ | $LD_{n-1}$ | $RC_{n-1,1}$ | $RC_{n-1,2}$ | … | $RC_{n-1,m-1}$ | $RC_{n-1,m}$ |
| $ST_n$ | $LD_n$ | $RC_{n,1}$ | $RC_{n,2}$ | … | $RC_{n,m-1}$ | $RC_{n,m}$ |

### 3.3.2 Step 2: Estimation of the stages of a service's life cycle progression

The HMM is employed to estimate the stages of a service's life cycle progression based on customer-oriented service maturity. To this end, we assume that the number of customer reviews is distributed according to a Poisson process with the event rate $\lambda$, but the rate changes over time along with the life cycle stages. This assumption has been widely used with considerable success to model the number of events in unit time (Burrel, 2001; Lee, Cho, Seol, and Park, 2012). Hence, we apply a Poisson mixture model to the customer review data, assuming the probability distribution of the data is modelled by a linear combination of Poisson distribution, as shown below.

$$p(y) = \sum_{k=1}^{K} P(k) p(y \mid k) \quad \text{-----Eq. (4)}$$

where $P(k)$ denotes a mixing parameter for the $k$th mixture component and $p(y|k)=Po(y;\lambda_k)$ is the component-conditional probability density function for the $k$th mixture component. The Bayesian information criterion (BIC) is used to find the optimal number of stages, which corresponds to the number of stages of the service life cycle.

The model parameters such as initial stage probability $\mathbf{\Pi} = [\pi_i]$, the stage transition probability $\mathbf{A}=[A_{ij}]$, and event rate $\boldsymbol{\lambda} = [\lambda_i]$, are estimated by maximizing the joint likelihood. The joint likelihood of an HMM over $N$ sequence data is defined as:

$$p(\{s_{1:T}^n, x_{1:T}^n\}_{n=1}^N \mid \theta) = \prod_{n=1}^{N} [p(s_1^n) p(x_1^n \mid s_1^n) \prod_{n=2}^{T} p(x_t^n \mid s_t^n) p(s_t^n \mid s_{t-1}^n)] \quad \text{-----Eq. (5)}$$

where $\theta=(\mathbf{A}, \mathbf{\Pi}, \boldsymbol{\lambda})$ and the subscript $n$ denotes the $n$th sequence. Specifically, the model parameters are estimated by using the Baum-Welch algorithm (Baum, Petrie, Soules, and Weiss, 1970) while the most probable stage sequence is obtained by using the Viterbi algorithm (Viterbi, 1967). Finally, the stages of a service's life cycle progression are represented by combining the stage sequence and the event rate at each stage.

### 3.3.3 Step 3: Identification of the future prospects of a service's life cycle progression

The future prospects of a service's life cycle progression are investigated based on the Markov property. In a Markovian process with $n$ exhaustive and mutually exclusive stages, the probability of a system moving from stage $i$ at time $t$-$1$ to stage $j$ at time $t$ is represented in the one-step stage transition probability matrix $\mathbf{A}$. For instance, if a service is currently in stage $s_i$, it moves to stage $s_j$ at the next unit time with a probability denoted by $\mathbf{A}_{ij}$. Similarly, a service can remain in the stage it is in, and this occurs with probability $\mathbf{A}_{ii}$. Given the transition matrix, the probability of a service transiting from stage $s_i$ to stage $s_j$ after $n$ unit times is given in the $ij$th entry of $n$-step stage transition probability matrix $\mathbf{A}^{(n)}$.

Based on this, we develop three indicators – life cycle progression, life cycle regression, and service life time – to represent the future behaviors of a service's life cycle progression. Firstly, the progression indicator is based on the probability of a service transiting to the next stages after $t$ unit times, and is defined as:

$$\text{Life cycle progression} = \sum_{j=i+1}^{N} \mathbf{A}_{ij}^{(t)} \text{ -----------------------------------------------------------Eq. (6)}$$

Secondly, the regression indicator is based on the probability of a service moving back to the previous stages after $t$ unit times, and is defined as:

$$\text{Life cycle regression} = \sum_{j=1}^{i-1} \mathbf{A}_{ij}^{(t)} \text{ -------------------------------------------------------------Eq. (7)}$$

The value of these two indicators ranges from 0 to 1, and approaches 1 if a service is likely to transit to the next stages and move back to the previous stages, respectively. Finally, the life time indicator represents the time expected to reach the end stage. This indicator is rooted in the determination of the probability of at least one passage from stage $i$ to stage $j$, defined as $f_{ij} = \sum_{n=1}^{\infty} f_{ij}^{(n)}$, where $f_{ij}^{(n)}$ is the probability of a first passage from stage $i$ to stage $j$ in $n$ unit times. An expression for $f_{ij}^{(n)}$ can be determined recursively as below:

$$\mathbf{A}_{ij}^{(n)} = f_{ij}^{(n)} + \sum_{k=1}^{n-1} f_{ij}^{(k)} \mathbf{A}_{ij}^{(n-k)} \text{ -----------------------------------------------------------------Eq. (8)}$$

Hence, the life time indicator is computed as:

$$\text{Service life time} = \sum_{n=1}^{\infty} n f_{ij}^{(n)} = (\mathbf{I} - \mathbf{M}_j)^{-1}\mathbf{1} \text{ ----------------------------------------------Eq. (9)}$$

where $i$ is the current stage and $j$ is the end stage. Here, Eq. (9) can be determined in a simpler way by using the following matrix-based formula (Taha, 2011) where **I**, $\mathbf{M}_j$, and **1** are the ($N$-1)-identity matrix, transition matrix **A** less $i$th row and $j$th column of target stage, and ($N$-1) column vector with all elements equal to 1.

### 3.4 Case study

A case study of mobile game services is presented to verify the feasibility and applicability of the proposed approach for the following reasons. First, mobile services are one of the fastest-changing areas with shortened life cycles (Lee et al., 2013), but little effort has been made to support the post-launch review stage of mobile service development processes. The Forrester research group predicted that the US market size of mobile services will reach $31 billion by 2016, and is increasing annually by 39% (Mulpuru, Evans, Sehgal, Ask, and Roberge, 2011). Among them, games are the most popular service category, representing 58% of all downloads in the mobile open markets (Agten, 2013). Second, customers may use the services for a very long time through just making a one-time purchase. In contrast, customers may actually not use the services after they purchase and download them. In this context, using the customer review data is more appropriate for measuring the customer-oriented service maturity. Finally, the life cycle in this category is expected to be shorter than that of any others, since competition among developers is fiercer and more competing services exist. The necessity of monitoring customer responses and managing competitor as well as in-company services, has come to the fore in this context. Therefore, we consider this case example to be appropriate for the suggested approach.

#### 3.4.1 Step1: Data collection and pre-processing

The Apple's App Store (http://www.iphoneappsplus.com) served as the source for data collection as follows. Firstly, the App Store represents the largest commercial and most dominant market in the world, which has more than 100,000 applications and where developers have earned $900 million in revenues (Ankeny, 2010). Secondly, the database is well organized in terms of search conditions and reliability, providing diverse information including customer reviews about each service with overall ratings and comments in electronic format.

A panel of experts was gathered from company A to identify the services of interest including direct competitors' and potential competitors' services. In addition, a Java-based web mining program was developed to download the customer review documents automatically from the website, since the number of customer review documents was huge. The enormous numbers of customer reviews inevitably require considerable time and resources if they are to be analyzed manually. A total of 45 mobile services with 153,084 customer reviews are collected, and then transformed into the structured database using Microsoft Office Access. The database includes service titles, launch dates, customer reviews, review dates, overall ratings, and others.

Finally, the customer review matrix was constructed from the database. The unit time was set to one week to capture the service life cycle reflecting immediate customer responses. It can be modified according to the purpose of the analysis, considering the trade-off between the most complex analysis (e.g. daily analysis), and a simple one with a tractable model. The resulting customer review matrix is a 46 by 80 matrix, but is not reported in its entirety owing to lack of space. A part of the customer review matrix is depicted in Table 3.

Table 3. Part of customer review matrix

| No. | ST | LD | Time $t_1$ | Time $t_2$ | … | Time $t_{51}$ | Time $t_{52}$ |
|---|---|---|---|---|---|---|---|
| 1 | 3D rollercoaster rush | 2009-06-16 | 46 | 35 | … | - | - |
| 2 | 20Q-mind reader | 2009-07-21 | 19 | 31 | … | - | - |
| 3 | Amateur surgeon | 2009-05-10 | 20 | 163 | … | - | - |
| … | … | … | … | … | … | … | … |
| 43 | Ishoot | 2008-12-08 | 11 | 6 | … | 32 | 64 |
| 44 | Unblock me | 2009-05-17 | 34 | 303 | … | - | - |
| 45 | Virtual pool | 2008-11-23 | 18 | 21 | … | 2 | 3 |

### 3.4.2 Step 2: Estimation of the stages of a service's life cycle progression

The model order was set to five based on the BIC after applying a Poisson model to the customer review data, which implies that the life cycle of mobile game services is best described with reference to the five different stages. The model parameters and the stage sequences were then estimated via Baum-Welch algorithms and Viterbi algorithms by using the 'RHmm' package implemented in R, as reported in Table 4. Note that the stages (from 1 to 4) are sorted so that the expected numbers of customer reviews therein are in ascending order while stage 5 represents the last stage of a service's life cycle which indicates the minimum customer attention.

The life cycle of mobile game services was derived by combining the stage sequence and event rate at each stage as exemplified in Figure 6. In the figure, services 21 and 30 present the different forms and shapes of life cycles. The most prominent difference is observed at the beginning of their life cycles.

Service 21 starts its progression at the first stage and enters the fourth stage of its life cycle immediately after its launch, while service 30 starts at the last stage which indicates the minimum customer attention, stagnates in the beginning, and enters the fourth stage after thirteen weeks. The lifetime and the duration at each stage are also different from each other. Service 21 illustrates a drastic increase and then a stepwise decrease, and ends its life cycle at the unit time 15, whereas service 30 is alive at the unit time 20 after its late growth.

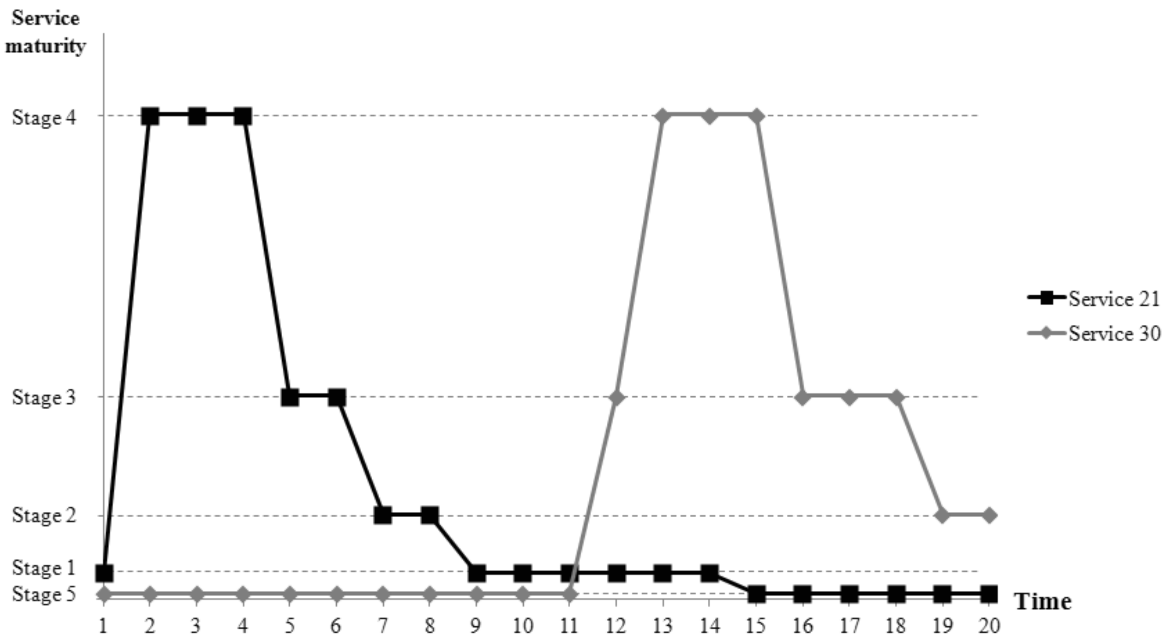



Figure 6. Life cycles for service 21 and 30

Going a step further, we conducted a cluster analysis to identify the main patterns of service life cycles. For this, we first extracted the seven factors that characterize the life cycle of a service such as age, duration at stage *i* ($dur_i$), and time to peak. The agglomerative hierarchical clustering (AHC) algorithm using Euclidean distances is then executed to group the life cycles of services, as shown below:

Euclidean distance $d(v_i, v_j) = \sqrt{\sum_{f=1}^{F}(v_{i,f} - v_{j,f})^2}$ --------------------------------------------Eq. (10)

where $v_i$ and $v_{i,f}$ represent the factor vector for life cycle *i* and the value of *f*th factor of the vector. Here, the number of groups is set to three, based on the dendrogram as shown in Figure 7.

Table 4. HMM parameters and stage sequences

(a) Initial stage probability vector

| Stage 1 | Stage 2 | Stage 3 | Stage 4 | Stage 5 |
|---|---|---|---|---|
| 1.00 | 0.00 | 0.00 | 0.00 | 0.00 |

(b) Stage transition probability matrix

| | Stage 1 | Stage 2 | Stage 3 | Stage 4 | Stage 5 |
|---|---|---|---|---|---|
| Stage 1 | 0.73 | 0.06 | 0.02 | 0.02 | 0.17 |
| Stage 2 | 0.23 | 0.70 | 0.06 | 0.01 | 0.00 |
| Stage 3 | 0.00 | 0.35 | 0.57 | 0.08 | 0.00 |
| Stage 4 | 0.00 | 0.01 | 0.39 | 0.60 | 0.00 |
| Stage 5 | 0.09 | 0.01 | 0.01 | 0.00 | 0.89 |

(c) Observation probability (event rate) matrix

| Stage 1 | Stage 2 | Stage 3 | Stage 4 | Stage 5 |
|---|---|---|---|---|
| 40.75 | 128.20 | 307.09 | 731.15 | 8.89 |

(d) Stage sequence

| No. | Time $t_1$ | Time $t_2$ | Time $t_3$ | … | Time $t_{50}$ | Time $t_{51}$ | Time $t_{52}$ |
|---|---|---|---|---|---|---|---|
| 1 | 1 | 1 | 1 | … | - | - | - |
| 2 | 5 | 1 | 4 | … | - | - | - |
| 3 | 5 | 2 | 2 | … | - | - | - |
| … | … | … | … | … | … | … | … |
| 43 | 5 | 5 | 5 | … | 1 | 1 | 5 |
| 44 | 1 | 3 | 3 | … | - | - | - |
| 45 | 5 | 5 | 1 | … | 5 | 5 | 5 |

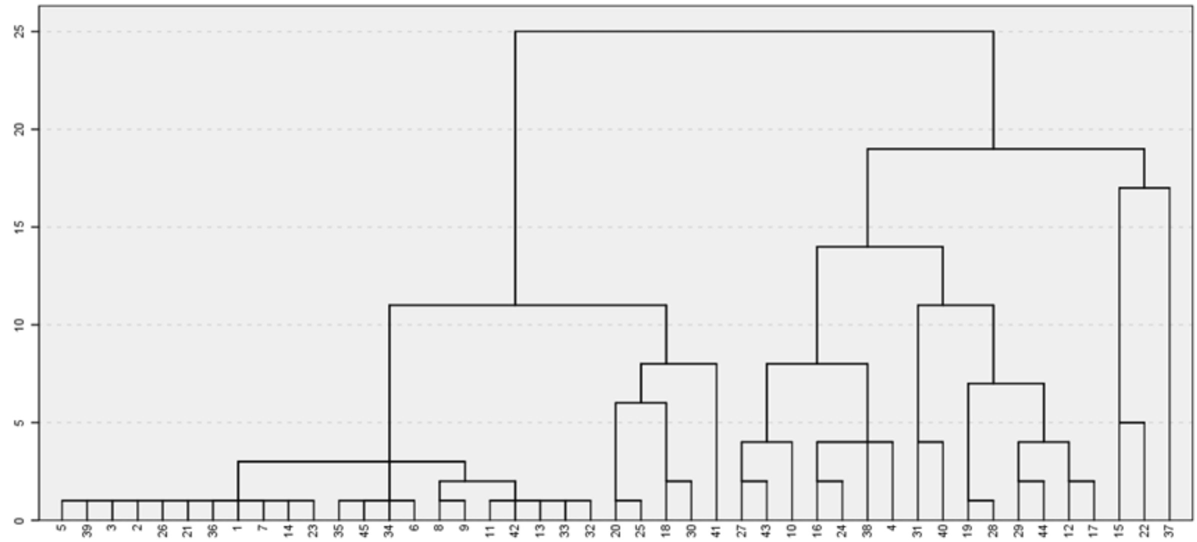


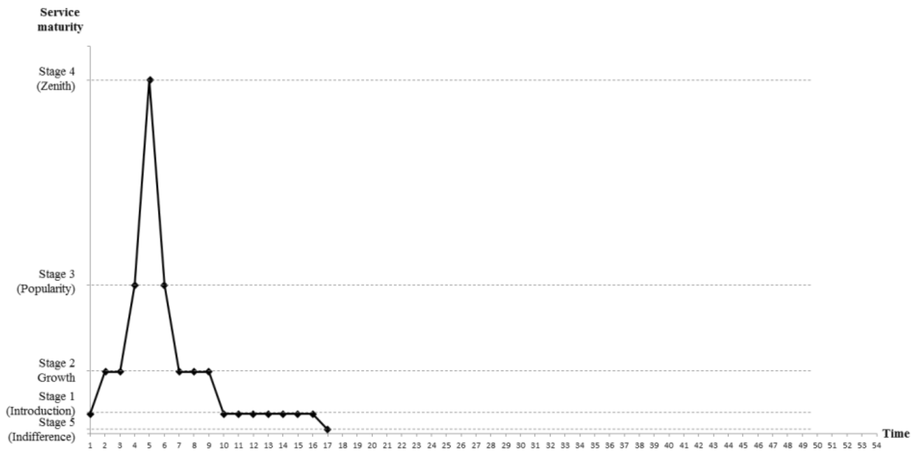
Figure 7. Dendrogram generated by the average-linkage AHC algorithm

The characteristics of life cycles of the mobile services are summarized in Table 5. Firstly, Cluster I, with 33 services, shows the shortest life time with minimal impact compared to other clusters. This pattern is characterized by a fast growth rate in the early stages and sudden death in the late stages of their life cycles. The average life time of the services in this category was found to be 19.40 weeks. Secondly, Cluster II, with 9 services, shows the stagnation in the beginning and slow death in the late stages of their life cycles. The services in this category generally take 10 weeks to reach the fourth stages and stay for 10.56 weeks in the fifth stages their life cycles. Finally, Cluster III, with 3 services, is found to be the community leader, enjoying longevity with the highest impact in the fourth stages of their life cycles. They generally spend 30.66 weeks in the third and fourth stages of their life cycles, and moreover, many of these services' life cycles do not come to an end in the data we used. The post-launch strategies (e.g. types of promotions and their timing) for these services need to be investigated in detail to run sustainable mobile game services.

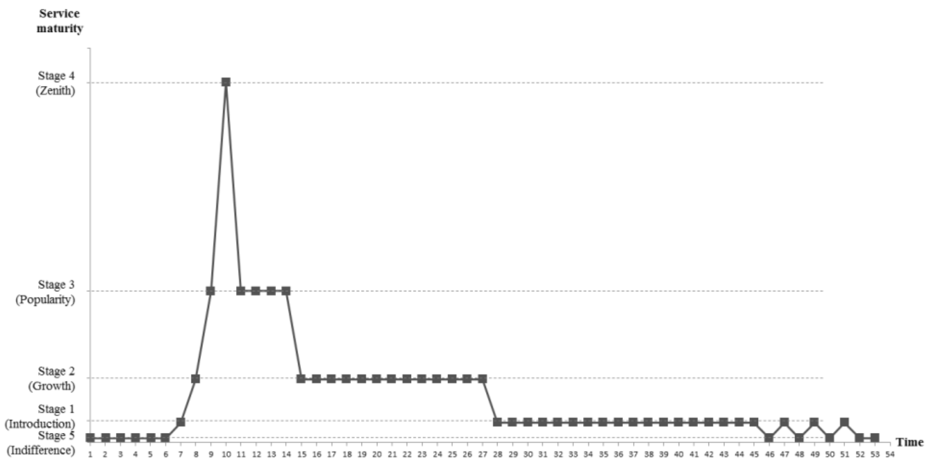
Table 5. Characteristics of the service life cycle patterns

| Cluster | Age | $Dur_1$ | Dur2 | $Dur_3$ | $Dur_4$ | $Dur_5$ | Time to peak | Patterns |
|---|---|---|---|---|---|---|---|---|
| I (33) | 19.40 | 7.52 | 4.55 | 2.00 | 0.67 | 4.70 | 5.00 | Ephemerality |
| II (9) | 52.78 | 22.78 | 13.56 | 5.00 | 0.89 | 10.56 | 10.00 | Early slump |
| III (3) | 54.33 | 3.67 | 20.00 | 17.33 | 13.33 | 0.00 | 3.00 | Community leader |

(a) Cluster I (Ephemerality)




(b) Cluster II (Early slump)




(c) Cluster III (Community leader)

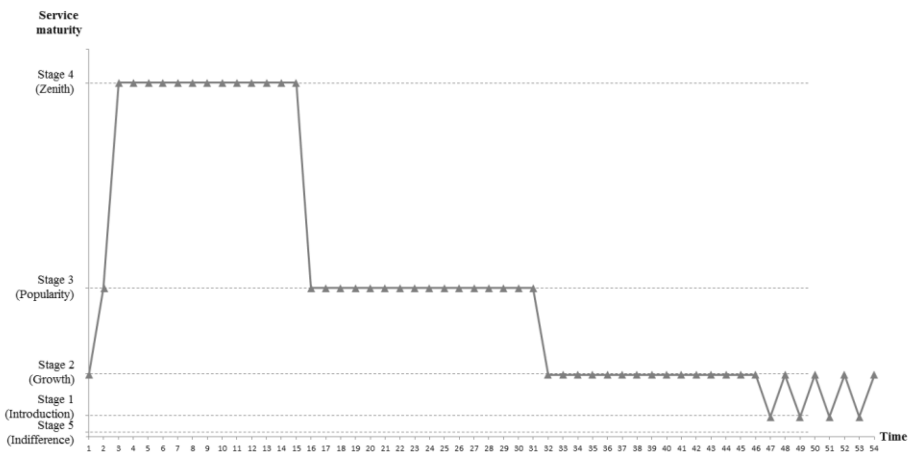



Figure 8. Life cycles for three clusters

### 3.4.3 Step 3: Identification of the future prospects of a service's life cycle progression

The future prospects of a service's life cycle progression are identified based on the three indicators reported in Table 6. First, in terms of the progression indicator, the services have less chance to move to the next stages when they are in the higher stages of their life cycles (except the fifth stage indicating the end of the service life cycle). For instance, if the service is currently in the first stage, there is a 10% chance it will move to the next stages, and this chance increases to around 30% over time. As for the service in the second stages, the chance increases until 6 weeks later and then decreases over time. It should also be noted that the services in the third stage become less likely to move to the next stages as time goes by. Second, with respect to the regression indicator, the probability for the service moving back to the previous stages increases over time, although the increasing ratios are different across the services' current stages. We calculated the difference between the progression and regression indicators over time to investigate the services' future prospects as they progresses in its life cycle, as Figure 8 reports. The value ranges from -1 to 1; it approaches 1 if the technology is likely to move to the next stages, but otherwise is close to -1. Generally, the probability for the service getting back to the previous stages is greater than that of transiting to the next stages, although the different decreasing ratios are found in the figure. From the result, we can conclude that the life cycle management in the third stages of the service's life cycle is the most critical to sustaining the services' popularity. Finally, in terms of the expected life time, it is identified that the higher stage a service is in its life cycle, the longer the service's life time is.

Table 6. Three indicators about a service's life cycle progression

(a)Progression indicator

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Stage 1 | 0.10 | 0.16 | 0.20 | 0.23 | 0.24 | 0.26 | 0.26 | 0.27 | 0.27 | 0.27 |
| Stage 2 | 0.07 | 0.11 | 0.13 | 0.14 | 0.14 | 0.15 | 0.14 | 0.14 | 0.14 | 0.14 |
| Stage 3 | 0.08 | 0.10 | 0.09 | 0.09 | 0.08 | 0.07 | 0.07 | 0.06 | 0.06 | 0.05 |
| Stage 5 | 0.11 | 0.19 | 0.26 | 0.31 | 0.35 | 0.38 | 0.41 | 0.43 | 0.45 | 0.46 |

(b) Regression indicator

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Stage 1 | 0.17 | 0.28 | 0.34 | 0.38 | 0.41 | 0.42 | 0.43 | 0.43 | 0.44 | 0.44 |
| Stage 2 | 0.23 | 0.37 | 0.46 | 0.51 | 0.55 | 0.58 | 0.61 | 0.62 | 0.64 | 0.65 |
| Stage 3 | 0.35 | 0.53 | 0.62 | 0.69 | 0.73 | 0.76 | 0.78 | 0.80 | 0.81 | 0.82 |
| Stage 4 | 0.40 | 0.61 | 0.73 | 0.80 | 0.85 | 0.88 | 0.90 | 0.91 | 0.92 | 0.93 |

(c) Life time indicator

| Stage | Stage 1 | Stage 2 | Stage 3 | Stage 4 |
|---|---|---|---|---|
| Expected life time | 10.20 | 15.69 | 19.10 | 21.51 |

### 3.5 Summary and discussions

This research has developed a stochastic service life cycle analysis based on customer-oriented service maturity. We first extracted customer reviews from the web and then employed the HMM to identify where a service is in its life cycle, and to give forecasts about its future progression. The case of mobile game services in App store was examined to illustrate the proposed approach. In the case of mobile services, some mobile services are transformed into new services through massive updates (e.g., Lineage M and Lineage M II). Since old and new services in many cases operate independently and the customers of old services need to sign up for a membership to start new services, these two services can be considered as distinct services. Nevertheless, the factors affecting the life cycles of the services such as cannibalization need further investigation.

In terms of model performance, a new method should be carefully deployed after testing its practical utility. Although the proposed approach provides information about the stages of a service's life cycle progression and its future prospects based on the theory of stochastic processes, which is well established, the results could be different from those of experts. For this reason, the experts assessed the informative value of the results of our analysis. The experts' judgments on our approach are summarized as follows. Firstly, it is confirmed that all the parameters and the stages of a service's life cycle progression were determined properly, so our method is of practical use. In particular, they concluded that the degree of service maturity derived from the number of customer reviews is meaningful to discretize the stages of a service's life cycle. In addition, it was identified that the stage sequences are estimated in the experts' expected ranges and fit their understanding of the life cycles of the mobile services employed. Second, the experts pointed out that a major strength of the proposed approach lies in its operational efficiency. The whole process of the proposed approach was automated from data collection and pre-processing to indicator analysis, and therefore can support continuous monitoring of the life cycles of a wide range of services, including in-company and competitor services. Finally, although the proposed approach is of practical use, the experts commented that the accuracy and reliability of the analysis could be improved if our method incorporates other variables such as customer ratings and the number of positive and negative reviews.

With regard to stage-customized strategies for mobile game services, different strategies are required in the different stages of service life cycles, although they may differ across organizational contexts. Based on the results of our analysis, the experts formulated the stage-customized post-launch strategies for the mobile game services. Firstly, it is most important to secure a significant number of customers in the beginning of the service life cycle (i.e. transitions from the first to the second stage) by implementing aggressive marketing strategies. The burst customer attention effect is very important in these stages–the more attention customers give in the beginning, the more successful the service is. For this, even the pre-launch and development phase marketing strategies using screenshots, video trailers, forums, and social media are required. Moreover, App Store optimization using proper keywords, tags, language localization, service icons, and description, should be implemented to increase the discoverability and customer attention. From a technical perspective, the system reliability and the user-friendly interface are of course found to be crucial. Secondly, the customer retention was found to be the most important factor in the third and fourth stages of the service life cycle. For this reason, regular content updates and appropriate promotions in the virtual service world should be provided. Moreover, in-app advertising and cross-promotion strategies need to be implemented to boost revenue. In these stages, it is important to negotiate special deals with various companies to get additional support. Finally, companies need to set up effective exit strategies in the decline stage of the service life cycle (i.e. transitions from the second to the fifth via the first stage). Here, it is important to have more customers migrate to the company's new and fresh services.

In addition, we conducted validity test to assess whether the results are robust across units of analysis. We used a biweekly unit instead of a weekly unit to evaluate how the results our method provides can be generalized to different data set. From the results as shown in Table7, it was found that the model parameters were estimated consistently for different units of analysis. Although there are slight differences in the results, when we consider that we used biweekly windows, the stages were also found to be estimated consistently. Therefore, we have confidence that the proposed method can be generalized in practice.

The contribution of the proposed approach is twofold. First, from an academic perspective, this study contributes to service engineering research by focusing on the operational back end of new service development processes. Although there have been significant attempts in developing scientific models, methods, and tools in the service area, previous research on service engineering has focused mainly on the fuzzy front end of new service development processes, and therefore cannot assist firms in establishing post-launch service strategies. In this respect, we developed a service life cycle analysis model using quantitative data and scientific models. The main advantages of our method lie in its ability to consider the idiosyncratic and intangible aspects of a service's progression and its operational efficiency to examine a wide range of services at acceptable levels of time and cost, further facilitating firms in customizing their decision making according to a service's life cycle stages. Second, from a practical standpoint, this study contributes to service life cycle analysis research by extending previous expert-centric approaches to a data-centric approach. The improvement on previous research is required, as expert-centric procedures too often demand considerable time and cost. Compared to the previous research, the proposed approach enhances the efficiency and reduces the burden of manual work, as all activities in the process have been systematized.

Table 7. HMM parameters and stage sequences

(a) Initial stage probability vector

| Stage 1 | Stage 2 | Stage 3 | Stage 4 | Stage 5 |
|---|---|---|---|---|
| 1.00 | 0.00 | 0.00 | 0.00 | 0.00 |

(b) Stage transition probability matrix

| | Stage 1 | Stage 2 | Stage 3 | Stage 4 | Stage 5 |
|---|---|---|---|---|---|
| Stage 1 | 0.66 | 0.08 | 0.01 | 0.01 | 0.24 |
| Stage 2 | 0.30 | 0.60 | 0.06 | 0.01 | 0.03 |
| Stage 3 | 0.00 | 0.50 | 0.38 | 0.12 | 0.00 |
| Stage 4 | 0.00 | 0.08 | 0.54 | 0.38 | 0.00 |
| Stage 5 | 0.10 | 0.04 | 0.05 | 0.00 | 0.81 |

(c) Observation probability (event rate) matrix

| Stage 1 | Stage 2 | Stage 3 | Stage 4 | Stage 5 |
|---|---|---|---|---|
| 78.19 | 252.21 | 653.41 | 1309.66 | 15.12 |

(d) Stage sequence

| No. | Time $t_1$ | Time $t_2$ | Time $t_3$ | … | Time $t_{50}$ | Time $t_{51}$ | Time $t_{52}$ |
|---|---|---|---|---|---|---|---|
| 1 | 1 | 1 | 1 | … | - | - | - |
| 2 | 3 | 1 | 4 | … | - | - | - |
| 3 | 4 | 4 | 3 | … | - | - | - |
| … | … | … | … | … | … | … | … |
| 43 | 3 | 3 | 3 | … | 1 | 1 | 5 |
| 44 | 3 | 3 | 5 | … | - | - | - |
| 45 | 5 | 4 | 3 | … | 5 | 5 | 5 |

Nevertheless, as this research is only at the explorative stage, there is room for further improvement to prove the usefulness of the proposed service life cycle analysis and extend its applicability. First, in regard to the data, the proposed approach does not yet model other indicators such as the number of purchases, the number of positive or negative customer reviews, and customer ratings. Multiple indicators need to be examined to improve its accuracy and reliability. Second, in terms of the scope of the analysis, identifying a service's progression in its life cycle and forecasting its future behaviors are not the ends in themselves for service life cycle analysis. Combining the suggested approach with other methods and subjects, such as pricing and service updates, will diversify the scope of the analysis and enhance the richness of potential information. Moreover, the proposed approach is not likely to be effective when a service is emerging or at a very early stage, due to the limited amount of available data. There is a need to develop more appropriate methods based on simulation techniques and the use of analogies for assessing the life cycles of disruptive services. Finally, with respect to the external validity, our case study is limited to the mobile game services in Apple's App Store. The validity of this approach necessitates further testing work by employing other indicators from a wider range of services, which is indispensable for gaining external validity. Nevertheless, we argue that the analytical power our approach offers a substantial contribution, both to current research and to future practice.

# 4 Service Failure Monitoring based on Customer Complaints

## 4.1 Introduction

Service quality management has become more strategically important as innovation cycles become shorter (Min et al., 2002) and the service market becomes more competitive (Eccles and Durand, 1998). Early research in the field focused on developing conceptual and qualitative models (e.g. Bojanic and Rosen, 1994; Buttle, 1996; Parasuraman et al., 1985), whereas recent academic interests have followed the industrial need for quantitative data use and the development of data-driven methods for service quality management (Mukherjee et al., 2018; Song et al., 2016). Recent models and methods for this purpose have proved useful for service quality management (e.g. Kang and Park, 2014; Song et al,. 2016). However, most prior models and methods have focused on examining static aspects of service quality (e.g. diagnosing and assessing service quality at a certain point in time), thereby offering limited practical assistance, especially in fast-changing service sectors[1] (Kim and Lee, 2017). Although there has been academic and industrial interest in the dynamic aspects of service quality (e.g. monitoring and managing the changes in service quality over time) (Dagger and Sweeney, 2007; Rust et al., 1999), a major question still remains in the literature as to the development of systematic models and methods to facilitate continuous service quality management.

We suggest that statistical process control (SPC)—which was originally developed to monitor, control, analyze, and improve production quality in real time during the manufacturing process (Macgregor and Kourti, 1995)—could be a good solution for continuous service quality management. Continuous service quality management based on SPC enables firms to maintain the desired degree of service quality while eliminating unnecessary quality checks, and provides numerical evidence of quality (Benton, 1991). However, despite its potential utility in service sectors, the value of SPC for continuous service quality management has rarely been studied. Although attempts have been made to integrate SPC into service quality management (Wood, 1994), the scope of analysis and potential implications were limited as they focused only on monitoring and managing physical aspects of service operations, such as waiting and response time, from the perspective of service providers rather than that of customers (Gardiner and Mitra, 1994; Mukherjee et al., 2018; Utley and Gaylord, 2009; Yang et al., 2012).

For continuous service quality management, SPC need to be modified and customized to take into account the distinct characteristics of services. Key to this problem are three crucial requirements that need to be addressed. First, in terms of quality indicators, prior literature has found

[1] One example is a mobile service whose updates are released with an extremely high frequency. For instance, the mobile services on Google Play are updated on average every 13 days (Comino, Manenti, and Mariuzzo 2016).

that customer complaints are crucial to service quality management as they lead to negative word of mouth (Berry and Parasuraman, 1992; Boshoff, 1997; Michel, 2001), customer dissatisfaction (Berry and Parasuraman, 1992; Michel, 2001), loyalty destruction (Miller et al., 2000), and customer defection (Michel, 2001; Roos, 1999). It has also been accepted in practice that customer complaints should be identified to examine the strengths and weaknesses of services, and then reflected in service redesign and improvement (Pyon et al., 2011). Hence, any approach should incorporate customer complaints into continuous service quality management. Second, with regard to the level of analysis, previous research has established that service quality is a multi-dimensional, higher-order construct (Grönroos, 1984; Jain and Gupta, 2004; Parasuraman et al., 1988). Therefore, constructing SPC at the service-feature level provides increased information and insight, facilitating close-look service complaints identification and recovery. Third, with respect to data and methods, conventional qualitative tools such as customer surveys and interviews are major means of collecting customer complaints; however, these approaches are time consuming and labor intensive. Recent studies have begun to use customer reviews to capture customer perceptions and feedbacks. If properly analyzed, these can provide organizations with rich and credible insight into customer complaints against their services. Moreover, the use of customer reviews for continuous service quality management is applicable to a wide range of services as many companies operate online platforms and forums to interact with customers (Bickart and Schindler, 2001).

Considering these requirements, we propose a sentiment analysis and SPC approach to identifying significant customer complaints that may need to be considered for service recovery and improvement. Sentiment analysis enables systematic measurement of customer reviews by quantifying the degree of customer satisfaction and dissatisfaction towards specific service features. SPC allows for continuous quality monitoring and early detection of significant customer complaints to prevent resulting service failures. An integration of these two methods makes it possible to monitor and manage customer complaints within an acceptable time and cost. Moreover, SPC analysis at the service-feature level provides a more detailed clue to poor service quality. The proposed approach was applied to a mobile game service. The case study demonstrated the effectiveness of the proposed approach for systematic and continuous monitoring of customer complaints by identifying out-of-control states.

The contributions of this research are two-fold. From an academic perspective, the application of SPC and sentiment analysis to the service sector extends the previous static assessments of service quality to continuous service quality monitoring and management. From a practical standpoint, this study presents a case study on the successful combination and application of SPC and sentiment analysis for real-world service quality monitoring. The systematic process and quantitative outcomes offered by our approach are expected to be valuable as a cost-effective practical tool for continuous service quality management.

The rest of this chapter is organized as follows: Section 4.2 presents the research background. Section 4.3 explains the proposed approach, which is then illustrated by a case study of a mobile game service in Section 4.4. Section 4.5 offers guidelines on the proposed approach and discusses the theoretical and practical implications. Finally, Section 4.6 concludes with a discussion of the study's limitations and suggests future research directions.

## 4.2 Background

### 4.2.1 Service quality measurement

A considerable amount of literature has been published on utilizing the 'voice of the customer' to facilitate service quality management (Kang and Park, 2014; Song et al., 2016). Many prior studies have argued that the 'voice of the customer' should be considered to understand the strengths and weaknesses of services, and it should be reflected in service redesign and improvement (Pyon et al., 2011). Among different types of the 'voice of the customer', it has been widely accepted that customer complaints are crucial to continuous service quality management. As such, companies have focused attention on customer complaints analysis, often using qualitative tools such as customer surveys and interviews (Chou et al., 2011). However, while such qualitative tools have proved useful for quality management in traditional service sectors including hotel and retailing services, they are time consuming and labor intensive, thereby offering limited practical assistance in fast-changing service sectors (Ziegler et al., 2008).

With the increased availability of online platforms and forums through which customers can post their opinions about services and share them with others, customer complaint analysis is no longer available solely with surveys or interviews, but also with user-generated content (Song et al., 2016). User-generated content, if properly analyzed, can provide organizations with rich and credible insight into customer opinions and perceptions on their services. In this respect, recognizing the potential benefits of customer reviews as a source of 'customer complaints', recent studies have developed customer review-based approaches to service quality management. We can summarize the major results of these studies as follows. Here, it is noteworthy that all studies reviewed here analyzed customer reviews at the service-feature level based on the service theory that service quality comprises several service features (i.e. dimensions). One stream of research has employed the overall ratings of customer reviews as a proxy for service quality or customer satisfaction (Duan et al., 2013; Guo et al., 2017; James et al., 2017). Although overall ratings of customer reviews are easily understood, many empirical studies have suggested that they are unreliable as a proxy of perceived service quality or customer satisfaction because customers may not be serious about the ratings and the system may contain some malicious spammers who give high ratings to low quality services (Fu et al., 2013; Liao et al., 2014; Pan et al., 2012).

In this context, the other stream of research has focused greater attention on the content of customer reviews, rather than ratings, to measure service quality and customer satisfaction (Gao et al., 2018; Kang and Park, 2014; Song et al., 2016; Xu and Li, 2016). However, as noted in the introduction, while prior studies have proved useful for many different purposes, they have been limited to the analysis of static aspects of service quality (e.g. measuring service quality at a specific point in time). A major question still remains in the literature as to how to monitor and manage dynamic aspects of service quality (e.g. monitoring customer complaints over time) for continuous service quality management. This question is our underlying motivation and is fully addressed in this study

### 4.2.2 Statistical process control

SPC is a tool for measuring and controlling product quality during the manufacturing process by confirming whether the process is in a 'state of statistical control' (Macgregor and Kourti, 1995). As exemplified in Figure 9, this method compares process output statistics (e.g. mean, range, and proportion of nonconformance) to their upper and lower limits to check whether they fit within expected and predictable quality levels. Originating from the field of product quality management, the applications of SPC have been extended to diverse domains, including environmental pollution monitoring (Gilbert, 1987), health care system surveillance (Boe et al., 2009; Melanson et al., 2009), trauma mortality monitoring (Clark et al., 1998), and crime rate monitoring (Anderson and Diaz, 1996).

SPC offers a number of advantages for continuous quality management. It enables firms to maintain a desired degree of service quality while eliminating any unnecessary quality checks, and provides numerical evidence of quality (Benton, 1991). Moreover, it can signal problems with service and provide the opportunity to take preventive actions to avoid the dissemination of problems (Rasouli and Zarei, 2016). Despite its potential utility as a quality management tool for use in service sectors, there have been only a few studies integrating SPC into service quality management. Gardiner and Mitra (1994) proposed a service control chart for banking services that monitors waiting time, number of customer arrivals, and number of waits exceeding three minutes. Similarly, Yang et al. (2012) presented a mean chart that also monitors service time in the banking services. Utley and Gaylord (2009) developed a residual control chart that monitors the total number of customers and total revenue of cellular services. Mukherjee et al. (2018) devised a non-parametric bivariate control chart that monitors response times and the number of answered calls within a specific time to manage service quality in a call center.

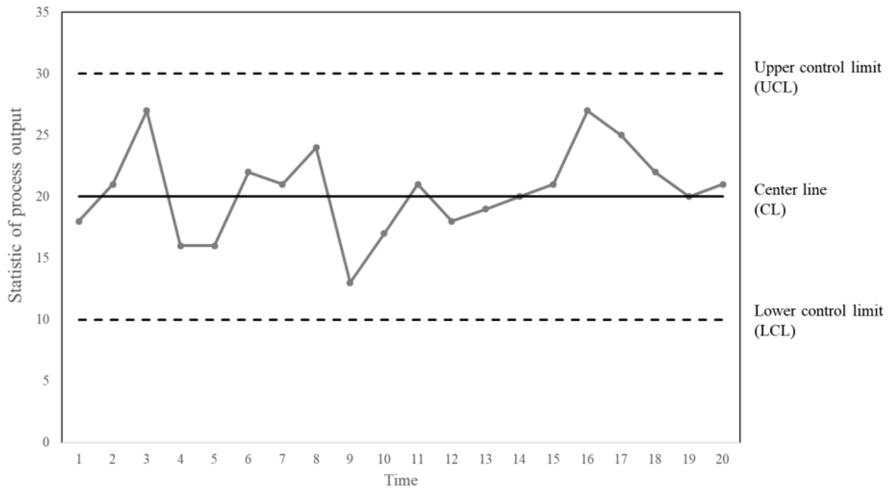



Figure 9. Example of an SPC chart

While these studies have proved useful for continuous service quality management, they have been limited to monitoring and managing physical aspects of service quality from the perspective of service providers. Recognizing the importance of the perspective of customers for continuous service quality management, Rasouli and Zarei (2016) employed a p-chart and Demerit chart with survey data to monitor patient dissatisfaction with hospital services. The proportion of dissatisfied patients was regarded as the proportion of nonconformities in the hospital service. Similarly, Chen et al. (2015) proposed a method for measuring service quality as the proportion of customers with complaints and provided a means to enable on-going monitoring of service quality using a p-chart and survey data.

Although previous studies proposed SPC approaches to service quality management using customer survey data, the benefits of SPC have not been fully achieved as these approaches are not free from the weaknesses of conventional qualitative methods for customer complaints analysis. It is currently evident that (1) customers should be at the center of continuous service quality management and (2) organizations should be able to monitor and identify customer complaints as soon as possible in order to reflect them in service redesign and improvement processes. Our study is in line with these efforts, suggesting a customer-review-based SPC approach to continuous service quality management.

### 4.3 Proposed method

Customer complaint analysis for continuous service quality management should monitor customer complaints against specific service features over time so as to facilitate responsive service recovery and improvement. As noted in the Section 4.2, customer-review-based approaches to service quality management have advantages in terms of the systematic identification of important service features from large amounts of customer review data, but they are limited to the analysis of static aspects of service quality (e.g. diagnosing and assessing service quality at a certain point in time). SPC-based approaches have the potential to examine the dynamic aspects of service quality (e.g. monitoring and managing the changes in service quality over time). However, they should be modified and customized to consider the inherent characteristics of services. This study integrates these two approaches to compensate for the weaknesses of any one approach. Table 8 summarizes the differences between previous approaches and our method.

Table 8. Comparisons of previous methods and the proposed approach

| Factor | Customer review-based service quality management | SPC-based service quality management | Proposed approach |
|---|---|---|---|
| Perspective | Customer | Service provider | Customer |
| Focus of analysis | Static aspects of service quality | Dynamic aspects of service quality | Dynamic aspects of service quality |
| Data | Customer reviews | Quality indicators of physical aspects of services (e.g. response time) | Customer reviews |
| Methods | Sentiment analysis and/or rating analysis | SPC | Sentiment analysis and SPC |
| Results and Implications | Assessing service quality at a certain point in time | Monitoring physical aspects of service quality over time | Monitoring customer complaints that may need to be considered in service recovery and improvement |

The proposed approach has advantages in that it enables early detection of significant customer complaints from customer reviews and can help prevent resulting service failures that might occur if the customer complaints are not resolved properly. Figure 10 shows the overall process of the proposed approach. The proposed approach is designed to be executed in four discrete steps: data collection and pre-processing, development of a service feature hierarchy with keyword dictionary, identification of customer complaints via sentiment analysis, and development of customer complaint charts through SPC.

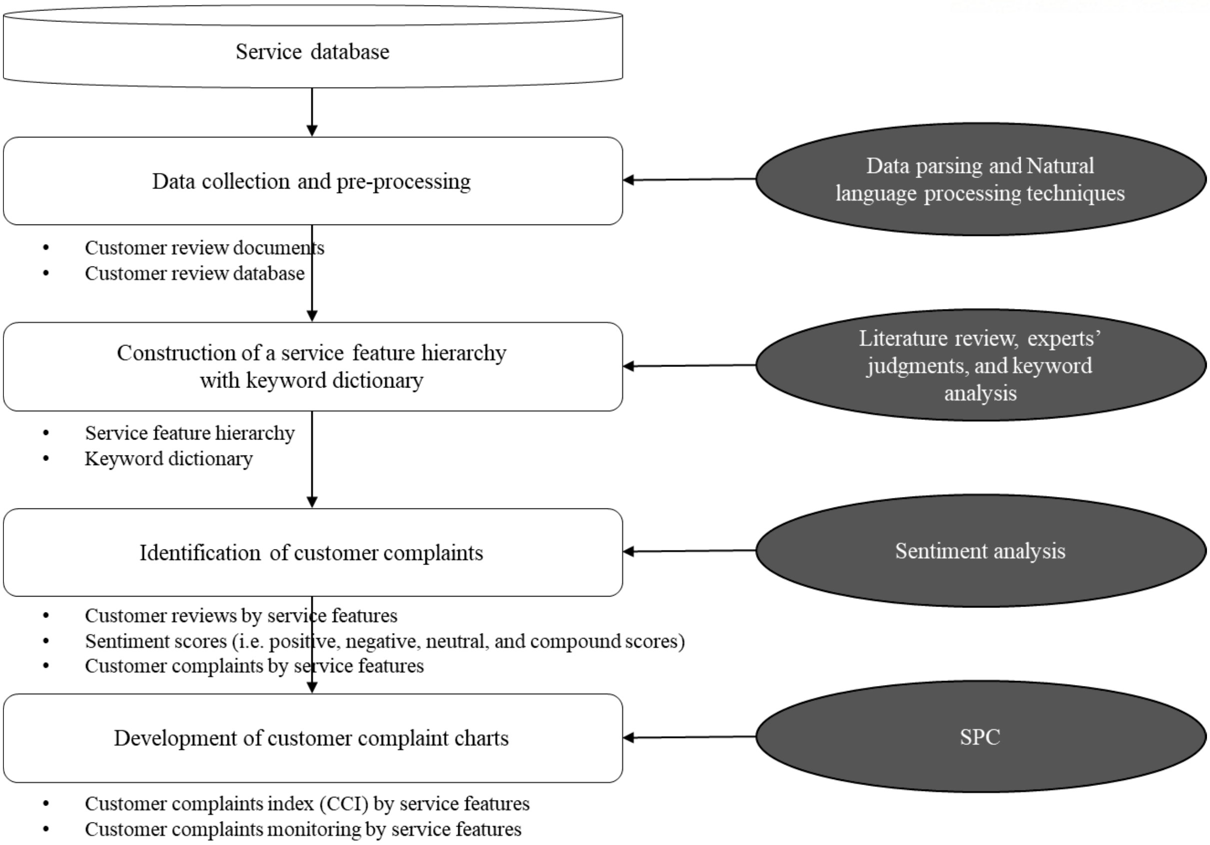



Figure 10. Overall process of the proposed approach

### 4.3.1 Step 1: Data collection and pre-processing

Customer review data can be obtained from many different sources, including online forums and platforms and social network services. Although the structure and content of customer reviews may differ across data sources, they generally provide information on customer profiles, overall ratings, review content, and review dates. While overall ratings are a basic measure that can be used to assess service quality and identify customer complaints (Xiang et al., 2015), the ratings, in many cases, are found to be unreliable as customers may not be serious about their ratings or are simply not familiar with the related field (Pan et al. 2012; Liao et al. 2014). In addition, the system may contain malicious spammers who give high ratings to low quality services (Fu et al., 2013). Moreover, researchers generally agree that service quality is a multi-dimensional, higher-order construct (Grönroos, 1984; Parasuraman et al., 1988). However, the ratings provide only abstract information about service quality, which is not sufficient to identify problematic dimensions (i.e. features) of service. For these reasons, review content and dates are employed in this study to monitor customer complaints over time.

Once the focal service is chosen, customer reviews are collected and transformed into a structured database while superfluous information, such as advertisements, is eliminated. Here, the structured items (e.g. review dates) are consistent in semantics and formats, whereas the unstructured items (e.g. review content) are textual and may have different structures or styles. For this reason, the review content is pre-processed for further analysis, as follows. First, the review content is tokenized by splitting the text into smaller tokens (e.g. words). Second, stop words (e.g. 'a' and 'the') are removed as they do not provide meaningful information. Third, part-of-speech (POS) tagging is used to assign POS tags (e.g. noun and verb) to tokens. Finally, a lemmatization technique is performed to reduce inflectional forms of a word (e.g. playing, plays, and played) and to return to the dictionary form of a word, which is known as the lemma (e.g. play). In Section 3.2, the results of POS tagging and lemmatization are used to construct a service feature hierarchy with a keyword dictionary, while the raw data are used to identify customer complaints via sentiment analysis in Section 3.3.

### 4.3.2 Step 2: Construction of a service feature hierarchy with keyword dictionary

As exemplified in Figure 11, this step constructs a service feature hierarchy with a keyword dictionary to assign customer reviews to related service features and eventually identify customer complaints at the service-feature level. Service features are defined as important dimensions or components that constitute a service and the specific issues discussed in customer reviews. Service feature hierarchy is referred to as the hierarchical structure of service features. A keyword dictionary is a collection of keywords that indicates specific service features in customer reviews.

There are two approaches to constructing a service feature hierarchy with a keyword dictionary: a top-down or bottom-up approach. The top-down approach develops a basic service feature hierarchy through experts' judgments and literature reviews, and then complements it with the results of text-mining techniques. The bottom-up approach constructs a basic feature hierarchy through text-mining techniques, such as latent semantic analysis and LDA, and then examines it through expert reviews and literature reviews. The choice between these two approaches strongly depends on the characteristics and quality of the data available. Given that customer reviews are short and informal text data, we adopt the top-down approach to constructing the service feature hierarchy with a keyword dictionary, as follows. First, service features, their hierarchical structures, and the keyword dictionary are defined through literature reviews and experts' judgments. Second, the keywords describing service features are extracted from customer reviews. For this, we consider the keywords in the form of nouns, noun phrases, and adjectives, as most service features are generally expressed as nouns and noun phrases, but some are related to adjectives (e.g. 'difficulty' in many cases is represented as 'difficult').

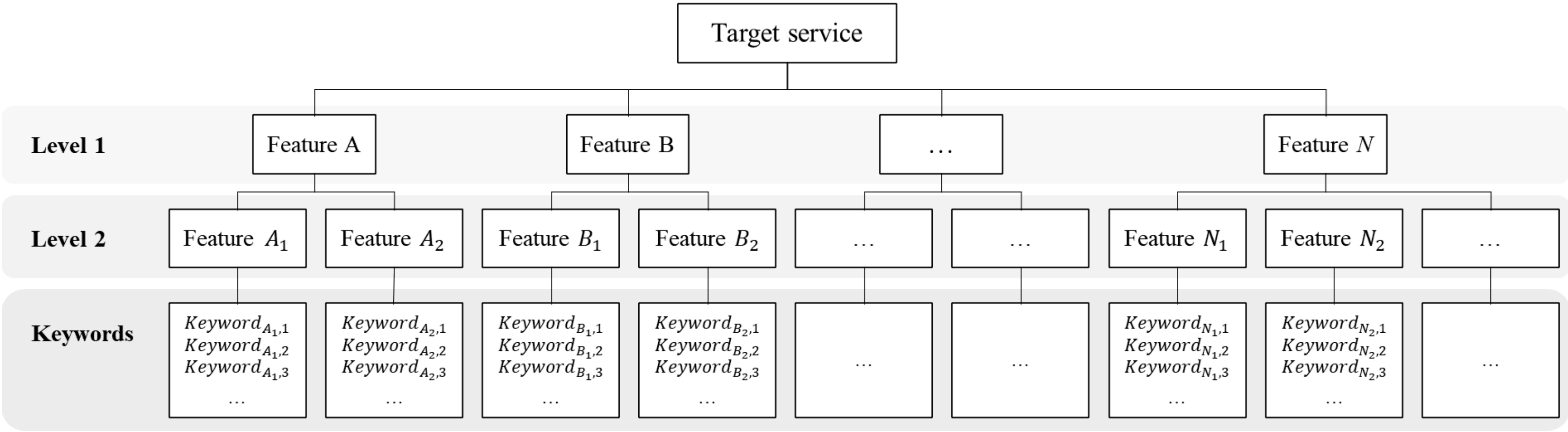



Figure 11. Form of the service feature hierarchy with keyword dictionary

Third, the keywords are rearranged by including their abbreviation and synonym. Finally, the basic service feature hierarchy with a keyword dictionary, which is developed through experts' judgments and literature reviews, is complemented by adding keywords related to specific features of the focal service. Here, service features can be newly added or removed according to the distribution of keywords.

### 4.3.3 Step 3: Identification of customer complaints

This step employs sentiment analysis—that is, a text mining technique to identify people's opinions, sentiment, or evaluation towards entities, such as brands, products, and services, by classifying the polarity of a given text (Hu and Liu, 2004)—to identify customer complaints from review content by quantifying the degree of customer satisfaction and dissatisfaction towards specific service features. There are two main approaches to sentiment analysis: the lexicon-based and text-classification approaches. The former uses a dictionary containing information on the polarities of sentiment-related words (Taboada et al., 2011). For example, such sentiment lexicons as *good, amazing*, and *wonderful* are positive sentiment words, while lexicons such as *poor, bad*, and *terrible* are negative sentiment words. The latter builds classifiers from labelled instances of texts, sentences, or phrases to measure the probability that a new instance belongs to a certain class (i.e. positive or negative) (Mullen and Collier, 2004). With respect to the trade-off between accuracy and universality, the text-classification approach may perform well in domains that the classifiers have been trained on; however, the performance may drop precipitously when applied to other domains (Taboada et al., 2011). The lexicon-based approach generally achieves high accuracy in comprehensive areas, but it is less accurate than text-classification approaches that take domain-specific characteristics into consideration. Moreover, the text-classification approach is not appropriate for this research as the overall ratings are unreliable, as noted in the prior section. Based on these considerations, this study adopts a lexicon-based sentiment analysis.

Previous studies have presented a variety of methods for lexicon-based sentiment analysis such as linguistic inquiry and word count (Tausczik and Pennebaker, 2010), semantic orientation calculator (SO-CAL) (Taboada et al., 2011), and positive and negative affect schedule (PANAS) (Goncalves et al., 2013). Among these, we employ the valence aware dictionary and sentiment reasoner (VADER) for the following reasons. First, this method performs well, even with short messages such as customer reviews, as the VADER's sentiment lexicon is attuned to sentiment in microblogs like Twitter (Part et al., 2017). Second, this method is applicable to a wide range of service sectors without any supplementary information because the VADER is a rule-based method that does not require any training data (Kim et al., 2016). Third, this method captures sentiments from emoticons, acronyms, and initialisms that customers frequently use to express their perceptions of services (Ribeiro et al., 2016). Finally, this method can be easily modified and customized according to organizational context, as it is an open source system (Ribeiro et al., 2016).

The VADER—which is also called human-validated sentiment analysis—is based on a gold-standard list of lexical features that has been constructed and validated manually. The lexical features are combined with consideration of five general rules that embody grammatical and syntactical conventions for expressing and emphasizing sentiment intensity: (1) Punctuations such as exclamation points increase the magnitude of the sentiment intensity; (2) capitalization increases the magnitude of the sentiment intensity; 3) degree modifiers impact sentiment intensity by either increasing or decreasing the intensity; 4) the contrastive conjunctions, such as *but* and *however*, signal a shift in sentiment polarity; and 5) the VADER catches nearly 90% of cases where negation flips the polarity of the text by examining the preceding tri-gram.

The procedure for identifying customer complaints using the VADER is as follows. First, customer reviews are assigned to specific service features via a string-matching technique that checks whether the customer reviews contain keywords related to the features. Second, the customer reviews are classified as positive or negative reviews by the VADER. For this, the four sentiment scores of $i$th customer review—which are positive, neutral, negative, and compound—are calculated by summing and adjusting the sentiment scores of words in the customer reviews according to the above five rules. As formulated in Eqs. (11)–(14), the positive, neutral, and negative scores represent the proportions of customer reviews that fall into each category, and the compound score is the normalized sentiment score between -1 (most extreme negative) and 1 (most extreme positive).

$$Positive\ score_i = \frac{Lexicon\ and\ rule\text{-}based\ positive\ score_i}{Total\ absolute\ sentiments_i} \text{-------------------------------------Eq. (11)}$$

$$Neutral\ score_i = \frac{The\ number\ of\ neutral\ words_i}{Total\ absolute\ sentiments_i} \text{-----------------------------------------------Eq. (12)}$$

$$Negative\ score_i = \frac{|Lexicon\ and\ rule\text{-}based\ negaitive\ score_i|}{Total\ absolute\ sentiments_i} \text{---------------------------------Eq. (13)}$$

$$Compoud\ score_i = Normalize(Lexicon\ and\ rule - based\ positive\ score_i + Lexicon\ and\ rule - based\ negaitive\ score_i), \text{---------------------Eq. (14)}$$

where the $total\ absolute\ sentiments_i$ is the sum of the lexicon and rule-based positive scores, the number of neutral words, and the absolute value of lexicon and rule-based negative scores for the $i$th customer review. Finally, the customer reviews with a compound score of less than 0 are identified as customer complaints.

### 4.3.4 Step 4: Development of customer complaint charts

This step develops customer complaints charts using SPC; these are tailored to monitor customer complaints that may need to be considered in service recovery and improvement processes. Three issues are critical to the design and development of a customer complaints chart. First, in terms of the use of statistics, we adopt a customer complaints index (CCI) by modifying the index used for service quality control (Chen and Yang, 2000; Rasouli and Zarei, 2016; Yang and Chen, 2000), as shown in Eq. (15).

$$CCI_{i,t} = \frac{N(NCR_{i,t})}{N(CR_{i,t})}, \quad \text{Eq. (15)}$$

where $NCR_{i,t}$ and $CR_{i,t}$ represent the number of negative customer reviews and total number of customer reviews, respectively, for the $i$th service feature at time period $t$. Each negative customer review is interpreted as a nonconformity. Many researchers have employed this ratio (i.e. the number of customers with complaints to the total number of customers encountered) to measure service quality because it is very straightforward and easy to apply (Yang and Chen, 2000). Utilizing the sentiment analysis with customer review data, the number of customers with complaints is estimated by the number of negative customer reviews. Accordingly, the upper control limit (UCL) and center line (CL) are defined, as shown in Eqs. (16) and (17).

$$CL_i = \frac{\sum_{t=1}^{T} NCR_{i,t}}{\sum_{t=1}^{T} CR_{i,t}} \quad \text{Eq. (16)}$$

$$UCL_{i,t} = CL_i + L\sqrt{\frac{CL_i(1-CL_i)}{CR_{i,t}}}, \quad \text{Eq. (17)}$$

where $T$ and $L$ denote the number of time periods and sensitivity parameter, respectively. After a clean set of process data is prepared, customer complaints are monitored, examining whether the CCI exceed the UCL. A CCI that plots outside of the UCL is interpreted as a signal that the service is out-of-control, and investigation and corrective actions are required to find and eliminate the cause. Second, in terms of the type of SPC, different charts can be developed according to the type of statistic. Interpreting each customer complaint as a non-conformity, we employ a variable-width p-chart, as the p-chart is effective in monitoring changes in the proportion of non-conformity over time and the variable-width chart addresses the different numbers of customer reviews per day. Finally, as far as the control parameters are concerned, two parameters (i.e. time period and $L$) should be determined manually to develop the customer complaints chart. The time period should be determined by considering how often an organization performs the proposed customer complaints analysis. For example, using a small value for time period (e.g. one week) may create more meaningful results in fast-changing service sectors by identifying customer complaints frequently. The sensitivity parameter should be determined by considering the objective and scope of customer complaints analysis. For example, if a company is only interested in major customer complaints, using a large value of $L$ (e.g. $L$=3 or 4) will provide a practical solution by restricting the scope of analysis to a manageable number of customer complaints. In contrast, if a company wants to identify minor customer complaints as well as major complaints, using a small value of $L$ (e.g. $L$=1 or 2) will identify more customer complaints.

An out-of-control state in the customer complaints chart is interpreted as a signal of service failure. If an out-of-control state is detected, customer complaints during that period should be further analyzed to determine whether the customer complaints are critical and lead to service failure, and if so, what has caused these customer complaints and how they can be resolved. Moreover, when an out-of-control state is caused by a real problem, the values of CL and UCL should be updated by deleting the data for that period to establish reliable control limits and continue to monitor customer complaints.

## 4.4 Case study

A mobile game service is chosen as a case study to verify the feasibility and applicability of the proposed approach. Mobile application services are one of the fastest growing and changing service sectors and are subject to intense competition (Kim et al., 2013). Among these, the game category is the most popular, representing 58% of all downloads and purchases in the mobile application market (Kim and Lee, 2017). To survive in a highly competitive mobile service market, it is important for industrial practitioners to identify customer complaints promptly and update their game services by adding new attributes or fixing bugs. So far, however, little effort has been made to support continuous quality management in the mobile service sector. Therefore, we consider this case example appropriate for the suggested approach.

### 4.4.1 Step 1: Data collection and pre-processing

The Apple App Store served as the source for data collection for the following reasons: First, this platform is the largest commercial, and most dominant, mobile application market in the world, and has more than 200,000 application services (Ankeny, 2010). Second, the platform is well structured in terms of search conditions and reliability. In particular, the customer reviews collected from this platform are considered valid as the Apple App Store allows only customers who have downloaded or purchased mobile application services to post reviews on those services (Kim et al., 2013).

Angry Birds 2—which was developed by ROVIO Entertainment in 2015 and downloaded more than one million times—was chosen as a test case, as this game service has been continuously updated for three years and six months to respond to customer complaints by adding new attributes and fixing bugs. We acquired and utilized a total of 2,010 customer reviews for the U.S. version of Angry Birds 2; these reviews were written in English and posted between July 31, 2017 and December 25, 2017. During that period, users actively posted reviews as ROVIO made a major update to launch a new service feature called clan battle and add new levels.

With respect to the review content, the 'natural language toolkit' (NLTK) in Python—which is a suite of libraries and programs for natural language processing for text written in English—was used for tokenization, stop words removal, POS tagging, and lemmatization. The resulting database includes both structured items, such as review number, review date, reviewer ID, overall rating, and version information, and unstructured items, such as title, content, and pre-processed content information, as shown in Table 9.

#### 4.4.2 Step 2: Construction of a service feature hierarchy with keyword dictionary

Researchers have identified mobile service features from customer reviews in several ways, as summarized in Table 10. Reviewing relevant literature, Ciurumelea et al. (2017) identified five types of mobile service features, that is, compatibility, usage, resources, pricing, and protection. Similarly, Maalej and Nabil (2015) compiled four types of service features—bug reports, feature requests, user experiences, and ratings. Khalid et al. (2015) manually identified 12 types of customer complaints from 6,390 customer reviews, namely, app crashing, compatibility, feature removal, feature request, functional error, hidden cost, interface design, network problem, privacy and ethical, resource heavy, uninteresting content, and unresponsive app. Mcllroy et al. (2016) also manually examined 7,456 reviews and added the update issue to the service features identified by Khalid et al. (2015). Fu et al. (2013) identified ten service features from customer reviews using the LDA, that is, attractiveness, stability, accuracy, compatibility, connectivity, cost, telephony, picture, media, and spam. Building upon these studies and experts' judgements, a basic service feature hierarchy comprising five first-level and ten second-level features was developed, covering the overall topics of the customer reviews without overlapping topics.

Table 9. Part of the customer review database

| No. | Date | Reviewer ID | Rating | Title | Review content (raw data) | Review content (Pre-processed data) | Version |
|---|---|---|---|---|---|---|---|
| 1 | 2017-07-31 | CPJ*** | 3 | A good pastime | It's Angry Birds, but on steroids. The graphics and sound effect make it fun to pass the time playing. | [(angry, JJ), (bird, NN), (steroid, JJ), (graphic, JJ), (sound, NN), effect, NN), (make, VBP), (fun, NN), (pas, NN), (time, NN), (play, NN)] | 2.14.0 |
| 2 | 2017-07-31 | love devastat*** | 1 | Bingo Bash Ad freezing whole game | The ad for Bingo Bash does not fully load. Only half the screen loads. There is no way to close it except to exit the app, which of course causes the current game to end… | [(ad, NN), (bingo, NN), (bash, NN), (fully, RB), (load, VBZ), (half, NN), (screen, JJ), (load, JJ), (way, NN), (close, JJ), (except, IN), (exit, NN), (app, JJ), (course, NN), (cause, NN), (current, JJ), (game, NN), (end, NN), …] | 2.14.0 |
| 3 | 2017-07-31 | Grrrranw*** | 3 | Some issues | I enjoy this game, but it seems to crash just before I am going to hit a milestone or win a great prize… | [(enjoy, JJ), (game, NN), (seem, VBP), (crash, NN), (go, VBP),( hit, VB), (milestone, RB), (win, JJ), (great, JJ), (prize, NN), …] | 2.14.0 |
| 4 | 2017-07-31 | player re*** | 5 | Awesome | Really awesome and fun game, addictive. | [(really, RB), (awesome, JJ), (fun, NN), (game, NN), (addictive, JJ)] | 2.14.0 |
| 5 | 2017-07-31 | Me*** | 5 | All the reviews are so negative, but... | This game is really fun and i've had no issues with it? weird, maybe they've fixed the bugs… | [(game, NN), (really, RB), (fun, JJ), (ive, JJ), (issue, NN), (it, PRP), (weird, VBD), (maybe, RB), (theyve, JJ), (fix, NN), (bug, NN), …] | 2.14.0 |
| 6 | 2017-07-31 | Amb*** | 2 | Locking up, can't win unless you pay | The game locks up all the time, it has other bugs that will cause you to lose. … | [(game, NN), (lock, NN), (time, NN), (bug, NN), (cause, NN), (lose, JJ), …] | 2.14.0 |
| 7 | 2017-07-31 | Squeek Hawk*** | 4 | Feat: Level all birds to 9! | "All my birds are at or above level 9, never got the feathers rewards... :( Other than that the game is great!" | [(bird, NN), (level, NN), (9, CD), (never, RB), (get, VB), (feather, RB), (reward, JJ), (game, NN), (great, JJ)] | 2.14.0 |
| … | … | … | … | … | … | … | … |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 2003 | 2017-12-25 | Pablo j*** | 2 | Bummer | Tells you that you can win a freebie and then it crashes. Freezes up when it tells you Almost There. … | [(tell, VB), (win, WRB), (freebie, JJ), (crash, NN), (freeze, NN), (wen, NN), (tell, VBP), (almost, RB), (there, RB), …] | 2.17.2 |
| 2004 | 2017-12-25 | B*** | 2 | Bleh | Too fancy and too messy for me. The simplicity and fun of the original is gone. | [(fancy, JJ), (messy, VB), (me, PRP), (simplicity, JJ), (fun, NN), (original, JJ), (go, VB)] | 2.17.2 |
| 2005 | 2017-12-25 | Cuzzieb*** | 1 | Cuzzieball | Everytime I beat another in the arena, up come a "2" and spinning gears in the right hand corner and will do nothing until I delete and reset this game! Fix it or I'm not gong to use the "play in the arena" option again!!!! | [(everytime, RB), (beat, NN), (another, DT), (arena, NN), (come, VBP), (2, CD), (spin, NN), (gear, NN), (right, JJ), (hand, NN), (corner, NN), (nothing, NN), (delete, JJ), (reset, NN), (game, NN), (fix, JJ), (im, NN), (gong, NN), (use, NN), (play, NN), (arena, NN), (option, NN), (again, RB)] | 2.17.2 |
| 2006 | 2017-12-25 | sslack1*** | 5 | Angry birds | Sweet | [(sweet, NN)] | 2.17.2 |
| 2007 | 2017-12-25 | Qued*** | 5 | Fun times | Just get rid of the ads | [(get, VB), (rid, JJ), (ad, NN)] | 2.17.2 |
| 2008 | 2017-12-25 | FelixTruv*** | 5 | Made it to Floor 71 | Okay... my previous rating nit-picked on the in-app purchases. But just a few minutes ago… | [(okay, RB), (previous, JJ), (rating, NN), (nitpicked, VBD), (inapp, JJ), (purchase, NN), (minute, NN), (ago, RB), …] | 2.17.2 |
| 2009 | 2017-12-25 | Cafa*** | 2 | Arena issues | When playing the arena the game locks up at bulb 5 continually. I loose feathers which delays progress in levels. Continually locking up. Please fix the bugs. | [(play, VB), (arena, JJ), (game, NN), (lock, NN), (bulb, VBD), (5, CD), (continually, RB), (loose, JJ), (feather, RBR), (delay, NN), (progress, NN), (level, NN), (continually, RB), (lock, VBZ), (up, RP), (please, JJ), (fix, NN), (bug, NN)] | 2.17.2 |
| 2010 | 2017-12-25 | Frogit to*** | 5 | Frogit | It's addictive I can't stop playing I have so much fun great game! | [(addictive, JJ), (cant, NN), (stop, VB), (play, NN), (much, JJ), (fun, NN), (great, JJ), (game, NN)] | 2.17.2 |

** Part of speech tags and descriptions: CD:cardinal number, DT=determiner, JJ=adjective, NN=noun (singular), PRP=personal pronoun, RB=adverb, RBR=adverb (comparative), VB=verb (base form), VBD=verb (past tense), VBP=verb (non-3rd person singular present), VBZ=verb (3rd person singular present), WRB=wh-adverb

The basic service feature hierarchy was complemented through a keyword analysis. For this, keywords describing the service features of Angry Birds 2 were identified by examining keywords in the form of nouns, noun phrases, or adjectives in the pre-processed review content. Second, these keywords were rearranged by including their abbreviation and synonyms via WordNet, which is a large lexical database of English. Finally, the keywords were assigned to each feature of basic service feature hierarchy. Here, a service feature (i.e. difficulty of game) is newly added to reflect the characteristics of mobile game services. Tables 10 and 11 report the keyword dictionaries and descriptions of the service feature hierarchy employed in this study.

Table 10. Keyword dictionaries employed in this study

| Service-feature hierarchy | | Keywords |
|---|---|---|
| Level 1 | Level 2 | |
| Compatibility | Version | edition, update, version |
| | Hardware | droid, galaxy, ios, ipad, iphone, nexus, phone, support |
| Usage | Attribute requests | arena, battle, button, rewards, flock, king, pig, power, star, tower |
| | Bug reporting | black, bug, crash, defect, error, fix, freeze, glitch, issue, network, problem, screen, solve, wifi |
| | Difficulty of game | challenge, difficult, level, rank, rating |
| | Spam | ad, advertisement, pop, video |
| Resources | Battery | battery, discharge |
| | Memory | storage, memory |
| Pricing | Price | dollar, expensive, free, gem, jewel, money, paid, pay, price, purchase, refund |
| Protection | Security | hackers, security |
| | Privacy | permission, privacy, snooping |

Table 11. Service feature hierarchy employed in this study

| Service feature (level 1) | Description | Service feature (level 2) | Description | Ciurumelea et al. (2017) | McIlroy et al. (2016) | Khalid et al. (2015) | Maalej and Nabil (2015) | Fu et al. (2013) |
|---|---|---|---|---|---|---|---|---|
| Compatibility | Issues related to version of the OS or the specific phone device | Version | Issues related to update or mobile app version | ✓ | ✓ | ✓ | | |
| | | Hardware | Issues related to a specific mobile phone device of OS | ✓ | ✓ | ✓ | | ✓ |
| Usage | Reports the things that are uncomfortable to use and things that user want to improve | Attribute requests | Issues related to additional attribute(s) or modification | ✓ | ✓ | ✓ | ✓ | |
| | | Bug reporting | Issues related to unexpected bug | | ✓ | ✓ | ✓ | |
| | | Difficulty of game | Issues related to difficulty of mobile game | Added by the authors | | | | |
| | | Spam | Issues related to advertisement | | | | | ✓ |
| Resources | Mentions the memory or battery usage | Battery | Issues related to battery usage | ✓ | ✓ | ✓ | | |
| | | Memory | Issues related to memory usage | ✓ | ✓ | ✓ | | |
| Pricing | Refers the licensing 5romodel, price of the app, or in-app purchase issues | Price | Issues related to the licensing model, price of the app, or in-app purchase | ✓ | ✓ | ✓ | | ✓ |
| Protection | States the security issues or user privacy | Security | Issues related to security or lack of it | ✓ | | | | |
| | | Privacy | Issues related to permissions and privacy | ✓ | ✓ | ✓ | | |

### 4.4.3 Step 3: Identification of customer complaints

The customer reviews were first coupled with relevant service features via a string-matching technique, as presented in Table 12. The attribute requests feature (e.g. making suggestions about clans and arenas) represents the largest portion of the customer reviews, and is followed by the bug reporting feature, which includes complaints about screen freezing or abnormal termination of the game. The protection feature is identified as the least relevant as there was only one relevant customer review. This is considered reasonable given the nature of mobile game services and the content of Angry Birds 2.

The customer reviews were classified into positive and negative reviews in accordance with the four sentiment scores (i.e. positive, neutral, negative, and compound scores), which were calculated by using the 'sentiment' package implemented in Python, as shown in Table 13. Out of 2,010 customer reviews, a total of 640 customer reviews with a compound score of less than 0 were identified as customer complaints.

Overall, the positive and negative reviews were well categorized, and the scores differed according to the five rules (i.e. exclamation point, capitalization, modifiers, contrastive conjunction, and negation) and sentiment words used in the customer reviews. For instance, in the case of the review saying 'I was instructed to update the game in order to continue playing. I updated and now I can't play because it is telling me to update…', the compound score was -0.061. However, in the case of the review saying 'App keeps freezing when you are playing, do you lose your archived levels. Very frustrating!', the compound score was -0.699, which was much lower than the previous review. Because of the intensifier *very* in front of the sentiment word *frustrating* as well as the exclamation point in the sentence, the negative score of the second review was higher than the first review.

Table 12. Number of customer reviews for each feature

| Service features | | Number of customer reviews |
|---|---|---|
| Level 1 | Level 2 | |
| Compatibility | Version | 265 |
| | Hardware | 103 |
| Usage | Attribute requests | 1,147 |
| | Bug reporting | 462 |
| | Difficulty of game | 380 |
| | Spam | 176 |
| Resources | Battery | 13 |
| | Memory | 2 |
| Pricing | Price | 377 |
| Protection | Security | 1 |
| | Privacy | 0 |

Table 13. Part of the results of sentiment analysis

| Review contents | Positive | Neutral | Negative | Compound | Classification |
|---|---|---|---|---|---|
| It's Angry Birds, but on steroids. The graphics and sound effect make it fun to pass the time playing. | 0.268 | 0.645 | 0.087 | **0.671** | Positive |
| Amazing and fun game I love it | 0.79 | 0.21 | 0.0 | **0.906** | Positive |
| I love playing the game it is so much fun and it is challenging it make your mind work thank you so much | 0.446 | 0.554 | 0. | **0.913** | Positive |
| It's cute and the upgrades are nice | 0.537 | 0.463 | 0.0 | **0.7** | Positive |
| The graphics on this latest version are amazing! Even on my iPad 2. | 0.254 | 0.746 | 0.0 | **0.624** | Positive |
| Love this game I find it quite challenging, you'll love it to. | 0.569 | 0.431 | 0 | **0.89** | Positive |
| This game is unfortunately addictive!!! I wake up playing it and go to bed playing it!!! Yes and I spend lots of money! I just wish the elevator portion is more friendlier giving us a break. Thanks for a great game!! | 0.44 | 0.514 | 0.046 | **0.963** | Positive |
| … | … | … | … | **…** | … |
| Fills the time; Allows me to play without constantly making purchases -or- annoying me with ad's. | 0.126 | 0.733 | 0.141 | **-0.077** | Negative |
| I had this game downloaded and was playing for almost two years. Then I had to delete the app and reinstall it on my iPhone because the game froze up on me. Now all of the coins and pearls that I earned are now gone. Furthermore, this app drains the battery on my phone really bad. I just deleted the app. The developers obviously don't know how to properly develop an app. | 0.025 | 0.922 | 0.053 | **-0.458** | Negative |
| I was instructed to update the game in order to continue playing. I updated and now I can't play because it is telling me to update. What's going on? | 0.065 | 0.862 | 0.073 | **-0.061** | Negative |
| App keeps freezing when you are playing, do you lose your archived levels. Very frustrating! | 0.088 | 0.54 | 0.372 | **-0.699** | Negative |
| Too hard to reach boss - run out of birds before getting to final room! | 0.074 | 0.811 | 0.114 | **-0.151** | Negative |
| Game has started freezing in the middle of the game. Is there a bug? | 0.0 | 0.896 | 0.104 | **-0.103** | Negative |
| It's now almost impossible to earn gems so your forced to buy them. The tower is rigged so forget winning anything off that. Let us earn gems again!! | 0.101 | 0.657 | 0.242 | **-0.634** | Negative |

### 4.4.4 Step 4: Development of customer complaint charts

Customer complaints charts for the overall status and specific service features were developed using the customer complaints index, which are not reported in their entirety owing to lack of space. Figures 12 and 13 present the customer complaints charts for the overall status and the attribute request feature of the service, respectively. The attribute request feature selected as an example is one of the most important features for quality management of a mobile game service as it is important to continuously update the game service by modifying existing attributes or adding new attributes to prevent customer churn. First, with respect to the customer complaints chart for the overall status of the service, shown in Figure 12, the value of the customer complaints index exceeds the UCL at times 28 (i.e. December 13th–17th) and 29 (i.e. December 18th–22nd), indicating that customer complaints against the overall status of the service were not under control. Further analysis of customer reviews registered during this period revealed that a compatibility issue with iPhone X was the cause of this out-of-control state. To confirm this, we investigated the update history of Angry Birds 2. There was a scheduled update to add a new attribute on December 11th, which caused a compatibility problem. The company performed an additional update to solve the compatibility problem on December 20th, 2017. We deleted the out-of-control states at times 28 and 29 and updated the CL and UCL. After recalculation, no other out-of-control states are found on the updated customer complaints chart. The value of the CL decreased from 0.296 to 0.288, and the UCL value was updated.

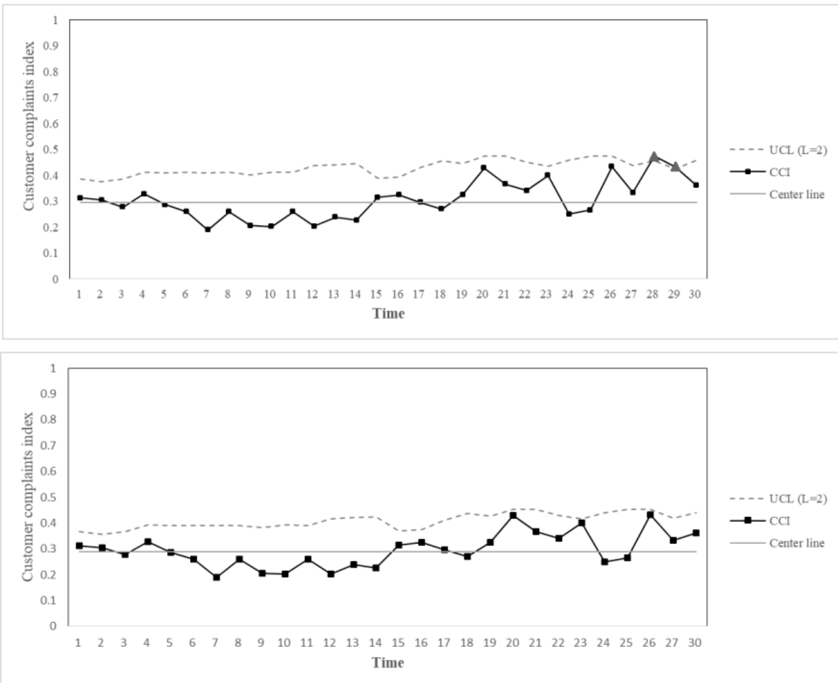



Figure 12. Customer complaints chart for the overall status

Second, as for the customer complaints charts for the attribute request feature (Figure 13), the value of the customer complaints index at time 23 (i.e. November 18th–22nd) exceeds the UCL. The customer reviews posted indicated that major customer complaints were related to the 'tower of fortune' attribute. Customers complained that they had to wait a long time or pay to reuse the tower of fortune. On December 6th, 2017, an update was made to allow customers to watch short video ads instead of waiting a long time or paying to reuse the tower of fortune. The customer complaints chart for the attribute request feature was updated by deleting the 23rd point, when the trouble took place, and no abnormal point was found in the updated chart. The value of CL decreased from 0.294 to 0.286, and the UCL values were updated.

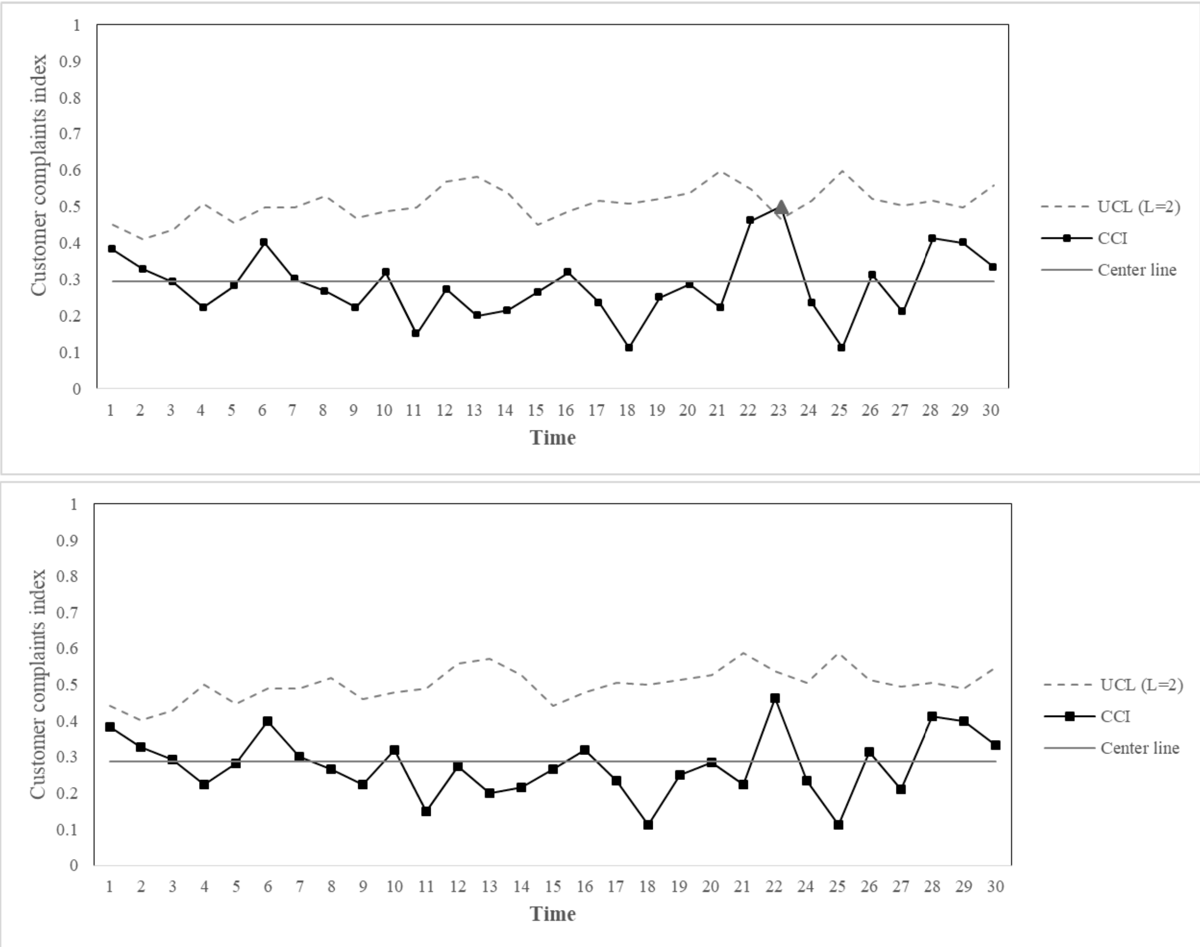



Figure 13. Customer complaints chart for the attributes request feature

Although the proposed approach is found to be useful for monitoring customer complaints that may need to be considered in service recovery and improvement processes, the results of the proposed approach may vary according to the value of control parameters (i.e. the time period and sensitivity parameters). For this reason, we conducted additional analysis using different values for the time period and sensitivity parameters to assess the robustness of the proposed approach. First, we compared customer complaints charts for the overall status of the service with different sensitivity parameters (i.e. *L=1, 2*, and *3*) and different time periods (i.e. time period=*5* and *10* days), respectively. As can be seen from Figure 14, the same period was identified as an out-of-control state under different time periods. The results also show that setting a smaller value for the sensitivity parameter to set the control limit more conservatively can detect even minor issues. By setting the sensitivity parameter to 1, we found additional minor issues related to unfair rewards, excessive commercial videos, and in-app purchases. We also examined the customer reviews during the CCI that did not exceed the UCL and confirmed that the service was in control during that period, although there were a few customer reviews expressing dissatisfaction.

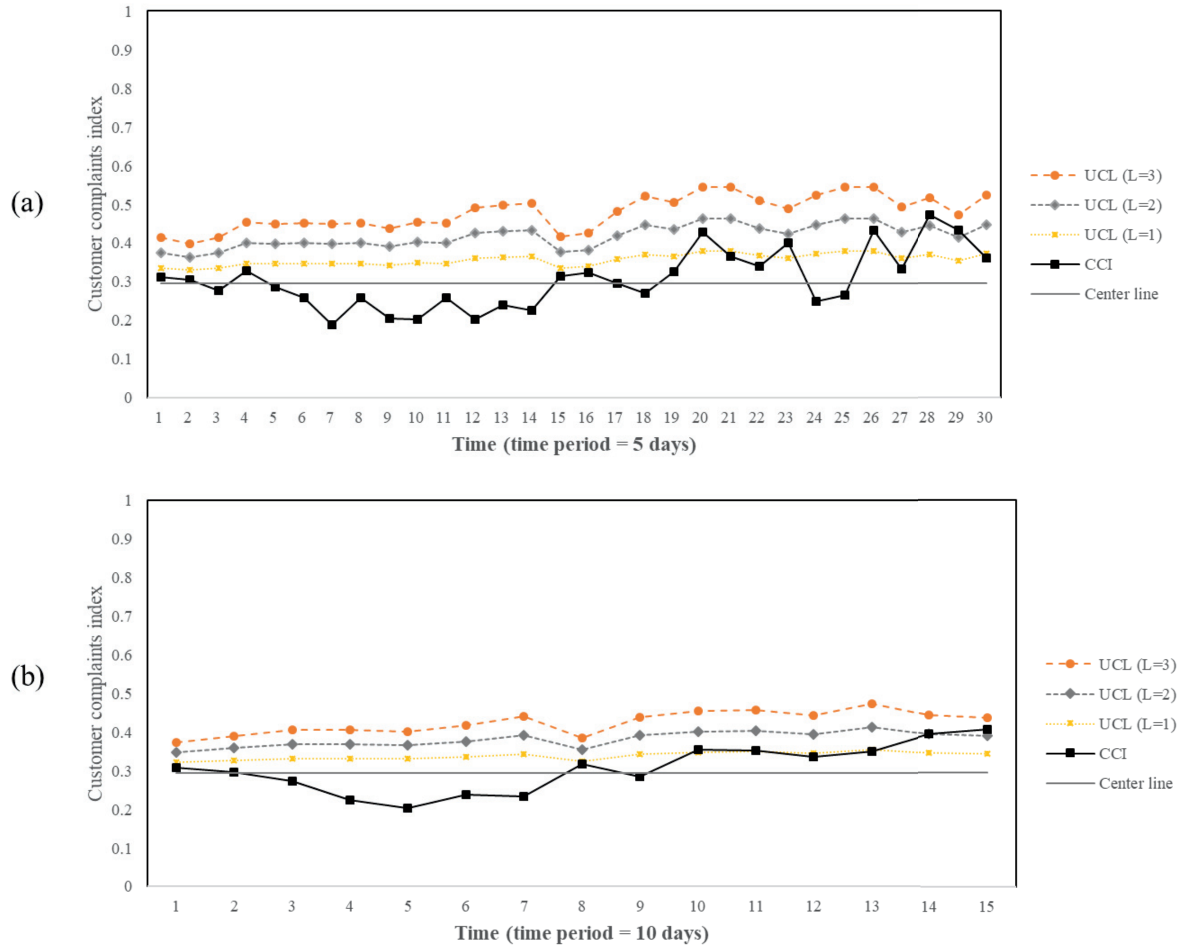



Figure 14. Customer complaints for the overall status with different values of control parameters

In the same way, we compared the customer complaints chart for the attribute request feature with different sensitivity parameters and different time periods, as shown in Figure 15. Although no out-of-control state was detected in the chart with the time period set to be 10 days, the customer complaints index at time 12 was close to the UCL. We also found additional minor issues related to specific attributes in the game such as arena, star ratings, and game friends by setting the sensitivity parameter to a lower value of 1. The results of additional analysis indicate that the proposed approach is considered reliable and robust across different values of control parameters, and they can be customized and deployed in different organizational contexts.

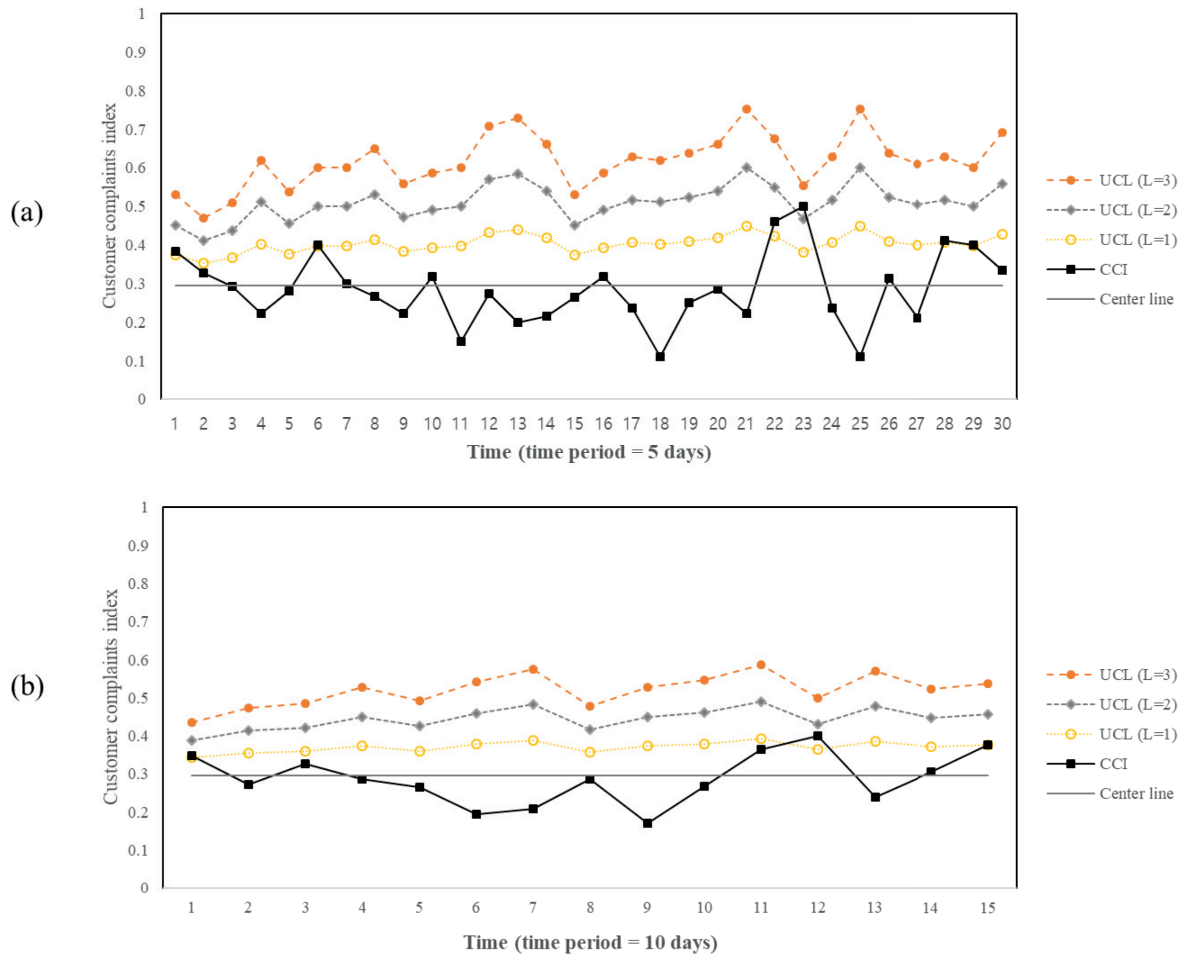



Figure 15. Customer complaints chart for the attribute request feature with different value of control parameters

## 4.5 Summary and discussions

This study provides a systematic means of customer complaints analysis for continuous service quality management, and, therefore, has novel implications for theory and practice. From a methodological perspective, a sentiment analysis method enables systematic identification of significant customer complaints from customer review data, and a statistical process control method allows prompt detection of significant customer complaints. The integration of these two methods makes it possible to monitor customer complaints at an acceptable time and cost.

In terms of implementation and customization, the proposed approach can be generalized to other service sectors, although our case study is specific to the mobile game service. However, several considerations should be examined before deploying the proposed method in practice. First, the objective of the proposed approach is to enhance the manual process of identifying customer complaints that may need to be considered in service recovery and improvement. Once the out-of-control state is detected, in-depth analysis of the relevant customer reviews should be performed to identify whether the customer complaints are critical, what caused the complaints, and how the complaints can be resolved. For this, the involvement of experts from different domains remains crucial to set up the service recovery and improvement strategies. Second, we classified customer reviews into two categories (i.e. positive and negative) and interpreted negative customer reviews as customer complaints. Instead, identifying customer reviews can be viewed as a multi-classification problem that sorts customer reviews into multiple categories according to the degree of customer dissatisfaction measured by sentiment scores. Moreover, customers give different feedback on specific service features. For example, some customers are satisfied with a service feature while the others are dissatisfied with the feature. In this context, the customer complaint index can be elaborated upon by considering both the number of positive and negative customer reviews for a specific service feature. Third, although the gold-standard list of lexical features equipped with the VADER has global scope and coverage, it may ignore important lexical features that are especially relevant to specialized services. The list should be modified and customized according to the service sector. Specifically, domain-specific lexical features can be added based on the keywords extracted from customer reviews. The scores of lexical features can be modified according to the service context. Fourth, different types of customer complaints charts using different SPC algorithms can be developed according to the objective of the analysis and the type of statistic. In particular, the proposed approach is not limited to the use of a single index, but it can incorporate multiple indicators of service quality. In the same vein, this study used a single variable control chart, but future studies could use multivariate control charts. Finally, the control parameters, such as the time period and sensitivity parameters, should be carefully determined according to the objective of the analysis. Moreover, they do not need to be fixed; instead, they can be variable, allowing flexible analysis to be performed along with events such as service updates and promotion.

With regard to validation of customer complaints identification, although significant customer complaints that had been reflected in service recovery and improvement processes were found in our case study, any approach that is developed should be carefully deployed in practice as there is no absolute confirmation regarding the validity and practicality of the proposed approach. In this respect, the performance of the proposed approach is strongly related to the ability to identify customer complaints from among customer reviews, which corresponds to a binary classification problem. For this reason, we examined the performance and reliability of the sentiment analysis using several quantitative metrics, as summarized in Table 14. Specifically, a total of 300 customer reviews were first randomly selected and classified into two categories (i.e. positive and negative reviews) by the authors and other experts. A confusion matrix was then constructed to compare the results of the sentiment analysis to those of experts, as shown in Table 14 (a). Finally, four quantitative indicators—which are accuracy, precision, recall, and $F_1$ score—were examined, as reported in Table 14 (b). Here, accuracy is defined as the percentage of correct classifications, as defined in Eq. (18). Precision indicates the number of correct results divided by the number of all returned results, as shown in Eq. (19). Recall measures the number of correct results divided by the number of results that should have been returned, as shown in Eq. (20). $F_1$ score represents the overall effectiveness of the sentiment analysis and is defined as the harmonic average of the precision and recall, as shown in Eq. (21). This score reaches its best value at 1 and worst at 0.

$$Accuracy = \frac{tp+tn}{tp+tn+fp+fn} \quad \text{Eq. (18)}$$

$$Precision = \frac{tn}{tn+fn} \quad \text{Eq. (19)}$$

$$Recall = \frac{tn}{tn+fp} \quad \text{Eq. (20)}$$

$$F_1 \ \ score = 2 \cdot \frac{precision \cdot recall}{precision+recall} \quad \text{Eq. (21)}$$

In the equations, $tp$, $tn$, $fp$, and $fn$ represent the number of positive reviews correctly classified, number of negative reviews correctly classified, number of negative reviews wrongly classified as positive, and number of positive reviews wrongly classified as negative. As can be seen from the Table 14 (b), the result of the performance evaluation indicates that the proposed approach provides effective and reliable performance in identifying customer complaints from among customer reviews.

This study proposed a sentiment analysis and SPC approach to identifying customer complaints from online customer reviews that may need to be considered in service recovery and improvement processes. A central tenet of the proposed approach is that customer reviews, if properly analyzed, can provide organizations with rich and credible insight into customer perceptions of their services. The case of a mobile game service confirmed that the proposed approach is valuable as a cost-effective, complementary tool for continuous service quality management. The case study also identified a way to improve the performance of the proposed approach and diversify the scope of analysis.

Despite its valuable contribution, this study has certain limitations that should be explored in future research. First, this study cannot provide information about silent but critical customer complaints as the proposed approach identifies only customer complaints that are described in customer reviews. For this, other data types, such as expert and lead customer review data need to be investigated to extend the scope of the analysis. Second, many issues remain as to how to improve the performance of the proposed approach. In particular, the lexicon features should be updated and customized according to the service sector. Different types of customer complaints charts should be developed according to the objective of the analysis. Third, the service features presented in this study are by no means fixed and exhaustive. Important features may differ across service industries, and thus, should be modified and customized by adding or removing features from the model. Fourth, the implications that can be derived from the analysis can be diversified by integrating other methods, such as the Kano model, and failure mode and effect analysis, into the proposed approach. Finally, this study considered a single case study, a mobile game service. Further testing on other service industries using different databases is essential to confirm the feasibility and validity of the proposed approach.

Table 14. Result of performance evaluation

(a) Confusion matrix

| | Predicted positive | Predicted negative |
|---|---|---|
| Actual positive | 178 | 9 |
| Actual negative | 31 | 82 |

(b) Summary of performance metrics

| Accuracy | Precision | Recall | $F_1$ score |
|---|---|---|---|
| 0.87 | 0.85 | 0.95 | 0.9 |

# 5 Service Benchmarking based on Customer Perception

## 5.1 Introduction

Benchmarking in service industries – defined as the continuous process of measuring service performance against the competitors recognized as industry leaders – is considered an indispensable activity for firms to create and sustain a competitive advantage (Fu et al., 2011). A variety of models and methods such as importance-performance analysis (IPA) (Albayrak, 2015; Bi et al., 2019; Taplin, 2012), quality function deployment (QFD) (Park et al., 2015), data envelopment analysis (DEA) (Lee and Kim, 2014) and the analytic hierarchy process (AHP) (Min et al., 2002), have been presented for measuring service performance and offering benchmarking strategies. However, while prior studies have proved quite useful for measuring the importance of service attributes and significantly improved the quality of analysis results, they rely heavily on customer survey data, thereby time-consuming and labor-intensive (Ziegler et al., 2008). Moreover, previous approaches might miss unexpected but important service attributes that are difficult to identify in the design of survey questionnaires. Furthermore, the quality and reliability of analysis results strongly depend on the contents, complexity, and length of survey questionnaires as well as the willingness of respondents to participate (Groves, 2006).

Highlighting possible avenues for methodological adaptation, recent academic interest has followed the need for the use of quantitative data and scientific methods. In particular, the dominant approach is offered by analysis of customer review data. Specifically, existing customer review-based approaches can largely be classified into two categories: one stream has utilized the rating information to identify the critical service attributes that distinguish a focal company’s service from those of its competitors (Mariani and Visani, 2020; Xia et al., 2019; Xia et al., 2020), whereas the other stream has focused more on the contents of customer reviews using text mining techniques to identify service attributes to diversify the scope of benchmarking guidelines (Bi et al., 2019; Hu and Trivedi, 2020). Although these approaches have enabled the identification of service attributes to be considered and the evaluation of service performance within acceptable limits of time and cost, they are limited to examining financial and operational aspects only in the benchmarking process. This is especially true in the selection of competitors (or best practices) as benchmarks. Taking the hotel industry as an example, the competitors were selected as the hotels of the same star rating. Despite the importance of intangible service attributes in customer satisfaction (Atkinson, 1988; Buttle, 1996; Lewis, 1989), the star rating systems tend to focus primarily on tangible criteria such as the availability and the size of facilities (Hensens et al., 2010) and become outdated since the systems stress some items that might no longer be important (e.g., a two-line telephone in every room) and overlook other items that might be more important to today's travelers (e.g., the atmosphere of hotel rooms and swimming pools) (Torres et al., 2014).

These drawbacks necessitate the development of new methods for service benchmarking. The review of previous studies suggests that three issues are central to this problem and need to be addressed. First, benchmarking service performance against competitors based on financial and operational indicators does not guarantee a sustainable competitive edge because of the inherent characteristics of services, such as intangibility and insubstantiality (Shamma and Hassan, 2013). At the core of service benchmarking are customer experience and/or value creation to provide a clear direction toward meeting the needs of sophisticated customers and improving service values (Koller and Salzberger, 2009). Hence, any approach that is proposed should offer benchmarking guidelines from the perspective of customers (e.g., identification of service attributes and selection of competitors based on customer experience and/or value creation). Second, benchmarking is a continuing process, not a one-time activity. This is especially important in recent business environments where the number of services (or competitors) is dramatically increasing (e.g., rooms available on Airbnb) and customer needs are becoming diversely segmented (e.g., business, airport, suites, residential, resort, casino, and convention hotels). Hence, any approach that is proposed should enable the quick analysis of the key aspects of the focal company's and competitors' services and support decision making with acceptable levels of time and cost using quantitative data and scientific methods (Bi et al., 2019). Finally, it is not sufficient to compare a focal company's service performance against its competitors. Organizations should implement as many plans derived via benchmarking processes as practical within resource limitations. Hence, any approach should provide specific and practical benchmarking guidelines according to customer experience and/or value creation (e.g., the priority of service attributes that should be improved and the directions of improvement).

We propose a data-driven approach to customer-oriented service benchmarking using large amounts of customer review data as a source of the comprehensive VoC. Specifically, we use both the ratings and the textual contents of customer review data since the ratings are a straightforward, unambiguous way of communicating the customer's overall assessment, yet considerable heterogeneity exists in customer interpretation and use of scales (Hu et al., 2017). The textual contents of customer reviews alleviate the ambiguity of the ratings by providing explanations and detailing the context for the ratings (Bi et al., 2019). The proposed approach includes (1) topic modeling to identify service attributes from customer review data; (2) index and sentiment analysis to measure the importance of service attributes and the focal company's performance in the same attribute; (3) clustering and the technique for order of preference by similarity to ideal solution (TOPSIS) to select competitors and best practices as benchmarks from the perspective of customers; and finally (4) importance-performance competitor analysis to develop a strategic action plan.

The proposed approach was applied to the customer review data collected from TripAdvisor. The case study of covering 26,934 customer reviews for 26 hotels confirms that the proposed approach enables quick identification of key aspects of the focal company's and competitors' services and assessment of the performance of the focal company and benchmarks within acceptable limits of time and cost. It is expected that the systematic process and quantitative outcomes offered by the proposed approach provide a valuable complementary tool for customer-oriented service benchmarking toward continuous service improvement.

## 5.2 Background

### 5.2.1 Customer survey-based approaches to service benchmarking

Prior studies have presented a variety of models and methods for measuring service performance and offering benchmarking guidelines. For instance, Min et al. (2002) identified a set of service benchmarks for monitoring service delivery processes and developed a service performance evaluation model using the analytic hierarchy process (AHP). Park et al. (2015) suggested a dual quality function deployment (QFD) by relocating the benchmarking matrix to the main body of the original QFD for comparative assessment of the activities of two firms (i.e., focal company and its competitor). Taplin (2012) presented a competitive importance-performance analysis, where the performance difference and the importance difference between the focal service and a competing service are located on the horizontal and vertical axis respectively. Similarly, Albayrak (2015) proposed an importance-performance competitor analysis that considers two types of gaps: the one is the gap between the importance of service attributes and the focal company's performance in the same service attribute, and the other is the gap between focal and competing companies' performance scores. Lee and Kim (2014) proposed the integrated approach of DEA and SERVPERF for identifying whom to benchmark and examining what degree service quality should be improved. Specifically, the proposed approach uses the five dimension values of SERVPERF models as outputs of the DEA to provide a single measure of overall service quality and benchmarking guidelines for inefficient decision-making units (i.e., service units).

However, while previous approaches have deepened our understanding of benchmarking in service industries, they are subject to certain limitations, as follows. First, previous approaches rely heavily on customer survey data, thereby time-consuming and labor-intensive (Ziegler et al., 2008). Second, the quality and reliability of analysis results strongly depend on survey questionnaires (e.g., contents, complexity, and length) and the willingness of respondents to participate (Groves, 2006).

Finally, previous approaches might miss unexpected but important service attributes that are difficult to identify in the design of service questionnaires. Hence, prior studies have faced serious challenges, especially when applied to the service sector where the number of services is huge, their complexity increasingly mounts, and customer needs change rapidly. These drawbacks provide our underlying motivation and are fully addressed in the proposed approach by employing large amounts of customer review data collected from the web as a source of the comprehensive VoC and developing a scientific method to enhance the efficiency and effectiveness of measuring service performance and offering benchmarking guidelines.

### 5.2.2 Customer review-based approaches to service benchmarking

The increased availability of online platforms and the introduction of intelligent computational algorithms have meant that benchmarking no longer depends solely on customer surveys but can exploit a great deal of information that can be collected on the web. Among others, customer review data has received increasing attention from researchers and industrial practitioners since it is not only publicly available, easily collected, low cost, spontaneous, and insightful, but also simple for firms to monitor and manage (Guo et al., 2017).

Existing customer review-based approaches can largely be classified into two categories. One stream of this research has utilized the rating information to identify the important service attributes and measure the competitiveness of service organizations. For instance, Xia et al. (2019) presented a kernel density estimation approach to identifying the critical service attributes that distinguish a hotel from its competitors and measuring the competitiveness of the hotel. Similarly, Xia et al. (2020) proposed an approach to measuring the competitiveness of a hotel brand from its competitors using the earth mover’s distance between the probability distributions of hotel feature ratings. Mariani and Visani (2020) developed a DEA model using three inputs (i.e. number of rooms, number of employees, net operating expenses) and two outputs (i.e. total revenues and online ratings) to measure the hotel’s operational efficiency and to offer benchmarking strategies. Although these studies have enabled the quick evaluation of service performance, as with customer-survey based approaches, they might miss unexpected but important service attributes since they measure the service quality using the ratings of pre-defined service attributes or financial and operational indicators.

Another approach has focused more on the contents of customer review data using text mining techniques to identify customer-oriented service attributes. Hu and Trivedi (2020) identified service attributes by annotating nouns and noun phrases though part-of-speech (POS) tagger and applying association rule mining to noun and noun phrases to find frequently mentioned service attributes in customer reviews. Bi et al. (2019) proposed a methodology for conducting the importance-performance analysis. Specifically, the service attributes were extracted through latent Dirichlet allocation (LDA) and the performance of each attribute was estimated through the support vector machine (SVM) based sentiment analysis. The importance of service attributes was measured via a neural network model linking the importance of service attributes derived via sentiment analysis and the overall rating of the service. However, while prior studies have proved useful for identifying important service attributes from customer review data, little attention has been paid to the selection of competitors (or best practices) as benchmarks in the benchmarking processes. The competitors were identified based solely on financial and/or operational indicators (e.g., star rating systems) (Albayrak, 2015; Bi et al., 2019; Fu et al., 2011; Min et al., 2002; Xia et al., 2019). Therefore, a major question remains regarding how to select competitors (or best practices) as benchmarks for customer-oriented service benchmarking.

Considering these issues, we present a data-driven approach to customer-oriented service benchmarking using large amounts of customer review data. The premise of this research is three-fold: (1) analysis of large-scale customer reviews can provide objective and reliable information on the key aspects of in-company and competitors' services (Bi et al., 2019); (2) analysis of the rating information of customer reviews can provide information about the overall customer satisfaction for service.; and finally (3) analysis of the contents of customer reviews can identify important service attributes and clues to the service performance based on customer experience and/or value creation.

### 5.2.3 Topic modeling

The topic model is an unsupervised machine learning technique that discovers the topic information in large-scale document collections or corpus (Blei et al., 2003). Approaches to topic modeling can be classified into two categories: probabilistic approaches (e.g. latent Dirichlet allocation (LDA)) and linear algebra approaches (e.g. non-negative matrix factorization (NMF)). First, LDA assumes that documents, a mixture of corpus-wide topics, exhibit multiple topics. Here topics are defined as a distribution of matrix of words, and words are drawn from one of those topics. LDA obtains the final outcomes of topic-word and document-topic distributions through a posterior maximization with Gibbs sampling. This kind of statistical learning is generally considered to perform quite well only on the condition that the corpus is statistically sufficient (Chen et al, 2019).

In contrast, NMF models the underlying components as coordinate axes and each document corresponds to a unique point in the latent linear space with a geometric perspective (Daniel et al., 1999). The text archive is usually firstly encoded in a term-document matrix with TF-IDF weights and then two non-negative matrices: term-topic and topic-document are sought with algorithms such as multiplicative update, in which each column of term-topic matrix can be viewed as a topic, and each column of topic-document can be treated as a compact embedding in the latent topic space. Chen et al. (2019) compared the derived topic quality of NMF and LDA on short text data and demonstrated that NMF is inclined to produce better topics than LDA. Considering that customer reviews are also short text documents, the proposed approach adopt NMF to identify service attributes. The graphical representation and equations are illustrated in the Section 5.3.2.

## 5.3 Proposed method

The overall process of the proposed approach is described in Figure 16. Given the complexities involved, the proposed approach is designed to be executed in five steps: (1) data collection and pre-processing; (2) identifying service attributes affecting the customer's perception of service performance with keyword dictionary; (3) measuring the importance of service attributes and the focal company's performance; (4) selecting competitors and best practices; and finally (5) developing a strategic action plan.

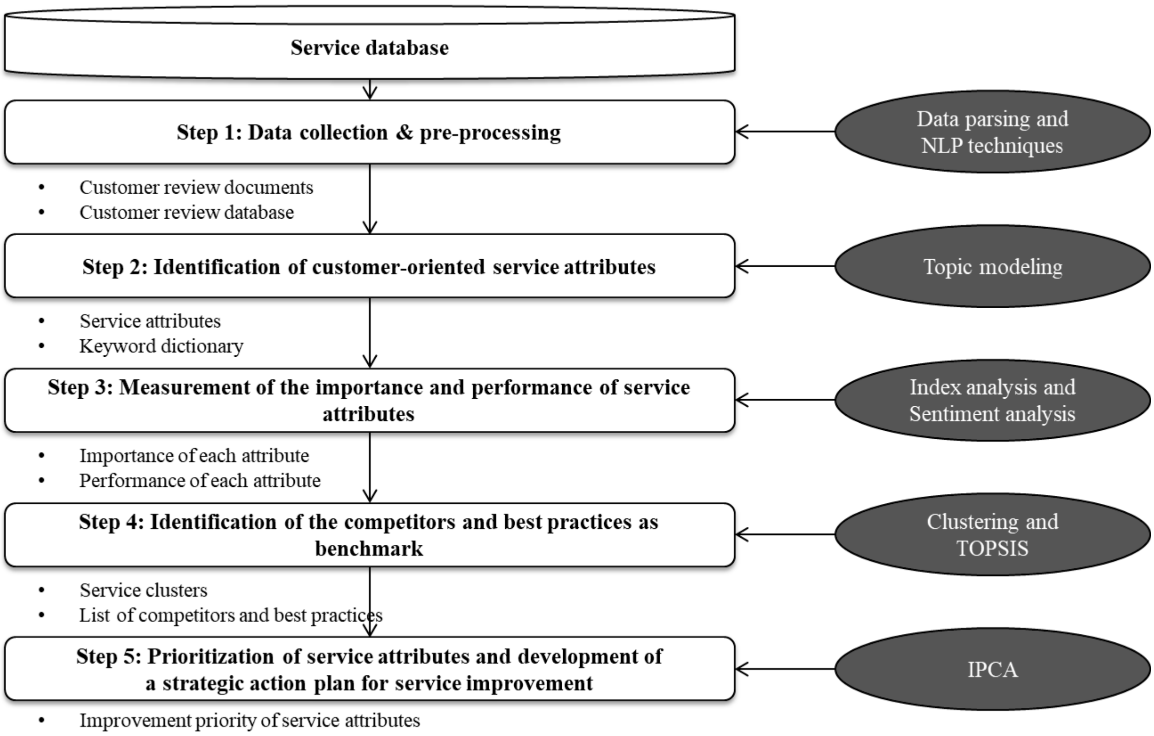



Figure 16. Overall process of the proposed approach

### 5.3.1 Step 1: Data collection and pre-processing

The customer review data can be obtained from various platforms such as online stores, forums, review sites, social media services, and blogs. Although the structure and contents of customer reviews may differ across the platforms, most platforms typically provide two components, i.e., ratings and textual contents that are the main inputs of the proposed approach. In this context, the customer reviews of focal service and relevant services are collected via web scrapping techniques and pre-processed via data parsing techniques. Here, the structured items that are consistent in semantics and formats (e.g., ratings and review date) can be easily stored in well-defined schemas such as relational databases while the textual contents that have different structures and styles need to be transformed for further analysis. For this, first, the textual contents are tokenized by splitting the text into smaller tokens. Second, all letters in the texts are converted to lowercase letters. Third, stop words, such as ‘a’ and ‘the’, and non-English words are removed. Fourth, part-of-speech (POS) tagging is performed to assign POS tags (e.g., noun and verb) to tokens. Finally, a lemmatization is executed to reduce inflectional forms of words and to return the dictionary forms of the words, which are known as the lemma. In Section 5.3.2 and 5.3.5, the results of POS tagging and lemmatization are used to construct a keyword dictionary and to develop a specific and practical action plan, while the raw data are used to extract sentence referring to specific attribute in Section 5.3.3.

### 5.3.2 Step 2: Identification of consumer-oriented service attributes

This step identifies service attributes affecting the customer’s perception of service performance with the keyword dictionary to assign customer reviews to the relevant service attributes and further to evaluate the service performance based on customer experience and/or value creation. Service attributes are defined as the dimensions and/or features that constitute a service, and the keyword dictionary is a collection of the keywords that indicate specific service attributes in customer reviews.

The process of identifying service attributes with the keyword dictionary consists of two sub-steps. First, the initial version of service attributes with the keyword dictionary is defined through topic modeling techniques. Approaches to topic modeling can be classified into two categories: probabilistic approaches (e.g. latent Dirichlet allocation (LDA)) and linear algebra approaches (e.g. non-negative matrix factorization (NMF)). The basic concept of probabilistic approaches is that a document is a mixture of topics and each word’s presence is attributable to one of the document’s topics with certain probabilities (Blei et al., 2001). The topic-word and document-topic distributions are obtained through a posterior maximization with Gibbs sampling. The linear algebra approaches project term-document matrix into a low dimensional subspace through linear or nonlinear transformation. In the topic space, each coordinate axis corresponds to a topic, and each document is represented as a linear combination of these topics (Lee and Seung, 1999). Prior empirical studies on the performance of these two approaches presented that linear algebra approaches tend to outperform than probabilistic approaches on short text datasets (Chen et al., 2019). Considering that customer reviews are also short text documents, we adopt linear algebra approaches to identifying service attributes.

The NMF algorithm is employed since it constructs a non-negative parts-based representation of data, which is natural for interpreting the textual representation (Tsarev et al., 2011). Given $N$ customer reviews, each review $d_n$ is linearly combined by $K$ components $u_k$ with coefficients $v_{kn}$, where $k = 1, 2, \dots, K$, and $n = 1, 2, \dots, N$ $(\text{i.e.}, d_n = \sum_{k=1}^{K} u_k v_{kn})$. This algorithm minimizes the square loss between the original term-document vector $d_n$ and the linear combinations $\sum_{k=1}^{K} u_k v_{kn}$, as shown in Eq. (22).

$$\min \sum_{n=1}^{N} \left\| d_n - \sum_{k=1}^{k} u_k v_{kn} \right\|_2^2 \qquad \text{s.t.} \begin{cases} v_{kn} \geq 0 \\ u_{mk} \geq 0 \end{cases} \text{------------------------------------------ Eq. (22)}$$

where $\|.\|_2^2$ denotes the square of a vector's L2-norm. The Eq. (22) is transformed into a compact version as shown in Eq. (23).

$$\min \| D - UV \|_F^2 \qquad \text{s.t.} \begin{cases} V \geq 0 \\ U \geq 0 \end{cases} \text{--------------------------------------------- Eq. (23)}$$

where $D = [d_1, d_2, \dots, d_n] \in R^{M \times N}, U = [u_1, u_2, \dots, u_K] \in R^{M \times K}$ and $V = [v_1, v_2, \dots, v_N] \in R^{K \times N}$ and $\|.\|_F^2$ represents the square of a matrix's Frobenius norm, as represented in Figure 17. As to the learning algorithm, the multiplicative update is designed as in Eqs. (24) and (25).

$$U \leftarrow U \frac{DV^T}{UVV^T} \text{------------------------------------------------------------------------------------------- Eq. (24)}$$

$$V \leftarrow V \frac{U^T D}{U^T UV} \text{------------------------------------------------------------------------------------------- Eq. (25)}$$

Eqs. (24) and (25) iterate until convergence and we obtain the term-topic matrix U and topic-document matrix V. The initial version of service attributes with the keyword dictionary is constructed by assigning a label to each topic, linking the labels of the topics to service attributes, and including the terms in the keyword dictionary.

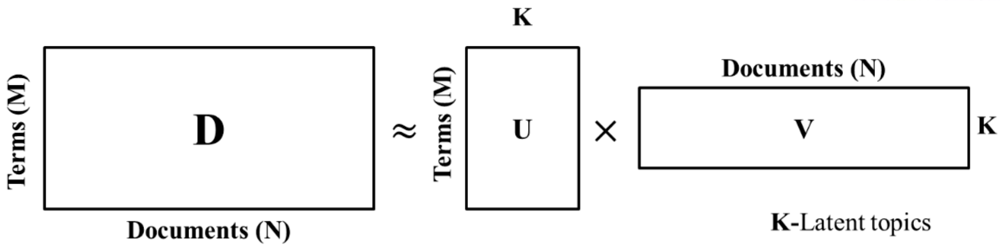



Figure 17. Graphical representation of non-negative matrix factorization (NMF)

Second, the initial version might miss the service attributes and terms that are not frequently occurred in customer reviews and might include some uninterpretable and duplicate topics, thereby needing manual examination by domain experts. For this reason, experts should adjust the initial version of service attributes with the keyword dictionary by identifying the interpretable topics, merging the topics with similar attributes, adding topics and terms, and deleting noise and duplicate terms. Here, we provide experts with the results of literature reviews and keyword analysis as supplementary materials. Although most service attributes are expressed as nouns and noun phrases, the forms of adjectives and verbs should be taken into account together. For instance, ‘cleanliness’ in many cases is represented as ‘clean’ or ‘dirty’ in customer reviews. Hence, the keyword analysis considers the forms of adjectives and verbs as well as abbreviation and synonym to diversify the scope and to enhance the reliability of the analysis.

#### 5.3.3 Step 3: Estimation of the importance and performance of each attribute

This step employs sentiment analysis to measure service performance at the service attribute-level using the contents of customer reviews and index analysis to identify the importance of service attributes using the ratings of customer reviews. First, in terms of the service performance at the service attribute-level, among different models for sentiment analysis (e.g., linguistic inquiry and word count (Tausczik and Pennebaker, 2010), semantic orientation calculator (SO-CAL) (Taboada et al., 2011), and positive and negative affect schedule (PANAS) (Gonçalves et al., 2013)), we employ the valence aware dictionary and sentiment reasoner (VADER) since this method was built on microblogs like Twitter (Part et al., 2015) and thus performs well with short texts such as customer reviews. Moreover, this method considers emoticons, acronyms, and initializes that occur frequently in customer reviews to measure sentiment intensity.

Specifically, the VADER is based on a gold-standard list of lexical features and examines five grammatical and syntactical conventions to measure the sentiment scores: (1) punctuations such as exclamation points increase the magnitude of the sentiment intensity; (2) capitalization increases the magnitude of the sentiment intensity; 3) degree modifiers impact sentiment intensity by either increasing or decreasing the intensity; 4) the contrastive conjunctions, such as but and however, signal a shift in sentiment polarity; and 5) the preceding tri-gram identifies the cases where negation flips the polarity of the text. Building on this concept, the procedure for measuring service performance at the service attribute-level consists of three sub-steps. First, customer reviews are linked to the relevant service attributes by matching the keyword dictionary constructed in the prior step, as shown in Table 15.

Table 15. Example of the customer review-service attribute matrix

| Service | Reviews | Attributes | | | Overall rating |
|---|---|---|---|---|---|
| | | Food | Facility | Wi-Fi | |
| 1 | 1 | Delicious breakfast | - | No Wi-Fi | 5 |
| | 2 | - | Swimming pool is awesome | Wi-Fi is good and free | 4 |
| | 3 | Terrible food | Perfect pool | - | 3 |
| | 4 | Breakfast was delicious | Noisy and dangerous gym | - | 5 |

Second, the sentiment scores of customer reviews for the relevant service attributes are calculated as the sentiment scores of the relevant sentences of the customer reviews corresponding to the relevant service attributes. Specifically, four different types of sentiment scores are measured at the service attribute-level as in Eqs. (26)-(29).

$$Positive\ score_{i,j} = \frac{Lexicon\ and\ rule\text{-}based\ positive\ score_{i,j}}{Total\ absolute\ sentiments_{i,j}} \text{------------------------------------Eq. (26)}$$

$$Neutral\ score_{i,j} = \frac{The\ number\ of\ neutral\ words_{i,j}}{Total\ absolute\ sentiments_{i,j}} \text{--------------------------------------------Eq. (27)}$$

$$Negative\ score_{i,j} = \frac{|Lexicon\ and\ rule\text{-}based\ negaitive\ score_{i,j}|}{Total\ absolute\ sentiments_{i,j}} \text{-------------------------------Eq. (28)}$$

$$Compoud\ score_{i,j} = Normalize(Lexicon\ and\ rule - based\ positive\ score_{i,j} + Lexicon\ and\ rule - based\ negaitive\ score_{i,j}), \text{-------------------Eq. (29)}$$

where $total\ absolute\ sentiments_{i,j}$ is the sum of the lexicon and rule-based positive scores, the number of neutral words, and the absolute value of lexicon and rule-based negative scores for the $i$th customer review for the $j$th attribute.

Third, given that the service performance cannot be ascertained fully from customer review data, customer reviews are classified as positive or negative at the service attribute-level to reduce the measurement error. For classification, we used the compound score that is the normalized overall sentiment score ranging from -1 (most extreme negative) to 1 (most extreme positive). The resulting customer review-attribute performance matrix is constructed as exemplified in Table 16. $P_{i.j,m}$ is defined as the performance of the $j$th service attribute in the $m$th customer review for $i$th service. $P_{i.j,m}$ is 1 if the compound score is positive, -1 if the compound score is negative, and 0 if $j$th service attribute is not mentioned in the $m$th customer review.

Table 16. Example of the customer review-attribute performance matrix

| Service | Reviews | Attributes | | | Overall rating |
|---|---|---|---|---|---|
| | | Food | Facility | Wi-Fi | |
| 1 | 1 | 1 | 0 | -1 | 5 |
| | 2 | 0 | 1 | 1 | 2 |
| | 3 | -1 | 1 | 0 | 3 |
| | 4 | 1 | -1 | 0 | 5 |

Finally, the service attribute performance matrix is constructed, as shown in Eq. (30).

$$P_{i,j} = \frac{\sum_{m=1}^{M} P_{i.j,m}}{r_{i,j}} \text{------------------------------------------------------------------------------------Eq. (30).}$$

where $r_{i,j}$ represents the number of customer reviews for service $i$ for service attribute $j$. For instance, the service attribute performance for food attribute in Table 17 is $P_{1,Food} = \frac{1-1+1}{3} = 0.33.$

Table 17. Example of the service attribute performance matrix

| Service | Attributes | | |
|---|---|---|---|
| | Food | Facility | Wi-Fi |
| 1 | 0.33 | 0.33 | 0 |

Second, two approaches can be used for estimating the importance of service attributes. First, the self-state approach is to estimate the explicit importance of service attributes through customer surveys. That is, customers are asked to evaluate the importance of service attributes using a Likert scale. Although it is simple and practical, this approach has a limitation in that the self-stated importance of service attributes is influenced by the performance of the service attributes (Lai and Hitchocock, 2016; Matzler et al., 2003). Second, the implicit approach estimates the importance of service attributes by liking the overall satisfaction and the performance of the service attributes using regression models (Danaher and Mattsson, 1994; Rust et al., 1994; Van et al., 2007) and conjoint analysis. However, the importance of service attributes derived by regression models is likely to be biased and misleading (Deng et al., 2008) since the assumptions associated with regression models are mostly violated in customer satisfaction research (Garver, 2002). In this context, new methods which are based on a conjoint analysis were proposed (Danaher, 1997; DeSarbo et al., 1994). Conjoint analysis has typically been used for new product development (Green and Srinivasan, 1990). However conjoint analysis has also been used in service industry setting to measure the relative importance of service attributes measured in customer satisfaction (Danaher, 1997; DeSarbo et al., 1994). Conjoint approach enable us to model a response surface for satisfaction across the full range of service attribute levels, at either the individual or overall service level (Danaher, 1997). Adopting the implicit approach and borrowing the concept of importance in conjoint analysis, we measure the importance of service attribute $j$ of service $i$ ($SAI_{i,j}$) as the difference in the mean of overall rating scores between positive and negative reviews, as shown Eq. (31).

$$SAI_{i,j} = \frac{\sum_{n=1}^{N} overall\ rating}{N_{ij}} - \frac{\sum_{m=1}^{M} overall\ rating}{M_{ij}} \text{------------------------------------------------Eq. (31)}$$

where $SAI_{i,j}$ represents the importance of service attribute $j$ of service $i$. $N$ and $M$ are the numbers of positive reviews and negative reviews, respectively. For example, the importance of service attribute *food* in Table 18 is $SAI_{1,Food} = \frac{10}{2} - \frac{3}{1} = 2$. If there is no positive (negative) review for service attribute $j$ of service $i$, the average of the overall rating scores for service $i$ is used instead. If there is no review for service attribute $j$ of service $i$, $SAI_{i,j}$ is set to 0. The constructed service attribute importance matrix is normalized using min-max feature scaling to range from 0 to 1 for each service, which is then used as the input of $k$-means clustering in Step 5.3.4.

Table 18. Example of the service attribute importance matrix

| Service | Attributes | | |
|---|---|---|---|
| | Food | Facility | Wi-Fi |
| 1 | 2 | -1.5 | -1 |

### 5.3.4 Step 4: Identification of the competitors and best practices as benchmark

This step employs $k$-means clustering to categorize services into groups according to the importance of service attributes and TOPSIS to select best practices as benchmarks. $k$-means clustering is a method that aims to partition $n$ observations into $k$ clusters in which each observation belongs to the cluster with the nearest mean that serves as a prototype of the cluster (Hartigan and Wong, 1979). Specifically, given a set of observation $(x_1, x_2, \ldots, x_n)$, they are partitioned into $k(\leq n)$ sets **S**=$\{S_1, S_2, \ldots, S_k\}$ to minimize the within-cluster sum of squares (WCSS), as shown in Eq. (32).

$$\arg\min \sum_{i=1}^{k} \sum_{x \in S_i} \|x - \mu_i\|^2 = \arg\min \sum_{i=1}^{k} |S_i| Var S_i \text{------------------------------------------Eq. (32)}$$

where $\mu_i$ is the mean of points in $S_i$. The elbow method – that plots the value of distortion produced by different values of $k$ and finds the optimal value of $k$ at which improvement in distortion declines the most – is used to determine the number of clusters. As a consequence, $k$ service groups with similar importance distributions of service attributes are identified and services belonging to a group are considered as competitors.

TOPSIS is a multi-criteria decision-making method based on the concept that the chosen alternative should have the shortest geometric distance from the positive ideal solution and the longest geometric distance from the negative ideal solution. TOPSIS has been used to evaluate and rank the performance of firms (Ertuğrul and Karakaşoğlu, 2009) and the sustainable business excellence (Metaxas et al., 2016). The advantages of this method lie in its simplicity and ability to consider a non-limited number of alternatives and criteria (Hung and Cheng, 2009). Moreover, this method measures and ranks the services considering the capacity limitation for each service attributes (Ibrahim et al., 2019). The process of ranking competitors (i.e., services belonging to a group) consists of seven steps. First, the service attribute-performance matrix for group $c$, $P_{c_i,j}$, is normalized to form matrix using the normalization method $n_{i,j} = P_{c_i,j} / \sum_{c_i=1}^{c_N} {P_{c_i,j}}^2$, where $c_N$ is the number of services in group $c$. Second, the weighted normalized decision matrix is constructed as $t_{c_i,j} = n_{c_i,j} \cdot w_j$, where $w_j = W_j / \sum_{j=1}^{J} W_j,\ j = 1, 2, \ldots, J$ so that $\sum_{j=1}^{J} w_j = 1$, and $W_j = \frac{\sum_{c_i=1}^{c_N} SAI_{c_i,j}}{c_N}, j = 1, 2, .., J$. Third, the best ($A_b$) and worst ($A_w$) hypothetical services are determined as in Eqs. (33) and (34).

$$A_w = \min\left(t_{c_i,j} \middle| c_i = 1,2,\ldots,c_N\right) \equiv \{t_{w,j} \mid j = 1,\ 2,\ \ldots,\ J\} \text{--------------------------------------Eq. (33)}$$

$$A_b = \max\left(t_{c_i,j} \middle| c_i = 1,2,\ldots,c_N\right) \equiv \{t_{b,j} \mid j = 1,\ 2,\ \ldots,\ J\} \text{--------------------------------------Eq. (34)}$$

Fourth, $L^2$-distances between the service $c_i$ and the best and worst hypothetical services are calculated as shown in Eqs. (35) and (36).

$$d_{c_i,w} = \sqrt{\sum_{j=1}^{J}(t_{c_i,j} - t_{w,j})^2}\ ,\ c_i = 1,\ 2,\ \dots,\ c_N \text{----------------------------------------------Eq. (35)}$$

$$d_{c_i,b} = \sqrt{\sum_{j=1}^{J}(t_{c_i,j} - t_{b,j})^2,}\ \ c_i = 1,\ 2,\ \dots,\ c_N \text{-----------------------------------------------Eq. (36)}$$

Fifth, the performance of services is calculated as $s_{c_i,w} = \frac{d_{c_i,w}}{d_{c_i,w}+d_{c_i,b}}$ . Finally, the best practices are identified as the services outperforming the focal service. If the focal service ranks first, then the services with high ranks are identified as the best practices.

### 5.3.5 Step 5: Prioritization of service attributes and development of strategic action plan for service improvement

This section determines the priority of the focal service's attributes for improvement via importance-performance competitor analysis (IPCA). IPCA has originated from importance-performance analysis (IPA) that formulates improvement priorities based on two dimensions: the importance of service attributes and the performance of focal service for the service attributes (customer satisfaction) (Martilla and James, 1977). To overcome the limitations of IPA (Chen, 2014; Keyt et al., 1994), IPCA considers the importance of service attributes and the performance of the focal service's and competitors' attributes together (Albayrak, 2015).

As Figure 18 shows, the IPCA plot examines two types of gaps: one is the gap between the focal service's performance and the importance for the attribute, and the other is the gap between focal and competing companies' performance scores for the attribute (i.e. performance difference). Specifically, the gap between an attributes' performance and importance is calculated as in Eq. (37).

$$GAP_{focal,j} = \overline{P_{focal,J}} - \overline{SAI_{focal,J}},\ \mathrm{j} = 1, 2, \dots, \mathrm{J} \text{-------------------------------------------------Eq. (37)}$$

where $\overline{P_{focal,J}}$ and $\overline{SAI_{focal,J}}$ represents the normalized performance and importance of the $j$th attribute for the focal service, i.e., $\overline{P_{focal,J}} = P_{focal,j}/\sum_{j=1}^{J} P_{focal,j}$ and $\overline{SAI_{focal,J}} = \frac{SAI_{focal,j}}{\sum_{j=1}^{J} SAI_{focal,j}}, \mathrm{j} = 1, 2, \dots, \mathrm{J}$. The performance difference between the focal service and benchmark services is calculated by Eq. (38)

$$PD_{focal,j} = P_{focal,j} - P_{benchmarks,j}, \ \mathrm{j} = 1, 2, \ldots, \mathrm{J} \text{----------------------------------------------Eq. (38)}$$

Here, $P_{benchmarks,j}$ denotes the mean value of benchmark services' performance (i.e., $P_{benchmarks,j} = \frac{\sum_{b_i=1}^{B_N} P_{b_i,j}}{B_N}$, where $\mathrm{j} = 1, 2, \ldots, \mathrm{J}$ and $B_N$ is the number of benchmarks selected in the preceding section.

Each attribute can be classified into one of the four areas of the IPCA plot: *solid competitive advantage*, *head-to-head competition*, *urgent action*, and *potential competitive edge*. First, the attributes associated with *solid competitive advantage* are regarded as the focal service's strengths showing higher performance than those of competitors and higher performance than the importance of attributes. The focal firm should keep its current performance level for these attributes. However, improving the quality of these attributes may not lead to a real advantage since the focal company presents higher performance than the importance of the attributes. Second, the focal firm presents lower performance than its competitors for the attributes in the *head-to-head competition* quadrant. Third, the attributes categorized as *urgent action* are the attributes that the focal company presents lower performance than competitors and their importance. These attributes considered as weaknesses where the focal company should take urgent action to improve them. Finally, the focal company outperforms competitors for the attributes classified as *potential competitive edge* although the focal company presents lower performance than the importance of the attributes. Improving the quality of these attributes might create a new competitive edge for the focal company.

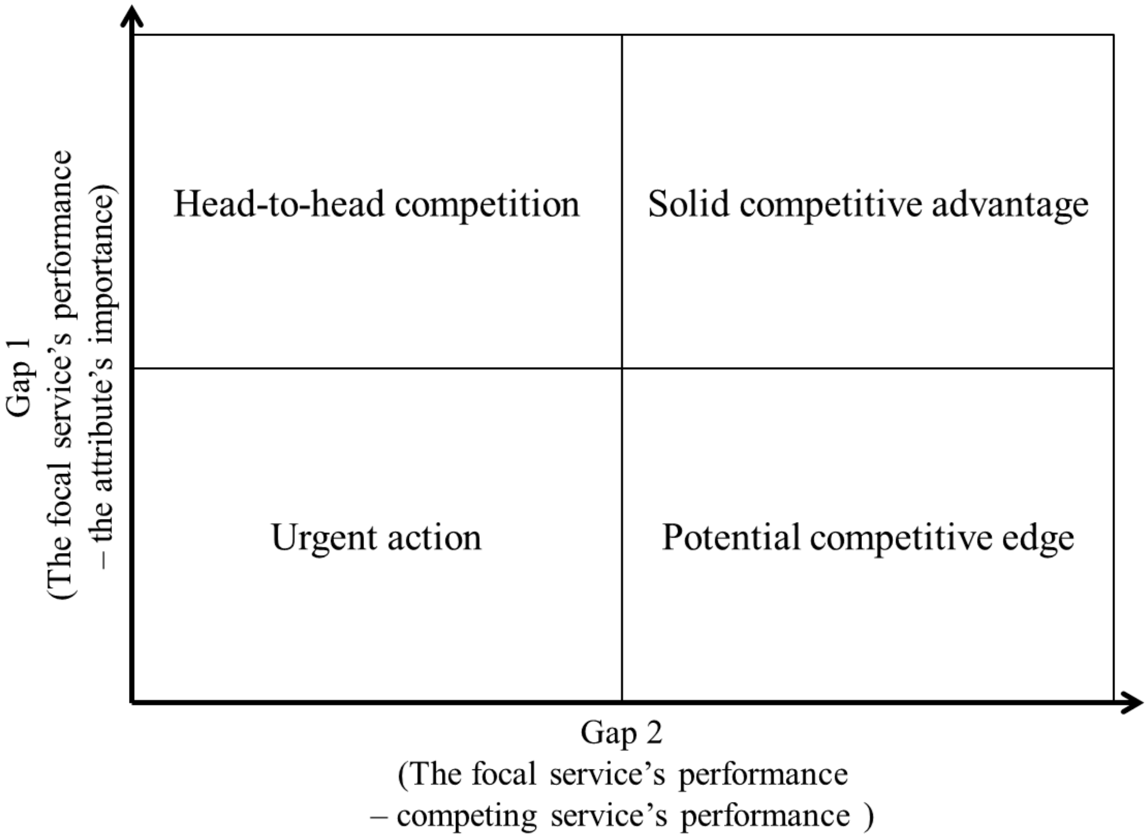



Figure 18. The importance-performance competitor analysis (IPCA) model

## 5.4 Case study

In this section, a case study of the hotels in Bangkok, the capital of Thailand, is presented to illustrate the feasibility of the proposed approach. Bangkok is one of the world's top tourist destination cities. Each year approximately 22.7 million international visitors arrive in Bangkok (Alexandra, 2019). The revenue of the hotels in Bangkok accounted for about 29.96% of the total hotel revenue in Thailand (National Statistical Office, 2013). However, there are more than 10,000 registered hotels in Bangkok, and these hotels strive to develop improvement strategies to survive in the oversaturated and fiercely competitive market (Sritama, 2018).

### 5.4.1 Step 1: Data collection and pre-processing

Customer review data were collected from *Tripadvisor* (https://www.tripadvisor.com) for the following two reasons. First, *Tripadvisor* is one of the world's most popular web sites for travel accommodations (Law, 2006). This web site had an average of more than 490 million monthly visitors and a total of more than 730 million reviews and opinions by the end of 2018 (Taecharungroj and

Mathayomchan, 2019). Second, the database is well organized in terms of search conditions and reliability, providing diverse information including overall ratings, dates of stay, and review comments in electronic format.

For 26 hotels, a total number of 26,934 online reviews posted between July 2016 and July 2019 were collected, as shown in Table 19. Each hotel has at least 100 reviews written in English posted during this period. For the review contents, the 'natural language toolkit' (NLTK) in python, which is a suite of libraries and programs for natural language processing for text written in English, was used for pre-processing the textual data such as tokenization, stop words removal, POS tagging, and lemmatization. The resulting customer review database includes the structural items, such as hotel identifiers, review identifiers, dates of stay, overall ratings, and the unstructured items, such as review contents and pre-processed content information. The database is not reported here in its entirety owing to lack of space, but a part of the database is shown in Table 20

Table 19. The number of customer reviews for 26 hotels

| Identifier | Number of reviews | Identifier | Number of reviews |
|---|---|---|---|
| S2-a | 741 | S4-b | 726 |
| S2-b | 312 | S4-c | 833 |
| S2-c | 638 | S4-d | 1,887 |
| S2-d | 416 | S4-e | 1,461 |
| S2-e | 115 | S4-f | 2,291 |
| S3-a | 324 | S4-g | 1,955 |
| S3-b | 320 | S5-a | 3,229 |
| S3-c | 363 | S5-b | 1,541 |
| S3-d | 450 | S5-c | 956 |
| S3-e | 193 | S5-d | 1,450 |
| S3-f | 316 | S5-e | 1,300 |
| S3-g | 298 | S5-f | 1,309 |
| S4-a | 1,180 | S5-g | 2,330 |

.

Table 20. Part of the customer review database

| Service No. | Review No. | Date of stay | Rating | Title | Review content (raw data) | Review content (Pre-processed data) |
|---|---|---|---|---|---|---|
| S2-a | 1 | 2019-07 | 5 | Lovely place! | This is a great place, friendly staff, conveniently located, safe and is highly recommended for anyone open to sharing. … | great place friendly staff conveniently locate safe highly recommend anyone open sharing … |
| S2-a | 2 | 2019-06 | 5 | Marvelous cabin territory | Thailand is a wonderful country and I get a kick out of the opportunity to explore each spot. A month prior I visited the said country and booked in Bed Station Hostel…. | thailand wonderful country get kick opportunity explore spot month prior visit say country book bed station hostel … |
| S2-a | 3 | 2019-06 | 5 | Best hotel in Bangkok | Lovely hotel that is clean and has good amenities such as a gym, PlayStation and tv, films, books, cheap water, a bar and events … | lovely hotel clean good amenity gym playstation tv film book cheap water bar event … |
| S2-a | 4 | 2019-05 | 4 | Bed Station Simply the Best | Located @strategist location, convinces, easy to connect to where ever i go and planned. Bed Station make me feel home and warm all times as when in enter in to the premises. … | locate strategist location convinces easy connect ever go planned bed station make feel home warm time enter premise … |
| S2-a | 5 | 2019-05 | 5 | Great small hostel | Awesome little place which is near all the major trains and also the canal for the boat into the old town. The rooms are phenomenal and a huge improvement from when i was in south east asia last about 15 years ago for 6 months. … | awesome little place near major train also canal boat old town room phenomenal huge improvement south east asia last year ago month … |
| S2-a | 6 | 2019-05 | 4 | Feel good | Would like to visit again next time if i have a trip in Thailand. The location is near BTS station that only needs walking about 2 mins to the elavator of BTS's. … | would like visit next time trip thailand location near bts station need walk min elavator bts … |
| … | … | … | ... | … | … | … |
| S5-g | 26929 | 2016-07 | 5 | Amazing stay | Walking into a gaw dropping lobby is just the beginning. Every single process here is very enjoyable from checking in to dining and even grabbing a taxi. … | walk gaw drop lobby begin every single process enjoyable checking din even grab taxi … |

| | | | | | | |
|---|---|---|---|---|---|---|
| S5-g | 26930 | 2016-07 | 5 | Beautiful service in land of smiles | I come to Bangkok frequently and Plaza Athenee is always my hotel of choice.Things that really get me (in a good way)1) Rooms are new and a few times I got upgraded.2) Staff greet me by name when I arrive at the hotel. Sure makes me feel good to be remembered. 3) Comfortable Club Lounge with great service. … | come bangkok frequently plaza athenee always hotel choice thing really get good way room new time get upgraded staff greet name arrive hotel sure make feel good remember comfortable club lounge great service … |
| S5-g | 26931 | 2016-07 | 1 | Almost an unpleasant night... | We were greeted by a lady in a black suit. She greeted us nicely, at first, asking if we had booked the table. I, then, told her my name. However, she replied that she couldn't find our name on her list. I am not going to exaggerate, but her change of attitude towards us was very noticeable. The tone of voice and facial expression, which was very polite before, turned rigid and almost snappy. … | greet lady black suit greet nicely first ask book table told name however reply find name list go exaggerate change attitude towards noticeable tone voice facial expression polite turn rigid almost snappy … |
| S5-g | 26932 | 2016-07 | 4 | Good service staff | I get a good service from the stuff(especially Seangwanich), they are friendly and understanding. The room is nice and the facility is well maintained … | get good service stuff especially seangwanich friendly understanding room nice facility well maintain |
| S5-g | 26933 | 2016-07 | 4 | Coffee at the Bakery | This is my first time visiting at the Bakery at Phaza Athenee Bangkok Hotel. Overall decoration is so fine and nice that you feel relaxing with a cup of Illy coffee, Italian blended coffee. … | first time visit bakery phaza athenee bangkok hotel overall decoration fine nice feel relax cup illy coffee italian blend coffee … |
| S5-g | 26934 | 2016-07 | 5 | Perfect Building, Perfect Staff, … | The hotel is incredibly beautiful, having previously been a palace for a Thai Prince. The architecture is stunning, and the service matches. The staff is helpful, courteous, and professional. … | hotel incredibly beautiful previously palace thai prince architecture stun service match staff helpful courteous professional … |

#### 5.4.2 Step 2: Identification of consumer-oriented service attributes

We defined the initial version of service attributes with the keyword dictionary from the review contents of the database using 'NMF' module in 'sklearn' package implemented in Python. Here, quantitative measurements such as perplexity and topic coherence can help determine the number of topics, yet qualitative experts' judgments are more flexible in practice (Chang et al., 2009). Moreover, the process should be conducted manually in that the number of topics may be subjective to the context of service benchmarking For instance, if a company carries out explorative research for radical changes, the service attributes should be defined at the macro-level. In contrast, if a company pursues minor changes, the service attributes should be defined at the micro-level, and specific service attributes will give a practical solution by including specific service attributes. Considering these issues, qualitative judgments by domain experts were employed to determine the number of topics, and the measurements of perplexity and topic coherence were provided as supplementary materials in this process. As a result, the number of topics was determined as 25, which produced the most interpretable results.

The labeling of topics was conducted by one expert and confirmed by two other experts to increase the reliability of the research. The task was based on the identification of a logical connection between the most frequent words and a topic. For instance, the topic name "Facilities" is labeled based on the words 'pool', 'swim', and 'gym'. Of the 25 topics, 6 uninterpretable and 4 repeated topics were removed after experts' examination. Finally, 7 service attributes were identified by grouping the remaining 15 topics according to their characteristics, as shown in Figure 19.

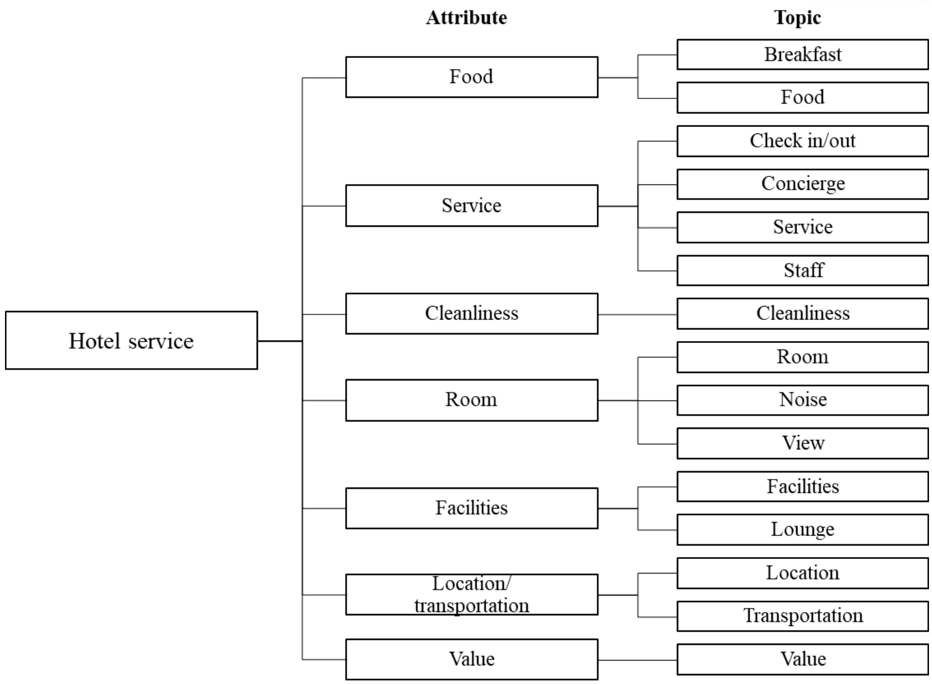



Figure 19. Service attributes and topics derived via NMF

The initial version of service attributes with the keyword dictionary was complemented through literature review and keyword analysis. Prior studies have presented the different attributes to describe key aspects of hotel services, as summarized in Table 22. For instance, using LDA techniques, Bi et al. (2019) identified nine attributes, that is, value, location/transport, room, cleanliness, service, food/drink, check-in/out, facility, and Wi-Fi/internet. Similarly, Guo et al. (2017) identified six attributes i.e., cleanliness, service, location, room, and value, and Mankad et al. (2016) identified amenities, location, transactions, value, and experience. Through literature review, Tsang and Qu (2000) identified 35 attributes including food, service, room, facility, location, cleanliness, value, and security through literature review. Chu and Choi (2000) identified hotel 26 attributes including food, service, room, facility, location, cleanliness, value, internet, and security via literature review and three focus group discussions. Building upon these studies, “Wi-Fi” and “security” attributes were added to the initial version of service attributes.

As for keyword analysis, first, the relevant keywords describing each attribute were identified by examining keywords in the form of nouns, noun phrases, adjectives, and verbs in the pre-processed contents. Second, the abbreviations and synonyms of keywords were identified via WordNet and. Finally, these keywords were added to the keyword dictionary for service each attribute, as shown in Table 21

Table 21. Keyword dictionary employed in this study

| Number | Attribute | Words |
|---|---|---|
| 1 | Food | Food, restaurant, buffet, delicious, dinner, breakfast, dish, fruit, drink, juice, dining, meal |
| 2 | Service | Concierge, staff, friendly, helpful, help, book, booking, check, ask, reception |
| 3 | Room | Bathroom, bedroom, noise, loud, loudly, bed, lamp, light, temperature, view |
| 4 | Facility | Facility, gym, pool, lounge, club, executive, lobby, bar, rooftop |
| 5 | Location | Bts, tuk, subway, train, taxi, walk, distance, mall, shop, airport, station, transport |
| 6 | Cleanliness | Clean, dirty, bug, cleanliness, cleaning, cleaner, dust, sheet, tidy |
| 7 | Value | Value, price, money, point, expensive, inexpensive, cost |
| 8 | Wi-Fi/Internet | Internet, wifi, wireless, computer, laptop, bandwidth |
| 9 | Security | Security, safe, lock, locker, guard, privacy, cctv |

Table 22. Service attributes employed in this study

| | Attribute | Description | Bi et al. (2019) | Guo et al. (2017) | Mankad et al. (2016) | Tsang & Qu (2000) | Chu & Choi (2000) |
|---|---|---|---|---|---|---|---|
| Initial service attributes | Food | Related to food and dining | √ | √ | | √ | √ |
| | Service | Related to service of staffs | √ | √ | √ | √ | √ |
| | Room | Related to room size, noise, and view | √ | √ | | √ | √ |
| | Facility | Related to lounge, gym, and pool | √ | √ | | √ | √ |
| | Location | Related to location and transportation | √ | √ | √ | √ | √ |
| | Cleanliness | Related to cleanliness and bug | √ | | | √ | √ |
| | Value | Related to value of room rate | √ | √ | √ | √ | √ |
| Complemented attributes by literature | Wi-Fi / Internet | Related to Wi-Fi and internet | √ | | √ | | √ |
| | Security | Related to privacy and safe | | | | √ | √ |

### 5.4.3 Step 3: Estimation of the importance and performance of each attribute

The contents of customer reviews may contain several sentences concerning different hotel attributes. For this reason, the contents of customer reviews were first split into sentences by using "sent_tokenize" module in "nltk" package in Python that divides a text into a list of sentences based on punctuation following sentences, abbreviation words, collocations, and words that start sentences. The customer reviews were then coupled with relevant attributes via a string-matching technique.

After assigning review sentences to the relevant attributes, the service performance and the importance of attributes were measured. First, in terms of the service performance at the service attribute-level, the four sentiment scores for each sentence of a customer review were measured by using "vader" module in "nltk" package, as shown in Table 23. For instance, for the review "Stayed for 5 nights, met a lot of people. Room is clean and tidy, shower clean, laundry services available. Biggest disappointment very boring and repetitive breakfast option - either cereal or toast", the *room* and *clean* attributes were classified as positive while the *food* attribute was classified as negative. Out of 88,989 sentences, a total of 73,261 and 15,728 sentences were identified as positive and negative sentences on the relevant service attributes according to their compound scores. The resulting customer review-attribute performance matrix is not reported here in its entirety owing to lack of space, but a part of the matrix is shown in Table 24. Moreover, the service attribute performance matrix was constructed by aggregating the attribute performance at the customer review level to the performance at the service level, as shown in Table 25.

Second, the importance of service attributes was measured based on the difference in the average of overall rating scores between positive and negative reviews and normalized to adjust different scales and to range from 0 to 1 as shown in Table 26. The proposed indicator is based on the premise that the overall rating scores correspond to the customer's overall satisfaction with the service. To confirm this, the Spearman correlation analysis between overall rating scores and compound scores calculated via VADER was performed. The Spearman correlation coefficient is a nonparametric measure of statistical dependence between two variables and applicable for both continuous and discrete ordinal variables (Croux and Dehon, 2010). It has a value between 1 and -1, where 1 is means that as the value of one variable increases, so does the value of the other variable. The result verified the positive relationship between overall ratings and compound scores with the 0.32 correlation coefficient and p-value less than 0.001. According to the average of the importance of attributes over 26 hotels in Bangkok, the attributes are considered important in order of *cleanliness, service, room, Wi-Fi, security, value, facility, food*, and *location*, which is in line with the results of existing studies that the satisfaction of customers staying at the hotels in Bangkok is most affected by cleanness (Cherdchamadol and Sriboonjit, 2017).

Table 23. Part of the results of sentiment analysis

| Review contents | Positive | Neutral | Negative | Compound | Classification |
|---|---|---|---|---|---|
| What I liked the Best about Red Planet Hotel room 517 was the efficient and courteous, thoughtful and pleasant attitude of staff. | 0.543 | 0.457 | 0 | 0.958 | Positive |
| Great service and friendly staff- were happy to help and leant us a plug adapter. | 0.632 | 0.368 | 0 | 0.958 | Positive |
| Our room was perfectly clean, the bed very comfortable, and everything is provided to make your stay relaxed, easy and soothing. | 0.557 | 0.443 | 0 | 0.958 | Positive |
| We came before check-in time but amazing stuff of reception there were very so kind and helpful. | 0.531 | 0.469 | 0 | 0.958 | Positive |
| The room was comfy - bed was kinda hard, but still nice - has a great balcony, nice bathroom with hot water in the shower and good A/C. | 0.413 | 0.562 | 0.026 | 0.957 | Positive |
| … | … | … | … | … | … |
| Breakfast is weak for a five-star hotel - there are no salmon, sparkling wines, a bad choice of cheeses and cold meat snacks. | 0 | 0.644 | 0.356 | -0.859 | Negative |
| The negative point was the dirty swimmingpool, we just couldn't swim in it and were not the only ones not to give it a try, too bad... | 0 | 0.704 | 0.296 | -0.878 | Negative |
| Negative points are the very poor standard of the machines in the gym, and the poor maintained pool area. | 0 | 0.611 | 0.389 | -0.881 | Negative |
| I bought ear plugs to try but made absolutely no difference the noise was so bad. | 0 | 0.404 | 0.596 | -0.888 | Negative |
| WiFi wrong (last name wrong) hence found out due to wrong last name - no spg member record and hence no points | 0 | 0.554 | 0.446 | -0.914 | Negative |

Table 24. Part of the customer review-attribute performance matrix

| Hotel | Reviews | Attributes | | | | | | | | | Overall rating |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Food | Service | Room | Facility | Location | Cleanliness | value | Wi-Fi | Security | |
| S2-a | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 5 |
| S2-a | 2 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 5 |
| S2-a | 3 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 5 |
| S2-a | 4 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 4 |
| S2-a | 5 | 0 | 1 | 1 | 0 | 1 | 1 | 0 | 0 | 1 | 5 |
| … | … | … | … | … | … | … | … | … | … | … | … |
| S5-g | 26930 | 1 | 1 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 5 |
| S5-g | 26931 | -1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| S5-g | 26932 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 4 |
| S5-g | 26933 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 4 |
| S5-g | 26934 | 1 | 1 | 0 | 1 | -1 | 0 | 0 | 0 | 0 | 5 |

Table 25. Service attribute performance matrix

| Hotel | Attributes | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Food | Service | Room | Facility | Location | Cleanliness | value | Wi-Fi | Security |
| S2-a | 0.546 | 0.832 | 0.568 | 0.696 | 0.548 | 0.932 | 0.591 | 0.851 | 0.439 |
| S2-b | 0.595 | 0.779 | 0.304 | 0.660 | 0.598 | 0.947 | 0.667 | 0.733 | 0.800 |
| S2-c | 0.303 | 0.731 | 0.445 | 0.306 | 0.329 | 0.863 | 0.525 | 0.420 | 0.846 |
| S2-d | 0.806 | 0.797 | 0.488 | 0.600 | 0.381 | 0.933 | 0.333 | 0.286 | 0.647 |
| S2-e | 0.082 | 0.425 | 0.067 | 0.000 | 0.160 | 0.733 | 0.405 | -0.176 | 0.000 |
| S3-a | 0.209 | 0.606 | 0.343 | 0.524 | 0.333 | 0.748 | 0.662 | 0.579 | 0.120 |
| S3-b | 0.449 | 0.714 | 0.337 | 0.617 | 0.525 | 0.905 | 0.553 | 0.652 | 0.583 |
| S3-c | 0.602 | 0.692 | 0.235 | 0.617 | 0.648 | 0.937 | 0.588 | 0.545 | 0.613 |
| S3-d | 0.442 | 0.489 | 0.475 | 0.429 | 0.269 | 0.782 | 0.504 | 0.556 | 0.306 |
| S3-e | 0.661 | 0.825 | 0.729 | 0.077 | 0.388 | 0.948 | 0.800 | 0.440 | 0.895 |
| S3-f | 0.307 | 0.685 | 0.139 | 0.429 | 0.383 | 0.697 | 0.558 | 0.730 | 0.111 |
| S3-g | 0.687 | 0.800 | 0.659 | 0.784 | 0.394 | 0.925 | 0.641 | 1.000 | 0.667 |
| S4-a | 0.456 | 0.588 | 0.173 | 0.510 | 0.434 | 0.633 | 0.474 | 0.436 | 0.206 |
| S4-b | 0.592 | 0.723 | 0.418 | 0.645 | 0.505 | 0.865 | 0.537 | 0.611 | 0.375 |
| S4-c | 0.615 | 0.746 | 0.487 | 0.727 | 0.395 | 0.843 | 0.612 | 0.542 | 0.745 |
| S4-d | 0.587 | 0.731 | 0.679 | 0.674 | 0.442 | 0.897 | 0.500 | 0.324 | 0.353 |
| S4-e | 0.668 | 0.758 | 0.661 | 0.737 | 0.473 | 0.732 | 0.498 | 0.143 | 0.273 |
| S4-f | 0.699 | 0.771 | 0.776 | 0.771 | 0.523 | 0.922 | 0.691 | 0.405 | 0.556 |
| S4-g | 0.703 | 0.780 | 0.717 | 0.742 | 0.550 | 0.862 | 0.675 | 0.714 | 0.576 |
| S5-a | 0.661 | 0.750 | 0.665 | 0.688 | 0.553 | 0.804 | 0.559 | 0.605 | 0.653 |
| S5-b | 0.654 | 0.720 | 0.673 | 0.674 | 0.462 | 0.902 | 0.563 | 0.593 | 0.525 |
| S5-c | 0.602 | 0.696 | 0.658 | 0.618 | 0.492 | 0.750 | 0.607 | 0.459 | 0.529 |
| S5-d | 0.687 | 0.793 | 0.629 | 0.500 | 0.611 | 0.853 | 0.694 | 0.418 | 0.379 |
| S5-e | 0.589 | 0.629 | 0.556 | 0.599 | 0.283 | 0.782 | 0.615 | 0.127 | 0.439 |
| S5-f | 0.631 | 0.761 | 0.589 | 0.633 | 0.473 | 0.948 | 0.623 | 0.581 | 0.449 |
| S5-g | 0.746 | 0.769 | 0.632 | 0.724 | 0.531 | 0.864 | 0.493 | 0.513 | 0.538 |

Table 26. Service attribute importance matrix

| Hotel | Attributes | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Food | Service | Room | Facility | Location | Cleanliness | value | Wi-Fi | Security |
| S2-a | 0.309 | 0.634 | 0.383 | 0.000 | 0.234 | 0.618 | 0.274 | 1.000 | 0.571 |
| S2-b | 0.369 | 0.655 | 0.670 | 0.000 | 0.284 | 0.813 | 0.437 | 1.000 | 0.168 |
| S2-c | 0.301 | 0.483 | 0.608 | 0.171 | 0.000 | 0.531 | 0.192 | 0.726 | 1.000 |
| S2-d | 0.105 | 0.033 | 0.120 | 0.097 | 0.000 | 1.000 | 0.292 | 0.274 | 0.551 |
| S2-e | 0.000 | 0.788 | 0.567 | 0.805 | 0.020 | 1.000 | 0.792 | 0.591 | 0.021 |
| S3-a | 0.175 | 0.235 | 0.667 | 0.363 | 0.124 | 0.787 | 0.000 | 0.389 | 1.000 |
| S3-b | 0.267 | 0.846 | 0.516 | 0.560 | 0.490 | 1.000 | 0.525 | 0.000 | 0.445 |
| S3-c | 0.229 | 0.527 | 0.271 | 0.015 | 0.196 | 1.000 | 0.000 | 0.024 | 0.283 |
| S3-d | 0.288 | 0.486 | 0.635 | 0.000 | 0.040 | 1.000 | 0.380 | 0.479 | 0.483 |
| S3-e | 0.667 | 1.000 | 0.712 | 0.595 | 0.678 | 0.299 | 0.729 | 0.517 | 0.000 |
| S3-f | 0.074 | 0.540 | 1.000 | 0.000 | 0.349 | 0.752 | 0.808 | 0.495 | 0.533 |
| S3-g | 0.151 | 0.103 | 0.060 | 0.079 | 0.056 | 0.205 | 0.111 | 1.000 | 0.000 |
| S4-a | 0.237 | 0.869 | 0.488 | 0.285 | 0.201 | 1.000 | 0.251 | 0.936 | 0.000 |
| S4-b | 0.309 | 0.768 | 0.569 | 0.128 | 0.183 | 1.000 | 0.756 | 0.000 | 0.496 |
| S4-c | 0.258 | 0.435 | 0.158 | 0.322 | 0.000 | 1.000 | 0.008 | 0.488 | 0.453 |
| S4-d | 0.253 | 0.531 | 0.549 | 0.283 | 0.000 | 1.000 | 0.319 | 0.714 | 0.746 |
| S4-e | 0.239 | 0.759 | 0.540 | 0.456 | 0.000 | 1.000 | 0.430 | 0.887 | 0.590 |
| S4-f | 0.344 | 0.596 | 0.577 | 0.483 | 0.000 | 0.841 | 0.281 | 0.498 | 1.000 |
| S4-g | 0.306 | 0.379 | 0.373 | 0.401 | 0.220 | 1.000 | 0.288 | 0.000 | 0.318 |
| S5-a | 0.075 | 0.328 | 0.211 | 0.213 | 0.000 | 1.000 | 0.332 | 0.035 | 0.200 |
| S5-b | 0.227 | 0.587 | 0.502 | 0.150 | 0.000 | 1.000 | 0.199 | 0.425 | 0.338 |
| S5-c | 0.346 | 0.523 | 0.533 | 0.321 | 0.081 | 1.000 | 0.550 | 0.020 | 0.000 |
| S5-d | 0.154 | 0.436 | 0.785 | 0.313 | 0.000 | 1.000 | 0.286 | 0.531 | 0.566 |
| S5-e | 0.516 | 0.812 | 0.694 | 0.612 | 0.373 | 1.000 | 0.556 | 0.774 | 0.000 |
| S5-f | 0.268 | 1.000 | 0.405 | 0.584 | 0.138 | 0.000 | 0.773 | 0.341 | 0.324 |
| S5-g | 0.320 | 0.504 | 0.521 | 0.275 | 0.159 | 1.000 | 0.425 | 0.428 | 0.000 |
| Average | 0.261 | 0.571 | 0.504 | 0.289 | 0.147 | 0.840 | 0.384 | 0.483 | 0.388 |

### 5.4.4 Step 4: Identification of the competitors and best practices as benchmark

We conducted *k*-means clustering using Euclidean distances to identify competitors based on customer perception. Here, the number of clusters was set to three based on the result of the elbow method, as shown in Figure 20. The characteristics of three clusters were derived by averaging the attribute importance scores over the hotels of each cluster, as shown in Figure 21. First, as for cluster 1 with 11 hotels, *cleanliness* is identified as the only important service attribute. Second, regarding cluster 2 with 9 hotels, the most distinctive attribute compared to those of other clusters was *security*. Although the importance of *cleanliness* is high, it is almost the same level as the average of the 26 hotels considered. Finally, two attributes, *service* and *Wi-Fi*, are considered important for cluster 3 with 6 hotels. Moreover, it is noteworthy the results of cluster analysis based on the attribute importance from the perspective of customers differ from those of star-rating systems based on financial operation indicators. All clusters have hotels with different star-ratings evenly, which indicates that selecting benchmarks using star-rating systems only might lead to the omission of competitors that must be considered in the service benchmarking process.

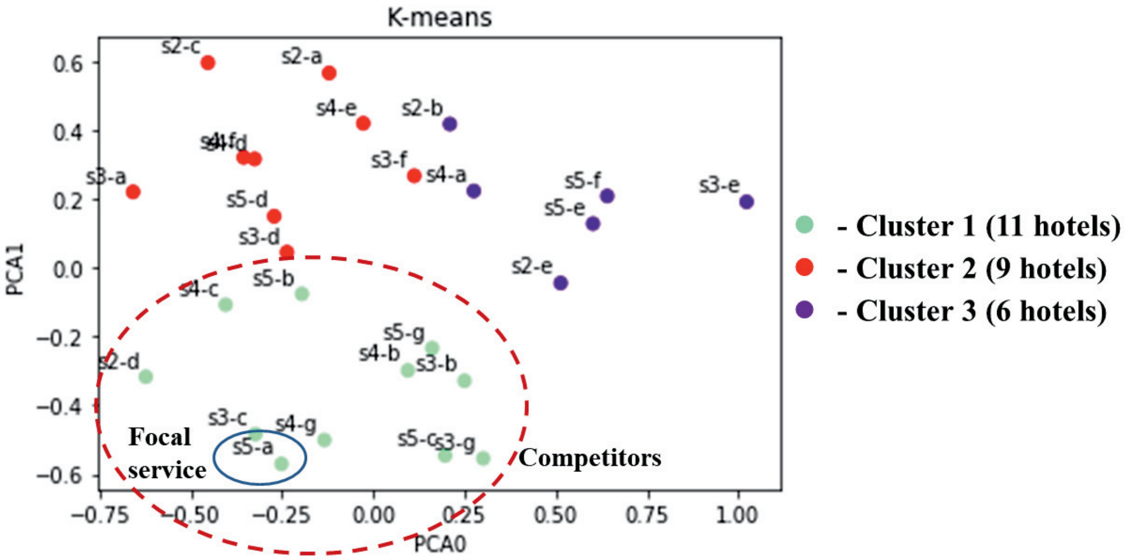



Figure 20. Identified groups for 26 hotels in Bangkok

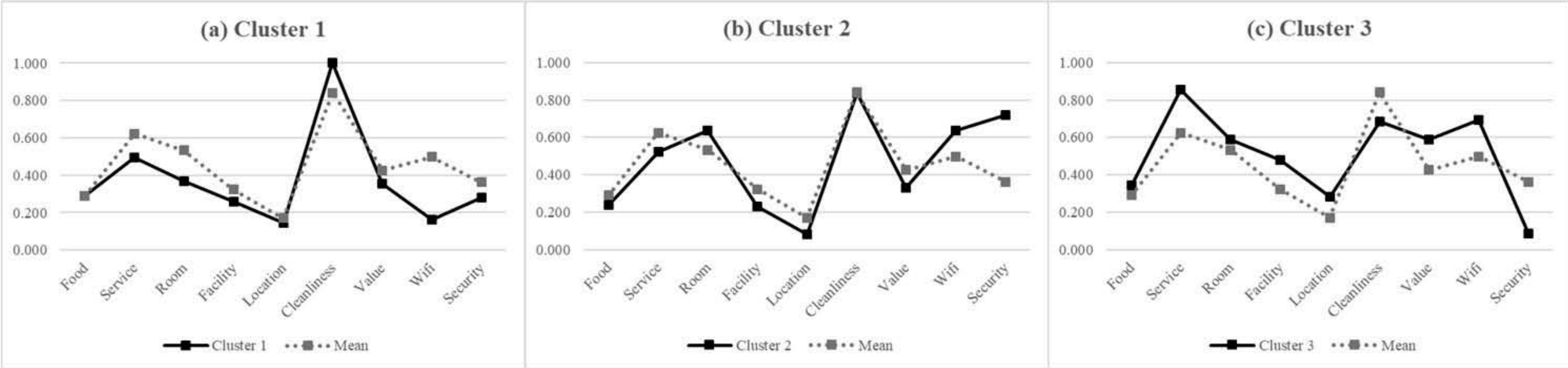



Figure 21. Average importance of attributes for three clusters: (a) cluster 1, (b) cluster 2, (c) cluster 3

To present specific use cases, we delineate the process of selecting benchmarks for cluster 1 with 11 hotels and the *Shangri-La hotel* (*S5-a*) as the focal service. TOPSIS was conducted to rank the hotels in the cluster based on service performance and attribute importance. First, the service attribute-performance matrix for the hotels in the cluster was constructed, as reported in Table 27(a). Second, the weighted normalized decision matrix was constructed, as represented in Table 27(b). Here, the average of attribute importance over the hotels in the cluster was employed as the weights, as shown in Table 27(c). Third, the best (worst) hypothetical hotel services were determined by combining the maximum (minimum) values of weighted normalized performance, as shown in Table 27(d). Fourth, $L^2$-distances between each hotel and the best (worst) hypothetical hotel service were calculated, as shown in Table 27(e). Fifth, the hotels are ranked according to their performance, as reported in Table 27(f). Finally, three hotels, *S3-g*, *S4-g*, and *S5-b*, were identified as benchmarks. In the case of *S3-g* that was ranked first in cluster 1, although this hotel is a three-star hotel, it was selected as the “Travelers’ choice” awarded by TripAdvisor only to the top 1% of the hotels in 2019, which supports our contention that the proposed approach is effective in measuring and ranking the performance of hotels.

Table 27. Matrices for the process of selecting benchmarks in cluster 1

(a) Service attribute-performance matrix for hotels in cluster 1

| Hotel | Attributes | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Food | Service | Room | Facility | Location | Cleanliness | value | Wi-Fi | Security |
| s2-d | 0.8065 | 0.7975 | 0.4878 | 0.6000 | 0.3814 | 0.9333 | 0.3333 | 0.2857 | 0.6471 |
| s3-b | 0.4490 | 0.7143 | 0.3371 | 0.6170 | 0.5253 | 0.9054 | 0.5529 | 0.6522 | 0.5833 |
| s3-c | 0.6015 | 0.6917 | 0.2353 | 0.6175 | 0.6481 | 0.9371 | 0.5882 | 0.5455 | 0.6129 |
| s3-g | 0.6869 | 0.8000 | 0.6591 | 0.7838 | 0.3937 | 0.9251 | 0.6410 | 1.0000 | 0.6667 |
| s4-b | 0.5920 | 0.7226 | 0.4182 | 0.6447 | 0.5045 | 0.8650 | 0.5367 | 0.6111 | 0.3750 |
| s4-c | 0.6148 | 0.7458 | 0.4866 | 0.7269 | 0.3953 | 0.8430 | 0.6115 | 0.5417 | 0.7455 |
| s4-g | 0.7034 | 0.7797 | 0.7173 | 0.7421 | 0.5500 | 0.8615 | 0.6747 | 0.7143 | 0.5758 |
| s5-a | 0.6613 | 0.7500 | 0.6655 | 0.6880 | 0.5527 | 0.8038 | 0.5587 | 0.6049 | 0.6526 |
| s5-b | 0.6540 | 0.7199 | 0.6734 | 0.6737 | 0.4620 | 0.9019 | 0.5625 | 0.5926 | 0.5250 |
| s5-c | 0.6019 | 0.6958 | 0.6584 | 0.6181 | 0.4924 | 0.7500 | 0.6067 | 0.4595 | 0.5294 |
| s5-g | 0.7457 | 0.7694 | 0.6321 | 0.7243 | 0.5307 | 0.8642 | 0.4925 | 0.5128 | 0.5385 |

(b) Weighted normalized decision matrix

| Hotel | Attributes | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Food | Service | Room | Facility | Location | Cleanliness | value | Wi-Fi | Security |
| S2-d | 0.0277 | 0.0466 | 0.0287 | 0.0196 | 0.0091 | 0.0943 | 0.0177 | 0.0108 | 0.0291 |
| S3-b | 0.0154 | 0.0417 | 0.0198 | 0.0201 | 0.0126 | 0.0915 | 0.0294 | 0.0247 | 0.0262 |
| S3-c | 0.0206 | 0.0404 | 0.0138 | 0.0201 | 0.0155 | 0.0947 | 0.0313 | 0.0206 | 0.0275 |
| S3-g | 0.0236 | 0.0467 | 0.0387 | 0.0256 | 0.0094 | 0.0934 | 0.0341 | 0.0378 | 0.0299 |
| S4-b | 0.0203 | 0.0422 | 0.0246 | 0.0210 | 0.0121 | 0.0874 | 0.0286 | 0.0231 | 0.0168 |
| S4-c | 0.0211 | 0.0435 | 0.0286 | 0.0237 | 0.0095 | 0.0851 | 0.0326 | 0.0205 | 0.0335 |
| S4-g | 0.0241 | 0.0455 | 0.0422 | 0.0242 | 0.0132 | 0.0870 | 0.0359 | 0.0270 | 0.0259 |
| S5-a | 0.0227 | 0.0438 | 0.0391 | 0.0224 | 0.0132 | 0.0812 | 0.0297 | 0.0229 | 0.0293 |
| S5-b | 0.0224 | 0.0420 | 0.0396 | 0.0220 | 0.0111 | 0.0911 | 0.0299 | 0.0224 | 0.0236 |
| S5-c | 0.0207 | 0.0406 | 0.0387 | 0.0201 | 0.0118 | 0.0758 | 0.0323 | 0.0174 | 0.0238 |
| S5-g | 0.0256 | 0.0449 | 0.0372 | 0.0236 | 0.0127 | 0.0873 | 0.0262 | 0.0194 | 0.0242 |

(c) Weight of attributes based on the importance of attributes

| | Attributes | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Food | Service | Room | Facility | Location | Cleanliness | value | Wi-Fi | Security |
| Importance | 0.074 | 0.144 | 0.110 | 0.073 | 0.040 | 0.293 | 0.100 | 0.077 | 0.088 |

(d) Best and worst hypothetical services

| | Attributes | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Food | Service | Room | Facility | Location | Cleanliness | value | Wi-Fi | Security |
| Best | 0.0277 | 0.0467 | 0.0422 | 0.0256 | 0.0155 | 0.0947 | 0.0359 | 0.0378 | 0.0335 |
| Worst | 0.0154 | 0.0404 | 0.0138 | 0.0196 | 0.0091 | 0.0758 | 0.0177 | 0.0108 | 0.0168 |

(e) Distance of each hotel from the best and worst hypothetical services

| Hotel ($c_i$) | S2-d | S3-b | S3-c | S3-g | S4-b | S4-c | S4-g | S5-a | S5-b | S5-c | S5-g |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $d_{c_i,b}$ | 0.0366 | 0.0315 | 0.0357 | 0.0091 | 0.0319 | 0.0261 | 0.0159 | 0.0227 | 0.0217 | 0.0320 | 0.0249 |
| $d_{c_i,w}$ | 0.0300 | 0.0267 | 0.0286 | 0.0474 | 0.0236 | 0.0309 | 0.0418 | 0.0347 | 0.0359 | 0.0309 | 0.0321 |

(f) Performance and ranking of hotels in cluster 1

| Hotel | S2-d | S3-b | S3-c | S3-g | S4-b | S4-c | S4-g | S5-a | S5-b | S5-c | S5-g |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Performance | 0.4508 | 0.4588 | 0.4452 | 0.8389 | 0.4255 | 0.5426 | 0.7237 | 0.6040 | 0.6229 | 0.4916 | 0.5635 |
| Ranking | 9 | 8 | 10 | 1 | 11 | 6 | 2 | 4 | 3 | 7 | 5 |

### 5.4.5 Step 5: Prioritization of service attributes and development of strategic action plan for service improvement

Two gaps between the focal service's performance and importance for attributes and between the focal service's performance and competing services' performance were first calculated as reported in Table 28. The IPCA plot was then developed based on these two gaps, as shown in Figure 22. The service attributes can be classified into one of four areas in the map: solid competitive advantage, head-to-head competition, urgent action, and potential competitive edge. Although the benchmarking guidelines may differ across the organizational context, the service attributes are generally managed in the direction shifting from urgent action and solid competitive advantage. First, two attributes, *cleanliness* and *value*, are located in the *urgent action* quadrant, indicating the performance of these attributes was lower than the importance of the attribute and lower than that of benchmarks. The quality of these attributes by examining the customer reviews for own service as well as benchmarks. Second, two attributes, *Wi-Fi* and *facility*, are positioned in the *head-to-head competition* quadrant, indicating that performances on the two attributes were higher than the importance of the attributes but lower than that of the benchmarks. *S5-a* should examine the quality improvement of these attributes could lead to organizational competitive advantages. If so, the directions of service improvement should be identified by examining the positive customer reviews of benchmarks. Third, *service* attribute is classified as *potential competitive edge*, representing that *S5-a* outperform benchmarks but present lower performance than the importance of attribute. The negative customer reviews for own service need to be closely investigated to identify the directions of service benchmarking. Finally, four attributes (i.e. *food, location, security, and room*) in the area of *solid competitive advantage* quadrant are considered satisfactory and need to be well-maintained.

Table 28. Result of IPCA for S5-a

| Attribute | Gap 1* | Gap 2** | [Priority] Quadrant |
|---|---|---|---|
| Food | 0.045 | -0.020 | [4] Solid competitive advantage |
| Service | 0.001 | -0.017 | [3] Potential competitive advantage |
| Room | 0.014 | -0.018 | [4] Solid competitive advantage |
| Facility | 0.017 | -0.045 | [2] Head-to-head competition |
| Location | 0.044 | 0.084 | [4] Solid competitive advantage |
| Cleanliness | -0.147 | -0.092 | [1] Urgent action |
| Value | -0.033 | -0.067 | [1] Urgent action |
| Wi-Fi | 0.045 | -0.164 | [2] Head-to-head competition |
| Security | 0.014 | 0.063 | [4] Solid competitive advantage |

* Gap 1 (The focal service's performance – the attribute's importance),
** Gap 2 (The focal service's performance – competing service's performance)

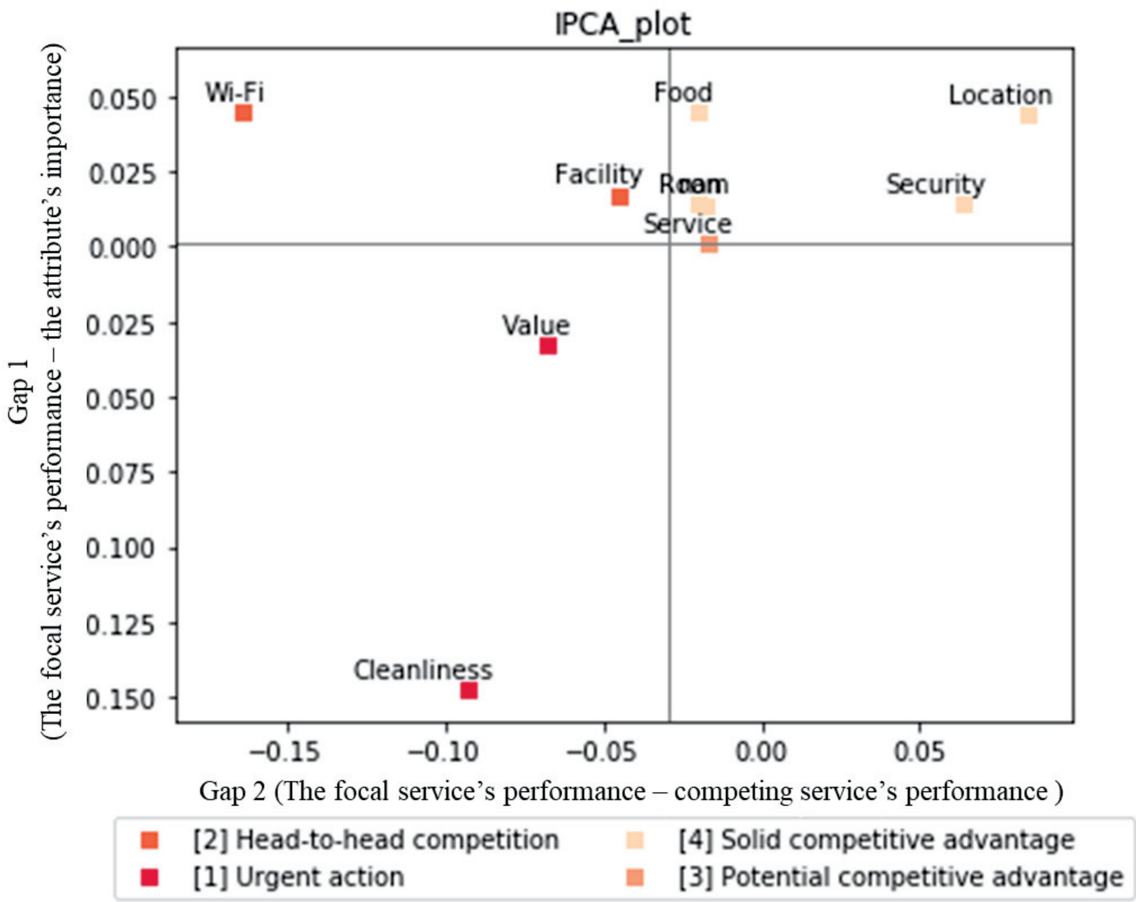



Figure 22. Importance-performance competitor analysis (IPCA) of S5-a

## 5.5 Summary and discussions

This study proposed a data-driven approach to customer-oriented service benchmarking using large amounts of customer review data as a source of the comprehensive VoC. The proposed approach includes (1) topic modeling to identify service attributes from customer reviews; (2) index and sentiment analysis to measure the importance of service attributes and the focal company's performance in the same attribute; (3) clustering and TOPSIS to select competitors and best practices as benchmarks from the perspective of customers; and finally (4) importance-performance competitor analysis to develop a strategic action plan. The case study of hotel services confirms that the proposed approach enables quick identification of key aspects of the focal company's and competitors' services and assessment of the performance of the focal company and benchmarks within acceptable limits of time and cost.

The contributions of this research are two-fold. From a theoretical perspective, this study contributes to service benchmarking research by extending previous financial and operational indicator-based approaches to customer-centric approaches. In particular, our approach identifies service attributes affecting the customer's perception of service quality, selects measure the focal service's and competitor's performance based on customer perception, and offers benchmarking guidelines based on the IPCA plot. It is expected that the systematic process and quantitative outcomes offered by the proposed approach provide a valuable complementary tool for customer-oriented service benchmarking toward continuous service improvement. Second, from a methodological perspective, this study integrates several methods such as topic modeling, sentiment analysis, clustering, and TOPSIS, enabling the quick analysis of the key aspects of the focal company's and competitors' services and support decision making with acceptable levels of time and cost using quantitative data and scientific methods. Moreover, the proposed approach systemizes the burden of experts' manual work when examining customer reviews. Although this study focused on service benchmarking, the proposed approach could be useful for various purposes such as developing promotion strategies.

With regard to validation of estimation of the importance and performance of each attribute, any approach that is developed should be carefully deployed in practice as there is no absolute confirmation regarding the validity and practicality of the proposed approach. In this respect, the performance of the proposed approach is strongly related to the ability to identify the sentiment of customer reviews for the relevant service attributes. For this reason, we examined the performance and reliability of the sentiment analysis using several quantitative metrics, as summarized in Table 29. Specifically, a total of 300 customer reviews were first randomly selected and classified into two categories (i.e. positive and negative reviews) by the authors and other experts. A confusion matrix was then constructed to compare the results of the sentiment analysis to those of experts, as shown in Table 29 (a). Finally, four quantitative indicators—which are accuracy, precision, recall, and $F_1$ score—were examined, as reported in Table 29 (b). Here, accuracy is defined as the percentage of correct classifications, as defined in Eq. (39). Precision indicates the number of correct results divided by the number of all returned results, as shown in Eq. (40). Recall measures the number of correct results divided by the number of results that should have been returned, as shown in Eq. (41). $F_1$ score represents the overall effectiveness of the sentiment analysis and is defined as the harmonic average of the precision and recall, as shown in Eq. (42). This score reaches its best value at 1 and worst at 0.

$$Accuracy = \frac{tp+tn}{tp+tn+fp+fn} \text{------------------------------------} \text{Eq. (39)}$$

$$Precision = \frac{tp}{tp+fp} \text{------------------------------------} \text{Eq. (40)}$$

$$Recall = \frac{tp}{tp+fn} \text{------------------------------------------} \text{Eq. (41)}$$

$$F_1 \ \ score = 2 \cdot \frac{precision \cdot recall}{precision+recall} \text{------------------------------------------} \text{Eq. (42)}$$

In the equations, $tp$, $tn$, $fp$, and $fn$ represent the number of positive reviews correctly classified, number of negative reviews correctly classified, number of negative reviews wrongly classified as positive, and number of positive reviews wrongly classified as negative. As can be seen from the Table 29 (b), the result of the performance evaluation indicates that the proposed approach provides effective and reliable performance in identifying the sentiment of customer reviews.

Table 29. Result of performance evaluation

(a) Confusion matrix

| | Predicted positive | Predicted negative |
|---|---|---|
| Actual positive | 195 | 32 |
| Actual negative | 8 | 65 |

(b) Summary of performance metrics

| Accuracy | Precision | Recall | $F_1$ score |
|---|---|---|---|
| 0.87 | 0.86 | 0.96 | 0.9 |

Despite its contributions, this study has certain limitations that should be explored in future research. First, the proposed approach still needs human intelligence and manual procedure particularly when defining service attributes and developing strategic action plans, although the approach can reduce experts' burden associated with service benchmarking and most parts of the approach can be automated. Second, although the case study verifies the feasibility and applicability of the proposed approach, further validation by domain experts is required to confirm the effectiveness of proposed approach such as the identification of competitors and measurement of the importance of each service attribute. Finally, this study considered a single case study, hotel services in Bangkok. Further testing on services across different industries is essential to confirm the external validity of the proposed approach. Nevertheless, the systematic processes and quantitative outcomes of the proposed approach offer a substantial contribution to both current research and future practice.

# 6 Conclusion

## 6.1 Summary and contributions

Continuous service improvement in the post-launch stages is considered one of the most crucial activities for companies to gain and maintain a competitive edge. The central tenets of the dissertation are three-fold: (1) customer reviews are a useful data source of the comprehensive VoC; (2) analysis of large-scale customer reviews can provide specific and practical guidelines on continuous service improvement, and finally (3) recent developments in machine learning and natural language processing techniques can contribute to service engineering research by offering information that cannot be produced easily by humans and can be transformed into service improvement strategies. Building on this notion, this dissertation develops models and methods for customer review analytics for continuous service improvement and applies the developed approaches to game and hotel service industries.

This dissertation deals with three main issues with critical problems for continuous service improvement in the post-launch stage. The issue is concerned with the dynamics of services given that different service improvement strategies should be offered along with different life cycle stages. We propose a stochastic service life cycle analysis to gauge where a service is in its life cycle and to give forecasts about its future prospects. The number of customer reviews is employed to customer attention-based service maturity and an HMM is used to estimate the probability of a service being at a certain stage of its life cycle. Based on this, three indicators are developed to represent the future prospects of a service's life cycle progression. The main advantages of the proposed approach lie in its ability to model different shapes of life cycles without any supplementary information and to examine a wide range of services at acceptable levels of time and cost. A case study of mobile game services in the Apple App Store is presented.

The second issue is associated with correcting existing defects for service improvement. We develop an integrated approach of sentiment analysis and SPC to monitoring customer complaints and to detecting potential service failures. The sentiment analysis method enables systematic identification of a customer satisfaction score from customer review data while statistical process control allows early detection of significant customer complaints and prevents service failures. The integration of two methods makes it possible to monitor customer complaints at acceptable levels of time and cost. A case study of a mobile game service is presented, offering a guideline for implementation and customization of the proposed approach.

The last issue is related to establish long-term strategies for service improvement. We present an approach to identifying competitors and best practices and developing guidelines on service benchmarking based on customer perception. The proposed approach includes (1) topic modeling to identify service attributes from customer reviews; (2) index and sentiment analysis to measure importance of service attributes and the focal company's performance in the same attribute; (3) clustering and the technique for order of preference by similarity to ideal solution (TOPSIS) to select competitors and best practices as benchmarks from the perspective of customers; and finally (4) importance-performance competitor analysis to develop a strategic action plan. The case study of hotel services confirms that the proposed approach is valuable as a complementary tool for customer-oriented service benchmarking.

The contributions of this dissertation are three-fold. Firstly, from a theoretical perspective, the dissertation contributes to service engineering research by developing quantitative models and methods for use in the post-launch stage. Of course, there have been similar attempts to apply machine learning and natural language processing techniques to customer review data in other disciplines such as computer science and applied statistics. However, prior studies view the customer reviews data as a data source of short texts and thus lack important details about value creation and implications in practice. Secondly, from a methodological perspective, previous approaches rely heavily on expert's opinions and/or customer surveys and thus become time-consuming and labor-intensive, as the number of services is increasing dramatically, customer needs shift rapidly, and service systems become large-scale and complex. The dissertation extends the previous expert- and customer survey-approaches to data-driven approaches, thereby enabling the quick analysis of the focal company's and competitors' services. Although this study has focused on service improvement in the post-launch stage, the proposed approaches could be useful for other purposes such as developing promotion and pricing strategies. Finally, from a practical standpoint, we automated the proposed approaches to allow non-specialists who are unfamiliar with customer review data and complex analytical models to benefit from the results of the analysis. It is expected that the proposed approaches and automated systems will assist the formation of post-launch service strategies and further serve as a starting point for developing more generic models.

## 6.2 Limitations and future research

Despite its contributions, this dissertation has certain limitations that should be complemented by future research. First, although customer reviews are considered as an ample and useful data source of VoC, this database faces reliability issues that stem from noisy data (e.g., fake reviews) and negative skewness of overall rating scores and sentiment of customers. Further algorithms should be incorporated into the analysis to detect noisy data and to correct the skewness problem. Second, other indices could be applied in the proposed approaches. In particular, as for service life cycle analysis, such indices as the sentiment scores (Song et al., 2016), the customer ratings (Zhang et al., 2010), and the depth/length (Mudambi and Schuff, 2010) of reviews could be employed to measure the customer attention-based service maturity. Third, many issues remain as to how to improve the performance of the proposed approach. As for sentiment analysis, although VADER was built on microblogs and found to perform well with short texts such as customer reviews, words may have different meanings across different service sectors. For this reason, the gold-standard list of lexical features of VADER should be modified and customized across analysis contexts. There are two ways of building sentiment lexicons: hand-craft elaboration (Taboada et al., 2011), and automatic construction on the basis of an external resource. Specifically, as for automatic construction, Turner et al. (2020) proposed the probabilistic approach, employing the overall rating scores as an external resource. The customer reviews with five or four rating were classified as positive reviews, and customer reviews with one or two rating were classified as negative reviews. First, the positive weight of a given word was measured by computing the quotient of its total number of occurrences across all customer reviews and the total number of words appearing in positive documents. Likewise, the negative weight of a word is the quotient of its total number of occurrences across all customer reviews and the total number of words appearing in negative customer reviews. Second, the probability of a word to be positive was computed by dividing its positive weight by the sum of its positive and negative weights. Similarly, the probability of a word to be negative is the result of its negative weight divided by the sum of its positive and negative weights. Finally, the sentiment score of a word is calculated as the difference between its probability of being positive and the probability of being negative. Regarding the SPC, as the number of attributes that require monitoring increases, the number of control charts increases accordingly. The univariate control chart can be extended to multivariate control charts enabling the assessment of multiple parameters together, although multivariate control charts have a disadvantage in terms of the interpretation of signals appearing on the control chart (Rogalewicz, 2012). Moreover, the type of charts should be modified according to analysis contexts as the procedure of establishing what really happened with the process can be very complicated. Fourth, although three case studies and qualitative validations verify the feasibility and applicability of the proposed approaches for customer-centric service improvement, there is limitation on the lack of quantitative validation. In future research, quantitative validation could be made by collaborating with service firms to utilize detailed and accurate information about services or by integrating other types of data such as implicit feedback data (e.g., click-through rate and browsing history). Fifth, our case study is limited to the mobile game app services and hotel services. The validity of developed approaches necessitates further testing work from a wider range of services. Finally, this dissertation focuses only on three issues for customer-centric continuous service improvement. Many research questions such as timing analysis of service updates, impact analysis of service failures, and identification of service recovery modes should be examined. Nevertheless, the systematic processes and quantitative outcomes of the proposed approaches offer a substantial contribution to both current research and future practice and service as a starting point for developing more general models.

# References

Agten, T. V. (2013). *Games: King of the mobile eco-system*. Distimo Publication.

Aguwa, C. C., Monplaisir, L., & Turgut, O. (2012). Voice of the customer: Customer satisfaction ratio based analysis. *Expert Systems with Applications*, *39*(11), 10112-10119.

Aitken, J., Childerhouse, P., & Towill, D. (2003). The impact of product life cycle on supply chain strategy. *International Journal of Production Economics*, *85*(2), 127–140.

Akehurst, G. (2009). User generated content: The use of blogs for tourism organisations and tourism consumers. *Service Business*, *3*(1), 51-61.

Alam, I., & Perry, C. (2002). A customer-oriented new service development process. *Journal of Services Marketing*, *16*(6), 515-534.

Albayrak, T. (2015). Importance Performance Competitor Analysis (IPCA): A study of hospitality companies. *International Journal of Hospitality Management*, *48*, 135-142.

Anderson, E. A., & Diaz, J. (1996). Using process control chart techniques to analyse crime rates in Houston, Texas. *Journal of the Operational research Society*, *47*(7), 871-881.

Ankeny, J. (2010). The app store that's never closed. *Harvard Business Review*, *38*(2), 22-27.

Atkinson, A. (1988). Answering the eternal question: what does the customer want. *Cornell Hotel and Restaurant Administration Quarterly*, *29*(2), 12-14.

Ballantyne, D. (1991). Coming to grips with service intangibles using quality management techniques.

Barras, R. (1986). Towards a theory of innovation in services. *Research Policy*, *15*(4), 161-173.

Bass, F. M. (1969). A new product growth model for consumer durables. *Management Science*, 50(12), 1825–1832.

Bauer, H. H., & Fischer, M. (2000). Product life cycle patterns for pharmaceuticals and their impact on R&D profitability of late mover products. *International Business Review*, *9*(6), 703-725.

Baum, L. E., Petrie, T., Soules, G., & Weiss, N. (1970). A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *The Annals of Mathematical Statistics*, *41*(1), 164–171.

Benton, W. C. (1991). Statistical process control and the Taguchi method: a comparative evaluation. *The international journal of production research*, *29*(9), 1761-1770.

Berry, L. L., & Parasuraman, A. (1992). Prescriptions for a service quality revolution in America. *Organizational Dynamics*, *20*(4), 5-15.

Bersimis, S., Psarakis, S., & Panaretos, J. (2007). Multivariate statistical process control charts: an overview. *Quality and Reliability engineering international*, *23*(5), 517-543.

Bi, J. W., Liu, Y., Fan, Z. P., & Zhang, J. (2019). Wisdom of crowds: Conducting importance-performance analysis (IPA) through online reviews. *Tourism Management*, *70*, 460-478.

Bickart, B., & Schindler, R. M. (2001). Internet forums as influential sources of consumer information. *Journal of Interactive Marketing*, *15*(3), 31–40.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, *3*(Jan), 993-1022.

Boe, D. T., Riley, W., & Parsons, H. (2009). Improving service delivery in a county health department WIC clinic: an application of statistical process control techniques. *American journal of public health*, *99*(9), 1619-1625.

Bojanic, D. C., & Drew Rosen, L. (1994). Measuring service quality in restaurants: an application of the SERVQUAL instrument. *Hospitality Research Journal*, *18*(1), 3-14.

Boshoff, C. (1997). An experimental study of service recovery options. *International Journal of service industry management*.

Bullinger, H. J., Fähnrich, K. P., & Meiren, T. (2003). Service engineering—Methodical development of new service products. *International Journal of Production Economics*, *85*(3), 275–287.

Burrel, Q. (2001). Stochastic modelling of the first–citation distribution. *Scientometrics*, *52*(1), 3–12.

Büschken, J., & Allenby, G. M. (2016). Sentence-based text analysis for customer reviews. *Marketing Science*, *35*(6), 953-975.

Buttle, F. (1996). SERVQUAL: review, critique, research agenda. *European Journal of marketing*, *30*(1), 8-32.

Camp, R. A., & Camp, R. A. (1989). *Entrepreneurs and politics in twentieth-century Mexico*. Oxford University Press on Demand.

Cantwell, R., Mirza, N., & Short, T. (1997). Continuous quality improvement efforts increase operating room efficiency. *Journal for Healthcare Quality*, *19*(6), 32-36.

Chan, F. T., Chan, H. K., Lau, H. C., & Ip, R. W. (2006). An AHP approach in benchmarking logistics performance of the postal industry. *Benchmarking: An International Journal*.

Chang, J., Gerrish, S., Wang, C., Boyd-Graber, J. L., & Blei, D. M. (2009). Reading tea leaves: How humans interpret topic models. In *Advances in neural information processing systems* (pp. 288-296).

Chen, K. S., Chang, T. C., Wang, K. J., & Huang, C. T. (2015). Developing control charts in monitoring service quality based on the number of customer complaints. *Total Quality Management & Business Excellence*, *26*(5-6), 675-689.

Chen, K. Y. (2014). Improving importance-performance analysis: The role of the zone of tolerance and competitor performance. The case of Taiwan's hot spring hotels. *Tourism Management*, *40*, 260-272.

Chen, S. H. (2016). Determining the service demands of an aging population by integrating QFD and FMEA method. *Quality & Quantity*, *50*(1), 283-298.

Chen, Y., Zhang, H., Liu, R., Ye, Z., & Lin, J. (2019). Experimental explorations on short text topic mining between LDA and NMF based Schemes. *Knowledge-Based Systems*, *163*, 1-13.

Cherdchamadol, P., & Sriboonjit, J. (2017). The factors influencing customer satisfaction with chain budget hotels in Bangkok by each traveler segment. *International Journal of Information, Business and Management*, *9*(1), 214.

Chou, C. C., Liu, L. J., Huang, S. F., Yih, J. M., & Han, T. C. (2011). An evaluation of airline service quality using the fuzzy weighted SERVQUAL method. *Applied Soft Computing*, *11*(2), 2117-2128.

Chu, R. K., & Choi, T. (2000). An importance-performance analysis of hotel selection factors in the Hong Kong hotel industry: a comparison of business and leisure travellers. *Tourism management*, *21*(4), 363-377.

Churchill Jr, G. A. (1979). A paradigm for developing better measures of marketing constructs. *Journal of marketing research*, *16*(1), 64-73.

Churchill, G. A. (1989). Stochastic models for heterogeneous DNA sequences. *Bulletin of Mathematical Biology*, *51*(1), 79–94.

Ciurumelea A, Schaufelbühl A, Panichella S, Gall HC (2017) Analyzing reviews and code of mobile apps for better release planning. In: Proceedings of the 24th International Conference on Software Analysis, Evolution and Reengineering (SANER), IEEE, pp 91–102

Clark, D. E., Cushing, B. M., & Bredenberg, C. E. (1998). Monitoring hospital trauma mortality using statistical process control methods. *Journal of the American College of Surgeons*, *186*(6), 630-635.

Comino S, Manenti FM, Mariuzzo F (2016) To upgrade or not to upgrade? the release of new versions to survive in the hypercompetitive app market. In: Proceedings of the International Workshop on App Market Analytics, pp 37–42

Cox, W. E. (1967). Product life cycles as marketing models. *Journal of Business*, 375–384.

Croux, C., & Dehon, C. (2010). Influence functions of the Spearman and Kendall correlation measures. *Statistical methods & applications*, *19*(4), 497-515.

Cunningham, M. T. (1969). The application of product life cycles to corporate strategy: Some research findings. *European Journal of Marketing*, *3*(1), 32–44.

Dagger, T. S., Sweeney, J. C., & Johnson, L. W. (2007). A hierarchical model of health service quality: scale development and investigation of an integrated model. *Journal of service research*, *10*(2), 123-142.

Danaher, P. J. (1997). Using conjoint analysis to determine the relative importance of service attributes measured in customer satisfaction surveys. *Journal of retailing*, *73*(2), 235-260.

Danaher, P. J., & Mattsson, J. (1994). Customer satisfaction during the service delivery process. *European journal of Marketing*.

Dave, K., Lawrence, S., & Pennock, D. M. (2003, May). Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. In *Proceedings of the 12th international conference on World Wide Web* (pp. 519-528).

Dean, J. (1950). *Pricing policies for new products*. Havard University Press.

Deng, W. J., Kuo, Y. F., & Chen, W. C. (2008). Revised importance–performance analysis: three-factor theory and benchmarking. *The Service Industries Journal*, *28*(1), 37-51.

DeSarbo, W. S., Huff, L., Rolandelli, M., & Choi, J. (1994). On the measurement of perceived service quality: a conjoint analysis approach. *Service quality: New directions in theory and practice*, 201-22.

Dickson, P. R. (2015). The adoption of customer service improvement practices. *Marketing Letters*, *26*(1), 1-15.

Duan W, Cao Q, Yu Y, Levy S (2013) Mining online user–generated content: using sentiment analysis technique to study hotel service quality. In: Proceedings of the 46th Hawaii International Conference on System Sciences, pp 3119–3128

Duckett, S., & Nijssen-Jordan, C. (2012). Using quality improvement methods at the system level to improve hospital emergency department treatment times. *Quality Management in Healthcare*, *21*(1), 29-33.

Eccles, G., & Durand, P. (1998). Complaining customers, service recovery and continuous improvement. *Managing Service Quality: An International Journal*, 8(1), 68–71

Edvardsson, B., Gustafsson, A., & Roos, I. (2005). Service portraits in service research: a critical review. *International journal of service industry management*, 16(1), 107-121

Ertuğrul, İ., & Karakaşoğlu, N. (2009). Performance evaluation of Turkish cement firms with fuzzy analytic hierarchy process and TOPSIS methods. *Expert Systems with Applications*, *36*(1), 702-715.

Fu, B., Lin, J., Li, L., Faloutsos, C., Hong, J., & Sadeh, N. (2013, August). Why people hate your app: Making sense of user feedback in a mobile app store. *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 1276-1284). ACM.

Fuglsang, L., Sundbo, J., & Sørensen, F. (2011). Dynamics of experience service innovation: innovation as a guided activity–results from a Danish survey. *The Service Industries Journal*, *31*(5), 661–677.

Gao, L., Porter, A. L., Wang, J., Fang, S., Zhang, X., Ma, T., Wang, W., & Huang, L. (2013). Technology life cycle analysis method based on patent documents. *Technological Forecasting and Social Change*, *80*(3), 398–407.

Gao, S., Tang, O., Wang, H., & Yin, P. (2018). Identifying competitors through comparative relation mining of online reviews in the restaurant industry. *International Journal of Hospitality Management*, *71*, 19-32.

Gardiner, S. C., & Mitra, A. (1994). Quality control procedures to determine staff allocation in a bank. *International Journal of Quality & Reliability Management*, 11(1):6–21

Garver, M. (2002). Using data mining for customer satisfaction research. *Marketing Research*, *14*(1), 8.

Gilbert RO (1987) Statistical methods for environmental pollution monitoring. John Wiley & Sons, New York

Gitto, S., & Mancuso, P. (2017). Improving airport services using sentiment analysis of the websites. *Tourism management perspectives*, *22*, 132-136.

Gonçalves, P., Benevenuto, F., & Cha, M. (2013). Panas-t: A pychometric scale for measuring sentiments on twitter. *arXiv preprint arXiv:1308.1857*. https://arxiv.org/abs/1308.1857 Accessed 17 July 2019

Grönroos C (1984) A service quality model and its marketing implications. *European journal of marketing*, 18(4), 36–44

Groves, R. M. (2006). Nonresponse rates and nonresponse bias in household surveys. *Public opinion quarterly*, *70*(5), 646-675.

Gu, Q., & Lago, P. (2007). A stakeholder–driven service life cycle model for SOA. *In 2nd International Workshop on Service oriented Software Engineering: in conjunction with the 6th ESEC/FSE Joint Meeting* (pp. 1–7). ACM.

Guo, Y., Barnes, S. J., & Jia, Q. (2017). Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation. *Tourism Management*, *59*, 467-483.

Hammer, C. (1981). Life cycle management. *Information & Management*, *4*(2), 71-80.

Hartigan, J. A., & Wong, M. A. (1979). Algorithm AS 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, *28*(1), 100-108.

Hartmann, J., Huppertz, J., Schamp, C., & Heitmann, M. (2019). Comparing automated text classification methods. *International Journal of Research in Marketing*, *36*(1), 20-38.

Haupt, R., Kloyer, M., & Lange, M. (2007). Patent indicators for the technology life cycle development. *Research Policy*, *36*(3), 387–398.

Hemmington, N., Kim, P. B., & Wang, C. (2018). Benchmarking hotel service quality using two-dimensional importance-performance benchmark vectors (IPBV). *Journal of Service Theory and Practice*.

Hensens, W., Struwing, M., & Dayan, O. (2010, October). Guest-review criteria on TripAdvisor compared to conventional hotel-rating systems to assess hotel quality. In *Passion for Hospitality Excellence: European Council on Hotel, Restaurant & Institutional Education 2010 Conference Proceedings* (pp. 1-12).

Hu M, Liu B (2004) Mining and summarizing customer reviews. In: Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, pp168–177

Hu, F., & Trivedi, R. H. (2020). Mapping hotel brand positioning and competitive landscapes by text-mining user-generated content. *International Journal of Hospitality Management*, *84*, 102317.

Hu, N., Pavlou, P. A., & Zhang, J. J. (2017). On Self-Selection Biases in Online Product Reviews. *MIS quarterly*, *41*(2), 449-471.

Hung, C. C., & Chen, L. H. (2009, March). A fuzzy TOPSIS decision making model with entropy weight under intuitionistic fuzzy environment. In *Proceedings of the international multiconference of engineers and computer scientists* (Vol. 1, pp. 13-16). IMECS Hong Kong.

Ibrahim, N. K., Hammed, H., Zaidan, A. A., Zaidan, B. B., Albahri, O. S., Alsalem, M. A., ... & Baqer, M. J. (2019). Multi-criteria evaluation and benchmarking for young learners' English language mobile applications in terms of LSRW skills. *IEEE Access*, *7*, 146620-146651.

Jain, S. K., & Gupta, G. (2004). Measuring service quality: SERVQUAL vs. SERVPERF scales. *Vikalpa*, *29*(2), 25-38.

James, T. L., Calderon, E. D. V., & Cook, D. F. (2017). Exploring patient perceptions of healthcare service quality through analysis of unstructured feedback. *Expert Systems with Applications*, *71*, 479-492.

Johns, N., Howcroft, B., & Drake, L. (1997). The use of data envelopment analysis to monitor hotel productivity. *Progress in tourism and hospitality research*, *3*(2), 119-127.

Johnston, R. (2005). Service operations management: from the roots up. *International Journal of Operations & Production Management*.

Kang, D., & Park, Y. (2014). Review–based measurement of customer satisfaction in mobile service: Sentiment analysis and VIKOR approach. *Expert Systems with Applications*, *41*(4), 1041–1050.

Keyt, J. C., Yavas, U., & Riecken, G. (1994). Importance-Performance Analysis. *International Journal of Retail & Distribution Management*.

Khalid, H., Shihab, E., Nagappan, M., & Hassan, A. E. (2014). What do mobile app users complain about?. *IEEE software*, *32*(3), 70-77.

Kim, J., & Lee, C. (2017). Stochastic service life cycle analysis using customer reviews. *The Service Industries Journal*, *37*(5-6), 296-316.

Kim, J., Lee, S., & Park, Y. (2013). User-centric service map for identifying new service opportunities from potential needs: A case of app store applications. *Creativity and Innovation Management*, *22*(3), 241-264.

Kim, Y., Kim, J., Kim, W., Im, J., Kim, T., Kang, S., & Kim, C. (2016). Predicting fluctuations in cryptocurrency transactions based on user comments and replies. *PloS one*, *11*(8).

Kingman-Brundage, J. (1989). The ABCs of service system blueprinting. *Designing a winning service strategy*, 30-43.

Koller, M., & Salzberger, T. (2009). Benchmarking in service marketing-a longitudinal analysis of the customer. *Benchmarking: An International Journal*, *16*(3), 401-414.

Kotler, P. (1965). Competitive strategies for new product marketing over the life cycle. *Management Science*, *12*(4), 104–119.

Kuusisto, J., Kuusisto, A., & Yli–Viitala, P. (2013). Service development tools in action. *The Service Industries Journal*, *33*(3–4), 352–365.

Lages, C. R., & Piercy, N. F. (2012). Key drivers of frontline employee generation of ideas for customer service improvement. *Journal of Service Research*, *15*(2), 215-230.

Lai, I. K. W., & Hitchcock, M. (2016). A comparison of service quality attributes for stand-alone and resort-based luxury hotels in Macau: 3-Dimensional importance-performance analysis. *Tourism Management*, *55*, 139-159.

Law, R., & Huang, T. (2006). How do travelers find their travel and hotel websites?. *Asia Pacific Journal of Tourism Research*, *11*(3), 239-246.

Lee, C., & Lee, H. (2015). Novelty–focussed document mapping to identify new service opportunities. *The Service Industries Journal*, *35*(6), 345–361.

Lee, C., Cho, Y., Seol, H., & Park, Y. (2012). A stochastic patent citation analysis approach to assessing future technological impacts. *Technological Forecasting and Social Change*, *79*(1), 16–29.

Lee, C., Lee, H., Seol, H., & Park, Y. (2012). Evaluation of new service concepts using rough set theory and group analytic hierarchy process. *Expert Systems with Applications*, *39*(3), 3404–3412.

Lee, C., Son, C., Yoon, B., & Park, Y. (2013). An instrument for discovering new mobile service opportunities. *International Journal of Mobile Communications*, *11*(4), 374–392.

Lee, C., Song, B., & Park, Y. (2009). Generation of new service concepts: A morphology analysis and genetic algorithm approach. *Expert Systems with Applications*, *36*(10), 12454–12460.

Lee, D., & Seung, H. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature*, *401*(6755), 788-791.

Lee, H., Lee, S., & Yoon, B. (2011). Technology clustering based on evolutionary patterns: The case of information and communications technologies. *Technological Forecasting and Social Change*, *78*(6), 953–967.

Lee, H., & Kim, C. (2014). Benchmarking of service quality with data envelopment analysis. *Expert Systems with Applications*, *41*(8), 3761-3768.

Lee, H., Kim, C., & Park, Y. (2010). Evaluation and management of new service concepts: An ANP–based portfolio approach. *Computers & Industrial Engineering*, *58*(4), 535–543.

Lee, S., Lee, H., & Yoon, B. (2012). Modeling and analyzing technology innovation in the energy sector: Patent–based HMM approach. *Computers & Industrial Engineering*, *63*(3), 564–577.

Lewis, B. R. (1989). Quality in the service sector: a review. *International Journal of Bank Marketing*.

Liao, H., Zeng, A., Xiao, R., Ren, Z. M., Chen, D. B., & Zhang, Y. C. (2014). Ranking reputation and quality in online rating systems. *PloS one*, *9*(5).

Liou, J. J., Tsai, C. Y., Lin, R. H., & Tzeng, G. H. (2011). A modified VIKOR multiple-criteria decision method for improving domestic airlines service quality. *Journal of Air Transport Management*, *17*(2), 57-61.

Luo, Y., & Tang, R. L. (2019). Understanding hidden dimensions in textual reviews on Airbnb: An application of modified latent aspect rating analysis (LARA). *International Journal of Hospitality Management*, *80*, 144-154.

Maalej W, Nabil H (2015) Bug report, feature request, or simply praise? on automatically classifying app reviews. In: Proceeding of 23rd International Requirements Engineering Conference (RE), pp 116–125

MacGregor, J. F., & Kourti, T. (1995). Statistical process control of multivariate processes. *Control Engineering Practice*, *3*(3), 403-414.

Maglio, P. P., & Spohrer, J. (2008). Fundamentals of service science. *Journal of the Academy of Marketing Science*, *36*(1), 18–20.

Mankad, S., Han, H. S., Goh, J., & Gavirneni, S. (2016). Understanding online hotel reviews through automated text analysis. *Service Science*, *8*(2), 124-138.

Mariani, M., & Baggio, R. (2020). The relevance of mixed methods for network analysis in tourism and hospitality research. *International Journal of Contemporary Hospitality Management*.

Martilla, J. A., & James, J. C. (1977). Importance-performance analysis. *Journal of marketing*, *41*(1), 77-79.

Matzler, K., Sauerwein, E., & Heischmidt, K. (2003). Importance-performance analysis revisited: the role of the factor structure of customer satisfaction. *The Service Industries Journal*, *23*(2), 112-129.

McIlroy, S., Ali, N., Khalid, H., & Hassan, A. E. (2016). Analyzing and automatically labelling the types of user issues that are raised in mobile app reviews. *Empirical Software Engineering*, *21*(3), 1067-1106.

Melanson, S. E., Goonan, E. M., Lobo, M. M., Baum, J. M., Paredes, J. D., Santos, K. S., ... & Tanasijevic, M. J. (2009). Applying Lean/Toyota production system principles to improve phlebotomy patient satisfaction and workflow. *American journal of clinical pathology*, *132*(6), 914-919.

Metaxas, I. N., Koulouriotis, D. E., & Spartalis, S. H. (2016). A multicriteria model on calculating the Sustainable Business Excellence Index of a firm with fuzzy AHP and TOPSIS. *Benchmarking: An International Journal, 23(6), 1522-1557*.

Michel, S. (2001). Analyzing service failures and recoveries: a process approach. *International journal of service industry management*. 12(1), 20–33

Miller, J. L., Craighead, C. W., & Karwan, K. R. (2000). Service recovery: a framework and empirical investigation. *Journal of operations Management*, *18*(4), 387-400.

Min, H., Min, H., & Chung, K. (2002). Dynamic benchmarking of hotel service quality. *Journal of*

*Services Marketing*, 16(4), 302-321

Min, H., Yun, J., & Geum, Y. (2018). Analyzing dynamic change in customer requirements: An approach using review-based Kano analysis. *Sustainability*, *10*(3), 746.

Mukherjee, A., Nath, P., & Pal, M. N. (2002). Performance benchmarking and strategic homogeneity of Indian banks. *International Journal of Bank Marketing*.

Mullen T, Collier N (2004) Sentiment analysis using support vector machines with diverse information sources. In: Proceedings of the 2004 conference on empirical methods in natural language processing, pp 412–418

Mulpuru, S., Evans, P., Sehgal, V., Ask, J. A., & Roberge, D. (2011). *Mobile commerce forecast: 2011 to 2016*. Forrester Research, Cambridge, MA.

National Statistical Office. (2013). The 2013 hotels and guest houses survey/ receipts of hotels and guest houses in 2013 by size of establishment and region. Retrieved May 15, 2016, from www.nso.go.th

Pan W, Xiang EW, Yang Q (2012) Transfer Learning in Collaborative Filtering with Uncertain Ratings. In: Proceeding of Twenty-Sixth AAAI Conference on Artificial Intelligence, pp 662–668

Parasuraman A, Zeithaml VA, Berry LL (1988) Servqual: A multiple–item scale for measuring consumer perceptions of service quality. J Retail 64(1):12–37

Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (1985). A conceptual model of service quality and its implications for future research. *Journal of marketing*, *49*(4), 41-50.

Park, H., Geum, Y., & Park, Y. (2015). A dual quality function deployment approach for benchmarking service quality. *Total Quality Management & Business Excellence*, *26*(5-6), 569-582.

Park, K., Kim, J., Park, J., Cha, M., Nam, J., Yoon, S., & Rhim, E. (2015, October). Mining the minds of customers from online chat logs. *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management* (pp. 1879-1882). ACM.

Popescu, A. M., & Etzioni, O. (2007). Extracting product features and opinions from reviews. In *Natural language processing and text mining* (pp. 9-28). Springer, London.

Potts, G. W. (1988). Exploit your product's service life cycle. *Harvard Business Review*, *66*(5), 32–36.

Pyon, C. U., Woo, J. Y., & Park, S. C. (2011). Service improvement by business process management using customer complaints in financial service industry. *Expert Systems with Applications*, *38*(4), 3267-3279.

Quan, X., Kit, C., Ge, Y., & Pan, S. J. (2015, June). Short and sparse text topic modeling via self-aggregation. In *Twenty-Fourth International Joint Conference on Artificial Intelligence*.

Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2), 257–285.

Rasouli, O., & Zarei, M. H. (2016). Monitoring and reducing patient dissatisfaction: a case study of an

Iranian public hospital. *Total Quality Management & Business Excellence*, *27*(5-6), 531-559.

Ribeiro, F. N., Araújo, M., Gonçalves, P., Gonçalves, M. A., & Benevenuto, F. (2016). Sentibench-a benchmark comparison of state-of-the-practice sentiment analysis methods. *EPJ Data Science*, *5*(1), 1-29.

Rink, D. R., & Swan, J. E. (1979). Product life cycle research: A literature review. *Journal of Business Research*, *7*(3), 219–242.

Rogalewicz, M. (2012). Some notes on multivariate statistical process control. *Management and Production Engineering Review*, *3*, 80-86.

Roos, I. (1999). Switching processes in customer relationships. *Journal of Service Research*, *2*(1), 68-85.

Rust, R. T., & Williams, D. C. (1994). How length of patronage affects the impact of customer satisfaction on repurchase intention. *Journal of Consumer Satisfaction, Dissatisfaction, and Complaining Behavior*, *7*, 107-113.

Rust, R. T., Inman, J. J., Jia, J., & Zahorik, A. (1999). What you don't know about customer-perceived quality: The role of customer expectation distributions. *Marketing Science*, *18*(1), 77-92.

Scheuing, E. E. (1969). The product life cycle as an aid in strategy decisions. *Management International Review*, 9(4/5), 111–124.

Schweitzer, E., & Aurich, J. C. (2010). Continuous improvement of industrial product-service systems. *CIRP Journal of Manufacturing Science and Technology*, *3*(2), 158-164.

Shamma, H., & Hassan, S. (2013). Customer-driven benchmarking. *Benchmarking: An International Journal*.

Son, C., Lee, C., Yoon, B., & Park, Y. (2015). GTM–based service map to identify new service opportunities. *International Journal of Mobile Communications*, *13*(2), 113–135.

Song, B., Lee, C., & Park, Y. (2013). Assessing the risks of service failures based on ripple effects: A Bayesian network approach. *International Journal of Production Economics*, *141*(2), 493–504.

Song, B., Lee, C., Yoon, B., & Park, Y. (2016). Diagnosing service quality using customer reviews: an index approach based on sentiment and gap analyses. *Service Business*, *10*(4), 775-798.

Stigler, J., Ziegler, F., Gieseke, A., Gebhardt, J. C. M., & Rief, M. (2011). The complex folding network of single calmodulin molecules. *Science*, *334*(6055), 512–516.

Storey, C., & Easingwood, C. J. (1999). Types of new product performance: Evidence from the consumer financial services sector. *Journal of Business Research*, *46*(2), 193–203.

Suchat Sritama. (2018) "Why are so many hotels for sale?". Bangkok Post.

Sundbo, J. (1997). Management of innovation in services. *Service Industries Journal*, *17*(3), 432–455.

Taboada, M., Brooke, J., Tofiloski, M., Voll, K., & Stede, M. (2011). Lexicon-based methods for sentiment analysis. *Computational linguistics*, *37*(2), 267-307.

Taecharungroj, V., & Mathayomchan, B. (2019). Analysing TripAdvisor reviews of tourist attractions in Phuket, Thailand. *Tourism Management*, *75*, 550-568.

Taha, H. A. (2011). *Operations research: An introduction*. (9th ed.). Pearson Education.

Tai, A. H., Ching, W. K., & Chan, L. Y. (2009). Detection of machine failure: Hidden Markov Model approach. *Computers & Industrial Engineering*, *57*(2), 608–619.

Talty, Alexandra. (2019) "Bangkok Is The Most Visited City In The World...Again". Forbes. Retrieved 20 May 2020.

Taplin, R. H. (2012). Competitive importance-performance analysis of an Australian wildlife park. *Tourism Management*, *33*(1), 29-37.

Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of language and social psychology*, *29*(1), 24-54.

Torres, E. N., Adler, H., & Behnke, C. (2014). Stars, diamonds, and other shiny things: The use of expert and consumer feedback in the hotel industry. *Journal of Hospitality and Tourism Management*, *21*, 34-43.

Tsang, N., & Qu, H. (2000). Service quality in China's hotel industry: a perspective from tourists and hotel managers. *International journal of contemporary hospitality management*, 12(5), 316-326.

Tsarev, D., Petrovskiy, M., & Mashechkin, I. (2011, December). Using NMF-based text summarization to improve supervised and unsupervised classification. In *2011 11th International Conference on Hybrid Intelligent Systems (HIS)* (pp. 185-189). IEEE.

Turner, Z., Labille, K., & Gauch, S. (2020). Lexicon-Based Sentiment Analysis for Stock Movement Prediction. *International Journal of Mechanical and Industrial Engineering*, *14*(5), 185-191.

Utley JS, Gaylord MJ (2009) Utley, J. S., & May, J. G. (2009). Monitoring service quality with residuals control charts. *Managing Service Quality: An International Journal*, 19(2):162–178

Van Den Ende, J. (2003). Modes of governance of new service development for mobile networks: A life cycle perspective. *Research Policy*, *32*(8), 1501–1518.

Van Ryzin, G. G., & Immerwahr, S. (2007). Importance-performance analysis of citizen satisfaction surveys. *Public Administration*, *85*(1), 215-226.

Verhoef, P. C. (2003). Understanding the effect of customer relationship management efforts on customer retention and customer share development. *Journal of Marketing*, *67*(4), 30–45.

Vinodhini, G., & Chandrasekaran, R. M. (2016). A comparative performance evaluation of neural network based approach for sentiment classification of online reviews. *Journal of King Saud University-Computer and Information Sciences*, *28*(1), 2-12.

Viterbi, A. J. (1967). Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Transactions on Information Theory*, *13*(2), 260–269.

Wemmerlöv, U. (1990). A taxonomy for service processes and its implications for system design. *International Journal of Service Industry Management*, *1*(3), 20-40.

Witell, L., Kristensson, P., Gustafsson, A., & Löfgren, M. (2011). Idea generation: Customer co–creation versus traditional market research techniques. *Journal of Service Management*, *22*(2), 140–159.

Wong, W., & Stamp, M. (2006). Hunting for metamorphic engines. *Journal in Computer Virology*, *2*(3), 211–229.

Wood, M. (1994). Statistical methods for monitoring service processes. *International Journal of Service Industry Management*, 5(4):53–68

Xia, H., Vu, H. Q., Lan, Q., Law, R., & Li, G. (2019). Identifying hotel competitiveness based on hotel feature ratings. *Journal of Hospitality Marketing & Management*, *28*(1), 81-100.

Xiang, Z., Schwartz, Z., Gerdes Jr, J. H., & Uysal, M. (2015). What can big data and text analytics tell us about hotel guest experience and satisfaction?. *International Journal of Hospitality Management*, *44*, 120-130.

Xiong, S., Wang, K., Ji, D., & Wang, B. (2018). A short text sentiment-topic model for product reviews. *Neurocomputing*, *297*, 94-102.

Xu, X., & Li, Y. (2016). The antecedents of customer satisfaction and dissatisfaction toward various types of hotels: A text mining approach. *International journal of hospitality management*, *55*, 57-69.

Yang, H. H., & Chen, K. S. (2000). A performance index approach to managing service quality. *Managing Service Quality: An International Journal*.

Yang, S. F., Cheng, T. C., Hung, Y. C., & W. Cheng, S. (2012). A new chart for monitoring service process mean. *Quality and Reliability Engineering International*, *28*(4), 377-386.

Yang, Z., & Fang, X. (2004). Online service quality dimensions and their relationships with satisfaction: A content analysis of customer reviews of securities brokerage services. *International Journal of Service Industry Management*, *15*(3), 302–326.

Ziegler, C. N., Skubacz, M., & Viermetz, M. (2008, December). Mining and exploring unstructured customer feedback data using language models and treemap visualizations. *Proceeding of the IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology* (Vol. 1, pp. 932-937). IEEE.

# Acknowledgments

반연리 산 속에 의연하게 자리 잡은 유니스트에 들어온 지 10년이 되었습니다. 이곳에서 저는 학문을 익히고, 교수님과 선·후배와의 만남을 통해 삶을 살아가는 방식도 배웠습니다. 그동안 유니스트는 저의 전부였습니다. 많은 분들의 응원과 도움 덕분에 박사학위 논문을 완성하였습니다. 논문을 마무리 하면서 저의 작은 연구 실적보다는 그동안 많은 도움을 주신 분들에 대한 고마움이 더 크고 깊게 느껴졌습니다. 이 글은 그분들에게 드리는 저의 감사인사입니다.

우선, 존경하는 임치현 교수님께 깊은 감사의 말씀을 올립니다. 부족함이 많은 박사 초년생인 저를 연구실에 받아주시고 지금까지 지도해주셨습니다. 교수님께서는 학문을 하는 연구자의 자세에서부터 제 학위논문이 완성되는 순간까지 한결같은 애정과 관심으로 가르침을 주셨습니다. 그리고 제 학위논문의 발전을 위해 아낌없는 조언을 해주시고 심사해 주신 우한균, 금영정, 김성일, 임성훈 교수님께도 진심으로 감사의 말씀을 드립니다. 또한 학부에 입학해 박사과정을 끝마치기까지 지켜보아 주시고 가르쳐 주셨던 유니스트 경영학부, 경영공학과 교수님들에게도 감사의 인사를 드립니다.

대학원 생활에서 가장 많은 시간을 함께했던 연구실 식구들에게도 감사의 인사를 전합니다. 학문적인 조언 뿐만 아니라 제 진로와 관련된 조언도 아낌없이 주셨던 김기훈, 정준각 박사님 정말 감사합니다. 선배로서 많은 도움을 주지 못해 미안하고 아쉬운 창헌, 기혁, 종경, 동기, 호진, 수혁, 현우, 예람에게도 고마움을 전합니다. 서로의 연구 내용을 나누고 농담도 주고받으며 울고 웃었던 그 시간들을 잊지 못할 것 같습니다. 대전에서 처음 만나 울산 생활을 함께한 현희와 정민이, 같이 사는 동안 많은 추억을 함께한 학부 룸메이트 민주와 대학원 룸메이트 현지언니, 저에게 소중한 존재인 연란언니와 민주, 언제 만나도 대학교 신입생 때로 돌아간 것 같은 OT 20조 친구들, 정말 고맙습니다.

세상 누구보다도 저를 사랑해 주시는 부모님, 제가 어떤 선택을 해도 항상 믿어주시고 응원해주셔서 정말 감사합니다. 나이 차이가 많이 난다는 이유로 엄마처럼 저를 보살펴 준 우리 가람이 언니, 정말 사랑합니다. 부족한 며느리를 항상 사랑으로 감싸주시는 시아버님, 시어머님에게도 감사의 인사를 드립니다. 늘 저에게 친절하고 상냥하게 대해 주시는 도련님과 동서, 모든 게 다 예쁘고 귀여운 조카 서윤이까지 가족 모두에게 사랑을 전합니다.

마지막으로, 저를 항상 행복하게 해주는 남편 이창용 박사에게 큰 사랑과 감사의 말을 전합니다. 남편의 한결 같은 사랑과 지지 덕분에 박사학위를 무사히 받을 수 있었습니다. 앞으로 남은 시간도 지금처럼 즐겁고 행복하게, 서로를 따뜻하게 바라보며 살고 싶습니다.

고마운 분들에게 보답하는 길은 미력이나마 제가 공부한 학문을 더 갈고 닦아 사회에 보탬을 주는 것이라 생각합니다. 여기에서 멈추지 않고 나날이 성장하는 사람이 되겠습니다. 감사합니다.

2020년 6월, 김주람 올림

# Curriculum Vitae

# Juram Kim

Department of Industrial Engineering
UNIST (Ulsan National Institute of Science and Technology)
50, UNIST-gil, Ulsan 44919, Republic of Korea
juram92@unist.ac.kr

## EDUCATION

Mar. 2015 ~ Present
**Ulsan National Institute of Science and Technology (UNIST)** Ulsan, Korea
Department of Management Engineering
*Integrated M. S. and Ph. D. Course*
*Advisor: Prof. Chiehyeon Lim*

Mar. 2011 ~ Feb. 2015
**Ulsan National Institute of Science and Technology (UNIST)** Ulsan, Korea
School of Business Administration
*B.S. in Business Administration*
(Honors: *Magna Cum Laude & Early Graduation of Excellence*)

## RESEARCH INTEREST

- Systematic business / service / technology intelligence analysis
- Applied data mining and machine learning techniques
- Data-driven service improvement

## PUBLICATIONS

1. **Kim, J**., & Lim, C., Service benchmarking based on customer perception, In Preparation.
2. Kim, S., **Kim, J.**, Cho, H., & Lim, C., Prediction for port logistics planning and operations through machine learning: A study on the container terminal traffic prediction with a Deep Forest model, In Preparation.
3. **Kim, J**., & Lim, C., Customer complaints monitoring with data analytics: An integrated method of sentiment and statistical process control analyses, *Industrial Management & Data Systems*, Under Review.
4. **Kim, J**., Kim, S., & Lee, C., Anticipating technological convergence: Link prediction using Wikipedia hyperlinks, *Technovation*, 79, 25-34, (2019).
5. **Kim, J.**, & Lee, C., Stochastic service life cycle analysis using customer reviews, *The Service Industries Journal,* 37(5-6), 296-316, (2017).

6. Lee, C., **Kim, J.**, Noh, M., Woo, H. G., & Gang, K., Patterns of technology life cycles: Stochastic analysis based on patent citations, *Technology Analysis & Strategic Management*, 29(1), 53-67, (2017).
7. Lee, C., **Kim, J.,** & Lee, S., Towards robust technology road mapping: How to diagnose the vulnerability of organisational plans, *Technological forecasting and social change*, 111, 164-175, (2016).
8. Lee, C., **Kim, J.**, Kwon, O., & Woo, H. G., Stochastic technology life cycle analysis using multiple patent indicators, *Technological Forecasting and Social Change*, 106, 53-64, (2016).

## R&D PROJECTS

1. **Development of a data-driven decision support system for customer-oriented service operations and improvement, *National research foundation of Korea*,** September 2019 – August 2020
**(Principal investigator)**
2. Development of outpatient waiting time prediction model and improvement of outpatient service processes, *Pusan National University Hospital*, September 2019 – July 2020
3. Development of a terminal congestion prediction service model to improve the efficiency of port operation, *Ulsan Port Authority*, January 2019 – December 2019
4. Development of data analytics methods to identify the sources of odor, *Ulsan Industry-University Convergence Campus*, October 2017 – March 2018
5. Benchmarking successful cases of world-leading systems to develop future directions of information analysis research, *Korea Institute of Science and Technology Information (KISTI)*, September 2014 – November 2014

## INTELLECTUAL PROPERTY RIGHTS

1. Lee, C, **Kim, J**, Kim, S. Method for prediction of converging technologies using link analysis, *Korea patent office*, 1020322580000 (2019.10.08)
2. Kwon, O., Lee, C., **Kim, J**., Noh, K. Apparatus for technology life analysis using multiple patent indicators, *Korea patent office*, 1016291780000 (2016.06.03)