Text-based Causality Modeling with Emotional Information Embedded in Hierarchic Topic Structure

Takurou Ogawa
College of Sustainable System Sciences,
School of Knowledge and Information Systems
Osaka Prefecture University
Japan
e-mail: saa01052@edu.osakafu-u.ac.jp

Ryosuke Saga
Department of Sustainable System Sciences
Graduate School of Humanities and Sustainable Systems
Osaka Prefecture University
Japan
e-mail: saga@cs.osakafu-u.ac.jp

***Abstract*—Service evaluation depends on various factors, such as assurance, responsiveness, and tangibles. Given that emotional satisfaction affects service satisfaction, analyzing both the evaluation and emotions is important in improving service. Previous studies have identified the evaluation factor and determined the degree of influence on the resulting evaluation. However, there is little effective analysis that reflects the influence of such a factor on emotion. In this study, we use hierarchal Latent Dirichlet Allocation and structural equation modeling (SEM) to express the causality relationships of service evaluation visually and quantitatively. Emotion obtained quantitatively by using sentiment analysis is newly applied to SEM to obtain knowledge reflecting the influence of emotion. As a result of the experiment, we can identify the causality of service and determine the influence of the evaluation factor and emotion quantitatively.**

Keywords-sentiment analysis; service analysis; SEM; hLDA; causal analysis

I. INTRODUCTION

In recent years, the service industry has grown rapidly such that in developed countries, there are so many markets that account for 60% to 70% of a country's gross domestic product (GDP). In the United States where GDP is the highest, the service industry's GDP is $ 15.52 trillion, accounting for 80% of the total GDP [1]. In addition, with the spread of smartphones, apps for various services (e.g., Twitter, navigation), the introduction of recommended hotels, and the rise of electronic services (e.g., Internet shopping) are rapidly increasing. With this background, the importance of services has grown in recent years. Service improvement is important as services are produced and consumed at the same time compared with products that are released and finished. Thus, analyzing the evaluation of the service in order to improve such service is important.

Service evaluation depends on various factors, such as assurance, responsiveness, and tangibles. For example, SERVQUAL evaluates the quality of service [2] with five-dimensional indicators, and Airport Service Quality (ASQ) [3] defines airport evaluation factors. As there are many factors in the evaluation of services, it is necessary to find out the evaluation factors to analyze the evaluation.

Generally, analyzing services is difficult because these have special features like Intangible, Heterogeneous, Inseparable, and Perishable (IHIP). However, there are several clues to analyze the services from the data (e.g., questionnaire). Especially, user review is useful because the review describes user experience of and perceived from the services. Therefore, it is possible to analyze the quality of service and the evaluation of service. Meanwhile, emotional satisfaction is also regarded as an important and attractive factor in service satisfaction. That is, customers experience different positive and negative emotions related to service, and these emotions influence service satisfaction [4]. Of course, these factors influence service evaluation and the emotions related to the service are implied in the user review; however, there is no study to identify and analyze evaluation factors together with emotional information.

This paper describes the method by which to perform causality analysis from text data, such as user review. In order to treat causal analysis, we use the topic-based approaches by applying a topic model to the review. In addition, the emotions for evaluation factors in the text are quantitatively determined using sentiment analysis technology. By applying topic and emotion information to structural equation modeling (SEM), we analyze the influence of each factor quantitatively.

The first contribution of this paper is that it obtains the knowledge reflecting emotional information from the user review by using sentiment analysis. Second, it understands the influence on the emotion of the evaluation factor based on the idea that emotions are essential for service evaluation factor analysis. By using SEM with path diagram, we can also analyze and understand the causality relationships among topics and their emotions associated with topics that are visually and quantitatively express.

Section 2 refers to the existing related research, section 3 explains the core technology of the analysis process, and section 4 describes analysis experiments using actual data.

II. LITERATURE REVIEW

In related research, SERVQUAL [2] measures the quality of service by measuring the gap between advance expectation and subsequent experience using five indicators prepared in advance. SURVPERF [5] measures the quality of service based on the subsequent experience alone. Related researches include a study that further increased the dimension from these five dimensions [6] and another that changed the dimension to measure the quality of electronic service [7].

Copyright (c) IARIA, 2019.

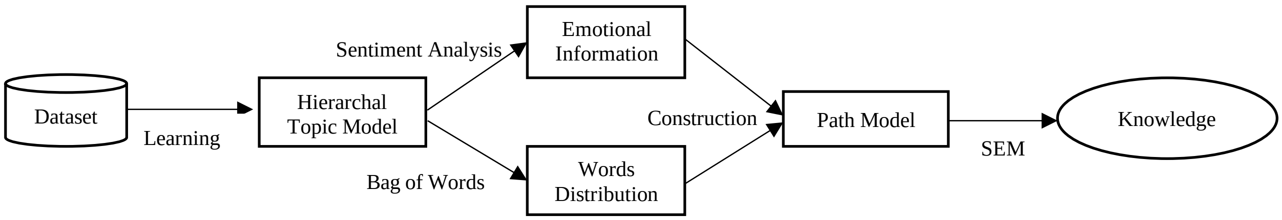

Figure 1. Analysis process

Asesthere are many evaluation indicators, it is difficult to measure all services by one standard because there are many types of services and their characteristics largely differ.

Meanwhile, related works on SEM include a study that has found relationships between customer loyalty and service quality [8] and another that has proposed a model to infer the purchase factor of the game by combining hierarchal Latent Dirichlet Allocation (hLDA) and SEM [9]. Another study made improvements to the SERVQUAL index and analyzed it with SEM [6]. These works, however, do not consider the emotion of the text.

Meanwhile, emotional satisfaction is largely believed to affect service satisfaction [4]. In relation to this, sentimental analysis is useful in comprehending and handling the emotional information. A study utilizes sentiment analysis and Latent Dirichlet Allocation (LDA) to evaluate the quality of airport services [10], while another determines the user's evaluation for each attribute by combining Airport Council International (ACI)-defined airport service quality attributes and sentiment analysis [11]. In these studies, emotion is considered one of the important factors in sales of services; thus it is essential to consider emotion. However, no study has proposed structural equation modeling that considers the emotion contained in text.

Therefore, the current paper proposes the model for SEM with emotion information. By using this model, we can acquire knowledge including emotion information visually.

III. Methodology

In this paper, the analysis is performed according to the process of Figure 1. First, topics are extracted by learning a topic model. Next, we find the emotion and topic distribution for that topic. Finally, a model is constructed based on these data and this is then analyzed by SEM so that can gain knowledge.

A. Topic Model

The topic model is a technology that tries to clarify the structure of a document group by inferring words contained in the topic based on the premise that the document group has a specific topic. In a topic model, a document is a collection of words probabilistically generated by the topic to which it belongs.

Topic models include different methods, such as latent semantic analysis (LSA) [12], LDA [13] and hLDA [14]. The LDA assumes a multi-topic model in which the document is based on mixed topics. LDA has a 1:*n* relationship between documents and topics, not 1:1 like LSA. LDA is considered to be a more natural model in documents, such as review texts that are written in one document about various aspects [13].

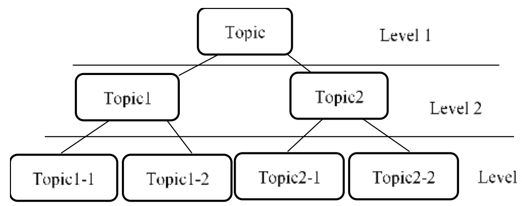

Figure 2. Hierarchal structure of topics

HLDA is an extended method of LDA and is a hierarchal model as shown in Figure 2. It has the property of automatically constructing relationships among hierarchical topics. As a learning result, a hierarchical model constructed hierarchically and a keyword group constituting each topic are generated together with their generation probabilities. The specific content of the topic can be inferred from the keyword groups of a topic. In this study, hLDA is used because it is a natural document model and the relationships between topics are defined automatically.

B. Sentiment Analysis

Sentiment analysis literally refers to the analysis of emotions. By using sentiment analysis, such as posted comments, one can determine whether consumers have negative or positive emotions and the strength of such emotions. Sentiment analysis can be performed on a per-document or per-sentence basis.

To embed emotion to SEM explained later, we have to recognize emotions on each topic for each review. In this study, we regard the average of emotion values ranging between -1(negative) and 1(positive) as document emotions by calculating Equation (1) as

$$E_{im}=\frac{1}{|T_i(S_m)|}\sum_{s\in T_i(S_m)} E(s), \qquad (1)$$

where E_{im} is the emotion about the topic T_i of the review R_m; S_m is a set of sentences in R_m and | | is the element number of a set; $T_i(S_m)$ represents the sentence set of S_m, including the topic *I*; and the function *E* recognizes the emotion of a sentence. If there is no sentence related to a topic, (1) calculates 0 and regards this sentiment about the topic as neutral. The longer the review, the more likely it is to include other topics. Therefore, it is possible to extract emotions related to topics more accurately by focusing only on sentences containing topics in reviews.

Here, valence aware dictionary for sentiment reasoning (VADER) [15] is used as function *E* in the equation. This

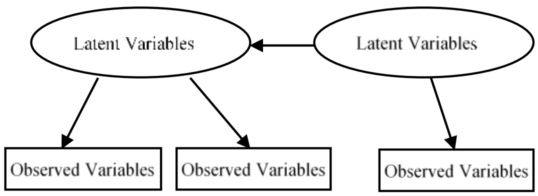

Figure 3. Path model of SEM

method is particularly accurate for sentiment analysis in social media. There are several studies that used VADER. One study analyzed the correlation of positive and negative user reviews of mobile apps before and after app update, respectively, by using VADER because VADER has the high precision in the social media field [16]. In VADER, the value of emotion is represented by -1 to 1 (the closer to -1 the more negative and the closer to 1 the more positive the emotion). Therefore, the E_{im} outputs the value between -1 and 1.

C. Strucuture Equation Modeling (SEM)

SEM [17] is a technology characterized by the use of factor analysis and regression analysis. Factor analysis is the idea that observed variables are based on some hidden factor, and the influence of the factor is to be determined by “correlation” (variance / covariance). Regression analysis is a technique for finding the relationship between a variable to be predicted (target variable) and a variable (explanatory variable, independent variable) that describes the target variable. In other words, SEM can be considered as a factor regression analysis.

The SEM can express causal relationships between variables visually and quantitatively by using a path model, as shown in Figure 3. A path model consists of three elements: latent variables, observed variables, and paths. Latent variables are factors that cannot be observed in actual. Observation variables can actually be observed and are essential for estimating a latent variable. In the path model. Latent variables are represented by ellipses and observation variables are represented by rectangles. The causal relationship between such items is represented by the path of the arrow, and the degree of influence is represented by the path coefficient.

D. Construct Path Model and Find Knowledge

Topics that cannot be observed directly are considered as latent variables serving as correspondence between SEM and topic model. The keywords that make up the topic, the emotion for the topics, and the rating values of each review are the observation variables. From the idea of the topic model that words are generated by topics, each topic is regarded as a factor and the paths from the topics are drawn to the keywords to which the topics are related. Moreover, the paths between topics are drawn from the upper topics to the lower ones based on the idea of the hierarchical structure of the hLDA topics.

Next, we explain the process of incorporating emotional information into the path model. Emotional information influences the intention of a model. Thus, we have to carefully determine how to incorporate emotional information. Generally, emotions for service are generated as perceived experience (after the service) or the expectation (before using the service). Therefore, the model is expressed by drawing a path to emotional information from each topic. When we draw a path from the topic to emotional information, the causal relationship between the emotion and the topic becomes clear. Moreover, rating evaluation is considered to be generated from the top-level topic that includes all elements. Therefore, by drawing the path from the top-level topic to the rating evaluation, the model can represent the causal relationship with the rating.

Furthermore, by comparing the values of path coefficients from the higher topics to the lower topics, it is possible to find an important factor for the rating. By paying attention to the path coefficient from the lowest topic to the keyword, we can find the degree of influence of more detailed factors. The path coefficients from each topic to emotion are large and the causal relationship with emotion could be expressed. By comparing the path coefficient from each topic to emotion, topics with a larger causal relationship with emotion can be found.

However, the path model of SEM is usually prone to model identification failure, especially if there are too many latent variables. Conversely, if the number of latent variables is less, the amount of information in the model may be too small for interpretation. As the topic is a latent variable in the path model, the number of topics must also be adjusted. We also need to remove unreliable paths and observation variables with relatively small influence.

IV. EXPERIMENTS

The purpose of this experiment is to confirm the feasibility of proposed approaches described in Section III.

TABLE I. DATA AND RESULT

Dataset Name	# of Reviews	GFI	AGFI	RMSEA	BIC
Hotel	8104	0.9025	0.8881	0.05525	9188
Airport	13444	0.9152	0.9005	0.05266	12950
App	5442	0.8979	0.8835	0.05960	6848
e-Commerce	19354	0.9213	0.9060	0.05446	19272

Copyright (c) IARIA, 2019.

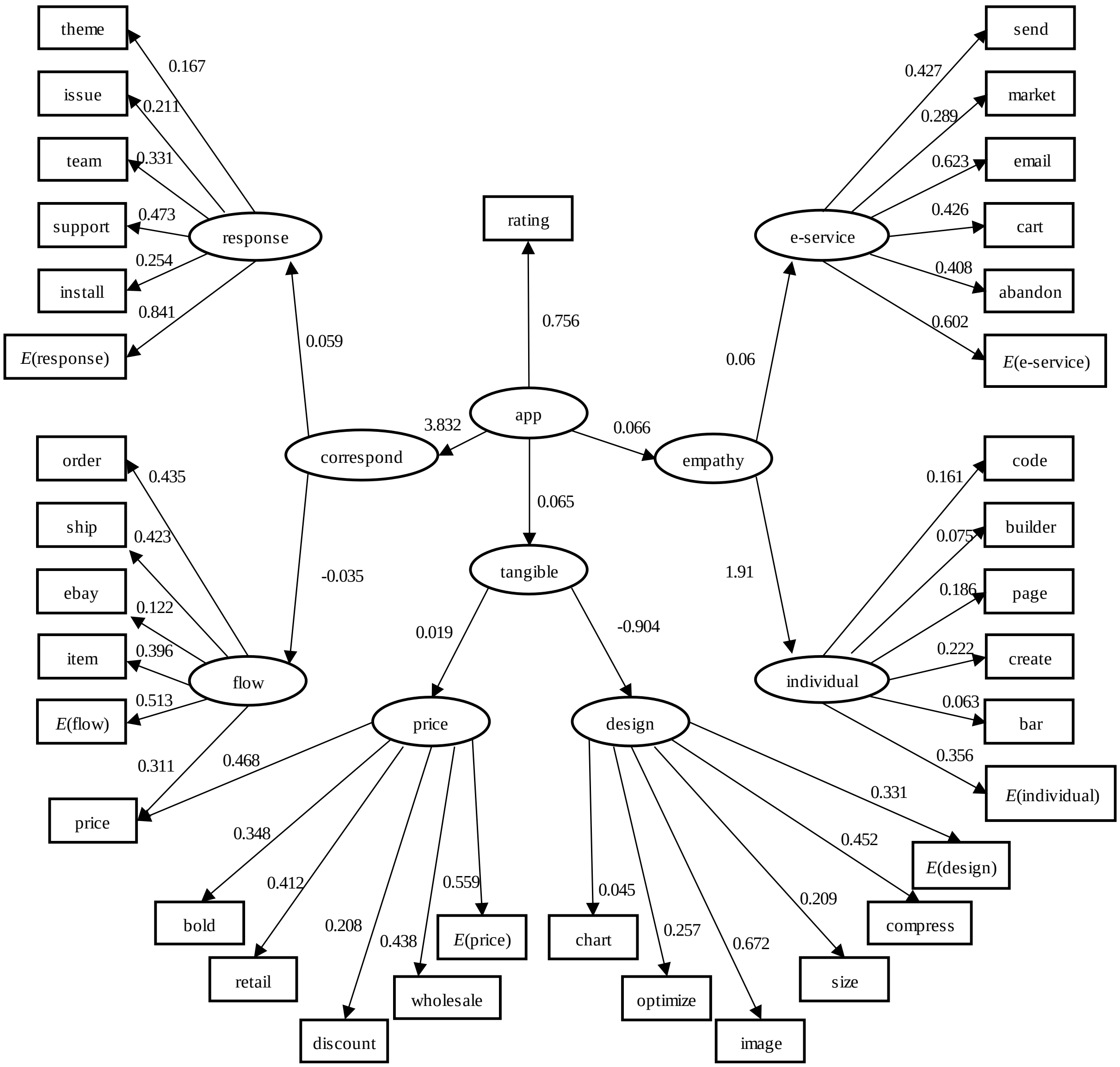

Figure 4. Analysis result of the app dataset

Furthermore, we consider the experimental results.

A. Dataset, Parameters, and Processing

In this analysis, the data must have text data and numerical evaluation data, and it is ideal to have as many review data as possible in order to apply the topic model. In addition, in order to characterize statistical data based on the concept of Bag of Words, the text of one review data must include many words. In this experiment, we employ user-reviews of the datasets published online by Kaggle and Github: the airport, hotel, app for shops and electronic services for purchasing clothes. Airport, app and electronic services reviews are collected by web scraping. Hotel reviews are provided by Datafiniti's Business Database. Each review has review text with a rating between 1 and 5 or 1 and 10. We also regard a review text as a document. In order to ensure that the topics and the appearance frequency of the feature words described are included in each document, only documents stated with more than 30 words are used. The app analyzes information from randomly extracted data. The number of reviews after these pre-processing is shown in Table 1. In this experiment, emotions on topics in the lowest level are determined for the construction of a path model. Moreover, T_i in (1) indicates a topic of the lowest level (i.e., topic in third level). Whether a sentence includes or does not include a topic is determined based on whether or not a keyword constituting the topic is included.

As criteria to evaluate the result, we use goodness of fit index (GFI), adjusted GFI (AGFI), root means square error of approximation (RMSEA), and Bayes information criterion (BIC) were used. As equations for GFI, AGFI, RMSEA, BIC,

$$GFI = \frac{tr((\Sigma(\hat{\theta})^{-1}(S - \Sigma(\hat{\theta})))^2)}{tr((\Sigma(\hat{\theta})^{-1} - S)^2)}, \quad (2)$$

where $\Sigma(\hat{\theta})$ is the estimated value of covariance matrix and

Copyright (c) IARIA, 2019. ISBN: 978-1-61208-757-3

S is value of the actual sample covariance matrix. $tr((A)^2)$ expresses $tr(AA')$,

$$AGFI = 1 - \frac{n(n+1)}{2df}(1 - GFI), \quad (3)$$

where n is the number of observed variables and df is degrees of freedom,

$$RMSEA = \sqrt{\frac{\max\left[\frac{\chi^2 - df}{N-1}, 0\right]}{df}}, \quad (4)$$

where N is the number of samples,

$$BIC = \chi^2 - df\log(N). \quad (5)$$

And as an equation to calculate degrees of freedom,

$$df = \frac{1}{2}n(n+1) - p, \quad (6)$$

where p is the number of variables in equation. Equation (2) expresses how well the total variance in the saturation model can be explained by the estimation model. A value between 0 and 1 is taken and the closer a value is to 1, the better the model becomes. A value of 0.9 or higher is desirable. GFI is unconditionally improved in fitness as model degrees of freedom decreases. Equation (3) corrects the shortcomings of GFI and penalizes models with many parameters and high complexity. The same value as GFI is taken, and the closer it is to 1 the better the resulting model. If the model is not complex, GFI and AGFI will be close values. Equation (4) is an index that expresses the difference between the model distribution and the true distribution. The fit is good with a value of 0.05 or less, and the fit is bad with 0.1 or more. Equation (5) estimates the posterior probability based on chi-square value when the model is selected. This is used to evaluate the balance between model suitability and the amount of information and is used in carrying out relative evaluation. It is better for the value to be smaller.

In this experiment, we used several packages and libraries: Mallet package [18] for hLDA, Python's nltk package with VADER technology [19] for sentiment analysis, and SEM package of R [20] for SEM analysis.

B. Result

Table 1 shows the calculation results of the evaluation indexes for each data and analyzed models. From Table 1, we could find that hotel, airport, and e-commerce models have a GFI of over 0.9 and AGFI maintains high levels. Moreover, none of the models have an RMSEA of less than 0.05 but the values are close to 0.05. It can be said that all of models fit well to the dataset and the constructed models are reliable from the viewpoint of these indices.

As an example, let us show the result of the app dataset in Figure 4. The words at the bottom of the model are those that make up the identified topics from the text data of the review using the topic extraction with hLDA. Here, the topics (latent variables) are estimated by authors from the words that make up each topic. For example, “response” is estimated because it has a large causal relationship with “support” and is considered to be a topic related to responses to actions, such as “install,” “team,” and “issue.” We were able to create a path model based on the hierarchical structure of a text data document group revealed by hLDA. Further, causal relationships can be analyzed by paying attention to arrow and values calculated by SEM between topics or between topics and words at the bottom of the model.

We focus on the “correspond” area with a large path coefficient from the top topic because this “correspond” can be considered as a topic having a large influence to the evaluation (rating). The “response” is also considered to be an important factor for evaluation because this path has a larger path coefficient after comparing between the two topics under “correspond.” Here, the path between the latent variable “response” and the value of the emotion has a large coefficient, implying that “response” influences the emotion strongly. Given that “correspond” has a strong path, therefore, it can be considered that the emotion of response also leads to evaluation.

In the same way, when we check the other paths to emotions, we could find the relationships with and influences to evaluation. From the figure, “response,” “flow,” “price,” and “e-service” have an effect of emotions (the paths over 0.5) and the “design” and “individual” did not. We are not certain whether the results agree or not, but this specific one indicates which topics lead to emotional satisfaction. In this way, it is possible to improve the service by quantitatively understanding the specific service factors that influence to the emotions and evaluation.

We summarize the following findings from the experiments:

- We obtained knowledge by analyzing service while considering emotions.
- We determined the impact on the rating of each topic.
- We obtained the causal relationship between each topic and emotion quantitatively and provided clues for further analyses.

V. Conclusion

In this paper, we analyzed the causal relationships in service by using SEM and emotional information. We constructed the path model by using hLDA and sentiment analysis between topics and emotions. The findings of the experiment using the user reviews of airports, hotels, shopping apps, and electronic services show the feasibility of our proposed model.

We also performed service analysis considering emotion and obtained knowledge reflecting emotional information from the user reviews. The consideration of emotional information is essential for service analysis, and the creation of path models with emotional information is considered effective in extracting information that helps increase service satisfaction. It is suggested that the analysis process in this paper may provide useful knowledge for service analysis and service improvement. On the one hand, this can be used by service providers in improving services and creating new services. Service providers can quantitatively find factors that have major impacts on the evaluation of services and customer emotions. On the other hand, it can be used by service users to efficiently grasp the outline of services that are not formed. Although we analyzed the indefinite service in the experiments, it can be applied to other things like tangible products. The potential applicability is high because analysis is performed from the text.

As future works, we have to consider three points: emotion expression. Firstly, we extracted emotion information of topics based on (1), but this equation does not consider the length of the sentence. Nevertheless, it enables us to accurately determine the emotion on the topic by considering the weight based on the sentence length. For example, longer sentences are more likely to include other topics. Therefore, it may be possible to extract emotions related topics more accurately by reducing the impact of such sentences on emotions of specific topics. Secondly, when two or more topics are included in one sentence, even if it is used in a contrasting sentence, such as "(Text about TOPIC A) but (Text about TOPIC B)," the same emotion value is calculated for the topic. If there is a conjunction (e.g., "but"), a more accurate emotion analysis can be performed by further processing, such as dividing. Finally, in this paper, the accuracy improvement and knowledge are obtained by constructing path models under different assumptions during the construction of the path model.

REFERENCES

[1] Central Intelligence Agency: *The World Factbook: GDP - Composition, by sector of origin*. [Online]. Available from: https://www.cia.gov/library/publications/the-world-factbook/fields/214.html, [retrieved: October, 2019].

[2] A. Parasuraman, V. A. Zeithaml, and L. L. Berry, "SERVQUAL: A Multiple-Item Scale for Measuring Consumer Perceptions of Service Quality", Journal of Retailing, vol. 64, No. 1, pp. 12-40, 1988.

[3] Airports Council International. ACI: *Airport Service Quality, ASQ. : The ASQ Barometer*. [Online]. Available from: https://aci.aero/customer-experience-asq/services/asq-barometer/, [retrieved: October, 2019].

[4] V. Liljander and T. Strandvik, "Emotions in service satisfaction", International Journal of Service Industry Management, vol. 8, Issue 2, pp. 148-169, 1997.

[5] J. J. Cronin and S. A. Taylor, "Measuring Service Quality: A Reexamination and Extension", Journal of Marketing, vol. 56, No. 3, pp. 55-68, July 1992.

[6] M. Ali and S. A. Raza, "Service quality perception and customer satisfaction in Islamic banks of Pakistan: the modified SERVQUAL model", Total Quality Management & Business Excellence, vol. 28, Issue 5-6, pp. 559-577, November 2015.

[7] P. K. Sari, A. Alamsyah, and S. Wibowo, "Measuring e-commerce service quality from online customer review using sentiment analysis", Journal of Physics: Conference Series, vol. 971, Issue 1, 2018.

[8] F. D. Orel and A. Kara, "Supermarket self-checkout service quality, customer satisfaction, and loyalty: Empirical evidence from an emerging market", Journal of Retailing and Consumer Services, vol. 21, Issue 2, pp, 118-129, March 2014.

[9] R. Kunimoto and R. Saga, "Causal Analysis of User's Game Software Evalution Using hLDA and SEM", IEEJ, vol. 135, Issue 6, pp. 602-610, 2015.

[10] K. Lee and C. You, "Assessment of airport service quality: A complementary approach to measure perceived service quality based on Google reviews", Journal of Air Transport Management, vol. 71, pp. 28-44, August 2018.

[11] L. Martin-Domingo, J. C. Martín, and G. Mandsberg, "Social media as a resource for sentiment analysis of Airport Service Quality(ASQ)", Journal of Air Transport Management, vol. 78, pp. 106-115, July 2019.

[12] S. Deerwester, S. T. Dumais, G. W. Furnas, T. K. Landauer, and R. Harshman, "Indexing by latent semantic analysis", Journal of The American Society for Information Science, vol. 41, Issue 6, pp. 391-407, 1990.

[13] D. M. Blei, A. Y. Ng, J. B. Edu, and M. I. Jordan, "Latent dirichlet allocation", The Journal of Machine Learning Research, No. 3, pp. 993-1022, 2003.

[14] D. M. Blei, T. L. Griffiths, M. I. Jordan, and J. B. Tenenbaum, "Hierarchical topic models and the nested chinese restaurant process", Proceedings of the 16th International Conference on Neural Information Processing Systems, pp. 17-24, 2003.

[15] C. J. Hutto and E. Gilbert, "VADER: a parsimonious rule-based model for sentiment analysis of social media text", International AAAI Conference on Web and Social Media, 2014.

[16] L. Xiaozhou, Z. Zheying, and S. Kostas, "Sentiment-aware Analysis of Mobile Apps User Reviews Regarding Particular Updates", Proceedings of The Thirteenth Intetnational Conference on Software Engineering Advances, pp. 99-107, 2018.

[17] C. J. Anderson and W. D. Gerbing, "Strucutural equation modeling in practice: A review and recommended two-step approach.", Psychologcal Bulletin, vol. 103, No. 3, pp. 411-423, May 1988.

[18] A Kachites, "Mallet: A Machine Learning for Language Toolkit", http://mallet.cs.umass.edu, [retrieved: October, 2019].

[19] "Natural Language Toolkit – NLTK 3.4.4 document", https://www.nltk.org, [retrieved: October, 2019].

[20] R. Ihaka, R. C. Gentleman, "The R Project for Statistical Computing", https://www.r-project.org, [retrieved: October, 2019].

Copyright (c) IARIA, 2019. ISBN: 978-1-61208-757-3