Study on Service Quality of Select Indian Banks: Usage of Data from Online Review Sites

Somnath Chakrabarti, Indian Institute of Management (IIM) Kashipur, India
Deepak Trehan, IIM Kashipur, India
Mayank Makhija, Global Consulting Firm, India

The European Business & Management Conference 2016
Official Conference Proceedings

Abstract
Measuring Service Quality has bas been an area of interest for researchers since 1980s. As the retail banking institutions become more customer centric, the focus on customer service quality is increasing across the world. Pre-existing service quality frameworks such as SERVPERF and SERVQUAL have been applied to evaluate the quality levels in banking. However, these frameworks are expensive, as these instruments need to be replicated across the bank branches. With this in consideration, through this study, we have explored a cost and time effective approach to approximate SERVPERF model based on sentiment analysis of online reviews on various social media sites. This paper proposes an innovative approach to measure service quality in a cost effective way. In this paper, our main objective is to analyze customer reviews to better understand bank's service quality and performance. We have collected large number of online reviews from a website for three private Indian banks. Our data set is distributed into three banks that have a similar proportion of reviews (33% each). For each bank, we have a similar mix of products- Credit Card(15-31%), Loan(60-71%) and CASA(9-11%). Further, we also note that the average ratings across banks and products reveals that customers feel differently about different products and banks.The reviews have been mapped to RATER dimensions of SERVPERF model, followed by calculating sentiments for each of these dimensions by adapting an upcoming field in informatics called as text analytics. Finally, a logistic regression model has been developed to understand importance of RATER dimensions in the mind of consumers. Our results show that improved sentiment on RATER dimensions especially on Tangibles and Responsiveness can lead to enhanced customer satisfaction.

Keywords: Service Marketing; Opinion Mining

Introduction

Indian banking industry has seen a paradigm shift due to liberalization and globalization measures initiated in 1991 in the country. Entry of private and foreign banks has resulted in increased competition to acquire and retain customers, making banks sensitive to needs of the customers. The banks are moving away from transactional to more relationship or customer focused operating models. Today, Banks are looking to become integral part of customers' lives by even moving to newer social media platforms such as Facebook. Leading banks in India now are at the forefront of developing solutions and facilitating the digital banking revolution which make banking simple and convenient for its customers through customized solutions to specific segments.

Banks in India have been waking up in recent times to the impact of social media on customers' perception of bank's products and services. Reserve Bank of India (RBI) (Deputy Governor) has suggested public sector banks to use social media to profile customers and to gain deeper insights into customers' creditworthiness [Reserve Bank of India 2011]. Keeping existing customers and acquiring new clients require study of customers' preferences in choosing banks (and their products) and identification of the critical effective factors for bank selection/rejection. Hence, marketers must gather as much information as they can about attitudes and preferences of customers. Therefore, it is very critical for banks to continuously monitor and improve their service levels. Moreover, banks also need to manage and maintain perception on online platforms, social media and review sites.

Measuring service quality has bas been an area of interest for researchers since 1980s. Out of the related models, many researchers have applied SERVPERF or SERVQUAL models to evaluate service quality for banks. In today's fast paced environment, collecting, processing and analyzing survey responses can be both time consuming and costly. Newman (2001) has studied the implementation of SERVQUAL by a major UK high street bank at an annual cost of one million pounds. Accordingly, SERVPERF implementation may be in the similar price range also. With this in consideration, through this study, a cost and time effective approach has been explored to approximate SERVPERF model based on sentiment analysis of online reviews.

Well established service quality literature and sentiment analysis literature have served as the foundation of this study. In this study, the main objective has been to analyze customer reviews to better understand bank's service quality and performance through development of the new model. We have collected large number of online reviews from a website for three leading private Indian banks. These large numbers of comments have helped us to develop a deeper understanding of underlying structure of the reviews. We have then drawn upon text mining literature and have deployed a novel sentiment analysis technique to understand what consumers are talking about in these reviews, and how this can be mapped to five theoretical dimensions of SERVPERF model. We have studied the correlation of the sentiment analysis results with the review ratings given by the users for each bank. We have tested the validity of the sentiment analysis technique and have then developed a logistic regression model to identify important service quality dimensions for Indian banking customers.

One of the most popular assessment tools is SERVQUAL developed by Parasuraman et al. (1988). In the model, service quality has been conceptualized within the disconfirmation paradigm and within the widely accepted American perspective on service quality in academics. They have defined service quality as the gap between customers' expectation from the service and their perception of service received (Parasuraman et al. 1988). Parasuraman et al. (1985) initially provided a list of ten determinants of service quality; access, communication, competence, courtesy, credibility, reliability, responsiveness, security, understanding and tangibles which they subsequently aggregated to five RATER dimensions as mentioned below.

1. Reliability - ability to perform services dependably and accurately
2. Assurance - knowledge and their ability to perform with confidence and trust
3. Tangibles – physical location, equipment, appearance is up-to date
4. Empathy – care and individual attention the firm provides to its customers
5. Responsiveness – willingness to help and respond to customer need

SERVQUAL model has been criticized (Buttle 1996) both on theoretical and on operational fronts. Some of the limitations of SERVQUAL model lie in its lack of applicability and generalizability across different service industries and in its psychometric properties such as the use of gap scores and the measurement of expectations. Accordingly, Cronan & Taylor (1992) have proposed to evaluate services only on the performance scale and not on expectations. They and other researchers have highlighted that the performance based measure has a higher degree of model fit, and have explained more of the variation in overall measure of service quality than the gap based SERVQUAL scale. In the process, they have reduced the overall instrument size to 22 items from original 44 items. They have shown that for banking industry, SERVPERF model performs better than SERVQUAL model. SERVPERF model has been deployed and tested extensively on banks across various cultures and countries. Most of these studies have showed importance of five dimensions of SERVPERF model in assessing service quality of the banks. As customer expectation is not always easy to measure from customer reviews, we have evaluated banks in the present study using SERVPERF model.

Opinion mining, popularly known as sentiment analysis, is an analytical technique that measures polarity (positive, negative or neutral) of the language. When a customer writes a review or posts a message or tweets about some topic, essentially the customer is expressing his/her sentiment about a topic for which he or she feels strongly about. There are two broad challenges in sentiment analysis – 1) identify product features, services, or topics customer is talking about 2) to decide whether these reviews, comments, or messages are positive or negative and to what degree. An important broad algorithm for estimating the sentiment in the comments and reviews is Lexicon based methods. In this approach, opinion based words, which are commonly used for expressing sentiments, are used to estimate sentiment associated with product features. This approach is comparatively simple to understand and implement. The numbers of positive and negative words near to the product feature are counted. If the number of positive words is greater than the negative ones, then we assign a positive opinion to the product feature otherwise negative. Hu & Liu (2004) have categorized approximately 6,800 words into positive and negative ones for sentiment analysis. Nielsen (2011) has created a list of opinion words by assigning strength or degree to the opinion words so as to negate the cancellation effect of positive and negative words. The opinion lexicon can also be built by bootstrapping process through a list of seed words using WordNet.

For this study, we have used lexicon based approach to evaluate performance of banks using the dimensions of SERVPERF model as a reference. Generally, when a customer writes a review, he or she tries to capture a recent moment of truth with the service provider. This essentially means that even when a customer may have formed an opinion about the service provider, the review may capture only two or three dimensions. Since, we are aggregating large number of reviews; all five RATER dimensions would essentially be captured for the service provider.

One purpose of this study is to demonstrate that a non-conventional approach can be studied to understand service quality levels of the banks. The methodology we have developed and refined utilizes sentiment analysis approach to determine the service levels from customers' perception point of view, as expressed in online reviews. The methodology follows a structured approach which starts with downloading data from online mediums, followed by consolidation of data from these sources. This study has used reviews from customer product or service ratings websites and forums (social media encompass wide range of online, word-of-mouth forums including the chosen forum here).

For our empirical analyses, we have collected data mainly from an Indian review site which also acts as an intermediary for financial products – bankbazaar.com. On this site, any user can become a member, and can write reviews including on banking products such as home loan, car loan and credit card. Apart from sharing his/her feedback or opinion, users also provide a rating to the review. This review rating has been taken as the satisfaction level of the customer.

Wenjing et. al (2013) have used the methodology of extracting the most frequent unigrams (single words such as Document) and bigrams (a pair of words such as phone call) and have created a mapping between these and RATER dimensions for travel and tourism industry. Whenever these words occur in a verbatim, they have assigned the corresponding dimension to the verbatim. We have replicated this methodology for assigning RATER dimensions to the reviews. Considering that opinion words, and product and service features vary with domain, it is imperative to build a list specific to banking domain to identify these words. While writing reviews, customers use opinion words such as "unresponsive", "obsolete" to express sentiment associated with a product or service. Two domain experts have identified the most frequent words that could be closely related to SERVPERF dimensions. In total 213 words (including variants) have been identified to maximize coverage of tagged reviews.

After assigning each review to one or more RATER dimensions, we have imported the data in a tool called R for sentiment analysis. The library called SYUZHET (an R package for the extraction of sentiment and sentiment based plot arcs from text) is installed for sentiment computation of each review. This library facilitates easy use of Lexicon based approach explained earlier. Library has functions to assign sentiment to each response using three lexicons including "afinn" developed by Finn {\AA} rup

Nielsen, "bing" developed by Minqing Hu and Bing Liu, and "nrc" developed by S. M. Mohammad and P. D. Turney (CRAN Project). In a lexicon based approach, pre-existing lexicons that contain words already tagged as positive, negative or neutral can be used to tag responses and to score them on a numerical scale from negative to positive. The magnitude on this scale depends on the lexicon used. We group ratings from [0,3) as Low Rating, [3,3] as Medium Rating, (3,5] as High Rating and average the sentiments obtained on the review level across these groups. These groupings are made because the number of reviews in some of the ratings is low and averages might be prone to biases if they are noted for individual ratings. The overall purpose of the averages across Low, Medium and High rating is to quickly validate if the calculated sentiments make sense in the context of ratings.

After assigning each response to one or more RATER dimensions and to sentiment scores, we have prepared the data for modeling. Since we are interested in testing whether RATER dimension sentiment scores are correlated with customer satisfaction, we postulate that it is sufficient to test whether polarity and magnitude of sentiment scores can predict if the rating has been good or bad. We define 'good' and 'bad' by grouping together any rating from [0, 2.5] (bad ratings) as 0 and (2.5, 5] (good ratings) as 1.

We have also tested whether there is a statistically significant interbank difference between RATER sentiment scores. The comparison helps us to identify points of parity (POP) and points of difference (POD) across Banks. We note that all banks focus on Tangibles (POP). We have also tested whether the differences in sentiments are statistically significant among the three Banks. This is done using the results of t-test.

We note that differences of means of sentiment scores between Bank A and Bank B are almost insignificant across dimensions except on reliability where the mean sentiment score of Bank B is higher than the same for Bank A in a statistically significant manner (probability: 0.047). We note that the differences are statistically significant for all dimensions except tangibles when Bank B is compared with Bank C (Scores for Bank B are higher than the same for Bank C). Similarly, there are statistically significant differences on dimensions of empathy and responsiveness when Bank A is compared with Bank C (Scores for Bank A are higher than the same for Bank C)

In this context, to test whether sentiment scores across RATER dimension are correlated with customer satisfaction of Indian customers, we have built the logistic regression model. The model highlights that tangibles and responsiveness are statistically significant variables. This could mean that focusing on tangibles is useful to get good customer satisfaction ratings. We also note that not all banks focus on responsiveness, although it contributes to good ratings.

The results of the logistic regression model (built using generalized linear model - glm) indicate that two of RATER dimensions scores (Average Responsiveness sentiment score & Average Tangibles sentiment score) are significant at 1% significance level. Our results show that improved sentiment on RATER dimensions especially on Tangibles and Responsiveness can lead to enhanced customer satisfaction. The developed model would help in providing parameters important for customers while deciding to rate banks. We have test the validity of the model by using primary metric like Receiver Operating Characteristics (ROC) Curve. To test the validity of the overall model, we have run it on the test set and get the ROC curve. Area under the Curve (AUC) above 0.6 is considered to be a good fit model. Our model has an AUC of 0.67 which indicates that the model is able to predict whether rating is 'good' or 'bad' with an accuracy significantly better than random chance.

Conclusion

This paper addresses the key issues of service quality and customer satisfaction as faced by the private Banking industry in India. This paper advances methodological thinking through applied research and through description of new model. The study leverages new technological applications to provide key consumer insights. In terms of academic implications of this research, this research paper develops an interdisciplinary approach to study SERVPERF model by combining service quality and text analytics paradigms. It also demonstrates how a novel text analytics approach can be used to develop a SERVPERF model which is considered to be a good fit model based on the metrics of ROC curve.

References

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

Cronin Jr, J. J. & Taylor, S. A. (1992). Measuring service quality: A reexamination and extension. *Journal of Marketing*, *56 (3)*, 55-68.

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* (pp. 168-177). KDD '04, August 22-25, Seattle, USA, ACM., http://userpages.umbc.edu/~juacubi1/733/presentation/Mining_and_Summarizing_Customer_Reviews.pdf, last accessed on February 11, 2016

Newman, K. (2001). Interrogating SERVQUAL: A critical assessment of service quality measurement in a high street retail bank. *International Journal of Bank Marketing*, *19(3),* 126-139.

Nielsen, F. Å. (2011). A new ANEW: Evaluation of a word list for sentiment analysis in microblogs. Cornell University Library, http://arxiv.org/abs/1103.2903, Submitted March 15, last accessed on October 27, 2015

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.

Parasuraman, A., Zeithaml, V.A., & Berry, L.L. (1988). SERVQUAL: A multiple item scale for measuring consumer perception of service quality. *Journal of Retailing*, 64(1), 12- 40

Reserve Bank of India RBI (2011). Connecting the Dots. Address by Deputy Governor on June 28, http://www.rbi.org.in/scripts/BS_SpeechesView.aspx?Id=580, last accessed on May 10, 2015

Wenjing, D., Cao, Q., Yu, Y., & Levy, S. (2013). Mining online user-generated content: using sentiment analysis technique to study hotel service quality. System Sciences (HICSS), 46th Hawaii International Conference on System Sciences, IEEE, 3119-3128, https://www.computer.org/csdl/proceedings/hicss/2013/4892/00/4892d119.pdf, last accessed on August 02, 2015
https://cran.r-project.org/web/packages/syuzhet/index.html, last accessed on May 13, 2016

Contact e mail: somnath.chakrabarti@iimkashipur.ac.in