

Attribute Reduction of Service Quality Based on Factor Analysis and Neighborhood Model

Lin Chen^{*1, a}, Xian Hu^{1, b}

¹College of Information Sciences, Beijing Language and Culture University, Beijing China

^achenlin@blcu.edu.cn, ^bsirly1990@yahoo.com.cn

Keywords: attribute reduction, factor analysis, neighborhood granulation and rough approximation, evaluation of service quality

Abstract. How to choose the attributes of service quality have become the foundation and the key for researches of service quality. In connection with the current situation and characteristics of the existing e-commerce service quality evaluation, we analyze the advantages and shortcomings of the widely used way, which is combining item-to-total correlation and factor analysis to reduce our service attribute scale. Then a method of neighborhood granulation and rough approximation for numerical attribute reduction is proposed. With the comparison of the two methods through specific examples of empirical research data, the validity and superiority of the application of neighborhood model for numerical attribute reduction are verified.

Introduction

With the rapid development of E-commerce, service quality has become a core element of the competitiveness of enterprises under e-commercial model. E-commerce in its early stage of development, the scope of the study mainly confined to the customers' interaction with the network, web-design and some other factors to build a quality conceptual model. As the topic of service quality is being widely studied, it has already become a hotspot. The choice of the attribute of service quality as its fundamental core point, has become crucial with the advancing of the research [1, 2, 4, 5, 11-14]. With rapid development of e-commerce, there are a range of studies on the dimensions, measures and attributes of e-service quality, designing some typical scales for measuring service quality (E-S-QUAL [5], WebQual [6], SITEQUAL [7] and e-TailQ [9]).

As the choice of the attributes of service quality, the method of factor analysis was widely used by academia [3, 4, 6, 7, 8, 9, 13, 14]. This method can achieve a certain sense of result when it applied in some practical cases. However, the result obtained in the attribute-selection of service quality is not stable on account for the limitation of the algorithm itself. By calculating the correlation and re-capturing the dimensions of service quality, we found the significant differences between the original scale and the new one. So we proposed a method for attribute selection of measuring customer-perceived service quality. Using specific examples, we established a neighborhood rough model [10, 11], choosing the proper neighborhood δ and variable accuracy, which can ensure to be effective to reduce the attributes and maintain high reliability.

According to numerous research achievements in the past years, the basic service quality scale in the research is a 28-item scale of seven dimensions: ease of use, practicality, reliability, responsiveness, interaction, empathy, and entertainment. The choice of college students from eleven universities in eight cities of China as the research object is due to the fact that college students have become a major consuming-group for on-line shopping in e-commerce. After collecting data through the importance evaluation of customer-perceived service quality in e-commerce (e.g. Taobao) with our questionnaires, we comprehensively analyze and compare the advantages and shortcomings of the methods of factor analysis and neighborhood rough model in selecting the attributes through how college students perceive service quality with numerical evaluation.

In the paper, a neighborhood rough set model is proposed based on the definitions of δ neighborhood and neighborhood relations in metric spaces. Each object in the universe is assigned with a δ neighborhood subset, called neighborhood granule. We use the dependency

function to evaluate the significance of numerical attributes and construct a forward greedy numerical attribute reduction algorithm to get the attributes selection. Experimental results show that neighborhood model can select a few attributes but keep, even improve dimensions of the scale in comparison with factor analysis.

Establishment of Preliminary Scale

We adopt survey research method so as to ensure the starting point of seven components e-commerce service quality, and explore 28 specific questions as the original table to gauge the service quality. As to the measure of the components of the service quality, the questionnaire was conducted, which includes the importance of customers' perception of service quality, evaluation of customers' satisfaction-extent and the basic statistics. We issued the web questionnaire prior to the formal questionnaire survey, improving the web questionnaire based on some advice reflected from some investigated objects, and we confirmed the real-time questionnaire.

- Ease of use:

- 1) Appealing and well organized web-site and easy-to remember realm name. (EOU1)
- 2) Intuitive and simple navigation, well-organized layout and good visual impression. (EOU2)
- 3) Easy use of the on-line transaction. (EOU3)
- 4) Ease use of the function offered (Quickly register, login and downloading). (EOU4)

- Practicality:

- 5) Abundant, accurate, and easy-to-check information. (PCL1)
- 6) Available to modify or defer the purchasing process at any time without commitment. (PCL2)
- 7) Accurate shopping records, free-evaluation of goods. (PCL3)
- 8) Willing to know about customers' needs (i.e., registering supply in short). (PCL4)
- 9) Reasonable prices shown with the items on the screen. (PCL5)

- Reliability:

- 10) Website being truthful about its offering.(RLB1)
- 11) Ensure of financial payment and delivery. (RLB2)
- 12) Protection of the financial and personal data of customers. (RLB3)
- 13) Information consistent with the entity and as your expectation. (RLB4)

- Interactive:

- 14) On-line consultation (i.e., set up corresponding chat). (ITA1)
- 15) Multiple interactive ways between customers and Web's administer. (ITA2)
- 16) Knowing more friends through Communication area and forum. (ITA3)
- 17) Offering reliable help and solving problems promptly. (ITA4)

- Responsiveness:

- 18) Timely responses to customers. (RPS1)
- 19) Prompt to deliver goods(Within 24 hours). (RPS2)
- 20) Receive the goods within 3 to 5 days. (RPS3)

- Empathy:

- 21) Prior-considered customers' benefit, having compensation policy. (EPT1)
- 22) Multiple delivery options. (EPT2)
- 23) Personalized design of the Web page content, recorded browsing etc. (EPT3)
- 24) Personalized products and services with the recorded information. (EPT4)
- 25) Address complaints friendly. (EPT5)

- Entertainment:

- 26) Cool and attractive Web site design, full of enjoyment during shopping. (ETM1)
- 27) Feedback and reward activities occasionally. (ETM2)
- 28) Amusing games offered on-line. (ETM3)

Sample Design and Data Collection

Data were collected using self-administered questionnaires according to service quality scale from the college students (167 of 175 copies are valid). The importance-extent to each statement about the dimensions of service quality was measured according to the students with their actual

experiences with a 5-point LIKERT-SCALE (5=fairly important, 4=relatively important, 3=unsure, 2=relatively unimportant, 1=absolutely unimportant). Data were analyzed using SPSS13.0 for reliability with Cronbach's Alpha, which is 0.798, indicating high reliability of the questionnaire.

Attributes Reduction Model

Factor Analysis . Factor analysis is a technique for data reduction. It can be used to express its basic data structure by researching the interior dependency relationship among the variables, pursuing the basic structure of the observed data and using few illusion variables. The original ones are observable showing-out variables, while the illusion ones are unobservable potential variables, which are called factors. Exploratory Factor Analysis (EFA), is a technique for finding out the basic structure of multiple observed variables and for reducing dimension-scale. Generally speaking, the principle of choosing the principal component is that the eigenvalue must be greater than 1. As to the cumulative percentage, which is always from 80% to 90%. A typical process of EFA should be,

- a) identify and collect observed variables
- b) obtain the covariance matrix
- c) test and verify the covariance matrix for EFA (significance level, anti-covariance matrix, KMO measure)
- d) choose the way to abstract the factors
- e) discovery the elements and elements loading
- f) confirm the number of abstracted factors (use Kaiser norm as the standard)
- g) interpret the abstracted factors

Principal component analysis is a kind of exploratory statistical analysis method that concentrates the information that scatters on a group of variables on several complex indexes. It turns several variables to few irrelevant principal components to describe the interior structure of the data set by using the thought of reducing dimensions. Assuming that there are P indexes in the practical problems that we are discussing, we regard these indexes as random variables and label them as X_1, X_2, \dots, X_p . The purpose of principal component analysis is to transfer this p indexes problem to a linear combination problem of p indexes and to make sure these new indexes F_1, F_2, \dots, F_k ($k \leq p$), fully reflect the information of the original indexes under the principle of keeping the main information and keep them independently. It has to meet the following conditions:

- a) The factor quadratic sum of each principal component must be 1, $u_{1i}^2 + u_{2i}^2 + \dots + u_{pi}^2 = 1$.
- b) The principal components are independent to each, $Cov(F_i, F_j) = 0, i \neq j, i, j = 1, 2, \dots, p$.
- c) The variance of the principal components decline one by one, and the importance also decline one by one, $Var(F_1) \geq Var(F_2) \geq \dots \geq Var(F_p)$.

Attributes Reduction of Factor Analysis . We do the attribute reduction based on the reduction process that PZB developed the SERVQUAL scale in 1988 and use statistical methods with item-to-total correlation and factor analysis to reduce the service quality attributes. The first step is to calculate the item-to-total correlation to asking items. This method is to compute the related coefficient of each item to any other items. From the view of scholars, once the item-to-total correlation less than 0.4, it means that the item can not indicate any perspective with any other items, and this item will be rejected. Table 1 is the item-to-total correlation to calculate 28 asking items. From the field of Corrected Item-Total Correlation (1) we can see that the item-to-correlation value of 2, 4, 12, 21, 28 is less than 0.4. So we need to delete them, after doing this, the answer is greater than 0.4 and the Cronbach's Alpha index rises to 0.840.

The second step of attribute reduction is factor analysis. We mix Exploratory Factor Analysis, principal component analysis and varimax with Kaiser normalization, a kind of rotation method. The principle of choosing the principal component is the eigenvalue is greater than 1 and the cumulative rate is higher than 80%. The single-item factors and the factors with loading value less than 0.5 will be erased. After doing for several times, the result can be seen in table 2, table 3.

Table 1. Item-total Statistics (A)

	Item	I-T C1	I-T C2		Item	I-T C1	I-T C2
Ease of Use	EOU1	.424	.524	Reliability	RLB1	.480	.448
	EOU2	-.058			RLB2	-.432	-.504
	EOU3	.591	.487		RLB3	.335	
	EOU4	-.054			RLB4	.458	.533
Interactive	ITA1	.753	.563	Empathy	EPT1	.048	
	ITA2	.770	.792		EPT2	.518	.535
	ITA3	.566	.463		EPT3	.758	.758
	ITA4	.488	.688		EPT4	.604	.604
Entertainment	ETM1	.484	.684	Practicality	EPT5	.713	.713
	ETM2	.685	.085		PCL1	.711	.732
	ETM3	.010			PCL2	.438	.518
Responsiveness	RPS1	.769	.652		PCL3	.691	.574
	RPS2	-.450	-.501		PCL4	.593	.467
	RPS3	.451	.414		PCL5	.479	.532

Table 2. Item-total Statistics (B)

COM	Initial Eigenvalues		Extraction S S Loadings		Rotation S S Loadings	
	Total	Cumulative %	Total	Cumulative %	Total	Cumulative %
1	7.462	39.273	7.462	39.273	7.398	38.937
2	4.280	61.799	4.280	61.799	4.131	60.680
3	1.857	71.574	1.857	71.574	1.996	71.185
4	1.340	78.624	1.340	78.624	1.335	78.213
5	1.077	84.292	1.077	84.292	1.155	84.292
6	0.840	88.714				

Table 3. Component Matrix and Rotated Component Matrix

Component Matrix						Rotated Component Matrix					
	Factors (Components)						Factors (Components)				
	1	2	3	4	5		1	2	3	4	5
23.	.987	-.062	.021	-.027	.052	23.	.990	.034	.004	-.028	.013
15.	.986	-.051	.018	-.016	.021	15.	.985	.042	.009	-.022	.046
18.	.986	-.051	.027	-.048	-.001	18.	.984	.040	.017	-.059	.060
14.	.979	-.047	.006	-.060	.015	14.	.977	.051	-.005	-.066	.046
5.	.968	-.093	-.021	.019	-.003	5.	.969	.003	-.035	.006	.078
25.	.955	-.089	.031	-.037	.080	25.	.963	.006	.003	-.033	-.021
7.	.927	-.080	-.031	.149	.028	7.	.929	.007	-.036	.142	.074
24.	.853	-.075	.034	.039	.013	24.	.856	-.002	.022	.028	.052
6.	.134	.937	-.151	-.021	.027	6.	.040	.952	.085	.036	.061
8.	-.022	.907	-.202	-.074	.063	8.	-.109	.927	.020	-.004	.009
3.	.115	.860	-.250	-.015	-.021	3.	.023	.894	-.024	.034	.116
9.	.203	.847	-.040	.062	.069	9.	.049	.850	.108	-.087	.018
17.	.131	.836	-.091	-.140	.031	17.	.124	.843	.171	.115	.021
26.	.029	.359	.923	.057	-.074	26.	.013	.121	.987	.002	-.024
16.	-.010	.392	.906	.062	-.096	16.	-.031	.150	.982	.004	.000
11.	.018	-.099	-.073	.812	-.031	11.	.023	-.143	-.033	.783	.202
10.	-.023	.251	-.030	.666	.351	10.	-.025	.225	.042	.738	-.183
22.	.231	.112	-.099	-.101	-.692	22.	-.005	-.076	.024	-.256	-.755
19.	-.068	-.126	.164	-.396	.662	19.	.168	.099	.004	-.240	.685

By analysis, we can see that the results of Principal Component Analysis and Rotation Method: Varimax with Kaiser Normalization are same. In the mean time, the Cronbach's Alpha index has increased to 0.859. We abstracted five factors from the table by doing the factor analysis, and their asking items and perspectives have a huge difference comparing with the original design. With

regards to the seven perspectives of the original scale, after factor analysis, each attributes minor term re-abstracted corresponding potential factors. But the interpretability of the components of each factor is really bad. The minor terms that belong to the same factor should originally indicate the same aspect of the attribute dimension. But in the consequence, we found minor terms that have different perspectives with those in the original scale, while their meanings are totally from different aspects. So this kind of consequence is unacceptable. Taking Factor (Component) 2 as an example, we can see it includes five items (6, 8, 3, 17, 9) where item-3 should be presented in Ease of use and item-17 should be presented in dimension of responsiveness. In fact, this problem can be found in many document literatures, which means the consequences that achieved by using this method are unstable. Based on this research, we can do an overall evaluation to method of principal component analysis.

a) It can eliminate the correlated impact among the evaluation indexes, therefore reduce the work of index choosing.

b) The weights that achieved by doing overall merit with principal components analysis are automatically generated with the mathematical transformation. They are objective.

c) The consequences of overall merit are unstable.

Neighborhood Model. Rough set theory was developed by Zdzislaw Pawlak in the early 1980's. It offers mathematical tools to discover patterns hidden in data. In the research, a neighborhood rough set model is proposed based on the definitions of δ neighborhood and neighborhood relations in metric spaces.

Neighborhood Decision System. *Definition 1* Let $NDT = \langle U, A, D \rangle$ is a neighborhood decision system, from D we can get U divided into N -equivalence classes: X_1, X_2, \dots, X_n , for $\forall B \subseteq A$, We can approximate D relating to B 's lower and upper approximations of X :

$$\begin{aligned}\underline{N}_B D &= \{x_i | \delta_B(x_i) \subseteq X, x_i \in U\} \\ \overline{N}_B D &= \{x_i | \delta(x_i) \cap X \neq \emptyset, x_i \in U\}\end{aligned}\quad (1)$$

The B -positive region of D is its lower-approximation: $POS_B(D) = \bigcup_{X \in U/D} \underline{N}_B(X)$. So B depends on

$$D \text{ in a degree } k \text{ denoted by } k = \gamma(B, D) = \frac{|POS_B(D)|}{|U|}. \quad (2)$$

Definition 2 The formula for the lower and upper approximations can be generalized to some arbitrary level of precision $\pi \in (0.5, 1]$ by means of the rough membership function

$$\underline{N}_\pi X = \{x | \mu_X^\pi(x) \geq \pi\}; \overline{N}_\pi X = \{x | \mu_X^\pi(x) > 1 - \pi\} \quad (3)$$

Note: The value of variable precision π for attributes reduction is useful to a certain extent, so we need to consider this parameter.

Forward Greedy Numerical Attribute Reduction Algorithm. *Definition 3* Let $NDT = \langle U, A, D \rangle$ is a neighborhood decision-making system, and $B \subseteq A, \forall a \in B$, then we define the significance of a relative to D as $SIG(a, B, D) = \gamma_{B \cup a}(D) - \gamma_B(D)$.

With this algorithm we use an empty set as a starting point, each time calculating the importance of all those remaining attributes. We have to select the most important attribute of all the attributes to our set defined before until all the remaining attributes with importance of zero, that is, adding any new attributes, the system is no longer dependent-relationship changes in value. We choose forward searching algorithm to ensure that the important attributes could be of the first to be added to the reduction-set, thus no loss of important features. *The Progress of numerical attribute reduction based on neighborhood granulation and rough approximation:*

Let $NDT = \langle U, A, D \rangle$ is a neighborhood decision-making system.

Step1: For $\forall a \in A$, compute the neighborhood-relation N_a

Step2: Define $\emptyset \rightarrow R \quad // \gamma_\emptyset(D) = 0$

Step3: For $\forall a_i \in A - R$, $SIG(a_i, B, D) = \gamma_{R \cup a_i}(D) - \gamma_R(D)$

Step4: Select the attribute a_k that is satisfying:

$$SIG(a_i, R, D) = \text{Max}_i(SIG(a_i, R, D))$$

Step5: If ($SIG(a_k, R, D) > 0$)

$$R \cup a_k \rightarrow R,$$

Go to Step3,

Else

Return R,

End.

Solution of Neighborhood Model . Let $NDT = \langle U, A, D \rangle$, where $U = \{u_1, u_2, \dots, u_n\}$ ($n=167$) respectively represent the survey data of the college students, and their evaluation of the important-extent of the 28 items in relation to reflect service quality as the condition-attribute set A, the mean of each evaluation constitutes the decision-attribute D. Let $\delta = 0.125$, that is, the diameter of the neighborhood is 0.25, choose the value of variable precision $\pi = 0.8$, according to our neighborhood decision systems and forward greedy numerical attribute reduction algorithm, programming with MATLAB realizes the attributes reduction of the scale.

Table 5. Reduction scale of neighborhood model

Ai	Aij	M	M'	Ai	Aij	M	M'
Ease of use	1. EOU1	3.75	3.953	Responsiveness	18. RPS1	4.20	3.91
	2. EOU2	4.11			19. RPS2	3.15	
	4. EOU4	4.00			20. RPS3	4.38	
Practicality	6. PCL2	4.16	4.135	Empathy	21. EPT1	4.12	3.603
	7. PCL3	4.11			22. EPT2	3.26	
Interactive	15. ITA2	4.18	3.913		Entertainment	26. ETM1	
	16. ITA3	3.44		27. ETM3		3.39	
	17. ITA4	4.12		28. ETM3		2.33	

We find that this method can reduce the number of features effectively and get a relatively high classification accuracy and the Cronbach's alpha increases to 0.876. Besides, the flexible combination of the neighborhood size and the choice of variable precision has made the calculated consequence has no deficit to the perspectives of the original design and reduced the redundant attributes. Therefore, the new scale can express all the meanings that the original one has in a more simple way. So we can believe this consequence is more stable than factor analysis, attribute reduction and scale perspectives regeneration. Moreover, this method does not need to do iterative computations and step-by-step reduction as factor analysis does leading to an error or an unstable consequence. This method generates rough set model based on the conceptions of the neighborhood of metric space and neighborhood relationship, pelletized the domain space based on the neighborhood of any objects in the domain space and uses lower approximation and upper approximation to approach any conception in the space, therefore accomplishes the granular computing for the space of real numbers. It is easy to gain reasonable consequence by choosing appropriate variable accuracy and using prior to the greedy algorithm based on the dependant function. The experimental data shows that the neighborhood rough set model could choose very few features while keep or even remarkably enhance the classification accuracy of the reduced data, which proves the rationality and effectiveness of this method.

Conclusions

With the rapid development of electronic commerce, the quality of service has become the core element of company competition ability under electronic commercial models. In this essay, we chose strategic point in allusion to the attributes of service quality, and bound with the current situation and characters of the electronic commercial service quality evaluation to analyze the advantage and disadvantage of the factor analysis and the method of reducing service quality attributes based on the item-to-total correlation that is widespread used right now, and then presented the reduction method of service quality attributes which is based on the neighborhood

granulating and rough approximation. By using basic neighborhood information grain to approximate and by generating prior to the greedy reduction algorithm with dependence function, this method can do feature selection to the attributes of the service quality. Through the analysis and comparison of the specific data, we proved the rationality, effectiveness and advantages of the reduction method of service quality attributes which is based on the neighborhood granulating and rough approximation, and we believe this method have a widespread application prospect.

Acknowledgment

This work is supported by the Nature Science Foundation of China (No. 60973148) and the Project of the Chinese Ministry of Education (No. 109028).

References

- [1] Gronroos, Christian, Strategic Management and Marketing in the Service Sector. Research Reports No.8, Swedish School of Economics and Business Administration, Helsinki. (1982)
- [2] Parasuraman, A., Zeithaml, V.A. and Berry, L.L., A Conceptual Model of Service Quality and Its Implications for Future Research, *Journal of Marketing*, Vol.49 (Fall), pp. 41-50. (1985)
- [3] Parasuraman, A., Zeithaml, V.A. and Berry, L.L., SERVQUAL, A Multiple-Item Scale for Measuring Consumer Perceptions of Service Quality. *Journal of Retailing*, Vol.64, 1 (Spring), pp.12-40. (1988)
- [4] Cronin, J.J. Jr., & Taylor, S.A., Measuring service quality: a reexamination extension. *Journal of Marketing*, 56(3), 55-68. (1992)
- [5] Parasuraman, A., Valarie A. Zeithaml & Malhotra, E-S-QUAL: A Multiple-Item Scale for Assessing Electronic Service Quality[J]. 7(3), 213-233. (2005)
- [6] Loiacono, E. T., Watson, R. T., & Goodhue, D. L., WebQual: a measure of Web site quality. *Marketing Theory and Applications*. (2002)
- [7] Yoo, B., & Donthu, N., Developing a Scale to Measure the Perceived Quality of an Internet Shopping Site (SITEQUAL). *Quarterly Journal of Electronic Commerce*, p31-46. (2001)
- [8] Santos, E-service quality: a model of virtual Managing Service Quality, 13(3), 233-246. (2003)
- [9] Collier, J.E. and Bienstock, C.C., Measuring service quality in e-retailing, *Journal of Service Research*, Vol. 8 No. 3, pp. 260-75. (2006)
- [10] Pawlak Z., *Rough Sets-Theoretical Aspects of Reasoning about Data*. Kluwer Academic, Dordrecht. (1991)
- [11] Qinghua Hu, Daren Yu, Zongxia Xie. Information-preserving hybrid data reduction based on fuzzy-rough techniques. *Pattern recognition letters*. 27 (5): 414-423. (2006),
- [12] Lixin Cui, *Service-Quality Evaluation Model*. Economic Daily Press, Beijing. (2003)
- [13] Kassim, N. and N. Abdullah, The effect of perceived service quality dimensions on customer satisfaction, trust, and loyalty in e-commerce settings. *Asia Pacific Journal of Marketing and Logistics*, 22(3): 351-371. (2010)
- [14] Wenying, W. and Q. Sun, Analysis the Factors Affecting Consumers to Use Internet Banking Services. *Proceeding of International Conference on Information Management and Engineering (ICIME)*: 180-184. (2010)