

# Identifying Crucial Website Quality Factors of Virtual Communities

Long-Sheng Chen<sup>1\*</sup>, and Pao-Chung Chang<sup>2</sup>

**Abstract**—Virtual communities are not only used to share information among their members with common interests, but also an avenue of entertainments, networking with friends, and shopping. Recently, the main research topics on them include social networking, success factors, reasons for participating, and consequences from joining in virtual communities. Relative few works to discuss what important quality factors are for attracting customers to join the virtual community websites. From the viewpoint of increasing margin of websites, to identify important quality characteristics for attracting more internet customers in virtual community is crucial for website enterprisers. In this study, we attempt to define the quality factors of virtual communities and then identify the key factors for attracting new members by using a feature selection technique. In addition, the importance–satisfaction model was also implemented for analyzing the collected data further. Finally, a real case study of “Facebook” has been provided to evaluate the effectiveness of the used methods.

**Index Terms**—Feature Selection, Quality Factors, Virtual Community, Customer Needs.

## I. INTRODUCTION

The virtual community has a 30-year history from the early Bulletin Board System (BBS) to Web 2.0 such as Twitter, Facebook, and My space. In 1993, the term “Virtual Community” was first wrote by Howard Rheingold who considered virtual community was a certain social group of people came from internet. Illum et al. [10] think that an online virtual community is defined as a group of people trying to achieve certain purposes, with a similar interest, interested in relationship building, transaction, and fantasy under certain rules by using new information technology as their means.

Virtual communities can be used for commercial purposes which include helping enterprisers to create new types of services, enhancing existing products and creating new divisions and capabilities [13], strengthening their positive image, establishing relationships with their customers and contributing to customer loyalty and sales [12]. Business potential is mainly used to create increased trust among a virtual community members combined with quality services that may improve customer loyalty [2].

Nowadays, the internet increasingly changing people’s

real lives from different aspects, especially the virtual community whose come forth broken through past conception of space and connected thousands of people from different place by the computer screen near at hand [3]. Some famous examples of virtual communities could be found in many websites such as “Facebook”, “Plurk”, “My space”, “LinkedIn” and “Twitter.” According to the official report of “Facebook”, more than 300 million users who ever visited their website or used their service until September, 2009. 57% users are male. The firstly heavy users belong to the young students whose ages are between 18 and 25 years old (32.8%). The second place is office workers whose ages are between 26 and 34 years old (18.4%). Therefore, virtual communities are not only used to share information among their members with common interests, but also an avenue of entertainments, networking with friends, and shopping.

Recently, virtual communities have been the object of research since their emergence. Major research topics on them include social networking, reasons for participating, and consequences from joining in virtual communities [10]. Spaulding [11] also indicated that success in virtual communities depends on an attitude of contribution, dedication of resources, building a critical mass, and matching community and business needs [11]. However, relative few works to discuss what important quality factors are for attracting customers to join the virtual community websites. From the viewpoint of increasing margin of websites, to identify important quality characteristics for attracting more internet customers in virtual community is crucial for website enterprisers.

In addition, we observed that users have some movements among different virtual communities. If a new virtual community website can create some innovative functions or applications, it will become popular immediately and attract lots of internet users to join. For example, to provide the function of posting home made video, YouTube attract lots of users from those traditional text or photo based blogs websites. Therefore, to identify important quality characteristics of virtual community websites are crucial for survival in commercial competitions. In this study, we attempt to define the quality factors of virtual communities and then identify the key factors for attracting new members. Decision tree (C4.5) has been employed to be the tool of feature selection. Importance–satisfaction model was also implemented for analyzing the collected data further. Finally, a real case study of “Facebook” has been provided to evaluate the effectiveness of the used methods.

Manuscript received December 8, 2009. This work was supported in part by the National Science Council of Taiwan, R.O.C. (Grant No. NSC 98-2410-H-324-007-MY2).

\*L.-S. Chen is with the Chaoyang University of Technology, Taichung 41349, Taiwan (phone: 886-4-23323000ext7752; fax: 886-4-23304902; e-mail: lschen@cyut.edu.tw).

P.-C. Chang is with the Chaoyang University of Technology, Taichung 41349, Taiwan (e-mail: bow1221@cyut.edu.tw).

II. LITERATURE REVIEW

A. Website Quality Models

Due to the rapid development of Web 2.0 service, the innovative website quality elements have become one of key factors for attracting internet customers. It's especially true in virtual community websites. Therefore, website service providers need to identify important e-service quality factors and keep improving them to increase customer satisfaction. In related works, researchers developed some models for measuring e-service quality of websites. In this section, we will introduce some famous models.

Table 1 WebQual 4.0 Model

e-service quality model	Dimension	Item	Content
WebQual 4.0 [9]	Usability	Usability	The main change is an increased emphasis on usability rather than site quality. The focus is on user experience rather than site characteristics.
		Site design	
	Information quality	Information	
		Service interaction quality	
Empathy			

Table 2 Perceived on-line service quality [7]

Quality factors	Description
Content (O1)	Accuracy Believability Timeliness Relevance Ease of understanding Appropriate level of detail Appropriate format
Usability (O2)	Easy to learn to operate Clear and understandable Easy to navigate Easy to use Attractive appearance Design appropriate for site Sense of competency Positive experience Feel in control Response time acceptable
Service interaction quality (O3)	Good reputation Security of personal information Sense of community Easy to communicate with the Organization Confident that goods or services will be delivered as promised Enjoyability or entertainment Managing and integrating roles and relationships Modify and update content Multi-lingual capability Push notifications
Transaction quality and safety (O4)	Transaction safety Transactions that are useful Transactions that save time or money Range of transactions More convenient than alternatives Platform-infrastructure reliability Technical security(e.g. virus protection)

WebQual model was firstly developed by Barnes & Vidgen in 1998. At first, they designed WebQual 1.0 for measuring the data and information quality in websites. Then, focusing on different considerations, several versions such as WebQual 2.0 (2000), WebQual 3.0 (2001) and WebQual 4.0 (2002) [7] have been created. Table 1 summarizes the dimensions and items of WebQual 4.0. Compared with versions 1.0~3.0, the main change in version 4.0 is an increased emphasis on usability rather than site quality. The focus is on user experience rather than site characteristics.

In 2007, Tate et al. also defined perceived e-service quality factors including content, usability, service interaction quality, transaction quality and safety. Table 2 provides these four main quality factors and detailed sub-items.

However, the first 2 models don't focus on virtual community websites. Therefore, eQual model which was developed by Barnes & Richard [9] in 2005 has been introduced. In eQual model as described in Table 3, four dimensions which involve 23 question items have been defined to measure e-service quality [9].

Table 3 The eQual Model

Category	No	Questions
Usability	1	I find the site easy to learn to operate
	2	My interaction with the site is clear and understandable
	3	I find the site easy to navigate
	4	I find the site easy to use
	5	The site has an attractive appearance
	6	The design is appropriate to the type of site
	7	The site conveys a sense of competency
	8	The site creates a positive experience for me
Information Quality	9	Provides accurate information
	10	Provides believable information
	11	Provides timely information
	12	Provides relevant information
	13	Provides easy to understand information
	14	Provides information at the right level of detail
	15	Presents the information in an appropriate format
Service Interaction	16	Has a good reputation
	17	It feels safe to complete transactions
	18	My personal information feels secure
	19	Creates a sense of personalization
	20	Conveys a sense of community
	21	Makes it easy to communicate with the organization
22	I feel confident that goods/services will be delivered as promised	
Overall	23	Overall view of the Web site

The fourth model shown as Table 4 was presented by Deng et al. [3] who focus on users' behaviors in virtual communities to define the functions (quality factors). This model can help researchers to understand the factors that can attract customers to join virtual community websites. To sum up, based on the mentioned above four models, we develop the quality factors of virtual community in this study.

Table 4 Descriptions of defined functions of virtual communities

No.	Defined functions of virtual communities
V1	Help users to find their own strings people
V2	Help users to create the management of the relevant contact people
V3	Remind the updates of the contacts
V4	Remind the users of the act which has direct connection to themselves
V5	Providing the webpage game which promotes cooperation or conflicts to the users
V6	Providing the professional application
V7	Providing with other application

**B. Feature Selection**

Feature selection is often applied in pattern classification, data mining, and machine learning. Reduction of pattern dimensionality via feature selection belongs to the most fundamental steps in data processing [8]. The main purpose of feature selection is to remove irrelevant or redundant attributes and improve the performance of data mining. Among many feature selection methods, GA, rough sets and decision trees have attracted much attention, and have become popular techniques for feature selection. In this study, decision tree algorithm has been employed for selecting important website quality factors.

A common understanding is that some learning algorithms have built-in feature selection, for example, ID3 [5], FRINGE and C4.5 [6]. When decision tree induction is used for feature selection, a tree is constructed from the given data. All attributes that don't appear in the tree are assumed to be irrelevant. The set of attributes appearing in the tree form the reduced subset of attributes [4].

**C. Importance-Satisfaction Analysis (I-S Model)**

The importance-satisfaction model (I-S Model) shown as Figure 1 [1] is a very easy business model for providing enterprisers a brief analysis. In this model, the horizontal dimension and the vertical dimension shows the degree of importance the satisfaction level of the quality factors, respectively. Then, the order pair (importance scale, satisfaction scale) can be located on the coordinates. The means of the importance scale and the satisfaction scale can be used to divide the coordinate into four areas:

**(I)Excellent area**

The quality factors located in this area are those that customers considered to be important, and for which the performance is satisfactory to customers. These quality factors are the major weapons for e-service providers.

**(II)To be improved area**

The quality factors listed in this area are those considered as important to customers but for which the performances have not met with expectation. The websites managers must focus on these factors and improve them immediately.

**(III) Surplus area**

The quality factors located at this area are not very important to customers, but the perceptions of customers are quite satisfactory. Therefore, the company can relocate the resources from surplus area to other areas. If the company

needs to cut costs, these attributes can be eliminated without incurring a significant negative impact.

**(IV) Care-free area**

These quality factors are those which customers have a lower satisfaction level and they also think they are not important. Therefore, enterprisers needn't to care about this kind of quality factors.

Although the importance-satisfaction model is a simple structure, it can provide much useful information about a company's quality performance. Therefore, the model has been employed in this study.

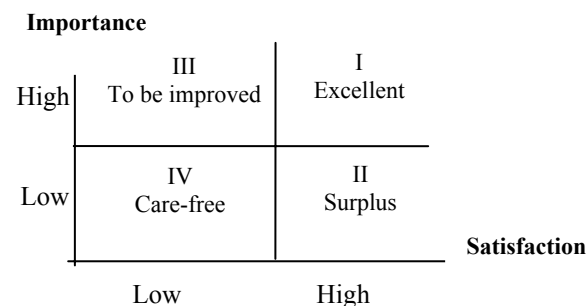


Figure 1 Importance-Satisfaction Model

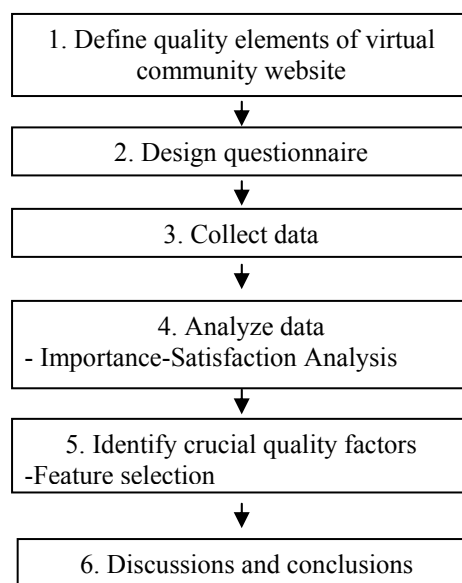


Figure 2 Research procedure

**III. RESEARCH PROCEDURE**

The procedure of this study can be illustrated as Figure 2. It contains 6 steps and they are defining quality factors of virtual community websites, designing questionnaire, collecting data, analyzing data, identifying crucial quality factors and drawing conclusions. In first step, we attempt to define the quality factors of virtual community websites based on related works such as eQual model [9]. If we define inappropriate quality factors, we cannot identify the crucial quality factors of customers. Then, we can design questionnaire according to these defined quality factors. In this step, every factor will expand to two questions to survey the perceived satisfaction and importance. In step 3, we

sample some representative customers to collect data. The purpose of step 4 is to analyze the collected data by applying I-S analysis. The results of I-S analysis can tell us the quality factors belong to which area in Figure 1.

In step 5, we implement feature selection technique to identify the crucial quality factors. In this step, the decision tree (DT) algorithm has been employed to select important quality attributes. DT is one of the most popular knowledge acquisition algorithms, and has been successfully applied in many areas. There are more than one decision tree algorithms, such as ID3 [5] and C4.5 [6]. This study uses C4.5. The core of C4.5 contains recursive partitioning of the training examples. Whenever a node is added to a tree, some subsets of the input features are used to pick the logical test at that node. The feature that results in the maximum information gain is selected for testing at that node. In other words, the algorithm chooses the “best” attribute to partition the data into individual classes at each node. After the test has been determined it is used to partition the examples, and the process is continued recursively until each subset contains examples of one class or satisfies some statistical criteria [14]. Finally, in step 6, we can draw some conclusions according to the results of steps 4~5.

#### IV. IMPLEMENTATIONS

Firstly, we need to define the quality factors of virtual community websites. Based on eQual model and other 2 models including “perceived on-line service quality” and “the functions of virtual communities”, 12 quality factors for virtual community websites have been defined. As shown in Table 5, they are “personal design”, “experience value accumulation”, “games”, “instant message”, “voice/video message”, “gathering persons who have the same habits”, “aficionado group”, “keep personal connection dark”, “transaction platform”, “online trade system”, “free cyberspace”, and “friend search”. Table 6 provides the detailed definitions of these 12 quality factors.

Table 5 The defined quality factors of virtual community websites

No	eQual	Perceived on-line Service Quality	The functions of virtual communities	Defined quality factors
1	Usability	O2	-	Personal design
2			V5	Experience value accumulation
3			V5	Games
4	Information quality	O1	V3	Instant message
5	Service interaction quality	O3	V7	Voice/video message
6			V1	Gathering persons who have the same habit
7				Fans' group
8			V1、V2	Keep personal connections dark
9				Transaction platform
10			O4	V6
11	-	O1	V6	Free cyber space
12			V1、V2	Friend search

Secondly, for each defined quality factor, a pair of

question items has been developed. Take quality factor Q1 for example, a pair of questions can be shown as bellow.

Quality factor 1: personal design

The website provides personal design functions which can allow users to change some settings to create individual features.

(1) If the website provides this quality, how would you feel?

- (A) Very satisfied (B) Satisfied (C) Neutral
- (D) Not satisfied (E) Not very satisfied

(2) Do you think the importance of this quality?

- (A) Very important (B) Important (C) Neutral
- (D) Not important (E) Not very important

Table 6 The description of defined quality factors

No.	Quality factors	Definition
Q1	Personal design	The website provides personal design functions which can allow users to change some settings to create individual features.
Q2	Experience value accumulation	The limits of authority for using some advanced functions in virtual communities are earned depending on the accumulated experiment value.
Q3	Games	The website provides the functions of games, mental tests, or fortune-telling, not only for entertainment, but also for interaction with friends.
Q4	Instant message	The website provides the instant message service which can allow online users to communicate to each other immediately.
Q5	Voice/video message	The website provides the voice/video message service which can leave a voice/video message to your friends who are not online.
Q6	Gathering persons who have the same habit	In this website, we can easily gather some users who have the same habit for some specific purposes such as a potluck party, or group purchase.
Q7	Fans' group	In this website, there are lots of fans' groups to be joined. For example, you can find many friends who love the same singers and exchange some useful information to each other.
Q8	Keep personal connections dark	The website can keep your connection network dark. Some guys who you do not like can not find you through email address, and the MSN friends' list.
Q9	Transaction platform	The website provides a transaction platform for users to trade virtual or physical products.
Q10	Online trade system	The website provides an official online trade system which users can exchange their experience values, virtual objects/treasure in games.
Q11	Free cyber space	Users can have free and unlimited cyberspace such as the capacity of mail box or internet hard disk.
Q12	Friend search	Users can easily add friend list by searching their email list, graduated schools, or MSN contact records, to expand the virtual community.

Thirdly, we take the famous website “Facebook” for example. 140 questionnaires have been issued and we got 102 returned. And merely 67 examples of 102 returned are valid. Then, we implement I-S analysis for these collected

data. Table 7 listed the results of importance-satisfaction analysis. From this Table, 5 quality factors are categorized into excellent. They are “experience value accumulation”, “gathering persons who have the same habit”, “fan’s group”, “transaction platform”, and “online trade system. It means customers think these four quality factors are satisfied and important.

Only one factor “free cyber space” is categorized as “Surplus”. That means customers are satisfied, but consider this service is not very important. The company can relocate the resources from this quality factor to someone else which has been considered important and has low level of satisfaction.

Others 6 quality factors such as “personal design”, “games”, “instant message”, “voice/video message”, “keep personal connections dark”, and “friend search” are considered as “care-free”. That means that customers have a lower satisfaction level and they also think they are not important. Therefore, enterprisers needn’t to care about this kind of quality factors.

Table 7 The results of Importance-Satisfaction Analysis

No.	Quality factors	I-S Analysis	No.	Quality factors	I-S Analysis
1	Personal design	Care-free	7	Fans’ group	Excellent
2	Experience value accumulation	Excellent	8	Keep personal connections dark	Care-free
3	Games	Care-free	9	Transaction platform	Excellent
4	Instant message	Care-free	10	Online trade system	Excellent
5	Voice/video message	Care-free	11	Free cyber space	Surplus
6	Gathering persons who have the same habit	Excellent	12	Friend search	Care-free

Table 8 The results of decision tree algorithm

Accuracy		Extracted factors
Fold 1	64.3%	Q1, Q2, Q3, Q5, Q9
Fold 2	61.5%	Q1, Q2, Q5, Q9
Fold 3	57.1%	Q1, Q2, Q5, Q8, Q9
Fold 4	69.2%	Q1, Q2, Q3, Q5, Q9, Q11
Mean	63.03%	Q1, Q2, Q3, Q5, Q8, Q9, Q11
S. D.	5.07%	Q11

Table 9 The results of decision tree algorithm

Factors	Tree 1	Tree 2	Tree 3	Tree 4	I-S Analysis
Q1	X	X	X	X	<b>Care-free</b>
Q2	X	X	X	X	<b>Excellent</b>
Q3	X			X	<b>Care-free</b>
Q4					Care-free
Q5	X	X	X	X	<b>Care-free</b>
Q6					Excellent
Q7					Excellent
Q8			X		<b>Care-free</b>
Q9	X	X	X	X	<b>Excellent</b>
Q10					Excellent
Q11				X	<b>Surplus</b>
Q12					Care-free

Note: “X” denotes the important attributes in a tree.

Next, we implement decision tree algorithm to extract important factors. We use the importance of 12 quality factors as input variables and the overall satisfaction toward surveyed website is the class label. 4-fold cross validation experiment has been employed. Therefore, 4 decision trees will be constructed. Table 8 provides the results of decision tree, including the classification accuracy and the extracted important attributes. In Table 8, the average classification accuracy is 63.3% and the standard deviation is 5.07%.

From Table 9, we can find different attributes are extracted by different trees. We summarize the results of these four trees and then 7 important quality factors have been extracted. They are “personal design”, “experience value”, “games”, “voice/video message”, “keep personal connections dark”, “transaction platform”, and “free cyber space”. According to the occurrence frequency, the top 4 are “personal design”, “experience value accumulation”, “voice/video message”, and “transaction platform”. Compared with the results of I-S analysis, 2 of 4 extracted factors are categorized as “excellent” and others are “care-free”.

## V. CONCLUSIONS

In this study, we apply feature selection technique and I-S analysis to identify important quality characteristics of virtual community websites. A real case study of “Facebook” has been provided to evaluate the effectiveness of the used methods. Depending on the discovered results, we draw some interesting conclusions. Firstly, I-S analysis is an easy tool which can provides useful information, but it needs both importance and satisfaction points in the same time. Feature selection techniques are powerful methods to extract important attributes based on the assumption of maximizing the overall classification accuracy. In other words, the extracted attributes are crucial for classifying data. Besides, feature selecting techniques cannot tell more useful information. Hence, if it’s possible, we suggest use both methods for getting better understanding of collected data.

Secondly, half of defined quality factors (6 quality characteristics) are categorized as “care-free” in I-S analysis. Some quality factors such as “games” and “friend search” are considered as successful factors of Facebook. But, the results indicated they are carefree by customers. It might be the potential direction of future research to discuss this issue.

## REFERENCES

- [1] C. C. Yang, “The Refined Kano’s Model and its Application,” *Total Quality Management*, vol. 16, Dec, 2005, pp. 1127–1137.
- [2] D. K. Kardaras, B. Karakostas, “Exploring the potential of virtual communities as a business model in banking: the customers’ view,” *International Journal of Web Based Communities*, vol. 3. 2007, pp. 316–331.
- [3] G. Deng, H. Zhou, P. Zhang, “Study of Incentive Mechanism and Internet Behavior of Virtual Communities,” *1st IEEE Symposium on Web Society*, Aug. 2009, pp. 213–218.
- [4] J. Ham, M. Kamber, *Data Mining: Concepts and Techniques*. San Diego: Academic Press, 2001.
- [5] J. R. Quinlan, “Induction of Decision Trees,” *Machine Learning*, vol. 1, Mar. 1986, pp. 81–106.
- [6] J. R. Quinlan, *C4.5: Programs for Machine Learning*. San Francisco: Morgan-Kaufmann, 1993.

- [7] M. Tate, J. Evermann, B. Hope, S. Barnes, "Perceived Service Quality in a University Web Portal: Revising the E-Qual Instrument," *40th Annual Hawaii International Conference on System Sciences*, Jan. 2007, pp. 147b–147b.
- [8] R. W. Swiniarski & A. Skowron, "Rough set methods in feature selection and recognition," *Pattern Recognition Letters*, vol. 24, Mar. 2003, pp. 833–849.
- [9] S. Barnes & R. Videgen, "Data Triangulation in Action: Using Comment Analysis to Refine Web Quality Metrics," *13th European Conference on Information System*, May. 2005, pp. 92–103.
- [10] S. F. Illum, S. H. Ivanov, Y. Liang, Using virtual communities in tourism research, *Tourism Management* (2009), doi:10.1016/j.tourman.2009.03.012
- [11] T. J. Spaulding, How can virtual communities create value for business? *Electron. Comm. Res. Appl.* (2009), doi:10.1016/j.elerap.2009.07.004
- [12] W. G. Kim, C. Lee, S. J. Hiemstra, "Effects of an online virtual community on customer loyalty and travel product purchases," *Tourism Management*, vol. 25, Jun. 2004, pp. 343–355.
- [13] Y. Wang, Q. Yu, D. R. Fesenmaier, "Defining the virtual tourist community: implications for tourism marketing," *Tourism Management*, vol. 23, Aug. 2002, pp. 407–417.
- [14] C.-T. Su and Y.-R. Shiue, "Intelligent scheduling controller for shop floor control systems: a hybrid genetic algorithm/decision tree learning approach," *International Journal of Production Research*, vol. 12, 2003, pp. 2619-2641.