Using Data Mining to Identify Customer Needs in Quality Function Deployment for Software Design

CHIH-HUNG HSU[1], SHIH -YUAN WANG [2], LIANG-TZUNG LIN[3]
[123]Department of Industrial Engineering and Management, Hsiuping Institute of Technology,
11, Gungye Road, Dali, Taichung 41249, Taiwan, R.O.C.

Abstract: - Software design is a high value-added technology, so the satisfaction of customer needs is critical issue for software company. The extraction of knowledge from large database has been successfully applied in a number of advanced fields by data mining. However, little research has been done in the quality function deployment of identifying future customer needs, using data mining. This study applied a time series-based data mining cycle, using sales questionnarie database, to identify future customer needs for software designers. Certain advantages may be observed when future customer needs are identified, using the data mining cycle. The need trend of each customer was found and satisfied with future customer needs in advance. The results of this study can provide an effective procedure of identifying the significant trends to satisfy customer needs for software company and enhance their competitiveness in the software marketplace.

Key-Words: - Data mining; Customer needs; Quality function deployment; Software design

1 Introduction

Software design is a high value-added technology, so the satisfaction of customer needs is critical issue for software designers. The term software includes various systems, such as large-scale on-line systems for the different businesses, basic software operating systems, application systems and others. Software satisfaction has traditionally been defined in terms of fitness for use [1]. A software product is believed to be fit for use if it performs to some level of user satisfaction, in terms of functionality and continuous operation [2]. In order to fulfill high customer satisfaction, the customer needs should be identified. Needs analysis of a software system is always considered as one of very important steps in the software development procedure [3].

On the other hand, Quality Function Deployment (QFD) is a Japanese development and design technology. QFD was first introduced by Akao in 1972 at Mitsubishi's Kobe shipyard site, and then Toyota and its suppliers developed it further for a rust prevention study [4]. After the concept of QFD was introduced in the US through auto manufacturers and parts suppliers [5], many US firms, such as AT&T, Digital Equipment, Ford, GM, Hewlett-Packard, Procter & Gamble, and Raychem, applied QFD to improving communication, product development [6, 7].

QFD has been widely applied to achieve customer needs and improve customer satisfaction in many fields. Some researchers defined QFD as follows: "This technology focuses and coordinates skills within an organization, first to design, then to manufacture and market products that customers want to purchase and will continue to purchase [8]." Some companies have claimed great success with QFD. Proponents assert that QFD has helped them reduce production costs, design time and cost; increase customer satisfaction and product quality [7, 9].

QFD is a cross-functional planning tool which is used to ensure that the voice of the customer is deployed throughout the product planning and design stages. QFD is used to encourage breakthrough thinking of new concepts and technology. Its use facilitates the process of concurrent engineering and encourages teamwork to work towards a common goal of ensuring customer satisfaction. Because the voice of the customer is essential, the House of Quality (HOQ) converts each customer need (CN) into one or more design requirements (DRs) in the first phase of QFD. The main goal of HOQ is to identify customer needs and weights for the product (WHATs) and then to convert these needs into design requirements (HOWs).

It has a great benefit that predicting future customer needs could help companies provide better products, enhance their competitiveness in marketplace, increase customer satisfaction. On the other hand, the application domain of data mining is quite broad and plausible in health insurance [10], surface roughness prediction [11], biomedical technology [12], risk prediction [13], human resource management [14], semiconductor manufacturing [15], production schedule [16], marketing [17] and others. However, little research has also been applied to identify customer needs and weights in QFD using data mining. This study applied a time series-based data mining cycle, using sales questionnarie database, to identify future customer needs for software designers. By applying the proposed approach, future customer needs can be found from a large database to enhance their competitiveness in the software marketplace.

2 Data Mining and Time Series

Berry and Linoff defined data mining as the analysis of huge amounts of data by automatic or semi-automatic means, in order to identify significant patterns or rules [18]. One of the most important data mining techniques is time series analysis. Time series data often arise when monitoring industrial processes or tracking corporate business metrics [19].

Time series analysis can be used to accomplish different goals:

(1).Descriptive analysis determines what trends and patterns a time series has by plotting or using more complex techniques.

(2).Spectral analysis is carried out to describe how variation in a time series may be accounted for by cyclic components. This may also be referred to as "Frequency Domain". With this an estimate of the spectrum over a range of frequencies can be obtained and periodic components in a noisy environment can be separated out [20].

(3).Forecasting can do just that - if a time series has behaved a certain way in the past, the future behavior can be predicted within certain confidence limits by building models.

(4).Intervention analysis can explain if there is a certain event that occurs that changes a time series. This technique is used a lot of the time in planned experimental analysis.

(5).Explanative analysis using one or more variable time series, a mechanism that results in a dependent time series can be estimated [21, 22].

One of the most important forecasting techniques is exponential smoothing analysis for time series analysis. Forecasts generated with this method are a weighted average of the past values of the variable. The weights decline for older observations. The rationale is that more recent observations are more inuential than older observations. The forecast for period t+1 calculated in period t is called Ft+1. Therefore, Ft is the forecast for period t calculated in period t-1. The forecast for period t+1 is,

$$F_{t+1} = \alpha A_t + (1 - \alpha) F_t$$

which represents a weighted average of the actual value (At) and the forecast (Ft) of the actual value (calculated at t-1). The higher the value of alpha the more weight is given to current values [23].

Thus, this study proposed a time series-based data mining cycle, in order to mine the patterns of weights for identifying future customer needs in QFD.

3 The Data Mining Procedure

This study uses data mining cycle to identify future customer needs with each respective step closely involved. The data mining cycle involves a series of activities, from defining the problem to evaluating and applying the results. The previous steps can be served as the baseline reference for the next step, and the steps for identifying customer needs in quality function deployment for software design are described below.

3.1 Defining the problem for data mining

Owing to unknown weights for future customer needs, a large anthropometric database was created for a professional software design company, based on many sales questionnaries, measured in each of 2500 operating needs of customers, according to four period questionnaries; this resulted in a huge amount of data.

The intent of this study was to explore and analyze a huge amount of data, by employing a time series-based data mining cycle in quality function deployment, so as to identify the weights within customer questionnaries in each period. Based on these the weights of customer needs, the future customer needs may be discovered and the results can be encouraged and beneficial for designers and company.

3.2 Data preparation and analysis

The data was processed, and analyzed, in order to enhance the efficiency and ensure the accuracy of the results [24]. Before mining the data, it had to be checked and processed, with all abnormal or missing data being separated out. As a result, of the 10,000 questionnaries, 44, which had missing or abnormal data, were deleted; this left a total of 9956 valid questionnaries regarding the operating needs of customers. There are six customer needs and five deign requirements for each questionnaire, as shown in Table 1 and Table 2. To simplify the space, each customer need is denoted as CN, and the notation of DR is used for each design requirement.

Table 1. Definitions of customer needs

	Customer Needs
CN1	Efficiently uses memory
CN2	Finishes operations reliably
CN3	Finishes operations quickly
CN4	Accurately presents the document
CN5	Small disk space needs
CN6	Promotes creativity

Table 2. Definitions of design requirements

	Deign Requirements
DR1	Difference of Hard Copy Form Presentation
DR2	Recommended Memory
DR3	Number of Creativity Tools From List
DR4	Recommended Disk Space
DR5	FEMA Rating of Design

3.3 Data mining by time series analysis

The weights of four period for each customer need are periodically computed in Table 3. The weight for each customer need is evaluated by a 1–10 scale, where a customer need with a lower value is not more important.

Table 3. Weights of customer needs

	Period 1	Period 2	Period 3	Period 4
CN1	5.1	4.9	6.2	6.9
CN2	6.7	6.9	7.6	8.2
CN3	6.5	7.2	6.5	7.3
CN4	5.4	4.5	2.9	4.8
CN5	3.2	3.4	7.9	8.4
CN6	7.1	6.5	2.6	2.3

On the other hand, it is essential for the company to reflect customer needs by corporate language and then fulfil those design requirements to satisfy customer needs. When customer needs are translated by HOWs, the company has to check the relationship between WHATs and HOWs.

QFD represent the respective strong (with a weight of 9), moderate (with a weight of 3), and weak relationship (with a weight of 1), while the blank is zero. Taking period 1 as an example, the matrix relationship between customer needs and deign requirements is shown in Table 4.

Table 4. The HOQ of period 1

	Period weight	DR1	DR2	DR3	DR4	DR5
CN1	5.1		9		3	1
CN2	6.7	1	1	1		9
CN3	6.5		3	1	1	
CN4	5.4	9	3		1	
CN5	3.2	1			9	3
CN6	7.1		3	9		1
DR important		58.5	109.6	77.1	56	82.1

Through checking the relationship between WHATs and HOWs, the matrix relationship between customer needs and deign requirements were determined. Subsequently, data mining was undertaken, using a time series-based data mining cycle, to mine the weights and determine the trend of each customer need for the next period.

According to the data mining cycle, the weights of customer needs in the next period (period 5) would be estimated as shown in Table 5. As shown, the weight of CN1 in the period 5 is 6.6; thus, these predicted weights for the customer needs were chosen for the next stage of processing. Furthermore, the forecast errors in all of the customer needs are less than the control limits of forecast. Thus, the exponential smoothing analysis is clearly quite accurate.

Table 5. Predicted weights for the customer needs

	Period 1	Period 2	Period 3	Period 4	Predicted
CN1	5.1	4.9	6.2	6.9	6.6
CN2	6.7	6.9	7.6	8.2	7.6
CN3	6.5	7.2	6.5	7.3	7.1
CN4	5.4	4.5	2.9	4.8	4.4
CN5	3.2	3.4	7.9	8.4	7.8
CN6	7.1	6.5	2.6	2.3	2.8

3.4 Evaluation and Application of Results

To gain a better insight into the predicted weights among the six customer needs resulting from the time series-based data mining cycle, a line plot was drawn of the weights of the customer needs. As can be seen in Figure 1 and Figure 2, the weights bear marked differences in the customer needs for each period. The trend for each customer need can be understood and controlled by software designers with the information in Figure 1 and Figure 2. The importance of each customer need can be considered to know customer future needs. The software can be designed and planned to satisfy with customer future needs in advance.

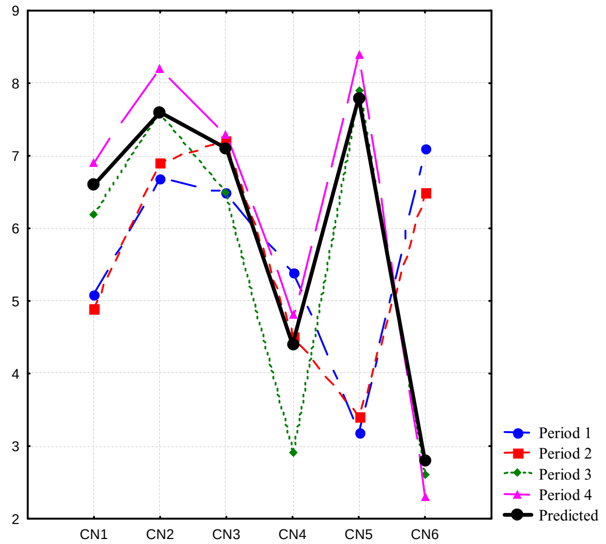

Fig. 1. The weight plot of customer needs for each period

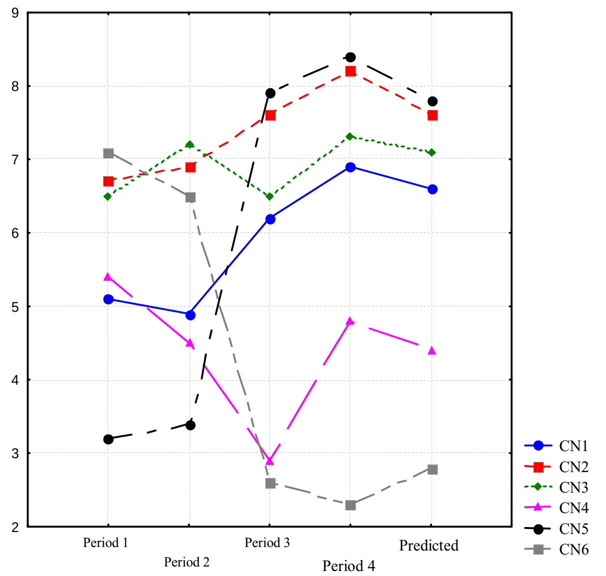

Fig. 2. The weight plot of periods for each customer need

In addition, the future trend of each design requirement to satisfy customer future needs can be analysed in Table 6. According to the future trend of each design requirement, DR4 and DR5 should be closely noticed since its importance has increased and could become the most important design requirement to satisfy customer needs in the future. On the other hand, DR3 has a similar importance in earlier periods, but the importance of DR3 has declined in the future, while the future trend of DR1 and DR2 has been quite steady. Different design requirement should be considered differently for software designers with the much more analysed information.

Table 6. Future trend of five deign needs

	DR1	DR2	DR3	DR4	DR5
Period 1	58.5	109.6	77.1	56	82.1
Period 2	50.8	105.6	72.6	57	83.7
Period 3	41.6	99.4	37.5	99.1	100.9
Period 4	59.8	113.5	36.2	108.4	108.2
Predicted	55	109.9	39.9	101.5	101.2

4 Conclusion

This study uses data minig cycle in QFD to analyse customer future needs. In addition, the time series-based data mining cycle can be applied to predict the weights and determine the trend of each customer need for the next period.

This is a powerful approach for evaluating the weights of each customer need, developing the software that fits customer needs and facilitate the software company competitiveness in the software marketplace.

The data mining cycle emphasizes the dataset information by repeating interaction activities. Since customer needs can change rapidly, the database of customer needs must be updated continually; therefore, the time series-based data mining cycle, proposed in this study, will continually update the database and continually identify the future customer needs for software designers. These revised deign requirements will exactly satisfy with customer needs, allowing software designers access to the latest customer needs, thus facilitating advanced software design.

Acknowledgments

The authors thank peer review of 3 independent qualified referees for their comments.

References:

[1] Dunn RH, Software quality assurance: A management perspective. *Quality Progress*,

21(7), 1988, pp.52-56.

[2] O'Brien DH, Software quality starts with the customer. *Quality Progress*, 30(6), 1991, pp.22-24.

[3] Liu XF, A quantitative approach for assessing the priorities of software quality requirements. *The Journal of Systems and Software*, 42, 1998, pp.105-113.

[4] Wasserman GS, On how to prioritise design requirements during the QFD planning process. *IIE Transaction*, 25(3), 1993, pp.59-65.

[5] Sullivan LP, Quality function deployment (QFD): the beginning, the end, and the problem in between. A collection of presentations and QFD case studies. American Supplier Institute, 1987.

[6] Ansari A and Modaress B, Quality function deployment: the role of suppliers. *International Journal of Purchase Management*, 30(4), 1994, pp.28-35.

[7] Griffin A, Evaluating QFD's use in US firms as a process for developing products. *Journal of Product Innovation Management*, 9, 1992, pp.171-187.

[8] Hauser JR and Clausing D, The house of quality. *Harvard Business Review*, 66(3), 1988, pp.63-73

[9] Wheelwright SC and Clark KB, Revolutionizing product development: Quantum leaps in speed, efficiency, and quality. New York: Free Press, 1992.

[10] Chas YM, Ho SH, Cho KW, Lee DH and Ji SH, Data mining approach to policy analysis in a health insurance domain, *International Journal of Medical Informatics*, 62, 2001, pp.103-111.

[11] Feng CX and Wang X, Development of empirical models for surface roughness prediction in finish turning, *International Journal of Advanced Manufacturing Technology*, 20, 2002, pp.348-356.

[12] Maddour M and Elloumi M, A data mining approach based on machine learning techniques to classify biological sequences, *Knowledge-Based Systems*, 15, 2002, pp.217-223.

[13] Becerra-Fernandez, I., Zanakis SH and alczak S, Knowledge discovery techniques for predicting country investment risk. *Computer and Industrial Engineering*, 43, 2002, pp.787-800.

[14] Min H and Emam A, Developing the profiles of truck drivers for their successful recruiDRent and retention: a data mining approach, *International Journal of Physical Distribution & Logistics Management*, 33, 2003, pp.149-162.

[15] Chien CF, Hsiao A and Wang I, Constructing semiconductor manufacturing performance indexes and applying data mining for manufacturing data analysis, *Journal of the Chinese Institute of Industrial Engineers,* 21, 2004, pp.313-327.

[16] Sha DY and Liu CH, Using data mining for due date assignment in a dynamic job shop environment, *International Journal of Advanced Manufacturing Technology*, 25, 2005, pp.1164-1174.

[17] Wong KW, Zhou S, Yang Q and Yeung MS, Mining customer value: from association rules to direct marketing, *Data Mining and Knowledge Discovery*, 11, 2005, pp.57-79.

[18] Berry M and Linoff G, Data Mining Techniques: for Marketing, Sales, and Customer Support, Wiley, New York, 1997.

[19] Giudici P, Applied Data Mining: Statistical Methods for Business and Industry, Wiley, England, 2003.

[20] Warner RM, Spectral analysis of time - series data. Guilford Press, New York, NY, USA, 1998.

[21] Chatfield C, The analysis of time series – an introduction. 5th. Chapman and Hall, London, UK, 1996.

[22] Box GEP, Jenkins G M, and Reinsel G C, Time series analysis – Forecasting and control. 3rd ed. Prentice Hall, Englewood Cliffs, NJ, USA, 1994.

[23] Harvey HC, Time series models. Halstead Press, New York, NY, USA, 1981.

[24] Pyle D, Data Preparation for Data Mining, Morgan Kaufmann, California, 1999.