MONITORING CUSTOMER SATISFACTION IN SERVICE INDUSTRY: A CLUSTER ANALYSIS APPROACH

MATUS HORVATH, ALEXANDRA MICHALKOVA

1 INTRODUCTION

The customer satisfaction can be defined as a customer's perception of the degree to which customers requirements have been fulfilled (ISO 9000:2005). This degree is one of the key performance indicators of organization and their quality system. Therefore, ISO 9001 standard include a requirement to regular customer satisfaction measuring. The common way of customers satisfaction measuring is survey. Unfortunately, this method of customer satisfaction measuring in practice deals with low response rate. This wants the organizations overcome with personal or phone gathering of the answers from customers. This way of answers gathering can be expensive when the survey sample is too big. This leads organizations to search for better methods of customer satisfaction analysis.

Special importance has customer satisfaction in service industries as the outcome measure of service quality (M. K. Brady et all 2002). Very common situation in service industries is widespread offer. In these cases doesn't exist an average customer, but customer groups with different use of service offered by the organization. The mentioned lacks of classic approach to customer satisfaction analysis can in such cases cause wrong conclusions. This paper deals with a customer satisfaction analysis based on analysis of information from customer and sale databases. These databases contain information about customer (e.g. sex, age) and information about the way how the customers use the services of the organizations (e.g. amount, type, number of complaints). The informatization of databases in organizations enabled quickly interconnecting of this information.

2 METHODOLOGY

The methodology of the approach can be split to four steps (Figure 1). First step is the information gathering from different databases. In this step, it is very important to choose right variables for creating data set for the next step. The data set must contain variables that provide information about customers (if possible) and their way of service use. For chosen variables are used historical data from selected time period. Authors recommend selection in approximately one month if possible. Use of historical data is necessary for the second part of

data set building. The key variable that is used for indicating of customer satisfaction is the actual service use status of each customer.

Second step deals with dividing the data set to the customer groups. For the effective dividing of the data set to the customer groups, we used cluster analysis. Cluster analysis refers to the grouping of records, observations, or cases into classes of similar objects – clusters. This implies that cluster is a collection of records that are similar to one another and dissimilar to records in other clusters (Larose, 2003). There exist several methods for cluster analysis. In this research, we used the hierarchical cluster analysis. Hierarchical cluster analysis is a nested sequence of clusters ranging from one to *N* clusters for a data set of size *N*. It groups data points into a tree of clusters, called a dendrogram, which shows how the clusters are related. The root of the tree represents one cluster, containing all data points, while, at the leaves of the tree, there are *N* clusters each containing one data point. There are two hierarchical approaches to how to create the dendrogram. The hierarchical divisive approach and more used hierarchical agglomerative approach. The hierarchical divisive approach start with one cluster that cover entire data set and progressively dividing single clusters into subclusters until each one contains only one data point. The hierarchical agglomerative approach operates in a bottom-up manner by performing a series of agglomerations in which small clusters, initially containing an individual data point, are merged together to form larger clusters. At each step of the agglomeration process, the two closest clusters are fused together. There are many variants to define the distance between two clusters. Single-link computation uses the shortest distance between two data points, each coming from one of two clusters. Centroid-based distance calculation uses the dissimilarity between the centres of two clusters. Ward distance calculation uses variance analysis, and the group-average method employs the average distance between all pairs of data points in two clusters (Pham & Afify, 2007). In the practical application of hierarchical cluster analysis, it is very common problem the selecting of the appropriate level of the dendrogram for the next part of the analysis. As the appropriate level, we selected a level after that is the regrouping rate very low.

Database	Cluster analysis	Exploratory data analysis	Results of the analysis
Include various data about customer from sale and quality departments.	Divide the customers to customers groups (clusters) according to they profile and way how the customers use the services.	Create profiles of the customer groups and determine the customer groups satisfaction level.	Corrective actions for increasing the customer's groups satisfaction level created for elimination of the lacks of services quality.

Figure 1 – Steps of the approach based on cluster analysis

The clusters from the selected level are inputs for the next part of the analysis. In exploratory data analysis is every cluster analysed for determination of the average customer group profile and the level of customer group satisfaction. For the customer group profile determination we recommend information as average age, percentage of genders and information about style of using the services of the organisation. In our example thereinafter (telecommunication organisation) it was the percentage of activated service package, number of minutes. For the measuring of the level of customers group satisfaction can be very good used information about the leaving of the customers. The results of the analysis can be use not just as precise measure of the customers satisfaction (different for each customers group) but also for generating corrective actions with the aim to increase the customer satisfaction level with change of the offered services.

3 EXAMPLE OF APPLICATION

Example was created as a part of bachelor thesis at the Technical University of Kosice. For example, of the mentioned approach to customers satisfaction analysis we used data set provide by UCI Machine Learning Repository. The data set contains real data from 5,000 customers of one telecommunication organisation in a one month. The data set includes 21 variables. After an analysis, we have decided to use for the analysis 14 variables described in Tab 1.

Table 1 – Description of used variables

Variable name	Variable description	Type of variable
ACCOUNT LENGHT	Length of service use.	Numerical - integer
INT´L PLAN	Indicate use or don't use the international plan.	Boolean– True/False
V´MAIL PLAN	Indicate use of voice mail plan.	Boolean– True/False
V´MAIL MESSAGES	Number of voice mail messages.	Numerical - integer
DAY MINS	Number of minutes used during a day.	Numerical - real
DAY CALLS	Number of calls during a day.	Numerical - integer
EVE MINS	Number of minutes used during an evening.	Numerical - real
EVE CALLS	Number of calls during an evening.	Numerical - integer
NIGHT MINS	Number of minutes used during a night.	Numerical - real
NIGHT CALLS	Number of calls during a night.	Numerical - integer
INT´L MINS	Number of minutes used for international calls.	Numerical - real
INT´L CALLS	Number of calls used for international calls.	Numerical - integer
CUSTSERV CALLS	Number of calls on customer service	Numerical - integer
CHURN	Indicate the loss of customer.	Boolean– True/False

For the cluster analysis we have used software R version 2.12.1 and function library Rattle version 2.6.8. As the method for hierarchical cluster analysis we use euclidean metric and ward agglomerative method. For cluster analysis we use all variables from data set beside churn rate. The result dendrogram is on Figure 2. For the next steps we used the level with 5 clusters. On Figure 2 are cluster numbered ascending according to customer leaving rate (variable churn).

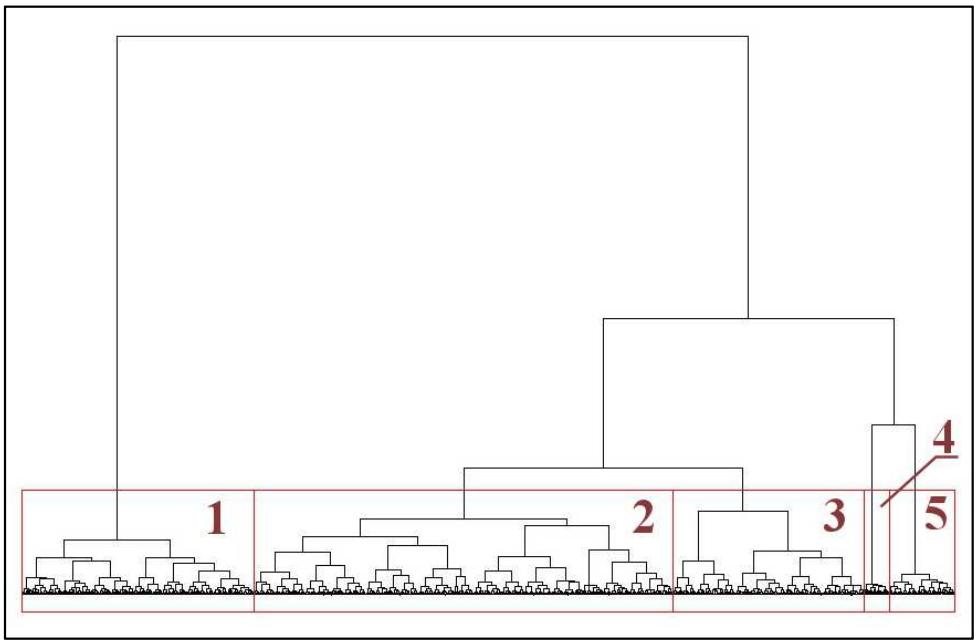

Figure 2 – Created dendrogram with 5 clusters

The results of the date exploratory analysis of 5 clusters are that we don't detect important differences in the amount of used minutes. The difference in the use of services is in the ratio of activated services and length of service use. In the Table 2, it is a summary of the findings. Based on the result of the cluster analysis the customer group with the most satisfied customer is the group one. In this group use 100 % of the customers voice mail service and nobody use international plan. The voice mail service is important as we can see (Tab. 2). The customer group 5 has the highest rate of customer leaving and the customer group 5 and 4 have just one difference in their customer profiles. The customer group 4 has 100% rate of activated voice mail service and the customer leaving rate is about 9 % lower as in customer group 5. That leads to the idea use this service as a marketing tool for increasing the customer satisfaction in all customer groups.

In the customer groups with the highest number of customers – 2 and 3 is the customer satisfaction rate at average 13 %. Lower customer satisfaction ratio achieved customer group 3, where are at average customers with the longest service use time and with the highest number of customer service calls. Because in this customer group we have rounded 31 % of all customers, it is necessary to

more deeply analyze the reasons. Service that is interconnecting two customer groups with highest customers leaving rate is the international plan. It is necessary to analyze the reasons of that high unsatisfied of the customers that use the service.

Table 2 – Description of customers groups

Customer group	Description	Customer leaving rate
1	• Customer with the second lowest average time of service use. • 100 % of activated voice mail plan service • 1f customers have activated international plan • 23.84% of all customers in the analysis.	4.70 %
2	• The lowest average time of service use in the analysis. • Neither of customers has activated voice mail plan or international plan services. • 35.38% of all customers in the analysis	12.61%
3	• The highest average time of service use. • Neither of customers has voice mail plan or international plan services. • The highest average number of call on customer service. • 31.32 % of all customers in the analysis	14.56%
4	• Customers with the second longest average time of service use • 100 % of customer have activated international plan and voice mail plan • 2.62% of all customers in the analysis	35.38%
5	• The middle average time of service use. • 100% customer has activated international plan • 6.84% of all customers in the analysis	44.73%

4 CONCLUSION

The aim of this paper was to introduce a specific approach to customer satisfaction analysis and the use of cluster analysis as a tool for dividing large heterogenic datasets to smaller parts according to their structure. Described technique can improve the customer satisfactory measuring process in organisations. This is very important in the service industry in which customer satisfactory measure is the only outcome measure of service quality. As can be shown in the result of the example of application, described technique is not a stand-alone approach and must be supported with other analysis with the aim to reveal the reason of low customer satisfaction.

AKNOWLEDGEMENT

This work is supported by the Scientific Grant Agency of the Ministry of Education, Science, Research and Sport of the Slovak Republic, project KEGA 009-4/2011: Creative Laboratory Education at Technical Faculties "CRELABTE".

REFERENCES

Blake C. L., Merz C. J. (1998), *Churn Data Set*, UCI Repository of Machine Learning Databases, University of California, Department of Information and Computer Science, Irvine, CA

International Organization for Standardization, (2005). *ISO 9000:2005 Quality management systems -- Fundamentals and vocabulary*, Geneva: ISO.

Kriška, J. (2008). *Hierarchická a nehierarchická zhluková analýza*. [online] Available: http://www2.fiit.stuba.sk/~kapustik/ZS/Clanky0607/kriska/index.html, [Accessed 15.9.2012].

Larose, D. T. (2003), *Discovering knowledge in data: an introduction to data mining*, John Wiley & Sons, Inc., Hoboken, New Jersey USA, 223 p., ISBN: 0-471-66657-2.

Pham D. T., Afify A. A. (2007), „Clustering techniques and their applications in engineering" in *Proc. IMechE Vol. 221 Part C: J. Mechanical Engineering Science,* Vol. 221, no. 11, pp. 1445- 1459.

ABOUT AUTHORS

Matus Horvath, Ing. – PhD. student, Technical University of Košice, e-mail: matus.horvath@tuke.sk, Letna 9, 042 00 Košice, Slovak Republic,
(Advisor: Edita Virčíková, Prof., PhD., Technical University of Košice, e-mail: edita.vircikova@tuke.sk, Letna 9, 042 00 Košice, Slovak Republic)

Alexandra Michalkova, Bc. – Msc. student, Technical University of Košice, Letna 9, 042 00 Košice, Slovak Republic