Customer Retention using Data Mining Techniques

Dr.R.Dhanapal
Easwari Engineering College
Affiliated to Anna University of Technology, Chennai-89.

S.Gayathri Subramanian
R.B.Gothi Jain College for Women
Affiliated to University of Madras, Chennai-52.

Jobin M Scaria
King Saud University
Kingdom of Saudi Arabia.

ABSTRACT
Customer retention represents a modern approach for quality in enterprises and organizations and serves the development of a truly customer-focused management and culture. Customer retention measures offer a meaningful and objective feedback about client's preferences and expectations. This paper presents an original methodological approach of customer satisfaction and retention evaluation, combining multicriteria preference desegregations analysis and rule induction data mining. Furthermore, it is examined whether the implementation of the two methodologies may offer a solution to the problem of missing data, in the initial data set.

Keywords
Rule-Induction Data Mining, Customer Satisfaction Measurement, Multicriteria Analysis.

1. INTRODUCTION
Customer Satisfaction and retention research is one of the fastest growing segments of the marketing field. Marketing and management sciences, nowadays, are focusing on the coordination of all the organization's activities in order to provide goods or services that can satisfy best specific needs of existing or potential customers.

To reinforce customer orientation on a day-to-day basis, a growing number of companies choose customer satisfaction as their main performance indicator. However, it is almost impossible to keep an entire company permanently motivated by a notion as abstract and intangible as customer satisfaction. Therefore, customer satisfaction must be translated into a number of measurable parameters directly linked to people's job-in other words factors that people can understand and influence [3].

The aim of this paper is to present an original methodological approach to the problem of customer satisfaction and customer retention evaluation, combining multicriteria preference disaggregation analysis and rule induction data mining. The two methodologies were applied to the results of a customer satisfaction survey. The main objectives of the paper are:

- to compare the results of the two methods
- to evaluate the homogeneity of the set of customers
- to overcome the problem of no response (missing data) in the data set.

The paper is organized into 3 sections. Section 2 presents briefly the basic principles of the two methods used: multicriteria preference disaggregation approach and rule induction approach, as well as the integration of the two approaches and the implemented methodological frame. Section 3 includes some conclusive remarks on the methodology proposed, as well as subjects for further research.

2. METHODOLOGICAL FRAME
2.1 MUSA (Multicriteria Satisfaction Analysis)
The MUSA (Multicriteria Satisfaction Analysis) is based on a preference disaggregation model. The aggregation of individual preferences into a collective value function is the main objective of this approach. More specifically, it is assumed that the customers' global satisfaction can be explained by a set of criteria or variables representing its characteristic dimensions (Figure 1).

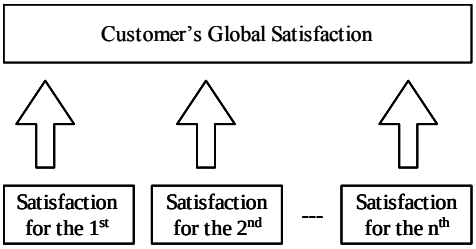

Figure1 : Aggregation of Customer's Judgement

The preference disaggregation methodology is an ordinal regression based approach [5,8] in the field of multicriteria analysis. It is used for the assessment of a set of marginal satisfaction functions in such a way that the global satisfaction criterion becomes as consistent as possible to customers' judgments.

According to the model, each customer is asked to express his/her judgements, namely his/her global satisfaction and his/her satisfaction with regard to the set of discrete criteria. The collected data is analyzed with the preference disaggregation model, respecting the ordinal and qualitative form of customers' judgements and preferences.

The main results of the method are [5,8,10] :

- global and partial satisfaction functions
- weights on the criteria (relative importance),
- average satisfaction indexes.

2.2 Rule Based Data Mining Techniques
The objective of data mining is to extract valuable information from one's data, to discover the 'hidden gold'. In Decision Support Management terminology, data mining can be defined

as 'a decision support process in which one search for patterns of information in data' [2].

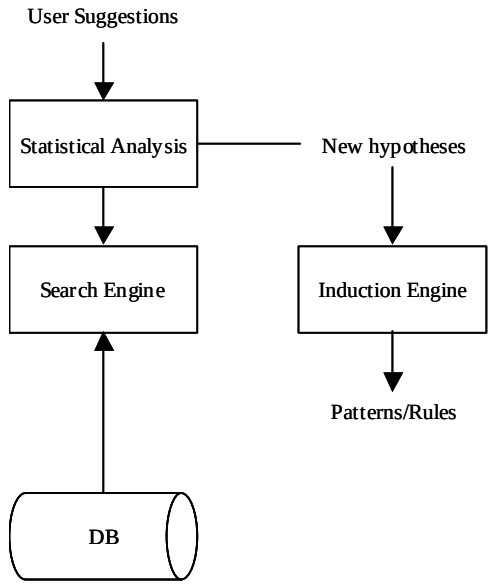

Figure 2: Rule Induction Process

Data mining techniques are based on data retention and data distillation. Rule induction models (Figure 2) belong to the logical, pattern distillation based approaches of data mining. These technologies extract patterns from data set and use them for various purposes, such as prediction of the value of a dependent field (Field to Predict). By automatically exploring the data set, the induction system forms hypotheses that lead to patterns. These patterns may be logic, equations or cross-tabulations. Logic can deal with both numeric and non-numeric data.

The central operator in a logical language is usually a variation on the 'if-then' statement. By supervised learning paradigm derive rules, of 'if-then' type, from data. Such rules relate an outcome of interest to a number of attributes. They are of the following form [1]:

> *if attribute1 = a and attribute2 = b then outcome = c (probability = .9)*

The rule's probability is the probability that for a random record satisfying the rule's condition(s), the rule's conclusion is also fulfilled [7].

Rules may easily go beyond attribute- value representations. They may have statements such as 'shipping state = receiving state'. Here, in attribute logic, we compare the values of the two fields, without naming any values. By expressing attribute-based patterns, rules have the advantage of being able to deal with numeric and non-numeric data (categorical fields).

2.3 Integrating Multicriteria and Rule-Induction Approach

The methodology, presented in this paper, combines the preference disaggregation model with the rule-induction process. The main stages of the methodology are described below (Figure 3):

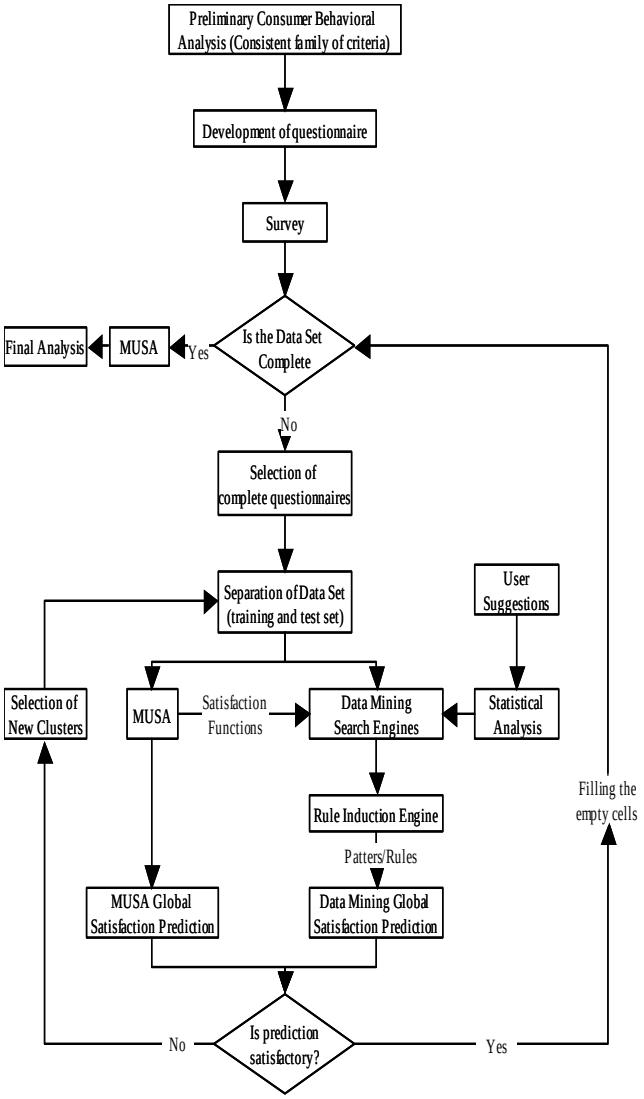

Figure 3: Customer Satisfaction Survey Process

- Preliminary analysis: customer satisfaction research objectives should be specified in this stage, in order to assess satisfaction dimensions (customers' consistent family of criteria).
- Questionnaire design and conducting survey: using results from the previous step, this stage refers to the development of the questionnaire, the determination of survey parameters and the survey conduction.
- Analysis: the two different approaches come to prediction. In case the prediction is not considered satisfactory, a new selection of clusters is made and the process of analysis restarts. In the opposite case (of satisfactory prediction), the predicted value is used to fill the empty cells in the data table. The empty cells correspond to cases of no response. The deriving filled data set is used by the preference desegregations method in order to perform final analysis.

3. CONCLUSIONS

The original methodology presented in this paper combines the preference disaggregation methodology with rule-induction data mining. The methodology is proposed as a potential solution to the problem of no response in the data set that may be due to insufficiently completed questionnaires.

The MUSA method evaluates the satisfaction added value curve with respect to customers' judgements. This curve normalized in [0, 100] shows the value received by customers for each level of the ordinal qualitative satisfaction scale.

The methodology has been applied to a pilot customer satisfaction survey for the Greek shipping sector. The main data set consists of 523 customers (test set: 100, training set: 423) and 5 criteria. Prediction level is quite satisfactory resulting that data mining techniques can be successfully combined with multiple criteria methods.

Using other customer characteristics, such as age, marital status, etc., the presented methodology may identify and analyze special group of customers. Moreover, the integration of ordinal data, instead of the satisfaction value estimations resulted from the preference-disaggregation model, may give better prediction in the rule-induction process.

4. REFERENCES

[1] Akeel Al-Attar, 1998, 'Data Mining– Beyond Algorithms', http://www.attar.com/tutor/ mining.htm.

[2] Berry, J. A. Michael; Linoff, Gordon, 1997, 'Data Mining Techniques: For Marketing, Sales, and Customer Support', John Wiley & Sons, Inc., Canada.

[3] Deschamps J.P. and P. Ranganath Nayak, 1995, 'Product Juggfernauts: How companies mobilize to generate a stream of market winners', Harvard Business School Press.

[4] Dr. Dhanapal R., Gayathri Subramanian, RajaGopal, HemaMalini, ,2010. Security Information Hiding in Data Mining on the basis of Privacy Preserving Technique, Volume 2. Issue 10, Journal of Computing.

[5] Grigoroudis E.; Siskos Y.; Saurais O., 1998, 'TELOS: A Customer Satisfaction Evaluation Software', Computers and Operations Research, (to appear).

[6] Jaquet-Lagrèze, Eric; Siskos, Jean, 1982, 'Assessing a set of additive utility functions for multi-criteria decision-making: The UTA method', European Journal of Operational Research, 10, pp.151-164.

[7] Meidan A., 1998, 'A data mining application for issuing predictions, summarizing the data and revealing interesting phenomena', http://www.wizsoft.com/why.html.

[8] Mihelis G.; Grigoroudis E.; and Siskos Y., 1998, 'Customer Satisfaction Measurement in the private Bank sector', European journal of Operational Reseaech, (to appear).

[9] Siskos Y and Yannacopoulos D., 1985, 'UTASTAR: an ordinal regression method building additive value functions', Investigaçao Operational, 5 (1), pp. 39-53.

[10] Siskos Y.; Grigoroudis E.; Zopounidis C.; Saurais O., 1998, 'Measuring Customer Satisfaction Using a Collective Preference Disaggregation Model', Journal of Global Optimization, 12, pp.175-195.