Data Mining Oriented CRM Systems Based on MUSASHI: C-MUSASHI*

Katsutoshi Yada[1], Yukinobu Hamuro[2], Naoki Katoh[3], Takashi Washio[4], Issey Fusamoto[1], Daisuke Fujishima[1], and Takaya Ikeda[1]

[1] Faculty of Commerce, Kansai University,
3-3-35, Yamate, Suita, Osaka 564-8680, Japan
{yada, da00587, da00591, da00039}@kansai-u.ac.jp
http://www2.ipcku.kansai-u.ac.jp/~yada/
[2] Department of Business administration, Osaka Sangyo University,
3-1-1 Nakagaito, Daito, Osaka, 567-8530 Japan
hamuro@adm.osaka-sandai.ac.jp
[3] Graduate School of Engineering, Kyoto University,
Kyoto University Katsura, Nishikyo-ku, Kyoto, 615-8540 Japan
naoki@archi.kyoto-u.ac.jp
[4] Institute for the Scientific and Industrial Research, Osaka University
8-1 Mihogaoka, Ibaraki, Osaka, 567-0047 Japan
washio@sanken.osaka-u.ac.jp

Abstract. MUSASHI is a set of commands which enables us to efficiently execute various types of data manipulations in a flexible manner, mainly aiming at data processing of huge amount of data required for data mining. Data format which MUSASHI can deal with is either an XML table written in XML or plain text file with table structure. In this paper we shall present a business application system of MUSASHI, called C-MUSASHI, dedicated to CRM oriented systems. Integrating a large amount of customer purchase histories in XML databases with the marketing tools and data mining technology based on MUSASHI, C-MUSASHI offers various basic tools for customer analysis and store management based on which data mining oriented CRM systems can be developed at extremely low cost. We apply C-MUSASHI to supermarkets and drugstores in Japan to discover useful knowledge for their marketing strategy and present possibility to construct useful CRM systems at extremely low cost by introducing MUSASHI.

1 Introduction

MUSASHI is an open-source software that provides a set of commands with which various types of data manipulations for a large amount of data mainly required for data mining in business field can be executed in an flexible manner

* Research of this paper is partly supported by the Grant-in-Aid for Scientific Research on Priority Areas (2) by the Ministry of Education, Science, Sports and Culture of Japan and the Kansai University Special Research fund, 2002.

S. Tsumoto et al. (Eds.): AM 2003, LNAI 3430, pp. 152–173, 2005.
© Springer-Verlag Berlin Heidelberg 2005

[11][3][4][5]. We have developed a data mining oriented CRM system that runs on MUSASHI by integrating several marketing tools and data mining technology. Discussing the cases regarding simple customer management in Japanese supermarkets and drugstores, we shall describe general outlines, components, and analytical tools for CRM system which we have developed.

With the progress of deflation in recent Japanese economy, retailers in Japan are now under competitive environment and severe pressure. Many of these enterprises are now trying to encompass and maintain their loyal customers through the introduction of FSP (Frequent Shoppers Program) [6][14]. FSP is defined as one of the CRM systems to accomplish effective sales promotion by accumulating purchase history of the customers with membership cards in its own database and by recognizing the nature and the behavior of the loyal customers. However, it is very rare that CRM system such as FSP has actually contributed to successful business activities of the enterprises in recent years.

There are several reasons why the existing CRM system cannot contribute to the acquisition of customers and to the attainment of competitive advantage in the business. First of all, the cost to construct CRM system is very high. In fact, some of the enterprises have actually spent a large amount of money merely for the construction of data warehouse to accumulate purchase history data of the customers and, as a result, no budget is left for carrying out customer analysis. Secondly, it happens very often that although data are actually accumulated, techniques, software and human resources in their firms to analyze these data are in shortage, and thus the analysis of the customers is not in progress. Therefore, in many cases, enterprises are simply accumulat-ing the data but do not carry out the analysis of the customers.

In this paper, we shall introduce a CRM system, named C-MUSASHI, which can be constructed at very low cost by the use of the open-source software MUSASHI, and thus can be adopted freely even by a small enterprise. C-MUSASHI consists of three components, basic tools for customer analysis, store management systems and data mining oriented CRM systems. These components have been developed through joint research activities with various types of enterprises. With C-MUSASHI, it is possible to carry out the analysis of the customers without investing a large amount of budget for building up a new analytical system. In this paper we will explain the components of C-MUSASHI and cases where C-MUSASHI is applied to a large amount of customer history data of supermarkets and drugstores in Japan to discover useful knowledge for marketing strategy.

2 C-MUSASHI in Retailers

2.1 MUSASHI and Operation Systems

MUSASHI, the Mining Utilities and System Architecture for Scalable processing of Historical data, is a data mining platform [3][5] that efficiently and flexibly processes large-scale data that has been described in XML data. One of its re-

markable advantages lies in the powerful and flexible ability to preprocess various amounts of raw data in the knowledge discovery process. The development of MUSASHI has been progressed as an open source software, and thus everybody can download it freely from [11].

MUSASHI has a set of small data processing commands designed for retrieving and processing large datasets efficiently for various purposes such as data extraction, cleaning, reporting and data mining. Such data processing can be executed simply by running MUSASHI commands as a shell script. These commands also includes various data mining commands such as sequential mining, association rule, decision tree, graph mining and clustering commands.

MUSASHI uses XML as a data structure to integrate multiple databases, by which various types of data can be represented. MUSASHI makes it feasible to carry out the flexible and low-cost integration of the structured and vast business data in companies.

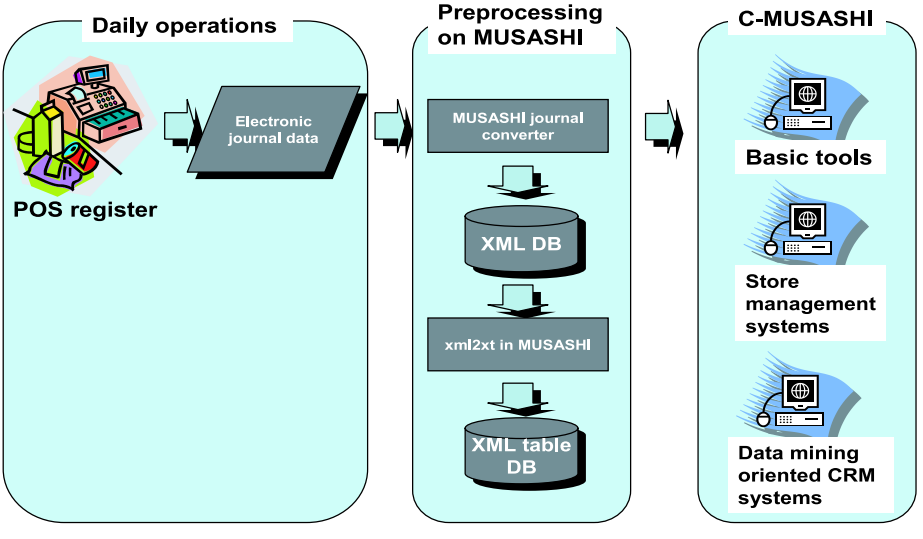

Fig. 1. C-MUSASHI system in the retailers

C-MUSASHI is a CRM system that runs on MUSASHI, by which it is possible to process the purchase history of a large number of customers and to analyze consumer behavior in detail for an efficient and effective customer control. Figure 1 shows the positioning of C-MUSASHI in a system for daily operation of the retailers. By using C-MUSASHI, everybody can build up a CRM system through the processes given below without introducing commercial data warehouse. POS (point of sales) registers used in recent years output the data called electronic journal, in which all operation logs are recorded. Store controller collects the electronic journals from the registers in the stores and accumulates them. The electronic journal data is then converted by "MUSASHI journal converter" to

XML data with minimal data loss. Figure 2 shows a sample of XML data that is converted from electronic journal data output. It is clear that operations in detail on POS cash registers are accumulated as XML data structure.

```
<?xml version="1.00" encoding="euc-jp"?>
<date="20011206" time=101545>
 <receipt="198765">
  <items>
   <JAN>4901984625422</JAN>
   <name>bread</name>
   <vol>1</vol>
   <unit>109 </unit>
  </items>
  <card>
   <customer>2101205787635 </customer>
   <name>Kandai Taro</name>
   <address>Osaka Susita 3-3-35</address>
  </customer>
  <items>
   <JAN>3053289502011</JAN>
   <name>milk</name>
   <vol>2</vol>
   <unit>149 </unit>
  </items>
 <accounts>
   <total>407</total>
   <deposit>1000</deposit>
   <balance>593</balance>
</receipt>
```

Fig. 2. XML data that is converted from electronic journal data

However, if all operation logs at POS registers are accumulated on XML data, the amount of data may become enormous which in turn leads to the decrease of the processing speed. In this respect, we define a table-type data structure called XML table (see Figure 3). It is easy for users to transform XML data into XML table data by using "xml2xt" command in MUSASHI. The XML table is an XML document such that the root element named <xmltbl> has two elements, <header> and <body>. The table data is described in the body element using a very simple text format with one record on each line, and each field within that record separated by a comma. Names and positional information relating to each of the fields are described in the <field> element. Therefore it is possible to access data via field names. The data title and comments are displayed in their respective <title> and <comment> fields. A system is built up by combining XML data such as operation logs with XML data and XML table data. Thus, by properly using pure XML and XML table depending on the purposes, it is to construct an efficient system with high degree of freedom.

```
<?xml version="1.0" encoding="euc-jp"?>
<xmltbl version="1.1">
<header>
<title>Sales Transactions</title>
<field no="1" name="customer"/>
<field no="2" name="date"/>
<field no="3" name="JAN"/>
<field no="4" name="vol"/>
</header>
<body><![CDATA[
2101205787635 20011206 4901984625422 1
2101205787635 20011206 3053289502011 2
]]></body>
</xmltbl>
```

Fig. 3. Sales data which has been converted into an XML table

2.2 Components of C-MUSASHI

C-MUSASHI is a business application for customer relationship management and is composed of three components, basic tools for customer analysis, store management systems and data mining oriented CRM systems. The first provides basic information necessary for the implementation of store management and customer management.

The second is a store management system based on the methodology "Store Portfolio Management (SPM)" which is newly proposed in this paper aiming to support the headquarter of chain store for controlling many stores. By using this system, it is possible to implement the appropriate strategy depending on the circumstance of each store. The third is a data mining oriented CRM system which discovers useful knowledge from customer purchase history data for the effective customer relationship management by using data mining techniques. It includes many data mining algorithms we developed to extract meaningful rules from a large amount of customer purchase history data. No other existing CRM system has the same functions and advanced data mining algorithms for customer relationship management. We will introduce these components in the subsequent sections.

3 Basic Tools for Customer Analysis

In this section, we shall introduce the tools for basic customer analysis in C-MUSASHI. Such tools are usually incorporated in existing CRM systems. Here we will explain some of such tools: decile analysis, RFM analysis, customer attrition analysis, and LTV (life time value) measurement. C-MUSASHI also has many other tools for customer analysis. We will present here only a part of them here.

3.1 Decile Analysis

In decile analysis, based on the ranking of the customers derived from the amount of purchase, customers are divided into ten groups of equal size, and then basic indices such as average amount of purchase, number of visits to the store, etc. are computed for each group [6][14]. From this report, it can be understood that all customers do not have an equal value for the store, but only a small fraction of the customers contribute to most of the profit in the store.

3.2 RFM Analysis

RFM analysis [14][8] is one of the tools most frequently used in the application purpose such as direct-mail marketing. The customers are classified according to three factors, i.e. recency of the last date of purchase, frequency of purchase, and monetary factor (purchase amount). Based on this classification, adequate sales promotion is executed for each customer group. For instance, in a supermarket, if a customer had the highest purchase frequency and the highest purchase amount, and did not visit to the store within one month, sufficient efforts must be made to bring back this customer from the stores of the competitors.

3.3 Customer Attrition Analysis

This analysis indicates which fraction of customers in a certain customer group would continuously visit the store in the next period (e.g. one month later) [8]. In other words, this is an analysis to predict how many customers will go away to the other stores. These output are also used for the calculation of LTV as described below.

3.4 LTV (Life Time Value)

LTV is a net present value of the profit which an average customer in a certain customer group brings to a store (an enterprise) within a given period [8][1]. It is calculated from the data such as sales amount of the customer group, customer maintaining rate, and discount rate such as the rate of interest on a national bond. Long-term customer strategy should be set up based on LTV, and it is an important factor relating to CRM system. However, the component for calculation of LTV prepared in C-MUSASHI is currently very simple and it must be customized depending on enterprises to use it.

These four tools are minimally required as well as very important for CRM in business field. It is possible to set up various types of marketing strategies based on the results of analysis. However, they are general and conventional, and then do not necessarily bring new knowledge to support differentiation strategy of the enterprise.

4 Store Management Systems Based on SPM

Store management systems in C-MUSASHI are to support for strategic planning of chain stores based on Store Portfolio Management (SPM) we proposed. SPM is the strategic store management methods based on Product Portfolio Management in strategic management theory in order to support for planning chain store strategy from the viewpoints of an overall firm to provide significant information of stores such as store ' s profitability and effectiveness of store sales promotion.

4.1 Product Portfolio Management in Strategic Management

Product Portfolio Management (PPM) is the strategic method for optimal allocation of management resources among multiple SBU (Strategic Business Units) in an enterprise to understand demand and supply of cash flow in each SBU [7]. It is not easy to completely understand demand and supply of cash flow in each SBU because of the multiplicity and the diversity of the circumstances of SBU. So "Growth-Share Matrix" has been employed in PPM in order to recognize clusters of SBU which exhibit the similar tendency of cash flow (See Figure 4). The growth-share matrix has two dimensions; the one is the growth dimension which is ratio of market growth in which each SBU belongs to, and the other is the share dimension which is relative market share of SBU to the rival company. These two dimensions come from the product lifecycle theory and the experience curve theory. Product lifecycle theory is to explain the dynamics of a product market by comparing a product to a living creature, and to propose that the growth process of product market is composed of four stages; introduction, growth, maturity and decline. In PPM the ratio of market growth in these four stages are associated with demand and supply of cash flow in each SBU. In the stage of introduction, the rate of market growth is high and the growth increases demand of cash flow to expand overall market and to penetrate their products into the market, while the market growth rate is low in stages of maturity and decline which decreases demand of cash flow. Based on the market growth dimension, it is possible to relatively evaluate which SBU demands cash flow in their enterprises.

Experience curve has been known as the linear relationship between the costs of production and the cumulative production quantity, which was first observed by a management consultant at the Boston Consulting Group in 1960s. It appears that the real value-added production cost decline by 20 to 30 percent for each doubling of cumulative production quantity. If a firm could acquire more experience of production by increasing its market share, it could achieve a cost advantage in its market (industry). Therefore high relative market share to a rival company which emerges a cost advantage is associated with capability of SBU to supply cash flow.

Four cells of the Growth-Share Matrix in Figure 4 are named such as "stars" "cash cows," "question marks" and "dogs". SBU in stars which has high market share and high growth rate, not only generates large amounts of cash, but also

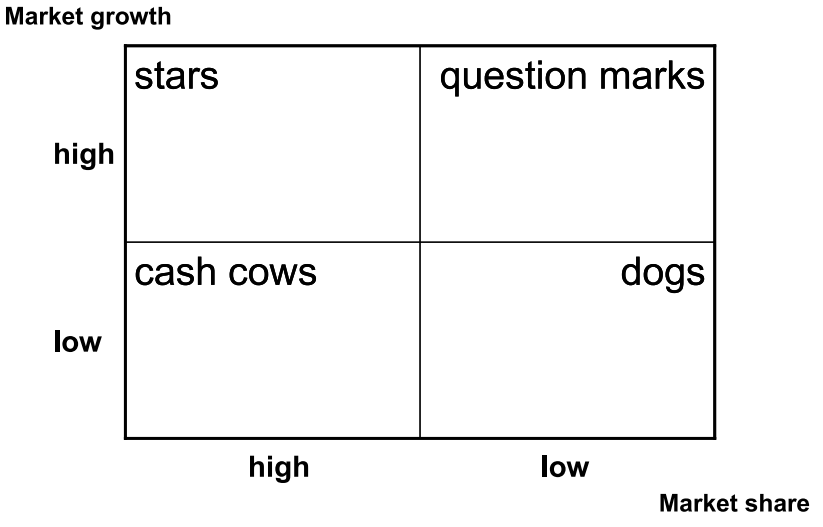

Fig. 4. The Growth-Share Matrix in PPM

consume large amounts of cash because of high growth market. "Cash cows" with high market share and low growth of market, exhibit the excess cash. "Question marks" which have low market share and high market growth consume large amounts of cash in order to gain market share and to become stars. "Dogs" with low market share and slow growth neither generate nor consume a large amount of cash. In PPM the strategy of each SBU is determined by whether its market share is high or low, and its ratio of market growth is high or not.

4.2 Store Portfolio Management for Chain Store Management

Based on the idea of the above PPM, we propose Store Portfolio Management (SPM) which is a methodology to support for strategic planning of chain stores management. The SPM provides useful information of each store such as store's profitability and effectiveness of store sales promotion from the viewpoints of the headquarter of an overall chain store. It is impossible to completely understand each situation of many stores and, even if it is possible, it is impossible to implement different marketing strategy corresponding to each store. The purpose of SPM is to plan different marketing strategy of each store cluster which has faced with the similar situation and markets. Here we the situations and markets of stores are measured by using various di-mensions.

Various kinds of evaluation dimensions can be employed in SPM such as profitability, sales amount, the number of visiting customers and an average of sales amount per customer (See Table 1). Users can select appropriate dimensions corresponding to their situations and industry. In SPM each evaluation criterion has been associated with each marketing strategy. For example, if a store is

Table 1. List of store's evaluation dimensions which can be used in SPM

Name of dimension	Function
Profitability	The ratio of profit to sales amount of a store
Sales amount	Sales amount per store and month
# of customers	The number of visiting customer in a month
Sales / a customer	An average of sales amount per customer in a month
Effectiveness of sales promotion	The ratio of waste sales promotion to all sales promotion of the store in a month

recognized as low profitability store, an appropriate pricing strategy has to be implemented to improve the profitability.

4.3 Store Evaluation

It is difficult for the marketing stuff of a chain store to categorize their stores into a number of clusters by using all the above dimensions because of low interpretability of the resulting clusters. According to our experience, it seems to be appropriate way to use two or three dimensions to clustering stores into some meaningful groups.

Store Evaluation by Using Two-dimensional Matrix. Figure 5 shows a typical case of store clusters by using two dimensions; profitability and sales amount. All stores have been classified into one of four clusters in comparison with an average of overall chain stores in Figure 5. Cluster A is "star store" cluster with high profitability and high sales amount. Stores in cluster B classified as "cash cows" exhibit high profitability and low sales amount. Stores in B have to increase market share in the local area by using more sales promotion strategies. Stores in cluster C classified as "Question marks" have low profitability and high sales amount and thus have to improve profitability. Cluster D is "dog store" with low profitability and sales amount. Marketing stuff of chain store has to consider whether these stores of D are to be scrapped or rebuilt under new store concept depending on the environment.

Figure 6 shows an example of store cluster of a drugstore chain in Japan by using above Profitability-Sales amount matrix. In this drugstore chain the number of stores classified into cluster A is small and the number of stores in cluster C or D is relatively large because this chain decreases competitiveness in each local area.

Store management system based on SPM which we developed has modules for analyzing the classified clusters in more details. Users can get useful information for the effective marketing strategy by using these modules. In fact, for this

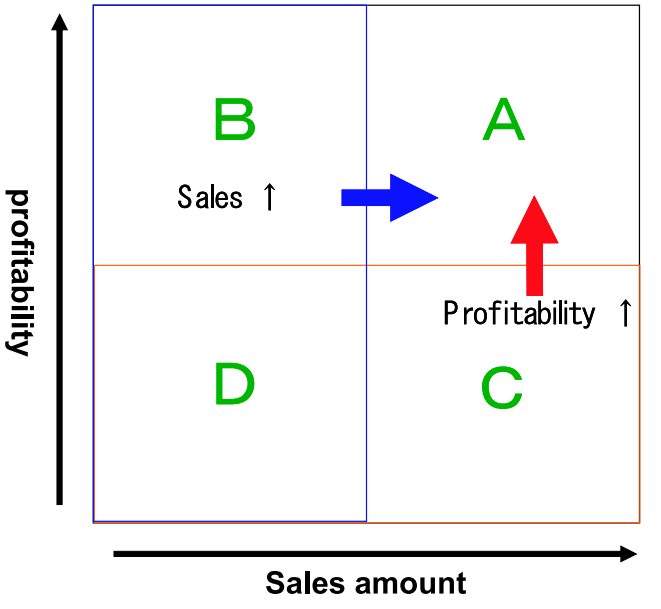

Fig. 5. A sample of store clusters by using Profitability-Sales amount matrix

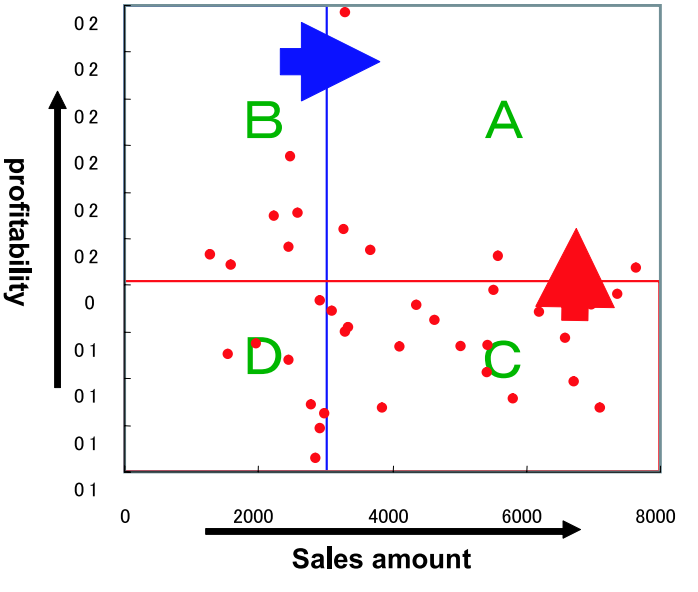

Fig. 6. An example of store clusters in a drugstore chain in Japan

drugstore chain, we found that the reason of the low profitability in these drugstores is ineffectiveness of sales promotion in each store. We then determined which product categories have to be improved for the effectiveness of sales promotion.One of the modules caluculates the number of days for each product category that the sales amount when sold at a discounted price is less than the average sales amount at regular price. Such promotion sales is called "waste sales promotion". Figure 7 shows the results obtained for three stores S_1, S_2, S_3. It is observed from the figure that the number of waste sales promotions in store A's soap category (See figure 7) is larger than that of other shops and the number of waste sales promotions in mouth care category of store B is large. The marketing stuff in the headquarters of drugstore can plan different sales strategy for each store to focus on the specific categories depending on these data.

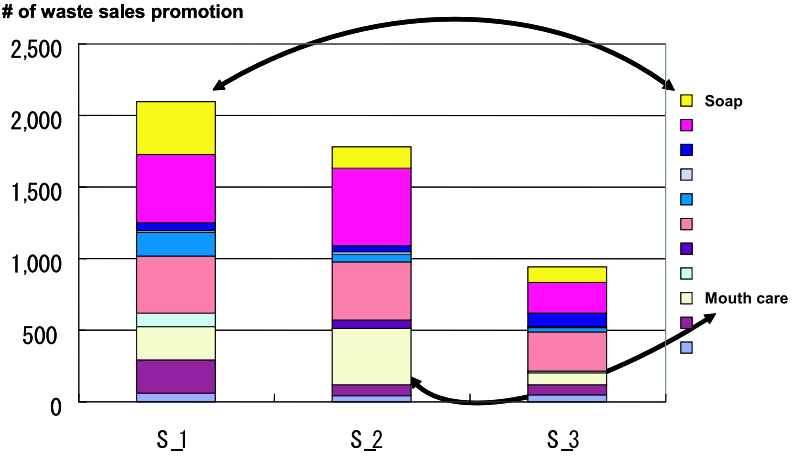

Fig. 7. The number of products with waste sales promotions of each store

Store Evaluation by Using Three-Dimensional Diagrams. Users can use more evaluation dimensions for analyzing situations of each store in detail. Figure 8 shows store evaluation by using threedimensions; profitability, an average of sales amount per customer and the number of visiting customers. It is observed in store S_1 that the number of visiting customers is large, but sales amount per customer and profitability is lower than other stores. We can guess that ratio of bargain hunters who purchase only discounted goods is high. In store S_2 the number of customers is small, but an average of sales amount per customer is very high compared to other stores. Store S_2 maybe has small number of loyal customer who contributes to almost profit of a store and acquire no sufficient newcomers from rival stores.

We can plan effective and appropriate store management strategy for each store clusters to understand each situation of stores by using the store manage-

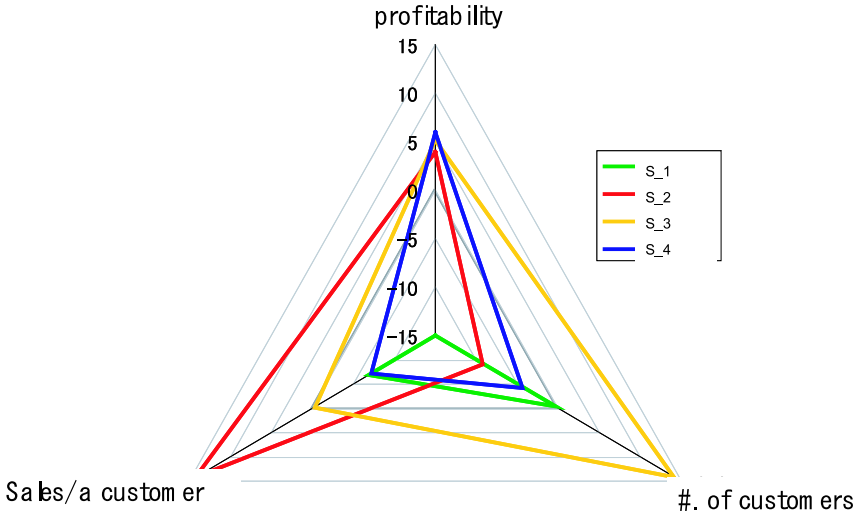

Fig. 8. Store evaluation by using three dimension diagrams

ment systems. Currently these store management systems can deal with the data of less than 100 stores and be applied to supermarket and drugstore chain.

5 Data Mining Oriented CRM Systems

In this section data mining oriented CRM system will be presented, which discovers new knowledge useful for implementing effective CRM strategy from the purchase data of the customers. General CRM systems commercially available simply comprise the functions of retrieval and aggregation as basic tools of C-MUSASHI, and there are very few CRM systems in which an analytical system that can deal with large-scale data equipped with data mining engine is available. In this section, we explain our system that can discover useful customer knowledge by integrating the data mining technique with CRM system.

5.1 Systems Structure and Four Modules of Data Mining Oriented CRM Systems

Figure 9 shows a structure of data mining oriented CRM system in C-MUSASHI. Customer purchase history data accumulated as XML table is preprocesses by a core system of MUSASHI. The preprocessed data is then provided as retail support information in two different ways.

In the first approach, the data preprocessed at the core of MUSASHI is received through WEB server. Then, the data is analyzed and provided to the

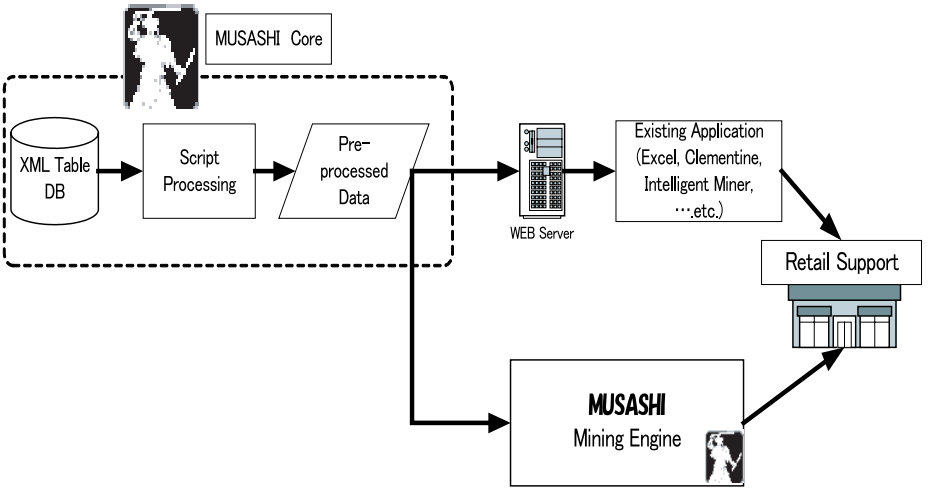

Fig. 9. Configuration of data mining oriented CRM systems

retail stores by using existing application software such as spread-sheet or data mining tools. In this case, C-MUSASHI is only in charge of preprocessing of a large amount of data.

In the second approach, the data is directly received from the core system of MUSASHI. Rules are extracted by the use of data mining engine in MUSASHI, and useful knowledge is obtained from them. In this case, C-MUSASHI carries out a series of processing to derive prediction model and useful knowledge. Whether one of these approaches or both should be adopted by the enterprise should be determined according to the existing analytical environment and daily business activities.

CRM system in C-MUSASHI which integrates the data mining technique consists of four modules corresponding to the life cycle of the customers [1][12][13]. Just as each product has its own life cycle, each customer has life cycle as a growth model. Figure 10 shows the time series change of the amount of money used by a typical customer. Just like the life cycle of the product, it appears that the customer life cycle has the stages of introduction, growth, maturation, and decline.

It is not that all customers should be treated on equal basis. Among the customers, there are bargain hunters who purchase only the commodities at the discounted price and also the loyal customers who make great contribution to the profit of the store. In the first stage, it is important to find the customers who will become loyal from among the new customers for effective sales promotion. So we developed an early discovery module to detect such customers in order to attract the customers who may bring higher profitability.

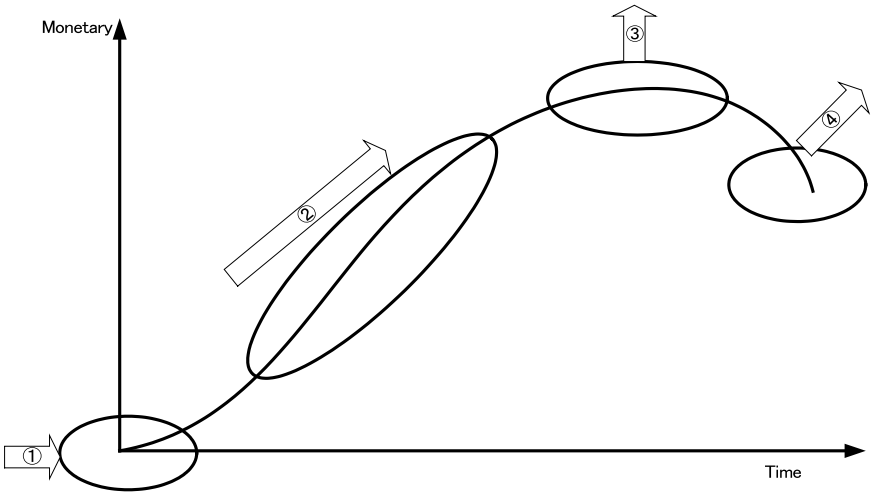

Fig. 10. Customer life cycle and four CRM modules

In the second stage, an analysis is required to promote new customers and to turn quasi-loyal customers to the loyal customers. For this purpose, we developed a decile switch analysis module. In the third stage, merchandise assortment is set up to meet the requirements of the loyal customers. A basket analysis module is prepared for the loyal customers in order to attain higher satisfaction and to raise the sales for them. In the final stage, for the purpose of preventing the loyal customers from going away to the other competitive stores, the analysis using the customer attrition analysis module is performed. Detailed description will be given below on these modules.

5.2 Early Discovery Module to Detect Loyal Customers from Newcomers

This module is a system for constructing a predictive model to discover from new customers within short time after the first visit to the store those who will become loyal customers in future and to acquire knowledge to capture these customers [9][10]. The user can select the preferred range of the customer groups classified by basic tools explained in Sections 3.1 and 3.2 and the period to be analyzed.

The explanatory attributes are prepared from the purchase data during the specified period such as one month or during the first five visits from the first one to the store. Sales ratio of the product category for each customer (the ratio of sales amount of the product category to total sales amount) is computed in the module. A model for predicting loyal customers will then be constructed from the above data joined together by using MUSASHI's command "xtclassify".

As a result, the model tells us which category of purchasing features these new prospective loyal customers have. Such information provides valuable implication as to how loyal customers are obtained from the competitive stores or to determine on which product category the emphasis should be put when a new store will be opened.

5.3 Decile Switch Analysis Module

Decile switch analysis module is a system to find out what kind of changes of purchase behavior of each of ten customer groups computed by decile analysis gives strong influence on the sales of the store. Given two periods, the following processing will be automatically started: First the customers who visited the store in the first period are classified into ten groups by decile analysis. The customers in each group are then classified into 10 groups again according to the decile analysis in the second period. When a customer did not visit the store in the second period, he/she does not belong to any decile group and thus classified as another group. Thus, the customers are classified into 11 groups. Since the customers are classified into 10 groups in the first stage, they are classified into 110 groups in total.

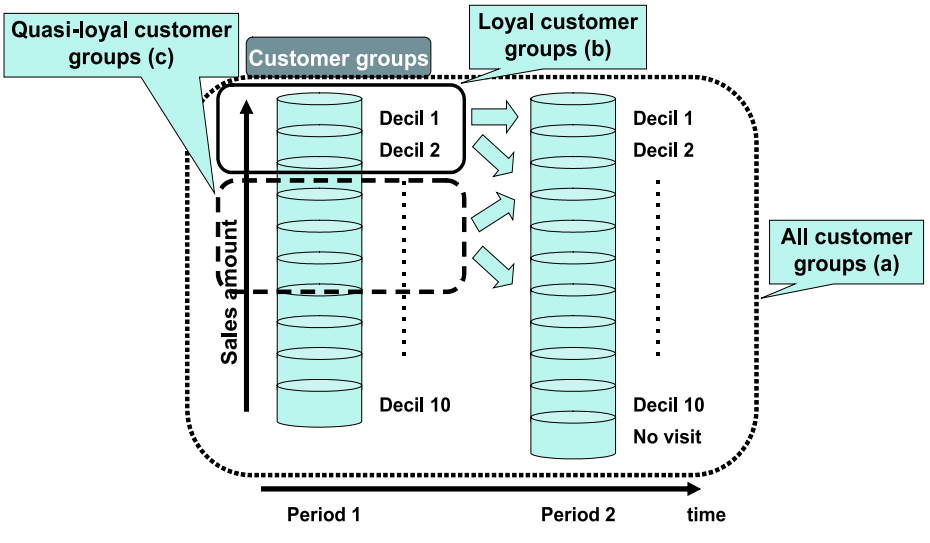

Fig. 11. The framework of decile switch analysis

For each of these customer groups, the difference between the purchase amount in the preceding period and that in the subsequent period is calculated. We then compute the changes of sales amount of which customer groups give strong influence on total sales amount of the store.

Next, judging from the above numerical values (influence on total sales amount of the store), the user decides which of the following data he/she wants to see, e.g., decile switch of (a) all customers, (b) loyal customers of the store, or (c) quasi-loyal customers (See Figure 11). If the user wants to see the decile switch of quasi-loyal customers, sales ratio of each product category for the relevant customer group in the preceding period is calculated, and a decision tree is generated, which shows the difference in the purchased categories between the quasi-loyal customers whose decile increased in the subsequent period (decile-up) and those whose decile value decreased (decile-down). Based on the rules obtained from the decision tree, the user can judge which product category should be recommended to quasi-loyal customers in order to increase the total sales of the store in future.

5.4 Basket Analysis Module of the Loyal Customer

For the purpose of increasing the sales amount of a store, it is the most important and also minimally required to keep loyal customers exclusively for a store. In general, the loyal customers have the tendency to continue to visit a particular store. As far as the merchandises and services to satisfy these customers are provided, it is easy to continuously keep these customers to the store than to make efforts to acquire the new customers from the rival stores. This module is to find out the merchandises preferred by loyal customers according to the result of the basket analysis on their purchase data [2].

From the results obtained by this module, it is possible not only to find out which product category the loyal customer prefers, but also to extract the most frequently purchased merchandise and to indicate the product belonging to C rank in ABC analysis. In the store control practiced in the past, if sales amount of the products preferred by the loyal customers is not very large, then the product often tends to disappear from the sales counter. Based on such information extracted from this module, the store manager can display the particular merchandise on the sales counter which loyal customer prefers and can pay special attention so that the merchandise will not be out of stock.

5.5 Module for Customer Attrition Analysis

Customer attrition analysis module is a system for extracting the purchase behavior of the loyal customers who left the store and to provide information for effective sales promotion in order to keep such loyal customers. When the user defines the loyal customers, the group of the customers is extracted, who had been loyal customers continuously for the past few months and had gone thereafter to the other stores. Using the sales ratio of product category preceding the attrition of the customers as explanatory variable, a classification model of the customer group is generated. By elucidating which group of customers is more easily diverted to the other store and which category of products these customers

had been purchasing, the store manager can obtain useful information on the improvement of merchandise lineup at the store to keep loyal customers.

Four modules explained above are briefly summarized in Table 2.

Table 2. The summary of four modules in data mining oriented CRM systems

Name of module	The purpose of module
Early discovery of loyal customer	To construct a prediction model to discover potential loyal customers from newcomers.
Decile switch analysis	To find out what kind of changes of purchase behavior of customer groups derived from decile analysis give strong effects on the sales of the store.
Basket analysis for loyal customer	To discover the merchandises preferred by loyal customers according to the result of the basket analysis to keep them to the store.
Customer attrition analysis	To extract the purchase behavior of loyal customers who left the store and to provide information for effective sales promotion in order to keep loyal customers.

6 Case Studies of C-MUSASHI in Japanese Supermarket

In this section we will discuss the cases of C-MUSASHI in real business world. Since we cannot deal with all of the cases of the four modules in this paper, we will introduce the cases of decile switch analysis and customer attrition analysis modules in a large-scale supermarket.

6.1 The Case of Customer Growth by Using Decile Switch Analysis Module in a Supermarket

There is a chain store of supermarket in Japan, selling a wide range of goods including fresh foods, groceries and medicine. Recently, the sales amount in almost all stores of them has been decreased because of the price reduction under local competitive market and deflation in Japan. However, one store gradually increased the sales amount more than those of the other stores during the period we are concerned with in this case. The purpose of this analysis is to find out which customer groups had positive effects on sales amount of the store and then to discover the feature of the product categories they purchased, in order to implement the effective sales promotions in other stores for increasing total sales of a chain store.

First, two periods, i.e. April and May of 2003, were set up for analysis by the marketing stuff in this supermarket. Figure 12 shows the number of customers

categorized in each decile switch group. There are many customers staying in decile 1 in both April and May, who are the most important loyal customers to the store. However they do not have so strong effects on sales of this store though the number of them is larger than that of the other decile switch groups.

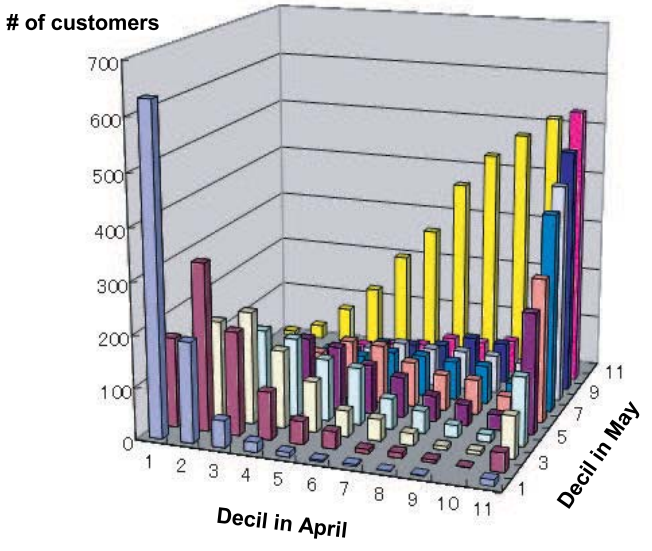

Fig. 12. The number of customers in each decile group

Figure 13 shows the changes of the purchase amounts in April and May of the customer groups classified according to the decile values of both periods. In the figure, the portion indicated by a circle shows that the sales for the quasi-loyal customer groups (the customers with decile 2-4) in April increases in May. From the figure, it is easy to observe that the sales increase of quasi-loyal customers makes great contribution to the increase of the total sales amount of the store.

Next, focusing on the quasi-loyal customers based on the above information, we carried out decile switch analysis by using decision tree to classify them into decile-up or decile-down customer groups. In the rules obtained from the decision tree, we found some interesting facts (See Figure 14). For instance, it was found that the customer who had exhibited higher purchase percentage of the product category such as milk, eggs, yoghurt, etc., which are easily perishable, shows high purchase amount in the subsequent period. Also, it was discovered that the customers who had been purchasing drugs such as medicine for colds or headache exhibited the increase in decile value in the subsequent period.

The store manager interpreted these rules as follows: If a customer is inclined to purchase daily foodstuffs at a store, total purchase amount of the customer

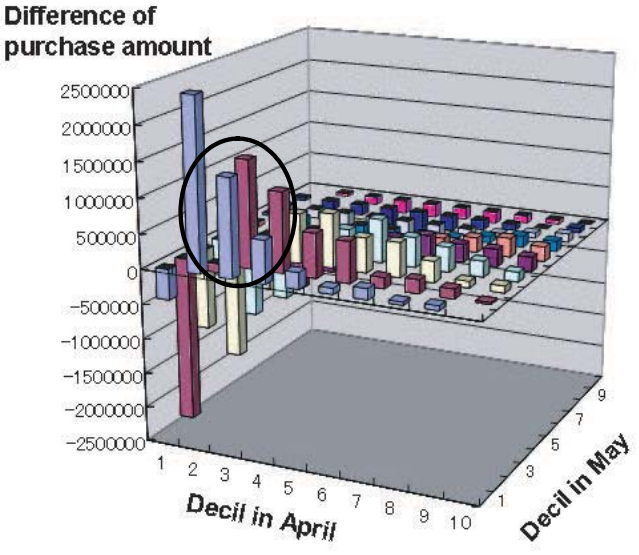

Fig. 13. Changes of sales amount for each decile group

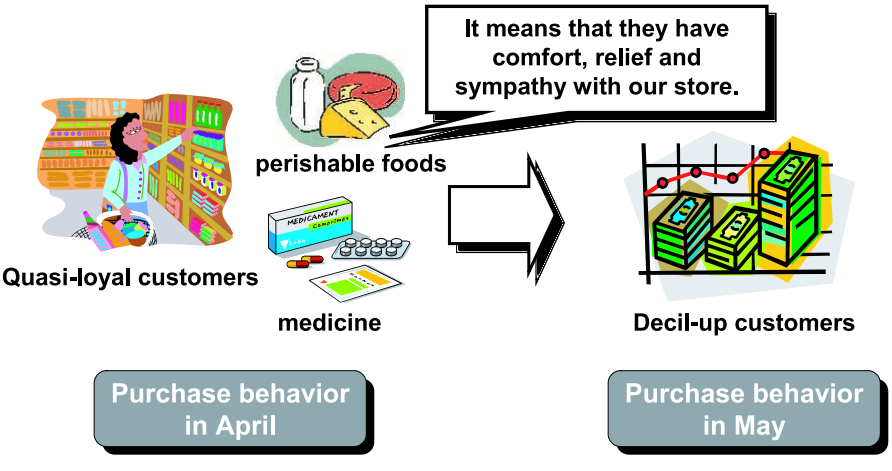

Fig. 14. The rules obtained from the decision tree by using decile switch analysis module

including other categories can be maintained at high level. As a result, the customer may have a sense of comfort, relief and sympathy with the store and would be more likely to buy other goods relating to health such as drugs. Based on such information, the store manager is carrying out sales promotion to keep the store in such atmosphere as to give the customers a sense of comfort, relief and sympathy to the store.

6.2 The Case of Customer Attrition Analysis

Due to the recent rapid change of economic circumstances, it has become important to manage loyal customers who contribute to profits of the store. The most important required term is to keep loyal customers of a store to increase the sales amount under these environments. However there exists a part of them who leave the store and switch to the rival store. The purpose of this analysis is to present analysis process to manage attrition customer by using customer attrition analysis module and to extract the purchase behavior of the loyal customers who left the store to provide useful information for effective sales promotion to keep such loyal customers. If we discover the distinctive purchase behavior of them before leaving the store, it is possible to implement sales promotion strategies to keep them to the store.

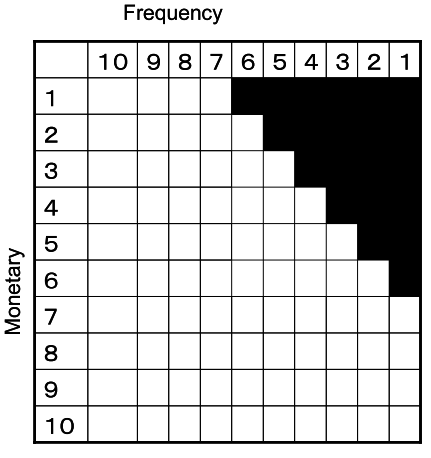

Fig. 15. The definition of loyal customer by using frequency and monetary dimensions

A user can define the loyal customer segments by using frequency and monetary dimensions (balck sells in Figure 15 illustrates an example of the loyal customer segments). After defining them, two customer groups are extracted; the first group called "fixed loyal customers" is defined as those who had been loyal customers continuously for the past four months in the target periods and the second called "attrition customers" is defined as those who had been loyal customers continuously for first three months and had gone to the other stores at the final month. We used four months of purchase data from Sep. 2002 through Dec. 2002 of a supermarket in Japan. The number of the fixed loyal customers is 1918, and that of attrition customers is 191. Using the sales ratios of rough product categories as explanatory variables, a classification model is generated which characterizes the distinction between the attrition customers and the fixed customers.

Figure 16 shows a part of results extracted from the derived decision tree. In this figure it was observed that the customer whose ratio of fruit category in his/her purchases had not been exceeding 5 percents and that of meat category had not been exceeding 15 percents, became an attrition customer after one month at the ratio of about 66 percents (44/67). Executing additional analysis, the marketing stuff in the store interpreted these rules as follows: The customers who have children have a tendency to leave the store. Depending on these findings, the new sales promotion strategy for a family with children has been planned in the store.

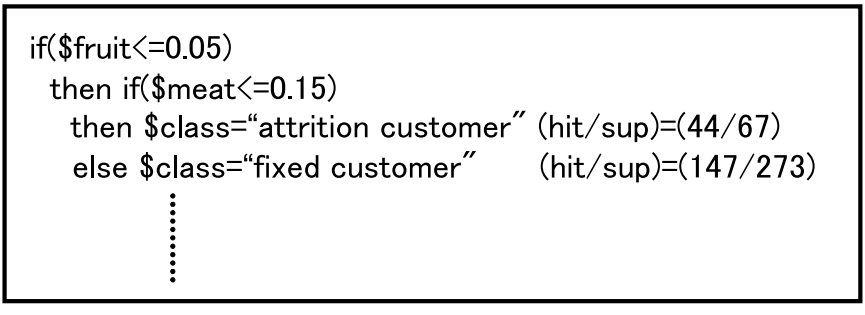

Fig. 16. A part of results extracted from purchase data of target customer groups

7 Conclusion

In this paper, we have introduced a CRM system called C-MUSASHI which can be constructed at very low cost by the use of the open-source software MUSASHI. We have explained components and software tools of C-MUSASHI. In particular, C-MUSASHI contains several data mining tools which can be used to analyze purchase behavior of customers in order to increase the sales amount of a retail store. However, we could not explain the details of all of the modules in this paper. In some of the modules, sufficient analysis cannot be carried out in actual business. We will try to offer these modules to the public as soon as possible so that those who are concerned in business field would have an advantage to use the modules. In future, we will continue to make improvement for the construction of effective CRM systems by incorporating the comments and advices from the experts in this field.

In C-MUSASHI, a typical decision tree tool and basket analysis tool were used as data mining technique. A number of useful data mining algorithms are now provided by the researchers. We will continuously try to utilize and incorporate these techniques into C-MUSASHI, and we will act as a bridge between the research field and actual business activities.

References

1. Blattberg, R. C., Getz, G., Thomas, J. S.: Customer Equity: Building and Managing Relationships as Valuable Assets. Harvard Business School Press. (2001)
2. Fujisawa, K., Hamuro, Y., Katoh, N., Tokuyama, T. and Yada, K.: Approximation of Optimal Two-Dimensional Association Rules for Categorical Attributes Using Semidefinite Programming. Lecture Notes in Artificial Intelligence **1721**. Proceedings of First International Conference DS'99. (1999) 148–159
3. Hamuro, Y., Katoh, N. and Yada, K.: Data Mining oriented System for Business Applications. Lecture Notes in Artificial Intelligence 1532. Proceedings of First International Conference DS'98. (1998) 441–442
4. Hamuro, Y., Katoh, N., Matsuda, Y. and Yada, K.: Mining Pharmacy Data Helps to Make Profits. Data Mining and Knowledge Discovery. **2** (1998) 391–398
5. Hamuro, Y., Katoh, N. and Yada, K.: MUSASHI: Flexible and Efficient Data Preprocessing Tool for KDD based on XML. DCAP2002 Workshop held in conjunction with ICDM2002. (2002) 38–49
6. Hawkins, G. E.: Building the Customer Specific Retail Enterprise. Breezy Heights Publishing. (1999)
7. Hedley, B.: Strategy and the business portfolio. Long Range Planning. **10** 1 (1977) 9–15
8. Hughes, A. M.: Strategic Database Marketing. The McGraw-Hill. (1994)
9. Ip, E., Johnson, J., Yada, K., Hamuro, Y., Katoh, N. and Cheung, S.: A Neural Network Application to Identify High-Value Customer for a Large Retail Store in Japan. Neural Networks in Business: Techniques and Applications. Idea Group Publishing. (2002) 55–69
10. Ip, E., Yada, K., Hamuro, Y. and Katoh, N.: A Data Mining System for Managing Customer Relationship. Proceedings of the 2000 Americas Conference on Information Systems. (2000) 101–105
11. MUSASHI: Mining Utilities and System Architecture for Scalable processing of Historical data. URL: http://musashi.sourceforge.jp/.
12. Reed, T.: Measure Your Customer Lifecycle. DM News 21. **33** (1999) 23
13. Rongstad, N.: Find Out How to Stop Customers from Leaving. Target Marketing 22. **7** (1999) 28–29
14. Woolf, B. P.: Customer Specific Marketing. Teal Books. (1993)