社団法人 情報処理学会 研究報告
IPSJ SIG Technical Report

2003－ＩＣＳ－133 (4)
2003／9／14

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

Katsutoshi YADA† Yukinobu HAMURO‡ Naoki KATOH*
Takashi WASHIO** Issey FUSAMOTO† and Daisuke FUJISHIMA†

† Faculty of Commerce, Kansai University 3-3-35 Yamate, Suita, Osaka, 564-8680 Japan
‡ Faculty of Business Administration, Osaka Sangyo University 3-1-1 Nakagaito, Daito, Osaka, 567-8530 Japan
*Graduate School of Engineering, Kyoto University, Yoshida-Honmachi, Sakyo-ku, Kyoto, 606-8501 Japan
**Institute for the Scientific and Industrial Research, Osaka University 8-1 Mihogaoka, Ibaraki, Osaka, 567-0047 Japan
E-mail: † {yada, da00587, da00591}@kansai-u.ac.jp, ‡ hamuro@adm.osaka-sandai.ac.jp, *naoki@archi.kyoto-u.ac.jp,
**washio,@sanken.osaka-u.ac.jp

Abstract MUSASHI is a set of commands that 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 data mining oriented CRM systems based on MUSASHI, which are integrated with the marketing tools and data mining technology. Everybody can construct useful CRM systems at extremely low cost by introducing MUSASHI.

Keyword MUSASHI, C-MUSASHI, CRM, Data Mining, Knowledge Discovery, XML

MUSASHI 上で動くデータマイニング指向 CRM システム: C-MUSASHI

矢田 勝俊† 羽室 行信‡ 加藤 直樹* 鷲尾 隆** 房本 一誠† 藤島 大輔†

† 関西大学商学部 〒564-8680 大阪府吹田市山手町 3-3-35
‡ 大阪産業大学経営学部 〒574-8530 大阪府大東市中垣内 3-1-1
*京都大学大学院工学研究科 〒606-8501 京都府京都市左京区吉田本町
**大阪大学産業科学研究所 〒567-0047 大阪府茨木市美穂ヶ丘 8-1

E-mail: † {yada, da00587, da00591}@ipcku.kansai-u.ac.jp, ‡ hamuro@adm.osaka-sandai.ac.jp, *naoki@archi.kyoto-u.ac.jp
**washio@sanken.osaka-u.ac.jp

あらまし MUSASHI とは、ビジネスにおけるデータマイニングに必要な大量データ処理のために開発されたコマンド群で、XML データを効率的に処理するためのオープンソフトウェアである。本論文では、MUSASHI 上で動く、いくつかのマーケティングツール、データマイニングツールを統合したデータマイニング指向の CRM システムを紹介する。我々は、MUSASHI を導入することによって、低コストで高度な CRM システムの構築が可能であることを明らかにする。

キーワード MUSASHI, C-MUSASHI, CRM, データマイニング, 知識発見, XML

1. Introduction

MUSASHI is a set of commands developed for the processing of a large amount of data required for data mining in business field and open-source software for achieving efficient processing of XML data [1][2][3][4]. 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 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 severe pressure. Many of these enterprises are now trying to encompass and maintain loyal customers through the introduction of FSP [5][6]. FSP (Frequent Shoppers Program) is defined as one of the CRM systems to accomplish effective sales promotion by accumulating purchase history of the customers in its own database and by recognizing the nature and the behavior of the loyal customers. However, it is very rare that CMS 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 data are actually accumulated while technique, software and human resources in their firms to analyze these data are in shortage, and the analysis of the customers is not in progress. Therefore, in many cases, the enterprises simply accumulate the data but do not carry out the analysis of the customers.

In this paper, we shall introduce a CRM system which can be constructed at very low cost by the use of the open-source software MUSASHI, which can be adopted freely even by a small enterprise. The components of the system comprise marketing tools and data mining technology, which we developed so far through joint research activities with various types of enterprises. Thus, it is possible to carry out the analysis of the customers without building up a new analytical system.

2. C-MUSASHI in Retailers

C-MUSASHI is defined as 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 customer control. Fig. 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 without introducing data warehouse through the processes given below. In POS 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 converted by "MUSASHI journal converter" to XML data with minimal data loss.

The framework of such system design provides two advantageous features. First, the loss of data is minimized. In the framework given above, no data will be lost basically. If new data is needed to discover new knowledge, the data can be provided from XML data if it can be obtained from the operation of POS registers.

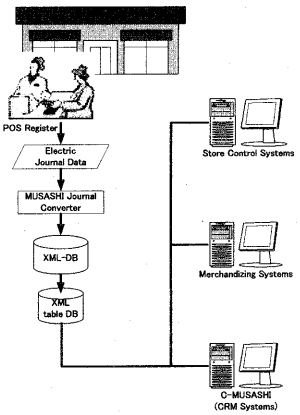

Fig. 1 C-MUSASHI system in the retailers

Secondly, the burden of the system design can be extensively reduced. Because all of the data are accumulated, necessary data can be easily extracted later. Therefore, these schemes can flexibly cope with the changes in the system.

However, if all operation logs at the 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*. 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, MUSASHI aims to construct an efficient system with high degree of freedom.

Based on the purchase data of the customers thus accumulated, the following systems are built up: store management system for the basic data processing in stores such as accounting information, merchandising system for commodity control and price control, and C-MUSASHI, which will be explained in the succeeding sections.

3. Components of C-MUSASHI

Software tools of C-MUSASHI are categorized into two groups: those for basic customer analysis and CRM systems using data mining technique. The former provides basic information necessary for the implementation of CRM strategy. The latter is a system to discover new knowledge to carry out effective CRM by using data mining technique.

In this section, we shall introduce the tools for basic customer analysis. Basic tools in C-MUSASHI have

several tools usually incorporated in general CRM systems: decil analysis, RFM analysis, customer attrition analysis, and LTV measurement. They are used for basic customer analysis. C-MUSASHI also has many other tools for customer analysis. We will present here only a part of them here.

3.1. Decil analysis

In decil analysis, based on the ranking of the customers derived from the amount of purchase, customers are divided into 10 groups with equal number of customers, and then basic indices such as average amount of purchase, number of visits to the store, etc. are computed for each group [5][6]. From this report, it can be understood that all customers do not have 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 [6][7] 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 what fraction of customers in a certain customer group would continuously visit the store in the next period (e.g. one month later) [7]. In other words, this is an analysis to indicate how many customers have gone away to the other stores. These numerical values 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 [7][8]. 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 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. CRM Systems Based on the data mining technique

In this section, CRM system based on the data mining technique will be presented, which discovers new knowledge useful for implementing effective CRM strategy from the purchase data of the customers. General CRM system commercially available simply comprises the processes of retrieval and aggregation for each customer group, and there are very few CRM systems in which analytical system that can deal with large-scale data equipped with data mining engine is available in actual business field. In this section, we explain our system that can discover useful customer knowledge by integrating the data mining technique with CRM system.

4.1. Systems structure of C-MUSASHI and four modules

Fig. 2 shows a structure of CRM system using 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 retail stores by 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.

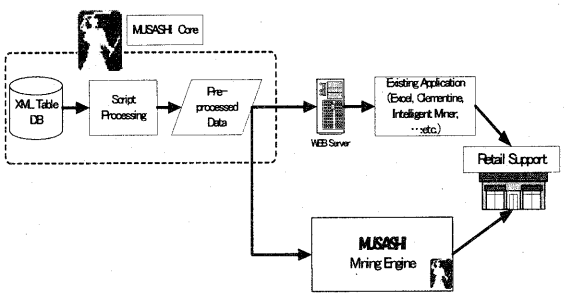

Fig. 2 System structure of C-MUSASHI

CRM system in C-MUSASHI which integrates the data mining technique consists of four modules corresponding to the life cycle of the customers [8][9][10]. Just as each product has its own life cycle, each customer has life cycle as a growth model. Fig. 3 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 customer life cycle has the stages of introduction, growth, maturation, and decline.

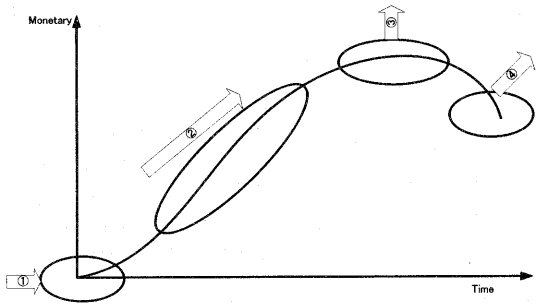

Fig. 3 Customer life cycle and four CRM modules

It is not that all customers should be treated on equal basis. Among the customers, there are bargain hunters which purchase only the commodities at the discounted price and also the loyal customers who give great contribution to the profit of the store. In the stage ①, the loyal customers are discriminated from among the new customers. We developed an early discovery module to detect loyal customers in order to attract the customers who may bring higher profitability.

In the stage ②, an analysis is required to promote new customers and quasi-loyal customers to turn them to the loyal customers. For this purpose, we developed a decil switch analysis module. In the stage ③, merchandise assortment is set up to meet the requirements of the loyal customers. A basket analysis module was prepared for the loyal customers in order to attain higher satisfaction and to raise the sales for them. In the final stage ④, for the purpose of preventing the loyal customers from going away to the other competitive stores, analysis is performed using the customer attrition analysis module. Detailed description will be given below on these modules.

4.2. Early discovery modules 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. The user can select the preferred range of the customer groups classified in Sections 3.1 and 3.2 and the period to be analyzed. The new customers are then classified into loyal customers and non-loyal ones.

The explanatory attributes are prepared from the purchase data during the specified period such as one-month or during the number of visits to the store from the first visit. Sales ratio of the product category for each customer (the ratio of sales amount of the product category to total sales amount) is generated in the module. When a model is built up by using the data mining engine of MUSASHI, a model for predicting loyal customers can be constructed from the above data joined together by using the model generating 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 when loyal customers are obtained from the competitive stores or when it must be determined on which product category the emphasis should be put when a new store will be opened.

4.3. Decil switch analysis module

Decil switch analysis module is a system to find out what kind of changes of purchase behavior of each customer group based on decil give strong influence on the sales of the store. Given two periods, the following processing will be automatically started: The changes of purchase behavior of the customers during the two periods are calculated, and the customers are classified into 110 customer groups according to the decil value of both periods. For instance, the customers who was classified as decil 3 in the preceding period are to be classified as one of decil 1 through 10 or as the customers who did not visit the store in the subsequent period. For

each of these customer groups, the difference between the purchase amount in the preceding period and that in the subsequent period is calculated, which makes it clear how the changes of sales amount of each of the 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., decil switch of all customers, loyal customers of the store, or quasi-loyal customers. If the user wants to see the decil switch of quasi-loyal customers, sales ratio of each product category for each 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 decil increased in the subsequent period (decil-up) and those whose decil value decreased (decil-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.

4.4. Basket analysis module of the loyal customer

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

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.

4.5. Customer attrition analysis module

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 regain 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 four 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.

4.6. The case of decil switch analysis module in a supermarket

Since we cannot discuss all of the cases of the above four modules in this paper, we will analyze the data of a large-scale supermarket and try to find out the possibility to promote quasi-loyal customers to loyal customers by using decil switch analysis module.

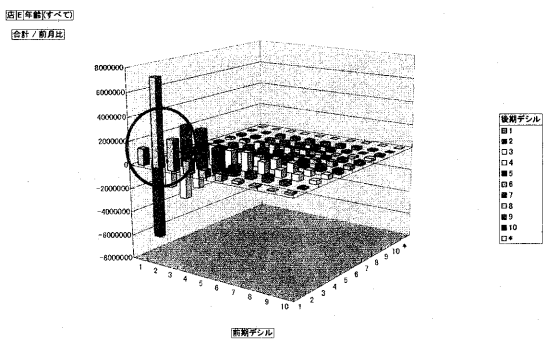

Fig. 4 Changes of sales amount for each decil switch group

In the supermarket used in this research, the sales increased more than those of the other stores during the period we are concerned with. The purpose of the analysis is to elucidate the reason in terms of features of the purchase behavior of the customers.

First, two periods, i.e. April and May of 2003, were set up for analysis. Fig. 4 shows the changes of the purchase amounts in April and May of the customer groups classified according to the decil values of both

periods. In the figure, the portion indicated by a circle shows that the sales for the quasi-loyal customers groups (the customers with decil 2 -4) in April increases in May. From the figure, it is clear 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, let us try to carry out decil switch analysis by using decision tree. In the rules obtained from the decision tree, there is some interesting information. For instance, it was found that the customer who had purchased higher 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 decil 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 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.

5. 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 ware 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 results and actual business activities.

Acknowledgements

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

References

[1] MUSASHI:Mining Utilities and System Architecture for Scalable processing of Historical data. URL: http://musashi.sourceforge.jp/

[2] Y. Hamuro, N. Katoh, and K. Yada, Data Mining oriented System for Business Applications, Lecture Notes in Artificial Intelligence 1532, Proceedings of First International Conference DS'98, pp.441-442, 1998.

[3] Y. Hamuro, N. Katoh, Y. Matsuda, and K. Yada, Mining Pharmacy Data Helps to Make Profits, Data Mining and Knowledge Discovery, Vol. 2 Issue 4, pp.391-398, December 1998.

[4] Y. Hamuro, N. Katoh and K. Yada, MUSASHI: Flexible and Efficient Data Preprocessing Tool for KDD based on XML, DCAP2002 Workshop held in conjunction with ICDM2002, 38-49, 2002.

[5] G. E. Hawkins, Building the Customer Specific Retail Enterprise, Breezy Heights Publishing, 1999.

[6] B. P. Woolf, Customer Specific Marketing, Teal Books, 1993.

[7] A. M. Hughes, Strategic Database Marketing, The McGraw-Hill, 1994.

[8] R. C. Blattberg, G. Getz and J. S. Thomas, Customer Equity: Building and Managing Relationships as Valuable Assets, Harvard Business School Press, 2001.

[9] T. Reed, Measure Your Customer Lifecycle, DM News 21, No.33, p.23, 1999.

[10] N. Rongstad, Find Out How to Stop Customers from Leaving, Target Marketing 22, No.7, pp.28-29, 1999.