Mining Pharmacy Data Helps to Make Profits

YUKINOBU HAMURO   hamuro@adm.osaka-sandai.ac.jp
Department of Business Administration, Osaka Industrial University, 3-1-1, Nakagaito, Daito, Osaka, Japan

NAOKI KATOH   naoki@archi.kyoto-u.ac.jp
Graduate School of Engineering, Kyoto University, Yoshida-Honmachi, Sakyo-ku, Kyoto, Japan

YASUYUKI MATSUDA   pharma@ba2.so-net.or.jp
G&G Pharma Inc., 2-177-2, Kitahanada-Cho, Sakai, Osaka, Japan

KATSUTOSHI YADA   yada@adm.osaka-sandai.ac.jp
Department of Business Administration, Osaka Industrial University, 3-1-1, Nakagaito, Daito, Osaka, Japan

Abstract.   *Pharma*, a drugstore chain in Japan, has been remarkably successful in the effective use of data mining. From over one tera bytes of sales data accumulated in databases, it has derived much interesting and useful knowledge that in turn has been applied to produce profits. In this paper, we shall explain several interesting cases of knowledge discovery at Pharma. We then discuss the innovative features of the data mining system developed in Pharma that led to meaningful knowledge discovery.

Keywords: data mining, knowledge discovery, pharmacy, point of sales

1.   Introduction

Due to the rapid development of modern computer technologies, much progress has been made in automating daily office work. This, in turn, has resulted in the accumulation of a significant amount of business data into databases. However, it does not seem that most companies have made full use of this accumulated data for strategic planning.

On the other hand, *knowledge discovery in databases* or *data mining* has become an active research area in which new technologies or methodologies are sought to automatically extract meaningful knowledge discovery from business data (Agrawal et al., 1993; Fukuda et al., 1996; Piatetsky-Shapiro, 1991; Uthurusamy et al., 1991). Although several successful cases of data mining systems in business have been reported (Apte et al., 1993; Carter et al., 1987; Hoschka et al., 1991; Manago et al., 1991; Schmitz et al., 1990), it is not clear how companies construct actual data mining systems and how they produce profits by using knowledge derived from them. In this paper, we shall introduce the case of Pharma[1], a drugstore chain in Japan, that has been remarkably successful in effectively using their data mining system. From a large amount of sales transaction data through POS (Point Of Sales)

terminals accumulated therein, Pharma has derived valuable knowledge that has increased retail sales and profits.

The organization of this paper is as follows; Section 2 discusses three successful cases of knowledge discoveries in Pharma. Section 3 discusses design concepts and innovative features of the data mining system of Pharma, and Section 4 concludes our discussion.

2. Successful cases of knowledge discovery

Pharma is a voluntary drugstore chain consisting of 1,230 retail shops across the country. Total annual sales reach approximately seventy billion yen and its membership club has more than 2.3 million customers. Pharma's organization is as illustrated in figure 1. The main tasks of Pharma headquarters are (i) to collect order data from retailers and send it on to wholesalers who deliver to the ordering retailer, (ii) to maintain customer and sales information, and (iii) to provide research reports on consumer behavior to manufacturers.

Over the past ten years, Pharma has accumulated an average of 60 giga bytes of sales data annually. From the knowledge gained in this large body of data, Pharma provides invaluable information about consumer behavior to manufactures, achieving favorable position in trade and competitive advantage. Thus, knowledge discovery can be regarded as the key business core of Pharma.

2.1. Analysis of store layout

One typical example of discovery through data mining is a new rule of associative purchasing concerning a pain reliever. Pain relievers were always candidates for discount sale, the initial
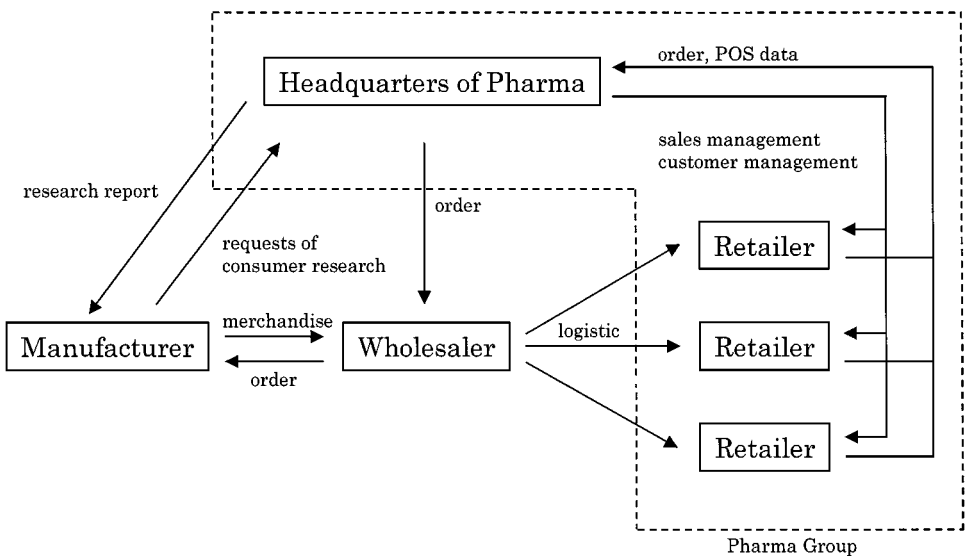

Figure 1.   Operation flow of Pharma.

motivation was to find a way in which to generate higher profits. By examining sales data it was discovered that this medicine has a high correlation with sanitary goods in terms of associated purchasing. Further analysis showed that by adding the attribute of goods location in stores to records of sales files, it was found that pain relievers can be sold 1.5 times more than before even at the regular price.

2.2. Knowledge discovery from deviation data

Since disposable pocket heaters, called *kairo* (in Japanese) were believed to sell mostly in winter, manufacturers formerly used to suspend their production line in summer. However, it was found that turnover continued into the summer at one store. This deviation data aroused interest. A simple examination revealed that *kairo* were sold off for clearance at that store. The more interesting and important question was why customers would buy them at that time. On further analysis of sales data at that store to find an association rule concerning associated purchasing, it was discovered that *kairo* have a very strong correlation with medicine for rheumatism. By pursuing the same analysis to sales data of all stores, it was finally found that *kairo* were also purchased in summer by women working in well air-conditioned offices.

Based on the above analysis, it became clear that *kairo* could sell well in summer. After providing this knowledge to manufacturers, they resumed production of *kairo* that eventually lead to the effective use of production capacity.

2.3. Effectiveness of free samples

Pharma has analyzed the effectiveness of distributing free samples of new products. This task was originally motivated by the inquiry from manufacturers of new products. The task was done in the following way. It is possible through POS terminals to acquire information of the sales volume of a particular commodity, but it is impossible to know who bought it and why it was sold. Manufacturers really want to know how distributing free samples affects its sales volume and the recognition level of customers because such information is essential to determining future planning for development of new products.

In order to cope with such requests from manufacturers, Pharma attached two additional functions to POS terminals, i.e., customer information management and questionnaire. Using the first function, it became possible to perform an analysis on the effectiveness of distributing free samples. This was done by classifying customers into two groups, depending upon whether samples were distributed or not, and then by comparing the sale volume of these two groups. In this way, it was learned that distribution of free samples is really effective.

Introducing a questionnaire function made it possible for detailed examination of consumer behavior. When starting to use this function, a number of questions are displayed on the screen equipped with a POS terminal. Entering answers into POS terminals will automatically collect detailed information concerning a purchased commodity such as customer consciousness and purchase motivation. Since manufacturers are extremely interested in obtaining this information, they remunerate stores for each questionnaire, which in turn gives incentive to clerks to use the questionnaire function.

Table 1. Pharma's advantage (cited from (Yada, 1996)).

	Ave. inventory turnover (rate per year)	Ave. sales (in million yen)
Retailers of Pharma	5–6	60–70
Other retailers	2–3	18–24

As discussed above, Pharma has made full use of discovered knowledge to make profits. As a result, Pharma has achieved a favorable position in trade and has attained a competitive advantage. The paper by Yada (1996) has verified Pharma's competitive advantage over other drugstore chains from several viewpoints (see Table 1).

3.   System characteristics

We shall discuss several interesting features of the data mining system that made it possible to discover the above knowledge. Currently, Pharma uses 27 UNIX workstations for daily routine business as well as data mining, and 37 lap-top personal computers for data transmission to and from their 1,230 member stores (see figure 2). When stores are closed, sales data accumulated during the day are sent to the headquarters through public communication network, and then stored in data files for the purpose of routine business and for data mining purpose. The total computer system of Pharma is designed to attain the following three goals:
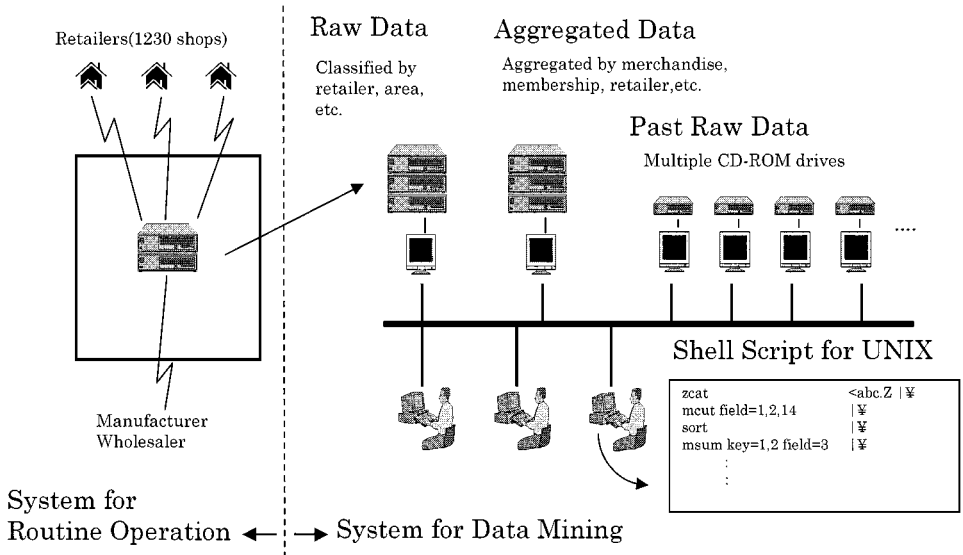
Figure 2. System configuration of Pharma.

```
100262 1994/12/01/10:59:12 1012 2010262010509 Y.Hamuro 780 1560 2 4987103030799 Vitumin_C 120 2510 TAKEDE ...
100262 1994/12/01/10:59:12 1012 2010262010509 Y.Hamuro 780 780 1 4903301447061 Diaper_ABC 64 10 P&C ...
13527 1994/11/06 22:10:12 101 Non Not_Member 198 198 1 0000000000232 Pharma_Coupon * Pharma ...
13527 1994/11/06 22:10:12 101 Non Not_Member 998 998 1 4901301006561 Anodyne_EFG 23 TAKEDE...
13527 1994/11/06 22:10:12 101 Non Not_Member  1358 2716 2 4901301006555 Napkin 80 p&C ...
13527 1994/11/06 22:10:12 101 Non Not_Member  98 298 1 4901301006554 Wrap_aaa 100 ABCDE ...
10548 1994/11/06 22:10:12 18106 2010262012541 K.Yada 180 180 1 4901301006561 Anodyne_EFG 23 TAKEDE ...
10548 1994/11/06 22:10:12 18106 2010262012541 K.Yada 298 298 1 4901301006554 Napkin 40 KAHO ...
                                              :
                                              :
```

From the leftside, Retailer Code, Sales Date, Sales Time, Receipt No., Membership Code, Name,
Unit Price, Price, Quantity, Merchandise Code, Merchandise Name, Size, Manufacturer Code, Manufacturer Name ...

Figure 3.   The structure of sales data.

  (i) All input data from POS terminals can be reproduced without losing any information
       to meet any future requests of data mining.
 (ii) The high efficiency of data processing can cope with a large amount of data.
(iii) The system can quickly cope with various requests of data mining.

   In order to attain these goals, Pharma designed its computer system with the following
innovative features:

(1) It is quite common in most companies to have a large amount of input data stored on
     computers for routine operation after aggregating it into compact form to save disk
     space. In addition, the main concern of current data mining research is how to derive
     useful knowledge from databases used for routine operation. With data mining in mind,
     however, we believe this approach suffers serious difficulty because some important
     information may be lost. Thus, in Pharma, all histories of input data are kept without
     any aggregations illustrated in figure 3. In fact, it would not have been possible to
     discover an association rule between pain relievers and sanitary goods as mentioned
     in Section 2.1, if data were aggregated (for instance, by summing up sales per item
     from sales transaction data). We believe that keeping all original data as history data
     is a key to the success of data mining. Although doing so is apparently costly from the
     viewpoint of disk space, it pays off as our research demonstrates.
(2) In order to attain goal (ii), the main effort was to avoid as much as possible communica-
     tion bottleneck and system overhead as much as possible. In principle Pharma's system
     is, therefore, designed so that programs or files are not shared by multiple users. Also,
     neither commercial online DBMS nor commercial application programs are used. As
     a result, computers can be used in stand-alone environments. On the other hand, the
     system contains a bunch of redundant programs and files, but such redundancy pays
     off in view of goal (ii). Notice that such redundancy does not lead to data inconsis-
     tency in Pharma's system because files for routine business are updated in batch mode.
     Furthermore, we try to use computers in single-task environment to extract the original
     computer power by avoiding unnecessary system programs. In addition, as will be
     mentioned below, all programs for data mining as well as routine business are written
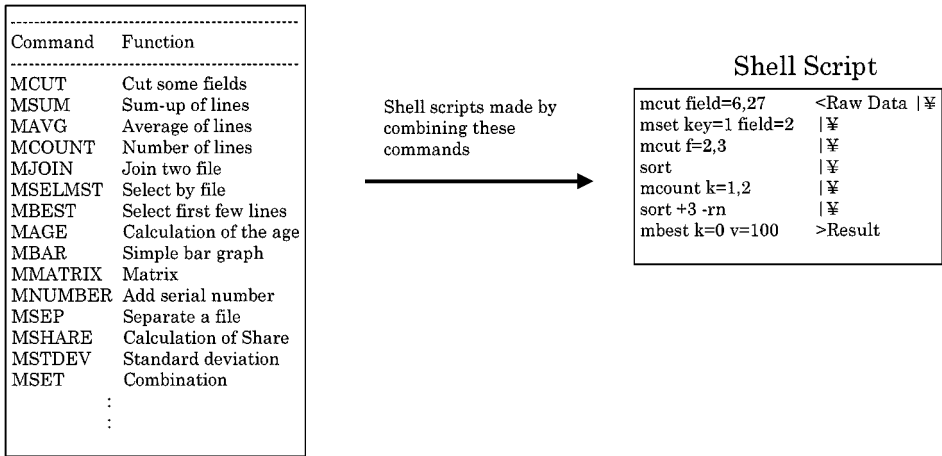   by using only UNIX shell commands and a set of originally developed commands.
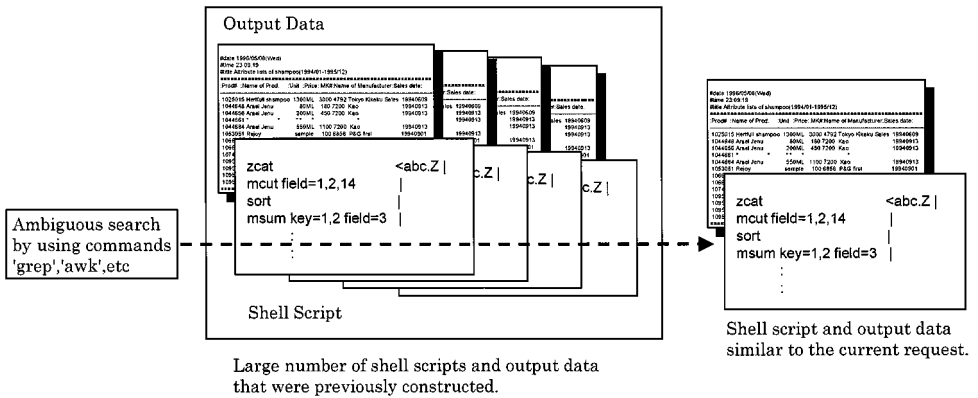
Figure 4. List of commands and shell script.

Figure 5. Retrieval of existing shell scripts and output data.

(3) The major principle underlying program development is to use UNIX shell commands as much as possible. For the functions that cannot be done by UNIX shell commands, about fifty commands were originally developed (see figure 4). All application programs for data mining as well as routine operation are written by combining these commands. Thus, the size of every program is very small. This helps to attain goals (ii) and (iii).

(4) Programs for data mining are easy to reuse. Since data mining becomes a daily routine operation, Pharma has accumulated a considerably large number of data mining programs. Thus, in order to quickly respond to a new data mining request and to effectively utilize such assets, a systematic tool for retrieving existing data mining programs that were previously constructed for purposes similar to the one requested has been developed (see figure 5). Since programs are written as shell scripts, this task can be

done with relative ease using UNIX commands such as 'grep'. We consider that this tool together with programs written in shell scripts can be viewed as a knowledge base system for data mining.

(5) In the information system of Pharma, the system for routine operation is tightly coupled with that for data mining. So it frequently happens that the system for routine operation needs to be modified by the requirements from the data mining system. Thus, it is very important to realize the operational system's flexibility so as to easily cope with data mining requests. For instance, all files in the whole system consist of variable-length records and are relatively easy to be augmented with new attributes. Such flexibility is essential to data mining tasks.

These features achieve high efficiency data mining processing at extremely low cost. In fact, the cost relating to the information system of Pharma is only 0.4% of annual sales.

4. Conclusion

We have discussed how the data mining system of Pharma produces profits and how it is constructed to increase its effectiveness and efficiency. For further development of our data mining system, it is important to construct a more effective knowledge base so that we can store and retrieve more detailed information about the data mining processes done in the past. In order to derive useful knowledge, we need to test a number of hypotheses against databases in a manner of trail and error. During this process, we believe easy retrieval of past information for similar purposes and simple access to results obtained previously, is very helpful. We are currently investigating how the experiences and ways of thinking of previous data miners are to be stored on computers as knowledge base.

Note

1. We have initiated a joint research project with Pharma on data mining and since all stored sales data of Pharma is open to data mining researchers for further investigation, please contact us if you are interested in joining us.

References

Agrawal, R., Imielinski, T., and Swami, A. 1993. Database mining: A performance perspective. IEEE Transactions on Knowledge and Data Engineering, 5:914–925.

Apte, C., Weiss, S., and Grout, G. 1993. Predicting defects in disk drive manufacturing: a case study in high-dimensional classification. Proceedings of the 9th Conference on Artificial Intelligence for Applications, pp. 212–218.

Carter, C. and Catlett, J. 1987. Assessing credit card applications using machine learning. IEEE Expert, Fall 1987:71–79.

Fukuda, T., Morimoto, Y., Morishita, S., and Tokuyama, T. 1996. Data mining using two-dimensional optimized association rules: scheme, algorithms, and visualization. Proceedings of the ACM SIGMOD Conference on Management of Data, pp. 13–23.

Hoschka, P. and Kloesgen, W. 1991. A Support System for Interpreting Statistical Data. In Knowledge Discovery in Databases, G. Piatetsky-Shapiro (Ed.). AAAI Press, pp. 325–345.

Manago, M. and Kodratoff, Y. 1991. Induction of decision tree from complex structured data. In Knowledge discovery in Databases, G. Piatetsky-Shapiro (Ed.). AAAI Press. pp. 289–306.

Matheus, C.J., Chan, P.K., and Piatetsky-Shapiro, G. 1993. Systems for knowledge discovery in databases, IEEE Transactions on Knowledge and Data Engineering, 5(6):903–913.

Piatetsky-Shapiro, G. (Ed.). 1991. Knowledge Discovery in Databases, AAAI Press.

Schmitz, J., Armstrong, G., and Little, J.D.C. 1990. CoverStory-automated news finding in marketing, DSS Transactions, Institute of Management Sciences, Providence, RI.

Uthurusamy, R., Fayyad, U.M., and Spangler, S. 1991. Learning useful rules from inconclusive data. In Knowledge Discovery in Databases, G. Piatetsky-Shapiro (Ed.). AAAI Press, pp. 141–157.

Yada, K. 1996. Strategic information systems and organizational capabilities—The examination of Pharma's business system (in Japanese), The Seiryodai Ronshu, 28(3):49–82. Kobe University of Commerce.

Yukinobu Hamuro received the M.A. degree from Kobe University of Commerce, Kobe, Japan, in 1991. He is currently an Assistant Professor in the Department of Business Administration, Osaka Industrial University, Osaka, Japan. His research interests include data mining in business, high performance computing in large text databases, and decision support systems.

Naoki Katoh received the B.E., M.E., and Ph.D. degrees from Kyoto University, Kyoto, Japan, in 1973, 1975 and 1981, respectively. From 1977 to 1981, he was with Department of Information System of the Center for Adult Diseases of Osaka, Osaka. From 1981 to 1997, he was with the Department of Management Science, Kobe University of Commerce, Kobe, Japan, where he was a Professor from 1990. He is currently a Professor in the Department of Architecture and Architectural Systems, Kyoto University, Kyoto, Japan. His research interests include combinatorial optimization algorithms, computational geometry, architectural information systems, and data mining. Dr. Katoh is a member of the Association for Computing Machinery, IEEE Computer Society, the Operations Research Society of Japan, Information Processing Society of Japan, the Japan Society for Industrial and Applied Mathematics, and Architectural Institute of Japan.

Yasuyuki Matsuda received the M.S. degree in Pharmaceutical Sciences from Osaka University, Osaka, Japan, in 1973. He was a founder of G&G Pharma Inc., Osaka, Japan. From 1982 to 1984, he was the president of G&G Pharma Inc. He is currently an adviser in Zett Inc., Osaka, Japan. His research interests are in all aspects of an effective use of information systems in business.

Katsutoshi Yada received the B.A. degree from Fukui University, Fukui, Japan, in 1992 and the M.A. from Kobe University of Commerce, Kobe, Japan, in 1994. He is currently an Assistant Professor in the Department of Business Administration, Osaka Industrial University, Osaka, Japan. His research interests focus on information strategy concerning with data mining and effects on organizations by information technology. Yada is a member of IEEE Computer Society, Organizational Science Society of Japan, Japan Business Management Society, and Japan Society for Management Information.