RCSSディスカッションペーパーシリーズ
第8号　2003年7月

Discussion Paper Series
No. 8 July, 2003

ISSN 1347-636X

Why Does the Data Mining Project Fail?

— Data Mining Lessons Learned from the Cases in the Business Field —

Katsutoshi Yada

RCSS

文部科学省私立大学学術フロンティア推進拠点
関西大学ソシオネットワーク戦略研究センター

Research Center of Socionetwork Strategies,
The Institute of Economic and Political Studies,
Kansai University
Suita, Osaka 564-8680 Japan
URL : http://www.rcss.kansai-u.ac.jp/
e-mail : keiseiken@jm.kansai-u.ac.jp
tel. 06-6368-1177
fax. 06-6330-3304

Why Does the Data Mining Project Fail?

— Data Mining Lessons Learned from the Cases in the Business Field —

Katsutoshi Yada

RCSS

文部科学省私立大学学術フロンティア推進拠点
関西大学ソシオネットワーク戦略研究センター

Research Center of Socionetwork Strategies,
The Institute of Economic and Political Studies,
Kansai University
Suita, Osaka 564-8680 Japan
URL : http://www.rcss.kansai-u.ac.jp/
e-mail : keiseiken@jm.kansai-u.ac.jp
tel. 06-6368-1177
fax. 06-6330-3304

Why Does the Data Mining Project Fail?

— Data Mining Lessons Learned from the Cases in the Business Field —

Katsutoshi Yada

Associate Professor, Dr.
Faculty of Commerce
Research Fellow of RCSS
The Institute of Economic and Political Strategy
Kansai University,
Suita, Osaka, Japan

Abstract

Some of the research studies give discussions on interpretability and user-oriented tools in the data mining and these are very useful in achieving the progress of the data mining in the business field. However, these research studies do not pay much attention to organizational process in the data mining project. In this paper, we attempt to discuss and elucidate, from the organizational viewpoint, why and how the project to introduce the data mining would fail in business field. The project for data mining includes very complicated organizing process, and its success often depends much on the effectiveness of the project management. Here, we will focus the discussion on the project management with successful and unsuccessful results and will evaluate the role of the project manager. In the successful data mining project as we have seen so far, we have found that the manager was well informed about technical and managerial problems to manage the project and was fully committed to it with the support from his corporation, which has ardent strategic intent in introducing the new information technology.

Key Words: data mining, failure, data mining lessons, dual triangle of dependences, innovation process approach, concurrent discovery process

1. Introduction

In recent years, attention has been focused on the data mining to discover useful knowledge from large-scale database in business field (Piatetsky-Shapiro 1991; Spangler 1999; Sung 1999; Ip 2002), and many Japanese corporations are now introducing or are planning to introduce the data mining. However, many of such corporations have failed in introducing the data mining and wasted a large amount of money for system investment (Hamuro 2002).

Since several years, we have carried out application study on the data mining in business field (Hamuro 1998; 2000; 2001; Yada 1998). Fortunately, during our industry-university cooperative research, we have often heard and experienced the cases of failure and the cases of success in introducing the data mining. As a result, we have learned many lessons from our experience. In particular, we have obtained many valuable suggestions, which might have been unavailable otherwise because it is customary that, when corporations fail, the cases of such corporate failure are not made public in most cases. What should we learn from these cases of failure?

Just like the cases of the existing studies, we have been discussing strategies to effectively introduce the data mining from both human and technical aspects (Yada

2002). We have made detailed studies on the methodology to extract the rules easily interpretable for the experts and on the effective method to create efficient business action (Hamuro2000).

However, the existing research studies have shown almost no interest in the project organization, which is organized to introduce the data mining into business despite of the fact that the failure in the development of the information system is brought out from the organizing process of the project. For the purpose of effectively introducing the data mining in the business field in future, we consider that important lessons may be learned by clearly defining the failure of data mining project and by elucidating the causes of such failure.

In this respect, we will analyze the organizing process of several cases based on innovation study and will discuss how data mining and system development should be introduced. As a result, we assert that the data mining project can achieve its goal through very complicated organizing process, and high-grade project management is required for the success. The discussions of the cases in this article are primarily concentrated on the projects to analyze sales data, while the lessons learned through discussion would be applicable for the cases in more extensive range.

The organization discussed in this paper is as follows: First, we intend to clearly define the failure of the data mining project to be discussed. Next, we will introduce some actual cases of failure or success cases. Finally, we will present the lessons learned from the cases based on innovation study.

2. Failure of the Data Mining Project

There are very few corporations, which are now trying to incorporate the data mining in daily business routine and to assign personnel to the sections or the department for such purpose. In many cases, it is incorporated in form of a temporary project. Data mining project (DMP) can be defined as a project to develop organizational information system, which includes daily business activities, organizational innovation, and system development for applying the data mining to the matters in business field. Before we discuss the reasons why DMP failed, we should clearly define the failure itself by clarifying specific circumstances and results, which characterize the failure of DMP. In this chapter, the failure of DMP will be defined and we will make it clear how and under what kind of organizing process such failure occurs.

2.1 Conventional studies on the failure of information system

Starting from the famous article of Ackoff titled "Management Misinformation Systems" (1967), a number of studies have appeared so far, describing the failure of information system. In the theory of failure of information system, there have been many discussions on the personal factors such as inability of the personnel (Dearden 1972) in charge of the information system, interaction between the user and the developer, or social and behavioral factors including organizational character and strategy of the corporation, which engages in the development of information system (Argyris 1971). According to Lyytinen & Hirschheim (1987), the discussions on the failure of information system can be understood in the frameworks as given below.

The first of the discussions on failure relates to "correspondence failure". This is a failure, which is characterized in that the developed information system does not satisfy the preset goal and the requirements. The second is "process failure", and it is a failure to build up the system within the limitation of time and cost. As a result, the project plan is suspended or abandoned. The third is "interaction failure". In this case, users are not satisfied well, and the system is utilized less frequently. The system often remains unused or it is used only partially. However, these concepts on the failure are simply limitative definitions relating to the failure, and these are not suitable as the concepts to explain the actual cases.

Under such circumstances, Lyytinen & Hirschheim (1987) found the cases where the information system could not cope with the expectation of a specific circle concerned, and they defined this as "expectation failure" and tried to elucidate the failure of information system in very wide range. It appears that the concept of expectation failure may have validity from theoretical viewpoint, while this does not seem to be appropriate definition of the failure because expectations of all persons concerned are discussed on the same and equal level and that no special consideration is given on organizational intent.

2.2 Terminal failure and triangle model of dependences

In this respect, Sauer (1993) extended the concept of the expectation failure and proposed to use the term "terminal failure". The terminal failure is defined as follows: When the supporters or the users are not satisfied, the support from them for the development of information system is often suspended. As a result, development, maintenance, operation, etc. of the information system are abandoned or brought to an end. Such definition has its premise on the development process of the information system called a model of "triangle of dependences".

The development of information system is initiated from corporate strategy and idea, and the development itself is executed by project organization. Because the information system is designed under the ever-changing environment, there is actually no information system, which has no defect. In order to correct the defect, the project organization carries out the development, while this involves the problem of cost. To pay the cost, the project organization expects and receives support from the supporters, who would enjoy benefit from the development information system. After evaluating the conveniences and the benefits as well as the possibilities of receiving such benefits to be obtained from the information system in future, the supporters would make decision to support. When this "triangle of dependences" is maintained, the development of the information system is guaranteed. If there is a problem in a part of the relation of this triangle, the development of the information system may fail. That is, if there is no more support from the supporters and the project organization is not able any more to develop the information system, the development is suspended or abandoned.

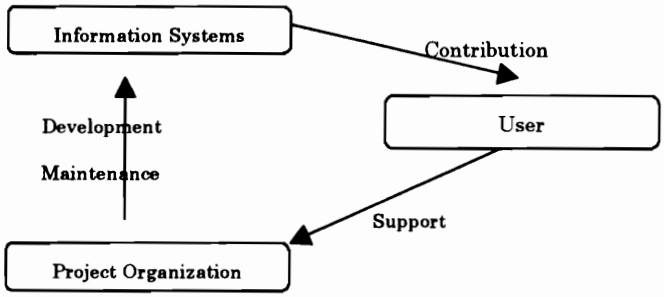

Fig. 1 Triangle of dependences

2.3 Failure of data mining project

When a corporation plans to introduce the data mining, a temporary project organization is set up in most cases. If the project were reviewed in detail, it would be possible to explain by the model of the triangle of dependences as described above. In the cases to be discussed in the subsequent chapter, a sales promotion staff of a retailer provided the support to introduce the data mining technique to the project organization in the expectation that sales proceeds would increase. The project organization gives conveniences and benefits to the supporters through the development of the data mining system. Therefore, the failure of the data mining project can be defined as follows: When the supporters are not satisfied with the conveniences and the benefits given to them from the development of the system or from the results of future

development, the project organization may not have the support from the supporter, and the system development may be suspended or abandoned.

For the introduction of data mining in actual business activities, this definition is more valid than the correspondence failure or the interaction failure as described above. The knowledge learned from the data mining cannot be predicted in advance, and the final result often differs from the purpose designed in the initial stage of the development. It also happens very often that important knowledge is discovered in the middle of the process of analysis, and project organization is re-organized along corporate strategy. Then, changes may occur in the framework of interaction between the system and the persons. These should not be regarded as failure but should be considered in the middle of the system development. In the present article, we regard the following condition as the failure of DMP: a condition where there is no more support for the development of data mining system and the project organization is suspended or abandoned.

3. The Case Study of Data Mining Projects

In the process of our cooperative research study on the data mining with many corporations, we have had chances to become aware of the cases of failure of DMP, which these corporations had previously experienced with the parties other than us. Also, among the cases of DMP, with which we were involved in the past, there were several cases, which ended regrettably in failure. From these valuable experiences, we are trying to learn lessons on the management of DMP. In the present article, we adopted a case study approach. By introducing and analyzing two cases of failure and one case of success, we will clarify the causes of DMP failure.

3.1 Failure Case 1 of DMP in a food manufacturer A

In spring, 2001, a corporation A (a food manufacturer) initiated the data mining project with the purpose of efficiently supporting the sales department of its own. The initial motivation was that the leader of the sale department was interested in the proposal and the presentation of a famous DM software company B. In the corporation A, an information system had been set up, by which the information such as quantities of commodities sold to the customer in each region as well as customer information and regional information were to be sent from each of the sales staffs assigned in the stores located nationwide to the head office. The current project was based on such an objective that the present sales operation had to be carried out more efficiently by

utilizing these data.

The corporation A concluded an agreement to purchase DM software with the company B, and a project to build up DM system was initiated to support the sales activities on the entire corporation level. The project was set up on temporary basis, and it was organized with 2 engineers from DM software company, a leader of the sale department, 3 sales staffs, and a leader and a staff from the system department of the corporation A. In all, 8 persons were participated in the project.

In the initial program of the project, data of a certain region had to be analyzed to set up the proto-type. Then, a system was to be set up on the entire corporation level according to the results. First, the system department was engaged in data extraction. During this period, the company B engaged in the analysis tried to collect the information at site. The system department subcontracted the data extraction to a company C, to which operation and maintenance of data warehouse had been entrusted in the past. When the raw data thus extracted was provided to the company B, the explanation of the data was given from the system department. This procedure required a period of about 3 months. Most of the time was needed for the preparation of the program to execute data extraction by the system department and the company C.

During the 3-month period, the company B performed analysis in various steps by using DM software. In this period, the company B exchanged the information relating to the data with the system department. Each time an inquiry was given to the system department, the investigation on the inquiry had to be commissioned to the company C, and vast amount of time was consumed for this purpose. However, the results of the analysis offered from the company B merely indicated general trends such as transition of sales proceeds, and no important information could be obtained. Later, similar procedure of analysis was repeated. When one year elapsed since the initiation of the project, the corporation A finally decided the suspension of DMP, and the investment of more than 20 million yen had been wasted for the project by that time.

3.2 Failure Case 2 of DMP in a manufacturer D (a daily necessaries manufacturer)

Several years ago, a company D, which manufactures daily necessaries such as sanitary goods, paper diapers, etc., started the analysis of sales data upon request from its customer company E (a drugstore chain). The data to be handled in the project was FSP data of the company E, which was recognizable by the customers, and it was attempted to discover new findings from the data and to carry out sales promotion more efficiently. For the introduction of the data mining, we decided to participate in the

project because we were well experienced in the analysis of sales data. From the company D, 2 marketing staffs and 2 sales staffs participated in the project. A system personnel of the company E, and our staffs also joined, and an industry-university cooperative project was initiated.

In the past, we had been engaged in the analysis of FSP data for many cases including drugstores and supermarkets. Regarding the contents of the data, we could reach full understanding within considerably short time by simply hearing and learning the routine procedure of the stores. However, the data extraction by the system personnel of the company E required nearly one and half months although the necessary data had been strictly selected in advance and the program and the procedure were very simple in nature. Similarly to the case as mentioned above, the company E entrusted a subcontractor with the maintenance of data warehouse, and much time was consumed for the giving and taking of information to and from the subcontractor.

As a subject of analysis, which would be advantageous and beneficial to both the manufacturer and the drugstore chain, we proposed to take an approach for the result of analysis named "customized coupon", i.e. a procedure to offer a different coupon to each different customer. In the past, the coupons used to be indiscriminately distributed to all customers, who visit the stores, and the system itself was very inefficient. POS register at the company E had been designed in such manner that, whenever FSP card of a specific customer was scanned over, an instruction to sales staff was given on the display of POS register. It was envisioned that more efficient sales promotion than in conventional type would be carried out by coupon distribution if this function of POS register and the idea of the customized coupon can be used in combination.

Three months after the initiation of the project, when we reported the results of analysis at the project meeting, the marketing staff, the sales staffs of the company D and the system staffs of the company E showed extreme interest in the results of analysis and they highly evaluated the usefulness of this approach. As the result of simulation study, extensive cost reduction for sales promotion activities was foreseen, and we placed much expectation on the execution of the project. However, the sales promotion approach as we proposed for the results of analysis was never put into practice. According to the report of the marketing staffs of the company D to us, the sales staffs in each of their regional stores had already set up annual sales promotion plan, and it was difficult to carry out new sales promotion activities by sudden change. Ironically, soon after we presented a study article describing this approach on analysis results, a mail-order firm using WEB succeeded to have more efficient sales promotion

activities by utilizing our approach on analysis results.

3.3 Success Case of DMP in a food manufacturer F

In 2002, we started an industry-university cooperative project with a food manufacturer F and a supermarket chain G to develop a new sales promotion approach based on the use of FSP data. The purpose of the project was to achieve the sales increase by proposing new types of stores based on customer sales data when the company F wanted to expand the existing line of commodities in the summer season.

We had a regular meeting once a week, and all members exchanged the information on the project with each other. One month was required for data extraction, but basic analysis was reported immediately thereafter. We could have deeper understanding on routine activities of the stores of the company G, and detailed consumer behavior analysis was carried out. After we performed analysis of various categories of customers by three times, the marketing staff of the manufacture F exhibited specific interest in the analysis of the associated purchase of the commodities of the company F by a group of excellent customers. According to the result of basket analysis using the association rule, a certain trend was found that the products of the company F had been purchased very often together with fresh vegetables among the excellent customers. Because this was consistent with his specialized knowledge, the market staff made a sales promotion proposal to carry out the associated display (to display the associated products adjacent to each other) for the fresh vegetables based on the data. In close cooperation with the sales staff, test marketing was carried out in the company G.

As the result of the above experiment, new customers have been successfully acquired, and the commodities of the company F showed extensive sales increase compared with the stores not participated in the experiment. Using this experiment result as proto-type, the marketing staff of the company G proposed the sales promotion approach in the supermarkets on nationwide scale, and similar sales promotion activities were developed in many supermarkets. There were 4 existing commodities and only one new commodity was added. Despite of this fact, the company F achieved sales increase of 50% more compared with the records in the previous year in the stores located nationwide. Since that time, the company has been continuously engaging in the data mining projects in corporation with us.

4. Innovation process approach to DMP management

In this chapter, we will discuss why DMP failed, and the lessons learned for future DMP management will be presented. We assert that DMP should be carried out through complicated organizing process and that the organization structure can be expressed in the form of "dual triangle of dependences". From the viewpoint of innovation process, important lessons necessary for the management of the complicated organizing process will be derived. We conclude that a heavy-weight class project manager is needed, who has high and rigorous sense of responsibility and authorization on the management of DMP, which is highly uncertain, and who exhibits superb leadership to promote the entire project.

4.1 "Dual triangle of dependences" in DMP

In actual business field, DMP is not so simple as it is assumed in the existing studies on the failure of the information system as described above. It can be established only through very complicated organizing process. To express the complexity of the process, we present a framework named "dual triangle of dependences" as shown in Fig. 2.

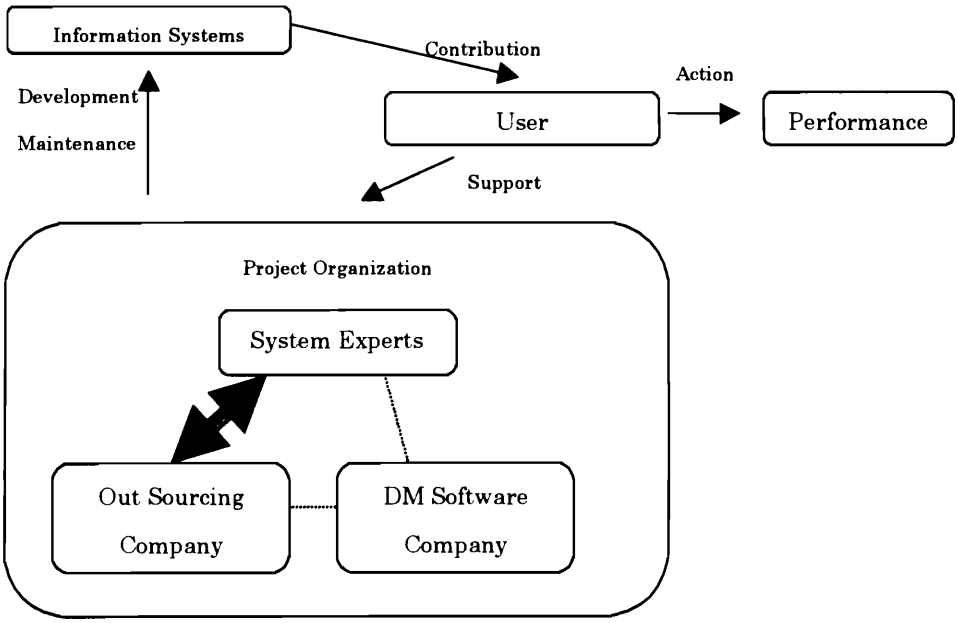

Fig. 2 Dual triangle of dependences

What clearly distinguishes the "triangle of dependences" as described above from our assertion is the part of the project organization. As it is evident from the case, the organization of DMP comprises three members, each of whom has a background

different from that of the other. The first is a personnel in charge of the system or a system expert, who assumes responsibility on the system to extract the data handled by the project. The second is a subcontractor. In many large corporations, the system department does not necessarily develop all information systems by itself. They usually entrust the work with other subcontractors, i.e. software developing firms. These subcontractors are engaged in construction and maintenance of data warehouse and offer services for efficient data extraction. Also, it is rather rare that the system department of a corporation originally has data mining technique. In almost all cases, the technique is introduced from outsiders. The third is a software firm, which has its own DM technique. In the latter two cases, the researchers working in universities are involved.

DMP comprises the members who have different backgrounds as given above. Unless all members cooperate organically with each other, it is not possible to successfully accomplish DMP. However, the degrees of intensity of the relationship between these members are different in each case, and it is difficult to attain the aimed goal as a single organization. The system staffs and the subcontractor have close relationship because the former engage in daily work in cooperation with the subcontractor. On the other hand, DM software company participates in the project only on temporary basis, and they do not have sufficient communication with the system staffs in most cases. Also, they have almost no close contact with the subcontractor, which maintains the data warehouse. In the corporation A of the Failure Case 1, DM software company did not have sufficient communication with the system staffs or the subcontractor and could not completely identify and grasp the data and the business rules. As a result, they could not discover new knowledge. As it is suggested by Yada (2002), in order to discover new knowledge from the data, the analyzer must have full understanding of the work, from which the data is derived.

Further, in case the introduction of the new technique may cause some changes in the existing power relationship, organizational resistance to the project may be induced (Markus & Pfeffer, 1983). In the corporation A of the Failure Case 1, the subcontractor thought that their work might be usurped when the data mining technique would be introduced, and it was found that the subcontractor intentionally delayed the work of data extraction on their part. In DMP, such power relationship should be controlled well, and adequate cooperation must be provided from the subcontractor.

These facts are generally well known to us. In the company D of the Failure Case 2, we provided sufficient information to the system staffs and the subcontractor

and thoroughly persuaded them that we would give no menace to them. We learned routine store work from them and could fully understand the data. In fact, we could discover new knowledge from the data. However, the idea thus produced was not practically executed. Why did this happen?

4.2 DMP as seen from innovation process approach

In this respect, we will evaluate the organizing process of DMP based on the innovation study. Fig. 3 shows a process of knowledge discovery in the existing research (Fayyad, 1996). In these models, it has been assumed that it might go back to the preceding step in the middle of the process. In the process of knowledge discovery, however, each step is highly independent and a sequential development process is assumed, in which the next (previous) step begins as soon as one step is completed. In the innovation study, such development process is called "sequential approach".

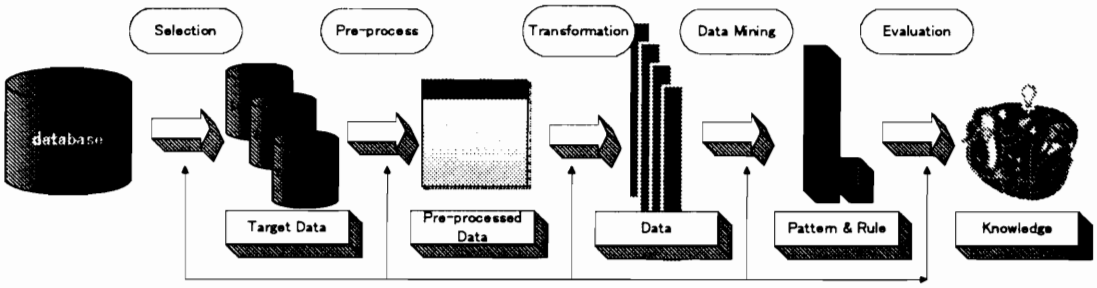

Fig. 3 Knowledge discovery process in the existing studies

In the management study, sequential approach has a certain definite purpose, and it is considered as more effective when the technique is less complicated (Clark 1991). In DMP, it is impossible to predict the results of analysis in advance, and even the purpose may be changed in the analytical process. Compared with the existing system, data mining technique is much more complicated. It is also based on high uncertainty, and algorithm may vary if the purpose is changed. In the success cases as described above, it was not a sequential development process as found in the knowledge discovery process in the cooperative project. The actual process is a concurrent type process as shown in Fig. 4. That is, all of the raw data are deposited at the university and all members are participating in the system development and analysis process. As a result, all processes including data selection, pre-processing, data transformation, data mining, etc. are carried out at the same time and in parallel.

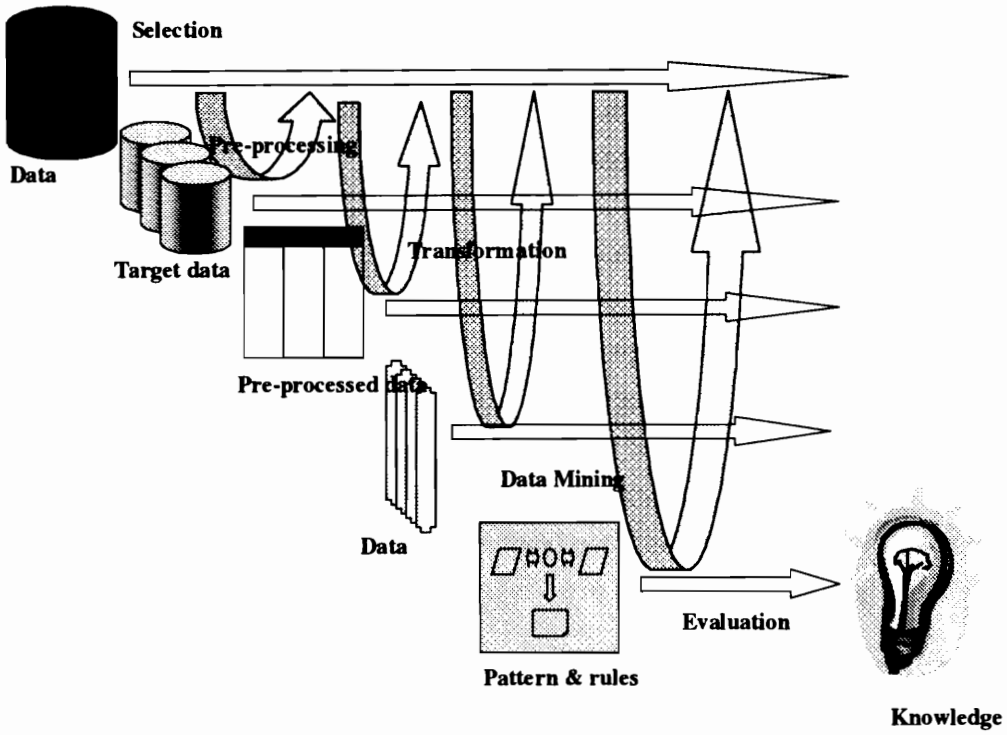

Fig. 4 Concurrent discovery process

In the Company F as we have seen in the Success Case, the project members gathered together in the study room in university or at the meeting in the company and discussed the extracted rules. Various types of algorithms were tried at the site, and the results were immediately evaluated by the members. To reinforce the interpretation of the rules, analysis of the associated purchase in each time zone and analysis on customer attributes of the associated purchase were additionally performed. For this purpose, the raw data were re-processed. All works were carried out at the same time and in parallel, and this was very similar to concurrent engineering, which is widely known in the study on the development of automobile. In this way, DMP is executed by very complicated organizing process.

4.3 What are the lessons learned?

Based on the concept of the heavy-weight class project manager (HWPM) in the innovation study of the management research, we have learned the importance of the project management. HWPM means project leaders with high responsibility and authorization, who are responsible for attaining the unification between various functional departments within the organization in an environment where the technique is highly complicated and customer requirements are indefinite and uncertain. They

must have not only the knowledge for the adjustment of internal organization but also technical knowledge and should be able to exert strong influence on the entire process of the project management. We have learned that the presence of the heavy-weight class project managers with the above characteristics is an important key to success in DMP.

4.3.1 Technical skill and management skill

Manager(s) (one or two managers) of DMP must have both the technical skill to handle large-scale data and the management skill to interpret and to use the extracted data.

In the company F of the Success Case, the project manager was a marketing staff, who had only general skill and knowledge in computer technique such as Excel, Word, etc. before the project was initiated. After the initiation of the project, he learned from us how to handle database, basic knowledge for analysis, data mining algorithm, etc. and has acquired technical skill. Therefore, at a glance of the association rule extracted by us, he could discover business chance, which was potentially contained in the rules. As seen in the corporation A of the Failure Case, when a staff having only technical skill engages in the analysis, it would be difficult to discover new knowledge, with which effective business action can be executed.

4.3.2 Commitment of the project manager

The project manager of DMP must have full responsibility and authorization and must commit himself in the execution of the new business action based on the data after elaborately persuading the persons concerned.

The project manager of the company D committed himself to his original work and to DMP. To him, DMP was a surplus work. In order to execute new sales promotion approach, he had to persuade all of financial department, sales staffs and retailers by himself. He knew that he would receive high evaluation if he simply performs his original proper job, which would be easier to carry out than troublesome DMP work. Therefore, he dared not positively engage in the new business action in accordance with the knowledge acquired by DMP.

The project manager of the company F entrusted his original proper job to his colleagues and was working on data analysis and sales promotion program every day until midnight. In actual step of business action, he had meetings repeatedly with the relevant sales staffs and the retailers and tried to persuade them about the importance and the profitability of the project. They trusted each other and they were functioning together as a team, and its core factor was the commitment of the leader.

4.3.3 Strategic intent

Strategic intent of the corporation provides supports to the commitment of the

project manager.

When DMP is started from an idea of a certain individual, it cannot gain full support from the corporation, and it fails in almost all cases. In the company F, customer-oriented marketing campaign was developed, and the analysis of customer data was a part of corporate strategy. Accordingly, the company F supported the project manager in entrusting his original job to the other colleagues and in providing the information to sales staffs in advance. When corporate strategy concurs with the purpose of DMP, the project manager would be able to acquire full support from the company and to lead the project to success.

5. Discussion

In the above, we have discussed the lessons, which could be learned from the knowledge discovery based on management data, and in particular, on sales data from the viewpoint of project management. In the past, it has been known that human factor plays an important role. In the present article, we discussed that DMP management is a key to success in the discovery of useful knowledge. Based on organization theory and innovation theory, we placed emphasis on the role of the project manager as a leader and clarified the meaning and the importance of the project manager. We assert that the discovery process can be made more efficient – not only by putting the importance on human role but also by establishing the management in the project organization.

References

Ackoff, R. L., "Management Misinformation systems, Management Science, Vol.14, No.4, 1967.

Argyris, C., Management Information Systems: the challenge to rationality and emotionality, Management Science, Vol.17, No.6, pp.275-292, 1971

Chung, H. M., and Gray, P. " Special Section: Data Mining," Journal of Management Information System (16:1), Summer 1999, pp. 11-16.

Clark, K. B. and Fujimoto, T., Product Development Performance: Strategym Organization, and Management in the World Auto Industry, Harvard Business School Press, 1991.

Dearden, J., "MIS is a mirage," Harvard Business Review, Jan-Feb, pp.90-99, 1972.

Fayyad, U., Piatetsky-Shapiro, G., and Smyth, P. "From Data Mining to Knowledge Discovery in Databases," AI Magazine 17, pp.1-34, 1996.

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

Hamuro, Y. Katoh, N. and Yada, K. "Discovering Interpretable Rules that Explain Customers' Brand Choice Behavior," Lecture Notes in Artificial Intelligence 1967, Proc. Third International Conference DS'2000, pp.263-267, 2000.

Hamuro, Y., Kawata, H., Katoh, N. and Yada, K. (2001)"A Machine Learning Algorithm for Analyzing String Patterns Helps to Discover Simple and Interpretable Business Rules from Purchase History," Progresses in Discovery Science, State-of-the-Art Surveys, LNCS, Springer-Verlag, pp.565-575.

Hamuro, Y., Katoh, N., Yada, K., "MUSASHI: Flexible and Efficient Data Preprocessing Tool for KDD on XML", Proc. of first international workshop on data cleaning and preprocessing, pp.38-49, 2002.

Lucas, H. C. Jr., Why Information systems Fail, Columbia University Press, New York, 1975.

E. Ip, J. Johnson, K. Yada, Y. Hamuro, N. Katoh and S. Cheung, "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, pp.55-69, 2002.

Lyytinen, K. and Hirschheim, R., Information Systems Failure: A Survey and Classification of the Empirical Literature, Oxford Surveys in Information Technology, Vol.4, pp.257-309, 1987.

Markus, M. L. and Pfeffer, J., Power and the Design and Implementation of Accounting and Control Systems, Organizations and Society, Vol.8, pp.205-218, 1983.

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

Sauer, C., Why Information Systems Fail: A Case Study Approach, Alfred Waller Ltd., 1993.

Spangler, W. E., May, J. H., and Vargas, L. G. "Choosing Data-mining Methods for Multiple Classification: Representational and Performance Measurement Implications for Decision Support," Journal of Management Information System (16:1), Summer 1999, pp. 37-62.

Yada, K., Katoh, N., Hamuro, Y., and Matsuda, Y., "Customer Profiling Makes Profits: How did a Japanese firm achieve competitive advantage through the knowledge creation?" Proceedings of The Practical Application of Knowledge Management 98, The Practical Application Company, pp.57-66, March 1998.

K. Yada, "The Future Direction of Active Mining in the Business World," Frontiers in Artificial Intelligence and Applications, Vol.79, IOS Press, pp.239-245, 2002.