A Study on the Construction of Knowledge Base in a Project Management System by Using SOM

KyungBae Yoon[1], JunHyeog Choi[2], ChangJong Wang[3]

[1]Dept. of Computer Science, Kimpo College,
San 14-1, Wolgot-Myun, Gimpo-Si, Gyonggi-Do, Rep. of Korea
Tel. +82-31-999-4202, Fax.: +82-31-989-4237
e-mail : kbyoon@kimpo.ac.kr
[2]Dept. of Computer Science, Kimpo College,
San 14-1, Wolgot-Myun, Gimpo-Si, Gyonggi-Do, Republic of Korea
Tel. +82-31-999-4164, Fax.: +82-31-989-4387
e-mail : jhchoi@kimpo.ac.kr
[3]Dept. of Computer Science & Engineering, Inha University,
253, YongHyun-Dong, Nam-Gu, Inchon, Rep. of Korea
Tel. +82-32-860-7441, Fax.: +82-32-860-7440
e-mail : cjwangse@dragon.inha.ac.kr

Abstract: Recent explosive increases in information volume have led to a rapid development or a change of information technology which stores, searches, and manages a vast amount of information. It is considered that an effective share and utilization of a large amount of digital information produced by work performances is a pivotal element which can make decisive contributions to a great success of business management. This common property of information reflects a changing social paradigm including a change of business processes. This paper is aimed at designing and embodying the construction of knowledge base in an efficient project management system using unsupervised data mining techniques in order to extract information and utilize it as knowledge about standard data (statistical data, template etc.,), size prediction and a danger precaution notice which are needed for a plan and a scheduling of a new project from data coming from already-established projects.

1. Introduction

The efficient management of a project depends on how to plan processes of a project, how to predict various problems occurring in the future, and how to use practically past experiences to expected solutions. However, an occurrence of a new project usually puts much restrictions on extracting a product of already-established similar projects. Therefore, there are some limitations in management or prediction for a new project to be carried out exactly and appropriately on the basis of past experiences.

The existing project management system brings about a problem of investing too much time and cost but low precision in making a plan, which exceeds the time and cost limit and lowers the re-use ratio of a common module unless past experiences are accumulated. So, a shortage of similar cases makes it difficult to solve some problems of a project.

This study aims to design and embody the construction of knowledge base in an efficient project management system using an autonomous nerve network so as to extract and utilize information data about standard data (statistical data, template etc.,), size prediction and a danger precaution notice needed for a plan and a scheduling of a new project from data coming from already-established projects••

2. Construction of Knowledge Base Using Unsupervised Data Mining Technique

With the experiences of numerous projects in the past being assisted, the knowledge base of past project cases should be constructed and Case Base Reasoning (CBR) should be made in order to search for the rules which are able to be applicable to a new project.

To solve the problems caused by a new project, a search of cases similar to past projects is to made and its data have to be applied. In other words, through knowledge base constructed for a new project, similar project cases are retrieved and some problems of a new project are reused. The reused project is added to knowledge base and revise an existing knowledge base. At last, the entire knowledge base is well retained and an occurrence of a new project will be completely dealt with.

To construct an excellent knowledge base, appropriate ways to construct project cases into data base, to retrieve similar project cases from data base, to apply similar past projects to a new project, and to attain the first project case should be taken into consideration.

In this study, 27 variables representing the characteristics of a project are used. After existing project cases are grouped together according to their pattern, the knowledge base will be constructed on the basis of these.

In ordinary knowledge base, all past projects are retrieved, grouped together, integrated according to their pattern, and they are finally applied to a new project. However, in this study, past project cases are grouped according to their similarity by SOM, an autonomous data mining technique and they are applied to retrieval of a new project. This project management is able to solve the problems caused by a present project with speed and accuracy. Case collection for the construction of the early knowledge base was made with simulated use of statistical probability distribution.

Table 1 shows input variables used in a model of SOM for projects grouping in input elements of an autonomous data mining for construction, retrieval, maintenance, and repair of knowledge base for a project management.

The project classification variables used in table 1 for the actual management of knowledge base bring a previous

normalization into practice and this normalized variables are used as an input variable of SOM.

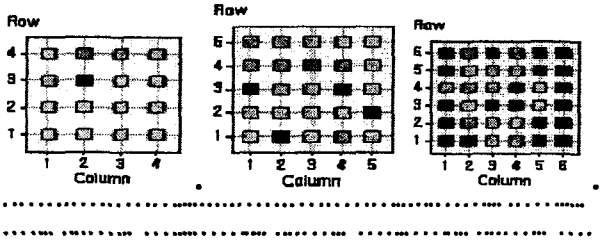
The knowledge base constructed by an existing project result is put into practical use for an optimal layout of a new project, when the early knowledge base construction is very important. But numerous cases are to be collected for the first construction of knowledge base which this study produced through a simulated experiment and found an analogous project pattern. To accomplish this, an autonomous data mining model is used and a simulated data output for each input variable of the model is earned by assigning probability distribution to each category with use of statistical discrete distribution. 1,010 project cases out of 27 variables are collected through a statistically simulated experiment.

3. Experiment and its Result

3.1 Experiment and its Result of Knowledge Base Construction for Project Management

Based on classification variables of 27 project cases, a project pattern is made through an autonomous data mining technique, SOM algorism. The initial weight value of feature maps produce and assign a column number from uniform distribution of 0 to 1. The early learning rate is 0.2 and is decreased in proportion to the times of study advance while the study in progress. This is intended to stabilize the width of renewal as a model is learned. The neighborhood size of a winner node includes all nodes in the entire feature maps by neighborhood function at an early stage, while a learning strategy of decreasing its size is taken as a study advances. This means that renewal of weight value begins to attract learning globally and widely at an early stage. But, when approaching weight value of a permissible range, a model is stabilized by decreasing the alteration range of renewal.

The learning is progressed by dividing the feature maps into 4*4, 5*5, and 6*6. Figures 1 through 3 represent SOM's grouping result with feature maps of 4*4, 5*5, and 6*6, respectively. Here, if the dimension of feature maps becomes larger, the number of groups forming from the result of learning is increased. The dark colored section in the figures represents a group with concentrated data, which is used for deciding the number of final groups and keeps on carrying out its groupings. The final grouping result put together by 3 figures is shown in Table 2.



As for the result of the above figures, the darker a color of each node is, the more the number of cases of each node is. The result shows that the number of final groups is increased in proportion to the size of figure maps.

Two groups are formed in case of 4*4 dimensional size of figure map. In addition, Four groups are formed in case of 5*5 dimensional size of figure map, and eight groups are formed in case of 6*6 dimensional size of figure map. The number of these groups becomes the number of project groups to be constructed as a pattern from project knowledge base. Experts can decide among groups number according to dimensional size of figure maps to apply in his or her business areas.

This grouping result is considered as a strategy to construct project pattern of knowledge base, but two groups are not thought to be proper division for a project management on account of a small number of patterned rules. Also, 8 project groups have so much detailed patterns that they give a difficulty to an effective knowledge base

management. To solve this, 4 groups pattern is used ultimately for an actual management and a rule and the dimension of figure maps is chosen as 5*5. At last, each project is assigned to 4 groups. The result is shown in Table 4.

To define each group, means of it about variables and total standard deviation of each variable are used in Table 5 and 6.

If a new project happens, the project management is accomplished in synthetic consideration of cases of project knowledge after finding the most similar cases on the basis of means and distribution of each variable using 27 project classification variables (X_1,...,X_{27}). The cases of success or failure stemming from a project performance result will be all used again for renewal of knowledge base. Through constant management of this constructed knowledge base, a new project performance and management will be made repeatedly by using the characteristics of the most similar project groups (4 groups in this study) to a newly-generating project.

4.2 Evaluation of Efficiency

The project-type Table 7 shows an efficiency evaluation testing of an existing method and a proposed one on a new project management. 27 classified items are used as input variables of 1 through 27 in Table 7.

First of all, a test to select the most similar project was done from the knowledge base of already-generated project cases constructed in section 4.1 about a new project.
The efficiency was evaluated from the viewpoint of accuracy about similar past project cases earned from the result made by a general method to retrieve all cases of knowledge base and a method of an autonomous data mining algorism proposed by this study. A homogeneity test is done to know how securely the cases retrieved by a concrete method of an efficiency comparison provide a new project management with necessary information. Heterogeneous cases are likely to belong to the result retrieved from knowledge base constructed to earn information for a new project management. It is confusing what should be applied to a present project and this information is difficult to use as a general rule for a new project performance. But, the retrieved cases, if homogeneous, will produce information which can be used as an effective rule for a new project management. X^2-test is chosen as a tool for this evaluation of efficiency. Another efficiency evaluation uses standard deviation which measures distribution of finally gathered cases. The larger a standard deviation is, the more heterogeneous it is. The former is an evaluation of efficiency by frequency; the

latter is an evaluation of efficiency by quantity. The existing method and th proposed one retrieved top 20 cases, respectively.

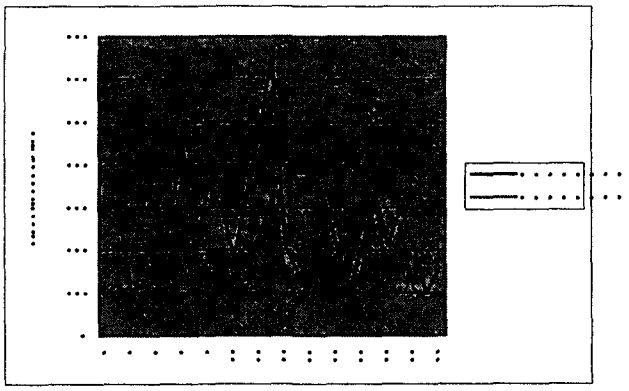
A homogeneity test was done to the top 20 cases retrieved by an existing method and a proposed one. The hypothesis for this test is given below.

$$H_0: C_1 = C_2 = \ldots = C_{20} \text{ vs } H_1: /H_0$$

H_0, null hypothesis, claims that all cases so alike are homogeneous; H_1, alternation hypothesis, says that the heterogeneous is included. The result of homogeneity test on the retrieved 20 project cases is given in Table 8.

The result shows that the homogeneity evaluation of the cases retrieved by a past method abandons H_0, null hypothesis with 95 percent of trust level and 0.02 of p-value. In other words, the heterogeneous project cases make it difficult for a project to get information needed for a stable general management. The result of top 20 project cases retrieved by a proposed method in this study shows that 0.16 of p-value can not abandon H_0 eral management. The result of top 20 project cases retrieved by a proposed method in this study shows that 0.16 of p-value can not abandon H_0, null hypothesis and homogeneity is maintained for the retrieved pojects. Also, through this homogeneous project cases, applicable rules can be found by objective and stable Imformation of a new project management.

Figure 4 is the result of standard deviation comparison over input variables of top 20 project cases retrieved by existing and proposed methods. Here, X axis means each input variable and Y axis means standard deviation. According to this figure, the standard deviation of algorism proposed in this study is smaller than that of a past method.



5. Conclusion and Further Research

This study constructed an effective knowledge base by using past project cases and used this to tackle the problems cased by management of a new project.

The construction of an early knowledge base was made with use of statistical probability distribution. The constructed cases were efficiently classified by an autonomous data mining technique. The retrieval of knowledge base of new project problems uses similarity based on means and standard deviation of each classification variable. This method makes groups of diverse and heterogeneous project cases according to their similarity, which intellectualizes the knowledge base. Also, this has an effect on decrease of retrieval time in which knowledge base of a new project is made. Further researches are scheduled to apply a variety of machine learning algorism including Bayesian learning to achieve construction, maintenance, and repair for more rapid and intelligent project management.

References

[1] Jiawei Han, Micheline Kamber, Data Mining: Concepts and Techniques, Morgan Kaufmann Publishers, 2001.

[2] T. Kohonen, Self-organized formation of topologically correct feature maps, Biological Cybernetics, 43, 59-69, 1982.

[3] T. Kohonen, Self-Organizing and Associative Memory, Springer-Verlag, Berlin, 1984.

[4] T. Kohonen, Self Organizing Maps, Springer, 1997.

[5] Tom M. Mitchell, Machine Learning, McGraw-Hill, 1997.

[6] Giarratano & Riley, Expert System, PWS, 1998.

[7] William J. Kennedy, Jr James E. Gentle, Statistical Computing, Marcel Dekker, INC., 1980.

[8] R.B.Chase & N.J.Aquilano, Production and Operations Management, Irwin, pp.502-504, 1989

[9] Jaeguk Kim et al, Study of Comprehensive Computerization Plan, Electronic Skills, Vol 31, pp. 6-17, 1997.

[10] Sunyong Lee, Theory of Production Management, Bummunsa, Seoul, pp.562, 1989.