Fourth International Conference on Construction in the 21st Century (CITC-IV)
"Accelerating Innovation in Engineering, Management and Technology"
July 11-13, 2007, Gold Coast, Australia

Ming Knowledge in KMS Performance Data - a Case Study of an A/E Consulting Firm

Wen-der Yu
Professor, Chung Hua University, Hsinchu, Taiwan

Cheng-te Lin
Master Student, Department of Civil Engineering, National Taiwan University, Taipei, Taiwan

Cheng-tien Yu
Ph.D. Student, Department of Civil Engineering, National Taiwan University, Taipei, Taiwan

Shen-jung Liu
Manager, Department of Business and Research, China Engineering Consultants Inc., Taipei, Taiwan

Huai-ching Luo
Section Chief, Department of Business and Research, China Engineering Consultants Inc., Taipei, Taiwan

Pei-lun Chang
Engineer, Department of Business and Research, China Engineering Consultants Inc., Taipei, Taiwan

Abstract
Knowledge Management System (KMS) has emerged as a popular approach not only for implementing knowledge management functions such as knowledge generation, storing, retrieval, and sharing, but also enabling an organization a tool to measure and monitor its intellectual property. Measurement of the KMS performance generates large amount of data that have profound implications for promotion of benefits resulted from the KMS both in terms of in the administration schemes and system modification schemes if the performance improvement patterns and rules can be found. In this paper, the Microsoft SQL Server® was adopted to perform Data Ming (DM) in order to dig out the abovementioned patterns and rules from the KMS performance data. Three DM techniques (Decision trees, Clusters, and Association Rules) were employed to mine the rules and patterns (performance improvement knowledge) existing in 654 historic performance data recorded from the KMS of a leading A/E consulting firm in Taiwan. Finally, 15 meaningful rules were obtained from decision trees; 5 useful patterns were identified from clusters; and 5 important regulations were concluded from 89 association rules. Performance improvement strategies are then inferred and planned based on the knowledge discovered from the historic performance data. It is concluded that the proposed method is systematic to follow by the KMS administrator; it is also effective for finding the knowledge for improvement of the performance of KMS.

Keywords
Knowledge Management; Data Mining; A/E consulting firm; Information systems; Intelligent decision support systems.

1. Introduction

Knowledge Management System (KMS) is a popular approach for knowledge management implementation in construction firms including contractors and A/E consultants. A KMS does not only provide a platform for knowledge generation, storing, retrieval, and sharing, but also enable an organization a tool to measure and monitor its intellectual property. In order to evaluate and improve the performance of the KMS, quantification methods for KMS performance were proposed in several previous works (Yu and Chang, 2005; Yu et al., 2006a; Yu et al., 2006b). From those works, it was found that profound implications may be inferred from the performance data recorded in daily knowledge management activities. Such implications can provide valuable strategies for promotion of benefits resulted from the KMS both in terms of administration and system modification schemes. The key to achieve such objectives is finding out the performance improvement knowledge. The Data Mining (DM) and Knowledge Discovery in Databases (KDD) are proven to be very effective in mining patterns and rules residing in large databases (Cabena et al., 1998; Fayyad and Uthurusamy, 1996; Han and Kamber, 2000; Leu et al. 2001).

In this paper, a methodology is proposed for mining knowledge of improvement strategy from the KMS performance data in a leading A/E consulting firm in Taiwan, the China Engineering Consultants Inc. (CECI). The proposed methodology combines three major elements: (1) a specialized data collection system for knowledge activity data of a KMS; (2) a quantitative model for measurement of the performance of KMS; and (3) commercial DM software—Microsoft SQL Server®— for performing DM tasks. 654 historic problem-solving cases are collected from the KMS of the case A/E firm for case study. Questionnaire surveys are conducted with the participants of problem-solving cases via a web-based internet questionnaire surveying system. The survey results are then converted into data in form acceptable for DM by the Microsoft SQL Server®. Three DM techniques (Decision trees, Clusters, and Association Rules) are employed to mine the rules and patterns existing in the performance data. Meaningful rules, useful patterns, and important association rules are found with DM. Performance improvement strategies are then inferred and planned.

The rest of this paper will be presented in the following manner: previous related works are reviewed in Section 2 to provide background of this paper; the methodology of knowledge discovery in KMS performance data is described in details in Section 3; then, a case study is conducted for mining of knowledge from KMS performance data of the case A/E firm; finally, findings from case are discussed and the concluded.

2. Review of related Previous Works

2.1 KMS in A/E Consulting Firms

Mezher et al. (2005) reported a work on a KMS in a mechanical and industrial engineering consulting firm in middle-east. Their paper concluded the process of building a knowledge management system in the Mechanical and Industrial Department at DAR AL HANDASAH, which is a leading consulting firm in the Middle East and the world. Finally, the paper concluded the lessons learned from the experience of building the knowledge management system and the steps needed to improve it. Other works related to KMS in A/E firms were reported by the authors of this paper (Yu and Chang, 2005; Yu et al., 2006a; Yu et al., 2006b), which focus on a specialized KMS for emergent problem-solving, namely SOS, of one of the top three A/E consulting firm in Taiwan. Those works analyzed the characteristics of knowledge management (KM) activities in the A/E consulting industry and how KMS can improve the competitiveness of the firm.

2.2 Performance Measurement of KMS

The most related work in literature on performance evaluation of a KMS was a work done by del-Rey-Chamorro et al. (2003) in Cambridge University. They developed an eight-step framework to create performance indicators for knowledge management solutions. del-Rey-Chamorro et al.'s work can be very useful for creating performance indicators of a KMS, however, their work was primarily developed based on the observations of KMS in manufacturing industry. Bassion et al. (2005) addressed that in developing a conceptual framework for measuring business performance in construction should take into account the organization's business objectives. Yu et al. (2006b) proposed a quantitative model for measuring time, man-hour, and cost benefits resulted from a KMS of an engineering consulting firm. In the paper, details of the proposed quantitative KMS benefit models are presented with a case study application to a A/E consulting firm. It was reported from their study that the average time benefit (TB) is 63%; the average man-hour benefit (MHB) is 73.8%; and the average cost benefit is 86.6%.

3. Methodology of Knowledge Discovery in KMS Performance Data

3.1 General KDD Procurdue

In this paper, the methodology of knowledge discovery in KMS performance data is based on the KDD procedure proposed by Han and Kamber (Han and Kamber, 2000) as depicted in Figure 1. A general process of KDD depicted in Figure 1 consists of the following detail steps: (1) Understanding the domain problem; (2) Extracting the target data set; (3) Data cleaning and pre-processing; (4) Data integration; (5) Data reduction and projection; (6) Choosing the function of data mining; (7) Choosing the data mining algorithm(s); (8) Data mining; (9) Interpretation; and (10) Using discovered knowledge—incorporating the discovered knowledge into the performance system, taking actions based on knowledge.

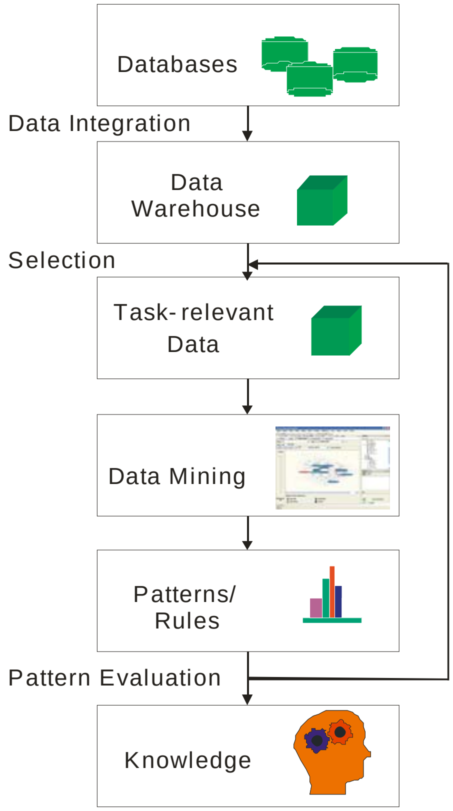

Figure 1: General process of KDD

3.2 Data Mining (DM)

Data mining is an interdisciplinary field with a general goal of predicting outcomes and uncovering relationships in data (Fayyad and Uthurusamy, 1996). DM is also the most critical step in the KDD process. While KDD refers to the overall process of turning low-level data into high-level knowledge, DM is the core mechanism that extracts useful knowledge from historical databases. It uses automated tools employing sophisticated algorithms to discover hidden patterns, associations, anomalies and/or structure from large amounts of data stored in data warehouses or other information repositories (Mitra et al. 2002). Data mining tasks can be descriptive, i.e., discovering interesting patterns describing the data, and predictive, i.e., predicting the behavior of the model based on available data (Furnkranz et al. 1997). Data mining involves fitting models to or determining patterns from observed data. The fitted models can be viewed as inferred knowledge. In this paper, the Microsoft SQL Server® is adopted for DM tasks on the KMS performance data. Even though the Microsoft SQL Server® provides nine different DM algorithms (Classification, Estimation, Prediction, Association rule, Clustering, Sequential Pattern, Decision tree, Neural Network, Time series), three of them (Decision Tree, Cluster, and Association Rule) are selected for this research after testing with the performance data.

3.3 Collection and Quantification of KMS Perofrmance Data

The SOS system in the case study was implemented since June the 1st of 2004. The period of data collection is from 2004/6/1 to 2006/9/26. Totally, 654 SOS problem-solving cases were collected. A web-based system, namely Knowledge-management-activity Survey Module (KSM) was developed for questionnaire survey and data collection. For every SOS case, both the questioner and responders were surveyed with KSM. The questionnaire was surveyed with the "questioner" for the following information:

(1) Whether the problem was solved or not? (2) Evaluation of the “level of contribution” (scale 1~5) of each solution from the responders; (3) Additional time required to develop the final solution; (4) The numbers of meetings, phone calls, and interviews required to develop similar solution via traditional approach; (5) Average time required for meetings, phone calls, and interviews required respectively to develop similar solution via traditional approach. The questionnaire was surveyed with the “responder” with the following information: (1) The time required developing the solution; (2) The time spent in-office and after-work, respectively, to develop the solution. Totally, 5,011 questionnaire surveys were conducted via KSM. Finally, 454 complete SOS cases and 3,250 valid responses were obtained.

3.4 Strategy Planning Methodology

A model for performance improvement strategy planning of construction knowledge management system was proposed in previous research (Yu et al., 2006a) and used for performance improvement strategy planning. The proposed method is called Performance Improvement Strategy Planning (PISP) model, which consists of four modules: (1) Knowledge-management-activity Surveying Module (KSM)—this module collects the information of knowledge creation activities of the participants; (2) Benefit Quantification Module (BQM) —this module quantifies the time, man-hour, and cost benefits of the KMS performance data; (3) Performance Data-mining Module (PDM) —the Microsoft SQL Server® is adopted to perform DM on the KMS performance data; and (4) Strategy Planning Module (SPM) —the improvement strategies of KMS are planned based on the patterns of high and low performance scenarios identified in PDM. The PISP model is adopted as the methodology for strategy planning in this paper.

4. Case Study

4.1 Background of Case A/E Firm and KMS

The case A/E firm is one of top three A/E firms in Taiwan. It was established in 1969 primarily for the purpose of promoting Taiwan's technology and assisting in the economic development of Taiwan and other developing countries. The number of full-time staffs of the firm is about 1,700. Among those around 800 are in-house staffs in headquarter located in Taipei, the other 900 are allocated in branches and site offices around the island. Headquarter, braches, and site offices are connected by Intranet. The structure of the case A/E firm consists of five business groups: (1) Civil Engineering Group; (2) Railway Engineering Group; (3) Electrical and Mechanical Engineering Group; (4) Construction Management Group; and Business and Administration Group. Each business group includes several functional departments. The annual revenue of case A/E firm is around 4 billion TWD (128 million USD). According to the information disclosed by the firm, more than 1,700 A/E projects were finished in the past thirty years. Totally volume (construction budget) of the finished projects exceeds 300 billion USD.

The implementation of KMS in the case A/E firm started five years ago. Unlike most of other examples of KMS implementation, the case A/E firm chose to develop the KMS completely by their own staffs without help of external consultants. Commercial software, Microsoft SharePoint® was adopted to develop the KMS. The system development took one year to complete the prototype. The prototype KMS began to operate after one year of the project commencement. It was found quickly that development of software KMS is not a tough job compared with the building of the culture and atmosphere for successful operation of the KMS. More that 40 communities of practice (COP) were established. The manager of COP is in charge of all activities for promotion of the knowledge creation in that COP. Incentives were provided by the company to stimulate the establishment of knowledge sharing atmosphere. The KMS has been modified quite a bit from its prototype three years ago. One of the most significant modifications was the introduction of SOS system for emergent problem solving. The case study will focus on the emergent problem-solving system of the case A/E firm, named SOS. The SOS system is a special design of COP, which provides a tentative forum for emergent problem encountered by engineers/managers. Once the problem is posed as SOS-problem, it is posted in the SOS board on the first page of the KMS for emergent discussions. Such arrangement forces every participant of KMS to take a look at the posed problem. So that it generally receives attentions and usually has a better chance to be solved by responders. Problems posed on the SOS board receive no response within one working day will be automatically removed and transferred to relevant COP. After then, it becomes regular problem for solving in COP.

4.2 Preparation of Data

The 454 complete SOS cases (from 3,250 valid responses) were transformed into quantitative performance records. Important information contained in each record includes the following: (1) Time benefit (TB, %); (2) Man-hour benefit (MHB, %); (3) Cost benefit (CB, %); (4) Time required by the questioner in developing final solution after the most favored response is obtained (FATS, hours.); (5) The satisfaction level of the most favored response (Max score, 1~5); (6) Total time required to solve problem by SOS system (NDS, days); (7) Total time required to solve the same problem by the traditional problem-solving approaches (NDT, days); (8) Time spent by the responder in responding the problem at work (W Duration, hours); (9) Time spent by the responder in responding the problem after (R Duration, hours); (10) others, such as average number of meetings, meeting time, phone calls, interviews in traditional approaches, etc.

The DM is performed with the Microsoft SQL Server® 2005. The general data preparation procedure consists of the following 7 steps: (1) Adding a new analysis service project; (2) Adding a new data source; (3) Linking the new data source; (4) Setting up data view of the new data source; (5) Selecting data table and data view; (6) Completing adding new data view; (7) Selecting a DM algorithm for data mining. In this paper, all nine DM algorithms provided by the Microsoft SQL Server ® 2005were tested with KMS performance data. Finally, three DM algorithms were adopted including Decision Tree, Clustering, and Association Rule.

4.3 Results of Data Mining

4.3.1 Decision trees

Procedure for mining of Decision Trees includes the following four steps: (1) Assigning dataset for DM to determine the variables for prediction; (2) Assessing and recognizing data type; (3) Entering DM structure name; (4) Decision Tree construction. After DM, six decision trees were constructed. Totally 28 rules were resulted from the Decision tree DM and 15 of them were found to be meaningful. The rules are named D1~D15.

Examples of rules obtained from Decision Trees are: (1) D2—the TB is very low as NDT is < 2.4 days; (2) D3—TB is average to very good for most of the cases with NDS < 2.24 days; (4) D7—in the data group of low TB (TB-), 65.79% of the cases are poor/very poor in TB; (5) D13—in the data group of high time benefit (TB+), TB is high as NDT is long (6.04 days).

4.3.2 Clusters

Procedure for mining of Clusters includes the following seven steps: (1) Assigning dataset for DM to determine the variables for prediction; (2) Assessing and recognizing data type; (3) Setting up variables; (4) Setting up parameters; (5) Viewing relationships between variables and parameters; (6) Viewing Clustering models; (7) Visualizing final Clustering model. After DM, 10 sets of clusters were obtained.

Totally 10 patterns were identified from the Clustering DM and 5 of them were found to be meaningful. The patterns are named C1~C5.

Examples of rules obtained from Clusters are: (1) C2—in the data group of high time benefit (TB+), NDS is highly correlated to FATS; (2) C3—in the data group of low time benefit (TB-), the CB is low and FATS is long (>2.8) when NDS is > 4 days; (3) C15—in the data group of low time benefit (TB-), the average problem resolved time (NDS) is long (4.79 days).

4.3.3 Association Rules

Procedure for mining of Association Rules includes the following seven steps: (1) Converting data type into transaction data; (2) Linking data table; (3) Selecting potential rules; (4) Setting up supports; (5) Determining screening criteria; (6) Screening out significant association rules. After DM, 89 association rules were obtained and 5 of them were found to be meaningful. The rules are named A1~A5.

Examples of rules obtained from Association Rules are: (1) A1—in the data group of high time benefit (TB+), the TB is very high as NDS is very short; and (2) A2—in the data group of high time benefit (TB+), the CB is very high as FATS is short.

4.4 Performance Improvement Strategy Planning

Based on the DM results described above, four strategies for KMS performance improvement are induced as follows:

Strategy I: Building Expert Map

Strategy descriptions: "In order to speed up the problem-solving process, a structured expert map should be built."

Facts: (1) A1—in the data group of high time benefit (TB+), the TB is very high as NDS is very short; (2) C2—in the data group of high time benefit (TB+), NDS is highly correlated to FATS.

Strategy II: Construction of Lesson-Learned File

Strategy descriptions: "In order to shorten the time required for developing final solution (FATS), the well-prepared lesson-learned file should be developed."

Facts: (1) A2—in the data group of high time benefit (TB+), the CB is very high as FATS is short; (2) C2—in the data group of high time benefit (TB+), NDS is highly correlated to FATS; (3) C3—in the data group of low time benefit (TB-), the CB is low and FATS is long (>2.8) when NDS is > 4 days.

Strategy III: Development of Pre-selection Mechanism

Strategy descriptions: "The problems with NDT>6 days are considered as priority for SOS; while the problems with NDT<2.8 days are not appropriate for SOS."

Facts: (1) D2—the TB is very low as NDT is < 2.4 days; (2) D13—in the data group of high time benefit (TB+), TB is high as NDT is long (6.04 days); (3) C2—in the data group of low time benefit (TB-), the CB is low and FATS is long (>2.8) when NDS is > 4 days.

Strategy IV: Setup of Warning System

Strategy descriptions: "The TB is good for problems with NDS < 1.46 days; the TB starts getting poor when NDS is > 2.4 days; the TB is negative for most of the problems with NDS >4.79 days. It is therefore helpful to set up a warning system for the time spots: (1) NDS =1.46 days—encouraging participants and promoting the discussions; (2) NDS =2.4 days—warning the questioner that the benefit of problem solving is getting worse, so action should be taken to improve the situation; (3) NDS =4.79 days—sending message to department leader that the problems can cause a negative impact if it is not solved right away; requiring the section leader to involve in the problem-solving."

Facts: (1) D7—in the data group of low TB (TB-), 65.79% of the cases are poor/very poor in TB; (2) D3—TB is average to very good for most of the cases with NDS < 2.24 days; (3) C15—in the data group of low time benefit (TB-), the average problem resolved time (NDS) is long (4.79 days).

5. Conclusions

Although many construction organizations have adopted knowledge management system (KMS) to facilitate the process of knowledge generation, storing, retrieval, and sharing. Previous research has also developed models of quantitative performance measurement for the KMS. However, effective methods for systematic performance improvement of the KMS have not yet developed. In this paper, a methodology is proposed for mining knowledge of improvement strategy from the KMS performance data in a leading A/E consulting firm in Taiwan, the China Engineering Consultants Inc. (CECI). The proposed methodology comprises of three major elements: (1) a specialized data collection system for collecting required data of a KMS; (2) a quantitative model for measurement of the performance of a KMS; and (3) the commercial DM software—Microsoft SQL Server®— for performing DM tasks. 654 historic problem-solving cases are selected from the KMS of the case A/E firm for case study. Totally, 5,011 questionnaire surveys are conducted with the participants of problem-solving cases. Finally, 454 complete problem-solving cases and 3,250 valid responses were obtained for performance evaluation. Three DM techniques (Decision trees, Clusters, and Association Rules) were employed for mining of the rules and patterns existing in the performance data. Meaningful rules, useful patterns, and important association rules are found by DM. It is summarized that 15 meaningful rules were obtained from decision trees; 5 useful patterns were identified from clusters; and 5 important regulations were concluded from 89 association rules. Four performance improvement strategies are then inferred and planned based on the knowledge discovered from the historic performance data. It is concluded that the proposed method is systematic to follow by the KMS administrator and also effective for finding out the required knowledge for improvement of the performance of KMS.

6. Acknowledgement

The founding of this research project was partially supported by the National Science Council, Taiwan, under project No. NSC 95-2221-E-216-049. Sincere appreciations are given to the sponsor by the authors. The valuable case study information presented in this paper was provided by CECI, Taipei. The authors would like to express sincere appreciations to the CECI, too.

7. References

Bassion, H. A., Price, A. D. F., and Hassan, T. M. (2005) "Building a conceptual framework for measuring business performance in construction: an empirical evaluation," *Construction Management and Economics*, 23(5), 495-507.

Cabena, P., Hadjinian, P., Stadler, R., Verhees, J., and Zanasi, A. (1998). *Discovering Data Mining: from Concept to Implementation*, Prentice-Hall, NJ.

del-Rey-Chamorro, F. M., Roy, R., van Wegen, B., and Steele, A. (2003) "A framework to create key performance indicators for knowledge management solutions," *Journal of Knowledge Management*, 7(2) 46–62.

Fayyad, U. and Uthurusamy, R. (1996). "Data mining and knowledge discovery in databases." *Commun. ACM*, Vol. 39, .pp. 24-27.

Furnkranz, J., Petrak, J., and Trappl, R. (1997) "Knowledge discovery in international conflict databases", *Applied Artificial Intelligence*, vol. 11, 91-118.

Han, J. and Kamber, K. (2000). *Data Mining: Concepts and Techniques*, Morgan Kaufmann Publishers, San Diego, U.S.A.

Leu, S.-S.; Chen, C.-N.; and Chang, S.-L. (2001) "Data mining for tunnel support stability: neural network approach," *Automation in Construction*, Vol. 10, No. 4, pp. 429-441.

Mezher, M.; Abdul-Malak, M. A., Ghosn, I.; and Ajam, M. (2005) "Knowledge Management in Mechanical and Industrial Engineering Consulting: A Case Study," Journal Management in Engineering, ASCE, Vol. 21, No. 3, pp. 138-147.

Mitra, S., Pal, S. K., and Mitra, P. (2002) "Data mining in soft computing framework: A survey", *IEEE Transactions on Neural Networks*, 13(1), 3-14.

Soibelman, S. and Kim, H. (2002) "Data Preparation Process for Construction Knowledge Generation through Knowledge Discovery in Databases," *Journal of Computing in Civil Engineering*, Vol. 16, No. 1, pp. 39-48.

Yu, W. D., and Chang, P. L. (2005) "Performance evaluation of the construction knowledge management system—a case study of an A/E consulting firm," *Proceedings of ICCEM 2005*, Session 4-D, Oct. 16~19, 2005, Seoul, Korea, 6 pp. 1058-1063.

Yu, W. D., Chang, P.L. and Liu, S. J. (2006a) "A Model of Performance Improvement Strategy Planning for a Construction Knowledge Management System," *Proceedings of The Fifth International Conference on Engineering Computational Technology (ECT 2006)*, Session VII—Modeling of Decision Making and Risk Analysis in Construction Management, Sept. 12~15, Las Palmas de Gran Canaria, Spain, 18 pp.

Yu, W. D., Chang, P.L. and Liu, S.J. (2006b) "Quantifying Benefits of Knowledge Management System: A Case Study of an Engineering Consulting Firm," *Proceedings of International Symposium on Automation and Robotics in Construction 2006 (ISARC 2006)*, Session A4 — Planning and Management (1), Oct. 3~5, 2006, Tokyo, Japan, 6 pp.