Mining knowledge management strategies from the performance data of cop

Wen-der Yu[1], Shih-ting Lin[2], Shen-jung Liu[3] and Pei-lun Chang[4]

[1]Professor, Institute of Const. Mgmt., Chung Hua Univ., Taiwan, wenderyu@chu.edu.tw
[2]Master Student, Institute of Const. Mgmt., Chung Hua Univ., Taiwan, m09616015@cc.chu.edu.tw
[3]Assistant Vice President, CECI Engineering Consultants, Inc., Taiwan, sjliu@ceci.com.tw
[4]Engineer, Depart. of Business & Research, CECI Engineering Consultants, Inc., Taiwan, peilun@ceci.com.tw

Abstract

Knowledge community of practice (CoP) is a popular approach for knowledge management implementation in construction organizations including contractors and A/E firms. In order to evaluate and improve the performance of the knowledge CoPs, quantification methods for performance measurement were proposed in previous researches. Profound implications may be inferred from the performance data recorded from daily knowledge management activities. Such implications provide directions of valuable strategies for administration schemes and system modifications. To achieve such goals, the performance improvement patterns and rules should be identified. In this paper MS SQL Server® was adopted to performed Data Ming (DM) tasks that dig out the abovementioned patterns and rules from the CoP performance data. Three DM techniques (Decision trees, Clusters, and Association Rules) were employed to mine the rules and patterns existing in the 4,892 historic performance data recorded from the CoPs of a leading A/E consulting firm in Taiwan. Performance improvement strategies are then inferred and planned based on the rules and patterns discovered.

Keywords: **Knowledge Management; Data Mining; Consulting firm; Performance Measurement, Strategy planning.**

1. Introduction

Knowledge Management System (KMS) is a popular approach for knowledge management implementation in construction organizations including contraction and A/E firms. 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 [1][2][3]. From those works, it was found that profound implications may be inferred from the performance data recorded in daily knowledge management activities. Such implications may indicate valuable strategies for increasing 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 [4][5][6][7].

In this paper, a case study is conducted on mining knowledge of improvement strategy from the performance data of a generic CoP in a leading A/E consulting firm in Taiwan, the CECI Engineering Consultants, Inc. (CECI). The proposed methodology combines two major elements: (1) a quantitative model for measurement of the performance of CoP; and (2) commercial DM software—Microsoft SQL Server®— for performing DM tasks. Totally 4,892 historic performance data were collected from nine selected CoPs of the case A/E firm for case study. Questionnaire surveys were conducted with the participants of CoP knowledge management (KM) activities via a web-based internet questionnaire surveying system. The survey results are then converted into data in the 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 CoP performance data is described in details in Section 3; then, a case study is conducted for mining of knowledge from CoP 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. [8] 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 [1][2][3], which focus on a specialized KMS for emergent problem-solving, namely SOS, of a 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. [9] 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. [10] 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. [3] 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 an 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%.

A similar research was reported by Yu et al. [11] in quantifying the performance of KM activities of a generic CoP of an A/E firm. In their research, the KM activities of a generic CoP can be classified into two categories: (1) knowledge sharing activities; (2) problem-solving activities. Quantitative models were developed for both of the two types of KM activities.

3. Methodology of Knowledge Discovery in CoP Performance Data

3.1 General KDD Procedure

In this paper, the methodology of knowledge discovery in CoP performance data is based on the KDD procedure proposed by Han and Kamber [6] 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.

3.2 Data Mining (DM)

Data mining is an interdisciplinary field with a general goal of predicting outcomes and uncovering relationships in data [5]. 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 [12]. Data mining tasks can be descriptive, i.e., discovering interesting patterns describing the data, and predictive, i.e., predicting the behaviour of the model based on available data [13]. 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.

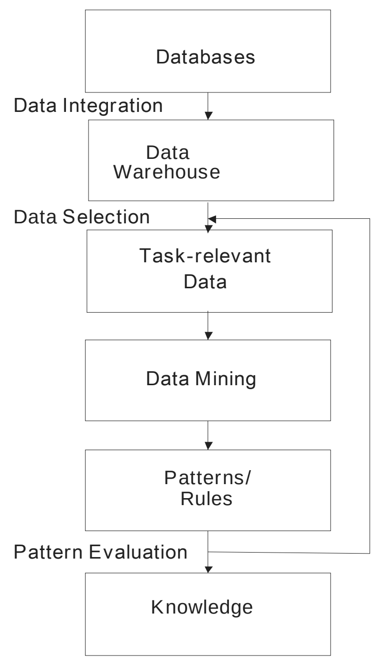

Figure 1 General process of KDD

3.3 Collection and Quantification of CoP Performance Data

A Knowledge Value Adding Model (KVAM) for quantitative performance measurement of KM activities in a CoP was proposed by Yu et al. [11]. The basic model of KVAM is depicted in Figure 2. There are two stages in a KM activity: (1) Raw knowledge creating process (RKCP) performed by the initiator of a KM activity; (2) Knowledge value adding process (KVAP)—performed by the participants/respondents of a topic in the CoP.

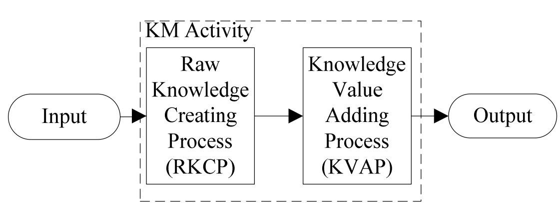

Figure 2 Basic model of KVAM [11]

It was propositioned by Yu et al. that "the amount of value of a KM activity is proportionate to the change of knowledge amount between Input and Output". Thus, the amount of Knowledge Value Added (KVA) of a KM activity can be measured by the difference between Input and Output of the process. This is calculated by multiplying the subjectively determined fuzzy terms of the two separate stages.

Two types of KM activities are identified: knowledge sharing and problem-solving. Fuzzy linguistic terms are defined to describe the RKCP and KVAP of the two types of KM activities. For knowledge sharing activities, five fuzzy terms are defined for RKCP. These fuzzy terms are called Raw Knowledge Creating Terms (RKCTs), including: (1) non-relevant; (2) data; (3) information; (4) knowledge; and (5) wisdom. Similarly, five fuzzy terms are defined for KVAP, Knowledge Value Adding Terms (KVATs), including: (1)

no-value-added; (2) get; (3) use; (4) learn; and (5) contribute. The similar definitions are given to fuzzy terms of the RKCP (RKCTs) and KVAP (KVATs), respectively, of the problem-solving activities. Details of the definitions for RKCTs and KVATs are referred to Yu et al. [11].The membership functions of the RKCTs and KVATs for both types of KM activities are assessed by questionnaires with the managers of the CoP. Techniques for determining the fuzzy means and sprays are used to construct the fuzzy membership functions [11].

4. Case Study

4.1 Background of Case A/E Firm and KMS

The CECI is one of leading 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 headquarters located in Taipei, the other 900 are allocated in branches and site offices around the island. Headquarters, 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 (5) 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.

4.2 Collection of CoP Performance Data

In order select appropriate CoPs with right cultural and enthusiasm in knowledge sharing, interviews were conducted by the research team during March 2007~May 2008 to meet with the managers of the CoPs. Finally, nine CoPs were selected: (1) Steel Community (associated with Structural Design Department); (2) Rail-Highway-Airport (associated with Transportation and Civil Department); (3) Supervision Art (associated with Construction Management Department); (4) Geotech (associated with Geotechnic Engineering Department); (5) Actuary (associated with Accounting Department); (6) Column-beam (associated with Structural Design Department); (7) Bridge (associated with Transportation Department); (8) Hydro-environ (associated with Hydraulic and Environmental Resources Department); and (9) Subway (associated with Railway and Mass Transportation Engineering Department).

Table 1 No. of surveyed KM cases

Sample statistics / CoP name	Total	
	Sampled	Valid
Steel Community	348	348
Rail-Highway-Airport	755	740
Supervision Art	674	663
Geotech	367	346
Actuary	182	177
Column-beam	111	107
Bridge	335	333
Hydro-environ	1876	1871
Subway	244	240
Total	4892	4825

The records of KM activities (including knowledge sharing and problem-solving activities) were collected from the nine CoPs from 2005/1 to 2007/12. Totally 4,892 KM cases were selected. Questionnaires were provided to the managers of the nine selected CoPs. See Table 1, among the 4,892 surveyed cases, 4,825 responses were valid with almost 99.9% of valid samples. Such high percentage of valid response was due to the effort of the research partner, CECI. An administrative mandate was ordered by the Business Research and Development to require all surveyed managers to participate in the questionnaire survey and respond timely. The data sets collected from questionnaire are then used for Kohonen learning to obtain the means of fuzzy membership functions associated with the associated RKCTs and KVATs.

4.3 Preparation of Data

The 4,892 historic CoP performance data were transformed into quantitative performance records by questionnaire surveys. Questionnaire surveys were conducted to the managers of the nine CoPs and all participants (initiators and responders) of the 4,825 KM records. After transformation, each KM activity performance record contains values of the seven attributes listed in Table 2.

Table 2 List of attributes in performance data

Item	Attribute name	Code	Description
1	CoP	ForumID	The name of CoP.
2	Article No.	TopicNo	Sequential No. of the article posted in a CoP. Totally, 4,892 articles are collected.
3	Technical code	TechCode	Classification code for the specialty of the articles.
4	Department	DeptCode	Department of the initiator. Totally, 41 departments.
5	RKCT	Kshare	Raw knowledge creating terms.
6	KVAT	Kapply	Knowledge value adding terms.
7	KVA	Kva	Knowledge-value-added calculated.

The Technical Code of Table 2 is a special classification system for all works and documents of the case firm. The Technical Code consists of three level: (1) Lifecycle code—first digit, describing the stage of the work in a project lifecycle; (2) Product & service code—digit 2 & 3, describing the product or service associated with the work; (3) function code—digit 4~6, describing the special function or technical domain of the work.

4.4 Performing Data Mining

The DM was performed with the Microsoft SQL Server® 2005. The general data DM 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.

4.5 Results of Data Mining

4.5.1 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 13 meaningful rules were resulted from the Decision tree DM. The rules are named D1~D13. An example of the mined decision trees is shown in Figure 3.

Two examples of rules obtained from Decision Trees are: (1) D1, D2—If TechCode = “060N25” (“water pollution” technical domain) or “060N00” (“environmental” technical domain), then the Kshare is "non -relevant"; (2) D9—If DeptCode is "Architecture" and TechCode = "571C35" ("RC bridge"), then Kva is "extremely high" (>=93).

4.5.2 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. Totally 14 meaningful patterns were identified from the Clustering DM. The patterns are named C1~C14. An example of the mined clusters is shown in Figure 4.

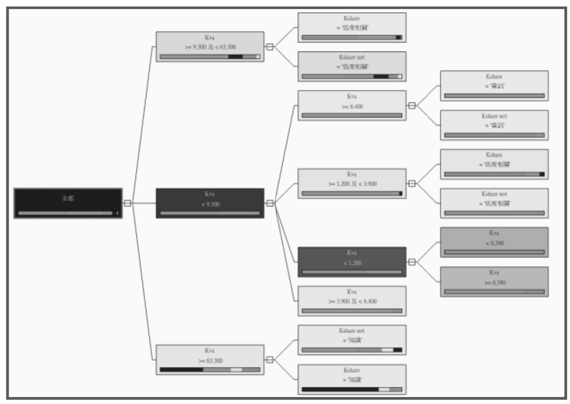

Figure 3 Example of mined decision tree

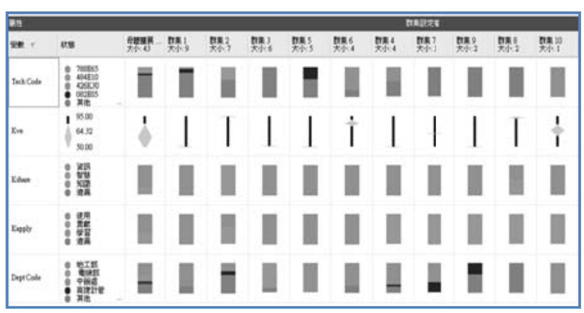

Figure 4 Viewer of mined clusters

Examples of rules obtained from Clusters are: (1) C1, 2, 4—low Kva clusters (Average Kva < 30) are identified as ForumID is "Hydro-environ" and DeptCode is "Hydraulic and Environmental Resources"; (2) C3—a high Kva cluster (Average Kva =83.26) is identified as ForumID is "Hydro-environ" and DeptCode is "Lioduey Supervision Office (a branch office of CECI in south Taiwan)"; (3) C8—a extremely high Kva cluster (Average Kva > 90) is identified as TechCode is "Foundation Engineering" and DeptCode is "Kaohsiung MRT Project Management"; (4) C10—a extremely high Kva cluster (Average Kva = 87) is identified as TechCode is "Railway Engineering" and DeptCode is "Electrical Engineering".

4.5.3 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, 11 meaningful association rules were obtained. The rules are named A1~A11. An example of the Window of the mined association rules is shown in Figure 5.

Examples of rules obtained from Association Rules are: (1) A3—If ForumID is "Steel Community" and Kshare is "Wisdom", then Kva is >= 80 (Support=20, Confidence=0.5); and (2) A4—If ForumID is "Bridge" and Kashare is "235C00" ("structural" technical domain), then Kva is >= 80 (Support=10, Confidence=0.5); (3) A5—If ForumID is "Geotech" and TechCode is "483I10" ("Viaduct rail" technical domain) and DeptCode is "MRT", then Kva is >= 80 (Support=20, Confidence=0.5);.

4.6 Performance Improvement Strategy Planning

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

Strategy I: Establish Screening Criteria for CoP

Strategy descriptions: "In order to eliminate the non-relevant articles, a screening system should be established. In the meanwhile, facility for uploading prevailing news and events should be provided".

Facts: D1, D2, C1, C2, C4, A1, and A2.

This strategy is induced based on observations of that the average Kva of "Hydro-environ" CoP is obviously lower than other CoPs. After reviewing the KM activities of that CoP, a great amount of articles posted on the CoP are related to prevailing news and events that the initiator would like to share with the CoP members. This situation has affected the members' attitude that tends to ignore the articles, and caused malfunction of the CoP. It is therefore recommended to screen out those articles from CoP and to provide the KM initiators (information providers) a facility to upload those articles directly onto the knowledge database of the KMS.

Figure 5 Window of the mined association rules

Strategy II: Establish a New Forum for Rail Related Disciplines

Strategy descriptions: "Railway and Mass Rapid Transportation (MRT) related issues contribute to high Kva values but those issues usually need be solved by multiple disciplines. It is recommended to establish a new CoP consisting of members from architectural, structural, geotechnic, electrical, and project management disciplines."

Facts: D11, C8, C10, and A7.

This strategy is recommended based on the observations that the Kaohsiung MRT project related articles were posted in sevral different CoPs and responded by the members with different backgrounds. Most of the KM activities were evaluated with high Kva. By interviewing with the KM initiators (questioners), it is found that the railway or MRT related engineering issues are highly diversified. Disciplines from different backgrounds are required to incorporate in the problem-solving process. This is actually the Medici Effect as reviewed previously.

Strategy III: Enhance the Lessons Learning Functions for Supervision Departments

Strategy descriptions: "Supervision departments and offices contribute significant amount of high value added KVAP activities. Lessons learning functions should be provided directly to those departments and offices to capture the valuable lessons and experiences from the construction sites."

Facts: D9, C8, C12, and A11.

This strategy is recommended based on the observations that the KVAP activities (Kshare) are evaluated relatively highly value-added in supervision works and issues. By reviewing the KM activities, it is found that those issues were encountered on construction sites and needed to be solved promptly. The results of problem-solving were also instantly fed back from the construction sites. Therefore, lesson-learning processes are actively performed in these departments and offices. However, due to the over-loaded works and shortage of staffs, the lessons-learned were mostly not recorded. This is absolutely a great loss of the firm. Facility and functions should be provided to those departments to enhance the lessons learning process in order to retain the valuable intellectual assets of the organization.

Strategy IV: Enhance Design and Construction Integration

Strategy descriptions: "Design and construction team usually benefit each other in KM activities. Mechanism should be established to enhance the design and construction integration."

Facts: D7, D9, D13, C3, C7, C8, C10, C12, C13, C14, and A5.

This strategy is recommended based on the observations that many high Kva KM activities (mostly with Kva values >= 80) were those initiated by the site engineers and responded by the design engineers, or reversely. It's obviously that integration between design and construction engineers is beneficial to knowledge value adding. Thus, mechanism (e.g., job rotation, seminars, and lessons-learned conference) can be established to enhance integration.

Conclusions

Many construction organizations have adopted Community of Practice (CoP) as an approach to facilitate the process of knowledge generation and sharing. However, most of previous efforts were spent on hardware and software investments. Prior research has developed models of quantitative performance measurement for a CoP, but there have been no effective method for systematic performance improvement of the CoP.

In this paper, a methodology is proposed to mine knowledge from the KM value-adding performance data for planning effective improvement strategies. The proposed methodology comprises of two major elements: (1) a quantitative model for measurement of the performance of KM activities in a CoP; and (2) the commercial DM software—Microsoft SQL Server®— for performing DM tasks. A case study was conducted for nine selected CoPs of a local leading consulting firm in Taiwan. Totally, 4,892 historic KM cases are sampled for questionnaire survey. Finally, 4,825 complete and valid responses were obtained for data mining.

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 13 meaningful rules were obtained from decision trees; 14 useful patterns were identified from clusters; and 11 important association rules were concluded. 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 provides the managers of KMS and the firm an effective tool to plan effective improvement strategies for their KM initiatives.

Acknowledgement

The founding of this research project was partially supported by the National Science Council, Taiwan, under project No. NSC 97-2221-E-216 -039. 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.

References

[1] 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.

[2] 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.

[3] 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.
[4] Cabena, P., Hadjinian, P., Stadler, R., Verhees, J., and Zanasi, A. (1998). Discovering Data Mining: from Concept to Implementation, Prentice-Hall, NJ.
[5] Fayyad, U. and Uthurusamy, R. (1996). "Data mining and knowledge discovery in databases." Commun. ACM, Vol. 39, .pp. 24-27.
[6] Han, J. and Kamber, K. (2000). Data Mining: Concepts and Techniques, Morgan Kaufmann Publishers, San Diego, U.S.A.
[7] 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.
[8] 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.
[9] 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.
[10] 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.
[11] Yu, W. D., Yao, H. H., Liu, S.J., and Chang, P.L. "Knowledge Value Adding Model for Quantitative Performance Evaluation of the Community of Practice in a Consulting Firm," Proceedings of The Sixth International Conference on Engineering Computational Technology (ECT 2008), Session VIII—Decision Making in Engineering Management, Sept. 2~5, Athens, Greece, 18 pp., 2008.
[12] 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.
[13] [Furnkranz, J., Petrak, J., and Trappl, R. (1997) "Knowledge discovery in international conflict databases", Applied Artificial Intelligence, vol. 11, 91-118.