Data Fusion and Modeling for Construction Management Knowledge Discovery

Lucio Soibelman *, Liang Y. Liu ** and Jianfeng Wu ***

* Assistant Professor, Department of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign, 205 N. Mathews Avenue, Urbana, IL 61801, U.S.A. soibelma@uiuc.edu

* Associate Professor, Department of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign, 205 N. Mathews Avenue, Urbana, IL 61801, U.S.A. lliu1@uiuc.edu

* Research Assistant, Department of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign, 205 N. Mathews Avenue, Urbana, IL 61801, U.S.A. jwu3@uiuc.edu

Abstract

Advances in construction data analysis techniques have provided useful tools to discover explicit knowledge on historical databases supporting project managers' decision making. However, in many situations, historical data are extracted and preprocessed for knowledge discovery based on time-consuming and problem-specific data preparation solutions, which often results in inefficiencies and inconsistencies. To overcome the problem, we are working on the development of a new *data fusion* methodology, which is designed to provide timely and consistent access to historical data for efficient and effective management knowledge discovery. The methodology is intended to be a new bridge between historical databases and data analysis techniques, which shields project managers from complex data preparation solutions, and enables them to use discovered knowledge for decision making more conveniently. This paper briefly describes the motivation, the background and the initial results of the ongoing research.

1. Introduction

In today's competitive market, the success of construction projects depends largely on project managers' capabilities to make corrective decisions during construction planning and control stages. Implicit experiences learned from previous projects by managers often play an important role in their decision making. For example, a construction expert is usually more capable to accurately predict production rates for activities under dynamic and complex conditions, based on his/her experiences as to how the values were changed with various conditions in previous projects. Unfortunately, obtaining experience is a highly subjective task limited by managers' personal experiences and capabilities to remember, understand, extract and use them. Moreover, the learning process for a novice manager usually takes many years, and sometimes is associated with a prohibitive cost due to incorrect management decisions.

In recent years, advances in data collection and analysis techniques have provided project managers with abundant data and powerful tools to improve current learning processes. On one hand, the availability of computerized databases increased in construction projects. Automatic data collection techniques, such as bar codes, radio frequency tags, mobile devices, and advanced scientific sensors, have been widely applied, which further increases the volumes of project databases. Also, many external sources have appeared to provide more comprehensive data about construction circumstances, e.g., weather data from National Climate Data Center (NCDC) and price index data from Engineering News-Record (ENR). On the other hand, latest data analysis techniques, such as Machine Learning (ML) and Data Mining (DM), have gained acceptance in construction projects. Compared with traditional data analysis techniques such as database queries and spreadsheets, these techniques are advantageous in providing rules and predictions on historical databases, instead of just data themselves. With the development of proper applications, ML/DM techniques could be useful to support project managers' decision making by providing more objective, comprehensive and explicit knowledge on large historical databases.

However, converting historical data into useful knowledge is a nontrivial process because historical databases are not always directly applicable for data analysis. Let's consider a typical scenario, in which a project manager wants to identify major reasons for the delays of a certain activity, say, installation of underground pipes. Even though computerized historical databases are available for the project manager, he/she cannot use them directly for data analysis for two major reasons. First, historical records of similar activities, such as their durations and circumstances, must be extracted from previous projects. Because the records are usually stored in heterogeneous databases employed by different projects, data extraction from them may require a lot of time and data query skills. Second, the extracted historical data must be cleaned, transformed, and formatted to satisfy the requirement for high-quality data by many data analysis algorithms. This is not an easy job either, because there are typically many data quality problems in historical databases, such as missing values for activity circumstances (e.g., soil conditions) or incorrect activity durations (e.g., -1 days). Therefore, in this scenario, proper data extraction and preparation operations must be applied on historical databases before useful knowledge can be identified in relevant historical records.

In common practices, due to the heterogeneous data formats and various data quality problems in historical databases, the majority of data extraction and preparation solutions are done manually by data analysts instead of project managers. Also, the solutions are specifically developed for predefined problems given by project managers. Such practices result in inefficiencies because a lot of time and effort must be repeatedly taken for data extraction and preparation for different problems (e.g., delay analysis for other activities). Moreover, since it is usually tedious and time-consuming for data analysts to repeat the solutions for each problem, they are more likely to employ ad hoc operations and even overlook some necessary steps, which cause inconsistencies in the output historical data. It is inconvenient for project managers, too, because they have to wait for data analysts to prepare and analyze data without any control of the process. Hence, a bridge between historical databases and data analysis techniques, which provide project managers with direct access to historical data for multiple problems, is essential to improve the applications of knowledge discovery for decision making. Otherwise, managers would still rely on their subjective and implicit, but directly available experiences.

2. Related Research Background

Research efforts in three different directions have provided potentials techniques to establish the efficient and consistent connection between historical databases and data analysis techniques: research in construction Knowledge Discovery in Databases (KDD) which includes data preparation as an integral part of complete knowledge discovery processes, and provides instructions and automatic tools for more efficient and consistent data preparation; research in construction data models which applies model-based techniques to improve efficiency and consistency of data exchange and retrieval in heterogeneous and distributed data sources; and recent research efforts in data integration approaches that are focused on different aspects of the connection between construction databases and data analysis techniques, e.g., data extraction and data organization. All three directions are discussed in detail below.

2.1 Research in Construction KDD

KDD, by definition "the nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data" (Fayyad et al. 1996), is used to extract knowledge from large amounts of data. It is an interdisciplinary field involving concepts from machine learning, database query, statistics, mathematics and visualization (Anand et al. 1998). It is also an iterative process consisting of a series of steps including domain and data understanding, data preparation, data mining, and finally, pattern evaluation and deployment (Chapman et al. 1999). In the areas of construction and facility management, KDD has been applied in recent research in response to the explosive growth of computerized historical databases (Buchheit et al. 2000; Soibelman and Kim 2002; Melhem and Cheng 2003). Compared with previous research in data analysis techniques which was focused on using machine learning and data mining algorithms, construction KDD research provides more complete and comprehensive instructions about knowledge discovery processes including data extraction and preparation.

It is also recognized in previous research (Buchheit 2002; Kim 2002) that data preparation is one of the most important and time-consuming steps in management knowledge discovery. Major operations in data preparation for construction KDD processes can be found in Kim's research (2002). Some automatic tools for data preparation tasks have been developed in previous research. For instance, Buchheit (2002) applied a multiple voting method to automatically assess and cleanse data quality problems in some transportation databases, and Melhem et al. (2003) introduced an application of automatic data selection methods for data analysis in a bridge inspection database.

Even though the developed frameworks and automatic tools can improve consistency and efficiency of data preparation to some extent, they cannot effectively eliminate inconsistencies and inefficiencies due to problem-specific data preparation solutions. In current construction KDD research, the data preparation step is still over-specified for predefined problems after domain and data understanding in earlier steps. Thus, preprocessed data and efforts in data preparation are often transitory with little reusability for other problems, which is a major cause for inefficiencies and inconsistencies. Moreover, although project managers are enabled to become involved in the KDD processes using the data preparation and analysis solutions, they usually lack time and data analysis skills to deal with the time-consuming and complex data preparation operations. As a result, they are still depending on data analysts for knowledge discovery in many cases, which incurs inconveniences for their applications of KDD processes as well.

2.2 Research in Construction Data Models

Shared project models are defined as collections of information that represent, describe or abstract Architecture/Engineering/Construction (AEC) projects (Fischer and Froese 1996). They are usually developed based on explicit construction domain knowledge for certain aspects of projects management, in order to provide domain-specific data integration solutions that are independent of particular applications and problems. Product models and process models, which represent relationships between construction products and processes respectively, are two major types of shared project models in current research. Examples of applications are described in extensive literature (Aouad et al. 1995; Fischer and Kam 2002; Eastman 1999; Wakefield and Sears, 1997; Martinez and Ioannou 1999; Choo et al., 1999; Karhu 2001). Moreover, Industry-wide data models have also been developed to support interoperation between computer applications used by various construction participants during different project stages, such as STEP (1993), IFC(2003), and aecXML (2002).

Research efforts in construction data models show the benefits of using mode-based techniques to provide efficient and consistent connections between heterogeneous data sources within a construction project. However, they are mainly focused on data exchange and retrieval to improve participants' decision making through enhanced communications and collaborations, lacking capabilities to support management knowledge discovery for several reasons:

- Most project models do not support data integration among multiple projects, while knowledge discovery processes are usually applied on historical databases from many previous projects that often employed heterogeneous data models.
- Shared project models are designed to support decision making in ongoing projects instead of data analysis for future projects. As a result, it is common that not all data relevant to data analysis are represented in project models and historical databases, and additional efforts must be included to incorporate external data, e.g., weather and economy data, for more accurate and comprehensive management knowledge discovery.
- Most project databases store construction data at the lowest levels of details for recording purposes. Therefore, summary data at different (especially higher) levels of details are difficult to obtain without tedious and time-consuming data queries and summarizations. As a result, timely and interactive data access to historical data for knowledge discovery processes is barely supported by such model-based databases.

These problems show that existing data integration models are usually not analysis-friendly enough to be directly applicable for building the connections between historical databases and data analysis techniques. There is a need to develop new construction data models, which not only extract historical data from heterogeneous databases, but also provide analysis-friendly data organization for project managers' knowledge discovery and decision making.

2.3 Research in New Data Integration Approaches

Besides model-based data integration techniques, research from other domains for data integration have also provided potential solutions for building the necessary connection between historical databases and data analysis techniques. Two major approaches within construction literature, which are focused on data extraction and organization respectively, are introduced below:

2.3.1 Wrapper/Mediator

An important assumption of research in data integration models is that project participants are subscribing a common construction data model for data exchanges and retrieval. However, this assumption is not always true in real projects. A wrapper between data sources and data query interfaces provides a potential alternative to address the data extraction problems in multiple and heterogeneous model-based databases. When a query is posted on a user interface, the wrapper is responsible for translating it into queries appropriate for each individual database. The results returned from the databases are merged in the wrapper and finally form a global answer set for the original query. O'Brien et al. (2002) developed the Scalable Extraction of Enterprise Knowledge (SEEK) project to provide data queries among heterogeneous databases owned by contractors, subcontractors and suppliers along a construction supply chain. Another wrapper-based initiative, Simple Access to Building Lifecycle Exchange (SABLE), is currently under development by a group of researchers and industrial practitioners to provide efficient and consistent data access to multiple IFC-based data servers.

A major advantage of the wrapper-based approach is its compatibility and extensibility to accommodate data in heterogeneous formats without losing important information in original sources. For instance, complex relationships between construction products and processes in construction data models can be kept intact while they are searched for data queries. Thus, the approach can be useful to be a neutral mediator between data sources based on heterogeneous construction data models. One of the disadvantages is that this approach requires complicated query writing and transformation at runtime, which may results in inefficiencies for data queries and analysis. This situation can become more serious when multi-level queries and summations are needed, or data sources are distributed in remote locations (Shen, Issa and O'Brien 2003).

2.3.2 Data Repository

As discussed before, one of the problems in current construction data models is that they store data at the lowest levels of details, lacking capabilities to support timely and interactive data access. A data repository that extracts, preprocesses, and reorganizes data from heterogeneous data sources into a single semantic data store may provide a solution for this data organization problem. Ahmad (2000) suggested that a subject-oriented, integrate, time-variant, and non-volatile data warehouse based on multidimensional models can provide higher performance for direct data queries and decision support on construction databases. Chau et al. (2003) developed a Construction Management Decision Support System (CMDSS) applying OLAP (On-Line Analytical Processing), data warehousing and visualization techniques.

The major advantage to the repository-based approach resides in its mechanism of preprocessed summarizations for multidimensional and multi-level data queries, which allow timely and interactive responses to users' queries and analysis. However, the approach usually organizes data only in decomposition-based data structures. As a result, other complex relationships between construction products and processes, such as interdependencies between project activities, may not be accommodated in the data repository. Consequently, the data analysis results could be inaccurate, because these interdependencies are also important for management knowledge discovery and decision making.

2.4 Summary: Problems and a New Need

Research efforts in construction KDD, data models and data integration approaches are useful to solve specific data integration, retrieval, and analysis problems to support project participants' decision making. However, these research efforts have their own limitations which make them insufficient to provide the necessary connection between historical databases and data analysis techniques. Data preparation solutions in current KDD research are still too problem-specific to support efficient and consistent knowledge discovery for multiple problems. Existing data integration models are not analysis-friendly enough to provide data organization for historical data from multiple projects. The two major data integration approaches in recent research, alone, also have disadvantages in supporting both efficient and accurate data analysis. Moreover, construction data models, which represent explicit domain knowledge, should be integrated into these approaches for more efficient and consistent data extraction and organization of historical data.

This situation motivated us to study on new construction domain-specific and analysis-friendly data models. A new methodology is being developed as the bridge between historical databases and data analysis techniques. Although data integration models and approaches are applied in this study, we named the methodology as "data fusion" because it does not only involve data retrieval from construction databases, but it also reorganizes and represents historical data in a new analysis-friendly way that is different from their original data structures.

3. Building the Bridge: Data Fusion Methodology

3.1 Definition

In the community of data collection and processing, data fusion is defined as "a multilevel, multifaceted process dealing with the automatic detection, association, correlation, estimation, and combination of data and information from single and multiple sources" (Carvalho et al. 2003). In many areas, the term "data fusion" refers specifically to multi-sensor data fusion, and it is widely used in sensor-based data/image collection and processing at different levels. Generally, high-level data fusion is concerned with combining the processed output from different data systems, while low-level data fusion deals with data collection applications and combines the original input from them into integrated data systems.

In this research, we use the general definition of "data fusion", but specifically limit it to the high-level implication. Accordingly, the new construction data fusion methodology is defined as a model-based methodology that deals with automatic or semi-automatic extraction, reorganization, and representation of historical data from previous projects and external sources, in order to support management knowledge discovery and decision making. This research is a significant departure from current research in data integration, which can be viewed as low-level "data fusion" and provides data sources for high-level "data fusion".

3.2 Research Goal

The goal of this research is to develop a new construction data fusion methodology to provide an efficient and consistent connection between historical databases and data analysis techniques. The application of the methodology shields project managers from complex data preparation solutions, and enables them to discover knowledge in historical data for decision making more conveniently.

Two major hypotheses are proposed for testing the main concepts of the methodology. First, we hypothesize that construction domain-specific data models, which are independent of particular data sources, data analysis techniques, and specific problems, can be built to support management knowledge discovery on historical databases. Second, we hypothesize that analysis-friendly organizations of historical data can improve the efficiency and effectiveness of current management knowledge discovery processes.

The scope of this ongoing research is limited to identifying and solving the fundamental problems in design, development and implementation of the new construction data fusion methodology. A prototype system is being developed and will be applied in sub-domains of construction management to validate the concepts.

3.3 Schematic Design

The new construction data fusion methodology is composed of three major components. As shown in Fig. 1, the domain-specific and analysis-friendly data fusion models construct the "main body" of the bridge between historical databases and data analysis techniques. Two additional parts, data extraction and data representation components are two "joints" connecting the bridge with both its ends. The three components are loosely coupled modules, which allow them to be flexibly and incrementally developed.

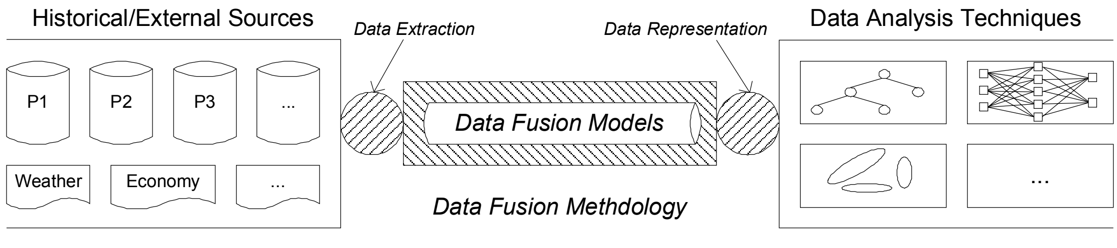

Figure 1. Schematic Design for the Data Fusion Methodology

A typical working procedure under the new construction data fusion methodology is:

- Data analysts collect, extract and prepare data from historical databases that are physically and semantically heterogeneous. The complex and time-consuming data extraction and preparations operations are implemented together only once for some abstract domain problems, such as relationships between activity performance and construction conditions.

- Extracted and preprocessed historical data are reorganized and stored in data structures based on analysis-friendly data fusion models.
- For specific problems in construction projects, e.g., delay analysis for a certain activity such as building drywalls, project managers can access the historical data directly to choose relevant historical records, and discover knowledge in them for decision making.

3.4 Functional Requirements for Data Fusion Models

The data fusion models organize historical data in some domain-specific and analysis-friendly ways to provide timely and consistent data access and support efficient and effective knowledge discovery. Currently, we have studied some major requirements for searching historical data in the data fusion models, based on the following observations in construction decision making:

- Despite the unique and changing nature of construction projects, project managers are able to capture "what-happened" in previous projects, i.e., relevant historical facts they experienced for decision making. Similarly, historical facts stored in the data fusion models should be organized in a way that relevant records can be identified correctly and efficiently for specific problems.
- "What-if" scenarios are important for knowledge discovery and decision making. The data fusion models should organize historical data to enable project managers to interactively explore "what-if" scenarios in different directions, such as adding/removing a factor for data analysis, or moving to different levels of details for selected factors. For example, if a project manager knows that "if it is hot, the average productivity of concrete work is reduced by 30%", he/she may want to ask more about "what if it is both raining and hot?" or "how are the productivity values of different activities in concrete work changed?"
- Sometimes project managers need to make more proactive decisions about "what-next" to predict possible problems and prepare for them. In this case, project managers need to search for problems based on given conditions (e.g., "if it is raining, which activities' durations would be influenced most?"), which is different from the first requirement to search facts for specific problems (e.g., reasons for delays of given activities). This requirement is being considered in the development of the data fusion models as well.

4. Data Fusion: A Preliminary Case Study

Currently, a preliminary case study based on the new construction data fusion methodology is under development. The details of the case study are introduced below.

4.1 Data Source and Requirement Analysis

In this case study, we use the RMS (Resident Management System) database provided by the U.S. Army Construction Engineering Research Laboratory (USA-CERL). The RMS database contains construction data from multiple projects. For each construction project, there are 55 data tables which contain data about both project performance and construction conditions. Weather data and cost index data from external sources (NCDC and ENR websites, respectively) are also included to accommodate more comprehensive construction conditions.

The case study is focused on a small sub-domain, which is to identify relationships between project performance (e.g., labor productivity, cost, and schedule) and construction conditions (e.g., managerial, technical, financial, and external conditions) for various activities at different levels of details (e.g., projects, activity groups, activities, work tasks, etc.). It is intended to support project mangers' data analysis in relationships between certain performance indicators and construction conditions for particular activities. Instead of preparing data for specific problems repeatedly in a previous study using the same database (Kim 2002), in this study, project managers are enabled to efficiently and consistently access the preprocessed historical data, find relevant historical records, and identify relationships in them.

4.2 Primary Design of the Data Fusion Model

In this case study, the problem domain is focused on construction activities. For this reason, we are looking into existing process models and data integration techniques, which can be combined together to represent historical data in a domain-specific and analysis-friendly way.

To represent historical data in a domain-specific way, we use IDEF0 (Integration Definition for Function Modeling) notations to describe the general relationships between project performance and construction conditions. IDEF0 was originally developed by the U.S. Air Force Knowledge Base Systems, Inc. and documented by the U.S. National Institute of Standards Technology (NIST) as Federal Information Processing Standard 183 (IDEF0 1993). The elements of the process model based on IDEF0 representation are shown as below (Fig. 2):

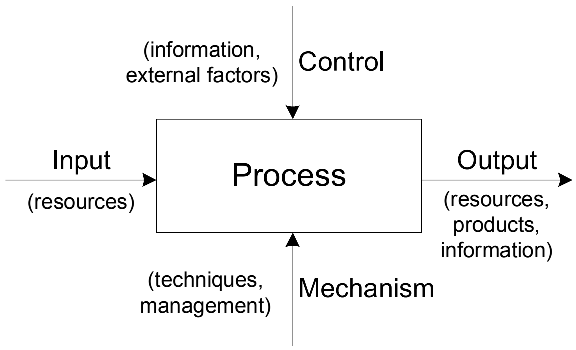

Figure 2. Elements of Construction Processes Described by IDEF0

In our vision, a construction process is implemented with inputs of resources such as money, manpower, machine and materials, and releases reusable resources, generates new products and information as outputs after its implementation for a certain time. The time, inputs and outputs are used to determine the performance, such as productivity and material usage. The controls, including information and external conditions, constrain the implementation. The technical and managerial mechanisms determine the way that the process is implemented. Both controls and mechanisms are viewed as construction conditions that influence the performance values.

Two relationships exist in construction processes for this ongoing case study. First, a process can be divided into several sub-processes, for instance, installing concrete wall can be divided into installing reinforcements, installing formwork, and placing concrete. In this way, the inputs, outputs, controls and mechanisms of the higher-level process are decomposed or distributed to its lower-level sub-processes. Also, the sub-processes can have their "internal" inputs, outputs, controls and mechanisms within the high-level process. Second, a process is dependent on another process if the former needs resources (as inputs) or information (as controls) from the latter. Fig. 3 shows a simple example of these two relationships.

Based on the basic definitions of construction processes, their elements and their relationships, our research team is applying the new data integration techniques to develop an analysis-friendly data fusion model for the case study. The new data fusion model is intended to support data integration from multiple projects in the RMS database, data combination of external data with RMS data, and data reorganization to support timely and flexible access to inputs, outputs, controls and mechanisms of construction processes at different levels of details. Advantages and disadvantages of both wrapper-based and repository-based approaches are being considered to develop the data model that accommodates both relationships between construction processes.

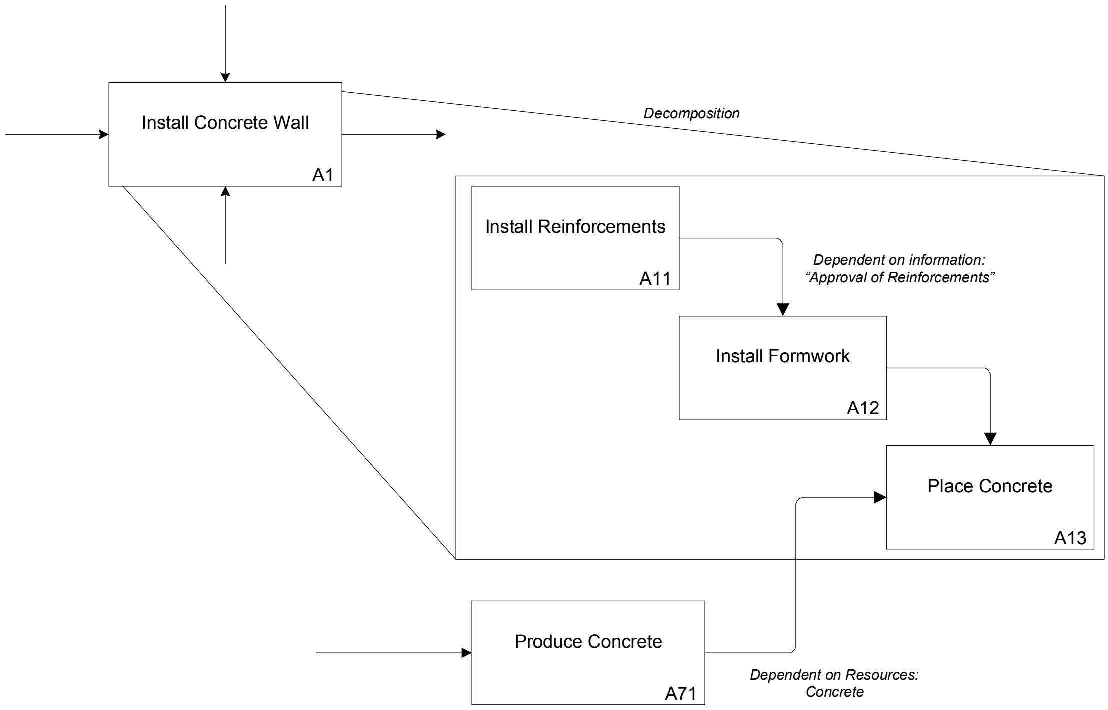

Figure 3. Relationships between Construction Processes

4.3 Data Fusion Prototype System

Besides data organization based on the data fusion model, the other two components of the data fusion prototype system are also being developed.

- Data extraction: common data preparation operations in the RMS database were identified, and an integrated development environment (IDE) has been developed to support these operations. The idea of the IDE is to provide a convenient environment for the data analysts so that they can follow the necessary data preparation steps instead of employing ad hoc operations or overlooking some of them. Currently, the operations supported by the IDE include: general information of data tables and attributes, such as automatic identification of relationships between data tables, and statistical information of attributes in a data table; data cleaning, e.g., removal or substitution of missing values, detection of abnormal values for attributes, identification of duplicated records, etc; and data transformation, for example, transformation of numerical attributes into categorical values, or vice versa.

- Data representation: a user interface, still under development, is intended to satisfy the three major functional requirements of the data fusion model (see details in section 3.4). First, for "what-happened" problems, project managers can look up historical records for specific activities with particular performance indicators and construction conditions, and apply data analysis tools to identify the relationships in them. Second, project managers are allowed to interactively browse activities at different levels of details, and choose performance values under various conditions for data analysis, in order to explore multiple "what-if" scenarios to make more reasonable decisions. Third, the interface also supports "what-next" queries by asking project managers to input specific conditions and a range of activities, and efficiently identifying activities at different levels that could be mostly influenced by the conditions, which helps managers to predict future problems and prepare well for them.

The case study is still under development and will be available in the near future. It is expected that both the benefits and challenges of the data fusion methodology can be identified in this case study. The implementation results will be helpful for future development of the data fusion models and prototype system for larger and more complex sub-domains.

5. Conclusion and Future Work

Efficient and consistent access to historical data for data analysis is essential for improving project managers' decision making based on knowledge discovery processes. However, data extraction and preparation solutions in previous construction KDD research are usually time-consuming and problem-specific, while existing data integration models are not analysis-friendly enough to establish the required connections between historical databases and data analysis techniques. By applying advanced data modeling and integration techniques, we are working on the development of a new data fusion methodology, which is intended to provide project managers with timely and consistent access to historical data, and allow them to efficiently and effectively discover knowledge for management decision making. Currently, a preliminary case study and prototype system is under development using concepts from the new methodology and will be available in the near future.

Based on the development of the case study and data fusion prototype system, our research team will continue to work in the following areas: further data collection and requirement analysis for more extensive understanding about problems and requirements for applications of historical databases and data analysis techniques in construction projects; design and development of the data fusion models and prototype system to apply concepts from the methodology in dissimilar sub-domains in construction management for case studies; and finally, generalization of the data fusion methodology for the construction management domain, and evaluation of the completeness and extensibility of the methodology using scientific methods.

References

aecXML – Architecture/Engineering/Construction XML (2002). http://www.iai-na.org/aecxml/

Ahmad, I. (2000). "Data warehousing in construction organizations." *Proceedings of the Construction Congress VI*, Orlando, FL, 194-203.

Anand, S. S. et al. (1998). "Discovering case knowledge using data mining." *Proceedings of the Second Pacific-Asia Conference in Knowledge Discovery and Data Mining (PAKDD)*, Melbourne, Australia, 25-35.

Aouad, G. et al. (1995). "The conceptual modeling of construction management information." *Automation in Construction*, 3(4), 267-282.

Buchheit, R.B. et al. "A knowledge discovery case study for the intelligent workplace." *Proceedings of the Eighth International Conference on Computing in Civil and Building Engineering*, Stanford, CA, 914-921.

Buchheit, R.B. (2002). "Vacuum: Automated procedures for assessing and cleansing civil infrastructure data." PhD Dissertation, Carnegie Mellon University, Pittsburgh, PA.

Carvalho, H. et al. (2003). "A general data fusion architecture." *Proceedings of the 6th International Conference on Information Fusion*, Cairns, Queensland, Australia.

Chapman et al. (1999). "The CRISP-DM process model." http://www.spss.it/download/pub-paper.pdf

Chau, K.W. et al. (2002). "Application of data warehouse and Decision Support System in construction management." *Automation in Construction*, 12(2), 213-224.

Choo, H.J. et al. (1999). "WorkPlan: Constraint-based database for work package scheduling." *Journal of Construction Engineering and Management*, ASCE, 125(3), 151-160.

ENR – Engineering News-Records, http://enr.construction.com/features/conEco/default.asp

Eastman, C. (1999). Building Product Models: Computer Environments Supporting Design and Construction. CRC Press, Boca Raton, FL.

Fayyad, U. M. et al. (1996). Advances in Knowledge Discovery and Data Mining. AAAI Press/ The MIT Press, Cambridge, MA

Fischer, M. and Froese, T. (1996). "Examples and characteristics of shared project models." *Journal of Computing in Civil Engineering*, ASCE, 10(3), 174-182.

Fischer, M. and Kam, C. (2002). "PM4D final report." CIFE Technical Report #143, Stanford University, Stanford, CA

IDEF0 – Integration Definition for Function Modeling. (1993). http://www.idef.com/idef0.html.

IFC – Industry Foundation Classes. (2003). http://iaiweb.lbl.gov/IFC/

NCDC – National Climate Data Center, http://www.ncdc.noaa.gov/oa/ncdc.html

Karhu, V. and Lahdenperä, P. (1999). "A formalised process model of current Finnish design and construction practice." *International Journal of Construction Information technology*, 7(1), 51-71.

Kim, H. (2000). "Knowledge discovery and machine learning in a construction project database." PhD Dissertation, University of Illinois at Urbana-Champaign, Urbana, IL.

Martinez, J.C. and Ioannou, P.G. (1999). "General-purpose systems for effective construction simulation." *Journal of Construction Engineering and Management*, ASCE, 125(4), 265-276

Melhem, H.G. and Cheng, Y. (2003). "Prediction of remaining service life of bridge decks using machine learning." *Journal of Computing in Civil Engineering*, ASCE, 17(1), 1-9.

Melhem, H.G. et al. (2003). "Wrapper methods for inductive learning: Example application to bridge decks." *Journal of Computing in Civil Engineering*, ASCE, 17(1), 46-57.

O'Brien, W. et al. (2002). "SEEK: Accomplishing enterprise information integration across heterogeneous sources." *Electronic Journal of Information Technology in Construction, Special Issue on Knowledge Management*, 7(2), 101-124

SABLE – Simple Access to Building Lifecycle Exchange, http://www.blis-project.org/~sable/

Shen, Z., Issa, R.R.A. and O'Brien, W. (2004). "A model for integrating construction design and schedule data."*Proceedings of the Fourth Joint International Symposium on Information Technology in Civil Engineering*, Nashville, TN

Soibelman, L. and Kim, H. (2002). "Data preparation process for construction knowledge generation through knowledge discovery in databases." *Journal of Computing in Civil Engineering*, ASCE, 16(1), 39-48.

STEP – Standard for the Exchange of Product Model Data. (1993). ISO 10303-22: Product Data Representation and Exchange, Part 22.

Teicholz, P. (1999). "White paper on AEC practice and research issues for the future." *Berkeley-Stanford CE&M Workshop: Defining a research agenda for AEC process/product development in 2000 and beyond*, August 1999, Stanford, CA.

Wakefield, R.R. and Sears, G.A. (1997). "Petri nets for simulation and modeling of construction systems." *Journal of Construction Engineering and Management*, ASCE, 123(2), 105-112