BERKELEY-STANFORD CE&M WORKSHOP: DEFINING A RESEARCH AGENDA FOR AEC PROCESS/PRODUCT DEVELOPMENT IN 2000 AND BEYOND

CONSTRUCTION KNOWLEDGE GENERATION AND DISSEMINATION

Lucio Soibelman
Soibelma@uiuc.edu
Assistant Professor
University of Illinois at Urbana-Champaign
Department of Civil and Environmental Engineering
NCSA Affiliate
3129C Newmark Civil Engineering Laboratory
205 North Mathews Avenue
Urbana, IL - 61801

Introduction:

Construction managers rely on past experience to be able to perform their most important tasks. "How much time is this task going to take?" or "How many nails do we need to build this panel?" are the types of questions that project managers face daily in their planning activities. Failure or success in developing good schedules, budgets and other project management tasks depends in the project manager ability to obtain reliable information to be able to answer these types of questions. Students and young practitioners tend to rely in information that are regional averages provided by different publishing companies, unlike experienced project manager that tend to rely heavily on their personal experience to adapt companies and published averages to the problem in question. In the beginning of my career as a construction manager I was lucky to work with my father, a project manager with more than 35 years of experience and more than 100 construction projects in his resume. I always admired his capability to add a reasoning process to the available information. He would tell me: "Consider an extra time in your schedule to build this wall. It will be built in the summer. We are expecting a rainy, humid and hot summer and the construction industry is in a boom. We are going to have a hard time hiring good bricklayers. Five years ago I built a wall with the same characteristics and we had a hot summer and difficulties hiring good labor, as a consequence our productivity was 25% below our average". Since that time I began to question myself over and over again: Do we really need to make mistakes to build our knowledge base? Is project experience the only way to transform managers in good project managers? Can we provide this kind of experience to students and young practitioners? Can we transform available project data in this kind of knowledge?

Many researchers in construction management study one narrow topic in great detail, seeking to improve the available scheduling algorithms, estimating spreadsheets and other project management tools. Such a "micro-scale" level of research is extremely

important providing the required tools for the project manager work. However, it is also important to step back from the process of tool development. Even with the best scheduling tool if we provide low quality input information we are going to produce a low quality schedule as output. As frequently said in project management: “garbage in, garbage out”.

At the same time that we are starving for knowledge we are drowning in information. The construction industry is seeing an explosive growth in its capabilities to both generate and collect data. Advances in scientific data collection, the introduction of bar codes for almost all-commercial products, and the computerization have generated a flood of data. Advances in data storage technology, such as faster, higher capacity, and cheaper storage devices (e.g. magnetic disks, CD-ROMS), better database management systems, and data warehousing technology, have allowed us to transform this enormous amount of data into a computerized database system. A construction project consists of a great amount of data such as data about labor productivity, materials, equipment, cost estimating, scheduling of activity duration, and so on. As the construction industry is adapting the new computer technologies in terms of hardware and software, computerized construction data are becoming more and more available. However, it is our reality that most of data in a construction project is used only for communication purposes and stored in a file or a database without being analyzed.

Research Issues

There is an urgent need to study this increasing amount of available data. Knowledge Discovery in Databases and Data Mining, technologies that combine techniques from machine learning, artificial intelligence, pattern recognition, statistics, databases and visualization to automatically extract concepts, concepts interrelationships, and patterns of interest from large databases, can be used to generate novel knowledge from large construction databases.

Data management started about three decades ago, when no data specific information was explicitly stored along with the data. Often data had to be stored more than once across the organization leading to inconsistencies and inefficiencies. Data Management Systems were introduced in the late 1960’s largely triggered by the Space Race. Constraints, such as data types, value ranges, dependencies, or generation languages were provided to ease application development at this time

Nowadays the explosive growth of many business, government, and scientific databases has far outpaced our ability to interpret and digest the data. Such volumes of data clearly overwhelm the traditional methods of data analysis such as spreadsheets and ad-hoc queries. The traditional methods can create informative reports from data, but cannot analyze the contents of those reports. A significant need exists for a new generation of techniques and tools with the ability to automatically assist humans in analyzing the mountains of data for useful knowledge.

Historically the notion of finding useful patterns in raw data has been given various names, including knowledge extraction, information discovery, information harvesting, data archeology, and data pattern processing. By the end of 1980s, a new term, knowledge discovery in databases (KDD), was coined to replace all of the old terms whose objective was to find patterns and similarities in raw data. Artificial intelligence and machine learning practitioners quickly adopted KDD and used it to cover the overall process of extracting knowledge from databases. The term, Data Mining has been used in this context for the process when the mining algorithms were applied. Recently, as a result of the increasing attention of vendors and the popular trade press in this area, the words data mining have been adapted and have come to mean, like KDD, the overall process of extracting knowledge from databases.

Research Needs

1. Develop improved methods to obtain novel knowledge from large construction databases. A process to select attributes, clean and prepare construction databases to the Knowledge Discovery in Databases process need to be developed. Existing machine learning, artificial intelligence, and data mining tools must be adapted to fit the special characteristics of construction databases

2. Improve access to past construction management experience and knowledge by practitioners and students. User friendly interfaces need to be created to allow easier and transparent use in the classroom and industry of the generated knowledge.

3. Use knowledge to improve the quality of information in construction. Construction knowledge and also knowledge about information transactions in construction, especially in knowledge intensive process like design, can improve the quality of construction management information systems

4. Develop methods to disseminate knowledge using construction information systems. The objective is to enact and support information systems in order to assure that accurate and relevant information and knowledge will be on the right place at the appropriate time and with the required quality

Personal Background

My construction industry experience began in Brazil during my first year as an undergraduate student when I began working as a trainee in a small construction company. Since the first day in the construction site I had the feeling that something was extremely wrong. Construction sites were unorganized, dirty, and dangerous and I always felt that that existed room for improvements. After graduation in 1984 I created a huge bureaucratic data management and control system for my construction sites. My intention was to input all available data in the computer using spreadsheets and databases. At that time I learned that there is a huge difference between data (numbers and more numbers) and information that helps us in the decision making process. The main problem that I could not solve at that moment was how to transform the available data in information.

After almost 10 years of construction practice I decided to return to the school as a masters student on construction management. My masters' research was a statistical work on construction material waste, its incidences, causes, and control. This work gave me deep understanding of the shortcomings of the available statistical tools in the process of transforming large amount of data in information and knowledge. After my masters I moved to U.S to work on my Ph.D. at the Massachusetts Institute of Technology. At MIT I decided to work with others tools available to generate knowledge from data. During my research I focused on different Machine Learning and Artificial Intelligence tools. I developed a distributed artificial intelligence computer system that works as a decision support tool for the preliminary design of tall buildings structures. This system applies an agent like approach developing competence modules, which is an expert at a particular small task oriented competence. The Internet was used as a communication backbone among the different systems that implement the reasoning mechanisms. The current implementation was developed with three modules: The classification module implemented by a decision tree agent, the past cases/experience module implemented by a Case Based Reasoning agent, and the adaptation module implemented by the Genetic Algorithm agent. My homepage: http://cenpc194.ce.uiuc.edu provides brief descriptions and links to current projects in my research.