Flexible Extraction of Practical Knowledge from Bridge Databases

Yoram Reich[1], Steven J. Fenves[2], Hon. M. ASCE, Eswaran Subrahmanian[3]

In the *Proceedings of the First Congress on Computing in Civil Engineering*
Washington, DC, June 20-22, 1994, pp. 1014-1021

ABSTRACT

Bridge databases contain significant information that can assist in future bridge management decisions. The effective utilization of this information requires facilities for the flexible extraction of usable knowledge in the form practitioners can use. We propose machine learning technology as a means to perform this knowledge extraction and to support the maintenance of bridge database information. An integrated system is proposed for supporting practitioners in the practical use of this technology.

INTRODUCTION

The premise of this research is that bridge databases, containing historical data on bridge performance and maintenance decisions, contain knowledge that can be used to enhance future decision making. The major impediment on using this vast amount of data in practice is the problem of making the knowledge embedded in the data explicit for decision makers. While the development of new database technologies, comprising the majority of present efforts in bridge management research, can provide efficient means for storing and retrieving information, it cannot help decision makers in comprehending this vast amount of information. A new technology, called machine learning (ML) (Michalski et al 1983), presents a significant potential for addressing the problem of transforming knowledge *implicit* in historical data into *explicit* knowledge for decision makers.

Briefly, ML techniques are methods that extract knowledge implicit in data and represent them in various forms such as decision trees or heuristic rules. Such knowledge can subsequently be used in solving problems. Several research projects have explored the use of ML in civil engineering (Arciszewski and Rossman 1992) and other relevant disciplines for: (1) evaluating structures during preliminary design (Arciszewski et al 1987); (2) learning from failures of structures (Stone and Blockley

[1]Senior Lecturer, Department of Solid Mechanics, Materials and Structures, Faculty of Engineering, Tel Aviv University, Tel Aviv 69978, Israel. Email: yoram@eng.tau.ac.il

[2]Sun Company University Professor, Department of Civil Engineering, Carnegie Mellon University, Pittsburgh, PA, 15213

[3]Senior Research Scientist, Engineering Design Research Center, Carnegie Mellon University, Pittsburgh, PA 15213

1989); (3) learning environmental evaluations (Julien et al 1992); and (4) building a cable-stayed bridge design assistant (Reich and Fenves 1995). The last effort has employed two ML techniques for extracting synthesis and redesign knowledge from a bridge database containing descriptions of designs.

The objective of facilitating the practical usability of ML techniques translates into two intertwined issues. The first issue deals with the development of new ML techniques appropriate for the data available in bridge databases and their required interactive mode of operation. The second issue deals with giving practitioners the ability to integrate learning in context by providing them with facilities for flexibly modeling their problems and executing programs suitable for solving their problems.

This paper discusses these issues and the impact that the technology presented has on the development of Bridge Management Systems (BMS) in general and bridge databases in particular.

BRIDGE MANAGEMENT: BACKGROUND

Our approach is based on a comprehensive view of the bridge management problem. It views the process of bridge design, maintenance, and replacement as a process generating an enormous amount of information. While current practice addresses parts of this information generation and management, our approach attempts to account for the *life-cycle flow* of this information.

Figure 1 outlines our model of the flow of information in bridge design and maintenance. The bold branches depict the functionality provided by our proposed approach, while the remaining branches describe the scope of present approaches to bridge information management.

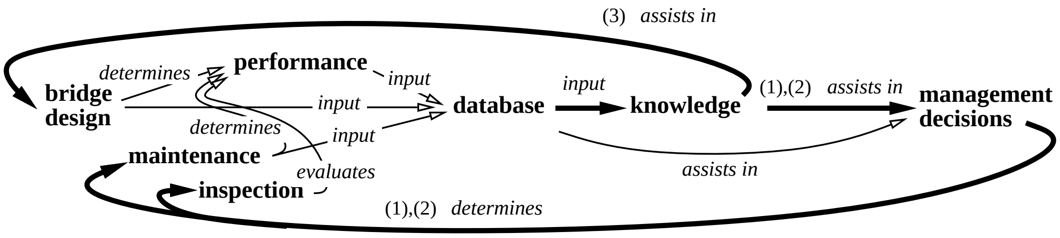

Figure 1. The life cycle flow of information of a bridge

Maintenance management for a deteriorated bridge involves bridge inspection and evaluation, and management decision making. The desire to improve management decision making has led to the establishment of a National Bridge Inventory (NBI) and to the development of computerized bridge management systems (BMS).

The NBI and other inventories in the U.S. contain information on about 550,000 bridges in a format that roughly follows a standard for coding bridge data (FHA 1979); although states and local agencies often use more elaborate information structures adapted to their own needs in their inventories.

Research directed towards improving the design and implementation of bridge management systems is continuously underway (TRR 1990a; TRR 1990b). The guidelines for bridge management systems based on (Hudson et al 1987) are currently in the final approval stages by AASHTO.[4]

These efforts are far from realizing their full potential. First, information stored in bridge inventories is almost never used effectively. Second, bridge management systems are developed based on few fixed algorithms such as: (1) the level of service concept (Isa Al-Subhi et al 1989); (2) statistical analysis for cost estimation (Saito and Sinha 1990); or (3) dynamic optimization (Jiang and Sinha 1989). Together, the ineffective and inflexible use of data can render any approach unsatisfactory.

EFFECTIVE USE OF MACHINE LEARNING IN BMSs

ML techniques can be used effectively on data stored in BMSs by creating knowledge that can be used in future management and design decision making. Knowledge that implicitly resides in bridge databases includes, among other, information about: (1) components of bridges that frequently need maintenance, therefore need to be inspected carefully; (2) historical consequences of maintenance decisions that may inform future decisions; and (3) components of bridges that significantly determine maintenance cost, therefore may inform future bridge designs (see the numbers on the branches in Figure 1). The significance of such knowledge for saving resources is unquestionable (OECD 1992).

As an illustration, Figure 2 contains two rules: one exemplifies the kinds of knowledge that can be extracted by using currently available ML techniques, and the second one that we intend to extract with a new ML technique. While the first rule is useful, it can only be generated if a significant number of bridges with similar cracking problems are recorded. In this rule, "feature-20" stands for the welded stiffening diaphragm that caused failure. Unfortunately, since the naming of bridge features in BMSs is arbitrary, the generation of this rule would be a rare occurrence. It can be generated only if the user of the learning program engages in some laborious data manipulation, knowledge extraction, and careful observation of the results to drive further data manipulation. Besides such careful data manipulation, the use of most available ML programs requires the conversion of parts of the data in the relational database,[5] into a single table that can then be used by ML techniques. Consequently, information embedded in structure of the relational database is lost, thus significantly limiting the scope of knowledge extraction possible.

In contrast to the first rule, the second rule does not refer to a specific idiosyncratic feature such as "feature-20," and can apply to bridges having only one steel girder span, while the rest of the spans are made of concrete. The second rule is based on a richer representation formalism than the first rule and its creation requires less data manipulation. We intend to develop a technique that creates such rules.

A system that learns from BMSs must provide facilities for handling a range of issues in applying ML programs to data (Reich et al 1993). In addition, since different agencies or users would need, to accommodate their own knowledge, bridge management policies or preferences in the learning process, the learning technique must allow for the use of *background knowledge*. Since the information in bridge databases evolves continually, the learning technique must be *incremental*. Finally, since the learning task is complex and different types of extracted knowledge may be requested, a new style of learning, *multistrategy learning* (Michalski and Tecuci 1992),

[4] Note that the BMS from which the guidelines were extracted failed in practice (NCHRP 1989).
[5] The use of relational databases is becoming the standard for BMSs.

Present functionality:

IF	site-minimum-temperature $< -10°F$
AND	year-bridge-built < 1960
AND	bridge-type is girder
AND	bridge-material is steel
AND	feature-20 is a welded stiffening diaphragm
THEN	inspect the welded connection of the girder web to feature-20

Desired functionality:

IF	site-minimum-temperature $< -10°F$
AND	bridge-age > 33 years
AND	bridge has a girder span whose material is steel
AND	the girder has a welded stiffening diaphragm
THEN	inspect the welds of the girder, web to the diaphragm

Figure 2. Examples of rules: Present and desired functionality

must be adopted. In particular, the use of several cooperative learning techniques will be most beneficial.

FLEXIBLE INTEGRATION OF MACHINE LEARNING IN CONTEXT

It is clear to us that no practitioner will use the ML tools discussed previously to their full advantage without an appropriate facility that manages the interaction with them. Such facility is not merely the product of *development*, but of research into *usability*, a major emerging research area in software engineering and information systems. We contend that since the task of extracting knowledge is an engineering task, it requires facilities for flexible modeling of data and information. Flexible modeling can allow engineers to model the kinds of knowledge and data they want to use for solving a particular problem and let these models perform the computations.[6]

n-dim is a tool developed to facilitate flexible modeling in diverse disciplines (Levy et al 1993; Subrahmanian et al 1993).[7] *n*-dim's underlying philosophy is based on numerous observational studies of engineers conducting design activities, a theory about shared memory in design (Konda et al 1992; Monarch et al 1993), and a theory about participatory design (Reich et al 1992). The third generation of *n*-dim is currently being built in an evolutionary prototyping mode.

n-dim does not impose any single representation or abstraction technique on the modeling activities of engineers. Such restrictions may render a sound and promising technique unusable, thus leading to its practical failure. Rather, *n*-dim uses a notion of conceptual information modeling that allows multiple classifications to be imposed over a corpus of information. Abstraction levels are imposed by the users, in whatever way they see fit. Observing the difficulties of diffusing information systems into departments of transportation (Lane and Hartgen 1990), in particular, and similar difficulties in other disciplines, we anticipate that the use of *n*-dim as an infrastruc-

[6] Of course, any algorithm mentioned before for manipulating BMSs' data can be incorporated in such models.

[7] *n*-dim is developed by a group that includes (in alphabetical order): R. F. Coyne, A. Dutoit, E. Gardner, S. L. Konda, S. N. Levy, I. A. Monarch, Y. Reich, E. Subrahmanian, M. Thomas, and A. W. Westerberg.

ture for the management of ML applications can result in an approach with higher chances to diffuse into practice.

The following example illustrates a user friendly contextualized use of ML techniques with n-dim (see Figure 3). Part (a) illustrates the overall model of a BMS created after the schema of the Allegheny County bridge database.[8] It contains models of the relations (i.e., the database conceptual schema) of the database as well as the database itself (`DBMS`). In addition, it contains models of tools available in the BMS including analytical tools and ML tools (`ML_Tools`). Part (b) illustrates the model of one of the relations written in a language derived from the `Frame` language. The ML tools model (not shown) includes several tools such as a model that executes the learning program CN2 (`CN2_Run`), depicted in part (c). In this particular run, only four relations were used and only part of the data was selected as input by the use of a query (`Query1` shown in part (d)). The links in the model `CN2_Run` denote data flow which will be automatically performed via operations attached to the links. Operational parameters for running the program are specified via a model as well (part (e)). Upon running the program, by double clicking on `CN2_Run`, CN2 creates a rule set (`Rules_case1`, part (f)) of which one is shown in part (g). The latter could be the first rule in Figure 2. We intend to develop a similar, more usable, environment for using the ML technique we intend to develop for extracting rules of the second kind in Figure 2.

The context of the rules generated is kept in the form of relations, queries, programs, and operational parameters used in the learning process. All can be searched for via the query language to be used in future learning situations. In such situations, old models can be retrieved, copied, modified, and reused. In addition, the rules created by one learning program can serve as the input of a decision procedure or even a ML program that uses knowledge while learning.[9] This can be done flexibly by creating a model similar to `CN2_Run`. The connection of output from one learning program to the input of another in a flexible manner is a key to successful multistrategy learning.

STATUS AND DISCUSSION

So far, we have experimented with some learning programs on the bridge data of the Allegheny County database and are in the process of integrating ML programs in n-dim.

Observing the data model in of the Allegheny County database and other proposals for bridge data modeling, we observe that these models may lead to problems in using the database. Part of the problems involve data dependencies in database relations, and an overall data model that is "unnatural" such as defining a single relation for bridge features containing up to 5 features each with a different name instead of two relations for bridge features and feature descriptions.

In general, the issue is what constitutes a good bridge database schema? The answer may be different than a schema that is normalized, as recommended in de-

[8] The database contains data on 350 bridges in the county. The data includes general information about each bridge, the present status of the bridge in terms of various limitation on usage, inspection information, repair information, and improvements planned.

[9] In general such use may create problems because the source of the rules is unknown; but in n-dim, the source is maintained with the rules' derivation.

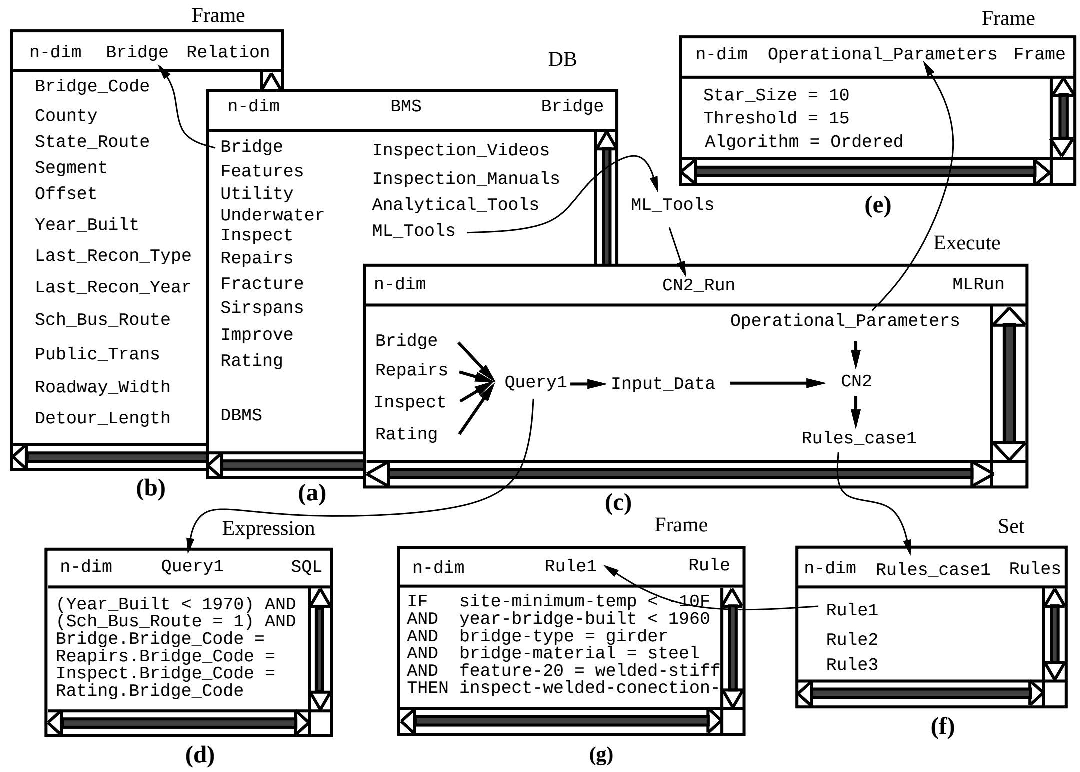

Figure 3. Flexible use of an existing ML technique

signing relational databases, as long as this schema supports the practice of bridge management. Thus, the schema needs to be developed based on the functionality it provides in actual practice. Part of the constraints on the structure will be imposed through the desire to use ML techniques on the data but, in addition, the ML techniques can assist in detecting data dependencies and enforcing integrity constraints (Schlimmer 1993). Thus, we see ML techniques as interacting closely with the practice of bridge management and the evolution of the bridge database schema. In this view, ML techniques not only extract knowledge from bridge databases but also support the maintenance of and addition of data to bridge databases. In order to support the aforementioned interaction, a facility is needed that supports flexible modifications of the conceptual schema while keeping track on the consequences of such modifications on practice.

Although not intended as such, *n*-dim can serve as an excellent framework for a complete BMS. It provides facilities for interacting with several commercial databases, query languages, storage of any material available in an electronic form (e.g., text, pictures, video, audio, etc.), and it allows for attaching tools that can operate on data, such as optimization or scheduling programs. Furthermore, *n*-dim supports the flexible organization of all information and tools. Finally, *n*-dim offers a distributed environment that can be used by multiple users and is being designed to scale up to handle significant amounts of data; this can allow practitioners from different locations or agencies to access a central databank while keeping their personal data separated.

REFERENCES

References

Arciszewski, T. and Rossman, L. A., editors (1992). *Knowledge Acquisition in Civil Engineering*, American Society of Civil Engineers, New York, NY.

Arciszewski, T., Mustafa, M., and Ziarko, W. (1987). "A methodology of design knowledge acquisition for use in learning expert systems." *International Journal of Man-Machine Studies*, 27(1):23–32.

FHA (1979). *Recording and Coding Guide for the Structure Inventory and Appraisal of the Nation's Bridges*, Federal Highway Administration, U.S. Department of Transportation, Washington, DC.

Hudson, S. W., Carmichael, R., Moser, L. O., Hudson, W. R., and Wilkes, W. J. (1987). "Bridge management systems." Final Report NCHRP Report 300, ARE Inc, December.

Isa Al-Subhi, K. M., Johnston, D., and Farid, F. (1989). "Optimizing system level bridge maintenance, rehabilitation and replacement decisions." Final Report FHWA-NC-89-001, North Carolina State University, FHWA, Jan.

Jiang, Y. and Sinha, K. (1989). "Dynamic optimization model for bridge management systems." *Transportation Research Record*, 1211:92–100.

Julien, B., Fenves, S. J., and Small, M. J. (1992). "Knowledge acquisition methods for environmental evaluation." *AI Applications*, 6(1):1–20.

Konda, S., Monarch, I., Sargent, P., and Subrahmanian, E. (1992). "Shared memory in design: A unifying theme for research and practice." *Research in Engineering Design*, 4(1):23–42.

Lane, J. S. and Hartgen, D. T. (1990). "Factors affecting adoption of information systems in state departments of transportation." In *Transportation Research Record No. 1271, Transportation Data and Information Systems: Current Applications and Needs*, pages 61–78, Washington, DC, Transportation Research Board, National Research Council.

Levy, S., Subrahmanian, E., Konda, S. L., Coyne, R. F., Westerberg, A. W., and Reich, Y. (1993). "An overview of the *n*-dim environment." Technical Report EDRC-05-65-93, Engineering Design Research Center, Carnegie Mellon University, Pittsburgh, PA.

Michalski, R. S. and Tecuci, G., editors (1992). *Machine Learning: A Multistrategy Approach. Vol IV*, Morgan Kaufman, San Mateo, CA.

Michalski, R. S., Carbonell, J. G., and Mitchell, T. M., editors (1983). *Machine Learning: An Artificial Intelligence Approach*, Tioga Press, Palo Alto, CA.

Monarch, I. A., Konda, S. L., Levy, S. N., Reich, Y., Subrahmanian, E., and Ulrich, C. (1993). "Shared memory in design: Theory and practice." In *Proceedings of the Invitational Workshop on Social Science Research, Technical Systems and Cooperative Work (Paris, France)*, pages 227–241, Paris, France, Départment Sciences Humaines et Sociales, CNRS.

NCHRP (1989). *Summary of Progress, December 31, 1992*, National Cooperative Highway Research Program, Trasportation Research Board, National Academy Press, Washington, DC.

OECD (1992). *Bridge Management*, OECD, Paris.

Reich, Y. and Fenves, S. J. (1995). "A system that learns to design cable-stayed bridges." *Journal of Structural Engineering, ASCE.* (in press).

Reich, Y., Konda, S., Monarch, I., and Subrahmanian, E. (1992). "Participation and design: An extended view." In Muller, M. J., Kuhn, S., and Meskill, J. A., editors, *PDC'92: Proceedings of the Participatory Design Conference (Cambridge, MA)*, pages 63–71, Palo Alto, CA, Computer Professionals for Social Responsibility.

Reich, Y., Konda, S., Levy, S. N., Monarch, I., and Subrahmanian, E. (1993). "New roles for machine learning in design." *Artificial Intelligence in Engineering*, 8(3):165–181. Special issue on Machine Learning in Design.

Saito, M. and Sinha, K. (1990). "Data collection and analysis of bridge rehabilitation and maintenance costs." *Transportation Research Record*, 1276.

Schlimmer, J. C. (1993). "Sepf-modeling databases: Learning and applying partial integrity constraints to detect database errors." *IEEE Expert*, 8(2):35–43.

Stone, J. and Blockley, D. (1989). "Towards machine learning from case histories." *Civil Engineering Systems*, 5(3):129–135.

Subrahmanian, E., Konda, S. L., Levy, S. N., Reich, Y., Westerberg, A. W., and Monarch, I. A. (1993). "Equations aren't enough: Informal modeling in design." *Artificial Intelligence in Engineering Design, Analysis, and Manufacturing*, 7(4):257–274.

TRR (1990). "Transportation Research Record No. 1271, Transportation Data and Information Systems: Current Applications and Needs..". Transportation Research Board, National Research Council, Washington, DC.

TRR (1990). "Transportation Research Record No. 1276, Maintenance Management 1990: Proceedings of a Workshop.". Transportation Research Board, National Research Council, Washington, DC.