December 1998

# A Framework for Designing Computer-Based Data-Driven Scientific Discovery Systems

Mukesh Rohatgi
*Old Dominion University*

# A Framework for Designing Computer-Based Data-Driven Scientific Discovery Systems

**Mukesh Rohatgi**
Information Systems and Decision Sciences Department
College of Business andPublic Administration
Old Dominion University

## Abstract

*The purpose of this paper is to propose a framework for designing computer-based data-driven discovery systems based on Human Learning Approach[1] (HLA) design philosophy suggested in Rohatgi (1994; 1997). The proposed framework is a generalization of an existing computer-based prototype system designed in accordance with the human learning approach and includes several hierarchically arranged learning modules responsible for acquisition and discovery of knowledge applicable to problem-solving and theory formation.*

## Introduction

The purpose of this paper is to propose an HLA-based (Human Learning Approach) framework for designing computer-based data-driven discovery systems. The proposed framework includes multiple elements which function as building blocks of a knowledge discovery system capable of learning conceptual knowledge in multiple domains. It is envisioned that systems built on the basis of the proposed framework will utilize multiple human learning processes to acquire and discover knowledge which can eventually be used for solving problems in dynamic environments.

The framework outlined in this paper is a generalization of an existing computer-based prototype system implemented in accordance with the HLA design philosophy (Rohatgi, 1994; 1997, also see the footnote). The prototype system, henceforth referred to as the AKBS (Adaptive Knowledge-Based System), was implemented to demonstrate the feasibility of "human learning approach" design method (for further details see Rohatgi, 1994). Furthermore, it was shown that the prototype system is capable of exhibiting learning behavior by achieving the following:

a. Acquisition of object descriptions;
b. Acquisition of associative knowledge to capture relationships between object types;
c. Learning of prototype-based concept descriptions;
d. Dynamic updating of category prototypes with the acquisition of new knowledge;
e. Discovering of complex concepts through conceptual combination;
f. Semantic interpretation of the complex concepts discovered through concept combination;
g. Dynamic updating of complex concepts to reflect the changes in the constituent concepts; and
h. Using previously acquired, learned, or discovered knowledge for problem-solving.

The prototype learning behavior characterized by features listed above was made possible by including multiple integrated learning processes derived from human learning behavior. An object learning process was implemented to acquire object-based descriptions in the form of attribute-value pairs. Relationship descriptions between object types were acquired with the help of an associative learning process. Object descriptions acquired by the object learning process were used by a prototype learning process to generate and update prototype-based concept descriptions which were then used by a complex concept generation process to discover complex concepts through conceptual combination. The process of discovering complex concepts was further continued by assigning suitable semantic interpretations for the discovered complex concepts and finally the knowledge learned and discovered thus far was eventfully used to solve problems related to object classification.

## The Framework

As mentioned previously, the objective of this paper is to generalize the learning mechanisms of the AKBS prototype to form a framework which can be used to design and implement data-driven discovery systems suitable for theory formation and problem-solving. The following generalized forms of AKBS learning processes referred to as required elements of the proposed framework can be integrated to develop systems that use a data-driven approach to knowledge discovery:

---

[1] Use of multiple learning processes derived from human learning behavior in the design of machine learning systems is referred to as the "Human Learning Approach." HLA-based systems use multiple human learning processes for acquisition, modification, and discovery of knowledge in the form of simple and complex concepts along with relationships that relate concepts with each other.

1. The object and associative learning processes used by the AKBS prototype can be generalized to form the basic required element of the framework which is a knowledge acquisition module capable of acquiring observable domain knowledge. Depending upon the scope of the system, the module can be designed either to selectively observe the environment, or to indiscriminately observe the environment. In the case of indiscriminate observation, a selection mechanism to extract relevant information from the input should also be implemented.
2. The second required element is a module capable of building concept descriptions based on observable knowledge made available by the knowledge acquisition module. This module corresponds to the prototype learning process of the AKBS system and should also be capable of creating a taxonomy of domain concepts.
3. The third required element is a module capable of generating nonobservable or theoretical terms derived from domain concepts. A theoretical term in the present context is an abstract higher-level knowledge structure derived from lower-level knowledge structures. Moreover, knowledge required for generating theoretical terms can be formulated in form of domain-independent heuristics. Under appropriate circumstances, domain-dependent heuristics can also be employed. Using the AKBS prototype system as an example, the complex concept generation process used for learning complex concepts resembles the implemented requirements of this module. Depending on the capabilities of the system, the theoretical terms can be recursively combined to generate higher-level forms of abstract theoretical terms.
4. The fourth required element is a module capable of discovering a suitable interpretation for the theoretical terms generated by the previous module. This element of the framework is analogous to the semantic interpretation process used by the AKBS system. The interpretation process results in identification of a relationship between the elements used to generate the theoretical term. When the constituents of the theoretical term are ground level concepts, the resulting interpretation can be called a simple proposition. On the other hand, if the constituents of a theoretical term are lower-level theoretical terms, then the interpretation would result in a higher-level abstract proposition. The set of all propositions discovered by the system can be said to represent a theory applicable to the domain currently being addressed by the system. Similar to the complex concept generation process, the knowledge used for discovering plausible relationships between the elements of a theoretical term can be represented in form of heuristics.
5. Finally, the theories and propositions postulated by the interpretation module should be tested by incorporating a problem solving module in the system. This module will use the acquired and discovered knowledge of the system to solve problems. Success in problem solving should lead to increased confidence in the knowledge used to solve the problem, while failures should have the opposite effect, and may require the knowledge to be revised. This module is the generalized form of problem-solving processes used by the AKBS system.

The framework as described above consists of a fundamental building block responsible for the acquisition of observable knowledge which leads to theory formation through learning and discovery with the help of modules described in items 2, 3, and 4. Validity of the discovered knowledge can be determined through successful problem-solving efforts based on acquired, learned or discovered knowledge. Furthermore, intelligence used for generation and interpretation of theoretical terms and propositions can be implemented with the help of heuristics. If the heuristics are innate to the system, then the system should be classified as a learning system. On the other hand, if the system is capable of learning the heuristics on its own, then it should be classified as a system capable of "learning to learn."

## Validation

Currently, learning systems that implement all aspects of the proposed framework do not exist. Hence, an example that can serve as the “proof of concept” for the proposed framework is not available. Although, the AKBS system described briefly in this paper (for details see Rohatgi, 1994) does provide a partial validation of the framework.

Features of the AKBS system that correspond to the framework elements have been described in the previous section. It can be surmised from this description that most of the framework elements are present in the AKBS system but their reach does not extend to the level required by the framework. Our plan for the future is to extend the learning capabilities of the AKBS system by implementing all the features incorporated in the proposed framework and use the learning behavior of the enhanced system to validate the framework.

## Conclusions

The purpose of this paper was to proposed a framework for designing computer-based data-driven scientific discovery systems. The proposed framework is based on human learning approach suggested in Rohatgi (1994) and is being presented as a generalization of an existing computer-based prototype system implemented in accordance with the dictates of the HLA design philosophy.

Most of the currently existing discovery systems described in machine learning literature (see Langley et al., 1987, for examples) tend to follow ad hoc approaches and almost none of them have taken any cues from human learning behavior. One system that closely mirrors the proposed framework is PI which is described in Holland et al. (1986) but unlike the framework proposed in this paper which is based on human learning approach, the PI system works with ad hoc formalisms developed to facilitate the implementation of the system.

In our opinion, the proposed framework is more robust and represents a domain-independent abstraction of the scientific discovery process. Its utility is further highlighted by the fact that it is based on human learning approach which has proven its adaptiveness and survivability over time.

## References


Holland, J. H., Holyoak, K. J., Nisbett, R. E., and Thagard, P. R., Induction: Processes of Inference, Learning and Discovery, MIT Press, Cambridge, MA, 1986.

Langley, P., Simon, H. A., Bradshaw, G. L., and Zytkow, J. M., Scientific Discovery: Computational Explorations of the Creative Processes, MIT Press, Cambridge, MA, 1987.

Rohatgi, M., "A human Learning Approach For Designing Adaptive Knowledge-Based Systems," Ph.D. Dissertation, Dept. of Information Systems, Texas Tech University, Lubbock, Texas, 1994.

Rohatgi, M., "A Human Learning Approach For Designing Adaptive Knowledge-Base Systems," Proceedings of the 1997 Americas Conference on Information Systems, Indianapolis, Indiana, August 1997.