Application of Empirical Discovery in Knowledge Acquisition

Jan M. Żytkow **Jieming Zhu**
Computer Science Department, Wichita State University
Wichita, KS 67208 U.S.A.
e-mail: zytkow@wsuiar.wsu.ukans.edu

ABSTRACT

There is no doubt that the most fundamental method of knowledge acquisition is discovery, but the AI subfield of Knowledge Acquisition neither studies nor uses discovery methods. We argue that machine discovery is approaching the stage at which it can be useful to knowledge acquisition in two ways: as a source of useful techniques, and as a model of unified knowledge representation and application. We present the discovery system FAHRENHEIT and we discuss its various real-world applications: automated experimentation and discovery in a chemistry laboratory, mining databases for useful knowledge, and others, demonstrating FAHRENHEIT's potential as a knowledge acquisition aid. Finally, we discuss the new developments in the area of discovering basic laws and hidden structure, and we note that automation of modeling would close the cycle of automated knowledge acquisition and application.

1 Introduction

Despite progress, the state of the art in knowledge acquisition is characterized by an amalgamation of incompatible techniques, lack of standards, and narrow applications (Boose & Gaines 1989). In our search for a solution to these problems, we should consider the best mechanism for knowledge acquisition developed by humanity: modern science and engineering. Science and engineering offer us a unifying knowledge organization spanning many levels: data, empirical regularities, models, basic laws, and principles. Science and engineering offer proven methods of acquiring and verifying knowledge. Although the scientific metaphor is becoming popular in the field of knowledge acquisition, and although researchers acknowledge that building a knowledge-based system resembles the construction of a scientific theory, the scientific method has been used only in a very limited way for knowledge acquisition and representation. There are several reasons. First, although separate elements of the scientific method have been reconstructed, many important elements are still missing. Second, although many attempts have been made at integration of various discovery capabilities (IDS: Nordhausen & Langley 1990; BLAGDEN: Sleeman, Stacey, Edwards, & Gray 1989; FAHRENHEIT: Zytkow 1987; Langley & Żytkow 1989), each integration is limited to only a few elements of the scientific method. Third, the applications of machine discovery are believed to be still in the research phase.

We concentrate on the third argument. Although no-one objects to the claim that science and the scientific method are worth understanding and automating, the feasibility of the task raises doubts, in both the short and long run. To show that practical results are possible we summarize the FAHRENHEIT project, which is focused on multi-dimensional experimentation and data analysis. We demonstrate the breadth and importance of real-world applications of FAHRENHEIT. Then, in response to the first argument, we summarize progress that has been made in the last few years on automation of other elements of scientific method. We concentrating especially on discovery of basic laws and discovery of hidden structure. Finally, we consider automation of knowledge application. Although automated model construction is the key to the acquisition of practical knowledge (Morik 1989; Żytkow & Lewenstam 1990), so far little has been done in that area and significant progress is needed.

In a short run, FAHRENHEIT can be viewed as another incompatible technique added to the repertoir of knowledge acquisition methods. But from a larger perspective, the results in the area of machine discovery fall into a master plan, as gradually reconstructed components of a unified method of experimental science. This approach is guaranteed to succeed because scientific method and scientific knowledge representation form a solid and proven theoretical platform for knowledge acquisition.

We concentrate on automated discovery rather than on user support, exemplified by systems such as BLIP (Morik 1989; Wrobel 1989), QuMAS (Mozetic 1987) and DISCIPLE (Kodratoff & Tecuci 1989).

2 Discovery of patterns in empirical data

In this section we review the research program in machine discovery based on FAHRENHEIT. Because we summarize a large research program and many results, the review will

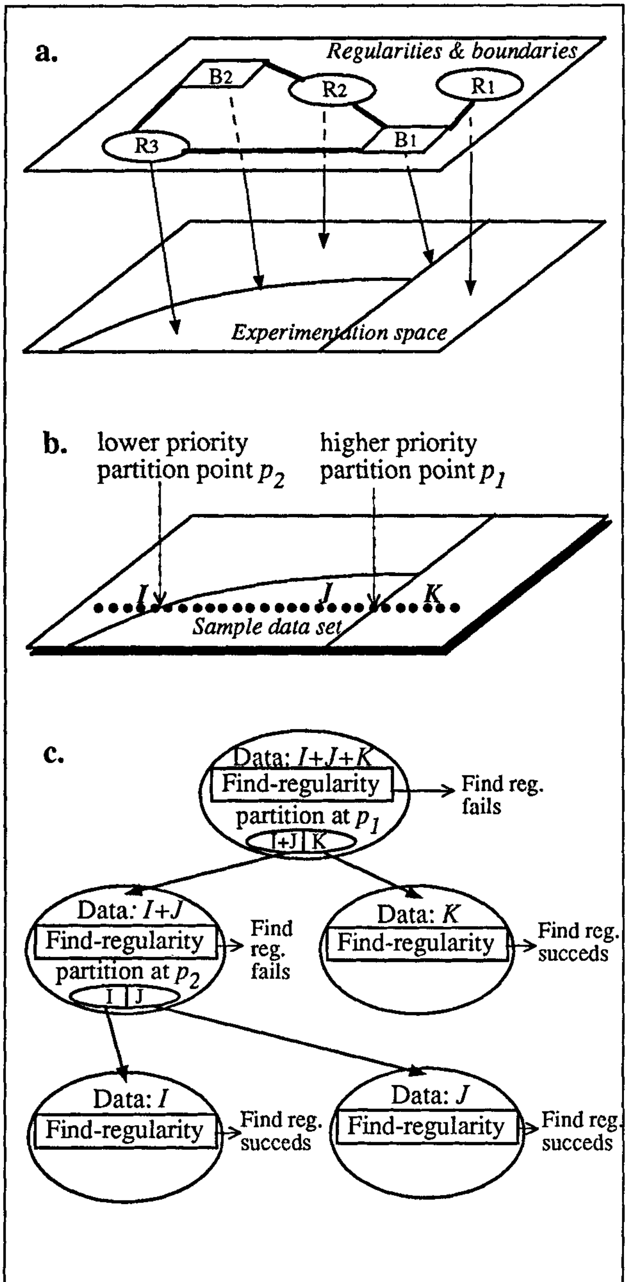

Figure 1. Finding multiple regularities.
(a) Knowledge representation for multiple regularities and their boundaries.
(b) A set of data that crosses regularity boundaries.
(c) The partitioning search on data in (b).

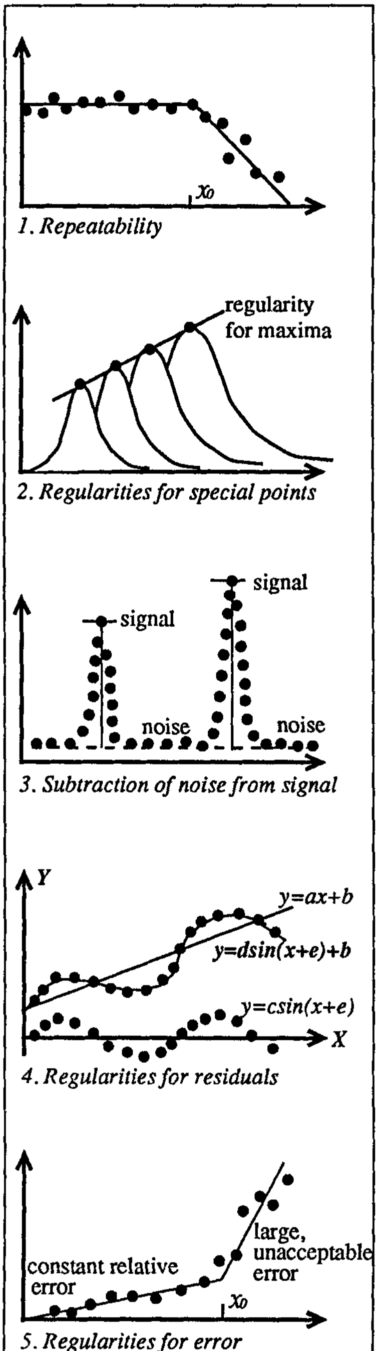

Figure 2. Applying BACON-1 (find-regularity) in new roles.

be done at the level of major components and related goals and we will make an unusual number of references to our own papers. FAHRENHEIT uses a well-known BACON system (Langley et al 1987), so we will use frequent comparisons with BACON.

2.1 The task of FAHRENHEIT

At the outset, FAHRENHEIT is given N independent variables $x_1, ..., x_N$ and one dependent variable y (also indicated as x_0), each limited in scope to a set of values $V_i, i = 0, 1, ..., N$.. The possible values of these variables form a cartesian product of $N + 1$ dimensions. FAHRENHEIT is given experimental control over the values of all independent variables and is supposed to find a regularity for the dependent variable y, which includes as many independent variables as possible. This has been the task of BACON, but while BACON has been able to succeed when there was one regularity in the whole block, FAHRENHEIT can discover several multidimensional regularities and the scope of each regularity within the block. The scope of each regularity is defined by a condition that must be satisfied by independent variables. Figure 1.a shows the space of two independent variables and three regularities R_1, R_2, R_3 divided by boundaries B_1 and B_2. The upper part of Figure 1.a shows the linked structure which is built by FAHRENHEIT to represent these regularities and their boundaries.

In addition to finding multiple regularities in data, FAHRENHEIT can also perform many other tasks in data analysis. We will discuss these in the next subsection.

2.2 Alternatives to regularity finding

Quantitative discovery systems were traditionally limited to regularity detection, whereas scientists are also interested in discovery of "special points" or patterns such as maxima, minima, discontinuities, and the like. Sometimes, finding a special point is more important than detecting a regularity. The maxima in data can indicate various chemical species; the maximum location indicates the type of ion, while the maximum height indicates the concentration. Discontinuity may indicate a phase change.

In addition to BACON's capability for detection of a single regularity for all data, FAHRENHEIT is capable of detection and analysis of such "patterns" in data as:

1. maxima and minima,
2. inflection points,
3. discontinuities,

4. changes of slope,

5. zeros and the values of x for which $y = a$,

6. boundaries of regularities.

This leads to a considerable growth of discovery power.

To understand the modularity of FAHRENHEIT it is important to notice that detection of special points is an alternative to regularity finding in formal terms of their inputs and outputs. Consider maxima as an example. A maximum is defined as a datapoint (x, y) which is higher (has the y value greater) than some points $x1$ and $x2$ on both sides of x by more than empirical error, and there are no points in between $x1$ and $x2$ higher than y. The input is a set of datapoints and their errors, in exactly the same format as required for the regularity finder. The output is a list of maxima. Each maximum is described by the location and error for location, and the height and error for the maximum height. The description of a regularity is analogous. It includes the values of parameters, such as slope and intercept, and the error of each parameter.

2.3 Goal structure

FAHRENHEIT combines several searches (Żytkow 1987). Each search corresponds to a particular goal. Our informal presentation is very close to the actual implementation of the system. We concentrate on the high level goals and we skip lower level details.

We will examine the circumstances in which different searches are called, starting from the situation in which BACON-3 has successfully completed the search for regularities for M independent variables, $M < N$ (N is the total number of independent variables):

1. Generalize to another dimension (cf. the Ordering search in the space of generalizations, Żytkow 1987). The instances of the generalization operator correspond to the remaining independent variables $x_{m+1}, ..., x_N$. They can be ordered by the use of the relevance relation between the dependent and independent variables).

2. Find a new non-investigated area in the M-dimensional subspace of independent variables. The Find-new-area search which is called to solve this problem returns several seeds which are M-tuples of independent values for which no regularity has yet been found. Each seed can be used to start the BACON-3 experiments in search for a new regularity for the given M independent variables and for y.

In the situation in which the Find-regularity search (BACON-1) for x_i or any module that detects special points has successfully returned to BACON-3 (Figure 3) the following goals can be invoked:

1. Generalize the regularity in the dimension x_i. Invoke the Boundary search (Żytkow 1987; Langley & Żytkow 1989) in order to find the scope of the regularity.

2. Apply Find-regularity to detect a regularity on maxima, or a regularity in the sequence of other special points.

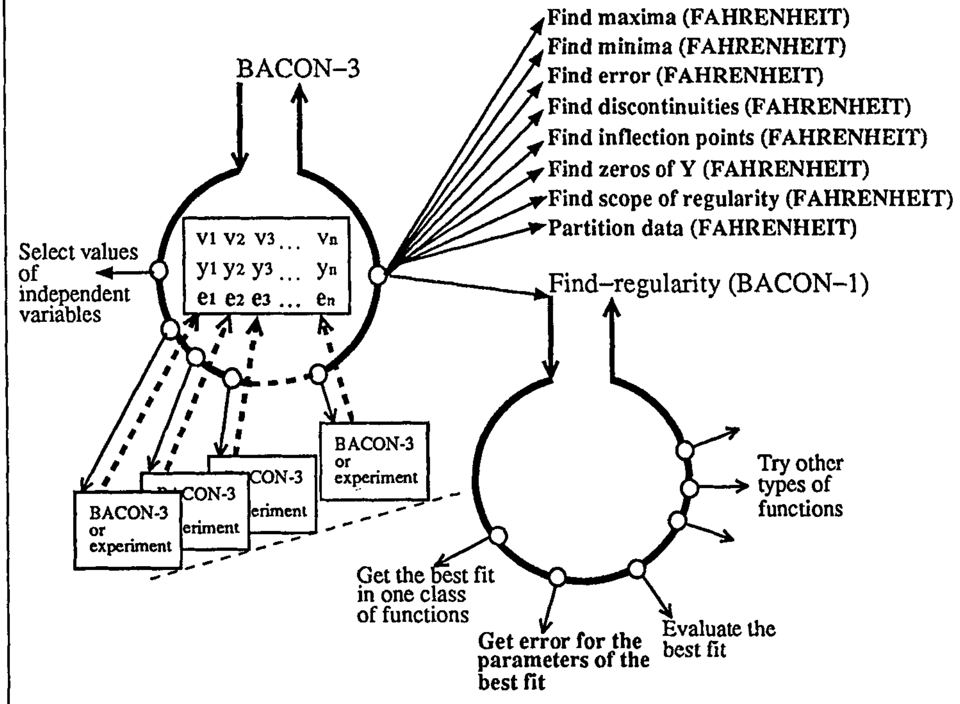

Figure 3. BACON-3 and BACON-1 searches in the FAHRENHEIT system. All additions to BACON are in the bold face.

In the case when the Find-regularity has been unsuccessful, FAHRENHEIT can take several steps.

1. Analyse the residuals, that is the differences $y - f(x)$ between predictions of the unsuccessful fit $f(x)$ and the actual data (Figure 2.4). Search for a regularity in residuals expands the curve fitting capability of FAHRENHEIT. The example in Figure 2.4 shows that Find-regularity, which can find a linear fit and $a \times sin(x+b)$, but not $a \times x \times sin(x+b) + c$, can discover the latter function as a result of the analysis of residuals.

2. Partition the data into several subsets and call Find-regularity on each. Figures 1.b and 1.c illustrate the Partition search in a situation in which the initial dataset

($I + J + K$ in Figure 1.b) belongs to three regularities. Find-regularity fails when called on the whole dataset. Search for the regularity in residuals also fails, but after the partitioning of data into subsets I, J, and K, three regularities are eventually detected (Figure 1.c).

The maxima, discontinuities, and the like, can be treated as patterns to be discovered in data, but they can be also treated instrumentally as hints for data partitioning. To consider all partitions of a large dataset into subsets would be a task of enormous complexity. The search can be reduced by hints about the likely partition points. Figures 1.b and 1.c illustrate our Partition search. Partition summarizes information about all types of special points. Each occurrence of a maximum, zero, or any other special point is counted as one reason for partitioning. As a result a list of likely partitioning points is returned, ordered by the number of reasons for partitioning data. In Figure 1.b, p_1 has the highest number of reasons, p_2 is the second. FAHRENHEIT partitions data gradually, starting at the points with the highest number of reasons. In our example, it uses p_1 first, and then p_2.

If the entire search for regularities in the given dataset for x and y fails, use Find-new-area for x (described earlier) then take a seed and call BACON-3 again.

2.4 Recursion and modularity in goal generation

BACON-3 is a recursive mechanism which detects an N-dimensional regularity step after step. One independent variable is varied at each step, and the successful search adds one dimension to the regularity. FAHRENHEIT expands that recursive mechanism to include new modules for special points detection. As we discussed in section 2.2, all those modules are compatible with Find-regularity in terms of their inputs and outputs. Each can be substituted in BACON-3 for the call to Find-regularity, as illustrated in Figure 3, and the results can be used at higher levels of BACON-3 recursion. This is an important theoretical result, because by addition of modules for special points detection we expand considerably the discovery power without substantial changes to the control mechanism.

The recursive mechanism allows one to mix and match various goals in data analysis. For instance, FAHRENHEIT can search for the regularities on minima, or for regularity on a boundary of a regularity on maxima. The latter could be interpreted as an equation for a surface of a phase change.

3 Applications of FAHRENHEIT

Not only the whole FAHRENHEIT system but also individual modules can be applied in numerous ways. Before we present various applications of FAHRENHEIT, let us focus on applications of the Find-regularity module, which looks for one or more regularities on the values of y as a function of x.

3.1 Different uses of one-dimensional regularity finder

The applications discussed below appear in the same order as the individual diagrams in Figure 2.

1. **Repeatability.** The scientific method requires that the repeatability analysis is conducted before the data can be used. If an experiment is repeated in the same circumstances, we expect the same results of measurements. In the real world, however, we cannot control all variables and some variables are expensive to control. We must accept limited repeatability, but it is essential to know the conditions within which experiments are approximately repeatable, and the measurement error specific to these conditions. We divide the independent variables given to FAHRENHEIT into two categories (Żytkow, Zhu, and Hussam 1990a): independent variables for which we want to build a theory and those independent variables which we want to abstract away but the values of which can be controlled. FAHRENHEIT starts from the repeatability analysis, concentrating on the latter class of variables. It performs experiments in which it varies the values of those variables, keeping constant the variables in the first category. Then it searches for regularities in data, paying particular attention to the constancy of the dependent variables. The scope of constancy is used as the range of repeatability. Figure 2.1 shows a considerable range of repeatability for all $x < x_0$. If no constant regularity can be discovered in a sequence of data for a particular independent variable, the value of that variable must be fixed to allow for repeatability.

2. **Regularities for special points.** Find-regularity can take a sequence of numbers that represent, for instance, heights of the maxima in data (plus error for each maximum height) as a function of an independent variable, and search for regularities on maxima.

3. **Noise estimate.** The heights of the maxima in Figure 2.3 could represent the concentration of several ions, but in order to obtain more precise data about con-

centration, we must find the noise and subtract it from the maxima heights. The noise can be detected as a regularity or regularities for the data points between peaks, indicated by a dashed line in Figure 2.3.

4. **Regularity for residuals.** Figure 2.4 illustrates the way in which the analysis of residuals expands the power of Find-regularity. Details have been described in section 2.3.

5. **Regularity for error.** As we discussed in item 1, above, repeatability analysis allows FAHRENHEIT to obtain the values of error. Find-regularity can detect a regularity on the size of error, such as constant absolute error, constant relative error, and the like. However, Find-regularity needs to know the error of error. The error of error is provided by the repeatability study as proportional to the error of standard deviation.

3.2 Discovery in a science laboratory

In virtually any physics or chemistry laboratory around the world we can find many tasks that can be interpreted as problems for FAHRENHEIT. All variables manipulated by the experimenter can be viewed as independent variables for FAHRENHEIT that form the N-dimensional product discussed in the introduction to section 2. The responses measured by the experimenter correspond to dependent variables.

FAHRENHEIT can be used as an automated system for data acquisition and analysis. Each independent variable must be physically interpreted by an output link to a particular manipulator, while each dependent variable by an input link from a sensor. This allows our discovery system to autonomously run a complete cycle in which it controls the experiments, collects data, and builds theories based on data analysis. Human intervention is reduced to the preparation of the initial experimental situation and occasional assistance.

We have conducted many experiments in the domain of differential pulse voltammetry, charging FAHRENHEIT with various tasks (Żytkow, Zhu, & Hussam 1990,1990a). Some experiments involved collection of many thousand data points and discovery of many regularities. The accuracy has been compatible with the accuracy achieved by human researchers. The values of error estimated in our tests are approximately the same as the values determined by a chemist and the regularities found by FAHRENHEIT were either equivalent to those of the chemist within empirical error or more accurate. In several cases our system detected a more complex and precise regularity than the chemist, or found a regularity (linear) in the cases in which the chemist did not look for it, believing that the results must be constant.

FAHRENHEIT returns the results in a much shorter time than human competitors. We found that what typically required several days of work for research assistants, FAHRENHEIT completed in 50 minutes.

Our results demonstrate that a quantitative discovery system can be used in a chemistry laboratory on an experimental problem of interest to a contemporary chemist and that a scientist might not only save enormously on time and effort spent on data analysis and derivation of empirical equations, but that the accuracy of results might improve.

3.3 Discovery of useful knowledge in databases

Any table of N attributes in a relational database allows us to define the N-dimensional space, which is the cartesian product of sets of values of all attributes. Finding regularities in such a block sounds like a typical task for FAHRENHEIT, but data in databases are sparse, no experiments can provide more data to allow for focussing on particularly interesting areas, and regularities are typically very poor, so they can be captured by contingency tables or weak correlations. Because of these differences, we constructed a descendent of FAHRENHEIT (Forty-niner: Żytkow & Baker 1991) and we applied it to the task of mining databases for useful Regularities. In a typical business, health-related, educational, scientific, or engineering enterprise, a large volume of data is available which represents knowledge accumulated over a long interval of time at considerable effort, and it is usually organized in a relational format. The data may be mined for useful trends and regularities. In a typical database, Forty-niner finds many regularities which are statistically significant, but much weaker than functional regularities satisfied within small error, which are entertained in physics and chemistry.

3.4 Experimental investigation of computational complexity

A computer program can be made available to FAHRENHEIT for the purpose of experimental study of computational complexity. Independent variables are the program parameters, while the dependent variables are the time or storage required for the computation. FAHRENHEIT can experiment with a given program in a similar way as it experiments with a physical situation, by changing the values of program parameters and recording the duration of each computation. Based on these data, FAHRENHEIT builds the theory that estimates computational complexity of the program in the same way as in the case of scientific data. The irregular time spent on garbage collection can cause problems which require the detection of outliers.

Phase changes are common in computer programs, similar to phase changes that occur in physical systems. Consider a parameter P which influences the program logarithmically starting from some threshold value p_0. For $p < p_0$, $C(p)$=const; for $p > p_0$, $C(p) = log(p)$. Both the Partition search in FAHRENHEIT (Figure 1.b, 1.c) and the Find-new-area search can handle phase changes in programs. For the worst case analysis, it will first find the maxima for time or storage, and then the regularity on maxima.

3.5 Analysis of abstract spaces

As a similar task, FAHRENHEIT can try to empirically analyse the function computed by a program. This may be very useful. When we define a particular altitude function to be used in hill climbing, the properties of that function relevant to the hill climbing search are often unclear. FAHRENHEIT can find useful information about the presence and distribution of local maxima, about their number and about the useful increment for the elementary step in hill climbing.

3.6 Automated knowledge acquisition by robots

Manipulatory skills can be developed as a result of experiments and theory formation. We applied FAHRENHEIT in a simulation in which a robot arm learns how to handle physical objects (Żytkow & Pachowicz 1989). Many regularities and many special points have been found, such as boundaries, maxima and minima. The center of gravity of solid objects has been discovered as a special point common in various experiments. The discovered theory can yield the rules which help to improve the efficiency and quality of future manipulations.

3.7 Discovery of patterns in sequences

Consider a sequence of numbers common to intelligence tests. MS.SPARC is another mutant descendent of FAHRENHEIT which handles regularities in sequences (Stefanski & Żytkow 1989). In order to avoid approximate solutions because they are not considered correct, MS.SPARC does not allow for any error in the fit.

3.8 Knowledge engineering aid

All previous applications can support expert knowledge acquisition. They can generate calibration curves and performance measurements such as complexity estimation. Rather than asking a human expert for his best guess of the values for certainty factors, we may apply FAHRENHEIT or Forty-niner on the relevant database to extract the precise values of coefficients, and to give a statistically sound estimation of the coefficient errors.

4 The cycle of knowledge discovery and application

Before we discuss our view of the full cycle of knowledge discovery and application, we briefly summarize the growth of discovery systems in the last three years. Many new systems have been developed, some of which are depicted at the top level in Figure 4. Several abilities lacking in earlier discovery systems have been introduced, primarily the ability to consider the empirical context of a law (IDS: Nordhausen and Langley, 1990; GALILEO: Zytkow 1990; Sleeman, Stacey, Edwards, and Gray, 1989), the ability to design experiments (KEKADA: Kulkarni and Simon, 1987; FAHRENHEIT: Zytkow, 1987; Langley and Żytkow 1989), the ability to represent objects, states and processes (Nordhausen and Langley, 1990; Żytkow 1990), and the ability to reason by analogy (Falkenhainer, 1987; Falkenhainer and Rajamoney, 1988). Several systems discover hidden components and their properties (REVOLVER: Rose 1989; GELL-MANN: Fischer and Zytkow 1990) or discover hidden properties of observable objects (BR-3: Kocabas 1991). Sleeman et al. (1989) propose an interesting search in the space of qualitative models of a chemical system. Progress has been made in the domain of discovering qualitative regularities (IDS; BLIP: Morik 1989, Wrobel 1989). Significant progress has been made also on the important issue of integration. Two systems, both descendents of BACON, reached a considerable integration: IDS and FAHRENHEIT, the latter augmented by the GALILEO system that generalizes knowledge by decomposing empirical equations into simpler expressions.

4.1 Transformation of empirical equations into basic laws

Many discovery systems, including BACON, FAHRENHEIT, and ABACUS (Falkenhainer & Michalski 1986), produce algebraic equations that summarize numerical data. Hereafter, we will call these BACON-like systems. Algebraic equations discovered by such

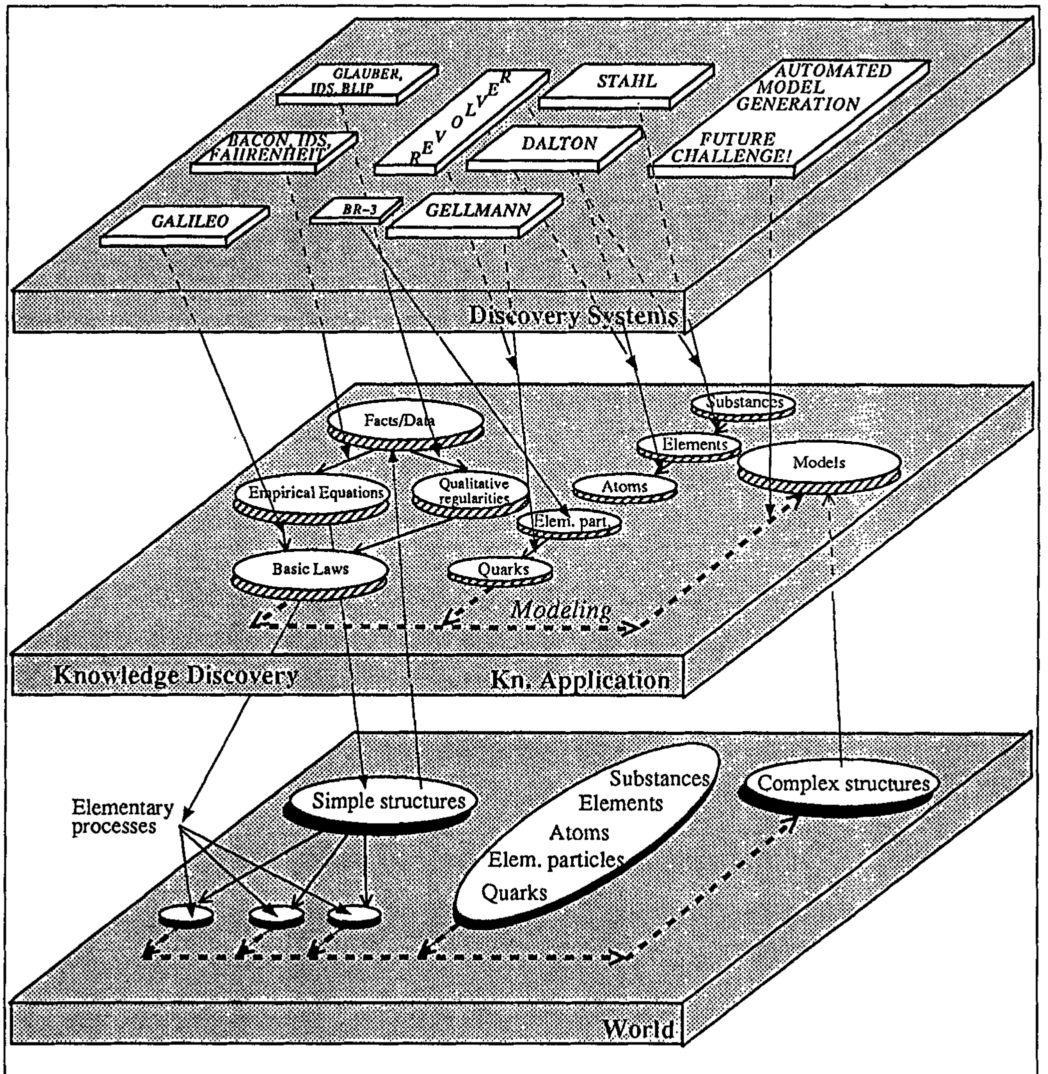

Figure 4. Overview of scientific discovery systems.

systems can be interpreted as quantitative laws of different sciences. Each law has been discovered by the empirical investigation of a particular situation. Physical situations can be varied in countless ways and different situations are usually described by different equations. If our discovery capability was limited to BACON-like systems we would have to discover the equations for each situation individually. GALILEO (Żytkow 1990) is a discovery system that has been developed to address this problem. It transforms equations generated by a BACON-like system into a form compatible with the structure of physical processes described by those equations, so that equations can be decomposed into expressions that describe elementary components of the physical structure, such as the expressions for kinetic or potential energy of a body, or energy disengaged during the change of phase. This type of analysis led scientists to a relatively small number of basic expressions which can be combined in countless ways to build models of new situations.

4.2 Discovery of hidden structure

Basic laws are not sufficient for knowledge application to a particular situation. Scientists must also know the structure for which to build an adequate model. Discovery of structure is the task complementary to finding laws and an autonomous discovery system must include both. Directly non-observable, hidden structure causes particular problems. Research on the discovery of hidden structure has resulted in many systems, each capable of dealing with limited, yet historically important cases in physics and chemistry (Figure 4). Among the systems that emerged from this research is REVOLVER (Rose 1989), which constitutes the first attempt at grasping generality in the process.

In the GELL-MANN system that discovers quarks (Fischer & Żytkow, 1990), hidden structure is described by hidden objects and their properties, by a pattern in which hidden objects combine to form observable particles, and by a particular combination of hidden objects for each observable particle. The earlier discovery systems, such as STAHL, DALTON, and REVOLVER did not postulate properties of hidden objects and have had a limited capability for postulating hidden objects. GELL-MANN finds the quark model that is accepted in physics and determines its uniqueness, which is critical for the model confirmation. Surprisingly, GELL-MANN has found a model that explains meson octet by two quarks and two antiquarks, and when allowed for multiples of 1/3 as strangeness values, it discovered the second quark model for the hadron octet. Both these new models do not, however, allow to reconstruct all particles which are explained by the accepted model.

4.3 Modeling

Knowledge of basic laws and knowledge of structure allow scientists efficient knowledge application in the form of model construction (Figure 4). To achieve substantial progress in the domain of **automated knowledge application, sharing, and interchange,** we must automate scientific modeling. Although modeling plays a major role in different areas of AI, applications are fragmented and the essence of scientific modeling is far from being understood. On average, scientists spend far more time on model construction than on development of basic theories, it is difficult to find a systematic account of modeling.

Żytkow (1990) argues that decomposition of knowledge into basic laws and basic structural components, and a capability for recombining them into models, will remedy the explosion in the size of the knowledge-bases. Moreover, Żytkow and Lewenstam (1990) provide a blueprint for automated model construction. Implementation of such a system would close the loop of automated knowledge discovery and application.

5 Conclusions

It is time to consider seriously machine discovery as a unifying schema for knowledge acquisition and representation. We demonstrated that FAHRENHEIT as a whole, and individual modules such as equation finder (BACON-1), can be used on many tasks, and that they can become tools for efficient and effective knowledge acquisition, especially when used in synergistic collaboration with a human knowledge engineer.

We noted the significant progress in automation of other discovery tasks, and we argued that the basic cycle of knowledge generation and application will be closed when we automate model generation.

Acknowledgement: The work described in this paper was supported by Office of Naval Research under the grants No. N00014-88-K-0226 and N00014-90-J-1603.

References

Boose, J.H., & Gaines, B.R. (1989) Knowledge Acquisition for Knowledge-Based Systems: Notes on the State-of-the-Art. *Machine Learning, 4*, 377-394.

Falkenhainer, B.C. (1987) Scientific Theory Formation Through Analogical Inference, *Proceedings of Fourth International Workshop on Machine Learning*, Los Altos, CA: Morgan Kaufmann Publ., 218-229.

Falkenhainer, B.C., & Michalski, R.S. (1986) Integrating Quantitative and Qualitative Discovery: The ABACUS System, *Machine Learning, 1*, 367-401.

Falkenhainer, B.C., & Rajamoney, S. (1988) The Interdependencies of Theory Formation, Revision, and Experimentation, *Proceedings of the Fifth International Conference on Machine Learning.* Morgan Kaufmann Publ.: Los Altos, CA, 353-366.

Fischer, P., & Żytkow, J. (1990) Discovering Quarks and Hidden Structure, in: Ras Z. ed. *Methodologies for Intelligent Systems 5*, Elsevier, New York, NY, 362-370.

Kocabas, S. (1991) Conflict Resolution as Discovery in Particle Physics. To be published in *Machine Learning, 6.*

Kodratoff, Y. & Tecuci, G. (1989) The Central Role of Explanations in DISCIPLE, in Morik K. ed. *Knowledge Representation and Organization in Machine Learning*, Springer-Verlag, Berlin, 135-147.

Kulkarni, D., & Simon, H.A. (1987) The Processes of Scientific Discovery: The Strategy of Experimentation, *Cognitive Science, 12*, 139-175.

Langley, P.W., Simon, H.A., Bradshaw, G., & Żytkow J.M. (1987) *Scientific Discovery; An Account of the Creative Processes.* Boston, MA: MIT Press.

Langley, P.W., & Żytkow, J.M. (1989) Data-Driven Approaches to Empirical Discovery. *Artificial Intelligence, 40*, 283-312.

Morik, K. (1989) Sloppy Modelling, In Morik K. ed. *Knowledge Representation and Organization in Machine Learning*, Springer-Verlag, Berlin, 107-134.

Mozetic, I. (1987) The role of abstraction in learning qualitative models. *Proc. Fourth Int. Workshop on Machine Learning.* Morgan Kaufmann, 242-255.

Nordhausen, B., & Langley, P. (1990). An Integrated Approach to Empirical Discovery. in: J.Shrager & P. Langley eds. *Computational Models of Scientific Discovery and Theory Formation*, Morgan Kaufmann Publishers, San Mateo, CA, 97-128.

Rose, D. (1989) Using Domain Knowledge to Aid Scientific Theory Revision. *Proceedings of the Sixth Int. Workshop on Machine Learning*, Morgan Kaufmann Publishers, San Mateo, CA, 272-277

Sleeman, D.H., Stacey, M.K., Edwards, P., & Gray, N.A.B. (1989) An Architecture for Theory-Driven Scientific Discovery, *Proceedings of EWSL-89.*

Stefanski P. & Żytkow Jan M. (1989) A Multisearch Approach to Sequence Prediction, in: Ras Z. ed. *Methodologies for Intelligent Systems 4*, New York, NY: Elsevier, 359-366.

Wrobel, S. (1989) Demand-driven Concept Formation. In Morik K. ed. *Knowledge Representation and Organization in Machine Learning*, Springer-Verlag, Berlin, 289-319.

Żytkow, J.M. (1987) Combining many searches in the FAHRENHEIT discovery system. *Proceedings of Fourth International Workshop on Machine Learning.* Los Altos, CA: Morgan Kaufmann, 281-287.

Żytkow, J.M. (1990) Deriving basic laws by analysis of processes and equations, in: J.Shrager & P. Langley eds. *Computational Models of Scientific Discovery and Theory Formation.* Morgan Kaufmann Publishers, San Mateo, CA, 129-156.

Żytkow, J. & Baker, J. (1991) Interactive Mining of Regularities in Databases, To be published in: Piatetsky-Shapiro, G. and Frawley, W. eds. *Knowledge Discovery in Databases,* The AAAI Press, Menlo Park, CA.

Żytkow, J.& Lewenstam A. (1990) Analytical Chemistry: the Science of Many Models, *Fresenius Journal for Analytical Chemistry*, 225-233.

Żytkow, J. & Pachowicz, P. (1989) Fusion of vision and touch for spatio-temporal reasoning in learning manipulation tasks, in: *Proceedings of SPIE's Advances in Intelligent Robotic Systems*, 404-415.

Żytkow, J.M., Zhu, J. & Hussam, A. (1990) Automated Discovery in a Chemistry Laboratory, *Proceedings of the AAAI-90*, AAAI Press, 889-894.

Żytkow, J.M., Zhu, J. & Hussam, A. (1990a) Determining Repeatability and Error in Experimental Results by a Discovery System, in Ras Z. ed. *Methodologies for Intelligent Systems 5*, Elsevier, New York, NY, 438-445.