

Automated Pattern Recognition: Self-Generating Expert Systems for the Future

Thomas L. Isenhour
Utah State University, Logan, UT 84322

Accepted: July 1, 1985

Chemometrics and pattern recognition had their start in chemistry in the late 1960's. The most recent review of the area by Michael DeLaney listed 438 journal articles and books. The three most important areas of future development will be Expert Systems, Relational Data Bases, and Robotics. It should now be possible to combine existing robotics and artificial intelligence software to create a system which will generate its own expert systems using relational data bases. The data will be in the chemical domain and the system I describe we are calling the Analytical Director. The Analytical Director will be an artificial intelligence/robotic expert system for the analytical laboratory. The Analytical Director will develop, test, implement and interpret chemical analysis procedures. It will learn from its own experience, the experience of others and communicate what it has learned to others. The Analytical Director will be a self-generating Expert System. I believe that such systems will, in the future, provide all the advantages of pattern recognition, expert systems and relational data bases in experimental settings. Problems will continue to be defined by human beings, but more and more, the laboratory will design, execute and evaluate its own experiments.

Key words: artificial intelligence; chemical analysis; expert systems; pattern recognition; relational data bases; robotics.

Chemometrics and pattern recognition had their start in chemistry in the late 1960's. Two areas of application appeared almost simultaneously, those being learning machines and project dendral, both applied to spectroscopic interpretation. The former originated in my research group at the University of Washington and have been carried on by us, and my two students, Peter Jurs and Bruce Kowalski, and the latter was developed at the Stanford Artificial Institute by a consortium from Chemistry and Computer Science. Over the past 15 years a variety of applications have occurred which include the following list and probably others: statistics, modeling and parameter estimation, resolution, calibration, signal processing, image analysis, factor analysis, pattern recognition, optimization, artificial intelli-

gence, graph theory and structure handling, and library searching.

The most recent review of the area by Michael DeLaney listed 438 journal articles and books. This was an *Analytical Chemistry* article which covered the last two years of activity. Clearly, pattern recognition and its applications now have an established place in chemistry.

The topic of this presentation is, however, not what has happened up to the point. Rather it is what will happen in the foreseeable future. I believe that the three most important areas of future development will be expert systems, relational data bases, and robotics. We will talk a little about each one of these and then go on to deal in detail with what may soon become one of the most, if not *the* most, sophisticated tool that the experimental chemist has ever acquired.

An expert system is, simply stated, a piece of software that behaves like an expert. The origin of expert systems were instruction manuals that told you what to do based upon what you encountered in a step by step fashion.

About the Author: Thomas L. Isenhour is with Utah State University's Department of Chemistry and Biochemistry.

Almost everyone is familiar with manuals on "How to Fix Your Chevy Station Wagon," etc. These are non-mysterious, and sometimes non-useful, written recipes designed to lead you by the hand through a repair, construction, gourmet meal preparation, etc., that an expert could do but you could not, at least on your own. Of course, the quality of such systems depends on the knowledge of the expert, and on the successful transfer of that knowledge from that expert to the author and from the author to you.

The modern expert system is usually a computer program that attempts the same thing, with the same limits of success. That is, the quality is dependent on the knowledge of the expert, the successful transfer of that knowledge from that expert to the author and from the author to the user. It is a little more, however, because the machine can use your input to do computations, etc., and while all of this could be done with a programmed manual, it certainly is faster by computer and potentially more accurate. There is less chance of you failing to follow the directions for correct use than in a written expert system.

Relational data bases are collections of data that interrelate along traditional and non-traditional lines. In a sense, the relational data base is the social scientist's dream. Given a set of descriptors for each entry, the relational data base allows very easy cross correlations such as, how many children in Miami who had braces before the age of 12 also have maternal grandparents alive in Manhattan. Again, a relational data base was always possible with pencil and paper, or even better with three-by-five cards, but it can be greatly facilitated in a computer. However, the use is still defined by the selection of the descriptors and the quality of the queries.

A robot, according to one handy dictionary, is "a machine devised to function in place of a human agent." This is quite a broad definition and could refer to an automatic ticket dispenser at a parking garage, an autopilot operating from an inertial guidance system in a jet airplane, or a computer interfaced with an autoanalyser at a clinical laboratory.

It is my contention that it is now possible to combine existing robotics and artificial intelligence software to create a system that will generate its own expert systems using relational data bases. The data will be in the chemical domain and the system I describe we are calling the Analytical Director. Simply stated, the Analytical Director will be an artificial intelligence/robotic expert system for the analytical laboratory. The Analytical Director will develop, test, implement, and interpret chemical analysis procedures. It will learn from its own experience, and that of others, and it will communicate what it has learned to others.

The subject of this research is neither automation nor the robot's roll in automation, but rather EXPERT laboratory management working through a combination of robotics and artificial intelligence. We propose to combine robotics and artificial intelligence into an expert system for the analytical

chemistry laboratory. We propose to demonstrate that an Analytical Director can develop, test, implement, and interpret chemical analysis procedures. It is a misconception that the best use of robots will be exhaustive testing of possible solutions to problems. While computationally exhaustive methods are often quite successful, they are rarely useful to analytical chemistry. Artificial intelligence is required for a real breakthrough in automated laboratory methodology.

Consider the following analytical problem which might arise from a relatively simple problem.

Given:

10 possible components to a mixture

10 reagents

10 possible temperatures

10 pH values

If each reaction combination were chemically independent, that is, if the results of any combination could be learned by a linear addition of the separate tests, then 10,000 procedures could be carried out to determine the entire system. This might be feasible if, for example, each test could be completed in one minute. (This would require just about one week of continuous work assuming the robot suffered no maintenance problems or other delays.)

However, chemical reactions are not usually independent. For example, if one of the components were Fe(III) and two of the reagents were CSN^- and citrate ion, there would clearly be complex equilibria interactions. If we redo the calculation considering from 1 to 10 possible components, from 1 to 10 possible reagents and any combination of 10 temperatures and 10 pH values, it requires 1.63×10^{15} tests. Again carried out at one minute intervals, assuming the robot could work through the entire set of procedures without interruption, 3.10×10^9 years would be needed. Experimental design methods might achieve a few orders of magnitude improvement but nothing like the 10 orders of magnitude necessary to make this approach feasible.

Furthermore, in real analytical situations the number of variables and dimensions is often much greater. It is clear that analytical chemistry cannot be done by exhaustive trial and error. Therefore, if robotics is to have any real effect upon the field an intelligent robot must be created that can choose meaningful experiments and profit from its experience, as well as the experience of others.

We propose to construct such a system and test it initially on a very limited analytical problem to prove that artificial intelligence can be used to seek efficient paths to complex analytical problems without resorting to exhaustive trial and error. To do so our first model system will be very simple, involving 3 ions, 5 reagents, 2 temperatures and 3 pH's. This system can be exhaustively tested with 4320 procedures. Assuming 1 minute tests, 3 days would be required. Exhaustive testing of this model system will produce a set of observations that will facilitate the development and test-

ing of generalized data structures and optimization routines to be used by the Analytical Director for more complex problems. For complex problems, the Analytical Director must develop artificial intelligence methods that circumvent the testing approach.

We have selected a developmental domain that is a closed system of simple analytical chemical problems. The domain will be wet and photometric analysis of simple cations. The set of manipulative skills required is purposely limited to the abilities of the Zymark system, and the chemical reactions and spectrophotometric measurements possible limited to those that can be performed with the available equipment. This way, we plan to be able to test thoroughly the creative capabilities of the artificial intelligence programs to be developed for the Analytical Director. We have further selected the domain of water analysis for an advanced test of the Analytical Director. Only by using an unbounded problem will we be able to demonstrate the true capability of the Analytical Director.

Given a set of standards, reagents, and manipulative skills, the Analytical Director will develop its own set of tests for each individual cation. These data will be stored in a relational data base keyed on ions, reagents, conditions and results of spectroscopic measurements. It will be assumed initially that no chemistry is known for these elements or reagents. After developing possible individual tests, these tests will be cross compared to identify likely interfering reactions. Tests will be characterized by their quality as defined by time, expense and reproducibility. As the best compromise of these components is a value judgment, an adjustable value coefficient will be developed. Then possible mixture methods will be systematically tested and compared for success. As this proceeds the relational data base will continue to expand. Finally, new unknowns will be introduced into the system to test the ability of the Analytical Director to adapt to new circumstances. At this point literature information will be added to the data base. The Analytical Director will thereby "learn" from the experience of others. Further, the Analytical Director will report developed procedures for possible use by others.

In summary, the Analytical Director will be a self-generating expert system. I believe that such systems will, in the future, provide all the advantages of pattern recognition, expert systems, and relational data bases in experimental settings. Problems will continue to be defined by human beings, but more and more the laboratory will design, execute, and evaluate its own experiments.