ACTUARIAL RESEARCH CLEARING HOUSE
1998 VOL. 1

TECHNOLOGIES USED IN MODELING

Arnold F. Shapiro, Thomas A. DeFilippo,
Katharine J. Phinney, and Jing Zhang[1]

Penn State University

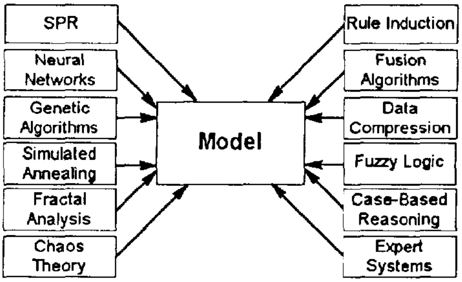
ABSTRACT

Last December, the Society of Actuaries held a conference whose focus was actuarial and financial modeling. During the course of that conference, one of the speakers mentioned a number of technologies his firm was using for modeling, which where either unfamiliar or only vaguely familiar to some of the audience. The purpose of this article is to define each of the technologies mentioned during that presentation and to briefly describe their distinguishing characteristics.

INTRODUCTION

The technologies discussed in this article are summarized in Figure 1.[2]

Figure 1
Technologies Used In Modeling



The rest of the article is devoted to a brief overview of each of these technologies. As a preliminary introduction, a descriptive statement of each follows:

---

[1]The authors are affiliated with the Penn State University. Arnold Shapiro is Professor of Actuarial Science and Insurance and Robert G. Schwartz Faculty Fellow. Thomas DeFilippo, Katharine Phinney, and Jing Zhang are students in the University Scholars Program.

[2]Adapted from Gorman (1996), p. 5.

- Statistical pattern recognition (SPR) identifies patterns in data and uses those patterns to reach conclusions;
- Neural networks represent an analogy of the human brain and the associated neural complex;
- Genetic algorithms are automated heuristics that perform optimization by emulating evolution in nature;
- Simulated annealing is a stochastic algorithm which minimizes numerical functions. Its distinguishing feature is that it involves a random process that eludes local minimums;
- Fractal analysis is the basis of the fractal market hypothesis;
- Chaos models incorporate memory effects;
- Data fusion is used to combine data from different sensors to facilitate target recognition;
- Data compression is used to maximize the amount of information contained in the data;
- Rule induction is a system which induces logical rules from historical data, and then applies the rules to make predictions on other given data;
- Fuzzy logic is a superset of conventional logic that has been extended to handle the concept of partial truths;
- Case-based reasoning is an approach to problem solving based on the retrieval and adaptation of cases; and
- Expert systems are designed to replicate the problem-solving capability associated with a specialized domain human expert.

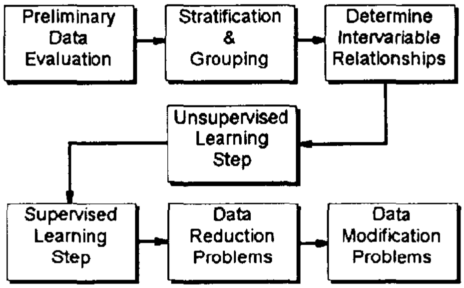
STATISTICAL PATTERN RECOGNITION

Statistical pattern recognition (SPR) is used to identify patterns in data and to use those patterns to reach conclusions. The technique is designed to facilitate data interpretation when confronted with large data sets and many parameters, and, as such, is a preliminary step for many of the technologies that follow.

Generally speaking, the steps of the methodology underlying the SPR approach may be depicted by the flowchart shown in Figure 2.[3] A brief description of these steps follows.

[3] Adapted from Wolff and Parsons (1983).

Figure 2
SPR Methodology



- Preliminary data evaluation. Screenings of variables are made, compatibility checked, and each variable is studied with respect to its distribution characteristics.

- Stratification and grouping. Investigate the stratification of the variables and a natural grouping for each. Check statistical differences in the groups obtained.

- Determine intervariable relationships. Study sets and pairs of variables for intervariable relationships, including correlations and regressions.

- Unsupervised learning. All variables are simultaneously analyzed for natural groupings in the data base.

- Supervised learning. This step contains supervised learning techniques which include checking the validity of the "natural" groupings found in Step 4. Another technique is to train the computer to recognize known category definitions for the data and to study what variables affect this grouping.

- Data reduction problems. Look at the problems inherent in data reduction techniques and how different subgroups of variables in the study define the categories.

- Data modification problems. Address the problems of data modification procedures.

Of course, the foregoing is only one possible approach.

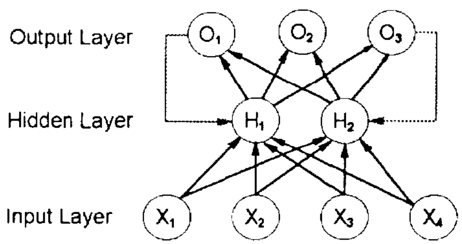
Statistical pattern recognition has been used in numerous applications. A computer-based "nearest-neighbor" pattern recognition technique could be used to emulate underwriters who observe applications and compare them with all the ones they have in their knowledge base. The program can be dynamic, in the sense that the criteria for prediction could be altered continuously as the result of learning. This is similar to case-based reasoning and expert systems.

NEURAL NETWORKS

Neural networks (NNs) are software programs that emulate the biological structure of the human brain and its associated neural complex and are used for pattern classification, prediction and financial analysis, and control and optimization.

An example of a three layer NN is depicted in Figure 3. Here, the first layer, the input layer, has four neurons, the second layer, the hidden processing layer,[4] has two neurons, and the third layer, the output layer, has three neurons. If the flow of information through the network is from the input to the output, it is known as a feed forward network. Moreover, if, as portrayed here, the process involves supervised learning, in the sense that inadequacies in the output are fed back through the network so that the algorithm can be improved, the NN is said to involve back-propagation.

Figure 3
Feedback Neural Network



The Learning Process of a Neural Network

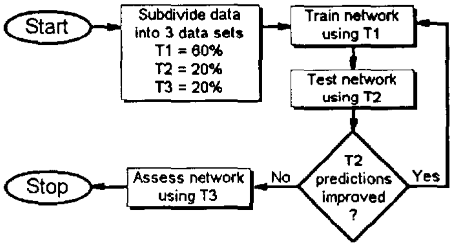
The characteristic feature of NNs is their ability to learn. The process by which this takes place involves training, testing, and validation, and is exemplified in Figure 4.[5] As indicated, the clean and scrubbed data is randomly subdivided into three subsets: T1, 60 percent, is used for training the network; T2, 20 percent, is used for testing the stopping rule; and T3, 20 percent, is used for testing the resulting network. The stopping rule reduces the likelihood that the network will become overtrained, by stopping the training on T1 when the predictive ability of the network, as measured on T2, is no longer improved.

---

[4]In essence, as observed by Brockett *et. al.* (1994), p. 408, the "hidden" layers in the model are conceptually similar to nonorthogonal latent factors in a factor analysis, providing a mutually depend summarization of the pertinent commonalities in the input data.

[5]This figure is based on a discussion of an application by Brockett *et. al.* (1994), p. 415.

Figure 4
The NN Learning Process

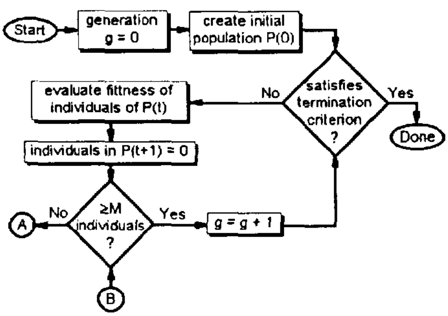


GENETIC ALGORITHMS

Genetic algorithms (GAs) are automated heuristics that perform optimization by emulating biological evolution. They are particularly well suited for solving problems that involve loose constraints, such as discontinuity, noise, high dimensionality, and multimodal objective functions.

GAs can be thought of as an automated, intelligent approach to trial and error, based on principles of natural selection. The flowchart in Figure 6 gives a representation of the process.

Figure 5
Flow Chart of Genetic Algorithm

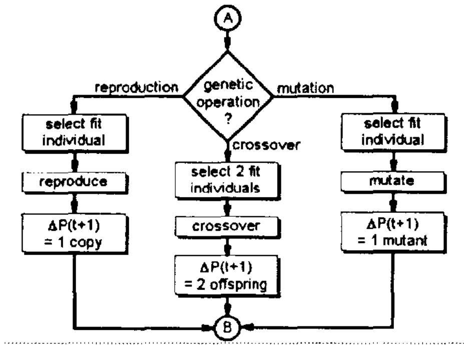


As indicated, GAs are iterative procedures, where each iteration represents a generation. The process starts with an initial population of solutions, which, typically, are randomly generated. From this initial population, the best solutions are "bred"with each other and the worse are discarded. The process ends when the termination criterion is satisfied.

A flowchart of the process for generating new populations of solutions is depicted in Figure 6.

As indicated, there are three ways to develop a new generation of solutions: reproduction, crossover, and mutation. Reproduction adds a copy of a fit individual to the next generation. Crossover emulates the process of creating children, and involves the creation of new individuals (children) from the two fit parents by a recombination of their genes (parameters). Under the process of mutation, a small number of gene values in the population are replaced with randomly generated values. This has the potential effect of introducing good gene values that may not have occurred in the initial population or which were eliminated during the iterations. In this representation, the process is repeated until the new generation has the same number of individuals as the current one.

Figure 6
Reproduction, Crossover and Mutation

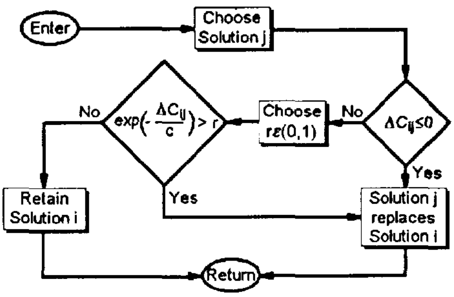


SIMULATED ANNEALING

Annealing is the physical process of heating a solid in a heat bath until it melts and then slowly cooling it down until it crystallizes into a state of perfect structure. The free energy (stress) of the solid is minimized during this process.

Simulated annealing is a heuristic which mimics the physical annealing process inasmuch as it allows solutions which increase the value of the objective to be accepted with a certain probability. Thus, unlike decent methods, in which only sequences which decrease the value of the objective function are accepted for further consideration, simulated annealing is a randomized improvement method which sometimes accepts a new sequence even though its objective value exceeds that of the old sequence. This procedure, which is known as hillclimbing, is represented in Figure 7.

Figure 7
Metropolis (Hillclimbing) Criterion



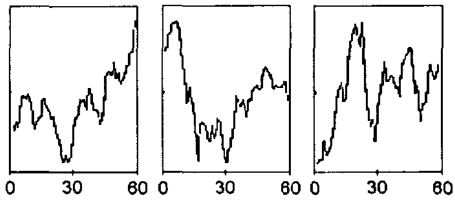
The Traveling Salesman Problem (TSP) [Lawler et. al. (1985)] is probably the best known problem in combinatorial optimization. The problem involves the pair (f,C), where f represent the potential tours of the salesman and C is a cost function which assigns a length of time to each tour. The problem is to find the tour for which C is a minimum, that is, the tour with the shortest length.

FRACTAL ANALYSIS

A simple explanation of a fractal is that it is an object in which the parts are in some way related to the whole. Trees are an example, since they branch according to a fractal scale.

An interesting context from an actuarial perspective involves fractal time series, that is, time series that are statistically self-similar with respect to time. This characteristic is exemplified in Figure 8, which shows daily, weekly, and monthly (not necessarily in that order) S&P 500 returns for 60 consecutive observations. With no scale on the X or Y axis, there is no way to tell which graph is which.

Figure 8
Daily, Weekly, and Monthly S&P 500 Returns
(Which is which?)



Fractal analysis and chaos theory (see next section) form the basis of the fractal market hypothesis [Peters (1991)], which holds that market stability depends on investors with a large range of investment horizons and that instability occurs when investors are predominantly in a short-term mode. This hypothesis provides a framework for short-term projections.

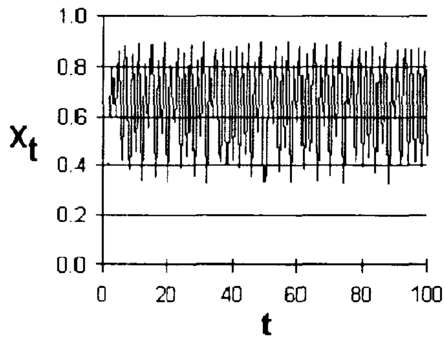
CHAOS THEORY

The essential message of chaos theory is that normal equations can produce random-looking results and, conversely, there may be order in random-looking data. These features of chaos theory are captured in the trend of the logistic equation [Baumol and Quandt (1985)]

$$x_{t+1} = x_t\, w\, (1 - x_t),\ 0<x<1,$$

as w varies between 2 and 4, where t is the number of iterations. When w = 2 in this nonlinear difference equation, x_t, quickly settles down to a stable value. However, as shown in Figure 9, when w = 3.6, the system loses all stability and the number of solutions is infinite. The result is chaos.

Figure 9
The Transition to Chaos

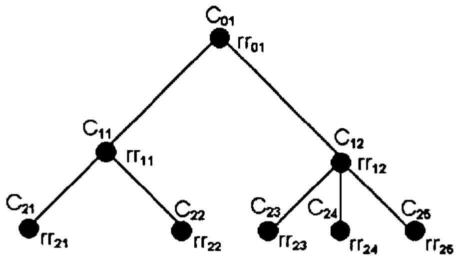


RULE INDUCTION

The purpose of rule induction is to extract implicit rules from a set of sample data. The essential feature of the technique is that it uses supervised clustering, based on observations of the data, to determine boundaries and extent of membership functions. The method is particularly suited to situations involving large data bases.

The procedure is easily explained with the diagram shown in Figure 10. Each of the nodes represents a classification of the data, C_{ij}, and a response rate for that classification, rr_{ij}, where I indicates a classification level and j represents a node within that level. C_{ij} is a subset of C_{kj} if i>k and C_{ij} dominates C_{kj}, if $rr_{ij} > rr_{kj}$. Thus, for example, if, in the figure, C_{01} was a preliminary acceptable classification of data, and $rr_{21} > rr_{01}$, classification C_{21} would also be an acceptable classification.

Figure 10
Classification v. Response Rate

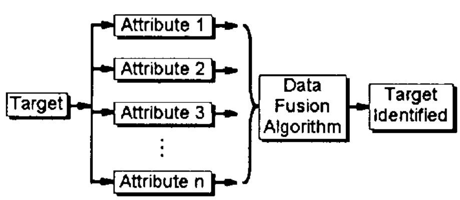


DATA FUSION ALGORITHMS

Data fusion algorithms (DFAs) are used to combine data from different sensors to facilitate target recognition. A simple example is a summation algorithm which merely sums the incoming data.

A general representation of the implementation of the technology is shown in Figure 11.

Figure 11
The Structure of Data Fusion Algorithms



Intuitively, DFAs reflect human cognition in so far as they take multiple data sources into account. For example, an insurance agent attempting to classify a potential client instinctively looks simultaneously at the many attributes of a client, such as income, age, and insurance needs.

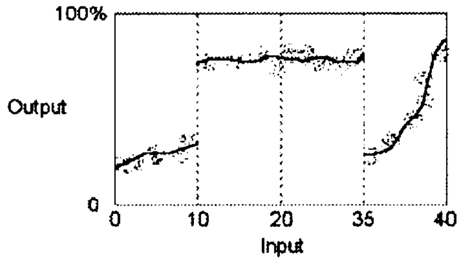
DATA COMPRESSION

Data compression is one of the first steps of the modeling process. The concern here, as it is in information theory in general, is to maximize the amount of information contained in the data.

One way to address this issue is to implement a process that operates on the raw state of input variables and brings them together into a concise set of aggregates. This might be accomplished, for example, by the use of nonlinear compression, a process related to factor analysis or principal components.

Segmentation is another approach to this problem. Here, the goal is to focus on segments in the population that have relatively constant behavior. An example is shown in Figure 12

Figure 12
Segmentation

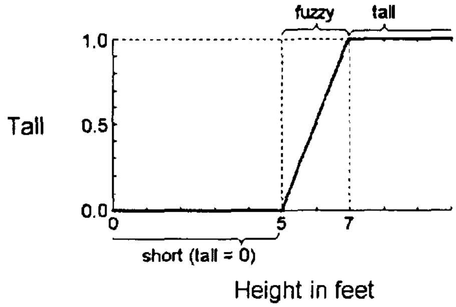


As indicated, aggregating may misrepresent the underlying structure of the data and reduce the resolution in the model. In contrast, if the domains are isolated, and the model is allowed to focus on relatively stationary behavior, the technique is more likely to extract the underlying information.

FUZZY LOGIC

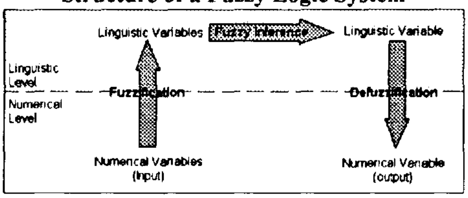
A fuzzy set is a class of objects with a continuum of grades of membership. Such a set, an example of which is shown in Figure 13, is characterized by a membership function which assigns to each object a grade of membership ranging between zero and one. In this case, which represents the set of tall people, individuals five feet high or less are assigned a membership grade of zero and those seven feet high and taller are assigned a grade of one. Between those heights the grade of membership is fuzzy. In this context, "tall" is an example of a linguistic variable.

Figure 13
The (Fuzzy) Set of Tall People



The essential structure of a fuzzy logic system is depicted in the flow chart shown in Figure 14.[6]

Figure 14
Structure of a Fuzzy Logic System



As indicated in the figure, the input of the system are numerical variables. These variables are passed through a fuzzification stage, where they are transformed to linguistic variables and subjected to inference rules. The logistics results are then transformed by a defuzzification stage into numerical values which become the output of the system.

CASE-BASED REASONING

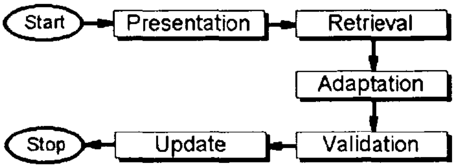
The distinguishing feature of case-based reasoning (CBR) is that it provides a solution to a problem by adapting a solution that was used to solve a similar problem in the past.

---

[6]Adapted from von Altrock (1997), p. 37.

Generally speaking, CBR can be thought of as the five phase process depicted in Figure 15.[7]

Figure 15
Case-based Reasoning Process



A brief explanation of each phase is as follows:

- Presentation. Present the characteristics of the problem as input to the system

- Retrieval. Retrieve the closest-matching cases from the case base (database of cases). This entails the use of an index library of some sort, which is essentially a search and retrieval facility, and may involve nearest-neighbor algorithms or decision trees.

- Adaptation. Generate a solution to the problem by adapting the closest-matching cases. The adaption module can involve derivational adaption, which creates a solution by using a past solution or structural adaption, which creates a solution by modifying a past solution.

- Validation. Validate the solution

- Update. If the validated solution is not represented in the case base, add it to the case base.

As with other processes of this kind, the extent to which these steps are implemented and the techniques used can vary considerably.

EXPERT SYSTEMS

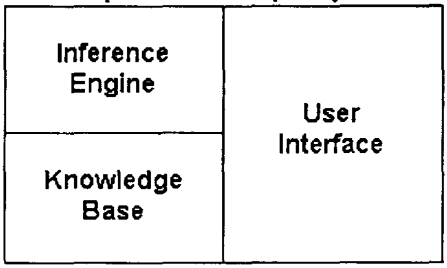
The final technology to be discussed, expert systems, probably is the one most widely known. This is not surprising since it was one of the first applications of artificial intelligence to crossover from the lab to commercial use.

---

[7] Adapted from Allen (1994). Originally adapted by Allen from DARPA (1989).

The main components of an expert system are show in Figure 16.

Figure 16
Components of an Expert System



Briefly, the knowledge base contains the rules and facts about a certain domain, the inference engine uses the encoded knowledge to form inferences and draw conclusions, and the user interface provides facilities for such things as data input, reporting, graphical displays, explanations and on-line help systems.

COMMENT

The purpose of this article was to present a brief overview of the key features of major current modeling technologies. Consequently, a number of issues were not addressed, including detail descriptions of the technologies and how they are implemented, the relationships between the technologies, how the technologies are combined to form models, and examples of applications. These issues will be addressed in future articles.

SELECTED REFERENCES

Allen, B. P. (1994) "Case-based reasoning: Business applications." *Communications of the ACM*. V37, Mar, p. 40-42.

Barber, J. C. (1995) "Genetic Algorithms as Tools for Optimization". *Risks and Rewards*, June.

Baumol, W. J., and R. E. Quandt. (1985) "Chaos Models and their Implications for Forecasting," *Eastern Economic Journal*, Vol. 11, No. 1, pp. 3-15

Bian, L. J. and Engelhardt, M. (1992) *Introduction to Probability and Mathematical Statistics*.

Brockett, P. L., W. W. Cooper, L. L. Golden, and U. Pitaktong. (1994) "A Neural Network Method for Obtaining an Early Warning of Insurer Insolvency," *The Journal of Risk and Insurance*, pp 402-424.

Day, R. H. (1996) "Complex Dynamics, Market Mediation and Stock Price Behavior," Actuarial and Financial Modeling Conference, December 16-17, 1996, Georgia State University.

Deboeck, G. Editor. (1994) *Trading on the Edge: Neural, Genetic, and Fuzzy Systems for chaotic financial markets.* (New York: John Wiley and Sons)

Derrig, R. A. and K. M. Ostoszewski. (1995) "Fuzzy Technique of Pattern Recognition in Risk Claim Classification". *The Journal of Risk and Insurance*. V62, pp 447-482.

Eberhardt, R. C. and R. W. Dobbins. *(1990). Neural Network PC Tools: A Practical Guide* (New York: Academic Press).

Forrest, S. (1996) "Genetic Algorithms". *ACM Computing Surveys*, March.

Francis, R. (1992) "Double Your Drives with Data Compression". *Datamation*, December.

Gleick, J. (1987) *Chaos - Making a New Science.* New York: Viking Press, p. 306.

Gorman, P. R. (1996) "Current Modeling Approaches: A Case Study," Actuarial and Financial Modeling Conference, December 16-17, Georgia State University.

Goldberg, D. E. *Genetic Algorithms in Search, Optimization, and Machine Learning*

Holland, J. H. (1975) Adaptation in Natural and Artificial Systems, Univ. Michigan Press, Ann Arbor

John, G. H., P. Miller, and R. Kerber. (1996) "Stock Selection Using Rule Induction". *IEEE Expert*, October.

Langley, P. and H. A. Simon. (1995) "Applications of Machine Learning and Rule Induction". *Communications of the ACM*, November.

Lawler, E. L., J. K. Lenstra, A. J. G. Rinnooy Kan, and D. B. Shmoys (Eds) (1985) *The Traveling Salesman Problem: A Guided Tour of combinatorial Optimization*, Wiley, Chichester.

Lewinson, L. "GeneHunter-GA Software from Ward". *PC AI Magazine*, March/April.

Macdonald, A. S. (1996) "Current Actuarial Modeling Practice and Related Issues and

Questions, Actuarial and Financial Modeling Conference, December 16-17, Georgia State University.

Massaro, D. W. and D. Friedman. (1990) "Models of integration given multiple sources of information" *Psychological Review*. V97, pp. 225-252

Masters, T. (1993) Practical Neural Network Recipes in C++, *Academic Press*.

Peters, E. E. (1991) Chaos and Order in the Capital Markets, *Wiley Finance Edition*.

Roberts, T. (1994) "Data Compression: Packing It In". *Compute*, January 1994: 52.

Smith, M. (1993) *Neural Networks for Statistical Modeling*, Van Nostrand Reinhold.

Von Altrock, C. (1997) *Fuzzy Logic and NeuroFuzzy Applications in Business and Finance.* Prentice-Hall.

Wolff, D. D. and M. L. Parsons. (1983) *Pattern Recognition Approach to Data Interpretation*, (New York: Plenum Press), pp. 223

Young, V. R. (1996) "Insurance Rate Changing: A Fuzzy Logic Approach," *The Journal of Risk and Insurance*, vol. 63, No. 3, 461.