NEW MEDIA: SOFTWARE

Neural Nets for Novices

John Wass

According to the company marketing insert, NeuroShell Easy was "designed for use by people who have no previous experience with neural networks." In this they succeed; these programs bring neural network software—useful for prediction and for classification of data contained in large data sets—into the hands of many users.

Neuroshell Easy:
Predictor
Classifier
Run-Time
by Ward Systems Group
Frederick, MD.
$395; $395; $295
Phone: (301) 662-7950
www.wardsystems.com

A branch of artificial intelligence (AI), neural networks mimic the human brain's method of problem solving by constructing artificial "neurons" in the process of analyzing data. Data can be entered in the form of numbers, patterns, or even sounds and images. Neurons are created by the program and connected in a straightforward, one-to-one manner as well as in complex arrays. The user is rarely aware of the number and relative importance (weighting) of these units. During analysis of a problem, a neural network can spontaneously grow (add new neurons) and reweight itself.

In normal operation, a user provides a data set (called inputs), and usually a single output, to "train" the network. The neural network "digests" the data, reweights the inputs, and adds layers of neurons as necessary to accurately predict the output. This prediction is then compared with the known output of an actual experiment. In older programs of this type, a user could then readjust weightings and algorithms to make predicted results move closer to observed results. Newer programs such as NeuroShell Easy, however, can run on autopilot, although the manufacturer can supply software to allow more user control.

NeuroShell Easy Predictor generates numeric estimates of output data from up to 150 distinct input variables. Predictor will support 1000 columns by 16,000 rows or 150 input variables and one output. The size of this database is far greater than most laboratory experimenters need, but is barely adequate for some business and social science applications. Output may be saved as standard ASCII types, .csv, .prn, .txt, or .dat files. The proprietary algorithm used (Turboprop2) is not based on standard back-propagation models, but uses a variant of the general regression neural networks algorithm.

There are two methods by which the new user acquires information on running a program. One is through a detailed tutorial that moves step-by-step through the performance of a complete analysis. The other employs an "Instructor," the AI analog of the Wizard function found on many statistical and spreadsheet programs. Unlike many of the Wizards, the Instructor is remarkably straightforward and simple to follow. Students can use one of three sample data files to learn the program and thereafter use their own data.

Two training strategies can be chosen. The "neural" option dynamically grows hidden neurons as needed, generalizes the model well, and trains more rapidly than the genetic algorithm. (A genetic algorithm assesses combinations of previously grown networks and recombines portions semi-randomly according to certain predefined rules, as with DNA recombination.) Once the network has been trained, the user can view the results as actual versus predicted data, actual versus learning level, or, for the genetic option, actual versus the relative importance of inputs. This last plot will be familiar to statisticians and quality engineers as the Pareto diagram, which charts the inputs in order of significance to the output. The actual versus predicted graph is interactive: A click anywhere on the graph will reveal the row number and the predicted and observed values at that point. A final diagnostic table lists the values along with the average error for the model.

NeuroShell Easy Classifier determines the probability that input patterns belong to certain defined categories. Once the Predictor is mastered, the Classifier operation will be simple, as both modules perform with essentially the same set of instructions by the user. The order of operation is the same as with Predictor: Select the data set, the input and output columns, the training strategy, and the graphics display type. In this case the program advises that the neural strategy is faster and produces no overfitting. The genetic algorithm, while slower, produces a probabilistic model that generalizes well.

During analysis of the small sample data set ($n = 50$ rows), the neural analysis ran in 4 seconds and the genetic algorithm in 29 seconds on a 166-MHz Pentium. During the analysis, the user can view a graph that displays the percent correct classification versus number of hidden neurons. The user may also choose to see probabilistic charts showing the percent correct classification rates for each class per input row, or a receiver operating characteristic curve to assist in separating one class from another. The genetic algorithm mode produces a useful learning graph of correct classifications per generation, but the business or scientific researcher may find the graph of importance of inputs much more illuminating.

NeuroShell Easy Run-Time is the simplest of the three programs to run. For the novice, use will be limited to calling and entering one or two data points. It offers the analyst a method of running a trained network on data already analyzed within NeuroShell. For example, if an Excel spreadsheet is open, Run-Time can be brought on-screen with one click and used to generate a prediction for a single row of inputs. The other half of this module is more useful to the programmer, as it provides an interface with languages such as Visual Basic, C, C++, Fortran, and Pascal.

In their effort to produce a user-friendly program with usable algorithms, the company has succeeded admirably. In several other areas it has failed to keep pace with highly desirable features, such as cutting and pasting with the clipboard, a feature present in most spreadsheets and statistical programs.

A constant source of complaints about neural networks is that they do not show equations or give any hint of the mechanisms by which they derive their predictions. NeuroShell Easy displays tables of predicted versus actual output, but only two or three graphics, to assist in interpreting the quality of the results. To a business determining buying patterns or investments, this may be enough, but to the researcher attempting to build a model or better understand a process, this omission becomes serious. Fine-tuning choices provide powerful tools that are very useful in selecting a final model. But these have, unfortunately, been left by the wayside for the sake of novice use. It is also a pity that NeuroShell has followed the all-too-common practice of taking one large, multipurpose program and breaking it up into smaller modules that, when purchased together, cost more than the original, more versatile program did.

These are 32-bit programs for Windows 95 and NT. The manufacturer says that these programs will run on a 486 with 16 MB of RAM, but a Pentium with at least 32 MB of RAM is highly recommended. Despite drawbacks for the more experienced analyst, NeuroShell Easy is recommended to new neural network users as a very simple-to-use program with powerful features for prediction and classification.

The author is at Abbott Laboratories, Abbott Park, IL 60045, USA. E-mail: John.Wass@add.ssw.abbott.com