Virtual DOE, Data Mining and Artificial Neural Networks

Six Sigma generates as much improvement by changing things as it does by reducing variability.

A key feature of Six Sigma is the integrated use of design of experiments and robust design of product and process. Experiments are an important part of this; they serve both to find the most robust design settings and to verify that the designs are, in fact, robust.

As beneficial and productive as experiments can be, the process of conducting them has its drawbacks. The workplace—be it a factory, a retail establishment or an office—revolves around a routine. The routine is the "real work" that must be done to generate the sales that, in turn, produce the revenues which keep the enterprise in existence. Experimenting, by its very nature, means disrupting the routine. Important things are changed to determine what effect they have on various important metrics. Often, these effects are unpleasant; that's why they weren't changed in the first place. The routine was established to steer a comfortable course that avoids the disruption and waste that results from making changes.

Unless we change, however, we can never improve. Six Sigma generates as much improvement by changing things as it does by reducing variability. It's part of the Six Sigma paradox, which states that to reduce variability in products and processes, we must encourage variability and experimentation.

It's possible to conduct "virtual" experiments using existing data and artificial neural network (neural net) software. Neural nets are a class of very powerful, general-purpose tools readily applied to prediction, classification and clustering. They have a proven track record in many data mining and decision-support applications. Neural nets have been applied across a broad range of industries, from predicting financial series to diagnosing medical conditions, from identifying fraudulent credit card transactions to predicting failure rates of engines.

Neural networks use a computer to model the neural connections in human brains. When used in well-defined domains, their ability to generalize and learn from data mimics our ability to learn from experience. However, there is a drawback. Unlike a well-planned and executed DOE, a neural network doesn't provide an explicit mathematical model of the process. For the most part, neural networks must be approached as black boxes with mysterious internal workings, much like the human brain itself.

All companies record important information. Such data represents potential value to the Six Sigma team. It contains information that can be used to evaluate process performance. If the data include information on process settings, for example, they may be matched up to identify possible cause-and-effect relationships and point the direction for improvement. The activity of sifting through a database for useful information is known as data mining. The process works as follows:

1. Create a detailed inventory of data available throughout the organization. The information systems department may already have compiled this information.
2. Determine the variables that apply to the process being improved and for which historical data exist.
3. Using a subset of the data, train the neural net to recognize relationships between patterns in the independent variables and patterns in the dependent variables.
4. Validate and test the neural net's predictive capacity with the remaining data.
5. Perform experimental designs as you would on a real process. However, instead of making changes to the actual process, make changes to the "virtual process" as represented by the neural net.
6. Once the sequential application of designed experiments has been completed using the neural net model, use the settings from the neural net as a starting point for conducting experiments on the actual process. If the experiment results confirm the results from the neural net, you can reduce sample sizes and move quickly along the path discovered by the virtual DOE process.

The entire virtual experimentation process helps answer the question, "Where are we?" It's important to recognize that neural net experiments aren't the same as live experiments; they are only simulations. However, the cost of doing them is minimal compared with experiments in the real world and the process of identifying input and output variables. Moreover, the process of deciding at which levels to test these variables will bear fruit when the team moves on to the real thing. Virtual experiments allow a great deal more "what if" analysis, which may stimulate creative thinking from team members.

This article is based on an excerpt from The Complete Guide to Six Sigma, *Quality Publishing LLC, scheduled for publication summer 1999.* *The article first appeared at* www.qualityamerica.com. *Reprinted by permission.*

About the author

Thomas Pyzdek is president and CEO of Pyzdek Management Inc. and a consultant in performance improvement.

Comments about this column can be e-mailed to Pyzdek at tpyzdek@qualitydigest.com, or visit his Web site at www.pyzdek.com qd