

The Fallacy of Automated Modeling

*By Jim Wheaton
Principal, Wheaton Group*

Original version of an article that appeared in "The DMA's 1996 Research Council Journal"

Regression as well as other traditional segmentation approaches have been under attack for quite some time. The battlegrounds are trade journals and conferences, and the opposition is the neural network companies.

This article does not explore the merits of neural networks versus traditional techniques. Nor does it question all of the proponents of the neural approach. Rather, the target is those who claim that neural networks allow the modeling process to be automated.

I frequently am frustrated when talking with modeling prospects. Many have read the neural network hype, and are predisposed to believe in the "push button" approach to segmentation. After all, automation will eliminate the need to interact with statisticians who ask difficult questions and speak in strange tongues.

It is my strong opinion that there never will be a substitute for the seasoned human analyst. This article will explain why, with ten examples – seven involving Exploratory Data Analysis and three involving Research Design – that are understandable to clients and prospects. (After all, the layman will never be convinced by mathematical equations!)

The following summarizes my message to all who ask about predictive modeling:

Whether the technique of choice is regression or neural nets (or, for that matter, tree analysis or genetic algorithms), what really separates the good models from the bad is the up-front work that must be done before the formal modeling process. (Also important is the back-end process of correctly implementing the model, time and time again, in a production environment.)

Before proceeding with the main body of this article, let me rephrase the previous paragraph in a way that I am sure will be controversial with certain factions of the neural network community:

If you want to get famous, promote the "push button" approach. If you want to build a great model, concentrate on the unglamorous up-front work – as well as the back-end implementation.

Seven Examples – Exploratory Data Analysis

Example #1

For a retailer with a \$70 average order size, a handful of customers on the analysis file had lifetime total dollars between ten and twenty thousand dollars.

In this instance, a well-designed "push button" approach will receive a passing grade. After all, automating the detection of outliers is a simple task. Each of the unusual customers in our example can be eliminated from the analysis, and the modeling process can proceed.

Even in this case, however, human intervention is ideal. Although the deletion of outliers is a good start towards a robust predictive model, the optimal approach is to determine the root cause of these outliers. And, root causes always fall into one of two categories, each of which calls for a different response:

Category #1

Especially within the world of retail, extreme outliers generally represent bad data rather than true purchase behavior. Human investigation in such instances often results in refinements to the data capture process. This will enhance the long-term quality of the database and, in turn, all subsequent models. Two examples:

- A sixty-six thousand dollar order was comprised of 111 \$600 items. This was a keying error because the true number of items was one rather than 111.

The analysis file record showed one additional order of this same item, one day later, reflecting the previous day's true purchase activity. Unfortunately – and in accordance with existing procedure – the corresponding "negative transaction" to cancel the previous 111 items had not been forwarded to the database.

This inspired a rethinking of existing procedure!

- A customer record showed hundreds of transactions within a six-month period, averaging only about \$70 but totaling about thirty thousand dollars. This was an intentional keying error (if such a thing is not oxymoronic). A lazy clerk realized that the Point of Sale data capture system was driven by reverse phone number lookup, and that the continuous entering of his own phone number would eliminate the need to question customers.

This resulted in the creation and dissemination of a formal disciplinary procedure for all employees who continued to engage in such behavior.

Category #2

Sometimes, however, extreme outliers represent true purchase behavior. In such instances, extremely loyal buyers have been identified whose behavior should be rewarded and encouraged.

Example #2

This example, as well as all subsequent ones, provides quite a bit more of a challenge for an automated modeling system:

A weak positive relationship was found between response and customer distance from the nearest store. In other words, the greater the distance, the higher the response.

Good retail customers generally don't live far from a store. The distance variable represented the straight-line distance between each customer's ZIP Centroid and the nearest store's ZIP Centroid. It was theorized that, for many customers, this was not a sufficiently refined calculation. Distance was recalculated by Carrier-Route Centroid, and the relationship to response went negative.

Example #3

A strong positive relationship was found between response and customer ownership of the private-label credit card.

The variable was left out of the model! At the time of the analysis file mailings, the credit card had just been introduced. Therefore, the small number of card owners were generally the client's most fervent buyers. However, by the time the model was to be put into production, the card ownership universe had expanded significantly. Therefore, the relationship of card ownership to response would have changed dramatically.

Example #4

A large number of historical orders on a "time 0" file for a September mailing were nine months old.

"Lumpy" order patterns generally are no big deal. After all, many businesses are seasonal. Christmas is a make-or-break period for just about everyone. So, our nine-month old orders were no big deal, right?

Wrong! The database was maintained by a direct marketer whose business was driven by syndication arrangements with several outside companies. The idea was to merge the customer history for each of these companies, and to use the resulting database to drive up-sell as well as cross-sell efforts with predictive models.

Unfortunately, one of the outside companies was not sensitive to the needs of sophisticated database marketing, and warehoused all of its order transactions for nine months before forwarding them to the syndicator. You can imagine the segmentation chaos caused by the many customers whose records reflected these compromised orders!

Example #5

The relationship of historical average order size to the dependant variable was slightly unusual.

The database had been built several years earlier, and had been "primed" by a large number of "warehoused" historical orders. Unfortunately, because of the lack of systematic data capture during this formative period, many of these historical orders did not contain a dollar amount. The manager of the database build, who had since left the company, had "plugged" these missing dollar amounts with the mean – along with a normally-distributed "plus or minus factor." In other words, he had created artificial data – a time bomb – that would be a challenge for any future analyst to detect.

The point here is that the unusual pattern in the relationship of historical average order size to response was quite subtle. After all, several subsequent years of systematic data capture had populated the analysis file with many legitimate order sizes. Would "push button" software ask the questions necessary to uncover the subset of bogus orders on the analysis file? I think not.

As an aside, neural network hype often supplements the "push button" pitch with the claim that such systems can detect subtle data patterns that are invisible to regression. The challenge, however, is how to determine whether a subtle pattern is an anomaly-driven mirage or the reflection of true, long-term buyer behavior. Here's where the science of predictive modeling gives way to the art of predictive modeling. And, within the realm of art, the judgment and experience of the human analyst is paramount.

Example #6

Here is an additional example that required a human being to ask some questions:

A customer model was built off an analysis file consisting of four mailings. For each mailing, the analysis file was interrogated for basic reasonableness (mail quantity, response rate, dollars per piece mailed, etc.). It was immediately apparent that something was wrong. Additional investigation revealed that the response information had been appended to the incorrect mailings.

The point here is that even automated modeling systems require accurate analysis files. And, it's very easy to generate an inaccurate file. The process of appending response information to the mail history ("time 0") file(s) often involves a number of complex steps that invite error. The only way to uncover analysis file problems is with the inquiring mind of an analyst.

Example #7

The following example required such an inquiring analyst. Although the client did not cooperate with the correct answer, at least the correct questions were asked. Ask whether an automated modeling system would have done the same:

For a customer model, interrogation of the analysis file revealed an unusually large percentage of individuals with only one order at the time of each mailing (single-buyers). The client's answer appeared reasonable: the business had enjoyed rapid and recent growth.

The real reason was discovered only later: about \$80 million dollars of transactions, representing about two and a half years of history, was unavailable when the database was constructed. The "live" model results suffered when who appeared to be the single-buyer inhabitants of the bottom deciles – who were in fact multi-buyers – ordered merchandise with a vengeance.

Three Examples – Research Design

The aforementioned seven examples illustrate the need for high-quality Exploratory Data Analysis, which can only be provided by a human being. The astute reader will notice, however, that none of these seven examples requires an analyst with an advanced degree in statistics. Instead, what is needed is a skilled "data detective." Theoretically, any astute database marketer qualifies as a data detective.

This brings us to the common but more modest contention that neural networks – although not the ticket to automated modeling – eliminate the need for a skilled analyst. You've heard the pitch – "Marketers, build your own models!"

The problem with this argument is that "practice makes perfect," in Exploratory Data Analysis as well as just about everything else in life. Many marketers are smart, but few can match the experience gained by a seasoned analyst who has built dozens – sometimes hundreds – of models.

Modeling experience encompasses not only Exploratory Data Analysis but also Research Design. Consider the following examples that could "trip up" the smartest marketer who "charges in" with neural network technology but no design experience:

Example #1

A predictive model was built for a cataloger in which everyone was eligible to be scored: multi-buyers, single-buyers, inactives, inquiries, and cross-sell candidates from other catalog titles within the overall corporate umbrella. Unfortunately, statistical models – whether traditional or neural networks – take the path of least resistance when segmenting by the probability of future response. Therefore, the result was a "sediment model," in which multi-buyers were the primary residents of

the top couple of deciles, single-buyers the residents of the next two, followed – sequentially – by inquiries, inactives, and cross-sell candidates.

Because the direct marketer already knew that multi-buyers generally perform better than single-buyers, who in turn generally perform better than inactives – and so on – the model essentially was worthless.

Example #2

A prospecting model was built for a fundraiser using individual/household-level overlay demographics. External validations of the model showed impressive segmentation power. Unfortunately, the fundraiser did not have access to net/net list rental arrangements, which meant that the names eliminated by the model would have to be paid for (as would not be the case with a ZIP-level model).

Consider, for example, a hypothetical list with a published cost of \$100/M, for which the model eliminated the bottom eight deciles. The actual, in-the-mail cost would be \$500/M, which clearly is not cost effective.

Fortunately, financial analysis was performed on the model before it was used in live mailings, and proved that under no realistic circumstances would the model ever be cost effective without net/net rental arrangements. In fact, the analysis suggested that there exists no realistic circumstance in which any individual/household-level prospecting model will ever work for any direct marketer without the existence of net/nets.

Example #3

A ZIP Code model was built to segment outside list rental prospects for a very targeted cataloger. Unfortunately, ZIP Code prospecting models generally do not display "lift," top 10% to average, of more than 140 (i.e., with an overall response rate of 1%, Decile 1 will not pull more than 1.4%). Because of the very circumscribed audience for this cataloger's product, only a handful of affordable rental lists were available, and all of those had response rates that were several times higher than average. As a result:

- For the handful of affordable rental lists, even Decile 10 (the worst) names performed above the mail/no mail cutoff.
- For all other rental lists, even Decile 1 (the best) names performed below the mail/no mail cutoff.

Therefore, the ZIP model, although statistically successful at differentiating responders from non-responders, was worthless from a business point of view.

Summary

There is no magical shortcut when building a predictive model. If you want good results, concentrate on the up-front work – sound Research Design and meticulous Exploratory Data Analysis – that must be done before undertaking the formal modeling process. Also, invest in the services of a seasoned analyst rather than a "push button" system.

I'll close with a final thought: Neural network proponents claim that their software is superior in recognizing underlying patterns within the data. If this is true, then the unglamorous, analyst-driven, up-front process will be even more critical. This is because neural networks will, by definition, do a superior job of identifying the spurious patterns that are inherent in bad data. Therefore, the resulting scoring algorithm will point us even farther away from our true target market!

In other words, there is no substitute in our business for a hard-boiled data detective!

Jim Wheaton is a Principal at Wheaton Group, and can be reached at 919-969-8859 or jim.wheaton@wheatongroup.com. The firm specializes in direct marketing consulting and data mining, data quality assessment and assurance, and the delivery of cost-effective data warehouses and marts. Jim is also a Co-Founder of Data University www.datauniversity.org.