September 2019

‘Neo-Classical’ Modeling to Solve Modern Business Problems

The Case for Data Science Artistry over Complexity

By Mike Kelly, Ph.D.

NAXION thinking

The Unfulfilled Promise of Complex Models

Business analysts can leverage a large and growing portfolio of advanced modeling techniques to predict key outcomes like customer spend or defection. These techniques include decision trees, random forests, structural equation modeling, neural networks, polynomial regression, regression splines, and support vector machines, among others. To novices, the options can be intimidating. Even data scientists are working hard to develop their points of view.

Advanced – *typically non-linear* – techniques have proliferated because they are assumed to deliver higher levels of predictive accuracy than what simpler, linear models can achieve. And advanced techniques are just the sort of hot, sexy work that gets young academics noticed, cited, and tenured.

But do complex models actually deliver better predictions?

It turns out that while they may outperform linear ones in the lab with artificial data sets, the incremental value is less clear-cut in the real world. Data scientists who've been tracking the evolution of complex models have noted that gains attributable to more advanced modeling developments can be surprisingly small – sometimes inconsequential.

Observations like these are prompting renewed appreciation for simple linear models in predictive modeling – and growing recognition that it's not necessarily a matter of "either-or". It turns out that we're able to strengthen models when we harness more advanced techniques to old-fashioned "linear" horsepower. In fact, a primary reason for the recent performance gains in neural networks has been the replacement of a non-linear function with a linear one (specifically "rectified linear units" or "ReLUs") in network architecture. This "back to the future" approach has improved overall accuracy and learning speed in diverse areas including language translation and image classification (e.g., facial recognition).

In a market research context, use of simple linear models can provide several advantages over complex models

Easier to understand and communicate	Linear model clarity makes it easier to see the implications for business decisions, facilitating translation of insights to action.
More democratic	Linear models are relatively easy to learn and apply, thus increasing the bandwidth of market analysts who might otherwise require specialized support from others in or outside the organization.
Highly effective with small data sets	More complex models such as neural networks typically require very large data sets (“Big Data”) to optimize performance.
Less likely to lead to wrong conclusions	Risk of model overfitting is significantly higher with non-linear models, so that even if they are slightly more “accurate”, they are also more likely to produce the wrong business decisions.
Less expensive and faster to apply	Linear models get most of the way toward accurate prediction, which means that investment in more complex models produces diminishing returns.
Flexible and adaptable	They can be modified to include non-linear components, to address multi-collinearity among predictors (e.g., through Kruskal regression) and to create complex, multi-stage models out of linear components (e.g., structural equation models).

But ‘Good’ is not Good *Enough*

Together, these advantages suggest that linear models will not be going away any time soon and that, in fact, the pendulum is poised to shift back a bit. We can expect added momentum in a DIY environment that features automated analytics platforms designed to expedite model-building and share the responsibility more broadly throughout organizations.

But that glass is only half-full. Linear models are not getting us the whole way there. Far from it. There is much we are unable to predict or explain about our customers, and an increasingly urgent need to better understand them.

So if complex models aren’t actually much more accurate than linear, does that mean we’re stuck with underperforming models overall? The answer is no. We can get higher horsepower from our engines when we hybridize and upgrade the fuel.

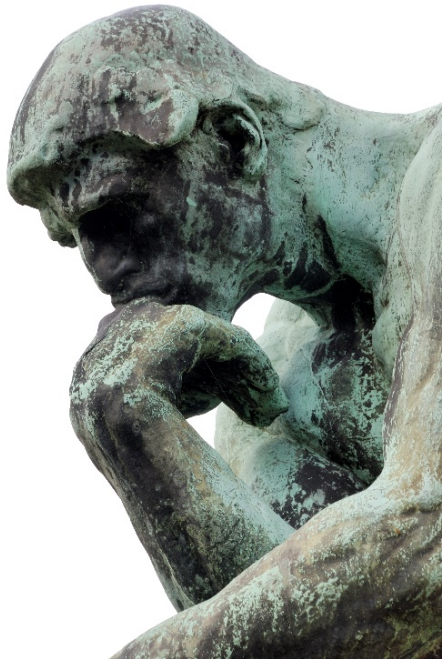
‘Feature Engineering’ to Improve Linear Models: Skilled Preparation, Great Ingredients

It’s important to keep in mind that there are two components to modeling: the analytic technique and predictive features. For reasons noted, academics have tended to focus their attention on optimizing the “machinery” and not the “fuel” but some of the best thinkers in the field are starting to talk about shifting the emphasis to inputs by building better predictors through “feature engineering”. In academic evaluations of modeling success using real world data sets, it turns out that the nature of the predictive features used as inputs can make a big difference.

At NAXION, our applied experience across hundreds of commercial modeling assignments bears out the academic conclusions. While we can typically squeeze out a bit more predictive accuracy by supplementing linear models with non-linear components, the return is substantially higher if we concentrate more on feature engineering.

For example, B2B models are commonly based on Dun & Bradstreet or other business databases. These databases contain many variables that can serve as predictors in modeling, such as revenues, number of employees, and industry. But we've found that model accuracy is significantly enhanced if we go beyond these raw variables to engineer new ones. *In fact, most of the predictors in our final B2B models consist of engineered rather than raw database variables.* Some heuristics we have found helpful in engineering predictive features include combining raw variables to create better indicators of key metrics, assigning businesses to industry categories customized for a particular market, and creating variables that score businesses relative to appropriate peers rather than on an absolute scale.

In consumer applications, use of created variables that integrate certain types of purchase behaviors or brand affinities can produce more illuminating driver models than some of the "natural-occurring" inputs from our survey data sets.



The Art in Artificial Intelligence

This "great ingredients" approach to modeling reinforces the importance of general business, marketing, and industry subject-matter expertise on modeling success because it is more artful than automated. In fact, while feature ***selection*** has long been automated in modeling, feature ***invention*** is not about to be automated any time soon. The best predictive features will, instead, be hand-carved using skills and intuition sharpened through experience. Thus, while we can continue to look forward to automation in many aspects of data science, further gains in modeling acuity will still rely heavily on industry knowledge and business acumen. What comes out the hopper will not take us as far as we need to get to guide critical business decisions.

About the Author

Mike Kelly, Ph.D. Mike is a Senior Group Director at NAXION who designs and manages major engagements for clients seeking to develop B2B and B2C business strategies and drive customer loyalty based on customer insight and advanced market analytics. He holds a doctorate in cognitive psychology from Cornell, and his highly regarded academic work in psycholinguistics as a UPenn faculty member continues to be a platform for decision science innovation. Mike writes frequently on use novel analytic techniques to wring greater nuance and predictive power from data of all kinds and sizes.

215.496.6842 mkelly@naxionthinking.com

About NAXION

NAXION is a broadly resourced, nimble boutique that relies on advanced research methods, data integration, and sector-focused experience to guide strategic business decisions that shape the destiny of brands. Our century-long history of innovation has helped to propel the insights discipline and continues to inspire NAXION's investments in the development and curation of emerging data science techniques. For information on what we're currently doing and how we might help you with your marketing challenges, please visit our website.