Interpreting Neural Network Loyalty Models.

Thomas R Osborn[1], Greg N Taylor[1] and Ian Roberts[2].

[1]Nelson Taylor Fox
Level 7, 1 York St
Sydney 2000

[2]Australia Post
Level 7, 90 Collins St
Melbourne 3000

Correspondence to: osborn@ntf.com.au

Abstract:

This paper has four threads which tie together the business of delivering the findings of loyalty studies to commercial clients. The threads emerged from a loyalty survey for which traditional analysis yielded no significant findings. The model problem arose from a lack of agreement between common assumptions made in traditional analysis (eg, linear, quasi-linear), and the semantics of the behaviour/belief structure underlying loyalty. The findings are applicable to other psychometric models derived from surveys, including choice, preference and rank preference, and other forms of declared intent models.

The threads:

1. *The need for sophisticated non-linear models to 'fit' complex customer and market behaviours.*
2. *The drawback of these advanced approaches is a loss of the ability to explain customer and market behaviours with simple 'main effect' co-efficients. Business must follow science in recognising the dangers of trying to summarise complex phenomena through simplistic and highly restrictive quantitative methods.*
3. *The authors argue a case for robust, sophisticated methods in conjunction with model simulation. Where commercial client have expectations presuming "driver strength"/"elasticity" coefficients, a Graphical User Interface Decision Support System (GUI/DSS) provides a generalised equivalent, with more precision and conditional dependencies.*
4. *Furthermore, scenario simulation on the GUI/DSS for purposive analysis of information in the survey.*

While this paper focusses on the example of a loyalty model in a service industry, the conclusions are just as applicable across the full gamut of marketing research and business analysis.

1. Introduction – Market Intelligence via a Loyalty Survey.

Businesses rely on several sources of information for marketing plans to cover such diverse decisions as introduction of new products, production scheduling, re-structuring pricing and re-engineering service delivery, meeting competitor challenges, determining optimal advertising expenditure and media flighting, and customer retention.

Qualitative studies (interviews and focus groups), staff experience and value judgements, sales records, observational research, advertising schedules and market intelligence inform the organisation of the kinds of attributes customers are sensitive ("drivers" of behaviour) and the current level of demand.

Surveys structured on preliminary qualitative studies gather information on customer choices, preferences, etc as a function of the "drivers". These provide a second stage of market intelligence, by *quantifying* reported intentions of respondents as representatives of the customer base.

Actual customer behaviour provides further information sources in the form of sales records, panel data and ultimately company bottom line. These data source provide valuable sources for validation of accepted company wisdom, and the correspondence between suggested behaviours from surveys and actual behaviours. Unfortunately, sales and panel data are "after the event", so their *analysis* can report outcomes of experiments on a live market.

With sufficiently precise and robust modelling, survey modelling and sales/price/advertising data modelling provide a simulation environment to flesh out the intrinsic value of a market position or value proposition, without laying taking a risk on a live market. The main imitation for this kinds of simulated marketing research is the precision and generality of the models themselves. We show in this paper that a weak model (developed on a less than ideal survey using traditional statistics for a low involvement category) was repairable using neural network technology with care.

Loyalty modelling focusses on the business goal of maintaining or growing a customer base, and diminishing defection of customers. In a competitive market, maintaining loyalty levels and managing defection is a worthy goal, as in general, the acquisition of new customers is much more expensive than retaining the old ones. Loyalty experiments on a live market are very ill-advised indeed. Good surveys are the preferred option.

This truism applies for businesses ranging from restaurants and gymnasiums, through to communications services and banks. Complications emerge from differential value of customers. These depend on their "segment" and size, etc within the customer base, as well as individual peculiarities of organisations which includes biases in perceptions, innate allegiances and histories (ie, habits). And then there are problems with surveys themselves…

Typical survey production *problems* include poor design, respondent fatigue, poor phraseology, differential response rates by customer segment, poor sampling population selection, and sometimes contamination of customer beliefs (by lead-in information, current events, etc). Design, sampling and phraseology issues are addressed by rigorous qualitative preparation, due care and experience within the particular market. Fatiguing surveys and low response rates are more problematic, since their significance may only be inferred indirectly. It is possible to produce an arbitrary number of responses by printing enough questionnaires, but the uniform sampling (in terms of response rate by population strata) is still an issue. Likewise, fatigue (from a dull survey) diminishes the quality of the respondent's answers.

This paper's findings start with the arrival survey response data from Loyalty Survey for a service industry client in a highly competitive market (to small business customers of highly variable size). Preliminary analysis indicated a low response rate, an *apparent* differential scale of response to driver variables, by market segment, including some extreme responses which may have been personal biases or have been fatigue related.

Traditional modelling (using generalised linear regression models - GLMs), across the customer base showed a very poor fit $r^2 \cong 0.02$ on a hold-out set (0.15 fitted). We outline below (Section 2) the weaknesses of linearity assumptions inherent in all GLMs relative to the likely topology of the response distribution for this survey. We also indicate the apparent strength of a GLM approach in providing Marketing Departments with interpretable "loyalty driver co-efficients". Unfortunately, using GLMs with this survey data gave "loyalty driver co-efficients" of no value whatsoever. Indeed, a marginally useable r^2 for fitting a GLM to the survey data was a false indicator of "co-efficient" magnitude, since the fitting process was specious (or accidental) as indicated by hold-out fit.

In Section 3, introduces a specialised neural network architecture which covers the GLM and provides extra flexibility. Many Marketing Researchers have tried neural models once or twice, and come across a wide variety of shortcomings. Primarily these arose from a lack of experience, or a lack of mathematical appreciation of the strengths and weaknesses of neural learning algorithms. These problems are akin to the challenge of using non-parametric (ie, non-bounded version space) statistical models, where the modelling amounts to selecting from a very large space of potential models, without a pre-specified functional form. These issue – related to over-fitting - may be addressed in several ways, each of which retains the flexibility of the potential solution, but penalises arbitrarily complex functional forms. The most familiar methods to achieve this are "early stopping" (via cross-validation), and constraints on the functional form. More exotic methods derive from information theoretic measurement applied to prediction.

Our specific method was to augment a GLM with a small number of “hidden neurons” which correspond to automatically derived variables, and then pruning of weak co-efficients to keep the underlying number of degrees of freedom in the model small, but with a posterior improvement on predictability of the model.

Section 4 covers the issues of communicating the model to the client – both useability and interpretability. Familiar concepts such as “loyalty driver co-efficient” depend on where on the *curve* the elasticity is samples. It takes only a little reflection to believe that the “co-efficient” notion may be locally constant (ie, within a local region in the space of survey independent variables), but different elsewhere. The question “what is the ‘price elasticity’ co-efficient”? is answered with “for what kind of customer”? Ie, the elasticity accelerates depending on non-linear response to key variables and cross-interactions between variables. An analogy to this in sales is the effects of “psychological pricing points” (such as “$1.99”) on retail sales. So the question has many answers depending on “which kind of customer” you ask. Luckily, the values fall within a reasonable range (about two to one). An example of interpretation of the models is to answer “what is the effective *loyalty co-efficient* for *‘speed of delivery’ for small businesses of size 50 to 200 employees*, and with *regular pricing* levels”?

Section 5 takes the model interpretability one step further, and provides client-driven simulations. Judicious segmentation and simple reporting on the corresponding sub-set of survey responses may address the example immediately above. If the service organisation wants to understand the consequences of increasing price for faster delivery, the simple reporting approach fails (as well as suffering from the small size of the sub-set). The NTF DSS/GUI offers the client a desktop access to answering “what if” questions. Specifically the GUI is configured to evaluate loyalty changes implicit from changes in changing some of the drivers. Indeed, given a language for articulating strategic positions, it is in principle possible to optimise loyalty (or its utility) based on our neural model, subject to some technical constraints on the application of the model for unsampled regions of the “driver” space.

2. Linearity Assumptions and Limitations of GLMs.

Regression has a long history back to Gauss in 1795 and Legendre in 1806 [Lindsey, 1997]. Since all computation was by hand, earlier regressions had closed form solutions, and were linear and polynomial regressions, with their generally well-understood limitions – very poor extrapolation, inability to fit non-linear models, underlying Gaussian assumption on the dependent variable, extreme sensitivity to data corruption, etc.

The exponential family of regressions (aka GLMs) was introduced by Fisher [Fisher, 1934]. The underlying modelling construct was a statistical variate of independent variables $\mathbf{X} = [x_1, x_2, \ldots x_k]$ to a dependent variable $\mathbf{Y}$ (possibly a vector), through a ‘wrapper’ function from the exponential family:

$$Variate(\mathbf{X}) = \sum_{i=1}^{k} a_i x_i + a_0 \text{ , where } \mathbf{Y} = Wrapper(Variate(\mathbf{X})) \text{ .}$$

Wrapper is a distributional form for the dependent variable and covers the exponential family, including Logistic, Gaussian (probit), Poisson, etc, as well as multinomials (for a unit vector corresponding to proportions). In general, these methods have the same extensive properties of linear regression, but cover more useful dependent variables such as proportions, head counts, responses of a pre-defined range, etc. They still suffer from extrapolation limitations, but less so than linear regression due to the bounded nature of the range [McCullagh and Nelder, 1989].

Modelling corresponds to estimating the parameters in the model: $[a_0, a_1, \ldots a_k]$. Estimation is either by “ordinary least squares” (minimisation of fitting error, presuming Gaussian error) or maximum likelihood estimation [MLE] (ie, using distributional assumptions about the underlying data, to maximise fitting agreement – in an optimal information gain sense).

The assumptions which concern us in this paper include the [quasi-]linear response of the “loyalty driver” variables, and the faithfulness of the underlying distributional assumptions inherent in MLE.

The deficiency with using a simple variate approach was evident when adjacent customer segments were models. Adjacent segments yielded different co-efficients. The population as a whole was not amenable to a one variate explanation. One approach would have been to segment the population and model segments separately. Given the small population size, and difficulty of melding models at segment boundaries, we opted for a more flexible, single model.

Our options were to write-off the survey, or stratify the responses and model each strata separately (with the huge amount of exploration inherent in that kind of approach), or apply neural (or non-parametric) modelling. We chose neural modelling, and have kept non-parametric modelling (eg, local kernel regression) open as a future option.

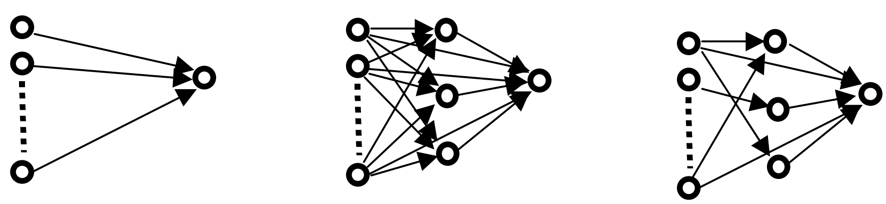
3. The Neural Network Solution.

We knew from exploratory data analysis on the survey that *locally* (within like cohorts and sub-ranges of "loyalty drivers") the survey respondents were well behaved, even though the extensive GLM models did not capture this. The simple interpretation is that variability between local regions weakened the GLM modelling too much.

However, there appeared to be sufficient consistency to propose that some regions of the dependent variable space were capable of "turning on" supplementary effects for an otherwise global effect.

Based on previous experience, we devised a neural network architecture which incorporated a GLM, but with three supplementary/derived variables. These variables were hidden neurons. [The reader should be cautioned that not all neural networks packages can support the approach outlined here – in essence to provide some supplementary correction to an underlying GLM model].

The architecture for 1a was 14 inputs (implying 15 co-efficients). For 1b, there were equivalent to 63 co-efficients. However, as the "wrapper"/transfer function for the hidden neurons was logistic, the dimensionality of the model was much lower. Effectively, each of the three neurons would turn on for a limited part of the input space. These three neurons modelled effective derived variables triggered by exceeding of critical (and learned) thresholds. For 1c, the implemented model, the number of "straight through" synapses was 4, while there were another fourteen surviving connections into the hidden layer of neurons, plus the three connections feeding from hidden neurons to the output. To manage over-fitting, a combination of weight decay and cross-validation was used. To improve the predictive value of the neural model, *after* the net's performance was measured, it was trained very briefly on a composite of all data sets, to pick up some extra signal available from cross-validation and test data [Haykin, 1999].



1a 1b 1c

Figure1: 1a: GLM (output neuron has an appropriate transfer function/wrapper). 1b GLM with three supplementary neurons. 1c GLM with three supplementary neurons, pruned. Pruning is automatic, and the model is re-trained after pruning to accommodate the "brain surgery".

The performance gain was quite astounding. The pruned architecture (version 1c) yielded an r^2 of 0.15 on hold-out data and 0.65 fitting. This converted an unuseable survey to a useable one. Furthermore, subsequent evaluation of the model for sub-regions of the independent variable space indicated even better fit for most of the model space, suggesting a small number of very irregular responses, however, we did not carry out any filtering to 'clean' the data further for re-modelling, since there were no formal grounds for rejecting outliers or irregular responses [eg, English, 1999].

The price of the performance boost was that we no longer had a *simple* equation which was understandable in traditional terms by Marketing Departments.

Readers may gain more practical knowledge of neural network modelling, with a results focus from (*inter alia*): Berry and Linoff, 1996; Bigus, 1996; Bishop, 1995; Orr and Muller, 1999; Hecht-Nielsen, 1992; and, Zurada, 1992. Haykin, 1999 provides a detailed coverage of MLE (EM algorithm) and gradient descent for practical problems.

4: A User-focussed Interface:

For a user, the expression of the underlying model, whether neural or traditional generalised regression, provides only limited insights for informing business decisions. Traditionally, for a GLM, the co-efficients to focus their discussion. We have observed in this survey, sufficient complexity and structure in the data do thwart a GLM, but a structure easily captured by a fairly simple neural extension. The remaining problem was to manage the client's expectations, including the framework they were used to for understanding their markets.

Clearly the standard report entailing many charts, and a single table of co-efficients was no longer viable. NTF's solution was a software solution wrapped around the neural model which provided the kinds of interpretation clients expected, and providing "what if" facilities. Indeed, it emerged that the client became sufficiently enamored with the simulation option that that comprise their primary use of the GUI.

Figure 2 shows the GUI in operation. In this screen capture the model computes the base and nominated experimental loyalty for three product types – ie, before and after. The current implementation simulates the loyalty response where each customer's loyalty is re-evaluated when their independent variables are supplemented as set via the sliders in the GUI. The output is the mean loyalty of the customer segment nominated by the client. This means the differential responsive of various customers factors into the overall loyalty measurement.

The client's traditional view of survey analysis – what they expect to see – is some indication of the responsiveness of loyalty to manipulation of the various drivers of loyalty. The GUI/DSS provides this directly via the magnitude of loyalty responses shown in the screen below, and also, by scenario and segment according to any designated segmentation.

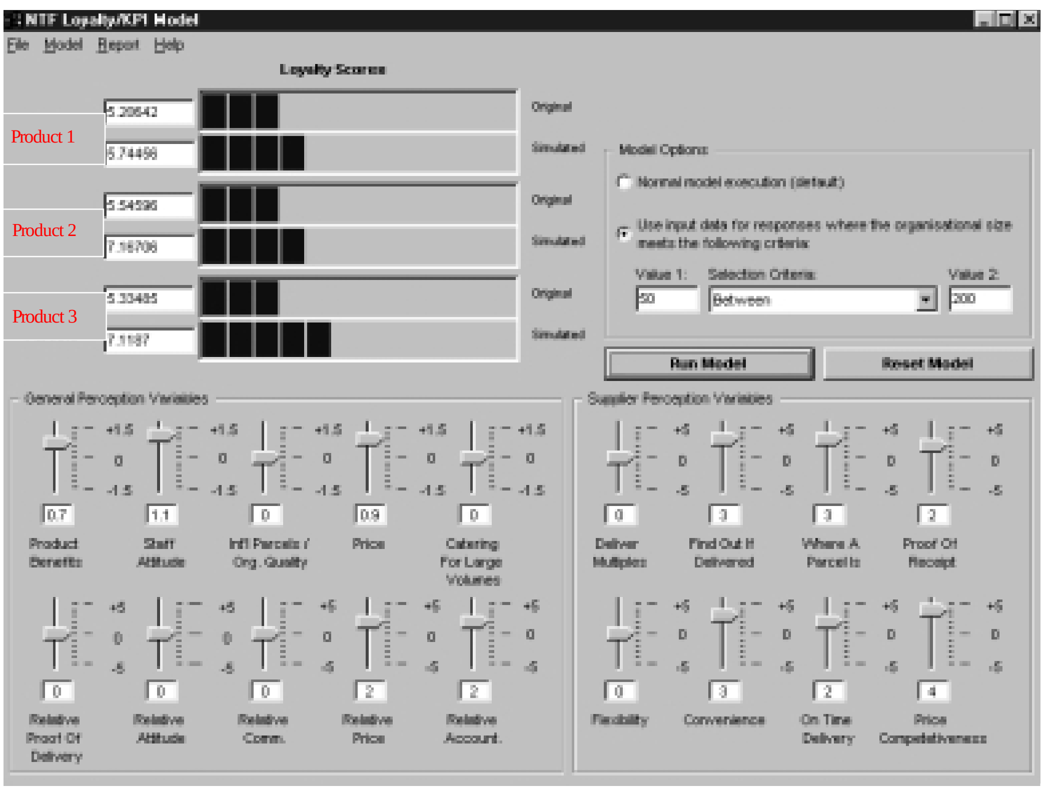


Figure 2: De-branded KPI/Loyalty GUI, main interface. The controlling sliders provide user driven simulation. Reporting pull-down presents other summary results. The small business size for this particular simulation is cut down to 50 or more and 200 or less employees.

5. Simulation Power to the user

The DSS provides the client with a direct estimate of loyalty under whatever scenarios they wish to investigate. The limitation of the this kind of simulated evaluation depends primarily on the strength of the modelling, but also on the client's appreciation of 'reasonableness of scenarios' they wish to evaluate. Subsequent developments in NTF's GUI/DSS for surveys will include optimisation algorithms and risk evaluation. To avoid recommending a totally unrealistic but "model optimal" solution to a company's bottom line problem, a "reality mask" must constrain model evaluation to realistic options, or to penalise unreality. For the client, this raises the question in their consciousness of "precisely what is a realistic option".

6. Acknowledgements

We wish to thank Dana Anspal for advice on Useability Engineering issues of this project, and acknowledge the skillful and dilligent development of the GUI by Bernard Leckning.

7. References:

Berry MJ and Linoff G, 1997, "Data Mining Techniques for Marketing, Sales and Customer Support", John Wiley and Sons.

Bigus JP, 1996, "Data Mining with Neural Networks", McGraw-Hill.

Bishop CM, 1995, "Neural Networks for Pattern Recognition", Oxford.

English LP, 1999, "Improving Data Warehouse and Business Information Quality", John Wiley and Sons.

Green PE and Wind Y, 1973, "Multiattribute Decisions in Marketing: A Measurement Approach", Dryden.

Hecht-Nielsen, R, 1990, "Neurocomputing", Addison-Wesley.

Lindsey JK, 1997, "Applying Generalised Linear Models", Springer.

McCullagh P and Nelder JA, 1989, "Generalised Linear Models", Chapman & Hall.

Orr GB and Muller K-R, 1998, "Neural Networks: Tricks of the Trade", Springer.