Neural Networks and Business Modelling - An Application of Neural Modelling Techniques to Prospect Profiling in the Telecommunications Industry

C.B. Kappert, S.W.F. Omta
University of Groningen, Faculty of Management and Organization
P.O. Box 800, 9700 AV Groningen, The Netherlands
datamind@euronet.nl

Abstract

Organizations operate in turbulent environments. Frequent and unexpected changes in competition, customer demands and technology increase the need for a dynamic, adaptive and knowledge-driven organization. Real-time data collected from operational processes and market research can be integrated and enriched towards business information. To reduce the data-overload produced by operational information systems, new data-analysis methods have been proposed to recognize patterns or relationships. Neural networks have been recently introduced to support these data mining and decision support applications. This article describes a practical application of neural networks to a specific area of business modelling, ie. prospect profiling. It is shown that neural networks can be used in this area. The benefits and pitfalls of neural networks in such data-mining applications are evaluated.

1. Introduction.

Organizations operate in turbulent environments. There is a 'complexification'[1] trend in business. Although competition forces companies to operate on a global scale (in order to profit from scale economies), customers still have to be satisfied in their individual needs. Accordingly, the customers are approached within small segments, or even at the individual level [10,12]. A shift is being made from transaction-oriented sales management towards management of customer relationships over time. Therefore, abstract product attributes like quality, image and excellent service are becoming more important at the expense of the material part of the marketing proposition [18]. This makes the definition of a product-market combination more complicated. Product life cycles are getting shorter, and the impact of innovations and non-linear product improvements decrease the time frame for accurate forecasts [6].

Managers look for possibilities to decrease the uncertainty in their decision making processes. By representing business processes in databases, they hope to gain insight on the essential parameters driving the business system. The databases produced a wealth of data. Until recently the data were mainly used for operational control, and not particularly suited for typical management tasks, such as strategy development and forecasting. Making these data available to management has been a focus for information technology projects in recent years [11].

The central problem is that the amount of operational data has exploded. Managers are facing an information overload [8]. The data stored in operational databases are sometimes ambiguous, redundant and unstructured. Reducement and structuring of these data towards business models which reflect the key business parameters is necessary. Although every manager has an intuitive mental model, this judgement-based model can be supported and augmented with (data) models created by information technology [3,25]. Mental models incorporate at most seven to ten factors, intuitively chosen and related in a linear way [21]. Research suggests that such models are too simple to represent the real world complexity.

The information technology (i.e. the datamining and decision support applications) has to take into account the specific characteristics of the real business world (data). First of all, the data stored in the oprational databases can be corrupted and incomplete. Incompleteness results from the technological, economical or even ethical constraints on data gathering. Corruption or noise is easily created during data gathering, manipulation and interpretation processes. Noise is defined as the (in principle) random variance of business parameters. Secondly, the context

[1] A word borrowed from the book by Casti [6]. 'In all cases, what we're looking for is some way to compress our observations into a small set of easily digestible rules, or models, that will serve as guides to what we should expect and what we shall do in an increasingly complex, hard-to-understand world' (p.3).

dependency of management decisions makes it impossible to estimate the impact of an isolated decision variable. Multiple effects can result from multiple causes acting in an indirect way. Thirdly, non-linear relationships will almost certainly occur, for example saturation effects in promotional campaigns, or the progression of sales over the product life cycle.

Recently, neural networks have been introduced to deal with the above mentioned characteristics of business modelling. This paper addresses the use of neural network technology for modelling research data in a new service development process in the telecommunications industry. In section 2 and 3 the neural modelling technique is discussed and compared to other business analysis methodologies. In section 4 we describe the case. Finally, section 5 gives some conclusions on the use of neural network technology for business modelling.

2. Deductive versus Inductive Data Analysis.

Data-analysis for decision support can be approached in two different ways: the deductive and the inductive method [20,28]. The deductive approach reasons from hypothetical constructs or relationships towards the individual data-elements. Based on empirical evidence, the hypothesis is either accepted, rejected or conditionally justified for statistical reasons. The analysis direction is from the general towards the specific. Although supported by a wealth of statistical procedures and causal analysis methods, the methodology may have adverse effects. Theoretically plausible models can be fitted to real world data which are not generated by the model specification, resulting in a bad fit. Alternative model solutions are sometimes missed because the model builder sticks to known statistical procedures, under which such model solutions can not be defined or found. For example, standard regression techniques fit a straight line to a dataset by using a one-step least-mean-squares approach. If the dataset contains strong non-linearities or noise, this is not an optimal procedure.

In the inductive or data-driven approach no relationship between data-elements is presumed initially. Instead, a random model is incrementally improved by reducing the error between the model output or forecast and the real world data. This process varies model parameters, such as weight and bias parameters in neural network modelling, in the direction which provides the steepest descent in the (multidimensional) model error surface (see figure 1). It is possible (without specific countermeasures) to model spurious relations or covariance, but the model building process is not influenced by theoretical or methodological assumptions.

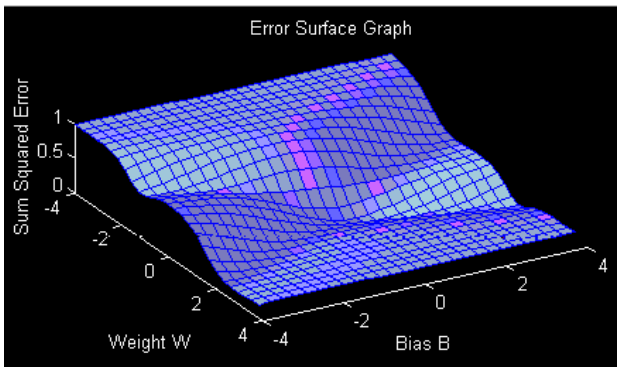

Figure 1. Error surface in weight-bias space

3. Neural Networks as inductive data analysis methodology.

Basically, the data-driven approach is the way neural networks perform their data-analysis and modelling task. Depending on the problem there are two kinds of networks: supervised and unsupervised. We limit the discussion to the supervised networks, because unsupervised networks were not used in this study. Technically speaking, a supervised neural network constructs an input-output mapping by means of a network of connected nodes. Each node, an artificial representation of the biological neuron, computes an output by means of a transfer function over a number of inputs which flow into the node [1,26] (see figure 2 and 2b).

The nodes are grouped in layers. Typically, we have an input layer, one or more hidden layer(s) and an output layer. Adding more than two hidden layers to the neural network architecture gives no advantage in most cases [4,14]. In statistical terminology, the input to a neural network is the independent variable set, the output of a neural network is the dependent variable (set). While the input layer and the output layer scale the variables, the hidden layer(s) perform(s) the (non)-linear mapping. This is done by varying the synaptic strengths, that is the relative weight of the connections between the nodes in the various layers. These synaptic strengths are optimized in an incremental process, in which the fit between the model generated data (by means of the neural input-output mapping) and the real world data is increased. This process, which is called (supervised) training, is initiated by a specific learning algorithm. It can be shown that a suitable configured network can learn any (smooth, i.e. continuous differentiable) input-output mapping [2,15,27]. This result is sometimes referred to as the computational completeness theorem or Kolmogorov theorem and provides the neural network with very powerful analytic

capabilities. Because neural modelling is based upon pattern recognition it is, in a statistical sense, a nonparametric approach[2] [17]. The choice of the functional form of the model is not determined up front. Features of the dataset are represented in the connection weights to and from the hidden layer(s), instead of being characterized by statistical parameters. The hidden nodes may be considered latent variables.

An advantage of neural models is that they cope with multi-collinearity in a natural way. By combining inputs in a specific node, covariation in the input space can be modelled and analyzed (see figure 2). This provides an additional advantage. Since no (hidden) node is solely responsible for representing a particular feature in the data, a neural network exhibits graceful degradation when the dataset is perturbed (for example, by noise)[24].

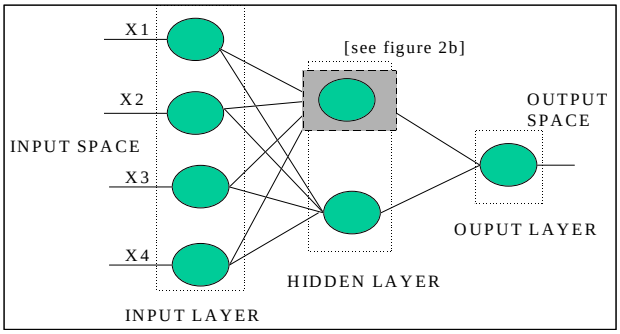

Figure 2. An example of a neural network architecture

The way neural networks perform their modelling task by incrementally generalizing over experimental data makes them suitable to deal with dynamical problems. New data can easily be incorporated into the model structure by training the network with the new data. Neural networks have therefore extensively been used to model dynamical systems, eg times series [22].

[2] The parametric model specification of statistical analysis boils down to various architectural options in neural network configurations. Choice of number of layers, neurons, transfer functions, thresholds and learning rules can influence neural network behaviour. There is not yet any theory specifying an optimal neural network architecture for a specific analysis problem [13].

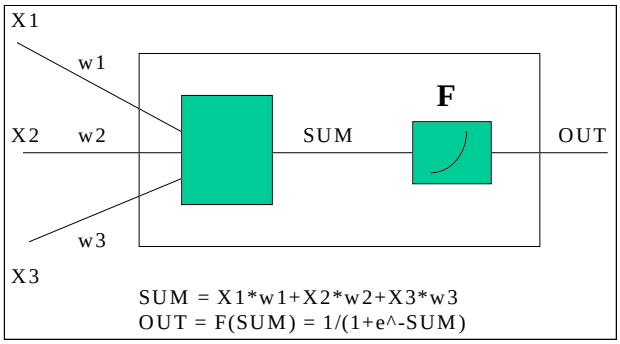

Figure 2b. Computation in an individual neuron.

The computational completeness theorem can lead to an overspecification of the model. This modelling error is called 'overfitting'. It means that the neural network model may perform well within the boundaries of the existing (training) dataset, but will produce faulty generalizations over new data. This an important problem, especially for inductive analysis systems [16]. With a procedure called 'parallel cross-validation', this risk can be reduced. A part of the dataset is split off as an independent validation set [4,14]. During training, the forecasting capability of the neural model is continuously tested on this 'test-set'. When the performance of the neural model on the test-set starts to degrade, a snapshot of the model on the training set is memorized, along with the value of the fit on the test-set. If further training does not produce a better value for the test-set fit, the memorized model becomes the 'best' model and is subsequently used. By taking this procedure, the final neural model is chosen on basis of the best forecasting ability to new data. This method provides advantages when dealing with noisy data. If noise is modelled, the performance of the neural model towards the test sample starts to degrade. By optimizing the fit on the test-set the 'best' model, given the level of noise, is taken.

The neural network training process is analytically complex. The multiple connections between the nodes with varying strength, the possibly non-linear transfer functions within nodes and the error-correction or training algorithms can produce mathematically intractable behaviour [26]. It is at start very difficult to specify to which model solution a neural network will converge. This has led to the conception that a neural network behaves like a black box. We do no take this position (see also [4]). Although neural networks are mathematically complex, the training process is strictly determined by algorithms. The resulting model can be submitted to statistical procedures to establish its reliability and the corresponding neural behaviour can be analyzed by means of sensitivity analysis. Part of the black-box conception is that analysis problems are sometimes ill-defined or

misunderstood. That problem cannot be circumvented with any other analysis methodology.

In summary, neural networks provide benefits such as adaptiveness, universal (function) approximation, robustness-to-noise and lowered sensitivity on multi-collinearity, which can prove useful in business modelling and decision support. The next section presents a business case in which neural networks are used to model the adoption of a new service by prospects in the telecommunications industry.

4. Prospect profiling for new services in the telecommunications industry.

The telecommunications industry is going through a period of turmoil [9,19]. Deregulation, technological innovations and shifting customer demands have created uncertainty for new business development management. Specifically, it has been notoriously difficult to target new services at current customers or new prospects. A mass marketing approach is, for cost-benefit reasons, not tractable and identification of the most 'attractive' prospects is therefore necessary. The likelihood of a prospect to adopt the new service can be represented in a scoring model. Such a scoring model of potential customers can be based on usage patterns on other comparable telecommunication services and socio-demographic attributes (such as risk-taking attitude, social interaction measures and age). Most of these data are already available in operational databases.

4.1. Research design.

The aim of the case study is to construct a scoring model for adopters of a new telecommunication service (for reasons of confidentiality we cannot disclose its name). The scoring model should identify the general characteristics which influence the adoption of the new service. In addition it should forecast the probability of adoption at the individual prospect level.

The data gathering stage consisted in the creation of a test market, in which a group of individuals tested the service for two months. 945 respondents took part in the trial and subsequently completed a questionnaire in which information was asked on personal and social characteristics along with specific telephone usage patterns. At the end of the trial period, 867 persons evaluated the new service as positive and would buy it when it would come available. 78 persons rated the new service negatively and would not buy the product if it came commercially available. This establishes a non-response rate of 8.2%, or a response rate of 91.8%. Although this represents a quite high response rate for new service adoption, several remarks can be made. Firstly, the way we propose to construct the scoring model is identical for situations in which the response rate is very much lower. In a sense, the magnitude of the response rate does not influence the methodology. Secondly, within this case substantial commercial savings can be reached by eliminating the worse 8.2% by means of a scoring model. Thirdly, a difference between stated adoption behaviour (the questionnaire response) and the real adoption behaviour can occur. We expect that the adoption rate will be much lower when the product comes commercially available.

The scoring model to be constructed would have to discriminate between persons with a positive attitude towards the product and persons with a more negative attitude. From an analytical standpoint, this is not an easy task. Firstly, the questionnaire variables may exhibit strong nonlinear behaviour. For example, age may have a bell-curve shaped effect on telephone usage behaviour. This represents a situation in which middle aged people use a service more often in comparison to younger and older people. Secondly, the independent variables may interact strongly. It is possible that sociodemographic variables, along with telephone usage factors drive the adoption of the new product. Thirdly, when asking respondents about their personal life style, noise may be created because of measurement difficulties like privacy, proudness etc. Each respondents gives a subjective interpretation to the questions. In addition, noise can be created because the respondent to the questionnaire is not solely responsible for telephone behaviour. This behaviour is determined at the household level rather than at the individual level.

These and other problematic issues are likely to be found in the completed questionnaires of the test market, contributing to a great deal of noise in the data. For these reasons, it would be very suprising if an analysis technique used to construct the model would perfectly discriminate between the adopters and non-adopters. Instead, we would like the model to produce a probability measure based on the independent variables which expresses the likelihood of adoption in the test market dataset. In the study, a threshold level was defined, above which a respondent was considered as an adopter, and under which the respondent would be considered a non-adopter. In first instance, this threshold equalled the mean likelyhood to buy. That is, if the model produced a output score higher than 867/945=0.91746, than this person was considered to have a positive attitude towards the service and a potentialdopter. Below average scores were considered to be the non-adopters. By comparing the

model output with the known real world output in the test market, the quality of the model is easily established.

Therefore, the dependent variable was defined on ratio-scale instead of the real world categorical scale (ie. adopter or nonadopter). An additional advantage is that the threshold level (and thereby the mean adoption rate) could be adjusted towards a level which optimizes the volume of adopters to be generated versus the costs to be made in order to reach that quantity of adopters. In direct-marketing terminology, a gains chart is constructed, in which the respondents are ranked on the basis of their adoption probabillity on the new service. It is possible to select the best prospects by taking the x% respondents with the highest probability. A substantial part of this group will indeed be adopters. By gradually decreasing the height of the probability, more and more non-adopters will be selected too. Therefore, a certain threshold level of probability has to be set to remain adequate percentages of adopters.

The dataset was split-up in two parts. 80% percent of the dataset was used to construct the model (the 'training-set'). In a random way, 20% was split off in order to provide an independent validation set (the 'test-set') to the model. By using the validation sample in neural network modelling we can estimate the generalizing ability of the model while building it. As discussed, the neural modelling process progresses in step-like fashion towards a final state, in which the forecasting ability to new data is optimized. This is a significant difference with regression techniques, in which the final model is tested on an independent sample.

Because the amount of independent variables can lead to complex models and statistical difficulties (overdimension), the dataset was reduced to a few key attributes. These variables described telephone usage. The variable set was constructed through the following questions in the questionnaire:

What was the duration of your last telephone talk [min.]
How many times did you use the phone last week
How many times did you receive a telephone call last week
How long were you on the phone last week [min.]

The ouput variable was constructed as follows:

Would you buy the product if it would come commercially available [yes/no]

There were two reasons for choosing these variables. The first reason is that the information on telephone usage is available through (operational) billing systems. A roll out procedure from the test market to the national market based on this information is therefore more easy than for any other variable set. If a commercially attractive model could be constructed on the test market data, the adoption probability of every customer of the telecommunications company can be established.

The second reason is that the quality of the response was reasonably good. Other attributes were impaired by some low quality response and missing values.

All selected independent variables were measured on a ratio-scale. The real world dependent variable has a nominal or categorical nature. This categorical value (yes/no or 1/0) was approximated by the neural model with a probability on ratio scale, taking values from the domain [0,1]. A supervised neural network was trained on this data set[3]. The answers to the four questions were given as input to the neural net, the response to the buy option was to be modelled as an output variable. After trying out some network architectures, the final network had a hidden layer containing 5 nodes. Some 1200 iterations were needed to train the network, after which the resulting model started to degenerate to a lower generalization performance. The training will take approximately 20 minutes on an 486-PC. Looking at the performance of the network on the training/model and test-set we noticed a few things.

The model-fit curve always increases as the neural network tries to fit to the 80% training sample (see figure 3). By expanding the network (increasing the number of hidden layers or nodes) we can eventually reach a 100% training fit (R^2 of 1). The network in this way memorizes the entire model dataset. But the resulting performance on the test-set will be poor. By choosing a smaller network configuration the error in generalizing to the test-set is reduced, but the resulting training fit may be lower because the neural network architecture cannot represent the complexity of the training data set. In this way, a balance has to be found in optimizing both test and training fit. The (final) neural network was trained to optimize for generalization ability, i.e. the test fit was maximized while training. The neural network was trained for considerable time and the best model was automatically saved whenever the test fit reached a maximum. Training was stopped at the moment when the improvement in training fit was negligible.

[3] The commercial neural software package '4THOUGHT' of Right Information Systems Ltd was used. This a multi-layer perceptron network system in which architectures with none, one or two hidden layers can be specified.

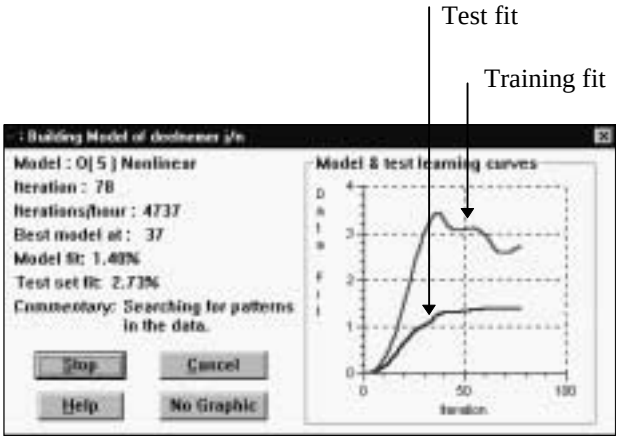

Figure 3. Training Set Fit (Model fit) and Test-set Fit while constructing the neural model.

4.2. Analysis.

Looking at the characteristics of the resulting neural network model, comparable fits on both the test sample and the training sample were found. This indicates that the network has found patterns in the training data which can (to a certain extent) be extrapolated to the test sample.

About 5% of the variation in the dependent variable could be modelled by using only four independent variables (see table 1). The low fit results from the way the scoring model was constructed. A 100% certainty in choosing between people who adopt the new service versus people who don't is not feasible and not particularly interesting. Instead a probability distribution which gives people who adopt a higher probability (above the mean probability) than people who don't will suffice. The model does not have to be perfect to be of a good quality. By looking at the gains chart[4] table, we see that this is the case (see figure 4). By selecting the 50% people with the lowest modelled probability of adopting the new service, already 85% of the people who did not adopt the new service are selected. By taking the worst 80%, nearly all non-adopters in the sample (97%) are identified. It can therefore be concluded that the neural network model is able to discriminate between the two customer groups. By taking the 20% customers with the highest modelled probability, only 3% non-respondents are identified.

[4] For details on the gains chart methodology see [23].

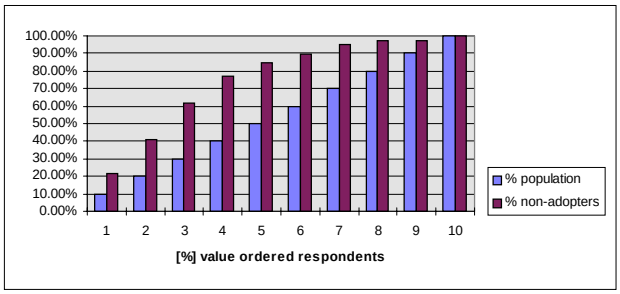

Figure 4. Gainchart of value ordered respondents (neural model).

The resulting model contained noise, non-linearities and multi-collinearity. The best model in terms of forecasting capability to unseen data was not the model with the highest training fit, the criterium on which traditional statistical techniques are built. Instead, somewhere in the iterative training process the neural network started to confuse noise with real patterns in the data, causing the test-set fit to decrease. By continuously validating against the test-set, a relatively noise-free neural model was identified.

By viewing a graph of one of the independent variables versus the dependent variable, non-linear relations are identified (see figure 5).

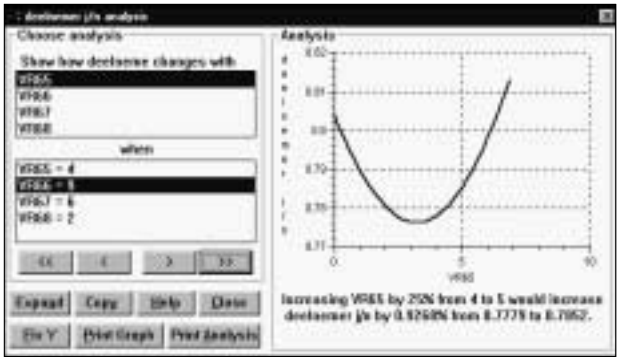

Figure 5. Nonlinear relationships with two dimensional analysis

A two-dimensional analysis can be dangerous when performing on a multi-dimensional model, in which various independent variables influence each other in their effect on the dependent variable. These interdependencies in the input space will not show in a two-dimensional analysis. When a sensitivity test was performed on the two-dimensional analysis it showed that this type of analysis was indeed dangerous. Multi-collinearity was present. A (traditional) way to tackle this problem is to (linearly) combine various independent variables into one factor, and use this factor as input for the model. This is

possible with the neural network method, by cascading independent variables through various input-space models[5]. The output of this analysis (resembling factor analysis) is subsequently used as input for the final model for the dependent variable, i.e. the adoption of the new service. But it is also possible to model those interdependencies directly, insofar the multi-collinearity is not too high.

In this study, a neural network without a hidden layer was used to simulate (log)linear regression. Without a hidden layer, the neural network is not capable of modelling non-linearities within the dataset. It was found that the resulting model performed worse on both the training set and the test-set (see table 1). The number of cases correctly classified may seem low, but results from the way the probability and the threshold level are defined within the scoring model. Although the adopter and the non-adopter group are in reality not of an equal size (867 versus 78), the scoring model will generate two groups of approximately the same size. One group is located above the mean probability to adopt (the adopters) and one group is located below that mean (the non-adopters). Because the modelled adopter and non-adopter group are of the same size, misclassifications will occur. This is the reason for taking a gains chart approach, in which the group with the lowest probability are considered non-adopters. This reduces the classification error significantly.

Table 1. Comparison of the quality of the neural network model with log-linear regression

Characteristic Model		Linear Method	Neural Method
T-statistic independent variables	Question 1	-1.6	-1.6
	Question 2	2.3	2.4
	Question 3	-0.3	-2.3
	Question 4	2.5	4.3
R-squared Model Set [80%]		1.19%	4.96%
R-squared Test Set [20%]		3.38%	5.26%
F-statistic		3.9	12.4
Correlations Model Values - Real Values		1.70%	5.00%
Number of cases correctly classified	Model	397	436
	Test	105	115
Number of cases incorrectly classified	Model	359	320
	Test	84	74

The gains chart for the log-linear model is shown in figure 6. The performance of this model is not as good as the neural model (cf. figure 4).

[5] Such a factor analysis can be performed by the use of unsupervised neural networks.

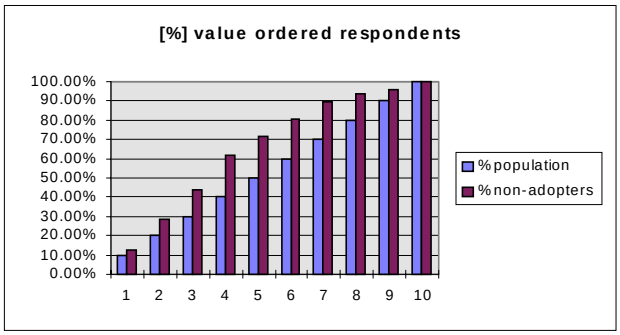

Figure 6. Gainchart of value ordered respondents (log-linear model).

4.3. Discussion.

The application of the neural net to the construction of the scoring model is a straightforward task. At start it was assumed that the problem at hand, the identification of adopters for a new service on basis of telephone usage, could show difficulties like noise, non-linearities and multi-collinearity. For this reason, a neural network was used to cope with these difficulties. The resulting model indeed showed these characteristics. A comparison with the traditional statistical technique of loglinear regression showed a superior neural network performance. In the problem at hand, neural nets proved that even with a small, imperfect variable set satisfactory and commercially attractive solutions can be found.

A substantial disadvantage of the neural network methodology is its newness and technical complexity. More cases in which neural networks are succesfully applied can increase the rate of adoption of this technique within the business and scientific community. This case study showed that an user-friendly and graphical interface for construction, analysis and testing of the neural model increased its acceptability. Moreover, the importance of a test procedure on an independent validation set is paramount. This improves the neural analysis technique itself in noisy datasets. It also improves the confidence of potential users of the model, if they see that the neural model is capable of generating accurate forecasts to new data.

5. Conclusions.

As the business environment moves towards increasing complexity, organizations try to keep up by assembling data and converting this data into information by suitable analysis techniques. Neural networks provide some interesting features when considering the difficulties encountered in modelling real business data. Although the

neural analysis techniques are very knowledge intensive, the advantages in terms of model performance can make the effort worthwhile. Neural nets can prove to be a viable option for datamining tasks. Management can use these techniques to extract knowledge from vast amounts of operational data.

This case study has shown that the neural modelling task can be performed within a time period comparable to traditional techniques. The results in terms of forecasting capability are substantially better. More cases should be studied using neural networks in order to validate the findings and gain knowledge and confidence in this new technique for business and research modelling[6].

Acknowledgements.

This study was made possible by a grant from Datamind Datamarketing Consultancy and the company where the research took place. We would like to thank Karma Sierts, Jo van Engelen and Pieter Terlouw for close collaboration. Richard Hoptroff and Right Information Systems are thanked for reviewing the technical aspects and for making neural network modelling possible.

References.

1. Anderson, James A., *An Introduction to Neural Networks*, MIT Press, Cambridge, Massachusetts, 1995.
2. Beal, R. and Jackson, T., *Neural Computing: An Introduction*, Adam Hilger, Bristol, 1991.
3. Bell, David E., Raiffa, Howard and Tversky, Amos, *Decision Making - Descriptive, normative, and prescriptive interactions*, Cambridge University Press, 1991.
4. Bishop, Christopher M., 'Neural Networks: A Pattern Recognition Perspective', *Technical Report*, NCRG/96/001, Department of Computer Science and Applied Mathematics, Aston University, January 1996.
5. Blayo, F., et. Al, *ELENA- Enhanced Learning for Evolutive Neural Architecture*, ESPRIT Basic Research Project Number 6891, Deliverable R3-B4-P Task B4: Benchmarks, 1995.
6. Brookes, Richard, *The New Marketing*, Gower Publishing Company Limited, Aldershot, England, 1988.
7. Casti, John L., *Complexification, Explaining a Paradoxical World through the Science of Suprise*, Harper Collins Publishers, New York, 1994.
8. Coates, David, Doherty, Neil and Alan French, 'The New Multivariate Jungle: Computer intensive Methods in Database Marketing', in: *Quantitative Methods in Marketing*, eds. Hooley, Graham.J and Hussey, Michael K., Academic Press, London, 1994, p 207-220.
9. Christoffersen, M., Henten, A., *Telecommunication - Limits to deregulation*, European Communication Policy Research Series, IOS Press, 1993.
10. Deighton, John, Peppers, Don and Rogers, Martha, 'Customer Transaction Databases: Present status and prospects', in: *The Marketing Information Revolution*, eds. Blattberg, R.C., Glazer, R. and Little, J.D.C., Harvard Business School Press, Boston, Massachusetts, 1994.
11. Han, Jiawei, Fu, Yongjian, Koperski, Krzysztof, Melli, Gabor, Wang, Wei and Zaiane, Osmar R., 'Knowledge Mining in Databases: An Integration of Machine Learning Methodologies with Database Technologies'', *Canadian Artificial Intelligence*, October 1995.
12. Hassan, S.S. and Kaynak, E., 'The Globalizing Consumer Markets: Issues and Concepts', in: *Globalization of Consumer Markets - Structures and Strategies*, eds. Hassan, S.S. and Kaynak, E., International Business Press, 1994b, pp. 19-25.
13. Haykin, Simon, *Neural Networks - A Comprehensive Foundation*, Macmillan College Publishing Group, 1994.
14. Hoptroff, Richard G., 'The Principles and Practice of Time Series Forecasting and Business Modelling Using Neural Nets', *Neural Computing Applications*, Vol. 1, No. 1, 1992.
15. Hornik, K., Stinchcombe, M. and White, H., 'Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks', *Neural Networks* 3(5), pp. 551-560.
16. Jensen, David D., 'Induction with Randomization Testing: Decision-Oriented Analysis of Large Data Sets', *PhD Thesis*, Sever Institute of Technology, Washington University, May 1992.
17. Kosko, Bart, *Neural Networks and Fuzzy Systems - a Dynamical Systems Approach to Machine Intelligence*, Prentice Hall, Englewood Cliffs, 1992.
18. Levitt, Theodore, *The Marketing Imagination*, The Free Press, New York, 1983.
19. Littlechild, S.C., *Elements of telecommunication Economics*, Institution of Electrical Engineers, Telecommunication series No. 7, 1979.
20. Losee, John, *Philosophy of Science*, Oxford University Press, third edition, 1993.
21. Miller, G.A., 'The Magic Number Seven Plus or Minus Two: Some Limits On Our Capacity for Processing Information', *Psychological Review*, 63, 81-97, 1956.
22. Moody, John, 'Economic Forecasting: Challenges and Neural Network Solutions', International Symposium on Artificial Neural Networks, Hsinchu, Taiwan, December 1995.
23. Roberts, Mary Lou and Berger, Paul D., *Direct Marketing Management*, Prentice Hall, Englewood Cliffs, 1989.
24. Rummelhart, D.E., McClelland and the PDP Research Group, *Parallel Distributed Processing*, Volume 1, The MIT Press, Cambridge, 1989.
25. Simon, H.A., *Economics, Bounded Rationality and the Cognitive Mind*, Edward Elgar Publishing Company, Vermont, 1992.
26. Wasserman, Philip D., *Neural Computing - Theory and Practice*, Van Nostrand Reinhold, New York, 1989.

[6] A Esprit research project compared various adaptive techniques with each other. See [5] for more details.

27. White, H., ‘Connectionist Nonparametric Regression: Multi-layer Feedforward Networks Can Learn Arbitrary Mappings’, *Neural Networks*, 3, 1990, p. 535-549.
28. Ziman, John, *Reliable Knowledge - An exploration of the grounds for belief in science*, Cambridge University Press, Cambridge, 1978.