

**NEURAL NETWORKS FOR TARGET SELECTION IN DIRECT  
MARKETING**

**ROB POTHARST, UZAY KAYMAK, WIM PIJLS**

ERIM REPORT SERIES <i>RESEARCH IN MANAGEMENT</i>	
ERIM Report Series reference number	ERS-2001-14-LIS
Publication	March 2001
Number of pages	16
Email address corresponding author	Potharst@few.eur.nl
Address	Erasmus Research Institute of Management (ERIM) Rotterdam School of Management / Faculteit Bedrijfskunde Erasmus Universiteit Rotterdam P.O.Box 1738 3000 DR Rotterdam, The Netherlands Phone: +31 10 408 1182 Fax: +31 10 408 9640 Email: info@erim.eur.nl Internet: <a href="http://www.erim.eur.nl">www.erim.eur.nl</a>

Bibliographic data and classifications of all the ERIM reports are also available on the ERIM website:  
[www.erim.eur.nl](http://www.erim.eur.nl)

# ERASMUS RESEARCH INSTITUTE OF MANAGEMENT

## REPORT SERIES *RESEARCH IN MANAGEMENT*

BIBLIOGRAPHIC DATA AND CLASSIFICATIONS		
Abstract	<p>Partly due to a growing interest in direct marketing, it has become an important application field for data mining. Many techniques have been applied to select the targets in commercial applications, such as statistical regression, regression trees, neural computing, fuzzy clustering and association rules. Modeling of charity donations has also recently been considered. The availability of a large number of techniques for analyzing the data may look overwhelming and ultimately unnecessary at first. However, the amount of data used in direct marketing is tremendous. Further, there are different types of data and likely strong nonlinear relations amongst different groups within the data. Therefore, it is unlikely that there will be a single method that can be used under all circumstances. For that reason, it is important to have access to a range of different target selection methods that can be used in a complementary fashion. In this respect, learning systems such as neural networks have the advantage that they can adapt to the nonlinearity in the data to capture the complex relations. This is an important motivation for applying neural networks for target selection. In this report, neural networks are applied to target selection in modeling of charity donations. Various stages of model building are described by using data from a large Dutch charity organization as a case. The results are compared with the results of more traditional methods for target selection such as logistic regression and CHAID.</p>	
Library of Congress Classification (LCC)	5001-6182	Business
	5201-5982	Business Science
	HF 5415.126	Direct Marketing
Journal of Economic Literature (JEL)	M	Business Administration and Business Economics
	M 11	Production Management
	R 4	Transportation Systems
	M 31 C 45	Marketing Neural Networks and Related Topics
European Business Schools Library Group (EBSLG)	85 A	Business General
	260 K	Logistics
	240 B	Information Systems Management
	280 K	Marketing Science (quantitative analysis)
Gemeenschappelijke Onderwerpsontsluiting (GOO)		
Classification GOO	85.00	Bedrijfskunde, Organisatiekunde: algemeen
	85.34	Logistiek management
	85.20	Bestuurlijke informatie, informatieverzorging
	85.40	Marketing
	54.72	Kunstmatige Intelligentie
Keywords GOO	Bedrijfskunde / Bedrijfseconomie	
	Bedrijfsprocessen, logistiek, management informatiesystemen	
	Direct Marketing, Data mining, Neurale Netwerken	
Free keywords	neural networks, target selection, direct marketing, direct mail, data mining.	

# Neural Networks for Target Selection in Direct Marketing

**Rob Potharst, Uzay Kaymak, Wim Pijls**  
**Erasmus University Rotterdam**  
**Faculty of Economics, Department of Computer Science**  
**P.O. Box 1738, 3000 DR Rotterdam, the Netherlands**  
**Email: {potharst,kaymak,pijls}@few.eur.nl**

## INTRODUCTION

Large amounts of data are nowadays available to companies about their customers. These data can be used to establish and maintain a direct relationship with the customers in order to target them individually for specific product offers and services from the company. Large databases of customer and market data are maintained for this purpose. The customers to be targeted in a specific campaign are selected from the database given different types of information such as demographic information and information on the customer's personal characteristics like profession, age and purchase history. Usually, the selected customers are contacted directly by mail promoting the new products or services. For this reason, this type of marketing is called *direct marketing*. Amongst others, a growing number of bank and insurance companies are adopting direct marketing as their main strategy for interacting with their customers. Apart from commercial firms and companies, charity organizations also apply direct marketing for fund raising. Charity organizations do not have customers in the regular sense of the word, but they must be able to trace people who are more likely to donate money in order to optimize their fund raising results. The targeted individuals are then contacted by mail preferentially in relation to other individuals in the database.

Partly due to the growing interest in direct marketing, it has become an important application field for data mining. Many techniques have been applied to select the targets in commercial applications, such as statistical regression (Bult and Wansbeek 1995), regression trees (Haughton and Oulabi 1993), neural computing (Zahavi and Levin 1997), fuzzy clustering (Setnes and Kaymak 2001) and association rules (Pijls and Potharst 2000). Modeling of charity donations has only recently been considered (Jonker *et al.* 2000). The availability of a large number of techniques for analyzing the data may look overwhelming and ultimately unnecessary at first. However, the amount of data used in direct marketing is tremendous. Further, there are different types of data and likely strong nonlinear relations amongst different groups within the data. Therefore, it is unlikely that there will be a single method that can be used under all circumstances. For that reason, it is important to have access to a range of different target selection methods that can be used in a complementary fashion. In this respect, learning systems such as neural networks have the advantage that they can adapt to the nonlinearity in the data to capture the complex relations. This is an important motivation for applying neural networks for target selection. In this chapter, neural networks are applied to target selection in modeling of charity donations. Various stages of model building are described by using data from a large Dutch charity organization as a case.

The outline of the chapter is as follows. The section on direct marketing explains briefly what direct marketing is and discusses the target selection problem in direct marketing. Target selection for a charity organization is also explained. The next section discusses how neural networks can be used for building target selection models that a charity organization can use. The section on data preparation considers how the actual data for training the neural networks

are obtained from the information stored in the organization's database. The actual model building steps are explained in the following section. The results of the neural network models are discussed afterwards, followed by a comparison of the results with some other target selection methods. Finally, the chapter concludes with a short discussion.

## ***DIRECT MARKETING***

In this section, a general description is given of direct marketing and the target selection problem in direct marketing. One or more media are used in direct marketing as advertising media to trigger a response on the part of the customer. An important characteristic of direct marketing is the possibility for individual targeting of customers, after which their response can be measured at an individual level. Hence, customer specific information can be collected about the purchase history and other related characteristics and then be used later on for selecting targeted individuals in specific campaigns.

### **Target selection**

For direct marketing companies, it does not often pay off to send a product offer to all customers in the database. The costs of such a full-scale mailing campaign soon become too large, and rise above the expected returns, since the product may only be interesting to a subset of the total customer base. For that reason, the customers who are most likely to be interested in the offer should be contacted. Moreover, sending many uninteresting product offers to customers leads to irritation as such mail is often considered "junk". Hence, targeting the mail only to interested customers is important also from the customer satisfaction point-of-view. The target selection problem is the determination of those customers in the customer database who would be interested in the offer being made.

Traditionally, marketers have divided the market into segments. A product offer would then be made to those people who are considered to be in the market segment for which the product was meant. Since computing power has become widely available at low costs, automation segmentation techniques have also been used to divide the customers into a number of segments. Some of these methods like CHi-squared Automatic Interaction Detection (CHAID) have been used in direct marketing extensively. Another approach to target selection is assigning an individual score to each customer, and to target those individuals that score above a threshold value. Regression methods, neural networks, association rules and some fuzzy set approaches fall into this category, essentially.

Different types of databases are available to a direct marketing company to be used in target selection. The quality of the database used for selection purposes is considered to be the most important aspect of a successful direct mailing campaign (Bult 1993). Usually, the internal customer database of the company provides the most reliable and relevant information regarding the customer behavior. Important information such as purchase history is typically stored in the internal databases. For data mining purposes, the purchase history is often represented in terms of the frequency of purchase, the recency of purchase and the monetary value involved with the purchase. Many target selection models are based on these so-called *RFM variables*, since they capture important customer-specific information. In addition to the internal database, many companies can acquire access to external databases for additional information. There are companies, which are specialized in collecting and maintaining data for sale to direct marketing companies in need of such data. One special category is the databases with demographic information. In the Netherlands, for example, there are

companies that provide demographic information on a regional scale according to the zip code of the address of people. The demographic information provides additional group-level information, which may be useful in determining targets with similar properties. These databases contain information like the average household size in a neighborhood, average income, percent of families with a given number of cars, etc. In general, the information that the RFM variables and the demographic variables convey are complementary, and the direct marketing company has to decide which type of data will provide the most relevant information for target selection purposes.

After target selection, the targeted customers are sent promotional material regarding the offer. The promotional material is expected to trigger a response on the recipient's side for a purchase or request for more information. Only some of the customers will respond to the product offer. These customers who respond to a specific mailing campaign are called the *responders* in the rest of this chapter. Those who do not respond are called *non-responders*. The percentage of responders in a typical mailing campaign for commercial purposes is very low (only a few percent). Small improvements by the target selection algorithm can thus be commercially very rewarding.

### **Target selection in a charity organization**

The target selection problem for charity organizations is similar to the target selection for direct marketing of commercial firms, although the specifics of the situation are different. Charity organizations collect donations from their customers for raising funds to various institutes and to their activities that agree with their goals. As such, they do not have a conventional product that they offer to the customers, but they must manage to draw the customers' attention long enough for them to make a donation. To this end, the charity organizations maintain a list of people who have donated money in the past, called the *supporters*. Periodically, the people in the list are sent a mail asking for a donation. Since not everyone donates money at every mailing, the charity organization must select those people in the database who are most likely to respond to the particular mailing. This is the target selection problem that faces charity organizations. The supporters who donate money in response to a specific mailing campaign are called the responders, similarly to responders in a commercial direct marketing campaign.

Despite the similarities, the target selection problem that faces a charity organization is different from target selection in direct marketing in that the expected returns per mailing are typically much lower. The response rate, however, is much higher, typically in the order of 20 to 30%. The mailing campaigns consist of sending relatively low-cost information and promotion letters to the supporters. Sending a mailing to all known supporters is thus an option. Part of these supporters, however, may be willing to donate money more often than others, and these people should be mailed more frequently in order to maximize the total donations received by the charity. Hence, such high promise supporters must be identified, and approached more frequently. The charity organizations can develop different mailing strategies by considering who should be mailed in a particular mailing campaign, how frequent the campaigns should be organized and how their promotional material should be organized to trigger a response.

In the case of the charity organization considered in this chapter, the database contains data regarding all the supporters of the charity. Anyone who has donated in the past (after the start of automatic recording in 1980's) is considered to be a supporter, and his or her data are recorded in the database. This data includes such information as a unique registration number,

the address and the zip code of the supporter, administrative data like how often a supporter can maximally receive promotional material. Further, campaign data are included such as whether the supporter is included in a particular campaign, and how much they have donated. The organization keeps track of different campaigns and the response to different campaigns through its unique campaign identification scheme. It should be noted that the database is a cumulative one, and it includes all supporters including the inactive ones (e.g. due to death) and the ones who donate periodically through automatic payment. Clearly, these special cases are not objects of any mailing campaign and should not be considered in the target selection problem. The mailing strategy of this particular organization consists of sending a mail to every active supporter once a year. Further, the most promising supporters are sent multiple mails a year up to a maximum of four, the frequency of mailing increasing if a supporter is considered more promising. This mailing strategy has influence on the way data are collected, and hence this influence must be considered in the modeling of the most promising targets in the database.

## **NEURAL NETWORK MODELING**

This section outlines the neural network approach to target selection. The way neural networks are used to obtain the target selection model is explained, followed by a consideration of network configuration. Then, a method for the analysis and evaluation of target selection models is described.

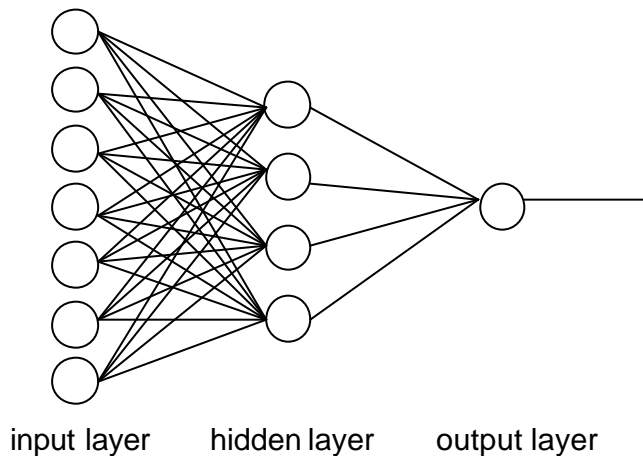
### **Target scoring**

When a segmentation method such as CHAID is used to determine the groups of supporters in the database who are most likely to respond (i.e. to donate), the data set is partitioned into disjoint groups, and the groups that are considered to be most likely to respond are mailed preferentially. Neural networks, however, do not segment data into disjoint groups. Hence, a different approach is needed to develop target selection models by using neural networks. Since target selection models try to differentiate responders from non-responders, one could consider formulating the problem as a classification problem, and develop a network that classifies each input pattern (i.e. each record regarding an individual) as a responder or non-responder. This interpretation could be encouraged by the relatively high response rate in the mailing campaigns of charity organizations. However, the classes are not well separated at all in target selection problems. Hence, finding a classifier that can separate responders from non-responders adequately is extremely difficult. Furthermore, the miss-classification costs are not symmetric. Miss-classifying a possible responder as a non-responder and not targeting that individual is more costly to the direct marketing company than miss-classifying a non-responder as a responder. Therefore, another method than pure classification should be used. One approach is to define a measure for the "likelihood" of a supporter to donate in response to a mailing. The supporters are then ordered according to this measure, and only the ones that score above a given threshold are mailed to. We will denote this method as *target scoring*. Regression models and neural network models for target selection fall into this category. For all practical purposes, the measure of "likelihood" can be interpreted as the probability of response for the particular supporter, although the probabilistic interpretation need not be the only one. The assumption is that the ordering of supporters according to the value of the measure is an estimate of the ordering according to the likelihood of response. The output of the neural network is thus a measure of likelihood of response. The goal of the training of the neural network is to determine a correct set of network parameters (i.e. weights) such that a good indication of the likelihood of response of the supporters is obtained, given the inputs to

the network. Naturally, these inputs must convey information about the characteristics of the supporters. In this approach, the neural network is essentially a nonlinear regression model.

### Network configuration

The complexity of the neural network should correspond to the degree of nonlinearity expected in the problem. In target selection problems, one models the likelihood of response at a single point in time, given various supporter characteristics and/or their past behavior. Since the system dynamics are not considered, recurrent neural networks need not be used. For practical purposes, feed-forward neural networks are sufficient for target selection problems. Important parameters that determine the complexity of feed-forward neural networks are the number of hidden layers and the number of neurons in each layer. Various methods can be used for selecting the correct range for the values of these parameters, such as growing and pruning (Bishop, 1995), heuristic search (experimentation) and optimization by using evolutionary computation (Harp *et al.*, 1989). In the target selection problem studied in this chapter, experiments have shown that a feed-forward network with one hidden layer as shown in Figure 1 is able to provide models with sufficient accuracy.



**Figure 1** Neural network architecture used for target selection. Bias inputs for the hidden layer and the output layer neurons are not shown.

The nonlinear activation function of the neurons also has an influence on the final neural network model for a given network structure and the number of neurons. The logistic sigmoid function

$$f(x) = \frac{1}{1 + e^{-x}},$$

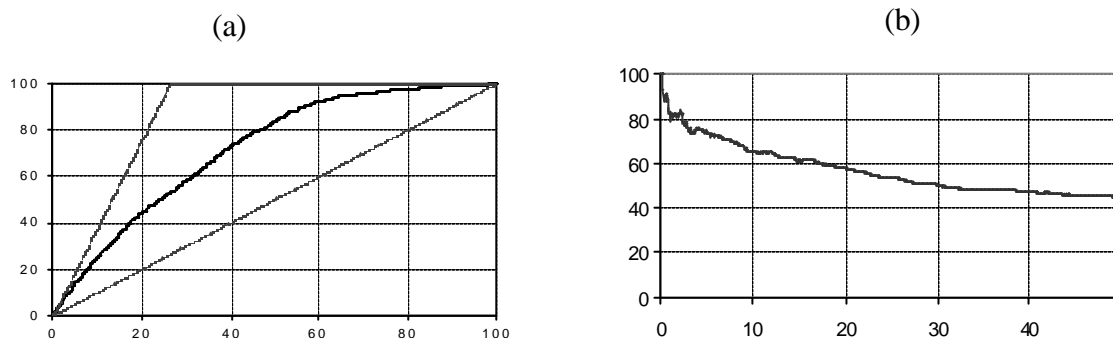
and the hyperbolic tangent sigmoid function

$$f(x) = \frac{e^x + e^{-x}}{e^x - e^{-x}}$$

are most often used. In this chapter, we prefer using the logistic sigmoid function, because the neural network output is expected to lie in the range  $[0,1]$  (likelihood of response). Moreover, a feed-forward network with one neuron in the hidden layer is then equivalent to logistic regression. Hence, the neural network model can truly be seen as a generalized nonlinear regression model.

## Model analysis and evaluation

Target selection models are typically evaluated by using the so-called *gain charts*. Figure 2a shows an example of a gain chart. A gain chart shows the gain made by using a derived model for target selection compared to a random selection of targets. It is assumed that the supporters are ordered according to the likelihood of response obtained from the model before plotting the gain chart. The horizontal axis shows the size of the mailed group as percentage of the total group of supporters. The y-axis shows the percentage of responders within the group selected for mailing, compared to the total number of responders in the considered data set. Thus, the gain chart shows how many of the responders in the analyzed campaign would have been targeted by the model, were the mailing sent to a subset of the total set. The model in Figure 2a shows that with a campaign directed to 20% of the supporters in the database, 44% of all responders could have been targeted, as opposed to only 20% in a randomly targeted mailing of the same size. A steeply increasing gain chart is desirable, since the model should ideally select only the people who really would have donated in response to the mailing. The gain chart corresponding to this ideal case is also shown in Figure 2a. Conversely, a straight line from the origin to the upper right corresponds to a random selection, where no model is used. The closer the gain chart of a target selection model to the ideal gain chart, the better that model is assumed to be. Given the gain chart for a model, the analyst can decide upon a threshold below which the supporters should be mailed. For example, if the strategy of the charity organization is to mail the best quartile of their supporter base in a certain mailing, the threshold is selected at 25%.



**Figure 2** An example gain chart (a), and an example hit probability chart (b).

Another way for studying the model output is to plot a *hit probability chart* as shown in Figure 2b. A hit probability chart shows what percentage of a selected population can be expected to donate. The x-axis shows a percentage of the total group selected for mailing, and the y-axis shows what percentage of that group is a responder for the particular mailing. Similar to the gain charts, the supporters must be ordered beforehand according to their likelihood of response obtained from the model output. The hit probability chart for a successful model starts with high values of hit probability. As the size of the selected group increases, the percentage of responders within the selected group will decrease until the hit percentage is equal to the percentage of responders within the total mailing considered in the campaign.

The gain charts and the hit probability charts are just tools for assessing the performance of the target selection models. They can not be used to assess the validity of the model obtained.



Usual neural network validation techniques are required before the performance of the models can be compared in terms of the gain charts and the hit probability charts. At any rate, the data must be split into at least one training set and one test set in order to be able to do basic model validation.

## **DATA PREPARATION**

Before any neural network modeling can begin, data must be prepared for training and validating the networks. The raw data in the database are hardly ever in a form directly usable for training neural networks. Features must be selected that are relevant for the desired model, and they must be computed from the raw data that are present in the database. Furthermore, suitable training and validation sets must be selected. This section explains how these steps are done. It also gives information regarding the raw data that were at our disposal for the case considered in this chapter.

### **Description of raw data**

The database available to us contains information regarding over 725000 supporters, of which about 675000 are actually considered for the mailing campaigns. In addition to the supporters known to the organization, mailing campaigns are also organized from time to time for possible supporters from databases external to the organization. Once people from these "external" campaigns donate money, they are included in the organization's own database. Hence, the database has grown over time, and the supporters have data regarding the campaigns after their first-ever-recorded donation. The database contains data regarding 26 mailing campaigns over a period of six years. For each mailing campaign, the supporters who were mailed in that campaign are recorded, as well as the amount they have donated (zero or more) in guilders. The mailing dates for each campaign are also known. Also recorded is the date at which the supporter has donated money in response to a particular mailing. Hence, a supporter may have donated multiple times in response to a particular campaign. In that case, the total amount donated together with the date of the last donation is recorded. The total recorded data amounts to a database of about 400MB.

### **Data set size**

It would be great if one could just submit the supporter database to some neural network software package, which processes it automatically until a nice robust neural network model for target selection is delivered. Unfortunately, the state of machine intelligence is far from this level, and the marketer has to take important decisions before the neural network models can be obtained. One of these decisions concerns the size of the data set to be used for the training and the validation of the models. Using the entire customer base for model building is inefficient for a number of reasons. For one thing, it would be very inefficient to build the models with the entire database that contains hundreds of thousands of clients, if a model of similar accuracy can be built by using a reasonably sized representative sample from the database, say 5000 clients. Secondly, the information most relevant for the models is often not directly encoded in the database. It must be decided upon separately by using valuable domain knowledge, and the relevant features that summarize the most important aspects of the available data must be selected. This process is also known as *feature selection*. These features are ultimately presented to the neural network for computing. Finally, since data are available for multiple (26 over a period of six years, in our case) mailing campaigns, it should be decided data regarding which campaigns are going to be used to obtain the target selection models.

In order to make things more manageable in our case study, we decided to restrict the size of the data sets used for building the neural network models. A random sample 10 000 supporter records have been taken and then split into two data sets: a training set of 5000 records and a test set of the same size. To assess the loss of accuracy that one suffers when re-sizing from 675 000 records to 10 000 records, we can make the following statement which is based on a statistical calculation. Suppose in the total population (the entire supporter database) there is a fraction of say 35% that satisfies some property that one is interested in. Then, this fraction will be somewhere between 34.07 % and 35.93% with probability 99% in the random sample. So, we lose about 1% of the accuracy of our results. We are able to bear this loss, however, especially when considering many other sources of uncertainty that surround a project like this.

### **Feature selection**

Feature selection can be seen as finding a set of variables that are most explanatory as inputs to a particular model. It is often also a data reduction process, since the information that is implicit in a database is extracted into a number of meaningful features, which can be expected to have an influence on the modeled behavior. In the target selection case, RFM-variables capture relevant information for modeling the response behavior of customers. Constructing such RFM-variables is common practice in direct marketing, see for instance (Bauer 1988). For our case, we constructed seven features from the supporter donation history data, of which two can be seen as recency features (R1 and R2), two as frequency features (F1 and F2), and three as features concerning monetary value (M1 to M3). Note that depending on the actual mailing campaign that is used for obtaining the modeling data, each of these features can be calculated for a different moment in time. We denote this moment as “*now*”. In other words, the features must be re-computed for every mailing campaign that is considered, i.e. whose response is being modeled. The following list shows the seven features that we have used.

R1: the number of weeks since the supporter responded to a mailing the last time before *now*

R2: the number of months since the supporter’s first-ever donation

F1: the fraction of the mailings the supporter has responded to

F2: the response time of the supporter in the period before *now*

M1: the average donated amount over all responded mailings before *now*

M2: the amount the supporter donated at his/her last response before *now*

M3: the average amount that the supporter has donated per year until *now*

Each of these features describes a different aspect of the donation behavior of the supporter and may be helpful in predicting the supporter’s future response behavior.

### **Data set selection**

The final step before the preparation of a reduced data set for training and validation of neural networks for target selection is deciding upon which mailing campaigns to use for building the models. In our case, which of the available 26 campaigns had to be used for modeling purposes? One might consider constructing a model for each mailing campaign, but we

decided against this for several reasons. First of all, we would have ended up with multiple models to base our predictions on. Secondly, the models are not likely to be of the same quality, since the more recent a campaign, the more historical data are available. Thirdly, as the database has grown over time, information about more supporters is available from the more recent campaigns. Of course, a larger mailing campaign contains more information than a smaller one, since only the responses of the participating supporters can be measured. Supporters that do not receive a mailing do not disclose their response behavior. We knew that the charity organization mails all active supporters in their database at least once a year. Therefore, we decided to base our model on the two most recent full mailing campaigns. The mailing from 1998 has been used to obtain a neural network model. The mailing from the most recent year in the database, 1999, has been used to assess the predictive power of the model obtained. In doing so, we used as much history information as possible and based the model on the most informative mailings.

Thus, we composed the following four data sets from the raw 5000 record training and test sets: a training set and a test set for the 1998 mailing, and a training and test set for the 1999 mailing. Each of these data sets contains data for the supporters on 8 variables: the 7 RFM features explained in the previous paragraph and calculated for the corresponding mailing, and the response (yes or no) to that mailing. The sizes of these data sets are shown in Table 1.

<b>Data set</b>	<b>Size</b>
Training set '98	4057
Test set '98	4080
Training set '99	4111
Test set '99	4131

**Table 1: The size of the final data sets.**

In each case, the number of records is less than 5000, since only those supporters have been included who have received the mailing.

## **MODEL BUILDING**

This section explains the details of the model building step of target selection case discussed in this chapter. The tools available for neural network modeling, data preparation and network training are discussed.

### **Available tools**

A multitude of training and modeling tools for neural networks are available to the analysts, nowadays. The choice ranges from stand-alone neural network packages to specialized extensions of large numerical modeling and quantitative analysis tools. Many packages can be downloaded from the Internet. Further, extensive product information is available on the Internet for a detailed analysis of pro's and con's of different packages. Stand-alone neural network packages provide an intuitive and often simple interface that simplifies the use of the packages considerably. Extensions to already existing quantitative analysis tools, however, may provide specific functionality not existing elsewhere, or they may provide a wider range of training methods, neuron types, network architectures, etc. Almost all general-purpose packages provide functionality for modeling with feed-forward neural networks where

variations of the backpropagation algorithm are used for the training. Most statistical analysis tools nowadays contain neural network modules, too. The advantage of these packages is that the analyst continues working in a familiar environment, without having to learn yet another interface for the new functionality. Moreover, the comparison of the new modeling technique to other techniques is simpler as all functionality is accessed in a single package.

For our case study, we have used a combination of Microsoft Excel and the statistical analysis package SPSS. Excel was used for data preparation, data analysis and feature selection. We used C++ for calculating the features, but this could have been done in Excel as well. The modeling was done in SPSS 10.0 by using the neural network package Neural Connection 2.1. The models have been validated in SPSS and then the results are exported to Excel for graphical visualization, such as producing the gain charts. Comparison to other techniques discussed in a later section was also done in SPSS.

### **Network training**

The neural network package Neural Connection 2.1 allows the analyst to build feed-forward neural networks with different configurations, using various training methods and different parameter settings. Default values for most options are provided that work reasonably well for a wide range of problems. In our modeling efforts, we have used as much as possible the default values, since extensive experimentation with different parameter settings may lead to small improvements, but at the cost of much valuable modeling time.

Perhaps the most important decision to be made for neural network modeling is the selection of a network configuration, i.e. the number of inputs, the number of outputs, the number of hidden layers and the number of neurons in the hidden layers. The number of inputs and outputs is part of the problem specification and feature selection. In our case, there are seven inputs (one corresponding to each RFM feature used in the model) and one output (a measure for the likelihood of response). As explained in previous sections, only one hidden layer is sufficient for most target selection problems. The main question is then determining the correct number of neurons in the hidden layer for a rich enough representation of the problem. Increasing the number of neurons will increase the ability of the network to model more complex relations between the inputs and the outputs, but the risk of over-training the model increases. If that happens, the generalization power of the neural network will be small, and its predictive capability will decrease.

Neural Connection has an automated network-growing algorithm for selecting the best number of neurons in the hidden layer. Roughly, this algorithm performs network training in multiple stages, where the training in each stage is done by using different subsets of the total training data set. By varying the number of neurons and the training data used at each stage, the algorithm estimates a correct number of neurons for the hidden layer. Running this algorithm on our data set has shown that four hidden layer neurons should be used for building our target selection model. We have also experimented by manually changing the number of neurons in the hidden layer, but the results did not improve significantly. Hence, we have used four neurons in the hidden layer and the network architecture looked like the one shown in Figure 1.

Other neural network parameters have been selected as follows. The activation function of the hidden layer neurons has been selected as the logistic sigmoid. The reasons for this choice have been motivated in a previous section. A linear activation function is used for the output neuron, since this reduces the complexity of the training. Conjugate gradient learning rule has been used, which is also a gradient-descent algorithm like backpropagation. With seven

neurons in the input layer, four hidden layer neurons and one neuron in the output layer, the network has in total 37 weight factors including the weights corresponding to the bias inputs of the neurons. These weight factors have been initialized randomly in the interval (-0.1,0.1) drawn from a uniformly distributed population. The stopping rules for the training have been left at their default values. Typically, the training stops after a given number of epochs or if a specified accuracy is achieved.

Note that both the input variables and the target variable have been normalized according to

$$\hat{x}_i = \frac{x_i - \mathbf{m}}{\mathbf{s}},$$

where  $x_i$  is one of the variables,  $\mathbf{m}$  is the mean of that variable's values in the data set,  $\mathbf{s}$  is its standard deviation. This normalization brings all variables to a similar range, which improves the stability of the training process: the network need not deal with weight values that are orders of magnitude different from one another.

## RESULTS

This section describes the two neural network models that we built: one based on the responses to the '98 mailing campaign and one based on the 99 mailing. The performance of both networks is measured on the test sets and compared. Special attention is paid to the performance of the '98 model when applied to the '99 data, since that could throw some light on the predictive power of a neural network model for future years. The results are illustrated with gain charts (GC's) and hit probability charts (HPC's).

The neural network that was trained on the '98 training set, is shown in Tables 2 and 3.

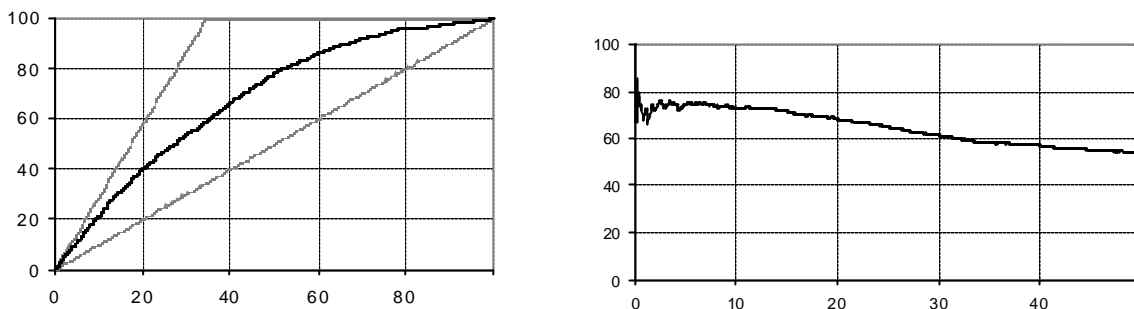
	<i>R1</i>	<i>R2</i>	<i>F1</i>	<i>F2</i>	<i>M1</i>	<i>M2</i>	<i>M3</i>	<i>Bias</i>
<i>H1</i>	-1.685016	+0.261069	+2.000788	-0.170105	+0.876737	+0.175668	+0.373603	+0.068256
<i>H2</i>	-1.320779	-0.106616	+1.638559	+0.013305	+0.670011	+0.007421	+0.335783	-0.110847
<i>H3</i>	-0.794401	+0.127476	+1.056931	-0.147008	+0.571066	-0.159912	+0.293345	-0.150575
<i>H4</i>	-0.958907	-0.052713	+1.427672	-0.057567	+0.501380	+0.137789	+0.334625	+0.069467

**Table 2: Weights of the '98 network input to hidden layer**

	<i>H1</i>	<i>H2</i>	<i>H3</i>	<i>H4</i>	<i>Bias</i>
<i>output</i>	+0.554891	+0.055592	+2.616742	-1.754274	-0.640994

**Table 3: Weights of the '98 network hidden to output layer**

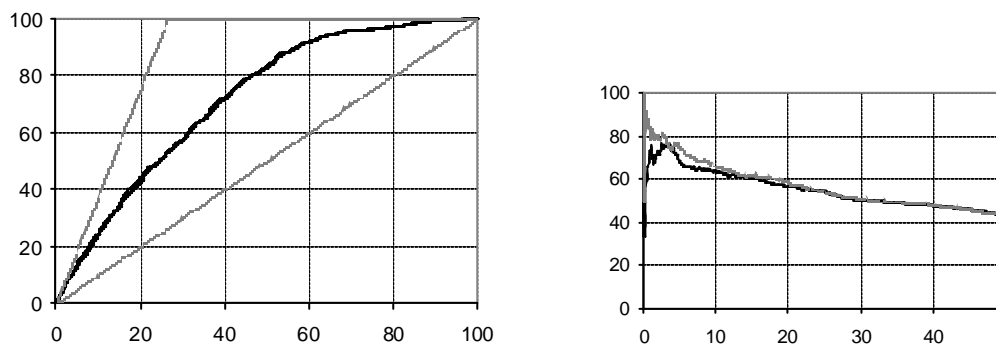
As can be seen, the network has 4 hidden nodes *H1*, ..., *H4*. The accuracy of this network on the independent '98 test set is 74.2% which is comparable to the accuracy of standard classification methods for this data set. To see how this neural network would perform when it would be used to select the target group in the '98 mailing, we used it to score all clients in the '98 test set and compared these scores with the actual responses. For this analysis we produced the gain chart and hit probability chart of Figure 3.



**Figure 3** Gain chart (a) and hit probability chart (b) for the '98 neural network on the '98 test set.

From the gain chart, it can be concluded that the neural network is a substantial improvement over a random selection of targets (the lower straight line), and especially in the lower region (0 – 5%) approaches the maximum possible gain (the upper straight line). From the hit probability chart we infer that for instance, when mailing 10% of the clients we will be able to generate a response of over 70%, while the average response for this data set is about 30%.

A similar network was trained on the '99 training set. This network also contains 4 hidden nodes and, although the individual weights differ substantially from the weights of the '98 network, the outcomes of both networks correlate highly on all four data sets. Therefore, it is not surprising that the accuracy of both networks on their respective test sets does not differ much (for the '99 network it is 77.2% on the '99 test set) and the gain charts and hit probability charts for the '99 network on the '99 test set look very similar to those of the '98 network. If both the '98 network and the '99 network are applied to the '99 data set, we get the GC's and HPC's of Figure 4.



**Figure 4** Gain charts (a) and hit probability charts (b) for the '98 and the '99 neural network on the '99 test set; the '99 network is gray in (b).

We can infer from the hit probability chart that the first few selects of the '98 network are non-responders: the curve starts at 0% but jumps to over 70% very quickly. From there onwards both curves behave very similar. The gain charts of both networks are practically overlapping. We conclude from these charts that the performance of the '98 network is not visibly different from that of the '99 network on the '99 data. Of course, this is very interesting since it implies that the neural network can predict very well from one year to the other. The question whether this remains so then the years are further apart, was not pursued any further in this study.

## COMPARATIVE STUDY

In this section we compare our neural network model for the target selection problem with two well-known methods for this problem, namely CHAID and logistic regression. CHAID is an example of a method that uses segmentation (Kass 1980), while logistic regression like a neural network model makes use of scores. Both methods are widely used in the direct mailing world (Bult 1993).

When we use CHAID on a data set, this data set is successively split up into parts or segments on the basis of different feature-values. CHAID tries to form groups that are internally homogeneous with respect to response rate. Thus, it produces segments of the data with widely different response rates. In the end, CHAID produces a tree-like structure whose leaves are the sought market segments. Subsequently, when we do target selection, we select as our targets those segments that have the highest response rates. For applying the CHAID method to target selection for the charity organization, we made use of the package Answer Tree 2.1, which is an add-in to the SPSS statistical package. We used most of the default settings for the CHAID module in Answer Tree, except for the minimum number of cases in a child node, which we set to 100. This number is still comparatively low for a data set of more than 4000 cases. The maximum depth of the tree was set to 3, which is equal to the default setting.

Like we did with our neural network models, we built a CHAID tree for both the '98 training set and for the '99 training set. The '98 tree has 18 leaves and uses the features R1, F1, F2 and M1 as splitting variables. In fact, the top split uses the feature F1 as splitting variable. The accuracy of this tree on the '98 test set is 72.6, which is similar to the comparable figure for the neural network model. However, when we select targets on the basis of this CHAID tree for the '98 and '99 test sets, the resulting GC's and HPC's clearly fall short of those of the

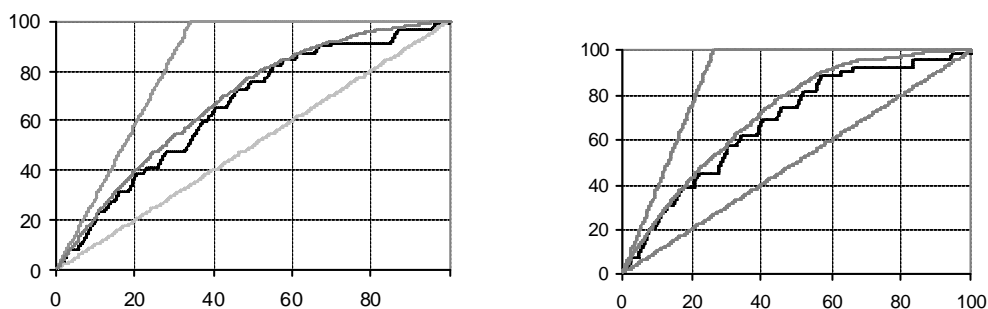
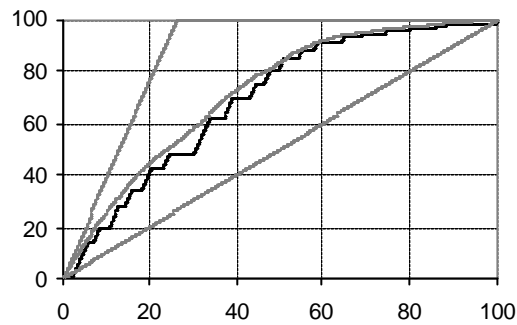


Figure 5 Comparison of the performance of the '98 neural network and the '98 CHAID model on the '98 test set (a) and the '99 test set (b).

neural models as seen in Figure 5.

We also build a CHAID tree using the '99 training set. This resulted in a tree with 20 leaves and it uses the features R1, F1, F2 and M3. Surprisingly, the top splitting feature, M3, is not used at all by the '98 CHAID model. However, the performance of this '99 CHAID model on the '99 test set is very similar to that of the '98 model, as can be seen from Figure 6.



**Figure 6 Comparison of the performance of the '99 neural network and the '99 CHAID model on the '99 test set .**

The piecewise linear character of the CHAID gain charts can be explained by the fact, that with the CHAID method all clients in the same market segment are treated alike, whereas in our neural model, each client gets a different response likelihood score. Presumably, the same explanation can be given for the fall back in performance of CHAID compared to the neural models: the granularity of the neural models is higher than that of the segmented CHAID models. In a neural model, each client's response likelihood is estimated, while in a CHAID model, all clients of a segment are treated alike.

When we use the logistic regression technique, we try to fit the following model to our data:

$$y = f( w_0 + w_1x_1 + w_2x_2 + \dots + w_nx_n ) + e,$$

where  $y$  is the response variable,  $f$  is the logistic sigmoid function, the symbols  $x_1, \dots, x_n$  stand for the feature values in the data set and  $e$  is an error term. Like in a neural network model, weights  $w_0, \dots, w_n$  are looked for that maximize the fit of this model to our data. From the above description it is clear that the logistic regression model can be seen as a special case of a feed-forward neural network, i.e. one with no hidden layer and one output node, using a logistic transfer function. Since logistic regression is a standard analysis technique in the SPSS 10.0 statistical package, we used that for our charity case.

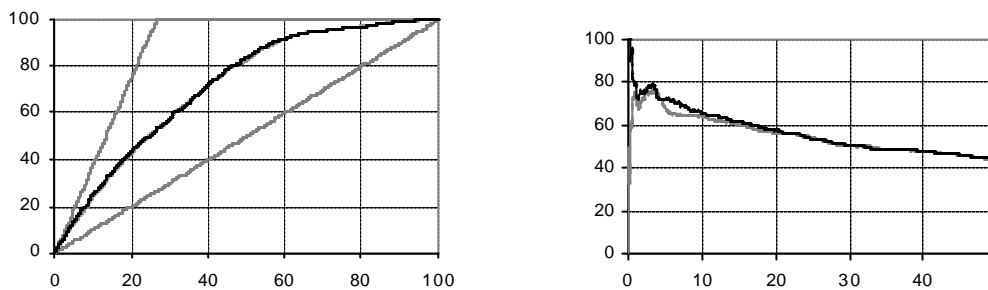
Again, we fitted two logistic regression models: one to the '98 training set and one to the '99 training set. The weights of both models are shown in Table 4, where the bias weight is the  $w_0$  from our formula above.



	<i>R1</i>	<i>R2</i>	<i>F1</i>	<i>F2</i>	<i>M1</i>	<i>M2</i>	<i>M3</i>	<i>Bias</i>
'98	-0.011	+0.084	+2.672	-0.001	-0.002	-0.003	+0.004	-1.654
'99	-0.012	+0.040	+2.117	-0.003	-0.006	-0.003	+0.005	-1.282

**Table 4: The weights of the '98 and '99 logistic regression models.**

The weights of the variables R2 and F1 appear to have a significant contribution to the models. It seems interesting to note that F1 also appears in both CHAID models, whereas R2 appears in neither one of the CHAID models. When we look at the performance of the '98 logistic regression model for the target selection problem, we read from Figure 7 that the gain charts almost overlap, while for the HPC's we notice a great similarity except for the very low values of the percentage selected. For those values the neural network is falling short, a phenomenon that was already noted in the Results section. For the '99 logistic regression model we found very similar results.



**Figure 7 Gain charts (a) and hit probability charts (b) of the '98 neural network (gray) and the '98 logistic regression model applied on the '99 test set.**

To conclude, we state that our neural network models perform markedly better than the CHAID models we generated, while they perform on an equal level with the logistic regression models we studied.

## CONCLUSIONS

We have applied neural networks to the target selection problem for some mailing campaigns of a large Dutch charity organization. Furthermore, we compared our neural network models with models built with some commonly used methods such as CHAID and logistic regression. Not only appear the neural models to perform at least as well as these others, but contrary to popular belief, it did not cost much effort to build them when using a modern modeling tool such as SPSS Neural Connection. As is often the case most of the effort had to be invested in the data preparation stage, especially feature selection.

## **ACKNOWLEDGEMENT**

We thank Jedid-Jah Jonker and Philip-Hans Franses for making available the data for this project and their pleasant and stimulating cooperation.

## **REFERENCES**

- Bauer, C.L. (1988). A direct mail customer purchase model, *Journal of Direct Marketing*, **2**, 16-24.
- Bishop, C.M. (1995). *Neural Networks for Pattern Recognition*. Clarendon Press, Oxford.
- Bult, J.R. (1993). *Target Selection for Direct Marketing*. Ph.D. Thesis, Rijksuniversiteit Groningen, the Netherlands.
- Bult, J.R. and T.J. Wansbeek (1995). Optimal selection for direct mail, *Marketing Science*, **14**, 378-394.
- Harp, S.A., T. Samad and A. Guha (1989). Towards the genetic synthesis of neural networks, in *Proceedings of the Third International Conference on Genetic Algorithms*, Morgan Kaufmann, San Mateo, CA, 360-369.
- Haughton, D. and S. Oulabi (1993). Direct marketing modeling with CART and CHAID, *Journal of Direct Marketing*, **7**, 16-26.
- Jonker, J.-J., R. Paap and P.H. Franses (2000). Modeling Charity Donations. Technical Report (2000-07/A), Econometric Institute, Erasmus University Rotterdam, the Netherlands.
- Kass, G.V. (1980). An exploratory technique for investigating large quantities of categorical data, *Appl. Statist.*, **29** (2), 119-127.
- Pijls, W. and R. Potharst (2000). Classification and Target Group Selection based upon Frequent Patterns, Technical Report ERS-2000-40-LIS, ERIM Report series, Erasmus University, the Netherlands. URL: <http://www.eur.nl/WebDOC/doc/erim/erimrs20001020162258.pdf>
- Setnes, M. and U. Kaymak (2001). Fuzzy modeling of client preference from large data sets: an application to target selection in direct marketing, to appear in *IEEE Transactions on Fuzzy Systems*.
- Zahavi, J. and N. Levin (1995), Issues and Problems in Applying Neural Computing to Target Marketing, *Journal of Direct Marketing*, **9** (3), 33-45.
- Zahavi, J. and N. Levin (1997). Applying neural computing to target marketing, *Journal of Direct Marketing*, **11** (1), 5-22.

# Publications in the Report Series Research\* in Management

ERIM Research Program: "Business Processes, Logistics and Information Systems"

## 2001

*Bankruptcy Prediction with Rough Sets*

Jan C. Bioch & Viara Popova

ERS-2001-11-LIS

*Neural Networks for Target Selection in Direct Marketing*

Rob Potharst, Uzay Kaymak & Wim Pijls

ERS-2001-14-LIS

*An Inventory Model with Dependent Product Demands and Returns*

Gudrun P. Kiesmüller & Erwin van der Laan

ERS-2001-16-LIS

## 2000

*A Greedy Heuristic for a Three-Level Multi-Period Single-Sourcing Problem*

H. Edwin Romeijn & Dolores Romero Morales

ERS-2000-04-LIS

*Integer Constraints for Train Series Connections*

Rob A. Zuidwijk & Leo G. Kroon

ERS-2000-05-LIS

*Competitive Exception Learning Using Fuzzy Frequency Distribution*

W-M. van den Bergh & J. van den Berg

ERS-2000-06-LIS

*Models and Algorithms for Integration of Vehicle and Crew Scheduling*

Richard Freling, Dennis Huisman & Albert P.M. Wagelmans

ERS-2000-14-LIS

*Managing Knowledge in a Distributed Decision Making Context: The Way Forward for Decision Support Systems*

Sajda Qureshi & Vlatka Hlupic

ERS-2000-16-LIS

*Adaptiveness in Virtual Teams: Organisational Challenges and Research Direction*

Sajda Qureshi & Doug Vogel

ERS-2000-20-LIS

*Assessment of Sustainable Development: a Novel Approach using Fuzzy Set Theory*

A.M.G. Cornelissen, J. van den Berg, W.J. Koops, M. Grossman & H.M.J. Udo

ERS-2000-23-LIS

*Applying an Integrated Approach to Vehicle and Crew Scheduling in Practice*

Richard Freling, Dennis Huisman & Albert P.M. Wagelmans

---

\* A complete overview of the ERIM Report Series Research in Management:  
<http://www.ers.erim.nl>

ERIM Research Programs:

LIS Business Processes, Logistics and Information Systems

ORG Organizing for Performance

MKT Decision Making in Marketing Management

F&A Financial Decision Making and Accounting

STR Strategic Renewal and the Dynamics of Firms, Networks and Industries

ERS-2000-31-LIS

*An NPV and AC analysis of a stochastic inventory system with joint manufacturing and remanufacturing*

Erwin van der Laan

ERS-2000-38-LIS

*Generalizing Refinement Operators to Learn Prenex Conjunctive Normal Forms*

Shan-Hwei Nienhuys-Cheng, Wim Van Laer, Jan Ramon & Luc De Raedt

ERS-2000-39-LIS

*Classification and Target Group Selection bases upon Frequent Patterns*

Wim Pijls & Rob Potharst

ERS-2000-40-LIS

*Average Costs versus Net Present Value: a Comparison for Multi-Source Inventory Models*

Erwin van der Laan & Ruud Teunter

ERS-2000-47-LIS

*Fuzzy Modeling of Client Preference in Data-Rich Marketing Environments*

Magne Setnes & Uzay Kaymak

ERS-2000-49-LIS

*Extended Fuzzy Clustering Algorithms*

Uzay Kaymak & Magne Setnes

ERS-2000-51-LIS

*Mining frequent itemsets in memory-resident databases*

Wim Pijls & Jan C. Bioch

ERS-2000-53-LIS

*Crew Scheduling for Netherlands Railways. "Destination: Customer"*

Leo Kroon & Matteo Fischetti

ERS-2000-56-LIS