Profiling your Customers using Bayesian Networks

Paola Sebastiani
Bayesware Limited and
Department of Mathematics,
Imperial College
180 Queen's Gate, London
SW7, United Kingdom
p.sebastiani@ic.ac.uk

Marco Ramoni
Bayesware Limited and
Knowledge Media Institute,
The Open University
Milton Keynes MK7 6AA,
United Kingdom
marco@bayesware.com

Alexander Crea
Knowledge Media Institute,
The Open University
Milton Keynes MK7 6AA,
United Kingdom
a.crea@open.ac.uk

ABSTRACT

This report describes a complete Knowledge Discovery session using Bayesware Discoverer, a program for the induction of Bayesian networks from incomplete data. We build two causal models to help an American Charitable Organization understand the characteristics of respondents to direct mail fund raising campaigns. The first model is a Bayesian network induced from the database of 96,376 Lapsed donors to the June '97 renewal mailing. The network describes the dependency of the probability of response to the renewal mail on a subset of the variables in the database. The second model is a Bayesian network representing the dependency of the dollar amount of the gift on the variables in the same reduced database. This model is induced from the 5% of cases in the database corresponding to the respondents to the renewal campaign. The two models are used for both predicting the expected gift of a donor and understanding the characteristics of donors. These two uses can help the charitable organization to maximize the profit.

Keywords

Bayesian Networks, Customer Profiling, Missing Data

1. INTRODUCTION

A typical problem of direct mail fund raising campaigns is the low response rate. Recent studies have shown that adding incentives or gifts in the mailing can increase the response rate. This is the strategy implemented by an American Charity in the June '97 renewal campaign. The mailing included a gift of personalized name and address labels plus an assortment of 10 note cards and envelopes. Each mail cost the charity 0.68 dollars and resulted in a response rate of about 5% in the group of so called lapsed donors, that is, individuals who made their last donation more than a year before the '97 renewal mail. Since the donations received by the respondents ranged between 2 and 200 dollars, and the median donation was 13 dollars, the fund raiser needs to decide when it is worth sending the renewal mail to a donor, on the basis of the information available about him from the in-house database. Furthermore, the charity is interested in strategies to recapture Lapsed Donors and, therefore, in making a profile from which it would be possible to understand motivations behind their lack of response.

Extending the approach of GainSmarts, the winner of the 1998 KDD Cup competition, we build two causal models. The first model (Response-net) captures the dependency of the probability of response to the mailing campaign on the independent variables in the database. The second network (Donation-net) models the dependency of the dollar amount of the gift and it is built by using only the 5% respondents to the '97 mailing campaign. The models are Bayesian networks [?] induced from data using Bayesware Discoverer a commercial product for the induction of Bayesian networks from possibly incomplete data produced by Bayesware Limited. Bayesware Discoverer induces Bayesian networks from complete data using Bayesian methods: The comparison of different networks is based on their posterior probability, that is, the revised network probability given the information provided by the data. The program implements the *Bound* and *Collapse* method to compute a first order approximation of the scoring metric when data are incomplete [?; ?; ?].

Bayesian networks provide a compact and easy-to-use representation of the probabilistic information conveyed by the data. The network structure is an effective way to communicate dependencies among the variables. Furthermore, one can easily investigate different relationships between the variables, as well as making prediction and explanation, by *querying* the network. This last task consists of computing the conditional probability of one variable, given that the value of some variables in the network are observed, by using one of the algorithms for probabilistic reasoning [?; ?]. For example, the network Response-net shows that the probability of a donation is directly affected by the wealth rating and the urbanicity level of the donor's neighborhood. Most likely to respond are those people who leave in a wealthy suburb neighborhood of high socio-economic status.

The network Donation-net shows that the same variables influence directly the dollar amount of the donation, although those who are most likely to respond are not those who make, on average, the largest donations. Apparently, donors tend to maintain the gift amount constant and their constancy is directly proportional to the number of times they responded to similar mail offers. Beside profiling donors, the two networks can also be used to compute the expected profit from each donor in the database, so that they offer an indication of whether it is in the interest of the Charity sending the renewal mail.

This report is structured as follows. We first give a description of the methodology implemented in Bayesware Discoverer and then describe the two steps of screening and cleaning of the data to produce the databases from which Bayesware Discoverer generated the two Bayesian networks. We then give the essential information to understand the modeling procedure implemented by Bayesware Discoverer. Findings are in the last section of this report.

2. BAYESIAN NETWORKS GENERATION USING BAYESWARE DISCOVERER

Bayesware Discoverer is a knowledge discovery system based on the enabling technology of Bayesian networks. It deploys a unified framework which regards the knowledge discovery process as the automated generation of Bayesian networks from data. The core of Bayesware Discoverer implements a novel methodology to discover Bayesian networks from possibly incomplete databases [?], a generalization of the well-known Bayesian methodology by [?] to learn Bayesian networks from data.

A Bayesian network [?] has two components: (1) a directed acyclic graph in which nodes represent stochastic variables and directed arcs represent conditional dependencies among variables; (2) a probability distribution for the network variables that decomposes according to conditional dependencies described by the directed acyclic graph [?]. A conditional dependency links a *child* variable to the set of its immediate predecessors in the graph, called its *parent* variables. Each conditional dependency is quantified by the conditional distributions of the child variable given the configurations of the parent variables. This graphical representation allows one to decompose the joint probability distribution of the variables in the network into local parents-child contributions thus yielding a significant saving of the probabilistic information required to specify the domain knowledge.

The induction of a Bayesian network from a database of cases $\mathcal{D}$ consists of the selection of the structure of dependencies among the variables $X_1, ..., X_v$ in $\mathcal{D}$ and the estimation of the probability distributions that quantify these dependencies. The Bayesian approach to solve these two problems regards both the set of possible Bayesian networks and associated conditional probabilities as parameters with prior distributions. Data are used to update the prior distributions in posterior distributions and lead one to choose the Bayesian network with the largest posterior probability. When all Bayesian networks are, a priori, equally likely, the posterior probability of a Bayesian network is proportional to a quantity called *marginal likelihood* and the choice between two Bayesian networks reduces to choosing the one having the largest marginal likelihood. Once a Bayesian network has been chosen, the conditional probabilities that quantify the dependencies in the network are estimated as adjusted relative frequencies of relevant cases [?]. Both the estimation and selection of Bayesian networks from a data set can be made computationally easy by taking advantage of the likelihood factorization induced by the decomposability of each Bayesian network and by adopting a prior distribution for the parameters that obeys the hyper Markov law [?]. In this way, the evaluation of parents-child dependencies can be performed locally, by using search algorithms [?].

Bayesware Discoverer implements several search strategies and in particular the K2 algorithm. The K2 algorithm works by selecting the, *a posteriori*, most probable Bayesian network from a subset of all the possible Bayesian networks. The subset of models is selected by the user who is asked to identify an order with which the variables in the data set are evaluated. The rank of each variable defines the set of variables that will be tested as possible parents: the higher the order of a variable, the larger the number of variables that will be tested as its possible parents. If the user does not specify an order, Bayesware Discoverer uses the order of appearance of the variables in the database to build an initial network that can be further explored to select other dependencies to be tested. The implementation of the K2 algorithm in Bayesware Discoverer starts from the highest ranked variable, say X_1, and computes, first, the marginal likelihood of the model that assumes no links pointing to X_1 from the other variables in the list.

The next step is the computation of the marginal likelihood of the models with one link only pointing to X_1 from the other variables in the list. If none of these dependencies has a marginal likelihood larger than that of the model without links pointing to X_1, the latter is taken as most probable model and the next variable in the list is evaluated.

If at least one of these models has a marginal likelihood larger than that the model without links pointing to X_1, the corresponding link is accepted and the search continues by trying adding two links pointing to X_1 and so on until the marginal likelihood does not increase any longer. Once the evaluation of one variable is terminated, the algorithm removes the variable X_1 from the list by replacing it with the second variable in the original list and repeats the same search. The fact that data are complete, as no entries are reported as unknown in the data set, is a key feature to maintain the induction of Bayesian networks computationally feasible. When data are incomplete under an ignorable missing data mechanism [?], the marginal likelihood of a Bayesian network becomes a mixture of marginal likelihood induced from the possible completions of the data, with the consequent loss of the decomposability properties described above. Bayesware Discoverer implements the approach of [?] to compute a first order approximation of the posterior probability of a Bayesian network.

This approximation is based on a novel estimation method — called *Bound* and *Collapse* [?] — to compute bounds on the set of estimates that are consistent with the data available, and to collapse these interval estimates into points by using the information provided by the user on the missing data mechanism. This first order approximation shares the same factorization of the posterior probability computed from complete data, so that the model search can be performed as if data were fully observed. The approximation works under the assumption that data are missing at random, that is, the probability that a value of a variable is missing is independent of the variables that are not fully observed in the data set [?].

As the current version of Bayesware Discoverer handles discrete variables only, continuous variables are discretized into a number of bins that can be chosen by the user and there are two possible discretization methods that one can choose from. Continuous variables can be discretized either into a number of equal length bins, or into a number of bins having approximately the same frequency of cases.

3. DATA MANIPULATION AND PREPROCESSING

The first step of the analysis was an accurate screening of the database to detect redundant variables — particularly variables that were apparently related to or explicitly derived from other variables in the database — to be removed from the database. The reason to remove such variables was that Bayesware Discoverer searches stochastic dependencies among variables and the presence of variables that are functionally related can mask genuine associations between other variables. For example, the date of birth — variable DOB— was removed as it provides essentially the same information as the age of the person. Similarly, variables in the history and promotion history files were removed in block to keep only independent summary variables derived from the original ones. Variables giving a detailed description of the family composition were removed to leave only a broader description of the family.

The database was then cleaned by removing all cases with entry errors as well as all variables with more than 99% of missing values and variables of which only one state was observed in the training set. (Given the methodology implemented by Bayesware Discoverer these variables would be considered constant anyway.)

All continuous variables were discretized into four bins of equal length. Before this step, variables having a skewed distribution (as the dollar amount of donations) were transformed in a logarithmic scale. Many integer-valued variables — as those indicating the number of known times the donor had responded to other types of mail order offers — were appropriately recorded and states observed with a low frequency were grouped in a unique state. The rationale behind this choice was trying to limit the number of sparse tables. We also decided to remove some nominal variables having a large number of categories that cannot be treated efficiently. A careful cleaning and several transformations were applied to the 285 variables reflecting characteristics of the donors neighborhood, as collected from the 1990 US Census. We noted that the database also reported variables that represented social, economic, demographic, urban and ethnic indicators of the donors neighborhood. Hence, we kept only these indicator variables thus reducing the original database by more than 50% of the variables. The database was then cleaned by removing all cases with entry errors as well as all variables with more than 50% of missing values and variables of which only one state was observed.

Globally, the screening and cleaning operation led to reduce the database of 468 variables into a database of 30 variables that can be divided into three groups. One group comprises variables with personal information about the donors, as age, gender, household income (Income), whether the donor gave the phone number (Hphone_D). The second group comprises variables with information about the donors neighborhood as socio-economic and urbanicity indicators (Domain1, Domain2, Cluster, Wealth1); the composition of the labor force in terms of percentage of employees of the federal, state and local government (Fedgov, Stategov, Localgov) and information about the presence of military veterans and employees of the Military (Vietvets, Malevets, WWiivets and Malemili). Finally, the third group comprises variables extracted from the history and promotion history file of the donors, that give details of the donations made by donors. For example, variables Minramnt, Maxramnt and Ramntall give the minimum, maximum and total dollar amount of donations. The variable Lastgift is the dollar amount of the last donation made by a donor, while Ngiftall is the total number of gifts made. Variables like Timelag and Odatedw give information about the time lag from the last donation and the first donation. The remaining variables provide further information about renewal mails received by donors and the number of donations made.

From this database, we then extracted a second database of about 4,000 cases containing only data of donors who made a donation in reply to the '97 renewal mail. The two data sets were used by Bayesware Discoverer to induce two Bayesian networks modeling the dependence of the probability of responding to the mailing campaign and the dollar amount of the gift. In all cases, we assumed that data were missing at random and we supposed that all Bayesian networks consistent with the order followed by the K2 algorithm were equally likely. We began by choosing an order in which the two target variables were tested as children of all other variables. This initial order let us have a first initial vision of the dependencies among the variables. We then repeated the model search by selecting different orders among the variables, reflecting different interesting dependencies to test.

Given the large number of variables in the data sets and the time constraint, we limited the search to models allowing two parents at most for each variable. Plausible larger dependencies were then individually evaluated by computing their marginal likelihood. The next section describes the two Bayesian networks that were eventually selected form the set of most likely models induced from the different orders. The final choice was based on both the overall marginal likelihood of the different models and the interpretability of the models.

4. RESULTS

This section describes the analytical process of understanding the knowledge extracted by Bayesware Discoverer and how this understanding can be improve the marketing strategy of the foundation.

4.1 Profiling Respondents

The Bayesian network Response-net in Figure 1 shows that the probability of a donation (variable Target-B in the top-left corner) is directly affected by the wealth rating (variable Wealth1) and the urbanicity level of the donor's neighborhood (variable Domain1). The dependence of Target-B on Wealth1 and Domain1 is exp(200) times more likely than the nearest scored model in which the dependence of Target-B is affected by the wealth rating and the variable cluster, that represents an indicator of the socio-economic and urbanicity level of the donor neighborhood.

Marginally, only 5% of those who received the renewal mail are likely to respond. Persons living in suburbs, cities or towns have a probability 5.2-5.3% of responding while donors living in rural or urban neighborhoods respond with probability 4.6-4.7%. The wealth rating of the donor neighborhood has a positive effect on the response rate of donors living in urban, suburban or city areas with donors living in wealthier neighborhoods being more likely to respond than donors living in poorer neighborhoods. The probability of responding raises up to 5.8% for donors living in wealth city neighborhoods.

The variable Domain1 is closely related to the variable Do-

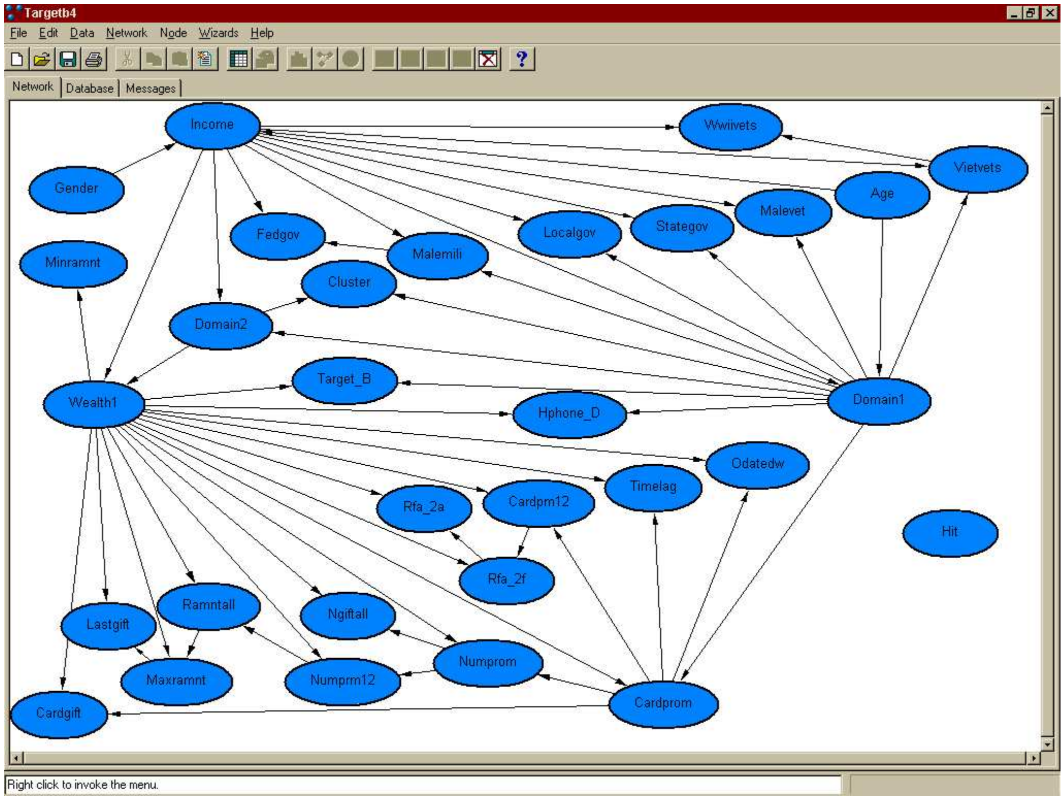

Figure 1: The Bayesian network Response-net induced from the data.

main2 that represents an indicator of the socio-economic status of the donor neighborhood and it shows that donors living in suburbs or city are more likely to live in neighborhoods having a highly rated socio-economic status. Therefore, they may be more sensitive to political and social issues. The model also shows that donors living in neighborhoods with a high presence of males active in the Military (Malemili) are more likely to respond. Again, since the charity collects funds for military veterans, this fact supports the hypothesis that sensitivity to the problem for which funds are collected has a large effect on the probability of response. On the other hand, the wealth rating of donors living in rural neighborhood has the opposite effect: the higher the wealth rating, the smaller the probability that the donor responds, and the least likely to respond (3.8%) are donors living in wealth rural areas. A curiosity is that persons living in rural and poor neighborhood are more likely to respond positively to mail including a gift than donors living in wealthy city neighborhood.

The household income (income) has a positive effect on the probability to respond, that increases with the donors' income. The data on donors' income are quite in agreement with the wealth rating of the donors neighborhood, so that, although the variable income had a large proportion of missing data, the hypothesis that data were missing at random is supported by this finding. The gender has essentially no effect on the likelihood to respond, while age has a negative effect, with older donors being less likely to respond. Older donors have, most likely, made a large number of donations over the years (Ngiftall), and the response rate is negatively related to the number and the total dollar amount of the donations. Given the fact that about 50% of the donors in the database is above 70 years of age, this result suggests that a way to improve the response rate is to target young donors and increase the database with young persons, sensitive to social issues.

By querying the network, we can profile respondents who are more likely to live in a wealth neighborhood, which is located in a suburb and they are less likely to have made a donation in the last 6 months than those who do not respond. One feature that discriminates respondents from non respondent is the household income, and respondents are 1.20 times more likely to be living in wealthy neighborhoods, and to be on higher income than non respondents.

4.2 Profiling Donors

The Bayesian network Donation-net in Figure 2 is the knowledge extracted from of donors to the '97 renewal mail. The network topology is very similar to that of the network Response-net and, again, the variables Wealth1 and Domain1 are those directly influencing the dollar amount of the donation (variable Logtar). This dependence is, at least, exp(50) times more likely than the others that were investigate during the search process. On average, 82% of donors make gifts between 1 and 13 dollars while 18% of donors make a donation between 14 and 100 dollars. Large donations are more likely to be made by people having a high income, usually living in suburbs, town or city in wealthy neighborhoods. Donors living in suburbs are expected to make the largest donations, exceeding 14 dollars. In particular, the probability of donations exceeding 20 dollars from donors living in suburbs range between 0.2 for donors on low household

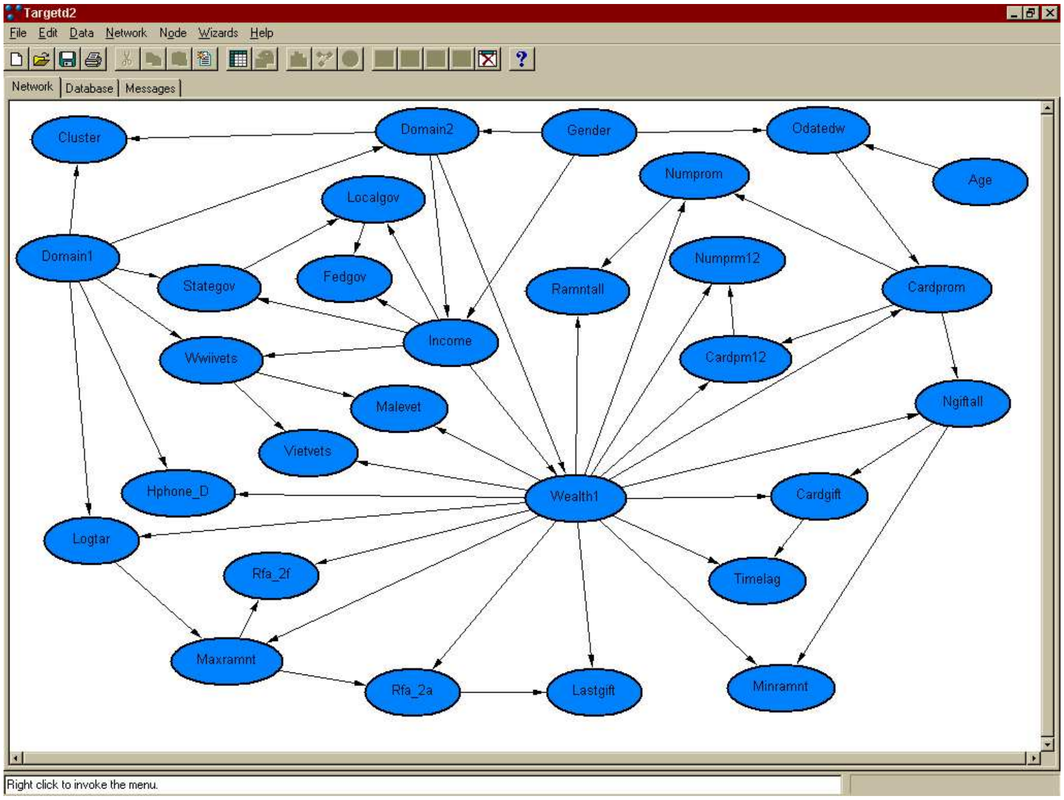

Figure 2: The Bayesian network Donation-net induced from the data available about donors. The variable Logtar — bottom right — is the log-donation.

income and 0.25 for donors on high income.

Donors on low income, living in rural neighborhoods, are expected to make a donation inferior to 10 dollars. However, donors who declare a low household income but leave in wealthy rural neighborhoods are more likely to make donations larger than 10 dollars. This behavior differs from the probability of responding to the renewal mail that is the smallest among donors with the same urbanicity and economic characteristics. Interestingly, the agreement between the declared household income and the wealth rating of the donor's neighborhood is largest for donors living in suburbs or city, while it is smallest for donors living in rural neighborhood. This finding, coupled with the effect of the income on the expected donation would suggest that the neighborhood wealth rating is a better indicator of the donors intention.

The donation amount is also affected by the dollar amount of the last gift prior the renewal mail (variable Lastgift). Donors appear to maintain the gift amount constant and their constancy is directly proportional to the number of times they responded to similar mail offers and their income. This finding is further confirmed, for example, by the dependency found between the dollar amount of the smallest and largest gift to date (Minramnt and Maxramnt), that are in direct proportion and have a large, indirect, effect on the donation distribution. Interestingly, the frequency of donation is inversely related to the donation amount made to the same charity, with frequent donations corresponding to gifts between 1 and 10 dollars and rare donations corresponding to gifts of more than 20 dollars. The donation is also negatively influenced by the number of card promotions received in the previous 12 months. Similarly, long time-lags between donations (Timelag) correspond to donors making large gifts.

From the network, we can profile the donor on the basis of the gift amount. For example, those who donate between 1 and 10 dollars are more likely to be females over 75 years of age, living in a household with low/medium income in either a town or rural neighborhood, and who donated an equivalent gift in the last donation. When the donations become larger, the probability that the donors live in a wealthy suburb neighborhood and has a high household income increases.

4.3 Profit Prediction

The two models Response-net and Donation-net can be used to predict the expected profit incurred in sending a renewal mail to a donor. Given information available about the lapsed donor, the network Response-net can be used to compute the probability that a lapsed donor responds to the renewal mail, say $p(\mathsf{Target}_B = 1)$. The network Donation-net can be used similarly to compute the expected donation by, first, computing the probability distribution of the donation amount, conditional on the information available about the lapsed donor, and then this distribution is used to calculate the expected donation $E(D)$. The expected profit is then computed as

$$P = -0.68 \times (1 - p(\mathsf{Target}_B = 1)) + p(\mathsf{Target}_B = 1) \times E(D)$$

and the decision of whether sending the renewal mail depends on P being positive. For example, a 73 years old lapsed donor living in a high rated socio-economic neighborhood, located in a suburb of medium wealth, who made his first donation 10 months before the renewal mail and made altogether 10 donations, has a probability of answering of 0.053. The expected donation however turns out to be 12 dollars so that the expected profit is 11 dollars thus suggesting that it is worthwhile sending the renewal mail. The lapsed donor features were selected from a test set and indeed the donor answered the renewal mail and made a donation.

5. CONCLUSIONS

This paper has shown an application of Bayesian methods to a Knowledge Discovery task. The goal of the analysis was making a profile of donors to help a Charity understand reasons behind the lack of response to renewal mail sent to donors who had made a donation in the past. The models extracted made a very reasonable profile of donors: essentially persons sensitive to social issues are more likely to make donations although the likelihood of a donation decreases over time. This finding suggests that a strategy to maintain a high response rate to direct fund raising is to continuously update the database of donors. Further information about Bayesware Discoverer is available from `www.bayesware.com`.

Acknowledgments

Bayesware, the Bayesware logo, and Bayesware Discoverer are trademarks of Bayesware Limited. GainSmart is trademark of Urbana Science.