

NON PROFIT FUND RAISING ANALYSIS USING DATA MINING

Nafisseh Heiat, College of Business, Montana State University-Billings, 1500 University Drive, Billings, MT 59101. 406-657-2224, nheiat@msubillings.edu

ABSTRACT

Data mining tools make it possible to apply mathematical models to the historical data to manipulate and discover new information. In this study, we are applying data mining techniques to fund-raising by using decision trees and artificial neural networks to classify donor and non-donors in an academic foundation.

INTRODUCTION

Fund raising has definitely come a long way from the pre-World War I spontaneous giving, and it is no longer a job for Boy/Girl scouts or even the local Church. Fundraising has gone way past professional and organizations and universities are trying to find ways of making their efforts more efficient and lucrative for their organizations. Gone are the days of randomly targeting individuals for donations and being happy with whatever is collected. Fundraising has become an organized project for most organizations which utilize every resource available to them to achieve a target goal. Giving USA 2010, the American Association of Fundraising Counsel reported that of the total charitable giving in the United States more than \$40 billion was given to public and private education, which ranks second only to religious organizations as a recipient of grants and gifts. Non-profit organizations are specially depended on donations and charities.

Colleges, universities, and other nonprofit organizations amass a great deal of information about the people they serve. What many organizations don't realize is that they can use this information to gain valuable insight to improve their advancement efforts. Applying effective business analytics, such as data mining and predictive modelling can yield significant benefits in cost savings and more productive projects.

The goal of this study is to use statistical and data mining methods to show how certain attributes relate to past patterns. Using attributes that strongly correlate with fund raiser's contribution, help us in developing scoring systems that are specific to organizations. Once a group has assigned a score to each person in the database, it can decide on whom to focus appeals and can limit mailings to donors with high probabilities of giving donations. Next section provides a brief literature review of data mining tools in general as well as the ones specifically developed for fund raisers. Section 3 describes the source of data and data preparation for this study. Two different data mining techniques, decision trees and artificial neural networks were used in this study to classify donors and non-donors. These techniques are described in methodology followed by findings and conclusions for the study.

LITERATURE REVIEW

According to Williams, G. (2005), history can be a powerful emotional trigger. It can be a potent marketing tool for universities, non-profit organizations and charities. History evokes memories which can result in a stronger emotional attachment to your organization. People want to feel connected to other people with similar interests, morals, values, goals and backgrounds.

According to Wylie, P. B., (2004) statistics can help organizations save money on fund raising mailings and telephone calls, improve rates of response to annual appeals, and identify people who may become major donors. He suggests developing a rating system to evaluate potential donors, and advises fund

raisers to pull pieces of demographic information from their databases, such as donors' age and marital status that might be linked to support for their group.

Data mining tools vary from query and reporting tools such as QBE or SQL to intelligent agents which utilize artificial intelligence tools such as neural networks, decision trees and fuzzy logic to help form the basis of information discovery and business intelligence. Multi-dimensional analysis tools, allow you to view multi-dimensional information from different perspectives. Statistical tools help apply various mathematical models to the information to manipulate and discover new information. Several management information systems software have been developed which come with data-mining tools and provide a user with capabilities to mine the data to produce useful information.

Helix Data Processing Consultants are one such group which has developed data-mining software specifically for fund raisers. Their system, called ANDAR/360 is a comprehensive software system that runs either on a standalone PC or a client/server network.

Harvard University has also developed a centrally operated fund-raising system that allows university institutions to share fund-raising information for the first time. The new Sybase system, called HOLDEN (Harvard Online Development Network), is expected to maximize the funds generated by the Harvard Development Office from the current donor pool by more precisely targeting existing resources and eliminating wasted efforts and redundancies across the university.

DATA PREPARATION

The **X** Foundation is a non-profit organization founded to enhance the overall quality of the academic foundations at a university. This is accomplished through a combination of solicitation, investment and management of monetary support for university programs.

As primary source of funding for the Foundation is donations made by alumni, it is important to identify which factors will predict the probability of donations to the Foundation. A database of alumni is maintained by the Foundation, tracking such basic information as name, gender, age, marital status, degree, address, years since graduation, BSED (Bachelor of Science in Education) and total amount of donations made to the Foundation.

The data used for this study was obtained from the above foundation. The alumni dataset acts as an address bank, and tracks such additional information as gender, birth date, class, major, and total donations. Additional information regarding current business, position, and salary are also tracked. While these may be predictors of one's ability to contribute to the Foundation, the information collected is sporadic, at best, which may skew the results of the data model. Thus, these factors will not be introduced as independent variables. Not all information was present for every individual in the database and to eliminate bias and ensure validity, data records with missing information were eliminated to obtain a sample database.

The original database contained 10871 records and 15 variables. After dropping all entries with missing data the remaining dataset contained 5993 records.

Before applying data mining to our data set, we used feature selection to determine those variables or features which are more important or closely related to the dependent variable. Feature selection methods not only help us avoid noise in the dataset, but also help us figure out which variable is more important to predict. The results of applying feature selection listed in Figure 1.

Best predictors for dependent variable Donation

Variable	P-Value
Years since Grad.	0.00000
Major	0.00000

Degree	0.00000
BSED	0.00000
Gender	0.00000

Figure 1

METHODOLOGY

Data Mining may be defined as the process of finding potentially useful patterns of information and relationships in data. More and more colleges, universities, and other nonprofit organizations are storing large amounts of data about the people they serve. As the quantity of data has accumulated, domain experts using manual analysis have not kept pace and have lost the ability to become familiar with the data in each case as the number of cases increases. Data visualization techniques can assist in the manual analysis of data, but ultimately the human factor becomes a bottleneck as an organization using a large database can receive hundreds or even thousands of matches to a simple query (Bentley 1997, Brachman, Khabaza, Kloesgen, and Simoudis, 1996 and Fayyad and Piatetsky-Shapiro 1996). Improved data and information handling capabilities have contributed to the rapid development of new opportunities for knowledge discovery. Interdisciplinary research on knowledge discovery in databases has emerged in this decade. Data mining, as automated pattern recognition, is a set of methods applied to knowledge discovery that attempts to uncover patterns that are difficult to detect with traditional statistical methods. Patterns are evaluated for how well they hold on unseen cases. Databases, data warehouses, and data repositories are becoming ubiquitous, but the knowledge and skills required to capitalize on these collections of data are not yet widespread. Innovative discovery-based approaches to data analysis warrant further attention (Hammergren 1996).

There are situations where organizations would like to search for patterns but human abilities are not well suited to search for those patterns. This usually involves the detection of “outliers”, pattern recognition over large data sets, classification, or clustering using statistical modelling. Data mining by discovering these patterns could provide new information (Meltzer 1998).

In this study I used two classification methods. The following is a brief description of these classification methods.

Decision Trees: Decision trees and rule induction are two most commonly used approaches to discovering logical patterns within data sets. Decision trees may be viewed as a simplistic approach to rule discovery because of the process used to discover patterns within data sets. Decision tree is built through a process known as binary recursive partitioning. This is an iterative process of splitting the data into partitions, and then splitting it up further on each of the branches. Initially, you start with a training set in which the classification label (say, "donor" or "non-donor") is known (pre-classified) for each record. All of the records in the training set are together in one big box. The algorithm then systematically tries breaking up the records into two parts, examining one variable at a time and splitting the records on the basis of a dividing line in that variable (say, YSG > 10 or YSG <=10). The object is to attain as homogeneous set of labels (say, "donor" or "non-donor") as possible in each partition. This splitting or partitioning is then applied to each of the new partitions. The process continues until no more useful splits can be found. The heart of the algorithm is the rule that determines the initial split rule (Parsaye 1996).

The process starts with a training set consisting of pre-classified records. Pre-classified means that the target field, or dependent variable, has a known class or label: "donor" or "non-donor. The goal is to

build a tree that distinguishes among the classes. For simplicity, assume that there are only two target classes and that each split is binary partitioning. The splitting criterion easily generalizes to multiple classes, and any multi-way partitioning can be achieved through repeated binary splits. To choose the best splitter at a node, the algorithm considers each input field in turn. In essence, each field is sorted. Then, every possible split is tried and considered, and the best split is the one which produces the largest decrease in diversity of the classification label within each partition. This is repeated for all fields, and the winner is chosen as the best splitter for that node. The process is continued at the next node and, in this manner, a full tree is generated.

Artificial Neural Networks (ANN): Artificial neural networks are defined as information processing systems inspired by the structure or architecture of the brain (Caudill & Butler, 1990). They are constructed from interconnecting processing elements, which are analogous to neurons. Neural networks function relatively similar to the human brain. Both the human brain and neural networks are operated by neurons, the only difference being that the human brain is controlled by biochemical processes whereas the neural networks that are binary digits. The activation process of the neural network can be either linear or logistic, depending on whether it is a numerical estimation problem (linear) or a classification problem (logistic). The neural net architecture may contain one or more middle layers that help model nonlinear relationships by assigning weights to the connections between the middle layer and the input and output nodes.

The two main techniques employed by neural networks are known as supervised learning and unsupervised learning. In unsupervised learning, the neural network requires no initial information regarding the correct classification of the data it is presented with. The neural network employing unsupervised learning is able to analyse a multi-dimensional data set in order to discover the natural clusters and sub-clusters that exist within that data. Neural networks using this technique are able to identify their own classification schemes based upon the structure of the data provided, thus reducing its dimensionality. Unsupervised pattern recognition is therefore sometimes called cluster analysis.

Supervised learning is essentially a two stage process; firstly training the neural network to recognize different classes of data by exposing it to a series of examples, and secondly, testing how well it has learned from these examples by supplying it with a previously unseen set of data. A trained neural network can be thought of as an "expert" in the category of information it has been given to analyse. It provides projections given new situations of interest and answers "what if" questions. There are disadvantages in using ANN. No explanation of the results is given i.e. it is difficult for the user to interpret the results. They are slow to train due to their iterative nature. Empirical studies have shown that if the data provided does not contain useful information within the context of the focus of the investigation, then the use of neural networks cannot generate such information any more than traditional analysis techniques can. However, it may well be the case that the use of neural networks for data mining allows this conclusion to be reached more quickly than might ordinarily be the case.

ANALYSIS

Applying effective business analytics, such as data mining and predictive modelling can yield significant benefits in cost savings and more productive projects. In this study we use statistical and data mining methods that show us how certain attributes relate to past patterns. For the first technique, decision trees, the basic design is a binary decision tree. The C&RT algorithm is a form of decision tree that can be used for either classification or estimation problems. C&RT algorithms continue making splits along the ranges of the predictor variables until some stopping function is satisfied. In our case we have to predict who is willing to donate to the foundation and the split keeps going on until we actually reach our target variable. Years since graduation, BSED, degree, major and gender are the independent variables used to predict the dependent variable, donation in this study.

Predictor importance (Split Input Data into Tr Response: Donation		
	Variable Rank	Importance
Years since Grad.	41	0.410567
BSED	4	0.038896
Degree	50	0.500191
Major	100	1.000000
Gender	4	0.042370

Figure 2

Figure 2 shows the results of applying the decision trees.

Figure 3 shows the same results graphically:

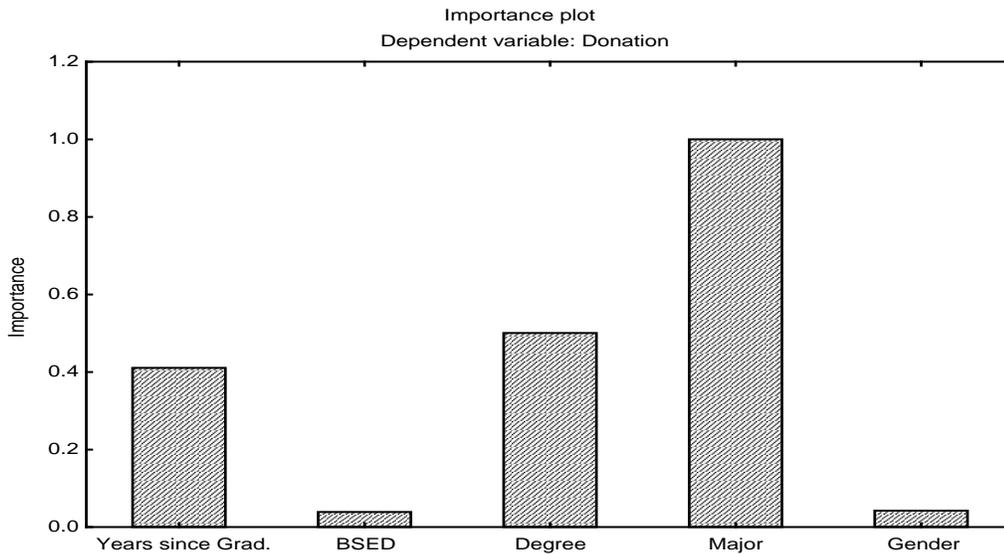


Figure 3

The Classification (Confusion) Matrix in Figure 4 indicates that the classification model is more accurate in classifying non-donors (76.30 %) compared to classifying donors (23.70 %).

In addition, from the decision tree we can induce these general rules: the longer a graduate has been graduated, the more they are willing to donate; major, Degree and BSED are also most important variables to use in our prediction.

In this study supervised multiperceptron ANN was used to classify donors and non-donors. For ANN again we used years since graduation, BSED, degree, major and gender as independent variables and donation as the dependent variable.

The results of Classification with neural network for the Foundation dataset are listed in Figure 5. 2429 cases were predicted accurately as being non-donors, 555 were predicted accurately as donors, 362 cases were predicted to be non-donors and ended up donation, and 997 were predicted to be donors and were not. One more time the prediction for non-donors is better than the prediction for donors weighing in at 87% predicted correctly for non-donors versus 35% predicted correctly for donors.

Classification matrix 1 (Split Input Data into Training and Testing Samples (Classification))				
Dependent variable: Donation				
Options: Categorical response, Analysis sample				
	Observed	Predicted 1	Predicted 0	Row Total
Number	1	866	1053	1919
Column Percentage		68.14%	25.73%	
Row Percentage		45.13%	54.87%	
Total Percentage		16.14%	19.63%	35.78%
Number	0	405	3040	3445
Column Percentage		31.86%	74.27%	
Row Percentage		11.76%	88.24%	
Total Percentage		7.55%	56.67%	64.22%
Count	All Groups	1271	4093	5364
Total Percent		23.70%	76.30%	

Figure 4

Donation (Classification summary)		
Samples: Train		
	Donation-0	Donation-1
Total	2791.000	1552.000
Correct	2429.000	555.000
Incorrect	362.000	997.000
Correct (%)	87.000	35.000
Incorrect (%)	12.000	64.000

Figure 5

CONCLUSIONS

The performance results of decision tree are very close with artificial neural network, the latter yielding the best performance. However, the 35% correct classification of donors is not good enough. One reason for this outcome is that the ratio of donors to non-donors is small in the data set. To improve results, equal number of donors and non-donors were used in the next analysis. This led to much better results in accurately classifying the donors. Once again the performance results of Decision Trees were very close with ANN, with the former yielding better results. Using equal Nos, improves the correct classification of donors from 35% to 90.75%. In future studies, other variables such as salary (or range of salary) and cost of misclassification should be included in the analysis to improve results further.

REFERENCES

References are available upon request.